Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

BAKALÁŘSKÁ PRÁCE

Blanka Jakůbková

Okupační rozdělení

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: RNDr. Zbyněk Pawlas, Ph.D.
Studijní program: Matematika
Studijní obor: Obecná matematika

2007

Mé poděkování patří vedoucímu práce, RNDr. Zbyňkovi Pawlasovi, Ph.D., za několikeré přečtení práce, rady a odpovědi na všetečné otázky.

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

V Praze dne 24. července 2007 Blanka Jakůbková

Obsah

Název práce: Okupační rozdělení
Autor: Blanka Jakůbková
Katedra: Katedra pravděpodobnosti a matematické statistiky
Vedoucí bakalářské práce: RNDr. Zbyněk Pawlas, Ph.D.
e-mail vedoucího: pawlas@karlin.mff.cuni.cz
Abstrakt: Tato práce se zabývá okupačními rozděleními. Jedná se o rozdělení počtu obsazených přihrádek v situacích, které se dají modelovat pomocí umisťování částic do přihrádek. Předložená práce je kompilací již známých výsledků, které jsou uspořádány od nejkonkrétnějšího okupačního problému spojeného s Maxwellovým–Boltzmannovým schématem přes problém komisí k modifikacím a dalším zobecněním problémů okupačního rozdělení. Většina uvedených výsledků je založena na kombinatorické pravděpodobnosti a byla známa již v 18. století. Práce je dále doplněna o řešené ilustrační příklady související s danou tématikou.
Klíčová slova: Okupační rozdělení, Maxwellův–Boltzmannův model, problém komisí.

Title: Occupancy distibutions
Author: Blanka Jakůbková
Department: Department of Probability and Mathematical Statistics
Supervisor: RNDr. Zbyněk Pawlas, Ph.D.
Supervisor's e-mail address: pawlas@karlin.mff.cuni.cz
Abstract: This bachelor thesis concerns the occupancy distribution. It is the distribution of quantity of occupied boxes in situations, which can be modeled as placing balls into boxes. Presented thesis is a compilation of known facts, which are ordered from the most concrete occupancy problem related to the Maxwell–Boltzmann distribution through the committee problem to the modifications and other generalizations of occupancy distribution problems. Most of the presented results are based on combinatorial probability and have been known since 18th century. Several solved illustrative problems related to the subject matter are appended to the thesis.
Keywords: Occupancy distribution, Maxwell–Boltzmann distribution, committee problems.

Úvod

Většina publikací věnovaných základům pravděpodobnosti nebo kombinatoriky obsahuje některou úlohu, která se dá modelovat pomocí tahů z urny, ve které je uloženo n různobarevných míčků. Předpokládá se, že všech n míčků má stejnou pravděpodobnost vytažení v každém tahu. Alternativně lze uvažovat n přihrádek, do kterých jsou náhodně umísťovány částice. Různé urnové modely se liší podle toho, jestli výběr probíhá s vracením nebo bez vracení a především podle toho, jakým způsobem se určí, kolik tahů je potřeba provést. Například pokud provádíme pevný počet r tahů s vracením, dostáváme Maxwellův–Boltzmannův model. Pokud naopak vytahujeme míčky (s vracením) tak dlouho, dokud jsme nevytáhli k různých barev, mluvíme o problému sběratele. Urnovým modelům a jejich aplikacím je věnována kniha [8].

Jednou z nejznámějších úloh spojenou s urnovým schématem je klasický okupační problém: Kolik různých barev míčků bylo vytaženo aspoň jednou v r tazích s vracením? V alternativní formulaci se ptáme, kolik z n přihrádek je obsazeno po umístění r částic. Každá částice je zařazena do dané přihrádky s pravděpodobností $1/n$ a nezávisle na ostatních částicích. Tento problém je uveden již v de Moivreově pojednání *De Mensura Sortis* z roku 1712 (problém 18), viz [5]. De Moivre ho zformuloval v řeči dvou hráčů, kteří hrají hazardní hru s kostkami. Při použití našeho značení se hledá pravděpodobnost, že při r hodech kostkou, která má n stěn, padne celkem m různých čísel alespoň jednou. Řešení, které de Moivre předkládá, je založeno na principu inkluze a exkluze. Pravděpodobnostní rozdělení počtu obsazených přihrádek (počtu vytažených barev míčků, počtu různých hodnot padlých na kostce) se nazývá okupační rozdělení. V tomto případě závisí na dvou parametrech (r a n).

Kromě počtu obsazených přihrádek je zde řada jiných charakteristik a otázek, které nás v souvislosti s okupačním problémem podle druhu aplikace mohou zajímat. Například se můžeme ptát na počet částic v jedné dané přihrádce nebo počet přihrádek obsahující daný počet částic. Seznam různých aplikací okupačního rozdělení lze nalézt v [3] nebo [9], str. 419–420. Z modernějších aplikací v informatice a přenosu dat uveďme [2] nebo [4].

V této práci se věnujeme zobecnění okupačního problému, které spočívá v tom, že v každém z r tahů se nevybírá pouze jeden míček, ale celkem w míčků (bez vracení). O tomto problému mluvíme jako o problému komisí podle interpretace pocházející z článku [10]. Podle [6] stejné zobecnění de Moivreova problému však uvažoval už Laplace v roce 1774: V osudí máme n lístků, které jsou očíslovány od 1 do n. V každém kole vytáhneme bez vracení w lístků, poté je všechny vrátíme zpět do osudí a taháme znovu. Hledá se pravděpodobnost, že po r kolech bylo právě m lístků vytaženo aspoň jedenkrát. Stejně jako v minulé úloze nás může zajímat rozdělení počtu vytažených lístků. Dostáváme tak zobecnění okupačního rozdělení, které má tentokrát tři parametry (r, n a w).

Práce je rozdělena do tří kapitol. V první je rozebrán klasický okupační problém a problém komisí. Druhá kapitola obsahuje řešené úlohy související s problémem komisí. Závěrečná kapitola je pak věnována různým modifikacím a zobecněním okupačních problémů.

Kapitola 1

Okupační problémy

1.1 Několik slov úvodem k problému komisí

Problémem komisí rozumíme následující úlohu: Skupina má n členů, z nichž w je rovnoměrně náhodně vybráno pro vytvoření komise. Jestliže je utvořeno r komisí, nezávisle na sobě, každá o velikosti w, chceme určit pravděpodobnost toho, že právě m členů přísluší do nějaké komise.

Počet způsobů, kterými můžeme vytvořit jednu komisi o velikosti w z celkového počtu n členů, je $\binom{n}{w}$ — jedná se o w-prvkovou kombinaci z n členů bez opakování, r komisí o velikosti w z celkem n členů vytvoříme $\binom{n}{w}^r$ různými způsoby — tohle je r-prvková variace s opakováním.

1.2 Maxwellův–Boltzmannův model

Popišme nejdříve nejjednodušší schéma, kdy položíme velikost komise $w = 1$. Pak mluvíme o Maxwellově–Boltzmannově modelu, kde n členů, ze kterých jsou komise tvořeny, odpovídá n přihrádkám, r komisí, každá mající jen jedno místo, odpovídá r předmětům náhodně vhozeným do přihrádek. Uvažujme tedy r částic, z nichž každá je právě v jedné z n přihrádek, a předpokládejme, že:

– Částice jsou rozlišitelné.

– Pro každou částici umíme určit přihrádku, v níž je částice umístěna.

– Stav systému lze udat tak, že pro každou částici určíme přihrádku, v níž je částice umístěna.

– Všechny stavy jsou stejně pravděpodobné.

Mějme nyní náhodné veličiny $K_1, \ldots, K_n$, které označují počty částic v přihrádkách $1, \ldots, n$. Nalezněme rozdělení náhodné veličiny K_i: ze symetrie je zřejmé, že toto rozdělení je pro všechny přihrádky stejné, vezměme proto v úvahu tu první. Do první přihrádky umístíme k částic, ostatních $r - k$ částic libovolně do zbývajících $n - 1$ přihrádek – to lze provést $\binom{r}{k}(n-1)^{r-k}$ způsoby.

Pravděpodobnost, že $K_1 = k$ je rovna:

$$\begin{aligned} P[K_1 = k] &= \binom{r}{k}\frac{(n-1)^{r-k}}{n^r} \\ &= \binom{r}{k}\left(\frac{1}{n}\right)^k\left(1-\frac{1}{n}\right)^{r-k}, \end{aligned} \tag{1.1}$$

což znamená, že K_1 má binomické rozdělení s parametry r a $1/n$.

Stanovme nyní společnou střední hodnotu náhodných veličin $K_1, \ldots, K_n$:

$$\begin{aligned} EK_1 &= \sum_{k=0}^{r} k\binom{r}{k}\left(\frac{1}{n}\right)^k\left(1-\frac{1}{n}\right)^{r-k} \\ &= \sum_{k=0}^{r} k\frac{r!}{(r-k)!k!}\left(\frac{1}{n}\right)^k\left(1-\frac{1}{n}\right)^{r-k} \\ &= \frac{r}{n}\sum_{k=1}^{r}\frac{(r-1)!}{(r-k)!(k-1)!}\left(\frac{1}{n}\right)^{k-1}\left(1-\frac{1}{n}\right)^{r-k} \\ &= \frac{r}{n}\sum_{k=1}^{r}\binom{r-1}{k-1}\left(\frac{1}{n}\right)^{k-1}\left(1-\frac{1}{n}\right)^{r-k} \\ &= \frac{r}{n}\sum_{j=0}^{r-1}\binom{r-1}{j}\left(\frac{1}{n}\right)^{j}\left(1-\frac{1}{n}\right)^{r-1-j} \\ &= \frac{r}{n}. \end{aligned} \tag{1.2}$$

Střední hodnotu můžeme interpretovat jako průměrný počet předmětů na jednu přihrádku.

Vyšetřeme sdružené rozdělení pravděpodobností náhodných veličin $K_1, \ldots, K_n$. Mějme nezáporná celá čísla $r_1, \ldots, r_n$, taková, že $r_1 + \cdots + r_n = r$. Do první přihrádky vybíráme r_1 částic z r, do druhé přihrádky vybíráme r_2 částic ze zbývajících $r - r_1$ atd. Proto dostáváme:

$$\begin{aligned} &P[K_1 = r_1, \ldots, K_n = r_n] \\ &= \frac{1}{n^r}\cdot\binom{r}{r_1}\binom{r-r_1}{r_2}\binom{r-r_1-r_2}{r_3}\cdots\binom{r-r_1-\ldots-r_{n-1}}{r_n} \\ &= \frac{1}{n^r}\cdot\binom{r}{r_1}\binom{r-r_1}{r_2}\binom{r-r_1-r_2}{r_3}\cdots\binom{r_n}{r_n} \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{n^r}\cdot\frac{r!}{(r-r_1)!r_1!}\cdot\frac{(r-r_1)!}{(r-r_1-r_2)!r_2!}\cdots\frac{r_n!}{r_n!}\\
&= \frac{1}{n^r}\cdot\frac{r!}{r_1!r_2!\dots r_n!},
\end{aligned}
$$

neboli $(K_1,\dots,K_n)$ má multinomické rozdělení s parametry r a $(1/n,\dots,1/n)$.

Nyní určeme pravděpodobnost q toho, že existuje prázdná přihrádka. Označme B_i, $i=1,\dots,n$, takový jev, že i-tá přihrádka je prázdná. Pak

$$
q = P\left(\bigcup_{i=1}^{n} B_i\right).
$$

Ovšem náhodné jevy $B_1,\dots,B_n$ nejsou neslučitelné, z toho důvodu platí:

$$
\begin{aligned}
P(B_i) &= \frac{(n-1)^r}{n^r}, i=1,\dots,n,\\
P(B_i\cap B_j) &= \frac{(n-2)^r}{n^r}, i\neq j, i,j=1,\dots,n,\\
&\dots\\
P(B_1\cap B_2\cap\dots\cap B_{n-1}) &= \frac{1}{n^r},\\
P(B_1\cap B_2\cap\dots\cap B_n) &= 0.
\end{aligned}
$$

Použijeme-li na předchozí princip inkluze a exkluze, dostáváme:

$$
\begin{aligned}
q &= \sum_{i=1}^{n} P(B_i) - \sum_{i<j}\sum P(B_i\cap B_j) + \\
&\quad + \sum\sum_{i<j<k}\sum P(B_i\cap B_j\cap B_k) - \cdots\\
&= \binom{n}{1}\left(1-\frac{1}{n}\right)^r - \binom{n}{2}\left(1-\frac{2}{n}\right)^r + \binom{n}{3}\left(1-\frac{3}{n}\right)^r - \cdots\\
&\quad \cdots(-1)^n\binom{n}{n-1}\left(1-\frac{n-1}{n}\right)^r\\
&= \sum_{i=1}^{n-1}(-1)^{i+1}\binom{n}{i}\left(1-\frac{i}{n}\right)^r.
\end{aligned}
$$

Označme $p_n(r,n)$ pravděpodobnost, že všech n přihrádek je obsazeno:

$$
\begin{aligned}
p_n(r,n) &= 1-q = 1-\sum_{i=1}^{n-1}(-1)^{i+1}\binom{n}{i}\left(1-\frac{i}{n}\right)^r\\
&= \sum_{i=0}^{n}(-1)^{i}\binom{n}{i}\left(1-\frac{i}{n}\right)^r. \qquad (1.3)
\end{aligned}
$$

Spočítejme pravděpodobnost $p_m(r,n)$ toho, že právě m přihrádek je obsazených, neboli právě $n-m$ přihrádek je prázdných, $m = 0, \ldots, n$. Těchto m přihrádek můžeme vybrat $\binom{n}{m}$ způsoby, r předmětů umístíme do m přihrádek tak, že každá z nich přihrádek je obsazena. Počet takových rozmístění je:

$$\begin{aligned} m^r \cdot p_m(r,m) &= m^r \sum_{k=0}^{m} (-1)^k \binom{m}{k} \left(1 - \frac{k}{m}\right)^r \\ &= \sum_{k=0}^{m} (-1)^k \binom{m}{k} (m-k)^r. \end{aligned}$$

Vydělením n^r najdeme pravděpodobnost $p_m(r,n)$, že právě m přihrádek je obsazených:

$$\begin{aligned} p_m(r,n) &= \frac{1}{n^r} \binom{n}{m} m^r p_m(r,m) \\ &= \frac{1}{n^r} \binom{n}{m} \sum_{k=0}^{m} (-1)^k \binom{m}{k} (m-k)^r. \end{aligned} \tag{1.4}$$

Rozdělení s pravděpodobnostmi $p_m(r,n)$ danými vzorcem (1.4), $m = 0, \ldots, n$, se nazývá okupační rozdělení. Má-li náhodná veličina X takovéto rozdělení, pak (podle [7], str. 252) její střední hodnota je:

$$\mathrm{E}X = n \left[1 - \left(1 - \frac{1}{n}\right)^r\right]$$

a rozptyl se rovná:

$$\mathrm{var}X = n \left(1 - \frac{1}{n}\right)^r + n\,(n-1) \left(1 - \frac{2}{n}\right)^r - n^2 \left(1 - \frac{1}{n}\right)^{2r}.$$

Uveďme ještě další odvození vzorce (1.4) pomocí vztahu pro pravděpodobnost složených jevů – odvození pochází od de Moivrea a je uvedeno v [9], str. 406. Uvažujme n jevů $E_1, \ldots, E_n$ a předpokládejme, že pravděpodobnosti současného výskytu libovolného počtu z nich jsou známy. Dále předpokládejme, že platí:

$$P[E_{i_1} \cap E_{i_2} \cap \ldots \cap E_{i_j}] = P[E_1 \cap E_2 \cap \ldots \cap E_j] \tag{1.5}$$

pro každou podmnožinu $\{i_1, \ldots, i_j\}$ množiny $\{1, \ldots, n\}$ a každé $j \in \{1, \ldots, n\}$.

Označme

$$S_j = \sum P[E_{i_1} \cap E_{i_2} \cap \ldots \cap E_{i_j}],$$

kde sčítáme přes všechny množiny indexů $\{i_1, i_2, \ldots, i_j\}$ takové, že $1 \leq i_1 < i_2 < \ldots < i_j \leq n$, $j = 1, 2, \ldots, n$; položme $S_0 = 1$. Tedy S_j je součet pravděpodobností (přes všechny j-tice), že nastane aspoň j daných událostí a použitím (1.5) pro S_j platí:

$$S_j = \binom{n}{j} P[E_1 \cap E_2 \cap \ldots \cap E_j].$$

Pak pravděpodobnost, že nastane aspoň m událostí z n je pomocí principu inkluze a exkluze:

$$\begin{aligned} P_m &= S_m - \binom{m}{1} S_{m+1} + \binom{m+1}{2} S_{m+2} - \cdots + (-1)^{n-m} \binom{n-1}{n-m} S_n \\ &= \sum_{i=0}^{n-m} (-1)^i \binom{m+i-1}{i} S_{m+i} \end{aligned}$$

a $P_0 = 1$.

Pravděpodobnost, že nastane právě m událostí z n je:

$$\begin{aligned} P_{[m]} &= P_m - P_{m+1} \\ &= S_m - \binom{m+1}{1} S_{m+1} + \binom{m+2}{2} S_{m+2} - \cdots + (-1)^{n-m} \binom{n}{n-m} S_n \\ &= \sum_{i=0}^{n-m} (-1)^i \binom{m+i}{i} S_{m+i}. \end{aligned} \tag{1.6}$$

Nechť nyní v konkrétním případě E_i označuje jev, že i-tá přihrádka je prázdná. Platí

$$P[E_1 \cap E_2 \cap \ldots \cap E_j] = \left(\frac{n-j}{n} \right)^r,$$

protože r předmětů umisťujeme do $n-j$ přihrádek. Pak lze S_j psát jako:

$$S_j = \binom{n}{j} \left(\frac{n-j}{n} \right)^r. \tag{1.7}$$

Dosadíme do (1.6) vzorec (1.7) za S_j a $n-m$ za m – hledáme pravděpodobnost, že právě $n-m$ přihrádek je prázdných, neboli právě m přihrádek je obsazených:

$$\begin{aligned} P_{[n-m]} &= \sum_{i=0}^{m} (-1)^i \binom{n-m+i}{i} S_{n-m+i} \\ &= \sum_{i=0}^{m} (-1)^i \binom{n-m+i}{i} \binom{n}{n-m+i} \left(\frac{m-i}{n} \right)^r \\ &= \sum_{i=0}^{m} (-1)^i \frac{(n-m+i)!}{(n-m)!i!} \cdot \frac{n!}{(m-i)!(n-m+i)!} \cdot \left(\frac{m-i}{n} \right)^r \\ &= \binom{n}{m} \sum_{i=0}^{m} (-1)^i \binom{m}{i} \left(\frac{m-i}{n} \right)^r. \end{aligned}$$

Výsledek je identický s (1.4).

1.3 Problém komisí

Dále uvažujme $w > 1$. Označme jako y_n počet způsobů vytvoření r komisí o w členech, $y_n = \binom{n}{w}^r$. Dále označme $\Delta y_k = y_{k+1} - y_k$, což můžeme interpretovat jako přírůstek počtu způsobů vybrání r komisí o velikosti w, jestliže do množiny k prvků, ze kterých komise tvoříme, přidáme jeden další prvek navíc. Je zřejmé, že tento další prvek musí být členem aspoň jedné z r komisí. Z tohoto důvodu Δy_k je počet způsobů vybrání komisí z $k+1$ členů, jestliže jeden určitý člen musí patřit do jedné nebo více komisí. Podobně $\Delta^2 y_k = \Delta(\Delta y_k) = \Delta y_{k+1} - \Delta y_k = y_{k+2} - 2y_{k+1} + y_k$ udává počet způsobů vybrání r komisí z $k+2$ členů, když dva určití členové jsou zahrnuti do jedné nebo více komisí. Počet způsobů vybrání komisí z $k+m$ členů, pokud každý z m určitých členů musí být členem aspoň jedné komise, je $\Delta^m y_k$. Obecně pro $\Delta^m y_k$ platí:

$$\begin{aligned}\Delta^m y_k &= \sum_{i=0}^{m} (-1)^i \binom{m}{i} y_{k+m-i} \\ &= \sum_{i=0}^{m} (-1)^{m-i} \binom{m}{i} y_{k+i} \\ &= \sum_{i=w}^{m} (-1)^{m-i} \binom{m}{i} \binom{k+i}{w}^r, \end{aligned} \tag{1.8}$$

protože platí, že $y_k = 0$, pokud $k < w$.

Uveďme, že $\Delta^m y_0$ je počet způsobů vybrání r komisí z m členů tak, že každý z nich je vybrán pro vytvoření jedné nebo více komisí.

Protože m prvků vybereme $\binom{n}{m}$ způsoby, tak počet způsobů vybrání komisí, aby právě m členů bylo použito, je $\binom{n}{m}\Delta^m y_0$. Pravděpodobnost p_m, že právě m jednotlivců bude členy komise, je (podle [8], str. 164):

$$\begin{aligned} p_m &= \frac{\binom{n}{m}\Delta^m y_0}{\binom{n}{w}^r} \\ &= \frac{\binom{n}{m}}{\binom{n}{w}^r} \sum_{i=w}^{m} (-1)^{m-i} \binom{m}{i} \binom{i}{w}^r \end{aligned}$$

$$= \frac{\binom{n}{m}}{\binom{n}{w}^r} \sum_{i=0}^{m-w} (-1)^i \binom{m}{i} \binom{m-i}{w}^r. \tag{1.9}$$

Pro $w = 1$ nám vyjde:

$$p_m = \frac{1}{n^r} \binom{n}{m} \sum_{i=0}^{m} (-1)^i \binom{m}{i} (m-i)^r,$$

a to odpovídá vzorci (1.4).

Odvození vzorce (1.9) je založeno na stejné myšlence jako původní řešení pocházející od de Moivrea. Stejně jako u Maxwellova–Boltzmannova modelu se v podstatě využívá princip inkluze a exkluze. Jiné možné odvození lze založit na vzorci (1.6).

Nechť $K_1, \ldots, K_n$ jsou náhodné veličiny, které označují, kolikrát byl který prvek vybrán pro vytvoření nějaké komise. Určeme rozdělení těchto náhodných veličin, které je pro všechny z nich stejné, vezměme proto v úvahu K_1. Pravděpodobnost, že vybereme jeden určitý prvek, je

$$p = \frac{\binom{n-1}{w-1}}{\binom{n}{w}} = \frac{(n-1)!}{(n-w)!(w-1)!} \cdot \frac{(n-w)!w!}{n!} = \frac{w}{n}.$$

Pravděpodobnost, že tento prvek vybereme právě k-krát, je

$$P[K_1 = k] \;=\; \binom{r}{k} \left(\frac{w}{n}\right)^k \left(1 - \frac{w}{n}\right)^{r-k}.$$

Dosadíme-li $w = 1$, dostáváme:

$$P[K_1 = k] = \binom{r}{k} \frac{(n-1)^{r-k}}{n^r},$$

což souhlasí se vzorcem (1.1).

Protože K_1 má binomické rozdělení (s parametry r a w/n), je střední hodnota rovna:

$$\mathrm{E}K_1 = \frac{rw}{n}.$$

Jestliže za w dosadíme 1, dostaneme

$$\mathrm{E}K_1 = \frac{r}{n},$$

což opět souhlasí se vzorcem (1.2) z Maxwellova–Boltzmannova modelu.

Na rozdíl od Maxwellova–Boltzmannova modelu je sdružené rozdělení náhodných veličin $K_1, \dots, K_n$ velmi komplikované.

Podobně jako v minulé podkapitole můžeme vyjádřit pravděpodobnost q, že existuje jednotlivec, který nebyl vybrán do žádné komise:

$$\begin{aligned} q &= 1 - p_n \\ &= 1 - \frac{1}{\binom{n}{w}^r} \sum_{i=0}^{n-w} (-1)^i \binom{n}{i} \binom{n-i}{w}^r \\ &= \frac{1}{\binom{n}{w}^r} \sum_{i=1}^{n-w} (-1)^{i+1} \binom{n}{i} \binom{n-i}{w}^r . \end{aligned}$$

Problém komisí může být dále zobecněn, pokud bereme v úvahu komise s proměnnou velikostí $w_i, i = 1, \dots, r$, viz [8], str. 165. V tomto případě máme

$$y_k = \prod_{i=1}^{r} \binom{k}{w_i}$$

a použitím (1.8) dostaneme

$$\Delta^m y_k = \sum_{j=\max w_i}^{m} (-1)^{m-j} \binom{m}{j} \prod_{i=1}^{r} \binom{k+j}{w_i}.$$

Pak pravděpodobnost p_m, že právě m prvků je členy nějaké komise, je pomocí (1.9) rovna:

$$p_m = \binom{n}{m} \sum_{j=\max w_i}^{m} (-1)^{m-j} \binom{m}{j} \prod_{i=1}^{r} \frac{\binom{j}{w_i}}{\binom{n}{w_i}}.$$

Problém komisí byl zformulován Mantelem a Pasternackem v [10]. Ti zobecnili úlohu známou jako chromozomový problém, kterou zformuloval Feller ve své knize [3], str. 102, cv. 16: Buňka obsahuje n chromozomů, mezi jakýmikoli dvěma z nich může nastat vzájemná výměna částí. Jestliže se přihodí r vzájemných výměn, chceme najít pravděpodobnost, že právě m chromozomů tuto výměnu vykoná.

V chromozomovém problému se vlastně jedná o speciální případ problému komisí, pokud položíme velikost komise $w = 2$ – je vybírána dvojice chromozomů, mezi kterými dochází k výměně. Ukazuje se, že v současné době tento problém není z biologického hlediska realistický.

Kapitola 2

Příklady

Příklad 1: Házíme dvanáctkrát šestistěnnou hrací kostkou. Jaká je pravděpodobnost, že v těchto hodech padnou všechna čísla od 1 do 6? Tento příklad je převzatý z [6], str. 423, cv. 18.

Řešení: Situaci si modelujme následovně: čísla od 1 do 6 představují přihrádky (tedy $n = 6$) a 12 hodů kostkou představuje předměty, které do přihrádek umisťujeme (tedy $r = 12$). Chceme určit pravděpodobnost, že všechny přihrádky jsou zaplněny – použijeme vzorec (1.3):

$$
\begin{aligned}
p_6(12,6) &= \sum_{i=0}^{6} (-1)^i \binom{6}{i} \left(1 - \frac{i}{6}\right)^{12} \\
&= \binom{6}{0}\left(1-\frac{0}{6}\right)^{12} - \binom{6}{1}\left(1-\frac{1}{6}\right)^{12} + \binom{6}{2}\left(1-\frac{2}{6}\right)^{12} - \binom{6}{3}\left(1-\frac{3}{6}\right)^{12} \\
&\quad + \binom{6}{4}\left(1-\frac{4}{6}\right)^{12} - \binom{6}{5}\left(1-\frac{5}{6}\right)^{12} + \binom{6}{6}\left(1-\frac{6}{6}\right)^{12} \\
&= 1 - 6\cdot\left(\frac{5}{6}\right)^{12} + 15\cdot\left(\frac{2}{3}\right)^{12} - 20\cdot\left(\frac{1}{2}\right)^{12} + 15\cdot\left(\frac{1}{3}\right)^{12} - 6\cdot\left(\frac{1}{6}\right)^{12} \\
&\doteq 0{,}437\,816.
\end{aligned}
$$

$\square$

Příklad 2: Předpokládejme, že rok má 365 dní. Mějme skupinu r lidí, každý z nich slaví své narozeniny rovnoměrně náhodně a nezávisle v průběhu celého roku. Jaká je pravděpodobnost, že počet dní, kdy někdo slaví narozeniny, je právě x? Příklad najdeme např. v [9], str. 415.

Řešení: Nechť X je taková náhodná veličina, která značí počet dní, kdy někdo slaví své narozeniny. Pak pravděpodobnost, že takových dní je právě x, tj. $P[X = x]$, je stejná jako pravděpodobnost jevu, že právě x přihrádek je obsazených v Maxwellově–Boltzmannově modelu, celkový počet přihrádek je $n = 365$

a počet předmětů je r (viz (1.4)):

$$\begin{aligned} P[X=x] &= p_x(r,365) \\ &= \binom{365}{x}\sum_{k=0}^{x}(-1)^k\binom{x}{k}\left(\frac{x-k}{365}\right)^r. \end{aligned}$$

□

Příklad 3: Disciplinární dvůr je tvořen šesti členy, tři z nich jsou náhodně vybráni pro nějaké zasedání. Jaká je šance, že každý z nich bude vybrán aspoň jednou, vybíráme-li členy pro pět zasedání? Příklad je řešen v [8], str. 166, př. 3.10.

Řešení: Vyřešme příklad jako problém komisí, kde $n = 6$, $w = 3$, $r = 5$, $m = 6$. Použijeme vzorec (1.9). Požadovaná pravděpodobnost p je:

$$p = \frac{\binom{n}{m}\Delta^m y_0}{\binom{n}{w}^r} = \frac{\binom{6}{6}\Delta^6 y_0}{\binom{6}{3}^5} = \frac{\Delta^6 y_0}{\binom{6}{3}^5},$$

kde podle (1.8)

$$\begin{aligned} \Delta^6 y_0 &= \sum_{i=3}^{6}(-1)^{6-i}\binom{6}{i}\binom{i}{3}^5 \\ &= -\binom{6}{3}\binom{3}{3}^5 + \binom{6}{4}\binom{4}{3}^5 - \binom{6}{5}\binom{5}{3}^5 + \binom{6}{6}\binom{6}{3}^5 \\ &= -20 + 15\cdot 4^5 - 6\cdot 10^5 + 20^5 \\ &= 2\ 615\ 340. \end{aligned}$$

Tudíž dostaneme požadovanou pravděpodobnost:

$$p \doteq \frac{2\ 615\ 340}{3\ 200\ 000} = 0{,}817.$$

□

Příklad 4: Za situace popsané v předchozím příkladě předpokládejme, že po prvním zasedání přibude do disciplinárního dvoru, z něhož vybíráme členy, další osoba. Jaká je pravděpodobnost p, že každý z těchto sedmi členů bude vybrán pro aspoň jedno z těchto pěti zasedání? Příklad je uveden v [8], str. 169 jako cv. 3.40.

Řešení: Počet trojic, které vybíráme pro první zasedání, je $\binom{6}{3}$. Pro druhé, třetí, čtvrté a páté zasedání vybíráme trojice $\binom{7}{3}^4$ způsoby, přičemž chceme, aby ti, co nebyli vybráni pro první zasedání, byli na druhém, třetím, čtvrtém nebo pátém aspoň jedenkrát. Spočítáme proto $\Delta^4 y_3$ (podle (1.8)):

$$\begin{aligned}
\Delta^4 y_3 &= \sum_{i=0}^{4}(-1)^i \binom{4}{i} y_{3+4-i} \\
&= \binom{4}{0} y_7 - \binom{4}{1} y_6 + \binom{4}{2} y_5 - \binom{4}{3} y_4 + \binom{4}{4} y_3 \\
&= \binom{7}{3}^4 - 4 \cdot \binom{6}{3}^4 + 6 \cdot \binom{5}{3}^4 - 4 \cdot \binom{4}{3}^4 + \binom{3}{3}^4 \\
&= 35^4 - 4 \cdot 20^4 + 6 \cdot 10^4 - 4 \cdot 4^4 + 1 \\
&= 919\,602.
\end{aligned}$$

Hledaná pravděpodobnost p je (podle (1.9)):

$$p = \frac{1}{\binom{6}{3}} \cdot \frac{\binom{6}{3} \Delta^4 y_3}{\binom{7}{3}^4} = \frac{919\,602}{1\,500\,625} \doteq 0{,}613.$$

□

Následující odvození uvedené v [8], str. 167, použijeme pro výpočet dalšího příkladu: Nechť je komise reprezentována maticí Ω, která je typu $r \times n$ a prvek $\omega_{i,j}$ je roven buď 1 nebo 0 podle toho, zda j-tý prvek, $j = 1, \ldots, n$, byl či nebyl vybrán pro i-tou komisi, $i = 1, \ldots, r$. Například uvažme případ s proměnnou velikostí komisí pro $n = 7, r = 3, w_1 = 2, w_2 = 3, w_3 = 4$:

$$\Omega = \begin{pmatrix} 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 \end{pmatrix}.$$

Označme symbolem n_j součet výrazů (prvků, členů) v j-tém sloupci a symbolem F_i počet takových sloupců, pro které je $n_j = i$, tj. F_i je počet členů, kteří patří do právě i komisí. V našem případě platí, že $F_0 = 2, F_1 = 2, F_2 = 2, F_3 = 1$.

Nechť

$$\mathcal{P} = \{(i_1, \ldots, i_k) | i_1 < \ldots < i_k : \{i_1, \ldots, i_k\} \subset \{1, \ldots, r\}\}.$$

Pak každé $p \in \mathcal{P}$ reprezentuje možné schéma, podle kterého daný člen přísluší do jednotlivých komisí. Například, pokud $r = 3$, pak $p = (1, 2)$ reprezentuje členství v komisi 1 a 2, ale ne 3, a v matici Ω odpovídá sloupcovému vektoru $(1, 1, 0)^T$. Nyní označme N_p jako počet členů se schématem p a N_0 jako počet členů, které nepřipadají na žádnou komisi. Je zřejmé, že každý prvek může mít pouze jedno schéma, tudíž $\sum\limits_{p\in\mathcal{P}} N_p + N_0 = n$. V našem případě pro $r = 3$ máme $F_0 = N_0, F_1 = N_1 + N_2 + N_3, F_2 = N_{12} + N_{23} + N_{13}, F_3 = N_{123}$. Protože celkový počet způsobů vybrání r komisí z n prvků o velikostech $w_1, w_2, \dots, w_r$ je $\prod\limits_{i=1}^{r} \binom{n}{w_i}$, obecně dostaneme (podle [8], str. 167):

$$P(N_p = j_p\ \forall p \in \mathcal{P};\ N_0 = j_0) = \left[\prod_{i=1}^{r} \binom{n}{w_i}\right]^{-1} \cdot \frac{n!}{j_0! \prod\limits_{p\in\mathcal{P}} j_p!}. \tag{2.1}$$

Příklad 5: Zadání příkladu je uvedeno v [1]: V jistém městě se snaží zjistit souvislost mezi znečistěním a zvýšeným výskytem astmatických záchvatů v určitých dnech. Tři astmatičtí pacienti jsou pozorováni po dobu 120 dnů. Předpokládá se, že při normálních „bezpečných“ venkovních podmínkách jsou jejich záchvaty náhodné. Za tohoto předpokladu se považuje za výjimečnou událost, když dva nebo více pacientů má záchvat ve stejný den. Problém je, jak přesně určit, jak moc výjimečná událost to je, aby sledování těchto dnů mohlo vést k informaci podporující hypotézu o souvislosti znečištění a výskytů astmatických záchvatů. Řekněme, že tito tři pacienti měli během sledovaného období 9, 10 a 22 záchvatů. Jaká je pravděpodobnost, že za „bezpečných“ podmínek má dva nebo více pacientů záchvat ve stejný den?

Řešení: Našemu značení odpovídá: $n = 120, r = 3, w_1 = 9, w_2 = 10, w_3 = 22$. Matice Ω má 3 řádky a 120 sloupců.

Nejdříve určeme pravděpodobnost, že se v matici Ω vyskytuje aspoň jednou aspoň jeden ze sloupcových vektorů $(1,1,0)^T$, $(1,0,1)^T$, $(0,1,1)^T$, $(1,1,1)^T$, tzn. chceme určit pravděpodobnost P_1, že $F_2 + F_3 = N_{12} + N_{23} + N_{13} + N_{123} \geq 1$. Přejděme k doplňkovému jevu a spočítejme pravděpodobnost Q, že $N_{12} + N_{23} + N_{13} + N_{123} = 0$ – to znamená, že se v matici Ω vyskytují jen vektory $(1,0,0)^T, (0,1,0)^T, (0,0,1)^T, (0,0,0)^T$, a to po řadě v počtu w_1, w_2, w_3 a $n - w_1 - w_2 - w_3$. Dosadíme-li do vzorce (2.1), dostáváme:

$$\begin{aligned} Q &= \left[\binom{n}{w_1}\binom{n}{w_2}\binom{n}{w_3}\right]^{-1} \cdot \frac{n!}{(n - w_1 - w_2 - w_3)!w_1!w_2!w_3!} \\ &= \left[\binom{120}{9}\binom{120}{10}\binom{120}{22}\right]^{-1} \cdot \frac{120!}{(120 - 9 - 10 - 22)!\,9!\,10!\,22!} \\ &\doteq 0{,}006\,593. \end{aligned}$$

Pak $P_1 = 1 - Q \doteq 0{,}993\,407$.

Spočítejme ještě pravděpodobnost P_2, že $N_{12} + N_{23} + N_{13} + N_{123} \geq 2$, tedy pravděpodobnost toho, že se záchvat aspoň u dvou pacientů vyskytne aspoň dvakrát během 120 dní. K tomu potřebujeme určit pravděpodobnost $\hat{P}_1$, že se záchvat aspoň u dvou pacientů objeví právě jednou. Platí totiž, že $P_2 = P_1 - \hat{P}_1$.

$$
\begin{aligned}
\hat{P}_1 &= P(N_{12} + N_{23} + N_{13} + N_{123} = 1) \\
&= P(N_{12} = 1, N_{13} = 0, N_{23} = 0, N_{123} = 0, \\
&\quad N_0 = n - w_1 - w_2 - w_3 + 1, N_1 = w_1 - 1, N_2 = w_2 - 1, N_3 = w_3) \\
&+ P(N_{12} = 0, N_{13} = 1, N_{23} = 0, N_{123} = 0, \\
&\quad N_0 = n - w_1 - w_2 - w_3 + 1, N_1 = w_1 - 1, N_2 = w_2, N_3 = w_3 - 1) \\
&+ P(N_{12} = 0, N_{13} = 0, N_{23} = 1, N_{123} = 0, \\
&\quad N_0 = n - w_1 - w_2 - w_3 + 1, N_1 = w_1, N_2 = w_2 - 1, N_3 = w_3 - 1) \\
&+ P(N_{12} = 0, N_{13} = 0, N_{23} = 0, N_{123} = 1, \\
&\quad N_0 = n - w_1 - w_2 - w_3 + 2, N_1 = w_1 - 1, N_2 = w_2 - 1, N_3 = w_3 - 1).
\end{aligned}
$$

Dosadíme-li hodnoty a použijeme-li opět vzorec (2.1), dostaneme:

$$
\begin{aligned}
\hat{P}_1 &= P(N_{12} = 1, N_{13} = 0, N_{23} = 0, N_{123} = 0, N_0 = 80, N_1 = 8, N_2 = 9, N_3 = 22) \\
&+ P(N_{12} = 0, N_{13} = 1, N_{23} = 0, N_{123} = 0, N_0 = 80, N_1 = 8, N_2 = 10, N_3 = 21) \\
&+ P(N_{12} = 0, N_{13} = 0, N_{23} = 1, N_{123} = 0, N_0 = 80, N_1 = 9, N_2 = 9, N_3 = 21) \\
&+ P(N_{12} = 0, N_{13} = 0, N_{23} = 0, N_{123} = 1, N_0 = 81, N_1 = 8, N_2 = 9, N_3 = 21) \\
&= \left[\binom{120}{9}\binom{120}{10}\binom{120}{22}\right]^{-1} \cdot 120! \cdot \\
&\quad \cdot \left(\frac{1}{80!\,8!\,9!\,22!} + \frac{1}{80!\,8!\,10!\,21!} + \frac{1}{80!\,9!\,9!\,21!} + \frac{1}{81!\,8!\,9!\,21!}\right) \\
&\doteq 0{,}043\,881.
\end{aligned}
$$

Pak $P_2 = P_1 - \hat{P}_1 \doteq 0{,}993\,407 - 0{,}043\,881 = 0{,}949\,526$.

V tabulce 2.1 jsou uvedeny pravděpodobnosti P_j, že $N_{12}+N_{23}+N_{13}+N_{123} \geq j$ pro $j = 0, \ldots, 15$. Výpočty jsou řešeny pomocí programu R, [13]. Hodnoty P_j jsou uvedeny i v [1], ovšem jen s třemi platnými číslicemi.

Střední hodnotu $\mathrm{E}(F_2 + F_3)$ počtu dnů, kdy došlo k alespoň dvěma záchvatům lze elegantně spočítat pomocí metod uvedených v podkapitole 3.2. Pro zajímavost uveďme, že vyjde $\mathrm{E}(F_2 + F_3) = 95/24 \doteq 3{,}958$. Z tabulky 2.1 je vidět, že medián náhodné veličiny $F_2 + F_3$ je 4.

□

Příklad 6: Představme si, že se chromozom skládá z n fragmentů. Křížíme dárce genomu s příslušníky nějakého kmenu. Kolik musíme provést křížení, aby došlo k výměně q procent genomu dárce s pravděpodobností alespoň p?

Řešení: Zadání můžeme interpretovat dvěma způsoby:

Interpretace 1: Každý fragment chromozomu dárce bude při křížení vyměněn s pravděpodobností 1/2. Situaci si můžeme představit tak, že máme n přihrádek,

j	P_j	j	P_j
0	$1{,}000000 \cdot 10^{0}$	8	$1{,}563827 \cdot 10^{-2}$
1	$9{,}934067 \cdot 10^{-1}$	9	$3{,}184225 \cdot 10^{-3}$
2	$9{,}495248 \cdot 10^{-1}$	10	$4{,}834451 \cdot 10^{-4}$
3	$8{,}208331 \cdot 10^{-1}$	11	$5{,}397229 \cdot 10^{-5}$
4	$6{,}003716 \cdot 10^{-1}$	12	$4{,}349514 \cdot 10^{-6}$
5	$3{,}537928 \cdot 10^{-1}$	13	$2{,}467859 \cdot 10^{-7}$
6	$1{,}631443 \cdot 10^{-1}$	14	$9{,}527486 \cdot 10^{-9}$
7	$5{,}789550 \cdot 10^{-2}$	15	$2{,}386008 \cdot 10^{-10}$

Tabulka 2.1: Pravděpodobnosti P_j, že došlo alespoň ke dvěma astmatickým záchvatům v alespoň j dnech.

každou obsadíme s pravděpodobností 1/2. Když toto obsazování zopakujeme r-krát, má každá přihrádka pravděpodobnost $1-(1/2)^r$, že bude obsazena. Hledáme nejmenší přirozené r, při kterém máme pravděpodobnost alespoň p, že q procent přihrádek je obsazeno. Nechť X je náhodná veličina označující počet prázdných přihrádek po r opakováních (tj. počet fragmentů chromozomu dárce, které se po r opakováních nevyměnily). Pak X má binomické rozdělení s parametry n a $(1/2)^r$. Chceme určit nejmenší přirozené r, aby platilo:

$$P\left(X \leq \left(1 - \frac{q}{100}\right) \cdot n\right) \geq p.$$

Pro hodnoty $n = 1000$, $q = 95$ a $p = 0{,}99$ vychází $r = 5$. Výpočet je proveden v programu R, [13].

Interpretace 2: Při každém křížení se vymění přesně polovina fragmentů (předpokládáme n sudé). Potom pravděpodobnost, že se po r kříženích změna projeví u m fragmentů, je dána pravděpodobností z problému komisí, velikost komise $w = \frac{n}{2}$, podle (1.9) máme:

$$p_m(r) = \frac{\binom{n}{m}}{\binom{n}{\frac{n}{2}}^r} \sum_{i=\frac{n}{2}}^{m} (-1)^{m-i} \binom{m}{i} \binom{i}{\frac{n}{2}}^r.$$

Hledáme nejmenší přirozené r, aby platilo:

$$\sum_{m=\lceil \frac{qn}{100} \rceil}^{n} p_m(r) \geq p.$$

Pro hodnoty stejné jako v předchozím případě vyjde po numerickém řešení opět $r = 5$. Tentokrát je výpočet prováděn v programu Mathematica, [14].

Kapitola 3

Modifikace a zobecnění problémů

3.1 Horní limit přihrádky

Problém s horním limitem přihrádky je uveden v [7], str. 252. Předpokládejme, že máme n přihrádek a každá z nich je schopna pojmout nejvýše s předmětů. Do přihrádek náhodně rozmístíme r předmětů tak, že pokud je nějaká přihrádka již plná, zaplňujeme zbývající přihrádky. Nechť K je náhodná veličina označující počet neprázdných přihrádek. Pak pravděpodobnost jevu, že neprázdných přihrádek je právě k, se rovná:

$$
\begin{aligned}
P[K = k] &= \frac{\binom{n}{k}}{\binom{ns}{r}} \left[\binom{sk}{r} - k\binom{s(k-1)}{r} + \binom{k}{2}\binom{s(k-2)}{r} + \cdots \right. \\
&\quad \left. + \quad (-1)^t \binom{k}{t}\binom{s(k-t)}{r} \right] \\
&= \frac{\binom{n}{k}}{\binom{ns}{r}} \sum_{i=0}^{t} (-1)^i \binom{k}{i}\binom{s(k-i)}{r},
\end{aligned}
$$

kde t je největší celé číslo takové, že $s(k-t) \geq r$.

Střední hodnota náhodné veličiny K se vypočítá:

$$
\mathrm{E}K = \sum_{k=0}^{n} k \frac{\binom{n}{k}}{\binom{ns}{r}} \sum_{i=0}^{t} (-1)^i \binom{k}{i}\binom{s(k-i)}{r}
$$

$$= \frac{n}{\binom{ns}{r}} \sum_{k=1}^{n} \binom{n-1}{k-1} \sum_{i=0}^{t} (-1)^i \binom{k}{i} \binom{s(k-i)}{r}.$$

Přeznačíme indexy: Označme $j = k - i$. Platí $s(k-t) \geq r$, a proto $k - i \geq \frac{r}{s}$, $i = 0, \ldots, t$. Čili $j \geq \left\lceil \frac{r}{s} \right\rceil$, $k \geq j$:

$$\mathrm{E}K = \frac{n}{\binom{ns}{r}} \sum_{j=\lceil \frac{r}{s} \rceil}^{n} \sum_{k=j}^{n} \binom{n-1}{k-1} (-1)^{k-j} \binom{k}{k-j} \binom{sj}{r}.$$

Ještě použijeme $l = n - k$:

$$\begin{aligned}
\mathrm{E}K &= \frac{n}{\binom{ns}{r}} \sum_{j=\lceil \frac{r}{s} \rceil}^{n} \binom{sj}{r} \sum_{l=0}^{n-j} \binom{n-1}{n-l-1} (-1)^{n-l-j} \binom{n-l}{n-l-j} \\
&= \frac{n}{\binom{ns}{r}} \sum_{j=\lceil \frac{r}{s} \rceil}^{n} \binom{sj}{r} \sum_{l=0}^{n-j} (-1)^{n-l-j} \cdot \frac{(n-1)!}{l!\,(n-l-1)!} \cdot \frac{(n-l)!}{j!\,(n-l-j)!} \\
&= \frac{n}{\binom{ns}{r}} \sum_{j=\lceil \frac{r}{s} \rceil}^{n} \binom{sj}{r} \sum_{l=0}^{n-j} (-1)^{n-l-j} \cdot \frac{(n-1)!\,(n-l)}{l!\,j!\,(n-l-j)!} \\
&= \frac{n}{\binom{ns}{r}} \sum_{j=\lceil \frac{r}{s} \rceil}^{n} \binom{sj}{r} \sum_{l=0}^{n-j} (-1)^{n-l-j} \cdot \frac{n-l}{n} \binom{n}{n-j} \binom{n-j}{l} \\
&= \frac{1}{\binom{ns}{r}} \sum_{j=\lceil \frac{r}{s} \rceil}^{n} \binom{sj}{r} \binom{n}{n-j} \sum_{l=0}^{n-j} (-1)^{n-l-j} \cdot (n-l) \binom{n-j}{l}.
\end{aligned}$$

Pro druhou sumu platí:

$$\begin{aligned}
&\sum_{l=0}^{n-j} (-1)^{n-l-j} \cdot (n-l) \binom{n-j}{l} \\
&= \sum_{l=0}^{n-j} (-1)^{n-l-j} \cdot n \binom{n-j}{l} - \sum_{l=0}^{n-j} (-1)^{n-l-j} \cdot l \binom{n-j}{l} \\
&= \sum_{l=0}^{n-j} (-1)^{n-l-j} \cdot n \binom{n-j}{l} - \sum_{l=1}^{n-j} (-1)^{n-l-j} \cdot (n-j) \binom{n-j-1}{l-1}
\end{aligned}$$

$$= \underbrace{n \cdot (1-1)^{n-j}}_{\substack{0 \text{ pro } j \neq n \\ n \text{ pro } j = n}} - \underbrace{(n-j) \cdot (1-1)^{n-j-1}}_{\substack{0 \text{ pro } j \neq n-1 \\ 1 \text{ pro } j = n-1}}.$$

Dosadíme do předchozího a dostaneme:

$$\begin{aligned} \mathrm{E}K &= \frac{1}{\binom{sn}{r}} \left[\binom{sn}{r} \cdot n - \binom{s(n-1)}{r} \cdot n \right] \\ &= n \left[1 - \frac{\binom{s(n-1)}{r}}{\binom{sn}{r}} \right]. \end{aligned}$$

3.2 Přihrádky s daným počtem předmětů

Další problém podobného charakteru je uveden v [12]: Celkem r předmětů je rovnoměrně náhodně rozděleno mezi neznámý počet přihrádek, mezi nimiž je zahrnuto N určitých přihrádek. Nechť p značí pravděpodobnost, se kterou je předmět umístěn do přihrádky. Označme F_k náhodnou veličinu, která udává počet přihrádek z N vybraných obsahujících právě k předmětů. Chceme určit střední hodnotu F_k.

Nejdříve uvažujme, že celkový počet přihrádek je N, tedy $p = 1/N$. Zadefinujme náhodnou veličinu X_{ij} následovně:

$$X_{ij} = \begin{cases} x, & \text{pokud je } j\text{-tý předmět v } i\text{-té přihrádce,} \\ 1 & \text{jinak,} \end{cases}$$

$i = 1, \dots, N$, $j = 1, \dots, r$.

Označme r_i jako počet předmětů v i-té přihrádce a počítejme:

$$\begin{aligned} \sum_{i=1}^{N} \prod_{j=1}^{r} X_{ij} &= \sum_{i=1}^{N} x^{r_i} \\ &= \sum_{k=0}^{r} F_k x^k. \end{aligned} \tag{3.1}$$

Platí:

$$\begin{aligned} \mathrm{E}X_{ij} &= x \cdot P[j\text{-tý předmět je v } i\text{-té přihrádce}] \\ &\quad + 1 \cdot P[j\text{-tý předmět není v } i\text{-té přihrádce}] \\ &= x \cdot \frac{1}{N} + 1 \cdot \frac{N-1}{N}. \end{aligned} \tag{3.2}$$

Aplikujeme střední hodnotu na (3.1) (využíváme nezávislost X_{ij} pro různá j) a dosadíme (3.2) a použijeme binomickou větu:

$$\begin{aligned}
\sum_{i=1}^{N}\prod_{j=1}^{r} EX_{ij} &= \sum_{k=0}^{r} EF_k x^k \\
&\Updownarrow \\
N\left(\frac{x}{N}+\frac{N-1}{N}\right)^r &= \sum_{k} EF_k x^k \\
&\Updownarrow \\
N\sum_{k=0}^{r}\binom{r}{k}\left(\frac{x}{N}\right)^k\left(\frac{N-1}{N}\right)^{r-k} &= \sum_{k=0}^{r} EF_k x^k.
\end{aligned}$$

Porovnáme koeficienty u x^k a vyjde nám:

$$\mathrm{E}F_k = N\binom{r}{k}\left(\frac{1}{N}\right)^k\left(1-\frac{1}{N}\right)^{r-k}.$$

V obecnějším případě uvažujme pravděpodobnost $p < 1/N$, že j-tý předmět je umístěn v i-té přihrádce. Použijeme stejný postup jako v minulém případě:

$$\begin{aligned}
EX_{ij} &= x\cdot p + 1\cdot(1-p) \\
\sum_{i=1}^{N}\prod_{j=1}^{r} EX_{ij} &= \sum_{k=0}^{r} EF_k x^k \\
&\Updownarrow \\
N\left(xp+(1-p)\right)^r &= \sum_{k=0}^{r} EF_k x^k \\
&\Updownarrow \\
N\sum_{k=0}^{r}\binom{r}{k} x^k p^k\left(1-p\right)^{r-k} &= \sum_{k=0}^{r} EF_k x^k.
\end{aligned}$$

Opět porovnáme koeficienty u x^k a vyjde nám:

$$\mathrm{E}F_k = N\binom{r}{k}p^k(1-p)^{r-k}.$$

Úlohu můžeme dále zobecnit pro problém komisí. Nechť F_k je náhodná veličina označující počet členů, kteří jsou právě v k komisích. Definujeme náhodnou veličinu X_{ij}:

$$X_{ij} = \begin{cases} x, & \text{pokud } i\text{-tý člen je v } j\text{-té komisi,} \\ 1 & \text{jinak.} \end{cases}$$

Střední hodnota X_{ij} se vypočítá:

$$\begin{aligned}
\mathrm{E}X_{ij} &= x \cdot P[i\text{-tý člen je v } j\text{-té komisi}] \\
&\quad + 1 \cdot P[i\text{-tý člen není v } j\text{-té komisi}] \\
&= x \cdot \frac{w}{N} + 1 \cdot \frac{N-w}{N}. \qquad (3.3)
\end{aligned}$$

Na (3.1), kde ovšem nyní r_i označuje počet komisí, v kolika je zastoupen i-tý člen, aplikujeme střední hodnotu, dosadíme (3.3) a použijeme binomickou větu:

$$\begin{aligned}
\sum_{i=1}^{N} \prod_{j=1}^{r} EX_{ij} &= \sum_{k=0}^{r} EF_k x^k \\
&\Updownarrow \\
N\left(x \cdot \frac{w}{N} + \frac{N-w}{N}\right)^r &= \sum_{k=0}^{r} EF_k x^k \\
&\Updownarrow \\
N \sum_{k=0}^{r} \binom{r}{k} \left(\frac{w}{N}\right)^k x^k \left(\frac{N-w}{N}\right)^{r-k} &= \sum_{k=0}^{r} EF_k x^k.
\end{aligned}$$

Porovnáme koeficienty u x^k a dostáváme:

$$\mathrm{E}F_k = N \binom{r}{k} \left(\frac{w}{N}\right)^k \left(1 - \frac{w}{N}\right)^{r-k}.$$

3.3 Členství ve skupinách

Tahle modifikace problému komisí je uvedena v [8], kap. 3.6.2, str. 170. Předpokládejme, že členové, ze kterých se má komise vybrat, se organizují v g disjunktních skupinách. Označme tyto skupiny symboly $\mathcal{G}_1, \ldots, \mathcal{G}_g$. Pravděpodobnost, že žádný člen ze skupiny $\mathcal{G}_h, h = 1, \ldots, g$, není v i-té komisi, $i = 1, \ldots, r$, je

$$\frac{\binom{n - n(\mathcal{G}_h)}{w_i}}{\binom{n}{w_i}},$$

kde $n(\mathcal{G}_h)$ značí počet členů skupiny $\mathcal{G}_h$.

Pravděpodobnost, že žádný člen ze skupiny $\mathcal{G}_h$ není členem žádné komise je

$$\mathcal{Q}_h = \prod_{i=1}^{r} \left\{ \frac{\binom{n - n(\mathcal{G}_h)}{w_i}}{\binom{n}{w_i}} \right\}.$$

Jestliže označíme množinu k skupin $\mathcal{G}_{h_1}, \ldots, \mathcal{G}_{h_k}$ symbolem $\mathcal{G}_{\mathbf{h}}^{(k)}$, pak pravděpodobnost, že žádný člen z těchto k skupin není členem nějaké komise, je

$$\mathcal{Q}_{\mathbf{h}}^{(k)} = \prod_{i=1}^{r} \left\{ \frac{\binom{n - n(\mathcal{G}_{\mathbf{h}}^{(k)})}{w_i}}{\binom{n}{w_i}} \right\},$$

kde $n(\mathcal{G}_{\mathbf{h}}^{(k)}) = \sum_{i=1}^{k} n(\mathcal{G}_{h_i})$ značí počet členů $\mathcal{G}_{\mathbf{h}}^{(k)}$. Nechť $S_k = \sum \cdots_{\mathbf{h}} \sum \mathcal{Q}_{\mathbf{h}}^{(k)}$, součet je přes všechna různá $\mathbf{h}$ o k složkách. Pak dostaneme pravděpodobnost P, že právě d skupin je zastoupeno v aspoň jedné komisi, neboli pravděpodobnost, že právě $g - d$ skupin není zastoupeno v žádné komisi:

$$\begin{aligned} P &= S_{g-d} - \binom{g-d+1}{g-d} S_{g-d+1} + \cdots + (-1)^d \binom{g}{g-d} S_g \\ &= \sum_{i=0}^{d} (-1)^i \binom{g-d+i}{g-d} S_{g-d+i}. \end{aligned} \tag{3.4}$$

Při odvození (3.4) se postupuje obdobně jako v (1.6).

Nyní uvažujme situaci, pro kterou platí:

- Počet členů ve skupině $\mathcal{G}_h$ přicházejících v úvahu pro členství v i-té komisi je $n_i(\mathcal{G}_h)$, kde n_i označuje celkový počet takových členů.
- Existuje r_i nominací pro i-tou komisi, které však mohou zahrnovat vícenásobné nominace pro některé členy, a konečná velikost komise w_i je proměnná.

Pak za předpokladu, že nominace jsou náhodné, pravděpodobnost, že pro i-tou komisi je nominován nějaký člen z $\mathcal{G}_{\mathbf{h}}^{(k)}$ je:

$$p_i(\mathcal{G}_{\mathbf{h}}^{(k)}) = \frac{n_i(\mathcal{G}_{\mathbf{h}}^{(k)})}{n_i}.$$

Pravděpodobnost, že žádný člen z $\mathcal{G}_{\mathbf{h}}^{(k)}$ není v žádné z r_i nominací, je:

$$\{q_i(\mathcal{G}_{\mathbf{h}}^{(k)})\}^{r_i},$$

kde

$$q_i(\mathcal{G}_{\mathbf{h}}^{(k)}) = 1 - p_i(\mathcal{G}_{\mathbf{h}}^{(k)}) = 1 - \frac{n_i(\mathcal{G}_{\mathbf{h}}^{(k)})}{n_i}.$$

Pro tento případ dostaneme

$$S_k = \sum_{\mathbf{h}} \cdots \sum \Big[\prod_{i=1}^{r} \{q_i(\mathcal{G}_{\mathbf{h}}^{(k)})\}^{r_i} \Big],$$

součet je přes všechna různá $\mathbf{h}$ o k složkách.

Ve speciálním případě, kdy každá skupina má tentýž počet n_i' členů vhodných pro členství v i-té komisi tak, že $n_i = gn_i'$, platí:

$$q_i(\mathcal{G}_{\mathbf{h}}^{(k)}) = 1 - \frac{kn_i'}{gn_i'} = \frac{g-k}{g}.$$

3.4 Znáhodněný problém komisí

Tento problém je uveden v [11] nebo v [8], kap. 3.6.3, str. 171. Předpokládejme, že v problému komisí je pravděpodobnost p, že osoba vybraná do komise členství odmítne. Potom pravděpodobnost, že z j-tice daných lidí jich y je vybráno a odmítne členství v komisi, zatímco zbylých $j-y$ není vybráno je:

$$\frac{\binom{j}{y} p^y \binom{n-j}{w-y}}{\binom{n}{w}}.$$

Tedy

$$p_j = \sum_{y=0}^{j} \frac{\binom{j}{y} p^y \binom{n-j}{w-y}}{\binom{n}{w}}$$

je pravděpodobnost, že j daných lidí neslouží v příslušné komisi. Označme E_i jev, že i-tý člověk není členem žádné komise. Tentokrát pro S_j platí:

$$S_j = \binom{n}{j} p_j^r. \tag{3.5}$$

Dosadíme (3.5) do (1.6) a po úpravě dostáváme pravděpodobnost, že právě m lidí je vybráno do aspoň jedné komise, neboli pravděpodobnost, že právě $n-m$ lidí není vybráno do žádné komise:

$$P_{[n-m]} \quad = \quad \sum_{i=0}^{n-(n-m)} (-1)^i \binom{n-m+i}{i} S_{n-m+i}$$

$$
\begin{aligned}
&= \sum_{i=0}^{m}(-1)^i \binom{n-m+i}{i}\binom{n}{n-m+i} p^r_{n-m+i} \\
&= \sum_{i=0}^{m}(-1)^i \frac{(n-m+i)!}{(n-m)!i!}\cdot\frac{n!}{(m-i)!(n-m+i)!} p^r_{n-m+i} \\
&= \binom{n}{m}\sum_{k=0}^{m}(-1)^{m-k}\binom{m}{m-k} p^r_{n-k} \\
&= \binom{n}{m}\sum_{k=0}^{m}(-1)^{m-k}\binom{m}{k} p^r_{n-k}.
\end{aligned}
$$

Toto okupační rozdělení má čtyři parametry: r, n, w, p.

3.5 Boseův–Einsteinův model

O posledních dvou modelech se lze dočíst v [15] nebo v [9].

Uvažujme r částic, z nichž každá je právě v jedné z n přihrádek. Předpokládejme, že:

– Částice jsou nerozlišitelné.

– Pro každou přihrádku známe počet částic, které jsou v ní umístěny.

– Stav systému je určen počty částic v jednotlivých přihrádkách.

– Všechny stavy jsou stejně pravděpodobné.

Boseovo–Einsteinovo rozmístění lze popsat jako neklesající posloupnost délky r prvků množiny $\{1, \ldots, n\}$:

$$
\underbrace{1\,1\ldots1}_{r_1}\underbrace{2\,2\ldots2}_{r_2}\;\ldots\;\underbrace{n\,n\ldots n}_{r_n},
$$

kde $\sum_{i=1}^{n} r_i = r$.

Takovýchto r-tic je stejně jako r-prvkových kombinací z n prvků s opakováním, tedy $\binom{r+n-1}{r} = \binom{r+n-1}{n-1}$.

Nechť náhodné veličiny $K_1, \ldots, K_n$ označují počet částic v přihrádkách $1, \ldots, n$. Takový jev, že v dané přihrádce je právě k částic, nastane tehdy, když zbývajících $r-k$ částic rozmístíme do zbývajících $n-1$ přihrádek. Proto dostáváme:

$$
P[K_1 = k] = \frac{\binom{r-k+n-2}{r-k}}{\binom{r+n-1}{r}}.
$$

Nyní určeme pravděpodobnost q toho, že v nějaké přihrádce není žádná částice. Mohou nastat dva případy: Pokud $r < n$, je určitě některá přihrádka prázdná a $q = 1$. Předpokládejme tedy, že $r \geq n$. Při výpočtu této pravděpodobnosti přejděme k jevu opačnému, tzn. určíme pravděpodobnost, že v každé přihrádce je aspoň jedna částice. Tento jev označme jako C. Pak vlastně rozdělujeme zbývajících $r - n$ částic:

$$\begin{aligned} P(C) &= \frac{\binom{(r-n)+n-1}{n-1}}{\binom{r+n-1}{n-1}} = \frac{\binom{r-1}{n-1}}{\binom{r+n-1}{n-1}}, \\ q &= 1 - P(C) = 1 - \frac{\binom{r-1}{n-1}}{\binom{r+n-1}{n-1}}. \end{aligned}$$

Podívejme se, jak vypadá okupační rozdělení v tomto případě. Pravděpodobnost, že právě m přihrádek bude obsazeno, je:

$$p_m = \frac{\binom{n}{m}\binom{r-1}{m-1}}{\binom{r+n-1}{n-1}}, \quad m = 1, \ldots, \min\{r, n\},$$

neboť $\binom{n}{m}$ způsoby můžeme vybrat m-tici obsazených přihrádek a do nich umisťujeme r částic tak, že v každé přihrádce je aspoň jedna částice.

3.6 Fermiův–Diracův model

Speciálním případem Boseova–Einsteinova modelu je model Fermiův–Diracův, použijeme-li Pauliho princip, tzn. že v každé přihrádce může být nejvýše jedna částice. Předpokládejme, že platí:

- Částice jsou nerozlišitelné.
- V každé přihrádce je nejvýše jedna částice.
- Stav systému je určen (neuspořádaným) seznamem přihrádek, které jsou obsazeny částicemi.
- Všechny stavy jsou stejně pravděpodobné.

Fermiův–Diracův model lze popsat pomocí uspořádaného r-rozměrného vektoru různých přirozených čísel z intervalu $[1, n]$, v tomto vektoru se každé číslo z intervalu $[1, n]$ vyskytuje nejvýše jednou. Takovýchto r-tic je stejně jako r-prvkových kombinací z n prvků, tedy $\binom{n}{r}$.

Označme D_j takový jev, že v j-té přihrádce, $j = 1, \ldots, n$, je umístěn nějaký (jeden) předmět. Pak platí:

$$P(D_j) = \frac{\binom{n-1}{r-1}}{\binom{n}{r}} = \frac{(n-1)!}{(n-r)!(r-1)!} \cdot \frac{(n-r)!r!}{n!} = \frac{r}{n}.$$

Pro pravděpodobnost q, že v j-té přihrádce není žádný předmět, platí:

$$q = 1 - P(D_j) = 1 - \frac{r}{n}.$$

Literatura

[1] P. J. Eicker, M. M. Siddiqui and P. W. Mielke, Jr. (1972): A matrix occupancy problem, *Ann. Math. Statist.* **43**, 988 –996.

[2] V. Eramo, M. Listanti, C. Nuzman and P. Whiting (2002): Optical switch dimensioning and the classical occupancy problem, *Internat. J. Commun.* **15**, 127–141.

[3] W. Feller (1957): *An Introduction to Probability Theory and Its Applications Vol. 1*, second edition, Wiley, New York.

[4] D. Gardy (2002): Occupancy urn models in the analysis of algorithms, *J. Statist. Plann. Inference* **101**, 95–105.

[5] A. Hald (1984): A. de Moivre: 'De Mensura Sortis' or 'On the measurement of chance', with a commentary, *Int. Statist. Rev.* **52**, 229–262.

[6] A. Hald (2003): *History of Probability and Statistics and Their Applications before 1750*, Wiley-Interscience.

[7] N. L. Johnson and S. Kotz (1969): *Discrete Distributions*, Wiley, New York.

[8] N. L. Johnson and S. Kotz (1977): *Urn Models and their Application. An Approach to Modern Discrete Probability Theory*, Wiley, New York.

[9] N. L. Johnson, S. Kotz and A. W. Kemp (1993): *Univariate Discrete Distributions*, second edition, Wiley, New York.

[10] N. Mantel and B. S. Pasternack (1968): A class of occupancy problems, *Am. Stat.* **22**, 23–24.

[11] D. A. Sprott (1969): A note on a class of occupancy problems, *Am. Stat.* **23**, 12–13.

[12] J. W. Tukey (1949): Moments of random group size distributions, *Ann. Math. Statist.* **20**, 523–539.

[13] R Development Core Team (2004): *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. URL: `http://www.R-project.org`

[14] Wolfram Research, Inc. (2005): *Mathematica, version 5.2.*

[15] K. Zvára, J. Štěpán (2002): *Pravděpodobnost a matematická statistika*, třetí vydání, Matfyzpress, Praha.