Univerzita Karlova v Praze

Matematicko-fyzikální fakulta

BAKALÁŘSKÁ PRÁCE

Veronika Nývltová

Problém sběratele kupónů

Katedra pravděpodobnosti a matematické statistiky

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

Praha 2012

Ráda bych zde vyjádřila svůj dík panu RNDr. Zbyňku Pawlasovi, Ph.D. za všechen svůj čas, který věnoval vedení práce, za podnětné nápady a důsledné připomínkování již napsaného textu a také za zapůjčení potřebné literatury.

Prohlašuji, že jsem tuto bakalářskou práci vypracovala samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova v Praze má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V .... dne ....

Název práce: Problém sběratele kupónů

Autor: Veronika Nývltová

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: RNDr. Zbyněk Pawlas, Ph.D., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: V bakalářské práci hledáme odpověď na otázku, kolik nákupů musíme uskutečnit, abychom získali sadu kartiček. Sadou rozumíme buď všechny typy kartiček, které výrobce přibaluje k výrobkům, nebo jen vybrané typy kartiček. Nejprve předpokládáme, že kartičky jsou přibalovány k výrobkům všechny se stejnou pravděpodobností. Počet potřebných nákupů je náhodný, zjišťujeme jeho střední hodnotu, rozptyl i pravděpodobnostní rozdělení. Studujeme limitní chování při počtu typů kartiček jdoucí k nekonečnu. Odpověď na stejnou otázku hledáme i v případě, že sbíráme několik sad – ať už úplných či neúplných – najednou. Pokud kartičky nemají stejnou pravděpodobnost obdržení, pak popisujeme střední hodnotu a rozptyl počtu nákupů, které musíme uskutečnit, abychom nasbírali několik úplných sad.

Klíčová slova: Gumbelovo rozdělení, limitní věty, Poissonizace, problém sběratele kupónů, Schurova konkávní funkce

Title: Coupon's collector problem

Author: Veronika Nývltová

Department: Department of Probability and Mathematical Statistics

Supervisor: RNDr. Zbyněk Pawlas, Ph.D., Department of Probability and Mathematical Statistics

Abstract: In the presented work we are looking for the answer to the question how many purchases must be made for obtaining a collection of cards. As a collection we understand all types of cards, which are packaged with products or we consider a collection of chosen types of these cards. First it is assumed that all cards are uniformly distributed. The number of required purchases is random and we derive its mean value, variance and probability distribution. We study limit behaviour when the number of types of cards is going to infinity. We are looking for the answer to the same question in the case of collecting several collections of cards at the same time. These collections could be complete or incomplete. In the case that cards are not uniformly distributed we describe mean value and variance of the number of purchases necessary for acquiring several collections of cards.

Keywords: Gumbel distribution, Limit theorems, Poissonization, Coupon's collector problem, Schur-concave function

Obsah

Úvod

Dobrým marketingovým tahem dnešních obchodníků je přidávání sběratelských kartiček k jednotlivým výrobkům či k celému nákupu. Podívejme se však na tento problém z pohledu matematika.

Je přirozené ptát se, kolik výrobků musíme zakoupit, abychom získali úplnou sbírku kartiček. Inverzní problém k této úloze je, kolik různých kartiček získáme, když nakoupíme daný počet výrobků. Tento problém se dá formulovat více způsoby a zabývali se jím již de Moivre, Laplace či Euler v 18. století. Například Problem 18 v de Moivrově spisu *De Mensura Sortis*, viz [6], formuluje úlohu dvou hráčů, kteří hází několikrát vícestěnnou kostkou a mají určený pevný počet různých stěn, které musí hráči padnout. Hledána je pravděpodobnost, že při daném počtu hodů kostkou padne určitý počet různých stěn. Kromě úlohy dvou hráčů o házení kostkami se na tuto úlohu můžeme koukat i jako na okupační problém. Ten spočívá ve vytahování čísel z osudí a umisťování míčků do přihrádek s vytaženým číslem: Vylosujeme číslo z osudí, do přihrádky s tímto číslem umístíme míček a číslo vrátíme do osudí. Je-li v přihrádce alespoň jeden míček, je „okupována“. Můžeme se ptát, kolik různých přihrádek je okupováno po určitém počtu tažení čísel. Jinak se můžeme ptát na situaci, kolik míčků potřebujeme, abychom zaplnili všechny přihrádky. Na podobném principu funguje i tzv. narozeninový problém, který řeší, jestli se ve stanovené skupině lidí najde dvojice, která slaví narozeniny týž den. V naší formulaci úlohy nás zajímá, kolik lidí potřebujeme, abychom pro každý den v roce měli osobu, která v ten den bude slavit narozeniny.

Úlohu ze začátku minulého odstavce formulujeme jako úlohu sbírání kartiček s obrázky, tzv. *Coupon's collector problem*, kdy budeme nakupovat výrobky, ve kterých jsou umístěny kartičky. Jde o posloupnost nezávislých nákupů, při kterých uvažujeme, že v každém nákupu máme pro každý typ kartičky stejnou pravděpodobnost jejího obdržení, tzn. nesnižuje se nám pravděpodobnost, že obdržíme typ kartičky, který již máme. V anglicky psané literatuře je tato úloha někdy známá také pod pojmem *The dixie cup problem*. To podle kelímků jedné společnosti prodávající zmrzlinu, která umisťovala obrázky známých hráčů na víčka svých obalů zmrzliny. Dnes se obrázky přibalují do žvýkaček, přidávají se k nákupu u pokladny, děti si je lepí do svého sběratelského alba a je okolo toho hodně marketingu. Zákazník by si měl spočítat, kolik výrobků musí koupit, aby získal úplnou sadu obrázků přibalených k výrobkům, a pak se rozhodnout, zda-li se mu vyplatí se do sbírání kartiček či obrázků vůbec pouštět. Výpočty, které zákazníkovi napomáhají k rozhodnutí, nalezne na následujících stranách, ale zda-li se mu takové nákupy kartiček vyplatí, si již musí zvážit sám.

Tato práce je inspirována kapitolou 5.3 v [2], ve které se autor zabývá případem, kdy přidávané kartičky jsou rovnoměrně rozdělené, tj. každá má stejnou pravděpodobnost obdržení. Stejný případ je podrobněji rozebírán v kapitole 1: V podkapitole 1.1 je uveden výpočet, kolik výrobků si bude muset zákazník ve střední hodnotě koupit, pokud chce jednu sadu kartiček. Pokud si přeje nasbírat jen vybrané typy kartiček, pak je výpočet pro tento případ uveden v podkapitole 1.2. Rozhodne-li se zákazník nakoupit jen určitý počet výrobků, pak pravděpodobnost, že mezi nimi bude úplná sbírka kartiček, najde v podkapitole 1.3. K úloze získávání jedné úplné sbírky je v podkapitole 1.4 ukázáno limitní rozdělení pro velký počet různých kartiček blížící se nekonečnu. V podkapitole 1.5 je popsána situace, ve které je již nasbírána jedna úplná sbírka a ptáme se, kolik typů kartiček chybí, abychom měli více kompletních sbírek. Pokud zákazníkovi nebude stačit jen jedna sada kartiček, ale bude si přát nasbírat hned několik sad najednou, pak výpočet střední hodnoty počtu zakoupených výrobků takových, že z nich lze sestavit několik sbírek, najde v podkapitole 1.6. Případ, kdy kartičky nejsou rovnoměrně rozdělené, je rozebírán v kapitole 2. Tam najdeme – podkapitola 2.1 – objasnění techniky vnoření do Poissonova procesu a spočtenou střední hodnotu a rozptyl počtu nákupů, které zákazník potřebuje k získání několika úplných sad kartiček. V podkapitole 2.2 je odvozen střední počet chybějících typů kartiček v nerovnoměrném případě a je zde určena výpočetně jednodušší horní a dolní mez.

V literatuře se dají najít různá další zobecnění základní úlohy sběratele, které už nejsou v této práci zmíněny. Například případ, kdy výrobce přibaluje hned několik kartiček do jednoho výrobku najednou, je zkoumán v [10] nebo v [16].

Některé vzorce a úlohy jsou doplněny o ilustrační výpočty a obrázky. Všechny byly provedeny pomocí programu Mathematica, viz [17].

1. Rovnoměrně náhodně rozdělené kartičky

1.1 Chceme získat jednu úplnou sbírku kartiček

Mějme sadu různých typů kartiček, které budeme umisťovat do balení výrobků. Nejprve předpokládejme, že počet typů kartiček v sadě je pevný a předem dán. Celkový počet různých typů kartiček označme n. Předpokládejme, že kartičky do výrobků přibalujeme rovnoměrně a náhodně. Zajímá nás, jaká je střední hodnota počtu výrobků, které zákazník musí zakoupit, aby měl doma úplnou sadu kartiček.

Předpokládejme, že již máme sbírku i různých typů kartiček. Označme X_i počet výrobků, které musíme dokoupit, abychom získali typ kartičky, který ještě nemáme ve sbírce. Pak

$$S_n = 1 + X_1 + X_2 + \cdots + X_{n-1}$$

značí celkový počet zakoupených výrobků potřebných k získání celé sbírky. Všimněme si, že X_i je počet pokusů, které potřebujeme k získání prvního úspěchu v řadě nezávislých pokusů. Pravděpodobnost, že získáme výrobek, který ještě nemáme, je

$$p_i = \frac{n-i}{n}, \quad i = 1, \ldots, n-1.$$

Tudíž pravděpodobnost, že X_i nabude hodnoty k, je

$$P(X_i = k) = p_i(1-p_i)^{k-1}$$

pro $k = 1, 2, \ldots$, a tedy X_i má *geometrické rozdělení* s parametrem p_i. Pravděpodobnosti p_i jsou navzájem různé, a tak X_i nejsou stejně rozdělené veličiny. Pro jednoduchost dalšího zápisu položme $X_0 = 1$.

Jelikož platí

$$\begin{aligned} EX_i &= \sum_{k=1}^{\infty} kP(X_i = k) = \sum_{k=1}^{\infty} kp_i(1-p_i)^{k-1} = p_i \sum_{k=1}^{\infty} k(1-p_i)^{k-1} = \\ &= -p_i \left[\sum_{k=0}^{\infty} (1-p_i)^k\right]' = -p_i \left[\frac{1}{(1-(1-p_i))}\right]' = -p_i \left[\frac{1}{p_i}\right]' = \frac{-p_i}{-p_i^2} = \frac{1}{p_i} = \\ &= \frac{1}{\frac{n-i}{n}} = \frac{n}{n-i}, \end{aligned}$$

dostaneme

$$\begin{aligned} ES_n &= E(1 + X_1 + X_2 + \cdots + X_{n-1}) = E1 + EX_1 + EX_2 + \cdots + EX_{n-1} = \\ &= 1 + \frac{n}{n-1} + \frac{n}{n-2} + \cdots + \frac{n}{2} + \frac{n}{1} = \\ &= n\left(\frac{1}{n} + \frac{1}{n-1} + \frac{1}{n-2} + \cdots + \frac{1}{2} + 1\right). \end{aligned}$$

Nyní můžeme zjistit, jak se chová střední hodnota počtu zakoupených výrobků potřebných k získání celé sady kartiček při zvětšujícím se počtu typů kartiček. Podle vzorce na straně 84 v [2] víme, že pro $n \to \infty$ je

$$ES_n = n\left(\ln n + \gamma + \frac{1}{2n} + \frac{1}{12n^2} + o(n^{-4})\right),$$

kde

$$\gamma = \lim_{n\to\infty}\left[\sum_{i=1}^{n}\frac{1}{i} - \ln n\right] = 0{,}577\,215\,66\ldots$$

je *Eulerova konstanta.*

Pro výpočet rozptylu X_i nejprve spočteme EX_i^2. Použijeme k tomu derivaci rovnosti

$$\sum_{k=1}^{\infty} k(1-p_i)^k = \frac{1-p_i}{(1-(1-p_i))^2}.$$

Ta po úpravě vyjde

$$-\sum_{k=0}^{\infty} k^2(1-p_i)^{k-1} = -\frac{2-p_i}{p_i^3}.$$

Pak platí

$$\begin{aligned} EX_i^2 &= \sum_{k=1}^{\infty} k^2 p_i(1-p_i)^{k-1} = p_i\sum_{k=1}^{\infty} k^2(1-p_i)^{k-1} = p_i\frac{2-p_i}{p_i^3} = \frac{2-p_i}{p_i^2} = \\ &= \frac{2-\frac{n-i}{n}}{(\frac{n-i}{n})^2} = \frac{n(n+i)}{(n-i)^2} = \frac{n^2+ni}{(n-i)^2}, \end{aligned}$$

z čehož máme

$$\operatorname{var} X_i = EX_i^2 - (EX_i)^2 = \frac{n^2+ni}{(n-i)^2} - \left(\frac{n}{n-i}\right)^2 = \frac{ni}{(n-i)^2}.$$

Jelikož náhodné veličiny X_i jsou nezávislé, můžeme rozptyl S_n spočítat jako

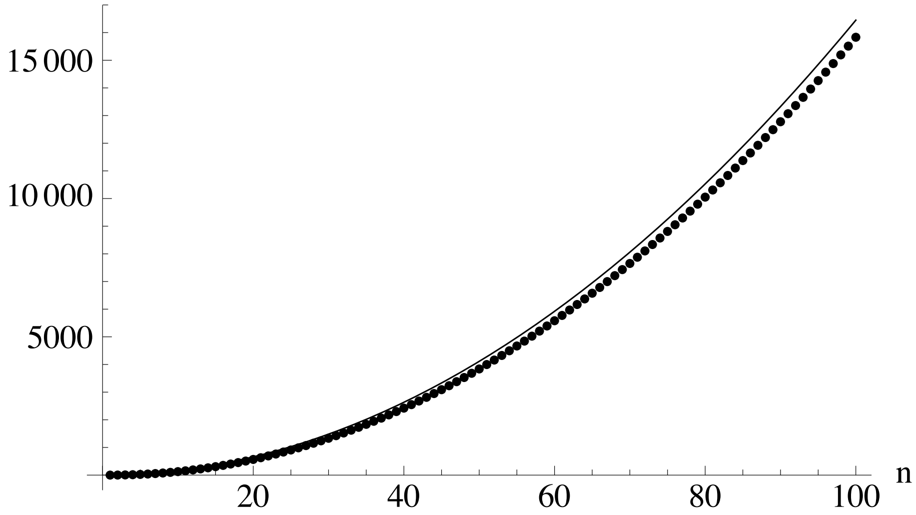
$$\operatorname{var} S_n = \sum_{i=0}^{n-1}\operatorname{var} X_i = \sum_{i=0}^{n-1}\frac{ni}{(n-i)^2} = n\sum_{i=0}^{n-1}\frac{i}{(n-i)^2}.$$

Protože rozptyl S_n můžeme rozepsat následovně

$$\begin{aligned} \operatorname{var} S_n &= n\sum_{i=0}^{n-1}\frac{i}{(n-i)^2} = n\sum_{i=1}^{n}\frac{n-i}{i^2} = n^2\sum_{i=1}^{n}\frac{1}{i^2} - n\sum_{i=1}^{n}\frac{1}{i} = \\ &= n^2\left(\frac{\pi^2}{6} - \sum_{i=n+1}^{\infty}\frac{1}{i^2}\right) - n\left[\ln n + \gamma + o(1)\right], \end{aligned}$$

a protože platí

$$\frac{1}{n+1} = \int_{n+1}^{\infty}\frac{1}{x^2}\,\mathrm{d}x \le \sum_{i=n+1}^{\infty}\frac{1}{i^2} \le \int_{n}^{\infty}\frac{1}{x^2}\,\mathrm{d}x = \frac{1}{n},$$



Obrázek 1.1: Rozptyl S_n pro $n = 23$ (tučně) v porovnání s členem $n^2 \frac{\pi^2}{6}$ (slabě)

je

$$\sum_{i=n+1}^{\infty} \frac{1}{i^2} = \frac{1}{n} + o\left(\frac{1}{n}\right), \quad n \to \infty,$$

tedy

$$\begin{aligned} \operatorname{var} S_n &= n^2 \left[\frac{\pi^2}{6} - \left(\frac{1}{n} + o\left(\frac{1}{n}\right)\right)\right] - n\left[\ln n + \gamma + o(1)\right] = \\ &= n^2 \frac{\pi^2}{6} - n \ln n - n(1+\gamma) + o(n), \quad n \to \infty. \end{aligned}$$

První člen tohoto rozvoje můžeme nazývat *hrubou aproximací*, viz [2], vzorec 5.13. Porovnání růstu rozptylu S_n s touto aproximací najdeme v obrázku 1.1.

1.2 Získáváme vybrané typy kartiček v jedné sadě

Budeme chtít jen předem daných l typů kartiček z celkového rozsahu n typů kartiček. Potřebný počet zakoupených výrobků k získání l-tice typů kartiček můžeme zapsat jako

$$S_{l,n} = X'_0 + X'_1 + \cdots + X'_{l-1},$$

kde X'_0 je počet výrobků, které musíme koupit, abychom dostali první typ kartičky z l-tice typů, kterou si přejeme. Potom X'_0 má geometrické rozdělení s parametrem $\frac{l}{n}$. Dále X'_1 je počet výrobků, které musíme zakoupit, abychom dostali typ kartičky různý od typu kartičky, kterou jsme dostali v předchozí várce nákupů, a zároveň to musí být typ kartičky ze žádané l-tice typů kartiček. Pak X'_1 má geometrické rozdělení s parametrem $\frac{l-1}{n}$. Obecně X'_i je počet výrobků, které musíme zakoupit, abychom poté, co již máme i typů kartiček z dané l-tice typů, dostali $(i+1)$-ní typ kartičky ze žádané l-tice typů. Pak X'_i má geometrické rozdělení s parametrem

$$p'_i = \frac{l-i}{n}, \quad i = 0, \ldots, l-1.$$

Střední hodnota počtu výrobků, které musíme zakoupit, abychom dostali typ kartičky, který ještě nemáme, ze žádané l-tice typů kartiček, je

$$EX_i' = \sum_{k=1}^{\infty} kP(X_i' = k) = \sum_{k=0}^{\infty} kp_i'(1-p_i')^{k-1} = \frac{1}{p_i'} = \frac{n}{l-i}.$$

Střední hodnota celkového počtu výrobků, které musíme koupit, abychom získali žádanou l-tici typů kartiček, je rovna

$$ES_{l,n} = E(X_0' + X_1' + \dots + X_{l-1}') = \sum_{i=0}^{l-1} EX_i' = \sum_{i=0}^{l-1} \frac{n}{l-i} = n\left(1 + \dots + \frac{1}{l}\right).$$

K výpočtu rozptylu $S_{l,n}$ použijeme následující rovnosti

$$EX_i'^2 = \sum_{k=1}^{\infty} k^2 p_i'(1-p_i')^{k-1} = \frac{2-p_i'}{p_i'^2} = \frac{2-\frac{l-i}{n}}{(\frac{l-i}{n})^2} = \frac{n^2(2n-l+i)}{n(l-i)^2} = n\frac{2n-l+i}{(l-i)^2},$$

$$\operatorname{var} X_i' = EX_i'^2 - (EX_i')^2 = n\frac{2n-l+i}{(l-i)^2} - \left(\frac{n}{l-i}\right)^2 = n\frac{n-l+i}{(l-i)^2},$$

pak rozptyl $S_{l,n}$ je roven:

$$\operatorname{var} S_{l,n} = \sum_{i=0}^{l-1} \operatorname{var} X_i' = n\sum_{i=0}^{l-1} \frac{n-l+i}{(l-i)^2}.$$

1.3 Souvislost s okupačním rozdělením

Uvažujme, že jsme zakoupili r výrobků. Zaveďme náhodnou veličinu $M(r)$, $r = 0, 1, \dots$ označující počet typů kartiček, které jsme získali do sbírky po r nákupech. Náhodná veličina $M(r)$ má tzv. *okupační rozdělení*. Název tohoto rozdělení pochází z interpretace úlohy v řeči urnových modelů, kdy $M(r)$ je počet obsazených (okupovaných) přihrádek při rovnoměrném náhodném umisťování r koulí do n přihrádek, více informací lze nalézt v [8]. Definujme náhodnou veličinu $N(r) = n - M(r)$ značící počet typů kartiček, které nám chybí ve sbírce po r nákupech.

Lemma 1.3.1. *Posloupnost $N(r), r = 0, 1, 2, \dots$ je Markovův řetězec a pravděpodobnosti přechodu jsou*

$$P\big(N(r) = m | N(r-1) = m\big) = 1 - \frac{m}{n}$$

a

$$P\big(N(r) = m-1 | N(r-1) = m\big) = \frac{m}{n}, \quad m = 0, 1, \dots, n.$$

Důkaz. Nejprve si uvědomíme, že markovská vlastnost je splněna, neboli dle [15], pro každé $j, i \in \{0, \dots, n\}$ platí:

$$P\big(N(r) = j | N(r-1) = i, N(r-2), \dots, N(0)\big) = P\big(N(r) = j | N(r-1) = i\big),$$

protože při znalosti celé historie počtu chybějících kartiček do času $r-1$ nám stačí pro určení podmíněné pravděpodobnosti znát $N(r-1)$.

Pro $N(r)$ nastane právě jedna ze dvou možností $N(r) = N(r-1)$ nebo $N(r) = N(r-1) - 1$. Nechť $m \in \{0, 1, \ldots, n\}$ je pevné. Pravděpodobnost $P\big(N(r) = m | N(r-1) = m\big)$ označuje pravděpodobnost situace, že při r-tém nákupu nekoupíme žádný z m typů, které ve sbírce nemáme. Pravděpodobnost, že při r-tém nákupu koupíme jeden typ z m typů je $\frac{m}{n}$, tj.

$$P\big(N(r) = m - 1 | N(r-1) = m\big) = \frac{m}{n}.$$

Pravděpodobnost jevu doplňkového je $1 - \frac{m}{n}$, tedy

$$P\big(N(r) = m | N(r-1) = m\big) = 1 - \frac{m}{n}.$$

□

Věta 1.3.2. *Definujeme-li posloupnost náhodných veličin*

$$T(r) = N(r)\left(\frac{n}{n-1}\right)^r, \quad r = 0, 1, 2, \ldots,$$

pak je tato posloupnost martingal.

Důkaz. Protože $N(r), r = 0, 1, 2, \ldots$ je Markovův řetězec, platí pro libovolné $m \in \{0, 1, \ldots, n\}$ následující rovnosti:

$$\begin{aligned} &E[N(r)|N(r-1) = m, N(r-2), \ldots, N(1), N(0)] = E[N(r)|N(r-1) = m] = \\ &= mP\big(N(r) = m|N(r-1) = m\big) + (m-1)P\big(N(r) = m-1|N(r-1) = m\big) = \\ &= m\left(1 - \frac{m}{n}\right) + (m-1)\frac{m}{n} = \frac{mn - m^2}{n} + \frac{m^2 - m}{n} = m\frac{n-1}{n}. \end{aligned}$$

Protože $N(r) \leq n$ pro každé $r = 0, 1, \ldots,$ je $EN(r) < \infty$, a protože $T(r)$ je vytvořena z $N(r)$ vynásobením kladným reálným číslem, je $E|T(r)| = ET(r) < \infty$. Dále platí:

$$\begin{aligned} &E\left[N(r)\left(\frac{n}{n-1}\right)^r \middle| N(r-1) = m\right] = \left(\frac{n}{n-1}\right)^r E\left[N(r)\big|N(r-1) = m\right] = \\ &= \left(\frac{n}{n-1}\right)^r m\frac{n-1}{n} = m\left(\frac{n}{n-1}\right)^{r-1}, \end{aligned}$$

což pro $t = m\left(\frac{n}{n-1}\right)^{r-1}$ dává

$$\begin{aligned} &E[T(r)\big|T(r-1) = t, \ldots, T(1), T(0)] = \\ &= E\left[N(r)\left(\frac{n}{n-1}\right)^r \middle| N(r-1)\left(\frac{n}{n-1}\right)^{r-1} = t, \ldots, N(0)\right] = \\ &= E\left[N(r)\left(\frac{n}{n-1}\right)^r \middle| N(r-1) = t\left(\frac{n-1}{n}\right)^{r-1}, \ldots, N(0)\right] = \\ &= E\left[N(r)\left(\frac{n}{n-1}\right)^r \middle| N(r-1) = t\left(\frac{n-1}{n}\right)^{r-1}\right] = t. \end{aligned}$$

Tedy pro libovolné $r = 1, 2, \dots$ máme rovnost

$$E[T(r) \big| T(r-1), \dots, T(1), T(0)] = T(r-1).$$

Nyní víme, že posloupnost $T(r)$, $r = 0, 1, 2, \dots$ je martingal.

□

Důsledek 1.3.3. *Platí*

$$EN(r) = n \left(\frac{n-1}{n} \right)^r, \quad r = 0, 1, 2, \dots$$

Důkaz. Pro $r = 0$ je

$$T(0) = N(0) \left(\frac{n}{n-1} \right)^0 = N(0) = n.$$

Protože posloupnost $T(r), r = 0, 1, 2, \dots$ je martingal, má konstantní střední hodnotu, která je rovna

$$ET(r) = ET(0) = En = n.$$

Protože platí

$$ET(r) = \left(\frac{n}{n-1} \right)^r EN(r),$$

ekvivalentní úpravou dostaneme

$$ET(r) \left(\frac{n-1}{n} \right)^r = EN(r),$$

kde $ET(r) = n$, což dává

$$EN(r) = n \left(\frac{n-1}{n} \right)^r.$$

□

Lemma 1.3.1, věta 1.3.2 a důsledek 1.3.3 jsou zformulovány v článku [10], kde také najdeme odvození dalších momentových charakteristik.

Ptejme se, jaká je pravděpodobnost, že zákazník zakoupil celou sbírku kartiček, koupí-li si r výrobků najednou, zde $r \geq n$. Značme tuto pravděpodobnost $P_{r,n} = P\big(M(r) = n\big) = P\big(N(r) = 0\big)$. Můžeme ji využít k zjištění rozdělení S_n, platí totiž

$$P(S_n = r) = P_{r,n} - P_{r-1,n}.$$

Věta 1.3.4. *Nechť $n \in \mathbb{N}$ a $r \in \mathbb{N}$, $r \geq n$. Pak*

$$P(S_n = r) = \sum_{i=0}^{n} (-1)^{i+1} \binom{n}{i} \left(1 - \frac{i}{n} \right)^{r-1} \frac{i}{n}.$$

Důkaz. Označme A_i situaci, že po nákupu r kartiček nám chybí i-tý typ kartičky. Protože při vybírání kartiček jde o variace s opakováním, kterých je n^r, je pravděpodobnost této situace rovna

$$P(A_i) = \frac{(n-1)^r}{n^r}$$

a pravděpodobnost, že nám chybí k typů kartiček, $k \leq n$, je rovna

$$P(A_1 \cap \cdots \cap A_k) = \frac{(n-k)^r}{n^r}.$$

Pravděpodobnost, že při koupi r výrobků najednou dostaneme kompletní sadu n kartiček, spočteme pomocí doplňkové pravděpodobnosti s využitím *principu inkluze a exkluze*, tj.

$$\begin{aligned}
1 - P_{r,n} = P(\bigcup_{i=1}^{n} A_i) &= \sum_{k=1}^{n} (-1)^{k-1} \sum_{I \in \binom{\{1,\dots,n\}}{k}} P(\bigcap_{i \in I} A_i) = \\
&= \sum_{k=1}^{n} (-1)^{k-1} \binom{n}{k} \frac{(n-k)^r}{n^r} = \sum_{k=1}^{n} (-1)^{k-1} \binom{n}{k} \left(1 - \frac{k}{n}\right)^r = \\
&= 1 - \left[\sum_{k=0}^{n} (-1)^k \binom{n}{k} \left(1 - \frac{k}{n}\right)^r\right],
\end{aligned}$$

kde $\binom{\{1,\dots,n\}}{k}$ je množina všech k-prvkových podmnožin množiny $\{1, \dots, n\}$.

Tedy

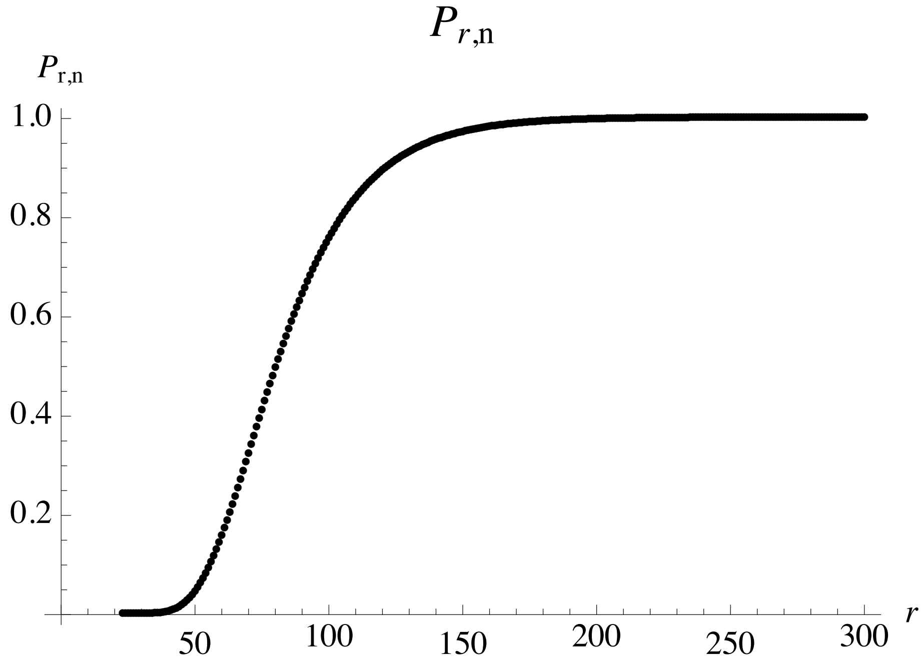
$$P_{r,n} = \sum_{i=0}^{n} (-1)^i \binom{n}{i} \left(1 - \frac{i}{n}\right)^r$$

a pro S_n platí:

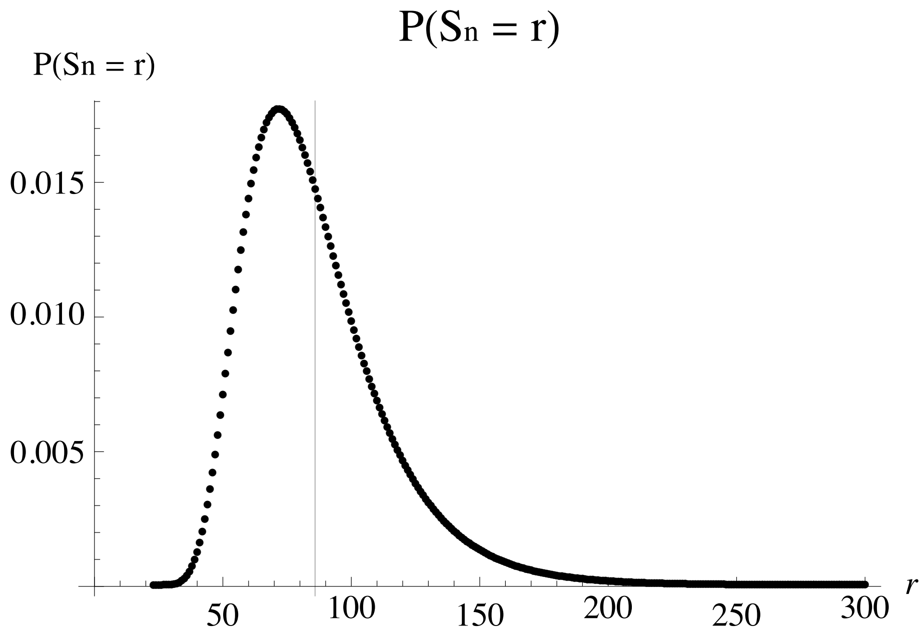
$$\begin{aligned}
P(S_n = r) = P_{r,n} - P_{r-1,n} &= \\
&= \sum_{i=0}^{n} (-1)^i \binom{n}{i} \left(1 - \frac{i}{n}\right)^r - \sum_{i=0}^{n} (-1)^i \binom{n}{i} \left(1 - \frac{i}{n}\right)^{r-1} = \\
&= \sum_{i=0}^{n} (-1)^i \binom{n}{i} \left(1 - \frac{i}{n}\right)^{r-1} \left(1 - \frac{i}{n} - 1\right) = \\
&= \sum_{i=0}^{n} (-1)^{i+1} \binom{n}{i} \left(1 - \frac{i}{n}\right)^{r-1} \frac{i}{n}.
\end{aligned}$$

□

Hodnoty $P_{r,n}$ jsou pro $n = 23$ znázorněny v obrázku 1.2 a hodnoty $P(S_n = r)$ jsou pro stejné n zobrazeny na obrázku 1.3 spolu se střední hodnotou. Pro $n = 23$ je $ES_n \doteq 85{,}89$. Můžeme si všimnout, že rozdělení S_n má kladnou šikmost a že nejpravděpodobnější doba čekání je kratší než očekávaná doba čekání, neboť modus S_n pro $n = 23$ je roven 72.



Obrázek 1.2: Pravděpodobnost obdržení celé sbírky při koupi r kartiček pro $n = 23$



Obrázek 1.3: Pravděpodobnost, že potřebujeme právě r výrobků k získání celé sady, s vyznačenou střední hodnotou S_n, $ES_n \doteq 85{,}89$, pro $n = 23$

1.4 Limitní chování počtu výrobků potřebných k získání jedné sady

V předchozí podkapitole jsme odvodili rozdělení náhodné veličiny S_n. Pro velká n je výpočet pravděpodobností $P(S_n = r)$ numericky náročnější. Následující věta dává limitní aproximaci pro velká n.

Věta 1.4.1. *Pro každé $x \in \mathbb{R}$ platí rovnost*

$$\lim_{n\to\infty} P\left(\frac{S_n - n\ln n}{n} < x\right) = \mathrm{e}^{-\,\mathrm{e}^{-x}}.$$

Důkaz. Důkaz provedeme podle [4].

Označme η_n, $n \in \mathbb{N}$, reálnou náhodnou veličinu

$$\eta_n = \frac{S_n - ES_n}{n}.$$

Nejprve vyjádřeme charakteristickou funkci X_j, které má geometrické rozdělení s parametry $p_j = \frac{n-j}{n}$, $j = 1, \dots, n-1$:

$$E\,\mathrm{e}^{\mathrm{i}tX_j} = \sum_{k=1}^{\infty} \mathrm{e}^{\mathrm{i}tk}\, p_j(1-p_j)^{k-1} = \frac{p_j}{1-p_j} \sum_{k=1}^{\infty} \left[\mathrm{e}^{\mathrm{i}t}(1-p_j)\right]^k, \quad t \in \mathbb{R}.$$

Protože pro $j = 1, \dots, n-1$ platí

$$|\,\mathrm{e}^{\mathrm{i}t}(1-p_j)| = |\,\mathrm{e}^{\mathrm{i}t}\,\frac{j}{n}| \le |\,\mathrm{e}^{\mathrm{i}t}\,||\frac{j}{n}| = \frac{j}{n} < 1,$$

můžeme sečíst geometrickou řadu

$$\sum_{k=1}^{\infty} \left[\mathrm{e}^{\mathrm{i}t}(1-p_j)\right]^k = \frac{\mathrm{e}^{\mathrm{i}t}\,(1-p_j)}{1-\mathrm{e}^{\mathrm{i}t}(1-p_j)},$$

pak pro $E\,\mathrm{e}^{\mathrm{i}tX_j}$ máme

$$E\,\mathrm{e}^{\mathrm{i}tX_j} = \frac{p_j}{1-p_j} \cdot \frac{\mathrm{e}^{\mathrm{i}t}\,(1-p_j)}{1-\mathrm{e}^{\mathrm{i}t}(1-p_j)} = \frac{p_j\,\mathrm{e}^{\mathrm{i}t}}{1-\mathrm{e}^{\mathrm{i}t}(1-p_j)} = \frac{\frac{n-j}{n}\,\mathrm{e}^{\mathrm{i}t}}{1-\mathrm{e}^{\mathrm{i}t}(1-\frac{n-j}{n})} =$$
$$= \frac{\frac{n-j}{n}\,\mathrm{e}^{\mathrm{i}t}}{1-\frac{j}{n}\,\mathrm{e}^{\mathrm{i}t}} = \frac{(n-j)\,\mathrm{e}^{\mathrm{i}t}}{n-j\,\mathrm{e}^{\mathrm{i}t}}.$$

Protože X_j jsou nezávislé, platí

$$E\,\mathrm{e}^{\mathrm{i}tS_n} = E\,\mathrm{e}^{\mathrm{i}t(1+X_1+\dots+X_{n-1})} = E\left[\mathrm{e}^{\mathrm{i}t}\,\mathrm{e}^{\mathrm{i}t(X_1+\dots+X_{n-1})}\right] = \mathrm{e}^{\mathrm{i}t}\,E\,\mathrm{e}^{\mathrm{i}t(X_1+\dots+X_{n-1})} =$$
$$= \mathrm{e}^{\mathrm{i}t} \prod_{j=1}^{n-1} E\,\mathrm{e}^{\mathrm{i}tX_j} = \mathrm{e}^{\mathrm{i}t} \prod_{j=1}^{n-1} \frac{(n-j)\,\mathrm{e}^{\mathrm{i}t}}{n-j\,\mathrm{e}^{\mathrm{i}t}} = \mathrm{e}^{\mathrm{i}tn} \prod_{j=1}^{n-1} \frac{n-j}{n-j\,\mathrm{e}^{\mathrm{i}t}}.$$

Pak charakteristická funkce η_n, označme ji $\varphi_n(t)$, je rovna

$$\varphi_n(t) = E\,\mathrm{e}^{\mathrm{i}t\eta_n} = E\,\mathrm{e}^{\mathrm{i}t\frac{S_n - ES_n}{n}} = E\left[\mathrm{e}^{\mathrm{i}\frac{t}{n}S_n}\,\mathrm{e}^{-\mathrm{i}\frac{t}{n}ES_n}\right] = \mathrm{e}^{-\mathrm{i}\frac{t}{n}ES_n}\,E\,\mathrm{e}^{\mathrm{i}\frac{t}{n}S_n} =$$

$$= \mathrm{e}^{-\mathrm{i}\frac{t}{n}n(\sum_{k=1}^{n}\frac{1}{k})}\,\mathrm{e}^{\mathrm{i}t}\prod_{j=1}^{n-1}\frac{n-j}{n-j\,\mathrm{e}^{\mathrm{i}\frac{t}{n}}} = \left(\prod_{k=1}^{n}\mathrm{e}^{-\frac{\mathrm{i}t}{k}}\right)\mathrm{e}^{\mathrm{i}t}\prod_{j=1}^{n-1}\frac{n-j}{n-j\,\mathrm{e}^{\mathrm{i}\frac{t}{n}}} =$$

$$= \prod_{k=1}^{n}\mathrm{e}^{-\frac{\mathrm{i}t}{k}}\prod_{j=0}^{n-1}\frac{(n-j)\,\mathrm{e}^{\mathrm{i}\frac{t}{n}}}{n-j\,\mathrm{e}^{\mathrm{i}\frac{t}{n}}} = \prod_{k=1}^{n}\mathrm{e}^{-\frac{\mathrm{i}t}{k}}\prod_{j=1}^{n}\frac{j\,\mathrm{e}^{\mathrm{i}\frac{t}{n}}}{n-(n-j)\,\mathrm{e}^{\mathrm{i}\frac{t}{n}}} =$$

$$= \prod_{k=1}^{n}\mathrm{e}^{-\frac{\mathrm{i}t}{k}}\prod_{j=1}^{n}\frac{j}{n\,\mathrm{e}^{-\mathrm{i}\frac{t}{n}}-(n-j)} = \prod_{k=1}^{n}\mathrm{e}^{-\frac{\mathrm{i}t}{k}}\prod_{j=1}^{n}\frac{j}{j+n(\mathrm{e}^{-\mathrm{i}\frac{t}{n}}-1)} =$$

$$= \prod_{k=1}^{n}\mathrm{e}^{-\frac{\mathrm{i}t}{k}}\prod_{j=1}^{n}\frac{1}{1+\frac{n}{j}(\mathrm{e}^{-\mathrm{i}\frac{t}{n}}-1)} = \frac{1}{\prod_{k=1}^{n}\mathrm{e}^{\frac{\mathrm{i}t}{k}}\left(1+\frac{n}{k}(\mathrm{e}^{-\mathrm{i}\frac{t}{n}}-1)\right)}.$$

Půjdeme-li s n do nekonečna, dostaneme

$$\lim_{n\to\infty}\varphi_n(t) = \lim_{n\to\infty}\frac{1}{\prod_{k=1}^{n}\mathrm{e}^{\frac{\mathrm{i}t}{k}}\left(1+\frac{n}{k}(\mathrm{e}^{-\mathrm{i}\frac{t}{n}}-1)\right)} =$$

$$= \lim_{n\to\infty}\frac{1}{\prod_{k=1}^{n}\mathrm{e}^{\frac{\mathrm{i}t}{k}}\left(1+\frac{n}{k}\left(\frac{\mathrm{e}^{\frac{-\mathrm{i}t}{n}}-1}{-\frac{\mathrm{i}t}{n}}\frac{-\mathrm{i}t}{n}\right)\right)} = \frac{1}{\prod_{k=1}^{\infty}\mathrm{e}^{\frac{\mathrm{i}t}{k}}\left(1+\frac{-\mathrm{i}t}{k}\right)} =$$

$$= \frac{1}{\prod_{k=1}^{\infty}\mathrm{e}^{\frac{\mathrm{i}t}{k}}\left(1-\frac{\mathrm{i}t}{k}\right)}. \tag{1.1}$$

V následujícím značme $\lim_{n\to\infty}\varphi_n(t)$ symbolem $\phi(t)$.

Nyní potřebujeme větu 14.21 z [11], která říká následující.

Tvrzení 1.4.2. *Označme $\hat{P}_X$ charakteristickou funkci náhodné veličiny X a označme $X_n \xrightarrow[n\to\infty]{\mathcal{D}} X$ konvergenci v distribuci náhodných veličin X_n k X. Nechť X_n jsou pro $n \in \mathbb{N}$ reálné náhodné veličiny a nechť $\hat{P}_{X_n}(t) \xrightarrow[n\to\infty]{} \psi(t)$ pro každé $t \in \mathbb{R}$. Je-li $\psi(t)$ spojitá v bodě 0, potom existuje reálná náhodná veličina X tak, že $X_n \xrightarrow[n\to\infty]{\mathcal{D}} X$ a $\hat{P}_X = \psi$.*

Použijeme tvrzení 1.4.2 pro $X_n = \eta_n$, $X = \eta$, $\hat{P}_{X_n}(t) = \varphi_n(t)$, $\psi(t) = \phi(t)$. Funkce $\phi(t)$ je spojitá v bodě 0:

$$\lim_{t\to 0}\phi(t) = \lim_{t\to 0}\frac{1}{\prod_{k=1}^{\infty}\mathrm{e}^{\frac{\mathrm{i}t}{k}}\left(1-\frac{\mathrm{i}t}{k}\right)} = \frac{1}{\prod_{k=1}^{\infty}\mathrm{e}^{0}} = \frac{1}{\prod_{k=1}^{\infty}1} = 1 = \phi(0),$$

a tedy existuje reálná náhodná veličina η taková, že $\eta_n \xrightarrow[n\to\infty]{\mathcal{D}} \eta$ a $\hat{P}_\eta = \phi$.

Použijeme-li ekvivalentní definici funkce Γ:

$$\Gamma(z) = \frac{\mathrm{e}^{-\gamma z}}{z}\prod_{k=1}^{\infty}\left(1+\frac{z}{k}\right)^{-1}\mathrm{e}^{\frac{z}{k}}$$

pro $z = -\mathrm{i}t$, pak použitím vztahu

$$z\Gamma(z) = \Gamma(1+z)$$

dostaneme z (1.1) rovnost

$$\phi(t) = \mathrm{e}^{-\gamma \mathrm{i}t}(-\mathrm{i}t)\frac{\mathrm{e}^{\gamma \mathrm{i}t}}{-\mathrm{i}t}\frac{1}{\prod_{k=1}^{\infty} \mathrm{e}^{\frac{\mathrm{i}t}{k}}\left(1-\frac{\mathrm{i}t}{k}\right)} = \mathrm{e}^{-\mathrm{i}t\gamma}\,\Gamma(1-\mathrm{i}t),$$

kde γ je Eulerova konstanta.

Nechť $F_1(x)$ je distribuční funkce *Gumbelova rozdělení*, tj. $F_1(x) = \mathrm{e}^{-\mathrm{e}^{-x}}$, $x \in \mathbb{R}$, pak příslušná charakteristická funkce má dle [5], strany 65, tvar

$$\int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}xt}\,\mathrm{d}F_1(x) = \Gamma(1-\mathrm{i}t), \quad t \in \mathbb{R}.$$

Tedy $\phi(t)$ je charakteristická funkce náhodné veličiny $\eta = X - \gamma$, kde X má Gumbelovo rozdělení a γ je Eulerova konstanta, jejíž hodnotu jsme připomněli v podkapitole 1.1.

Následně využijeme větu 13.12 z [11]:

Tvrzení 1.4.3. *Pro reálné náhodné veličiny X_n, $n \in \mathbb{N}$ a X s příslušnými distribučními funkcemi F_n a F platí: $X_n \xrightarrow[n\to\infty]{\mathcal{D}} X$ právě tehdy, když $\lim_{n\to\infty} F_n(x) = F(x)$ pro každý $x \in \mathbb{R}$ bod spojitosti F.*

Podle tvrzení 1.4.2 víme, že $\eta_n = \frac{S_n - ES_n}{n} \xrightarrow[n\to\infty]{\mathcal{D}} \eta$. Protože $P(\eta < x) = P(X - \gamma < x) = P(X < x + \gamma)$, je distribuční funkce η rovna funkci $\mathrm{e}^{-\mathrm{e}^{-(x+\gamma)}}$. Tato funkce je spojitá v každém bodě $x \in \mathbb{R}$, a tedy pro každé $x \in \mathbb{R}$ máme podle tvrzení 1.4.3:

$$\lim_{n\to\infty} P\left(\frac{S_n - ES_n}{n} < x\right) = P(\eta < x).$$

Rovnost platila pro každé $x \in \mathbb{R}$, tedy platí i pro každé $x - \gamma \in \mathbb{R}$, pak

$$\lim_{n\to\infty} P\left(\frac{S_n - ES_n}{n} < x - \gamma\right) = P(\eta < x - \gamma) = F_1(x)$$

a také platí následující vztahy

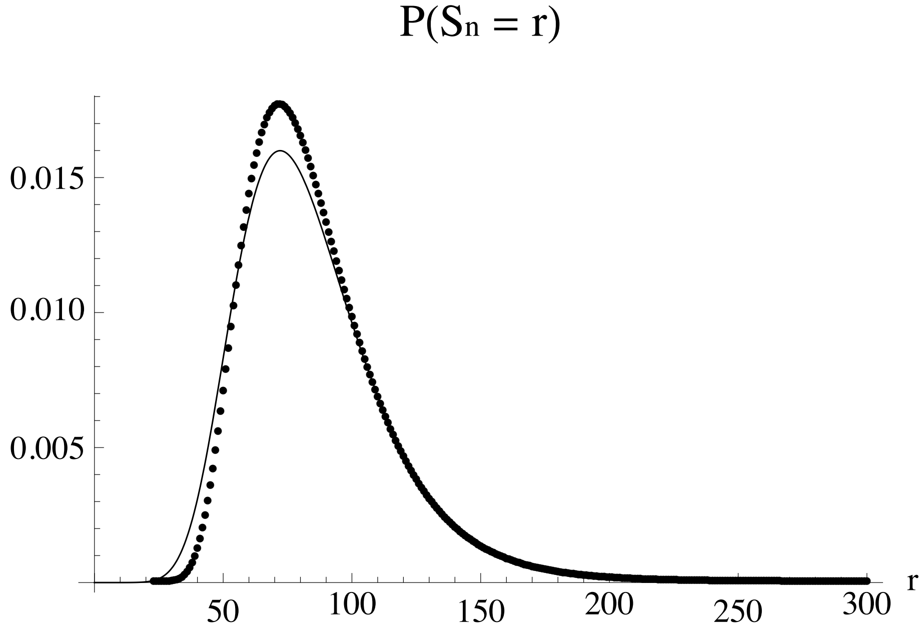
$$\begin{aligned}
\lim_{n\to\infty} P\left(\frac{S_n - n[\ln n + \gamma + o(1)]}{n} < x - \gamma\right) &= F_1(x),\\
\lim_{n\to\infty} P\left(\frac{S_n - n\ln n - n\gamma - o(n)}{n} + \gamma < x\right) &= F_1(x),\\
\lim_{n\to\infty} P\left(\frac{S_n - n\ln n - n\gamma - o(n) + n\gamma}{n} < x\right) &= F_1(x),\\
\lim_{n\to\infty} P\left(\frac{S_n - n\ln n}{n} < x\right) &= F_1(x).
\end{aligned}$$

Tedy pro každé $x \in \mathbb{R}$ máme rovnost

$$\lim_{n\to\infty} P\left(\frac{S_n - n\ln n}{n} < x\right) = \mathrm{e}^{-\mathrm{e}^{-x}}.$$

□

Nechť $f_1(x)$ je hustota Gumbelova rozdělení. Pro ilustraci limitní věty srovnejme hodnoty $P(S_n = r)$ s funkcí $g(r) = f_1(\frac{r - n \ln n}{n})$. Výsledek je vykreslen pro případ $n = 23$ na obrázku 1.4. Tučně je na něm znázorněn diskrétní graf pro $P(S_n = r)$ a slabou čárou je na něm znázorněna vhodně normovaná hustota $g(r)$ Gumbelova rozdělení.



Obrázek 1.4: Porovnání rozdělení S_{23} (tučně) a příslušné limitní aproximace hustotou Gumbelova rozdělení (slabě)

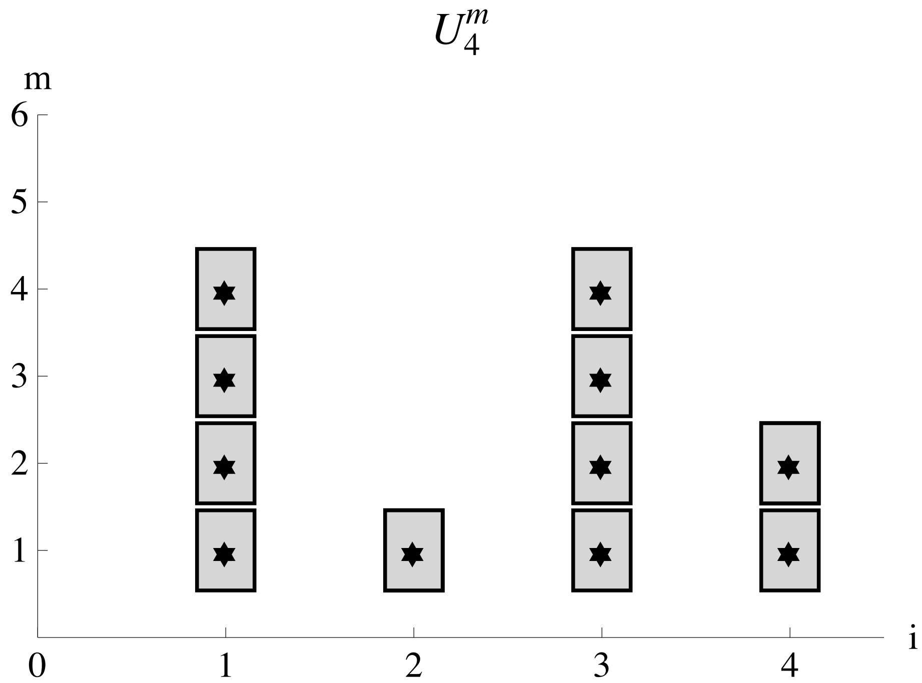
1.5 Počet chybějících typů kartiček v neúplných sadách

Nyní uvažujme okamžik, kdy jsme právě nasbírali jednu úplnou sadu kartiček, tzn. od každého typu kartičky máme alespoň jednu kartičku. Budeme chtít zjistit, kolik kartiček máme ve sbírce od i-tého typu, $i = 1, \ldots, n$. Nazývejme m-tou sadou, $m = 1, 2, \ldots$, množinu typů kartiček, které jsme již obdrželi m-krát. Postupujme podle článku [1].

Označme U_n^m počet typů kartiček, které ještě nemáme v m-té sadě, kde uvažujeme $m \geq 2$, neboť první sadu již máme nasbíránu celou. Pro n pevné je U_n^m neklesající posloupností v m.

Příklad. *Nakoupili jsme* 11 *výrobků s kartičkami, z nichž čtyři kartičky byly typu* 1*, jedna kartička byla typu* 2*, čtyři kartičky byly typu* 3 *a dvě kartičky byly typu* 4*. Kartičku typu* 2 *jsme získali jako poslední. Předpokládejme, že více typů kartiček výrobce do výrobků nepřibaluje a spočtěme* U_n^m.

Řešení. Máme čtyři různé typy kartiček, tzn. $n = 4$, a od každého typu alespoň jednu kartičku. Takovouto situaci můžeme zaznamenat v obrázku 1.5.



Obrázek 1.5: Ilustrace k příkladu U_n^m pro $n = 4$

Prázdné místo nad sloupcem odpovídajícím typu kartičky říká, že nám tento typ kartičky chybí k získání úplné m-té sady. Na jednotlivých hladinách odpovídajících $m = 2, 3, \ldots$ snadno spočteme hodnoty U_4^m:

$$U_4^2 = 1,$$
$$U_4^3 = 2,$$
$$U_4^4 = 2$$

a pro každé $j \geq 5$ je $U_4^j = 4$.

$\triangle$

Věta 1.5.1. *Pro každé $m \geq 2$ platí*

$$EU_n^m = \sum_{i=1}^{n} \binom{n}{i} \frac{(-1)^{i+1}}{i^{m-1}}.$$

Důkaz. Označíme-li A_i^m jev, kdy máme ve sbírce alespoň m kartiček od i-tého typu kartičky, pak U_n^m můžeme ekvivalentně napsat jako

$$U_n^m = n - \sum_{i=1}^{n} \mathbf{1}(A_i^m) = \sum_{i=1}^{n} \left[1 - \mathbf{1}(A_i^m)\right], \tag{1.2}$$

kde $\mathbf{1}(A_i^m)$ je indikátor jevu A_i^m, tj.

$$\mathbf{1}(A_i^m) = \begin{cases} 1, & \text{jestliže } A_i^m \text{ nastane,} \\ 0, & \text{jestliže } A_i^m \text{ nenastane.} \end{cases}$$

Nalezení kartičky ve výrobku má pro i-tý typ kartičky pravděpodobnost $\frac{1}{n}$. Pak pravděpodobnost, že nastane A_i^m je stejná pro každé $i \in \{1, \ldots, n\}$ a střední hodnotu U_n^m můžeme spočíst následovně

$$\begin{aligned} EU_n^m &= E\left(n - \sum_{i=1}^{n} \mathbf{1}(A_i^m)\right) = n - \sum_{i=1}^{n} E\mathbf{1}(A_i^m) = n - \sum_{i=1}^{n} P(A_i^m) = \\ &= n - nP(A_n^m) = n\left[1 - P(A_n^m)\right]. \end{aligned}$$

Nechť $B_{m,i}^n$ značí jev, kdy ve sbírce už máme alespoň m kartiček od n-tého typu kartičky a ještě jsme nekoupili výrobek s první kartičkou i-tého typu. Pak

$$P(A_n^m) = P\left(\bigcup_{i=1}^{n-1} B_{m,i}^n\right)$$

a pro $m \geq 2$ pomocí principu inkluze a exkluze máme

$$\begin{aligned} P(A_n^m) &= \sum_{i=1}^{n-1} (-1)^{i+1} \sum_{I \in \binom{\{1,\dots,n-1\}}{i}} P\left(\bigcap_{l \in I} B_{m,l}^n\right) = \\ &= \sum_{i=1}^{n-1} (-1)^{i+1} \binom{n-1}{i} \left(\frac{1}{i+1}\right)^m, \end{aligned}$$

protože vybíráme jednu kartičku n-tého typu z $(i+1)$ typů kartiček m-krát, je pravděpodobnost průniku $(\frac{1}{i+1})^m$.

Tedy

$$\begin{aligned} EU_n^m &= n\left[1 - P(A_n^m)\right] = n\left[1 - \sum_{i=1}^{n-1} (-1)^{i+1} \binom{n-1}{i} \left(\frac{1}{i+1}\right)^m\right] = \\ &= n + \sum_{i=1}^{n-1} (-1)^{i+2} \binom{n}{i} (n-i) \left(\frac{1}{i+1}\right)^m = \\ &= \sum_{i=0}^{n-1} (-1)^{i+2} \binom{n}{i} (n-i) \left(\frac{1}{i+1}\right)^m = \\ &= \sum_{k=1}^{n} (-1)^{k+1} \binom{n}{k} k \left(\frac{1}{k}\right)^m = \sum_{k=1}^{n} (-1)^{k+1} \binom{n}{k} \frac{1}{k^{m-1}}, \end{aligned}$$

což jsme chtěli ukázat.

□

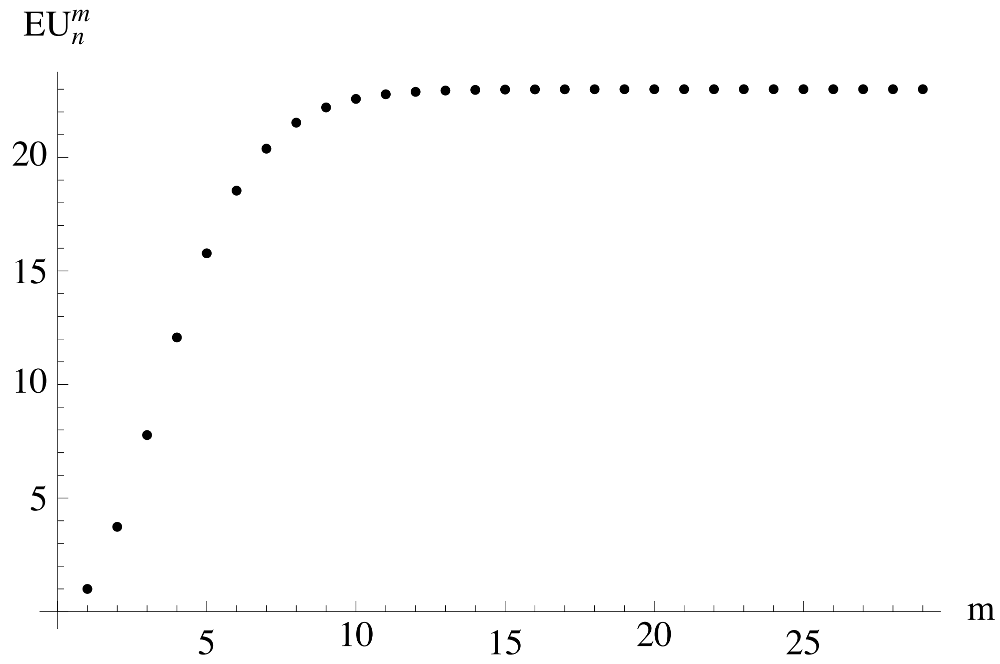
Poznámka. *Platí:*

$$EU_n^m \xrightarrow[m \to \infty]{} n.$$

Důkaz. První člen $(i = 1)$ součtu ve větě 1.5.1 je roven n. Všichni ostatní sčítanci jdou pro $m \to \infty$ k 0. Z *aritmetiky limit* neklesající posloupnost EU_n^m konverguje k n pro $m \to \infty$.

□

Průběh EU_n^m pro případ $n = 23$ je vykreslen v obrázku 1.6.



Obrázek 1.6: Hodnoty EU_n^m v závislosti na m při pevném $n = 23$

1.6 Několik sad kartiček

Uvažujme nyní situaci, ve které chceme nasbírat více sad kartiček, avšak nebudeme nutně chtít úplné sady o n typech kartiček, budeme chtít nasbírat jen sady vybraných typů kartiček. Chtějme tedy jen l, $l \leq n$, typů kartiček a každý typ z této l-tice chtějme nasbírat m-krát. Důležitým případem této situace je situace, kdy chceme nasbírat úplných m sad o n různých typech kartiček, tj. $l = n$.

Označme počet výrobků, které musí zákazník koupit, aby měl doma úplných m sad tvořených l typy kartiček, v závislosti na rozsahu jedné sady symbolem $S_{l,n}^m$. Ve speciálním případě $l = n$ značme $S_{n,n}^m = S_n^m$. Chceme-li jen l, $l \leq n$ typů kartiček ze sady o celkovém počtu n typů, pak bez újmy na obecnosti můžeme označit kartičky, které chceme, $1, 2, \ldots, l$ a kartičky, které nejsou v námi vybrané l-tici, $l + 1, \ldots, n$. Pak pro jakýkoliv počet takovýchto sad platí následující věta.

Věta 1.6.1. *Pro každé $m \in \mathbb{N}$, $n \in \mathbb{N}$, $l \in \mathbb{N}$ takové, že $l \leq n$, platí*

$$ES_{l,n}^m = n \int_0^\infty \left[1 - (1 - \delta_m(t)\,\mathrm{e}^{-t})^l\right] \mathrm{d}t,$$

kde

$$\delta_m(t) = \sum_{k=0}^{m-1} \frac{t^k}{k!}, \quad t \in \mathbb{R}.$$

Speciálně pro $l = n$ platí

$$ES_n^m = n \int_0^\infty [1 - (1 - \delta_m(t)\,\mathrm{e}^{-t})^n]\,\mathrm{d}t.$$

Důkaz. Definujme B_i jevy takové, že po koupi i kartiček stále nemáme celou sbírku m l-tic, avšak při koupi $(i+1)$-ní kartičky již celou takovou sbírku získáme. Pak B_i jsou po dvou disjunktní a $\bigcup_{i=0}^\infty B_i$ dává celý pravděpodobnostní prostor.

Označme p_i pravděpodobnost, že po nákupu i kartiček stále nebudeme mít m úplných l-tic kartiček. Označíme-li N_i počet možností, jak koupit i výrobků, aniž bychom mezi nimi získali kompletních m sad l-tic, pak

$$p_i = \frac{N_i}{n^i}.$$

Střední hodnotu $ES_{l,n}^m$ můžeme spočíst následovně

$$ES_{l,n}^m = \sum_{i=0}^{\infty}(i+1)P(B_i) = \sum_{i=0}^{\infty}(i+1)(p_i - p_{i+1}) =$$
$$= p_0 - p_1 + 2p_1 - 2p_2 + 3p_2 - 3p_3 + \cdots = \sum_{i=0}^{\infty} p_i.$$

Budeme-li nakupované typy kartiček značit $x_1, \ldots, x_n$, pak k dalším výpočtům můžeme použít *multinomickou větu* ze strany 81 v [13], která říká, že pro každé $x_1, \ldots, x_n \in \mathbb{R}$ a pro každé $k \in \mathbb{N}$ platí vztah

$$(x_1 + x_2 + \cdots + x_n)^k = \sum_{\substack{k_1+\cdots+k_n=k \\ k_1,\ldots,k_n \geq 0}} \binom{k}{k_1,\ldots,k_n} x_1^{k_1}\cdots x_n^{k_n}, \tag{1.3}$$

kde $\binom{k}{k_1,\ldots,k_n}$ je *multinomický koeficient*, který je roven

$$\binom{k}{k_1,\ldots,k_n} = \frac{k!}{k_1!\cdots k_n!}.$$

Odebereme-li na pravé straně výrazu (1.3) všechny sčítance, které mají mocniny u všech činitelů sbírané l-tice kartiček větší nebo rovny m, odstraníme tím všechny způsoby, jak získat kompletních m sad sbírané l-tice typů kartiček. Pak tedy N_i můžeme vyjádřit jako součet výrazu

$$\sum_{\substack{k_1+\cdots+k_n=i \\ k_1,\ldots,k_n \geq 0 \\ \min_{j\in\{1,\ldots,l\}} k_j < m}} \binom{i}{k_1,\ldots,k_n} x_1^{k_1}\cdots x_n^{k_n}$$

při dosazení $(1,\ldots,1)$ za $(x_1,\ldots,x_n)$, tj.

$$N_i = \sum_{\substack{k_1+\cdots+k_n=i \\ k_1,\ldots,k_n \geq 0 \\ \min_{j\in\{1,\ldots,l\}} k_j < m}} \binom{i}{k_1,\ldots,k_n} = \sum_{\substack{k_1+\cdots+k_n=i \\ k_1,\ldots,k_n \geq 0 \\ \min_{j\in\{1,\ldots,l\}} k_j < m}} \frac{i!}{k_1!\cdots k_n!},$$

a střední hodnotu $ES_{l,n}^m$ můžeme spočítat jako

$$ES_{l,n}^m = \sum_{i=0}^{\infty} p_i = \sum_{i=0}^{\infty} \frac{N_i}{n^i} = \sum_{i=0}^{\infty} \frac{1}{n^i} \sum_{\substack{k_1+\cdots+k_n=i \\ k_1,\ldots,k_n \geq 0 \\ \min_{j\in\{1,\ldots,l\}} k_j < m}} \frac{i!}{k_1!\cdots k_n!} =$$
$$= \sum_{i=0}^{\infty} \frac{1}{n^i} \sum_{\substack{k_1+\cdots+k_n=i \\ k_1,\ldots,k_n \geq 0 \\ \min_{j\in\{1,\ldots,l\}} k_j < m}} \binom{i}{k_1,\ldots,k_n}.$$

Položme

$$F = \mathrm{e}^{x_1+\cdots+x_n} - [\mathrm{e}^{x_1} - \delta_m(x_1)] \cdots [\mathrm{e}^{x_l} - \delta_m(x_l)] \, \mathrm{e}^{x_{l+1}+\cdots+x_n},$$

kde $\delta_m(t)$ je definováno ve znění věty. Pak F můžeme ekvivalentně vyjádřit jako

$$\begin{aligned}
F &= \mathrm{e}^{x_1} \cdots \mathrm{e}^{x_n} - \sum_{k_1=m}^{\infty} \frac{x_1^{k_1}}{k_1!} \cdots \sum_{k_l=m}^{\infty} \frac{x_l^{k_l}}{k_l!} \, \mathrm{e}^{x_{l+1}} \cdots \mathrm{e}^{x_n} = \\
&= \sum_{k_1=0}^{\infty} \frac{x_1^{k_1}}{k_1!} \cdots \sum_{k_n=0}^{\infty} \frac{x_n^{k_n}}{k_n!} - \sum_{k_1=m}^{\infty} \frac{x_1^{k_1}}{k_1!} \cdots \sum_{k_l=m}^{\infty} \frac{x_l^{k_l}}{k_l!} \sum_{k_{l+1}=0}^{\infty} \frac{x_{l+1}^{k_{l+1}}}{k_{l+1}!} \cdots \sum_{k_n=0}^{\infty} \frac{x_n^{k_n}}{k_n!} = \\
&= \sum_{k_1=0}^{\infty} \cdots \sum_{k_n=0}^{\infty} \frac{x_1^{k_1} \cdots x_n^{k_n}}{k_1! \cdots k_n!} - \sum_{k_1=m}^{\infty} \cdots \sum_{k_l=m}^{\infty} \sum_{k_{l+1}=0}^{\infty} \cdots \sum_{k_n=0}^{\infty} \frac{x_1^{k_1} \cdots x_n^{k_n}}{k_1! \cdots k_n!} = \\
&= \sum_{\substack{k_1,\ldots,k_n \geq 0 \\ \min_{j \in \{1,\ldots,l\}} k_j < m}} \frac{x_1^{k_1} \cdots x_n^{k_n}}{k_1! \cdots k_n!} = \sum_{k=0}^{\infty} \frac{1}{k!} \sum_{\substack{k_1+\cdots+k_n=k \\ k_1,\ldots,k_n \geq 0 \\ \min_{j \in \{1,\ldots,l\}} k_j < m}} \frac{k!}{k_1! \cdots k_n!} x_1^{k_1} \cdots x_n^{k_n}.
\end{aligned}$$

Vidíme, že pokud ve výrazu F dosadíme $(1, \ldots, 1)$ za $(x_1, \ldots, x_n)$, pak se od $ES_{l,n}^m$ liší jen jmenovatelem.

Analogicky jako jsme zjednodušili F, upravíme výraz

$$\begin{aligned}
& n \int_0^{\infty} \left[\mathrm{e}^{t(x_1+\cdots+x_n)} - (\mathrm{e}^{tx_1} - \delta_m(tx_1)) \cdots (\mathrm{e}^{tx_l} - \delta_m(tx_l)) \, \mathrm{e}^{tx_{l+1}} \cdots \mathrm{e}^{tx_n}\right] \mathrm{e}^{-nt} \, \mathrm{d}t = \\
&= n \int_0^{\infty} \left[\mathrm{e}^{tx_1} \cdots \mathrm{e}^{tx_n} - \left(\sum_{k=m}^{\infty} \frac{(tx_1)^k}{k!} \cdots \sum_{k=m}^{\infty} \frac{(tx_l)^k}{k!}\right) \mathrm{e}^{tx_{l+1}} \cdots \mathrm{e}^{tx_n}\right] \mathrm{e}^{-nt} \, \mathrm{d}t = \\
&= n \int_0^{\infty} \left[\sum_{k=0}^{\infty} \sum_{\substack{k_1+\cdots+k_n=k \\ k_1,\ldots,k_n \geq 0 \\ \min_{j \in \{1,\ldots,l\}} k_j < m}} \frac{(tx_1)^{k_1} \cdots (tx_n)^{k_n}}{k_1! \cdots k_n!}\right] \mathrm{e}^{-nt} \, \mathrm{d}t = \\
&= \sum_{k=0}^{\infty} \sum_{\substack{k_1+\cdots+k_n=k \\ k_1,\ldots,k_n \geq 0 \\ \min_{j \in \{1,\ldots,l\}} k_j < m}} \frac{x_1^{k_1} \cdots x_n^{k_n}}{k_1! \cdots k_n!} k! \, n \int_0^{\infty} \frac{t^k}{k!} \mathrm{e}^{-nt} \, \mathrm{d}t.
\end{aligned}$$

Protože platí vztah

$$n \int_0^{\infty} \frac{t^i}{i!} \mathrm{e}^{-nt} \, \mathrm{d}t = \frac{1}{n^i}, \quad i = 0, 1, 2, \ldots,$$

můžeme jej ještě upravit

$$\begin{aligned}
& \sum_{k=0}^{\infty} \sum_{\substack{k_1+\cdots+k_n=k \\ k_1,\ldots,k_n \geq 0 \\ \min_{j \in \{1,\ldots,l\}} k_j < m}} \frac{x_1^{k_1} \cdots x_n^{k_n}}{k_1! \cdots k_n!} k! \, n \int_0^{\infty} \frac{t^k}{k!} \mathrm{e}^{-nt} \, \mathrm{d}t = \\
&= \sum_{k=0}^{\infty} \sum_{\substack{k_1+\cdots+k_n=k \\ k_1,\ldots,k_n \geq 0 \\ \min_{j \in \{1,\ldots,l\}} k_j < m}} \frac{x_1^{k_1} \cdots x_n^{k_n}}{k_1! \cdots k_n!} k! \frac{1}{n^k}.
\end{aligned}$$

Dosazením $(1,\dots,1)$ za $(x_1,\dots,x_n)$ dostaneme

$$\sum_{k=0}^{\infty}\frac{1}{n^k}\sum_{\substack{k_1+\dots+k_n=k\\ k_1,\dots,k_n\geq 0\\ \min_{j\in\{1,\dots,l\}}k_j<m}}\frac{k!}{k_1!\cdots k_n!}=\sum_{k=0}^{\infty}\frac{1}{n^k}\sum_{\substack{k_1+\dots+k_n=k\\ k_1,\dots,k_n\geq 0\\ \min_{j\in\{1,\dots,l\}}k_j<m}}\binom{k}{k_1,\dots k_n},$$

což je $ES_{l,n}^m$. Neboli

$$\begin{aligned}
ES_{l,n}^m &= n\int_0^\infty \left[\mathrm{e}^{(1+\dots+1)t}-(\mathrm{e}^t-\delta_m(t))^l\,\mathrm{e}^{(n-l)t}\right]\mathrm{e}^{-nt}\,\mathrm{d}t=\\
&= n\int_0^\infty \left[\mathrm{e}^{nt}-\mathrm{e}^{lt}\left(1-\mathrm{e}^{-t}\,\delta_m(t)\right)^l\mathrm{e}^{(n-l)t}\right]\mathrm{e}^{-nt}\,\mathrm{d}t\\
&= n\int_0^\infty \left[1-\left(1-\mathrm{e}^{-t}\,\delta_m(t)\right)^l\right]\mathrm{d}t.
\end{aligned}$$

Speciálně pro $l=n$ máme

$$\begin{aligned}
ES_n^m &= n\int_0^\infty \left[\mathrm{e}^{(1+\dots+1)t}-(\mathrm{e}^t-\delta_m(t))^n\right]\mathrm{e}^{-nt}\,\mathrm{d}t=\\
&= n\int_0^\infty \left[\mathrm{e}^{nt}-\mathrm{e}^{nt}(1-\delta_m(t)\,\mathrm{e}^{-t})^n\right]\mathrm{e}^{-nt}\,\mathrm{d}t=\\
&= n\int_0^\infty \left[1-(1-\delta_m(t)\,\mathrm{e}^{-t})^n\right]\mathrm{d}t.
\end{aligned}$$

□

Důkaz jsme provedli podle článku [14], kde je dokázán pro speciální případ $l=n$.

Pro případ $m=1$ je $S_n^1=S_n$ a věta 1.6.1 dává stejný výsledek jako v podkapitole 1.1:

$$\begin{aligned}
ES_n^1 &= n\int_0^\infty \left[1-(1-\delta_1(t)\,\mathrm{e}^{-t})^n\right]\mathrm{d}t = n\int_0^\infty \left[1-(1-\mathrm{e}^{-t})^n\right]\mathrm{d}t=\\
&= n\int_0^\infty (1-1+\mathrm{e}^{-t})\sum_{k=0}^{n-1}(1-\mathrm{e}^{-t})^k\,\mathrm{d}t = n\sum_{k=0}^{n-1}\int_0^\infty \mathrm{e}^{-t}(1-\mathrm{e}^{-t})^k\,\mathrm{d}t=\\
&= n\sum_{k=0}^{n-1}\int_0^1 u^k\,\mathrm{d}u = n\sum_{k=0}^{n-1}\frac{1}{k+1}=n\left(1+\frac{1}{2}+\dots+\frac{1}{n}\right)=ES_n.
\end{aligned}$$

Dívejme se na ES_n^m pro pevné m jako na funkci nejprve v proměnné n, pak pro pevné n bude ES_n^m funkcí proměnné m. Hodnoty této funkce pro malá m a malá n můžeme numericky spočíst. Výsledky zaznamenáme do tabulky 1.1.

Věta 1.6.1 dává vyjádření střední hodnoty ES_n^m pro pevná $m\in\mathbb{N}$ a $n\in\mathbb{N}$. Asymptotické chování je studováno v [14], kde je dokázáno, že pro m pevné a $n\to\infty$ platí

$$ES_n^m = n\left[\ln n+(m-1)\ln\ln n + C_m + o(1)\right].$$

Přitom C_m má podle [4] tvar

$$C_m=\gamma-\ln(m-1)!.$$

Numerické hodnoty ES_n^m					
$m\backslash n$	5	10	15	20	25
1	11,417	29,290	49,773	71,955	95,399
2	19,041	46,230	76,480	108,697	142,368
3	25,987	61,366	100,149	141,104	183,658
4	32,603	75,633	122,354	171,420	222,209
5	39,015	89,362	143,653	200,442	259,062

Tabulka 1.1: Numericky vyjádřené ES_n^m pro malé m a malé n

V [4] je kromě limitního chování střední hodnoty navíc zkoumáno limitní chování celého rozdělení.

Pro každé $x \in \mathbb{R}$ platí

$$\lim_{n\to\infty} P\left(\frac{S_n^m}{n} < \ln n + (m-1)\ln\ln n + x\right) = \exp\left(-\frac{\mathrm{e}^{-x}}{(m-1)!}\right).$$

Jedná se o zobecnění věty 1.4.1. Položíme-li $m = 1$, je $S_n^1 = S_n$, pak pro každé $x \in \mathbb{R}$ platí

$$\lim_{n\to\infty} P\left(\frac{S_n}{n} < \ln n + x\right) = \mathrm{e}^{-\mathrm{e}^{-x}},$$

tedy stejně jako v podkapitole 1.4.

Jiná metoda k řešení úloh souvisejících s problémem sběratele je založena na vnoření do Poissonova procesu a v článku [7] ji Lars Holst využívá k důkazu vztahů z této podkapitoly. Vnoření do Poissonova procesu bude objasněno v podkapitole 2.1.

2. Nerovnoměrně náhodně rozdělené kartičky

2.1 Poissonizace

Nyní budeme náhodně nakupovat výrobky s kartičkami různých typů a v nákupech nepolevíme, ani když nasbíráme úplnou sadu n kartiček. Nákupy kartiček uvažujeme v nepravidelných časových okamžicích a tyto okamžiky budeme považovat za okamžiky Poissonova procesu, a tak nákupy kartiček vnoříme do Poissonova procesu dle [7]. Výhodou této metody je, že nemusíme předpokládat pouze rovnoměrné rozdělení, ale i-tou kartičku uvažujeme s pravděpodobností p_i, $p_i > 0$, kde $\sum_{i=1}^{n} p_i = 1$.

Mějme Poissonův proces $\mathcal{P}$ s intenzitou 1. Dle strany 101 v [15] víme, že veličiny $Z_1, Z_2, \ldots$ označující doby mezi nákupy kartiček jsou nezávislé náhodné veličiny, které mají exponenciální rozdělení se stejným parametrem jako je intenzita Poissonova procesu, tedy s parametrem 1.

Uvažujme Poissonův proces nákupů i-tých typů kartiček, označme jej $\mathcal{P}_i$ pro $i = 1, \ldots, n$. Pak jeho intenzita je p_i. Námi sledovaný proces $\mathcal{P}$ nákupu kartiček je pak složením n nezávislých Poissonových procesů nákupů jednotlivých typů kartiček.

Ptáme se, kolik kartiček musíme nakoupit, abychom získali m sbírek všech typů kartiček, kde $m \in \mathbb{N}$ je pevně dáno. Pro každé $i = 1, \ldots, n$, zavedeme náhodnou veličinu T_i označující čas, ve kterém nastane událost nákupu m-té kartičky i-tého typu do sbírky. Pak T_i má *Erlangovo rozdělení* s parametry m a p_i, které má dle strany 55 v [5] hustotu

$$f_{T_i}(t) = \frac{p_i^m t^{m-1}}{(m-1)!} \mathrm{e}^{-p_i t}, \quad t \geq 0,$$

distribuční funkci rovnou

$$F_{T_i}(t) = 1 - \mathrm{e}^{-p_i t} \sum_{j=0}^{m-1} \frac{(p_i t)^j}{j!}, \quad t \geq 0$$

a první dva momenty rovny

$$\begin{aligned} ET_i &= \frac{m}{p_i}, \\ ET_i^2 &= \frac{m(m+1)}{p_i^2}. \end{aligned}$$

Připomeňme, že S_n^m označuje počet nákupů potřebných k získání m celých sbírek kartiček. Označíme-li

$$T_{n:n} = \max(T_1, \ldots, T_n),$$

pak $T_{n:n}$ je doba, kdy jsme obdrželi m sbírek, a můžeme ji vyjádřit jako součet S_n^m dob mezi nákupy

$$T_{n:n} = \sum_{j=1}^{S_n^m} Z_j.$$

Pro nezáporné náhodné veličiny platí:

$$T_{n:n} = \max(T_1, \dots, T_n) \leq T_1 + \dots + T_n, \tag{2.1}$$

z čehož pro střední hodnotu plyne

$$ET_{n:n} \leq E(T_1 + \dots + T_n) = ET_1 + \dots + ET_n = \frac{m}{p_1} + \dots + \frac{m}{p_n} = \sum_{i=1}^{n} \frac{m}{p_i}.$$

Odhadli jsme ji konečným součtem konečných sčítanců, a proto je $ET_{n:n}$ konečná. Pro druhý moment $T_{n:n}$ z odhadu (2.1) a z nezávislosti $T_1, \dots, T_n$ máme:

$$\begin{aligned} ET_{n:n}^2 \leq E(T_1 + \dots + T_n)^2 &= \sum_{i=1}^{n} ET_i^2 + \sum_{i \neq j} ET_iT_j = \sum_{i=1}^{n} ET_i^2 + \sum_{i \neq j} ET_iET_j = \\ &= \sum_{i=1}^{n} \frac{m(m+1)}{p_i^2} + \sum_{i \neq j} \frac{m^2}{p_ip_j}. \end{aligned}$$

Ten jsme odhadli konečným součtem konečných sčítanců, tedy i druhý moment $T_{n:n}$ je konečný.

Protože S_n^m a $Z_1, Z_2, \dots$ jsou nezávislé, můžeme střední hodnotu $T_{n:n}$ spočíst pomocí podmíněné střední hodnoty a tvrzení 7.21 v [11] následovně:

$$ET_{n:n} = E\left[E\left[T_{n:n}|S_n^m\right]\right] = E\left[E\left[\sum_{j=1}^{S_n^m} Z_j|S_n^m\right]\right] = E\left[S_n^m \cdot EZ_1\right] = ES_n^m, \tag{2.2}$$

neboť $EZ_1 = 1$.

Ověřili jsme, že $ET_{n:n}^2$ je konečný, tudíž dle věty 3.26 v [3] platí:

$$\operatorname{var} T_{n:n} = \operatorname{var}[E(T_{n:n}|S_n^m)] + E[\operatorname{var}(T_{n:n}|S_n^m)].$$

Podle tvrzení 7.21 v [11] víme, že platí vztahy

$$\begin{aligned} E[T_{n:n}|S_n^m] &= S_n^m EZ_1, \\ \operatorname{var}[T_{n:n}|S_n^m] &= S_n^m \operatorname{var} Z_1. \end{aligned}$$

Tedy máme

$$\operatorname{var} T_{n:n} = \operatorname{var}[S_n^m EZ_1] + E[S_n^m \operatorname{var} Z_1] = (EZ_1)^2 \operatorname{var} S_n^m + \operatorname{var} Z_1 ES_n^m.$$

Protože Z_1 má exponenciální rozdělení s parametrem 1, víme, že platí $EZ_1 = 1$ a $\operatorname{var} Z_1 = 1$, což dává rovnost

$$\operatorname{var} T_{n:n} = \operatorname{var} S_n^m + ES_n^m,$$

a protože $ES_n^m = ET_{n:n}$, viz vztah (2.2), dostáváme

$$\operatorname{var} T_{n:n} = \operatorname{var} S_n^m + ET_{n:n}. \tag{2.3}$$

Z těchto vztahů můžeme formulovat zobecnění věty 1.6.1, ve které byly pravděpodobnosti rovnoměrné, tj. platilo $p_1 = \dots = p_n = \frac{1}{n}$.

Věta 2.1.1. *První dva momenty S_n^m jsou konečné a platí:*

$$ES_n^m = \int_0^\infty \left[1 - \prod_{k=1}^{n} \left(1 - \mathrm{e}^{-p_k t} \sum_{j=0}^{m-1} \frac{(p_k t)^j}{j!}\right)\right] \mathrm{d}t$$

a

$$\operatorname{var} S_n^m = 2 \int_0^\infty t \left[1 - \prod_{k=1}^{n} \left(1 - \mathrm{e}^{-p_k t} \sum_{j=0}^{m-1} \frac{(p_k t)^j}{j!}\right)\right] \mathrm{d}t - ES_n^m(ES_n^m + 1).$$

Důkaz. Konečnost momentů S_n^m plyne z konečnosti momentů $T_{n:n}$ a předchozích vztahů (2.2) a (2.3).

Jelikož $T_1, \ldots, T_n$ jsou nezávislé s Erlangovým rozdělením ([5], strana 55), pro distribuční funkci $T_{n:n}$ platí:

$$\begin{aligned}
F_{T_{n:n}}(t) &= P(T_{n:n} \leq t) = P(\max_{j \in \{1,\ldots,n\}} T_j \leq t) = P(T_1 \leq t, \ldots, T_n \leq t) = \\
&= P(T_1 \leq t) \cdots P(T_n \leq t) = \\
&= \left(1 - \mathrm{e}^{-p_1 t} \sum_{j=0}^{m-1} \frac{(p_1 t)^j}{j!}\right) \cdots \left(1 - \mathrm{e}^{-p_n t} \sum_{j=0}^{m-1} \frac{(p_n t)^j}{j!}\right) = \\
&= \prod_{k=1}^{n} \left(1 - \mathrm{e}^{-p_k t} \sum_{j=0}^{m-1} \frac{(p_k t)^j}{j!}\right).
\end{aligned}$$

Pak protože $T_{n:n}$ je reálná nezáporná náhodná veličina, můžeme dle lemmatu 5.7 v [11] počítat

$$ET_{n:n} = \int_0^\infty [1 - F_{T_{n:n}}(t)]\, \mathrm{d}t = \int_0^\infty \left[1 - \prod_{k=1}^{n} \left(1 - \mathrm{e}^{-p_k t} \sum_{j=0}^{m-1} \frac{(p_k t)^j}{j!}\right)\right] \mathrm{d}t.$$

Celkem máme:

$$ES_n^m = ET_{n:n} = \int_0^\infty \left[1 - \prod_{k=1}^{n} \left(1 - \mathrm{e}^{-p_k t} \sum_{j=0}^{m-1} \frac{(p_k t)^j}{j!}\right)\right] \mathrm{d}t.$$

Podobně nyní vyjádříme rozptyl. Dle lemmatu 3.4 v [9] platí:

$$ET_{n:n}^2 = 2 \int_0^\infty t[1 - F_{T_{n:n}}(t)]\, \mathrm{d}t,$$

pak pro rozptyl S_n^m ze vztahu (2.3) máme:

$$\begin{aligned}
\operatorname{var} S_n^m &= \operatorname{var} T_{n:n} - ET_{n:n} = ET_{n:n}^2 - (ET_{n:n})^2 - ET_{n:n} = \\
&= 2 \int_0^\infty t \left[1 - \prod_{k=1}^{n} \left(1 - \mathrm{e}^{-p_k t} \sum_{j=0}^{m-1} \frac{(p_k t)^j}{j!}\right)\right] \mathrm{d}t - (ES_n^m)^2 - ES_n^m = \\
&= 2 \int_0^\infty t \left[1 - \prod_{k=1}^{n} \left(1 - \mathrm{e}^{-p_k t} \sum_{j=0}^{m-1} \frac{(p_k t)^j}{j!}\right)\right] \mathrm{d}t - ES_n^m(ES_n^m + 1).
\end{aligned}$$

□

2.2 Dolní a horní aproximace středního počtu chybějících kartiček

Připomeňme značení z podkapitoly 1.5, že ve sbírce, která již obsahuje alespoň jednu kartičku každého druhu, je A_i^m jev, kdy jsme nasbírali alespoň m kartiček od i-tého typu kartičky. Nechť m je pevné a nechť T_i je doba nákupu m-té kartičky i-tého typu. Doplněk jevu A_i^m znamená, že ve chvíli, kdy jsme právě nasbírali celou jednu sadu, máme méně než m kartiček i-tého typu. Neboli v čase $T_i = t$ už musíme mít alespoň jednu kartičku od každého typu různého od i, což má pravděpodobnost

$$\prod_{k \neq i} \left(1 - \mathrm{e}^{-p_k t}\right).$$

Proto pro $m \geq 2$ je

$$\begin{aligned} 1 - P(A_i^m) &= \int_0^\infty f_{T_i}(t) \prod_{k \neq i} (1 - \mathrm{e}^{-p_k t}) \, \mathrm{d}t = \\ &= \int_0^\infty p_i \, \mathrm{e}^{-p_i t} \frac{(p_i t)^{m-1}}{(m-1)!} \prod_{k \neq i} (1 - \mathrm{e}^{-p_k t}) \, \mathrm{d}t. \end{aligned} \tag{2.4}$$

Poznamenejme, že U_n^m bylo definováno v podkapitole 1.5. Pro vyjádření střední hodnoty U_n^m použijeme vztah (1.2):

$$\begin{aligned} EU_n^m &= E\left(n - \sum_{i=1}^n \mathbf{1}(A_i^m)\right) = n - \sum_{i=1}^n E\mathbf{1}(A_i^m) = n - \sum_{i=1}^n P(A_i^m) = \\ &= \sum_{i=1}^n [1 - P(A_n^m)] = \sum_{i=1}^n \left[\int_0^\infty p_i \, \mathrm{e}^{-p_i t} \frac{(p_i t)^{m-1}}{(m-1)!} \prod_{k \neq i} (1 - \mathrm{e}^{-p_k t}) \, \mathrm{d}t \right]. \end{aligned}$$

V této podkapitole budeme chtít najít vhodnou horní a dolní mez pro EU_n^m. Postupujme podle článku [1]. Nejprve definujme Schurovu konkávní funkci, neboť její vlastnosti se nám hodí pro nalezení horní meze pro EU_n^m.

Definice. *Nechť $x, y \in \mathbb{R}^n$. Uspořádáme-li složky vektorů x a y sestupně, tj. $x_{[1]} \geq \dots \geq x_{[n]}$ pro vektor x a $y_{[1]} \geq \dots \geq y_{[n]}$ pro vektor y, pak řekneme, že x a y jsou v relaci $x \prec y$, jestliže splňují vztahy:*

$$\begin{aligned} \sum_{j=1}^k x_{[j]} &\leq \sum_{j=1}^k y_{[j]} \quad \text{pro } k = 1, \dots, n-1, \\ \sum_{j=1}^n x_{[j]} &= \sum_{j=1}^n y_{[j]}. \end{aligned}$$

Reálnou funkci ϕ definovanou na množině $\mathbb{A} \subset \mathbb{R}^n$ nazveme Schurovou konkávní funkcí na množině $\mathbb{A}$, jestliže platí implikace

$$x \prec y \text{ na } \mathbb{A} \Rightarrow \phi(x) \geq \phi(y). \tag{2.5}$$

Nyní připomeňme vlastnost symetrie: Aby ϕ byla symetrická, musí její definiční obor být symetrický a pro každou permutaci $\Pi : \{1, \dots, n\} \to \{1, \dots, n\}$ musí platit rovnost

$$\phi(x_1, \dots, x_n) = \phi(x_{\Pi(1)}, \dots, x_{\Pi(n)}).$$

Poznámka. *Nechť $I \subset \mathbb{R}$. Jestliže ϕ je Schurova konkávní funkce na I^n, pak je symetrickou funkcí na I^n.*

Důkaz. Permutování na množině I^n, kde $I \subset \mathbb{R}$, nemění relaci $\prec$, neboť $x \prec \Pi x$ a zároveň $\Pi x \prec x$ pro každou permutační matici Π. Uspořádání $x \prec \Pi x$ implikuje $\phi(x) \geq \phi(\Pi x)$ a uspořádání $\Pi x \prec x$ implikuje $\phi(\Pi x) \geq \phi(x)$. Celkem tedy máme, že $\phi(x) = \phi(\Pi x)$ pro všechny permutační matice Π.

□

Budeme potřebovat některé vlastnosti Schurovy konkávní funkce, konkrétně větu 3.A.4 z [12].

Tvrzení 2.2.1. *Nechť $I \subset \mathbb{R}$ je otevřený interval a nechť $\phi : I^n \to \mathbb{R}$ je spojitě diferencovatelná funkce. Pak ϕ je Schurova konkávní funkce na n-rozměrném intervalu I^n právě tehdy, když ϕ je symetrická funkce na I^n a zároveň $\phi_{(i)}(x) = \frac{\partial \phi(x)}{\partial x_i}$ je neklesající v $i = 1, \dots, n$ pro všechna $x \in I^n$ splňující $x_1 \geq \dots \geq x_n$.*

Ekvivalentně: ϕ je Schurova konkávní funkce na I^n právě tehdy, když ϕ je symetrická funkce na I^n a zároveň pro všechna $x \in I^n$ platí:

$$(x_1 - x_2)\left[\phi_{(1)}(x) - \phi_{(2)}(x)\right] \leq 0. \tag{2.6}$$

Lemma 2.2.2. *Funkce $\phi(y) = \prod_{j=1}^{r}(1 - \mathrm{e}^{-\mathrm{e}^{y_j}})$ je Schurova konkávní funkce proměnné $y = (y_1, \dots, y_r)$.*

Důkaz. Funkce $\phi(y)$ je definována pro každé $y \in \mathbb{R}^r$ a prohození pořadí činitelů v součinu součin neovlivní, tedy funkce ϕ je symetrickou funkcí.

Nyní si připravíme vztahy pro následné ověření Schurovy podmínky (2.6). Zderivujeme-li funkci $1 - \mathrm{e}^{-\mathrm{e}^u}$ podle u, dostaneme

$$\frac{\partial}{\partial u}(1 - \mathrm{e}^{-\mathrm{e}^u}) = -\mathrm{e}^{-\mathrm{e}^u}(-\mathrm{e}^u) = \mathrm{e}^u\,\mathrm{e}^{-\mathrm{e}^u}, \quad u \in \mathbb{R}.$$

Tedy pro $\phi(y_1, \dots, y_r) = \prod_{j=1}^{r}(1 - \mathrm{e}^{-\mathrm{e}^{y_j}})$ máme

$$\frac{\partial \phi}{\partial y_i}(y) = \frac{\partial}{\partial y_i}\prod_{j=1}^{r}(1 - \mathrm{e}^{-\mathrm{e}^{y_j}}) = \mathrm{e}^{y_i}\,\mathrm{e}^{-\mathrm{e}^{y_i}}\prod_{\substack{k=1\\k\neq i}}^{r}(1 - \mathrm{e}^{-\mathrm{e}^{y_k}}).$$

Zderivujme funkci $f(u) = \frac{\mathrm{e}^u\,\mathrm{e}^{-\mathrm{e}^u}}{1-\mathrm{e}^{-\mathrm{e}^u}}$ podle u, dostaneme

$$\frac{\partial}{\partial u}f(u) = \frac{\partial}{\partial u}\frac{\mathrm{e}^u\,\mathrm{e}^{-\mathrm{e}^u}}{1 - \mathrm{e}^{-\mathrm{e}^u}} = \mathrm{e}^u\,\frac{\mathrm{e}^{-\mathrm{e}^u}(1 - \mathrm{e}^{-\mathrm{e}^u}) - \mathrm{e}^u\,\mathrm{e}^{-\mathrm{e}^u}(1 - \mathrm{e}^{-\mathrm{e}^u}) - \mathrm{e}^u\,\mathrm{e}^{-2\,\mathrm{e}^u}}{(1 - \mathrm{e}^{-\mathrm{e}^u})^2} =$$

$$= \mathrm{e}^u\,\mathrm{e}^{-\mathrm{e}^u}\,\frac{1 - \mathrm{e}^{-\mathrm{e}^u} - \mathrm{e}^u}{(1 - \mathrm{e}^{-\mathrm{e}^u})^2}.$$

Činitelé e^{-e^u}, e^u a $\frac{1}{(1-e^{-e^u})^2}$ jsou zřejmě kladní a protože pro každé reálné $x > 0$ platí nerovnost $1 - e^{-x} < x$, pro $x = e^u$ platí $1 - e^{-e^u} < e^u$, a tedy funkce $1 - e^{-e^u} - e^u$ je záporná. Znaménko derivace f je záporné, tudíž funkce $\frac{e^u e^{-e^u}}{1-e^{-e^u}}$ je klesající. Tedy pro $y_1 < y_2$ platí

$$\frac{e^{y_1} e^{-e^{y_1}}}{1 - e^{-e^{y_1}}} > \frac{e^{y_2} e^{-e^{y_2}}}{1 - e^{-e^{y_2}}}.$$

Po úpravě dostaneme

$$e^{y_1} e^{-e^{y_1}}(1 - e^{-e^{y_2}}) > e^{y_2} e^{-e^{y_2}}(1 - e^{-e^{y_1}})$$

a převedením na levou stranu získáme

$$e^{y_1} e^{-e^{y_1}}(1 - e^{-e^{y_2}}) - e^{y_2} e^{-e^{y_2}}(1 - e^{-e^{y_1}}) > 0. \tag{2.7}$$

Nyní ověříme Schurovu podmínku

$$(y_1 - y_2)\left[\phi_{(1)}(y) - \phi_{(2)}(y)\right] \leq 0$$

pro všechna $y \in \mathbb{R}^r$. Zvolme $y \in \mathbb{R}^r$ pevné, pak

$$\begin{aligned}
&(y_1 - y_2)\left[\phi_{(1)}(y) - \phi_{(2)}(y)\right] = \\
&= (y_1 - y_2)\left[e^{y_1} e^{-e^{y_1}} \prod_{k=2}^{r}(1 - e^{-e^{y_k}}) - e^{y_2} e^{-e^{y_2}} \prod_{\substack{k=1\\k\neq 2}}^{r}(1 - e^{-e^{y_k}})\right] = \\
&= (y_1 - y_2)\Bigg[e^{y_1} e^{-e^{y_1}}(1 - e^{-e^{y_2}}) \prod_{k=3}^{r}(1 - e^{-e^{y_k}}) - \\
&\qquad - e^{y_2} e^{-e^{y_2}}(1 - e^{-e^{y_1}}) \prod_{k=3}^{r}(1 - e^{-e^{y_k}})\Bigg] = \\
&= (y_1 - y_2)\left[e^{y_1} e^{-e^{y_1}}(1 - e^{-e^{y_2}}) - e^{y_2} e^{-e^{y_2}}(1 - e^{-e^{y_1}})\right] \prod_{k=3}^{r}(1 - e^{-e^{y_k}}),
\end{aligned} \tag{2.8}$$

neboť

$$\phi_{(1)}(y) = \frac{\partial \phi}{\partial y_1}(y) = e^{y_1} e^{-e^{y_1}} \prod_{k=2}^{r}(1 - e^{-e^{y_k}})$$

a

$$\phi_{(2)}(y) = \frac{\partial \phi}{\partial y_2}(y) = e^{y_2} e^{-e^{y_2}} \prod_{\substack{k=1\\k\neq 2}}^{r}(1 - e^{-e^{y_k}}).$$

Pro každé reálné x platí $0 > -e^x$, tedy i vztah $e^0 > e^{-e^x}$, což převedením na jednu stranu dává $1 - e^{-e^x} > 0$. Součin kladných funkcí je kladná funkce, tudíž $\prod_{k=3}^{r}(1 - e^{-e^{y_k}})$ je kladná funkce. Pro $y_1 < y_2$ je rozdíl $y_1 - y_2 < 0$ a pro funkci $e^{y_1} e^{-e^{y_1}}(1 - e^{-e^{y_2}}) - e^{y_2} e^{-e^{y_2}}(1 - e^{-e^{y_1}})$ máme z (2.7), že je kladná pro $y_1 < y_2$. Jelikož (2.8) zachovává znaménko při prohození proměnných y_1 a y_2, dostáváme pro každé $y_1, y_2 \in \mathbb{R}, y_1 \neq y_2$:

$$(y_1 - y_2)\left[\phi_{(1)}(y) - \phi_{(2)}(y)\right] < 0.$$

Tedy Schurova podmínka je splněna.

□

Uvažujme situaci, kdy pravděpodobnosti nákupu jednotlivých typů kartiček jsou skoro rovnoměrně rozdělené, tj. $(p_1, \ldots, p_n)$ jsou blízko $(\frac{1}{n}, \ldots, \frac{1}{n})$. Nyní se budeme snažit aproximovat EU_n^m, neboť výpočet k tomu potřebného integrálu (2.4) je poněkud složitý.

Věta 2.2.3. *Označme*

$$v(a_1, \ldots, a_n) = \sum_{r=0}^{n-1} \binom{n-1}{r} (-1)^r \sum_{i=1}^{n} \left(\frac{p_i}{p_i + ra_i} \right)^m ,$$

pak pro každé $n \in \mathbb{N}$ a každé $m \geq 2$ platí:

$$v(q_1, \ldots, q_n) \leq EU_n^m \leq v(g_1, \ldots, g_n),$$

kde $q_i = \min_{k \neq i} p_k$ a $g_i = \left(\prod_{k \neq i} p_k \right)^{\frac{1}{n-1}}, \quad i = 1, \ldots, n.$

Důkaz. Je-li $q_i = \min_{k \neq i} p_k$, pak $q_i \leq p_i$ pro každé $i = 1, \ldots, n$, a tedy součin $\prod_{k \neq i} (1 - \mathrm{e}^{-p_k t})$ můžeme odhadnout zdola výrazem $(1 - \mathrm{e}^{-q_i t})^{n-1}$, neboť počet činitelů v součinu je $n - 1$. Pro vektory $x = (\ln g_i t, \ldots, \ln g_i t) \in \mathbb{R}^{n-1}$ a $y = (\ln p_1 t, \ldots, \ln p_{i-1} t, \ln p_{i+1} t, \ldots, \ln p_n t) \in \mathbb{R}^{n-1}$ platí

$$\sum_{j=1}^{n-1} x_{[j]} = \sum_{j=1}^{n-1} \ln g_j t = (n-1) \ln g_j t = (n-1) \ln \left[\left(\prod_{k \neq i} p_k \right)^{\frac{1}{n-1}} t \right] =$$

$$= (n-1) \ln \left(\prod_{k \neq i} p_k \right)^{\frac{1}{n-1}} + (n-1) \ln t = \ln \left(\prod_{k \neq i} p_k \right) + \ln t^{n-1} =$$

$$= \sum_{k \neq i} \ln p_k + \ln t^{n-1} = \sum_{k \neq i} \ln p_k t = \sum_{j=1}^{n-1} y_{[j]},$$

a tedy splňují druhou podmínku definice pro relaci $x \prec y$, Nerovnost součtu složek ukážeme pro $r < n$ následovně: Jelikož platí rovnost

$$\ln g_i = \ln \left(\prod_{k \neq i} p_k \right)^{\frac{1}{n-1}} = \frac{1}{n-1} \sum_{k \neq i} \ln p_k,$$

pak pro $p_1 \geq \cdots \geq p_{n-1}$ je

$$\ln p_1 \geq \frac{1}{n-1} \sum_{k \neq i} \ln p_k,$$

$$\ln p_1 + \ln p_2 \geq \frac{2}{n-1} \sum_{k \neq i} \ln p_k,$$

obecně pro $r < n$ dostáváme nerovnost

$$\ln p_1 + \cdots + \ln p_r \geq \frac{r}{n-1} \sum_{k \neq i} \ln p_k,$$

která dává
$$\sum_{j=1}^{r} y_{[j]} \geq \sum_{j=1}^{r} x_{[j]}.$$
Dohromady máme $x \prec y$. Funkce $\phi(y) = \prod_{j=1}^{r}(1 - \mathrm{e}^{-\mathrm{e}^{y_j}})$ je pro $y = (\ln p_1 t, \ldots, \ln p_{i-1} t, \ln p_{i+1} t, \ldots, \ln p_n t)$ rovna
$$\phi(y) = \prod_{k \neq i}(1 - \mathrm{e}^{-\mathrm{e}^{\ln p_k t}}) = \prod_{k \neq i}(1 - \mathrm{e}^{-p_k t})$$
a pro $x = (\ln g_i t, \ldots, \ln g_i t)$ platí
$$\phi(x) = \prod_{k=1}^{n-1}(1 - \mathrm{e}^{-\mathrm{e}^{\ln g_i t}}) = \prod_{k=1}^{n-1}(1 - \mathrm{e}^{-g_i t}) = (1 - \mathrm{e}^{-g_i t})^{n-1}.$$
Použitím lemmatu 2.2.2 a vztahu 2.5 pro $x \prec y$ máme:
$$\phi(y) = \prod_{k \neq i}(1 - \mathrm{e}^{-p_k t}) \leq (1 - \mathrm{e}^{-g_i t})^{n-1} = \phi(x).$$
Celkem jsme dostali nerovnosti:
$$(1 - \mathrm{e}^{-q_i t})^{n-1} \leq \prod_{k \neq i}(1 - \mathrm{e}^{-p_k t}) \leq (1 - \mathrm{e}^{-g_i t})^{n-1}.$$

Použitím horní meze součinu, dostaneme pro $1 - P(A_i^m)$ nerovnost
$$\begin{aligned}
1 - P(A_i^m) &= \int_0^\infty p_i \,\mathrm{e}^{-p_i t} \frac{(p_i t)^{m-1}}{(m-1)!} \prod_{k \neq i}(1 - \mathrm{e}^{-p_k t}) \,\mathrm{d}t \leq \\
&\leq \int_0^\infty p_i \,\mathrm{e}^{-p_i t} \frac{(p_i t)^{m-1}}{(m-1)!} \left(1 - \mathrm{e}^{-g_i t}\right)^{n-1} \,\mathrm{d}t = \\
&= \int_0^\infty p_i \,\mathrm{e}^{-p_i t} \frac{(p_i t)^{m-1}}{(m-1)!} \sum_{r=0}^{n-1} \binom{n-1}{r} \left(-\mathrm{e}^{-g_i t}\right)^r \,\mathrm{d}t = \\
&= \sum_{r=0}^{n-1} \binom{n-1}{r} (-1)^r \int_0^\infty p_i \,\mathrm{e}^{-p_i t - r g_i t} \frac{(p_i t)^{m-1}}{(m-1)!} \,\mathrm{d}t = \\
&= \sum_{r=0}^{n-1} \binom{n-1}{r} (-1)^r p_i^m \int_0^\infty \mathrm{e}^{-(p_i + r g_i)t} \frac{t^{m-1}}{(m-1)!} \,\mathrm{d}t = \\
&= \sum_{r=0}^{n-1} \binom{n-1}{r} (-1)^r \left(\frac{p_i}{p_i + r g_i}\right)^m \int_0^\infty (p_i + r g_i) \,\mathrm{e}^{-(p_i + r g_i)t} \frac{[(p_i + r g_i)t]^{m-1}}{(m-1)!} \,\mathrm{d}t.
\end{aligned}$$
Označíme-li $\lambda = p_i + r g_i$, pak integrál
$$\int_0^\infty (p_i + r g_i) \,\mathrm{e}^{-(p_i + r g_i)t} \frac{[(p_i + r g_i)t]^{m-1}}{(m-1)!} \,\mathrm{d}t = \int_0^\infty \lambda \,\mathrm{e}^{-\lambda t} \frac{(\lambda t)^{m-1}}{(m-1)!} \,\mathrm{d}t,$$
je roven jedné, jelikož integrand je hustota Erlangova rozdělení s parametry m a λ.

Dohromady jsme pro $1 - P(A_i^m)$ dostali následující horní mez

$$1 - P(A_i^m) \leq \sum_{r=0}^{n-1} \binom{n-1}{r} (-1)^r \left(\frac{p_i}{p_i + r g_i} \right)^m .$$

Analogicky upravíme dolní mez pro $1 - P(A_i^m)$, kde označíme $\kappa = p_i + r q_i$,

$$\begin{aligned}
1 - P(A_i^m) &= \int_0^\infty p_i \, \mathrm{e}^{-p_i t} \frac{(p_i t)^{m-1}}{(m-1)!} \prod_{k \neq i} (1 - \mathrm{e}^{-p_k t}) \, \mathrm{d}t \geq \\
&\geq \int_0^\infty p_i \, \mathrm{e}^{-p_i t} \frac{(p_i t)^{m-1}}{(m-1)!} \left(1 - \mathrm{e}^{-q_i t}\right)^{n-1} \, \mathrm{d}t = \\
&= \int_0^\infty p_i \, \mathrm{e}^{-p_i t} \frac{(p_i t)^{m-1}}{(m-1)!} \sum_{r=0}^{n-1} \binom{n-1}{r} (- \mathrm{e}^{-q_i t})^r \, \mathrm{d}t = \\
&= \sum_{r=0}^{n-1} \binom{n-1}{r} (-1)^r \int_0^\infty p_i \, \mathrm{e}^{-(p_i + r q_i)t} \frac{(p_i t)^{m-1}}{(m-1)!} \, \mathrm{d}t = \\
&= \sum_{r=0}^{n-1} \binom{n-1}{r} (-1)^r \left(\frac{p_i}{p_i + r q_i} \right)^m \int_0^\infty \kappa \, \mathrm{e}^{-\kappa t} \frac{(\kappa t)^{m-1}}{(m-1)!} \, \mathrm{d}t = \\
&= \sum_{r=0}^{n-1} \binom{n-1}{r} (-1)^r \left(\frac{p_i}{p_i + r q_i} \right)^m .
\end{aligned}$$

Celkem jsme dostali nerovnosti

$$\sum_{r=0}^{n-1} \binom{n-1}{r} (-1)^r \left(\frac{p_i}{p_i + r q_i} \right)^m \leq 1 - P(A_i^m) \leq \sum_{r=0}^{n-1} \binom{n-1}{r} (-1)^r \left(\frac{p_i}{p_i + r g_i} \right)^m .$$

A protože pro $m \geq 2$ platí rovnost $EU_n^m = \sum_{i=1}^n [1 - P(A_i^m)]$, dostáváme

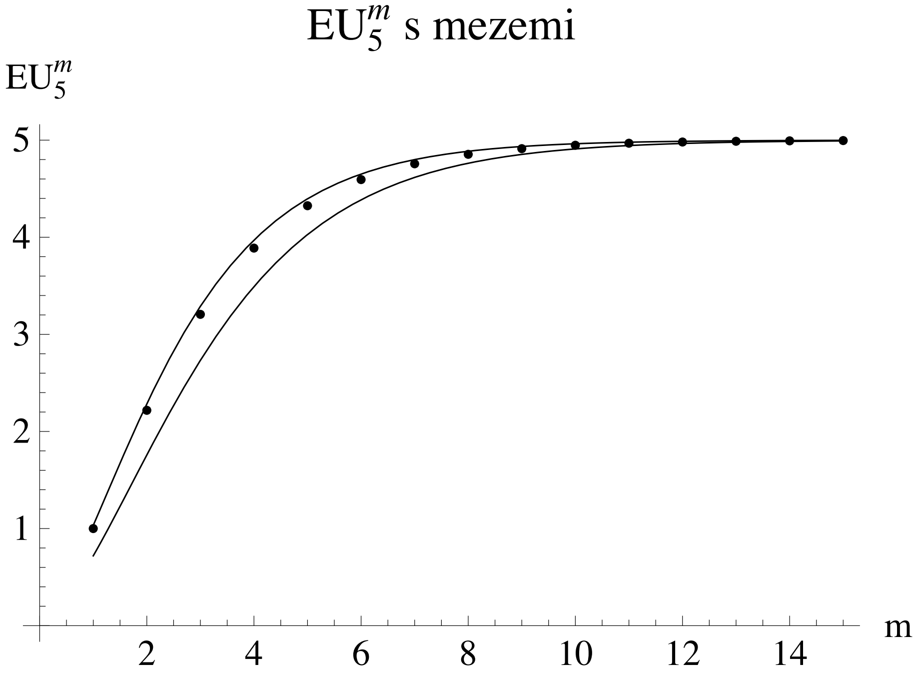
$$\sum_{i=1}^n \sum_{r=0}^{n-1} \binom{n-1}{r} (-1)^r \left(\frac{p_i}{p_i + r q_i} \right)^m \leq EU_n^m \leq \sum_{i=1}^n \sum_{r=0}^{n-1} \binom{n-1}{r} (-1)^r \left(\frac{p_i}{p_i + r g_i} \right)^m ,$$

což můžeme ještě upravit na následující meze

$$v(q_1, \ldots, q_n) \leq EU_n^m \leq v(g_1, \ldots, g_n).$$

$\square$

V obrázku 2.1 znázorníme konkrétní hodnoty EU_n^m pro $n = 5$ a pravděpodobnosti $(p_1, \ldots, p_n)$ rovny 0,2; 0,15; 0,15; 0,25; 0,25 a vykreslíme i horní a dolní mez pro EU_n^m z věty 2.2.3.



Obrázek 2.1: Hodnoty EU_n^m společně s mezemi pro $n = 5$ a $(p_1, \ldots, p_5) = (0{,}2; 0{,}15; 0{,}15; 0{,}25; 0{,}25)$

Závěr

Problém sbírání kartiček za účelem získání jedné nebo více sad kartiček či jen neúplné sady kartiček byl rozebírán v celé práci.

Nejprve jsme se zabývali problémem sběratele kupónů v diskrétním případě, viz kapitola 1. Všechny kartičky nám v celé této kapitole přicházely se stejnou pravděpodobností v diskrétních pravidelných časech a k veškerým výpočtům nám stačily metody klasické pravděpodobnosti. Jednu podkapitolu jsme věnovali limitním výpočtům pro počet kartiček jdoucí k nekonečnu a zjistili jsme, že v takovém případě má celkový počet zakoupených výrobků potřebných k získání celé sbírky kartiček Gumbelovo rozdělení.

V kapitole 2 jsme použili metodu vnoření do Poissonova procesu. Uvažujeme v ní, že nákupy kartiček neprobíhají v diskrétních okamžicích, ale uvažujeme je ve spojitém časovém úseku, ve kterém jsou doby nákupů náhodně rozmístěny. Tato metoda se ukázala býti výhodnější, neboť se dá přirozeně použít i pro obecnější případ získávání kartiček, tj. v případě, kdy obdržení jednotlivých kartiček nemá stejnou pravděpodobnost. Zjistili jsme, že tato metoda převádí klasické úlohy na studium pořádkových statistik.

Kdybychom si přáli další zobecnění problému sběratele kupónů, mohli bychom se zabývat případem, kdy sběratel kartičky dostává nikoliv po jedné kartičce, jak je uvažováno v celé práci, ale po větším počtu kartiček najednou. Mohli bychom také zkoumat optimální strategii, pokud by byla možnost dokoupit si za jistou cenu chybějící kartičky, více viz [16], nebo uvažovat případ, kdy by mohl sběratel do jisté míry ovlivnit, jakou kartičku si vybere.

Literatura

[1] Adler, I.; Oren, S.; Ross, S. M.: The coupon-collector's problem revisited. *Journal of Applied Probability*, ročník 40, č. 2, 2003: s. 513–518.

[2] Anděl, J.: *Matematika náhody*. Praha: MATFYZPRESS, 2003.

[3] Anděl, J.: *Základy matematické statistiky*. Praha: MATFYZPRESS, třetí vydání, 2011.

[4] Erdös, P.; Rényi, A.: On a classical problem of probability theory. *A Magyar Tudományos Akadémia Matematikai Kutató Intézetének Közleményei*, ročník 6, č. 1-2, 1961: s. 215–220.

[5] Evans, M.; Hastings, N.; Peacock, B.: *Statistical Distributions*. New York: Wiley, druhé vydání, 1993.

[6] Hald, A.; de Moivre, A.; McClintock, B.: A. de Moivre: 'De Mensura Sortis' or 'On the Measurement of Chance'. *International Statistical Review / Revue Internationale de Statistique*, ročník 52, č. 3, 1984: s. 229–262.

[7] Holst, L.: On birthday, collectors', occupancy and other classical urn problems. *International Statistical Review*, ročník 54, č. 1, 1986: s. 15–27.

[8] Johnson, N. L.; Kotz, S.; Kemp, A. W.: *Univariate Discrete Distributions*. New York: Wiley, druhé vydání, 1993.

[9] Kallenberg, O.: *Foundations of Modern Probability*. New York: Springer, druhé vydání, 2002.

[10] Kan, N. D.: Martingale approach to the coupon collection problem. *Journal of Mathematical Sciences*, ročník 127, č. 1, 2005: s. 1737–1744.

[11] Lachout, P.: *Teorie pravděpodobnosti*. Praha: Karolinum, 1998.

[12] Marshall, A. W.; Olkin, I.: *Inequalities: Theory of Majorization and Its Applications*. New York: Academic Press, 1979.

[13] Matoušek, J.; Nešetřil, J.: *Kapitoly z diskrétní matematiky*. Praha: Karolinum, čtvrté vydání, 2009.

[14] Newman, D. J.; Shepp, L.: The double dixie cup problem. *The American Mathematical Monthly*, ročník 67, č. 1, 1960: s. 58–61.

[15] Prášková, Z.; Lachout, P.: *Základy náhodných procesů*. Praha: Karolinum, 2001.

[16] Stadje, W.: The collector's problem with group drawings. *Advances in Applied Probability*, ročník 22, č. 4, 1990: s. 866–882.

[17] Wolfram Research, Inc.: *Mathematica, Version 7.0*. Champaign, Illinois: Wolfram Research, Inc., 2008.