



**MATEMATICKO-FYZIKÁLNÍ  
FAKULTA**  
Univerzita Karlova

## **BAKALÁŘSKÁ PRÁCE**

Diana Pavlovičová

# **Geometrické rozdělení a jeho mnohorozměrné rozšíření**

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: doc. RNDr. Daniel Hlubinka, Ph.D.

Studijní program: Obecná matematika

Praha 2022

Prohlašuji, že jsem tuto bakalářskou práci vypracovala samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů. Tato práce nebyla využita k získání jiného nebo stejného titulu.

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 má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 autorského zákona.

V ..... dne .....

Podpis autora

Chtěla bych poděkovat především docentu Danielu Hlubinkovi za cenné rady a skvělou spolupráci. Dále bych chtěla poděkovat své rodině a přátelům za podporu, zejména mému dvojčeti Ladě za to, že mě na celé této dlouhé cestě provázela nejen jako korektorka, ale také jako nekonečný zdroj smíchu a pochopení.

Název práce: Geometrické rozdělení a jeho mnohorozměrné rozšíření

Autor: Diana Pavlovičová

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: doc. RNDr. Daniel Hlubinka, Ph.D., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: V této práci se zabýváme vícerozměrným geometrickým rozdělením, především jeho dvojrozměrnou variantou. Nejprve uvedeme základní definici, ve které uvažujeme dva typy neúspěchů. Dále spočteme některé základní popisné charakteristiky tohoto rozdělení. Poté se zaměříme na jinou verzi dvojrozměrného geometrického rozdělení, kterou odvodíme pomocí podmiňování a u které opět uvedeme některé popisné charakteristiky. Tuto verzi dále rozšíříme na případ, kdy uvažujeme tři typy neúspěchů. Získané výsledky dále přímo zobecníme pro případ vícerozměrného negativně binomického rozdělení. V poslední kapitole se zaměříme na odhady parametrů jednoduchého dvojrozměrného geometrického rozdělení a uvedeme jednoduchou simulaci, na které demonstrujeme kvalitu těchto odhadů.

Klíčová slova: geometrické rozdělení, negativně binomické rozdělení, mnohorozměrné rozdělení

Title: Geometric distribution and its multivariate version

Author: Diana Pavlovičová

Department: Department of Probability and Mathematical Statistics

Supervisor: doc. RNDr. Daniel Hlubinka, Ph.D., Department of Probability and Mathematical Statistics

Abstract: In this work we will discuss the basics of a multivariate geometric distribution, especially its two-dimensional version. First of all, we establish a fundamental definition in which we consider two types of failures. Next, we compute some of its properties. We then focus on a different version of the two-dimensional case which we obtain by conditioning and for which we again compute its properties. We extend this approach to the case where we consider three types of failures. We further generalize the obtained results for the case of a multivariate negative binomial distribution. Lastly, we focus on the estimates of the parameters of the fundamental two-dimensional version of the multivariate geometric distribution and present a simple simulation in which we demonstrate the accuracy of the obtained estimates.

Keywords: geometric distribution, negative binomial distribution, multivariate distribution

# Obsah

Úvod	2
<b>1 Problematika vícerozměrných rozdělení</b>	<b>3</b>
<b>2 Definice a základní vlastnosti</b>	<b>4</b>
2.1 Momenty a vytvořující funkce . . . . .	4
<b>3 Okamžik nastání prvního úspěchu</b>	<b>9</b>
3.1 Součet počtu neúspěchů . . . . .	9
3.1.1 Rozšíření na negativně binomické rozdělení . . . . .	10
3.2 Kombinované typy neúspěchů . . . . .	10
3.2.1 Momenty . . . . .	12
3.2.2 Rozšíření na negativně binomické rozdělení . . . . .	12
<b>4 Odhady parametrů</b>	<b>14</b>
4.1 Odhady metodou maximální věrohodnosti . . . . .	14
4.2 Vlastnosti odhadů . . . . .	15
4.3 Intervaly spolehlivosti . . . . .	16
4.4 Odhady parametrů pro hody kostkou . . . . .	17
<b>Závěr</b>	<b>21</b>
<b>Seznam použité literatury</b>	<b>22</b>

# Úvod

V této práci se budeme podrobně zabývat přístupy k nalezení vícerozměrného geometrického rozdělení, jednotlivými definicemi tohoto rozdělení a jeho základními popisnými charakteristikami, a to ve dvojrozměrném případě. Podkladem pro osnovu práce je článek Davy a Rayner (1996). V daném článku je celá problematika popsána bez výpočtů, formálních definic a hlubších komentářů.

V první kapitole se zaměříme na problematiku vícerozměrných rozdělení, jejich odvození z jiných rozdělení a také jejich dalších zobecnění. Pro lepší pochopení celé problematiky dále uvedeme dva motivační příklady. V neposlední řadě připomeneme některé speciální vlastnosti geometrického rozdělení.

Ve druhé kapitole zprvu formálně definujeme dvojrozměrné geometrické rozdělení a dále z této definice odvodíme momenty a vytvářející funkci tohoto rozdělení. Výše zmíněný článek se v úvodu odkazuje na definici vícerozměrného negativně binomického rozdělení odvozeného v článku Guldberg (1934), ze které je následně odvozeno geometrické rozdělení jako speciální případ. V této práci definici odvodíme sami, intuitivně, nikoli jako speciální případ jiného rozdělení, a podrobně ukážeme veškeré výpočty.

Ve třetí kapitole se seznámíme s verzí dvojrozměrného geometrického rozdělení vzniklou podmiňováním. Poprvé byl tento přístup zmíněn v Sibuya a kol. (1964), kdy se opět jednalo pouze o negativně binomické rozdělení. Spočteme vytvářející funkci, ze které odvodíme marginální rozdělení jednotlivých veličin. Také zde provedeme přímé rozšíření na negativně binomické rozdělení. V další části této kapitoly se zaměříme na případ, ve kterém uvažujeme tři typy neúspěchů, a také zde poznatky rozšíříme na případ negativně binomického rozdělení.

V poslední kapitole nalezneme odhady parametrů dvojrozměrného geometrického rozdělení metodou maximální věrohodnosti. Následně se zaměříme na vlastnosti těchto odhadů a intervaly spolehlivosti. V poslední části práce provedeme jednoduchou simulaci pro konkrétní situaci, ve které se pokusíme předvést spolehlivost nalezených odhadů.

# 1. Problematika vícerozměrných rozdělení

Jednorozměrné geometrické rozdělení modeluje počet neúspěchů před prvním úspěchem v nezávislých bernoulliiovských pokusech. Jedná se o jedno ze základních diskretních pravděpodobnostních rozdělení. Ekvivalentně na něj lze nahlížet jako na diskretní verzi exponenciálního rozdělení, které modeluje dobu čekání na nějakou událost ve spojitém čase.

Podobně jako multinomické rozdělení je zobecněním binomického rozdělení, je vícerozměrné geometrické rozdělení rozšířením geometrického rozdělení, kdy neuvažujeme pouze 2 stavy - úspěch a neúspěch - nýbrž více stavů, v našem případě více typů neúspěchů. Problematiku ilustrujeme na dvou základních příkladech.

**Příklad.** *Uvažujme hody spravedlivou kostkou. Hra skončí ve chvíli, kdy na kostce padne šestka. Předpokládejme, že šestka padla v  $n$ -tém kole. Kolik padlo sudých čísel menších než šest? Kolik lichých? Kolik padlo čísel menších nebo rovných než tři? Kolik větších?*

V tomto příkladu vidíme, že mohou nastat tři stavy: padne šestka, padne sudé číslo menší než šest, padne liché číslo. Jedná se tedy o situaci, ve které uvažujeme dva typy neúspěchů. Podobně máme dva typy neúspěchů rozdělené podle toho, zda na kostce padlo číslo větší než tři (ale ne šestka), nebo číslo menší či rovné třem.

**Příklad.** *Mějme urnu s kuličkami čtyř barev: červené, zelené, černé a bílé. V každém kole vytáhneme jednu kuličku. Pokud je bílá, hra končí. Pokud není, vrátíme ji zpět a taháme další. Kolikrát jsme před vytažením bílé kuličky vytáhli zelenou, červenou, nebo černou kuličku?*

Zde si můžeme všimnout, že uvažujeme 4 možné stavy podle toho, jakou barvu kuličky vytáhneme. Jedná se tedy o případ, kdy máme 3 typy neúspěchů podle toho, jestli jsme v daném kole vytáhli červenou, zelenou, nebo černou kuličku.

Je známo, že exponenciální rozdělení je jediné z rodiny spojitých rozdělení, které je bez paměti. Podobně je tomu i v rodině diskretních rozdělení u rozdělení geometrického. To znamená, že v daném okamžiku je pravděpodobnost dosažení úspěchu nezávislá na tom, co se dělo v minulosti. Formálněji, mějme náhodnou veličinu  $X$  udávající počet neúspěchů před prvním úspěchem. Potom pro rozdělení veličiny  $X$  platí rovnost

$$P(X \geq n + m | X \geq m) = P(X \geq n)$$

pro  $n, m \in \mathbb{N}_0$ .

## 2. Definice a základní vlastnosti

Nejprve si odvodíme základní vlastnosti dvojrozměrného geometrického rozdělení, tedy rozdělení se dvěma typy neúspěchů. Předpokládejme, že v každém pokusu musí nastat právě jedna ze tří událostí  $U$  (úspěch),  $N_1$  (neúspěch I. typu) a  $N_2$  (neúspěch II. typu), přičemž pro nějaká kladná  $p_1$  a  $p_2$  splňující  $p_1 + p_2 < 1$  platí

$$P(N_1) = p_1, \quad P(N_2) = p_2, \quad P(U) = 1 - (p_1 + p_2) =: p_0.$$

Výsledky jednotlivých pokusů jsou navzájem nezávislé náhodné veličiny. Označme  $X_1$  náhodnou veličinu udávající počet neúspěchů I. typu před prvním úspěchem a  $X_2$  náhodnou veličinu udávající počet neúspěchů II. typu před prvním úspěchem. Rozdělení  $(X_1, X_2)^\top$  náhodného vektoru počtu neúspěchů jednotlivých druhů před prvním úspěchem popisuje následující definice.

**Definice 1.** *Náhodný vektor  $(X_1, X_2)^\top$  má dvojrozměrné geometrické rozdělení (anglicky **bivariate geometric distribution**), které je dáno vzorcem*

$$P(X_1 = n, X_2 = m) = \binom{n+m}{n} p_1^n p_2^m p_0, \quad (2.1)$$

pro  $n, m \in \mathbb{N}_0$ ,  $p_1, p_2 \in (0, 1)$ , kde  $p_0 = 1 - (p_1 + p_2)$ .

### 2.1 Momenty a vytvořující funkce

Mějme náhodný vektor  $(X_1, X_2)^\top$  z definice 1. V této části určíme jeho střední hodnotu, varianční matici a vytvořující funkci.

Pro střední hodnotu vektoru platí  $E(X_1, X_2)^\top = (E X_1, E X_2)^\top$  a proto stačí spočítat jen  $E X_1$ :

$$\begin{aligned} E X_1 &= \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \binom{n+m}{n} \left(\frac{p_1}{p_1+p_2}\right)^n \left(\frac{p_2}{p_1+p_2}\right)^m (p_1+p_2)^{n+m} p_0 \cdot n \\ &= \sum_{k=0}^{\infty} \sum_{l=0}^k \binom{k}{l} \left(\frac{p_1}{p_1+p_2}\right)^l \left(\frac{p_2}{p_1+p_2}\right)^{k-l} (p_1+p_2)^k p_0 \cdot l \\ &= \sum_{k=0}^{\infty} (p_1+p_2)^k p_0 \sum_{l=0}^k \binom{k}{l} \left(\frac{p_1}{p_1+p_2}\right)^l \left(\frac{p_2}{p_1+p_2}\right)^{k-l} \cdot l \\ &= \frac{p_1}{p_1+p_2} p_0 \sum_{k=0}^{\infty} (p_1+p_2)^k \cdot k \\ &= \frac{p_1}{p_1+p_2} p_0 (p_1+p_2) \sum_{k=1}^{\infty} (p_1+p_2)^{k-1} \cdot k = p_1 p_0 \frac{1}{p_0^2} \\ &= \frac{p_1}{p_0}. \end{aligned}$$

Pro čtvrtou rovnost jsme využili známého faktu

$$\sum_{l=0}^k \binom{k}{l} \left(\frac{p_1}{p_1+p_2}\right)^l \left(\frac{p_2}{p_1+p_2}\right)^{k-l} \cdot l = k \frac{p_1}{p_1+p_2},$$

který plyne ze vzorce pro střední hodnotu binomického rozdělení  $Bi(k, \frac{p_1}{p_1+p_2})$ , a dále jsme zbylou sumu zderivovali a využili vzorec pro součet geometrické řady. Podobně pro  $X_2$  dostáváme  $EX_2 = \frac{p_2}{p_0}$ . Střední hodnota náhodného vektoru  $(X_1, X_2)^\top$  je tedy rovna

$$\left(\frac{p_1}{p_0}, \frac{p_2}{p_0}\right)^\top.$$

Z jednoduchého geometrického rozdělení plyne  $E(X_1 + X_2) = (1 - p_0)/p_0 = (p_1 + p_2)/p_0$ , což je v souladu s dosaženým výsledkem.

Pro kovarianční matici

$$\mathbb{K}_{X_1, X_2} = \begin{pmatrix} \text{var}(X_1) & \text{cov}(X_1, X_2) \\ \text{cov}(X_1, X_2) & \text{var}(X_2) \end{pmatrix}$$

dále využijeme vzorec  $\text{var} X_1 = E(X_1(X_1 - 1)) + E(X_1 - (E X_1)^2)$ . Nejprve určíme

$$\begin{aligned} E(X_1(X_1 - 1)) &= \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \binom{m+n}{n} p_1^n p_2^m p_0 \cdot n(n-1) \\ &= \sum_{k=0}^{\infty} \sum_{l=0}^k \binom{k}{l} \left(\frac{p_1}{p_1+p_2}\right)^l \left(\frac{p_2}{p_1+p_2}\right)^{k-l} (p_1+p_2)^k p_0 \cdot l(l-1) \\ &= \sum_{k=0}^{\infty} (p_1+p_2)^k p_0 \underbrace{\sum_{l=0}^k \binom{k}{l} \left(\frac{p_1}{p_1+p_2}\right)^l \left(\frac{p_2}{p_1+p_2}\right)^{k-l}}_{=(p_1/(p_1+p_2))^2 \cdot k(k-1)} \cdot l(l-1) \\ &= \sum_{k=2}^{\infty} (p_1+p_2)^k p_0 \left(\frac{p_1}{p_1+p_2}\right)^2 \cdot k(k-1) \\ &= \left(\frac{p_1}{p_1+p_2}\right)^2 p_0 (p_1+p_2)^2 \sum_{k=2}^{\infty} k(k-1) (p_1+p_2)^{k-2} \\ &= \left(\frac{p_1}{p_1+p_2}\right)^2 p_0 (p_1+p_2)^2 \frac{2}{p_0^3} \\ &= \frac{2p_1^2}{p_0^2}. \end{aligned}$$

Výraz ve svorce totiž lze vypočítat následovně:

$$\begin{aligned}
& \sum_{l=0}^k \binom{k}{l} \left( \frac{p_1}{p_1 + p_2} \right)^l \left( \frac{p_2}{p_1 + p_2} \right)^{k-l} \cdot l(l-1) \\
&= \sum_{l=0}^k \frac{k!}{l!(k-l)!} \left( \frac{p_1}{p_1 + p_2} \right)^l \left( \frac{p_2}{p_1 + p_2} \right)^{k-l} \cdot l(l-1) \\
&= \left( \frac{p_1}{p_1 + p_2} \right)^2 k(k-1) \sum_{l=2}^k \frac{(k-2)!}{(l-2)!(k-l)!} \left( \frac{p_1}{p_1 + p_2} \right)^{l-2} \left( \frac{p_2}{p_1 + p_2} \right)^{k-l} \quad (2.2) \\
&\stackrel{(\Delta)}{=} \left( \frac{p_1}{p_1 + p_2} \right)^2 k(k-1) \sum_{l=0}^k \binom{k}{l} \left( \frac{p_1}{p_1 + p_2} \right)^l \left( \frac{p_2}{p_1 + p_2} \right)^{k-l} \\
&= \left( \frac{p_1}{p_1 + p_2} \right)^2 k(k-1) \left( \frac{p_1}{p_1 + p_2} + \frac{p_2}{p_1 + p_2} \right)^k \\
&= \left( \frac{p_1}{p_1 + p_2} \right)^2 k(k-1),
\end{aligned}$$

kde jsme v  $(\Delta)$  pouze posunuli vzniklou sumu a následně využili binomickou větu. Celkem tedy dostáváme:

$$\text{var } X_1 = \frac{2p_1^2}{p_0^2} + \frac{p_1}{p_0} - \frac{p_1^2}{p_0^2} = \frac{p_1(1-p_2)}{p_0^2}.$$

Obdobně získáme

$$\text{var } X_2 = \frac{p_2(1-p_1)}{p_0^2}.$$

Nyní spočítáme kovarianci:

$$\begin{aligned}
\mathbb{E} X_1 X_2 &= \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \binom{m+n}{n} \left( \frac{p_1}{p_1 + p_2} \right)^n \left( \frac{p_2}{p_1 + p_2} \right)^m (p_1 + p_2)^{n+m} p_0 \cdot nm \\
&= \sum_{k=0}^{\infty} \sum_{l=0}^k \binom{k}{l} \left( \frac{p_1}{p_1 + p_2} \right)^l \left( \frac{p_2}{p_1 + p_2} \right)^{k-l} (p_1 + p_2)^k p_0 \cdot l(k-l) \\
&= \sum_{k=0}^{\infty} (p_1 + p_2)^k p_0 \left[ \underbrace{k \cdot \sum_{l=0}^k l \binom{k}{l} \frac{p_1^l p_2^{k-l}}{(p_1 + p_2)^k}}_{(*)} - \underbrace{\sum_{l=0}^k l^2 \binom{k}{l} \frac{p_1^l p_2^{k-l}}{(p_1 + p_2)^k}}_{(**)} \right].
\end{aligned}$$

Pro  $(*)$  platí

$$k \cdot \sum_{l=0}^k l \binom{k}{l} \left( \frac{p_1}{p_1 + p_2} \right)^l \left( \frac{p_2}{p_1 + p_2} \right)^{k-l} = k^2 \frac{p_1}{p_1 + p_2},$$

kde jsme využili znalosti střední hodnoty binomického rozdělení (v našem případě

se jedná o  $Bi(k, \frac{p_1}{p_1+p_2})$ ). Dále pro (\*\*) počítejme

$$\begin{aligned}
\sum_{l=0}^k l^2 \binom{k}{l} \left(\frac{p_1}{p_1+p_2}\right)^l \left(\frac{p_2}{p_1+p_2}\right)^{k-l} &= \sum_{l=0}^k [l + l(l-1)] \binom{k}{l} \frac{p_1^l p_2^{k-l}}{(p_1+p_2)^k} \\
&= \sum_{l=0}^k l \binom{k}{l} \frac{p_1^l p_2^{k-l}}{(p_1+p_2)^k} + \sum_{l=0}^k l(l-1) \binom{k}{l} \frac{p_1^l p_2^{k-l}}{(p_1+p_2)^k} \\
&= k \frac{p_1}{p_1+p_2} + k(k-1) \left(\frac{p_1}{p_1+p_2}\right)^2,
\end{aligned}$$

kde jsme opět využili znalost střední hodnoty binomického rozdělení s výše zmíněnými parametry a výpočet 2.2. Sečtením získáme

$$\begin{aligned}
\mathbb{E} X_1 X_2 &= \sum_{k=1}^{\infty} (p_1+p_2)^k p_0 \left[ k^2 \frac{p_1}{p_1+p_2} - k \frac{p_1}{p_1+p_2} - k(k-1) \left(\frac{p_1}{p_1+p_2}\right)^2 \right] \\
&= \sum_{k=1}^{\infty} (p_1+p_2)^k p_0 k(k-1) \frac{p_1 p_2}{(p_1+p_2)^2} \\
&= \frac{p_1 p_2}{(p_1+p_2)^2} p_0 (p_1+p_2)^2 \sum_{k=2}^{\infty} k(k-1) (p_1+p_2)^{k-2} \\
&= p_1 p_2 p_0 \frac{2}{p_0^3} \\
&= \frac{2p_1 p_2}{p_0^2}.
\end{aligned}$$

Celkem dostáváme

$$\begin{aligned}
\text{cov}(X_1, X_2) &= \mathbb{E} X_1 X_2 - \mathbb{E} X_1 \mathbb{E} X_2 \\
&= \frac{2p_1 p_2}{p_0^2} - \frac{p_1}{p_0} \frac{p_2}{p_0} \\
&= \frac{p_1 p_2}{p_0^2}.
\end{aligned}$$

Kovarianční matice má tedy tvar

$$\mathbb{K}_{X_1, X_2} = \frac{1}{p_0^2} \cdot \begin{pmatrix} p_1(1-p_2) & p_1 p_2 \\ p_1 p_2 & p_2(1-p_1) \end{pmatrix}.$$

Opět se snadno přesvědčíme, že  $\text{var}(X_1 + X_2) = (1-p_0)/p_0^2$ .

Nyní se zaměříme na pravděpodobnostní vytvořující funkci náhodného vektoru z definice 1. Tato funkce je užitečná pro popsání všech hodnot, kterých může náhodný vektor nabývat.

**Definice 2.** *Pravděpodobnostní vytvořující funkce* diskrétního náhodného vektoru  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^\top$  je definována jako

$$P_{\mathbf{Y}}(\mathbf{s}) = \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} \cdots \sum_{k_n=0}^{\infty} P(Y_1 = k_1, Y_2 = k_2, \dots, Y_n = k_n) s_1^{k_1} s_2^{k_2} \cdots s_n^{k_n}, \quad (2.3)$$

kde  $\mathbf{s} = (s_1, s_2, \dots, s_n)$ ,  $\mathbf{s} \in \mathbb{R}^n$  takové, že výsledný výraz konverguje.

Nejprve spočítáme vytvořující funkci binomického rozdělení. Pro náhodnou veličinu  $Z$ , která má binomické rozdělení  $Bi(n, p)$ , platí

$$\mathbb{P}(Z = k) = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k},$$

pro její vytvořující funkci získáme tedy

$$P_Z(s) = \sum_{k=0}^{\infty} \mathbb{P}(Z = k) s^k = \sum_{k=0}^n \binom{n}{k} (1-p)^{n-k} p^k s^k = [(1-p) + ps]^n,$$

kde jsme v poslední rovnosti využili binomickou větu.

Pro náhodný vektor  $(X_1, X_2)^\top$  z definice 1 můžeme nyní počítat

$$\begin{aligned} & P_{X_1, X_2}(s_1, s_2) \\ &= \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \binom{m+n}{n} \left(\frac{p_1}{p_1+p_2}\right)^n \left(\frac{p_2}{p_1+p_2}\right)^m (p_1+p_2)^{m+n} p_0 s_1^n s_2^m \\ &= \sum_{k=0}^{\infty} (p_1+p_2)^k p_0 s_2^k \sum_{l=0}^k \binom{k}{l} \left(\frac{p_1}{p_1+p_2}\right)^l \left(\frac{p_2}{p_1+p_2}\right)^{k-l} \left(\frac{s_1}{s_2}\right)^l \\ &= p_0 \sum_{k=0}^{\infty} (p_1+p_2)^k s_2^k \left(\frac{p_2}{p_1+p_2} + \frac{p_1}{p_1+p_2} \cdot \frac{s_1}{s_2}\right)^k \\ &= p_0 \sum_{k=0}^{\infty} \left[ (p_1+p_2) s_2 \frac{p_2 + p_1 s_1/s_2}{p_1+p_2} \right]^k \\ &= \frac{p_0}{1 - p_1 s_1 - p_2 s_2}. \end{aligned} \tag{2.4}$$

# 3. Okamžik nastání prvního úspěchu

## 3.1 Součet počtu neúspěchů

Uvažujme nyní stejný model jako výše, ve kterém máme tři typy událostí, které mohou nastat:  $U$ ,  $N_1$ ,  $N_2$  s pravděpodobnostmi postupně  $p_0$ ,  $p_1$ ,  $p_2$ . Sledujeme-li pokusy do  $n$ -tého kola, kde  $n$  je pevně dané, potom rozdělení náhodného vektoru  $(X_1, X_2)^\top$  odpovídá rozdělení multinomickému. Je-li totiž  $n$  pevné, pak pouze sledujeme, kolikrát jaká událost nastala a počet událostí typu  $U$  můžeme pak snadno dopočítat. My ovšem chceme pokus provádět pouze do prvního úspěchu.

Označme  $N$  počet pokusů předcházejících prvnímu úspěchu, tedy celkový počet neúspěchů bez ohledu na jejich typ. Potom má náhodný vektor  $(X_1, N)^\top$  sdružené rozdělení dané vzorcem

$$P(X_1 = x, N = n) = \binom{n}{x} p_1^x p_2^{n-x} p_0, \quad (3.1)$$

kde  $n \in \mathbb{N}_0$  a  $x \in \{0, 1, \dots, n\}$ .

Odsud spočítáme pravděpodobnostní vytvořující funkci

$$\begin{aligned} P_{X_1, N}(s_1, s_2) &= \sum_{n=0}^{\infty} \sum_{x=0}^n \binom{n}{x} p_1^x p_2^{n-x} p_0 s_1^x s_2^n \\ &= \sum_{n=0}^{\infty} p_0 s_2^n (p_1 s_1 + p_2)^n \\ &= \frac{p_0}{1 - p_1 s_1 s_2 - p_2 s_2}. \end{aligned}$$

Dosazením postupně  $s_2 = 1$ ,  $s_1 = 1$  získáme marginální rozdělení jednotlivých složek vektoru  $(X_1, N)^\top$ . Pro náhodnou veličinu  $X_1$  získáme

$$\begin{aligned} P_{X_1}(s_1) &= \frac{p_0}{1 - p_1 s_1 - p_2} \\ &= \frac{p_0}{p_0 + p_1 - p_1 s_1} \\ &= \frac{p_0}{p_0 + p_1(1 - s_1)} \\ &= \frac{\frac{p_0}{p_0 + p_1}}{1 - s_1 \left(1 - \frac{p_0}{p_0 + p_1}\right)}. \end{aligned}$$

Odsud vidíme, že veličina  $X_1$  má geometrické rozdělení s parametrem  $\frac{p_0}{p_0 + p_1}$ . Obdobně pro náhodnou veličinu  $N$  máme

$$P_N(s_2) = \frac{p_0}{1 - p_1 s_2 - p_2 s_2} = \frac{p_0}{1 - s_2(p_1 + p_2)} = \frac{p_0}{1 - s_2(1 - p_0)},$$

což odpovídá vytvořující funkci geometrického rozdělení s parametrem  $p_0$ .

Nyní zkoumejme podmíněné rozdělení  $X_1$  za znalosti celkového počtu neúspěchů, tedy hodnoty náhodné veličiny  $N$ . Ukazuje se, že za podmínky  $N = n$  je

rozdělení veličiny  $X_1$  binomické s parametry  $(n, p_1/(1 - p_0))$ . Z definice podmíněného rozdělení máme totiž

$$\begin{aligned} \mathbb{P}(X_1 = x | N = n) &= \frac{\mathbb{P}(X_1 = x, N = n)}{\mathbb{P}(N = n)} = \frac{\binom{n}{x} p_1^x p_2^{n-x} p_0}{(1 - p_0)^n p_0} \\ &= \binom{n}{x} \left( \frac{p_1}{1 - p_0} \right)^x \left( 1 - \frac{p_1}{1 - p_0} \right)^{n-x}. \end{aligned}$$

Sdružené rozdělení náhodného vektoru  $(X_1, X_2)^\top$  je poté dáno vzorcem 2.1 z definice dvojrozměrného geometrického rozdělení.

### 3.1.1 Rozšíření na negativně binomické rozdělení

Jak jsme konstatovali v první kapitole, geometrické rozdělení je speciálním případem negativně binomického rozdělení. Výše zmíněné dvojrozměrné geometrické rozdělení lze tedy snadno rozšířit na dvojrozměrný případ negativně binomického rozdělení, jehož rozdělení získáme ze vzorce 3.1, kdy budeme postupovat do okamžiku nastání  $r$ -tého úspěchu. Získáváme tak vzorec

$$\mathbb{P}(X_1 = x, N = n) = \binom{n+r-1}{r-1} \binom{n}{x} p_1^x p_2^{n-x} p_0^r.$$

Pravděpodobnostní vytvořující funkci lze pak získat z pravděpodobnostní vytvořující funkce pro dvojrozměrné geometrické rozdělení, a to sčítáním do  $r$ -tého úspěchu. Celou situaci si totiž můžeme rozdělit na nezávislá čekání na jednotlivé úspěchy. Je-li  $S = \sum_{k=1}^n X_k$  náhodná veličina definovaná jako součet nezávislých stejně rozdělených náhodných veličin, pak pro její vytvořující funkci platí  $P_S(s) = [P_{X_1}(s)]^n$ . V našem případě udávají veličiny  $X_i$ ,  $i \in \{1, 2, \dots, n\}$  počet neúspěchů před prvním úspěchem v  $i$ -tém kole. Díky 2.4 tak získáváme

$$P_{X_1, N}(s_1, s_2) = \left( \frac{p_0}{1 - p_1 s_1 s_2 - p_2 s_2} \right)^r.$$

## 3.2 Kombinované typy neúspěchů

Uvažujme čtyři možné stavy, které mohou nastat:  $U$ ,  $N_1$ ,  $N_2$ ,  $N_3$  s pravděpodobnostmi postupně  $p_0$ ,  $p_1$ ,  $p_2$ ,  $p_3$ . Dále označme  $X_{i,j}$  náhodnou veličinu udávající počet výstupů  $N_i$  nebo  $N_j$ ,  $i \neq j$ . Pokus provádíme do prvního dosažení stavu  $U$ , kterým rozumíme úspěch. Podobně jako výše označme  $N$  počet neúspěchů před prvním úspěchem. Potom platí

$$\begin{aligned} \mathbb{P}(X_{1,2} = x, X_{2,3} = y) &= \sum_{k=0}^{\min(x,y)} \mathbb{P}(X_1 = x - k, X_2 = k, X_3 = y - k) \\ &= \sum_{k=0}^{\min(x,y)} \binom{x}{k} \binom{x+y-k}{y-k} p_1^{x-k} p_2^k p_3^{y-k} p_0, \end{aligned}$$

kde  $x, y \in \mathbb{N}_0$ .

**Poznámka.** Všimněme si, že pokud  $p_1 = 0$ , pak musí nutně platit  $X_1 = 0$ , a tedy  $X_{1,2} = X_2$ , a potom  $X_{2,3} = N$ . Za těchto podmínek jsme tedy v případě ze sekce 3.1. Pokud  $p_2 = 0$ , pak  $X_2 = 0$ , a tedy  $X_{1,2} = X_1$ ,  $X_{2,3} = X_3$ . Za těchto podmínek jsme tedy v případě z kapitoly 2.

Nyní budeme zkoumat marginální rozdělení náhodných veličin  $X_{1,2}$  a  $X_{2,3}$ . Podobně jako v sekci 3.1 určíme pravděpodobnostní vytvořující funkci sdruženého náhodného vektoru  $(X_{1,2}, X_{2,3})^\top$ . K tomu budeme potřebovat následující tvrzení.

**Tvrzení 1.** Necht  $K$  je diskrétní náhodná veličina,  $\{Y_i\}$ ,  $i \in \{1, 2, \dots\}$  jsou nezávislé stejně rozdělené náhodné veličiny a necht je veličina  $K$  s nimi nezávislá. Označíme-li  $S_K = \sum_{i=1}^K Y_i$ , pak pro  $P_{Y_1}$  vytvořující funkci náhodné veličiny  $Y_1$ ,  $P_K$  vytvořující funkci náhodné veličiny  $K$  platí

$$P_{S_K}(s) = P_K(P_{Y_1}(s)).$$

*Důkaz.* Z definice pravděpodobnostní vytvořující funkce počítejme

$$\begin{aligned} P_{S_K}(s) &= \mathbf{E} s^{S_K} = \mathbf{E} \left( \mathbf{E} \left[ s^{S_K} | K \right] \right) = \sum_{k=0}^{\infty} \mathbf{E} \left[ s^{S_K} | K = k \right] \mathbf{P}(K = k) \\ &= \sum_{k=0}^{\infty} \mathbf{E} \left[ s^{S_k} | K = k \right] \mathbf{P}(K = k) = \sum_{k=0}^{\infty} \mathbf{E} (s^{Y_1})^k \mathbf{P}(K = k) \\ &= \sum_{k=0}^{\infty} [P_{Y_1}(s)]^k \mathbf{P}(K = k) = P_K(P_{Y_1}(s)), \end{aligned}$$

kde jsme ve třetí rovnosti využili větu o úplné pravděpodobnosti a v páté rovnosti nezávislost  $K$  a  $\{Y_i\}_{i=1}^{\infty}$  a toho, že veličiny  $\{Y_i\}_{i=1}^{\infty}$  jsou stejně rozdělené a nezávislé. □

Nejprve se zaměříme na vytvořující funkci náhodného vektoru  $(X_{1,2}, X_{2,3})^\top$  podmíněnou jevem  $N = n$ . Ze sekce 3.1 víme, že  $N$  má geometrické rozdělení s parametrem  $p_0$ . Podmíněné rozdělení náhodného vektoru  $(X_{1,2}, X_{2,3})^\top$  za podmínky  $N = n$  pak z definice podmíněného rozdělení je

$$\begin{aligned} &\mathbf{P}(X_{1,2} = x, X_{2,3} = y | N = n) \\ &= \frac{\mathbf{P}(X_1 = n - y, X_2 = x + y - n, X_3 = n - x, N = n)}{\mathbf{P}(N = n)} \\ &= \binom{n}{x} \binom{x}{n - y} \frac{p_1^{n-y} p_2^{x+y-n} p_3^{n-x}}{(1 - p_0)^n}, \end{aligned}$$

kde  $x \in \{0, 1, \dots, n\}$  a  $y \in \{n - x, \dots, n\}$ .

Odsud můžeme pak určit pravděpodobnostní vytvořující funkci

$$\begin{aligned}
P_{X_{1,2}, X_{2,3} | N=n}(s_1, s_2) &= \sum_{x=0}^n \sum_{y=n-x}^n \binom{n}{x} \binom{x}{n-y} \frac{p_1^{n-y} p_2^{x+y-n} p_3^{n-x}}{(1-p_0)^n} s_1^x s_2^y \\
&= \frac{1}{(1-p_0)^n} \sum_{x=0}^n \binom{n}{x} s_1^x (p_3 s_2)^{n-x} \sum_{l=0}^x \binom{x}{l} p_1^{x-l} (p_2 s_2)^l \\
&= \frac{1}{(1-p_0)^n} \sum_{x=0}^n \binom{n}{x} (p_1 s_1 + p_2 s_1 s_2)^x (p_3 s_2)^{n-x} \\
&= \left( \frac{p_1 s_1 + p_2 s_1 s_2 + p_3 s_2}{1-p_0} \right)^n.
\end{aligned} \tag{3.2}$$

Z tvrzení 1 a výpočtu 3.2 pak dostáváme

$$\begin{aligned}
P_{X_{1,2}, X_{2,3}(s_1, s_2)} &= P_N \left( \frac{p_1 s_1 + p_2 s_1 s_2 + p_3 s_2}{1-p_0} \right) \\
&= \frac{p_0}{1 - p_1 s_1 - p_2 s_1 s_2 - p_3 s_2}.
\end{aligned}$$

Marginální rozdělení náhodných veličin  $X_{1,2}$  a  $X_{2,3}$  získáme obdobně jako výše dosazením postupně  $s_2 = 1$ ,  $s_1 = 1$  do pravděpodobnostní vytvořující funkce sdruženého náhodného vektoru  $(X_{1,2}, X_{2,3})^\top$ . Získáme postupně

$$P_{X_{1,2}}(s_1) = \frac{p_0}{1 - s_1(p_1 + p_2) - p_3} = \frac{\frac{p_0}{1-p_3}}{1 - s_1 \left(1 - \frac{p_0}{1-p_3}\right)}$$

a

$$P_{X_{2,3}}(s_2) = \frac{p_0}{1 - p_1 - s_2(p_2 + p_3)} = \frac{\frac{p_0}{1-p_1}}{1 - s_2 \left(1 - \frac{p_0}{1-p_1}\right)}.$$

Z těchto vyjádření můžeme nahlédnout, že veličiny  $X_{1,2}$  a  $X_{2,3}$  mají geometrické rozdělení s parametrem  $p_0/(1-p_3)$  a  $p_0/(1-p_1)$  v tomto pořadí.

### 3.2.1 Momenty

Díky znalosti marginálních rozdělení náhodných veličin  $X_{1,2}$  a  $X_{2,3}$  můžeme snadno určit jejich střední hodnoty a rozptyly. Dostáváme tak

$$\mathbb{E} X_{i,j} = \frac{p_i + p_j}{p_0} \quad \text{a} \quad \text{var} X_{i,j} = \frac{(p_i + p_j)(p_i + p_j + p_0)}{p_0^2},$$

kde  $(i,j) = (1,2)$  nebo  $(i,j) = (2,3)$ .

### 3.2.2 Rozšíření na negativně binomické rozdělení

Obdobně jako na konci sekce 3.1 můžeme získané výsledky aplikovat na negativně binomické rozdělení, které si i zde můžeme interpretovat jako součet  $r$  nezávislých náhodných veličin s geometrickým rozdělením. Vytvořující funkce bude mít tvar

$$P_{X_{1,2}, X_{2,3}}(s_1, s_2) = \left( \frac{p_0}{1 - p_1 s_1 - p_2 s_1 s_2 - p_3 s_2} \right)^r$$

a pro pravděpodobnostní rozdělení platí

$$P(X_{1,2} = x, X_{2,3} = y) = \sum_{k=0}^{\min(x,y)} \binom{x+y-k+r-1}{r-1} \binom{x+y-k}{x-k} \binom{y}{k} p_1^{x-k} p_2^k p_3^{y-k} p_0^r,$$

kde  $x, y \in \mathbb{N}_0$ .

# 4. Odhady parametrů

## 4.1 Odhady metodou maximální věrohodnosti

V této části se budeme zabývat odhady parametrů dvojrozměrného geometrického rozdělení. K jejich nalezení využijeme metodu maximální věrohodnosti. Připomeňme, že známe-li pravděpodobnosti neúspěchů jednotlivých druhů, ku příkladu  $p_1$  a  $p_2$ , pak pravděpodobnost úspěchu  $p_0$  můžeme dopočítat z jejich znalosti jako  $p_0 = 1 - p_1 - p_2$ . My ovšem  $p_0$  budeme považovat za další parametr našeho rozdělení právě z toho důvodu, že odpovídá pravděpodobnosti úspěchu. Pokud nás totiž zajímá pouze to, s jakou pravděpodobností nastane úspěch, můžeme použít odhad parametru  $p_0$  přímo a nemusíme jej dopočítávat z výše zmíněné rovnosti. Navíc při reálných simulacích dochází k zaokrouhlování odhadů parametrů. Při dopočítávání  $p_0$  bychom se tak mohli dopouštět větší chyby, než při výpočtu přímo z odhadu pro  $p_0$ .

Mějme nyní náhodný výběr  $X_1, X_2, \dots, X_n$  z dvojrozměrného geometrického rozdělení s parametry  $p_1, p_2$  (a  $p_0$ ). Označme  $\boldsymbol{\theta} = (p_0, p_1, p_2)$ . Označme náhodný vektor  $\mathbf{X} = (X_1, X_2, \dots, X_n)^\top$ . Dále označme  $X_{+1} = \sum_{i=1}^n X_{i,1}$ , podobně pro  $X_{+2}$ . Pro sdruženou hustotu náhodného vektoru  $\mathbf{X}$  platí

$$f_{\mathbf{X}}(\mathbf{x}, \boldsymbol{\theta}) = \prod_{i=1}^n \binom{x_{i,1} + x_{i,2}}{x_{i,1}} p_1^{x_{i,1}} p_2^{x_{i,2}} p_0 = p_1^{x_{+1}} p_2^{x_{+2}} p_0^n \prod_{i=1}^n \binom{x_{i,1} + x_{i,2}}{x_{i,1}}$$

Dále označme  $\mathcal{L}_n(\boldsymbol{\theta}) = \log(f_{\mathbf{X}}(\mathbf{X}, \boldsymbol{\theta}))$ , potom

$$\mathcal{L}_n(\boldsymbol{\theta}) = X_{+1} \log(p_1) + X_{+2} \log(p_2) + n \cdot \log(p_0) + \sum_{i=1}^n \binom{X_{i,1} + X_{i,2}}{X_{i,1}}.$$

Nyní využijeme Lagrangeovy multiplikátory. V našem případě přidáme podmínku  $p_0 + p_1 + p_2 = 1$ . Pišme tedy

$$\mathfrak{L}_n(\boldsymbol{\theta}) = \mathcal{L}_n(\boldsymbol{\theta}) - \lambda(p_0 + p_1 + p_2 - 1).$$

Nyní funkci  $\mathfrak{L}_n(\boldsymbol{\theta})$  zderivujeme podle všech parametrů a dané výrazy položíme rovny nule. Pro  $p_1$  tak například dostáváme rovnost

$$\frac{\partial \mathfrak{L}_n(\boldsymbol{\theta})}{\partial p_1} = \frac{X_{+1}}{p_1} - \lambda = 0,$$

odkud dostáváme

$$p_1 = \frac{X_{+1}}{\lambda}, \tag{4.1}$$

pro  $p_0$  a  $p_2$  obdobným výpočtem dostáváme

$$p_0 = \frac{n}{\lambda}, \quad p_2 = \frac{X_{+2}}{\lambda}.$$

Hodnotu  $\lambda$  určíme z těchto rovností a z vazebné podmínky:

$$1 = p_0 + p_1 + p_2 = \frac{1}{\lambda}(X_{+1} + X_{+2} + n) \implies \lambda = X_{+1} + X_{+2} + n.$$

Pro  $p_1$  tak dostáváme z rovnosti 4.1 odhad

$$\widehat{p}_{1n} = \frac{X_{+1}}{X_{+1} + X_{+2} + n},$$

obdobně pro  $p_2$  získáváme odhad

$$\widehat{p}_{2n} = \frac{X_{+2}}{X_{+1} + X_{+2} + n}$$

a nakonec pro  $p_0$  je odhad metodou maximální věrohodnosti roven

$$\widehat{p}_{0n} = \frac{n}{X_{+1} + X_{+2} + n}.$$

Zjistili jsme tedy, že odhad vektorového parametru  $\theta$  metodou maximální věrohodnosti je roven

$$\widehat{\theta}_n^\top = \left( \frac{n}{X_{+1} + X_{+2} + n}, \frac{X_{+1}}{X_{+1} + X_{+2} + n}, \frac{X_{+2}}{X_{+1} + X_{+2} + n} \right)^\top.$$

Pokud tedy máme k dispozici  $n$  pozorování, pak pravděpodobnost úspěchu získáme jako tento počet pozorování; v případě geometrického rozdělení se jedná o počet nastalých úspěchů, dělený celkovým počtem kol, která k těmto  $n$  úspěchům vedla, tedy počtem všech pozorovaných událostí. To souhlasí také s intuicí, která nám říká, že jednotlivé pravděpodobnosti můžeme odhadnout jako počet událostí daného typu dělený celkovým počtem všech událostí.

## 4.2 Vlastnosti odhadů

Zaměříme se nyní na vlastnosti získaných odhadů. O odhadech získaných metodou maximální věrohodnosti je obecně známo, že bývají (slabě) konzistentní a zpravidla nejsou nestranné. Zaměříme se nyní na konzistenci odhadů jednotlivých parametrů. Ze silného zákona velkých čísel pro nezávislé stejně rozdělené náhodné veličiny dostáváme, že

$$\frac{1}{n} \sum_{i=1}^n X_{i,1} = \bar{X}_{n,1} \xrightarrow[n \rightarrow \infty]{P} \frac{p_1}{p_0},$$

a podobně

$$\frac{1}{n} \sum_{i=1}^n X_{i,2} = \bar{X}_{n,2} \xrightarrow[n \rightarrow \infty]{P} \frac{p_2}{p_0}.$$

Potom pro funkci  $g : (0, 1)^2 \mapsto \mathbb{R}$ , pro niž platí  $g(x, y) = x/(x + y + 1)$  a která je spojitá na svém definičním oboru, získáváme z věty o spojitě transformaci

$$\widehat{p}_{1n} = g(\bar{X}_{n,1}, \bar{X}_{n,2}) \xrightarrow[n \rightarrow \infty]{P} g\left(\frac{p_1}{p_0}, \frac{p_2}{p_0}\right) = \frac{\frac{p_1}{p_0}}{\frac{p_1}{p_0} + \frac{p_2}{p_0} + 1} = \frac{p_1 p_0}{p_0(p_0 + p_1 + p_2)} = p_1.$$

Vidíme tedy, že odhad  $\widehat{p}_{1n}$  je konzistentním odhadem parametru  $p_1$ . Obdobně díky funkci  $h : (0, 1)^2 \mapsto \mathbb{R}$  definované předpisem  $h(x, y) = y/(x + y + 1)$  získáme z věty o spojitě transformaci konzistenci  $\widehat{p}_{2n}$ . Konzistenci odhadu  $\widehat{p}_{0n}$  pak obdržíme díky

funkci  $k : (0, 1)^2 \mapsto \mathbb{R}$ , pro niž  $k(x, y) = 1/(x + y + 1)$ . Odsud již dostáváme, že odhad  $\widehat{\boldsymbol{\theta}}_n$  je konzistentním odhadem  $\boldsymbol{\theta}$ .

Z vlastností výběrového průměru víme, že

$$\mathbb{E} \bar{X}_{n,1} = \frac{p_1}{p_0}, \quad \mathbb{E} \bar{X}_{n,2} = \frac{p_2}{p_0}.$$

Potom pro  $p_0$  z Jensenovy nerovnosti pro funkci  $k(x, y)$  definovanou výše platí

$$\mathbb{E} \widehat{p}_{0n} = \mathbb{E} k(\bar{X}_{n,1}, \bar{X}_{n,2}) > k(\mathbb{E} \bar{X}_{n,1}, \mathbb{E} \bar{X}_{n,2}) = p_0,$$

a tedy vidíme, že odhad  $\widehat{p}_{0n}$  není nestranný.

Dále si uvědomme, že pokud by byl odhad  $\widehat{p}_{1n}$  nestranný, pak by i  $\widehat{p}_{2n}$  musel být nestranný, protože např. střední hodnotu odhadu  $\widehat{p}_{2n}$  lze snadno získat ze střední hodnoty odhadu  $\widehat{p}_{1n}$  prohozením  $p_1$  a  $p_2$ . Aplikujme nyní na obě strany rovnosti

$$\widehat{p}_{0n} + \widehat{p}_{1n} + \widehat{p}_{2n} = 1$$

střední hodnotu. Z linearity střední hodnoty pak dostáváme

$$\mathbb{E} \widehat{p}_{0n} + \mathbb{E} \widehat{p}_{1n} + \mathbb{E} \widehat{p}_{2n} = \mathbb{E} 1 = 1.$$

Ovšem  $\mathbb{E} \widehat{p}_{0n} \neq p_0$ , tedy nutně  $\mathbb{E} \widehat{p}_{1n} \neq p_1$  a  $\mathbb{E} \widehat{p}_{2n} \neq p_2$ . Odhady tedy nejsou nestranné a tedy  $\widehat{\boldsymbol{\theta}}_n$  není nestranným odhadem  $\boldsymbol{\theta}$ .

### 4.3 Intervalv spolehlivosti

Pro jednotlivé odhady parametrů nyní určíme (asymptotické) intervaly spolehlivosti. Z centrální limitní věty pro náhodný výběr z dvojrozměrného geometrického rozdělení dostáváme

$$\sqrt{n} \left( \begin{pmatrix} \bar{X}_{n,1} \\ \bar{X}_{n,2} \end{pmatrix} - \begin{pmatrix} \frac{p_1}{p_0} \\ \frac{p_2}{p_0} \end{pmatrix} \right) \xrightarrow[n \rightarrow \infty]{d} N_2 \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \frac{1}{p_0^2} \begin{pmatrix} p_1(1-p_2) & p_1 p_2 \\ p_1 p_2 & p_2(1-p_1) \end{pmatrix} \right)$$

Dále využijeme delta metodu k nalezení asymptotického rozdělení pro jednotlivé odhady parametrů. Označíme-li  $\mathbf{X} = (X_1, X_2)^\top$  náhodný vektor s dvojrozměrným geometrickým rozdělením, pak asymptotická rozdělení neznámých parametrů budou obecně tvaru

$$\sqrt{n} \left( f(\bar{X}_{n,1}, \bar{X}_{n,2}) - f(\mathbb{E} \mathbf{X}) \right) \xrightarrow[n \rightarrow \infty]{d} N \left( 0, (\mathbb{D}f(\mathbb{E} \mathbf{X}))^\top \mathbb{K}_{X_1, X_2} (\mathbb{D}f(\mathbb{E} \mathbf{X})) \right), \quad (4.2)$$

kde  $\mathbb{K}_{X_1, X_2}$  je kovarianční matice náhodného vektoru  $\mathbf{X}$ ,  $f : (0, 1)^2 \rightarrow \mathbb{R}$  je nějaká vhodně zvolená funkce a  $\mathbb{D}f$  je Jacobiho matice této funkce.

Uvažujme nejprve pro  $\widehat{p}_{1n}$  funkci  $g(x, y) = x/(x + y + 1)$ . Pro Jacobiho matici této funkce platí

$$(\mathbb{D}g(x, y))^\top = \left( \frac{y+1}{(x+y+1)^2}, \frac{-x}{(x+y+1)^2} \right)$$

a dále po dosazení středních hodnot

$$\left( \mathbb{D}g \left( \frac{p_1}{p_0}, \frac{p_2}{p_0} \right) \right)^\top = \left( p_0(p_0 + p_2), -p_0 p_1 \right).$$

dosazením do vzorce 4.2 dostáváme

$$\sqrt{n}(\widehat{p}_{1n} - p_1) \xrightarrow[n \rightarrow \infty]{d} N(0, p_0 p_1 (1 - p_1)).$$

Odsud snadno nahlédneme, že interval spolehlivosti pro  $p_1$  je

$$\left( \widehat{p}_{1n} - \frac{u_{1-\alpha/2} \sqrt{\widehat{p}_{0n} \widehat{p}_{1n} (1 - \widehat{p}_{1n})}}{\sqrt{n}}, \widehat{p}_{1n} + \frac{u_{1-\alpha/2} \sqrt{\widehat{p}_{0n} \widehat{p}_{1n} (1 - \widehat{p}_{1n})}}{\sqrt{n}} \right),$$

kde  $u_{1-\alpha/2}$  je  $(1 - \alpha/2)$ -kvantil normovaného normálního rozdělení.

Obdobným způsobem lze spočítat asymptotická rozdělení a intervaly spolehlivosti pro zbylé parametry pomocí funkcí  $h(x, y) = y/(x + y + 1)$  pro parametr  $p_2$  a  $k(x, y) = 1/(x + y + 1)$  pro parametr  $p_0$ . Obecně pak pro  $i = 0, 1, 2$  dostáváme interval spolehlivosti pro  $p_i$  tvaru

$$\left( \widehat{p}_{in} - \frac{u_{1-\alpha/2} \sqrt{\widehat{p}_{0n} \widehat{p}_{in} (1 - \widehat{p}_{in})}}{\sqrt{n}}, \widehat{p}_{in} + \frac{u_{1-\alpha/2} \sqrt{\widehat{p}_{0n} \widehat{p}_{in} (1 - \widehat{p}_{in})}}{\sqrt{n}} \right).$$

## 4.4 Odhady parametrů pro hody kostkou

V závěrečné části kapitoly o odhadech parametrů se podíváme na konkrétní příklad odhadování parametrů v situaci, kdy házíme kostkou tak dlouho, dokud nepadne první šestka. Pomocí výše uvedených odhadů parametrů  $p_0, p_1, p_2$  spočítáme v naší konkrétní situaci uvedené pravděpodobnosti.

Mějme spravedlivou kostku a uvažujme tři možné stavy, které mohou nastat: padla šestka ( $U$ ), padlo sudé číslo menší než šest ( $N_1$ ) a padlo liché číslo ( $N_2$ ). Hra končí v okamžiku, kdy padne první šestka. Očekávané pravděpodobnosti jednotlivých stavů jsou

$$P(U = 1) = p_0 = \frac{1}{6}, \quad P(N_1 = 1) = p_1 = \frac{1}{3}, \quad P(N_2 = 1) = p_2 = \frac{1}{2}.$$

Tuto hru provedeme stokrát, přičemž u každé zaznamenané počet padlých sudých čísel menších než šest a počet padlých lichých čísel před první padlou šestkou. Nakonec pak dosadíme do vzorečků odvozených v sekci 4.2, čímž získáme kýžené odhady.

**Poznámka.** *Hody kostkou byly simulovány v programu PyCharm.*<sup>1</sup>

V proběhlé simulaci padlo ve 100 hrách celkem 181 sudých čísel menších než 6 a 275 lichých čísel. Ze vzorečků pro odhady dostáváme tak postupně

$$\widehat{p}_{0n} = \frac{100}{181 + 275 + 100} = \frac{25}{139} \doteq 0,18$$

$$\widehat{p}_{1n} = \frac{181}{181 + 275 + 100} = \frac{181}{556} \doteq 0,33$$

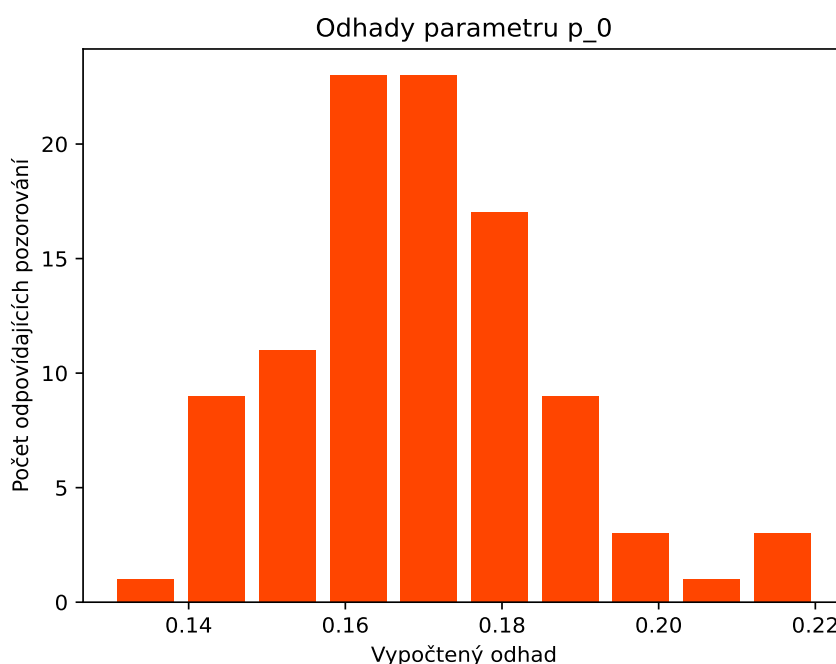
$$\widehat{p}_{2n} = \frac{275}{181 + 275 + 100} = \frac{275}{556} \doteq 0,49.$$

<sup>1</sup>Verze programu: PyCharm 2019.3.5 (Community Edition) Build PC-193.7288.30, built on May 7, 2020 Runtime version: 11.0.6+8-b520.66 amd64 VM: OpenJDK 64-Bit Server VM by JetBrains s.r.o Windows 10 10.0 GC: ParNew, ConcurrentMarkSweep Memory: 968M Cores: 8

Tyto odhady se od skutečných pravděpodobností liší jen zanedbatelně. Vidíme tedy, že odvozené odhady parametrů jsou poměrně spolehlivé a v praxi tak mohou sloužit jako dobrá aproximace skutečných neznámých hodnot parametrů.

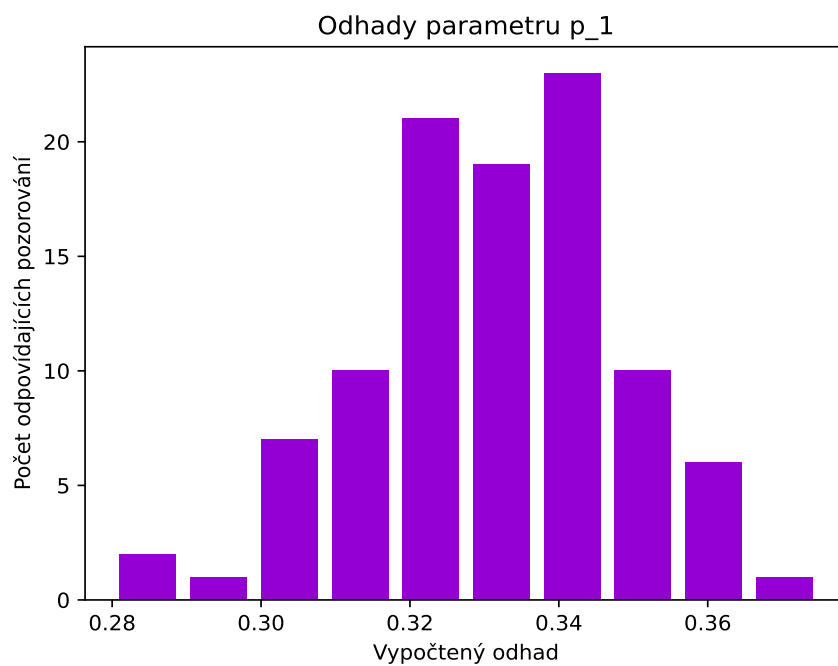
Dosažením do vzorce pro intervalový odhad dostáváme pro  $p_1$  95% interval spolehlivosti přibližně  $(0,29; 0,36)$ . Odsud vidíme, že tento intervalový odhad skutečnou hodnotu parametru pokrývá, a může tedy také sloužit jako spolehlivá aproximace v reálných situacích, kdy skutečnou hodnotu parametru neznáme. Obdobně pro  $p_2$  dostáváme 95% interval spolehlivosti  $(0,45; 0,54)$  a pro  $p_0$  máme interval  $(0,15; 0,21)$ .

Nyní celý tento pokus provedeme stokrát a jednotlivé četnosti vypočtených hodnot odhadů parametrů zaneseme do histogramů.<sup>2</sup>

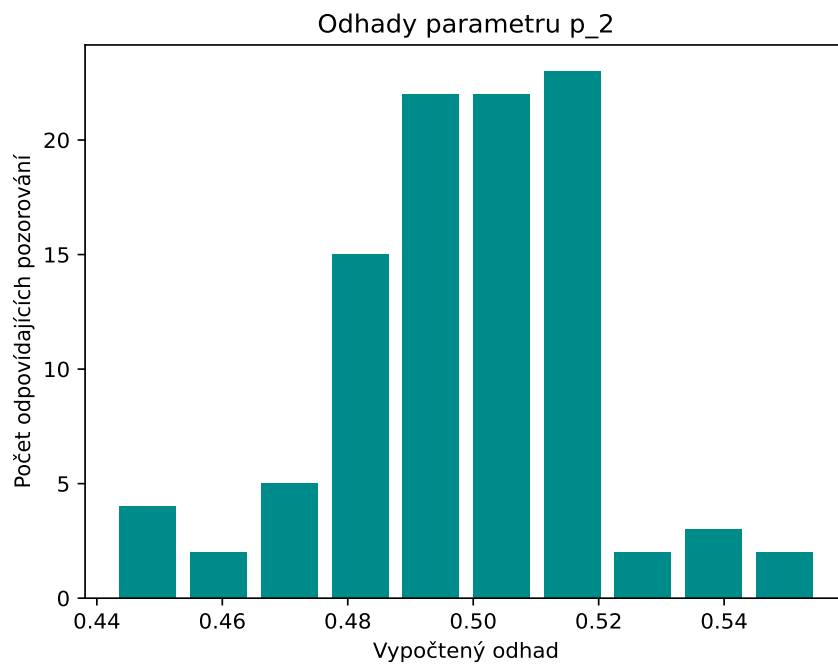


Obrázek 4.1: Četnosti vypočtených hodnot odhadu parametru  $p_0$ .

<sup>2</sup>Zdrojový kód k provedené simulaci lze nalézt pod následujícím odkazem v souboru *thesis-code*: <https://github.com/DianaPavlovic/Diana.git>.



Obrázek 4.2: Četnosti vypočtených hodnot odhadu parametru  $p_1$ .



Obrázek 4.3: Četnosti vypočtených hodnot odhadu parametru  $p_2$ .

V histogramu 4.1 četností odhadů parametru  $p_0$  vidíme, že nejčastější hodnota odhadu je přibližně mezi hodnotami 0,16 a 0,17. Skutečná hodnota parametru  $p_0$  v situaci, kdy házíme spravedlivou kostkou, je rovna  $1/6 \doteq 0,166$ , tedy odhad parametru funguje velmi spolehlivě.

Obdobně v histogramu 4.2 pro odhad parametru  $p_1$  vidíme, že většina pozorovaných hodnot se pohybuje mezi hodnotami 0,32 a 0,35. Skutečná hodnota  $p_1$  je rovna  $1/3 \doteq 0,333$ , tedy i tento odhad parametru je poměrně přesný.

Nakonec v histogramu 4.3 vidíme, že naměřené hodnoty se pohybují mezi hodnotami 0,48 až 0,52, přičemž skutečná hodnota parametru je rovna přesně 0,5.

Z výše popsané simulace vidíme, že odhady jsou velmi přesné a mohou nám tak dát velmi dobrou představu o skutečné hodnotě parametrů v situaci, kdy tyto parametry neznáme.

# Závěr

Tato práce umožňuje vhled do problematiky vícerozměrného geometrického rozdělení. Je zřejmé, že podobných rozšíření je možno udělat nepřeborné množství a že v práci odkrýváme jen malou část rozsáhlé problematiky. Zde jsme se zaměřili jen na ta rozšíření, která intuitivně vyplývají ze základní verze dvojrozměrného geometrického rozdělení.

Doufáme, že čtenáře tato práce obohatila a že mu příjemným způsobem odkryla svět vícerozměrných rozdělení.

# Seznam použité literatury

- DAVY, P. a RAYNER, J. (1996). Multivariate geometric distributions. Communications in Statistics - Theory and Methods, **25**(12), 2971–2987. doi: 10.1080/03610929608831881. URL <https://doi.org/10.1080/03610929608831881>.
- GULDBERG, A. (1934). On discontinuous frequency functions of two variables. Scandinavian Actuarial Journal, **1934**(1), 89–117.
- SIBUYA, M., YOSHIMURA, I. a SHIMIZU, R. (1964). Negative multinomial distribution. Annals of the Institute of Statistical Mathematics, **16**(1), 409–426.