MASARYKOVA UNIVERZITA
PŘÍRODOVĚDECKÁ FAKULTA
ÚSTAV MATEMATIKY A STATISTIKY

Bakalářská práce

BRNO 2015 MILENA TOPALOVIĆ

MASARYKOVA UNIVERZITA
PŘÍRODOVĚDECKÁ FAKULTA
ÚSTAV MATEMATIKY A STATISTIKY

Sbírka řešených příkladů z pravděpodobnosti a statistiky

Bakalářská práce

Milena Topalović

Vedoucí práce: Mgr. Jan Koláček, Ph.D. **Brno 2015**

Bibliografický záznam

Autor:	Milena Topalović Přírodovědecká fakulta, Masarykova univerzita Ústav matematiky a statistiky
Název práce:	Sbírka řešených příkladů z pravděpodobnosti a statistiky
Studijní program:	Matematika
Studijní obor:	Finanční a pojistná matematika
Vedoucí práce:	Mgr. Jan Koláček, Ph.D.
Akademický rok:	2014/15
Počet stran:	x + 101
Klíčová slova:	náhodná veličina; střední hodnota; rozptyl; kvantil; kovariance; korelace; náhodný vektor; centrální limitní věta; normální rozdělení; odhad; testování hypotéz

Bibliographic Entry

Author:	Milena Topalović Faculty of Science, Masaryk University Department of Mathematics and Statistics
Title of Thesis:	The collection of solved examples from theory of probability and mathematical statistics
Degree Programme:	Mathematics
Field of Study:	Financial and Actuarial Mathematics
Supervisor:	Mgr. Jan Koláček, Ph.D.
Academic Year:	2014/15
Number of Pages:	$x + 101$
Keywords:	random variable; the mean; variance; quantile; covariance; correlation; random vector; central limit theorem; normal distribution; estimate; hypothesis testing

Abstrakt

Cílem této bakalářské práce je sestavit sbírku řešených příkladů z pravděpodobnosti a statistiky. Sbírka bude doplněním k již existujícím učebním textům, vytvořeným k předmětu M4122. V textu jsou uvedeny teoretické základy, které jsou využity k vyřešení příkladů. Součástí jsou rovněž nevyřešené příklady, sloužící k procvičování probrané látky.

Abstract

The aim of this Bachelor thesis is to compile a collection of exercises in probability and statistics. The collection will be added as supplement to existing textbooks, created for the course M4122. Theretical foundations has been introduced within the text, which are than used as reference for solving problems. Also included are unsolved examples, used to practice the subject materia.

MASARYKOVA UNIVERZITA
Přírodovědecká fakulta

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

Akademický rok: 2014/2015

Ústav: Ústav matematiky a statistiky

Studentka: Milena Topalović

Program: Matematika

Obor: Finanční a pojistná matematika

Ředitel *Ústavu matematiky a statistiky* PřF MU Vám ve smyslu Studijního a zkušebního řádu MU určuje bakalářskou práci s tématem:

Téma práce: Sbírka řešených příkladů z pravděpodobnosti a statistiky

Téma práce anglicky: The collection of solved examples from theory of probability and mathematical statistics

Oficiální zadání:
Studentka vytvoří sbírku řešených příkladů z pravděpodobnosti a matematické statistiky. Tato sbírka bude doplněním učebního textu k předmětu M4122.

Literatura:

AGRESTI, Alan a Christine A. FRANKLIN. *Statistics :the art and science of learning from data.* Upper Saddle River, NJ: Pearson Prentice Hall, 2006. xxv, 693 s. ISBN 0-13-045536-9.

FORBELSKÁ, Marie a Jan KOLÁČEK. *Pravděpodobnost a statistika II.* 1. vyd. Brno: Masarykova univerzita, 2013. Elportál. ISBN 978-80-210-6711-0.

FIELD, Andy a Jeremy MILES. *Discovering statistics using R.* : SAGE, 2012. ISBN 978-1-4462-0045-2.

MELOUN, Milan a Jiří MILITKÝ. *Interaktivní statistická analýza dat.* 2012. ISBN 978-80-246-2173-9.

Jazyk závěrečné práce:

Vedoucí práce: Mgr. Jan Koláček, Ph.D.

Datum zadání práce: 2. 6. 2014

V Brně dne: 29. 10. 2014

Souhlasím se zadáním (podpis, datum): 7. 11. 2014

Milena Topalović	Koláček	v. z. Paul Horák
Milena Topalović **studentka**	Mgr. Jan Koláček, Ph.D. **vedoucí práce**	prof. RNDr. Jiří Rosický, DrSc. **ředitel Ústavu matematiky a statistiky**

Poděkování

Mé poděkování patří Mgr. Janu Koláčkovi, Ph.D. za odborné vedení, trpělivost a ochotu, kterou mi v průběhu zpracování bakalářské práce věnoval. Děkuji také Mgr. Janu Karafiátovi za věnovaný čas a pomoc s gramatickou kontrolou práce.

Prohlášení

Prohlašuji, že jsem svoji bakalářskou práci vypracovala samostatně s využitím informačních zdrojů, které jsou v práci citovány.

Brno 27. května 2015 ....

Milena Topalović

Obsah

Úvod

Tato bakalářská práce vnikla především proto, aby poskytla posluchačům předmětu M4122 podrobný postup řešení příkladů podobných těm, které jsou součástí cvičení. Sbírka byla vytvořena jako doplňující učební text, není tedy určena pro samostatné studium. Tvoří ji pět kapitol a celkem 54 podrobně řešených a 33 nevyřešených příkladů s výsledky. Na začátku každé kapitoly je základní seznámení s látkou, která bude v této kapitole probraná. Během kapitoly se postupně věnujeme teorii, která je nezbytná k vyřešení zadaných úloh. Teorie je čerpána ze skript Mgr. Jana Koláčka, Ph.D. a paní RNDr. Marie Forbelské, Ph.D., která jsou uvedena v literatuře. Příklady jsou postupně vyřešeny a vysvětleny. Na konci každé kapitoly se nachází cvičení s úlohami, sloužící studentům k opakování a procvičení probrané látky.

Na úplném začátku první kapitoly se seznámíme se základními definicemi náhodných veličin a jejich funkcemi, dále pak probereme číselné charakteristiky rozdělení pravděpodobností, tedy střední hodnotu, rozptyl, kvantily, kovariance a korelační koeficient. Na konci kapitoly budeme řešit úkoly na charakteristiky náhodných vektorů. Tyto charakteristiky budeme dále používat v celé práci. Druhá kapitola se zabývá centrálními limitními větami: Markovova a Čebyševova nerovnost se používají především v dokazování vět, my však ukážeme, jak lze s jejich pomocí vyřešit určité úlohy. Do této kapitoly patří také centrální limitní věta. Ve třetí kapitole pouze připomeneme definici normálního rozdělení a uvedeme definice rozděleních z něj odvozených, která budeme potřebovat v další části sbírky. Čtvrtou kapitolu věnujeme teorii odhadů, tzn. nestrannosti a konzistenci odhadů, metodám konstrukce bodových odhadů, a na konci internalovým odhadům. Pátá kapitola částečně navazuje na čtvrtou ve smyslu, že při testování statistických hypotéz budeme používat intervalové odhady. Tato kapitola je tedy věnována testnování hypotéz. V příloze jsou pak uvedeny kvantilové tabulky pro určitá rozdělení.

Všechny definice a věty v této práci byly čerpány z [1], [2], [3] a [7]. Inspiraci pro tvorbu některých příladů jsme nalezli v [6], [8], [9] a [10]. Tabulky v příloze byly čerpány z [5]. Bakalářská práce byla zpracována v systému LaTeX, s výjimkou tabulek v příloze práce, které byly zpracovány v programu Excel, a obrázků, které byly zpracovány v programu TikzEdt.

Přehled použitého značení

Pro snazší orientaci v textu zde čtenáři předkládáme přehled značení, které v práci není definované.

$\mathscr{A}$	jevová σ-algebra na Ω
$\mathscr{B}$	borelovská množinová σ-algebra na $\mathbb{R}$
$\mathbb{N}$	množina všech přirozených čísel
$\mathbb{R}$	množina všech reálných čísel
M	spočetná množina reálných čísel
P	pravděpodobnost
ω	elementární jev
Ω	prostor elementárních jevů
Θ	množina možných hodnot parametru θ
$(\Omega, \mathscr{A})$	jevové pole
$(\Omega, \mathscr{A}, P)$	pravděpodobnostní prostor
$A(\theta)$	alternativní rozdělení s parametrem θ
$Bi(n; \theta)$	binomické rozdělení s parametry n a θ
$Po(\lambda)$	Poissonovo rozdělení s parametrem λ
$Ge(\theta)$	geometrické rozdělení s parametrem θ
$NeBi(n; \theta)$	negativně binomické rozdělení s parametry n a θ
$Ro(a; b)$	rovnoměrné rozdělení s parametry a a b
$Ex(\lambda)$	exponenciální rozdělení s parametrem λ

Kapitola 1

Číselné charakteristiky rozdělení pravděpodobností

Distribuční a pravděpodobnostní funkce diskrétní náhodné veličiny a distribuční funkce a hustota náhodné veličiny spojitého typu představují celkové charakteristiky těchto veličin. Nicméně u spousty praktických problémů není ani potřeba charakterizovat náhodnou veličinu v celosti. Většinou nám stačí pouze spočítat některé číselné charakteristiky, což nám významně usnadní práci, a ty nám pak ukazují na důležité vlastnosti náhodných veličin. Postupně se seznámíme se všemi těmito číselnými charakteristikami, ale ještě před tím se seznámíme se základními definicemi náhodných veličin a jejich funkcemi, které je popisují. Připomínám, že zde je uvedená pouze teorie nutná k vyřešení zadaných úloh. Zbývající teorii naleznete v uvedené literatuře.

Definice 1.1. *Nechť $(\Omega, \mathscr{A}, P)$ je pravděpodobnostní prostor, $X : \Omega \to \mathbb{R}$ je takové zobrazení, že pro $\forall x \in \mathbb{R}$ platí*

$$\{\omega \in \Omega : X(\omega) \leq x\} = X^{-1}\big((-\infty, x]\big) \in \mathscr{A}.$$

*Pak X nazýváme **náhodnou veličinou** vzhledem k jevovému poli $(\Omega, \mathscr{A})$.*

Definice 1.2. *Nechť X je náhodná veličina definovaná na pravděpodobnostním prostoru $(\Omega, \mathscr{A}, P)$. Pak funkci*

$$F_X(x) = P(X \leq x),$$

*kde $x \in \mathbb{R}$, nazýváme **distribuční funkcí** náhodné veličiny X.*

Definice 1.3. *Nechť X je náhodná veličina definovaná na pravděpodobnostním prostoru $(\Omega, \mathscr{A}, P)$ a F_X její distribuční funkce. Pak množinovou funkci P_X, definovanou vztahem*

$$P_X(A) = P(X \in A),$$

*kde $A \in \mathscr{B}$ a $\mathscr{B}$ je borelovská množinová σ-algebra na $\mathbb{R}$, nazýváme **rozdělení pravděpodobností** náhodné veličiny X.*

Poznámka: Definici borelovské množinové σ-algebry $\mathscr{B}$ viz ve [1, str. 11].

Definice 1.4. *Řekneme, že náhodná veličina X je* ***diskrétního typu****, pokud existuje nejvýše spočetná množina $M \subset \mathbb{R}$ taková, že platí*

$$P_X(M) = 1,$$

a budeme ji značit $X \sim (M, p_X)$.

Definice 1.5. *Nechť X je diskrétní náhodná veličina. Pak funkci*

$$p_X(x) = P(X = x),$$

kde $x \in M$, nazýváme ***pravděpodobnostní funkcí*** *diskrétní náhodné veličiny X a množinu M* ***oborem hodnot*** *X.*

Definice 1.6. *Řekneme, že náhodná veličina X definovaná na $(\Omega, \mathscr{A}, P)$ je* ***absolutně spojitého typu****, jestliže existuje nezáporná integrovatelná funkce f_X taková, že rozdělení pravděpodobností*

$$P_X(B) = \int_B f_X(x)\mathrm{d}x \text{ pro } \forall B \in \mathscr{B},$$

a budeme je značit $X \sim f_X(x)$. Funkci f_X nazýváme ***hustotou rozdělení pravděpodobností*** *náhodné veličiny X absolutně spojitého typu.*

Když jsme se seznámili s definicemi náhodných veličin, můžeme nyní popsat i jejich číselné charakteristiky. Pojem, který má nejrozšířenější aplikaci především ve statistice, ale i v jiných oborech, je *střední hodnota*. Ostatními významnými pojmy jsou *rozptyl* a *směrodatná odchylka, kovariance, korelační koeficient* a *kvantily*. Všechny tyto vlastnosti mají aplikovatelnost při počítání mnohých úkolů.

1.1 Střední hodnota a rozptyl náhodné veličiny

Střední hodnota je matematická hodnota, která představuje očekávanou (střední, průměrnou) hodnotu náhodné veličiny.

Definice 1.1.1. *Nechť X je náhodná veličina definovaná na $(\Omega, \mathscr{A}, P)$ a nechť existuje integrál $\int\limits_{\Omega} X(\omega)\mathrm{d}P(\omega) < \infty$. Potom číslo*

$$E(X) = \int\limits_{\Omega} X(\omega)\mathrm{d}P(\omega)$$

*nazýváme **střední hodnotou** náhodné veličiny X. Pokud uvedený integrál není konečný nebo neexistuje, říkáme, že střední hodnota náhodné veličiny X neexistuje.*

Věta 1.1.2. *VLASTNOSTI STŘEDNÍ HODNOTY. Nechť X, X_1, X_2 jsou náhodné veličiny definované na pravděpodobnostním prostoru $(\Omega, \mathscr{A}, P)$, $a \in \mathbb{R}$ je konstanta. Potom platí*

1. *Jestliže $P(X = a) = 1 \Leftrightarrow E(X) = a$;*
2. *$E(aX) = aE(X)$;*
3. *$E(X_1 + X_2) = E(X_1) + E(X_2)$;*
4. *Nechť X_1 a X_2 jsou **nezávislé** náhodné veličiny $\Rightarrow E(X_1X_2) = E(X_1)E(X_2)$.*

Důkaz: Vlastnosti střední hodnoty plynou přímo z vlastností integrálů a integrovatelných funkcí.

Důsledek 1.1.3.

1. *Je-li náhodná veličina **diskrétního typu**, potom platí*

$$E(X) = \sum_{x \in M} x p_X(x),$$

za předpokladu, že případná nekonečná řada absolutně konverguje.

2. *Je-li náhodná veličina **absolutně spojitého typu**, potom platí*

$$E(X) = \int\limits_{-\infty}^{\infty} x f_X(x)\mathrm{d}x,$$

za předpokladu, že nekonečný Riemannův integrál absolutně konverguje.

Nyní uvedeme několik příkladů střední hodnoty.

Příklad 1. Kamarád nám nabídl, že si s námi zahraje hru: házení kostkou. V případě, že hodíme sudé číslo, obdržíme od kamaráda 3 Kč. Jestliže hodíme pětku, dáme kamarádovi 1 Kč. Pokud padne jednička, zaplatíme 5 Kč; trojka nikomu zisk nepřinese. Vyplatí se nám hrát takovou hru?

Řešení: Abychom se dozvěděli, zdali v takové hře získáme nějaké peníze, nejjednodušší způsob, jak to zjistit, je, že spočítáme očekávanou (střední) hodnotu zisku. Označme jako X náhodnou veličinu, která udává počet korun, jež získáme z jedné hry. Pravděpodobnost, že hodíme jakékoliv číslo je $\frac{1}{6}$. Lépe to můžeme vidět v následující tabulce:

číslo	1	2	3	4	5	6
X	-5	3	0	3	-1	3
$p_X(x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Tabulka 1.1. Pravděpodobnostní tabulka.

Potom lze střední hodnotu zisku lehce spočítat:

$$E(X) = \sum_{i=1}^{6} x_i p_X(x_i) = (-5)\cdot\frac{1}{6} + 3\cdot\frac{1}{6} + 0\cdot\frac{1}{6} + 3\cdot\frac{1}{6} + (-1)\cdot\frac{1}{6} + 3\cdot\frac{1}{6} = \frac{1}{2} = 0,5.$$

Očekávaný zisk při jedné hře je $0,5$ Kč, tedy 5 Kč při 10 zopakovaných hrách.

Příklad 2. Nechť X a Y jsou náhodné veličiny definované následujícím způsobem:

$$p_X(x) = \begin{cases} 0,1 & \text{pro } x = -10;\ 10, \\ 0,2 & \text{pro } x = 0;\ 5, \\ 0,4 & \text{pro } x = 20, \\ 0 & \text{jinak}, \end{cases}$$

$$p_Y(y) = \begin{cases} 0,1 & \text{pro } y = -6;\ 0, \\ 0,2 & \text{pro } y = 9, \\ 0,3 & \text{pro } y = -2;\ 3, \\ 0 & \text{jinak}. \end{cases}$$

Spočítejte:

a) $E(X)$;

b) $E(X^2)$;

c) $E(3X+2Y)$;

d) $E(X^3 - 4Y^2)$.

Řešení:

a) Stejně jako v předchozím příkladu spočítáme střední hodnotu:

$$E(X) = \sum_{i=1}^{5} x_i p_X(x_i) = (-10)\cdot 0,1 + 10\cdot 0,1 + 0\cdot 0,2 + 5\cdot 0,2 + 20\cdot 0,4 = 9.$$

b) Tento případ se mírně liší od předchozího. Podívejme se jak:

$$E(X^2) = \sum_{i=1}^{5} x_i^2 p_X(x_i) = (-10)^2 \cdot 0,1 + 10^2 \cdot 0,1 + 0^2 \cdot 0,2 + 5^2 \cdot 0,2 + 20^2 \cdot 0,4 = \\ = 165.$$

Takovým způsobem lze samozřejmě spočítat také $E(X^3)$, $E(X^4)$, ...

c) Zde už použijeme vlastnosti střední hodnoty:

$$E(3X + 2Y) = 3E(X) + 2E(Y).$$

Musíme tedy najít střední hodnotu $E(Y)$. Lze lehce spočítat, že $E(Y) = 1,5$. Střední hodnotu $E(X)$ už máme, takže můžeme pokračovat dál:

$$E(3X + 2Y) = 3E(X) + 2E(Y) = 3 \cdot 9 + 2 \cdot 1,5 = 30.$$

d) V tomto případě se jedná o kombinaci dvou předchozích případů:

$$E(X^3 - 4Y^2) = E(X^3) - 4(Y^2).$$

Spočítáme postupně:

$$E(X^3) = \sum_{i=1}^{5} x_i^3 p_X(x_i) = \\ = (-10)^3 \cdot 0,1 + 10^3 \cdot 0,1 + 0^3 \cdot 0,2 + 5^3 \cdot 0,2 + 20^3 \cdot 0,4 = 3225,$$

$$E(Y^2) = \sum_{i=1}^{5} y_i^2 p_Y(y_i) = \\ = (-6)^2 \cdot 0,1 + 0^2 \cdot 0,1 + 9^2 \cdot 0,2 + (-2)^2 \cdot 0,3 + 3^2 \cdot 0,3 = 23,7.$$

Nyní můžeme spočítat:

$$E(X^3 - 4Y^2) = E(X^3) - 4(Y^2) = 3225 - 4 \cdot 23,7 = 3130,2.$$

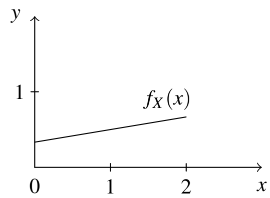
Příklad 3. Doba životnosti X_A opotřebovaného přístroje A (dána v rocích) má rozdělení s hustotou:

$$f_X(x) = \begin{cases} \frac{1}{3}\left(\frac{1}{2}x + 1\right) & \text{pro } 0 \le x \le 2, \\ 0 & \text{jinak.} \end{cases}$$

Ověřte, zda je $f_X(x)$ opravdu hustotou a spočítejte střední dobu životnosti přístroje A.

Řešení:

Nezáporná funkce $f_X(x)$ je hustotou, jestliže $\int\limits_{-\infty}^{\infty} f_X(x)\mathrm{d}x = 1$. Ověřme tento požadavek:



Obrázek 1.1. Hustota.

$$\int\limits_{-\infty}^{\infty} f_X(x)\mathrm{d}x = \int\limits_{0}^{2} \frac{1}{3}\left(\frac{1}{2}x+1\right)\mathrm{d}x = \left[\frac{1}{12}x^2+\frac{1}{3}x\right]_0^2 = \frac{1}{3}+\frac{2}{3} = 1.$$

Funkce $f_X(x)$ je opravdu hustotou. Jak víme z definice, střední hodnotu náhodné veličiny absolutně spojitého typu počítáme pomoci integrálu:

$$E(X_A) = \int\limits_{-\infty}^{\infty} x f_X(x)\mathrm{d}x = \int\limits_{0}^{2} x\cdot\frac{1}{3}\left(\frac{1}{2}x+1\right)\mathrm{d}x = \int\limits_{0}^{2}\left(\frac{1}{6}x^2+\frac{1}{3}x\right)\mathrm{d}x =$$
$$= \left[\frac{1}{18}x^3+\frac{1}{6}x^2\right]_0^2 = \frac{4}{9}+\frac{2}{3} = \frac{10}{9}.$$

Střední doba životnosti přístroje A je tedy $\frac{10}{9}$ roku.

V některých případech nestačí znát pouze střední hodnotu nějaké náhodné veličiny, což můžeme vidět na následujícím příkladu:

Příklad 4. Když řekneme, že průměrná teplota je v nějakém městě 15 °C, máme dojem, že je tam příjemné klima, ale to také může znamenat, že je v létě 40 °C a v zimě -10 °C.

Proto kromě střední hodnoty nějaké náhodné veličiny potřebujeme vědět, jaká je odchylka, tj. jaký je rozptyl možných hodnot náhodné veličiny kolem střední (očekávané) hodnoty.

Uvedeme definici a vlastnosti rozptylu, a potom na několika příkladech ukážeme, jak se počítá a jak ho lze využít v praxi.

Definice 1.1.4. *Nechť X je náhodná veličina definovaná na $(\Omega,\mathscr{A},P)$. Potom číslo*

$$\mu_k = E\big(X - E(X)\big)^k$$

nazýváme k-tým ***centrálním momentem*** *náhodné veličiny X za předpokladu, že uvedené střední hodnoty pro k=1, 2, ... existují.*

Definice 1.1.5.

1. *Řekneme, že náhodná veličina X má konečný* ***druhý obecný moment****, jestliže*

$$\mu_2' = E(X^2) < \infty.$$

2. *Druhý centrální moment nazýváme* ***rozptyl*** *a značíme*

$$D(X) = E\big(X - E(X)\big)^2 = \mu_2.$$

3. *Číslo*

$$\sigma = \sqrt{D(X)}$$

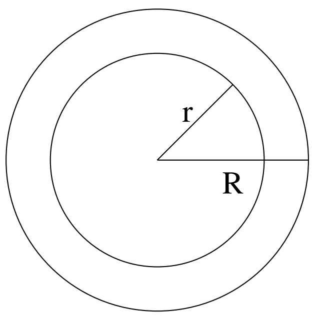
nazýváme ***směrodatnou odchylkou*** *náhodné veličiny X.*

Věta 1.1.6. *VLASTNOSTI ROZPTYLU. Nechť X, X_1, X_2 jsou náhodné veličiny definované na $(\Omega, \mathscr{A}, P)$ s konečnými druhými momenty, $a \in \mathbb{R}$ je konstanta. Potom platí*

1. $P(X = a) = 1 \Rightarrow D(X) = 0$;
2. $D(X) = E(X^2) - E^2(X)$;
3. $D(aX) = a^2 D(X)$;
4. $D(X_1 + X_2) = D(X_1) + D(X_2)$, *přičemž X_1 a X_2 jsou* ***nezávislé*** *náhodné veličiny.*

Důkaz: Viz [1, str. 75].

Příklad 5. Honza se rozhodl vyzkoušet své schopnosti ve střelbě lukem na terč. Pravděpodobnost, že netrefí terč, a tudíž nezíská žadné body je 50 %. Terč je rozdělen tak, že poloměr vnitřního kruhu r je $\frac{\sqrt{2}}{2}R$, kde R je poloměr terče. Vnitřní část (kruh) přináší dva body, vnější část jeden bod. Spočítejte očekávaný počet bodů při deseti pokusech a také o kolik se odchýlí možné hodnoty od očekávané.



Obrázek 1.2. Terč.

Řešení: Nechť X je náhodná veličina, která představuje počet bodů. Nejprve musíme zjistit, jaký je poměr vnitřní a vnější části. Spočítáme obě plochy, když víme, že $r = \frac{R\sqrt{2}}{2}$:

$$P_1 = r^2\pi = R^2\frac{\pi}{2},$$
$$P_2 = R^2\pi - r^2\pi = R^2\pi - R^2\frac{\pi}{2} = R^2\frac{\pi}{2}.$$

Vidíme tudíž, že se plochy rovnají, což znamená, že je pravděpodobnost zásahu obou částí stejná. Rozepíšeme tedy pravděpodobnosti dosažení každého počtu bodů:

$$p_X(x) = \begin{cases} 0,5 & \text{pro } x = 0, \\ 0,25 & \text{pro } x = 1, 2, \\ 0 & \text{jinak.} \end{cases}$$

Jak jsme už viděli, očekávaný počet bodů spočítáme pomocí definice střední hodnoty:

$$E(X) = \sum_{i=1}^{3} x_i p_X(x_i) = 0 \cdot 0,5 + 1 \cdot 0,25 + 2 \cdot 0,25 = 0,75.$$

Očekávaný počet bodů při 10 pokusech je 7,5. Dále podle definice spočítáme rozptyl, abychom z něj dostali hledanou odchylku:

$$D(X) = E\big(X - E(X)\big)^2 = E(X^2) - E^2(X) =$$
$$= 0^2 \cdot 0,5 + 1^2 \cdot 0,25 + 2^2 \cdot 0,25 + 0,75^2 = 1,25 + 0,5625 = 1,8125.$$

Rozptyl označuje kvadratickou odchylku od střední hodnoty. Směrodatná odchylka je:

$$\sigma = \sqrt{D(X)} = \sqrt{1,8125} = 1,34629.$$

Příklad 6. Spočítejte $D(3X+2)$ náhodné veličiny X definované v příkladu 3 na straně 5.

Řešení: Použijeme vlastnost rozptylu:

$$D(3X+2) = 9D(X).$$

Potřebujeme tedy spočítat pouze rozptyl $D(X)$. Abychom jej spočítali, musíme nejprve vypočítat $E(X^2)$:

$$E(X^2) = \int_0^2 x^2 \cdot \frac{1}{3}\left(\frac{1}{2}x + 1\right)\mathrm{d}x = \int_0^2 \left(\frac{1}{6}x^3 + \frac{1}{3}x^2\right)\mathrm{d}x =$$
$$= \left[\frac{1}{24}x^4 + \frac{1}{9}x^3\right]_0^2 = \frac{2}{3} + \frac{8}{9} = \frac{14}{9},$$

$$D(3X+2) = 9D(X) = 9\big(E(X^2) - E^2(X)\big) = 9\left(\frac{14}{9} - \left(\frac{10}{9}\right)^2\right) = \frac{26}{9}.$$

Nyní definujeme následující funkci a několik jejich vlastností, které nám významně usnadní počítání mnohých úkolů, což můžeme vidět na příkladu č. 7 na straně 9.

Definice 1.1.7. *Funkce Γ je pro $a > 0$ definovaná předpisem*

$$\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} \mathrm{d}x.$$

Věta 1.1.8. *VLASTNOSTI Γ FUNKCE. Její nejčastěji používané vlastnosti pro $a > 0$, $n \in \mathbb{N}$ jsou*

1. $\Gamma(a+1) = a\Gamma(a)$;
2. $\Gamma\left(\frac{1}{2}\right) = \pi$;
3. $\Gamma(n) = (n-1)!$.

Důkaz: Viz [1, str. 38].

Příklad 7. Doba X do vybití baterie určovaná v rocích se řídí rozdělením s hustotou:

$$f_X(x) = \begin{cases} 2e^{-2x} & \text{pro } x > 0, \\ 0 & \text{pro } x \leq 0. \end{cases}$$

Spočítejte její střední hodnotu životnosti.

Řešení:

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) \mathrm{d}x = \int_0^\infty 2x e^{-2x} \mathrm{d}x \left| \begin{matrix} 2x = t \\ 2\mathrm{d}x = \mathrm{d}t \end{matrix} \right| = \frac{1}{2} \int_0^\infty t e^{-t} \mathrm{d}t = \frac{1}{2}\Gamma(2) = \frac{1}{2} \cdot 1! = \frac{1}{2}.$$

Střední hodnota její životnosti je tedy půl roku.

V praxi se opakovaně setkáváme s některými rozděleními diskrétních a spojitých náhodných veličin. Znalost jejich středních hodnot a rozptylů může mít velký význam při řešení mnohých úloh. Některá rozdělení nemají střední hodnotu ani rozptyl, což ukážeme na následujícím příkladu. Poté za účelem snadnějšího počítání uvedeme již zmíněné charakteristiky.

Příklad 8. Dokažte, že střední hodnota a rozptyl standardního Cauchyho rozdělení neexistují.

Řešení: Cauchyho rozdělení pravděpodobnosti je definováno hustotou v následujícím tvaru:

$$f_X(x;a;b) = \frac{1}{\pi\big(b^2+(x-a)^2\big)} \quad \text{pro } x \in \mathbb{R},$$

přičemž $a \in \mathbb{R},\ b > 0$ jsou daná čísla.

Pro $a = 0$ a $b = 1$ dostaneme tzv. standardní Cauchyho rozdělení, které je speciálním případem Studentova t-rozdělení s jedním stupněm volnosti. Zapisujeme $X \sim t(1)$, anebo $X = \frac{U_1}{U_2}$, kde $U_i \sim N(0,1)$ pro $i = 1,2$ (více ve 3. kapitole). Potom dostáváme hustotu vyjádřenou vztahem:

$$f_X(x) = \frac{1}{\pi(1+x^2)} \quad \text{pro } x \in \mathbb{R}.$$

Počítáme tedy její střední hodnotu a rozptyl:

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)\mathrm{d}x = \int_{-\infty}^{\infty} \frac{x}{\pi(1+x^2)}\mathrm{d}x.$$

Jedná se tedy o lichou funkci. Jeji integrál se rovná nule, kdy $x \in (-a,a),\ a \in \mathbb{R}$, ale neexistuje, pokud a je nevlastní bod. Pokud daný integrál neexistuje, plyne z toho, že hledaná střední hodnota náhodné veličiny a zároveň její rozptyl neexistují.

$X \sim$	$A(\theta)$	$Bi(n;\theta)$	$Po(\lambda)$	$Ge(\theta)$	$NeBi(n;\theta)$
$E(X)$	θ	$n\theta$	λ	$\frac{1-\theta}{\theta}$	$n\frac{1-\theta}{\theta}$
$D(X)$	$\theta(1-\theta)$	$n\theta(1-\theta)$	λ	$\frac{1-\theta}{\theta^2}$	$n\frac{1-\theta}{\theta^2}$

Tabulka 1.2. Střední hodnota a rozptyl důležitých diskrétních rozdělení.

$X \sim$	$Ro(a;b)$	$Ex(\lambda)$	$N(\mu;\sigma^2)$	$\chi^2(n)$	$t(n)$	$F(n_1;n_2)$
$E(X)$	$\frac{a+b}{2}$	$\frac{1}{\lambda}$	μ	n	0 [1]	$\frac{n_1}{n_1-2}$ [2]
$D(X)$	$\frac{(b-a)^2}{12}$	$\frac{1}{\lambda^2}$	σ^2	$2n$	$\frac{n}{n-2}$ [3]	$\frac{2n_1^2(n_1+n_2-2)}{n_2(n_1-2)^2(n_1-4)}$ [4]

Tabulka 1.3. Střední hodnota a rozptyl důležitých spojitých rozdělení.

Příklad 9. Nechť X_1 a X_2 jsou stochasticky nezávislé náhodné veličiny z exponenciálního rozdělení. Vypočítejte střední hodnotu náhodné veličiny Y, kde $Y = min\{X_1, X_2\}$.

Řešení: Víme, že exponenciální rozdělení má následující rozdělení a distribuční funkce:

$$f_{X_i}(x_i) = \begin{cases} \lambda e^{-\lambda x_i} & \text{pro } x_i > 0, \\ 0 & \text{jinak,} \end{cases} \qquad \text{a} \qquad F_{X_i}(x_i) = \begin{cases} 1 - e^{-\lambda x_i} & \text{pro } 0 < x_i < 1, \\ 0 & \text{jinak.} \end{cases}$$

kde $i = 1, 2$. Nyní můžeme počítat:

$$\begin{aligned} F_Y(y) &= P(Y < y) = P(min\{X_1, X_2\} < y) = 1 - P\big(min\{X_1, X_2\} > y\big) = \\ &= 1 - P(X_1 > y \wedge X_2 > y) = 1 - P(X_1 > y) \cdot P(X_2 > y) = \\ &= 1 - \big(1 - P(X_1 < y)\big) \cdot \big((1 - P(X_2 < y)\big) = 1 - \big(1 - F_{X_1}(y)\big) \cdot \big(1 - F_{X_2}(y)\big) = \\ &= 1 - \big(1 - (1 - e^{-\lambda y})\big) \cdot \big(1 - (1 - e^{-\lambda y})\big) = 1 - e^{-\lambda y} \cdot e^{-\lambda y} = 1 - e^{-2\lambda y}. \end{aligned}$$

Z toho plyne, že:

$$f_Y(y) = F_Y'(y) = (1 - e^{-2\lambda y})' = 2\lambda e^{-2\lambda y}.$$

Konečně můžeme spočítat střední hodnotu:

$$\begin{aligned} E(Y) &= \int_{-\infty}^{\infty} y f_Y(y) \mathrm{d}y = \int_0^{\infty} 2\lambda y e^{-2\lambda y} \left| \begin{matrix} 2\lambda y = t \\ 2\lambda \mathrm{d}y = \mathrm{d}t \end{matrix} \right| = \\ &= \frac{1}{2\lambda} \int_0^{\infty} t e^{-t} \mathrm{d}t = \frac{1}{2\lambda} \Gamma(2) = \frac{1}{2\lambda} 1! = \frac{1}{2\lambda}. \end{aligned}$$

Využitím Γ funkce jsme dostali střední hodnotu rovnou $\frac{1}{2\lambda}$.

[1] Platí pouze pro $n > 1$, jinak neexistuje.
[2] Platí pouze pro $n_1 > 2$, jinak neexistuje.
[3] Platí pouze pro $n > 2$, jinak neexistuje.
[4] Platí pouze pro $n_1 > 4$, jinak neexistuje.

1.2 Kvantily

V této podkapitole uvedeme jednak definici kvantilu, jednak kvantily, které jsou velmi často používané: medián, dolní a horní kvartil.

Definice 1.2.1. *Nechť F_X je distribuční funkcí a $\alpha \in (0,1)$. Potom funkce*

$$F_X^{-1}(\alpha) = Q(\alpha) = \inf\{x \in \mathbb{R} : F_X(x) \geq \alpha\}.$$

se nazývá ***kvantilová funkce*** *a číslo*

$$x_\alpha = Q(\alpha)$$

se nazývá ***α-kvantilem*** *rozdělení s distribuční funkcí $F_X(x)$.*

Poznámka 1.2.2. *Mezi často používané kvantily patří*

1. $x_{0,25}$ – ***dolní kvartil****;*
2. $x_{0,5}$ – ***medián****;*
3. $x_{0,75}$ – ***horní kvartil****.*

Odhady kvantilů jsou dobře využitelné v matematické statistice. Zde se budeme zabývat spíše „teoretickými" příklady. Příklady na odhady kvantilů budeme řešit v kapitole 4.

Příklad 10. Kuba letí letadlem z města A do města B, přičemž obě města jsou na stejné zeměpisné výšce. Město A se nachází na zeměpisné šířce 6° a město B na zeměpisné šířce 49°. Využitím definice kvantilu spočítejte, na jaké zeměpisné šířce se Kuba nachází, pokud letadlo uletělo 17 % cesty.

Řešení: Vzhledem k tomu, že se města nachází na stejné zeměpisné výšce, hustotu vzdáleností mezi těmito městy si můžeme představit funkcí rovnoměrného rozdělení:

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & \text{pro } x \in (a,b),\ a < b, \\ 0 & \text{jinak.} \end{cases}$$

Víme, že její distribuční funkce má tvar:

$$F_X(x) = \begin{cases} 0 & \text{pro } x \leq a, \\ \dfrac{x-a}{b-a} & \text{pro } x \in (a,b),\ a < b, \\ 1 & \text{pro } x \geq b. \end{cases}$$

Podle definice kvantilu je α-kvantil roven inverzní distribuční funkci v bodě α. Zapisujeme

$$F_X^{-1}(\alpha) = x_\alpha = Q(\alpha),$$

z čehož plyne, že

$$F_X(x_\alpha) = \alpha.$$

Nyní můžeme pokračovat ve výpočtu:

$$\begin{aligned}
&F_X(x_\alpha) = \frac{x_\alpha - a}{b-a} = \alpha,\\
&x_\alpha - a = \alpha(b-a),\\
&x_\alpha = \alpha(b-a)+a.
\end{aligned}$$

My však potřebujeme spočítat, kde se nachází Kuba, když urazil 17 % cesty, tzn. sedmnáctý kvantil:

$$\begin{aligned}
&x_{0,17} = 0,17(49-6)+6,\\
&x_{0,17} = 13,31.
\end{aligned}$$

Vidíme, že po přeletu 17 % cesty se Kuba nachází na zeměpisné šířce $13,31^\circ$.

Příklad 11. Spočítejte medián, dolní a horní kvartil náhodné veličiny X s následující distribuční funkcí $F_X(x)$:

$$F_X(x) = \begin{cases} x - \frac{1}{4}x^2 & \text{pro } x \in (0,2),\\ 0 & \text{pro } x \in (-\infty, 0],\\ 1 & \text{pro } x \in [2,\infty). \end{cases}$$

Řešení: Nejprve spočítáme kvantil ve všeobecném tvaru, potom za α dosadíme odpovídající čísla:

$$\begin{aligned}
&F_X(x_\alpha) = \alpha,\\
&x_\alpha - \frac{1}{4}x_\alpha^2 = \alpha,\\
&1 - \left(1 - \frac{1}{2}x_\alpha\right)^2 = \alpha,\\
&\left(1 - \frac{1}{2}x_\alpha\right)^2 = 1 - \alpha,\\
&1 - \frac{1}{2}x_\alpha = \pm\sqrt{1-\alpha},\\
&x_\alpha = 2 \pm 2\sqrt{1-\alpha}.
\end{aligned}$$

Protože počítáme funkce za $x \in (0, 2)$, x nemůže být větší jak 2. Tudíž nám zůstává pouze jedno řešení:

$$x_\alpha = 2 - 2\sqrt{1-\alpha}.$$

Nyní můžeme dosadit za α čísla, jejichž kvantily počítáme:

$$x_{0,25} = 2 - 2\sqrt{1-0,25} \approx 0,268,$$
$$x_{0,5} = 2 - 2\sqrt{1-0,5} \approx 0,586,$$
$$x_{0,75} = 2 - 2\sqrt{1-0,75} = 1.$$

Za α lze samozřejmě v případě potřeby dosadit jakékoliv číslo z intervalu $(0, 1)$.

1.3 Kovariance a korelační koeficient

V této podkapitole se budeme zabývat kovariancí a korelačním koeficientem (dále jen korelace). Kovariance je střední hodnota součinu odchylek dvou náhodných veličin od jejich středních hodnot, zatímco korelace ukazuje na stupeň závislosti dvou náhodných veličin.

V celé této podkapitole budeme předpokládat, že náhodné veličiny mají konečné druhé momenty.

Definice 1.3.1. Kovariancí *dvou náhodných veličin X a Y nazýváme číslo $C(X,Y) = E\big(X-E(X)\big)\big(Y-E(Y)\big)$ a číslo $R(X,Y) = \dfrac{C(X,Y)}{\sqrt{D(X)D(Y)}}$ nazýváme **korelační koeficient**.*

Věta 1.3.2.

1. *Nechť $(X,Y)' \sim (M, p_{X,Y}(x,y))$ jsou náhodné veličiny **diskrétního typu**, potom platí*

$$C(X,Y) = \sum_{(x,y)\in M} \big(x-E(X)\big)\big(y-E(Y)\big)p_{X,Y}(x,y).$$

2. *Nechť $(X,Y)' \sim f_{X,Y}(x,y)$ jsou náhodné veličiny **absolutně spojitého typu**, potom platí*

$$C(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \big(x-E(X)\big)\big(y-E(Y)\big)\mathrm{d}F_{X,Y}(x,y).$$

Důkaz: Viz [1, str. 77].

Věta 1.3.3. *VLASTNOSTI KOVARIANCE A KORELACE. Nechť X a Y jsou náhodné veličiny,* $a_1, a_2, b_1, b_2 \in \mathbb{R}$ *jsou koeficienty. Potom platí*

1. $C(X,X) = D(X)$;
   $R(X,X) = 1$;
2. $C(X,Y) = C(Y,X)$;
   $R(X,Y) = R(Y,X)$;
3. $C(X,Y) = E(XY) - E(X)E(Y)$;
4. *Jsou-li náhodné veličiny X a Y nezávislé, pak* $C(X,Y) = R(X,Y) = 0$;
   POZOR: Obráceně neplatí!
5. $|C(X,Y)| \leq \sqrt{D(X)D(Y)}$;
   $|R(X,Y)| \leq 1$;
6. $C(a_1 + a_2X, b_1 + b_2Y) = a_2b_2C(X,Y)$;
7. $R(a_1 + a_2X, b_1 + b_2Y) = R(X,Y)sign(a_2b_2)$, *je-li* $a_1 \neq 0$ *a* $b_1 \neq 0$;
8. $D(X+Y) = D(X) + D(Y) + 2C(X,Y)$.

Důkaz: Viz [1, str. 79].

Z vlastností kovariancí a korelací můžeme vidět, že korelace nabývá hodnoty v intervalu $[-1,1]$. Korelace v hodnotě -1 znamená zcela nepřímou závislost, naproti tomu hodnota 1 znamená, že veličiny jsou ve zcela přímé závislosti. Avšak nulová korelace ukazuje na to, že mezi statistickými veličinami neexistuje žádná lineární závislost. Také vidíme, že nezávislost náhodných veličin implikuje nulovou kovarianci, zatímco nulová kovariance neimplikuje nezávislost náhodných veličin, ale nepřítomnost jakéhokoliv lineárního vztahu mezi nimi.

Předtím, než začneme řešit příklady na kovarianci a korelaci, uvedeme definici náhodného vektoru:

Definice 1.3.4. *Nechť* $\boldsymbol{X} = (X_1, ..., X_n)' : \Omega \to \mathbb{R}^n$ *je zobrazení definované na pravděpodobnostním prostoru* $(\Omega, \mathscr{A}, P)$ *takové, že pro* $\forall \boldsymbol{x} \in \mathbb{R}^n$ *platí*

$$\{\omega \in \Omega : \boldsymbol{X}(\omega) \leq \boldsymbol{x}\} \in \mathscr{A}.$$

Pak $\boldsymbol{X}$ *nazýváme* ***n-rozměrným náhodným vektorem****.*

Podívejme se na několik příkladů kovariancí a korelací:

Příklad 12. Spočítejte kovarianci mezi náhodnými veličinami U a V, jestliže $U = X + Y$ a $V = Y - X$, přičemž X a Y jsou nezávislé náhodné veličiny.

Řešení: Z nezávislosti dvou náhodných veličin plyne nulová kovariance. Postupně počítáme:

$$
\begin{aligned}
R(U,V) &= \frac{C(U,V)}{\sqrt{D(U)D(V)}},\\
C(U,V) &= R(U,V)\sqrt{D(U)D(V)} = R(X+Y,Y-X)\sqrt{D(X+Y)D(Y-X)} =\\
&= \big(R(X,Y)-R(X,X)+R(Y,Y)-R(Y,X)\big)\sqrt{\big(D(X)+D(Y)\big)^2} =\\
&= (1-1)\cdot\big(D(X)+D(Y)\big) = 0.
\end{aligned}
$$

Tady vidíme, že mezi náhodnými veličinami U a V neexistuje lineární vztah.

Ještě předtím, než se dostaneme k následujícímu příkladu, připomeňme si dvě věty. První popisuje marginální pravděpodobnostní funkci a marginální hustotu, druhá jejich vlastnosti v případě nezávlislosti náhodných veličin, kterou v následujícím příkladu budeme používat.

Věta 1.3.5. *Pro přirozené $k < n$ mějme indexy $\{i_1, ..., i_k\} \subset \{1, ..., n\}$ a $\{j_1, ..., j_{n-k}\} = \{1, ..., n\} \setminus \{i_1, ..., i_k\}$.*

1. *Nechť $\mathbf{X} \sim (M, p_{\mathbf{X}})$. Pak marginální náhodný vektor $\mathbf{X}^*$ má marginální pravděpodobnostní funkci rovnu*

   $$p^*_{\mathbf{X}}(\mathbf{x}^*) = p^*_{\mathbf{X}}(x_{i_1}, ..., x_{i_k}) = P(\mathbf{X}^* = \mathbf{x}^*) = \sum_{x_{j_1}\in M_{j_1}} \cdots \sum_{x_{j_{n-k}}\in M_{j_{n-k}}} p_{\mathbf{X}}(x_1, ..., x_n),$$

   kde $M = M_1 \times \cdots \times M_n$, přičemž M_i je obor hodnot náhodné veličiny X_i, pro $i = 1, ..., n$.

2. *Nechť $\mathbf{X} \sim f_{\mathbf{X}}(\mathbf{x})$. Pak marginální náhodný vektor $\mathbf{X}^*$ má marginální hustotu rovnu*

   $$f^*_{\mathbf{X}}(\mathbf{x}^*) = f^*_{\mathbf{X}}(x_{i_1}, ..., x_{i_k}) = P(\mathbf{X}^* = \mathbf{x}^*) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\mathbf{X}}(x_1, ..., x_n)\mathrm{d}x_{j_1} ... \mathrm{d}x_{j_{n-k}}.$$

Důkaz: Viz [1, str. 47].

Věta 1.3.6.

1. *Mějme diskrétní náhodný vektor* $\boldsymbol{X} = (X_1, ..., X_n)' \sim (M, p_{\mathbf{X}})$. *Pak* $X_1, ..., X_n$ *jsou nezávislé, právě když*

$$p_{\mathbf{X}}(x_1, ..., x_n) = \prod_{i=1}^{n} p_{X_i}(x_i) \quad \textit{pro} \quad \forall \mathbf{x} = (x_1, ..., x_n)' \in \mathbb{R}^n,$$

kde pro $i = 1, ..., n$ *je* $p_{X_i}(x_i)$ *marginální pravděpodobnostní funkce náhodné veličiny* X_i.

2. *Mějme absolutně spojitý náhodný vektor* $\boldsymbol{X} = (X_1, ..., X_n)'$ *se sdruženou hustotou* $f_{\mathbf{X}}(x_1, ..., x_n)$. *Pak* $X_1, ..., X_n$ *jsou nezávislé, právě když*

$$f_{\mathbf{X}}(x_1, ..., x_n) = \prod_{i=1}^{n} f_{X_i}(x_i) \quad \textit{pro} \quad \textit{s.v.} \quad \mathbf{x} = (x_1, ..., x_n)' \in \mathbb{R}^n,$$

kde pro $i = 1, ..., n$ *je* $f_{X_i}(x_i)$ *marginální hustota náhodné veličiny* X_i.

Důkaz: Viz [1, str. 50].

Příklad 13. Máme k dispozici dvě kostky, kterými házíme ve stejný čas: červenou a modrou. Pokud červenou kostkou hodíme sudé číslo, obdržíme za něj jeden bod, zatímco za liché číslo dostaneme 2 body. Když házíme modrou kostkou a hodíme číslo 1, obdržíme 1 bod, pokud hodíme 2 nebo 3, obdržíme 2 body a za čísla 4, 5 nebo 6 dostaneme 3 body. Pravděpodobnostní funkce náhodného vektoru $\mathbf{X} = (X, Y)'$, kde první složka označuje počet bodů dosažených červenou kostkou a druhá složka ukazuje na počet bodů získaných modrou kostkou, počítá pravděpodobnost dosažení bodů jedním hodem. Dokažte, že mezi těmito dvěma veličinami neexistuje žádná lineární závislost.

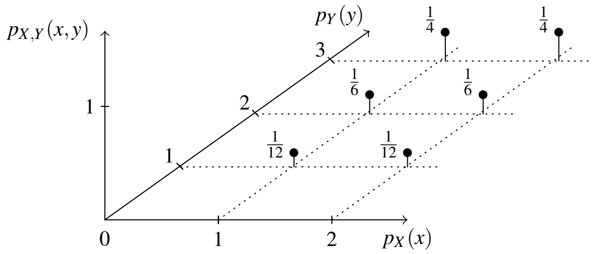
Řešení: Máme dokázat, že zadané náhodné veličiny jsou lineárně nezávislé mezi sebou. Potom platí:

$$p_{X,Y}(x, y) = p_X(x) p_Y(y).$$

Hodnoty pravděpodobnostní funkce ukážeme nejlépe pomocí pravděpodobnostní tabulky:

$X \backslash Y$	1	2	3	$p_X(x)$
1	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{2}$
2	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{2}$
$p_Y(y)$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$	1

Tabulka 1.4. Pravděpodobnostní tabulka.



Obrázek 1.3. Pravděpodobnostní funkce.

Zapíšeme zvlášť pravděpodobnostní funkce:

$$p_X(x) = \begin{cases} \frac{1}{2} & \text{pro } x = 1, \\ \frac{1}{2} & \text{pro } x = 2, \\ 0 & \text{jinak,} \end{cases} \qquad \text{a} \qquad p_Y(y) = \begin{cases} \frac{1}{6} & \text{pro } y = 1, \\ \frac{1}{3} & \text{pro } y = 2, \\ \frac{1}{2} & \text{pro } y = 3, \\ 0 & \text{jinak.} \end{cases}$$

Abychom určili závislost mezi veličinami, musíme spočítat jejich korelaci. Jak jsme již viděli na předchozím příkladu, předtím než spočítáme korelaci, musíme nejprve spočítat jejich střední hodnoty, rozptyly a kovarianci:

$$E(X) = \sum_{i=1}^{2} x_i p_X(x_i) = 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{2} = \frac{3}{2},$$

$$E(Y) = \sum_{j=1}^{3} y_j p_Y(y_j) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{3} + 3 \cdot \frac{1}{2} = \frac{7}{3},$$

$$\begin{aligned} E(XY) &= \sum_{i=1}^{2} \sum_{j=1}^{3} x_i y_j p_{X,Y}(x_i, y_j) = \\ &= 1 \cdot 1 \cdot \frac{1}{12} + 1 \cdot 2 \cdot \frac{1}{6} + 1 \cdot 3 \cdot \frac{1}{4} + 2 \cdot 1 \cdot \frac{1}{12} + 2 \cdot 2 \cdot \frac{1}{6} + 2 \cdot 3 \cdot \frac{1}{4} = \frac{7}{2}, \end{aligned}$$

$$C(X,Y) = E(XY) - E(X)E(Y) = \frac{7}{2} - \frac{3}{2} \cdot \frac{7}{3} = 0 \Longrightarrow R(X,Y) = 0.$$

Případ nulové kovariance nám implikuje nulovou korelaci, čímž můžeme dokázat, že mezi danými veličinami neexistuje lineární závislost, aniž bychom počítali ostatní veličiny.

Příklad 14. Spočítejte konstantu c tak, aby náhodné veličiny X a Y byly ve zcela nepřímé závislosti. Náhodný vektor má následující simultánní hustotu:

$$f_{X,Y}(x,y)=\begin{cases}\frac{1}{4}\left(e^{-(1-x)}+cy\right) & \text{pro } x,y\in[-1,1],\\ 0 & \text{jinak.}\end{cases}$$

Řešení: Už bylo zmíněno, že zcela nepřímá závislost dvou náhodných veličin znamená, že korelace mezi těmito dvěma veličinami je -1. Abychom byli schopni spočítat korelaci, nejprve musíme spočítat střední hodnoty, rozptyl a kovarianci mezi těmito náhodnými veličinami, ale ještě před tím spočítáme jejich marginální hustoty:

$$f_X(x)=\int_{-\infty}^{\infty}f_{X,Y}(x,y)\mathrm{d}y=\int_{-1}^{1}\frac{1}{4}\left(e^{-(1-x)}+cy\right)\mathrm{d}y=\frac{1}{4}\int_{-1}^{1}e^{x-1}\mathrm{d}y+\frac{1}{4}\int_{-1}^{1}cy\mathrm{d}y=$$
$$=\frac{1}{4}\left[e^{x-1}y\right]_{-1}^{1}+\frac{1}{4}\left[\frac{cy^2}{2}\right]_{-1}^{1}=\frac{e^{x-1}}{4}+\frac{e^{x-1}}{4}+\frac{c}{8}-\frac{c}{8}=\frac{e^{x-1}}{2},$$

$$f_Y(y)=\int_{-\infty}^{\infty}f_{X,Y}(x,y)\mathrm{d}x=\int_{-1}^{1}\frac{1}{4}\left(e^{-(1-x)}+cy\right)\mathrm{d}x=\frac{1}{4}\int_{-1}^{1}e^{x-1}\mathrm{d}x+\frac{1}{4}\int_{-1}^{1}cy\mathrm{d}x=$$
$$=\frac{1}{4}\left[e^{x-1}\right]_{-1}^{1}+\frac{1}{4}\left[cyx\right]_{-1}^{1}=\frac{1}{4}-\frac{1}{4e^2}+\frac{cy}{4}+\frac{cy}{4}=\frac{e^2+2e^2cy-1}{4e^2}.$$

Teď, když známe jejich marginální hustoty, můžeme spočítat jejich střední hodnoty a rozptyly:

$$E(X)=\int_{-\infty}^{\infty}xf_X(x)\mathrm{d}x=\int_{-1}^{1}\frac{xe^{x-1}}{2}\mathrm{d}x=\frac{1}{2}\int_{-1}^{1}xe^{x-1}\mathrm{d}x\left|\begin{matrix}u=x & v'=e^{x-1}\\ u'=1 & v=e^{x-1}\end{matrix}\right|=$$
$$=\frac{1}{2}\left(xe^{x-1}\Big|_{-1}^{1}-\int_{-1}^{1}e^{x-1}\mathrm{d}x\right)=\frac{1}{2}\left(1+\frac{1}{e^2}\right)-\frac{1}{2}\left[e^{x-1}\right]_{-1}^{1}=$$
$$=\frac{e^2+1}{2e^2}-\frac{1}{2}\left(1-\frac{1}{e^2}\right)=\frac{e^2+1}{2e^2}-\frac{e^2-1}{2e^2}=\frac{1}{e^2},$$

$$E(X^2)=\int_{-\infty}^{\infty}x^2f_X(x)\mathrm{d}x=\frac{1}{2}\int_{-1}^{1}x^2e^{x-1}\mathrm{d}x\left|\begin{matrix}u=x^2 & v'=e^{x-1}\\ u'=2x & v=e^{x-1}\end{matrix}\right|=$$
$$=\frac{1}{2}\left(x^2e^{x-1}\Big|_{-1}^{1}-\int_{-1}^{1}2xe^{x-1}\mathrm{d}x\right)=\frac{1}{2}\left(1-\frac{1}{e^2}\right)-\int_{-1}^{1}xe^{x-1}\mathrm{d}x=$$
$$=\frac{1}{2}-\frac{1}{2e^2}-\frac{2}{e^2}=\frac{e^2-5}{2e^2},$$

$$E(Y)=\int_{-\infty}^{\infty} y f_Y(y)\mathrm{d}y=\int_{-1}^{1}\frac{y(e^2+2e^2cy-1)}{4e^2}\mathrm{d}y=\int_{-1}^{1}\frac{y}{4}\mathrm{d}y+\int_{-1}^{1}\frac{cy^2}{2}\mathrm{d}y-\int_{-1}^{1}\frac{y}{4e^2}\mathrm{d}y=$$
$$=\left[\frac{y^2}{8}\right]_{-1}^{1}+\left[\frac{cy^3}{6}\right]_{-1}^{1}-\left[\frac{y^2}{8e^2}\right]_{-1}^{1}=\frac{1}{8}-\frac{1}{8}+\frac{c}{6}+\frac{c}{6}-\frac{1}{8e^2}+\frac{1}{8e^2}=\frac{c}{3},$$

$$E(Y^2)=\int_{-\infty}^{\infty} y^2 f_Y(y)\mathrm{d}y=\int_{-1}^{1}\frac{y^2(e^2+2e^2cy-1)}{4e^2}\mathrm{d}y=$$
$$=\int_{-1}^{1}\frac{y^2}{4}\mathrm{d}y+\int_{-1}^{1}\frac{cy^3}{2}\mathrm{d}y-\int_{-1}^{1}\frac{y^2}{4e^2}\mathrm{d}y=$$
$$=\left[\frac{y^3}{12}\right]_{-1}^{1}+\left[\frac{cy^4}{8}\right]_{-1}^{1}-\left[\frac{y^3}{12e^2}\right]_{-1}^{1}=\frac{1}{12}+\frac{1}{12}+\frac{c}{8}-\frac{c}{8}-\frac{1}{12e^2}-\frac{1}{12e^2}=$$
$$=\frac{e^2-1}{6e^2},$$

$$E(XY)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y=\int_{-1}^{1}\int_{-1}^{1}\frac{1}{4}xy\left(e^{x-1}+cy\right)\mathrm{d}x\mathrm{d}y=$$
$$=\int_{-1}^{1}\left[\int_{-1}^{1}\frac{1}{4}xy\left(e^{x-1}+cy\right)\mathrm{d}y\right]\mathrm{d}x=\int_{-1}^{1}\left(\int_{-1}^{1}\frac{xe^{x-1}y}{4}\mathrm{d}y+\int_{-1}^{1}\frac{cxy^2}{4}\mathrm{d}y\right)\mathrm{d}x=$$
$$=\int_{-1}^{1}\left\{\left[\frac{xe^{x-1}y^2}{8}\right]_{-1}^{1}+\left[\frac{cxy^3}{12}\right]_{-1}^{1}\right\}\mathrm{d}x=\int_{-1}^{1}\left(\frac{xe^{x-1}}{8}-\frac{xe^{x-1}}{8}+\frac{cx}{12}+\frac{cx}{12}\right)\mathrm{d}x=$$
$$=\int_{-1}^{1}\frac{cx}{6}\mathrm{d}x=\left[\frac{cx^2}{12}\right]_{-1}^{1}=\frac{c}{12}-\frac{c}{12}=0,$$

$$D(X)=E(X^2)-E^2(X)=\frac{e^2-5}{2e^2}-\left(\frac{1}{e^2}\right)^2=\frac{e^4-5e^2-2}{2e^4},$$

$$D(Y)=E(Y^2)-E^2(Y)=\frac{e^2-1}{6e^2}-\left(\frac{c}{3}\right)^2=\frac{3e^2-2e^2c^2-3}{18e^2}.$$

Nyní můžeme snadno spočítat kovarianci a korelaci mezi nimi:

$$C(X,Y)=E(XY)-E(X)E(Y)=0-\frac{1}{e^2}\cdot\frac{c}{3}=-\frac{c}{3e^2}.$$

Teď konečně spočítáme korelaci, z čehož plyne hledaná konstanta c:

$$R(X,Y) = \frac{C(X,Y)}{\sqrt{D(X)D(Y)}} = \frac{-\frac{c}{3e^2}}{\sqrt{\frac{e^4-5e^2-2}{2e^4}\cdot\frac{3e^2-2e^2c^2-3}{18e^2}}} = -1 \Longrightarrow$$
$$\Longrightarrow \frac{c}{3e^2} = \sqrt{\frac{e^4-5e^2-2}{2e^4}\cdot\frac{3e^2-2e^2c^2-3}{18e^2}}.$$

Protože $3e^2$ je kladné číslo, musí platit $c \geq 0$. Získanou rovnici umocníme a vyřešíme:

$$\frac{c^2}{9e^4} = \frac{(e^4-5e^2-2)\cdot(3e^2-2e^2c^2-3)}{36e^6},$$
$$c^2 = \frac{3e^6-2e^6c^2-3e^4-15e^4+10e^4c^2+15e^2-6e^2+4e^2c^2+6}{4e^2},$$
$$c^2 = \frac{3}{4}e^4 - \frac{9}{2}e^2 + \frac{9}{4} + c^2\left(-\frac{1}{2}e^4 + \frac{5}{2}e^2 + 1\right) + \frac{3}{2e^2},$$
$$c^2\left(\frac{1}{2}e^4 - \frac{5}{2}e^2\right) = \frac{3}{4}e^4 - \frac{9}{2}e^2 + \frac{9}{4} + \frac{3}{2e^2},$$
$$8,83c^2 \cong 10,15,$$
$$c^2 \cong 1,15,$$
$$c \cong \pm 1,07.$$

Protože víme, že hledaná konstanta $c \geq 0$, existuje jenom jediné řešení; c se přibližně rovná $1,07$.

1.4 Charakteristiky náhodných vektorů

V předchozí podkapitole jsme uvedli pojem náhodného vektoru. Zde uvidíme, jaké jsou jeho číselné charakteristiky a na příkladu ukážeme, jak se počítají.

Definice 1.4.1. *Nechť* $\boldsymbol{X} = (X_1, ..., X_n)' : \Omega \to \mathbb{R}^n$ *je náhodný vektor definovaný na pravděpodobnostním prostoru* $(\Omega, \mathscr{A}, P)$. *Potom* ***střední hodnotu náhodného vektoru X*** *nazýváme vektor*

$$E(\mathbf{X}) = \big(E(X_1), ..., E(X_n)\big)'.$$

Definice 1.4.2. *Nechť* $\boldsymbol{X} = (X_1, ..., X_n)' : \Omega \to \mathbb{R}^n$ *je náhodný vektor definovaný na pravděpodobnostním prostoru* $(\Omega, \mathscr{A}, P)$. *Potom* ***varianční (kovarianční) maticí náhodného vektoru X*** *nazýváme matici*

$$D(\mathbf{X}) = var(\mathbf{X}) = cov(\mathbf{X},\mathbf{X}) = C(\mathbf{X},\mathbf{X}) = \big(C(X_i, X_j)\big)_{\substack{i=1,...,n \\ j=1,...,n}}.$$

> ***Definice 1.4.3.*** *Nechť* $\mathbf{X} = (X_1, ..., X_n)' : \Omega \to \mathbb{R}^n$ *je náhodný vektor definovaný na pravděpodobnostním prostoru* $(\Omega, \mathscr{A}, P)$. *Potom* ***korelační maticí náhodného vektoru X*** *nazýváme matici*
>
> $$R(\mathbf{X}) = cor(\mathbf{X}) = R(\mathbf{X},\mathbf{X}) = \big(R(X_i, X_j)\big)_{\substack{i=1,...,n \\ j=1,...,n}}.$$

Příklad 15. Ve třídě je 25 žáků; 11 chlapců a 14 děvčat. Učitelka má náhodně vybrat jedno dítě, které bude mít na starosti třídní nástěnku. Máme náhodný vektor $\mathbf{X} = (X_1, X_2)'$, který je definován takto:

$$X_1 = \begin{cases} 1 & \text{vybere-li chlapce,} \\ 0 & \text{jinak,} \end{cases} \quad \text{a} \quad X_2 = \begin{cases} 1 & \text{vybere-li dívku,} \\ 0 & \text{jinak.} \end{cases}$$

Spočítejte varianční a korelační matice vektoru $\mathbf{X} = (X_1, X_2)'$.

Řešení: U příkladu 13 ne straně 17 jsme už uváděli pravděpodobnostní tabulku, a to uděláme také zde. Víme, že pravděpodobnost výběru chlapce je $\frac{11}{25}$, zatímco pravděpodobnost výběru dívky je $\frac{14}{25}$:

$X_1 \backslash X_2$	0	1	$p_{X_1}(x_1)$
0	0	$\frac{14}{25}$	$\frac{14}{25}$
1	$\frac{11}{25}$	0	$\frac{11}{25}$
$p_{X_2}(x_2)$	$\frac{11}{25}$	$\frac{14}{25}$	1

Tabulka 1.5. Pravděpodobnostní tabulka.

Z toho lze snadno spočítat:

$$E(X_1) = \sum_{i=1}^{2} x_{1_i} p_{X_1}(x_{1_i}) = 1 \cdot \frac{11}{25} + 0 \cdot \frac{14}{25} = \frac{11}{25},$$

a

$$E(X_2) = \sum_{i=1}^{2} x_{2_i} p_{X_2}(x_{2_i}) = 0 \cdot \frac{11}{25} + 1 \cdot \frac{14}{25} = \frac{14}{25}.$$

Stejně:

$$E(X_1^2) = \sum_{i=1}^{2} x_{1_i}^2 p_{X_1}(x_{1_i}) = 1 \cdot \frac{11}{25} + 0 \cdot \frac{14}{25} = \frac{11}{25},$$

a

$$E(X_2^2) = \sum_{i=1}^{2} x_{2_i}^2 p_{X_2}(x_{2_i}) = 0 \cdot \frac{11}{25} + 1 \cdot \frac{14}{25} = \frac{14}{25}.$$

Je zřejmé, že $E(X_1X_2) = 0$, neboť pravděpodobnost, že najednou vybere dvě děti, je nulová. Potom lze spočítat:

$$C(X_1, X_2) = E(X_1X_2) - E(X_1)E(X_2) = 0 - \frac{11}{25} \cdot \frac{14}{25} = -\frac{154}{625} = C(X_2, X_1).$$

Potřebujeme ještě rozptyly:

$$D(X_1) = E(X_1^2) - E^2(X_1) = \frac{11}{25} - \frac{121}{625} = \frac{154}{625},$$

$$D(X_2) = E(X_2^2) - E^2(X_2) = \frac{14}{25} - \frac{196}{625} = \frac{154}{625}.$$

Nyní můžeme spočítat korelace mezi nimi:

$$R(X_1, X_2) = \frac{C(X_1, X_2)}{\sqrt{D(X_1)D(X_2)}} = -1 = R(X_2, X_1).$$

Konečně máme vše, abychom sestavili kovarianční a korelační matice:

$$cov(\mathbf{X}) = \frac{154}{625}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}, \qquad \text{a} \qquad cor(\mathbf{X}) = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$

Všimněme si, že se diagonální prvky kovarianční matice rovnají rozptylům a korelační matice rovnají jedné, neboť $R(X, X) = 1$.

1.5 Cvičení

1. Nechť X je náhodná veličina, která udává počet rubů při 3 hodech mincí. Spočítejte střední hodnotu a rozptyl náhodné veličiny.

   [$E(X) = 1,5$; $D(X) = 3$]

2. Dva hráči hází kostkou. První hráč před hodem zaplatí nějakou sumu. Druhý hráč po svém hodu zaplatí prvnímu tolik peněz, kolik hodil na kostce. Kolik peněz má zaplatit první hráč, aby hra byla férová?

   [$E(X) = 3,5$]

3. Prodavač zmrzliny utrží 1200 Kč, když je pěkné počasí, a 400 Kč, když je počasí špatné. Kolik prodavač utrží, když je pravděpodobnost, že bude špatné počasí 35 %?

$[E(X) = 920]$

4. Televizory s různými poruchami přinášejí do dílny na opravy. Nechť X je náhodná veličina udávající čas opravy televizoru, s následující distribuční funkcí:

$$F_X(x) = \begin{cases} 1 - e^{-x} & \text{pro } x \geq 0, \\ 0 & \text{pro } x < 0. \end{cases}$$

Nalezněte očekávanou hodnotu a rozptyl této náhodné veličiny.

$[E(X) = 1;\ D(X) = 1]$

5. Nechť je X náhodná veličina definovaná následujícím způsobem:

$$p_X(x) = \begin{cases} \frac{1}{10} & \text{pro } x = 4, \\ \frac{6}{10} & \text{pro } x = 5, \\ \frac{3}{10} & \text{pro } x = 2, \\ 0 & \text{jinak.} \end{cases}$$

Spočítejte:

a) $E(3X + 4)$;

b) $E(X^3)$;

c) $D(1 - 2X)$.

[a)$E(3X + 4) = 16$; b)$E(X^3) = 83,8$; c)$D(1 - 2X) = 7,2$]

6. Dokažte, že $D(X) = E(X^2) - E^2(X)$.

7. Dokažte, že se střední hodnota a rozptyl náhodné veličiny X

a) binomického rozdělení $Bi(n;\ \theta)$, rovnají $E(X) = n\theta$ a $D(X) = n\theta(1 - \theta)$;

b) exponencialního rozdělení $Ex(\lambda)$, rovnají $E(X) = \frac{1}{\lambda}$ a $D(X) = \frac{1}{\lambda^2}$;

c) normálního rozdělení $N(\mu;\ \sigma^2)$, rovnají $E(X) = \mu$ a $D(X) = \sigma^2$.

8. V krabičce máme 4 bílé a 5 černých koulí. X je náhodná veličina, která udává počet vytažených bílých koulí ze 3 pokusů s vracením koulí do krabičky. Nakreslete distribuční funkci a použitím definice kvantilu spočítejte $x_{0,15},\ x_{0,25},\ x_{0,83}\ a\ x_{0,99}$.

$[x_{0,15} = 1;\ x_{0,25} = 1;\ x_{0,83} = 2;\ x_{0,99} = 3]$

9. Nechť $X \sim Ex(\lambda)$. Spočítejte horní a dolní kvartil, pro $\lambda = 2$.

$$[x_{0,25} = 0,1438;\ x_{0,75} = 0,6931]$$

10. Nechť X, Y a Z jsou vzájemně nezávislé náhodné veličiny. Spočítejte
   a) střední hodnotu transformované náhodné veličiny $U = X^2 + XY + Y^2$, pokud $E(X) = 4$, $E(Y) = -6$, $D(X) = 1$ $D(Y) = 4$ a $R(X,Y) = 0,3$;
   b) korelační koeficient $R(V,W)$, kde $V = X + Z$ a $W = Z - 3Y$, přičemž $E(X) = 4,\ E(Y) = 2,\ E(Z) = 7,\ D(X) = 5,\ D(Y) = 1\ a\ D(Z) = 9$.

$$[\text{a)}E(U) = 33,6;\ \text{b)}R(V,W) = \frac{3\sqrt{7}}{14}]$$

11. Náhodný vektor $\mathbf{X} = (X,Y)'$ má následující simultánní hustotu:

$$f_{X,Y}(x,y) = \begin{cases} 4xy & \text{pro } x,y \in [0,\,1], \\ 0 & \text{jinak.} \end{cases}$$

   Spočítejte kovarianci a korelaci mezi náhodnými veličinami X a Y.

$$[C(X,Y) = 0;\ R(X,Y) = 0]$$

12. Nechť $\mathbf{X} = (X_1, X_2)'$ je náhodný vektor, jehož hodnoty pravděpodobnostní funkce jsou dané pravděpodobnostní tabulkou:

$X_1 \backslash X_2$	0	1
0	$\frac{1}{6}$	$\frac{1}{4}$
1	$\frac{1}{3}$	$\frac{1}{4}$

Tabulka 1.6. Pravděpodobnostní tabulka.

   Nalezněte kovarianci a korelaci mezi náhodnými veličinami $X_1\ a\ X_2$.

$$[C(X_1,X_2) = -\frac{1}{24};\ R(X_1,X_2) = -\frac{\sqrt{35}}{35}]$$

13. Na stole máme balíček karet lícem dolů. Náhodně vybereme jednu kartu. Náhodné veličiny $X_1,\ X_2\ a\ X_3$ jsou definovány následujícím způsobem:

$$X_1 = \begin{cases} 1 & \text{vybereme-li esa,} \\ 0 & \text{jinak,} \end{cases}$$

$$X_2 = \begin{cases} 1 & \text{vybereme-li černou dámu,} \\ 0 & \text{jinak,} \end{cases}$$

$$X_3 = \begin{cases} 1 & \text{vybereme-li červenou kartu,} \\ 0 & \text{jinak.} \end{cases}$$

Vypočtěte kovarianční a korelační matice náhodného vektoru $\mathbf{X} = (X_1, X_2, X_3)'$.

$$[cov(\mathbf{X}) = \begin{pmatrix} \frac{12}{169} & -\frac{1}{338} & -\frac{1}{26} \\ -\frac{1}{338} & \frac{25}{676} & -\frac{1}{52} \\ -\frac{1}{26} & -\frac{1}{52} & \frac{1}{4} \end{pmatrix};\ cor(\mathbf{X}) = \begin{pmatrix} 1 & -\frac{\sqrt{3}}{30} & -\frac{\sqrt{3}}{6} \\ -\frac{\sqrt{3}}{30} & 1 & -\frac{1}{5} \\ -\frac{\sqrt{3}}{6} & -\frac{1}{5} & 1 \end{pmatrix}]$$

Kapitola 2

Limitní věty

V této kapitole se seznámíme s limitními větami, které mají jak teoretický, tak praktický význam v pravděpodobnosti a statistice. Jedná se o *zákon velkých čísel* a *centrální limitní větu*. Zákon velkých čísel má spíše význam v teorii, proto se jím nebudeme zabývat (čtenáři se o něm mohou dozvědět v [1]). Centrální limitní věta (CLV) je velmi užitečná při řešení mnoha úkolů, ale ještě před ní se zmíníme o dvou nerovnostech, které se často používají v dokazování CLV a s jejichž pomocí si ukážeme, jak lze vyřešit některé úlohy. Jde o *Markovovu* a *Čebyševovu nerovnost*.

2.1 Markovova a Čebyševova nerovnost

Nejprve je důležité zmínit, že tímto způsobem lze nalézt pouze **odhad**, nikoliv přesné řešení. Právě proto nemůžeme s jistotou říct, že řešení je vždy spolehlivé, i když pomocí těchto dvou nerovností lze snadno vyřešit některé úlohy. Snadnost vyřešení, jakož i nespolehlivost řešení nejlépe uvidíme na prvních dvou příkladech.

Věta 2.1.1. *Nechť X je náhodná veličina definována na pravděpodobnostním prostoru $(\Omega, \mathscr{A}, P)$ a nechť existují $E(X)$ a $D(X)$. Potom pro libovolné $\varepsilon > 0$ platí* ***Čebyševova nerovnost***

$$P(|X - E(X)| > \varepsilon) \leq \frac{D(X)}{\varepsilon^2}.$$

Důkaz: Viz [1, str. 77].

Věta 2.1.2. *Nechť X je náhodná veličina definována na pravděpodobnostním prostoru $(\Omega, \mathscr{A}, P)$ a nechť existuje $E(X)$, pričemž $P(X > 0) = 1$. Potom pro libovolné $\varepsilon > 0$ platí* ***Markovova nerovnost***

$$P(X > \varepsilon) \leq \frac{E(X)}{\varepsilon}.$$

Důkaz: Plyne z důkazu Čebyševovy nerovnosti (návod: místo rozptylu počítáme střední hodnotu).

Nyní na dvou velmi podobných příkladech ukážeme, že řešení není vždy spolehlivé:

Příklad 1. Průměrný počet cestujících vlakem z Brna do Prahy je 2000 lidí denně. Jaká je pravděpodobnost, že v jednom dni bude cestovat 5000 lidí?

Řešení: V zadání nemáme dáno odchýlení počtu lidí od průměrného počtu, takže víme pouze střední hodnotu a úlohu vyřešíme použitím Markovovy nerovnosti (pro $\varepsilon = 5000$):

$$P(X > \varepsilon) \leq \frac{E(X)}{\varepsilon},$$
$$P(X > 5000) \leq \frac{2000}{5000} = 0,4.$$

Pravděpodobnost, že denně bude cestovat 5000 lidí, je menší než $0,4$.

Nyní uvedeme ještě jeden téměř stejný příklad, pouze jej nepatrně upravíme:

Příklad 2. Průměrný počet cestujících vlakem z Brna do Prahy je 2000 lidí denně, přičemž víme, že rozptyl počtu je 1300 lidí. Jaká je pravděpodobnost, že v jednom dni bude cestovat víc než 5000 lidí?

Řešení: Teď máme daný rozptyl, takže kromě střední hodnoty $E(X) = 2000$ víme také $D(X) = 1300$. Nyní ulohu vyřešíme použitím Čebyševovy nerovnosti (pro $\varepsilon = 5000$):

$$P(|X - E(X)| > \varepsilon) \leq \frac{D(X)}{\varepsilon^2},$$
$$P(|X - 2000| > 5000) \leq \frac{1300}{5000^2} = 0,000052.$$

V prvním příkladu jsme dostali pravděpodobnost menší jak $0,4$, zatímco v druhém menší jak $0,000052$. Můžeme říct, že druhé řešení je mnohem reálnější než první.

Příklad 3. Průměrný kadeřník ostříhá denně 10 lidí. Spočítejte:

a) pravděpodobnost, že ve středu ostříhá alespoň 12 lidí;

b) pravděpodobnost, že v sobotu ostříhá maximálně 18 lidí.

Řešení:

a) Ptáme se, jaká je pravděpodobnost, že počet ostříhaných lidí bude větší nebo roven 12, tzn. větší než 11:

$$P(X > \varepsilon) \le \frac{E(X)}{\varepsilon},$$
$$P(X > 11) \le \frac{10}{11} = 0,\overline{90}.$$

b) Teď se hledaná pravděpodobnost vztahuje na méně než 18 lidí včetně.

$$P(X \le \varepsilon) \le \frac{E(X)}{\varepsilon},$$
$$1 - P(X > \varepsilon) \le \frac{E(X)}{\varepsilon},$$
$$P(X > \varepsilon) \ge 1 - \frac{E(X)}{\varepsilon},$$
$$P(X > 18) \ge 1 - \frac{10}{18} = 0,\overline{4}.$$

Tímto příkladem jsme ukázali, jak se počítá pravděpodobnost v závislosti na zadané maximální nebo minimální hodnotě.

Vzhledem k malé spolehlivosti řešení získaného pomocí těchto dvou nerovností, používáme tento způsob tehdy, když nevíme, jaké rozdělení náhodná veličina má. V takovém případě nemáme žádný jiný způsob, než úkol „nějak“ vyřešit. Podívejme se, co se stane když známe rozdělení náhodné veličiny:

Příklad 4. Nechť X je náhodná veličina se střední hodnotou $E(X) = \frac{a+b}{2}$ a rozptylem $D(X) = \frac{(b-a)^2}{12}$. Spočítejte:

a) $P\left(|X - E(X)| > \dfrac{b-a}{3}\right)$, pro každé $b > a$;

b) $P\left(|X - E(X)| > \dfrac{b-a}{3}\right)$, pro každé $b > a$, přičemž $X \sim Ro(a;b)$.

Řešení:

a) Protože nevíme, jaké má náhodná veličina X rozdělení, pravděpodobnost spočítáme pomocí Čebyševovy nerovnosti:

$$P(|X - E(X)| > \varepsilon) \le \frac{D(X)}{\varepsilon^2},$$
$$P\left(\left|X - \frac{a+b}{2}\right| > \frac{b-a}{3}\right) \le \frac{9(b-a)^2}{12(b-a)^2} = 0,75.$$

b) Nyní víme, že náhodná veličina má rovnoměrné rozdělení, a proto tento příklad vyřešíme jinak:

$$\begin{aligned}
P(|X-E(X)|>\varepsilon) &= P\left(\left|X-\frac{a+b}{2}\right|>\frac{b-a}{3}\right) = 1-P\left(\left|X-\frac{a+b}{2}\right|\le\frac{b-a}{3}\right) = \\
&= 1-P\left(-\frac{b-a}{3}\le X-\frac{a+b}{2}\le\frac{b-a}{3}\right) = \\
&= 1-P\Bigg(\underbrace{\frac{5a+b}{6}}_{a'}\le X\le\underbrace{\frac{a+5b}{6}}_{b'}\Bigg) = 1-\int_{a'}^{b'} f_X(x)\mathrm{d}x = \\
&= 1-\int_{\frac{5a+b}{6}}^{\frac{a+5b}{6}}\frac{1}{b-a}\mathrm{d}x = 1-\left[\frac{x-a}{b-a}\right]_{\frac{5a+b}{6}}^{\frac{a+5b}{6}} = \\
&= 1-\frac{\frac{a+5b}{6}-a}{b-a}+\frac{\frac{5a+b}{6}-a}{b-a} = 1-\frac{4(b-a)}{6(b-a)} = 1-\frac{2}{3} = 0,\overline{3}.
\end{aligned}$$

Vidíme, že pravděpodobnost je zde téměř $2,5$-krát menší než v případě řešení a).

2.2 Centrální limitní věta

Až dosud jsme pracovali pouze s jednotlivými náhodnými veličinami nebo vektory. *Centrální limitní věta* vyžaduje posloupnost náhodných veličin (nebo náhodných vektorů), které mají stejné rozdělení. Ukážeme si vlastnosti těchto posloupností, které nebudou záviset na počátečním rozdělení zmíněných náhodných veličin (nebo vektorů) a platí při posloupnostech s n náhodných pokusů.

Ještě předtím, než přejdeme k CLV, uvedeme příklad, s jehož pomocí zavedeme pojem *centrované* a *standardizované* náhodné veličiny:

Příklad 5. Nechť X je náhodná veličina se střední hodnotou $E(X)=\mu$ a rozptylem $D(X)=\sigma^2$.

a) Nechť $C=X-\mu$, spočítejte $E(C)$ a $D(C)$;

b) Nechť $U=\frac{X-\mu}{\sigma}$, spočítejte $E(U)$ a $D(U)$.

Řešení:

a) Podívejme se, jak bude vypadat řešení:

$E(C)=E(X-\mu)=E(X)-E(\mu)=\mu-\mu=0,.$

$D(C)=D(X-\mu)=D(X)+D(\mu)=\sigma^2+0=\sigma^2$

b) Podobně:

$$E(U) = E\left(\frac{X-\mu}{\sigma}\right) = E\left(\frac{X}{\sigma}\right) - E\left(\frac{\mu}{\sigma}\right) = \frac{1}{\sigma}\mu - \frac{\mu}{\sigma} = 0,$$

$$D(U) = D\left(\frac{X-\mu}{\sigma}\right) = D\left(\frac{X}{\sigma}\right) + D\left(\frac{\mu}{\sigma}\right) = \frac{1}{\sigma^2}\sigma^2 + 0 = 1.$$

Nyní můžeme definovat pojem *centrované* a *standardizované náhodné veličiny*:

> ***Definice 2.2.1.*** *Nechť $\{X_n\}_{n=1}^{\infty}$ je posloupnost nezávislých náhodných veličin definovaných na pravděpodobnostním prostoru $(\Omega, \mathscr{A}, P)$, a nechť $E(X_i) = \mu_i$ a $D(X_i) = \sigma_i^2$, pro $i = 1, ..., n$. Potom pro $i = 1, ..., n$ říkáme, že náhodná veličina*
>
> 1. *$C_i = X_i - \mu_i$ je **centrovaná** $\Rightarrow E(C_i) = 0$ a $D(C_i) = \sigma_i^2$;*
> 2. *$U_i = \dfrac{X_i - \mu_i}{\sigma_i}$ je **standardizovaná** $\Rightarrow E(U_i) = 0$ a $D(U_i) = 1$.*

Teď přejdeme k příkladu, který nám ukáže, jak vypadá *standardizovaný průměr*:

Příklad 6. Pokud je $\overline{X}_n$ průměr vzájemně nezávislých náhodných veličin se střední hodnotou $E(X_i) = \mu_i$ a rozptylem $D(X_i) = \sigma_i^2$ pro $i = 1, \ldots, n$, spočítejte standardizovaný průměr $U_{\overline{X}_n}$.

Řešení: Označíme průměr $\overline{X}_n = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$ a spočítáme jeho střední hodnotu a rozptyl:

$$E(\overline{X}_n) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n} \mu_i = \frac{1}{n}(\mu_1 + ... + \mu_n),$$

$$D(\overline{X}_n) = \frac{1}{n^2}\sum_{i=1}^{n} D(X_i) = \frac{1}{n^2}\sum_{i=1}^{n} \sigma_i^2 = \frac{1}{n^2}(\sigma_1^2 + ... + \sigma_n^2).$$

Z definice standardizované náhodné veličiny máme:

$$U_{\overline{X}_n} = \frac{\overline{X}_n - E(\overline{X}_n)}{\sqrt{D(\overline{X}_n)}} = \frac{\frac{1}{n}\sum\limits_{i=1}^{n} X_i - \frac{1}{n}\sum\limits_{i=1}^{n} \mu_i}{\sqrt{\frac{1}{n^2}\sum\limits_{i=1}^{n} \sigma_i^2}} = \frac{\frac{1}{n}\sum\limits_{i=1}^{n} (X_i - \mu_i)}{\frac{1}{n}\sqrt{\sum\limits_{i=1}^{n} \sigma_i^2}} = \frac{\sum\limits_{i=1}^{n} (X_i - \mu_i)}{\sqrt{\sum\limits_{i=1}^{n} \sigma_i^2}}.$$

Pokud $E(X_i) = \mu$ a $D(X_i) = \sigma^2$ pro $i = 1, ..., n$, potom je

$$U_{\overline{X}_n} = \frac{\sum\limits_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} = \frac{\overline{X_n} - \mu}{\sigma}\sqrt{n}.$$

Nyní uvedeme první verzi CLV, která říká, že standardizovaný průměr $U_{\overline{X}_n}$ nezávislých náhodných veličin konverguje k normálnímu rozdělení.

Věta 2.2.2. *LINDEBERGOVA-LÉVYHO CLV. Nechť $\{X_n\}_{n=1}^{\infty}$ je posloupnost nezávislých náhodných veličin definovaných na pravděpodobnostním prostoru $(\Omega, \mathscr{A}, P)$ se stejným rozdělením se střední hodnotou μ a rozptylem σ^2. Potom náhodné veličiny*

$$U_{\overline{X}_n} = \frac{(\overline{X}_n - \mu)\sqrt{n}}{\sigma}$$

mají ***asymptoticky standardizované normální rozdělení*** *$N(0,1)$. Označujeme*

$$U_{\overline{X}_n} \overset{A}{\sim} N(0;1).$$

Důkaz: Viz [1, str. 92].

Poznámka 2.2.3. *Nechť $\Phi(u)$ je distribuční funkce standardizovaného normálního rozdělení. Potom platí*

1. $\Phi(-u) = 1 - \Phi(u)$;
2. $P(U_{\overline{X}_n} \leq u) = \Phi(u) = 1 - \Phi(-u) = 1 - P(U_{\overline{X}_n} \leq -u) = P(U_{\overline{X}_n} > -u)$.

Nejlépe tuto problematiku objasníme na příkladech:

Příklad 7. Doba potřebná k objevení a odstranění poruchy stroje má střední hodnotu $E(X) = 20$ minut a rozptyl $D(X) = 225$ minut. Jaká je pravděpodobnost, že doba potřebná k objevení a opravení 50 vzájemně nezávislých poruch nepřekročí 15 hodin?

Řešení: Nejprve označíme potřebné veličiny:

X_i ... doba potřebná k nalezení a opravení i-té poruchy (pro $i = 1, \ldots, 50$);
Y_n ... suma všech dob potřebných k nalezení a opravení poruch;
Z ... náhodný jev, že doba objevení a opravení všech poruch nepřekročí 15 hodin (tj. 900 minut).

Zde sice nevíme, jaké rozdělení mají tyto náhodné veličiny, ale víme, že všechny jej mají stejné. Proto na základě Lindebergovy-Lévyho CLV můžeme tato rozdělení aproximovat normálním rozdělením. Střední hodnota náhodné veličiny Y_n je tedy

$$\mu = E(Y_n) = \sum_{i=1}^{50} E(X_i) = 50 \cdot 20 = 1000,$$

a rozptylem

$\sigma^2 = D(Y_n) \overset{\text{nez.}}{=} \sum_{i=1}^{50} D(X_i) = 50 \cdot 225 = 11250.$

Nyní již není problém určit hledanou pravděpodobnost:

$$P(Y_n \in Z) = P(Y_n \leq 900) = P\left(\frac{Y_n - 1000}{\sqrt{11250}} \leq \frac{900 - 1000}{\sqrt{11250}}\right) = P(U_{\overline{X}_n} \leq -0,9428) =$$
$$= \Phi(-0,9428) = 1 - \Phi(0,9428) = 1 - 0,82639 = 0,17361.$$

Vidíme tedy, že pravděpodobnost, že hledaná doba nepřekročí 15 hodin, je 17 %.

Příklad 8. Životnost elektrické žehličky Philips má exponenciální rozdělení se střední hodnotou $E(X) = 3$ roky.

a) Odhadněte pravděpodobnost, že průměrná životnost 200 prodaných žehliček převýší 42 měsíců.

b) Jaká má být záruční doba, aby pravděpodobnost překročení průměrné životnosti 100 žehliček byla maximálně 5 %?

c) Kolik musíme vzít žehliček, aby pravděpodobnost překročení průměrné životnosti přes 42 měsíců byla nejvíce 95 %?

Řešení:

a) Pravděpodobnost jsme už odhadovali. Podívejme se, jaká bude v tomto případě:

X_i ... životnost i-té Philips žehličky (pro $i = 1, ..., 200$);
Z ... náhodný jev, že průměrná životnost 200 prodaných žehliček převýší 42 měsíců (tj. 3,5 roků).

Víme, že střední hodnota exponenciálního rozdělení je $E(X) = \dfrac{1}{\lambda} = 3$. Z toho plyne, že $\lambda = \dfrac{1}{3}$ a $D(X) = \dfrac{1}{\lambda^2} = 9$.

Vzhledem k tomu, že se jedná o průměr náhodných veličin, střední hodnota zůstane stejná, tj.

$$\mu = E(\overline{X}_n) = \frac{1}{n} \sum_{i=1}^{200} E(X_i) = 3,$$

zatímco rozptyl bude

$$\sigma^2 = D(\overline{X}_n) \overset{\text{nez.}}{=} \frac{1}{n^2} \sum_{i=1}^{200} D(X_i) = \frac{9}{200} = 0,045.$$

$$P(\overline{X}_n \in Z) = P(\overline{X}_n \leq 3,5) = P\left(\frac{\overline{X}_n - 3}{\sqrt{0,045}} \leq \frac{3,5-3}{\sqrt{0,045}}\right) = P(U_{\overline{X}_n} \leq 2,357) =$$
$$= \Phi(2,357) = 0,99061.$$

Zde je hledaná pravděpodobnost příliš vysoká, o něco více než 99 %.

b) Podívejme se, jak bude vypadat řešení tentokrát:

X_i ... životnost i-té Philips žehličky (pro $i = 1,\dots,100$);
Z_a ... odhadovaná záruční doba;
Z ... náhodný jev, že pravděpodobnost překročení průměrné životnosti 100 žehliček bude 0,05.

Střední hodnota bude stejná jako dříve, $\mu = 3$, ale rozptyl bude

$$\sigma^2 = D(\overline{X}_n) \overset{\text{nez.}}{=} \frac{1}{n^2}\sum_{i=1}^{100} D(X_i) = \frac{9}{100} = 0,09.$$

$$P(\overline{X}_n \in Z) = P(\overline{X}_n \leq Z_a) = P\left(\frac{\overline{X}_n - 3}{\sqrt{0,09}} \leq \frac{Z_a - 3}{\sqrt{0,09}}\right) = P\left(U_{\overline{X}_n} \leq \frac{Z_a - 3}{0,3}\right) \leq 0,05,$$

$$\Phi\left(\frac{Z_a - 3}{0,3}\right) \leq 0,05,$$

$$\frac{Z_a - 3}{0,3} \leq u_{0,05} = -u_{0,95} = -1,645,$$

$$Z_a \leq 2,5065.$$

Vidíme tedy, že hledaná záruka musí být 2 a půl roku, což je 30 měsíců.

c) Nyní odhadujeme počet žehliček, aby pravděpodobnost, že průměrná životnost překročí 42 měsíců byla nejvíce 95 %:

n ... odhadovaný počet žehliček;
X_i ... životnost i-té Philips žehličky (pro $i = 1, \dots, n$);
Z ... náhodný jev, že pravděpodobnost, že průměrná životnost překročí 42 měsíců, bude 95 %.

Rozptyl je nyní $\sigma^2 = \frac{9}{n}$, střední hodnota zůstává stejná.

$$P(\overline{X}_n \in Z) = P\left(\frac{\overline{X}_n - 3}{\sqrt{9}}\sqrt{n} \leq \frac{3,5-3}{\sqrt{9}}\sqrt{n}\right) = P\left(U_{\overline{X}_n} \leq \frac{3,5-3}{3}\sqrt{n}\right) \leq 0,95,$$

$$\Phi\left(\frac{3,5-3}{3}\sqrt{n}\right) \leq 0,95,$$

$$\frac{3,5-3}{3}\sqrt{n} \leq u_{0,95} = 1,645,$$

$$\sqrt{n} \leq 9,87 \Rightarrow n \leq 97,4169.$$

Počet žehliček je tedy 97.

Teď se podívejme na druhou verzi CLV, tzv. *Moivre-Laplaceovu větu.*

> ***Věta 2.2.4.*** *MOIVRE-LAPLACEOVA INTEGRÁLNÍ VĚTA. Nechť Y_n je náhodná veličina, která udává počet úspěchů v posloupnosti $\{X_i\}_{i=1}^n$ nezávislých alternativních pokusů s pravděpodobností úspěchu θ. Potom náhodné veličiny*
>
> $$U_{\overline{X}_n} = \frac{Y_n - n\theta}{\sqrt{n\theta(1-\theta)}} \overset{A}{\sim} N(0;1).$$

Důkaz: Viz [1, str. 94].

Jinými slovy, náhodná veličina Y_n je binomická náhodná veličina. Z toho je zřejmá její střední hodnota $E(Y_n) = n\theta$ a rozptyl $D(Y_n) = n\theta(1-\theta)$.

Poznámka: Zde používáme tzv. **opravu na spojitost**:
$P(X \leq a) = P(X < a+1) \Rightarrow P(X \leq a+0,5)$,
kde X je diskrétní náhodná veličina a $a \in \mathbb{N}$.

Uvidíme to na následujícím příkladu:

Příklad 9. Přijímací zkoušky se zúčastnilo celkem 9000 studentů v ramci celé univerzity. Jaká je pravděpodobnost, že přijímací zkoušku úspěšně složí maximálně 6900 studentů? Pravděpodobnost, že student zkoušku složí, je 75 %.

Řešení: Je to binomické rozdělení s $n = 9000$ a $\theta = 0,75$.

Y_n ... počet úspěšně složených přijímacích zkoušek;
B ... náhodný jev, že přijímací zkoušku úspěšně složí maximálně 6900 studentů.

Potřebujeme spočítat $P(Y_n \in B)$, kde $Y_n \sim Bi(9000; 0,75)$, přičemž $E(Y_n) = n\theta = 6750$ a $D(Y_n) = n\theta(1-\theta) = 1687,5$:

$$\begin{aligned} P(Y_n \in B) &= P(Y_n \leq 6900) = P(Y_n < 6901) = P(Y_n \leq 6900,5) = \\ &= P\left(\frac{Y_n - E(Y_n)}{\sqrt{D(Y_n)}} \leq \frac{6900,5 - E(Y_n)}{\sqrt{D(Y_n)}}\right) = P\left(U_{\overline{X}_n} \leq \frac{6900,5-6750}{\sqrt{1687,5}}\right) = \\ &= P(U_{\overline{X}_n} \leq 3,6636) = \Phi(3,66) = 0,99987. \end{aligned}$$

Hledaná pravděpodobnost je tedy 99 %.

Nyní se zaměříme na další možné příklady, ale ještě předtím uvedeme jednu větu:

> ***Poznámka 2.2.5.*** *Nechť u_α jsou kvantily standardizovaného normálního rozdělení, kde $\alpha \in (0; 0,5)$. Potom platí*
>
> $$u_\alpha = 1 - u_{1-\alpha}.$$

Příklad 10. Do obchodu přišlo 560 lidí. 190 z nich ukončilo nakupování za 15 minut anebo méně. 470 lidí zůstalo 40 minut nebo méně. Spočítejte střední hodnotu a rozptyl doby nákupu v obchodě.

Řešení: Označme

X ... délka trvání nákupu;
$X \sim N(\mu; \sigma^2)$.

Víme, že $P(X \leq 15) = 0,34$ a $P(X \leq 40) = 0,84$. Z toho dostáváme dvě rovnice o dvou neznámých:

$$
\begin{aligned}
&P\left(\frac{X-\mu}{\sigma} \leq \frac{15-\mu}{\sigma}\right) = 0,34, && P\left(\frac{X-\mu}{\sigma} \leq \frac{40-\mu}{\sigma}\right) = 0,84,\\
&P\left(U_{\overline{X}_n} \leq \frac{15-\mu}{\sigma}\right) = 0,34, && P\left(U_{\overline{X}_n} \leq \frac{40-\mu}{\sigma}\right) = 0,84,\\
&\Phi\left(\frac{15-\mu}{\sigma}\right) = 0,34, \qquad\qquad \text{a} && \Phi\left(\frac{40-\mu}{\sigma}\right) = 0,84,\\
&\frac{15-\mu}{\sigma} = u_{0,34} = -u_{0,66}, && \frac{40-\mu}{\sigma} = u_{0,84},\\
&\frac{15-\mu}{\sigma} = -0,412, && \frac{40-\mu}{\sigma} = 0,994.
\end{aligned}
$$

Z toho snadno dostaneme, že $\mu = 22,325$ a $\sigma^2 = 316,128$.

Příklad 11. Chceme slavit narozeniny v restauraci, ve které můžeme zarezervovat pouze 20 míst, a chtěli bychom pozvat 17 kamarádů a 12 kamarádek. Avšak ve stejný večer se hraje důležitý zápas, takže pravděpodobnost, že přijde kamarád, je 40 %. Kamarádka přijde s jistotou na 90 %. Samozřejmě nechceme, aby se stalo, že přijde více lidí, než máme zarezervovaných míst. Můžeme zarezervovat pouze 20 míst s rizikem 1 %, že přijde více než 20 lidí?

Řešení: Na základě uvedených pravděpodobností chceme spočítat, kolik lidí přijde s jistotou 99 %. Pravděpodobnost, že někdo přijde, ať už kamarád, nebo kamarádka je

$$\theta = \frac{17 \cdot 0,4 + 12 \cdot 0,9}{29} = 0,6068.$$

Tudíž máme binomické rozdělení s neznámým parametrem n a $\theta = 0,6068$. Označme:

X ... celkový počet lidí, kteří přijdou;
Z ... pravděpodobnost, že přijde méně než 20 lidí.

Střední hodnota a rozptyl jsou tedy:

$E(X) = n\theta = 0,6068n,$
$D(X) = n(1-\theta)\theta = 0,2385n.$

$$P(X \in Z) = P(X < 20) = P(X \leq 19) = P(X \leq 19,5) =$$
$$= P\left(\frac{X - E(X)}{\sqrt{D(X)}} \leq \frac{20 - E(X)}{\sqrt{D(X)}}\right) = P\left(U_{\overline{X}_n} \leq \frac{20 - 0,6068n}{\sqrt{0,2385n}}\right) = 0,99,$$

$$\frac{20 - 0,6068n}{\sqrt{0,2385n}} = u_{0,99} = 2,326,$$

$$20 - 0,6068n = 1,1359\sqrt{n}.$$

Uvedeme substituci $\sqrt{n} = t$ a dostáváme kvadratickou rovnici:

$$0,6068t^2 + 1,1359t - 20 = 0.$$

Protože t musí být kladné, vyřešením této rovnice, dostáváme pouze jedno řešení: $t = 4,8808$, takže $n = 23,8222 \Rightarrow n = 23$.
Nemůžeme tedy počítat s tím, že přijde pouze 20 lidí, takže zarezervujeme nějakou jinou restauraci.

2.3 Cvičení

1. Nechť X je náhodná veličina, která udává počet návštěvníků Brněnské muzejní noci v roce 2015.

   a) Odhadněte pravděpodobnost, že v roce 2015 bude méně než 7000 návštěvníků, pokud je střední hodnota $E(X) = 5500$;

   b) Odhadněte pravděpodobnost, že muzejní noc navštíví mezi 3000 a 4000 návštěvníky, pokud je střední hodnota $E(X) = 3500$ a rozptyl $D(X) = 500$.

[a) $0,214$; b) $0,998$]

2. Balíčky važící 1kg třešní mají střední hodnotu 120 třešní a rozptyl 900 třešní. Jaká je pravděpodobnost, že celkový počet třešní ve 20 nakoupených balíčcích nepřekročí 2500 třešní?

   $[0,77035]$

3. Náhodně jsme vybrali 1000 aut a podívali jsme se na ujetou vzdálenost v uplynulém roce. 80 z nich ujelo více než 35000 km, zatímco 700 z nich ujelo více než 10000 km. Spočítejte střední hodnotu a rozptyl ujetých kilometrů.

   $[\mu = 16791,08;\ \sigma^2 = 12960,08^2]$

4. Podle statistiky chodí 65 % studentů do menzy každodenně.

   (a) Jaká je pravděpodobnost, že maximálně 1150 studentů chodí do menzy, pokud jsme náhodně vybrali 1700 studentů?

   (b) Kolik studentů musíme vybrat, aby pravděpodobnost, že alespoň 1000 chodí do menzy byla 99 %?

   [a) $0,98956$; b) 1473]

5. Pravděpodobnost úspěchu prvního servisu u průměrného hráče tenisu je 48 %. Kolik prvních servisů musí hráč odpálit, aby pravděpodobnost 70 úspěšných servisů byla 90 %?

   [129]

Kapitola 3

Normální a odvozená rozdělení

Tato kapitola nám slouží pouze k tomu, abychom se seznámili s rozděleními odvozenými z normálního rozdělení. V tabulce 1.3. jsme uvedli střední hodnoty a rozptyly těchto rozdělení, nyní je definujeme a ukážeme vlastnosti. Jde o *standardizované normální rozdělení,* χ^2 *rozdělení, Studentovo t rozdělení* a *Fisherovo-Snedecorovo F rozdělení*. Tato rozdělení budeme potřebovat především v dalších kapitolách.

Na začátku si ještě připomeňme, jaký tvar má hustota normálního rozdělení a jaké rozdělení má transformovaná náhodná veličina:

Definice 3.1. *Nechť X je náhodná veličina, která má* ***normální rozdělení*** *s parametry $\mu \in \mathbb{R}$ a $\sigma^2 > 0$. Potom její hustota má tvar*

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right), \ \textit{pro}\ \forall x \in \mathbb{R}.$$

Zapisujeme $X \sim N(\mu; \sigma^2)$.

Věta 3.2. *Nechť X je náhodná veličina, která má normální rozdělení $X \sim N(\mu; \sigma^2)$, a nechť $a, b \in \mathbb{R}$, $b \neq 0$ jsou konstanty. Potom lineární transformace náhodné veličiny $Y = a + bX$ má normální rozdělení, a to*

$$Y = a + bX \sim N(a + b\mu; b^2\sigma^2).$$

Speciálně náhodná veličina

$$U = \frac{X-\mu}{\sigma} \sim N(0; 1)$$

má standardizované normální rozdělení.

Důkaz: Viz [1, str. 61].

O standardizovaném normálním rozdělení jsme se již zmínili dříve, podívejme se tedy na ta další:

Definice 3.3. *Řekneme, že náhodná veličina X má* χ^2 ***rozdělení*** *s* $\nu > 0$ *stupni volnosti, pokud její hustota má tvar*

$$f_X(x) = \begin{cases} \dfrac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})} x^{\frac{\nu}{2}-1} e^{-\frac{1}{2}x} & \text{pro } x \geq 0, \\ 0 & \text{jinak.} \end{cases}$$

Zapisujeme $X \sim \chi^2(\nu)$.

Věta 3.4. *Nechť* $U_1, \ldots, U_n$ *jsou nezávislé náhodné veličiny se standardizovaným normálním rozdělením, tj.*

$$U_i \sim N(0;1), \text{ pro } i = 1, \ldots, n.$$

Pak náhodná veličina

$$K = \sum_{i=1}^{n} U_i^2 \sim \chi^2(n)$$

má χ^2 *rozdělení o n stupních volnosti.*

Důkaz: Viz [1, str. 64].

Než přejdeme k příkladům, uvedeme ještě ostatní odvozená rozdělení:

Definice 3.5. *Řekneme, že náhodná veličina X má* **Studentovo t *rozdělení*** *s* $\nu > 0$ *stupni volnosti, pokud její hustota má tvar*

$$f_X(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{\nu}{2}\right)} \nu^{-\frac{1}{2}} \left(\frac{x^2}{\nu} + 1\right)^{-\frac{\nu+1}{2}}, \text{ pro } x \in \mathbb{R}.$$

Zapisujeme $X \sim t(\nu)$.

Věta 3.6. *Nechť náhodné veličiny* $U \sim N(0;1)$ *a* $K \sim \chi^2(\nu)$ *jsou nezávislé. Pak náhodná veličina*

$$T = \frac{U}{\sqrt{\frac{K}{\nu}}} \sim t(\nu)$$

má Studentovo t rozdělení o ν *stupních volnosti.*

Důkaz: Viz [1, str. 66].

Definice 3.7. *Řekneme, že náhodná veličina X má* **Fisherovo-Snedecorovo F** ***rozdělení*** *s $\nu_1 > 0$ a $\nu_2 > 0$ stupních volnosti, pokud její hustota má tvar*

$$f_X(x) = \begin{cases} \dfrac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \left(\dfrac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} y^{\frac{\nu_1}{2}-1} \left(\dfrac{\nu_1}{\nu_2} y + 1\right)^{-\frac{\nu_1+\nu_2}{2}} & \text{pro } y \geq 0, \\ 0 & \text{jinak.} \end{cases}.$$

Zapisujeme $X \sim F(\nu_1; \nu_2)$.

Věta 3.8. *Nechť náhodné veličiny $K_1 \sim \chi^2(\nu_1)$ a $K_2 \sim \chi^2(\nu_2)$ jsou nezávislé. Pak náhodná veličina*

$$F = \frac{K_1 \nu_2}{K_2 \nu_1} \sim F(\nu_1; \nu_2)$$

má Fisherovo-Snedecorovo F rozdělení o ν_1 a ν_2 stupních volnosti.

Důkaz: Viz [1, str. 67].

V následujícím příkladu probereme všechna zmíněná rozdělení.

Příklad 1. Nechť X_i jsou nezávislé náhodné veličiny z normálního rozdělení se střední hodnotou $E(X_i) = 0$ a rozptylem $D(X_i) = 1$ pro $i = 1, \ldots, 5$. Určete, jaké rozdělení pravděpodobnosti mají následující transformované náhodné veličiny:

a) $Y_1 = 2X_1 + X_2 - X_3 + 4$;

b) $Y_2 = \dfrac{\sqrt{8}X_1 + 2X_4 - 3}{\sqrt{3}}$;

c) $Y_3 = \displaystyle\sum_{i=1}^{5} X_i^2$;

d) $Y_4 = X_2^2 + \dfrac{X_1^2 - 2X_1X_3 + X_3^2}{2}$;

e) $Y_5 = \dfrac{2X_1 + X_2}{\sqrt{\sum_{i=3}^{7} X_i^2}}$;

f) $Y_6 = \dfrac{3X_1^2}{X_2^2 + X_3^2 + X_4^2}$.

Řešení: Postupně počítáme:

a) Zde se očividně jedná o normální rozdělení. Spočítejme pouze střední hodnotu a rozptyl:

$$E(Y_1) = E(2X_1 + X_2 - X_3 + 4) = 2E(X_1) + E(X_2) - E(X_3) + E(4) = 4,$$

$$D(Y_1) = D(2X_1 + X_2 - X_3 + 4) = 4D(X_1) + D(X_2) + D(X_3) + D(4) = 6.$$

Z toho plyne $Y_1 \sim N(4; 6)$.

b) Nyní spočítáme střední hodnotu a rozptyl:

$$E(Y_2) = E\left(\frac{\sqrt{8}X_1 + 2X_4 - 3}{\sqrt{3}}\right) = \frac{1}{\sqrt{3}}\big(\sqrt{8}E(X_1) + 2E(X_4) - E(3)\big) = -\sqrt{3},$$

$$D(Y_2) = D\left(\frac{\sqrt{8}X_1 + 2X_4 - 3}{\sqrt{3}}\right) = \frac{1}{3}\big(8D(X_1) + 4D(X_4) + D(3)\big) = 4.$$

Zde máme opět $Y_2 \sim N(-\sqrt{3}; 4)$.

c) Tento případ je jednoduchý, neboť řešení plyne přímo z definice:

$$Y_3 = X_1^2 + X_2^2 + X_3^2 + X_4^2 + X_5^2 \sim \chi^2(5).$$

d) Zde už musíme výraz mírně upravit:

$$Y_4 = X_2^2 + \frac{X_1^2 - 2X_1X_3 + X_3^2}{2} = X_2^2 + \left(\frac{X_1 - X_3}{\sqrt{2}}\right)^2.$$

Spočítejme nyní střední hodnotu a rozptyl výrazu v závorce:

$$E\left(\frac{X_1 - X_3}{\sqrt{2}}\right) = \frac{1}{\sqrt{2}}\big(E(X_1) - E(X_3)\big) = 0,$$

$$D\left(\frac{X_1 - X_3}{\sqrt{2}}\right) = \frac{1}{2}\big(D(X_1) + D(X_3)\big) = 1,$$

$$\frac{X_1 - X_3}{\sqrt{2}} \sim N(0; 1).$$

Z toho plyne $Y_4 \sim \chi^2(2)$.

e) $Y_5 = \dfrac{2X_1 + X_2}{\sqrt{\sum\limits_{i=3}^{7} X_i^2}}$.

Určeme nejprve rozdělení čitatele:

$E(2X_1 + X_2) = 2E(X_1) + E(X_2) = 0,$

$D(2X_1 + X_2) = 4D(X_1) + D(X_2) = 5.$

Tedy $2X_1 + X_2 \sim N(0; 5)$.

Vydělíme-li celý zlomek s $\sqrt{5}$, potom:

$\dfrac{2X_1 + X_2}{\sqrt{5}} \sim N(0; 1)$ a $\sum\limits_{i=3}^{7} X_i^2 \sim \chi^2(5)$, tedy celkem $Y_5 \sim t(5)$.

f) Zde je očividné, že

$X_1^2 \sim \chi^2(1),$

$X_2^2 + X_3^2 + X_4^2 \sim \chi^2(3),$

což celkem dává $Y_6 \sim F(1; 3)$.

3.1 Cvičení

1. Nechť $X_i \sim N(0; 1)$ jsou nezávislé náhodné veličiny pro $i = 1, \ldots, 15$. Určete rozdělení pravděpodobnosti následujících náhodných veličin:

   a) $Y_1 = X_1 + 2X_2 + 3X_3$;

   b) $Y_2 = 3X_2 - X_1 + 4 - 2X_3$;

   c) $Y_3 = \dfrac{4X_1 + 3X_3 - \sqrt{11}X_5 + 12}{\sqrt{6}}$;

   d) $Y_4 = X_2^2 + X_4^2 + X_6^2$;

   e) $Y_5 = X_1^2 + \dfrac{1}{2} \sum\limits_{i=2}^{5} X_i^2 + X_2X_3 - X_4X_5$;

f) $Y_6 = \dfrac{2\sum\limits_{i=1}^{3} X_i}{\sqrt{\sum\limits_{j=4}^{15} X_j^2}}$;

g) $Y_7 = \dfrac{X_1^2}{X_2^2}$;

h) $Y_8 = \dfrac{X_2^2 + X_3^2 + X_4^2}{X_1^2 + \dfrac{1}{2}\sum\limits_{i=5}^{8} X_i^2 + X_5 X_6 - X_7 X_8}$.

[a) $Y_1 \sim N(0;\, 14)$; b) $Y_2 \sim N(4;\, 14)$; c) $Y_3 \sim N(2\sqrt{6};\, 6)$; d) $Y_4 \sim \chi^2(3)$;
e) $Y_5 \sim \chi^2(3)$; f) $Y_6 \sim t(12)$; g) $Y_7 \sim F(1;\, 1)$; h) $Y_8 \sim F(3;\, 3)$]

Kapitola 4

Teorie odhadu

V této kapitole se začneme zabývat statistickými metodami k řešení úloh. Na rozdíl od teorie pravděpodobnosti, kde se předpokládá, že jsou pravděpodobnostní prostor $(\Omega, \mathscr{A}, P)$ a rozdělení pravděpodobnosti náhodných veličin známé, v matematické statistice máme n nezávislých pozorování hodnot sledované náhodné veličiny X a jejich výsledky, tj. máme $x_1 = X(\omega_1), ..., x_n = X(\omega_n)$, kde $\omega_1, ..., \omega_n \in \Omega$. Na základě těchto pozorování jsme potom schopni udělat výpověď o rozdělení zkoumané náhodné veličiny. Předtím, než se dostaneme k řešení úkolů, uvedeme základní používané pojmy: *náhodný výběr, statistiku* a *výběrové charakteristiky*.

Definice 4.1. *Náhodný vektor* $\boldsymbol{X}_n = (X_1, ..., X_n)'$ *nazýváme* ***náhodným výběrem*** *z rozdělení pravděpodobnosti P, pokud*

1. $X_1, ..., X_n$ *jsou nezávislé náhodné veličiny;*
2. $X_1, ..., X_n$ *mají stejné rozdělení pravděpodobnosti P.*

Libovolný bod $\boldsymbol{x}_n = (x_1, ..., x_n)'$*, kde* x_i *je realizace náhodné veličiny* X_i*, pro* $i = 1, ..., n$*, budeme nazývat* ***realizací náhodného výběru*** $\boldsymbol{X}_n = (X_1, ..., X_n)'$*. Číslo n nazýváme jako* ***rozsah náhodného výběru.***

Definice 4.2. *Libovolnou náhodnou veličinu* T_n*, která je funkcí náhodného výběru* $\boldsymbol{X}_n = (X_1, ..., X_n)'$*, budeme nazývat* ***statistikou****, tj.* $T_n = T(X_1, ..., X_n)'$.

Definice 4.3. *VÝBĚROVÉ CHARAKTERISTIKY. Nechť* $\mathbf{X}_n = (X_1, ..., X_n)'$ *je náhodný výběr rozsahu n z rozdělení s distribuční funkcí* $F(x;\theta)$*, kde* $\theta \in \Theta$*, a* $\Theta \subset \mathbb{R}^n$ *je parametrický prostor. Potom statistika*

1. $\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$ *se nazývá* ***výběrový průměr****;*
2. $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$ *se nazývá* ***výběrový rozptyl****;*
3. $S = \sqrt{S^2}$ *se nazývá* ***výběrová směrodatná odchylka****;*
4. $F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty,x]}(X_i)$ *se nazývá* ***výběrová (empirická) distribuční funkce****.*

Příkladem empirické distribuční funkce se nebudeme zabývat (viz [2, str. 2]).

Nyní, když jsme se seznámili se základními pojmy, můžeme uvést také definice odhadů parametrů a ukázat, jak se počítají.

4.1 Nestrannost a konzistence odhadů

V této podkapitole se budeme zabývat *bodovými odhady*, což v praxi znamená najít nějakou statistiku T_n tak, aby nejlépe aproximovala parametr θ. Ještě než se zaměříme na *nestranné* a *konzistentní odhady*, ukážeme si příklad odhadu kvantilu:

Příklad 1. Následující tabulka udává mzdy zaměstnanců, počet zaměstnanců, jakož i kumulativní četnost (vyjádřenou v procentech) zaměstnanců v podniku. Rozložte mzdový interval na 4 stejné intervaly podle počtu pracovníků, kteří patří do příslušného intervalu. Odhadněte také procento pracovníků, kteří jsou „bohatí“, jestliže je bohatství určeno měsíčnou mzdou 25000 Kč.

Mzdy (v Kč)	Počet zaměstnanců	Kumulativní četnost v %
méně než 10001	45	4,5
10001 – 12000	55	10
12001 – 15000	110	21
15001 – 18000	120	33
18001 – 21000	150	48
21001 – 24000	200	68
24001 – 30000	180	86
30001 – 50000	140	100
Celkem	1000	–

Tabulka 4.1. Zaměstnanci.

Řešení: Rozdělit nějakou uspořádanou řadu na 4 „stejně dlouhé“ znamená nalézt medián, dolní a horní kvartil. Rozdělíme tedy všechny pracovníky do 4 skupin tak, že do první skupiny bude patřit čtvrtina zaměstnanců s nejmenšími mzdami atd. Z tabulky vidíme, že 25 % patří do čtvrté skupiny pracovníků s mzdovým intervalem 15001 − 18000. To znamená, že nejvyšší plat, který 25 % pracovníků dostává, se nachází v tomto intervalu.

Pro přibližný výpočet kvartilů použijeme lineární interpolaci, kde $x_{0,25}$ bude rozdělovat interval, ve kterém se nachází ve stejném poměru, jako hledaná mzda rozděluje příslušný interval.

$$\frac{\hat{x}_{0,25} - 15000}{18000 - 15000} = \frac{25 - 21}{33 - 21}.$$

Z toho lze lehce spočítat $\hat{x}_{0,25}$:
$$\begin{aligned} 12(\hat{x}_{0,25} - 15000) &= 12000, \\ \hat{x}_{0,25} - 15000 &= 1000, \\ \hat{x}_{0,25} &= 16000. \end{aligned}$$

Podobně zjistíme, že:

$$\begin{aligned} &\frac{\hat{x}_{0,50} - 21000}{24000 - 21000} = \frac{50 - 48}{78 - 48}, \\ &30(\hat{x}_{0,50} - 21000) = 6000, \\ &\hat{x}_{0,50} - 21000 = 200, \\ &\hat{x}_{0,50} = 21200, \end{aligned}$$

a

$$\begin{aligned} &\frac{\hat{x}_{0,75} - 24000}{30000 - 24000} = \frac{75 - 68}{86 - 68}, \\ &18(\hat{x}_{0,75} - 24000) = 42000, \\ &\hat{x}_{0,75} - 24000 = 2333,\overline{3}, \\ &\hat{x}_{0,75} = 26333,\overline{3}. \end{aligned}$$

Odhadli jsme tedy, že 25 % zaměstnanců má plat menší jak 16000, polovina dostává méně než 21200 a tři čtvrtiny méně než $26333,\overline{3}$.

Nyní ještě odhadneme, jaké procento zaměstnaných dostává více než 25000. Teď známe mzdu, hledáme kvantil:

$$\begin{aligned} &\frac{25000 - 24000}{30000 - 24000} = \frac{\hat{x} - 68}{86 - 68}, \\ &6000(\hat{x} - 68) = 18000, \\ &\hat{x} - 68 = 3, \\ &\hat{x} = 71. \end{aligned}$$

Konečně máme procento pracovníků, kteří „nejsou bohatí“, tzn. že „bohatých“ pracovníků je 29 %, což je 290 ze 1000.

Můžeme ještě spočítat, že průměrná mzda je 22180 Kč. Vidíme tedy, jak tyto dvě hodnoty můžou být odlišné.

Nyní uvedeme definice nestrannosti a konzistence odhadu parametrů (ostatní naleznete v [2]):

Definice 4.1.1. *Nechť $\boldsymbol{X}_n = (X_1, ..., X_n)'$ je náhodný výběr rozsahu n z rozdělení pravděpodobnosti P_θ, kde θ je vektor neznámých parametrů. Nechť $\gamma(\theta)$ je daná parametrická funkce. Potom je statistika $T_n = T(X_1, ..., X_n)'$*

1. ***nestranným*** *(nevychýleným) odhadem parametrické funkce $\gamma(\theta)$, pokud pro $\forall\theta \in \Theta$ platí $E_\theta(T_n) = \gamma(\theta)$;*
2. ***asymptoticky nestranným*** *odhadem parametrické funkce $\gamma(\theta)$, pokud pro $\forall\theta \in \Theta$ platí $\lim\limits_{n\to\infty} E_\theta(T_n) = \gamma(\theta)$;*
3. ***(slabě) konzistentním*** *odhadem parametrické funkce $\gamma(\theta)$, pokud pro $\forall\varepsilon > 0$ platí $\lim\limits_{n\to\infty} P_\theta(|T_n - \gamma(\theta)| > \varepsilon) = \mathbf{0}$, tj. $T_n \xrightarrow{P_\theta} \gamma(\theta)$.*

Na základě následujících dvou příkladů můžeme odvodit závěr o nestranných odhadech středních hodnot a rozptylů:

Příklad 2. Nechť $\mathbf{X}_n = (X_1, ..., X_n)'$ je náhodný výběr z rozdělení pravděpodobnosti P_θ se střední hodnotou μ. Určete zda je výběrový průměr $\overline{X}_n$ nestranným odhadem střední hodnoty μ.

Řešení: Zde máme statistiku $T_n = \overline{X}_n = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$, a počítáme její střední hodnotu $E_\theta T_n$:

$$E_\theta T_n = E_\theta\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}E_\theta\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n}n\mu = \mu.$$

Výběrový průměr je tedy nestranným odhadem střední hodnoty.

Příklad 3. Nechť $\mathbf{X}_n = (X_1, \dots, X_n)'$ je náhodný výběr z rozdělení pravděpodobnosti P_θ se střední hodnotou μ a rozptylem σ^2. Dokažte, že výběrový rozptyl S^2 je nestranným odhadem rozptylu σ^2.

Řešení: Nyní máme statistiku $T_n = S^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i - \overline{X}_n)^2$, a znovu počítáme její střední hodnotu $E_\theta(T_n)$:

$$
\begin{aligned}
E_\theta(T_n) &= E_\theta\left(\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X}_n)^2\right)=\frac{1}{n-1}E_\theta\left(\sum_{i=1}^{n}\big((X_i-\mu)+(\mu-\overline{X}_n)\big)^2\right)=\\
&=\frac{1}{n-1}E_\theta\left(\sum_{i=1}^{n}(X_i-\mu)^2+2(X_i-\mu)(\mu-\overline{X}_n)+(\mu-\overline{X}_n)^2\right)=\\
&=\frac{1}{n-1}\sum_{i=1}^{n}\left(\underbrace{E_\theta(X_i-\mu)^2}_{D(X_i)=\sigma^2}+2E_\theta(X_i-\mu)(\mu-\overline{X}_n)+\underbrace{E_\theta(\overline{X}_n-\mu)^2}_{D(\overline{X}_n)=\frac{\sigma^2}{n}}\right)=\\
&=\frac{1}{n-1}\sum_{i=1}^{n}\left(\sigma^2+2E_\theta(X_i\mu-X_i\overline{X}_n-\mu^2+\mu\overline{X}_n)+\frac{\sigma^2}{n}\right)=\\
&=\frac{1}{n-1}\sum_{i=1}^{n}\left(\sigma^2+2E_\theta(X_i\mu)-2E_\theta(X_i\overline{X}_n)-2E_\theta(\mu^2)+2E_\theta(\mu\overline{X}_n)+\frac{\sigma^2}{n}\right)=\\
&=\frac{1}{n-1}\sum_{i=1}^{n}\left(\sigma^2+2\mu^2-2E_\theta\left(X_i\left(\frac{1}{n}\sum_{j=1}^{n}X_j\right)\right)-2\mu^2+2\mu^2+\frac{\sigma^2}{n}\right)=\\
&=\frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{(n+1)\sigma^2}{n}+2\mu^2-2E_\theta\left(X_i\left(\frac{1}{n}\sum_{\substack{j=1\\ j\neq i}}^{n}X_j+\frac{1}{n}X_i\right)\right)\right)=\\
&=\frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{(n+1)\sigma^2}{n}+2\mu^2-2E_\theta\left(\frac{1}{n}X_i\sum_{\substack{j=1\\ j\neq i}}^{n}X_j+\frac{1}{n}X_i^2\right)\right)=\\
&=\frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{(n+1)\sigma^2}{n}+2\mu^2-2\frac{1}{n}E_\theta(X_i)\cdot E_\theta\sum_{\substack{j=1\\ j\neq i}}^{n}X_j-2\frac{\sigma^2+\mu^2}{n}\right)=\\
&=\frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{(n+1)\sigma^2}{n}+2\mu^2-2\frac{1}{n}\mu(n-1)\mu-2\frac{\sigma^2+\mu^2}{n}\right)=\\
&=\frac{1}{n-1}\big((n+1)\sigma^2+2n\mu^2-2(n-1)\mu^2-2\sigma^2-2\mu^2\big)=\\
&=\frac{1}{n-1}\big(n\sigma^2+\sigma^2+2n\mu^2-2n\mu^2+2\mu^2-2\sigma^2-2\mu^2\big)=\\
&=\frac{(n-1)\sigma^2}{n-1}=\sigma^2.
\end{aligned}
$$

Vidíme tedy, že výběrový průměr je nestranným odhadem střední hodnoty, stejně tak výběrový rozptyl je nestranným odhadem rozptylu.

Příklad 4. Nechť $\mathbf{X}_n=(X_1,\ldots,X_n)'$ je náhodný výběr z rozdělení pravděpodobnosti $Ge(\theta)$. Určete konstantu k tak, aby statistika $T_n=k\sum\limits_{i=2}^{n}(X_i-X_{i-1})^2$ byla nestranným odhadem střední hodnoty $\frac{1-\theta}{\theta}$. Potom určete, zda je statistika T_n asymptoticky nestranným odhadem rozptylu $\frac{1-\theta}{\theta^2}$ pro $k=\frac{1}{n}$.

Řešení: Aby statistika T_n byla nestranným odhadem parametru θ, musí platit

$E_\theta(T_n) = \theta$. Na základě toho můžeme spočítat konstantu k:

$$\begin{aligned}
E_\theta(T_n) &= E_\theta\left(k\sum_{i=2}^{n}(X_i - X_{i-1})^2\right) = kE_\theta\left(\sum_{i=2}^{n}\Big(\big(X_i - E(X_i)\big) - \big(X_{i-1} - E(X_i)\big)\Big)^2\right) = \\
&= k\sum_{i=2}^{n}E_\theta\Big(\big(X_i - E(X_i)\big)^2 - 2\big(X_i - E(X_i)\big)\big(X_{i-1} - E(X_i)\big) + \\
&+ \big(X_{i-1} - E(X_i)\big)^2\Big) = k\left(\sum_{i=2}^{n}\frac{1-\theta}{\theta^2} - 2\cdot 0 + \sum_{i=2}^{n}\frac{1-\theta}{\theta^2}\right) = \\
&= 2k(n-1)\frac{1-\theta}{\theta^2} = \frac{1-\theta}{\theta} \Rightarrow k = \frac{\theta}{2(n-1)}.
\end{aligned}$$

Dostali jsme výsledný tvar statistiky $T_n = \frac{\theta}{2(n-1)}\sum\limits_{i=2}^{n}(X_i - X_{i-1})^2$.

Pro $k = \frac{1}{n}$ je střední hodnota $E_\theta(T_n)$ následující:

$$E_\theta(T_n) = E_\theta\left(\frac{1}{n}\sum_{i=2}^{n}(X_i - X_{i-1})^2\right) = \frac{1}{n}E_\theta\left(\sum_{i=2}^{n}(X_i - X_{i-1})^2\right) = \frac{2(n-1)(1-\theta)}{n\theta^2}.$$

Potom je limita:

$$\lim_{n\to\infty} E_\theta(T_n) = \lim_{n\to\infty}\frac{2(n-1)(1-\theta)}{n\theta^2} = \frac{-2(1-\theta)}{\theta^2}.$$

Statistika T zde není asymptotickým nestranným odhadem rozptylu $\frac{1-\theta}{\theta^2}$.

Nyní uvedeme větu o konzistentním odhadu, která nám velmi pomůže ve výpočtu:

Věta 4.1.2. *Nechť statistika $T_n = T(X_1, ..., X_n)'$ je nestranný nebo asymptoticky nestranný odhad parametrické funkce $\gamma(\theta)$ a platí*

$$\lim_{n\to\infty} D_\theta(T_n) = 0.$$

Pak je statistika $T_n = T(X_1, ..., X_n)'$ konzistentním odhadem parametrické funkce $\gamma(\theta)$.

Důkaz: Viz [2, str. 7].

Příklad 5. Určete, zda je statistika $T_n = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$ konzistentním odhadem parametru $\frac{1}{\lambda}$ v náhodném výběru $\mathbf{X}_n = (X_1, ..., X_n)'$ z exponencialního rozdělení $Ex(\lambda)$.

Řešení: Nejprve ověřme nestrannost či asymptotickou nestrannost odhadu:

$$E_\theta(T_n) = E_\theta\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E_\theta(X_i) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\lambda} = \frac{1}{\lambda}.$$

Vidíme tedy, že statistika T je nestranným odhadem parametru $\frac{1}{\lambda}$. Podle předchozí věty určeme, zda je konzistentním odhadem. Spočítejme nejprve rozptyl:

$$D_\theta(T_n) = D_\theta\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} D_\theta(X_i) = \frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{\lambda^2} = \frac{1}{\lambda^2 n}.$$

Určeme ještě limitu:

$$\lim_{n\to\infty} D_\theta(T_n) = \lim_{n\to\infty}\frac{1}{\lambda^2 n} = 0.$$

Můžeme říct, že statistika T_n je konzistentním odhadem parametru $\frac{1}{\lambda}$.

Nyní uvedeme ještě jednu definici, která říká, jak máme vybrat ten „nejlepší“ odhad, když máme více možností odhadů:

> ***Definice 4.1.3.*** *Nechť T_n je nestranný odhad parametrické funkce $\gamma(\theta)$ a pro všechna $\theta \in \Theta$ platí*
> $$D_\theta(T_n) \le D_\theta(T_n^*),$$
> *kde T_n^* je libovolný nestranný odhad parametru $\gamma(\theta)$. Potom odhad T_n nazveme **nejlepším nestranným odhadem** parametrické funkce $\gamma(\theta)$.*

Podívejme se na příklad:

Příklad 6. Máme náhodný výběr $\mathbf{X}_5 = (X_1, ..., X_5)'$ se střední hodnotou μ a rozptylem σ^2 a následující statistiky:

$$T_1 = X_4,$$

$$T_2 = \frac{2X_1 - X_2}{2},$$

$$T_3 = \frac{X_1 + X_2 + X_5}{5},$$

$$T_4 = \frac{X_1 + 2X_2 + X_5}{4},$$

$$T_5 = 2X_1 - X_2,$$

$$T_6 = \frac{X_2 + X_5}{2}.$$

Najděte nejlepší lineární nestranný odhad střední hodnoty μ.

Řešení: Aby statistiky T_i, pro $i = 1, ..., 5$ byly nestranným odhadem střední hodnoty μ, musí platit $E_\theta(T_i) = \mu$:

$$E_\theta(T_1) = E_\theta(X_4) = \mu,$$

$$E_\theta(T_2) = E_\theta\left(\frac{2X_1 - X_2}{2}\right) = \frac{\mu}{2},$$

$$E_\theta(T_3) = E_\theta\left(\frac{X_1 + X_2 + X_5}{5}\right) = \frac{3\mu}{5},$$

$$E_\theta(T_4) = E_\theta\left(\frac{X_1 + 2X_2 + X_5}{4}\right) = \mu,$$

$$E_\theta(T_5) = E_\theta(2X_1 - X_2) = \mu,$$

$$E_\theta(T_6) = E_\theta\left(\frac{X_2 + X_5}{2}\right) = \mu.$$

Statistiky T_1, T_4, T_5 a T_6 jsou nestranným odhadem střední hodnoty μ. Ověřme ještě podmínku z předchozí definice, tedy která z nich má nejmenší rozptyl:

$$D_\theta(T_1) = D_\theta(X_4) = \sigma^2,$$

$$D_\theta(T_4) = D_\theta\left(\frac{X_1 + 2X_2 + X_5}{4}\right) = \frac{6\sigma^2}{16} = \frac{3\sigma^2}{8},$$

$$D_\theta(T_5) = D_\theta(2X_1 - X_2) = 5\sigma^2,$$

$$D_\theta(T_6) = D_\theta\left(\frac{X_2 + X_5}{2}\right) = \frac{2\sigma^2}{4} = \frac{\sigma^2}{2}.$$

Nejmenší rozptyl je $D_\theta(T_4) = \frac{3\sigma^2}{8}$, takže statistika T_4 je nejlepším nestranným lineárním odhadem strední hodnoty μ.

4.2 Konstrukce bodových odhadů

Zmínili jsme se již o tom, co je to bodový odhad a jaké vlastnosti pozorujeme. V této podkapitole ukážeme dvě metody, jak lze takový odhad zkonstruovat. Jde o *momentovou metodu* a *metodu maximální věrohodnosti*.

4.2.1 Momentová metoda

Tato metoda je velice jednoduchá, i když výsledky nejsou příliš „kvalitní“. Spočívá v porovnání **výběrových obecných momentů**, definovaných vztahem $M_k' = \frac{1}{n}\sum\limits_{i=1}^{n} X_i^k$ pro $k = 1, 2, ...$, a **obecných momentů** $\mu_k'(\theta)$ pro $k = 1, 2, ...$, za předpokladu, že pro náhodný výběr obecné momenty existují.

Příklad 7. Nechť $\mathbf{X}_n$ je náhodný výběr z exponenciálního rozdělení $Ex(\lambda)$. Momentovou metodou odhadněte parametr λ.

Řešení: Protože odhadujeme jeden parametr, stačí nám porovnání pouze prvních momentů:

$$\mu_1' = E(X_i)^1 = \frac{1}{\lambda}, \text{ pro } i = 1, ..., n,$$
$$M_1' = \frac{1}{n}\sum_{i=1}^{n} X_i^1 = \overline{X}.$$

Když tyto dvě hodnoty porovnáme, dostáváme odhad $\hat{\lambda} = \frac{1}{\overline{X}}$.

Příklad 8. Nechť $\mathbf{X}_n$ je náhodný výběr z rovnoměrného rozdělení $Ro(a; b)$. Momentovou metodou odhadněte parametry a a b.

Řešení: Na rozdíl od předchozího příkladu, kde jsme odhadovali pouze jeden parametr, zde odhadujeme dva, takže budeme potřebovat alespoň dvě rovnice:

$$\mu_1' = E(X_i)^1 = \frac{a+b}{2}, \text{ pro } i = 1, ..., n,$$
$$M_1' = \frac{1}{n}\sum_{i=1}^{n} X_i^1,$$
$$\mu_2' = E(X_i)^2 = \int_{-\infty}^{\infty} x^2 f_X(x)\mathrm{d}x = \int_a^b \frac{x^2}{b-a}\mathrm{d}x = \frac{b^3-a^3}{3(b-a)} = \frac{a^2+ab+b^2}{3}, \text{ pro } i = 1, ..., n,$$
$$M_2' = \frac{1}{n}\sum_{i=1}^{n} X_i^2,$$

$$\frac{a+b}{2} = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

$$\frac{a^2+ab+b^2}{3} = \frac{1}{n}\sum_{i=1}^{n} X_i^2.$$

Upravme obě rovnice následujícím způsobem:

$$\left(\frac{a+b}{2}\right)^2 = \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2, \qquad \text{a} \qquad a^2+ab+b^2 = \frac{3}{n}\sum_{i=1}^{n} X_i^2.$$
$$a^2+2ab+b^2 = \frac{4}{n^2}\left(\sum_{i=1}^{n} X_i\right)^2,$$

Odečteme druhou rovnici od první a dostaneme:

$$ab = \frac{4}{n^2}\left(\sum_{i=1}^{n} X_i\right)^2 - \frac{3}{n}\sum_{i-1}^{n} X_i^2.$$

Z toho vyjádříme a, dosadíme do první původní rovnice a úpravou dostaneme:

$$b^2 - b\frac{2}{n}\sum_{i=1}^{n} X_i + \frac{4}{n^2}\left(\sum_{i=1}^{n} X_i\right)^2 - \frac{3}{n}\sum_{i-1}^{n} X_i^2 = 0.$$

Vyřešením této kvadratické rovnice dostáváme:

$$\hat{b}_{1,2} = \frac{1}{n}\sum_{i=1}^{n} X_i \pm \sqrt{\frac{3}{n}\sum_{i=1}^{n} X_i^2 - \frac{3}{n^2}\left(\sum_{i=1}^{n} X_i\right)^2}.$$

Dosazením do původní rovnice získáme:

$$\hat{a}_{1,2} = \frac{1}{n}\sum_{i=1}^{n} X_i \mp \sqrt{\frac{3}{n}\sum_{i=1}^{n} X_i^2 - \frac{3}{n^2}\left(\sum_{i=1}^{n} X_i\right)^2}.$$

4.2.2 Metoda maximální věrohodnosti

Na rozdíl od předchozí metody je tato metoda často využívána právě proto, že poskytuje „kvalitní“ výsledky. Zde budeme pracovat s tzv. **věrohodnostní funkcí náhodného výběru**, definovanou jako $L(\theta; x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i; \theta)$. Hledáme tedy maximálně věrohodný odhad:

> ***Definice 4.2.2.1.*** *Odhad $\hat{\theta}_{MLE}$ nazveme* ***maximálně věrohodným****, jestliže pro $\forall\theta \in \Theta$ platí*
> $$L(\hat{\theta}_{MLE}; x_1, \ldots, x_n) \geq L(\theta; x_1, \ldots, x_n).$$

V tomto případě je snadnější pracovat s logaritmem věrohodnostní funkce

$l(\theta;\mathbf{x}) = \ln L(\theta;\mathbf{x})$. Odhad $\hat{\theta}_{MLE}$ dostaneme vyřešením systému rovnic

$$\frac{\partial}{\partial \theta_j} l(\theta;\mathbf{x}) = \frac{\partial}{\partial \theta_j} \sum_{i=1}^{n} \ln f_{X_i}(x_i;\, \theta_1,\, ...,\, \theta_m) = 0, \text{ pro } j = 1,\, ...,\, m.$$

Udělejme příklad č. 7 (str. 53) touto metodou a porovnejme odhady.

Příklad 9. Nechť $\mathbf{X}_n$ je náhodný výběr z exponenciálního rozdělení $Ex(\lambda)$. Metodou maximální věrohodnosti odhadněte parametr λ.

Řešení: Nejprve „vytvořme" věrohodnostní funkci, kterou potom zlogaritmujeme a zderivujeme:

$$L(\lambda;\mathbf{x}) = \prod_{i=1}^{n} f_{X_i}(\lambda;x_i) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i},$$

$$l(\lambda;\mathbf{x}) = \ln L(\lambda;\mathbf{x}) = \ln\left(\lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}\right) = \ln \lambda^n + \ln e^{-\lambda \sum_{i=1}^{n} x_i} = n \ln \lambda - \lambda \sum_{i=1}^{n} x_i,$$

$$\frac{\partial}{\partial \lambda} l(\lambda;\mathbf{x}) = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0 \Rightarrow \hat{\lambda}_{MLE} = \frac{1}{\overline{X}}.$$

Touto metodou jsme dostali stejný výsledek jako u metody předchozí.

Příklad 10. Nechť $\mathbf{X}_n$ je náhodný výběr z negativně binomického rozdělení $NeBi(n;\, \theta)$. Metodou maximální věrohodnosti odhadněte parametr θ.

Řešení: Připomeňme si nejprve, jak vypadá pravděpodobnostní funkce negativně binomického rozdělení:

$$p_X(x) = \begin{cases} \binom{n+x-1}{x} \theta^n (1-\theta)^x & \text{pro } x \geq 0,\ \theta \in (0,\, 1), \\ 0 & \text{jinak.} \end{cases}$$

Další postup je už známý:

$$L(\theta;\mathbf{x}) = \prod_{i=1}^{n} p_{X_i}(\theta;x_i) = \prod_{i=1}^{n} \binom{n+x_i-1}{x_i} \theta^n (1-\theta)^{x_i} =$$

$$= \left(\prod_{i=1}^{n} \binom{n+x_i-1}{x_i}\right) \theta^{n^2} (1-\theta)^{\sum_{i=1}^{n} x_i},$$

$$l(\theta;\mathbf{x}) = \ln L(\theta;\mathbf{x}) = \ln\left[\left(\prod_{i=1}^{n}\binom{n+x_i-1}{x_i}\right)\theta^{n^2}(1-\theta)^{\sum\limits_{i=1}^{n}x_i}\right] =$$
$$= \ln\prod_{i=1}^{n}\binom{n+x_i-1}{x_i} + \ln\theta^{n^2} + \ln(1-\theta)^{\sum\limits_{i=1}^{n}x_i} =$$
$$= \sum_{i=1}^{n}\ln\binom{n+x_i-1}{x_i} + n^2\ln\theta + \sum_{i=1}^{n}x_i\ln(1-\theta),$$
$$\frac{\partial}{\partial\theta}l(\theta;\mathbf{x}) = \frac{n^2}{\theta} - \frac{1}{1-\theta}\sum_{i=1}^{n}x_i = 0 \Rightarrow \hat{\theta}_{MLE} = \frac{n}{n+\overline{X}}.$$

Odhad je tedy $\hat{\theta}_{MLE} = \frac{n}{n+\overline{X}}$.

Dříve, než uděláme další příklad, vzpomeňme si na funkci Γ z první kapitoly a zadefinujeme další funkci (**digamma** funkce), kterou použijeme k vyřešení některých příkladů a je definovaná jako logaritmická derivace funkce Γ.

Definice 4.2.2.2. *Funkce Ψ je definovaná předpisem*

$$\Psi(m) = \frac{\partial}{\partial m}\ln\Gamma(m).$$

Další příklad nebudeme řešit, jenom nám ukáže, jak lze využít digamma funkci. Tento příklad nelze vyřešit v obecné formě, pouze numericky, využitím některé z numerických metod.

Příklad 11. Nechť $\mathbf{X}_n$ je náhodný výběr z rozdělení s hustotou:

$$f_X(x) = \begin{cases} \dfrac{\lambda^m x^{m-1}}{(m-1)!}e^{-\lambda x} & \text{pro } x \geq 0, m \in \mathbb{N}, \\ 0 & \text{jinak.} \end{cases}$$

Najděte odhady parametrů λ a m metodou maximální věrohodnosti.

Řešení: Využijeme následující vlastnost: $\Gamma(m) = (m-1)!$:

$$L(\lambda;m;\mathbf{x}) = \prod_{i=1}^{n}f_{X_i}(\lambda;m;x_i) = \prod_{i=1}^{n}\frac{\lambda^m x_i^{m-1}}{\Gamma(m)}e^{-\lambda x_i} = \frac{\lambda^{mn}\prod\limits_{i=1}^{n}x_i^{m-1}}{\big(\Gamma(m)\big)^n}e^{-\lambda\sum\limits_{i=1}^{n}x_i},$$

$$l(\lambda; m; \mathbf{x}) = \ln L(\lambda; m; \mathbf{x}) = \ln\left(\frac{\lambda^{mn}\prod\limits_{i=1}^{n} x_i^{m-1}}{\big(\Gamma(m)\big)^n} e^{-\lambda\sum\limits_{i=1}^{n} x_i}\right) =$$

$$= \ln\lambda^{mn} + \ln\prod_{i=1}^{n} x_i^{m-1} - \ln\big(\Gamma(m)\big)^n + \ln e^{-\lambda\sum\limits_{i=1}^{n} x_i} =$$

$$= mn\ln\lambda + (m-1)\sum_{i=1}^{n}\ln x_i - n\ln\Gamma(m) - \lambda\sum_{i=1}^{n} x_i.$$

Nyní odhadujeme dva parametry, takže budeme mít dvě derivace:

1) $\displaystyle \frac{\partial}{\partial\lambda} l(\lambda; m; \mathbf{x}) = \frac{mn}{\lambda} - \sum_{i=1}^{n} x_i = 0 \Rightarrow \lambda = \frac{m}{\overline{X}};$

2) $\displaystyle \frac{\partial}{\partial m} l(\lambda; m; \mathbf{x}) = n\ln\lambda + \sum_{i=1}^{n}\ln x_i - n\Psi(m) = 0 \Rightarrow \Psi(m) = \ln\lambda + \frac{1}{n}\sum_{i=1}^{n}\ln x_i.$

Z těchto dvou rovnic lze tedy numericky volbou vhodného iteračního procesu odhadnout parametry λ a m (např. Newtonovou metodou; viz [4]).

4.3 Intervalové odhady

Dosud jsme se zabývali bodovými odhady parametrů, přesněji, odhad parametrické funkce $\gamma(\theta)$ jsme určovali jedním číslem. Zde přejdeme k *intervalovým odhadům* parametrické funkce $\gamma(\theta)$. To znamená, že vytvoříme interval, jehož krajní body jsou statistiky, tak, že skutečná hodnota parametrické funkce $\gamma(\theta)$ je s velkou spolehlivostí uvnitř tohoto intervalu. Podívejme se na definici intervalového odhadu:

Definice 4.3.1. *Nechť* $\mathbf{X}_n = (X_1, \dots, X_n)'$ *je náhodný výběr z rozdělení pravděpodobnosti s distribuční funkcí* $F(x;\theta)$, $\theta \in \Theta$. *Dále, nechť* $\gamma(\theta)$ *je parametrická funkce,* $\alpha \in (0,1)$ *a* $D = D_n(\mathbf{X}_n)$, $H = H_n(\mathbf{X}_n)$ *jsou statistiky. Potom interval* $[D; H]$ *nazveme* $100(1-\alpha)\%$ ***intervalem spolehlivosti*** *pro parametrickou funkci* $\gamma(\theta)$, *jestliže*

$$P\big(D_n(\mathbf{X}_n) \le \gamma(\theta) \le H_n(\mathbf{X}_n)\big) = 1-\alpha.$$

Jestliže

$$P\big(D_n(\mathbf{X}_n) \le \gamma(\theta)\big) = 1-\alpha,$$

potom statistiku $D = D_n(\mathbf{X}_n)$ *nazýváme* ***dolním odhadem*** *parametrické funkce* $\gamma(\theta)$ *se spolehlivostí* $1-\alpha$. *Jestliže*

$$P\big(\gamma(\theta) \le H_n(\mathbf{X}_n)\big) = 1-\alpha,$$

potom statistiku $H = H_n(\mathbf{X}_n)$ *nazýváme* ***horním odhadem*** *parametrické funkce* $\gamma(\theta)$ *se spolehlivostí* $1-\alpha$.

Kvantily jsme se už zabývali ve 2. kapitole. V následující tabulce uvedeme ditribuční funkce rozdělení, kterými se zde budeme zabývat, jakož i jejich kvantily a vlastnosti těchto kvantilů, jejichž hodnoty naleznete v tabulce (viz přílohy).

$X \sim$	$N(0,1)$	$\chi^2(n)$	$t(n)$	$F(n,m)$
$distr.fce$	Φ	G_n	H_n	$Q_{n,m}$
$kvantil$	u_α	$\chi^2_\alpha(n)$	$t_\alpha(n)$	$F_\alpha(n,m)$
$vlastnosti$	$u_\alpha = -u_{1-\alpha}$	$-$	$t_\alpha = -t_{1-\alpha}$	$F_\alpha(n,m) = \frac{1}{F_{1-\alpha}(m,n)}$

Tabulka 4.2. Kvantily důležitých rozdělení.

Konstrukce intervalových odhadů začíná nalezením vhodné **pivotové statistiky** pro příslušný parametr. Ohraničíme ji správnými kvantily a z ní vyjádříme hledaný parametr. Postupně se seznámíme se všemi pivotovými statistikami a na příkladě ukážeme, jak odhadneme konkrétní parametr ve zvoleném intervalu.

4.3.1 Intervalové odhady parametrů normálního rozdělení

V této části se budeme zabývat statistikami, jejichž náhodné výběry mají normální rozdělení.

Definice 4.3.1.1. *Nechť $\mathbf{X}_n$ je náhodný výběr z $N(\mu;\sigma^2)$. Potom, pro neznámé μ, když je σ^2 známé, je pivotová statistika*

$$U = \frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0;1),$$

kde $\overline{X}_n$ je výběrový průměr. Potom

$$[D,H] = \left[\overline{X}_n - u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \overline{X}_n + u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right]$$

je $100(1-\alpha)\%$ interval spolehlivosti pro střední hodnotu μ při známém σ^2,

$$D = \overline{X}_n - u_{1-\alpha}\frac{\sigma}{\sqrt{n}}$$

*je **dolní odhad** střední hodnoty μ při známém σ^2 se spolehlivostí $1-\alpha$,*

$$H = \overline{X}_n + u_{1-\alpha}\frac{\sigma}{\sqrt{n}}$$

*je **horní odhad** střední hodnoty μ při známém σ^2 se spolehlivostí $1-\alpha$.*

Poznámka: Interval spolehlivosti není ani nutné znát, lze jej snadno odvodit z pivotové statistiky.

Příklad 12. Výrobci pracího prášku přináší na trh nové balení, s nímž spotřebitel obdrží aviváž zdarma. Hmotnost uvedená výrobcem je 5 kg. Prvních 15 vyrobených balení vážilo následující hmotnosti: 5,02; 5; 4,97; 4,99; 5; 4,99; 4,99; 4,98; 5,01; 5,01; 5; 4,96; 5; 5; 4,99. Výrobní linka je nastavená na 5 kg s povolenou směrodatnou odchylkou 0,05 kg. Sestrojte 95 % interval pro střední hodnotu za předpokladu, že se hmotnost řídí normálním rozdělením.

Řešení: Zde máme sestrojit interval spolehlivosti pro střední hodnotu, přičemž rozptyl známe ($\sigma^2 = 0,05^2$). Využijeme pivotovou statistiku U, ale ještě předtím spočítáme výběrový průměr:

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{15} X_i = 4,994.$$

Potřebujeme vytvořit 95 procentní interval, tzn. že $100(1-\alpha)\% = 95\% \Rightarrow \alpha = 0,05$. Nyní můžeme vytvořit interval, který dosadíme do pivotové statistiky:

$$P\left(u_{\frac{\alpha}{2}} \leq \frac{\overline{X}_n - \mu}{\sigma}\sqrt{n} \leq u_{1-\frac{\alpha}{2}}\right) = 0,95,$$

$$P\left(u_{0,025}\frac{\sigma}{\sqrt{n}} \leq \overline{X}_n - \mu \leq u_{0,975}\frac{\sigma}{\sqrt{n}}\right) = 0,95,$$

$$P\left(\overline{X}_n - u_{0,975}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X}_n - u_{0,025}\frac{\sigma}{\sqrt{n}}\right) = 0,95,$$

$$P\left(\overline{X}_n - u_{0,975}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X}_n + u_{0,975}\frac{\sigma}{\sqrt{n}}\right) = 0,95,$$

$$P\left(4,994 - \frac{1,96 \cdot 0,05}{\sqrt{15}} \leq \mu \leq 4,994 + \frac{1,96 \cdot 0,05}{\sqrt{15}}\right) = 0,95,$$

$$P\big(4,9686 \leq \mu \leq 5,0193\big) = 0,95.$$

Střední hodnota hmotnosti balení je s pravděpodobností 95 % v intervalu
$[D,H] = [4,9686; 5,0193]$.

Definice 4.3.1.2. *Nechť $\mathbf{X}_n$ je náhodný výběr z $N(\mu;\sigma^2)$. Potom, pro neznámé μ, když je σ^2 neznámé, je pivotová statistika*

$$T = \frac{\overline{X}_n - \mu}{\frac{S}{\sqrt{n}}} \sim t(n-1),$$

kde $\overline{X}_n$ je výběrový průměr a S je výběrová směrodatná odchylka. Potom

$$[D,H] = \left[\overline{X}_n - t_{1-\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}}, \overline{X}_n + t_{1-\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}}\right]$$

je $100(1-\alpha)\%$ interval spolehlivosti pro střední hodnotu μ při neznámém σ^2,

$$D = \overline{X}_n - t_{1-\alpha}(n-1)\frac{S}{\sqrt{n}}$$

*je **dolní odhad** střední hodnoty μ při neznámém σ^2 se spolehlivostí $1-\alpha$,*

$$H = \overline{X}_n + t_{1-\alpha}(n-1)\frac{S}{\sqrt{n}}$$

*je **horní odhad** střední hodnoty μ při neznámém σ^2 se spolehlivostí $1-\alpha$.*

Příklad 13. Na vybrané pobočce v bance je měřena délka uzavírání běžného účtu, která má normální rozdělení, s následující výsledky (v minutách): 5,4; 7,7; 5,4; 6,2; 6,8; 6,7; 4,9; 8,1; 7,9; 5,3; 8,5; 6,5; 6,3; 6,5; 6,4; 5,4. Sestrojte 95 % interval spolehlivosti pro očekávánou délku uzavírání smluv.

Řešení: Nyní neznáme rozptyl, proto použijeme příslušnou pivotovou statistiku T. Spočítejme výběrový průměr a směrodatnou odchylku:

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{16} X_i = 6,5,$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{16}(X_i - \overline{X}_n)^2 = 1,1906 \Rightarrow S = 1,0911.$$

Zde je rovněž $\alpha = 0,05$. Dosadíme:

$$P\left(t_{\frac{\alpha}{2}}(n-1) \leq \frac{\overline{X}_n - \mu}{S}\sqrt{n} \leq t_{1-\frac{\alpha}{2}}(n-1)\right) = 0,95,$$

$$P\left(t_{0,025}(15)\frac{S}{\sqrt{n}} \leq \overline{X}_n - \mu \leq t_{0,975}(15)\frac{S}{\sqrt{n}}\right) = 0,95,$$

$$P\left(\overline{X}_n - t_{0,975}(15)\frac{S}{\sqrt{n}} \leq \mu \leq \overline{X}_n + t_{0,975}(15)\frac{S}{\sqrt{n}}\right) = 0,95,$$

$$P\left(6,5 - \frac{2,131 \cdot 1,0911}{\sqrt{16}} \leq \mu \leq 6,5 + \frac{2,131 \cdot 1,0911}{\sqrt{16}}\right) = 0,95,$$

$$P\big(5,9187 \leq \mu \leq 7,0812\big) = 0,95.$$

S pravděpodobností 95 % je tedy očekáváná střední hodnota délky uzavírání smlouvy v intervalu $[D, H] = [5,9187; 7,0812]$.

Definice 4.3.1.3. *Nechť* $\mathbf{X}_n$ *je náhodný výběr z* $N(\mu; \sigma^2)$. *Potom, pro neznámé* σ^2, *když je* μ *neznámé, je pivotová statistika*

$$K = (n-1)\frac{S^2}{\sigma^2} \sim \chi^2(n-1),$$

kde S^2 *je výběrový rozptyl. Potom*

$$[D, H] = \left[\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}, \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}\right]$$

je $100(1-\alpha)\%$ *interval spolehlivosti pro rozptyl* σ^2 *při neznámém* μ,

$$D = \frac{(n-1)S^2}{\chi^2_{1-\alpha}(n-1)}$$

je ***dolní odhad*** *rozptylu* σ^2 *při neznámém* μ *se spolehlivostí* $1-\alpha$,

$$H = \frac{(n-1)S^2}{\chi^2_{\alpha}(n-1)}$$

je ***horní odhad*** *rozptylu* σ^2 *při neznámém* μ *se spolehlivostí* $1-\alpha$.

Příklad 14. Sestrojte 90 % interval spolehlivosti pro rozptyl σ^2, pokud jsme ze 30 údajů při kontrole dodržování termínu splatnosti úvěru v bance spočítali výběrovou směrodatnou odchylku $S = 0,23$. Termín splatnosti úvěru má normální rozdělení.

Řešení: Na rozdíl od předchozích dvou příkladů, kde jsme odhadovali interval pro střední hodnotu, zde máme odhadnout interval spolehlivosti pro rozptyl. Zde je $\alpha = 0,1$. Využijeme statistiku K:

$$P\left(\chi^2_{\frac{\alpha}{2}}(n-1) \leq (n-1)\frac{S^2}{\sigma^2} \leq \chi^2_{1-\frac{\alpha}{2}}(n-1)\right) = 0,9,$$

$$P\left(\frac{\chi^2_{0,05}(29)}{(n-1)S^2} \leq \frac{1}{\sigma^2} \leq \frac{\chi^2_{0,95}(29)}{(n-1)S^2}\right) = 0,9,$$

$$P\left(\frac{(n-1)S^2}{\chi^2_{0,95}(29)} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{0,05}(29)}\right) = 0,9,$$

$$P\left(\frac{29 \cdot 0,0529}{42,6} \leq \sigma^2 \leq \frac{29 \cdot 0,0529}{17,7}\right) = 0,9,$$

$$P\big(0,0360 \leq \sigma^2 \leq 0,0866\big) = 0,9.$$

Hledaný 90 % interval spolehlivosti pro σ^2 je tedy $[D, H] = [0,0360; 0,0886]$.

Definice 4.3.1.4. *Nechť $\mathbf{X}_n$ je náhodný výběr z $N(\mu;\sigma^2)$. Potom, pro neznámé σ^2, když je μ známé, je pivotová statistika*

$$Q = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 \sim \chi^2(n).$$

Potom

$$[D, H] = \left[\frac{\sum\limits_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{1-\frac{\alpha}{2}}(n)}, \frac{\sum\limits_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{\frac{\alpha}{2}}(n)}\right]$$

je $100(1-\alpha)$% interval spolehlivosti pro rozptyl σ^2 při známém μ,

$$D = \frac{\sum\limits_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{1-\alpha}(n)}$$

*je **dolní odhad** rozptylu σ^2 při neznámém μ se spolehlivostí $1-\alpha$,*

$$H = \frac{\sum\limits_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{\alpha}(n)}$$

*je **horní odhad** rozptylu σ^2 při známém μ se spolehlivostí $1-\alpha$.*

Příklad 15. Odhadněte 99 % interval spolehlivosti pro rozptyl σ^2, jestliže jsme během 4 dnů sledovali vývoj hodnoty portfolia s normálním rozdělením a máme následující výsledky: 415,9; 418,8; 418,4; 419,1. Střední hodnota je $\mu = 416,7$.

Řešení: Nyní známe střední hodnotu, potřebujeme interval spolehlivosti pro rozptyl. Použijeme pivotovou statistiku Q:

$$P\left(\chi^2_{\frac{\alpha}{2}}(n) \le \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 \le \chi^2_{1-\frac{\alpha}{2}}(n)\right) = 0,99,$$

$$P\left(\frac{\chi^2_{0,005}(4)}{13,7} \le \frac{1}{\sigma^2} \le \frac{\chi^2_{0,995}(4)}{13,7}\right) = 0,99,$$

$$P\left(\frac{13,7}{14,9} \le \sigma^2 \le \frac{13,7}{0,207}\right) = 0,99,$$

$$P\big(0,9194 \le \sigma^2 \le 66,1835\big) = 0,99.$$

Hledaný 99 % interval spolehlivosti pro σ^2 je tedy $[D,H] = [0,9194; 66,1835]$.

Doposud jsme počítali oboustanné intervaly spolehlivosti. Nyní si ukážeme, jak se počítají dolní a horní odhady.

Příklad 16. Náhodný výběr $\mathbf{X}_n$ má normální rozdělení a představuje počet odpracovaných dnů za měsíc. Údaje jsou následující: 12; 23; 21; 17; 19; 16; 13; 25; 16. Určete 95 % pravostranný interval spolehlivosti pro střední hodnotu μ.

Řešení: Hledáme horní odhad pro neznámou střední hodnotu při neznámém rozptylu. Použijeme pivotovou statistiku T:

$$P\left(t_\alpha(n-1) \le \frac{\overline{X}_n-\mu}{S}\sqrt{n}\right) = 0,95,$$

$$P\left(t_\alpha(n-1)\frac{S}{\sqrt{n}} \le \overline{X}_n-\mu\right) = 0,95,$$

$$P\left(\mu \le \overline{X}_n - t_\alpha(n-1)\frac{S}{\sqrt{n}}\right) = 0,95,$$

$$P\left(\mu \le \overline{X}_n + t_{1-\alpha}(n-1)\frac{S}{\sqrt{n}}\right) = 0,95,$$

$$P\left(\mu \le 18 + t_{0,95}(8)\frac{4,3874}{\sqrt{9}}\right) = 0,95,$$

$$P(\mu \le 20,7201) = 0,95.$$

Horní odhad pro střední hodnotu se spolehlivostí 95 % je $H = 20,7201$.

Příklad 17. Sedm balíčků cukru má následující hmotnosti v gramech: 993; 1001; 999;

1005; 1000; 997; 995. Odhadněte 95 % levostranný interval spolehlivosti pro rozptyl σ^2, jestliže hmotnost má normální rozdělení.

Řešení: Zde potřebujeme pivotovou statistiku pro odhad rozptylu při neznámé střední hodnotě:

$$P\left((n-1)\frac{S^2}{\sigma^2} \le \chi^2_{1-\alpha}(n-1)\right) = 0,95,$$

$$P\left(\frac{(n-1)S^2}{\chi^2_{0,95}(6)} \le \sigma^2\right) = 0,95,$$

$$P\left(\frac{6 \cdot 15,9523}{12,6} \le \sigma^2\right) = 0,95,$$

$$P\big(7,5963 \le \sigma^2\big) = 0,95.$$

Dolní odhad pro rozptyl se spolehlivostí 95 % je $D = 7,5963$.

Zatím jsme se zabývali intervalovými odhady pro jeden náhodný výběr. V další části ukážeme, jak určit interval spolehlivosti pro *dva nezávislé výběry*.

Definice 4.3.1.5. *Nechť $\boldsymbol{X}_{n_1}$ je náhodný výběr rozsahu n_1 z normálního rozdělení $N(\mu_1; \sigma_1^2)$, $\overline{X}_{n_1}$ je jeho výběrový průměr. Dále nechť $\boldsymbol{X}_{n_2}$ je náhodný výběr rozsahu n_2 z normálního rozdělení $N(\mu_2; \sigma_2^2)$, $\overline{X}_{n_2}$ je jeho výběrový průměr. Předpokládejme, že oba výběry jsou nezávislé. Potom, pro neznámý rozdíl středních hodnot $\mu_1 - \mu_2$, když jsou σ_1^2 a σ_2^2 známé, je statistika*

$$U_{\overline{X}_{n_1}-\overline{X}_{n_2}} = \frac{\overline{X}_{n_1} - \overline{X}_{n_2} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0; 1).$$

Potom

$$[D, H] = \left[\overline{X}_{n_1} - \overline{X}_{n_2} - u_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, \overline{X}_{n_1} - \overline{X}_{n_2} + u_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right]$$

je $100(1-\alpha)\%$ interval spolehlivosti pro rozdíl středních hodnot $\mu_1 - \mu_2$ když jsou σ_1^2 a σ_2^2 známé,

$$D = \overline{X}_{n_1} - \overline{X}_{n_2} - u_{1-\alpha}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

*je **dolní odhad** rozdílu středních hodnot $\mu_1 - \mu_2$, když jsou σ_1^2 a σ_2^2 známé se spolehlivostí $1-\alpha$,*

$$H = \overline{X}_{n_1} - \overline{X}_{n_2} + u_{1-\alpha}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

*je **horní odhad** rozdílu středních hodnot $\mu_1 - \mu_2$, když jsou σ_1^2 a σ_2^2 známé se spolehlivostí $1-\alpha$.*

Příklad 18. Zkoušku dělalo 9 studentů; 5 mužů a 4 ženy. Jejich výsledky byly následující: 63; 75; 78; 80; 93 a 66; 79; 81; 96. Spočítejte 95 % interval spolehlivosti pro rozdíl středních hodnot výsledků mužů a žen u zkoušky, pokud se výsledky řídí normálním rozdělením. Rozptyly jsou $\sigma_1^2 = 81$ a $\sigma_2^2 = 49$.

Řešení: Využijeme statistiku $U_{\overline{X}_{n_1}-\overline{X}_{n_2}}$ pro rozdíl středních hodnot, kde $\alpha = 0{,}005$. Spočítejme výběrové průměry:

$$\overline{X}_{n_1} = \frac{1}{5}\sum_{i=1}^{5} X_{1_i} = 77{,}8,$$

$$\overline{X}_{n_2} = \frac{1}{4}\sum_{j=1}^{4} X_{2_j} = 80{,}5.$$

Nyní máme všechno, abychom vytvořili interval:

$$P\left(u_{\frac{\alpha}{2}} \leq \frac{\overline{X}_{n_1} - \overline{X}_{n_2} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \leq u_{1-\frac{\alpha}{2}}\right) = 0,95,$$

$$P\left(\overline{X}_{n_1} - \overline{X}_{n_2} - u_{0,975}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq \right.$$

$$\left. \leq \overline{X}_{n_1} - \overline{X}_{n_2} + u_{0,975}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right) = 0,95,$$

$$P\big(-13,1543 \leq \mu_1 - \mu_2 \leq 7,7543\big) = 0,95.$$

Pro rozdíl středních hodnot $\mu_1 - \mu_2$ je 95 % interval spolehlivosti $[D, H] = [-13,1543; 7,7543]$.

Definice 4.3.1.6. *Nechť* $\boldsymbol{X}_{n_1}$ *je náhodný výběr rozsahu* n_1 *z normálního rozdělení* $N(\mu_1;\sigma_1^2)$, $\overline{X}_{n_1}$ *je jeho výběrový průměr a* S_1^2 *jeho výběrový rozptyl. Dále nechť* $\boldsymbol{X}_{n_2}$ *je náhodný výběr rozsahu* n_2 *z normálního rozdělení* $N(\mu_2;\sigma_2^2)$, $\overline{X}_{n_2}$ *je jeho výběrový průměr a* S_2^2 *jeho výběrový rozptyl. Předpokládejme, že oba výběry jsou nezávislé. Potom, pro neznámý rozdíl středních hodnot* $\mu_1-\mu_2$, *když jsou* σ_1^2 *a* σ_2^2 *neznámé, ale platí* $\sigma_1^2=\sigma_2^2=\sigma^2$, *je statistika*

$$T_{\overline{X}_{n_1}-\overline{X}_{n_2}}=\frac{\overline{X}_{n_1}-\overline{X}_{n_2}-(\mu_1-\mu_2)}{S_{12}}\sqrt{\frac{n_1n_2}{n_1+n_2}}\sim t(n_1+n_2-2),$$

kde

$$S_{12}^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}.$$

Potom

$$[D,H]=\left[\overline{X}_{n_1}-\overline{X}_{n_2}-t_{1-\frac{\alpha}{2}}(n_1+n_2-2)S_{12}\sqrt{\frac{n_1+n_2}{n_1n_2}},\right.$$

$$\left.\overline{X}_{n_1}-\overline{X}_{n_2}+t_{1-\frac{\alpha}{2}}(n_1+n_2-2)S_{12}\sqrt{\frac{n_1+n_2}{n_1n_2}}\right]$$

je $100(1-\alpha)\%$ *interval spolehlivosti pro rozdíl středních hodnot* $\mu_1-\mu_2$ *když jsou* σ_1^2 *a* σ_2^2 *neznámé, ale platí* $\sigma_1^2=\sigma_2^2=\sigma^2$,

$$D=\overline{X}_{n_1}-\overline{X}_{n_2}-t_{1-\alpha}(n_1+n_2-2)S_{12}\sqrt{\frac{n_1+n_2}{n_1n_2}}$$

je ***dolní odhad*** *rozdílu středních hodnot* $\mu_1-\mu_2$, *když jsou* σ_1^2 *a* σ_2^2 *neznámé, ale platí* $\sigma_1^2=\sigma_2^2=\sigma^2$, *se spolehlivostí* $1-\alpha$,

$$H=\overline{X}_{n_1}-\overline{X}_{n_2}+t_{1-\alpha}(n_1+n_2-2)S_{12}\sqrt{\frac{n_1+n_2}{n_1n_2}}$$

je ***horní odhad*** *rozdílu středních hodnot* $\mu_1-\mu_2$, *když jsou* σ_1^2 *a* σ_2^2 *neznámé, ale platí* $\sigma_1^2=\sigma_2^2=\sigma^2$, *se spolehlivostí* $1-\alpha$.

Příklad 19. Vyřešte předchozí příklad za předpokladu, že se rozptyly rovnají $\sigma_1^2=\sigma_2^2$.

Řešení: Zde kromě výběrových průměrů, které jsme už vypočítali v předchozím příkladě, potřebujeme rovněž výběrové rozptyly:

$$S_1^2=\frac{1}{4}\sum_{i=1}^{5}(X_{1_i}-\overline{X}_{n_1})^2=115{,}7,$$

$$S_2^2=\frac{1}{3}\sum_{j=1}^{4}(X_{2_j}-\overline{X}_{n_2})^2=151,$$

$$S_{12}^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} = 130,8285.$$

Použijeme statistiku $T_{\overline{X}_{n_1}-\overline{X}_{n_2}}$:

$$P\left(t_{\frac{\alpha}{2}} \leq \frac{\overline{X}_{n_1}-\overline{X}_{n_2}-(\mu_1-\mu_2)}{S_{12}}\sqrt{\frac{n_1 n_2}{n_1+n_2}} \leq t_{1-\frac{\alpha}{2}}\right) = 0,95,$$

$$P\left(\overline{X}_{n_1}-\overline{X}_{n_2}-t_{0,975}(n_1+n_2-2)S_{12}\sqrt{\frac{n_1+n_2}{n_1 n_2}} \leq \mu_1-\mu_2 \leq \right.$$

$$\left. \leq \overline{X}_{n_1}-\overline{X}_{n_2}+t_{0,975}(n_1+n_2-2)S_{12}\sqrt{\frac{n_1+n_2}{n_1 n_2}}\right) = 0,95,$$

$$P\left(-20,8463 \leq \mu_1-\mu_2 \leq 15,4463\right) = 0,95.$$

95 % interval spolehlivosti pro rozdíl středních hodnot $\mu_1-\mu_2$ je tedy $[D,H] = [-20,8463;\ 15,4463]$.

Definice 4.3.1.7. *Nechť $\mathbf{X}_{n_1}$ je náhodný výběr rozsahu n_1 z normálního rozdělení $N(\mu_1;\sigma_1^2)$, $\overline{X}_{n_1}$ je jeho výběrový průměr a S_1^2 jeho výběrový rozptyl. Dále nechť $\mathbf{X}_{n_2}$ je náhodný výběr rozsahu n_2 z normálního rozdělení $N(\mu_2;\sigma_2^2)$, $\overline{X}_{n_2}$ je jeho výběrový průměr a S_2^2 jeho výběrový rozptyl. Předpokládejme, že oba výběry jsou nezávislé. Potom, pro neznámý podíl rozptylů $\frac{\sigma_1^2}{\sigma_2^2}$, když jsou μ_1, μ_2, σ_1^2 a σ_2^2 neznámé, je statistika*

$$F_{\sigma_1^2/\sigma_2^2}=\frac{S_1^2}{S_2^2}\cdot\frac{\sigma_2^2}{\sigma_1^2}\sim F(n_1-1;n_2-1).$$

Potom

$$[D,H]=\left[\frac{S_1^2}{S_2^2}\cdot\frac{1}{F_{1-\frac{\alpha}{2}}(n_1-1,n_2-1)},\frac{S_1^2}{S_2^2}\cdot\frac{1}{F_{\frac{\alpha}{2}}(n_1-1,n_2-1)}\right]$$

je $100(1-\alpha)\%$ interval spolehlivosti pro neznámý podíl rozptylů $\frac{\sigma_1^2}{\sigma_2^2}$, když jsou μ_1, μ_2, σ_1^2 a σ_2^2 neznámé,

$$D=\frac{S_1^2}{S_2^2}\cdot\frac{1}{F_{1-\alpha}(n_1-1,n_2-1)}$$

*je **dolní odhad** pro neznámý podíl rozptylů $\frac{\sigma_1^2}{\sigma_2^2}$, když jsou μ_1, μ_2, σ_1^2 a σ_2^2 neznámé se spolehlivostí $1-\alpha$,*

$$H=\frac{S_1^2}{S_2^2}\cdot\frac{1}{F_{\alpha}(n_1-1,n_2-1)}$$

*je **horní odhad** pro neznámý podíl rozptylů $\frac{\sigma_1^2}{\sigma_2^2}$, když jsou μ_1, μ_2, σ_1^2 a σ_2^2 neznámé se spolehlivostí $1-\alpha$.*

Příklad 20. Při zkoumání inteligence u dětí ve dvou skupinách na mateřské škole jsme obdrželi následující výsledky s normálním rozdělením: 71; 76; 82; 90; 93; 101; 103; 112 a 75; 77; 80; 83; 92; 95; 99; 100. Spočítejte 99 % interval spolehlivosti pro podíl rozptylů.

Řešení: Při počítání hledaného intervalu použijeme statistiku $F_{\sigma_1^2/\sigma_2^2}$, ale nejdříve vypočítáme výběrové průměry a rozptyly:

$$\overline{X}_{n_1}=\frac{1}{8}\sum_{i=1}^{8}X_{1_i}=91,$$

$$S_1^2=\frac{1}{7}\sum_{i=1}^{8}(X_{1_i}-\overline{X}_{n_1})^2=199{,}4285,$$

$$\overline{X}_{n_2}=\frac{1}{8}\sum_{j=1}^{8}X_{2_j}=87{,}625,$$

$$S_2^2 = \frac{1}{7}\sum_{j=1}^{8}(X_{2_j} - \overline{X}_{n_2})^2 = 101,125.$$

Získané hodnoty dosadíme do statistiky:

$$P\left(F_{\frac{\alpha}{2}}(n_1-1,n_2-1) \leq \frac{S_1^2}{S_2^2}\cdot\frac{\sigma_2^2}{\sigma_1^2} \leq F_{1-\frac{\alpha}{2}}(n_1-1,n_2-1)\right) = 0,99,$$

$$P\left(\frac{S_1^2}{S_2^2}\frac{1}{F_{1-\frac{\alpha}{2}}(n_1-1,n_2-1)} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{S_1^2}{S_2^2}\frac{1}{F_{\frac{\alpha}{2}}(n_1-1,n_2-1)}\right) = 0,99,$$

$$P\left(\frac{S_1^2}{S_2^2}\frac{1}{F_{0,995}(7,7)} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{S_1^2}{S_2^2}F_{0,995}(n_2,n_1)\right) = 0,99,$$

$$P\left(0,2219 \leq \frac{\sigma_1^2}{\sigma_2^2} \leq 17,5220\right) = 0,99.$$

Dostali jsme 95 % interval spolehlivosti pro podíl rozptylů $[D, H] = [0,2219;\ 17,5220]$.

Posledním případem, kterým se zde budeme zabývat, jsou tzv. *párové výběry*.

Věta 4.3.1.8. *Nechť* $\boldsymbol{X}_1 = (X_1, Y_1)', \ldots, \boldsymbol{X}_n = (X_n, Y_n)'$ *je náhodný výběr z dvourozměrného normálního rozdělení* $N_2(\mu;\Sigma)$*, s parametry* $\mu = \begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}$ *a* $\Sigma = \begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}$*, kde* $\mu_1,\ \mu_2 \in \mathbb{R}$*,* $\sigma_1^2 > 0$*,* $\sigma_2^2 > 0$ *a* $\rho \in (0,\ 1)$*. Pro* $i = 1, \ldots, n$ *označme*

$$Z_i = X_i - Y_i,$$

dále

$$\overline{Z}_n = \frac{1}{n}\sum_{i=1}^{n} Z_i$$

je výběrový průměr a

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Z_i - \overline{Z}_n)^2$$

výběrový rozptyl. Potom

$$[D, H] = \left[\overline{Z}_n - t_{1-\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}},\ \overline{Z}_n + t_{1-\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}}\right]$$

je intervalový odhad parametrické funkce $\mu = \mu_1 - \mu_2$ *o spolehlivosti* $1-\alpha$*.*

Důkaz: Viz [2, str. 62].

Spočítejme poslední příklad v této části:

Příklad 21. Nechť $\mathbf{X}_n = (73; 61; 105; 92; 56; 77; 68)'$ je náhodný výběr z normálního rozdělení, který udává váhy zkoumaných žen, vyjádřené v kilogramech, dříve než začaly užívat čaj na hubnutí. Po 7 měsících užívání čaje, se ženy znovu vážily s následujícími výsledky: $\mathbf{Y}_n = (65; 57; 92; 86; 52; 71; 60)'$. Spočítejte 95 % interval spolehlivosti pro rozdíl vah zkoumaných žen.

Řešení: Spočítejme všechny hodnoty uvedené v předchozí větě:

$$\mathbf{Z}_n = (8;\ 4;\ 13;\ 6;\ 4;\ 6;\ 8)',$$

$$\overline{Z}_n = \frac{1}{7}\sum_{i=1}^{7} Z_i = 7,$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Z_i - \overline{Z_n})^2 = 12.$$

Vytvoříme interval:

$$P\left(t_{\frac{\alpha}{2}}(n-1) \le \frac{\overline{Z}_n - \mu}{S}\sqrt{n} \le t_{1-\frac{\alpha}{2}}(n-1)\right) = 0,95,$$

$$P\left(\overline{Z}_n - t_{0,975}(6)\frac{S}{\sqrt{n}} \le \mu \le \overline{Z}_n + t_{0,975}(6)\frac{S}{\sqrt{n}}\right) = 0,95,$$

$$P\big(3,7961 \le \mu \le 10,2038\big) = 0,95.$$

95 % interval spolehlivosti pro rozdíl vah je tedy $[D, H] = [3,7961;\ 10,2038]$.

4.3.2 Intervalové odhady založené na centrální limitní větě

Doposud jsme se zabývali odhady parametrů normálního rozdělení. Nyní uvedeme intervaly spolehlivosti pro parametry z dalších rozdělení, založených na centrální limitní větě.

Věta 4.3.2.1. *Nechť* $\boldsymbol{X}_n = (X_1, ..., X_n)'$ *je náhodný výběr z rozdělení s konečnými druhými momenty, kde* $\mu(\theta) = E(X_i)$ *a* $\sigma^2(\theta) = D(X_i)$*, pro* $i = 1, ..., n$ *a* $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ *je výběrový průměr. Nechť* S_*^2 *je* ***(slabě) konzistentním odhadem*** *rozptylu* $\sigma^2(\theta)$. *Potom, pro neznámou střední hodnotu* $\mu(\theta)$*, je statistika*

$$U_* = \frac{\overline{X}_n - \mu(\theta)}{S_*}\sqrt{n} \overset{A}{\sim} N(0,1).$$

Potom

$$[D, H] = \left[\overline{X}_n - u_{1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}}, \overline{X}_n + u_{1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right],$$

kde S^2 *je výběrový rozptyl, je intervalovým odhadem o* ***asymptotické*** *spolehlivosti* $1-\alpha$ *pro neznámou střední hodnotu* $\mu(\theta)$,

$$D = \overline{X}_n - u_{1-\alpha}\frac{S}{\sqrt{n}}$$

je ***dolní odhad*** *střední hodnoty* $\mu(\theta)$*, o asymptotické spolehlivosti* $1-\alpha$,

$$H = \overline{X}_n + u_{1-\alpha}\frac{S}{\sqrt{n}}$$

je ***horní odhad*** *střední hodnoty* $\mu(\theta)$ *o asymptotické spolehlivosti* $1-\alpha$.

Důkaz: Viz [2, str. 63].

Dříve jsme se již zmínili o tom, že alternativní rozdělení při dostatečně velkém počtu opakování má přibližně normální rozdělení se střední hodnotou θ a rozptylem $\theta(1-\theta)$. Protože $\overline{X}_n$ je konzistentním odhadem střední hodnoty θ, můžeme provést aproximaci při použití statistiky U_* ke konstrukci intervalového odhadu střední hodnoty. Podívejme se, jak v tomto případě bude vypadat interval spolehlivosti:

Důsledek 4.3.2.2. *Nechť $\boldsymbol{X}_n = (X_1, ..., X_n)'$ je náhodný výběr s alternativním rozdělením $A(\theta)$. Potom*

$$[D, H] = \left[\overline{X}_n - u_{1-\frac{\alpha}{2}}\sqrt{\frac{\overline{X}_n(1-\overline{X}_n)}{n}}, \overline{X}_n + u_{1-\frac{\alpha}{2}}\sqrt{\frac{\overline{X}_n(1-\overline{X}_n)}{n}}\right]$$

je intervalovým odhadem o asymptotické spolehlivosti $1-\alpha$ pro neznámý parametr θ,

$$D = \overline{X}_n - u_{1-\alpha}\sqrt{\frac{\overline{X}_n(1-\overline{X}_n)}{n}}$$

*je **dolní odhad** neznámého parametru θ o asymptotické spolehlivosti $1-\alpha$,*

$$H = \overline{X}_n + u_{1-\alpha}\sqrt{\frac{\overline{X}_n(1-\overline{X}_n)}{n}}$$

*je **horní odhad** neznámého parametru θ o asymptotické spolehlivosti $1-\alpha$.*

Stejně tak aproximujeme střední hodnotu a rozptyl u Poissonova rozdělení při tvoření intervalového odhadu, kde se oba rovnají λ:

Důsledek 4.3.2.3. *Nechť $\boldsymbol{X}_n = (X_1, ..., X_n)'$ je náhodný výběr s Poissonovým rozdělením $Po(\lambda)$. Potom*

$$[D, H] = \left[\overline{X}_n - u_{1-\frac{\alpha}{2}}\sqrt{\frac{\overline{X}_n}{n}}, \overline{X}_n + u_{1-\frac{\alpha}{2}}\sqrt{\frac{\overline{X}_n}{n}}\right]$$

je intervalovým odhadem o asymptotické spolehlivosti $1-\alpha$ pro neznámý parametr $\lambda \in (0,\infty)$,

$$D = \overline{X}_n - u_{1-\alpha}\sqrt{\frac{\overline{X}_n}{n}}$$

*je **dolní odhad** neznámého parametru λ o asymptotické spolehlivosti $1-\alpha$,*

$$H = \overline{X}_n + u_{1-\alpha}\sqrt{\frac{\overline{X}_n}{n}}$$

*je **horní odhad** neznámého parametru λ o asymptotické spolehlivosti $1-\alpha$.*

Podívejme se, jak to vypadá na příkladech:

Příklad 22. Při provádění ankety k novému tarifu v telekomunikační společnosti bylo zkoumáno 150 lidí. 126 z nich mělo dobré zkušenosti s tarifem. Určete 95 % interval spolehlivosti pro podíl nespokojených klientů.

Řešení: Máme alternativní rozdělení, kde je náhodná veličina definovaná jako

$$X_i = \begin{cases} 1 & \text{když i-tý zákazník není spokojený,} \\ 0 & \text{když i-tý zákazník je spokojený.} \end{cases}$$

pro $i = 1, ..., 150$. $\overline{X}_n$ je v tomto případě $\overline{X}_n = \frac{24}{150} = 0,16$. Využijme odpovídající statistiku, když je střední hodnota θ a rozptyl $\theta(1-\theta)$:

$$P\left(u_{\frac{\alpha}{2}} \leq \frac{\overline{X}_n - \theta}{\sqrt{\theta(1-\theta)}}\sqrt{n} \leq u_{1-\frac{\alpha}{2}}\right) = 0,95.$$

Provedeme dříve zmíněnou aproximaci:

$$P\left(u_{\frac{\alpha}{2}} \leq \frac{\overline{X}_n - \theta}{\sqrt{\overline{X}_n(1-\overline{X}_n)}}\sqrt{n} \leq u_{1-\frac{\alpha}{2}}\right) = 0,95,$$

$$P\left(\overline{X}_n - u_{0,975}\sqrt{\frac{\overline{X}_n(1-\overline{X}_n)}{n}} \leq \theta \leq \overline{X}_n + u_{0,975}\sqrt{\frac{\overline{X}_n(1-\overline{X}_n)}{n}}\right) = 0,95,$$

$$P\big(0,1013 \leq \theta \leq 0,2186\big) = 0,95.$$

Můžeme s 95 % pravděpodobností říct, že je procento nespokojených klientů v intervalu $[D, H] = [0,1013; 0,2186]$, tj. mezi 10,13 % a 21,86 %.

Příklad 23. Vedoucí plaveckého stadionu posunul zavírací dobu v letní sezoně z 20:00 na 21:00 hodinu. Během letní sezony, která začíní 21. června a končí 22. října, přišlo během této poslední hodiny 651 plavců. Odhadněte 90 % interval spolehlivosti pro střední hodnotu příchodu plavců během poslední hodiny, jestliže počet plavců přicházejících v určitém čase má Poissonovo rozdělení.

Řešení: Letní sezona v našem případě trvá 93 dnů. Podíl plavců, kteří přijdou během poslední hodiny za den $\overline{X}_n = \frac{651}{93} = 7$. Vytvořme odpovídající interval a dosaďme jej, přičemž střední hodnota a rozptyl se rovnají λ:

$$P\left(u_{\frac{\alpha}{2}} \leq \frac{\overline{X}_n - \lambda}{\sqrt{\lambda}}\sqrt{n} \leq u_{1-\frac{\alpha}{2}}\right) = 0,9.$$

Opět provedeme aproximaci:

$$P\left(u_{\frac{\alpha}{2}} \leq \frac{\overline{X}_n - \lambda}{\sqrt{\overline{X}_n}}\sqrt{n} \leq u_{1-\frac{\alpha}{2}}\right) = 0,9,$$

$$P\left(\overline{X}_n - u_{0,95}\sqrt{\frac{\overline{X}_n}{n}} \leq \lambda \leq \overline{X}_n + u_{0,95}\sqrt{\frac{\overline{X}_n}{n}}\right) = 0,9,$$

$$P\big(6,5486 \leq \lambda \leq 7,4513\big) = 0,9.$$

Zde je hledaný interval spolehlivosti pro střední hodnotu $[D,H] = [6,5486; 7,4513]$.

4.4 Cvičení

1. Náhodně jsme vybrali 12 čtenářů v knihovně, a ptali jsme se jich, jak dlouho průměrně čtou denně. Odpovědi byly následující (v minutách): 30; 120; 60; 90; 15; 45; 60; 60; 30; 90; 60; 45. Vypočítejte:

   a) výběrový průměr;

   b) výběrový rozptyl;

   c) výběrovou směrodatnou odchylku denního čtení.

   [a) $\overline{X}_n = 58,75$; b) $S^2 = 877,84$; c) $S = 29,6283$]

2. Nechť $\mathbf{X}_n \sim Ro(0;\, \tau)$ je náhodný výběr. Dokažte, že statistika $T_n = 2\overline{X}_n$ je nestranným odhadem parametru τ.

3. Metodou maximální věrohodnosti odhadněte parametr θ náhodného výběru $\mathbf{X}_n$, které je z geometrického rozdělení s pravděpodobnostní funkcí:

$$p(x) = \begin{cases} (1-\theta)^x\theta & \text{pro } x \in \mathbb{N}_0,\ \theta \in (0,\, 1), \\ 0 & \text{jinak.} \end{cases}$$

   $[\hat{X}_{MLE} = \dfrac{1}{1+\overline{X}_n}]$

4. Vyřešte předchozí příklad momentovou metodou a porovnejte výsledky.

   $[\hat{X} = \dfrac{1}{1+\overline{X}_n}]$

5. Při výrobě zahradních hadic se kontroluje jejich tloušťka. Z výsledků 15 měření máme následující údaje (v mm): 10; 9; 10; 11; 9; 12; 11; 10; 10; 13; 12; 8; 14; 11; 9. Předpokládejme, že sledovaná veličina ná normální rozdělení.

   a) Odhadněte 95 % interval spolehlivosti pro střední hodnotu tloušťky hradic, pokud je směrodatná odchylka $\sigma = 2,5$.

b) Kontrola výrobního procesu teprve začíná a směrodatnou odchylku σ ještě neznáme. Odhadněte střední hodnotu tloušťky hadic.

[a) $[D, H] = [9,7998; 11,4001]$; b) $[D, H] = [9,8910; 11,3089]$]

6. Kolik žen bychom museli vybrat, pokud chceme odhadnout jejich střední výšku se spolehlivostí 99 % při směrodatné odchylce 6 cm? Dovolujeme si chybu $\pm 0,5$cm. Výšky se řídí normálním rozdělením.

[$n = 956$]

7. Při testování spotřeby určitého typu motoru při rychlosti 50km/hod, byly zjištěny následující hodnoty (l/100km): 3; 2,9; 3,2; 3,5; 2,8; 3,1; 3; 3,1; 2,4. Odhadněte rozptyl a sestrojte 90 % interval spolehlivosti, jestliže předpokládáme normální rozdělení.

[$[D, H] = [0,0335; 0,1904]$]

8. Během jednoho týdne byl počítán počet lidí jezdících v MHD bez jízdenky. Výsledky mají normální rozdělení a jsou následující: 15; 3; 7; 10; 18; 23; 5. Odhadněte 95 % pravostranný interval spolehlivosti pro rozptyl σ^2, když je $\mu = 12$.

[$H = 149,7695$]

9. Náhodně jsme vybrali 28 studentů, mezi kterými bylo 12 studentů a 16 studentek, kteří odpovídali na otázku, jak dlouho se učí denně ve zkouškovém období. Spočítali jsme, že se studenti učí průměrně 6 hodin denně, zatímco studentky se průměrně učí 8 hodin denně. Předpokládáme, že se výsledky řídí normálním rozdělením. Spočítejte 95 % interval spolehlivosti pro rozdíl středních hodnot, pokud $\sigma_1 = 7$ a $\sigma_2 = 4$.

[$[D, H] = [-6,4190; 2,4190]$]

10. Na rodičovské schůzce mělo 30 rodičů hlasovat, zda souhlasí, že dětí pojedou na výlet do Vídně. Každý rodič musel hlasovat, tedy říct zda souhlasí, nebo ne. Ze 30 rodičů výlet podpořilo právě 18 z nich. Spočítejte 90 % interval spolehlivosti pro podíl rodičů hlasujících pro výlet.

[$[D, H] = [0,4528; 0,7471]$]

Kapitola 5

Testování statistických hypotéz

Poslední kapitolu věnujeme *testování hypotéz*. Částečně navazujeme na předchozí kapitolu v tom smyslu, že budeme používat pivotové statistiky. Tato kapitola je pouze úvodem do testování statistických hypotéz, které se budou během dalších kurzů v rámci studia doplňovat a rozšiřovat.
Máme náhodný výběr $\mathbf{X}_n = (X_1, ..., X_n)'$ rozsahu n z rozdělení pravděpodobnosti s distribuční funkcí $F(x;\theta)$, které závisí na parametru $\theta = (\theta_1, ..., \theta_m)' \in \Theta$, kde $\Theta \subset \mathbb{R}^m$ je parametrický prostor. Otázkou je, zda parametr θ patří do nějaké neprázdné podmnožiny $\Theta_0 \subset \Theta$? Protože nemůžeme s jistotou říct, že patří, nazveme tvrzení $\theta \in \Theta_0$ **nulovou hypotézou**. Stručně označme $H_0 : \theta \in \Theta_0 \subset \Theta$. Zbývající možností je, že patří do podmnožiny $\Theta_1 = \Theta \setminus \Theta_0$, kterou nazveme **alternativní hypotézou**. Označme $H_1 : \theta \in \Theta_1 = \Theta \setminus \Theta_0$. Dále, je-li množina Θ_0 jednobodová, pak říkáme, že hypotéza H_0 je jednoduchá. V opačném případě je hypotéza H_0 složená. Stejně to platí i pro alternativní hypotézu. O platnosti nulové hypotézy rozhodujeme na základě náhodného výběru $\mathbf{X}_n$ tak, že platnost hypotézy H_0 buď zamítneme nebo nezamítneme. Na testování použijeme tzv. **testovací statistiku** $T_n = T(\mathbf{X}_n)$. Na nákladě realizace náhodného výběru spočítáme **hodnotu testovací statistiky** $t_n = T(\mathbf{x}_n)$. Zvolíme vhodný **kritický obor** W, pak, je-li $t_n \in W$, hypotézu $\theta \in \Theta_0$ zamítáme; pokud ne, hypotézu nezamítáme. Může se stát, že hypotézu zamítneme, i když je správná, v tomto případě se dopustíme **chyby prvního druhu**. Jestliže hypotézu nezamítneme, i když není správná, dopustíme se **chyby druhého druhu**.
Poznámka: Pakliže není v textu uvedeno jinak, považujeme za $\alpha = 0,05$. Číslo α se nazývá **hladinou významnosti**.
Podrobnou teorii naleznete v [2].
Ve 4. kapitole jsme již odvodili intervalové odhady, pomocí kterých dostáváme celou řadu kritických oborů testů. Přehled takto získaných testů uvidíme v následujících tabulkách:

H_0	H_1	Hypotézu H_0 zamítáme, pokud	Předpoklady
$\mu = \mu_0$	$\mu \neq \mu_0$	$\lvert\overline{X} - \mu_0\rvert\sqrt{n} \geq \sigma u_{1-\frac{\alpha}{2}}$	σ^2 známé
$\mu = \mu_0$	$\mu > \mu_0$	$(\overline{X} - \mu_0)\sqrt{n} \geq \sigma u_{1-\alpha}$	σ^2 známé
$\mu = \mu_0$	$\mu < \mu_0$	$(\overline{X} - \mu_0)\sqrt{n} \leq -\sigma u_{1-\alpha}$	σ^2 známé
$\mu = \mu_0$	$\mu \neq \mu_0$	$\lvert\overline{X} - \mu_0\rvert\sqrt{n} \geq St_{1-\frac{\alpha}{2}}(n-1)$	σ^2 neznámé
$\mu = \mu_0$	$\mu > \mu_0$	$(\overline{X} - \mu_0)\sqrt{n} \geq St_{1-\alpha}(n-1)$	σ^2 neznámé
$\mu = \mu_0$	$\mu < \mu_0$	$(\overline{X} - \mu_0)\sqrt{n} \leq -St_{1-\alpha}(n-1)$	σ^2 neznámé
$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$\frac{(n-1)S^2}{\sigma_0^2} \notin \left(\chi^2_{\frac{\alpha}{2}}(n-1), \chi^2_{1-\frac{\alpha}{2}}(n-1)\right)$	μ neznámé
$\sigma^2 = \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$\frac{(n-1)S^2}{\sigma_0^2} \geq \chi^2_{1-\alpha}(n-1)$	μ neznámé
$\sigma^2 = \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$\frac{(n-1)S^2}{\sigma_0^2} \leq \chi^2_{\alpha}(n-1)$	μ neznámé

Tabulka 5.1.[1] Tabulka testů pro jeden náhodný výběr z normálního rozdělení.

H_0	H_1	Hypotézu H_0 zamítáme, pokud	Předpoklady
$\mu_1 = \mu_2$	$\mu_1 \neq \mu_2$	$\lvert\overline{X}_{n_1} - \overline{X}_{n_2}\rvert \geq u_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	σ_1^2 a σ_2^2 známé
$\mu_1 = \mu_2$	$\mu_1 \neq \mu_2$	$\lvert\overline{X}_{n_1} - \overline{X}_{n_2}\rvert \geq t_{1-\frac{\alpha}{2}}(n_1 + n_2 - 2)S_{12}\sqrt{\frac{n_1+n_2}{n_1 n_2}}$	$\sigma_1^2 = \sigma_2^2$ neznámé
$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 \neq \sigma_2^2$	$\frac{S_1^2}{S_2^2} \notin \left(F_{\frac{\alpha}{2}}(n_1-1, n_2-1),\right.$ $\left.F_{1-\frac{\alpha}{2}}(n_1-1, n_2-1)\right)$	μ_1 a μ_2 neznámé

Tabulka 5.2.[2] Tabulka testů pro dva nezávislé náhodné výběry z normálního rozdělení.

H_0	H_1	Hypotézu H_0 zamítáme, pokud	Předpoklady
$\mu = \mu_0$	$\mu \neq \mu_0$	$\frac{\lvert\overline{X}-\mu_0\rvert}{S_*}\sqrt{n} \geq u_{1-\frac{\alpha}{2}}$	$0 < \sigma^2(\theta) < \infty$
$\mu = \mu_0$	$\mu \neq \mu_0$	$\frac{\lvert\overline{X}-\mu_0\rvert}{\sqrt{\overline{X}}}\sqrt{n} \geq u_{1-\frac{\alpha}{2}}$	$\mathbf{X}_n \sim Po(\mu)$
$p = p_0$	$p \neq p_0$	$\frac{\lvert\overline{X}-p_0\rvert}{\sqrt{p_0(1-p_0)}}\sqrt{n} \geq u_{1-\frac{\alpha}{2}}$	$\mathbf{X}_n \sim A(p)$

Tabulka 5.3.[3] Tabulka asymptotických testů pro náhodné výběry.

Na prvním příkladě ukážeme všeobecný postup, jak se provádí testování statistických hypotéz, které většinou budeme používat i k řešení ostatních úloh.

[1] Převzato z [2, str. 75].
[2] Převzato z [2, str. 75].
[3] Převzato z [2, str. 75].

Příklad 1.[4] Zástupci ekologického hnutí vystupují proti výstavbě nové továrny v oblasti, ve které je životní prostředí poznamenáno průmyslovou činností. Jedním z argumentů, který používají, je nízká porodní váha novorozenců v dané oblasti. Průměrná hmotnost 40 náhodně vybraných novorozenců narozených v této oblasti byla 3010 g. Má smysl použít tento argument proti výstavbě nové továrny, jestliže porodní váha zdravé populace má normální rozdělení se střední hodnotou 3300 g a směrodatnou odchylkou 476 g?

Řešení: Zaprvé, musíme formulovat nulovou a alternativní hypotézu. Nulová hypotéza je taková, že vycházíme ze současného stavu a ověřujeme její platnost. Zde předpokládáme, že střední hodnota porodní váhy je 3300 g, takže $H_0 : \mu = 3300$. Cílem je zamítnout hypotézu, a proto za alternativní zvolíme to, co chceme dokázat; zde je to nižší porodní váha, takže $H_1 : \mu < 3300$.

Zadruhé, musíme vybrat vhodné testovací kritérium. Volba záleží na hypotéze, kterou testujeme a rozdělení, které má sledovaná veličina. Zde testujeme hypotézu o střední hodnotě normálního rozdělení, přičemž rozptyl známe, takže použijeme už známou pivotovou statistiku U.

Zatřetí, zvolíme hladinu významnosti α. Zde není přesně uvedeno, jaká je hladina významnosti, a proto bereme 5 %, tj. $\alpha = 0,05$.

Začtvrté, musíme stanovit kritický obor W. Stanovujeme ho podle alternativní hypotézy. Zde máme levostrannou hypotézu, protože obor možných hodnot parametru μ je vymezen nalevo od 3300. Kritický obor je tedy
$W = \{-\infty; u_{0,05}\} = \{-\infty; -u_{0,95}\} = \{-\infty; -1,645\}$.

Nyní nám ještě zůstává spočítat hodnotu testovacího kritéria a udělat závěr:

$$U = \frac{\overline{X}_n - \mu}{\sigma}\sqrt{n} = \frac{3010 - 3300}{476}\sqrt{40} = -3,8531.$$

V tomto případě patří hodnota testovacího kritéria do kritického oboru:

$U \in W \Rightarrow H_0$ zamítáme na hladině významnosti $\alpha = 0,05$.

Na 5 % hladině významnosti jsme potvrdili, že porodní váha novorozenců je nižší než u zdravé populace, takže zástupci ekologického hnutí můžou použít tento argument proti výstavbě nové továrny.

Příklad 2. Výrobce kávovarů tvrdí, že stroj je nastavený tak, že při jedné přípravě kávy natočí 30 ml se směrodatnou odchylkou 5 ml. Předpokládejme, že objem kávy je náhodná veličina, která se řídí normálním rozdělením. Chceme zkontrolovat, zda při změně kávovaru nedošlo ke změně směrodatné odchylky. Při 10 přípravách kávy jsme dostali následující objemy (v ml):

[4] Převzato z [10, str. 82].

31; 28; 36; 32; 27; 30; 25; 24; 31; 33.

Řešení: Chceme ověřit tvrzení výrobce, že přesnost kávovaru měřená směrodatnou odchylkou je 5ml, proti hypotéze, že se přesnost po výměně kávovaru změnila:

$$H_0 : \sigma = 5,$$
$$H_1 : \sigma \neq 5.$$

Tyto testy jsou ekvivalentní hypotézám:

$$H_0 : \sigma^2 = 25,$$
$$H_1 : \sigma^2 \neq 25.$$

Odhadujeme rozptyl při neznámé střední hodnotě, takže za testovací kritérium vybereme pivotovou statistiku K.

Zvolíme $\alpha = 0,05$.

Kritický obor se řídí dvoustrannou hypotézou, takže
$W = \{0; \chi^2_{0,025}(9)\} \cup \{\chi^2_{0,975}(9); \infty\} = \{0; 2,7\} \cup \{19; \infty\}$.
POZOR: χ^2 nemůže mít záporné kvantily!

Spočítejme ještě výběrový průměr, rozptyl a hodnotu testovacího kritéria:

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i = 29,7,$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 = 13,7889,$$

$$K = (n-1)\frac{S^2}{\sigma^2} = \frac{9 \cdot 13,7889}{25} = 4,964.$$

$K \notin W \Rightarrow H_0$ nezamítáme na hladině významnosti $\alpha = 0,05$.

Naším testem se nepodařilo zamítnout tvrzení výrobce o udávané přesnosti kávovaru.

Testy z předchozích dvou příkladů byly jednovýběrové v normálním rozdělení. Další bude dvouvýběrový v normálním rozdělení. Statistiky zůstávají stejné jako v předchozí kapitole.

Příklad 3. Ve školních novinách bylo napsáno, že dívky nejrychleji rostou v období mezi 11 a 13 roky, zatímco chlapci nejrychleji rostou mezi 13 a 15 roky. Můžeme potvrdit teorii, že v období 12 let jsou děvčata vyšší než chlapci, jestliže jsme ve stejné škole

náhodně vybrali žáky 6. třídy a měřením jejich výšek dostali následující údaje (v cm) a výšky se řídí normálním rozdělením:
chlapci ($\mathbf{X}_{n_1}$): 132; 128; 134; 129; 139; 133; 141; 130; 122;
děvčata ($\mathbf{X}_{n_2}$): 138; 142; 157; 124; 134; 132; 139?

Řešení: Zadané údaje představují dva náhodné výběry. Z toho je očividné, že budeme dělat dvouvýběrový test. Ověřujeme, zda jsou děvčata v daném období vyšší než chlapci.

$$H_0 : \mu_1 = \mu_2 \Rightarrow \mu_1 - \mu_2 = 0,$$
$$H_1 : \mu_1 < \mu_2 \Rightarrow \mu_1 - \mu_2 < 0.$$

Odhadujeme rozdíl středních hodnot při neznámých rozptylech (aniž víme, zda jsou stejné). Takovou statistiku jsme zatím neřešili. Proto zde nejprve zavedeme jednu „pomocnou“ hypotézu o rovnosti rozptylů.

$$H_0 : \sigma_1^2 = \sigma_2^2 \Rightarrow H_0 : \frac{\sigma_2^2}{\sigma_1^2} = 1,$$
$$H_1 : \sigma_1^2 \neq \sigma_2^2 \Rightarrow H_1 : \frac{\sigma_2^2}{\sigma_1^2} \neq 1.$$

V ramci pomocné hypotézy použijeme $F_{\sigma_1^2/\sigma_2^2}$ statistiku.

Za α zvolíme 5 %.

Kritický obor se nyní řídí dvoustrannou hypotézou, takže bude

$$W = \{0; F_{0,025}(8;6)\} \cup \{F_{0,975}(8;6); \infty\} = \left\{0; \frac{1}{F_{0,975}(6;8)}\right\} \cup \left\{F_{0,975}(8;6); \infty\right\} =$$
$$= \{0; 0,2149\} \cup \{5,6; \infty\}.$$

Pro statistiku $F_{\sigma_1^2/\sigma_2^2}$ potřebujeme spočítat výběrové rozptyly:

$$\overline{X}_{n_1} = 132,$$
$$\overline{X}_{n_2} = 138.$$

$$S_1^2 = 33,$$
$$S_2^2 = 104,3334.$$

$$F_{\sigma_1^2/\sigma_2^2} = \frac{S_1^2}{S_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} = \frac{104,3334}{33} \cdot 1 = 3,1616.$$

$F_{\sigma_1^2/\sigma_2^2} \notin W \Rightarrow H_0$ nezamítáme na hladině významnosti $\alpha = 0,05$.

Věříme, že $\sigma_1^2 = \sigma_2^2$, protože H_0 jsme nezamítli. Ale POZOR: formálně to nemáme dostatečně podepřeno!

Pokračujeme v našem testu. Nyní můžeme zvolit statistiku $T_{\overline{X}_{n_1}-\overline{X}_{n_2}}$.

Za α znovu zvolíme 5 %.

Nový kritický obor je tedy
$W = \{-\infty; t_{0,05}(14)\} = \{-\infty; -t_{0,95}(14)\} = \{-\infty; -1,761\}$.

Pro tuto statistiku potřebujeme ještě spočítat S_{12}^2:

$$S_{12}^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} = \frac{8 \cdot 33 + 6 \cdot 104,3334}{14} = 63,5714,$$

$$T_{\overline{X}_{n_1}-\overline{X}_{n_2}} = \frac{\overline{X}_{n_1} - \overline{X}_{n_2} - (\mu_1 - \mu_2)}{S_{12}} \sqrt{\frac{n_1 n_2}{n_1+n_2}} = \frac{132 - 138 - 0}{\sqrt{63,5714}} \sqrt{\frac{9 \cdot 7}{9+7}} = -1,4932.$$

$T_{\overline{X}_{n_1}-\overline{X}_{n_2}} \notin W \Rightarrow H_0$ nezamítáme na hladině významnosti $\alpha = 0,05$.

Na základě našeho testu nemůžeme odvodit závěr, že děvčata jsou v tomto období opravdu vyšší než kluci.

Na posledním příkladu uvidíme, jak se hypotézy testují při párových výběrech.

Příklad 4. Majitel vlastní dvě restaurace se stejnou nabídkou, přičemž první se nachází na periferii města, druhá v centru. Zákazníci preferují chodit na oběd do první restaurace, a to proto, že si myslí, že je levnější. Náhodně jsme vybrali 10 zákazníků, kteří chodí na oběd do první restaurace, a získali tak částky, které za obědy zaplatili. Zároveň jsme získali částky, které by zaplatili, kdyby chodili do druhé restaurace na stejný oběd. Částky jsou následující:
První restaurace: 109; 118; 131; 45; 76; 73; 110; 118; 189; 46.
Druhá restaurace: 118; 141; 151; 48; 81; 89; 114; 126; 194; 64.
Můžeme tvrdit s rizikem omylu 1 %, že ceny v druhé restauraci jsou vyšší než v první, pokud ceny v obou restauracích mají normální rozdělení?

Řešení: Máme dva náhodné výběry $\mathbf{X}_n = (X_1, ..., X_n)'$ a $\mathbf{Y}_n = (Y_1, ..., Y_n)'$, kde X_i je cena obědu náhodně vybraného zákazníků v první restauraci a Y_i cena v druhé, pro $i = 1, ..., n$, tj. $i = 1, ..., 10$. Jde o případ, kde náhodné veličiny X_i a Y_i jsou pozorovány u stejné jednotky za jiných podmínek, a proto má smysl uvažovat rozdíl hodnot každého páru. Mluvíme tedy o párovém testu.
Úkolem je tedy provést test o rovnosti cen v obou restauracích, proti hypotéze, že jsou ceny v první restauraci nižší než v případě druhé restaurace.

$H_0 : \mu_1 = \mu_2 \Rightarrow \mu_1 - \mu_2 = 0,$
$H_1 : \mu_1 < \mu_2 \Rightarrow \mu_1 - \mu_2 < 0.$

Protože jde o párové výběry, zavedeme „nový“ náhodný výběr $\mathbf{Z}_n$ a střední hodnotu μ, které se rovnají rozdílu předchozích dvou:

$$\mathbf{Z}_n = \mathbf{X}_n - \mathbf{Y}_n = (-9;\ -23;\ -20;\ -3;\ -5;\ -16;\ -4;\ -8;\ -5;\ -18)',$$

$$\mu = \mu_X - \mu_Y.$$

Z toho nám plynou nové hypotézy ekvivalentní už zadaným:

$H_0 : \mu = 0,$
$H_1 : \mu < 0.$

Použijeme statistiku pro párové výběry T.

Na rozdíl od předchozích dvou příkladů zde máme zadané $\alpha = 0{,}01$.

Kritický obor tvoříme na základě levostranného testu, tedy $W = \{-\infty; t_{0{,}01}(9)\} = \{-\infty; -t_{0{,}99}(9)\} = \{-\infty; -2{,}821\}$.

Spočítejme ještě výběrový průměr a rozptyl, které jsou potřebné k vypočítání hodnoty testovacího kritéria:

$$\overline{Z}_n = \frac{1}{n}\sum_{i=1}^{n} Z_i = -11{,}1,$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Z_i - \overline{Z}_n)^2 = 55{,}2112,$$

$$S = \sqrt{55{,}2112} = 7{,}4304,$$

$$T = \frac{\overline{Z}_n - \mu}{S}\sqrt{n} = \frac{-11{,}1 - 0}{7{,}4304}\sqrt{10} = -4{,}7240.$$

$T \in W \Rightarrow H_0$ zamítáme na hladině významnosti $\alpha = 0{,}01$.

Tímto testem jsme zamítli tvrzení, že se ceny v restauracích rovnají, takže podle našeho testu jsou ceny opravdu nižší v první restauraci než ve druhé.

5.1 Cvičení

1. Trolejbusy projíždějící centrem města mají průměrnou rychlost s normálním rozdělením 20 km/hod. Vedoucí MHD rozhodli, že změní trasu trolejbusů, aby zvýšili jejich průměrnou rychlost. Na nové trase byly naměřeny následující rychlosti v náhodně vybraných hodinách: 23; 19; 27; 24, 17; 20; 21. Bylo toto rozhodnutí správné?

   [Rozhodnutí nebylo správné.]

2. Výrobce elektrických strojků tvrdí, že použitím nové výrobní technologie prodlouží průměrnou výdrž baterie ze 100 hodin na 103 hodiny. Tato veličina má normální rozdělení s rozptylem $\sigma^2 = 16$. Na základě 12 testovaných strojků jsme zjistili, že průměrná výdrž baterie je 102 hodiny.

   a) Je tvrzení výrobce, že se průměrná výdrž baterie zvýší, správné?
   b) Uvedl výrobce správný rozptyl, pokud je výdrž testovaných strojků následující: 99; 102; 107; 103; 100; 101; 98; 110; 103; 100; 101; 100? Víme, že $\mu = 100$.

   [Tvrzení není správné; Neuvedl správný rozptyl.]

3. U osmi náhodně vybraných zákazníků byly zjištěny následující doby čekání (ve dnech) na objednání v kadeřnickém salonu: 7; 9; 2; 13; 4; 6; 7; 10. Paní kadeřnice tvrdí, že střední hodnota čekání jejích zákazníků na objednání není větší jak 7 dnů. Je toto tvrzení správné?

   [Tvrzení není správné.]

4. Směrodatná odchylka ročních teplot v konkrétním městě, měřena v období 100 let je 8° C. Měřena je rovněž střední denní teplota 15. dne v měsíci během posledních 15 let a je spočítaná směrodatná odchylka 8° C. Jestliže předpokládáme, že teploty mají normální rozdělení, můžeme na hladině významnosti 1 % tvrdit, že se směrodatná odchylka teplot v posledních 15 let zmenšila?

   [Nemůžeme tvrdit.]

Závěr

Cílem této bakalářské práce bylo vytvořit sbírku vyřešených příkladů z pravděpodobnosti a statistiky, která umožní posluchačům předmětu M4122 snadnější přípravu k zápočtovým testům a závěrečným zkouškám.

Na základě vlastních zkušeností z tohoto kurzu jsem cítila, že vypracování takové sbírky by přivítali všichni studenti, kteří kurz navštěvovali. V každé kapitole jsem se snažila shrnout základní teorii, která je v souladu s přednáškami Mgr. Jana Koláčka, Ph.D. a která by umožnila studentům snazší řešení zadaných úloh. Příklady jsou navrženy tak, aby odpovídaly úkolům, které jsou řešeny ve cvičeních. I když jsem před zahájením práce měla dojem, že její psaní nebude obtížné, vytvoření takové sbírky nebylo vůbec snadné. Bylo to především z toho důvodu, že jsem se snažila sama vytvořit příklady, které by byly podobné těm, co se počítají na cvičeních, ovšem ne úplně stejné. Jedním z největších problémů bylo formulovat zadání tak, aby nebylo úplně odtržené od reality i s ohledem na konečný výsledek. To mohu připsat své vlastní nezkušenosti při vytváření úkolů. Nejtěžší část práce spočívala ve vymýšlení „teoretických“ příkladů, jednak kvůli již zmíněnému záměrnému vynechávání příkladů vypracovaných na přednáškách a cvičeních, jednak je takových příkladů i v literatuře velmi málo, někde nejsou dokonce obsaženy vůbec. Proto může mít čtenář pocit, že těchto příkladů není v práci dostatečné množství a že některé z nich jsou jednodušší. U ostatních příkladů jsem se snažila, aby byly zajímavé a podporovaly kritické myšlení.

Přílohy

Příloha 1. Distribuční funkce normálního rozdělení.
Příloha 2. Distribuční funkce normálního rozdělení - pokračování.
Příloha 3. Kvantily normovaného normálního rozdělení.
Příloha 4. Kvantily χ^2 rozdělení.
Příloha 5. Kvantily χ^2 rozdělení - pokračování.
Příloha 6. Kvantily Studentova t rozdělení.
Příloha 7. Kvantily $F_{0,95}$ Fisherovo-Snedecorova rozdělení.
Příloha 8. Kvantily $F_{0,95}$ Fisherovo-Snedecorova rozdělení - pokračování.
Příloha 9. Kvantily $F_{0,975}$ Fisherovo-Snedecorova rozdělení.
Příloha 10. Kvantily $F_{0,975}$ Fisherovo-Snedecorova rozdělení - pokračování.
Příloha 11. Kvantily $F_{0,99}$ Fisherovo-Snedecorova rozdělení.
Příloha 12. Kvantily $F_{0,99}$ Fisherovo-Snedecorova rozdělení - pokračování.
Příloha 13. Kvantily $F_{0,995}$ Fisherovo-Snedecorova rozdělení.
Příloha 14. Kvantily $F_{0,995}$ Fisherovo-Snedecorova rozdělení - pokračování.

u	Φ(u)	u	Φ(u)	u	Φ(u)	u	Φ(u)
0,00	0,50000	0,40	0,65542	0,80	0,78814	1,20	0,88493
0,01	0,50399	0,41	0,65910	0,81	0,79103	1,21	0,88686
0,02	0,50798	0,42	0,66276	0,82	0,79389	1,22	0,88877
0,03	0,51197	0,43	0,66640	0,83	0,79673	1,23	0,89065
0,04	0,51595	0,44	0,67003	0,84	0,79955	1,24	0,89251
0,05	0,51994	0,45	0,67364	0,85	0,80234	1,25	0,89435
0,06	0,52392	0,46	0,67724	0,86	0,80511	1,26	0,89617
0,07	0,52790	0,47	0,68082	0,87	0,80785	1,27	0,89796
0,08	0,53188	0,48	0,68439	0,88	0,81057	1,28	0,89973
0,09	0,53586	0,49	0,68793	0,89	0,81327	1,29	0,90147
0,10	0,53983	0,50	0,69146	0,90	0,81594	1,30	0,90320
0,11	0,54380	0,51	0,69497	0,91	0,81859	1,31	0,90490
0,12	0,54776	0,52	0,69847	0,92	0,82121	1,32	0,90658
0,13	0,55172	0,53	0,70194	0,93	0,82381	1,33	0,90824
0,14	0,55567	0,54	0,70540	0,94	0,82639	1,34	0,90988
0,15	0,55962	0,55	0,70884	0,95	0,82894	1,35	0,91149
0,16	0,56356	0,56	0,71226	0,96	0,83147	1,36	0,91309
0,17	0,56749	0,57	0,71566	0,97	0,83398	1,37	0,91466
0,18	0,57142	0,58	0,71904	0,98	0,83646	1,38	0,91621
0,19	0,57535	0,59	0,72240	0,99	0,83891	1,39	0,91774
0,20	0,57926	0,60	0,72575	1,00	0,84134	1,40	0,91924
0,21	0,58318	0,61	0,72907	1,01	0,84375	1,41	0,92073
0,22	0,58706	0,62	0,73237	1,02	0,84614	1,42	0,92220
0,23	0,59095	0,63	0,73565	1,03	0,84850	1,43	0,92364
0,24	0,59483	0,64	0,73891	1,04	0,85083	1,44	0,92507
0,25	0,59871	0,65	0,74215	1,05	0,85314	1,45	0,92647
0,26	0,60257	0,66	0,74537	1,06	0,85543	1,46	0,92786
0,27	0,60642	0,67	0,74857	1,07	0,85769	1,47	0,92922
0,28	0,61026	0,68	0,75175	1,08	0,85993	1,48	0,93056
0,29	0,61409	0,69	0,75490	1,09	0,86214	1,49	0,93189
0,30	0,61791	0,70	0,75804	1,10	0,86433	1,50	0,93319
0,31	0,62172	0,71	0,76115	1,11	0,86650	1,51	0,93448
0,32	0,62552	0,72	0,76424	1,12	0,86864	1,52	0,93574
0,33	0,62930	0,73	0,76730	1,13	0,87076	1,53	0,93699
0,34	0,63307	0,74	0,77035	1,14	0,87286	1,54	0,93822
0,35	0,63683	0,75	0,77337	1,15	0,87493	1,55	0,93943
0,36	0,64058	0,76	0,77637	1,16	0,87698	1,56	0,94062
0,37	0,64431	0,77	0,77935	1,17	0,87900	1,57	0,94179
0,38	0,64803	0,78	0,78230	1,18	0,88100	1,58	0,94295
0,39	0,65173	0,79	0,78524	1,19	0,88298	1,59	0,94408

Příloha 1. Distribuční funkce normálního rozdělení.

u	Φ(u)	u	Φ(u)	u	Φ(u)	u	Φ(u)
1,60	0,94520	2,00	0,97725	2,40	0,99180	3,10	0,99903
1,61	0,94630	2,01	0,97778	2,41	0,99202	3,12	0,99910
1,62	0,94738	2,02	0,97831	2,42	0,99224	3,14	0,99916
1,63	0,94845	2,03	0,97882	2,43	0,99245	3,16	0,99921
1,64	0,94950	2,04	0,97932	2,44	0,99266	3,18	0,99926
1,65	0,95053	2,05	0,97982	2,45	0,99286	3,20	0,99931
1,66	0,95154	2,06	0,98030	2,46	0,99305	3,22	0,99936
1,67	0,95254	2,07	0,98077	2,47	0,99324	3,24	0,99940
1,68	0,95352	2,08	0,98124	2,48	0,99343	3,26	0,99944
1,69	0,95449	2,09	0,98169	2,49	0,99361	3,28	0,99948
1,70	0,95543	2,10	0,98214	2,50	0,99379	3,30	0,99952
1,71	0,95637	2,11	0,98257	2,52	0,99413	3,32	0,99955
1,72	0,95728	2,12	0,98300	2,54	0,99446	3,34	0,99958
1,73	0,95818	2,13	0,98341	2,56	0,99477	3,36	0,99961
1,74	0,95907	2,14	0,98382	2,58	0,99506	3,38	0,99964
1,75	0,95994	2,15	0,98422	2,60	0,99534	3,40	0,99966
1,76	0,96080	2,16	0,98461	2,62	0,99560	3,42	0,99969
1,77	0,96164	2,17	0,98500	2,64	0,99585	3,44	0,99971
1,78	0,96246	2,18	0,98537	2,66	0,99609	3,46	0,99973
1,79	0,96327	2,19	0,98574	2,68	0,99632	3,48	0,99975
1,80	0,96407	2,20	0,98610	2,70	0,99653	3,50	0,99977
1,81	0,96485	2,21	0,98645	2,72	0,99674	3,55	0,99981
1,82	0,96562	2,22	0,98679	2,74	0,99693	3,60	0,99984
1,83	0,96638	2,23	0,98713	2,76	0,99711	3,65	0,99987
1,84	0,96712	2,24	0,98745	2,78	0,99728	3,70	0,99989
1,85	0,96784	2,25	0,98778	2,80	0,99744	3,75	0,99991
1,86	0,96856	2,26	0,98809	2,82	0,99760	3,80	0,99993
1,87	0,96926	2,27	0,98840	2,84	0,99774	3,85	0,99994
1,88	0,96995	2,28	0,98870	2,86	0,99788	3,90	0,99995
1,89	0,97062	2,29	0,98899	2,88	0,99801	3,95	0,99996
1,90	0,97128	2,30	0,98928	2,90	0,99813	4,00	0,99997
1,91	0,97193	2,31	0,98956	2,92	0,99825	4,05	0,99997
1,92	0,97257	2,32	0,98983	2,94	0,99836	4,10	0,99998
1,93	0,97320	2,33	0,99010	2,96	0,99846	4,15	0,99998
1,94	0,97381	2,34	0,99036	2,98	0,99856	4,20	0,99999
1,95	0,97441	2,35	0,99061	3,00	0,99865	4,25	0,99999
1,96	0,97500	2,36	0,99086	3,02	0,99874	4,30	0,99999
1,97	0,97558	2,37	0,99111	3,04	0,99882	4,35	0,99999
1,98	0,97615	2,38	0,99134	3,06	0,99890	4,40	0,99999
1,99	0,97670	2,39	0,99158	3,08	0,99897	4,45	1,00000

Příloha 2. Distribuční funkce normálního rozdělení - pokračování.

P	u_P	P	u_P	P	u_P	P	u_P
0,500	0,000	0,750	0,674	0,950	1,645	0,975	1,960
0,510	0,025	0,760	0,706	0,951	1,655	0,976	1,970
0,520	0,050	0,770	0,739	0,952	1,665	0,977	1,995
0,530	0,075	0,780	0,772	0,953	1,675	0,978	2,014
0,540	0,100	0,790	0,806	0,954	1,685	0,979	2,034
0,550	0,126	0,800	0,842	0,955	1,695	0,980	2,054
0,560	0,151	0,810	0,878	0,956	1,706	0,981	2,075
0,570	0,176	0,820	0,915	0,957	1,717	0,982	2,097
0,580	0,202	0,830	0,954	0,958	1,728	0,983	2,120
0,590	0,228	0,840	0,994	0,959	1,739	0,984	2,144
0,600	0,253	0,850	1,036	0,960	1,751	0,985	2,170
0,610	0,279	0,860	1,080	0,961	1,762	0,986	2,197
0,620	0,305	0,870	1,126	0,962	1,774	0,987	2,226
0,630	0,332	0,880	1,175	0,963	1,787	0,988	2,257
0,640	0,358	0,890	1,227	0,964	1,799	0,989	2,290
0,650	0,385	0,900	1,282	0,965	1,812	0,990	2,326
0,660	0,412	0,905	1,311	0,966	1,825	0,991	2,366
0,670	0,440	0,910	1,341	0,967	1,838	0,992	2,409
0,680	0,468	0,915	1,372	0,968	1,852	0,993	2,457
0,690	0,496	0,920	1,405	0,969	1,866	0,994	2,512
0,700	0,524	0,925	1,440	0,970	1,881	0,995	2,576
0,710	0,553	0,930	1,476	0,971	1,896	0,996	2,652
0,720	0,583	0,935	1,514	0,972	1,911	0,997	2,748
0,730	0,613	0,940	1,555	0,973	1,927	0,998	2,878
0,740	0,643	0,945	1,598	0,974	1,943	0,999	3,090

Příloha 3. Kvantily normovaného normálního rozdělení.

v\P	0,0005	0,001	0,005	0,01	0,025	0,05	0,1
1	0,0639	0,0516	0,0439	0,0316	0,0398	0,0239	0,0158
2	0,0210	0,0220	0,0100	0,0201	0,0506	0,1030	0,2110
3	0,0153	0,0243	0,0717	0,1150	0,2160	0,3520	0,5840
4	0,0639	0,0908	0,2070	0,2970	0,4840	0,7110	1,0600
5	0,1580	0,2100	0,4120	0,5540	0,8310	1,1500	1,6100
6	0,2990	0,3810	0,6760	0,8720	1,2400	1,6400	2,2000
7	0,4850	0,5980	0,9890	1,2400	1,6900	2,1700	2,8300
8	0,7100	0,8570	1,3400	1,6500	2,1800	2,7300	3,4900
9	0,9720	1,1500	1,7300	2,0900	2,7000	3,3300	4,1700
10	1,2600	1,4800	2,1600	2,5600	3,2500	3,9400	4,8700
11	1,5900	1,8300	2,6000	3,0500	3,8200	4,5700	5,5800
12	1,9300	2,2100	3,0700	3,5700	4,4000	5,2300	6,3000
13	2,3100	2,6200	3,5700	4,1100	5,0100	5,8900	7,0400
14	2,7000	3,0400	4,0700	4,6600	5,6300	6,5700	7,7900
15	3,1100	3,4800	4,6000	5,2300	6,2600	7,2600	8,5500
16	3,5400	3,9400	5,1400	5,8100	6,9100	7,9600	9,3100
17	3,9800	4,4200	5,7000	6,4100	7,5600	8,6700	10,100
18	4,4400	4,9000	6,2600	7,0100	8,2300	9,3900	10,900
19	4,9100	5,4100	6,8400	7,6300	8,9100	10,100	11,700
20	5,4000	5,9200	7,4300	8,2600	9,3900	10,900	12,400
21	5,9000	6,4500	8,0300	8,9000	10,300	11,600	13,200
22	6,4000	6,9800	8,6400	9,5400	11,000	12,300	14,000
23	6,9200	7,5300	9,2600	10,200	11,700	13,100	14,800
24	7,4500	8,0800	9,8900	10,900	12,400	13,800	15,700
25	7,9900	8,6500	10,500	11,500	13,100	14,600	16,500
26	8,5400	9,2200	11,200	12,200	13,800	15,400	17,300
27	9,0900	9,8000	11,800	12,900	14,600	16,200	18,100
28	9,6600	10,400	12,500	13,600	15,300	16,900	18,900
29	10,200	11,000	13,100	14,300	16,000	17,700	19,800
30	10,800	11,600	13,800	15,000	16,800	18,500	20,600

Příloha 4. Kvantily χ^2 rozdělení.

v\P	0,9	0,95	0,975	0,99	0,995	0,999	0,9995
1	2,71	3,84	5,02	6,63	7,88	10,80	12,10
2	4,61	5,99	7,38	9,21	10,60	13,80	15,20
3	6,25	7,81	9,35	11,30	12,80	16,30	17,70
4	7,78	9,49	11,10	13,30	14,90	18,50	20,00
5	9,24	11,10	12,80	15,10	16,70	20,50	22,10
6	10,60	12,60	14,40	16,80	18,50	22,50	24,10
7	12,00	14,10	16,00	18,50	20,30	24,30	26,00
8	13,40	15,50	17,50	20,10	22,00	26,10	27,90
9	14,70	16,90	19,00	21,70	23,60	27,90	29,70
10	16,00	18,30	20,50	23,20	25,20	29,60	31,40
11	17,30	19,70	21,90	24,70	26,80	31,30	33,10
12	18,50	21,00	23,30	26,20	28,30	32,90	34,80
13	19,80	22,40	24,70	27,70	29,80	34,50	36,50
14	21,00	23,70	26,10	29,10	31,30	36,10	38,10
15	22,30	25,00	27,50	30,60	32,80	37,70	39,70
16	23,50	26,30	28,80	32,00	34,30	39,30	41,30
17	24,80	27,60	30,20	33,40	35,70	40,80	42,90
18	26,00	28,90	31,50	34,80	37,20	42,30	44,40
19	27,20	30,10	32,90	36,20	38,60	43,80	46,00
20	28,40	31,40	34,20	37,60	40,00	452,00	47,50
21	29,60	32,70	35,50	38,90	41,40	46,80	49,00
22	30,90	33,90	36,80	40,30	42,80	48,30	50,50
23	32,00	35,20	38,10	41,60	44,20	49,70	52,00
24	33,20	36,40	39,40	43,00	45,60	51,20	53,50
25	34,40	37,70	40,60	44,30	46,90	52,60	54,90
26	35,60	38,90	41,90	45,60	48,30	54,10	56,40
27	36,70	40,10	43,20	47,00	49,60	55,50	57,90
28	37,90	41,30	44,50	48,30	51,00	56,90	59,30
29	39,10	42,60	45,70	49,60	52,30	58,30	60,70
30	40,30	43,80	47,00	50,90	53,70	59,70	62,20

Příloha 5. Kvantily χ^2 rozdělení - pokračování.

ν\P	0,9	0,95	0,975	0,99	0,995
1	3,078	6,314	12,706	31,821	63,657
2	1,886	2,920	4,303	6,965	9,925
3	1,638	2,353	3,182	4,541	5,841
4	1,553	2,132	2,776	3,747	4,604
5	1,476	2,015	2,571	3,365	4,032
6	1,440	1,943	2,447	3,143	3,707
7	1,415	1,895	2,365	2,998	3,499
8	1,397	1,860	2,306	2,896	3,355
9	1,383	1,833	2,262	2,821	3,250
10	1,372	1,812	2,228	2,764	3,169
11	1,363	1,796	2,201	2,718	3,106
12	1,356	1,782	2,179	2,681	3,055
13	1,350	1,771	2,160	2,650	3,012
14	1,345	1,761	2,145	2,624	2,977
15	1,341	1,753	2,131	2,602	2,947
16	1,337	1,746	2,120	2,583	2,921
17	1,333	1,740	2,110	2,567	2,898
18	1,330	1,734	2,101	2,552	2,878
19	1,328	1,729	2,093	2,539	2,861
20	1,325	1,725	2,086	2,528	2,845
21	1,323	1,721	2,080	2,518	2,831
22	1,321	1,717	2,074	2,508	2,819
23	1,319	1,714	2,069	2,500	2,807
24	1,318	1,711	2,064	2,492	2,797
25	1,316	1,708	2,060	2,485	2,787
26	1,315	1,706	2,056	2,479	2,779
27	1,314	1,703	2,052	2,473	2,771
28	1,313	1,701	2,048	2,467	2,763
29	1,311	1,699	2,045	2,462	2,756
30	1,310	1,697	2,042	2,457	2,750

Příloha 6. Kvantily Studentova *t* rozdělení.

$v_2 \backslash v_1$	1	2	3	4	5	6	7	8	9
1	161,450	199,500	215,710	224,580	230,160	233,990	236,770	238,880	240,540
2	18,513	19,000	19,164	19,247	19,296	19,330	19,353	19,371	19,385
3	10,128	9,552	9,277	9,117	9,014	8,941	8,887	8,845	8,812
4	7,709	6,944	6,591	6,388	6,256	6,163	6,094	6,041	5,999
5	6,608	5,786	5,410	5,192	5,050	4,950	4,876	4,818	4,773
6	5,987	5,143	4,757	4,534	4,387	4,284	4,207	4,147	4,099
7	5,591	4,737	4,347	4,120	3,972	3,866	3,787	3,726	3,677
8	5,318	4,459	4,066	3,838	3,688	3,581	3,501	3,438	3,388
9	5,117	4,257	3,863	3,633	3,482	2,274	3,293	3,230	3,179
10	4,965	4,103	3,708	3,478	3,326	3,217	3,136	3,072	3,020
11	4,844	3,982	3,587	3,357	3,204	3,095	3,012	2,948	2,896
12	4,747	3,885	3,490	3,259	3,106	2,996	2,913	2,849	2,796
13	4,667	3,806	3,411	3,179	3,025	2,915	2,832	2,767	2,714
14	4,600	3,739	3,344	3,112	2,958	2,848	2,764	2,699	2,646
15	4,543	3,682	3,287	3,056	2,901	2,791	2,707	2,641	2,588
16	4,494	3,634	3,239	3,007	2,852	2,741	2,657	2,591	2,538
17	4,451	3,592	3,197	2,965	2,810	2,699	2,614	2,548	2,494
18	4,414	3,555	3,160	2,928	2,773	2,661	2,577	2,510	2,456
19	4,381	3,522	3,127	2,895	2,740	2,628	2,544	2,477	2,423
20	4,351	3,493	3,098	2,866	2,711	2,599	2,514	2,447	2,393
21	4,325	3,467	3,073	2,840	2,685	2,573	2,488	2,421	2,366
22	4,301	3,443	3,049	2,817	2,661	2,549	2,464	2,397	2,342
23	4,279	3,422	3,028	2,796	2,640	2,528	2,442	2,375	2,320
24	4,260	3,403	3,009	2,776	2,621	2,508	2,423	2,355	2,300
25	4,242	3,385	2,991	2,759	2,603	2,490	2,405	2,337	2,282
26	4,225	3,369	2,975	2,743	2,587	2,474	2,388	2,321	2,266
27	4,210	3,354	2,960	2,728	2,572	2,459	2,373	2,305	2,250
28	4,196	3,340	2,947	2,714	2,558	2,445	2,359	2,291	2,236
29	4,183	3,328	2,934	2,701	2,545	2,432	2,346	2,278	2,223
30	4,171	3,316	2,922	2,690	2,534	2,421	2,334	2,266	2,211
40	4,085	3,232	2,839	2,606	2,450	2,336	2,249	2,180	2,124
60	4,001	3,150	2,758	2,525	2,368	2,254	2,167	2,097	2,040
120	3,920	3,072	2,680	2,447	2,290	2,175	2,087	2,016	1,959
∞	3,842	2,996	2,605	2,372	2,214	2,099	2,010	1,938	1,880

Příloha 7. Kvantily $F_{0,95}$ Fisherovo-Snedecorova rozdělení.

$v_2 \backslash v_1$	10	12	15	20	24	30	40	60	120	∞
1	241,880	243,910	245,950	248,010	249,050	250,090	251,140	252,200	253,250	254,320
2	19,396	19,413	19,429	19,446	19,454	19,462	19,471	19,479	19,487	19,496
3	8,786	8,745	8,703	8,660	8,639	8,617	8,594	8,572	8,549	8,527
4	5,964	5,912	5,858	5,803	5,774	5,746	5,717	5,688	5,658	5,628
5	4,735	4,678	4,619	4,558	4,527	4,496	4,464	4,431	4,398	4,365
6	4,060	4,000	3,938	3,874	3,842	3,808	3,774	3,740	3,705	3,669
7	3,637	3,575	3,511	3,445	3,411	3,376	3,340	3,304	3,267	3,230
8	3,347	3,284	3,218	3,150	3,115	3,079	3,043	3,005	2,967	2,928
9	3,137	3,073	3,006	2,937	2,901	2,864	2,826	2,787	2,748	2,707
10	2,978	2,913	2,845	2,774	2,737	2,700	2,661	2,621	2,580	2,538
11	2,854	2,788	2,719	2,646	2,609	2,571	2,531	2,490	2,448	2,405
12	2,753	2,687	2,617	2,544	2,506	2,466	2,426	2,384	2,341	2,296
13	2,671	2,604	2,533	2,459	2,420	2,380	2,339	2,297	2,252	2,206
14	2,602	2,534	2,463	2,388	2,349	2,308	2,266	2,223	2,178	2,131
15	2,544	2,475	2,404	2,328	2,288	2,247	2,204	2,160	2,114	2,066
16	2,494	2,425	2,352	2,276	2,235	2,194	2,151	2,106	2,059	2,010
17	2,450	2,381	2,308	2,230	2,190	2,148	2,104	2,058	2,011	1,960
18	2,412	2,342	2,269	2,191	2,150	2,107	2,063	2,017	1,968	1,917
19	2,378	2,308	2,234	2,156	2,114	2,071	2,026	1,980	1,930	1,878
20	2,348	2,278	2,203	2,124	2,083	2,039	1,994	1,946	1,896	1,843
21	2,321	2,250	2,176	2,096	2,054	2,010	1,965	1,917	1,866	1,812
22	2,297	2,226	2,151	2,071	2,028	1,984	1,938	1,890	1,838	1,783
23	2,275	2,204	2,128	2,048	2,005	1,961	1,914	1,865	1,813	1,757
24	2,255	2,183	2,108	2,027	1,984	1,939	1,892	1,842	1,790	1,733
25	2,237	2,165	2,089	2,008	1,964	1,919	1,872	1,822	1,768	1,711
26	2,220	2,148	2,072	1,990	1,946	1,901	1,853	1,803	1,749	1,691
27	2,204	2,132	2,056	1,974	1,930	1,884	1,836	1,785	1,731	1,672
28	2,190	2,118	2,041	1,959	1,915	1,869	1,820	1,769	1,714	1,654
29	2,177	2,105	2,028	1,945	1,901	1,854	1,806	1,754	1,698	1,638
30	2,165	2,092	2,015	1,932	1,887	1,841	1,792	1,740	1,684	1,622
40	2,077	2,004	1,925	1,839	1,793	1,744	1,693	1,637	1,577	1,509
60	1,993	1,917	1,836	1,748	1,700	1,649	1,594	1,534	1,467	1,389
120	1,911	1,834	1,751	1,659	1,608	1,554	1,495	1,429	1,352	1,254
∞	1,831	1,752	1,666	1,571	1,517	1,459	1,394	1,318	1,221	1,000

Příloha 8. Kvantily $F_{0,95}$ Fisherovo-Snedecorova rozdělení - pokračování.

$v_2 \backslash v_1$	1	2	3	4	5	6	7	8	9
1	647,790	799,500	864,160	899,580	921,850	937,110	948,220	956,660	963,280
2	38,506	39,000	39,165	39,248	39,298	39,331	39,355	39,373	39,387
3	17,443	16,044	15,439	15,101	14,885	14,735	14,624	14,540	14,473
4	12,218	10,649	9,979	9,605	9,365	9,197	9,074	8,980	8,905
5	10,007	8,434	7,764	7,388	7,146	6,978	6,853	6,757	6,681
6	8,813	7,260	6,599	6,227	5,988	5,820	5,696	5,600	5,523
7	8,073	6,542	5,890	5,523	5,285	5,119	4,995	4,899	4,823
8	7,571	6,060	5,416	5,053	4,817	4,652	4,529	4,433	4,357
9	7,209	5,715	5,078	4,718	4,484	4,320	4,197	4,102	4,026
10	6,937	5,456	4,826	4,468	4,236	4,072	3,950	3,855	3,779
11	6,724	5,256	4,630	4,275	4,044	3,881	3,759	3,664	3,588
12	6,554	5,096	4,474	4,121	3,891	3,728	3,607	3,512	3,436
13	6,414	4,965	4,347	3,996	3,767	3,604	3,483	3,388	3,312
14	6,298	4,857	4,242	3,892	3,663	3,501	3,380	3,285	3,209
15	6,200	4,765	4,153	3,804	3,576	3,415	3,293	3,199	3,123
16	6,115	4,687	4,077	3,729	3,502	3,341	3,219	3,125	3,049
17	6,042	4,619	4,011	3,665	3,438	3,277	3,156	3,061	2,985
18	5,978	4,560	3,954	3,608	3,382	3,221	3,100	3,005	2,929
19	5,922	4,508	3,903	3,559	3,333	3,172	3,051	2,956	2,880
20	5,872	4,461	3,859	3,515	3,289	3,128	3,007	2,913	2,837
21	5,827	4,420	3,819	3,475	3,250	3,090	2,969	2,874	2,798
22	5,786	4,383	3,783	3,440	3,215	3,055	2,934	2,839	2,763
23	5,750	4,349	3,751	3,408	3,184	3,023	2,902	2,808	2,731
24	5,717	4,319	3,721	3,379	3,155	2,995	2,874	2,779	2,703
25	5,686	4,291	3,694	3,353	3,129	2,969	2,848	2,753	2,677
26	5,659	4,266	3,670	3,329	3,105	2,945	2,824	2,729	2,653
27	5,633	4,242	3,647	3,307	3,083	2,923	2,802	2,707	2,631
28	5,610	4,221	3,626	3,286	3,063	2,903	2,782	2,687	2,611
29	5,588	4,201	3,607	3,267	3,044	2,884	2,763	2,669	2,592
30	5,568	4,182	3,589	3,250	3,027	2,867	2,746	2,651	2,557
40	5,424	4,051	3,463	3,126	2,904	2,744	2,624	2,529	2,452
60	5,286	3,925	3,343	3,008	2,786	2,627	2,507	2,412	2,334
120	5,152	3,805	3,227	2,894	2,674	2,515	2,395	2,299	2,222
∞	5,024	3,689	3,116	2,786	2,567	2,408	2,288	2,192	2,114

Příloha 9. Kvantily $F_{0,975}$ Fisherovo-Snedecorova rozdělení.

$v_2 \backslash v_1$	10	12	15	20	24	30	40	60	120	∞
1	968,930	976,710	984,870	993,100	997,250	1001,400	1005,600	1009,800	1014,000	1018,300
2	39,398	39,415	39,431	39,448	39,456	39,465	39,473	39,481	39,490	39,498
3	14,419	14,337	14,253	14,167	14,124	14,081	14,037	13,992	13,947	13,902
4	8,844	8,751	8,657	8,560	8,511	8,461	8,411	8,360	8,309	8,257
5	6,619	6,525	6,428	6,329	6,278	6,227	6,175	6,125	6,069	6,012
6	5,461	5,366	5,269	5,168	5,117	5,065	5,013	4,959	4,905	4,849
7	4,761	4,666	4,568	4,467	4,415	4,362	4,309	4,256	4,199	4,142
8	4,295	4,200	4,101	4,000	3,947	3,894	3,840	3,784	3,728	3,670
9	3,964	3,868	3,769	3,667	3,614	3,560	3,506	3,449	3,392	3,333
10	3,717	3,621	3,522	3,419	3,365	3,311	3,255	3,198	3,140	3,080
11	3,526	3,430	3,330	3,226	3,173	3,118	3,061	3,004	2,944	2,883
12	3,374	3,277	3,177	3,073	3,019	2,963	2,906	2,848	2,787	2,725
13	3,250	3,153	3,053	2,948	2,893	2,837	2,788	2,720	2,659	2,596
14	3,147	3,050	2,949	2,844	2,789	2,732	2,674	2,614	2,552	2,487
15	3,060	2,963	2,862	2,756	2,701	2,644	2,585	2,524	2,461	2,395
16	2,986	2,889	2,788	2,681	2,625	2,568	2,509	2,447	2,383	2,316
17	2,922	2,825	2,723	2,616	2,560	2,502	2,442	2,380	2,315	2,247
18	2,866	2,769	2,667	2,559	2,503	2,445	2,384	2,321	2,256	2,187
19	2,817	2,720	2,617	2,509	2,452	2,394	2,333	2,270	2,203	2,133
20	2,774	2,676	2,573	2,465	2,408	2,349	2,287	2,223	2,156	2,085
21	2,735	2,637	2,534	2,425	2,368	2,308	2,247	2,182	2,114	2,042
22	2,700	2,602	2,498	2,389	2,332	2,272	2,210	2,145	2,076	2,003
23	2,668	2,570	2,467	2,357	2,299	2,239	2,176	2,111	2,042	1,968
24	2,640	2,541	2,437	2,327	2,269	2,209	2,146	2,080	2,010	1,935
25	2,614	2,515	2,411	2,301	2,242	2,182	2,118	2,052	1,981	1,906
26	2,590	2,491	2,387	2,276	2,217	2,157	2,093	2,026	1,955	1,878
27	2,568	2,469	2,364	2,253	2,195	2,133	2,069	2,002	1,930	1,853
28	2,547	2,448	2,344	2,232	2,174	2,112	2,048	1,980	1,907	1,829
29	2,529	2,430	2,325	2,213	2,154	2,092	2,028	1,959	1,886	1,807
30	2,511	2,412	2,307	2,195	2,136	2,074	2,009	1,940	1,866	1,787
40	2,388	2,288	2,182	2,068	2,007	1,943	1,875	1,803	1,724	1,637
60	2,270	2,169	2,061	1,945	1,882	1,815	1,744	1,667	1,581	1,482
120	2,157	2,055	1,945	1,825	1,760	1,690	1,614	1,530	1,433	1,310
∞	2,048	1,945	1,833	1,709	1,640	1,556	1,484	1,388	1,268	1,000

Příloha 10. Kvantily $F_{0,975}$ Fisherovo-Snedecorova rozdělení - pokračování.

$v_2 \backslash v_1$	1	2	3	4	5	6	7	8	9
1	4052,200	4999,500	5403,500	5624,600	5763,700	5859,000	5928,300	5981,600	6022,500
2	98,503	99,000	99,166	99,249	99,299	99,332	99,356	99,374	99,388
3	34,116	30,817	29,457	28,710	28,237	27,911	27,672	27,489	27,345
4	21,198	18,000	16,694	15,977	15,522	15,207	14,976	14,799	14,639
5	16,258	13,274	12,060	11,392	10,967	10,672	10,456	10,289	10,158
6	13,745	10,925	9,780	9,148	8,746	8,466	8,260	8,102	7,976
7	12,246	9,547	8,451	7,847	7,460	7,191	6,993	6,840	6,719
8	11,259	8,649	7,591	7,006	6,632	6,371	6,178	6,029	5,911
9	10,561	8,022	6,992	6,422	6,057	5,802	5,613	5,467	5,351
10	10,044	7,559	6,552	5,994	5,636	5,386	5,200	5,057	4,942
11	9,646	7,206	6,217	5,668	5,316	5,069	4,886	4,745	4,632
12	9,330	6,927	5,953	5,412	5,064	4,821	4,640	4,499	4,388
13	9,074	6,701	5,739	5,205	4,862	4,620	4,441	4,302	4,191
14	8,862	6,515	5,564	5,035	4,695	4,456	4,278	4,140	4,030
15	8,683	6,359	5,417	4,893	4,556	4,318	4,142	4,005	3,895
16	8,531	6,226	5,292	4,773	4,437	4,202	4,026	3,890	3,780
17	8,400	6,112	5,185	4,669	4,336	4,102	3,927	3,791	3,682
18	8,285	6,013	5,092	4,579	4,428	4,015	3,841	3,705	3,597
19	8,185	5,926	5,010	4,500	4,171	3,939	3,765	3,631	3,523
20	8,096	5,849	4,938	4,431	4,103	3,871	3,699	3,564	3,457
21	8,017	5,780	4,874	4,369	4,042	3,812	3,640	3,506	3,398
22	7,945	5,719	7,817	4,313	3,988	3,758	3,587	3,453	3,346
23	7,881	5,664	4,765	4,264	3,939	3,710	3,539	3,406	3,299
24	7,823	5,614	4,718	4,218	3,895	3,667	3,496	3,363	3,256
25	7,770	5,568	4,676	4,177	3,855	3,627	3,457	3,324	3,217
26	7,721	5,526	4,637	4,140	3,818	3,591	3,421	3,288	3,182
27	7,677	5,488	4,601	4,106	3,785	3,558	3,388	3,256	3,149
28	7,636	5,453	4,568	4,074	3,754	3,528	3,358	3,226	3,120
29	7,598	5,421	4,538	4,045	3,725	3,500	3,330	3,198	3,092
30	7,563	5,390	4,510	4,018	3,699	3,474	3,305	3,173	3,067
40	7,314	5,179	4,313	3,828	3,514	3,291	3,124	2,993	2,888
60	7,077	4,977	4,126	3,649	3,339	3,119	2,953	2,823	2,719
120	6,851	4,787	3,949	3,480	3,174	2,956	2,792	2,663	2,559
∞	6,635	4,605	3,782	3,319	3,017	2,802	2,639	2,511	2,407

Příloha 11. Kvantily $F_{0,99}$ Fisherovo-Snedecorova rozdělení.

$v_2 \backslash v_1$	10	12	15	20	24	30	40	60	120	∞
1	6055,800	6106,300	6157,300	6208,700	6234,600	6260,700	6286,800	6313,000	6339,400	6366,000
2	99,399	99,416	99,432	99,449	99,458	99,466	99,474	99,483	99,491	99,501
3	27,229	27,052	26,872	26,690	26,598	26,505	26,411	26,316	26,221	26,125
4	14,546	14,374	14,198	14,020	13,929	13,838	13,745	13,652	13,558	13,463
5	10,051	9,888	9,722	9,553	9,467	9,379	9,291	9,202	9,112	9,020
6	7,874	7,718	7,559	7,396	7,313	7,229	7,143	7,057	6,969	6,880
7	6,620	6,469	6,314	6,155	6,074	5,992	5,908	5,824	5,737	5,650
8	5,814	5,667	5,515	5,359	5,279	5,198	5,116	5,032	4,946	4,859
9	5,257	5,111	4,962	4,808	4,729	4,649	4,567	4,483	4,398	4,311
10	4,849	4,706	4,558	4,405	7,327	4,247	4,165	4,082	3,997	3,909
11	4,539	4,397	4,251	4,099	4,021	3,941	3,860	3,776	3,690	3,603
12	4,296	4,155	4,010	3,858	3,781	3,701	3,619	3,536	3,449	3,361
13	4,100	3,960	3,815	3,665	3,587	3,507	3,425	3,341	3,255	3,165
14	3,939	3,800	3,656	3,505	3,427	3,348	3,266	3,181	3,094	3,004
15	3,805	3,666	3,522	3,372	3,294	3,214	3,132	3,047	2,960	2,868
16	3,691	3,553	3,409	3,259	3,181	3,101	3,018	2,933	2,845	2,753
17	3,593	3,455	3,312	3,162	3,084	3,003	2,921	2,835	2,746	2,653
18	3,508	3,371	3,227	3,077	2,999	2,919	2,835	2,749	2,660	2,566
19	3,434	3,297	3,153	3,003	2,925	2,844	2,761	2,674	2,584	2,489
20	3,368	3,231	3,088	2,938	2,859	2,779	2,695	2,608	2,517	2,421
21	3,310	3,173	3,030	2,880	2,801	2,720	2,636	2,548	2,457	2,360
22	3,258	3,121	2,978	2,827	2,749	2,668	2,583	2,495	2,403	2,306
23	3,211	3,074	2,931	2,781	2,702	2,620	2,536	2,447	2,354	2,256
24	3,168	3,032	2,889	2,738	2,659	2,577	2,492	2,404	2,310	2,211
25	3,129	2,993	2,850	2,699	2,620	2,538	2,453	2,364	2,270	2,169
26	3,094	2,958	2,815	2,664	2,585	2,503	2,417	2,327	2,233	2,132
27	3,062	2,926	2,783	2,632	2,552	2,470	2,384	2,294	2,198	2,097
28	3,032	2,896	2,753	2,602	2,522	2,440	2,354	2,263	2,167	2,064
29	3,005	2,869	2,726	2,574	2,495	2,412	2,325	2,234	2,138	2,034
30	2,979	2,843	2,700	2,549	2,469	2,386	2,299	2,208	2,111	2,006
40	2,801	2,665	2,522	2,369	2,288	2,203	2,114	2,019	1,917	1,805
60	2,632	2,496	2,352	2,198	2,115	2,029	1,936	1,836	1,726	1,601
120	2,472	2,336	2,192	2,035	1,950	1,860	1,763	1,656	1,533	1,381
∞	2,321	2,185	2,039	1,878	1,791	1,696	1,592	1,473	1,325	1,000

Příloha 12. Kvantily $F_{0,99}$ Fisherovo-Snedecorova rozdělení - pokračování.

$v_2 \backslash v_1$	1	2	3	4	5	6	7	8	9
1	16211	20000	21615	22500	23056	23437	23715	23925	24091
2	198,500	199,000	199,170	199,250	199,300	199,330	199,360	199,370	199,390
3	55,552	49,799	47,467	46,196	45,392	44,838	44,434	44,126	43,882
4	31,333	26,284	24,259	23,155	22,456	21,975	21,622	21,352	21,139
5	22,785	18,314	16,530	15,556	14,940	14,513	14,200	13,961	13,772
6	18,635	14,544	12,917	12,028	11,464	11,073	10,786	10,566	10,391
7	16,236	12,404	10,882	10,050	9,522	9,155	8,885	8,678	8,514
8	14,688	11,042	9,597	8,805	8,330	7,952	7,694	7,496	7,339
9	13,614	10,107	8,717	7,956	7,471	7,134	6,885	6,693	6,541
10	12,826	9,427	8,081	7,343	6,872	6,545	6,303	6,116	5,968
11	12,226	8,912	7,600	6,881	6,422	6,102	5,865	5,682	5,537
12	11,754	8,510	7,226	6,521	6,071	5,757	5,525	5,345	5,202
13	11,374	8,187	6,926	6,234	5,791	5,482	5,253	5,076	4,935
14	11,060	7,922	6,680	5,998	5,562	5,257	5,031	4,857	4,717
15	10,798	7,701	6,476	5,803	5,372	5,071	4,847	4,674	4,536
16	10,575	7,514	6,303	5,638	5,212	4,913	4,692	4,521	4,384
17	10,384	7,354	6,156	5,497	5,075	4,779	4,559	4,389	4,254
18	10,218	7,215	6,028	5,375	4,956	4,663	4,445	4,276	4,141
19	10,073	7,094	5,916	5,268	4,853	4,561	4,345	4,177	4,043
20	9,944	6,987	5,818	5,174	4,762	4,472	4,257	4,090	3,956
21	9,830	6,891	5,730	5,091	4,681	4,393	4,179	4,013	3,880
22	9,727	6,806	5,652	5,017	4,609	4,323	4,109	3,944	3,812
23	9,635	6,730	5,582	4,950	4,544	4,259	4,047	3,882	3,750
24	9,551	6,661	5,519	4,890	4,486	4,202	3,991	3,826	3,695
25	9,475	6,598	5,462	4,835	4,433	4,150	3,939	3,776	3,645
26	9,406	6,541	5,409	4,785	4,384	4,103	3,893	3,730	3,599
27	9,342	6,489	5,361	4,740	4,340	4,059	3,850	3,688	3,557
28	9,284	6,440	5,317	4,698	4,300	4,020	3,811	3,649	3,519
29	9,230	6,396	5,276	4,659	4,262	3,983	3,775	3,613	3,483
30	9,180	6,355	5,239	4,623	4,228	3,949	3,742	3,580	3,451
40	8,828	6,066	4,976	4,374	3,986	3,713	3,509	3,350	3,222
60	8,495	5,795	4,729	4,140	3,760	3,492	3,291	3,134	3,008
120	8,179	5,539	4,497	3,921	3,548	3,285	3,087	2,933	2,808
∞	7,879	5,298	4,279	3,715	3,350	3,091	2,897	2,744	2,621

Příloha 13. Kvantily $F_{0,995}$ Fisherovo-Snedecorova rozdělení.

$v_2 \backslash v_1$	10	12	15	20	24	30	40	60	120	∞
1	24224	24426	24630	24836	24940	25044	25148	25253	25359	25465
2	199,400	199,420	199,430	199,450	199,460	199,470	199,470	199,480	199,490	199,510
3	43,686	43,387	43,085	42,778	42,622	42,466	42,308	42,149	41,989	41,829
4	20,967	20,705	20,438	20,167	20,030	19,892	19,752	19,611	19,468	19,325
5	13,618	13,384	13,146	12,903	12,780	12,656	12,530	12,402	12,274	12,144
6	10,250	10,034	9,814	9,589	9,474	9,358	9,241	9,122	9,002	8,879
7	8,380	8,176	7,968	7,754	7,645	7,735	7,423	7,309	7,193	7,076
8	7,211	7,015	6,814	6,608	6,503	6,396	6,288	6,177	6,065	5,951
9	6,417	6,227	6,033	5,832	5,729	5,625	5,519	5,410	5,300	5,188
10	5,847	5,661	5,471	5,274	5,173	5,071	4,966	4,859	4,750	4,639
11	5,418	5,236	5,049	4,855	4,756	4,654	4,551	4,445	4,337	4,226
12	5,086	4,906	4,721	4,530	4,432	4,331	4,228	4,123	4,015	3,904
13	4,820	4,643	4,460	4,270	4,173	4,073	3,970	3,866	3,758	3,647
14	4,603	4,428	4,247	4,059	3,961	3,862	3,760	3,655	3,547	3,436
15	4,424	4,250	4,070	3,883	3,786	3,687	3,585	3,480	3,372	3,260
16	4,272	4,099	3,921	3,734	3,638	3,538	3,437	3,332	3,224	3,112
17	4,142	3,971	3,793	3,607	3,511	3,412	3,311	3,206	3,097	2,984
18	4,031	3,860	3,683	3,498	3,402	3,303	3,201	3,096	2,987	2,873
19	3,933	3,763	3,587	3,402	3,306	3,208	3,106	3,000	2,891	2,776
20	3,847	3,678	3,502	3,318	3,222	3,123	3,022	2,916	2,806	2,690
21	3,771	3,602	3,427	3,243	3,147	3,049	2,947	2,841	2,730	2,614
22	3,703	3,535	3,360	3,176	3,081	2,982	2,880	2,774	2,663	2,546
23	3,642	3,475	3,300	3,117	3,021	2,922	2,820	2,713	2,602	2,484
24	3,587	3,420	3,246	3,062	2,967	2,868	2,765	2,659	2,546	2,428
25	3,537	3,370	3,196	3,013	2,918	2,819	2,716	2,609	2,496	2,377
26	3,492	3,325	3,152	2,969	2,873	2,774	2,671	2,563	2,450	2,330
27	3,450	3,284	3,110	2,928	2,832	2,733	2,630	2,522	2,408	2,287
28	3,412	3,246	3,073	2,890	2,794	2,695	2,592	2,483	2,369	2,247
29	3,377	3,211	3,038	2,855	2,759	2,660	2,557	2,448	2,333	2,210
30	3,344	3,179	3,006	2,823	2,727	2,628	2,524	2,415	2,300	2,176
40	3,117	2,953	2,781	2,598	2,502	2,402	2,296	2,184	3,064	1,932
60	2,904	2,742	2,571	2,387	2,290	2,187	2,079	1,962	1,834	1,688
120	2,705	2,544	2,373	2,188	2,089	1,984	1,871	1,747	1,606	1,431
∞	2,519	2,358	2,187	2,000	1,898	1,789	1,669	1,533	1,364	1,000

Příloha 14. Kvantily $F_{0,995}$ Fisherovo-Snedecorova rozdělení - pokračování.

Seznam použité literatury

[1] FORBELSKÁ, Marie, Jan KOLÁČEK. *Pravděpodobnost a statistika I*. 1. vyd. Brno: Masaryková univerzita, 2013, 94 s. Dostupné z WWW: `https://is.muni.cz/auth/el/1431/jaro2015/M4122/um/M3121_text.pdf`.

[2] FORBELSKÁ, Marie, Jan KOLÁČEK. *Pravděpodobnost a statistika II*. 1. vyd. Brno: Masaryková univerzita, 2013, 94 s. Dostupné z WWW: `https://is.muni.cz/auth/el/1431/jaro2014/M4122/um/M4122_text.pdf`.

[3] FORBELSKÁ, Marie. *Učební materiály k předmětu Lineární statistické modely I*. Brno: Masarykova univerzita, 75 s.

[4] HOROVÁ, Ivana, Jiří ZELINKA. *Učební materiály k předmětu Numerické metody*. Brno: Masarykova univerzita, 293 s.

[5] *Statistika, tabulky*. Praha: Vysoká škola ekonomická v Praze, 16 s. Dostupné z WWW: `http://statistika.vse.cz/download/materialy/tabulky.pdf`.

[6] KOVAČEVIĆ, Ivana. *Verovatnoća i statistika sa zbirkom zadataka*. 1. vyd. Bělehrad: Univerzitet Singidunum, 2011, 212 s. ISBN 978-86-7912-378-7.

[7] MAREK, Luboš. *Pravděpodobnost*. 1. vyd. Praha: Professional Publishing, 2012, 249 s. ISBN 978-80-7431-087-4.

[8] ANDĚL, Jiří. *Základy matematické statistiky*. Vyd. 1. Praha: Matfyzpress, 2005, 358 s. ISBN 80-86732-40-1.

[9] ARLTOVÁ, Markéta. *Příklady k předmětu Statistika A*. Vyd. 2. Praha: Oeconomica, 2004, 197 s. ISBN 80-245-0730-7.

[10] SISÁKOVÁ, Jaroslava, Alena TARTAĽOVÁ a Tomáš ŽELINSKÝ. *Pravdepodobnosť a štatistika: riešené a neriešené úlohy*. 1. vyd. Košice: Elfa, 2008, 132 s. ISBN 978-80-8086-097-4.