Helena Koutková, Ivo Moll

ÚVOD DO PRAVDĚPODOBNOSTI A MATEMATICKÉ STATISTIKY

BRNO

listopad 2000

OBSAH

Základní označení a pojmy

Je-li pojem označený symbolem "$*$", znamená to, že v závěru této kapitoly jsou uvedeny některé jeho vlastnosti. Je-li pojem označený "$**$", znamená to, že je definován v textu skripta.

$det\boldsymbol{A}$		determinant matice $\boldsymbol{A}$
$\boldsymbol{A}^T$		matice transponovaná k matici $\boldsymbol{A}$
$\boldsymbol{A}^{-1}$		matice inverzní k matici $\boldsymbol{A}$
$\boldsymbol{I}_n$		jednotková matice typu (n/n)
$\mathbb{R}$		množina reálných čísel
$\mathbb{N}$		množina přirozených čísel
$A_1 \times A_2 \times \cdots \times A_n$	*	kartézský součin množin $A_1, A_2, \ldots, A_n$
A^n		$\underbrace{A \times A \times \cdots \times A}_{n-\text{krát}}$
$\sup M$	*	supremum množiny M
$\max M$		maximum množiny M
e		základ přirozených logaritmů
π		Ludolfovo číslo
$n!$		faktoriál přirozeného čísla n $n! = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 3 \cdot 2 \cdot 1$
$\binom{n}{k}$		kombinační číslo definované vztahem $\dfrac{n!}{k!(n-k)!}$
$f : H \longmapsto K$		funkce z množiny H do množiny K
$F(-\infty)$		nevlastní limita funkce F v mínus nekonečnu
$F(\infty)$		nevlastní limita funkce F v nekonečnu
$F(a-0)$		limita funkce F v bodě a zleva
$E(X)$	**	střední hodnota náhodné veličiny X
$D(X)$	**	rozptyl náhodné veličiny X
$Mo(X) = x^\circ$	**	modus náhodné veličiny X
$x(\alpha)$	**	$100\,\alpha$ procentní kvantil náhodné veličiny X
$C(X,Y)$	**	kovarince náhodných veličin X a Y
$\rho(X,Y)$	**	koeficient korelace náhodných veličin X a Y
Γ	**	gama funkce
$N(\mu, \sigma^2)$	**	normální rozdělení s parametry μ, σ^2
Φ	**	distribuční funkce $N(0,1)$
φ	**	hustota $N(0,1)$
$u(\alpha)$	**	$100\,\alpha$ - procentní kvantil $N(0,1)$
$\chi^2(n)$	**	χ^2 - rozdělení [Pearsonovo rozdělení] s n stupni volnosti
$\chi^2(n;\alpha)$	**	$100\,\alpha$ - procentní kvantil $\chi^2(n)$

$t(n)$	** t - rozdělení [Studentovo rozdělení] s n stupni volnosti
$t(n;\alpha)$	** 100 α - procentní kvantil $t(n)$
$F(n_1, n_2)$	** F - [Fisherovo - Snedecorovo] rozdělení s n_1. n_2 stupni volnosti
$F(n_1, n_2; \alpha)$	** 100 α - procentní kvantil $F(n_1, n_2)$
$D(n;\alpha)$	** kritická hodnota Kolmogorovova testu

<u>Kartézský součin množin</u>

Jsou-li $A_1, A_2, \ldots, A_n$ libovolné neprázdné množiny, potom $A_1 \times A_2 \times \cdots \times A_n$ je množina, jejíž prvky jsou všechny možné n-tice vytvořené tak, že první člen n-tice je prvek množiny A_1, druhý člen n-tice je prvek množiny A_2, ..., n-tý člen n-tice je prvek množiny A_n. Zapsáno formálně

$$A_1 \times A_2 \times \cdots \times A_n = \{(a_1, a_2, \ldots, a_n);\ a_1 \in A_1, a_2 \in A_2, \ldots a_n \in A_n\}.$$

Např.

$$\langle 0, 1\rangle \times \langle 0, 2\rangle = \{(x, y);\ x \in \langle 0, 1\rangle\,,\ y \in \langle 0, 2\rangle\}.$$

<u>Spočetná množina</u>

Množinu M považujeme za spočetnou, má-li tolik prvků, kolik je přirozených čísel. Přesněji řečeno: množina M je spočetná, když existuje vzájemně jednoznačné zobrazení množiny N na množinu M.
Např. množina celých čísel je spočetná; neprázdný interval (a, b) není spočetná množina.

<u>Supremum množiny</u>

Buď M neprázdná podmnožina množiny reálných čísel. Reálné číslo ξ (pokud existuje) se nazývá supremum množiny M (značíme $\xi = \sup M$), jestliže

1) pro každé $a \in M$ je $a \leq \xi$,
2) číslo ξ je nejmenší z čísel mající vlastnost 1).

Např. množina $\langle 0, 0.4)$ stejně jako množina $\langle 0, 0.4\rangle$ má supremum 0.4. Množina $\langle 0, \infty)$ nemá supremum.

Poznámka

Definice, věty i příklady jsou v rámci každé kapitoly číslovány zvlášť tak, že např. v kapitole 2 musí po příkladu 2.1 následovat příklad 2.2.

Odkaz na větu 3.1 v části Teorie pravděpodobnosti [Matematická statistika] znamená, že se odvoláváme na větu 3.1 z části Teorie pravděpodobnosti [Matematická statistika]. Odkaz na větu P - 3.1 v části Matematická statistika znamená, že se odvoláváme na větu 3.1 z části Teorie pravděpodobnosti.

Alternativní možnost uvádíme v hranatých závorkách, části názvosloví, které se někdy vynechávají, jsou v závorkách kulatých. Konec příkladu je značen ♣, konec důkazu je označen $\square$.

TEORIE PRAVDĚPODOBNOSTI

Předmětem teorie pravděpodobnosti je zkoumání zákonitostí náhody. Náhodu lze charakterizovat jako souhrn drobných vlivů, které nelze zcela nebo vůbec zjistit. Právě náhoda je příčinou toho, že výsledek některých činností nebo procesů nejsme schopni s jistotou předpovědět. Dále se budeme takovými činnostmi a procesy zabývat.

1. Pokus, náhodná veličina, náhodný vektor, výběrová rozdělovací funkce, histogram

S pojmem pokus jsme se již všichni setkali a běžně jej používáme. Pravdou ale je, že bychom již asi hůře formulovali, co si pod tímto pojmem představujeme. Někteří autoři definují pokus následovně.

Definice 1.1. *Pokus je jakákoliv činnost nebo proces, které jsou libovolně mnohokrát opakovatelné a které se uskutečňují za pevně vymezené soustavy podmínek.*

Podmínky, o nichž hovoří definice pokusu, jsou nenáhodné a bývají často explicitně či implicitně obsaženy již v názvu pokusu. Jsou to tzv. definiční podmínky pokusu. Definice pokusu ovšem připouští, že na výsledek pokusu působí ještě další faktory (vlivy), obecně náhodné. Tyto faktory nejsou mezi definičními podmínkami pokusu. Každý z nich musí nabývat alespoň dvou různých stavů, jinak by patřil mezi nenáhodné vlivy. Dále se omezíme na ty pokusy, u kterých hraje roli vliv náhody. Správně bychom je měli nazývat náhodné pokusy, ale slovo náhodné budeme vynechávat.

Je-li např. pokus nazván hod hrací kostkou, je jeho názvem již vyslovena (nebo nevyslovena?) řada podmínek, které jej definují. Předpokládá se mimo jiné, že kostka je krychlového tvaru, je z homogenního materiálu, je hozena určitým způsobem atd. Mezi náhodné vlivy zařadíme např. polohu kostky v ruce (kelímku, házecím stroji, ...) při počátku hodu, rychlost a směr hodu.

Zabývat se budeme jen takovými pokusy, u kterých je výsledkem každého opakování n-tice reálných čísel, kde n je pevné přirozené číslo.

Množinu všech teoreticky možných výsledků pokusu značíme Ω a nazýváme základní prostor pokusu. Každou n-tici reálných čísel lze chápat jako bod prostoru $\mathbb{R}^n$, proto je Ω podmnožinou $\mathbb{R}^n$.

Řekli jsme, že výsledek pokusu nelze s jistotou předvídat. Tento neznámý vysledek (tj. výsledek pokusu do té doby, než pokus provedeme) budeme nazývat náhodný vektor a značit $(X_1, X_2, \dots, X_n)$. Když provedeme pokus a zapíšeme výsledek, dostaneme n-tici reálných čísel $(x_1, x_2, \dots, x_n)$, tzv. realizaci náhodného vektoru. Složky náhodného vektoru tedy značíme velkými písmeny, složky realizace náhodného vektoru pak odpovídajícími malými písmeny. Množinu všech teoreticky možných výsledků pokusu Ω nazýváme v této souvislosti obor hodnot náhodného vektoru. Je-li $n = 1$, nazýváme náhodný vektor náhodná veličina nebo proměnná a značíme X, její realizaci pak x.

Uveďme některé příklady pokusů, náhodných veličin a náhodných vektorů:

Příklad 1.1. Uvažujme hod hrací kostkou a sledujme počet ok, které padnou při jednom hodu touto kostkou. Je to činnost, kterou lze libovolně mnohokrát opakovat a výsledkem každého opakování je jedno z čísel 1, 2, ..., 6. Jedná se tedy o pokus se základním prostorem $\Omega = \{1, 2, \dots, 6\}$. Neznámým výsledkem pokusu je zde náhodná veličina X, tj. počet ok, které padnou při jednom hodu hrací kostkou.

♣

Příklad 1.2. Náhodně vybíráme bod z intervalu $\langle 0, 2 \rangle$. Jedná se zřejmě o pokus. Nechť X je x-ová souřadnice bodu, který bude vybrán. Zřejmě X je náhodná veličina s oborem hodnot $\Omega = \langle 0, 2 \rangle$.

♣

Příklad 1.3. Vyberme konkrétní křižovatku samostatně řízenou světelným signalizačním zařízením podle pevně nastaveného signalizačního plánu. Zapisujme počet zastavených, počet přibrzděných a počet volně projetých vozidel za hodinu. Je to opět činnost, kterou lze libolně mnohokrát opakovat a výsledkem každého opakování je trojice celých nezáporných čísel. Jde tedy o pokus. Nechť X [Y, resp. Z] je počet vozidel, které zastaví [přibrzdí, resp. volně projedou] během hodiny na dané křižovatce. Potom (X, Y, Z) je náhodný vektor s oborem hodnot $\Omega \subseteq \{0, 1, 2, \dots\} \times \{0, 1, 2, \dots\} \times \{0, 1, 2, \dots\}$.

♣

Příklad 1.4. Vyrábějme betonové krychle o hraně délky deset centimetrů, každou z nich nejprve zvažme a potom podrobme tlakové zkoušce pevnosti. Zapišme obě výsledné hodnoty jako dvojici reálných čísel (x, y), kde x je hmotnost krychle (v kg) a y je minimální tlak potřebný k jejímu rozdrcení (v MPa). O množině Ω možných výsledků je zřejmé, že je podmnožinou množiny $(0, \infty) \times (0, \infty)$ a tedy podmnožinou $\mathbb{R}^2$. Jde o pokus. Nechť X je neznámý výsledek vážení a Y je neznámý výsledek tlakové zkoušky. Potom (X, Y) je náhodný vektor s oborem hodnot Ω. Příkladem realizace náhodného vektoru (X, Y) je uspořádaná dvojice $(2.3, 31.8)$.

♣

Příklad 1.5. Pro zjištění kvality nově vyvinutého měřicího přístroje provádíme měření mezi dvěma nivelačními body. Správná hodnota výsledku je známá předem. Zapisujeme chybu měření. Množina možných výsledků Ω je zřejmě podmnožina

$\mathbb{R}$, jedná se tedy o pokus. Nechť X je náhodná chyba měření, potom X je náhodná veličina s oborem hodnot Ω.

♣

Obsahem zbývající části kapitoly je motivace pojmů pravděpodobnost a rozdělovací funkce. Definice těchto pojmů budou uvedeny až v kapitole třetí a druhé.

Podobně jako s pojmem pokus jsme se všichni již setkali s pojmem pravděpodobnost a v nejjednodušších případech ji umíme vypočítat. Každý z nás asi ví, že pravděpodobnost padnutí sudého počtu ok při jednom hodu hrací kostkou je 1/2. K tomuto závěru dospějeme na základě toho, že kostka je homogenní, tj. každý výsledek pokusu je stejně možný a základní prostor pokusu je konečná množina. Samozřejmě, že tato skutečnost neznamená, že když např. desetkrát hodíme kostkou, musí sudý počet ok padnout pětkrát. Kdybychom ale mnohokrát hodili kostkou, měl by v přibližně polovině hodů padnout sudý počet ok. Podobně pravděpodobnost, že v příkladu 1.2 vybereme bod z intervalu $< 0, 1 >$ za předpokladu, že každý bod ze základního prostoru má stejnou „šanci“ být vybrán je 1/2. Kromě stejné šance jsme využili skutečnosti, že základní prostor je podmnožinou $\mathbb{R}$ konečné délky. Pravděpodobnost, že počet volně projetých vozidel křižovatkou během hodiny je maximálně 75, resp. pravděpodobnost, že pevnost materiálu je maximálně 31 MPa již samozřejmě tak triviálními úvahami nejsme schopni spočítat.

Z uvedených případů je snad patrné, že nás budou v technických aplikacích spíše zajímat ty situace, kdy nelze použít žádnou z výše uvedených úvah a ani s intuicí nevystačíme.

Zavedeme tedy pojem pravděpodobnosti tak, jak se v technických aplikacích pravděpodobnost prakticky počítá, tj. pomocí tzv. rozdělovací funkce. Ukážeme, že námi uvedená definice zahrnuje výše uvedené případy, ale i mnohem obecnější situace.

Nejprve je ale zapotřebí zavést pojem rozdělovací funkce. Uvědomme si, že výroky „Buď dán pokus se základním prostorem Ω.“ a „Buď dán náhodný vektor $(X_1, X_2, \ldots, X_n)$ s oborem hodnot Ω.“ vlastně znamenají totéž.

Omezme se nyní na případ náhodné veličiny X s oborem hodnot $\Omega \subset \mathbb{R}$. Naznačíme, jak si lze představit rozdělovací funkci v souvislosti s tzv. výběrovou rozdělovací funkcí náhodné veličiny X.

Proveďme n-krát pokus a označme $x_{\min}$ nejmenší a $x_{\max}$ největší ze zjištěných hodnot. Interval $\langle x_{\min}, x_{\max} \rangle$ nazýváme variační obor. Rozložme nyní variační obor na k disjunktních tříd Ω_j, $j = 1, 2, \ldots, k$. Při rozkladu dbáme zpravidla těchto zásad.

1. Je-li mezi zjištěnými hodnotami jen malý počet navzájem různých hodnot, volíme každou hodnotu za třídu Ω_j. Tato situace odpovídá případu, kde je Ω konečná nebo spočetná množina. Mluvíme o tzv. prostém třídění.

2. Je-li mezi zjištěnými hodnotami značně velký počet různých hodnot volíme za třídy Ω_j intervaly. Tato situace odpovídá případu, kde je Ω nespočetná množina, např. interval. Přitom délky intervalů volíme zpravidla stejné. Doporučuje se, aby intervalů bylo $5 - 20$ podle počtu pokusů n. Mluvíme o tzv. skupinovém třídění.

Nechť n_j značí počet výsledků, které padly do třídy Ω_j. Číslo n_j nazýváme absolutní četnost třídy Ω_j. Potom zřejmě

$$n_1 + n_2 + \cdots + n_k = n.$$

Číslo $f_j = \frac{n_j}{n}$ nazýváme relativní četnost třídy Ω_j. Zřejmě

$$f_1 + f_2 + \cdots + f_k = 1.$$

Označme dále $m(\Omega_j)$ počet prvků [délku] třídy Ω_j v případě prostého [skupinového] třídění.

Definujme nyní reálnou funkci g_n předpisem

$$g_n(x) = \begin{cases} \dfrac{n_j}{n \cdot m(\Omega_j)} & \text{pro} \quad x \in \Omega_j,\ j = 1, \ldots, k \\ 0 & \text{jinak} \end{cases}.$$

Funkci g_n budeme nazývat výběrová rozdělovací funkce náhodné veličiny X, její graf budeme nazývat histogram.

Zvětšujme počet n provedení pokusu do nekonečna a nechť (proti úmluvě) také počet k tříd Ω_j stoupá až na počet rovný počtu bodů množiny Ω. Potom posloupnost funkcí g_n konverguje někdy (nezkoumejme jak a kdy) k limitní funkci - označme ji g. Tuto funkci budeme nazývat rozdělovací funkce náhodné veličiny X. V dalším se budeme zabývat pouze takovými pokusy, u nichž konvergence nastává.

Příklad 1.6. V příkladu 1.1 zapisujme počet ok, které padnou při jednom hodu hrací kostkou. Stokrát jsme hodili touto hrací kostkou. Jedno a dvě oka padla 15-krát, tři, pět a šest ok padlo 17-krát a čtyři oka 19-krát. Určete výběrovou rozdělovací funkci a histogram.

Řešení. Nechť X je počet ok při jednom hodu hrací kostkou, potom je obor hodnot Ω náhodné veličiny X množina $\{1,2,3,4,5,6\}$. Jedná se o konečnou množinu, zvolíme tedy prosté třídění a položíme $\Omega_j = j,\ j = 1, \ldots, 6$. Výsledky třídění znázorníme formou tabulky rozdělení četností (tabulka 1.1)

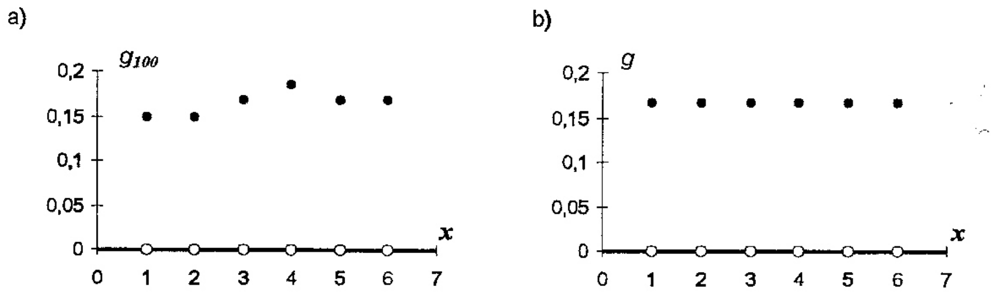
Třída		n_j	f_j
Ω_1	1	15	0.15
Ω_2	2	15	0.15
Ω_3	3	17	0.17
Ω_4	4	19	0.19
Ω_5	5	17	0.17
Ω_6	6	17	0.17
Součet		100	1.00

Tabulka 1.1: Rozdělení četností počtu ok při hodu hrací kostkou

Vzhledem k tomu, že $m(\Omega_j) = 1$, dostáváme výběrovou rozdělovací funkci

$$g_{100}(x) = \begin{cases} 0.15 & \text{pro} \quad x = 1, 2 \\ 0.17 & \text{pro} \quad x = 3, 5, 6 \\ 0.19 & \text{pro} \quad x = 4 \\ 0 & \text{jinak} \end{cases},$$

což lze znázornit graficky pomocí histogramu (obrázek 1.1 a))



Obrázek 1.1: a) Histogram počtu ok pro 100 hodů a 6 tříd
b) Rozdělovací funkce počtu ok při jednom hodu hrací kostkou

Uvědomme si, že číslo n_j lze považovat za „odhad šance", že výsledek pokusu bude Ω_j, tj., že na kostce padne j ok, za úmluvy, že jistotu vyjadřujeme jedničkou a nemožnost nulou. Dále zřejmě

$$\sum_{x\in\Omega} g_n(x) = 1.$$

Limitní rozdělovací funkce při házení homogenní kostkou by asi měla být funkce

$$g(x) = \begin{cases} \dfrac{1}{6} & \text{pro} \quad x = 1, 2, \ldots, 6 \\ 0 & \text{jinak} \end{cases},$$

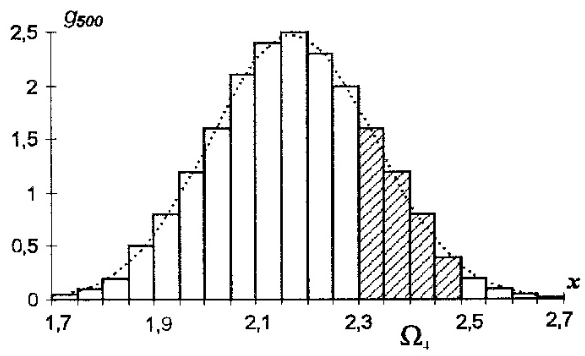
jejíž graf je na obrázku 1.1 b). Šanci, že padne číslo x pak vyjadřuje zřejme $g(x)$ ideálně a opět platí

$$\sum_{x\in\Omega} g(x) = 1.$$

♣

Příklad 1.7. V příkladu 1.4 zapisujme pouze výsledek vážení (v kg) Dejme tomu, že máme $n = 100$ krychlí vyrobených za stejných podmínek a po jejich zvážení jsme obdrželi následující výsledky (v kg):

na výsledek z původního Ω_4 ideálně.



Obrázek 1.3: Histogram váhy betonové kostky pro 500 měření a 20 tříd

♣

Častěji se požaduje, aby výška obdélníků nad třídami Ω_j byla rovna relativní, resp. absolutní četnosti příslušné třídy Ω_j. Hovoříme pak o histogramu relativních, resp. absolutních četností. V případě skupinového třídění pak tyto histogramy už samozřejmě nezaručují, že obsah obrazce ohraničeného osou x a histogramem je jedna. Podobně v případě prostého třídění není již součet výšek „obdélníků" roven jedné.

Tabulku rozdělení četností a histogram četností lze vytvořit i pomocí programu EXCEL (v nabídce Nástroje $\longmapsto$ Analýza dat ... $\longmapsto$ Histogram) nebo pomocí statistického software STATGRAPHICS (v nabídce STATS $\longmapsto$ Descriptive Methods $\longmapsto$ Descriptive Methods $\longmapsto$ Frequency Tabulation, Frequency Histogram).

Charakter pokusu, a tedy i náhodné veličiny, lze hodnotit z mnoha pohledů; nás budou zajímat dva.

První hledisko: Určitě nastane jedna z možností:

Ia) Všechny rozdělovací funkce jsou identické nebo se liší tak málo, že jsme ochotni je za identické považovat.

Ib) Alespoň dvě rozdělovací funkce jsou různé.

Má-li náhodná veličina vlastnost Ia, znamená to, že náhodné faktory působí při velkém počtu opakování pokusu v souhrnu jakýmsi způsobem zákonitě. Tuto vlastnost lidé již mnohokrát zkoušeli prakticky prověřit, literatura uvádí například dvacet čtyři tisíc hodů stejnou mincí (s výsledkem 12012 padnutí jedné a 11988 druhé strany mince).

V dalším se budem zabývat výhradně pokusy s vlastností Ia. To má za důsledek, že ke každé náhodné veličině - a vlastně teprve nyní při vlastnosti Ia bychom ji měli nazývat náhodnou - bude existovat jediná rozdělovací funkce g.

Druhé hledisko: Buď X náhodná veličina s rozdělovací funkcí g a oborem hodnot Ω. Lehce nahlédneme, že musí nastat jedna ze tří následujících situací:

IIa) Základní prostor Ω je tvořen konečně mnoha nebo spočetně mnoha body a platí:

$$\sum_{x \in \Omega} g(x) = 1.$$

IIb) Základní prostor Ω je taková podmožina $\mathbb{R}$, která má nenulovou délku,tj. existuje

$$\int_{\Omega} dx$$

a je různý od nuly. Rozdělovací funkce g je taková, že

$$\int_{\Omega} g(x) dx = 1.$$

IIc) Neplatí IIa ani IIb.

Má-li náhodná veličina vlastnost IIa, budeme ji nazývat diskrétní, v případě vlastnosti IIb budeme hovořit o spojité náhodné veličině. Jinými náhodnými veličinami než spojitými a diskrétními se zabývat nebudeme.

Analogicky jako výše bychom mohli zavést pojem výběrové rozdělovací funkce náhodného vektoru $(X_1, X_2, \ldots, X_n)$ a provést i další úvahy. V tomto případě je ale základní prostor Ω podmnožinou $\mathbb{R}^n$ a výklad by nebyl tak názorný, nemluvě o grafickém znázornění, které má význam pro $n \leq 2$.

2. Diskrétní a spojitý náhodný vektor, pravděpodobnostní funkce a hustota

V této kapitole budeme definovat pojmy z jejího nadpisu a uvedeme některé jejich vlastnosti. Nejprve začneme případem náhodné veličiny a pak přejdeme na náhodný vektor.

Připomeňme, že jsme v kapitole 1 náhodnou veličinu s vlastností IIa chtěli nazývat diskrétní. V tomto případě potom většina autorů používá název pravděpodobnostní funkce místo rozdělovací funkce. Tedy je možné definovat diskrétní náhodnou veličinu a pravděpodobnostní funkci například takto:

Definice 2.1. *Buď X náhodná veličina s oborem hodnot Ω. Řekneme, že náhodná veličina X je diskrétní (s pravděpodobnostní funkcí p), když*

a) Ω je nejvýše spočetná množina,

b) funkce $p\colon \mathbb{R} \longrightarrow \mathbb{R}$ má následující vlastnosti:

1) $p(x) \geq 0$ pro každé $x \in \mathbb{R}$,

2) $p(x) > 0$ právě tehdy, když $x \in \Omega$,

(2.1) 3) $\displaystyle\sum_{x\in\Omega} p(x) = 1.$

Funkci p nazýváme pravděpodobnostní funkce náhodné veličiny X. Značíme $X \sim p$.

Zcela obdobně s uvážením vlastnosti IIb z kapitoly 1 obdržíme definici spojité náhodné veličiny a rozdělovací funkce, která v tomto případě dostane název hustota náhodné veličiny.

Definice 2.2. *Buď X náhodná veličina s oborem hodnot Ω. Řekneme, že náhodná veličina X je spojitá (s hustotou f), když*

a) existuje

$$\int_{\Omega} dx$$

a je nenulový,

b) funkce $f : \mathbb{R} \longrightarrow \mathbb{R}$ má následující vlastnosti:

1) $f(x) \geq 0$ pro každé $x \in \mathbb{R}$,

2) $f(x) > 0$ právě tehdy, když $x \in \Omega$,

(2.2) 3) $\displaystyle\int_{\Omega} f(x)\, dx = 1.$

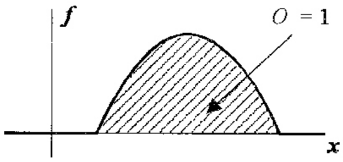
Funkci f nazýváme hustota (pravděpodobnosti) náhodné veličiny X. Značíme $X \sim f$.

Definice 2.3. *Pravděpodobnostní funkci a hustotu nazýváme rozdělovací funkce. Když známe rozdělovací funkci, říkáme, že známe zákon rozložení (rozdělení).*

Poznámka 2.1.

1) Vztah (2.1) říká, že součet existuje a je roven jedné. Korektnost vztahu pro konečný obor hodnot je zřejmá. Vzhledem k tomu, že v případě spočetného oboru hodnot jde o součet nekonečné řady s kladnými členy, není fakt, že neznáme pořadí sčítanců, na závadu a vztah je znovu korektní.

2) Vlastnost a) z definice 2.2 si můžeme představit tak, že „délka“ oboru hodnot Ω je nenulová. Vlastnost b 3) z téže definice říká, že inegrál existuje a je roven jedné. Vzhledem k tomu, že hustota je na $\mathbb{R} - \Omega$ nulová a na Ω kladná, lze si tuto vlastnost představit tak, že obsah O obrazce ohraničeného osou x a grafem hustoty f je roven jedné (viz obrázek 2.1).



Obrázek 2.1

3) Již u těchto definic stojí za povšimnutí, že tam kde se vyskytuje v diskrétním případě „suma“ a pravděpodobnostní funkce, je ve spojitém případě „integrál“ a hustota.

4) Aby nějaká funkce byla rozdělovací funkce nějaké náhodné veličiny, musí v diskrétním [spojitém] případě splňovat podmínky z definice 2.1 [2.2].

Příklad 2.1. Náhodná veličina X má pravděpodobnostní funkci

$$\text{a) } p(x) = \begin{cases} c\,(x+1) & \text{pro } x = 1,2,3 \\ 0 & \text{jinak} \end{cases}$$

$$\text{b) } p(x) = \begin{cases} c\left(\dfrac{1}{2}\right)^{x} & \text{pro } x = 2,3,\dots \\ 0 & \text{jinak} \end{cases}.$$

Určete konstantu c.

Řešení. V obou případech vyjdeme ze vztahu (2.1).

a) Obor hodnot Ω náhodné veličiny X je množina $\{1,2,3\}$. Musí tedy platit

$$\sum_{x=1}^{3} c\,(x+1) = 1.$$

Protože

$$\sum_{x=1}^{3} c\,(x+1) = 9\,c,$$

je $c = 1/9$. Poněvadž $c > 0$, jsou splněny i ostatní podmínky z definice 2.1 a dostáváme

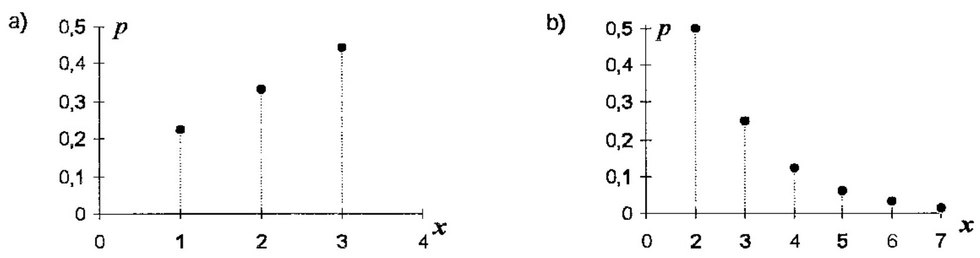
$$p(x) = \begin{cases} \dfrac{1}{9}(x+1) & \text{pro } x = 1,2,3 \\ 0 & \text{jinak} \end{cases}.$$

Pravděpodobnostní funkci můžeme vyjádřit tabulkou, ve které uvedeme realizace x náhodné veličiny X a hodnoty pravděpodobnostní funkce $p(x)$ v těchto bodech (tabulka 2.1).

x	1	2	3
$p(x)$	2/9	3/9	4/9

Tabulka 2.1

Graf pravděpodobnostní funkce je na obrázku 2.2 a). Pro zjednodušení zobrazujeme pouze hodnoty p pro $x \in \Omega$, tj. nenulové hodnoty pravděpodobnostní funkce, což budeme dělat i nadále.



Obrázek 2.2

b) Obor hodnot Ω náhodné veličiny X je množina $\{2, 3, \ldots.\}$. Musí tedy platit

$$\sum_{x=2}^{\infty} c\left(\frac{1}{2}\right)^x = 1.$$

Na základě výpočtu součtu geometrické řady dostáváme

$$\sum_{x=2}^{\infty} c\left(\frac{1}{2}\right)^x = c\,\frac{\frac{1}{4}}{1-\frac{1}{2}} = \frac{1}{2}\,c,$$

tedy $c = 2$. Poněvadž $c > 0$, jsou splněny všechny podmínky z definice pravděpodobnostní funkce a dostáváme

$$p(x) = \begin{cases} \left(\frac{1}{2}\right)^{x-1} & \text{pro } x = 2, 3, \ldots \\ 0 & \text{jinak} \end{cases}.$$

Hodnoty pravděpodobnostní funkce jsou uvedeny v tabulce 2.2, graf pravděpodobnostní funkce je na obrázku 2.2 b).

x	2	3	4	5	6	.	.	.
$p(x)$	1/2	1/4	1/8	1/16	1/32	.	.	.

Tabulka 2.2

♣

Příklad 2.2. Náhodná veličina X má hustotu

$$f(x) = \begin{cases} c\,(1+x^2) & \text{pro } \ x \in \langle -1, 1\rangle \\ 0 & \text{jinak} \end{cases}.$$

Určete konstantu c.

Řešení. Obor hodnot Ω náhodné veličiny X je interval $\langle -1, 1\rangle$. Aby f byla hustota náhodné veličiny X musí platit vztah (2.2), tedy

$$\int_{-1}^{1} c\,(1+x^2)\,dx = 1.$$

Protože

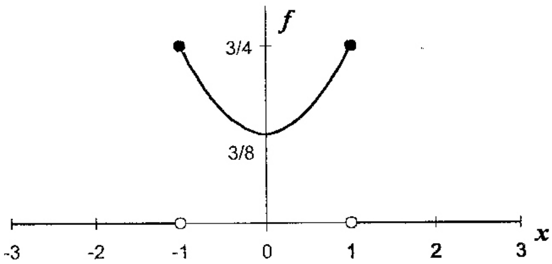
$$\int_{-1}^{1} c\,(1+x^2)\,dx = 2c\int_{0}^{1} (1+x^2)\,dx = c\left[\frac{x^3}{3}+x\right]_0^1 = \frac{8}{3}\,c,$$

dostáváme $c = 3/8$.

Protože $c > 0$ a $\int\limits_{\Omega} dx = 2 \neq 0$, jsou splněny i ostatní podmínky z definice 2.2 a tedy

$$f(x) = \begin{cases} \dfrac{3}{8}\,(1+x^2) & \text{pro } \ x \in \langle -1, 1\rangle \\ 0 & \text{jinak} \end{cases}.$$

Graf hustoty f je na obrázku 2.3.



Obrázek 2.3

♣

Příklad 2.3. Určete konstanty A a α tak, aby funkce

$$f(x) = \begin{cases} \dfrac{1}{\alpha}\cdot \mathrm{e}^{-\dfrac{x-A}{\alpha}} & \text{pro } \ x > A \\ 0 & \text{jinak} \end{cases}$$

byla hustota pravděpodobnosti náhodné veličiny X.

Řešení. Obor hodnot Ω náhodné veličiny X je interval (A, ∞). Vyjdeme opět ze vztahu (2.2), potom musí platit

$$\int_A^\infty \frac{1}{\alpha} \cdot \mathrm{e}^{-\frac{x-A}{\alpha}}\, dx = 1$$

Pro výpočet integrálu zavedeme substituci

$$u = -\frac{x-A}{\alpha} \Rightarrow -\alpha\, du = dx$$

Je-li

$$\begin{array}{l} 1)\ \alpha > 0 \ \text{ potom } \ u(A) = 0,\ u(\infty) = -\infty, \\ 2)\ \alpha < 0 \ \text{ potom } \ u(A) = 0,\ u(\infty) = \infty. \end{array}$$

Potom

$$1) \int_A^\infty \frac{1}{\alpha} \cdot \mathrm{e}^{-\frac{x-A}{\alpha}}\, dx = -\int_0^{-\infty} \mathrm{e}^u\, du = -\left[\mathrm{e}^u\right]_0^{-\infty} = 1,$$

$$2) \int_A^\infty \frac{1}{\alpha} \cdot \mathrm{e}^{-\frac{x-A}{\alpha}}\, dx = -\int_0^{\infty} \mathrm{e}^u\, du = -\left[\mathrm{e}^u\right]_0^{\infty} = -\infty.$$

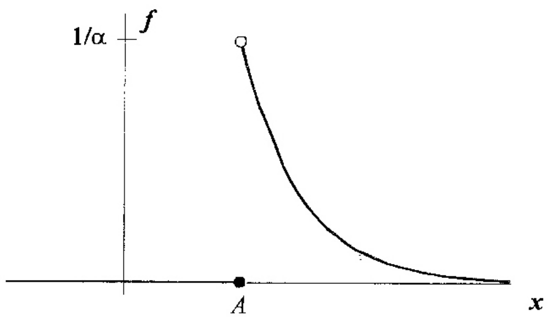
Požadovaný vztah je tedy splněn pouze v případě 1), tj. pro každé A, $\alpha \in \mathbb{R}$; $\alpha > 0$. Aby $f(x)$ byla hustota, musí ještě platit

$$\int_\Omega dx \neq 0 \quad \text{a} \quad f(x) > 0 \quad \text{pro každé } x > A.$$

Zřejmě

$$\int_A^\infty dx = \infty \neq 0 \quad \text{a} \quad \frac{1}{\alpha} \cdot \mathrm{e}^{-\frac{x-A}{\alpha}} > 0 \quad \Leftrightarrow \quad \alpha > 0,\ A \in \mathbb{R},\ x \in \mathbb{R}.$$

Tedy f je hustota pro libovolné A a kladné α. Graf hustoty f je na obrázku 2.4.



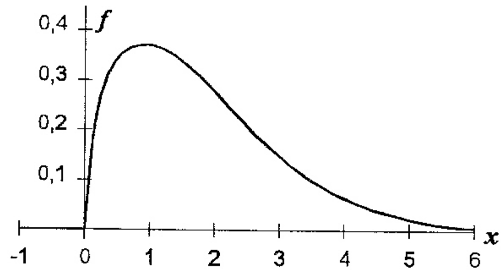
Obrázek 2.4

♣

Poznámka 2.2. Z příkladu 2.3 je vidět, že konstanty A a α již nejsme schopni na základě vlastností hustoty dále více specifikovat. Tyto neznámé konstanty v rozdělení budeme nazývat parametry. Jejich hodnotu lze pouze v praktických situacích odhadnout, ale to je otázka až matematické statistiky. O náhodné veličině X z příkladu 2.3 říkáme, že má <u>exponenciální rozdělení s parametry A a α</u>, $A \in \mathbb{R}, \alpha > 0$, což značíme $X \sim E(A, \alpha)$. Toto rozdělení má širokou použitelnost např. v teorii spolehlivosti a životnosti. Rozdělení $E(0, \alpha)$ popisuje dobře rozložení doby života zařízení, u kterých dochází k poruše ze zcela náhodných (vnějších) příčin a nikoliv zákonitě v důsledku opotřebení např. mechanického, únavy materiálu, případně v důsledku skrytých vad. Doba života (tj. doba do poruchy) zařízení, u kterých se projevuje mechanické opotřebení a únava materiálu má tzv. <u>Weibulovo rozdělení s parametry δ a c, $\delta > 0, c > 0$</u>, které značíme $W(\delta, c)$. Hustota Weibulova rozdělení je

$$f(x) = \begin{cases} \dfrac{c\,x^{c-1}}{\delta^c}\,\mathrm{e}^{-\left(\frac{x}{\delta}\right)^c} & \text{pro } x > 0\,, \\ 0 & \text{jinak} \end{cases}$$

jejíž graf je pro $c = 1.5$ a $\delta = 2$ na obrázku 2.5.



Obrázek 2.5

Dále vyslovíme definice diskrétního a spojitého náhodného vektoru a pravděpodobnostní funkce, resp. hustoty. Tyto definice budou analogické definicím diskrétní, resp. spojité náhodné veličiny. Stačí si pouze uvědomit, že obor hodnot Ω náhodného vektoru $(X_1, X_2, \ldots, X_n)$ je podmnožinou $\mathbb{R}^n$ a tudíž zřejmě rozdělovací funkce nebude funkcí jedné proměnné, ale n proměnných $x_1, x_2, \ldots x_n$.

Definice 2.4. *Buď $(X_1, X_2, \ldots, X_n)$ náhodný vektor s oborem hodnot Ω. Řekneme, že náhodný vektor $(X_1, X_2, \ldots, X_n)$ je <u>diskrétní (s pravděpodobnostní funkcí p)</u>, když*

a) Ω je nejvýše spočetná množina,

b) funkce $p\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ má následující vlastnosti:

1) $p(x_1, x_2, \ldots, x_n) \geq 0$ pro každé $(x_1, x_2, \ldots x_n) \in \mathbb{R}^n$,

2) $p(x_1, x_2, \ldots, x_n) > 0$ právě tehdy, když $(x_1, x_2, \ldots x_n) \in \Omega$,

(2.3) *3)* $$\sum_{(x_1, x_2, \ldots, x_n) \in \Omega} p(x_1, x_2, \ldots, x_n) = 1.$$

♣

Funkci p nazýváme pravděpodobnostní funkce *náhodného vektoru $(X_1, \dots, X_n)$. Značíme $(X_1, \dots, X_n) \sim p$.*

Definice 2.5. *Buď $(X_1, X_2, \dots, X_n)$ náhodný vektor s oborem hodnot Ω. Řekneme, že náhodný vektor $(X_1, X_2, \dots, X_n)$ je* spojitý (s hustotou f)*, když*

a) existuje

$$\int \cdots \int_{\Omega} dx_1 dx_2 \dots dx_n$$

a je nenulový,

b) funkce $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ má následující vlastnosti:

1) $f(x_1, x_2, \dots, x_n) \geq$ pro každé $(x_1, x_2, \dots x_n) \in \mathbb{R}^n$,

2) $f(x_1, x_2, \dots, x_n) > 0$ právě tehdy, když $(x_1, x_2, \dots x_n) \in \Omega$,

(2.4) 3) $\int \cdots \int_{\Omega} f(x_1, x_2, \dots, x_n)\, dx_1 dx_2 \dots dx_n = 1$.

Funkci f nazýváme hustota (pravděpodobnosti) *náhodného vektoru $(X_1, \dots, X_n)$. Značíme $(X_1, \dots, X_n) \sim f$.*

Pokud čtenáři dělá problémy pracovat s funkcí n proměnných a n - rozměrnou sumou, resp. integrálem, doporučujeme mu, aby si předchozí dvě definice přepsal pro případ $n = 2$ a zamyslel se v případě definice 2.5 nad geometrickým významem požadovaných vlastností.

Příklad 2.4. Určete konstantu c tak, aby funkce

$$g(x, y) = \begin{cases} c\,(x^2 + y) & \text{pro } (x, y) \in \{0, 1, 2\} \times \{1, 2, 3\} \\ 0 & \text{jinak} \end{cases}$$

byla pravděpodobnostní funkce náhodného vektoru (X, Y).

Řešení. Obor hodnot Ω náhodného vektoru (X, Y) je množina $\{0, 1, 2\} \times \{1, 2, 3\}$. Aby $g(x, y)$ byla pravděpodobnostní funkce, musí platit vztah (2.3), tj.

$$\sum_{x=0}^{2} \sum_{y=1}^{3} c\,(x^2 + y) = 1.$$

Protože

$$\sum_{x=0}^{2} \sum_{y=1}^{3} c\,(x^2 + y) = c \sum_{x=0}^{2} \Big[(x^2 + 1) + (x^2 + 2) + (x^2 + 3)\Big] = 33\, c,$$

je $c = 1/33$. Poněvadž $c > 0$, jsou splněny i ostatní podmínky z definice 2.4 a dostáváme

$$g(x, y) = \begin{cases} \dfrac{1}{33}\,(x^2 + y) & \text{pro } (x, y) \in \{0, 1, 2\} \times \{1, 2, 3\} \\ 0 & \text{jinak} \end{cases}.$$

Pro vyjádření pravděpodobnostní funkce můžeme využít i tabulky, ve které jsou uvedeny hodnoty pravděpodobnostní funkce $p(x,y)$ náhodného vektoru (X,Y) v bodech $(x,y) \in \Omega$.

$x \setminus y$	1	2	3
0	1/33	2/33	3/33
1	2/33	3/33	4/33
2	5/33	6/33	7/33

Tabulka 2.3

♣

Příklad 2.4. Určete konstantu c tak, aby funkce

$$g(x,y) = \begin{cases} c\,(x+2y) & \text{pro } (x,y) \in \langle 0,1\rangle \times \langle 0,2\rangle \\ 0 & \text{jinak} \end{cases}$$

byla hustota náhodného vektoru (X,Y).

Řešení. Obor hodnot Ω náhodného vektoru (X,Y) je množina $\langle 0,1\rangle \times \langle 0,2\rangle$. Jedná se o obdélník, jehož obsah je 2. Tedy $\int\limits_{\Omega} dxdy = 2 \neq 0$. Aby $g(x,y)$ byla hustota náhodného vektoru (X,Y), musí podle definice 2.5 platit vztah (2.4), tedy

$$\iint\limits_{\Omega} c\,(x+2y)\,dxdy = 1.$$

Protože

$$\iint\limits_{\Omega} c\,(x+2y)dxdy = \int\limits_0^2 \Big[\int\limits_0^1 c\,(x+2y)\,dx\Big]\,dy$$
$$= c\int\limits_0^2 \Big[\frac{x^2}{2} + 2xy\Big]_0^1 dy = c\int\limits_0^2 (2y + \frac{1}{2})\,dy = 5\,c,$$

je $c = \dfrac{1}{5}$. Vzhledem k tomu, že $c > 0$, jsou splněny i zbývající podmínky z definice hustoty a dostáváme

$$g(x,y) = \begin{cases} \frac{1}{5}\,(x+2y) & \text{pro } (x,y) \in \langle 0,1\rangle \times \langle 0,2\rangle \\ 0 & \text{jinak} \end{cases}.$$

♣

3. Pravděpodobnost

V této kapitole budeme definovat pojem pravděpodobnosti tak, jak se v technických aplikacích prakticky počítá, tj. pomocí rozdělovací funkce zavedené v předchozí kapitole. Nejprve se opět omezíme na případ náhodné veličiny. Zobecnění pro náhodný vektor pak provedeme zcela analogicky jako v předchozí kapitole.

V kapitole 1 jsme dospěli k závěru, že „šanci“, že diskrétní náhodná veličina X nabude hodnoty x, vyjadřuje ideálně hodnota rozdělovací funkce v bodě x. V případě spojité náhodné veličiny jsme „šanci“, že tato veličina nabude hodnoty z určitého intervalu, vyjádřili obsahem plochy nad tímto intervalem zhora ohraničené grafem rozdělovací funkce. Je tedy přirozené definovat pravděpodobnost následovně:

Definice 3.1. *Buď X diskrétní [spojitá] náhodná veličina s oborem hodnot Ω a rozdělovací funkcí g. Buď A libovolná podmnožina $\mathbb{R}$ [podmnožina $\mathbb{R}$, pro kterou $\int_A dx$ existuje]. Potom A nazýváme náhodný jev.*

Pravděpodobnost, že nastane jev A, tj. pravděpodobnost, že náhodná veličina X nabude hodnoty z množiny A, značíme

$$P(X \in A)$$

a definujeme

$$P(X \in A) = \sum_{x \in A \cap \Omega} g(x). \tag{3.1}$$

$$\left[P(X \in A) = \int_{\Omega \cap A} g(x)\, dx\right]. \tag{3.2}$$

Předpisem (3.1) [(3.2)] je definováno zobrazení množiny všech náhodných jevů do množiny reálných čísel $\mathbb{R}$. Toto zobrazení nazýváme pravděpodobnost (příslušná náhodné veličině X).

Poznámka 3.1.

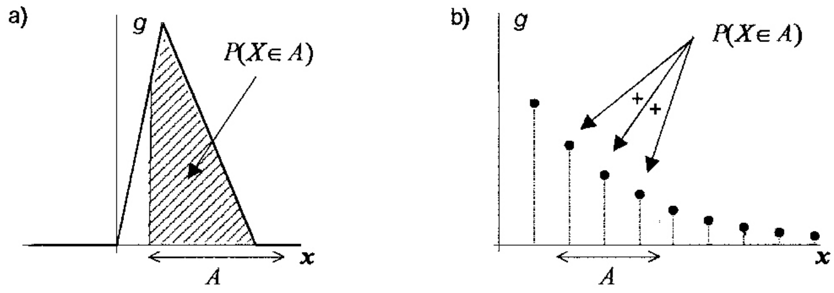
1) Nechť X je diskrétní [spojitá] náhodná veličina s rozdělovací funkcí g. Výpočet $P(X \in A)$ je graficky znázorněn na obrázku 3.1 a) [b)].

2) Opět si všimněme, že v diskrétním případě pracujeme se „sumou“ a pravděpodobnostní funkcí, ve spojitém případě s „integrálem“ a hustotou.

3) Protože hustota g je na $\mathbb{R} - \Omega$ nulová, platí

$$P(X \in A) = \int_A g(x)\, dx. \tag{3.3}$$

PDF-Xchange Click to buy NOW! www.docu-track.com



Obrázek 3.1: Grafické znázornění výpočtu pravděpodobnosti z rozdělovací funkce

Z definice pravděpodobnosti a z vlastností nekonečných řad reálných čísel v diskrétním případě, stejně jako z definice pravděpodobnosti a vlastností integrálů v případě spojitém, lze téměř okamžitě dokázat následující větu:

Věta 3.1. Buď P pravděpodobnost příslušná náhodné veličině X s oborem hodnot Ω. Potom:

1) Pro každý jev A platí $0 \leq P(X \in A) \leq 1$.
2) $P(X \in \Omega) = 1$.
3) Buďte $A_1, A_2, A_3, \ldots$ jevy navzájem disjunktní (tj. takové, že pro $i \neq j$ platí $A_i \cap A_j = \emptyset$). Potom

$$P\Big(X \in \bigcup_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} P(X \in A_i).$$

4) $P(X \in \emptyset) = 0$.
5) $P(X \notin A) = 1 - P(X \in A)$.
6) Pro každé dva jevy A, B ze vztahu $A \subset B$ plyne $P(X \in A) \leq P(X \in B)$.
7) Pro každé dva jevy A, B platí

$$P(X \in A \cup B) = P(X \in A) + P(X \in B) - P(X \in A \cap B).$$

8) Pro každé dva jevy A, B, $A \subset B$ platí

$$P(X \in B - A) = P(X \in B) - P(X \in A).$$

Poznámka 3.2.

1) Je-li např. $A = (a, b\rangle$, potom místo $P(X \in A)$ píšeme častěji $P(a < X \leq b)$. Je-li $a = -\infty$, píšeme poslední vztah ve tvaru $P(X \leq b)$. Podobné označení použijeme i v jiných alternativách. Uvedené označení využijeme zejména při zavedení a aplikacích pojmu distribuční funkce náhodné veličiny.

2) Nechť $X \sim g$ a $c \in \mathbb{R}$, označme $A = \{c\}$.

Je-li X spojitá náhodná veličina, potom ze vztahu (3.3) plyne

$$P(X = c) = \int_A g(x)\,dx = \int_c^c g(x)\,dx = 0.$$

Je-li X diskrétní náhodná veličina, potom ze vztahu (3.1) dostáváme

$$P(X = c) = \sum_{x \in A \cap \Omega} g(x) = \begin{cases} \sum\limits_{x \in A} g(x) = g(c) & \text{pro } \; c \in \Omega \\ \sum\limits_{x \in \emptyset} g(x) = 0 & \text{pro } \; c \notin \Omega \end{cases}.$$

Poněvadž je podle definice 2.1 $g(c) = 0$ právě tehdy, když $c \in \mathbb{R} - \Omega$, dostaneme v posledním případě $P(X = c) = g(c)$.

Platí tedy

$$P(X = c) = \begin{cases} g(c) & \text{v případě diskrétní náhodné veličiny } X \\ 0 & \text{v případě spojité náhodné veličiny } X \end{cases}.$$

Tedy pravděpodobnostní funkce $g(x)$ náhodné veličiny X udává pravděpodobnost, že náhodná veličina X nabude hodnoty x. Pravděpodobnost, že spojitá náhodná veličina nabude konstantní hodnoty, je nula.

3) Z výše uvedeného vyplývá: Je-li X spojitá [diskrétní] náhodná veličina, pak pravděpodobnost, že nabude hodnoty z množiny A nezávisí [obecně závisí] na tom, zda A obsahuje nebo neobsahuje hranici.

4) Využití věty 3.1 mnohdy může zjednodušit výpočty pravděpodobnosti.

5) Vlastností 1, 2 a 3 z věty 3.1 se často užívá jako definičních vlastností pravděpodobnosti i pro obecnější případy, než kterými se zabýváme v těchto skriptech.

V následujících několika příkladech uvedeme jednoduché ukázky aplikací předchozí teorie.

Příklad 3.1. Náhodná veličina X má rozdělovací funkci g

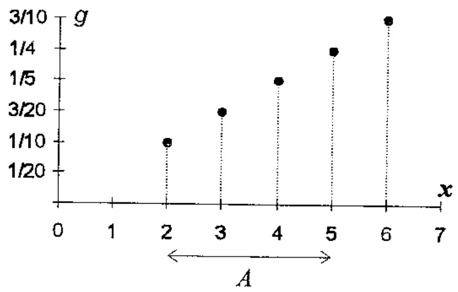
a) $g(x) = \begin{cases} \dfrac{x}{20} & \text{pro } \; x = 2, 3, 4, 5, 6 \\ 0 & \text{jinak} \end{cases},$ diskrétní

b) $g(x) = \begin{cases} \dfrac{x}{18} & \text{pro } \; x \in (0, 6\rangle \\ 0 & \text{jinak} \end{cases}.$ spojitá

Určete následující pravděpodobnosti $P(X \in \langle 2, 5\rangle)$, $P(X \in (2, 5\rangle)$, $P(X \in (2, 5))$, $P(X \in \langle 2, 5))$, $P(X \leq 3)$, $P(X < 3)$, $P(X = 5)$, $P(X = 3.5)$, $P(X > 3)$, $P(X \geq 3)$, $P(X \neq 5)$, $P(X \leq 3 \vee X \geq 5)$.

Řešení. V obou případech se jedná o výpočet pravděpodobnosti, když známe rozdělovací funkci. Musíme tedy určit, zda se jedná o diskrétní nebou spojitou náhodnou veličinu a podle toho použít vztah (3.1) nebo (3.2). Dále samozřejmě můžeme pro usnadnění výpočtů použít větu 3.1.

a) Poněvadž X je diskrétní veličina s oborem hodnot $\Omega = \{2,3,4,5,6\}$, dostáváme ze vztahu (3.1) (viz obrázek 3.2)



Obrázek 3.2: Rozdělovací funkce g z příkladu 3.1 a)

$$
\begin{aligned}
P(X \in \langle 2,5 \rangle) &= \Big| A = \{x \in \mathbb{R};\ 2 \le x \le 5\} \Big| = P(X \in A) \\
&= \sum_{x \in A\, \cap \Omega} g(x) = \Big| A \cap \Omega = \{2,3,4,5\} \Big| \\
&= g(2) + g(3) + g(4) + g(5) = 0.7.
\end{aligned}
$$

Podobně

$$
\begin{aligned}
&P(X \in (2,5\rangle) = g(3) + g(4) + g(5) = 0.6, \\
&P(X \in (2,5)) = g(3) + g(4) = 0.35, \\
&P(X \in \langle 2,5)) = g(2) + g(3) + g(4) = 0.45, \\
&P(X \le 3) = g(2) + g(3) = 0.25, \\
&P(X < 3) = g(2) = 0.1, \\
&P(X = 5) = g(5) = 0.25, \\
&P(X = 3.5) = g(3.5) = 0.
\end{aligned}
$$

Ze vztahu (3.1) nebo využitím věty 3.1 5) dostaneme

$$
P(X > 3) = \left\{ \begin{array}{l} g(4) + g(5) + g(6) \\ 1 - P(X \le 3) \end{array} \right\} = 0.75,
$$

$$
P(X \ge 3) = \left\{ \begin{array}{l} g(3) + g(4) + g(5) + g(6) \\ 1 - P(X < 3) \end{array} \right\} = 0.9,
$$

$$
P(X \ne 5) = \left\{ \begin{array}{l} g(2) + g(3) + g(4) + g(6) \\ 1 - P(X = 5) \end{array} \right\} = 0.75.
$$

Využitím věty 3.1 3) nebo 5) dostaneme

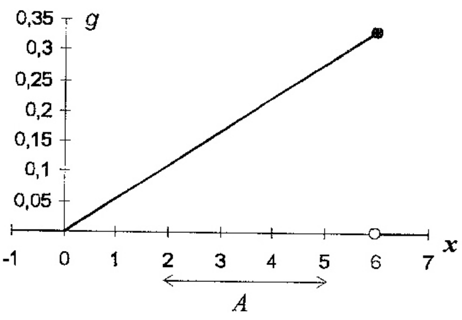
$$
P(X \le 3 \vee X \ge 5) = \left\{ \begin{array}{l} P(X \le 3) + P(X \ge 5) \\ 1 - P(X = 4) \end{array} \right\} = 0.8.
$$

PDF-XChange Click to buy NOW! www.docu-track.com

b) Poněvadž X je spojitá náhodná veličina s oborem hodnot $\Omega = (0, 6\rangle$, dostáváme ze vztahu (3.2) (viz obrázek 3.3)

$$P(X \in \langle 2,5\rangle) = \Big| A = \{x \in \mathbb{R};\ 2 \leq x \leq 5\} \Big| = P(X \in A)$$

$$= \int\limits_{A \cap \Omega} g(x)\,dx = \Big| A \cap \Omega = A \Big| = \int\limits_2^5 \frac{x}{18}\,dx = \frac{1}{36}\Big[x^2\Big]_2^5 = 0.58\overline{3}.$$



Obrázek 3.3: Rozdělovací funkce g z příkladu 3.1 b)

Podobně s využitím věty 3.1 3) 5) a poznámky 3.2 2) je

$$P(X \in (2,5\rangle) = P(X \in (2,5)) = P(X \in \langle 2,5)) = P(X \in \langle 2,5\rangle) = 0.58\overline{3},$$

$$P(X \leq 3) = \int\limits_0^3 \frac{x}{18}\,dx = 0.25,$$

$$P(X < 3) = P(X \leq 3) = 0.25,$$

$$P(X = 5) = 0,$$

$$P(X = 3.5) = 0,$$

$$P(X > 3) = \left\{ \begin{array}{l} 1 - P(X \leq 3) \\ \displaystyle\int\limits_3^6 \frac{x}{18}\,dx \end{array} \right\} = 0.75,$$

$$P(X \geq 3) = \left\{ \begin{array}{l} P(X > 3) \\ 1 - P(X < 3) \end{array} \right\} = 0.75,$$

$$P(X \neq 5) = \left\{ \begin{array}{l} 1 - P(X = 5) \\ \displaystyle\int\limits_0^5 \frac{x}{18}\,dx + \int\limits_5^6 \frac{x}{18}\,dx \end{array} \right\} = 1,$$

$$P(X \le 3 \vee X \ge 5) = \left\{ \begin{array}{l} P(X \le 3) + P(X \ge 5) = 0.25 + \int\limits_5^6 \frac{x}{18}\, dx \\ \\ 1 - P(X \in (3,5)) = 1 - \int\limits_3^5 \frac{x}{18}\, dx \end{array} \right\} = 0.\bar{5}.$$

♣

Příklad 3.2. Je známo, že náhodná chyba měření má normální rozdělení s parametry $\mu = 1[j], \sigma^2 = 0.01[j^2]$ (viz definice 9.2.1). Vypočtěte, s jakou pravděpodobností bude absolutní hodnota náhodné chyby měření zmenšené o systematickou chybu (tj. $\mu = 1[j]$) maximálně 0.2 [j].

Řešení. Náhodná veličina X je spojitá a pro její hustotu f platí (viz definice 9.2.1)

$$f(x) = \frac{1}{\sqrt{2\pi} \cdot 0.1} e^{-\frac{(x-1)^2}{2 \cdot 0.01}} \qquad \text{pro } x \in \mathbb{R}.$$

Potom

$$P\big(|X-1| \le 0.2\big) = P(0.8 \le X \le 1.2) = 2P(1 \le X \le 1.2)$$
$$= 2 \int\limits_1^{1.2} \frac{1}{\sqrt{2\pi} \cdot 0.1} e^{-\frac{(x-1)^2}{2 \cdot 0.01}}\, dx = 2I.$$

Pro výpočet integrálu I zavedeme substituci

$$t = \frac{x-1}{0.1} \Rightarrow 0.1\, dt = dx, \ 0 < t < 2.$$

Odtud

$$I = \frac{1}{\sqrt{2\pi}} \cdot \int\limits_0^2 e^{-\frac{t^2}{2}}\, dt.$$

Pomocí Taylorova rozvoje integrované funkce a na základě vlastností mocniných řad dostáváme

$$I = \frac{1}{\sqrt{2\pi}} \cdot \int\limits_0^2 \sum_{k=0}^{\infty} (-1)^k \cdot \frac{t^{2k}}{2^k k!}\, dt$$
$$= \frac{1}{\sqrt{2\pi}} \cdot \sum_{k=0}^{\infty} \left[(-1)^k \cdot \frac{t^{2k+1}}{2^k k!(2k+1)} \right]_0^2$$
$$= \frac{1}{\sqrt{2\pi}} \cdot \sum_{k=0}^{\infty} (-1)^k \cdot \frac{2^{k+1}}{k!(2k+1)} = 0.47724.$$

Hledaná pravděpodobnost je přibližně 0.95448 (s přesností na 5 desetinných míst).

♣

Příklad 3.3. Hrací kostka je konstruovaná tak, ža jednička padá třikrát častěji než šestka a šestka padá stejně často jako dvojka, trojka, čtyřka a pětka. Určete pravděpodobnost, že padne liché číslo.

Řešení. Nechť X je počet ok při jednom hodu hrací kostkou, potom obor hodnot Ω náhodné veličiny X je množina $\{1,2,3,4,5,6\}$, tedy X je diskrétní náhodná veličina. Když určíme pravděpodobnostní funkci p této náhodné veličiny, bude

$$P\big(X \in \{1,3,5\}\big) = p(1) + p(3) + p(5).$$

Pravděpodobnostní funkci p stačí určit pro $x \in \Omega$, neboť všude jinde je nulová. Víme, že $P(X = x) = p(x)$ a ze zadání plyne, že

$$\begin{aligned} p(1) &= 3\,p(6), \\ p(6) &= p(2) = p(3) = p(4) = p(5). \end{aligned}$$

Odtud a ze vztahu (2.1) dostáváme $8\,p(6) = 1$, tedy $p(6) = 0.125$. Potom

$$X \sim p(x) = \begin{cases} 0.375 & \text{pro } x = 1 \\ 0.125 & \text{pro } x = 2,3,4,5,6 \\ 0 & \text{jinak} \end{cases}.$$

Hledaná pravděpodobnost je tedy 0.625.

♣

Příklad 3.4. Mezi dvěma místy je potrubí o délce L km. Z této délky je prvních k km vedeno po povrchu a zbývajících $L - k$ km pod zemí. Na potrubí vznikne porucha. Jaká je pravděpodobnost, že k poruše došlo u nadzemní části potrubí?

Řešení. Nechť X je vzdálenost místa poruchy od začátku potrubí. Obor hodnot Ω náhodné veličiny X je interval $\langle 0, L\rangle$ a X je spojitá náhodná veličina. Předpokládejme, že šance, že dojde k poruše potrubí v určitém rozmezí sledovaného potrubí závisí pouze na „délce" tohoto rozmezí a ne na umístění v trase potrubí. Potom pro hustotu f náhodné veličiny X platí

$$X \sim f(x) = \begin{cases} c & \text{pro } x \in \langle 0, L\rangle \\ 0 & \text{jinak} \end{cases}.$$

Vzhledem ke vztahu (2.2) musí platit

$$\int_0^L c\,dx = 1,$$

odtud $c = 1/L$ a tedy

$$X \sim f(x) = \begin{cases} \dfrac{1}{L} & \text{pro } x \in \langle 0, L\rangle \\ 0 & \text{jinak} \end{cases}.$$

Potom pro hledanou pravděpodobnost podle vztahu (3.2) dostáváme $P\big(X \leq k\big) = k/L$.

Poznamenejme, že ke stejnému výsledku bychom dospěli i v případě, kdyby bylo pod zemí libovolných k km potrubí.

♣

Poznámka 3.3. O náhodné veličině X z příkladu 3.2 říkáme, že má tzv. rovnoměrné rozdělení s parametry 0 a L. Obecně řekneme, že má náhodná veličina rovnoměrné rozdělení s parametry a a b, $a < b$, když má hustotu

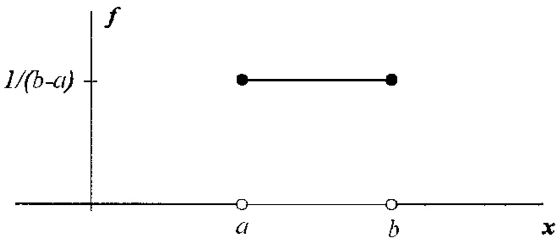
$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{pro } x \in \langle a, b\rangle \\ 0 & \text{jinak} \end{cases} .$$

Značíme $X \sim R(a, b)$.
Rovnoměrné rozdělení má např.:

1) doba čekání na nějakou událost, která se opakuje v pravidelných časových intervalech $\langle a, b\rangle$ - např. doba čekání na trolejbus, který jezdí v pravidelných desetiminutových intervalech má rozdělení $R(0, 10)$.

2) chyby zaokrouhlování - má-li např. stupnice buzoly dílky po 1°, potom má chyba zaokrouhlování rozdělení $R(-0.5, 0.5)$.

Graf hustoty f rozdělení $R(a, b)$ je na obrázku 3.4.



Obrázek 3.4: Hustota rozdělení $R(a, b)$

Dále se budeme zabývat náhodným vektorem $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$. Definice pravděpodobnosti v tomto případě bude zřejmě analogická definici 3.1, stačí si pouze uvědomit, že $\Omega \subset \mathbb{R}^n$ a rozdělovací funkce g je funkcí n proměnných. Místo „sumy“ a „integrálu“ budeme tedy zřejmě pracovat s n-rozměrnou „sumou“ a n-rozměrným „integrálem“. Tato analogie nás provází celou pravděpodobností.

Definice 3.2. *Buď $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ diskrétní [spojitý] náhodný vektor s oborem hodnot Ω a rozdělovací funkcí funkcí g. Buď A libovolná podmnožina $\mathbb{R}^n$ [podmnožina $\mathbb{R}^n$, pro kterou $\int \cdots \int_A dx_1 dx_2 \ldots dx_n$ existuje]. Potom A nazýváme náhodný jev.*

Pravděpodobnost, že náhodný vektor $\boldsymbol{X}$ nabude hodnoty z množiny A značíme

$$P(\boldsymbol{X} \in A)$$

a definujeme

$$P(\boldsymbol{X} \in A) = \sum_{(x_1, x_2, \ldots, x_n) \in A \cap \Omega} g(x_1, x_2, \ldots, x_n) \tag{3.4}$$

$$\left[P(\boldsymbol{X} \in A) = \int_{\Omega \cap A} \cdots \int g(x_1, x_2, \ldots, x_n)\, dx_1 dx_2 \ldots dx_n \right]. \tag{3.5}$$

Vztahem (3.4) [(3.5)] je definováno zobrazení množiny všech náhodných jevů do množiny reálných čísel $\mathbb{R}$. Toto zobrazení nazýváme pravděpodobnost (příslušná náhodnému vektoru $\boldsymbol{X}$).

Poznámka 3.4. Pro náhodný vektor $\boldsymbol{X}$ platí věta analogická větě 3.1, která nám i v tomto případě umožňuje zjednodušení výpočtu pravděpodobnosti. Stačí si jenom uvědomit, že $A, B, A_1, A_2, A_3, \dots$ jsou podmnožiny $\mathbb{R}^n$.

Poznámka 3.5.

1) Nechť P je pravděpodobnost příslušná náhodnému vektoru $(X_1, \dots, X_n)$. Je-li např. $A = (a_1, b_1\rangle \times (a_2, b_2\rangle \times \cdots \times (a_n, b_n\rangle$, potom místo $P\big((X_1, \dots, X_n) \in A\big)$ píšeme častěji

$$P\big(a_1 < X_1 \le b_1, a_2 < X_2 \le b_2, \dots, a_n < X_n \le b_n\big).$$

Speciálně pro $a_1 = a_2 = \cdots = a_n = -\infty$ píšeme poslední vztah ve tvaru

$$P\big(X_1 \le b_1, X_2 \le b_2, \dots, X_n \le b_n\big).$$

2) Má-li náhodný vektor $(X_1, \dots, X_n)$ rozdělovací funkci g a $c = (c_1, \dots, c_n) \in \mathbb{R}^n$, potom podobně jako v poznámce 3.2 2) platí

$$P\big(X_1 = c_1, \dots, X_n = c_n\big) = \begin{cases} g(c_1, \dots, c_n) & \text{v diskrétním případě} \\ 0 & \text{ve spojitém případě} \end{cases}.$$

Příklad 3.5. Náhodný vektor (X, Y) má rozdělovací funkci

$$\text{a) } g(x, y) = \begin{cases} \dfrac{4}{3}\left(\dfrac{1}{3}\right)^{x+y} & \text{pro } (x, y) \in \{0, 1, 2, \dots\} \times \{1, 2, 3, \dots\} \\ 0 & \text{jinak} \end{cases},$$

$$\text{b) } g(x, y) = \begin{cases} \dfrac{1}{3}xy & \text{pro } (x, y) \in (0, 1\rangle \times \langle 2, 4\rangle \\ 0 & \text{jinak} \end{cases}.$$

Vypočtěte $P(X \le 0.5, Y \le 3)$, $P(X \le 0.5, Y < 3)$, $P(X = 1, Y = 3)$, $P(X \le 1, Y = 3)$, $P(X = 1, Y \le 3)$, $P\big((X, Y) \in H\big)$, kde H je vnitřek trojúhelníku ABC s vrcholy $A = (-1, 0)$, $B = (1, 0)$, $C = (1, 4)$.

Řešení. V obou případech se jedná o výpočet pravděpodobnosti, když známe rozdělovací funkci. Musíme tedy určit, zda se jedná o diskrétní nebo spojitý náhodný vektor a podle toho použít vztah (3.4) nebo (3.5).

a) Obor hodnot Ω náhodného vektoru (X, Y) je množina

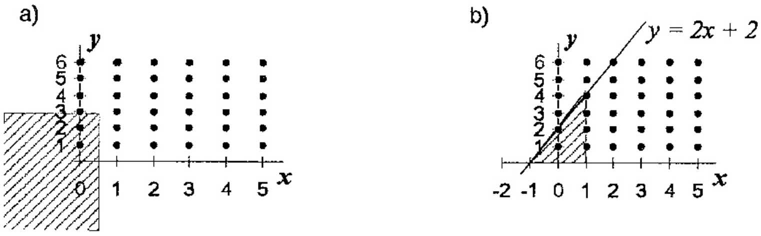
$$\{0, 1, 2, \dots\} \times \{1, 2, 3, \dots\},$$

jedná se tedy o diskrétní náhodný vektor. Potom podle vztahu (3.4) dostaneme

$$\begin{aligned} P(X \le 0.5, Y \le 3) &= \left| A = \{(x, y) \in \mathbb{R}^2;\ x \le 0.5, y \le 3\} \right| \\ &= P\big((X, Y) \in A\big) = \sum_{(x,y) \in A \cap \Omega} g(x, y). \end{aligned}$$

Zřejmě (viz obrázek 3.5 a))

$$A \cap \Omega = \{(0,1), (0,2), (0,3)\}.$$



Obrázek 3.5

Odtud

$$P(X \leq 0.5, Y \leq 3) = g(0,1)+g(0,2)+g(0,3) = \frac{4}{3}\left[\frac{1}{3}+\left(\frac{1}{3}\right)^2+\left(\frac{1}{3}\right)^3\right] \doteq 0.642.$$

Podobně

$$P(X \leq 0.5, Y < 3) = g(0,1) + g(0,2) = \frac{4}{3}\left[\frac{1}{3} + \left(\frac{1}{3}\right)^2\right] \doteq 0.593,$$

$$P(X = 1, Y = 3) = g(1,3) = \frac{4}{3}\left(\frac{1}{3}\right)^4 \doteq 0.016,$$

$$P(X \leq 1, Y = 3) = g(0,3) + g(1,3) \doteq 0.066,$$

$$P(X = 1, Y \leq 3) = g(1,1) + g(1,2) + g(1,3) \doteq 0.214,$$

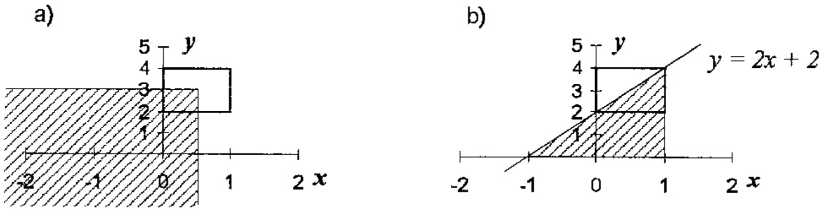
$P((X,Y) \in H) = g(0,1) \doteq 0.444$ (viz obrázek 3.5 b))

b) Obor hodnot Ω náhodného vektoru (X,Y) je množina $\langle 0,1\rangle \times \langle 2,4\rangle$, jedná se tedy o spojitý náhodný vektor. Podle vztahu (3.5) dostaneme

$$P(X \leq 0.5, Y \leq 3) = \left| A = \{(x,y) \in \mathbb{R}^2;\ x \leq 0.5,\ y \leq 3\} \right|$$
$$= P((X,Y) \in A) = \iint_{A\cap\Omega} g(x,y)\,dx dy.$$

Zřejmě (viz obrázek 3.6 a))

$$A \cap \Omega = \{(x,y) \in \mathbb{R}^2;\ 0 \leq x \leq 0.5,\ 2 \leq y \leq 3\}.$$



Obrázek 3.6

Odtud

$$P(X \leq 0.5, Y \leq 3) = \int_0^{0.5} \Big[\int_2^3 \frac{1}{3} xy \, dy \Big] \, dx = \frac{1}{3} \int_0^{0.5} x \Big[\frac{y^2}{2} \Big]_0^3 \, dx$$

$$= \frac{5}{6} \int_0^{0.5} x \, dx \doteq 0.104.$$

Podobně

$$P(X \leq 0.5, Y < 3) = P(X \leq 0.5, Y \leq 3) \doteq 0.104,$$

$$P(X = 1, Y = 3) = \int_1^1 \Big[\int_3^3 \frac{1}{3} xy \, dy \Big] \, dx = 0,$$

$$P(X \leq 1, Y = 3) = \int_0^1 \Big[\int_3^3 \frac{1}{3} xy \, dy \Big] \, dx = 0,$$

$$P(X = 1, Y \leq 3) = \int_1^1 \Big[\int_2^3 \frac{1}{3} xy \, dy \Big] \, dx = 0.$$

Dále

$$P((X,Y) \in H) = \int_{H \cap \Omega} g(x,y) \, dx dy.$$

Zřejmě (viz obrázek 3.6 b))

$$H \cap \Omega = \{(x,y) \in \mathbb{R}^2;\ 0 \leq x \leq 1,\ 2 \leq y \leq 2x + 2\},$$

potom

$$P((X,Y) \in H) = \int_0^1 \Big[\int_2^{2x+2} \frac{1}{3} xy \, dy \Big] \, dx = \frac{1}{3} \int_0^1 x \Big[\frac{y^2}{2} \Big]_2^{2x+2} \, dx$$

$$= \frac{4}{6} \int_0^1 (x^3 + 2x^2) \, dx \doteq 0.611.$$

♣

Příklad 3.6. Dvojice součástek má dobu života popsanou hustotou

$$f(x,y) = \begin{cases} \frac{1}{4} \cdot e^{\frac{-x-y}{2}} & \text{pro } x > 0,\ y > 0 \\ 0 & \text{jinak} \end{cases}.$$

PDF-XChange Click to buy NOW! www.docu-track.com

Vypočtěte pravděpodobnost, že

a) obě součástky budou mít dobu života maximálně jednu časovou jednotku,

b) druhá součástka má dobu života maximálně jednu časovou jednotku,

c) druhá součástka přežije první.

Řešení. Nechť X je doba života první součástky a Y je doba života druhé součástky. Obor hodnot náhodného vektoru (X, Y) je množina $(0, \infty) \times (0, \infty)$, jedná se tedy o spojitý náhodný vektor Při výpočtu pravděpodobnosti použijeme vztah (3.5).

$$\text{a)}\quad P(X \le 1, Y \le 1) = \Big| A = \{(x,y) \in \mathbb{R}^2;\ x \le 1, y \le 1\} \Big|$$
$$= P\big((X,Y) \in A\big) = \iint\limits_{A \cap \Omega} f(x,y)\, dx dy.$$

Zřejmě
$$A \cap \Omega = \{(x,y) \in \mathbb{R}^2;\ 0 < x \le 1,\ 0 < y \le 1\}$$
a tedy
$$P(X \le 1, Y \le 1) = \int_0^1 \Big[\int_0^1 \frac{1}{4} \cdot \mathrm{e}^{\frac{-x-y}{2}}\, dy \Big]\, dx.$$

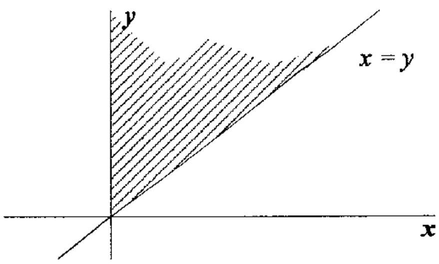
Výpočtem integrálu (po zavedení vhodné substituce) dostaneme, že hledaná pravděpodobnost je přibližně *0.155*.

b) Máme spočítat pravděpodobnost $P(Y \le 1)$. Platí
$$P(Y \le 1) = P(X \le \infty, Y \le 1) = \int_0^\infty \Big[\int_0^1 \frac{1}{4} \cdot \mathrm{e}^{\frac{-x-y}{2}}\, dy \Big]\, dx.$$

Výpočtem integrálu dostaneme, že hledaná pravděpodobnost je přibližně 0.393.

$$\text{c)}\quad P(X < Y) = P(Y - X > 0) = \Big| C = \{(x,y) \in \mathbb{R}^2;\ y > x\} \Big|$$
$$= P\big((X,Y) \in C\big) = \iint\limits_{C \cap \Omega} f(x,y)\, dx dy.$$

Zřejmě (viz obrázek 3.7)
$$C \cap \Omega = \{(x,y) \in \mathbb{R}^2;\ x < y < \infty,\ 0 < x < \infty\}.$$



Obrázek 3.7

Potom
$$P(Y > X) = \int_0^{\infty} \Big[\int_x^{\infty} \frac{1}{4} \cdot e^{\frac{-x-y}{2}} \, dy \Big] \, dx.$$
Výpočtem integrálu dostaneme, že hledaná pravděpodobnost je 0.5.

♣

Příklad 3.7.

a) Ke zkoušce bylo vydáno 50 otázek, student tahá jednu z nich. Ví, že se na výbornou naučil odpověď na 30 otázek, na dvojku 12, na trojku 3 a o 5-ti otázkách neví nic. Zkráceně řečeno, jsou jeho znalosti (30 - 12 - 3 - 5). Nechť T je číslo tažené otázky. Obor hodnot Ω_1 náhodné veličiny T je množina $\{1, 2, \ldots, 50\}$ a T je tedy diskrétní náhodná veličina s pravděpodobnostní funkcí
$$g(x) = \begin{cases} \frac{1}{50} & \text{pro } x = 1, 2, \ldots, 50 \\ 0 & \text{jinak} \end{cases}.$$
Jedná se o tzv. <u>klasické rozdělení s parametrem 50</u>. Nechť jsou otázky číslovány vzestupně s klesajícími studentovými znalostmi. Potom pravděpodobnost, že student udělá zkoušku je
$$P(T \leq 45) = g(1) + \cdots + g(45) = \frac{45}{50} = 0.9.$$

b) Nechť je zkouška organizována tak, že 50 otázek je rozděleno na 20 z teorie pravděpodobnosti a 30 z matematické statistiky. Nechť studentovi znalosti z pravděpodobnosti jsou (10 - 5 - 2 - 3) a ze statistiky (20 - 7 - 1 - 2). K úspěšnému složení zkoušky je nutné získat alespoň „za tři“ z obou částí. Jaká bude nyní pravděpodobnost, že student zkoušku složí?

Nechť jsou otázky v obou částech číslovány od jedné. Potom čísla tažené dvojice otázek tvoří náhodný vektor, označme jej (X, Y). Obor hodnot náhodného vektoru (X, Y) je množina $\Omega_2 = \{1, 2, \ldots, 20\} \times \{1, 2, \ldots, 30\}$, která obsahuje $20 \cdot 30 = 600$ bodů. Pravděpodobnostní funkce p náhodného vektoru (X, Y) je tedy
$$p(x, y) = \begin{cases} \frac{1}{600} & \text{pro } (x, y) \in \Omega_2 \\ 0 & \text{jinak} \end{cases}.$$
Lehce nahlédneme, že těch dvojic otázek, které zajišťují úspěšný průběh zkoušky je $17 \cdot 28 = 476$. Pravděpodobnost složení zkoušky je nyní rovna
$$\frac{476}{600} \doteq 0.793.$$

Označme nyní V_a a V_b výsledek zkoušky studenta v situacích a) a b). V_a a V_b jsou diskrétní náhodné veličiny, jejich pravděpodobnostní funkce p_a a p_b jsou uvedeny v následující tabulce:

v	1	2	3	4
$p_a(v)$	$\frac{30}{50}$	$\frac{12}{50}$	$\frac{3}{50}$	$\frac{5}{50}$
$p_b(v)$	$\frac{200}{600}$	$\frac{255}{600}$	$\frac{21}{600}$	$\frac{124}{600}$

Tabulka 3.1

Určení funkce p_a je zřejmé; v případě b) při nestejných výsledcích obou částí (pokud žádný z nich není nevyhovující) dává zkoušející známku průměrnou a případně zaokrouhluje na horší. Tedy například

$$P(V_b = 3) = P\big((X,Y) \in \{(2,3),(3,2),(3,3)\}\big) = \frac{21}{600}.$$

♣

K uvedení posledního příkladu nás vedly zejména dva důvody. První z nich je snaha zařadit do předkládaného způsobu zavedení pojmu pravděpodobnosti středoškolské znalosti typu „pravděpodobnost je podíl počtu příznivých případů ku počtu případů možných". Druhý důvod vyplyne lépe ze zápisu obou pravděpodobnostních funkcí ve formě tabulky 3.2.

v	1	2	3	4
$p_a(v)$	0.60	0.24	0.06	0.10
$p_b(v)$	0.33	0.42	0.04	0.21

Tabulka 3.2

Uvědomme si, že obě dvě pravděpodobnostní funkce vyjadřují šanci téhož studenta na získání jednotlivých známek v závislosti na tom, jakou strategii zkoušející zvolí. A zde je už onen druhý důvod.

Pro zdárný výsledek jakéhokoliv experimentu může být rozhodující strategie jeho provádění. Touto problematikou se zabývá celá matematická disciplina zvaná plánování experimentu.

4. Distribuční funkce

Definice 4.1. *Buď P pravděpodobnost příslušná náhodné veličině X. Buď F reálná funkce jedné reálné proměnné definovaná vztahem*

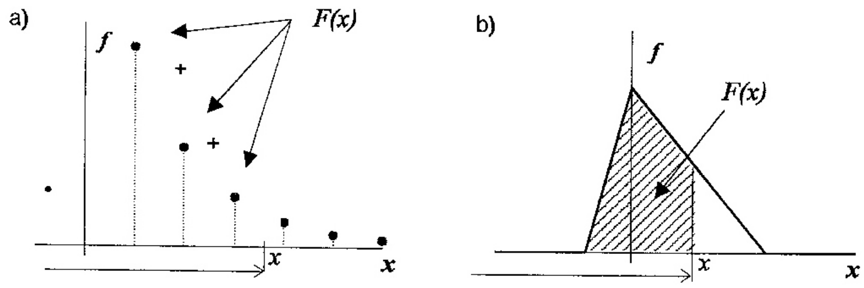
$$F(x) = P(X \le x). \tag{4.1}$$

Funkci F nazýváme distribuční funkce náhodné veličiny X. Značíme $X \sim F$.

Poznámka 4.1.

1) Distribuční funkci, na rozdíl od rozdělovací funkce, značíme vždy velkým písmenem, rozdělovací funkci pak stejným malým písmenem.

2) Nechť X je diskrétní [spojitá] náhodná veličina s rozdělovací funkcí f. Výpočet distribuční funkce F je graficky znázorněn na obrázku 4.1 a) [b)] .



Obrázek 4.1: Grafické znázornění výpočtu distribuční funkce z rozdělovací funkce

Příklad 4.1. Náhodná veličina X má rozdělovací funkci

$$\text{a) } g(x) = \begin{cases} \dfrac{x^2}{10} & \text{pro } x = -2, -1, 1, 2 \\ 0 & \text{jinak} \end{cases},$$

$$\text{b) } g(x) = \begin{cases} \dfrac{1}{6}\left(x^2 + 1\right) & \text{pro } x \in \langle -1, 2 \rangle \\ 0 & \text{jinak} \end{cases}.$$

Určete distribuční funkci G náhodné veličiny X.

Řešení. V obou případech vyjdeme ze vztahu (4.1). Protože známe rozdělovací funkci g náhodné veličiny X, jedná se o výpočet pravděpodobnosti na základě vztahu (3.1), resp. (3.3) podle toho, zda je X diskrétní nebo spojitá náhodná veličina.

a) Obor hodnot Ω náhodné veličiny X je množina $\{-2,-1,1,2\}$ a tedy X je diskrétní náhodná veličina (graf její pravděpodobnostní funkce je na obrázku 4.2 a)). Pro $x \in \mathbb{R}$ zřejmě platí

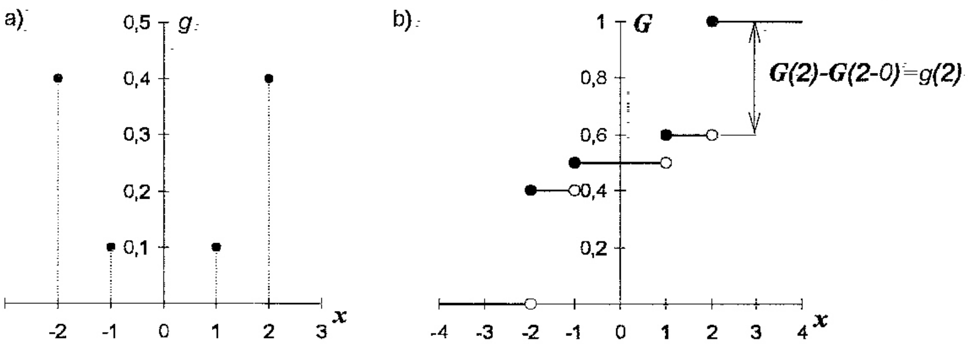
$$G(x) = P(X \leq x) = \sum_{t \leq x,\, t \in \Omega} g(t).$$

Potom např.

$$G(1.0) = P(X \leq 1.0) = \sum_{t \leq 1.0,\, t \in \Omega} g(t) = g(-2) + g(-1) + g(1) = 0.6,$$

$$G(1.7) = P(X \leq 1.7) = \sum_{t \leq 1.7,\, t \in \Omega} g(t) = g(-2) + g(-1) + g(1) = 0.6,$$

$$G(2.0) = P(X \leq 2.0) = \sum_{t \leq 2.0,\, t \in \Omega} g(t) = g(-2) + g(-1) + g(1) + g(2) = 1.0.$$



Obrázek 4.2: a) Pravděpodobnostní funkce g náhodné veličiny X z příkladu 4.1 a)
b) Distribuční funkce G náhodné veličiny X z příkladu 4.1 a)

Je zřejmé, že G bude schodovitá funkce se skoky v bodech $x \in \Omega$. Výška skoku v bodě x bude rovna $g(x) = P(X = x)$. Spočítejme tedy nejprve hodnoty G pro $x \in \Omega$. Dostaneme

$$G(-2) = P(X \leq -2) = 0.4,$$

$$G(-1) = P(X \leq -1) = 0.5,$$

$$G(1) = P(X \leq 1) = 0.6,$$

$$G(2) = P(X \leq 2) = 1.0.$$

Potom určíme hodnoty G pro zbývající $x \in \mathbb{R}$.

Pro $x < -2$ dostaneme

$$G(x) = 0,$$

podobně pro $x \in (-2,-1)$ je

$G(x) = g(-2) = 0.4.$

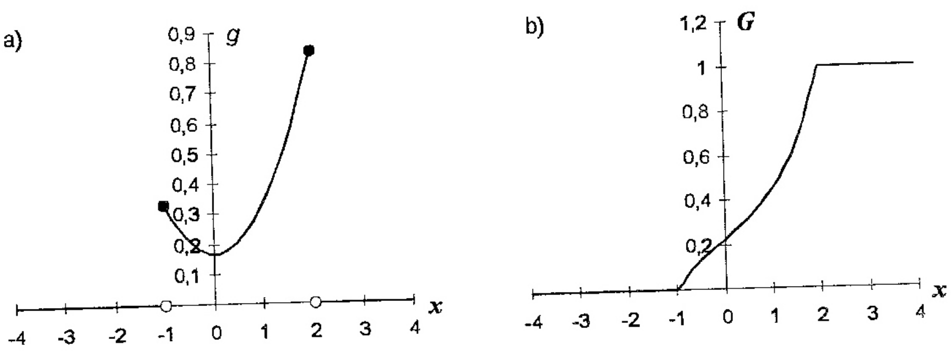
Analogicky vyjádříme $G(x)$ postupně pro $x \in (-1,1)$, $x \in (1,2)$, $x \in (2,\infty)$. Po výpočtu dostaneme

$$G(x) = \begin{cases} 0 & \text{pro } x \in (-\infty,-2) \\ 0.4 & \text{pro } x \in \langle -2,-1) \\ 0.5 & \text{pro } x \in \langle -1,1) \\ 0.6 & \text{pro } x \in \langle 1,2) \\ 1 & \text{pro } x \in \langle 2,\infty) \end{cases} .$$

Graf distribuční funkce G je na obrázku 4.2 b).

b) Obor hodnot Ω náhodné veličiny X je množina $\langle -1,2\rangle$, tedy X je spojitá náhodná veličina (graf její hustoty je na obrázku 4.3 a)). Potom

$$G(x) = P(X \le x) = \int_{-\infty}^{x} g(t)\,dt.$$



Obrázek 4.3: a) Hustota náhodné veličiny X z příkladu 4.1 b)
b) Distribuční funkce náhodné veličiny X z příkladu 4.1 b)

Stejně jako v a) určíme $G(x)$ nejprve pro $x \in \Omega$. Z posledního vztahu dostaneme

$$G(x) = \int_{-\infty}^{x} g(t)\,dt = \int_{-\infty}^{-1} 0\,dt + \int_{-1}^{x} \frac{1}{6}(t^2+1)\,dt$$
$$= 0 + \frac{1}{6}\left[\frac{t^3}{3} + t\right]_{-1}^{x} = \frac{1}{6}\left(\frac{x^3}{3} + x + \frac{4}{3}\right).$$

Dále určíme $G(x)$ pro zbývající $x \in \mathbb{R}$.

Pro $x < -1$ je $G(x) = \int_{-\infty}^{x} g(t)\,dt = \int_{-\infty}^{x} 0\,dt = 0.$

Pro $x > 2$ je $G(x) = \int\limits_{-\infty}^{x} g(t)\,dt = \int\limits_{-\infty}^{-1} 0dt + \int\limits_{-1}^{2} \frac{1}{6}(t^2+1)\,dt + \int\limits_{2}^{x} 0\,dt = 1.$

Tedy

$$G(x) = \begin{cases} 0 & \text{pro } -\infty < x < -1 \\ \frac{1}{6}\left(\frac{x^3}{3} + x + \frac{4}{3}\right) & \text{pro } -1 \leq x \leq 2 \\ 1 & \text{pro } 2 < x < \infty \end{cases}.$$

Graf této funkce je na obrázku 4.3 b).

Poznámka 4.2. Všimněme si, že:

1) V případě diskrétní náhodné veličiny X je distribuční funkce G schodovitá se skoky v bodech $x \in \Omega$. V případě spojité [diskrétní] náhodné veličiny je distribuční funkce spojitá [zprava spojitá]. Tedy z průběhu distribuční funkce poznáme, zda se jedná o spojitou nebo diskrétní náhodnou veličinu.

2) Distribuční funkce náhodné veličiny X je nekonstantní na oboru hodnot této náhodné veličiny.

3) Hodnota distribuční funkce je v největší dolní [nejmenší horní] hranici Ω nula [jedna] .

4) Když známe distribuční funkci G náhodné veličiny X, lze určit rozdělovací funkci g. Z obrázku 4.2 vidíme, že v diskrétním případě je $g(x) = G(x) - G(x-0)$. Ze vztahu $G(x) = \int_{-\infty}^{x} g(t)\,dt$ pak ve spojitém případě dostaneme $g(x) = \frac{d}{dx}G(x)$.

5) Podle věty 3.1 8) platí $P(X < x) = P(X \leq x) - P(X = x)$. Potom

$$P(X < x) = \begin{cases} G(x) - [G(x) - G(x-0)] = G(x-0) & \text{v diskrétním případě} \\ G(x) & \text{ve spojitém případě} \end{cases}.$$

6) Podle věty 3.1 6) a 1) je distribuční funkce neklesající a pro každé $x \in \mathbb{R}$ platí

$$0 \leq G(x) \leq 1.$$

7) Pro distribuční funkci G náhodné veličiny X platí

$$G(-\infty) = P(X \leq -\infty) = 0,$$
$$G(\infty) = P(X \leq \infty) = 1.$$

Ukázali jsme si určité vlastnosti distribuční funkce a jak vypočítat distribuční funkci, když známe rozdělovací funkci, resp. jak vypočítat rozdělovací funkci, když známe distribuční funkci. V následující větě tyto vlastnosti a vztahy uvedeme. Důkaz věty neprovádíme, vyplývá z definice 4.1 a věty 3.1, podobně jako naše úvahy výše.

Věta 4.1. Buď X náhodná veličina s distribuční funkcí G. Potom:

a) G je neklesající funkce.

b) G je zprava spojitá.

c) $G(-\infty) = 0$.

PDF-XChange
Click to buy NOW!
www.docu-track.com

d) $G(\infty) = 1$.

e) Je-li X diskrétní [spojitá] s rozdělovací funkcí g a oborem hodnot Ω, potom:

$$1)\ G(x) = \sum_{t \le x, t \in \Omega} g(t) \quad \left[G(x) = \int_{-\infty}^{x} g(t)\,dt\right],$$

$$2)\ g(x) = G(x) - G(x-0) \quad \left[g(x) = \frac{d}{dx}G(x)\right],$$

$$3)\ P(X < x) = G(x-0) \quad \left[P(X < x) = G(x)\right],$$

$$4)\ P(X = x) = g(x) \quad \left[P(X = x) = 0\right].$$

Příklad 4.2. Náhodná veličina má distribuční funkci

$$\text{a) } G(x) = \begin{cases} 0 & \text{pro } x \in (-\infty, 1) \\ 0.1 & \text{pro } x \in \langle 1, 2) \\ 0.4 & \text{pro } x \in \langle 2, 3) \\ 0.8 & \text{pro } x \in \langle 3, 4) \\ 1 & \text{pro } x \in \langle 4, \infty) \end{cases},$$

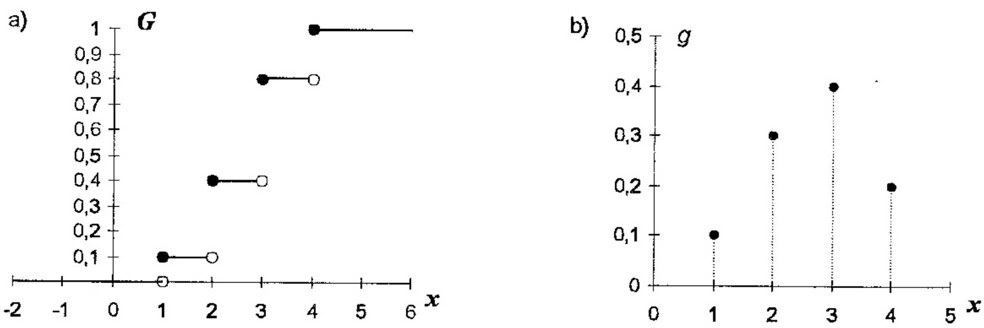
$$\text{b) } G(x) = \begin{cases} 0 & \text{pro } x \le 0 \\ 1 - \mathrm{e}^{-\frac{x}{5}} & \text{pro } x > 0 \end{cases}.$$

Určete:

1) Následující pravděpodobnosti $P(X \le 2)$, $P(X \le 2.7)$, $P(X < 2)$, $P(X < 2.7)$, $P(X = 3)$, $P(X = 3.5)$, $P(X > 2)$, $P(X \ge 2)$, $P(X \in (1,4\rangle)$, $P(X \in (1,4))$, $P(X \in \langle 1,4\rangle)$, $P(X \in \langle 1,4))$, $P(X \in \langle -1,3))$.

2) Rozdělovací funkci g náhodné veličiny X.

Řešení. a) 1) Obor hodnot Ω náhodné veličiny X je množina $\{1, 2, 3, 4\}$ (viz obrázek 4.4 a)), X je tedy diskrétní náhodná veličina. Máme počítat $P(X \in A)$, když známe distribuční funkci G náhodné veličiny X. Vyjdeme ze vztahu (4.1) a věty 4.1 e).



Obrázek 4.4: a) Distribuční funkce náhodné veličiny X z příkladu 4.2 a)
b) Pravděpodobnostní funkce náhodné veličiny X z příkladu 4.2 a)

Potom dostaneme

$$
\begin{aligned}
&P(X \leq 2) = G(2) = 0.4,\\
&P(X \leq 2.7) = G(2.7) = 0.4,\\
&P(X < 2) = G(2-0) = 0.1,\\
&P(X < 2.7) = G(2.7-0) = 0.4,\\
&P(X = 3) = G(3) - G(3-0) = 0.4,\\
&P(X = 3.5) = G(3.5) - G(3.5-0) = G(3.5) - G(3.5) = 0.
\end{aligned}
$$

Z věty 3.1 5) dostáváme

$$
\begin{aligned}
&P(X > 2) = 1 - P(X \leq 2) = 1 - G(2) = 0.6,\\
&P(X \geq 2) = 1 - P(X < 2) = 1 - G(2-0) = 0.9.
\end{aligned}
$$

Z věty 3.1 8) plyne

$$
\begin{aligned}
&P(X \in (1,4\rangle) = P(X \leq 4) - P(X \leq 1) = G(4) - G(1) = 0.9,\\
&P(X \in (1,4)) = P(X < 4) - P(X \leq 1) = G(4-0) - G(1) = 0.7,\\
&P(X \in \langle 1,4\rangle) = P(X \leq 4) - P(X < 1) = G(4) - G(1-0) = 1,\\
&P(X \in \langle 1,4)) = P(X < 4) - P(X < 1) = G(4-0) - G(1-0) = 0.8,\\
&P(X \in \langle -1,3)) = P(X < 3) - P(X < -1) = G(3-0) - G(-1-0) = 0.4.
\end{aligned}
$$

2) Pravděpodobnostní funkci g stačí určit pro $x \in \Omega$, protože všude jinde je nulová. Vyjdeme z věty 4.1 e) vztahu

$$g(x) = G(x) - G(x-0),$$

potom dostaneme

$$
g(x) = \begin{cases} 0.1 & \text{pro } x = 1\\ 0.3 & \text{pro } x = 2\\ 0.4 & \text{pro } x = 3\\ 0.2 & \text{pro } x = 4 \end{cases}.
$$

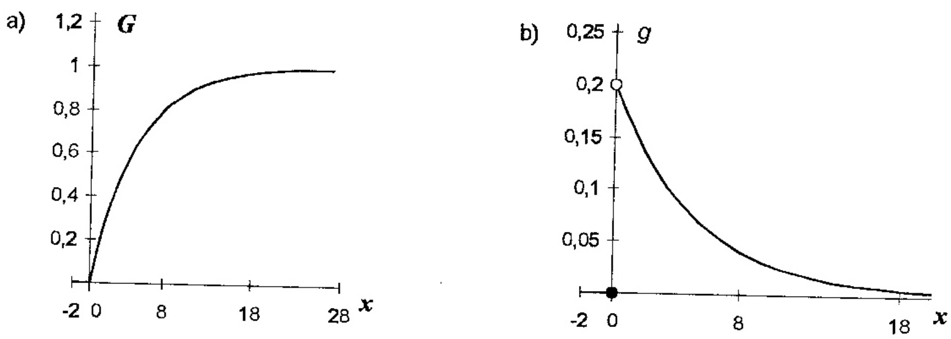
Graf pravděpodobnostní funkce g je na obrázku 4.4 b).

b) Obor hodnot Ω náhodné veličiny X je množina $(0,\infty)$ (viz obrázek 4.5 a)). X je tedy spojitá náhodná veličina. Máme počítat $P(X \in A)$, když známe distribuční funkci G náhodné veličiny X. Vyjdeme ze vztahu (4.1) a věty 4.1 e) a uvědomíme si fakt, který známe již z kapitoly 3, že pravděpodobnost, že náhodná veličina X nabude hodnoty z množiny A nezávisí na tom, zda A obsahuje nebo neobsahuje hranici. Potom dostaneme

$$
\begin{aligned}
&P(X \leq 2) = G(2) = 1 - e^{-\frac{2}{5}} \doteq 0.330,\\
&P(X \leq 2.7) = G(2.7) \doteq 0.417,
\end{aligned}
$$

PDF-XChange Click to buy NOW! www.docu-track.com

$$P(X < 2) = G(2) \doteq 0.330,$$



Obrázek 4.5: a) Distribuční funkce náhodné veličiny X z příkladu 4.2 b)
b) Hustota náhodné veličiny X z příkladu 4.2 b)

$$P(X < 2.7) = G(2.7) \doteq 0.417,$$
$$P(X = 3) = 0,$$
$$P(X = 3.5) = 0.$$

Z věty 3.1 5) dostaneme

$$P(X > 2) = 1 - P(X \leq 2) = 1 - G(2) \doteq 0.670,$$
$$P(X \geq 2) = 1 - P(X < 2) = 1 - G(2) \doteq 0.670.$$

Z věty 3.1 8) plyne

$$P(X \in (1,4\rangle) = P(X \leq 4) - P(X \leq 1) = G(4) - G(1) \doteq 0.369,$$
$$P(X \in (1,4)) = P(X \in \langle 1,4\rangle) = P(X \in \langle 1,4)) \doteq 0.369,$$
$$P(X \in \langle -1,3)) = P(X < 3) - P(X < -1) = G(3) - G(-1) \doteq 0.451.$$

2) Hustotu g stačí určit pro $x \in \Omega$, protože všude jinde je nulová. Vyjdeme z věty 4.1 e) vztahu

$$g(x) = \frac{d}{dx} G(x),$$

potom dostaneme

$$g(x) = \frac{d}{dx}(1 - \mathrm{e}^{-\frac{x}{5}}) = -\frac{d}{dx}(\mathrm{e}^{-\frac{x}{5}}) = \frac{1}{5}\mathrm{e}^{-\frac{x}{5}} \qquad \text{pro } x > 0.$$

Tedy

$$g(x) = \begin{cases} 0 & \text{pro } x \leq 0 \\ \frac{1}{5}\mathrm{e}^{-\frac{x}{5}} & \text{pro } x > 0 \end{cases}.$$

Graf hustoty g je na obrázku 4.5 b).

♣

Poznámka 4.3. Když známe distribuční funkci náhodné veličiny X, je výpočet pravděpodobnosti, že náhodná veličina X nabude hodnoty z nějakého intervalu $A \subset \mathbb{R}$, jednodušší než v případě znalosti rozdělovací funkce. Nemusíme totiž

„integrovat“ nebo „sečítat“ rozdělovací funkci, ale pravděpodobnost dostaneme „odečítáním“ hodnot, resp. limit zleva distribuční funkce. V diskrétním případě tak odpadá pracnost výpočtu, ve spojitém případě pak eventuální komplikace s výpočtem integrálu. V technických aplikacích se totiž převážně pracuje s takovými hustotami, které nelze ani analyticky integrovat (viz příklad 3.2) a v případě, že tyto hustoty lze analyticky integrovat, bývá výpočet velmi komplikovaný. Z tohoto důvodu jsou jednak hodnoty distribuční funkce nejpoužívanějších rozdělení tabelovány (viz [13]), jednak je počítá každý statistický software (např. STATGRAPHICS, ale i tabulkový procesor EXCEL).

Příklad 4.3. Nechť má náhodná veličina X distribuční funkci

$$F(x) = A + B \operatorname{arctg} x \quad \text{pro } x \in \mathbb{R}.$$

Určete

1) konstanty A a B,
2) $P\big(X \in (0,1)\big)$,
3) hustotu $f(x)$ náhodné veličiny X.

Řešení.

1) Podle věty 4.1 c, d) je $F(\infty) = 1$ a $F(-\infty) = 0$, tedy

$$\left.\begin{aligned} A + B\frac{\pi}{2} &= 1 \\ A - B\frac{\pi}{2} &= 0 \end{aligned}\right\} \Rightarrow A = \frac{1}{2},\ B = \frac{1}{\pi} \Rightarrow F(x) = \frac{1}{2} + \frac{1}{\pi}\operatorname{arctg} x \quad \text{pro } x \in \mathbb{R}.$$

2) $P\big(X \in (0,1)\big) = P\big(X < 1\big) - P\big(X \leq 0\big) = F(1) - F(0) = \dfrac{1}{4}.$

3) $f(x) = \dfrac{d}{dx}F(x) = \dfrac{1}{\pi}\dfrac{1}{1+x^2} \quad \text{pro } x \in \mathbb{R}.$

♣

Poznámka 4.4. Zamysleme se nyní nad pojmem zákon rozložení náhodné veličiny. Víme, že znát zákon rozložení je totéž, jako znát rozdělovací funkci. Známe-li rozdělovací funkci, známe podle definice 3.1 i pravděpodobnost. Známe-li pravděpodobnost, potom podle definice 4.1 známe distribuční funkci. Ze znalosti distribuční funkce podle věty 4.1 e) umíme určit rozdělovací funkci. Vidíme, že ze zadání kterékoliv z informací o chování náhody tj.

- rozdělovací funkce
- pravděpodobnosti
- distribuční funkce

lze určit i ostatní dvě. Znalost libovolné z těchto tří funkcí je tedy ekvivalentní znalosti zákona rozložení náhodné veličiny.

Ve zbývající části kapitoly se budeme opět zabývat náhodným vektorem. Zobecněním definice 4.1 dostaneme:

Definice 4.2. *Buď P pravděpodobnost příslušná náhodnému vektoru $(X_1, \dots, X_n)$. Buď F reálná funkce n reálných proměnných definovaná vztahem*

$$F(x_1, x_2, \dots, x_n) = P(X_1 \le x_1, X_2 \le x_2, \dots, X_n \le x_n).$$

Funkci F nazýváme distribuční funkce náhodného vektoru $(X_1, X_2, \dots, X_n)$. Značíme $(X_1, X_2, \dots, X_n) \sim F$.

Z definice 4.2 a s využitím vlastností pravděpodobnosti (viz věta 3.1) lze dokázat následující větu, která je analogická větě 4.1.

Věta 4.2. Buď $(X_1, X_2, \dots, X_n)$ náhodný vektor s distribuční funkcí G. Potom:

a) G je vzhledem ke kterékoliv své proměnné neklesající.

b) G je vzhledem ke kterékoliv své proměnné zprava spojitá.

c) $\lim\limits_{x_i \to -\infty} G(x_1, x_2, \dots, x_n) = 0$ pro každé $i \in \{1, 2, \dots, n\}$.

d) $\lim\limits_{(x_1, x_2, \dots, x_n) \to (\infty, \infty, \dots, \infty,)} G(x_1, x_2, \dots, x_n) = 1.$

e) Je-li $(X_1, X_2, \dots, X_n)$ diskrétní s pravděpodobnostní funkcí g a oborem hodnot Ω, potom

1) $G(x_1, x_2, \dots, x_n) =$
$$= \sum_{(t_1, t_2, \dots, t_n) \in \Omega \cap (-\infty, x_1\rangle \times (-\infty, x_2\rangle \times \dots \times (-\infty, x_n\rangle} g(t_1, t_2, \dots, t_n),$$

2) $g(x_1, x_2, \dots, x_n) = G(x_1, x_2, \dots, x_n) -$
$$- \lim_{(h_1, h_2, \dots, h_n) \to (0, 0, \dots, 0)+} G(x_1 - h_1, x_2 - h_2, \dots, x_n - h_n),$$

3) $P(X_1 < x_1, X_2 < x_2, \dots, X_n < x_n) =$
$$= \lim_{(h_1, h_2, \dots, h_n) \to (0, 0, \dots, 0)+} G(x_1 - h_1, x_2 - h_2, \dots, x_n - h_n),$$

4) $P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = g(x_1, x_2, \dots, x_n).$

f) Je-li $(X_1, X_2, \dots, X_n)$ spojitý s hustotou g, potom

1) $G(x_1, x_2, \dots, x_n) = \int\limits_{-\infty}^{x_1} \cdots \int\limits_{-\infty}^{x_n} g(t_1, t_2, \dots, t_n)\, dt_1 dt_2 \dots dt_n,$

2) $g(x_1, x_2 \dots, x_n) = \dfrac{\partial^n}{\partial x_1 \partial x_2 \dots \partial x_n} G(x_1, x_2, \dots, x_n),$

3) $P(X_1 < x_1, X_2 < x_2, \dots, X_n < x_n) = G(x_1, x_2, \dots, x_n),$

4) $P(X_1 = x_1,\ X_2 = x_2, \dots, X_n = x_n) = 0.$

Poznámka 4.5. Pro náhodný vektor platí poznámka analogická poznámce 4.4. Pouze místo zákona rozložení náhodné veličiny uvažujeme zákon rozložení náhodného vektoru a vycházíme z definice 3.2, definice 4.2 a věty 4.2 e) a f).

Příklad 4.5. Hustota náhodného vektoru (X,Y) je dána vztahem

$$f(x,y)=\begin{cases}\frac{1}{6}\Big(\frac{x}{2}+\frac{y}{3}\Big) & \text{pro}\quad 0<x<2,\ 0<y<3\\ 0 & \text{jinak}\end{cases}.$$

Určete:

1) Distribuční funkci F náhodného vektoru (X,Y).

2) Pravděpodobnost, že náhodná veličina X nabude hodnoty z intervalu $(0,1)$ a současně náhodná veličina Y nabude hodnoty z intervalu $\langle 2,3\rangle$ jednak pomocí hustoty f, jednak pomocí distribuční funkce F.

Řešení. Obor hodnot Ω náhodného vektoru (X,Y) je obdélník $(0,2)\times(0,3)$, jedná se tedy o spojitý náhodný vektor. Při řešení příkladu využijeme větu 4.2.

1) Vyjděme z věty 4.2 f) vztahu 1). Potom

$$F(x,y)=\int_{-\infty}^{x}\Big[\int_{-\infty}^{y}f(u,v)\,dv\Big]du.$$

Dostaneme

$$F(x,y)=\begin{cases}\int_0^x\Big[\int_0^y\frac{1}{6}\Big(\frac{u}{2}+\frac{v}{3}\Big)dv\Big]du=\frac{1}{12}\Big(\frac{x^2y}{2}+\frac{xy^2}{3}\Big) & \text{pro } 0<x<2,\ 0<y<3\\ \int_0^2\Big[\int_0^y\frac{1}{6}\Big(\frac{u}{2}+\frac{v}{3}\Big)dv\Big]du=\frac{1}{6}\Big(y+\frac{y^2}{3}\Big) & \text{pro } x\ge 2,\ 0<y<3\\ \int_0^x\Big[\int_0^3\frac{1}{6}\Big(\frac{u}{2}+\frac{v}{3}\Big)dv\Big]du=\frac{1}{4}\Big(\frac{x^2}{2}+x\Big) & \text{pro } 0<x<2,\ y\ge 3\\ 1 & \text{pro } x\ge 2,\ y\ge 3\\ 0 & \text{jinak}\end{cases}$$

2) Pomocí hustoty f ze vztahu (3.5) dostaneme

$$P(0<X<1,\ 2\le Y\le 3)=\int_0^1\Big[\int_2^3\frac{1}{6}\Big(\frac{x}{2}+\frac{y}{3}\Big)dy\Big]dx=\frac{13}{72}.$$

Pomocí distribuční funkce

$$P(0<X<1,\ 2\le Y\le 3)=F(1,3)-F(0,3)-F(1,2)+F(0,2)=\frac{13}{72}.$$

♣

PDF-XChange Click to buy NOW! www.docu-track.com

5. Zákon rozložení marginálního náhodného vektoru

Definice 5.1. *Vynecháme-li z náhodného vektoru $(X_1, X_2, \ldots, X_n)$ alespoň jednu a nejvýše $n-1$ složek při nezměněném pořadí, získáme tzv. marginální (částečný) náhodný vektor. Náhodný vektor $(X_1, X_2, \ldots, X_n)$ pak v této souvislosti nazýváme simultánní (sdružený). Distribuční funkci a rozdělovací funkci marginálního [simultánního] náhodného vektoru doplňujeme někdy přívlastkem marginální [simultánní] .*

Příklad 5.1. Je-li (X_1, X_2, X_3) náhodný vektor, pak všechny marginální náhodné vektory jsou X_1, X_2, X_3, (X_1, X_2), (X_1, X_3), (X_2, X_3).

♣

Poznámka 5.1. Známe-li zákon rozložení simultánního náhodného vektoru, lze určit zákon rozložení všech marginálních náhodných vektorů. Opak obecně neplatí. Postup ukážeme na triviálních příkladech.

Příklad 5.2. Ve druhém ročníku na fakultě studuje, dejme tomu, 1000 studentů. Nechť jejich výsledky z matematiky a deskriptivní geometrie za I. semestr byly k určitému datu takové, jako udává tabulka 5.1

M \ DG	1	2	3	4	$\sum$
1	43	59	100	6	208
2	67	71	152	25	315
3	18	125	180	30	353
4	5	20	50	49	124
$\sum$	133	275	482	110	1000

Tabulka 5.1: Výsledky u zkoušek

Vyberme náhodně jednoho ze studentů a nechť náhodný vektor (X, Y) je
$(X, Y) =$ (známka z matematiky, známka z deskriptivní geometrie).
Určete rozdělovací funkce náhodných veličin X a Y.

Řešení. Obor hodnot $\Omega_{(X,Y)}$ náhodného vektoru (X, Y) je množina $\{1, 2, 3, 4, \} \times \{1, 2, 3, 4, \}$, jedná se tedy o diskrétní náhodný vektor. Pravděpodobnostní funkce g náhodného vektoru (X, Y) má zřejmě hodnoty z tabulky 5.2.

Obor hodnot Ω_X náhodné veličiny X je množina $\{1, 2, 3, 4\}$. X je tedy opět diskrétní náhodná veličina. Označme g_1 její pravděpodobnostní funkci. Potom

$$X \sim g_1(x) = P(X = x).$$

Pravděpodobnostní funkci g_1 stačí určit pro $x \in \Omega_X$, protože pro $x \in \mathbb{R} - \Omega_X$ je nulová. Postupně dostaneme

$$
\begin{aligned}
g_1(1) &= P(X=1) = P(X=1, Y \leq \infty) \\
&= \left| A = \{(x,y) \in \mathbb{R}^2;\ x=1, y \leq \infty\} \right| = \sum_{(x,y) \in A \cap \Omega} g(x,y) \\
&= \left| A \cap \Omega = \{(1,y) \in \mathbb{R}^2;\ g(1,y) \neq 0\} = \{(1,1),(1,2),(1,3),(1,4)\} \right| \\
&= g(1,1) + g(1,2) + g(1,3) + g(1,4) = 0.208 \quad \Big(= \sum_{\{y \in \mathbb{R};\ g(1,y) \neq 0\}} g(1,y) \Big).
\end{aligned}
$$

$x \setminus y$	1	2	3	4	$\sum$
1	0.043	0.059	0.100	0.006	0.208
2	0.067	0.071	0.152	0.025	0.315
3	0.018	0.125	0.180	0.030	0.353
4	0.005	0.020	0.050	0.049	0.124
$\sum$	0.133	0.275	0.482	0.110	1.000

Tabulka 5.2: Hodnoty pravděpodobnostní funkce $g(x,y)$

Podobně

$$
\begin{aligned}
g_1(2) &= P(X=2) = P(X=2, Y \leq \infty) \\
&= g(2,1) + g(2,2) + g(2,3) + g(2,4) = 0.315 \quad \Big(= \sum_{\{y \in \mathbb{R};\ g(2,y) \neq 0\}} g(2,y) \Big).
\end{aligned}
$$

Stejně bychom spočítali $g_1(3)$ a $g_1(4)$. Náhodná veličina X má pravděpodobnostní funkci danou tabulkou 5.3.

x	1	2	3	4
$g_1(x)$	0.208	0.315	0.353	0.124

Tabulka 5.3: Hodnoty $g_1(x)$

Všimněme si významu posledního sloupce v tabulce 5.2.

Obecně pro $x \in \Omega_X$ zřejmě platí

$$
g_1(x) = \sum_{\{y \in \mathbb{R};\ g(x,y) \neq 0\}} g(x,y),
$$

tedy

$$
X \sim g_1(x) = \begin{cases} \displaystyle\sum_{\{y \in \mathbb{R};\ g(x,y) \neq 0\}} g(x,y) & \text{pro } x \in \Omega_X \\ 0 & \text{jinak} \end{cases} \tag{5.1}
$$

Podobné úvahy bychom mohli provést pro náhodnou veličinu Y. Obor hodnot Ω_Y náhodné veličiny Y je opět množina $\{1,2,3,4\}$ a Y je tedy diskrétní náhodná veličina. Pro pravděpodobnostní funkci g_2 náhodné veličiny Y platí

$$Y \sim g_2(y) = P(Y = y) = P(X \leq \infty, Y = y)$$

Pravděpodobnostní funkci g_2 stačí určit pouze pro $y \in \Omega_Y$. Např.

$$\begin{aligned} g_2(1) = P(Y = 1) = P(X \leq \infty, Y = 1) &= \sum_{\{x \in \mathbb{R};\, g(x,1) \neq 0\}} g(x,1) \\ &= g(1,1) + g(2,1) + g(3,1) + g(4,1) \doteq 0.133. \end{aligned}$$

Obecně

$$(5.2) \qquad Y \sim g_2(y) = \begin{cases} \displaystyle\sum_{\{x \in \mathbb{R};\, g(x,y) \neq 0\}} g(x,y) & \text{pro } y \in \Omega_Y \\ 0 & \text{jinak} \end{cases}$$

Náhodná veličina Y má pravděpodobnostní funkci danou tabulkou 5.4

y	1	2	3	4
$g_2(y)$	0.133	0.275	0.482	0.110

Tabulka 5.4: Hodnoty $g_2(y)$

Všimněme si významu posledního řádku v tabulce 5.2.

♣

Poznámka 5.2. Uvažujme nyní spojitý náhodný vektor (X,Y) s hustotou g a oborem hodnot $\Omega_{(X,Y)}$. Označme opět g_1 $[g_2]$ rozdělovací funkci náhodné veličiny X $[Y]$ a Ω_X $[\Omega_Y]$ obor hodnot náhodné veličiny X $[Y]$. Z "analogie mezi diskrétním a spojitým případem" by asi nyní každý napsal

$$(5.3) \qquad X \sim g_1(x) = \begin{cases} \displaystyle\int_{\{y \in \mathbb{R};\, g(x,y) \neq 0\}} g(x,y)\, dy & \text{pro } x \in \Omega_X \\ 0 & \text{jinak} \end{cases},$$

$$(5.4) \qquad Y \sim g_2(y) = \begin{cases} \displaystyle\int_{\{x \in \mathbb{R};\, g(x,y) \neq 0\}} g(x,y)\, dx & \text{pro } y \in \Omega_Y \\ 0 & \text{jinak} \end{cases}.$$

Odvoďme nyní hustotu g_2 náhodné veličiny Y, tj. vztah (5.4). Vyjděme z distribuční funkce G_2 náhodné veličiny Y. Dostaneme

$$Y \sim G_2(y) = P(Y \leq y) = P(X \leq \infty, Y \leq y) = G(\infty, y).$$

Vyjádřeme distribuční funkce G_2 a G pomocí hustot, dostaneme

$$\int_{-\infty}^{y} g_2(v)\,dv = \int_{-\infty}^{y} \Big[\int_{-\infty}^{\infty} g(u,v)\,du \Big]\,dv.$$

Odtud

$$g_2(v) = \int_{-\infty}^{\infty} g(u,v)\,du$$

a tedy pro $y \in \Omega_Y$

$$\begin{aligned} g_2(y) &= \int_{-\infty}^{\infty} g(x,y)\,dx = \int_{\{x\in\mathbb{R};\ g(x,y)\neq 0\}} g(x,y)\,dx + \int_{\{x\in\mathbb{R};\ g(x,y)=0\}} g(x,y)\,dx \\ &= \int_{\{x\in\mathbb{R};\ g(x,y)\neq 0\}} g(x,y)\,dx. \end{aligned}$$

Čímž je vztah (5.4) dokázán.

Analogicky bychom dokázali vztah (5.3).

Příklad 5.3. Z kruhu o poloměru jedna a středu v počátku je náhodně vybírán bod tak, že každý bod kruhu má stejnou šanci dostat se do výběru. Nechť náhodný vektor (X,Y) znamená souřadnice vybraného bodu. Určete rozdělovací funkce náhodných veličin X a Y.

Řešení. Obor hodnot $\Omega_{(X,Y)}$ náhodného vektoru (X,Y) je množina

$$\{(x,y) \in \mathbb{R}^2;\ x^2 + y^2 \leq 1\},$$

jedná se tedy o spojitý náhodný vektor. Vzhledem k tomu, že každý bod z oboru hodnot má stejnou šanci být vybrán, musí pro hustotu g náhodného vektoru (X,Y) platit

$$g(x,y) = \begin{cases} c & \text{pro } (x,y) \in \Omega_{(X,Y)} \\ 0 & \text{jinak} \end{cases}.$$

Z vlastností hustoty (definice 2.5) dostáváme

$$1 = \iint_{\Omega_{(X,Y)}} g(x,y)\,dxdy = \iint_{\Omega_{(X,Y)}} c\,dxdy = \pi c,$$

odtud $c = 1/\pi$ a tedy

$$g(x,y) = \begin{cases} \dfrac{1}{\pi} & \text{pro } (x,y) \in \Omega_{(X,Y)} \\ 0 & \text{jinak} \end{cases}.$$

Obor hodnot Ω_X náhodné veličiny X je množina $\langle -1, 1\rangle$, jedná se tedy o spojitou náhodnou veličinu a pro její hustotu g_1 platí vztah (5.3).

Pro $x \in \Omega_X$ je (viz obrázek 5.1)

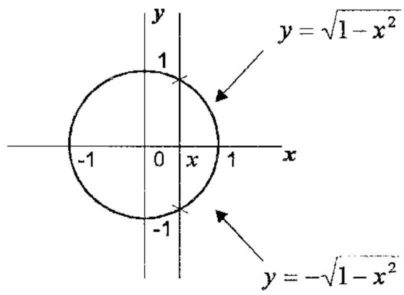
$$\{y \in \mathbb{R};\ g(x,y) \neq 0\} = \{y \in \mathbb{R};\ -\sqrt{1-x^2} \leq y \leq \sqrt{1-x^2}\}.$$

Tedy

$$X \sim g_1(x) = \begin{cases} \displaystyle\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi}\, dy & \text{pro } x \in \langle -1, 1\rangle \\ 0 & \text{jinak} \end{cases},$$

odtud

$$X \sim g_1(x) = \begin{cases} \dfrac{2}{\pi}\sqrt{1-x^2} & \text{pro } x \in \langle -1, 1\rangle \\ 0 & \text{jinak} \end{cases}.$$



Obrázek 5.1

Podobně obor hodnot Ω_Y náhodné veličiny Y je množina $\langle -1, 1\rangle$, jedná se tedy o spojitou náhodnou veličinu a pro její hustotu g_2 platí vztah (5.4). Pro $y \in \Omega_Y$ je

$$\{x \in \mathbb{R};\ g(x,y) \neq 0\} = \{x \in \mathbb{R};\ -\sqrt{1-y^2} \leq x \leq \sqrt{1-y^2}\}.$$

Tedy

$$Y \sim g_2(y) = \begin{cases} \displaystyle\int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} \frac{1}{\pi}\, dx & \text{pro } y \in \langle -1, 1\rangle \\ 0 & \text{jinak} \end{cases},$$

odtud

$$X \sim g_2(y) = \begin{cases} \dfrac{2}{\pi}\sqrt{1-y^2} & \text{pro } y \in \langle -1, 1\rangle \\ 0 & \text{jinak} \end{cases}.$$

♣

Zobecnění postupů z příkladů 5.2 a 5.3 pro náhodný vektor $(X_1, X_2, \ldots, X_n)$ uvádí následující věta. Máme-li marginální náhodný vektor $(X_{i_1}, X_{i_2}, \ldots, X_{i_k})$, $1 \leq i_1 < i_2 < \cdots < i_k \leq n$, $1 \leq k < n$, lze vhodným přečíslováním proměnných dosáhnout toho, že marginální vektor bude tvořen prvními složkami vektoru simultánního, tedy půjde o vektor $(X_1, X_2, \ldots, X_k)$. O jednu úroveň se tím sníží potřeba indexování.

Věta 5.1. Buď $(X_1, X_2, \ldots, X_n)$ simultánní náhodný vektor s distribuční funkcí G a rozdělovací funkcí g. Buď $1 \leq k < n$. Označme distribuční funkci marginálního náhodného vektoru $(X_1, X_2, \ldots, X_k)$ jako $G_{1,2,\ldots,k}$, rozdělovací funkci jako $g_{1,2,\ldots,k}$ a obor hodnot $\Omega_{(X_1, X_2, \ldots, X_k)}$. Potom

$$G_{1,2,\ldots,k}(x_1, x_2, \ldots, x_k) = G(x_1, x_2, \ldots, x_k, \infty, \infty, \ldots, \infty).$$

Dále platí

a) Nechť $(X_1, X_2, \ldots, X_n)$ je spojitý, potom

$$g_{1,2,\ldots,k}(x_1, x_2, \ldots, x_k) = \\ = \begin{cases} \displaystyle\int\cdots\int_{\{(x_{k+1},\ldots,x_n)\in\mathbb{R}^{n-k};\, g(x_1,\ldots,x_n)\neq 0\}} g(x_1, \ldots, x_n)\, dx_{k+1} \ldots dx_n & \\ \qquad\qquad \text{pro } (x_1, \ldots, x_k) \in \Omega_{(X_1,\ldots,X_k)} \\ 0 \qquad\qquad \text{jinak} \end{cases}.$$

b) Nechť $(X_1, X_2, \ldots, X_n)$ je diskrétní, potom

$$g_{1,2,\ldots,k}(x_1, x_2, \ldots x_k) = \\ = \begin{cases} \displaystyle\sum_{\{(x_{k+1},\ldots,x_n)\in\mathbb{R}^{n-k};\, g(x_1,\ldots,x_n)\neq 0\}} g(x_1, \ldots, x_n) & \\ \qquad\qquad \text{pro } (x_1, \ldots, x_k) \in \Omega_{(X_1,\ldots,X_k)} \\ 0 \qquad\qquad \text{jinak} \end{cases}.$$

Poznámka 5.3. Věta 5.1 vlastně říká, že při výpočtu hustoty [pravděpodobnostní funkce] marginálního náhodného vektoru integrujeme [sečítáme] "nenulovou" hustotu [nenulové hodnoty pravděpodobnostní funkce] simultánního náhodného vektoru podle zbývajících proměnných [přes zbývající proměnné].

Příklad 5.4. Náhodný vektor (X, Y, Z) má hustotu

$$g(x, y, z) = \begin{cases} \frac{8}{49}\, xyz & \text{pro } (x, y, z) \in \langle 1, 2\rangle \times \langle 3, 4\rangle \times \langle 3, 4\rangle \\ 0 & \text{jinak} \end{cases}.$$

Určete hustotu náhodné veličiny Y a náhodného vektoru (X, Z).

Řešení. Obor hodnot $\Omega_{(X,Y,Z)}$ náhodného vektoru (X, Y, Z) je množina $\langle 1, 2\rangle \times \langle 3, 4\rangle \times \langle 3, 4\rangle$, jedná se tedy o spojitý náhodný vektor.

Obor hodnot Ω_Y náhodné veličiny Y je množina $\langle 3, 4\rangle$ a pro hustotu g_2 náhodné veličiny Y podle věty 5.1 platí

$$g_2(y) = \begin{cases} \displaystyle\iint_{\{(x,z)\in\mathbb{R}^2;\ g(x,y,z)\neq 0\}} g(x, y, z)\, dx dz & \text{pro } y \in \Omega_Y \\ 0 & \text{jinak} \end{cases}.$$

Protože

$$\{(x,z) \in \mathbb{R}^2;\ g(x,y,z) \neq 0\} = \{(x,z) \in \mathbb{R}^2;\ 1 \leq x \leq 2,\ 3 \leq z \leq 4\},$$

je

$$\iint\limits_{\{(x,z)\in\mathbb{R}^2;\ g(x,y,z)\neq 0\}} g(x,y,z)\,dx\,dz = \int_3^4 \Big[\int_1^2 \frac{8}{49}\,xyz\,dx\Big]\,dz.$$

Výpočtem posledního integrálu dostaneme

$$g_2(y) = \begin{cases} \frac{2}{7}\,y & \text{pro } y \in \langle 3,4\rangle \\ 0 & \text{jinak} \end{cases}.$$

Obor hodnot $\Omega_{(X,Z)}$ náhodného vektoru (X,Z) je množina $\langle 1,2\rangle \times \langle 3,4\rangle$ a pro hustotu $g_{1,3}$ náhodného vektoru (X,Z) platí

$$g_{1,3}(x,z) = \begin{cases} \displaystyle\int\limits_{\{y\in\mathbb{R};\ g(x,y,z)\neq 0\}} g(x,y,z)\,dy & \text{pro } (x,z) \in \Omega_{(X,Z)} \\ 0 & \text{jinak} \end{cases}.$$

Protože

$$\{y \in \mathbb{R};\ g(x,y,z) \neq 0\} = \{y \in \mathbb{R};\ 3 \leq y \leq 4\},$$

dostaneme

$$\int\limits_{\{y\in\mathbb{R};\ g(x,y,z)\neq 0\}} g(x,y,z)\,dy = \int_3^4 \frac{8}{49}\,xyz\,dy.$$

Tedy

$$(X,Z) \sim g_{1,3}(x,z) = \begin{cases} \frac{4}{7}\,xz & \text{pro } (x,z) \in \langle 1,2\rangle \times \langle 3,4\rangle \\ 0 & \text{jinak} \end{cases}.$$

♣

6. Nezávislé náhodné veličiny

Než zavedeme definici nezávislých náhodných veličin $X_1, \ldots, X_n$, uvedeme dva motivující příklady, ve kterých ukážeme podstatu nezávislosti.

Příklad 6.1.

a) Nechť X je neznámá délka hrany krychle. Potom pro objem Y krychle platí $Y = X^3$. Náhodná veličina Y je tedy funkcí náhodné veličiny X. Veličiny X a Y jsou závislé, jedná se o tzv. funkční závislost.

b) Nechť X je váha náhodně vybraného studenta, Y je výška tohoto studenta a Z jeho měsíční kapesné. Potom zřejmě náhodnou veličinu Y nelze vyjádřit jako funkci náhodné veličiny X a tedy veličiny X a Y nejsou funkčně závislé. Přesto cítíme, že mezi Y a X nějaká závislost existuje. Pravděpodobnostní chování náhodné veličiny Y závisí totiž na tom, jakých hodnot nabyla náhodná veličina X, tj. pravděpodobnost, že náhodná veličina Y nabude hodnoty z nějakého intervalu závisí na tom, jakých hodnot nabyla náhodná veličina X. V tomto případě budeme hovořit o tzv. stochastické závislosti. Teorie pravděpodobnosti se právě zabývá závislostmi, resp. nezávislostmi posledního typu, přitom se slovo stochastické někdy vynechává. Mezi veličinami X a Z opět neexistuje funkční závislost, ale dokonce je zřejmé, že ani stochastická závislost. Takové veličiny budeme nazývat nezávislé.

Intuitivně tedy cítíme, co budeme rozumět stochastickou nezávislostí, ale s intuicí samozřejmě v technických aplikacích nevystačíme. Potřebujeme nějaký definiční vztah, který jednak odpovídá naší intuitivní představě a jednak umožňuje ověřit, zda jsou veličiny nezávislé.

♣

Příklad 6.2. Vraťme se k příkladu 5.2. Zajímá nás, zda pravděpodobnostní chování náhodné veličiny Y závisí na tom, jakých hodnot nabyla náhodná veličina X. Nechť $Y|X = 1$ značí známku z deskriptivní geometrie těch studentů, kteří mají z matematiky výbornou. Zřejmě se jedná o diskrétní náhodnou veličinu s oborem hodnot $\{1, 2, 3, 4\}$. Pravděpodobnostní funkci u této veličiny získáme z tabulky 5.1, dostaneme tabulku 6.1

$y\|X = 1$	1	2	3	4
$u(y)$	43/208	59/208	100/208	6/208

Tabulka 6.1: Hodnoty pravděpodobnostní funkce $u(y)$.

Neboť např.

$$u(1) = P(Y = 1|X = 1) = \frac{43}{208},$$

$$u(2) = P(Y = 2|X = 1) = \frac{59}{208}.$$

PDF-XChange Click to buy NOW! www.docu-track.com

Obecně

$$u(y) = P(Y = y|X = 1),$$

což vlastně značí pravděpodobnost, že náhodná veličina Y nabude hodnoty y za podmínky, že náhodná veličina X nabyla hodnoty 1.
Dále zřejmě

$$u(1) = P(Y = 1|X = 1) = \frac{43}{208} = \frac{\frac{43}{1000}}{\frac{208}{1000}} = \frac{P(X = 1, Y = 1)}{P(X = 1)}.$$

odtud

$$P(Y = 1|X = 1) = \frac{g(1,1)}{g_1(1)}.$$

Podobně

$$u(2) = P(Y = 2|X = 1) = \frac{59}{208} = \frac{\frac{59}{1000}}{\frac{208}{1000}} = \frac{P(X = 1, Y = 2)}{P(X = 1)} = \frac{g(1,2)}{g_1(1)}.$$

Obecně

$$u(y) = P(Y = y|X = 1) = \frac{P(X = 1, Y = y)}{P(X = 1)} = \frac{g(1,y)}{g_1(1)} \quad \text{pro} \quad y \in \mathbb{R}.$$

Tedy pro pevné $x \in \Omega_x$ dostáváme

$$(6.1) \qquad P(Y = y|X = x) = \frac{P(X = x, Y = y)}{P(X = x)} = \frac{g(x,y)}{g_1(x)} \quad \text{pro} \quad y \in \mathbb{R}.$$

Pravděpodobnostní chování náhodné veličiny Y zřejmě nezávisí na tom, jakých hodnot nabyla náhodná veličina X, když

$$P(Y = y|X = x) = P(Y = y) = g_2(y) \quad \text{pro každé} \quad y \in \mathbb{R},\ x \in \Omega_X.$$

Odtud a ze vztahu (6.1) dostaneme požadavek nezávislosti Y na X ve tvaru

$$\frac{g(x,y)}{g_1(x)} = g_2(y) \quad \text{pro každé} \quad y \in \mathbb{R},\ x \in \Omega_X,$$

tj. chceme, aby platilo

$$g(x,y) = g_1(x)\,g_2(y) \quad \text{pro každé} \quad y \in \mathbb{R},\ x \in \Omega_X.$$

Protože pro $x \notin \Omega_X$ je $g_1(x) = 0$ a $g(x,y) = 0$, lze poslední vztah napsat ve tvaru

$$(6.2) \qquad g(x,y) = g_1(x)\,g_2(y) \quad \text{pro každé} \quad (x,y) \in \mathbb{R}^2.$$

Z tabulky 5.2 vyplývá, že $g(1,1) \neq g_1(1)\,g_2(1)$, tedy vztah (6.2) neplatí, tj. pravděpodobnostní chování náhodné veličiny Y závisí na tom, jakých hodnot nabyla náhodná veličina X.

Podobné úvahy bychom mohli provést pro zjištění faktu, zda pravděpodobnostní chování náhodné veličiny X závisí na tom jakých hodnot nabyla náhodná veličina Y. Úplně stejně jako výše bychom ukázali, že pravděpodobnostní chování náhodné veličiny X nezávisí na tom jakých hodnot nabyla náhodná veličina Y právě tehdy, když platí

$$g(x,y) = g_1(x)\,g_2(y) \quad \text{pro každé } (x,y) \in \mathbb{R}^2,$$

tj. vztah (6.2).

Z uvedeného příkladu vyplývá:

1) Místo nezávislosti Y na X resp. X na Y můžeme hovořit o nezávislosti X a Y.

2) Nezávislost X a Y lze v diskrétním případě ověřit pomocí vztahu (6.2). Podobné úvahy bychom mohli provést pro případ spojitých náhodných veličin a přes distribuční funkce G, G_1, G_2 bychom opět dospěli k požadavku tvaru (6.2) pro hustoty $g_1\,[g_2]$ náhodných veličin $X\,[Y]$. Logická se tedy zdá následující definice nezávislosti.

Definice 6.1. *Nechť (X,Y) je náhodný vektor s rozdělovací funkcí g. Označme $g_1[g_2]$ rozdělovací funkci náhodné veličiny X $[Y]$. Řekneme, že náhodné veličiny X a Y jsou stochasticky nezávislé, když platí*

$$g(x,y) = g_1(x)\,g_2(y) \quad \textit{pro každé } (x,y) \in \mathbb{R}^2.$$

Poznámka 6.1. Známe-li tedy rozdělovací funkci g náhodného vektoru (X,Y), lze určit podle definice 6.1, zda jsou veličiny X a Y nezávislé. Podle kapitoly 5 spočítáme rozdělovací funkce g_1, g_2 náhodných veličin X, Y a pak ověříme, zda platí definiční vztah nezávislosti.

Bez důkazu uvedeme následující větu:

Věta 6.1. Nechť (X,Y) je náhodný vektor s distribuční funkcí G. Označme $G_1[G_2]$ distribuční funkci náhodné veličiny X $[Y]$. Potom jsou veličiny X a Y stochasticky nezávislé právě tehdy, když

$$G(x,y) = G_1(x)\,G_2(y) \quad \text{pro každé } (x,y) \in \mathbb{R}^2.$$

Příklad 6.3. Určete, zda jsou veličiny X a Y v příkladu 5.3 nezávislé.

Řešení. V příkladu 5.3 máme

$$(X,Y) \sim g(x,y) = \begin{cases} \dfrac{1}{\pi} & \text{pro } (x,y) \in \Omega_{(X,Y)} \\ 0 & \text{jinak} \end{cases},$$

kde $\Omega_{(X,Y)} = \{(x,y) \in \mathbb{R}^2;\ x^2 + y^2 \leq 1\}$,

$$X \sim g_1(x) = \begin{cases} \dfrac{2}{\pi}\sqrt{1-x^2} & \text{pro } x \in \langle -1, 1\rangle \\ 0 & \text{jinak} \end{cases};$$

$$Y \sim g_2(y) = \begin{cases} \dfrac{2}{\pi}\sqrt{1-y^2} & \text{pro } \ y \in \langle -1, 1\rangle \\ 0 & \text{jinak} \end{cases}.$$

Veličiny X a Y budou podle definice 6.1 nezávislé, když

$$g(x,y) = g_1(x)\, g_2(y) \quad \text{pro každé } \ (x,y) \in \mathbb{R}^2.$$

Zřejmě $g(0,0) \neq g_1(0) \cdot g_2(0)$. Tedy veličiny X a Y nejsou nezávislé. ♣

Příklad 6.4. Náhodný vektor (X, Y) má rozdělovací funkci g

a) danou tabulkou

$x \setminus y$	1	2
0	0.03	0.07
1	0.27	0.63

b) danou tabulkou

$x \setminus y$	1	2	3
0	0.2	0.1	0.1
1	0.3	0.2	0.1

c)

$$g(x,y) = \begin{cases} 4x(y-1) & \text{pro } \ (x,y) \in (0,1) \times (1,2) \\ 0 & \text{jinak} \end{cases}.$$

Zjistěte, zda jsou veličiny X a Y nezávislé.

Řešení. Ve všech případech vyjdeme z definice nezávislosti, tj. označme $\Omega_{(X,Y)}$ obor hodnot náhodného vektoru (X, Y), $g_1\, [g_2]$ rozdělovací funkci náhodné veličiny $X\, [Y]$. Veličiny X a Y budou nezávislé právě tehdy, když

$$g(x,y) = g_1(x)\, g_2(y) \quad \text{pro každé } \quad (x,y) \in \mathbb{R}^2.$$

a) Obor hodnot $\Omega_{(X,Y)}$ náhodného vektory (X, Y) je množina $\{0,1\} \times \{1,2\}$, jedná se tedy o diskrétní náhodný vektor. Podle kapitoly 5 dostaneme pravděpodobnostní funkci $g_1 [g_2]$ náhodné veličiny X $[Y]$ - viz tabulka 6.2.

$x \setminus y$	1	2	$g_1(x)$
0	0.03	0.07	0.10
1	0.27	0.63	0.90
$g_2(y)$	0.30	0.70	1.00

Tabulka 6.2

Odtud

$$g(0,1) = g_1(0)\, g_2(1),$$

$$g(0,2) = g_1(0)\,g_2(2),$$
$$g(1,1) = g_1(1)\,g_2(1),$$
$$g(1,2) = g_1(1)\,g_2(2).$$

Tedy

$$g(x,y) = g_1(x)\,g_2(y) \quad \text{pro každé} \quad (x,y) \in \mathbb{R}^2$$

a veličiny X a Y jsou nezávislé.

b) Obor hodnot $\Omega_{(X,Y)}$ náhodného vektoru (X,Y) je množina $\{0,1\} \times \{1,2,3\}$, jedná se opět o diskrétní náhodný vektor a postupujeme stejně jako v a). Pravděpodobnostní funkce $g_1[g_2]$ náhodné veličiny X $[Y]$ opět dopočítáme v tabulce, dostaneme tabulku 6.3.

$x \setminus y$	1	2	3	$g_1(x)$
0	0.2	0.1	0.1	0.4
1	0.3	0.2	0.1	0.6
$g_2(y)$	0.5	0.3	0.2	1.0

Tabulka 6.3

Potom

$$g(0,1) = g_1(0)\,g_2(1),$$
$$g(0,2) \neq g_1(0)\,g_2(2).$$

Požadovaný vztah není splněn pro bod $(0,2)$ a tedy náhodné veličiny X, Y nejsou nezávislé.

c) Obor hodnot $\Omega_{(X,Y)}$ náhodného vektory (X,Y) je množina $(0,1) \times (1,2)$, jedná se tedy o spojitý náhodný vektor. Podle kapitoly 5 postupně dostaneme

$$X \sim g_1(x) = \begin{cases} \displaystyle\int_{\{y\in\mathbb{R};\, g(x,y)\neq 0\}} g(x,y)\,dy & \text{pro } \ x \in \Omega_X \\ 0 & \text{jinak} \end{cases} \Rightarrow$$

$$\Rightarrow g_1(x) = \begin{cases} \displaystyle\int_1^2 4x(y-1)\,dy = 2x & \text{pro } \ x \in (0,1) \\ 0 & \text{jinak} \end{cases}.$$

Podobně

$$Y \sim g_2(y) = \begin{cases} \displaystyle\int_{\{x\in\mathbb{R};\, g(x,y)\neq 0\}} g(x,y)\,dx & \text{pro } \ y \in \Omega_Y \\ 0 & \text{jinak} \end{cases} \Rightarrow$$

$$\Rightarrow g_2(y) = \begin{cases} \displaystyle\int_0^1 4x(y-1)\,dx = 2(y-1) & \text{pro } \ y \in (1,2) \\ 0 & \text{jinak} \end{cases}.$$

Protože

$$4x(y-1) = 2x \cdot 2(y-1) \quad \text{pro každé } (x,y) \in \Omega_{(X,Y)},$$

jsou veličiny X a Y nezávislé.

♣

Poznámka 6.2. Nechť (X, Y) značí

$$X = \begin{cases} 0 & \text{respondent není spokojený ve své práci} \\ 1 & \text{respondent je spokojený ve své práci} \end{cases},$$

$$Y = \begin{cases} 1 & \text{respondent je muž} \\ 2 & \text{respondent je žena} \end{cases}.$$

Nechť náhodný vektor (X, Y) má rozdělovací funkci z tabulky 6.2. Potom spokojenost v práci není závislá na pohlaví.

Zobecněním definice 6.1 obdržíme definici nezávislosti náhodných veličin X_1, $X_2, \dots, X_n$.

Definice 6.2. *Nechť $(X_1, X_2, \dots, X_n)$ je náhodný vektor s rozdělovací funkcí g. Označme g_i rozdělovací funkci náhodné veličiny X_i pro $i = 1, 2, \dots, n$. Řekneme, že náhodné veličiny $X_1, X_2, \dots, X_n$ jsou stochasticky nezávislé, když platí*

$$g(x_1, x_2, \dots, x_n) = g_1(x_1)\, g_2(x_2) \dots g_n(x_n) \quad \text{pro každé } (x_1, x_2, \dots, x_n) \in \mathbb{R}^n. \tag{6.3}$$

Věta 6.2. Nechť $(X_1, X_2, \dots, X_n)$ je náhodný vektor s distribuční funkcí G. Označme G_i distribuční funkci náhodné veličiny X_i pro $i = 1, 2, \dots, n$. Potom jsou náhodné veličiny $X_1, X_2, \dots, X_n$ nezávislé právě tehdy, když platí

$$G(x_1, x_2, \dots, x_n) = G_1(x_1)\, G_2(x_2) \dots G_n(x_n) \quad \text{pro každé } (x_1, x_2, \dots, x_n) \in \mathbb{R}^n.$$

Příklad 6.5. Buď $X_1, X_2, \dots, X_n$ nezávislé náhodné veličiny, $I_1, I_2, \dots, I_n$ reálné intervaly. Dokažte, že platí

$$P(X_1 \in I_1, X_2 \in I_2, \dots, X_n \in I_n) = P(X_1 \in I_1)\, P(X_2 \in I_2) \dots P(X_n \in I_n).$$

Řešení. Důkaz provedeme pro spojité náhodné veličiny $X_1, X_2, \dots, X_n$. Buď tedy $(X_1, X_2, \dots, X_n)$ spojitý náhodný vektor s hustotou g. Označme g_i hustotu náhodné veličiny X_i pro $i = 1, 2, \dots, n$. Potom podle (6.3) je

$$g(x_1, x_2, \dots, x_n) = g_1(x_1)\, g_2(x_2) \dots g_n(x_n) \quad \text{pro každé } (x_1, x_2, \dots, x_n) \in \mathbb{R}^n.$$

Odtud

$$
\begin{aligned}
P(X_1 \in I_1, X_2 \in I_2, \ldots, X_n \in I_n) &= \int\limits_{I_1 \times I_2 \times \cdots \times I_n} \!\!\cdots\!\! \int g(x_1, x_2, \ldots, x_n)\, dx_1 dx_2 \ldots dx_n \\
&= \int\limits_{I_1 \times I_2 \times \cdots \times I_n} \!\!\cdots\!\! \int g_1(x_1)\, g_2(x_2) \ldots g_n(x_n)\, dx_1 dx_2 \ldots dx_n \\
&= \int\limits_{I_1} g_1(x_1)\, dx_1 \int\limits_{I_2} g_2(x_2)\, dx_2 \ldots \int\limits_{I_n} g_n(x_n)\, dx_n \\
&= P(X_1 \in I_1)\, P(X_2 \in I_2) \ldots P(X_n \in I_n).
\end{aligned}
$$

♣

Příklad 6.6. Předpokládejme, že životnost výrobku (v letech) je náhodná veličina X s distribuční funkcí

$$
F(x) = \begin{cases} 0 & \text{pro } x \le 0 \\ 1 - e^{-\dfrac{x}{10}} & \text{pro } x > 0 \end{cases}.
$$

Jaká je pravděpodobnost, že:

a) U dvou stejných výrobků dojde k první poruše až po dvou letech?

b) Alespoň jeden ze dvou výrobků vydrží bez poruchy alespoň dva roky?

Řešení. Označme X_i životnost i-tého výrobku, $i = 1, 2$. Potom mají náhodné veličiny X_1, X_2 stejné rozdělení jako náhodná veličina X a jsou zřejmě nezávislé. Na základě příkladu 6.5 dostáváme

a)

$$
\begin{aligned}
P(X_1 > 2, X_2 > 2) &= P(X_1 > 2)\, P(X_2 > 2) \\
&= \left[P(X > 2)\right]^2 = \left[1 - P(X \le 2)\right]^2 \\
&= \left[1 - F(2)\right]^2 = \left[e^{-\frac{1}{5}}\right]^2 \doteq 0.670.
\end{aligned}
$$

b)

$$
\begin{aligned}
P(X_1 > 2 \vee X_2 > 2) &= P(X_1 > 2) + P(X_2 > 2) - \\
&- P(X_1 > 2, X_2 > 2) = 2\left[1 - F(2)\right] - 0.670 \\
&= 2e^{-\frac{1}{5}} - 0.670 \doteq 0.967.
\end{aligned}
$$

♣

Příklad 6.7. Vraťte se k příkladu 3.6 z kapitoly 3. Ověřte, že jsou veličiny X a Y nezávislé a na základě této skutečnosti vypočtěte opět pravděpodobnosti v tomto příkladu hledané.

♣

7. Transformace náhodných veličin a vektorů

Při řešení některých pravděpodobnostních úloh se setkáváme se situací, kdy známe zákon rozdělení náhodné veličiny X a hledáme zákon rozdělení náhodné veličiny Y, která je funkcí veličiny X, tj.

$$Y = h(X),$$

kde h je reálná funkce jedné reálné proměnné definovaná na oboru hodnot náhodné veličiny X. O náhodné veličině Y potom říkáme, že vznikla transformací h náhodné veličiny X.

Je zřejmé, že libovolnou transformací h diskrétní veličiny musí vzniknout diskrétní náhodná veličina. Různými transformacemi spojité náhodné veličiny však nemusí vzniknout veličina spojitá.

Situaci si lze jednoduše představit v případě diskrétní náhodné veličiny.

Příklad 7.1. Vraťme se k příkladu P - 3.7 a nechť zkoušející zvolil strategii b). Dejme tomu, že se rodiče se studentem domluvili následovně

a) za jedničku nebo dvojku dostane 400 Kč, za trojku nebo nevyhovující nedostane nic,

b) za jedničku dostane 400 Kč, za dvojku 200 Kč, za trojku nic a za nevyhovující dá rodičům 200 Kč.

Určete rozdělení změny studentova finančního stavu.

Řešení. Náhodnou veličinu výsledek zkoušky V_b označme nyní bez indexu jako V a její pravděpodobnostní funkci p_b označme také bez indexu jako p.

a) Označme Z_a změnu studentova finančního stavu. Obor hodnot Ω_a náhodné veličiny Z_a je zřejmě množina $\{0, 400\}$. Jedná se tedy o diskrétní náhodnou veličinu, která vznikla transformací náhodné veličiny V - a to podle tabulky 7.1.

v	1	2	3	4
z	400	400	0	0

Tabulka 7.1

Potom $Z_a = h_a(V)$.

Pravděpodobnostní funkci q_a náhodné veličiny Z_a stačí určit pro $z \in \Omega_a$, protože všude jinde je nulová. Dostaneme

$$\begin{aligned} q_a(400) &= P(Z_a = 400) = P(V = 1 \vee V = 2) \\ &= P(V = 1) + P(V = 2) = p(1) + p(2) = 0.75, \\ q_a(0) &= P(Z_a = 0) = P(V = 3 \vee V = 4) \\ &= P(V = 3) + P(V = 4) = p(3) + p(4) = 0.25. \end{aligned}$$

Tedy

$$Z_a \sim q_a(z) = \begin{cases} 0.75 & \text{pro } z = 400 \\ 0.25 & \text{pro } z = 0 \\ 0 & \text{jinak} \end{cases} .$$

b) Označme v tomto případě změnu studentova finančního stavu jako Z_b. Obor hodnot Ω_b náhodné veličiny Z_b je zřejmě množina $\{-200, 0, 200, 400\}$. Jedná se tedy o diskrétní náhodnou veličinu, která vznikla transformací náhodné veličiny V – to podle tabulky 7.2.

v	1	2	3	4
z	400	300	0	-200

Tabulka 7.2

Tedy $Z_b = h_b(V)$. Veličina Z_b má zřejmě pravděpodobnostní funkci q_b danou tabulkou 7.3.

z	-200	0	200	400
$q_b(z)$	0.21	0.04	0.42	0.33

Tabulka 7.3: Hodnoty pravděpodobnostní funkce $q_b(z)$.

Protože je funkce h_b prostá na množině Ω_b, existuje k ní inverzní funkce h_b^{-1}, tj. platí $v = h_b^{-1}(z)$. Vztah mezi pravděpodobnostními funkcemi p a q_b veličin V a Z_b můžeme v tomto případě vyjádřit následovně. Pro $z \in \Omega_b$ platí

$$q_b(z) = P(Z_b = z) = P(h_b(V) = z) = P(V = h_b^{-1}(z)) = p(h_b^{-1}(z)).$$

Tedy

$$q_b(z) = \begin{cases} p(h_b^{-1}(z)) & \text{pro } z \in \Omega_b \\ 0 & \text{jinak} \end{cases} .$$

♣

Poznámka 7.1. V příkladu 7.1 b) jsme ukázali, že platí následující věta:

Věta 7.1. Nechť X a Y jsou dvě diskrétní náhodné veličiny a nechť náhodná veličina Y vznikla transformací h náhodné veličiny X, tj. $Y = h(X)$. Označme Ω_X $[\Omega_Y]$ obor hodnot náhodné veličiny X $[Y]$ a p $[q]$ pravděpodobnostní funkci náhodné veličiny X $[Y]$. Je-li funkce h prostá na množině Ω_x, potom pro pravděpodobnostní funkci q náhodné veličiny Y platí

$$q(y) = \begin{cases} p(h^{-1}(y)) & \text{pro } y \in \Omega_Y \\ 0 & \text{jinak} \end{cases} ,$$

kde h^{-1} je funke inverzní k funkci h.

Poznámka 7.2. Připomeňme si, že inverzní funkci h^{-1} k funkci h získáme tak, že z rovnice $y = h(x)$ vyjádříme x. Výsledek označíme $x = h^{-1}(y)$.

PDF-XChange
Click to buy NOW!
www.docu-track.com

Příklad 7.2. Uvažujme kulovou částici o průměru r. Řezná rovina protne tuto částici v kruhu o průměru Z. Označme T vzdálenost řezné roviny od středu částice. Nechť T je náhodná veličina mající hustotu

$$f(t) = \begin{cases} \dfrac{2}{r} & \text{pro} \quad 0 \le t \le \dfrac{r}{2} \\ 0 & \text{jinak} \end{cases}.$$

Stanovte hustotu veličiny Z.

Řešení. Pro realizace z a t náhodných veličin Z a T zřejmě platí (nakreslete si obrázek!)

$$\left(\frac{z}{2}\right)^2 = \left(\frac{r}{2}\right)^2 - t^2,$$

odtud

$$z^2 = r^2 - 4t^2,$$

a tedy

$$z = \sqrt{r^2 - 4t^2},$$

protože $z \ge 0$. Tedy

$$Z = \sqrt{r^2 - 4T^2} = h(T), \tag{1}$$

tj. náhodná veličina Z vznikla transformací $z = h(t) = \sqrt{r^2 - 4t^2}$ náhodné veličiny T.

Obor hodnot Ω_T náhodné veličiny T je intevrval $\langle 0, r/2\rangle$, obor hodnot Ω_Z náhodné veličiny Z je interval $\langle 0, r\rangle$. Obě náhodné veličiny jsou tedy spojité náhodné veličiny.

Označme F distribuční funkci náhodné veličiny T, g [G] hustotu [distribuční funkci] náhodné veličiny Z. Hustotu a distribuční funkci náhodné veličiny Z stačí určit pro $z \in \Omega_Z$, protože pro $z \notin \Omega_Z$ obě funkce známe.

Vyjdeme z definičního vztahu distribuční funkce náhodné veličiny Z, tj. vztahu

$$G(z) = P\big(Z \le z\big). \tag{2}$$

Pravděpodobnost ve vztahu (2) vyjádříme pomocí náhodné veličiny T. Ze vztahu (1) dostaneme

$$\begin{aligned} G(z) &= P\big(\sqrt{r^2 - 4T^2} \le z\big) = P\big(r^2 - 4T^2 \le z^2\big) \\ &= P\Big(T^2 \ge \frac{1}{4}(r^2 - z^2)\Big) \\ &= P\Big(T \ge \frac{1}{2}\sqrt{r^2 - z^2} \vee T \le -\frac{1}{2}\sqrt{r^2 - z^2}\Big) \\ &= P\Big(T \ge \frac{1}{2}\sqrt{r^2 - z^2}\Big) + P\Big(T \le -\frac{1}{2}\sqrt{r^2 - z^2}\Big) \\ &= P\Big(T \ge \frac{1}{2}\sqrt{r^2 - z^2}\Big). \end{aligned}$$

(Protože $P\Big(T \leq -\frac{1}{2}\sqrt{r^2-z^2}\Big) = 0.$)
Tedy

$$(3) \qquad G(z) = \begin{cases} 0 & \text{pro } z < 0 \\ P\Big(T \geq \frac{1}{2}\sqrt{r^2-z^2}\Big) & \text{pro } z \in \langle 0, r\rangle \\ 1 & \text{pro } z > r \end{cases} .$$

Pomocí distribuční funkce F náhodné veličiny T pak ze vztahu (3) dostaneme

$$(4) \quad G(z) = 1 - P\Big(T < \frac{1}{2}\sqrt{r^2-z^2}\Big) = 1 - F\Big(\frac{1}{2}\sqrt{r^2-z^2}\Big) \quad \text{pro každé } z \in \Omega_Z.$$

Vyjádřili jsme tedy neznámou distribuční funkci náhodné veličiny Z pomocí známé distribuční funkce náhodné veličiny T.

Nyní pro hustotu g náhodné veličiny Z ze vztahu (4) dostáváme

$$\begin{aligned} g(z) &= \frac{d}{dz}G(z) = \frac{d}{dz}\left[1 - F\Big(\frac{1}{2}\sqrt{r^2-z^2}\Big)\right] \\ &= -\frac{d}{dz}F\Big(\frac{1}{2}\sqrt{r^2-z^2}\Big) = \left|u = u(z) = \frac{1}{2}\sqrt{r^2-z^2}\right| = -\frac{d}{du}F(u)\,\frac{d}{dz}u(z) \\ &= -f(u)\,\frac{1}{4}\,(r^2-z^2)^{-\frac{1}{2}}\,(-2z) = \frac{1}{2}\,f(u)\,\frac{z}{\sqrt{r^2-z^2}} \\ &= \frac{1}{2}\,f\Big(\frac{1}{2}\sqrt{r^2-z^2}\Big)\,\frac{z}{\sqrt{r^2-z^2}}. \end{aligned}$$

Tedy

$$(5) \qquad g(z) = \frac{1}{2}\,\frac{z}{\sqrt{r^2-z^2}}\,f\Big(\frac{1}{2}\sqrt{r^2-z^2}\Big) \quad \text{pro } z \in \Omega_Z.$$

Zbývá vyjádřit hustotu $f(\frac{1}{2}\sqrt{r^2-z^2})$. Pro $z \in \Omega_Z$ je

$$\frac{1}{2}\,\sqrt{r^2-z^2} \in \left\langle 0, \frac{r}{2}\right\rangle$$

a tedy pro $z \in \Omega_Z$ je

$$f\Big(\frac{1}{2}\,\sqrt{r^2-z^2}\Big) = \frac{2}{r}.$$

Odtud a ze vztahu (5) dostaneme

$$g(z) = \frac{1}{2}\,\frac{z}{\sqrt{r^2-z^2}}\,\frac{2}{r} = \frac{z}{r\sqrt{r^2-z^2}} \quad \text{pro } z \in \Omega_Z.$$

Tedy

$$g(z) = \begin{cases} \dfrac{z}{r\sqrt{r^2-z^2}} & \text{pro } 0 < z < r \\ 0 & \text{jinak} \end{cases} .$$

♣

PDF-Xchange Click to buy NOW! www.docu-track.com

Příklad 7.3. Náhodná chyba měření X má normální rozdělení s parametry μ a σ^2, tj. X má hustotu

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}\,\mathrm{e}^{-\dfrac{(x-\mu)^2}{2\sigma^2}} \qquad \text{pro } x \in \mathbb{R}$$

(viz definice P - 9.2.1). Určete rozdělení absolutní hodnoty náhodné chyby měření.

Řešení. Označme $Y = |X|$, tedy veličina Y vznikla transformací $y = h(x) = |x|$ náhodné veličiny X. Obor hodnot Ω_X náhodné veličiny X je $\mathbb{R}$. Obor hodnot Ω_Y náhodné veličiny Y je tedy interval $\langle 0, \infty)$, neboť

$$-\infty < x < \infty \Rightarrow 0 \leq |x| < \infty \Rightarrow 0 \leq y < \infty.$$

Označme F distribuční funkci náhodné veličiny X, g $[G]$ hustotu [distribuční funkci] náhodné veličiny Y. Hustotu a distribuční funkci náhodné veličiny Y stačí určit pro $y \in \Omega_Y$. Nejprve opět vyjádříme distribuční funkci G náhodné veličiny Y pomocí distribuční funkce F náhodné veličiny X tak, že vyjdeme z definičního vztahu distribuční funkce náhodné veličiny Y. Dostaneme

$$\begin{aligned} G(y) &= P\big(Y \leq y\big) = P\big(|X| \leq y\big) = P\big(-y \leq X \leq y\big) \\ &= F(y) - F(-y) \quad \text{pro } y \in \Omega_Y. \end{aligned}$$

Potom

$$\begin{aligned} g(y) &= \frac{d}{dy}G(y) = \frac{d}{dy}\big[F(y) - F(-y)\big] = \frac{d}{dy}F(y) - \frac{d}{dy}F(-y) \\ &= \Big| u = u(y) = -y \Big| = f(y) - \frac{d}{du}F(u)\,\frac{d}{dy}u(y) \\ &= f(y) + f(u) = f(y) + f(-y) \quad \text{pro } y \in \Omega_Y. \end{aligned}$$

Nyní zbývá vyjádřit $f(y)$, resp. $f(-y)$.
Pro $y \in \langle 0, \infty)$ je

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma}\,\mathrm{e}^{-\dfrac{(y-\mu)^2}{2\sigma^2}}.$$

Pro $y \in \langle 0, \infty)$ je

$$f(-y) = \frac{1}{\sqrt{2\pi}\sigma}\,\mathrm{e}^{-\dfrac{(y+\mu)^2}{2\sigma^2}}.$$

Tedy

$$g(y) = \begin{cases} \dfrac{1}{\sqrt{2\pi}\sigma}\left[\mathrm{e}^{-\dfrac{(y-\mu)^2}{2\sigma^2}} + \mathrm{e}^{-\dfrac{(y+\mu)^2}{2\sigma^2}}\right] & \text{pro } y \in \langle 0, \infty) \\ 0 & \text{jinak} \end{cases}.$$

♣

Příklad 7.4. Pro náhodnou veličinu Y platí $Y = |X|$, kde X je náhodná veličina s hustotou

$$f(x) = \begin{cases} \frac{1}{3} & \text{pro } x \in \langle -1, 2\rangle \\ 0 & \text{jinak} \end{cases}.$$

Určete rozdělení náhodné veličiny Y.

Řešení. Náhodná veličina Y vznikla transformací $y = h(x) = |x|$ náhodné veličiny X. Obor hodnot Ω_X náhodné veličiny X je interval $\langle -1, 2\rangle$. Odtud

$$-1 \leq x \leq 2 \Rightarrow 0 \leq |x| \leq 2 \Rightarrow 0 \leq y \leq 2.$$

Tedy obor hodnot Ω_Y náhodné veličiny Y je interval $\langle 0, 2\rangle$.

Označme F distribuční funkci náhodné veličiny X, g [G] hustotu [distribuční funkci] náhodné veličiny Y. Hustotu a distribuční funkci náhodné veličiny Y stačí určit pro $y \in \Omega_Y$. Stejně jako v příkladu 7.3 dostaneme

$$G(y) = F(y) - F(-y) \quad \text{pro } y \in \Omega_Y,$$
$$g(y) = f(y) + f(-y) \quad \text{pro } y \in \Omega_Y.$$

Nyní zbývá vyjádřit $f(y)$, resp. $f(-y)$ pro $y \in \Omega_Y = \langle 0, 2\rangle$.

Je-li

$$y \in \langle 0, 2\rangle \wedge y \in \langle -1, 2\rangle \Rightarrow f(y) = \frac{1}{3}.$$

Je-li

$$y \in \langle 0, 2\rangle \wedge -y \in \langle -1, 2\rangle \Rightarrow f(-y) = \frac{1}{3}.$$

Tedy

$$\text{pro} \quad y \in \langle 0, 2\rangle \quad \text{je} \quad f(y) = \frac{1}{3},$$
$$\text{pro} \quad y \in \langle 0, 1\rangle \quad \text{je} \quad f(-y) = \frac{1}{3}.$$

Odtud

$$g(y) = \begin{cases} \frac{2}{3} & \text{pro } y \in \langle 0, 1\rangle \\ \frac{1}{3} & \text{pro } y \in (1, 2\rangle \\ 0 & \text{jinak} \end{cases}.$$

♣

Pro hustotu transformované spojité náhodné veličiny platí věta analogická větě 7.1.

Věta 7.2. Nechť X a Y jsou dvě spojité náhodné veličiny a nechť náhodná veličina Y vznikla transformací h náhodné veličiny X, tj. $Y = h(X)$. Označme Ω_X [Ω_Y] obor hodnot náhodné veličiny X [Y] a f [g] hustotu náhodné veličiny X [Y].

Je-li funkce h prostá na množině Ω_X a má-li uvnitř množiny Ω_X spojitou první derivaci, potom pro hustotu g náhodné veličiny Y platí

$$g(y) = \begin{cases} f\big(h^{-1}(y)\big) \left|\dfrac{d}{dy}h^{-1}(y)\right| & \text{pro } y \in \Omega_Y \\ 0 & \text{jinak} \end{cases},$$

kde h^{-1} je inverzní funkce k funkci h.

Důkaz. Označme F distribuční funkci náhodné veličiny X, g $[G]$ hustotu [distribuční funkci] náhodné veličiny Y. Hustotu a distribuční funkci náhodné veličiny Y stačí určit pro $y \in \Omega_Y$. Předpokládejme, že funkce h je na množině Ω_X klesající. Potom pro $y \in \Omega_Y$ platí

$$\begin{aligned} G(y) &= P(Y \le y) = P\big(h(X) \le y\big) = P\big(X \ge h^{-1}(y)\big) \\ &= 1 - P\big(X < h^{-1}(y)\big) = 1 - F\big(h^{-1}(y)\big). \end{aligned}$$

Odtud

$$\begin{aligned} g(y) &= \frac{d}{dy}G(y) = \frac{d}{dy}\big[1 - F\big(h^{-1}(y)\big)\big] = -\frac{d}{dy}F\big(h^{-1}(y)\big) \\ &= \left|x = h^{-1}(y)\right| = -\frac{d}{dx}F(x)\,\frac{d}{dy}h^{-1}(y) = -f(x)\,\frac{d}{dy}h^{-1}(y) \qquad (6) \\ &= -f\big(h^{-1}(y)\big)\,\frac{d}{dy}h^{-1}(y) \quad \text{pro } y \in \Omega_Y. \end{aligned}$$

Protože je h klesající, je i h^{-1} klesající (této skutečnosti jsme využili již výše) a tedy

$$\frac{d}{dy}h^{-1}(y) < 0 \Rightarrow \left|\frac{d}{dy}h^{-1}(y)\right| = -\frac{d}{dy}h^{-1}(y)$$

Odtud a ze vztahu (6) dostáváme

$$g(y) = \begin{cases} f\big(h^{-1}(y)\big) \left|\dfrac{d}{dy}h^{-1}(y)\right| & \text{pro } y \in \Omega_Y \\ 0 & \text{jinak} \end{cases}.$$

Důkaz pro h rostoucí je analogický. □

Příklad 7.5. Řešte předchozí příklady pomocí věty 7.2, pokud to lze.

Řešení. V příkladu 7.2 vznikla náhodná veličina Z transformací $z = h(t) = \sqrt{r^2 - 4t^2}$ náhodné veličiny T. Jedná se o funkci prostou na $\Omega_T = \langle 0, r/2\rangle$, která je na Ω_T klesající. Lze tedy použít větu 7.2. Inverzní funkci h^{-1} k funkci h dostaneme, když z rovnice

$$z = h(t) = \sqrt{r^2 - 4t^2} \quad \text{pro } t \in \left\langle 0, \frac{r}{2}\right\rangle$$

vyjádříme t. Dostaneme

$$z^2 = r^2 - 4t^2 \Rightarrow t^2 = \frac{1}{4}\left(r^2 - z^2\right) \Rightarrow t_{1,2} = \pm\frac{1}{2}\sqrt{r^2 - z^2}.$$

Protože $t \in \langle 0, r/2\rangle$, máme

$$t = \frac{1}{2}\sqrt{r^2 - z^2} = h^{-1}(z).$$

Potom podle věty 7.2 pro $z \in \Omega_Z$ platí

$$\begin{aligned} Z \sim g(z) &= f\big(h^{-1}(z)\big)\left|\frac{d}{dz}h^{-1}(z)\right| \\ &= f\Big(\frac{1}{2}\sqrt{r^2 - z^2}\Big)\left|\frac{d}{dz}\Big(\frac{1}{2}\sqrt{r^2 - z^2}\Big)\right| \\ &= \left|-\frac{1}{2}\,\frac{z}{\sqrt{r^2 - z^2}}\right| f\Big(\frac{1}{2}\sqrt{r^2 - z^2}\Big) \\ &= \frac{1}{2}\,\frac{z}{\sqrt{r^2 - z^2}}\, f\Big(\frac{1}{2}\sqrt{r^2 - z^2}\Big). \end{aligned}$$

Další postup je stejný jako v příkladu 7.2.

V příkladech 7.3 a 7.4 není transformační funkce h prostá na oboru hodnot transformované náhodné veličiny X, nelze tedy přímo použít větu 7.2.

♣

Dále se budeme zabývat transformací náhodného vektoru. Uvažujme tedy náhodný vektor $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ s oborem hodnot $\Omega_{\boldsymbol{X}}$. Nechť

$$\begin{aligned} y_1 &= h_1(x_1, x_2, \ldots, x_n) \\ y_2 &= h_2(x_1, x_2, \ldots, x_n) \\ &\vdots \\ y_m &= h_m(x_1, x_2, \ldots, x_n) \end{aligned}$$

je m reálných funkcí n reálných proměnných definovaných na množině $\Omega_{\boldsymbol{X}}$. Nechť

$$\begin{aligned} Y_1 &= h_1(X_1, X_2, \ldots, X_n) \\ Y_2 &= h_2(X_1, X_2, \ldots, X_n) \\ &\vdots \\ Y_m &= h_m(X_1, X_2, \ldots, X_n) \end{aligned} \tag{7}$$

potom řekneme, že náhodný vektor $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_m)$ vznikl transformací $\boldsymbol{h} = (h_1, h_2, \ldots, h_m)$ náhodného vektoru $\boldsymbol{X}$.

Budeme se zabývat pouze situací, kdy jsou náhodné vektory $\boldsymbol{X}$ a $\boldsymbol{Y}$ spojité. Budeme předpokládat, že známe hustotu f náhodného vektoru $\boldsymbol{X}$ a budeme chtít určit hustotu g náhodného vektoru $\boldsymbol{Y}$. Obecný postup je následující:

1. Vyjdeme z definičního vztahu distribuční funkce G náhodného vektoru $\boldsymbol{Y}$, tj.

$$\begin{aligned} \boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_m) &\sim G(y_1, y_2, \ldots, y_m) \\ &= P(Y_1 \le y_1, Y_2 \le y_2, \ldots, Y_m \le y_m). \end{aligned} \tag{8}$$

Pravděpodobnost ve vztahu (8) vyjádříme pomocí náhodného vektoru $\boldsymbol{X}$, tj. veličiny $Y_1, Y_2, \ldots, Y_m$ vyjádříme pomocí vztahů (7). Potom

$$G(y_1, y_2, \ldots, y_m)$$
$$= P\Big(h_1(X_1, \ldots, X_n) \leq y_1, h_2(X_1, \ldots, X_n) \leq y_2, \ldots, h_m(X_1, \ldots, X_n) \leq y_m\Big).$$

Označme nyní

$$A = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : \; h_1(x_1, \ldots, x_n) \leq y_1, \ldots, h_m(x_1, \ldots, x_n) \leq y_m\}.$$

Dostaneme

$$G(y_1, y_2, \ldots, y_m) = \int \cdots \int_A f(x_1, x_2, \ldots, x_n)\, dx_1\, dx_2 \ldots dx_n.$$

2. Pro hustotu g náhodného vektoru $\boldsymbol{Y}$ pak zřejmě platí

$$g(y_1, y_2, \ldots, y_m) = \frac{\partial^m}{\partial y_1 \partial y_2 \ldots \partial y_m} G(y_1, y_2, \ldots, y_m).$$

Příklad 7.6. Určete hustotu délky průvodiče bodu B, když jeho souřadnice (X_1, X_2) mají hustotu

$$f(x_1, x_2) = \frac{1}{2\pi a^2}\, \mathrm{e}^{-\dfrac{x_1^2 + x_2^2}{2a^2}} \qquad \text{pro} \;\; (x_1, x_2) \in \mathbb{R}^2.$$

Řešení. Pro délku R průvodiče bodu B zřejmě platí

$$R = \sqrt{X_1^2 + X_2^2}.$$

(Nakreslete si obrázek!)
Tedy R je spojitá náhodná veličina s oborem hodnot $\Omega_R = \langle 0, \infty)$, která vznikla transformací $h = h(x_1, x_2) = \sqrt{x_1^2 + x_2^2}$ náhodného vektoru (X_1, X_2).

Označme g $[G]$ hustotu [distribuční funkci] náhodné veličiny R. Potom pro $r \geq 0$ dostaneme

$$G(r) = P(R \leq r) = P\Big(\sqrt{X_1^2 + X_2^2} \leq r\Big).$$

Označme

$$A = \Big\{(x_1, x_2) \in \mathbb{R}^2 : \; x_1^2 + x_2^2 \leq r^2\Big\}.$$

Potom

$$P\Big(\sqrt{X_1^2 + X_2^2} \leq r\Big) = \iint_A f(x_1, x_2)\, dx_1\, dx_2 = \iint_A \frac{1}{2\pi a^2}\, \mathrm{e}^{-\dfrac{x_1^2 + x_2^2}{2a^2}}\, dx_1\, dx_2.$$

Pro výpočet integrálu použijeme substituci do polárních souřadnic, tj.

$$\begin{aligned} x_1 &= \rho\cos\varphi & \qquad A^*: \quad & 0 \le \rho \le r \\ x_2 &= \rho\sin\varphi & & 0 \le \varphi \le 2\pi \\ |J| &= \rho \end{aligned}$$

Odtud

$$\begin{aligned} \iint\limits_{A} \frac{1}{2\pi a^2}\, e^{-\dfrac{x_1^2+x_2^2}{2a^2}}\, dx_1\, dx_2 &= \frac{1}{2\pi a^2} \iint\limits_{A^*} \rho\, e^{-\dfrac{\rho^2}{2a^2}}\, d\rho\, d\varphi \\ &= \frac{1}{2\pi a^2} \int_0^r \Big[\int_0^{2\pi} \rho\, e^{-\dfrac{\rho^2}{2a^2}}\, d\varphi\Big] d\rho = \frac{1}{2\pi a^2} \int_0^r \rho\, e^{-\dfrac{\rho^2}{2a^2}} \Big[\varphi\Big]_0^{2\pi} d\rho \\ &= \frac{1}{a^2} \int_0^r \rho\, e^{-\dfrac{\rho^2}{2a^2}}\, d\rho. \end{aligned}$$

Výpočtem posledního integrálu (zavedením substituce $t = -\rho^2/(2a^2)$) dostaneme

$$G(r) = 1 - e^{-\dfrac{r^2}{2a^2}} \quad \text{pro} \;\; r \ge 0.$$

Z posledního vztahu pak vyjádříme hustotu g náhodné veličiny R. Zřejmě pro $r \ge 0$ je

$$g(r) = \frac{d}{dr} G(r) = \frac{r}{a^2}\, e^{-\dfrac{r^2}{2a^2}}.$$

Tedy

$$g(r) = \begin{cases} \dfrac{r}{a^2}\, e^{-\dfrac{r^2}{2a^2}} & \text{pro} \;\; r \ge 0 \\ 0 & \text{jinak.} \end{cases}$$

Jedná se o tzv. Rayleighovo rozdělení.

♣

Výpočet hustoty náhodného vektoru, který vznikl transformací spojitého náhodného vektoru, může podstatně zjednodušit následující věta, která je analogická větě 7.2.

Věta 7.3. Nechť $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ a $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)$ jsou dva spojité náhodné vektory a nechť náhodný vektor $\boldsymbol{Y}$ vznikl transformací $h = (h_1, h_2, \ldots, h_n)$ náhodného vektoru $\boldsymbol{X}$, tj.

$$\begin{aligned} Y_1 &= h_1(X_1, X_2, \ldots, X_n) \\ Y_2 &= h_2(X_1, X_2, \ldots, X_n) \\ &\;\vdots \\ Y_n &= h_n(X_1, X_2, \ldots, X_n) \end{aligned}$$

Označme $\Omega_{\boldsymbol{X}}$ [$\Omega_{\boldsymbol{Y}}$] obor hodnot náhodného vektoru $\boldsymbol{X}$ [$\boldsymbol{Y}$] a f [g] hustotu náhodného vektoru $\boldsymbol{X}$ [$\boldsymbol{Y}$].

I. Předpokládejme, že soustava rovnic

$$\begin{aligned} y_1 &= h_1(x_1, x_2, \ldots, x_n) \\ y_2 &= h_2(x_1, x_2, \ldots, x_n) \\ &\vdots \\ y_n &= h_n(x_1, x_2, \ldots, x_n) \end{aligned}$$

má pro každé $(y_1, y_2, \ldots, y_n) \in \Omega_{\boldsymbol{Y}}$ právě jedno řešení $(x_1, x_2, \ldots, x_n)$. Označme pro toto řešení

$$\begin{aligned} x_1 &= h_1^{-1}(y_1, y_2, \ldots, y_n) \\ x_2 &= h_2^{-1}(y_1, y_2, \ldots, y_n) \\ &\vdots \\ x_n &= h_n^{-1}(y_1, y_2, \ldots, y_n) \end{aligned}$$

II. Předpokádejme, že funkce $h_1^{-1}, h_2^{-1}, \ldots, h_n^{-1}$ mají uvnitř $\Omega_{\boldsymbol{Y}}$ spojité parciální derivace prvního řádu a že determinant (tzv. jakobián)

$$J(y_1, y_2, \ldots, y_n) = \det \begin{bmatrix} \frac{\partial}{\partial y_1} h_1^{-1} & \frac{\partial}{\partial y_2} h_1^{-1} & \cdots & \frac{\partial}{\partial y_n} h_1^{-1} \\ \frac{\partial}{\partial y_1} h_2^{-1} & \frac{\partial}{\partial y_2} h_2^{-1} & \cdots & \frac{\partial}{\partial y_n} h_2^{-1} \\ \vdots & & \ddots & \vdots \\ \frac{\partial}{\partial y_1} h_n^{-1} & \frac{\partial}{\partial y_2} h_n^{-1} & \cdots & \frac{\partial}{\partial y_n} h_n^{-1} \end{bmatrix}$$

je pro každé $(y_1, y_2, \ldots, y_n) \in \Omega_{\boldsymbol{Y}}$ nenulový.

Potom

$$g(y_1, \ldots, y_n) = \begin{cases} f\Big(h_1^{-1}(y_1, \ldots, y_n), \ldots, h_n^{-1}(y_1, \ldots, y_n)\Big) \Big| J(y_1, \ldots, y_n) \Big| & \text{pro } (y_1, \ldots, y_n) \in \Omega_{\boldsymbol{Y}} \\ 0 & \text{jinak} \end{cases}.$$

Příklad 7.7. Náhodný vektor (X_1, X_2) má hustotu

$$f(x_1, x_2) = \begin{cases} x_1 e^{-x_1 - x_2} & \text{pro } (x_1, x_2) \in (0, \infty) \times (0, \infty) \\ 0 & \text{jinak} \end{cases}.$$

Nechť

$$Y_1 = \frac{X_1}{X_1 + X_2}, \quad Y_2 = X_1 + X_2.$$

Určete hustotu náhodného vektoru (Y_1, Y_2) a zjistěte, zda jsou veličiny Y_1 a Y_2 nezávislé.

Řešení. Náhodný vektor (Y_1, Y_2) je zřejmě spojitý náhodný vektor, který vznikl transformací $\boldsymbol{h} = (h_1, h_2)$ náhodného vektoru (X_1, X_2), kde

$$\begin{aligned} y_1 &= h_1(x_1, x_2) = \frac{x_1}{x_1 + x_2} \\ y_2 &= h_2(x_1, x_2) = x_1 + x_2. \end{aligned} \tag{9}$$

Ověříme, zda lze pro výpočet hustoty g náhodného vektoru (Y_1, Y_2) použít větu 7.3.

I. Obor hodnot $\Omega_{(Y_1,Y_2)}$ náhodného vektoru (Y_1, Y_2) je zřejmě množina

$$\Omega_{(Y_1,Y_2)} = \left\{(y_1, y_2) \in \mathbb{R}^2 : \ 0 < y_1 < 1, \ 0 < y_2 < \infty\right\}.$$

Soustava rovnic (9) má zřejmě pro každé $(y_1, y_2) \in \Omega_{(Y_1,Y_2)}$ právě jedno řešení (x_1, x_2), kde

$$\begin{aligned} x_1 &= y_1 y_2 \\ x_2 &= y_2 - y_1 y_2. \end{aligned}$$

Označme v souladu s větou 7.3

$$\begin{aligned} x_1 &= y_1 y_2 = h_1^{-1}(y_1, y_2) \\ x_2 &= y_2 - y_1 y_2 = h_2^{-1}(y_1, y_2). \end{aligned}$$

II. Funkce h_1^{-1}, h_2^{-1} mají spojité parciální derivace prvního řádu v $\mathbb{R}^2$ a tedy i na množině $\Omega_{(Y_1,Y_2)}$.
Zřejmě platí

$$J(y_1, y_2) = \det \begin{bmatrix} \dfrac{\partial}{\partial y_1} h_1^{-1} & \dfrac{\partial}{\partial y_2} h_1^{-1} \\ \dfrac{\partial}{\partial y_1} h_2^{-1} & \dfrac{\partial}{\partial y_2} h_2^{-1} \end{bmatrix} = \det \begin{bmatrix} y_2 & y_1 \\ -y_2 & 1 - y_1 \end{bmatrix} = y_2.$$

Tedy

$$J(y_1, y_2) \neq 0 \quad \text{pro každé } (y_1, y_2) \in \Omega_{(Y_1,Y_2)}.$$

Pro výpočet hustoty g náhodného vektoru (Y_1, Y_2) lze tedy použít větu 7.3. Odtud pro $(y_1, y_2) \in \Omega_{(Y_1,Y_2)}$ dostáváme

$$\begin{aligned} g(y_1, y_2) &= f\Big(h_1^{-1}(y_1, y_2), h_2^{-1}(y_1, y_2)\Big) \Big|J(y_1, y_2)\Big| \\ &= f\big(y_1 y_2, y_2 - y_1 y_2\big) \big|y_2\big| = y_1 y_2^2 \mathrm{e}^{-y_2}, \end{aligned}$$

tj.

$$g(y_1, y_2) = \begin{cases} y_1 y_2^2 \mathrm{e}^{-y_2} & \text{pro } (y_1, y_2) \in (0,1) \times (0, \infty) \\ 0 & \text{jinak} \end{cases}.$$

Zbývá ověřit, zda jsou náhodné veličiny Y_1 a Y_2 stochasticky nezávislé. Podle definice nezávislosti (viz definice 6.1) jsou veličiny Y_1 a Y_2 nezávislé, když je rozdělovací funkce náhodného vektoru (Y_1, Y_2) rovna součinu rozdělovacích funkcí veličin Y_1 a Y_2. Označme tedy g_1 [g_2] rozdělovací funkci náhodné veličiny Y_1 [Y_2]. Podle kapitoly 5 dostaneme

$$g_1(y_1) = \begin{cases} \int\limits_{-\infty}^{\infty} g(y_1, y_2) dy_2 = \int\limits_{0}^{\infty} y_1 y_2^2 \,\mathrm{e}^{-y_2}\, dy_2 = 2y_1 & \text{pro } y_1 \in (0,1) \\ 0 & \text{jinak} \end{cases},$$

$$g_2(y_2) = \begin{cases} \int\limits_{-\infty}^{\infty} g(y_1, y_2) dy_1 = \int\limits_{0}^{1} y_1 y_2^2 \,\mathrm{e}^{-y_2}\, dy_1 = \frac{1}{2} y_2^2 \,\mathrm{e}^{-y_2} & \text{pro } y_2 \in (0,\infty) \\ 0 & \text{jinak} \end{cases}.$$

Platí tedy

$$g(y_1, y_2) = g_1(y_1)\, g_2(y_2) \quad \text{pro každé } (y_1, y_2) \in \mathbb{R}^2,$$

tj. veličiny Y_1 a Y_2 jsou nezávislé.

♣

Příklad 7.8. Určete hustotu náhodné veličiny $Y = X_1 + X_2$, kde X_1 a X_2 jsou nezávislé náhodné veličiny, které mají rovnoměrné rozdělení s parametry 0 a 1.

Řešení. Víme, že pro $i = 1, 2$ platí

$$X_i \sim f_i(x_i) = \begin{cases} 1 & \text{pro } x_i \in \langle 0, 1\rangle \\ 0 & \text{jinak} \end{cases}$$

(viz poznámka 3.3).
Vzhledem k nezávislosti veličin X_1 a X_2 dostaneme, že náhodný vektor (X_1, X_2) má hustotu

$$f(x_1, x_2) = f_1(x_1) f_2(x_2) = \begin{cases} 1 & \text{pro } (x_1, x_2) \in \langle 0,1\rangle \times \langle 0,1\rangle \\ 0 & \text{jinak} \end{cases}.$$

Náhodná veličina Y je zřejmě spojitá náhodná veličina s oborem hodnot $\Omega_Y = \langle 0, 2\rangle$, která vznikla transformací $h = h(x_1, x_2) = x_1 + x_2$ náhodného vektoru (X_1, X_2). Máme určit hustotu náhodné veličiny Y. Vzhledem k tomu, že transformovaný vektor má oproti původnímu pouze jednu složku, nelze přímo použít větu 7.3. Budeme tedy postupovat stejně jako v příkladu 7.6. Na závěr pak ukážeme, že lze i v tomto případě vhodným zavedením další proměnné použít i větu 7.3.

Označme tedy g [G] hustotu [distribuční funkci] náhodné veličiny Y. Pro $y \in \Omega_Y = \langle 0, 2\rangle$ dostáváme

$$G(y) = P\big(Y \le y\big) = P\big(X_1 + X_2 \le y\big).$$

Označme

$$A = \big\{(x_1, x_2) \in \mathbb{R}^2 :\ x_1 + x_2 \le y\big\}.$$

Potom

$$G(y) = \iint\limits_{A} f(x_1, x_2)\, dx_1 dx_2.$$

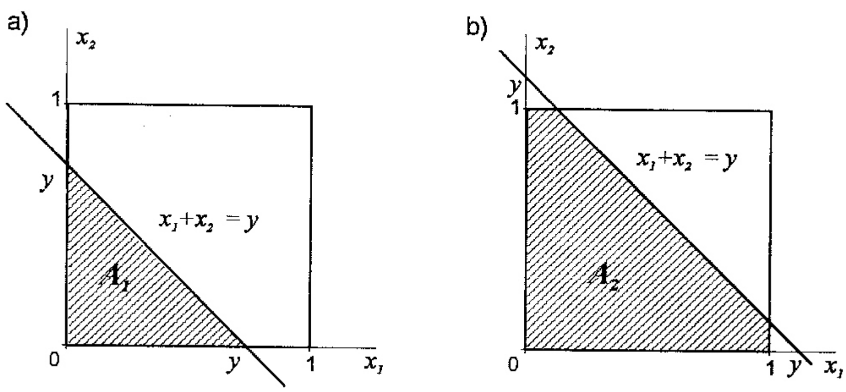
PDF-Xchange
Click to buy NOW!
www.docu-track.com

Pro $y \in \langle 0, 1\rangle$ dostaneme (viz obrázek 7.1 a))

$$G(y) = \iint\limits_{A_1} 1\, dx_1 dx_2 = \frac{y^2}{2}.$$

Je-li $y \in (1, 2\rangle$, potom (viz obrázek 7.1 b))

$$G(y) = \iint\limits_{A_2} 1\, dx_1 dx_2 = 1 - \frac{(2-y)^2}{2}.$$



Obrázek 7.1

Odtud

$$g(y) = \frac{d}{dy} G(y) = \begin{cases} y & \text{pro } \; y \in \langle 0, 1\rangle \\ 2 - y & \text{pro } \; y \in (0, 2\rangle \end{cases}$$

a tedy

$$g(y) = \begin{cases} y & \text{pro } \; y \in \langle 0, 1\rangle \\ 2 - y & \text{pro } \; y \in (0, 2\rangle \, . \\ 0 & \text{jinak} \end{cases}$$

Pro využití věty 7.3 zavedeme náhodnou veličinu $Y_1 = X_1$. Potom náhodný vektor (Y_1, Y) vznikl transformací $\boldsymbol{h} = (h_1, h_2)$ náhodného vektoru (X_1, X_2). kde

(10)
$$\begin{aligned} y_1 &= h_1(x_1, x_2) = x_1 \\ y &= h_2(x_1, x_2) = x_1 + x_2 \end{aligned}$$

Určíme-li hustotu $q(y_1, y)$ náhodného vektoru (Y_1, Y), potom pro hustotu g náhodné veličiny Y zřejmě platí

$$g(y) = \int_{-\infty}^{\infty} q(y_1, y)\, dy_1 \quad \text{pro } \; y \in \Omega_Y.$$

Ověříme tedy, zda lze pro určení hustoty q náhodného vektoru (Y_1, Y) použít větu 7.3.

Obor hodnot $\Omega_{(Y_1,Y)}$ náhodného vektoru (Y_1, Y) je zřejmě množina

$$\Omega_{(Y_1,Y)} = \{(y_1, y) \in \mathbb{R}^2 : \ 0 \le y_1 \le 1, y_1 \le y \le 1 + y_1)\}.$$

Řešením soustavy (10) dostáváme

$$\begin{aligned} x_1 &= y_1 = h_1^{-1}(x_1, x_2) \\ x_2 &= y - y_1 = h_2^{-1}(x_1, x_2) \end{aligned}.$$

Vzhledem k tomu, že

$$J(y_1, y) = \det \begin{bmatrix} \frac{\partial}{\partial y_1} h_1^{-1} & \frac{\partial}{\partial y} h_1^{-1} \\ \frac{\partial}{\partial y_1} h_2^{-1} & \frac{\partial}{\partial y} h_2^{-1} \end{bmatrix} = \det \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} = 1,$$

lze použít větu 7.3. Potom pro $(y_1, y) \in \Omega_{(Y_1,Y)}$ je

$$q(y_1, y) = f\Big(h_1^{-1}(y_1, y), h_2^{-1}(y_1, y)\Big)\Big|J(y_1, y)\Big| = f(y_1, y - y_1) = 1.$$

Odtud pro $y \in \Omega_Y$ dostáváme

$$g(y) = \int_{-\infty}^{\infty} q(y_1, y) dy_1 = \begin{cases} \int_0^y 1\, dy_1 = y & \text{pro } \ y \in \langle 0, 1\rangle \\ \int_{y-1}^1 1\, dy_1 = 2 - y & \text{pro } \ y \in (1, 2\rangle \end{cases}.$$

(Nakreslete si obrázek $\Omega_{(Y_1,Y)}$!)
Dostáváme tedy samozřejmě stejný výsledek jako při obecném postupu. ♣

PDF-XChange Click to buy NOW! www.docu-track.com

8. Číselné charakteristiky náhodných veličin a ~~vektorů~~

Známe-li rozdělovací funkci náhodné veličiny X, víme o ní z pravděpodobnostního hlediska vše. Často je zapotřebí shrnout informaci o pravděpodobnostním chování náhodné veličiny X do několika čísel tzv. číselných charakteristik náhodné veličiny.

K nejčastěji používaným číselným chrakteristikám náhodné veličiny patří její střední hodnota a rozptyl, jejichž definice nyní uvedeme.

Definice 8.1. *Buď X spojitá [diskrétní] náhodná veličina s rozdělovací funkcí g a oborem hodnot Ω. Číslo $E(X)$ (pokud existuje) definované vztahem*

$$E(X) = \int_{\Omega} x\, g(x)\, dx \tag{8.1}$$

$$\left[E(X) = \sum_{x\in\Omega} x\, g(x)\right] \tag{8.2}$$

se nazývá střední nebo očekávaná hodnota náhodné veličiny X.

Číslo $D(X)$ (pokud existuje) definované vztahem

$$D(X) = E\left([X - E(X)]^2\right) \tag{8.3}$$

se nazývá rozptyl nebo disperse náhodné veličiny X.

Poznámka 8.1.

1) Pro střední hodnotu a rozptyl náhodné veličiny X se často používá označení μ a σ^2.

2) Střední hodnota a rozptyl nemusí existovat. Např. náhodná veličina X, která má hustotu

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

tzv. Cauchyovo rozdělení), nemá ani střední hodnotu ani rozptyl.

3) Střední hodnotu náhodné veličiny X si můžeme představit jako $x-$ovou souřadnici těžiště hmotného tělesa Ω, jehož měrná hmotnost je rovna rozdělovací funkci g.

4) Střední hodnota náhodné veličiny X patří k tzv. charakteristikám polohy (pokud existuje). Je to jakýsi „střed" rozdělení, okolo kterého kolísají realizace náhodné veličiny X.

PDF-Xchange Click to buy NOW! www.docu-track.com

5) Rozptyl náhodné veličiny X patří k tzv. charakteristikám variability. Náhodná veličina $[X - E(X)]^2$ představuje kvadratickou odchylku náhodné veličiny X od čísla $E(X)$, rozptyl $D(X)$ je střední hodnota této odchylky. Odtud plyne: Čím je $D(X)$ větší, tím více jsou realizace náhodné veličiny X rozptýlenější okolo $E(X)$. Rozptyl $D(X)$ je tedy míra rozptýlenosti realizací náhodné veličiny X okolo její střední hodnoty $E(X)$.

6) Vzhledem k tomu, že náhodná veličina $[X - E(X)]^2$ nabývá pouze nezáporných hodnot a $D(X)$ je střední hodnota této náhodné veličiny, musí platit $D(X) \geq 0$. Má tedy smysl $\sqrt{D(X)}$. Číslo $\sqrt{D(X)}$ nazýváme směrodatná odchylka náhodné veličiny X. Směrodatná odchylka náhodné veličiny X patří samozřejmě také k charakteristikám variability. Důvod, proč se v technických aplikacích používá častěji než rozptyl, je ten, že má stejné jednotky jako náhodná veličina X. Často se také používá tzv. variační koeficient náhodné veličiny X, tj. číslo

$$V(X) = \frac{\sqrt{D(X)}}{E(X)}.$$

Jedná se o bezrozměrnou charakteristiku variability, která vynásobena stem udává, kolik procent střední hodnoty představuje směrodatná odchylka.

Příklad 8.1. Náhodná veličina X má rozdělovací funkci

$$\text{a) } g(x) = \begin{cases} \dfrac{x}{3} & \text{pro } x \in \{1,2\} \\ 0 & \text{jinak} \end{cases},$$

$$\text{b) } g(x) = \begin{cases} 3x^2 & \text{pro } x \in (0,1\rangle \\ 0 & \text{jinak} \end{cases}.$$

Určete střední hodnotu náhodné veličiny X.

Řešení. V obou případech vyjdeme z definice 8.1. Musíme zjistit, zda se jedná o diskrétní nebo spojitou náhodnou veličinu.

a) Obor hodnot Ω náhodné veličiny X je množina $\{1,2\}$, jedná se tedy o diskrétní náhodnou veličinu. Potom podle vztahu (8.1) dostaneme:

$$E(X) = \sum_{x \in \Omega} x\, g(x) = \sum_{x=1}^{2} x \cdot \frac{1}{3} x = \frac{1}{3} \sum_{x=1}^{2} x^2 = \frac{5}{3}.$$

b) Obor hodnot Ω náhodné veličiny X je množina $(0,1\rangle$, jedná se tedy o spojitou náhodnou veličinu. Podle vztahu (8.2) je

$$E(X) = \int_{\Omega} x\, g(x)\, dx = \int_0^1 x \cdot 3x^2\, dx = 3 \int_0^1 x^3\, dx = \frac{3}{4}.$$

♣

Příklad 8.2. Automatická linka vyrábí při normálním nastavení zmetek s pravděpodobností p. Seřízení linky se provede vždy, když vyrobí zmetek. Určete střední hodnotu počtu součástek, které linka vyrobí mezi dvěma po sobě jdoucími seřízeními.

Řešení. Nechť X je počet součástek vyrobených mezi dvěma po sobě jdoucími seřízeními. Obor hodnot Ω náhodné veličiny X je množina $\{1, 2, \dots\}$. X je tedy diskrétní náhodná veličina. Pro pravděpodobnostní funkci g náhodné veličiny X zřejmě platí

$$g(x) = P(X = x) = \begin{cases} (1-p)^{x-1}p & \text{pro } x = 1, 2, \dots \\ 0 & \text{jinak} \end{cases}.$$

Potom

$$\begin{aligned} E(X) &= \sum_{x \in \Omega} x\, g(x) = \sum_{x=1}^{\infty} x\,(1-p)^{x-1} p \\ &= p\left[\sum_{x=0}^{\infty}(1-p)^x + \sum_{x=1}^{\infty}(1-p)^x + \sum_{x=2}^{\infty}(1-p)^x + \dots\right]. \end{aligned}$$

Na základě součtu geometrických řad s kvocientem $q = 1 - p$ dostaneme

$$\begin{aligned} E(X) &= p\left[\frac{1}{1-(1-p)} + \frac{1-p}{1-(1-p)} + \frac{(1-p)^2}{1-(1-p)} + \dots\right] \\ &= p\left[\frac{1}{p} + \frac{1-p}{p} + \frac{(1-p)^2}{p} + \dots\right] \\ &= 1 + (1-p) + (1-p)^2 + \dots = \frac{1}{1-(1-p)} = \frac{1}{p}. \end{aligned}$$

Je-li tedy např. $p = 0.01$, lze očekávat, že linka vyrobí mezi dvěma seřízeními 100 součástek.

♣

Uvažujme nyní náhodnou veličinu X s rozdělovací funkcí g a nechť náhodná veličina Y vznikla transformací h náhodné veličiny X, tj. $Y = h(X)$. Střední hodnotu náhodné veličiny Y lze počítat přímo z její definice tak, že nejprve ze známé rozdělovací funkce g náhodné veličiny X určíme rozdělovací funkci r veličiny Y. Snadnější je však postup, který umožňuje následující věta, uváděná bez důkazu.

Věta 8.1. Buď X spojitá [diskrétní] náhodná veličina s rozdělovací funkcí g a oborem hodnot Ω. Uvažujme funkci $h(X)$ náhodné veličiny X. Potom

$$E\big(h(X)\big) = \int_{\Omega} h(x)\, g(x)\, dx \tag{8.4}$$

$$\left[E\big(h(X)\big) = \sum_{x \in \Omega} h(x)\, g(x)\right]. \tag{8.5}$$

Poznámka 8.2. Větu 8.1 lze vyslovit i pro náhodný vektor $(X_1, X_2, \dots, X_n)$. My zde uvedeme pouze případ $n = 2$. Případ, kdy $n > 2$, ponecháváme čtenáři.

PDF-Xchange Click to buy NOW! www.docu-track.com

Věta 8.2. Buď (X_1, X_2) spojitý [diskrétní] náhodný vektor s rozdělovací funkcí g a oborem hodnot Ω. Uvažujme funkci $h(X_1, X_2)$ náhodných veličin X_1, X_2. Potom

$$E\big(h(X_1, X_2)\big) = \iint\limits_{\Omega} h(x_1, x_2)\, g(x_1, x_2)\, dx_1 dx_2 \tag{8.6}$$

$$\Big[E\big(h(X_1, X_2)\big) = \sum_{(x_1,x_2)\in\Omega} h(x_1, x_2)\, g(x_1, x_2)\Big]. \tag{8.7}$$

Poznámka 8.3.

1) Výše uvedené věty patří k nejvyužívanějším tvrzením o střední hodnotě náhodné veličiny; usnadňují totiž podstatně výpočty.

2) Z těchto vět vyplývá následující věta o pravidlech pro výpočet střední hodnoty a rozptylu náhodných veličin.

Věta 8.3. Jsou-li X a Y náhodné veličiny s konečnou střední hodnotou a konečným rozptylem a a, b reálná čísla, potom platí:

1. $E(aX + b) = aE(X) + b$ (speciálně $E(b) = b$).
2. $E(aX + bY) = aE(X) + bE(Y)$.
3. Jsou-li X a Y nezávislé, potom $E(XY) = E(X)E(Y)$.
4. $D(X) = E(X^2) - [E(X)]^2$.
5. $D(aX + b) = a^2 D(X)$ (speciálně $D(b) = 0$).
6. Jsou-li X a Y nezávislé, potom $D(aX+bY) = a^2 D(X) + b^2 D(Y)$.

Důkaz: Větu dokážeme pro spojité náhodné veličiny X a Y. Pro diskrétní náhodné veličiny by byl důkaz analogický, pouze místo integrálů bychom pracovali s řadami. Nejprve dokážeme pravidla 1, 4 a 5. Nechť X má hustotu g a obor hodnot Ω.

1. Položme ve větě 8.1 $h(X) = aX + b$, potom

$$\begin{aligned} E(aX + b) &= \int\limits_{\Omega} (ax + b)\, g(x)\, dx = a \int\limits_{\Omega} x\, g(x)\, dx + b \int\limits_{\Omega} g(x)\, dx \\ &= aE(X) + b \cdot 1 = aE(X) + b. \end{aligned}$$

Je-li $a = 0$, dostaneme $E(b) = b$.

4. Podle definice 8.1 je

$$D(X) = E\left([X - E(X)]^2\right).$$

Položme nyní ve větě 8.1 $h(X) = [X - E(X)]^2$, potom

$$\begin{aligned} D(X) &= \int\limits_{\Omega} [x - E(X)]^2\, g(x)\, dx = \int\limits_{\Omega} \left(x^2 - 2xE(X) + [E(X)]^2\right) g(x)\, dx \\ &= \int\limits_{\Omega} x^2\, g(x)\, dx - 2E(X) \int\limits_{\Omega} x\, g(x)\, dx + [E(X)]^2 \int\limits_{\Omega} g(x)\, dx \\ &= E(X^2) - 2E(X)E(X) + [E(X)]^2 = E(X^2) - [E(X)]^2. \end{aligned}$$

PDF-Xchange Click to buy NOW! www.docu-track.com

5. Vyjdeme z definice rozptylu a využijeme již dokázanou vlastnost 1, dostaneme

$$\begin{aligned} D(aX+b) &= E\big([aX+b-E(aX+b)]^2\big) = E\big([aX+b-aE(X)-b]^2\big) \\ &= E\big([aX-aE(X)]^2\big) = a^2E([X-E(X)]^2) = a^2D(X). \end{aligned}$$

Je-li $a=0$, dostaneme $D(b)=0$.

Zbývající pravidla 2, 3 a 6 dokážeme analogicky, pouze místo náhodné veličiny X uvažujeme náhodný vektor (X,Y) s hustotou g a oborem hodnot Ω a využijeme větu 8.2.

2. Podle vztahu (8.6) je

$$\begin{aligned} E(aX+bY) &= \iint_\Omega (ax+by)\,g(x,y)\,dxdy \\ &= a\iint_\Omega x\,g(x,y)\,dxdy + b\iint_\Omega y\,g(x,y)\,dxdy = aE(X)+bE(Y). \end{aligned}$$

3. Označme $g_1[g_2]$ hustotu náhodné veličiny X $[Y]$. Vzhledem k tomu, že jsou veličiny X a Y nezávislé, musí podle věty 6.1 platit

$$g(x,y) = g_1(x)\,g_2(y) \quad \text{pro každé } (x,y)\in\mathbb{R}^2.$$

Odtud a z věty 8.2 dostáváme

$$\begin{aligned} E(XY) &= \iint_\Omega xy\,g(x,y)\,dxdy = \iint_\Omega xy\,g_1(x)\,g_2(y)\,dxdy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\,g_1(x)\,g_2(y)\,dxdy = \int_{-\infty}^{\infty} x\,g_1(x)\,dx \int_{-\infty}^{\infty} y\,g_2(y)dy \\ &= E(X)E(Y). \end{aligned}$$

Vztah 6 bychom dokázali postupně využitím pravidel 5, 2 a 3 této věty. Důkaz ponecháváme čtenáři. □

Poznámka 8.4.

1) Pro výpočet rozptylu náhodné veličiny bývá výhodnější místo definice 8.1 použít vzorec 4 z věty 8.3.

2) Pravidla 2. a 6. z věty 8.3 lze zobecnit. Jsou-li $X_1, X_2, \ldots, X_n$ náhodné veličiny, pro které existují číselné charakteristiky dále uvedené a $a_1, a_2, \ldots, a_n$ reálná čísla, potom

a) $E\Big(\sum\limits_{i=1}^n a_iX_i\Big) = \sum\limits_{i=1}^n a_iE(X_i)$,

b) $D\Big(\sum\limits_{i=1}^n a_iX_i\Big) = \sum\limits_{i=1}^n a_i^2D(X_i)$ pro $X_1, X_2, \ldots, X_n$ nezávislé.

Příklad 8.3. Vypočítejte rozptyl náhodné veličiny X z příkladu 8.1.

Řešení. S využitím poznámky 8.4 1) a věty 8.1 dostaneme

$$\text{a) } D(X) = E(X^2) - [E(X)]^2 = \sum_{x\in\Omega} x^2\, g(x) - \left(\frac{5}{3}\right)^2 = \sum_{x=1}^{2} x^2 \cdot \frac{1}{3}x - \frac{25}{9} = \frac{2}{9}.$$

$$\text{b) } D(X) = E(X^2) - [E(X)]^2 = \int_\Omega x^2\, g(x)\, dx - \left(\frac{3}{4}\right)^2 = \int_0^1 x^2 \cdot 3x^2\, dx - \frac{9}{16}$$

$$= \frac{3}{80}.$$

♣

Příklad 8.4. Uvažujme nezávislé náhodné veličiny X a Y, pro které platí

$$E(X) = 1,\ D(X) = 2,\ E(Y) = -2,\ D(Y) = 3.$$

Určete střední hodnotu a rozptyl náhodných veličin $2X+1$, $X+2Y$, $X-Y$.

Řešení. Ve všech případech vyjdeme z věty 8.3. Postupně dostaneme

$$\begin{aligned}
&E(2X+1) = 2E(X) + 1 = 3,\\
&D(2X+1) = 4D(X) = 8,\\
&E(X+2Y) = E(X) + 2E(Y) = -3,\\
&D(X+2Y) = D(X) + 4D(Y) = 14,\\
&E(X-Y) = E\left(X + (-1)Y\right) = E(X) - E(Y) = 3,\\
&D(X-Y) = D\left(X + (-1)Y\right) = D(X) + D(Y) = 5.
\end{aligned}$$

♣

Poznámka 8.5. Uvědomme si, že jsme pro výpočet střední hodnoty součtu náhodných veličin, na rozdíl od výpočtu rozptylu, nepotřebovali informaci o nezávislosti těchto veličin.

Příklad 8.5. Buď X náhodná veličina s konečnou střední hodnotou $E(X) = \mu$ a konečným nenulovým rozptylem $D(X) = \sigma^2$. Nechť

$$U = \frac{X-\mu}{\sigma}.$$

Dokažte, že platí: $E(U) = 0$, $D(U) = 1$.

Řešení. Podle věty 8.3 postupně dostáváme

$$\begin{aligned}
&E(U) = \frac{1}{\sigma}\, E(X-\mu) = \frac{1}{\sigma}\left[E(X) - \mu\right] = 0,\\
&D(U) = \frac{1}{\sigma^2}\, D(X-\mu) = \frac{1}{\sigma^2}\, D(X) = \frac{1}{\sigma^2}\,\sigma^2 = 1.
\end{aligned}$$

♣

Poznámka 8.6. Náhodná veličina, která má střední hodnotu nula a rozptyl jedna se nazývá normovaná náhodná veličina. Příklad 8.5 říká, že každou náhodnou veličinu, která má konečný nenulový rozptyl a konečnou střední hodnotu, lze normovat.

Příklad 8.6. Buď $X_1, X_2, \ldots, X_n$ nezávislé náhodné veličiny se stejnou střední hodnotou μ a stejným rozptylem σ^2. Buď

$$M = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Dokažte, že platí: $E(M) = \mu$, $D(M) = \sigma^2/n$.

Řešení. S využitím poznámky 8.4 2) postupně dostáváme:

$$E(M) = \frac{1}{n}E\Big(\sum_{i=1}^{n} X_i\Big) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \mu,$$

$$D(M) = \frac{1}{n^2}D\Big(\sum_{i=1}^{n} X_i\Big) = \frac{1}{n^2}\sum_{i=1}^{n} D(X_i) = \frac{\sigma^2}{n}.$$

♣

Poznámka 8.7. Jestliže n-krát nezávisle a se stejnou přesností σ^2 měříme tutéž veličinu μ, pak má podle předchozího příkladu aritmetický průměr těchto měření střední hodnotu μ a rozptyl σ^2/n.

Příklad 8.7. Přesně vypočtená částka daně je zaokrouhlena na celé koruny směrem nahoru. Odhadněte částku, která by byla získána navíc, pokud by se místo na celé koruny zaokrouhlovalo nahoru na celé desetikoruny při pěti milionech daňových poplatníků.

Řešení. Označme pro $i = 1, 2, \ldots, 5$ mil.

X_i = nezaokrouhlená daň i-tého poplatníka,

Y_i = daň i-tého poplatníka zaokrouhlená nahoru na celé koruny,

Z_i = daň i-tého poplatníka zaokrouhlená nahoru na celé desetikoruny.

Potom $Z_i - Y_i$ je částka, která by se získala navíc u i-tého poplatníka a $E(Z_i - Y_i)$ je očekávaná hodnota této částky. Zřejmě

$$Z_i - Y_i = (Z_i - X_i) - (Y_i - X_i)$$

a tedy

$$E(Z_i - Y_i) = E[(Z_i - X_i)] - E[(Y_i - X_i)].$$

Náhodné veličiny $(Z_i - X_i)$ a $(Y_i - X_i)$ mají rovnoměrné rozdělení s parametry 0, 10 a 0, 1 (viz poznámka 3.3). Tedy $E(Z_i - X_i) = 5, E(Y_i - X_i) = 0.5$. Odtud $E(Z_i - Y_i) = 4.5$. Očekávané zvýšení při pěti milionech poplatníků by tedy bylo

$$E\Big(\sum_{i=1}^{5\text{ mil.}} (Z_i - Y_i)\Big) = \sum_{i=1}^{5\text{ mil.}} E[(Z_i - Y_i)] = 22500000 \text{ [Kč]}.$$

♣

Příklad 8.8. Celková doba provozu vysavače v domácnosti během roku (při jednotce 100 hodin) je náhodná veličina $Y = 0.7X^2$, kde náhodná veličina X má hustotu

$$g(x) = \begin{cases} x & \text{pro } x \in (0,1\rangle \\ 2-x & \text{pro } x \in (1,2\rangle \\ 0 & \text{jinak} \end{cases}.$$

Určete střední hodnotu a směrodatnou odchylku doby provozu vysavače.

Řešení. Obor hodnot Ω náhodné veličiny X je množina $(0,2\rangle$, jedná se tedy o spojitou náhodnou veličinu. Obor hodnot náhodné veličiny Y je množina $(0, 2.8\rangle$.
Podle věty 8.3 platí

$$E(Y) = E(0.7X^2) = 0.7E(X^2).$$

Z věty 8.1 dostáváme

$$E(X^2) = \int_{\Omega} x^2\, g(x)\, dx = \int_0^1 x^2 \cdot x\, dx + \int_1^2 x^2 \cdot (2-x)\, dx = 1.16\overline{6},$$

odtud

$$E(Y) = 0.81\overline{6}.$$

Tedy střední doba provozu je přibližně 82 hodin.

Podobně podle věty 8.3 je

$$D(Y) = D(0.7X^2) = 0.49D(X^2).$$

Z věty 8.3 a 8.1 dostaneme

$$D(X^2) = E(X^4) - \left[E(X^2)\right]^2 = E(X^4) - 1.16\overline{6}^2,$$

$$E(X^4) = \int_{\Omega} x^4\, g(x)\, dx = \int_0^1 x^4 \cdot x\, dx + \int_1^2 x^4 \cdot (2-x)\, dx \doteq 2.06\overline{6}.$$

Odtud

$$D(X^2) = 0.70\overline{5} \Rightarrow D(Y) \doteq 0.346 \Rightarrow \sqrt{D(Y)} \doteq 0.588.$$

Tedy směrodatná odchylka doby provozu je přibližně 59 hodin. ♣

Příklad 8.9. Náhodný vektor (X,Y) má hustotu z příkladu 6.4 c) a b). Určete $E(XY)$, $E(Y)$, $D(Y)$.

Řešení.

c) Vyjdeme z věty 8.2, dostaneme

$$E(XY) = \iint_{\Omega_{(X,Y)}} xy\, g(x,y)\, dx dy = \int_1^2 \left[\int_0^1 xy \cdot 4x(y-1)\, dx\right] dy = \frac{10}{9}.$$

Střední hodnotu $E(Y)$ a rozptyl $D(Y)$ můžeme určit dvěma způsoby.

I. způsob: Nejprve vyjádříme hustotu g_2 náhodné veličiny Y, potom

$$E(Y) = \int\limits_{\Omega_Y} y\, g_2(y)\, dy,$$

$$D(Y) = E(Y^2) - [E(Y)]^2 = \int\limits_{\Omega_Y} y^2\, g_2(y)\, dy - [E(Y)]^2.$$

V příkladu 6.4 jsme dostali

$$Y \sim g_2(y) = \begin{cases} 2(y-1) & \text{pro } y \in (1,2) \\ 0 & \text{jinak} \end{cases},$$

odtud

$$E(Y) = \int\limits_1^2 y \cdot 2(y-1)\, dy = \frac{5}{3},$$

$$D(Y) = \int\limits_1^2 y^2 \cdot 2(y-1)\, dy - \left(\frac{5}{3}\right)^2 = \frac{1}{18}.$$

II. způsob: Nebudeme vyjadřovat hustotu náhodné veličiny Y, ale vyjdeme z věty 8.2 a dané číselné charakteristiky budeme počítat přímo. Dostaneme

$$E(Y) = \iint\limits_{\Omega} y\, g(x,y)\, dx dy = \int\limits_1^2 \Big[\int\limits_0^1 y \cdot 4x(y-1)\, dx\Big]\, dy = \frac{5}{3},$$

$$D(Y) = E(Y^2) - [E(Y)]^2 = \iint\limits_{\Omega} y^2\, g(x,y)\, dx dy - \left(\frac{5}{3}\right)^2$$

$$= \int\limits_1^2 \Big[\int\limits_0^1 y^2 \cdot 4x(y-1)\, dx\Big]\, dy - \left(\frac{5}{3}\right)^2 = \frac{1}{18}.$$

b) Postupujeme stejně jako v c), ale vzhledem k tomu, že se jedná o diskrétní náhodný vektor počítáme místo integrálů součty. Dostaneme

$$E(XY) = \sum_{(x,y)\in\Omega_{(X,Y)}} xy\, g(x,y) = \sum_{x=0}^{1}\sum_{y=1}^{3} xy\, g(x,y)$$

$$= 1\, g(1,1) + 2\, g(1,2) + 3\, g(1,3) = 1.0.$$

Při výpočtu $E(Y)$ a $D(Y)$ můžeme opět postupovat dvěma způsoby.

I. způsob: Pravděpodobnostní funkci g_2 náhodné veličiny Y máme vypočítanou v tabulce 6.3. Odtud

$$E(Y) = \sum_{y\in\Omega_Y} y\, g_2(y) = \sum_{y=1}^{3} y\, g_2(y) = 1.7,$$

$$D(Y) = E(Y^2) - [E(Y)]^2 = \sum_{y\in\Omega_Y} y^2\, g_2(y) - 1.7^2 = 0.61.$$

II. způsob: Podle věty 8.2 dostaneme

$$
\begin{aligned}
E(Y) &= \sum_{(x,y)\in\Omega_{(X,Y)}} y\, g(x,y) \\
&= g(0,1) + g(1,1) + 2\, g(0,2) + 2\, g(1,2) + 3\, g(0,3) + 3\, g(1,3) = 1.7, \\
D(Y) &= E(Y^2) - [E(Y)]^2 = \sum_{(x,y)\in\Omega_{(X,Y)}} y^2\, g(x,y) - 1.7^2 \\
&= \sum_{x=0}^{1}\sum_{y=1}^{3} y^2\, g(x,y) - 1.7^2 = 0.61.
\end{aligned}
$$

♣

Poznámka 8.8.

1) Někdy se zavádí řada dalších charakteristik. V definičních vztazích budeme vždy předpokládat existenci, konečnost, případně nenulovost potřebných charakteristik. Charakteristika

$$\mu_k = \mu_k(X) = E(X^k), \quad \text{pro} \quad k = 1, 2, \dots$$

se nazývá <u>k-tý obecný moment</u> náhodné veličiny X. Charakteristika

$$\mu_k^0 = \mu_k^0(X) = E\big([X - E(X)]^k\big), \quad \text{pro} \quad k = 1, 2, \dots$$

se nazývá <u>k-tý centrální moment</u> náhodné veličiny X. Charakteristika

$$\gamma_1 = \gamma_1(X) = \frac{\mu_3^0}{\sqrt{[D(X)]^3}}$$

se nazývá <u>koeficient šikmosti (asymetrie)</u> náhodné veličiny X. Charakteristika

$$\gamma_2 = \gamma_2(X) = \frac{\mu_4^0}{\sqrt{[D(X)]^4}} - 3$$

se nazývá <u>koeficient špičatosti (excesu)</u> náhodné veličiny X.

Výpočet těchto číselných charakteristik zjednodušuje fakt, že k-tý centrální moment lze vyjádřit pomocí j-tých obecných momentů, $j = 1, \dots, k$. Zřejmě

$$
\begin{aligned}
\mu_1 &= E(X), \\
\mu_2^0 &= E\big([X - E(X)]^2\big) = D(X).
\end{aligned}
$$

Potom např.

$$
\begin{aligned}
\mu_1^0 &= E\big(X - E(X)\big) = 0, \\
\mu_2^0 &= E(X^2) - [E(X)]^2 = \mu_2 - {\mu_1}^2, \\
\mu_3^0 &= E\big([X - E(X)]^3\big) = \mu_3 - 3\mu_1\mu_2 + 2{\mu_1}^3, \\
\mu_4^0 &= E\big([X - E(X)]^4\big) = \mu_4 - 4\mu_3\mu_1 + 6{\mu_1}^2\mu_2 - 3{\mu_1}^3.
\end{aligned}
$$

PDF-XChange Click to buy NOW! www.docu-track.com

Odvození posledních dvou vztahů plyne z pravidel o počítání středních hodnot (věta 8.3) a známých vzorců $(a-b)^3$, $(a-b)^4$. Pro výpočet koeficientu šikmosti a špičatosti stačí spočítat j-tý obecný moment náhodné veličiny X pro $j = 1, 2, 3, 4$.

2) Koeficient šikmosti má následující význam: Pro symetrické rozdělení platí $\gamma_1(X) = 0$. Je-li rozdělení protáhlejší směrem napravo než směrem nalevo je $\gamma_1(X) > 0$ a naopak pro rozdělení protáhlejší směrem nalevo než napravo je $\gamma_1(X) < 0$.

Koeficient špičatosti má následující význam: Pro normální rozdělení (viz definice 9.2.1) je $\gamma_2(X) = 0$. Má-li veličina X symetrické rozdělení a je-li $\gamma_1(X) > 0$ $[\gamma_1(X) < 0]$, znamená to, že na svých koncích je rozdělovací funkce větší [menší] než hustota normálního rozdělení se stejnou střední hodnotou a rozptylem.

3) Číselné charakteristiky, které lze vyjádřit jako funkce obecných momentů, se nazývají momentové charakteristiky. Všechny dosud probrané charakteristiky jsou momentové.

Další velmi důležitou charakteristikou je tzv. $100\,\alpha$ procentní kvantil náhodné veličiny X, jehož definici pro spojitou náhodnou veličinu uvedeme.

Definice 8.2. *Buď X spojitá náhodná veličina s distribuční funkcí F, $\alpha \in (0,1)$. Číslo $x(\alpha)$ se nazývá $100\,\alpha$ procentní kvantil náhodné veličiny X, jestliže platí*

$$F(x(\alpha)) = \alpha. \tag{8.8}$$

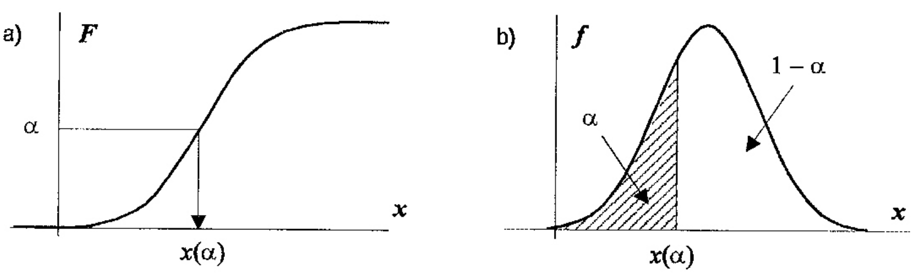
Poznámka 8.9.

1) Z definice $100\,\alpha$ procentního kvantilu vyplývá, že

$$\alpha = F(x(\alpha)) = P(X \le x(\alpha)).$$

Tedy v průměru pro $100\,\alpha\%$ realizací náhodné veličiny X platí, že jsou menší nebo rovny číslu $x(\alpha)$ a zbývajících $100(1-\alpha)\%$ realizací je větších nebo rovných tomuto číslu.

2) Nechť X má distribuční funkci $F(x)$, potom geometrické znázornění $100\,\alpha$ procentního kvantilu ukazuje obrázek 8.1 a).



Obrázek 8.1

PDF-XChange Click to buy NOW! www.docu-track.com

3) Nechť X má hustotu f, potom pro $100\,\alpha$ procentní kvantil platí

$$\alpha = F(x(\alpha)) = P(X \leq x(\alpha)) = \int_{-\infty}^{x(\alpha)} f(x)\,dx. \tag{8.9}$$

Tedy přímka $x = x(\alpha)$ rozděluje oblast ohraničenou hustotou f a osou x na dvě části, jejichž obsahy jsou v poměru $\alpha : (1-\alpha)$ (viz obrázek 8.1 b)).

4) Některé kvanily mají speciální názvy. Nejčastěji se používají

$x(0.5)$	50%-ní kvantil	tzv. medián
$x(0.25)$	25%-ní kvantil	tzv. dolní kvartil
$x(0.75)$	75%-ní kvantil	tzv. horní kvartil .

Rozdíl $x(0.75) - x(0.25)$ se nazývá mezikvartilová odchylka.

5) Kvantily patří k charakteristikám polohy, mezikvartilová odchylka patří k charakteristikám variability. Obě dvě charakteristiky nejsou momentové.

6) S kvantily rozdělení budeme pracovat v matematické statistice. Tam se kvantily nepočítají, neboť ty nejdůležitější jsou buďto tabelovámy (viz [13]) nebo je počítá každý statistický software.

Příklad 8.10. Náhodná veličina X má

a) distribuční funkci

$$F(x) = \begin{cases} 0 & \text{pro } \ x \leq 1 \\ (x-1)^4 & \text{pro } \ 1 < x < 2\,, \\ 1 & \text{pro } \ 2 \leq x \end{cases}$$

b) hustotu

$$f(x) = \begin{cases} 1 - |x| & \text{pro } \ x \in (-1, 1) \\ 0 & \text{jinak} \end{cases}.$$

Určete medián, dolní a horní kvartil a mezikvartilové rozpětí náhodné veličiny X.

Řešení.

a) Hledané kvantily musí být v intervalu $(1, 2)$. Odtud pro medián $x(0.5)$ musí podle vztahu (8.8) platit

$$0.5 = F(x(0.5)) = [x(0.5) - 1]^4.$$

Rovnice

$$[x(0.5) - 1]^4 = 0.5$$

má pouze dva reálné kořeny (zbývající dva jsou komplexně sdružené). Potom

$$x_{1,2}(0.5) - 1 = \pm\sqrt[4]{0.5} \Rightarrow x_{1,2} = 1 \pm \sqrt[4]{0.5} \Rightarrow x_{1,2} \doteq \begin{cases} 1.841 \\ 0.159 \end{cases}.$$

Druhý kořen není z intervalu $(1, 2)$, tedy medián je 1.841.

Podobně dostaneme

$x(0.25) \doteq 1.707,$

$x(0.75) \doteq 1.931.$

Mezikvartilové rozpětí je 0.224.

b) Hledané kvantily musí ležet v intervalu $(-1, 1)$. Vzhledem k tomu, že hustota f je symetrická podle osy y je $x(0.5) = 0$. Pro horní kvartil $x(0.75)$ podle vztahu (8.9) platí

$$0.75 = F\big((x(0.75)\big) = \int_{-\infty}^{x(0.75)} f(x)\,dx = \int_{-1}^{x(0.75)} (1 - |x|)\,dx$$

$$= 0.5 + \int_{0}^{x(0.75)} (1 - x)\,dx = 0.5 + \left[x - \frac{x^2}{2}\right]_0^{x(0.75)}$$

$$= 0.5 + \left[x(0.75) - \frac{x^2(0.75)}{2}\right] \Rightarrow 0.25 = x(0.75) - \frac{x^2(0.75)}{2} \Rightarrow$$

$$\Rightarrow x^2(0.75) - 2x(0.75) + 0.5 = 0 \Rightarrow x_{1,2}(0.75) = \frac{2 \pm \sqrt{2}}{2} \doteq \begin{cases} 1.707 \\ 0.293 \end{cases}.$$

Tedy $x(0.75) = 0.293$.

Podobně $x(0.25) = -0.293$. Přitom dolní kvartil jsme nemuseli počítat, protože vzhledem k symetričnosti rozdělení musí platit $x(0.25) = -x(0.75)$. Navíc jsme v tomto triviálním případě nemuseli ani integrovat, ale výsledek dostaneme porovnáváním obsahů vhodných trojúhelníků.

♣

Poznámka 8.10. Další často používanou charakteristikou náhodné veličiny X je tzv. modus nebo modální hodnota náhodné veličiny X. Tuto charakteristiku značíme $\mathrm{Mo}(X) = x^\circ$.

1. Jestliže náhodná veličina X je diskrétní s pravděpodobnostní funkcí g, pak její modus je číslo x°, pro které platí

$$g(x^\circ) = \max_{x \in \mathbb{R}} g(x).$$

Modus x° diskrétní náhodné veličiny X je tedy nejpravděpodobnější hodnota náhodné veličiny X.

2. Jestliže náhodná veličina X je spojitá s hustotou g, pak její modus je číslo x° (pokud existuje), ve kterém funkce $g(x)$ nabývá lokálního maxima.

Má-li náhodná veličina právě jeden modus, pak se její rozdělení nazývá modální. Má-li náhodná veličina více než jeden modus, mluvíme o polymodálním rozdělení.

Např. rovnoměrné rozdělení (viz poznámka 3.3) je polymodální a náhodné veličiny z příkladu 3.1 mají modální rozdělení.

V kapitole 6 jsme zavedli pojem stochastické nezávislosti, resp. závislosti. S pojmem stochastické závislosti dvou náhodných veličin souvisí čísla zvaná kovariance a koeficient korelace. Těmto charakteristikám se budeme věnovat ve zbývající části kapitoly.

PDF-XChange
Click to buy NOW!
www.docu-track.com

Definice 8.3. *Buď X, Y náhodné veličiny. Číslo $C(X, Y)$ (pokud existuje) definované vztahem*

$$C(X,Y) = E\big([X - E(X)][Y - E(Y)]\big) \tag{8.10}$$

se nazývá kovariance náhodných veličin X, Y.

Je-li $C(X, Y) \neq 0$, pak náhodné veličiny X, Y nazýváme korelačně závislými (korelovanými).

Je-li $C(X, Y) = 0$, pak veličiny X, Y nazýváme korelačně nezávislými (nekorelovanými).

Věta 8.4. Buď X, Y náhodné veličiny, pro které existují charakteristiky dále uvedené, a a b reálná čísla, potom platí následující vztahy:

1. $C(X, Y) = E(XY) - E(X)E(Y)$.
2. $D(aX + bY) = a^2 D(X) + b^2 D(Y) + 2ab\, C(X, Y)$.

Důkaz:

1. Zřejmě je

$$\begin{aligned} C(X,Y) &= E\big([X - E(X)][Y - E(Y)]\big) \\ &= E\big(XY - XE(Y) - E(X)Y + E(X)E(Y)\big) \\ &= E(XY) - E(X)E(Y) - E(X)E(Y) + E(X)E(Y) \\ &= E(XY) - E(X)E(Y). \end{aligned}$$

2. Pro rozptyl $D(aX + bY)$ platí

$$\begin{aligned} D(aX + bY) &= E\big([(aX + bY) - E(aX + bY)]^2\big) \\ &= E\big([aX + bY - aE(X) - bE(Y)]^2\big) \\ &= E\big([a(X - E(X)) + b(Y - E(Y))]^2\big) \\ &= E\big(a^2[X - E(X)]^2 + b^2[Y - E(Y)]^2 + 2ab\,[X - E(X)][Y - E(Y)]\big) \\ &= a^2 E\big([X - E(X)]^2\big) + b^2 E\big([Y - E(Y)]^2\big) \\ &+ 2ab\, E\big([X - E(X)][Y - E(Y)]\big) \\ &= a^2 D(X) + b^2 D(Y) + 2ab\, C(X, Y). \end{aligned}$$

Poznamenejme, že pro stochasticky nezávislé veličiny je $C(X, Y) = 0$, takže výsledný vzorec je vztah 6 ve větě 8.3. $\square$

Poznámka 8.11.

1) Ze vztahu (8.10) plyne

$$C(X, Y) = C(Y, X),$$
$$C(X, X) = D(X).$$

2) Pro výpočet kovariance bývá výhodnější místo definice 8.3 využít vzorec 1 z věty 8.4.

3) Jsou-li veličiny X, Y stochasticky nezávislé, potom podle věty 8.3 vztahu 3 je $E(XY) = E(X)E(Y)$ a tedy ze vztahu 1 věty 8.4 dostáváme $C(X, Y) = 0$. Jsou-li tedy náhodné veličiny X, Y stochasticky nezávislé, potom jsou korelačně nezávislé. Opak obecně neplatí, tj. jsou-li veličiny X, Y korelačně nezávislé, mohou ale nemusí být stochasticky nezávislé (viz příklad 8.11). Jsou-li ale veličiny X, Y korelačně závislé (tj. $C(X, Y) \neq 0$), potom jsou veličiny X, Y stochasticky závislé.

Příklad 8.11. Uvažujme náhodný vektor (X, Y), jehož hustota je

$$g(x, y) = \begin{cases} \dfrac{1}{\pi} & \text{pro } (x, y) \in \Omega_{(X,Y)} \\ 0 & \text{jinak} \end{cases},$$

kde $\Omega_{(X,Y)} = \{(x, y) \in \mathbb{R};\ x^2 + y^2 \leq 1\}$. Ukažte, že jsou náhodné veličiny X, Y stochasticky závislé, i když jsou korelačně nezávislé.

Řešení. V příkladu 5.3 jsme určili hustoty g_1, g_2 marginálních náhodných veličin X, Y:

$$X \sim g_1(x) = \begin{cases} \dfrac{2}{\pi}\sqrt{1 - x^2} & \text{pro } x \in \langle -1, 1 \rangle \\ 0 & \text{jinak} \end{cases},$$

$$Y \sim g_2(y) = \begin{cases} \dfrac{2}{\pi}\sqrt{1 - y^2} & \text{pro } y \in \langle -1, 1 \rangle \\ 0 & \text{jinak} \end{cases}.$$

V příkladu 6.2 jsme ukázali, že jsou veličiny X, Y stochasticky závislé. Ukážeme, že jsou však korelačně nezávislé. Především $E(X) = 0, E(Y) = 0$, protože hustoty g_1, g_2 náhodných veličin X, Y jsou symetrické podle příslušných souřadnicových os x, y. Potom je

$$C(X, Y) = E(XY) = \iint\limits_{\Omega_{(X,Y)}} xy\, g(x, y)\, dx dy = \frac{1}{\pi} \iint\limits_{\Omega_{(X,Y)}} xy\, dx dy.$$

Vzhledem k tomu, že je funkce $z = xy$ symetrická podle počátku (0,0,0) souřadnic, je poslední integrál nulový. Veličiny X, Y jsou tedy korelačně nezávislé.

♣

Definice 8.4. *Buď X, Y náhodné veličiny. Číslo $\rho(X, Y)$ (pokud existuje) definované vztahem*

$$\rho(X, Y) = \frac{C(X, Y)}{\sqrt{D(X)D(Y)}} \tag{8.11}$$

se nazývá korelační koeficient (koeficient korelace) náhodných veličin X, Y.

Číslo $\rho^2(X, Y)$ se nazývá koeficient determinace.

Příklad 8.12. Buď X, Y náhodné veličiny, určete konstantu k tak, aby náhodná veličina $U = Y - kX$ měla minimální rozptyl.

Řešení. Ze vztahu 2 věty 8.4 dostáváme

$$D(U) = D(Y - kX) = D(Y) + k^2 D(X) - 2k\, C(X,Y).$$

Protože

$$C(X,Y) = \rho(X,Y)\sqrt{D(X)D(Y)},$$

můžeme předchozí výsledek psát ve tvaru

$$D(U) = k^2 D(X) - 2k\,\rho(X,Y)\sqrt{D(X)D(Y)} + D(Y).$$

Funkce

$$f(k) = k^2 D(X) - 2k\,\rho(X,Y)\sqrt{D(X)D(Y)} + D(Y)$$

představuje vzhledem k parametru k polynom druhého stupně. Minima může nabýt pouze pro taková čísla k, pro něž $f'(k) = 0$, nebo-li pro která

$$2k\,D(X) - 2\,\rho(X,Y)\sqrt{D(X)D(Y)} = 0.$$

Odtud pro $D(X) \neq 0$ je

$$k = \rho(X,Y)\sqrt{\frac{D(Y)}{D(X)}}.$$

Protože

$$f''(k) = 2D(X) > 0,$$

jde skutečně o minimum. Příslušný minimální rozptyl $D_{\min}(U)$ tedy je

$$\begin{aligned} D_{\min}(U) &= \rho^2(X,Y)D(Y) + D(Y) - 2\,\rho^2(X,Y)D(Y) \\ &= D(Y)[1 - \rho^2(X,Y)]. \end{aligned} \tag{8.12}$$

♣

Vlastnosti korelačního koeficientu uvádí následující věta.

Věta 8.5. Buď X, Y náhodné veličiny s konečnou střední hodnotou a konečným nenulovým rozptylem. Potom platí:

1. $|\rho(X,Y)| \leq 1$.
2. $|\rho(X,Y)| = 1 \Leftrightarrow Y = kX + q$ pro nějaká čísla $k, q, k \neq 0$.
3. Náhodné veličiny X, Y jsou korelačně závislé právě tehdy, když $\rho(X,Y) \neq 0$. Přitom pro $\rho(X,Y) > 0$ mluvíme o kladné korelaci, kdežto při $\rho(X,Y) < 0$ o záporné korelaci náhodých veličin X, Y.

Důkaz:

1. První tvrzení plyne ze vztahu (8.12). Protože $D_{\min}(U) \geq 0, D(Y) > 0$ musí být $1 - \rho^2(X,Y) \geq 0$, odtud $|\rho(X,Y)| \leq 1$.

2. Je-li $\rho^2(X,Y) = 1$, pak z téhož vzorce plyne $D_{\min}(U) = 0$, neboli náhodná veličina U je konstantní, tj. $Y - kX = q$, kde $q = E(U)$, odtud plyne tvrzení.

3. Třetí tvrzení plyne z definice 8.3 a 8.4.

Tím je věta dokázána. □

Poznámka 8.12. Z předchozí věty plyne, že je korelační koeficient mírou lineární závislosti dvou veličin.

Příklad 8.13. V cisterně je na začátku dne $Y \cdot 1000$ l petroleje, kde Y je náhodná veličina. Během dne se odebere z cisterny $X \cdot 1000$ l petroje, kde X je opět náhodná veličina, přičemž se cisterna nedoplňuje. Předpokládejme, že náhodný vektor (X, Y) má hustotu

$$g(x,y) = \begin{cases} 2 & \text{pro } 0 < x < y < 1 \\ 0 & \text{jinak} \end{cases}.$$

Určete, zda jsou veličiny X a Y lineárně závislé a v případě, že nejsou lineárně závislé, ověřte, zda jsou nezávislé.

Řešení. Podle vlastnosti 2 z věty 8.5 jsou veličiny X, Y lineárně závislé právě když

$$|\rho(X,Y)| = 1.$$

Budeme tedy počítat korelační koeficient $\rho(X,Y)$ náhodných veličin X, Y. Vyjdeme z definice 8.4, tj.

$$\rho(X,Y) = \frac{C(X,Y)}{\sqrt{D(X)D(Y)}}.$$

Pro výpočet $D(X), D(Y), C(X,Y)$ použijeme výpočtové vzorce

$$D(X) = E(X^2) - [E(X)]^2,$$

$$D(Y) = E(Y^2) - [E(Y)]^2,$$

$$C(X,Y) = E(XY) - E(X)E(Y).$$

Musíme tedy vypočítat

$$E(X),\ E(X^2),\ E(Y),\ E(Y^2),\ E(XY).$$

Obor hodnot $\Omega_{(X,Y)}$ náhodného vektoru (X, Y) je množina

$$\{(x,y) \in \mathbb{R}^2;\ 0 < x < 1, x < y < 1\},$$

jedná se tedy o spojitý náhodný vektor. Označme postupně g_1, g_2 hustoty náhodných veličin X, Y a Ω_X, Ω_Y obory hodnot těchto veličin. Podle kapitoly 6 platí

$$X \sim g_1(x) = \begin{cases} \displaystyle\int\limits_{\{y \in \mathbb{R};\ g(x,y) \neq 0\}} g(x,y)\, dy & \text{pro } x \in \Omega_X \\ 0 & \text{jinak} \end{cases},$$

tedy

$$X \sim g_1(x) = \begin{cases} \displaystyle\int_x^1 2\, dy = 2(1-x) & \text{pro } x \in (0,1) \\ 0 & \text{jinak} \end{cases}.$$

Podobně

$$Y \sim g_2(y) = \begin{cases} \int\limits_0^y 2\,dx = 2y & \text{pro } y \in (0,1) \\ 0 & \text{jinak} \end{cases}.$$

Potom

$$E(X) = \int\limits_{\Omega_X} x\,g_1(x)\,dx = \int\limits_0^1 x \cdot 2(1-x)\,dx = \frac{1}{3},$$

$$E(X^2) = \int\limits_{\Omega_X} x^2\,g_1(x)\,dx = \int\limits_0^1 x^2 \cdot 2(1-x)\,dx = \frac{3}{18},$$

$$E(Y) = \int\limits_{\Omega_Y} y\,g_2(y)\,dx = \int\limits_0^1 y \cdot 2y\,dx = \frac{2}{3},$$

$$E(Y^2) = \int\limits_{\Omega_Y} y^2\,g_2(y)\,dx = \int\limits_0^1 y^2 \cdot 2y\,dx = \frac{9}{18},$$

$$E(XY) = \iint\limits_{\Omega_{(X,Y)}} xy\,g(x,y)\,dxdy = \int\limits_0^1 \Big[\int\limits_x^1 xy \cdot 2\,dy\Big]\,dx = \frac{1}{4}.$$

Odtud

$$D(X) = \frac{1}{18},\ D(Y) = \frac{1}{18},\ C(XY) = \frac{1}{36}.$$

Potom

$$\rho(X,Y) = \frac{1}{2}.$$

Protože $|\rho(X,Y)| \neq 1$, nejsou veličiny X, Y lineárně závislé. Protože $\rho(X,Y) \neq 0$, jsou veličiny X, Y stochasticky závislé.

♣

Příklad 8.14. Ověřte, zda jsou veličiny X, Y lineárně závislé. V případě, že nejsou lineárně závislé, ověřte zda jsou stochasticky nezávislé. Náhodný vektor (X, Y) má rozdělovací funkci danou tabulkou

a)

$x\backslash y$	2	5
0	0.2	0
1	0	0.8

b)

$x\backslash y$	0	1	2
0	0.42	0.12	0.06
1	0.28	0.08	0.04

Řešení. V obou případech postupujeme analogicky jako v příkladu 8.13. Vzhledem k tomu, že se jedná o diskrétní náhodné vektory budeme místo integrování sečítat.

a) Obor hodnot $\Omega_{(X,Y)}$ náhodného vektoru (X,Y) je množina $\{(0,2),(1,5)\}$. Označme postupně g_1, g_2 pravděpodobnostní funkce náhodných veličin X, Y a příslušné obory hodnot Ω_X, Ω_Y. Pravděpodobností fukci g_1, g_2 dostaneme dopočítáním v tabulce (viz kapitola 5), dostaneme tabulku 8.1

$x \backslash y$	2	5	$g_1(x)$
0	0.2	0	0.2
1	0	0.8	0.8
$g_2(y)$	0.2	0.8	1.0

Tabulka 8.1

Potom

$$E(X) = \sum_{x\in\Omega_X} x\, g_1(x) = \sum_{x=0}^{1} x\, g_1(x) = 0.8,$$

$$E(X^2) = \sum_{x\in\Omega_X} x^2\, g_1(x) = \sum_{x=0}^{1} x^2\, g_1(x) = 0.8.$$

Podobně

$$E(Y) = 4.4,$$

$$E(Y^2) = 20.08.$$

Dále

$$E(XY) = \sum_{(x,y)\in\Omega} xy\, g(x,y) = \sum_{x=0}^{1}\sum_{y=2}^{3} xy\, g(x,y) = 4.$$

Odtud

$$D(X) = 0.16,\ D(Y) = 1.44,\ C(X,Y) = 0.48.$$

Tedy

$$\rho(X,Y) = 1.$$

Protože $|\rho(X,Y)| = 1$, dostáváme, že jsou veličiny X, Y lineárně závislé, přičemž se jedná o přímou závislost.

b) Postupujeme stejně jako v a), dostaneme $\rho(X,Y) = 0$. Protože $|\rho(X,Y)| \neq 1$, nejsou veličiny X, Y lineárně závislé. Tyto veličiny jsou nekorelované a tudíž z korelačního koeficientu nelze určit, zda jsou nezávislé. Protože platí (ověřte)

$$g(x,y) = g_1(x)\, g_2(y) \quad \text{pro každé } (x,y) \in \mathbb{R}^2,$$

jsou podle definice nezávislosti (viz definice 6.1) veličiny X, Y nezávislé.

♣

Příklad 8.15. Investor rozložil svoje prostředky rovným dílem do dvou projektů. Náhodné veličiny X_1, X_2 jsou výnosy obou investic. Střední hodnoty výnosů pro obě investice jsou 10 tis. Kč a směrodatné odchylky (rizika) jsou v obou případech rovny 2 tis. Kč. Vypočtěte střední hodnotu a riziko průměrného výnosu těchto investic. Předpokládejte, že výnosy obou investic mají korelační koeficienty postupně -0.8, 0 a 0.5.

Řešení. Pro průměrný výnos těchto investic platí:

$$Y = \frac{1}{2}(X_1 + X_2).$$

Tedy

$$E(Y) = \frac{1}{2}\big[E(X_1) + E(X_2)\big] = 10 \text{ [tis. Kč]}$$

bez ohledu na stupeň závislosti obou výnosů.

Zcela jiná situace ale je, pokud jde o riziko, tj. $\sqrt{D(Y)}$. Platí

$$\begin{aligned} D(Y) &= D\Big[\frac{1}{2}(X_1 + X_2)\Big] = \frac{1}{4}D(X_1 + X_2) \\ &= \frac{1}{4}\big[D(X_1) + D(X_2) + 2C(X_1, X_2)\big] \\ &= \frac{1}{4}\big[D(X_1) + D(X_2) + 2\sqrt{D(X_1)D(X_2)}\rho(X_1, X_2)\big]. \end{aligned}$$

Odtud pro $\rho(X_1, X_2) = -0.8$ dostáváme

$$D(Y) = \frac{1}{4}\big[4 + 4 + 2\sqrt{4 \cdot 4} \cdot (-0.8)\big] = 0.4 \Rightarrow \sqrt{D(Y)} \doteq 0.63 \text{ [tis. Kč]}.$$

Podobně

pro $\rho(X_1, X_2) = 0$ dostaneme $\sqrt{D(Y)} \doteq 1.41$ [tis. Kč],
pro $\rho(X_1, X_2) = 0.5$ dostaneme $\sqrt{D(Y)} \doteq 1.73$ [tis. Kč].

Vidíme tedy, že strategií, směřující k minimalizaci rizika při investování je rozptýlení investované částky na větší počet dílčích částí, které je třeba alokovat (umístit) do investičních příležitostí, jejichž výnosy jsou co nejvíce negativně korelovány.

♣

Uvažujme dále náhodný vektor $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ s rozdělovací funkcí g a oborem hodnot Ω. Číselné charakteristiky jeho složek $X_i, i = 1, 2, \ldots, n$, jsou číselné charakteristiky náhodných veličin. Při jejich výpočtu používáme zobecnění věty 8.2. Tak např. pro k-tý obecný moment náhodné veličiny X_i, tj. $E(X_i^k)$ platí

$$E(X_i^k) = \begin{cases} \int_\Omega \cdots \int x_i^k g(x_1, \ldots, x_n)\, dx_1 \ldots dx_n & \text{pro spojitý náhodný vektor} \\ \displaystyle\sum_{(x_1, \ldots, x_n) \in \Omega} x_i^k g(x_1, \ldots, x_n) & \text{pro diskrétní náhodný vektor} \end{cases}.$$

Vzhledem k tomu, že centrální momenty lze vyjádřit jako funkci obecných momentů můžeme nyní určit všechny momentové charakteristiky marginálních veličin $X_i, i = 1, \ldots, n$.

Vektor $\big(E(X_1), E(X_2), \ldots, E(X_n)\big)$ středních hodnot náhodných veličin X_1, $X_2, \ldots, X_n$ se nazývá střední hodnota náhodného vektoru $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$.

Kromě charakteristik jednotlivých veličin $X_i, i = 1, \ldots, n$, nás zajímají charakteristiky marginálních vektorů (X_i, X_j), $i, j = 1, \ldots, n$. Především již dříve definovaná kovariance $C_{ij} = C(X_i, X_j)$ a korelační koeficient $\rho_{ij} = \rho(X_i, X_j)$. Při jejich výpočtu použijeme opět zobecnění věty 8.2. Zřejmě

$$C_{ij} = C(X_i, X_j) = E\big([X_i - E(X_i)][X_j - E(X_j)]\big) = E(X_iX_j) - E(X_i)E(X_j),$$

kde

$$E(X_iX_j) = \begin{cases} \int\limits_{\Omega}\cdots\int x_ix_jg(x_1, \ldots, x_n)\,dx_1 \ldots dx_n & \text{pro spojitý náhodný vektor} \\ \sum\limits_{(x_1,\ldots,x_n)\in\Omega} x_ix_jg(x_1, \ldots, x_n) & \text{pro diskrétní náhodný vektor} \end{cases}.$$

Zřejmě platí

1. $C_{ij} = C_{ji}$, $\rho_{ij} = \rho_{ji}$,
2. $C_{ii} = D(X_i)$, $\rho_{ii} = 1$.

Matice

$$cov(\boldsymbol{X}) = \begin{bmatrix} C_{11} & C_{12} & \ldots & C_{1n} \\ C_{21} & C_{22} & \ldots & C_{2n} \\ \vdots & & \ddots & \vdots \\ C_{n1} & C_{n2} & \ldots & C_{nn} \end{bmatrix}$$

se nazývá kovarianční (varianční) matice náhodného vektoru $\boldsymbol{X}$.

Matice

$$cor(\boldsymbol{X}) = \begin{bmatrix} \rho_{11} & \rho_{12} & \ldots & \rho_{1n} \\ \rho_{21} & \rho_{22} & \ldots & \rho_{2n} \\ \vdots & & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \ldots & \rho_{nn} \end{bmatrix}$$

se nazývá korelační matice náhodného vektoru $\boldsymbol{X}$.

Obě matice jsou symetrické, typu (n, n). Kovarianční matice má na hlavní diagonále rozptyly $D(X_i)$ náhodných veličin $X_i, i = 1, 2, \ldots, n$. Korelační matice má na hlavní diagonále jedničky.

9. Některá důležitá rozdělení

V aplikacích se můžeme setkat s celou řadou rozdělení. Ta, která se vyskytují nejčastěji, mají speciální názvy. Již dříve jsme se seznámili s klasickým, rovnoměrným a exponenciálním rozdělením. Nyní se seznámíme s některými dalšími.

9.1. Diskrétní rozdělení

Definice 9.1.1. *Řekneme, že náhodná veličina X má binomické rozdělení (je binomická náhodná proměnná) s parametry n, p, jestliže má pravděpodobnostní funkci*

$$q(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{pro } x = 0, 1, \ldots, n \\ 0 & \text{jinak} \end{cases}, \tag{9.1.1}$$

kde $n \in \mathbb{N}$, $p \in (0,1)$. Značíme $X \sim Bi(n,p)$.

Věta 9.1.1. Platí

$$X \sim Bi(n,p) \Rightarrow E(X) = np,\ D(X) = np(1-p). \tag{9.1.2}$$

Poznámka 9.1.1.

1) Rozdělením $Bi(n,p)$ se řídí počet výskytů jevu A v n nezávislých pokusech, jestliže pravděpodobnost výskytu tohoto jevu v každém jednotlivém pokusu je rovna témuž číslu p. Pravděpodobnost, že v prvních x pokusech nastane jev A, zatímco ve zbývajících $n-x$ pokusech jev A nenastane, je vzhledem k nezávislosti pokusů rovna $p^x(1-p)^{n-x}$. Počet různých možností, při nichž v n pokusech nastane jev A právě x-krát, je dán kombinačním číslem $\binom{n}{x}$. Tudíž pravděpodobnost, že v n nezávislých pokusech nastane jev A právě x-krát, je dána výrazem (9.1.1).

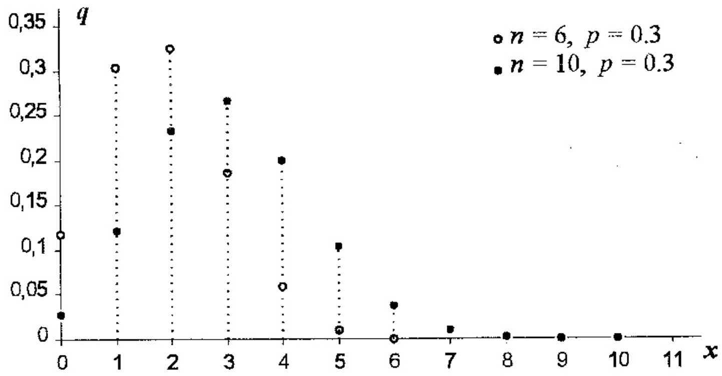
2) Příkladem náhodné veličiny X, která má rozdělení $Bi(n,p)$ je počet úspěšných zkoušek přístroje z n nezávislých zkoušek, počet vadných výrobků z n nezávisle vyrobených výrobků apod.

3) Je-li $X \sim Bi(n,p))$, potom pro modus x° náhodné veličiny X platí

$$(n+1)p - 1 \le x^\circ \le (n+1)p. \tag{9.1.3}$$

Je-li $(n+1)p$ přirozené číslo, má rozdělení dvě modální hodnoty $x^\circ = (n+1)p - 1$, $x^\circ = (n+1)p$, jinak má právě jednu.

4) Graf pravděpodobnostní funkce q rozdělení $Bi(n,p)$ je na obrázku 9.1.1.



Obrázek 9.1.1: Pravděpodobnostní funkce q rozdělení $Bi(n,p)$

5) Speciálním případem rozdělení $Bi(n,p)$ je rozdělení $Bi(1,p)$, které nazýváme alternativní rozdělení s parametrem p a značíme $A(p)$. Pro toto rozdělení ze vztahu (9.1.1) dostáváme

$$X \sim A(p) \sim q(x) = \begin{cases} \binom{1}{x} p^x (1-p)^{1-x} & \text{pro } x = 0, 1 \\ 0 & \text{jinak} \end{cases}.$$

Tedy

$$X \sim A(p) \sim q(x) = \begin{cases} p & \text{pro } x = 1 \\ 1-p & \text{pro } x = 0 \\ 0 & \text{jinak} \end{cases}.$$

Příklad 9.1.1. Na betonárce je pět míchaček betonu, z nichž každá má koeficient využití 0.8. Předpokládejme, že míchačky pracují nezávisle na sobě. Určete

a) pravděpodobnost, že v průběhu normální směny budou pracovat: 1) pouze dvě míchačky, 2) alespoň dvě míchačky, 3) nejvýše dvě míchačky,

b) očekávaný počet pracujících míchaček,

c) nejpravděpodobnější počet pracujících míchaček.

Řešení. Nechť X je počet míchaček z pěti, které budou pracovat v průběhu normální pracovní směny. Zřejmě $X \sim Bi(n,p)$, kde $n = 5, p = 0.8$. Ze vztahu (9.1.1) dostáváme

$$X \sim q(x) = \begin{cases} \binom{5}{x} 0.8^x 0.2^{5-x} & \text{pro } x = 0, 1, \ldots, 5 \\ 0 & \text{jinak} \end{cases}.$$

Odtud

a) 1) $P(X=2) = q(2) = \binom{5}{2} 0.8^2 0.2^3 = \frac{5!}{3!2!} 0.8^2 0.2^3 \doteq 0.051.$

$$2)\ P(X \geq 2) = 1 - P(X < 2) = 1 - [q(0) + q(1)] = 1 - \left[\binom{5}{0} 0.8^0 0.2^5 + \binom{5}{1} 0.8^1 0.2^4\right] \doteq 0.993.$$

$$3)\ P(X \leq 2) = q(0) + q(1) + q(2) = \binom{5}{0} 0.8^0 0.2^5 + \binom{5}{1} 0.8^1 0.2^4 + \binom{5}{2} 0.8^2 0.2^3 \doteq 0.058.$$

b) Ze vztahu (9.1.2) dostáváme $E(X) = np = 4$. Lze tedy očekávat, že budou pracovat čtyři míchačky.

c) Podle vztahu (9.1.3) musí platit:

$$6 \cdot 0.8 - 1 \leq x^\circ \leq 6 \cdot 0.8 \Leftrightarrow 3.8 \leq x^\circ \leq 4.8 \Rightarrow x^\circ = 4.$$

Nejpravděpodobněji budou pracovat čtyři míchačky.

♣

Příklad 9.1.2. Pravděpodobnost, že je množství odebraného elektrického proudu během dne v určitém podniku v normě (tj. nepřesáhne během dne plánovanou spotřebu), je 0.75. Určete, během kolika dnů bude odběr proudu alespoň jeden den nad normou s pravděpodobností alespoň 0.95.

Řešení. Nechť X je počet dní, ve kterých je spotřeba proudu během n dní nad normou. Zřejmě $X \sim Bi(n, 0.25)$. Máme určit počet dní n tak, aby

$$P(X \geq 1) \geq 0.95.$$

Podle vztahu (9.1.1) platí

$$X \sim q(x) = \begin{cases} \binom{n}{x} 0.25^x 0.75^{n-x} & \text{pro } x = 0, 1, \ldots, n \\ 0 & \text{jinak} \end{cases}.$$

Potom

$$P(X \geq 1) = 1 - P(X < 1) = 1 - q(0) = 1 - \binom{n}{0} 0.25^0 0.75^n = 1 - 0.75^n.$$

Tedy

$$P(X \geq 1) \geq 0.95 \Leftrightarrow 1 - 0.75^n \geq 0.95 \Leftrightarrow 0.75^n \leq 0.05 \Leftrightarrow$$
$$\Leftrightarrow n \ln 0.75 \leq \ln 0.05 \Leftrightarrow n \geq \frac{\ln 0.05}{\ln 0.75} \Rightarrow n \geq 10.41.$$

S pravděpodobností alespoň 0.95 bude během jedenácti dní odběr proudu alespoň jeden den nad normou.

♣

Definice 9.1.2. *Řekneme, že náhodná veličina X má* Poissonovo rozdělení *(je* Poissonova *náhodná proměnná)* s parametrem λ, *jestliže má pravděpodobnostní funkci*

$$q(x) = \begin{cases} \dfrac{\lambda^x e^{-\lambda}}{x!} & \textit{pro } x = 0, 1, \dots \\ 0 & \textit{jinak} \end{cases}, \tag{9.1.4}$$

kde $\lambda > 0$. *Značíme* $X \sim Po(\lambda)$.

Věta 9.1.2. Platí

$$X \sim Po(\lambda) \Rightarrow E(X) = \lambda,\ D(X) = \lambda. \tag{9.1.5}$$

Poznámka 9.1.2.

1) Rozdělení $Po(\lambda)$ má náhodná veličina X, která udává počet výskytů jevu A za časové období t, jestliže

a) jev A může nastat v libovolném okamžiku sledovaného časového období,

b) počet výskytů jevu A závisí pouze na délce tohoto časového období a ne na jeho počátku ani na tom, kolikrát jev A nastoupil před jeho počátkem,

c) pravděpodobnost, že jev A nastoupí více než jednou v časovém období délky t, konverguje k nule rychleji než t,

d) $\lambda = kt$, kde k je průměrný počet výskytů jevu A za časovou jednotku a t je sledované časové období.

Místo časového období můžeme výše uvažovat například i počet výskytů jevu A v určité ploše, objemu, hmotnosti apod.

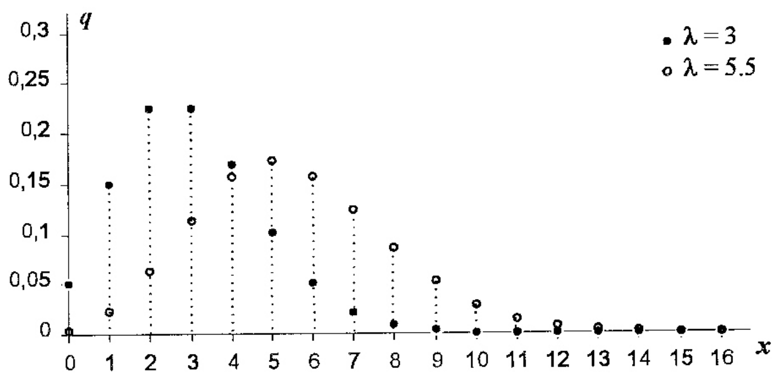
2) Rozdělení $Po(\lambda)$ aproximuje rozdělení $Bi(n,p)$. Je-li $p < 0.1$ a $n > 30$, potom $Bi(n,p) \doteq Po(\lambda)$, kde $\lambda = np$.

3) Poissonovým rozdělením se řídí např. počet poruch zařízení za směnu, počet signálů v telefonní ústředně, počet kazů v jednotce plochy nebo objemu.

4) Je-li $X \sim Po(\lambda)$, pak pro modus x° náhodné veličiny X platí

$$\lambda - 1 \le x^\circ \le \lambda. \tag{9.1.6}$$

5) Graf pravděpodobnostní funkce q rozdělení $Po(\lambda)$ je na obrázku 9.1.2.



Obrázek 9.1.2: Pravděpodobnostní funkce q rozdělení $Po(\lambda)$

Příklad 9.1.3. Je známo, že na 1000 m vozovky připadá průměrně pět nerovností překračujících povolenou normativní hodnotu. Jaká je pravděpodobnost, že na úseku 100 m

a) nebude ani jedna nerovnost,
b) bude nerovnost,
c) budou alespoň tři nerovnosti.

Řešení. Nechť X je počet nerovností na 100 m vozovky. Předpokládejme, že $X \sim Po(\lambda)$. Potom $\lambda = kt = \dfrac{5}{1000}100 = 0.5$. Tedy $X \sim Po(0.5)$. Ze vztahu (9.1.4) dostáváme

$$X \sim q(x) = \begin{cases} \dfrac{0.5^x \mathrm{e}^{-0.5}}{x!} & \text{pro } x = 0, 1, \dots \\ 0 & \text{jinak} \end{cases}.$$

Odtud

$$\text{a) } P(X=0) = q(0) = \frac{\mathrm{e}^{-0.5}}{0!} = \frac{1}{\sqrt{\mathrm{e}}} \doteq 0.607,$$

$$\text{b) } P(X \geq 1) = 1 - q(0) \doteq 0.396,$$

$$\begin{aligned} \text{c) } P(X \geq 3) &= 1 - P(X \leq 3) = 1 - \Big[q(0) + q(1) + q(2)\Big] \\ &= 1 - \mathrm{e}^{-0.5}\left[\frac{0.5^0}{0!} + \frac{0.5^1}{1!} + \frac{0.5^2}{2!}\right] \doteq 0.014. \end{aligned}$$

♣

Příklad 9.1.4. Při přejímací kontrole bylo z tisíce dílců náhodně kontrolováno padesát. Jaká je pravděpodobnost, že mezi kontrolovanými dílci není ani jeden vadný dílec, jestliže je známo, že v celém souboru jsou čtyři vadné.

Řešení. Nechť X je počet vadných dílců mezi padesáti dílci. Potom $X \sim Bi(n,p)$, kde $n = 50, p = 0.004$. Tedy podle vztahu (9.1.1) platí

$$X \sim q(x) = \begin{cases} \dbinom{50}{x} 0.004^x 0.996^{50-x} & \text{pro } x = 0, 1, \dots, 50 \\ 0 & \text{jinak} \end{cases}.$$

Odtud

$$P(X=0) = q(0) = \binom{50}{0} 0.004^0 0.996^{50} \doteq 0.8184.$$

Vzhledem k tomu, že $n > 30$ a $p < 0.1$, lze v souladu s poznámkou 9.1.2 2) použít aproximaci Poissonovým rozdělením. Platí $Bi(n,p) \doteq Po(\lambda)$, kde $\lambda = np = 0.2$. Tedy ze vztahu (9.1.4) dostáváme

$$X \sim q(x) \doteq \begin{cases} \dfrac{0.2^x \mathrm{e}^{-0.2}}{x!} & \text{pro } x = 0.1, \dots \\ 0 & \text{jinak} \end{cases},$$

odtud

$$P(X=0) = q(0) \doteq \mathrm{e}^{-0.2} \doteq 0.8187.$$

Hledaná pravděpodobnost je přibližně 0.82.

Příklad 9.1.5. Ze zkušenosti víme, že při správném chodu stroje je v průměru 0.09% výrobků vadných. Ke stroji nastoupil nový pracovník. Z pěti tisíc kusů, které vyrobil během týdne bylo 11 zmetků. Zjistěte, zda je jeho práce vyhovující, to znamená, není-li počet zmetků vyšší, než lze odůvodnit náhodou.

Řešení. Vypočítáme pravděpodobnost, že by bylo vyrobeno 11 zmetků za předpokladu, že nový pracovník je pracovníkem průměrným. Nechť tedy X je počet zmetků z 5000 kusů. Zřejmě $X \sim Bi(5000, p)$, kde je za výše uvedeného předpokladu $p = 0.0009$. Vzhledem k poznámce 9.1.2 2) lze rozdělení $Bi(5000, 0.0009)$ aproximovat rozdělením $Po(4.5)$. Ze vztahu (9.1.4) dostáváme

$$X \sim q(x) \doteq \begin{cases} \dfrac{4.5^x \mathrm{e}^{-4.5}}{x!} & \text{pro } x = 0, 1, \dots \\ 0 & \text{jinak} \end{cases}.$$

Odtud

$$P(X = 11) = q(11) \doteq \frac{4.5^{11} \mathrm{e}^{-4.5}}{11!} \doteq 0.004.$$

Tento výsledek svědčí o tom, že vysoký počet zmetků prakticky nelze vysvětlit působením náhodných činitelů a že tedy práce nového pracovníka není vyhovující.

Z aproximace Poissonovým rozdělením navíc plyne, že střední hodnota počtu zmetků z pěti tisíc kusů by měla být 4.5 a nejpravděpodobnější počet zmetků by měl být 4.

♣

9.2. Spojitá rozdělení

Definice 9.2.1. *Řekneme, že náhodná veličina X má Laplaceovo - Gaussovo nebo normální rozdělení (je normální náhodná proměnná) s parametry μ, σ^2, jestliže má hustotu*

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{pro } x \in \mathbb{R}, \tag{9.2.1}$$

kde $\mu \in \mathbb{R}$, $\sigma > 0$. Značíme $X \sim N(\mu, \sigma^2)$.

Věta 9.2.1. Platí

$$X \sim N(\mu, \sigma^2) \Rightarrow E(X) = \mu, \ D(X) = \sigma^2. \tag{9.2.2}$$

Poznámka 9.2.1.

1) Normální rozdělení je nejdůležitějším rozdělením spojité náhodné veličiny. Obecně lze říci, že toto rozdělení je použitelné všude tam, kde na kolísání náhodné veličiny působí velký počet nepatrných a vzájemně nezávislých vlivů. (Např. při sériové výrobě ovlivňuje rozměry výrobků kolísající kvalita surovin, nestejnoměrnost

strojového zpracování, různá pozornost dělníka apod.). Rozdělení $N(\mu, \sigma^2)$ mají například náhodné chyby měření a některé technické a fyzikální veličiny (délka, váha, pevnost,...).

2) Rozdělením $N(\mu, \sigma^2)$ lze aproximovat za jistých podmínek řadu dalších i diskrétních rozdělení. Např. $Bi(n,p) \doteq N(np, np(1-p))$ v případě, že $np(1-p) > 9$.

3) Jsou-li $X_1, \dots, X_n$ $(n > 30)$ nezávislé a stejně rozdělené náhodné veličiny, potom má aritmetický průměr těchto veličin přibližně normální rozdělení. Označíme-li $E(X_i) = \mu$, $D(X_i) = \sigma^2$ pro $i = 1, \dots, n$, potom

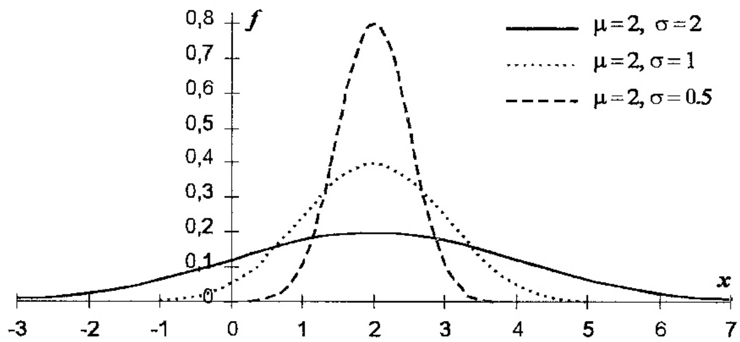
$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \mathrel{\dot\sim} N(E(\overline{X}), D(\overline{X})).$$

Z příkladu 8.6 plyne, že

$$E(\overline{X}) = \mu, \quad D(\overline{X}) = \frac{\sigma^2}{n}.$$

Této skutečnosti se často využívá v matematické statistice.

4) Graf hustoty f rozdělení $N(\mu, \sigma^2)$ je na obrázku 9.2.1. Graf hustoty f je souměrný podle přímky $x = \mu$, jedná se tedy o symetrické rozdělení. Funkce f má maximum v bodě $x = \mu$, které je rovno $1/(\sqrt{2\pi}\sigma)$, inflexní body jsou $x_{1,2} = \mu \pm \sigma$.



Obrázek 9.2.1: Graf hustoty f rozdělení $N(\mu, \sigma^2)$

Věta 9.2.2. Lineární kombinace nezávislých normálních náhodných veličin je normální náhodná veličina.

Poznámka 9.2.2. Buď $X_1, \dots, X_n$ nezávislé náhodné veličiny, $X_i \sim N(\mu_i, \sigma_i^2)$ pro $i = 1, \dots, n$ a $c_0, c_1, \dots, c_n$ reálná čísla, pro která platí $\sum_{i=1}^{n} c_i^2 > 0$. Lineární kombinace veličin $X_1, \dots, X_n$ je náhodná veličina

$$X = c_0 + c_1 X_1 + \cdots + c_n X_n.$$

Podle věty 9.2.2 má náhodná veličina X rozdělení $N\big(E(X), D(X)\big)$. Pro $E(X)$ a $D(X)$ platí podle poznámky 8.4 2)

$$\begin{aligned} E(X) &= E(c_0 + c_1X_1 + \cdots + c_nX_n) = c_0 + c_1E(X_1) + \cdots + c_nE(X_n) \\ &= c_0 + c_1\mu_1 + \cdots + c_n\mu_n, \\ D(X) &= D(c_0 + c_1X_1 + \cdots + c_nX_n) = c_1^2D(X_1) + \cdots + c_n^2D(X_n) \\ &= c_1^2\sigma_1^2 + \cdots + c_n^2\sigma^2. \end{aligned}$$

Poznámka 9.2.3. (Důsledek věty 9.2.2)

1) Je-li $X \sim N(\mu, \sigma^2)$, potom $U = (X - \mu)/\sigma \sim N(0, 1)$. Náhodná veličina, která má rozdělení $N(0, 1)$ se nazývá normovaná normální náhodná veličina. Rozdělovací, resp. distribuční funkci veličiny $N(0, 1)$ značíme φ, resp. Φ. Ze vztahu (9.2.1) dostaneme

$$U \sim N(0,1) \sim \varphi(u) = \frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-\frac{u^2}{2}} \quad \text{pro} \ \ u \in \mathbb{R},$$

$$U \sim N(0,1) \sim \Phi(u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-\frac{t^2}{2}}\,dt \quad \text{pro} \ \ u \in \mathbb{R}.$$

2) Jsou-li veličiny $X_1, \ldots, X_n$ nezávislé a $X_i \sim N(\mu, \sigma^2)$, $i = 1, \ldots, n$, potom

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu, \frac{\sigma^2}{n}).$$

Věta 9.2.3. Je-li $X \sim N(\mu, \sigma^2)$ s distribuční funkcí F a hustotou f, $U \sim N(0, 1)$ s distribuční funkcí Φ a hustotou φ, potom

1. $F(x) = \Phi\Big(\dfrac{x-\mu}{\sigma}\Big)$ 2. $f(x) = \dfrac{1}{\sigma}\,\varphi\Big(\dfrac{x-\mu}{\sigma}\Big)$
3. $\varphi(-x) = \varphi(x)$ 4. $\Phi(-x) = 1 - \Phi(x)$.

Důkaz

1. Vztah 1 vyplývá z poznámky 9.2.3 1). Má-li totiž náhodná veličina X rozdělení $N(\mu, \sigma^2)$, potom má náhodná veličina $U = (X - \mu)/\sigma$ rozdělení $N(0, 1)$, tedy $X = \sigma U + \mu$. Odtud

$$F(x) = P\big(X \le x\big) = P\big(\sigma U + \mu \le x\big) = P\Big(U \le \frac{x-\mu}{\sigma}\Big) = \Phi\Big(\frac{x-\mu}{\sigma}\Big).$$

2. Vztah 2 dostaneme derivací vztahu 1, neboť

$$\begin{aligned} f(x) &= \frac{d}{dx}F(x) = \frac{d}{dx}\Phi\Big(\frac{x-\mu}{\sigma}\Big) = \Big|v = v(x) = \frac{x-\mu}{\sigma}\Big| \\ &= \frac{d}{dv}\Phi(v)\frac{d}{dx}v(x) = \varphi(v)\,\frac{1}{\sigma} = \frac{1}{\sigma}\,\varphi\Big(\frac{x-\mu}{\sigma}\Big). \end{aligned}$$

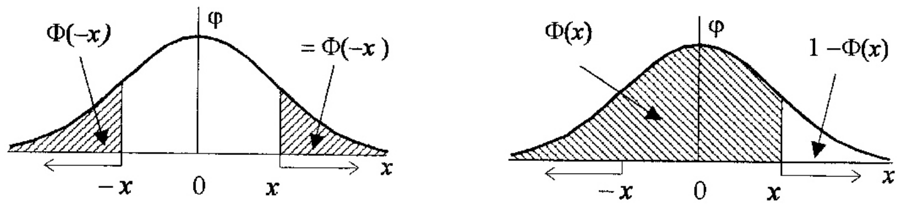
3. Připomeňme, že platí

$$U \sim N(0,1) \sim \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad \text{pro } x \in \mathbb{R},$$

odtud $\varphi(-x) = \varphi)(x)$.

4. Vzhledem k symetričnosti rozdělení platí (viz obrázek 9.2.2)

$$\Phi(-x) = P\big(X \leq -x\big) = P\big(X > x\big) = 1 - P\big(X < x\big) = 1 - \Phi(x).$$



Obrázek 9.2.2

Poznámka 9.2.4.

a) Tabelovány jsou:

1) Hodnoty distribuční funkce Φ (viz tab. 1a, 1b nebo [13]) a hustoty φ normované normální náhodné veličiny $N(0,1)$, a to v nezáporných bodech (viz [13]). Pro práci s tabulkami musíme tedy znát převodní vztahy z věty 9.2.3.

2) $100\,\alpha$ procentní kvantily $u(\alpha)$ veličiny $N(0,1)$ pro $\alpha > 0.5$ (tab. 2 nebo [13]). Pro $\alpha < 0.5$ je $u(\alpha) < 0$ a platí

$$u(\alpha) = -u(1-\alpha),$$

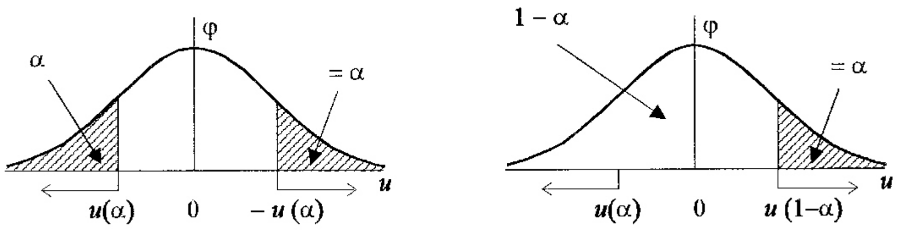
neboť (viz obrázek 9.2.3)

$$\begin{aligned} \alpha &= P\big(U \leq u(\alpha)\big) = P\big(U > -u(\alpha)\big) = 1 - P\big(U \leq -u(\alpha)\big) \\ &\Rightarrow P\big(U \leq -u(\alpha)\big) = 1 - \alpha \Rightarrow -u(\alpha) = u(1-\alpha) \Rightarrow u(\alpha) = -u(1-\alpha). \end{aligned}$$

b) EXCEL počítá:

1) Hodnoty distribuční funkce normální náhodné veličiny s libovolnými parametry.

2) Kvantily normální náhodné veličiny s libovolnými parametry.



Obrázek 9.2.3

Příklad 9.2.1. Náhodná veličina U má normované normální rozdělení. Určete
a) $P(U \leq 2), P(U \leq -1), P(U \geq 3), P(-2 \leq U \leq 1)$,
b) 95%-ní a 1%-ní kvantil náhodné veličiny U.

Řešení. Platí $U \sim N(0,1) \sim \Phi$. Pomocí tabulek (tab. 1a, 1b) postupně dostaneme

a) $P(U \leq 2) = \Phi(2) = 0.9773$,

$$P(U \leq -1) = \Phi(-1) = 1 - \Phi(1) = 1 - 0.8413 = 0.1587,$$

$$P(U \geq 3) = 1 - P(U < 3) = 1 - \Phi(3) = 1 - 0.9987 = 0.0013,$$

$$P(-2 \leq U \leq 1) = P(U \leq 1) - P(U < -2) = \Phi(1) - \Phi(-2) = \Phi(1) - [1 - \Phi(2)] = 0.8413 - 1 + 0.9773 = 0.8186.$$

b) Pomocí tabulek (tab. 2) a poznámky 9.2.4 dostaneme

$$u(0.95) = 1.645, \; u(0.01) = -u(0.99) = -2.326.$$

♣

Příklad 9.2.2. Náhodná veličina X má rozdělení $N(-2,9)$. Určete $P(|X| > 1)$, $P(X \geq -5)$.

Řešení. Ze vztahu 1 věty 9.2.4 plyne

$$X \sim N(-2,9) \sim F(x) = \Phi\Big(\frac{x+2}{3}\Big).$$

Odtud pomocí tabulek (tab. 1a, 1b) postupně dostaneme:

$$\begin{aligned} P(|X| > 1) &= P(X > 1 \vee X < -1) = P(X > 1) + P(X < -1) \\ &= 1 - P(X < 1) + P(X < -1) = 1 - F(1) + F(-1) \\ &\doteq 1 - \Phi(1) + \Phi(0.33) = 1 - 0.8413 + 0.6293 = 0.788, \end{aligned}$$

$$\begin{aligned} P(X \geq -5) &= 1 - P(X < -5) = 1 - F(-5) = 1 - \Phi(-1) \\ &= 1 - [1 - \Phi(1)] = \Phi(1) = 0.8413. \end{aligned}$$

♣

Příklad 9.2.3. $X \sim N(\mu, \sigma^2)$. Určete

a) $P(\mu - \sigma \leq X \leq \mu + \sigma)$,

b) $P(\mu - 2\sigma \leq X \leq \mu + 2\sigma)$,

c) $P(\mu - 3\sigma \leq X \leq \mu + 3\sigma)$.

Řešení. Platí

$$X \sim N(\mu, \sigma^2) \sim F(x) = \Phi\Big(\frac{x-\mu}{\sigma}\Big).$$

Odtud

$$\text{a) } P(\mu - \sigma \leq X \leq \mu + \sigma) = F(\mu+\sigma) - F(\mu-\sigma) = \Phi(1) - \Phi(-1) = 2\Phi(1) - 1$$
$$= 2 \cdot 0.8413 - 1 = 0.6826.$$

Podobně

$$\text{b) } P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = 2\Phi(2) - 1 = 0.9546,$$

$$\text{c) } P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = 2\Phi(3) - 1 = 0.9974.$$

♣

Poznámka 9.2.5. Výsledek z příkladu 9.2.3 c) se nazývá pravidlo tří sigma. Má-li náhodná veličina X rozdělení $N(\mu, \sigma^2)$, pak lze téměř jistě očekávat, že nabude hodnoty z intervalu $\langle \mu - 3\sigma, \mu + 3\sigma \rangle$.

Příklad 9.2.4. Hmotnost výrobku je vyhovující, jestliže je v mezích 68-69 g. Za standardních podmínek má váha výrobku normální rozdělení se střední hodnotou 68.3 g a směrodatnou odchylkou 0.2 g.

1) Jaká je pravděpodobnost, že váha náhodně vybraného výrobku bude v předepsaných mezích?

2) Jakou maximální váhu bude mít výrobek s pravděpodobností 0.99?

Řešení. Nechť X je váha výrobku, potom

$$X \sim N(68.3, 0.04) \sim F(x) = \Phi\Big(\frac{x - 68.3}{0.2}\Big).$$

Odtud

$$1)\ P(68 < X < 69) = F(69) - F(68) = \Phi(3.5) - \Phi(-1.5) = \Phi(3.5) + \Phi(1.5) - 1$$
$$= 0.9998 + 0.9332 - 1 = 0.933.$$

Váha výrobku bude v předepsaných mezích s pravděpodobností 0.933.

2) Máme určit váhu x tak, aby

$$P(X \leq x) = 0.99.$$

Odtud pomocí tabulek (tab. 2) dostaneme

$$0.99 = P(X \leq x) = F(x) = \Phi\Big(\frac{x - 68.3}{0.2}\Big) \Rightarrow \frac{x - 68.3}{0.2} = u(0.99)$$
$$\Rightarrow \frac{x - 68.3}{0.2} = 2.326 \Rightarrow x \doteq 68.8[g].$$

Výrobek bude mít s pravděpodobností 0.99 maximální váhu 68.8 g.

Příklad 9.2.5. Automat na nápoje je seřízen tak, že plní šálky po 200 ml požadovaného nápoje. Předpokládejme, že množství požadovaného nápoje má přibližně normální rozdělení se směrodatnou odchylkou 15 ml. Určete:

1) Kolik procent šálků bude obsahovat více než 224 ml.
2) Kolik procent šálků bude obsahovat 191-209 ml.
3) Kolik procent šálků z tisíce pravděpodobně přeteče, když budou použity 230 ml šálky.
4) Jaké maximální množství nápoje obsahuje 10 procent nejméně naplněných šálků.

Řešení. Nechť X je množství nápoje v šálku. Potom

$$X \sim N(200, 15^2) \sim F(x) = \Phi\Big(\frac{x - 200}{15}\Big).$$

1) $P(X \geq 224) = 1 - F(224) \doteq 1 - \Phi(1.6) = 0.055.$

5.5% šálků bude obsahovat více než 224 ml.

2) $P(191 < X < 209) = F(209) - F(191) = 2\Phi(0.6) - 1 = 0.452.$

Požadované množství bude obsahovat 45.2% šálků.

3) $P(X > 230) = 1 - F(230) = 1 - \Phi(2) = 0.023.$

Pravděpodobně přeteče 23 šálků.

4) Máme určit číslo x tak, aby platilo

$$P(X \leq x) = 0.10.$$

Tedy

$$0.10 = F(x) = \Phi\Big(\frac{x - 200}{15}\Big) \Rightarrow \frac{x - 200}{15} = u(0.10) = -u(0.90) = -1.282$$
$$\Rightarrow x \doteq 180.77.$$

Hledané maximální množství je tedy přibližně 181 ml.

♣

Příklad 9.2.6. Střední hodnota života určitého typu přístroje je 10 let a směrodatná odchylka 2 roky. Předpokládejme, že doba života přístroje má přibližně normální rozdělení. Určete:

1) Kolik procent přístrojů se porouchá do čtyř, osmi, resp. šestnácti let.
2) Do jaké doby se porouchá 10, 75, resp. 90 procent přístrojů.
3) Kolik procent přístrojů bude mít dobu života 7-13 let.
4) Délku záruční doby tak, aby byla reklamována pouze tři procenta přístojů.

Řešení. Nechť X je doba života přístrojů. Potom

$$X \dot\sim N(10, 2^2) \sim F(x) = \Phi\Big(\frac{x - 10}{2}\Big).$$

1) $P(X \leq 4) = F(4) = \Phi(-3) = 0.001$

PDF-XChange Click to buy NOW! www.docu-track.com

Podobně

$$P(X \leq 8) = 0.159,$$

$$P(X \leq 16) = 0.999.$$

Porouchá se tedy postupně 0.1% 15.9% a 99.9% přístrojů.

2) Počítáme stejně jako v příkladu 9.2.5 4). Dostaneme postupně 7.5, 11.4, 12.6 let.

3) $P(7 \leq X \leq 13) = F(13) - F(7) = \Phi 1.5) - \Phi(-1.5) = 2\Phi(1.5) - 1 = 0.866.$

Požadovanou dobu života bude mít 86.6% přístojů.

4) Máme určit číslo x tak, aby

$$P(X \leq x) = 0.03.$$

Dostaneme $x \doteq 6.238$. Tedy záruční doba by měla být 6.24 let.

$\clubsuit$

Příklad 9.2.7. Teploměry se balí automaticky do pouzder, přičemž délky v mm mají u teploměrů rozdělení $N(150, 2.53)$ a u pouzder rozdělení $N(155, 0.36)$. Je-li pouzdro krátké, stroj se zastaví a je nutná ruční operace. Během směny se vyrobí tři tisíce kusů. Jaký počet zastávek stroje lze očekávat během směny?

Řešení. Označme délku teploměru jako náhodnou veličinu X_1 a délku pouzdra jako veličinu X_2. Rozdíl $X = X_2 - X_1$ má potom vzhledem k nezávislosti veličin X_1 a X_2 normální rozdělení se střední hodnotou $\mu = \mu_2 - \mu_1 = 155 - 150 = 5$ a rozptylem $\sigma^2 = \sigma_1^2 + \sigma_2^2 = 2.53 + 0.36 = 2.89$. Tedy

$$X \sim N(5, 2.89) \sim F(x) = \Phi\left(\frac{x-5}{\sqrt{2.89}}\right).$$

Pravděpodobnost, že pouzdro bude kratší než teploměr je ekvivalentní s pravděpodobností, že náhodná veličina X nabude záporné hodnoty, tedy

$$P(X < 0) = F(0) = \Phi\left(\frac{-5}{\sqrt{2.89}}\right) \doteq \Phi(-2.94) = 1 - \Phi(2.94) = 0.00164.$$

Očekávaný počet zastávek stroje během jedné směny (střední hodnota počtu) je tedy $3000 \cdot 0.00164 \doteq 5$.

$\clubsuit$

Příklad 9.2.8. Měřící přístroj je zatížen jednak systematickou chybou 0.5, jednak náhodnou chybou. Náhodné chyby mají normální rozdělení se směrodatnou odchylkou 0.1.

1) V jakých mezích lze očekávat odchylku od správné hodnoty při jednom měření s pravděpodobností 0.95?

2) Kolik měření musíme provést, aby se aritmetický průměr těchto měření zmenšený o systematickou chybu odchyloval od správné hodnoty maximálně o 0.05 s pravděpodobností alespoň 0.95?

Řešení.

1) Nechť X je náhodná chyba měření, Y je neznámý výsledek měření veličiny jejíž skutečná hodnota je μ. Potom $Y = \mu + 0.5 + X$. Celková chyba měření je $Y - \mu = 0.5 + X$. Jestliže určíme konstantu k tak, aby

$$P(-k \le X \le k) = 0.95,$$

potom bude celková chyba měření v intervalu $\langle 0.5 - k, 0.5 + k\rangle$ s pravděpodobností 0.95. Zřejmě

$$X \sim N(0, 0.01) \sim F(x) = \Phi\left(\frac{x}{0.1}\right).$$

Odtud

$$0.95 = P(-k \le X \le k) = F(k) - F(-k) = \Phi\left(\frac{k}{0.1}\right) - \Phi\left(-\frac{k}{0.1}\right) = 2\Phi\left(\frac{k}{0.1}\right) - 1$$
$$\Rightarrow \Phi\left(\frac{k}{0.1}\right) = 0.975 \Rightarrow \frac{k}{0.1} = u(0.975)$$
$$\Rightarrow \frac{k}{0.1} = 1.960 \Rightarrow k = 0.196.$$

Tedy celková chyba měření bude s předepsanou pravděpodobností v intervalu $\langle 0.304, 0.696\rangle$

b) Nechť Y_i je neznámý výsledek i-tého měření veličiny μ, X_i náhodná chyba i-tého měření, $i = 1, \ldots, n$. Potom

$$Y_i = \mu + 0.5 + X_i,\ i = 1, \ldots, n,\ X_i \sim N(0, 0.01).$$

Označme

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i, \quad \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Chceme určit n tak, aby

$$P(|\overline{Y} - 0.5 - \mu| \le 0.05) \ge 0.95.$$

Zřejmě

$$\overline{Y} - 0.5 - \mu = \frac{1}{n}\sum_{i=1}^{n} Y_i - 0.5 - \mu = \frac{1}{n}\sum_{i=1}^{n}(\mu + 0.5 + X_i) - 0.5 - \mu =$$
$$= \mu + 0.5 + \frac{1}{n}\sum_{i=1}^{n} X_i - 0.5 - \mu = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$$

Tedy požadujeme aby:

$$0.95 \le P(|\overline{X}| \le 0.05).$$

Platí

$$\overline{X} \sim N\left(0, \frac{0.01}{n}\right) \sim F(x) = \Phi\left(\frac{x}{0.1}\sqrt{n}\right),$$

odtud

$$
\begin{aligned}
0.95 &\le P(-0.05 \le \overline{X} \le 0.05) = F(0.05) - F(-0.05) \\
&= \Phi(0.5\sqrt{n}) - \Phi(-0.5\sqrt{n}) = 2\Phi(0.5\sqrt{n}) - 1 \Rightarrow \Phi(0.5\sqrt{n}) \ge 0.975 \\
&\Rightarrow 0.5\sqrt{n} \ge u(0.975) \Rightarrow 0.5\sqrt{n} \ge 1.960 \Rightarrow n \ge 15.3664.
\end{aligned}
$$

Musíme tedy provést alespoň 16 měření.

♣

Na závěr této kapitoly uvedeme přehled rozdělení odvozených z normálního rozdělení. Tato rozdělení mají důležité uplatnění v teorii matematické statistiky (při intervalových odhadech parametrů a testování hypotéz).

Poznámka 9.2.6.

1) Jsou-li $X_1, X_2, \dots, X_n$ nezávislé normované normální náhodné veličiny, potom má náhodná veličina

$$X = X_1^2 + X_2^2 + \cdots + X_n^2$$

tzv. Pearsonovo nebo chí-kvadrát rozdělení s n stupni volnosti. Značíme $X \sim \chi^2(n)$.

2) Jsou-li X_1, X_2 nezávislé náhodné veličiny, $X_1 \sim N(0,1), X_2 \sim \chi^2(n)$, potom má náhodná veličina

$$X = \frac{X_1}{\sqrt{\dfrac{X_2}{n}}}$$

tzv. Studentovo nebo t-rozdělení s n stupni volnosti. Značíme $X \sim t(n)$.

3) Jsou-li X_1, X_2 nezávislé náhodné veličiny, $X_1 \sim \chi^2(n_1), X_2 \sim \chi^2(n_2)$, potom má náhodná veličina

$$X = \frac{\dfrac{X_1}{n_1}}{\dfrac{X_2}{n_2}}$$

tzv. Fisherovo-Snedecorovo nebo F-rozdělení s n_1, n_2 stupni volnosti. Značíme $X \sim F(n_1, n_2)$.

V tabulce 9.2.1 uvádíme hustoty výše uvedených rozdělení. $\Gamma(a)$ zde značí tzv. gama funkci, která je definovaná vztahem

$$\Gamma(a) = \int_0^\infty t^{a-1} \mathrm{e}^{-t} dt \quad \text{pro} \quad a > 0.$$

Při výpočtu Γ-funkce vystačíme z následujícími vlastnostmi:

$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$,

$\Gamma(1) = 1,$
$\Gamma(a) = (a-1)\Gamma(a-1)$ pro $a > 0,$
$\Gamma(a) = (a-1)!$ pro $a \in \mathbb{N}.$
Γ-funkci tedy můžeme chápat jako zobecnění pojmu faktoriál.

Název rozdělení Označení Parametry	Hustota
χ^2 – rozdělení $\chi^2(n)$ $n \in \mathbb{N}$	$f(x) = \begin{cases} \dfrac{x^{\frac{n}{2}-1}e^{-\frac{x}{2}}}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} & \text{pro } x > 0 \\ 0 & \text{jinak} \end{cases}$
t – rozdělení $t(n)$ $n \in \mathbb{N}$	$f(x) = \dfrac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{k}{2})\sqrt{\pi n}}(1+\frac{x^2}{n})^{-\frac{k+1}{2}} \quad \text{pro } x \in \mathbb{R}$
F – rozdělení $F(n_1, n_2)$ $n_1, n_2 \in \mathbb{N}$	$f(x) = \begin{cases} \dfrac{\Gamma(\frac{n_1+n_2}{2})}{\Gamma(\frac{n_1}{2})\Gamma(\frac{n_2}{2})}(\frac{n_1}{n_2})^{\frac{n_1}{2}} x^{\frac{n_1}{2}-1}(1+\frac{n_1}{n_2}x)^{-\frac{n_1+n_2}{2}} & \text{pro } x > 0 \\ 0 & \text{jinak} \end{cases}$

Tabulka 9.2.1: Hustoty rozdělení

Na obrázcích 9.2.4 - 9.2.6 jsou grafy hustot výše uvedených rozdělení.

V tabulce 9.2.2 uvádíme označení kvantilů, střední hodnoty a rozptyly výše uvedených rozdělení.

Rozdělení	$100\alpha\%$–ní kvantil	Střední hodnota	Rozptyl
$\chi^2(n)$	$\chi^2(n;\alpha)$	n	$2n$
$t(n)$	$t(n;\alpha)$	$0 \ (n > 1)$	$\dfrac{n}{n-2} \quad (n > 2)$
$F(n_1, n_2)$	$F(n_1, n_2; \alpha)$	$\dfrac{n_1}{n_2 - 2} \ (n_2 > 1)$	$\dfrac{2n^2(n_1 - n_2 - 2)}{n_1(n_2-2)^2(n_2-4)} \ (n_2 > 4)$

Tabulka 9.2.2: Označení kvantilů, střední hodnoty a rozptyly

Poznámka 9.2.7.

a) Pro práci s tabulkami je nutné znát:

1. Kvantily $\chi^2(n;\alpha)$ jsou někdy tabelovány pouze pro $n \leq 30$. Pro $n > 30$ platí:

$$\chi^2(n;\alpha) \doteq \frac{1}{2}\left[\sqrt{(2n-1)} + u(\alpha)\right]^2.$$

2. Kvantily $t(n,\alpha)$ jsou tabelovány vždy pro $\alpha \geq 0.5$ a někdy pouze pro $n \leq 30$. Pro $\alpha < 0.5$ platí:

$$t(n;\alpha) = -t(n;1-\alpha).$$

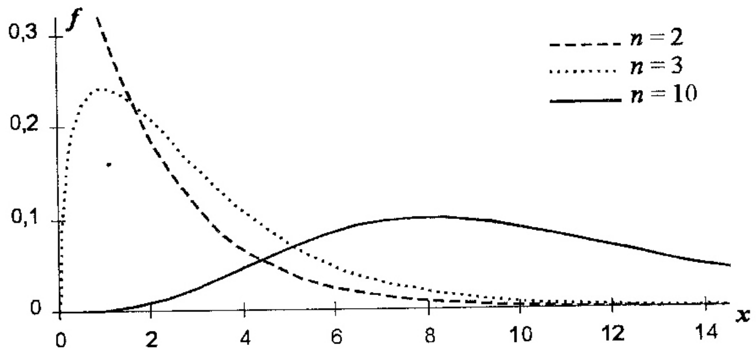
Pro $n > 30$ platí:

$$t(n;\alpha) \doteq u(\alpha).$$

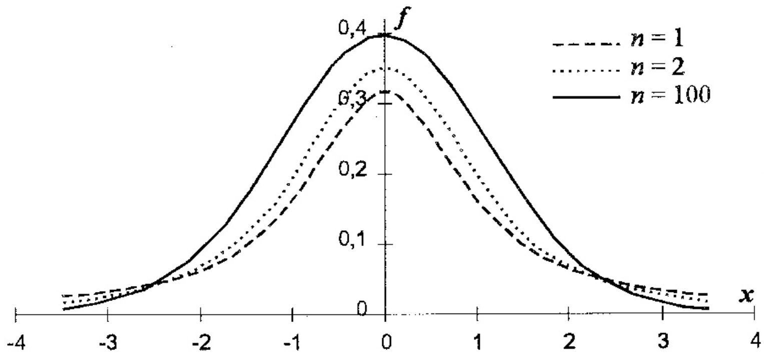
3. Kvantily $F(n_1,n_2;\alpha)$ jsou tabelovány pro $\alpha \geq 0.5$. Pro $\alpha < 0.5$ platí:

$$F(n_1,n_2;\alpha) = \frac{1}{F(n_2,n_1;1-\alpha)}.$$

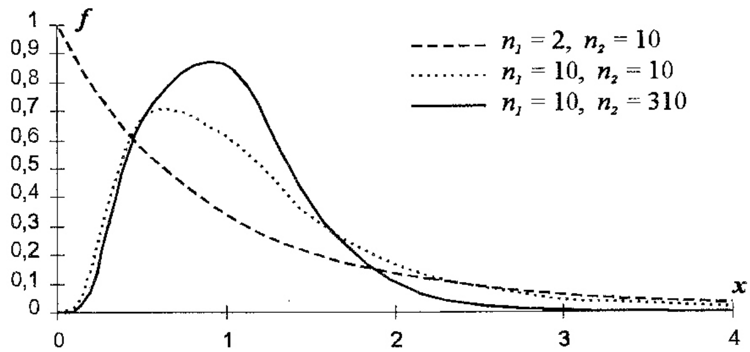
b) EXCEL počítá kvantily $\chi^2(n;\alpha), t(n;\alpha), F(n_1,n_2;\alpha)$ pro libovolné přípustné hodnoty parametrů.



Obrázek 9.2.4: Graf hustoty f rozdělení $\chi^2(n)$



Obrázek 9.2.5: Graf hustoty f rozdělení $t(n)$



Obrázek 9.2.6: Graf hustoty f rozdělení $F(n_1, n_2)$

PDF-XChange Click to buy NOW! www.docu-track.com

MATEMATICKÁ STATISTIKA

1. Úvod

Víme, že náhodná veličina X je určena, jestliže známe její rozdělení. V teorii pravděpodobnosti jsme předpokládali, že rozdělení náhodné veličiny X známe a zabývali jsme se studiem jeho vlastností. V technických aplikacích je však rozdělení náhodné veličiny X neznámé a je třeba o něm získat nějaké poznatky. Získávání těchto poznatků je velmi usnadněno, je-li předem známý typ rozdělení. V tomto případě pouze odhadujeme některé neznámé parametry rozdělení nebo ověřujeme svoje domněnky (hypotézy) o jejich skutečných hodnotách. Obecněji, pokud neznáme typ rozdělení X, odhadujeme jeho samotný tvar, resp. ověřujeme hypotézu o skutečném tvaru rozdělení. Tyto úlohy (a řadu jiných) řešíme pomocí metod matematické statistiky.

Přirozená otázka je, na základě čeho budeme získávat popsané poznatky o rozdělení náhodné veličiny X. Odpověď na ni nalezneme v následujícím typicky školském příkladu.

Příklad 1.1. Mějme hrací kostku a označme X počet „šestek", které padnou v jednom hodu touto kostkou. Potom je zřejmě základní prostor Ω množina $\{1, 0\}$ a pravděpodobnostní funkce q náhodné veličiny X je

$$q = q(x) = P(X = x) = \begin{cases} \dfrac{1}{6} & \text{pro } x = 1 \\ \dfrac{5}{6} & \text{pro } x = 0\,. \\ 0 & \text{jinak} \end{cases}$$

Tedy $X \sim A(1/6)$ (viz poznámka P - 9.1.1 5)). Známe tedy rozdělení náhodné veličiny X.

Mějme nyní hrací kostku „falešnou" a nechť X má stejný význam jako výše. Náhodná veličina X je opět alternativní, ale neznáme parametr p (protože je kostka falešná, nevíme s jakou pravděpodobností padne při jednom hodu číslo 6). Označme g pravděpodobnostní funkci náhodné veličiny X, potom

$$g = g(x; p) = \begin{cases} p & \text{pro } x = 1 \\ 1-p & \text{pro } x = 0\,, \\ 0 & \text{jinak} \end{cases}$$

kde p je nějaké číslo z intervalu (0,1).

Známe tedy tvar rozdělení X (víme, že je alternativní), neznáme však hodnotu parametru p. O parametru víme pouze, že musí ležet v intervalu $(0, 1)$. Kdybychom věděli, že kostka byla „úspěšně" upravena, pak $p \in (1/6, 1)$. Chceme nyní odhadnout neznámý parametr p. Každého asi napadne, aby n-krát hodil touto kostkou a spočítal, kolikrát mu v těchto n hodech padlo číslo 6. Za odhad parametru p by pak volil počet šestek ku celkovému počtu hodů, tj. n. Pokusme se nyní naše počínání vyjádřit v terminologii náhodných veličin (jakožto neznámých výsledků pokusů).

Označme X_i počet šestek v i-tém hodu falešnou kostkou ($i = 1, \dots, n$). Náhodná veličina X_i ($i = 1, \dots, n$) zřejmě může nabýt pouze hodnoty 1 nebo 0. Přitom hodnoty 1 nabude, jestliže číslo 6 v i-tém hodu padne, a hodnotu 0 nabude, jestliže číslo 6 v i-tém hodu nepadne. Tedy X_i ($i = 1, \dots, n$) je alternativní náhodná proměnná. Pokud jsme nezměnili při házení podmínky, tj. v našem případě kostku (což by bylo v dané úloze ověřování „kvality" kostky zcela nesmyslné), je $X_i \sim A(p)$ ($i = 1, \dots, n$) - tedy všechny veličiny $X_1, \dots, X_n$ mají alternativní rozdělení se stejným parametrem p. Dále si uvědomme, že výsledek X_i ($i = 1, \dots, n$) v i-tém hodu bude zřejmě nezávislý na výsledcích v ostatních hodech.

Máme tedy uspořádanou n-tici $(X_1, \dots, X_n)$ nezávislých a stejně rozdělených náhodných veličin -tzv. náhodný výběr z rozdělení $X \sim A(p)$. Před provedením „experimentu", tj. než n-krát hodíme kostkou a zapíšeme výsledek, víme pouze, že výsledek musí být náhodný vektor $(X_1, \dots, X_n)$ s výše uvedenými vlastnostmi. Po provedení experimentu a zapsání výsledku dostaneme konkrétní n-tici $(x_1, \dots, x_n)$ reálných čísel, kde $x_i \in \{1, 0\}$ - tzv. realizaci náhodného výběru $(X_1, \dots, X_n)$.

Jestliže nám např. číslo 6 padlo pouze v prvním a předposledním hodu, dostaneme po zapsání výsledek (1,0, ... ,0,1,0), jestliže číslo 6 padlo ve třetím hodu a posledních dvou hodech, dostaneme (0,0,1,0, ... ,0,1,1).

Řekli jsme, že za odhad parametru p vezmeme počet úspěšných hodů ku celkovému počtu hodů, tedy v prvním případě realizace $2/n$, v druhém případě $3/n$. Před provedením experimentu pak odhad parametru p můžeme psát ve tvaru $(X_1 + X_2 + \dots + X_n)/n$.

♣

Ukázali jsme tedy, že jisté poznatky o neznámém rozdělení náhodné veličiny X lze vyvodit z tzv. náhodného výběru z rozdělení X, přesněji řečeno z jeho realizace.

V následující kapitole uvedeme definici těchto dvou základních pojmů matematické statistiky.

2. Náhodný výběr

Definice 2.1. *Buď X náhodná veličina, $(X_1, X_2, \ldots, X_n)$ náhodný vektor s vlastnostmi:*

a) veličiny $X_1, X_2, \ldots, X_n$ mají stejný zákon rozdělení jako X

b) veličiny $X_1, X_2, \ldots, X_n$ jsou nezávislé.

Potom náhodný vektor $(X_1, X_2, \ldots, X_n)$ nazýváme <u>*náhodný výběr z (rozdělení) X (o rozsahu n)*</u>*.*

Obor hodnot náhodného výběru nazýváme <u>*výběrový prostor*</u> *a značíme S. Každý prvek výběrového prostoru nazýváme* <u>*realizace náhodného výběru*</u>*, značíme $(x_1, x_2, \ldots, x_n)$.*

Poznámka 2.1.

1. Buď Ω obor hodnot náhodné veličiny X. Je-li $(X_1, X_2, \ldots, X_n)$ náhodný výběr z X, potom výběrový prostor je množina Ω^n.

2. Uvažujme nějaký pokus, jehož výsledkem je (do té doby, než jej známe) náhodná veličina X. Opakujme tento pokus n-krát ($n \geq 1$). Označme X_i výsledek (do té doby než jej známe) v i-tém pokusu ($i = 1, 2, \ldots, n$). Jestliže výsledek v i-tém pokusu ($i = 1, 2, \ldots, n$) nezávisí na výsledcích ostatních pokusů (přesněji řečeno, je-li rozdělovací funkce náhodného vektoru $(X_1, X_2, \ldots, X_n)$ součinem rozdělovacích funkcí náhodných veličin X_i ($i = 1, 2, \ldots, n$)) a jestliže nedochází během experimentu ke změně podmínek (tj. X_i má stejné rozdělení jako X), bude $(X_1, X_2, \ldots, X_n)$ náhodný výběr z X. Tedy náhodný výběr $(X_1, X_2, \ldots, X_n)$ je výsledek n nezávislých opakování téhož pokusu, za nezměněných definičních podmínek pokusu. Realizace $(x_1, x_2, \ldots, x_n)$ náhodného výběru z X je pak známá hodnota výsledku po provedení popsaného experimentu (tedy n-tice reálných čísel).

Příklad 2.1. Pozorování hustoty provozu (vyjádřené např. počtem projíždějících vozidel za hodinu) na jednom místě ve stejnou dobu (např. v ranní dopravní špičce) konané v průběhu n pracovních dnů stejného ročního období poskytne náhodný výběr rozsahu n; zpravidla lze předpokládat, že jde o náhodný výběr z Poissonova rozdělení. *Všimněme si zde všech vyjmenovaných podmínek: jedno roční období,* stejná denní doba, pracovní dny; spojit např. pozorování z různých ročních období by znamenalo porušit podmínku stálosti podmínek a způsobilo by, že pozorování by nemělo stejné rozdělení, neboť charakter provozu bývá v zimě jiný než v létě. Měli bychom pak co dělat s několika náhodnými výběry z Poissonova rozdělení s různými hodnotami parametru λ.

♣

Pro rozdělení náhodného výběru platí následující věta.

Věta 2.1. Buď $(X_1, X_2, \ldots, X_n)$ náhodný výběr z rozdělení X.

a) Nechť X má distribuční funkci G. Potom náhodný výběr $(X_1, X_2, \ldots, X_n)$ má distribuční funkci

$$H(x_1, x_2, \ldots, x_n) = G(x_1)\, G(x_2)\, \ldots\, G(x_n).$$

b) Nechť X má rozdělovací funkci g. Potom náhodný výběr $(X_1, X_2, \ldots, X_n)$ má rozdělovací funkci

$$h(x_1, x_2, \ldots, x_n) = g(x_1)\, g(x_2)\, \ldots\, g(x_n).$$

Důkaz: Plyne přímo z definice nezávislosti náhodných veličin. □

Poznámka 2.2. Z věty 2.1 plyne:

1. Známe-li tvar rozdělení náhodné veličiny X, pak známe i tvar rozdělení náhodného výběru z X.
2. Jestliže rozdělení náhodné veličiny X závisí na nějakých neznámých konstantách (parametrech), pak na těchto parametrech závisí i rozdělení náhodného výběru z X.

Příklad 2.2. *Určete rozdělovací funkci náhodného výběru z rozdělení*

a) alternativního

b) normálního.

Řešení.

a) Víme, že $X \sim A(p)$, kde $p \in (0, 1)$. Pravděpodobnostní funkci g náhodné veličiny X můžeme psát ve tvaru

$$g = g(x; p) = \begin{cases} p^x (1-p)^{1-x} & \text{pro } x \in \{0, 1\} \\ 0 & \text{jinak} \end{cases}, \quad \text{kde} \quad p \in (0, 1).$$

Buď $(X_1, X_2, \ldots, X_n)$ náhodný výběr z rozdělení X. Obor hodnot Ω náhodné veličiny X je množina $\{0, 1\}$, výběrový prostor S je tedy množina

$$S = \{0, 1\}^n = \{(x_1, x_2, \ldots, x_n);\ x_i \in \{0, 1\},\ i = 1, 2, \ldots, n\}.$$

Potom pravděpodobnostní funkce h náhodného výběru $(X_1, X_2, \ldots, X_n)$ z X je (podle věty 2.1)

$$\begin{aligned} h = h(x_1, x_2, \ldots, x_n; p) &= g(x_1; p)\, g(x_2; p)\, \ldots\, g(x_n; p) \\ &= \begin{cases} p^{\sum_{i=1}^n x_i} \cdot (1-p)^{n - \sum_{i=1}^n x_i} & \text{pro } (x_1, x_2, \ldots, x_n) \in S \\ 0 & \text{jinak} \end{cases}, \end{aligned}$$

kde $p \in (0, 1)$.

b) Hustota f náhodné veličiny $X \sim N(\mu, \sigma^2)$ je pro $x \in \mathbb{R}$

$$f = f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\, \mathrm{e}^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \text{ kde } (\mu, \sigma^2) \in (-\infty, \infty) \times (0, \infty).$$

Buď $(X_1, X_2, \ldots, X_n)$ náhodný výběr z rozdělení X. Obor hodnot Ω náhodné veličiny X je $\mathbb{R}$, výběrový prostor je tedy množina $\mathbb{R}^n$. Potom hustota s náhodného výběru $(X_1, X_2, \ldots, X_n)$ z X je pro $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$

$$
\begin{aligned}
s &= s(x_1, x_2, \ldots, x_n; \mu, \sigma^2) = f(x_1; \mu, \sigma^2)\, f(x_2; \mu, \sigma^2)\, \ldots\, f(x_n; \mu, \sigma^2) \\
&= \frac{1}{\left(\sqrt{2\pi}\sigma\right)^n}\, \mathrm{e}^{-\frac{1}{2\sigma^2}\left[(x_1-\mu)^2 + (x_2-\mu)^2 + \cdots + (x_n-\mu)^2\right]},
\end{aligned}
$$

kde $(\mu, \sigma^2) \in (-\infty, \infty) \times (0, \infty)$.
Při označení

$$
\boldsymbol{\mu} = (\mu, \mu, \ldots, \mu), \ \boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_n, \ \boldsymbol{x} = (x_1, x_2, \ldots, x_n),
$$

lze psát hustotu s v běžněji užívaném maticovém tvaru

$$
s = \frac{1}{\left(\sqrt{2\pi}\right)^n \sqrt{\det \Sigma}}\, \mathrm{e}^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})^T}.
$$

♣

3. Bodový odhad parametrů rozdělení

V dalším se budeme zabývat situací, kdy zkoumáme náhodnou veličinu X s rozdělovací funkcí f. Budeme předpokládat, že známe rozdělovací funkci f, až na nějaké konstanty θ_i $(i = 1, 2, \dots, m)$ - tzv. parametry. Označme pro $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_m)$ jako $\boldsymbol{\Theta}$ množinu teoreticky možných hodnot vektorového parametru $\boldsymbol{\theta}$, kterou nazýváme parametrický prostor. Potom budeme psát

$$f = f(x; \boldsymbol{\theta}),\ \boldsymbol{\theta} \in \boldsymbol{\Theta}.$$

Uvědomme si, že pro konkrétní hodnotu vektorového parametru $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ obdržíme konkrétní rozdělovací funkci náhodné veličiny X.

Příklad 3.1. Víme, že pravděpodobnostní funkce g alternativní náhodné proměnné s parametrem p je

$$g = g(x; p) = \begin{cases} p^x (1-p)^{1-x} & \text{pro } x \in \{0, 1\} \\ 0 & \text{jinak} \end{cases}.$$

V příkladu 1.1. o „falešné" hrací kostce máme tedy pouze jeden neznámý parametr p, tj. $\boldsymbol{\theta} = \theta_1 = \theta = p$. Protože $p \in (0, 1)$ je $\boldsymbol{\Theta} = (0, 1)$. Víme-li, že těžiště hrací kostky bylo posunuto směrem k číslu 1, potom zřejmě $p \in (1/6, 1)$, tj. $\boldsymbol{\Theta} = (1/6, 1)$.
♣

Příklad 3.2.

a) Při měření vzdálenosti konkrétním odzkoušeným přístrojem, který nevykazuje systematickou chybu, je výsledek pokusu X normální náhodná veličina s neznámou střední hodnotou - skutečnou vzdáleností a (většinou) známým rozptylem (vyjadřujícím přesnost přístroje). Potom pro hustotu f náhodné veličiny X platí

$$f = f(x; \mu) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \quad \text{pro } x \in \mathbb{R},$$

kde $\mu \in (0, \infty)$. Tedy $\boldsymbol{\theta} = \theta_1 = \theta = \mu$, $\boldsymbol{\Theta} = (0, \infty)$.

b) Při měření neodzkoušeným přístrojem bude neznámý i rozptyl, potom

$$f = f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \quad \text{pro } x \in \mathbb{R},$$

kde $(\mu, \sigma^2) \in (0, \infty) \times (0, \infty)$. Tedy $\boldsymbol{\theta} = (\theta_1, \theta_2) = (\mu, \sigma^2)$ a $\boldsymbol{\Theta} = (0, \infty) \times (0, \infty)$.
♣

Poznámka 3.1. Dále se budeme zabývat odhadem parametru θ_i. K tomu je zapotřebí:

1. říci, co to je odhad parametru,
2. uvést „rozumné" požadavky kladené na odhad parametru,
3. seznámit se s metodami, pomocí kterých odhady hledáme.

My se zde budeme zabývat pouze otázkami 1 a 2. Problematiku 3 najdeme např. v [2].

Definice 3.1. *Nechť má náhodná veličina X rozdělovací funkci $f = f(x;\boldsymbol{\theta})$, kde $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m) \in \boldsymbol{\Theta}$ je neznámý parametr. Buď $(X_1, X_2, \dots, X_n)$ náhodný výběr z X, Ω^n výběrový prostor, $t(x_1, x_2, \dots, x_n)$ reálná funkce n reálných proměnných definovaná na Ω^n.*

Potom náhodnou veličinu

$$T = t(X_1, X_2, \dots, X_n)$$

nazýváme <u>statistika</u> (vždy) nebo <u>(bodový) odhad</u> (podle souvislostí) <u>neznámého parametru θ_i</u> pro $i \in \{1, 2, \dots, m\}$.

Je-li $(x_1, x_2, \dots, x_n)$ realizace náhodného výběru, potom číslo

$$t = t(x_1, x_2, \dots, x_n)$$

nazýváme <u>realizace statistiky</u> nebo <u>realizace bodového odhadu parametru θ_i</u>.

Poznámka 3.2.

1. Víme-li tedy, že náhodný výběr „provádíme" proto, abychom získali číselnou informaci o poloze neznámého parametru θ_i, potom libovolnou možnou funkci z náhodného výběru nazýváme bodový odhad parametru ať už je „dobrý" z jakéhokoliv hlediska nebo není, ať už je smysluplný či není.

2. Pojem statistiky byl v definici 3.1 zaveden proto, že ji budeme používat i k jiným účelům a nejenom pro bodový odhad parametru rozdělení.

Příklad 3.3. V příkladu 1.1 jsme pracovali s náhodným výběrem z alternativního rozdělení s parametrem p a za odhad parametru p jsme zvolili statistiku $T_1 = \frac{1}{n}\sum_{i=1}^{n} X_i$, kterou můžeme považovat za smysluplný odhad. Na první pohled je zřejmé, že statistiky $T_2 = 2X_2 - X_1$, $T_3 = X_1 + X_2 + \dots + X_n$ nepatří mezi smysluplné odhady parametru p a $T_4 = \ln X_1 + \ln X_2 + \dots + \ln X_n$, $T_5 = \frac{p}{n}\sum_{i=1}^{n} X_i$ nejsou ani statistiky.

♣

Kdy bude odhad neznámého parametru smysluplný? To lze posuzovat z mnoha hledisek. Uvědomme si, že odhad $T = t(X_1, X_2, \dots, X_n)$ parametru θ_i je náhodná veličina. V technické praxi nás samozřejmě zajímá číselná hodnota odhadu, tj. realizace odhadu, kterou získáme na základě realizace náhodného výběru. Je zřejmé, že z každé realizace náhodného výběru můžeme obecně dostat jinou realizaci odhadu, tj. jinou číselnou hodnotu. Nejčastěji se vyskytuje požadavek, aby realizace t odhadu T parametru θ_i kolísaly okolo skutečné hodnoty parametru θ_i, ať je hodnota vektorového parametru $\boldsymbol{\theta}$ jakákoliv, tj. aby byl odhad T parametru θ_i <u>nestranný</u> (<u>nevychýlený</u>). Uvědomíme-li si nyní, že realizace náhodné veličiny T

PDF-XChange Click to buy NOW! www.docu-track.com

kolísají okolo její střední hodnoty (pokud existuje), budeme požadovat, aby střední hodnota odhadu T byla rovna skutečné hodnotě parametru θ_i.

Než uvedeme definici nevychýleného odhadu parametru rozdělení, uvědomme si ješte , že uvažujeme náhodnou veličinu X s rozdělovací funkcí $f = f(x;\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta$. Potom má podle věty 2.1 náhodný výběr $(X_1, X_2, \ldots, X_n)$ z X rozdělovací funkci

$$h = h(x_1, x_2 \ldots, x_n; \boldsymbol{\theta}) = f(x_1;\boldsymbol{\theta})\, f(x_2;\boldsymbol{\theta}) \ldots f(x_n;\boldsymbol{\theta}),\ \boldsymbol{\theta} \in \boldsymbol{\Theta}.$$

Pro konkrétní hodnotu $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ tedy obdržíme konkrétní rozdělení náhodného výběru $(X_1, X_2, \ldots, X_n)$ z X a tedy i konkrétní zákon rozdělení statistiky $T = t(X_1, X_2, \ldots, X_n)$. Dále budeme značit $E_{\boldsymbol{\theta}}(T)$ a $D_{\boldsymbol{\theta}}(T)$ střední hodnotu a rozptyl statistiky T, které obdržíme pro konkrétní hodnotu $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

Definice 3.2. *Nechť má náhodná veličina X rozdělovací funkci $f = f(x;\boldsymbol{\theta})$, kde $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m) \in \boldsymbol{\Theta}$ je neznámý parametr. Buď $(X_1, X_2, \ldots, X_n)$ náhodný výběr z X, $T = t(X_1, X_2, \ldots, X_n)$ statistika.*

Řekneme, že statistika T je nestranným nebo nevychýleným odhadem neznámého parametru θ_i, když

$$E_{\boldsymbol{\theta}}(T) = \theta_i \quad \text{pro každé } \boldsymbol{\theta} \in \boldsymbol{\Theta}.$$

Příklad 3.4. Zjistěte, zda jsou statistiky T_1, T_2 a T_3 z příkladu 3.3 nestrannými odhady parametru p alternativního rozdělení.

Řešení. Uvažujeme náhodnou veličinu $X \sim A(p)$, tj.

$$X \sim g = g(x;p) = \begin{cases} p & \text{pro } x = 1 \\ 1-p & \text{pro } x = 0 \\ 0 & \text{jinak} \end{cases},\ p \in (0,1)\,.$$

Podle definice 3.2 je statistika $T = t(X_1, \ldots, X_n)$ nestranným odhadem parametru p, když

$$E_p(T) = p \quad \text{pro každé } p \in (0,1)\,.$$

$(X_1, \ldots, X_n)$ je náhodný výběr z X a tedy $E_p(X_i) = E_p(X)$ pro $i = 1, \ldots, n$, protože veličiny $X_1, \ldots, X_n$ mají stejné rozdělení jako X. Zřejmě $E_p(X) = p$ a tedy $E_p(X_i) = p$ pro $i = 1, \ldots, n$. Potom

$$E_p(T_1) = E_p\Big(\frac{1}{n}\sum_{i=1}^{n} X_i\Big) = \frac{1}{n}\sum_{i=1}^{n} E_p(X_i) = \frac{1}{n}\sum_{i=1}^{n} p = p \quad \text{pro každé } p \in (0,1)\,.$$

Statistika T_1 je nestranným odhadem parametru p.
Podobně

$$E_p(T_2) = E_p(2X_2 - X_1) = 2p - p = p \quad \text{pro každé } p \in (0,1)\,.$$

A tedy i T_2 je nestranný odhad parametru p.

$$E_p(T_3) = E_p(X_1 + \cdots + X_n) = np \quad \text{pro každé } p \in (0,1)\,.$$

Statistika T_3 tedy není pro $n > 1$ nestranným odhadem parametru p.

♣

PDF-XChange
Click to buy NOW!
www.docu-track.com

Poznámka 3.3.

1. Z příkladu 3.4 je vidět, že T_1 a T_2 jsou nestrannými odhady parametru p. Přesto T_2 není vhodný. Nestrannost odhadu parametru p sice zaručuje, že jeho realizace kolísají okolo skutečné hodnoty parametru p, ale samozřejmě vhodnost nestranného odhadu závisí na tom, jaká je rozptýlenost realizací odhadu okolo skutečné hodnoty parametru.

2. Existují-li nestranné odhady parametru θ_i, pak samozřejmě budeme chtít použít ten nestranný odhad, který má ze všech nestranných odhadů parametru θ_i nejmenší rozptyl. Takový odhad (pokud existuje) budeme nazývat nejlepší nestranný odhad parametru θ_i.

Definice 3.3. *Nechť má náhodná veličina X rozdělovací funkci $f = f(x; \boldsymbol{\theta})$, kde $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m) \in \boldsymbol{\Theta}$ je neznámý parametr. Buď $(X_1, X_2, \ldots, X_n)$ náhodný výběr z X. Je-li $T = t(X_1, X_2, \ldots, X_n)$ nestranný odhad parametru θ_i a jestliže pro každý jiný nestranný odhad $T^* = t^*(X_1, \ldots, X_n)$ parametru θ_i platí*

$$D_{\boldsymbol{\theta}}(T) \leq D_{\boldsymbol{\theta}}(T^*) \quad \text{pro každé} \,; \boldsymbol{\theta} \in \boldsymbol{\Theta},$$

nazývá se statistika T <u>nejlepší nestranný odhad parametru θ_i</u>.

Příklad 3.5. Najděte nejlepší nestranný lineární odhad střední hodnoty normální náhodné veličiny X, která má neznámou střední hodnotu i rozptyl.

Řešení. Nechť $X \sim N(\mu, \sigma^2), (X_1, X_2, \ldots, X_n)$ je náhodný výběr z X. Potom $\boldsymbol{\theta} = (\theta_1, \theta_2) = (\mu, \sigma^2)$, $\boldsymbol{\Theta} = (-\infty, \infty) \times (0, \infty)$.
Máme najít odhad střední hodnoty μ, tj. statistiku $T = t(X_1, X_2, \ldots, X_n)$. Poněvadž požadujeme, aby odhad byl lineární, musí být

$$t(X_1, X_2, \ldots, X_n) = c_1 X_1 + c_2 X_2 + \cdots + c_n X_n, \tag{3.2}$$

kde $c_1, c_2, \ldots, c_n$ jsou reálná čísla, z nichž alespoň jedno je různé od nuly.
Vyjádřeme nyní střední hodnotu a rozptyl odhadu (3.2). Poněvadž $(X_1, \ldots, X_n)$ je náhodný výběr z X, jsou veličiny $X_1, X_2, \ldots, X_n$ nezávislé a všechny mají normální rozdělení se stejnými parametry μ, σ^2. Odsud podle poznámky P - 8.4 2) a věty P - 8.4 dostáváme

$$E_{\boldsymbol{\theta}}(c_1 X_1 + c_2 X_2 + \cdots + c_n X_n) = (c_1 + c_2 + \cdots + c_n)\mu, \tag{3.3}$$

$$D_{\boldsymbol{\theta}}(c_1 X_1 + c_2 X_2 + \cdots + c_n X_n) = (c_1^2 + c_2^2 + \cdots + c_n^2)\sigma^2. \tag{3.4}$$

Ze všech lineárních funkcí (3.2) máme určit ty, které jsou nestrannými odhady μ. Z definice 3.2 a vztahu (3.3) plyne, že pro tyto funkce musí platit

$$(c_1 + c_2 + \cdots + c_n)\mu = \mu \quad \text{pro každé } \boldsymbol{\theta} \in \boldsymbol{\Theta}. \tag{3.5}$$

Požadavek nevychýlenosti (3.5) lze tedy přepsat do tvaru

$$c_1 + c_2 + \cdots + c_n = 1. \tag{3.6}$$

Nyní mezi všemi nestrannnými lineárními odhady, tj. odhady určenými vztahy (3.2) a (3.6) hledáme nejlepší, tj. takový, pro nějž bude hodnota výrazu (3.4) minimální při každé hodnotě $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Hledáme tedy minimum funkce $c_1^2+c_2^2+\cdots+c_n^2$ za podmínky $c_1 + c_2 + \cdots + c_n = 1$. Běžnými metodami obdržíme minimum pro $c_1 = c_2 = \cdots = c_n = 1/n$. Tedy hledaný odhad je

$$t(X_1, X_2, \ldots, X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

♣

Poznamenejme, že parametr σ^2 se na výpočtu ani výsledku, stejně jako předpoklad normality veličiny X nepodílel.

Příklad 3.6. Aplikujte předchozí postup na vyřešení zadání: Máme dva různé výsledky (x_1, x_2) měření téže veličiny s různými přesnostmi, tj. x_1 je realizace veličiny $X_1 \sim N(\mu, \sigma_1^2)$ a x_2 je realizace veličiny $X_2 \sim N(\mu, \sigma_2^2)$. Veličiny X_1, X_2 jsou nezávislé, $\sigma_1^2 \neq \sigma_2^2$ jsou známá čísla. Určete nejlepší nevychýlené lineární zpracování dat za účelem odhadnutí neznámého μ.

Řešení. Označme G lineární zpracování dat, potom

$$G = c_1X_1 + c_2X_2,$$

kde c_1, c_2 jsou reálná čísla.

Požadujeme, aby G bylo nevychýlené zpracování, tj. musí platit

$$E_\mu(c_1X_1 + c_2X_2) = \mu \quad \text{pro každé } \mu,$$

odtud

$$c_1 + c_2 = 1. \tag{3.7}$$

Hledáme nejlepší nevychýlené zpracování, tedy

$$D_\mu(c_1X_1 + c_2X_2)$$

musí být minimální pro každé μ při současném splnění podmínky (3.7), tj.

$$c_1^2\sigma_1^2 + c_2^2\sigma_2^2$$

musí být minimální při současném splnění podmínky (3.7).

Hledáme tedy absolutní minimum funkce $H(c_1, c_2) = c_1^2\sigma_1^2 + c_2^2\sigma_2^2$ při současném splnění podmínky (3.7), tj. při splnění podmínky $c_2 = 1 - c_1$. Stačí tedy najít absolutní minimum funkce

$$K(c_1) = H(c_1, 1 - c_1) = c_1^2\sigma_1^2 + (1 - c_1)^2\sigma_2^2.$$

Vzhledem k tomu, že

$$\frac{dK(c_1)}{dc_1} = 2c_1\sigma_1^2 - 2(1 - c_1)\sigma_2^2,$$

dostaneme, že stacionární bod funkce $K(c_1)$ je

$$c_1 = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}.$$

Protože

$$\frac{d^2 K(c_1)}{dc_1^2} = 2\sigma_1^2 + 2\sigma_2^2 > 0 \quad \text{pro každé } c_1,$$

nabývá funkce $K(c_1)$ ve svém stacionárním bodě absolutní minimum. Z podmínky (3.7) dostáváme

$$c_2 = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}.$$

Nejlepší lineární zpracování dat je tedy

$$G = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} X_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} X_2.$$

♣

K další často požadované vlastnosti odhadu patří tzv. konzistence odhadu.

Definice 3.4. *Buď X náhodná veličina s rozdělovací funkcí $f = f(x;\boldsymbol{\theta})$, kde $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m) \in \boldsymbol{\Theta}$ je neznámý parametr, $(X_1, X_2, \ldots, X_n)$ náhodný výběr z X. Odhad $T = T_n = t(X_1, X_2, \ldots, X_n)$ parametru θ_i nazýváme konzistentní odhad parametru θ_i, jestliže pro každé $\epsilon > 0$ a pro každé $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ platí*

$$\lim_{n\to\infty} P_{\boldsymbol{\theta}}\big(|T_n - \theta_i| < \epsilon\big) = 1.$$

Poznámka 3.4.

1) Je-li T_n konzistentní odhad parametru θ_i, pak z definice limity a definice 3.4 plyne, že pro libovolné $\epsilon > 0$ a libovolné $\delta > 0$ existuje rozsah výběru n_0 tak, že pro každý rozsah výběru n větší než n_0 platí

$$P_{\boldsymbol{\theta}}\big(|T_n - \theta_i| < \epsilon\big) > 1 - \delta,$$

tedy

$$P_{\boldsymbol{\theta}}\big(T_n - \epsilon < \theta_i < T_n + \epsilon\big) > 1 - \delta,$$

tj. s „dostatečně" velkou pravděpodobností bude T_n „dostatečně" blízko skutečné hodnotě parametru θ_i. Kdybychom tedy do konzistentního odhadu parametru θ_i zpracovali náhodný výběr nekonečného rozsahu, dostali bychom „jistě" skutečnou hodnotu parametru θ_i.

2) Pro ověření konzistence odhadu lze využít následující větu, kterou uvádíme bez důkazu.

Věta 3.1. Buď X náhodná veličina s rozdělovací funkcí $f = f(x;\boldsymbol{\theta})$, kde $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m) \in \boldsymbol{\Theta}$ je neznámý parametr, $(X_1, X_2, \ldots, X_n)$ náhodný výběr z X. Odhad $T_n = t(X_1, X_2, \ldots, X_n)$ parametru θ_i je konzistentní odhad parametru θ_i, jestliže pro každé $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ platí

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

PDF-Xchange Click to buy NOW! www.docu-track.com

Příklad 3.7. Zjistěte, zda je statistika $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ konzistentním odhadem střední hodnoty μ normálního rozdělení.

Řešení. Buď $(X_1, X_2, \ldots, X_n)$ náhodný výběr z rozdělení $X \sim N(\mu, \sigma^2)$. Potom $\boldsymbol{\theta} = (\theta_1, \theta_2) = (\mu, \sigma^2), \Theta = (-\infty, \infty) \times (0, \infty)$. Označme $T_n = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Podle příkladu P - 8.6 platí

$$E_{\boldsymbol{\theta}}(\overline{X}) = \mu,$$
$$D_{\boldsymbol{\theta}}(\overline{X}) = \frac{\sigma^2}{n}.$$

Odtud pro každé $\boldsymbol{\theta} \in \Theta$

$$\lim_{n\to\infty} E_{\boldsymbol{\theta}}(T_n) = \lim_{n\to\infty} E_{\boldsymbol{\theta}}(\overline{X}) = \lim_{n\to\infty} \mu = \mu,$$
$$\lim_{n\to\infty} D_{\boldsymbol{\theta}}(T_n) = \lim_{n\to\infty} D_{\boldsymbol{\theta}}(\overline{X}) = \lim_{n\to\infty} \frac{\sigma^2}{n} = 0.$$

Tedy $\overline{X}$ je konzistentní odhad střední hodnoty μ. Poznamenejme opět, že se na výpočtu nepodílel předpoklad normálního rozdělení.

♣

Na závěr kapitoly uvedeme větu o odhadu střední hodnoty a rozptylu náhodné veličiny X, jejíž první část jsme dokázali v příkladu 3.5.

Věta 3.2. Buď X náhodná veličina s konečnou střední hodnotou μ a konečným rozptylem σ^2, $(X_1, X_2, \ldots, X_n)$ náhodný výběr z X. Potom je statistika

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

nejlepším nestranným lineárním odhadem střední hodnoty μ náhodné veličiny X. Statistiky

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2,$$
$$S_0^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$$

jsou nestranné odhady rozptylu σ^2 téže náhodné veličiny X.

Poznámka 3.5.

1. Statistiku S^2 používáme pro odhad neznámého rozptylu v situaci, kdy je neznámá i střední hodnota; statistiku S_0^2 používáme pro odhad rozptylu při známé střední hodnotě.
2. Věta 3.2 platí pro náhodnou veličinu X s libovolným rozdělením a ne nutně normálním.
3. Pro odhad směrodatné odchylky σ používáme statistiku S, resp. S_0.
4. Odhady $\overline{X}, S^2, S_0^2, S, S_0$ jsou konzistentní odhady příslušných parametrů.

5. Pro odhad rozptylu σ^2 se někdy používá i statistika

$$M = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

Zřejmě $M = \dfrac{n-1}{n} S^2$ a tedy

$$E(M) = \frac{n-1}{n}\sigma^2 = \sigma^2 - \frac{1}{n}\sigma^2.$$

Tedy statistika M není nestranný odhad σ^2 (je to doleva vychýlený odhad). Je ale konzistentním odhadem σ^2 (dokažte!) a navíc platí

$$E(M - \sigma^2)^2 \leq E(S^2 - \sigma^2)^2 = D(S^2).$$

Statistika M má tedy menší střední čtvercovou odchylku od σ^2 než statistika S^2.

6. Pro výpočet realizací odhadů μ a σ^2 z věty 3.2 lze použít jakýkoliv statistický software nebo EXCEL bez dosazování do vzorců. Na nižší úrovni pak kalkulačky, které umožňují statistické výpočty. Realizaci $\overline{x}$ spočítáme pomocí programu na výpočet $\overline{x}$. Na některých kalkulačkách se vyskytuje dvojice s a σ, na některých σ_{n-1} a σ_n. Platí pro ně následující vztahy

$$s = \sigma_{n-1} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

$$\sigma = \sigma_n = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2} \quad (= \sqrt{m})$$

Výpočet realizace s_0 nelze na kalkulačkách ani na žádném mně známém software přímo spočítat. Musí se využít vztah z věty 3.2.

7) Když je rozsah n náhodného výběru z X velký, třídíme realizaci náhodného výběru pro přehlednost i další analýzu do k disjunktních tříd Ω_j $(j = 1, \ldots, k)$. Toto třídění je subjektivní, i když existují jistá objektivní pravidla (viz kapitola P - 1). Nebudeme se zde ale tříděním detailně zabývat. Vzhledem k jeho pracnosti by stejně každý použil nějaký vhodný software (např. EXCEL, UNISTAT, STATGRAPHICS). Zmíníme se pouze o odhadu číselných charakteristik z roztříděných dat.

Připomeňme, že jsme označili n_j počet hodnot, které padly do j-té třídy Ω_j - tzv. absolutní četnost j-té třídy $(j = 1, \ldots, k)$. Zřejmě

$$n_1 + \cdots + n_k = n.$$

Označme $\overline{x}_j$ střed j-té třídy. Hodnoty, které padly do j-té třídy nahradíme středem $\overline{x}_j$. Potom

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \doteq \frac{1}{n}\sum_{j=1}^{k} n_j \overline{x}_j,$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 \doteq \frac{1}{n-1}\sum_{j=1}^{k} n_j(\overline{x}_j - \overline{x})^2,$$

$$s_0^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2 \doteq \frac{1}{n}\sum_{j=1}^{k} n_j(\overline{x}_j - \mu)^2.$$

V diskrétním případě, tj. v případě prostého třídění, platí tyto vzorce přesně.

8) Zřejmě platí

$$\sum_{i=1}^{n}(x_i - \overline{x})^2 = \sum_{i=1}^{n}(x_i^2 - 2x_i\overline{x} + \overline{x}^2) = \sum_{i=1}^{n} x_i^2 - 2\overline{x}\cdot\sum_{i=1}^{n} x_i + \sum_{i=1}^{n}\overline{x}^2 =$$

$$= \sum_{i=1}^{n} x_i^2 - 2\overline{x}\cdot n\cdot\overline{x} + n\cdot\overline{x}^2 = \sum_{i=1}^{n} x_i^2 - n\cdot\overline{x}^2.$$

Podobně

$$\sum_{i=1}^{n}(x_i - \mu)^2 = \sum_{i=1}^{n} x_i^2 - 2n\overline{x}\mu + n\mu^2.$$

Odtud a z věty 3.2 dostáváme

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\overline{x}^2\right),$$

podobně

$$s_0^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - 2\overline{x}\mu + \mu^2.$$

V případě roztříděných dat

$$s^2 \doteq \frac{1}{n-1}\left(\sum_{j=1}^{k} n_j\overline{x}_j^2 - n\overline{x}^2\right),$$

$$s_0^2 \doteq \frac{1}{n}\sum_{j=1}^{k} n_j\overline{x}_j^2 - 2\overline{x}\mu + \mu^2.$$

Podobným výsledkům je věnováno ve starší statistické literatuře poměrně mnoho prostoru; v souvislosti s rozvojem výpočetní techniky význam pro běžného uživatele statistiky klesá.

PDF-Xchange Click to buy NOW! www.docu-track.com

PDF-XChange Click to buy NOW! www.docu-track.com

Příklad 3.8. Při měření veličiny konstantní délky byly zjištěny následující chyby měření v mm

1, -2, -1, 0, -3, 2, -1, -1, 1, -1.

Určete

1) realizaci bodového odhadu střední hodnoty a rozptylu chyby měření,

2) realizaci bodového odhadu rozptylu chyby měření, když je známo, že měřící přístroj vykazuje systematickou chybu -1 mm,

3) realizaci bodového odhadu střední hodnoty chyby měření, když je známo, že přesnost měřícího přístroje vyjádřená směrodatnou odchylkou je 2 mm.

Řešení. Náhodnou veličinou X je zde náhodná chyba měření. Označme $E(X) = \mu$, $D(X) = \sigma^2$. K dispozici máme realizaci náhodného výběru z X o rozsahu $n = 10$. Máme najít realizace odhadů $\widehat{\mu}$ a $\widehat{\sigma^2}$ parametrů μ a σ^2 v jednotlivých situacích.

1) Jde o situaci, kdy ani μ ani σ^2 neznáme. Podle věty 3.2 dostaneme

$$\widehat{\mu} = \overline{x} = \frac{1}{10}\sum_{i=1}^{10} x_i = -0.50 \text{ [mm]},$$
$$\widehat{\sigma^2} = s^2 = \frac{1}{9}\sum_{i=1}^{10}(x_i + 0.5)^2 \doteq 2.28 \text{ [mm}^2\text{]}.$$

Přímo (tj. bez dosazování) s využitím kalkulačky podle poznámky 3.5 6)

$$\widehat{\mu} = \overline{x} = -0.50 \text{ [mm]},$$
$$\widehat{\sigma} = s = \sigma_{n-1} \doteq 1.51 \text{ [mm]} \Rightarrow \widehat{\sigma^2} \doteq 2.28 \text{ [mm}^2\text{]}.$$

Nebo s využitím poznámky 3.5 8) a skutečnosti, že $\sum_{i=1}^{10} x_i = 23$, dostaneme

$$\widehat{\sigma^2} = \frac{1}{9}[23 - 10 \cdot (-0.5)^2] \doteq 2.28 \text{ [mm}^2\text{]}.$$

2) Jde o situaci, kdy víme, že $\mu = -1$ mm. Podle věty 3.2 je

$$\widehat{\sigma^2} = s_0^2 = \frac{1}{10}\sum_{i=1}^{10}(x_i + 1)^2 = 2.30 \text{ [mm}^2\text{]}.$$

(Na kalkulačce lze využít nabídky $\sum x^2$.)
Podle poznámky 3.5 8)

$$\widehat{\sigma^2} = s_0^2 = \frac{1}{10}23 - 2 \cdot (-0.5) \cdot (-1) + (-1)^2 = 2.30 \text{ [mm}^2\text{]}.$$

3) Víme, že $\sigma = 2$ mm, tato informace ale neovlivňuje odhad μ, tedy $\widehat{\mu} = \overline{x} = -0.5$ mm .

♣

Příklad 3.9. Při stavbě betonové konstrukce bylo odebráno 40 vzorků betonové směsi. Po 28 dnech vykázaly kostky tuto krychlovou pevnost v MPa:

23.5,	28.0,	25.1,	30.8,	27.1,	29.3,	32.5,	33.8,	30.4,	26.2,
30.8,	29.2,	30.9,	28.6,	27.5,	28.0,	31.2,	28.2,	30.7,	28.8,
32.7,	29.0,	31.9,	25.4,	32.6,	27.4,	33.1,	29.6,	29.7,	30.3,
26.8,	30.4,	25.6,	34.0,	34.8,	27.2,	31.5,	32.3,	29.7,	32.4

Určete realizaci bodového odhadu směrodatné odchylky pevnosti betonu

1) při neznámé střední hodnotě pevnosti
2) víte-li, že střední hodnota pevnosti je 28.0 MPa.

Řešení. Náhodnou veličinou X je zde pevnost betonu, označme $E(X) = \mu, D(X) = \sigma^2$. Vzhledem k tomu, že rozsah výběru $n = 40$ je velký, roztřídíme realizaci náhodného výběru z X do k tříd Ω_j $(j = 1, \dots, k)$ (viz kapitola P - 1). Za třídy Ω_j volíme intervaly (stejné délky d), protože se jedná o spojitou náhodnou veličinu. Počet k intervalů zvolíme $k \doteq \sqrt{n} = \sqrt{40} \doteq 6$. Variační obor je interval $\langle x_{min}, x_{max}\rangle = \langle 23.5, 34.8\rangle$. Zřejmě $\langle 23, 35\rangle \supset \langle 23.5, 34.8\rangle$, délku d intervalů tedy zvolme $d = \dfrac{35-23}{6} = 2$. Výsledek třídění pak ukazuje následující tabulka.

Třída	$\langle t_{j-1}, t_j)$	n_j	$\overline{x}_j$
Ω_1	23 – 25	1	24
Ω_2	25 – 27	5	26
Ω_3	27 – 29	10	28
Ω_4	29 – 31	12	30
Ω_5	31 – 33	8	32
Ω_6	33 – 35	4	34

Tabulka 3.1: Rozdělení absolutních četností pevnosti betonu

Odhad $\widehat{\sigma}$ směrodatné odchylky σ budeme počítat z takto roztříděných dat.

1) Podle poznámky 3.5 7) dostaneme

$$\widehat{\mu} = \overline{x} \doteq \frac{1}{40}\sum_{j=1}^{6} n_j \overline{x}_j = 29.650 \text{ [MPa]},$$

$$\widehat{\sigma^2} = s^2 \doteq \frac{1}{39}\sum_{j=1}^{6} n_j(\overline{x}_j - 29.650)^2 \doteq 6.335 \text{ [MPa}^2\text{]} \Rightarrow \widehat{\sigma} = s \doteq 2.517 \text{ [MPa]}.$$

Nebo s využitím kalkulačky přímo (bez dosazování) je

$$\widehat{\sigma} = \sigma_{n-1} \doteq 2.517 \text{ [MPa]}.$$

(Hodnotu $\overline{x}_j$ uložíme n_j-krát na většině kalkulaček tak, že $\overline{x}_j \times n_j$ uložíme pomocí nabídky DATA.)

2) Podle poznámky 3.5 7) dostaneme

$$s_0^2 \doteq \frac{1}{40}\sum_{j=1}^{6} n_j(\overline{x}_j - 28)^2 \doteq 8.9 \text{ [MPa}^2\text{]} \Rightarrow \widehat{\sigma} = s_0 \doteq 2.98 \text{ [MPa]}.$$

(Na kalkulačce lze opět využít nabídku $\sum x^2$.)

4. Intervalový odhad parametrů rozdělení

Položme si přirozenou otázku: Jaká je pravděpodobnost, že realizace bodového odhadu neznámého parametru rozdělení bude shodná s jeho skutečnou hodnotou? Zjistíme, že ve většině případů nulová. (Neboť bodový odhad parametru je náhodná veličina; je-li např. tato veličina spojitá, pak pravděpodobnost, že nabude konstantní hodnoty, je vždy rovna nule.) Proto kromě bodového odhadu parametru θ_i hledáme tzv. intervalový odhad parametru θ_i, tj. takový interval, který s předem danou pravděpodobností překryje skutečnou hodnotu tohoto parametru.

Buď X náhodná veličina s rozdělovací funkcí $f = f(x;\boldsymbol{\theta})$; $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Nechť je skutečná hodnota parametru rovna $\boldsymbol{\theta}$. Potom pravděpodobnost, že náhodná veličina X nabude hodnoty z množiny B značíme jako $P_{\boldsymbol{\theta}}(X \in B)$.

Definice 4.1. *Buď X náhodná veličina s rozdělovací funkcí $f = f(x;\boldsymbol{\theta})$ kde $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_m) \in \boldsymbol{\Theta}$ a nechť $(X_1, X_2, \dots, X_n)$ je náhodný výběr z X. Buďte*

$$T_1 = t_1(X_1, X_2, \dots, X_n), \quad T_2 = t_2(X_1, X_2, \dots, X_n)$$

dvě statistiky, pro které na celém výběrovém prostoru platí

$$t_1(x_1, x_2, \dots, x_n) \leq t_2(x_1, x_2, \dots, x_n).$$

Buď α reálné číslo z intervalu $(0,1)$.

Řekneme, že interval (dvojice statistik) $\langle T_1, T_2 \rangle$ je 100 · (1 − α) procentní intervalový odhad (parametru) θ_i nebo 100 · (1 − α) procentní interval spolehlivosti pro (parametr) θ_i ($i \in \{1, \dots, m\}$), když pro každé $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ platí

$$P_{\boldsymbol{\theta}}\big(t_1(X_1, X_2, \dots, X_n) \leq \Theta_i \leq t_2(X_1, X_2, \dots, X_n)\big) \geq 1 - \alpha.$$

Číslo $1-\alpha$ nazýváme spolehlivost odhadu, číslo α riziko odhadu.

Dosadíme-li realizaci $(x_1, x_2, \dots, x_n)$ náhodného výběru $(X_1, X_2, \dots, X_n)$ z X do intervalového odhadu $\langle t_1(X_1, X_2, \dots, X_n), t_2(X_1, X_2, \dots, X_n)\rangle$, dostaneme reálný interval $\langle t_1, t_2\rangle = \langle t_1(x_1, x_2, \dots, x_n), t_2(x_1, x_2, \dots, x_n)\rangle$, tzv. realizaci intervalového odhadu.

Poznámka 4.1. Konce intervalu spolehlivosti tvoří náhodné veličiny, proto mluvíme o náhodném intervalu. Pravděpodobnost, že tento interval překryje skutečnou hodnotu parametru θ_i, je větší nebo rovna danému číslu $1-\alpha$ (a podle definice nezávisí na $\boldsymbol{\theta}$). Je zapotřebí si uvědomit správný výklad intervalového odhadu. Nemůže se totiž říci, že pravděpodobnost padnutí Θ_i do intervalu spolehlivosti je alespoň $1-\alpha$. Parametr θ_i je pevné číslo (i když neznámé), které nemůže nikam padat. Náhodný je interval spolehlivosti, jehož realizace buď θ_i obsahuje, nebo nikoliv. Zařídili jsme však, aby pravděpodobnost, že interval spolehlivosti obsahuje

θ_i, byla alespoň $1-\alpha$ (tj. předepsaně velká). Vypočteme-li tedy velký počet realizací $100 \cdot (1-\alpha)$ procentního intervalového odhadu parametru θ_i, pak průměrně alespoň $100(1-\alpha)\%$ těchto realizací překrývá skutečnou hodnotu parametru θ_i.

Dosavadní technické normy vyžadují většinou počítat 99%-ní nebo 95%-ní intervaly spolehlivosti (tj. volbu $\alpha = 0.01$ nebo $\alpha = 0.05$).

Dále se budeme převážně zabývat normální náhodnou veličinou. Z tohoto důvodu na tomto místě uvedeme následující větu, kterou budeme často používat.

Věta 4.1. Buď $(X_1, \ldots, X_n)$ náhodný výběr z rozdělení $X \sim N(\mu, \sigma^2)$. Potom

1. $\dfrac{\overline{X}-\mu}{\sigma}\sqrt{n} \sim N(0,1)$,
2. $\dfrac{\overline{X}-\mu}{S}\sqrt{n} \sim t(n-1)$,
3. $\dfrac{nS_0^2}{\sigma^2} \sim \chi^2(n)$,
4. $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$.

Poznámka 4.2. Vztah 1 věty 4.1 vyplývá z věty P - 9.2.3 (viz poznámka P - 9.2.3 2)). Ze zbývajících vztahů ukážeme platnost vztahu 3. Buď tedy $(X_1, \ldots, X_n)$ náhodný výběr z rozdělení $X \sim N(\mu, \sigma^2)$. Podle věty 3.2 je

$$S_0^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2.$$

Podle poznámky P - 9.2.3 platí

$$\frac{X_i-\mu}{\sigma} \sim N(0,1)$$

a podle poznámky P - 9.2.6

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

Zřejmě

$$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 = \frac{nS_0^2}{\sigma^2},$$

odtud plyne vztah 3.

Příklad 4.1. Odvoďte $100 \cdot (1-\alpha)$ procentní intervalový odhad střední hodnoty μ normální náhodné veličiny X se známým rozptylem $\sigma^2 > 0$.

Řešení. Buď $(X_1, X_2, \ldots, X_n)$ náhodný výběr z rozdělení X, tj. z rozdělení $N(\mu, \sigma^2)$. Poněvadž $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ je nestranným odhadem parametru μ, zdá se být rozumné hledat kladné konstanty k_1 a k_2 tak, aby interval

$$\langle \overline{X}-k_1, \overline{X}+k_2 \rangle \tag{4.1}$$

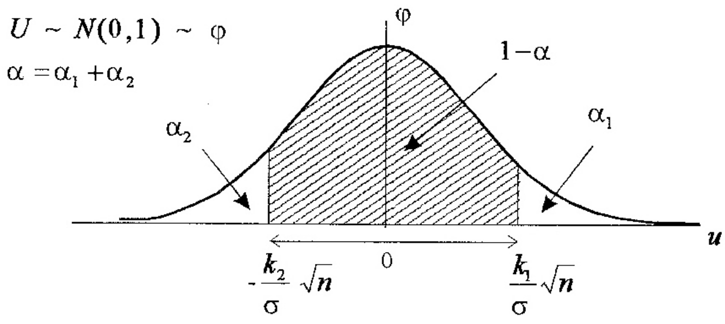
překrýval skutečnou střední hodnotu μ s velkou pravděpodobností. Určíme-li konstanty k_1, k_2 tak, aby platilo

$$(4.2) \qquad P_\mu\big(\overline{X} - k_1 \le \mu \le \overline{X} + k_2\big) \ge 1 - \alpha \quad \text{pro dané } \sigma^2 \text{ a každé } \mu,$$

bude interval (4.1) $100 \cdot (1-\alpha)$ procentním intervalovým odhadem parametru μ. Podle vztahu 1 věty 4.1 platí

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

a rozdělení U nezávisí na μ a σ^2.



Obrázek 4.1

Vyjádřeme tedy pravděpodobnost ve vztahu (4.2) pomocí náhodné veličiny U. Postupně dostaneme

$$\begin{aligned}
P_\mu\big(-k_1 \le \mu - \overline{X} \le k_2\big) &= P_\mu\big(-k_2 \le \overline{X} - \mu \le k_1\big) \\
&= P_\mu\Big(-\frac{k_2}{\sigma}\sqrt{n} \le \frac{\overline{X} - \mu}{\sigma}\sqrt{n} \le \frac{k_1}{\sigma}\sqrt{n}\Big) \\
&= P\Big(-\frac{k_2}{\sigma}\sqrt{n} \le U \le \frac{k_1}{\sigma}\sqrt{n}\Big) \\
&= 1 - P\Big(U \notin \langle -\frac{k_2}{\sigma}\sqrt{n}, \frac{k_1}{\sigma}\sqrt{n}\rangle\Big).
\end{aligned}$$

Vidíme, že pravděpodobnost ve vztahu (4.2) je určena jednoznačně. Stačí tedy nyní určit konstanty k_1, k_2 tak, aby platilo

$$1 - \alpha = 1 - P\Big(U \notin \langle -\frac{k_2}{\sigma}\sqrt{n}, \frac{k_1}{\sigma}\sqrt{n}\rangle\Big).$$

Odtud

$$\begin{aligned}
\alpha &= P\Big(U \notin \langle -\frac{k_2}{\sigma}\sqrt{n}, \frac{k_1}{\sigma}\sqrt{n}\rangle\Big) = P\Big(U < -\frac{k_2}{\sigma}\sqrt{n} \vee U > \frac{k_1}{\sigma}\sqrt{n}\Big) \\
&= P\Big(U < -\frac{k_2}{\sigma}\sqrt{n}\Big) + P\Big(U > \frac{k_1}{\sigma}\sqrt{n}\Big).
\end{aligned}$$

Rozdělme nyní riziko α na dvě nezáporná čísla α_1, α_2, tj.

$$\alpha = \alpha_1 + \alpha_2, \ \alpha_1 \geq 0, \ \alpha_2 \geq 0.$$

Nyní stačí určit čísla k_1, k_2 tak, aby platilo (viz obrázek 4.1)

$$P\Big(U < -\frac{k_2}{\sigma}\sqrt{n}\Big) = \alpha_2, \quad P\Big(U > \frac{k_1}{\sigma}\sqrt{n}\Big) = \alpha_1. \tag{4.3}$$

Ze vztahu (4.3) dostaneme

$$\frac{k_1}{\sigma}\sqrt{n} = u(1-\alpha_1) \Rightarrow k_1 = u(1-\alpha_1)\frac{\sigma}{\sqrt{n}},$$

$$-\frac{k_2}{\sigma}\sqrt{n} = u(\alpha_2) = -u(1-\alpha_2) \Rightarrow k_2 = u(1-\alpha_2)\frac{\sigma}{\sqrt{n}}.$$

Hledaný $100 \cdot (1-\alpha)$ procentní intervalový odhad parametru μ je

$$\Big\langle \overline{X} - u(1-\alpha_1)\frac{\sigma}{\sqrt{n}}, \overline{X} + u(1-\alpha_2)\frac{\sigma}{\sqrt{n}}\Big\rangle.$$

♣

Dokázali jsme následující větu:

Věta 4.2. Buď $(X_1, X_2, \ldots, X_n)$ náhodný výběr z rozdělení $X \sim N(\mu, \sigma^2)$, σ^2 známé, μ neznámé, $\alpha \in (0,1)$ dané číslo; α_1, α_2 nezáporná čísla, $\alpha_1 + \alpha_2 = \alpha$. Potom je interval

$$\Big\langle \overline{X} - u(1-\alpha_1)\frac{\sigma}{\sqrt{n}}, \overline{X} + u(1-\alpha_2)\frac{\sigma}{\sqrt{n}}\Big\rangle$$

$100 \cdot (1-\alpha)$ procentním intervalovým odhadem střední hodnoty μ.

Poznámka 4.3. Riziko α lze rozdělit na α_1, α_2 nekonečně mnoha způsoby. Běžně se používají pouze tři:

1. $\alpha_1 = \alpha_2 = \dfrac{\alpha}{2}$
2. $\alpha_1 = 0$, $\alpha_2 = \alpha$
3. $\alpha_1 = \alpha$, $\alpha_2 = 0$.

Jestliže platí

$$\alpha_1 > 0, \ \alpha_2 > 0,$$

dostaneme tzv. oboustranný intervalový odhad. Tedy při prvním způsobu dělení rizika α dostaneme oboustranný intervalový odhad. Jestliže volíme druhý způsob, pak zřejmě

$$u(1-\alpha_1) = u(1) = \infty, \ u(1-\alpha_2) = u(1-\alpha).$$

Dostaneme tedy intervalový odhad

$$\Big(-\infty, \overline{X} + u(1-\alpha)\frac{\sigma}{\sqrt{n}}\Big\rangle \tag{4.4}$$

Ve třetím případě, pak zcela analogicky dostáváme intervalový odhad

$$\Big\langle \overline{X} - u(1-\alpha)\frac{\sigma}{\sqrt{n}}, \infty\Big) \tag{4.5}$$

Intervalové odhady (4.4) a (4.5) jsou tzv. jednostranné intervalové odhady, odhad (4.4) je tzv. pravostranný nebo horní odhad, odhad (4.5) je tzv. levostranný nebo dolní odhad.

Příklad 4.2.

1. Střední hodnota chyby měření určitým přístrojem bude odhadována oboustranným intervalovým odhadem.

2. Střední hodnota pevnosti nějakého konstrukčního materiálu bude při zkouškách odhadována pomocí dolního intervalového odhadu (nezáleží na tom, zda materiál má lepší vlastnosti než požaduje konstruktér).

3. Střední hodnota podílu zmetků v dodávce výrobků bude při přejímce odhadována horním intervalovým odhadem (odběrateli nezáleží na dolním odhadu podílu zmetků).

♣

Věta 4.3. Buď $(X_1, X_2, \ldots, X_n)$ náhodný výběr z rozdělení $X \sim N(\mu, \sigma^2), \sigma^2$ neznámé, μ známé, $\alpha \in (0,1)$ dané číslo, α_1, α_2 nezáporná čísla, $\alpha_1 + \alpha_2 = \alpha$. Potom je interval

$$\left\langle \frac{nS_0^2}{\chi^2(n; 1-\alpha_1)}, \frac{nS_0^2}{\chi^2(n;\alpha_2)} \right\rangle$$

$100 \cdot (1-\alpha)$ procentním intervalovým odhadem rozptylu σ^2.

Důkaz. Buď $(X_1, X_2, \ldots, X_n)$ náhodný výběr z X, tj. z rozdělení $N(\mu, \sigma^2)$. Poněvadž

$$S_0^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$$

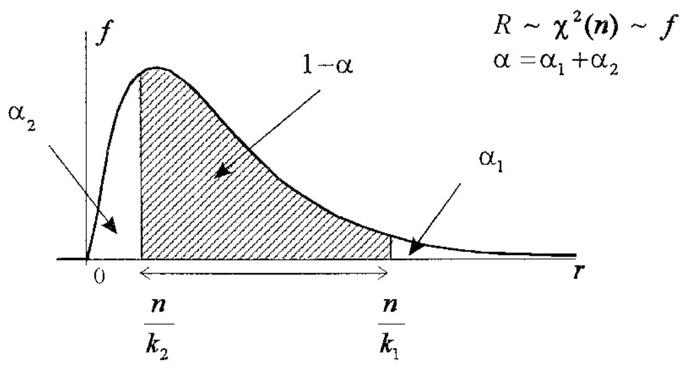
je nestranným odhadem parametru σ^2 (viz věta 3.2), překrývá interval

$$\left\langle S_0^2 k_1, S_0^2 k_2 \right\rangle, \tag{4.6}$$

kde $k_1 < k_2$ jsou kladné konstanty, skutečnou hodnotu parametru σ^2 s jistou pravděpodobností. Určíme-li konstanty k_1, k_2 tak, aby platilo

$$P_\sigma\left(S_0^2 k_1 \le \sigma^2 \le S_0^2 k_2\right) \ge 1-\alpha \quad \text{pro dané } \mu \text{ a každé } \sigma^2, \tag{4.7}$$

bude interval (4.6) $100 \cdot (1-\alpha)$ procentním intervalovým odhadem rozptylu σ^2.



Obrázek 4.2

Podle vztahu 3 věty 4.1 platí

$$R = \frac{nS_0^2}{\sigma^2} \sim \chi^2(n)$$

a rozdělení R nezávisí na μ a σ^2. Vyjádřeme pravděpodobnost ve vztahu (4.7) pomocí náhodné veličiny R a určeme konstanty k_1, k_2. Postupně pro dané μ a každé σ^2 dostáváme

$$1 - \alpha \geq P_\sigma\left(k_1 \leq \frac{\sigma^2}{S_0^2} \leq k_2\right) = P_\sigma\left(\frac{1}{k_2} \leq \frac{S_0^2}{\sigma^2} \leq \frac{1}{k_1}\right)$$
$$= P_\sigma\left(\frac{n}{k_2} \leq \frac{nS_0^2}{\sigma^2} \leq \frac{n}{k_1}\right) = P\left(\frac{n}{k_2} \leq R \leq \frac{n}{k_1}\right).$$

Zcela analogicky jako v příkladu 4.1 rozdělíme riziko α na dvě nezáporná čísla α_1, α_2 (viz obrázek 4.2), dostaneme

$$\frac{n}{k_2} = \chi^2(n; \alpha_2) \Rightarrow k_2 = \frac{n}{\chi^2(n; \alpha_2)},$$
$$\frac{n}{k_1} = \chi^2(n; 1 - \alpha_1) \Rightarrow k_1 = \frac{n}{\chi^2(n; 1 - \alpha_1)}.$$

Hledaný $100 \cdot (1 - \alpha)$ procentní intervalový odhad parametru σ^2 je tedy

$$\left\langle \frac{nS_0^2}{\chi^2(n; 1 - \alpha_1)}, \frac{nS_0^2}{\chi^2(n; \alpha_2)} \right\rangle.$$

Položíme-li $\alpha_1 = 0, \alpha_2 = \alpha$, potom je

$$\chi^2(n; 1 - \alpha_1) = \chi^2(n; 1) = \infty, \; \chi^2(n; \alpha_2) = \chi^2(n; \alpha)$$

a tedy

$$\left(0, \frac{nS_0^2}{\chi^2(n; \alpha)}\right\rangle$$

je $100 \cdot (1 - \alpha)$ procentní horní intervalový odhad σ^2.
Podobně, pro $\alpha_2 = 0, \alpha_1 = \alpha$ dostaneme

$$\left\langle \frac{nS_0^2}{\chi^2(n; 1 - \alpha)}, \infty\right),$$

což je $100 \cdot (1 - \alpha)$ procentní dolní intervalový odhad σ^2. □

Poznámka 4.4. Doposud jsme se zabývali situací, kdy známe právě jeden z parametrů normálního rozdělení. Zbývá vyřešit ještě situaci, kdy neznáme ani jeden z parametrů μ, σ^2. Při hledání intervalového odhadu střední hodnoty μ postupujeme v tomto případě analogicky jako v příkladu 4.1. Stačí si uvědomit, že nestranným odhadem μ zůstává $\overline{X}$ a využít vztah 2 věty 4.1. Tj. pracovat s veličinou $R = \frac{\overline{X}-\mu}{S}\sqrt{n}$, která má rozdělení $t(n-1)$. Při hledání intervalového odhadu σ^2 pak vyjdeme z nestranného odhadu σ^2, tj. statistiky S^2 a využijeme vztah 4 věty 4.1. Postupujeme zcela analogicky jako v důkazu věty 4.3. Výsledky shrnuje následující věta.

Věta 4.4. Buď $(X_1, X_2, \ldots, X_n)$ náhodný výběr z rozdělení $X \sim N(\mu, \sigma^2), \sigma^2$ neznámé a μ neznámé, $\alpha \in (0,1)$ dané číslo, α_1, α_2 nezáporná čísla, $\alpha_1 + \alpha_2 = \alpha$. Potom je interval

$$\left\langle \overline{X} - t(n-1; 1-\alpha_1)\frac{S}{\sqrt{n}}, \overline{X} + t(n-1; 1-\alpha_2)\frac{S}{\sqrt{n}} \right\rangle$$

$100 \cdot (1-\alpha)$ procentním intervalovým odhadem střední hodnoty μ a interval

$$\left\langle \frac{(n-1)S^2}{\chi^2(n-1; 1-\alpha_1)}, \frac{(n-1)S^2}{\chi^2(n-1; \alpha_2)} \right\rangle$$

je $100 \cdot (1-\alpha)$ procentním intervalovým odhadem rozptylu σ^2.

Horní [dolní] intervalové odhady dostaneme opět volbou

$$\alpha_1 = 0, \; \alpha_2 = \alpha \qquad [\alpha_1 = \alpha, \; \alpha_2 = 0].$$

Poznámka 4.5. V technických aplikacích se většinou $100 \cdot (1-\alpha)$ procentní intervalové odhady neznámých parametrů neodvozují, ale stačí je najít v příslušné statistické nebo technické literatuře. Pro výpočet realizace intervalového odhadu pak stačí zvolit spolehlivost odhadu $1-\alpha$, vypočítat realizace příslušných statistik v intervalových odhadech se vyskytujících a potřebné kvantily najít ve statistických tabulkách.

Realizace intervalových odhadů střední hodnoty a rozptylu normální náhodné veličiny při obou neznámých parametrech počítá STATGRAPHICS v nabídce STATS $\longmapsto$ Estimation and Testing $\longmapsto$ One-Sample Analysis. EXCEL počítá realizaci intervalového odhadu střední hodnoty normálního rozdělení se známým rozptylem v nabídce průvodce funkcí $f_x \longmapsto$ funkce statistické $\longmapsto$ CONFIDENCE.

Intervalový odhad pro střední hodnotu normálního rozdělení lze použít i v případě, kdy náhodná veličina X nemá normální rozdělení a rozsah realizace je větší než 30 (viz poznámka P - 9.2.1 3)). Intervalový odhad pro rozptyl normálního rozdělení lze použít pouze v případě náhodného výběru z normálního rozdělení.

Příklad 4.3. Určete realizaci 99%-ního intervalového odhadu neznámých parametrů v jednotlivých situacích v příkladu 3.8 za předpokladu, že náhodná chyba měření má normální rozdělení.

Řešení. V příkladu 3.8 je X náhodná chyba měření, $E(X) = \mu, D(X) = \sigma^2$. Nyní navíc předpokládáme, že $X \sim N(\mu, \sigma^2)$. Pro intervalový odhad neznámých parametrů můžeme tedy využít výsledků této kapitoly.

1. V této situaci neznáme ani μ ani σ^2, proto využijeme větu 4.4. $100 \cdot (1-\alpha)\%$-ní intervalový odhad střední hodnoty μ je interval

$$\left\langle \overline{X} - t(n-1; 1-\alpha_1)\frac{S}{\sqrt{n}}, \overline{X} + t(n-1; 1-\alpha_2)\frac{S}{\sqrt{n}} \right\rangle.$$

Realizace statistik $\overline{X}$ a S (tj. realizace odhadů μ a σ^2 při neznámém μ a σ) máme spočítané v příkladu 3.8, tedy $\widehat{\mu} = \overline{x} = -0.50$ mm, $\widehat{\sigma} = s \doteq \sqrt{2.28} \doteq 1.51$ mm.

Hledáme realizaci 99%-ního (oboustranného) intervalového odhadu, tj. $\alpha = 0.01$. Rozdělením rizika α dostaneme

$$\alpha_1 = \alpha_2 = \frac{\alpha}{2} = 0.005.$$

Odtud (viz tabulky kvantilů t - rozdělení)

$$t(n-1; 1-\alpha_1) = t(9; 0.995) = 3.250,$$
$$t(n-1; 1-\alpha_1) = t(9; 0.995) = 3.250.$$

Realizace 99%-ního intervalového odhadu střední hodnoty náhodné chyby měření je tedy

$$\left\langle -0.50 - 3.250 \cdot \frac{1.51}{\sqrt{10}}, -0.50 + 3.250 \cdot \frac{1.51}{\sqrt{10}} \right\rangle \doteq \langle -2.05, 1.05 \rangle .$$

Interval $\langle -2.05 \text{ mm}, 1.05 \text{ mm} \rangle$ téměř jistě překrývá skutečnou střední hodnotu náhodné chyby měření.

Podobně $100 \cdot (1-\alpha)$%-ním intervalovým odhadem rozptylu σ^2 je interval

$$\left\langle \frac{(n-1)S^2}{\chi^2(n-1; 1-\alpha_1)}, \frac{(n-1)S^2}{\chi^2(n-1; \alpha_2)} \right\rangle.$$

Stejně jako výše rozdělíme riziko α a příslušné kvantily najdeme v tabulkách. Dostaneme

$$\alpha = 0.01 \Rightarrow \alpha_1 = \alpha_2 = \frac{\alpha}{2} = 0.005,$$
$$\chi^2(n-1; 1-\alpha_1) = \chi^2(9; 0.995) = 23.59,$$
$$\chi^2(n-1; \alpha_2) = \chi^2(9; 0.005) = 1.735,$$

Realizace 99% - ního intervalového odhadu rozptylu náhodné chyby měření je tedy

$$\left\langle \frac{9 \cdot 2.28}{23.59}, \frac{9 \cdot 2.28}{1.735} \right\rangle \doteq \langle 0.87, 11.68 \rangle .$$

Rozptyl náhodné chyby měření téměř jistě překrývá interval $\langle 0.87 \text{ mm}^2, 11.68 \text{ mm}^2 \rangle$.

2) V této situaci známe střední hodnotu chyby měření, víme, že $\mu = -1$ mm. Zbývá tedy najít 99%-ní intervalový odhad rozptylu σ^2. Podle věty 4.3 je $100 \cdot (1-\alpha)$%-ní intervalový odhad parametru σ^2 při známém μ interval

$$\left\langle \frac{nS_0^2}{\chi^2(n; 1-\alpha_1)}, \frac{nS_0^2}{\chi^2(n; \alpha_2)} \right\rangle.$$

Realizaci statistiky S_0^2 (tj. realizaci odhadu rozptylu σ^2 při známém μ) máme spočítanou v příkladu 3.8, tedy

$$\widehat{\sigma^2} = s_0^2 = 2.30 \text{ mm}^2$$

Analogicky jako výše dostaneme

$$\alpha = 0.01 \Rightarrow \alpha_1 = \alpha_2 = \frac{\alpha}{2} = 0.005,$$
$$\chi^2(n; 1-\alpha_1) = \chi^2(10; 0.995) = 25.19,$$
$$\chi^2(n; \alpha_2) = \chi^2(10; 0.005) = 2.156,$$

$$\left\langle \frac{10 \cdot 2.30}{25.19}, \frac{10 \cdot 2.30}{2.156} \right\rangle \doteq \langle 0.91, 10.67 \rangle .$$

Rozptyl náhodné chyby měření téměř jistě překrývá interval $\langle 0.91\ \text{mm}^2, 10.67\ \text{mm}^2 \rangle$.

3. V poslední situaci známe rozptyl náhodné chyby měření. Z věty 4.2 dostaneme, že realizace hledaného intervalového odhadu pro střední hodnotu náhodné chyby měření je interval $\langle -2.13\ \text{mm}, 1.13\ \text{mm} \rangle$.

♣

Příklad 4.4. Určete 95%-ní dolní odhad střední hodnoty a 95%-ní horní odhad směrodatné odchylky pevnosti betonu v příkladu 3.9 za předpokladu, že má pevnost betonu přibližně normální rozdělení a neznáme ani jeden z parametrů.

Řešení. Předpokládáme tedy, že pevnost betonu $X \sim N(\mu, \sigma^2)$, kde ani jeden z parametrů neznáme. K dispozici máme realizaci náhodného výběru z X o rozsahu $n = 40$. V příkladu 3.9 1) byly spočteny realizace odhadů neznámých parametrů

$$\widehat{\mu} = \overline{x} = 29.650\ \text{Mpa},$$
$$\widehat{\sigma} = s \doteq 2.517\ \text{MPa} \Rightarrow \widehat{\sigma^2} = s^2 \doteq 6.335\ \text{MPa}^2.$$

Při konstrukci hledaných intervalových odhadů využijeme větu 4.4. $100{\cdot}(1-\alpha)$%-ním intervalovým odhadem střední hodnoty μ je interval

$$\left\langle \overline{X} - t(n-1; 1-\alpha_1)\frac{S}{\sqrt{n}}, \overline{X} + t(n-1; 1-\alpha_2)\frac{S}{\sqrt{n}} \right\rangle.$$

Vzhledem k tomu, že hledáme dolní intervalový odhad střední hodnoty, dostaneme

$$\alpha = 0.05 \Rightarrow \alpha_1 = \alpha = 0.05,\ \alpha_2 = 0$$
$$t(n-1; 1-\alpha_1) = t(39; 0.95) = 1.960,$$
$$t(n-1; 1-\alpha_2) = t(39; 1) = \infty.$$

Realizace 95%-ního dolního intervalového odhadu střední hodnoty pevnosti betonu je tedy interval $\langle 28.870, \infty \rangle$, tj. střední hodnota pevnosti betonu je téměř jistě větší než 28.870 MPa.

Podobně $100{\cdot}(1-\alpha)$%-ním intervalovým odhadem rozptylu σ^2 je interval

$$\left\langle \frac{(n-1)S^2}{\chi^2(n-1; 1-\alpha_1)}, \frac{(n-1)S^2}{\chi^2(n-1; \alpha_2)} \right\rangle.$$

Analogicky jako výše dostaneme

$$\alpha = 0.05 \Rightarrow \alpha_1 = 0, \alpha_2 = \alpha = 0.05$$
$$\chi^2(n-1; 1-\alpha_1) = \chi^2(39; 1) = \infty,$$
$$\chi^2(n-1; \alpha_2) = \chi^2(39; 0.05) = 25.692.$$

Tedy interval

$$\left(0, \frac{39 \cdot 6.335}{25.692}\right\rangle \doteq (0, 9.616\rangle$$

je realizací 95%-ního horního intervalového odhadu σ^2, tj. interval $(0, \sqrt{9.616}\rangle$ je realizací 95%-ního horního intervalového odhadu σ, tj. směrodatná odchylka pevnosti betonu je téměř jistě menší než 3.101 MPa.

♣

Poznámka 4.6. V posledním příkladu jsme pro výpočet intervalového odhadu střední hodnoty pevnosti (na rozdíl od intervalového odhadu rozptylu) nepotřebovali informaci o rozdělení náhodné veličiny X, protože rozsah realizace výběru z X byl větší než 30 (viz poznámka 4.5).

Příklad 4.5. Během 21 dnů byly v určitém regionu registrovány počty nehod za den. Výsledky jsou v následujícící tabulce

Počet nehod	0	1	2	3	4	5
n_j	4	8	5	1	2	1

Určete

a) realizaci bodového odhadu střední hodnoty a směrodatné odchylky počtu nehod za den,

b) realizaci 99%-ního horního odhadu střední hodnoty počtu nehod za den, je-li známo, že počet nehod za den má přibližně Poissonovo rozdělení.

Řešení. Náhodnou veličinou X je zde počet nehod za den. Máme k dispozici realizaci náhodného výběru z X o rozsahu $n = 21$, která je roztříděna do šesti tříd.

a) Pro výpočet realizací odhadů $\widehat{E(X)}$ a $\sqrt{\widehat{D(X)}}$ střední hodnoty $E(X)$ a směrodatné odchylky $\sqrt{D(X)}$ nepotřebujeme informaci o rozdělení náhodné veličiny X. Dostaneme

$$\widehat{E(X)} = \overline{x} \doteq 1.619,$$
$$\sqrt{\widehat{D(X)}} = s = \sigma_{n-1} = 1.396.$$

b) Náhodná veličina X nemá normální rozdělení a rozsah náhodného výběru z X není větší než 30. Nemůžeme tedy pro výpočet realizace intervalového odhadu $E(X)$ použít intervalový odhad střední hodnoty normálního rozdělení. Tento intervalový odhad najdeme (bez odvozování) v literatuře. Víme, že $X \sim Po(\lambda)$ a tedy $E(X) = \lambda$. $100 \cdot (1-\alpha)$%-ní intervalový odhad parametru λ Poissonova rozdělení je (viz např. [14]) interval

$$\left\langle \frac{\chi^2(2n\overline{X};\alpha_1)}{2n}, \frac{\chi^2(2n\overline{X}+2;1-\alpha_2)}{2n} \right\rangle.$$

Při výpočtu realizace 99%-ního horního intervalového odhadu parametru λ již postupujeme analogicky jako v případě výpočtu realizace intervalového odhadu neznámých parametrů normálního rozdělení. Postupně dostaneme

$$\begin{aligned}
&\alpha = 0.01 \Rightarrow \alpha_1 = 0,\ \alpha_2 = \alpha = 0.01,\\
&\chi^2(2n\overline{x};\alpha_1) = \chi^2(68;0) = 0,\\
&\chi^2(2n\overline{x}+2;1-\alpha_2) = \chi^2(70;0.99) = 101.232.
\end{aligned}$$

Hledaná realizace 99%-ního horního intervalového odhadu parametru λ je tedy interval

$$\left\langle 0, \frac{10.232}{12} \right\rangle \doteq \langle 0, 2.410 \rangle.$$

Střední hodnota počtu nehod za den je tedy téměř jistě nejvýše 2.410 nehod.

♣

5. Testování statistických hypotéz

5.1. Základní pojmy

Další oblastí statistických metod je testování statistických hypotéz. Statistická hypotéza je tvrzení, které se týká rozdělení náhodné veličiny X. Statistické hypotézy se tedy mohou týkat buď typu rozdělení X nebo hodnot neznámého parametru rozdělení X (v případě, že jsme si typem rozdělení X jisti).

Pokud je hypotéza formulována tak, že jednoznačně určuje rozdělení X, nazýváme ji jednoduchou hypotézou. Hypotéza, která rozdělení X jednoznačně nespecifikuje, se nazývá složená hypotéza.

Příklady jednoduchých hypotéz jsou např. tvrzení:

a) X má rozdělení $N(3, 4)$

b) normální rozdělení X s rozptylem 9 má střední hodnotu $\mu = 5$.

Příklady složených hypotéz jsou např. tvrzení:

a) X má t-rozdělení s n stupni volnosti (n je neznámé)

b) střední hodnota μ normálního rozdělení X je větší než 7.

Hypotézy a) z výše uvedených příkladů se týkají tvaru rozdělení, hypotézy b) se týkají parametrů rozdělení.

Při testování hypotéz klademe proti sobě vždy dvě hypotézy. Jednu z nich nazýváme testovanou nebo nulovou hypotézou a značíme ji H_0; druhou, popírající platnost testované hypotézy, nazýváme alternativní hypotéza (k hypotéze H_0) a značíme ji H. Alternativní hypotéza nemusí být nutně opačná k hypotéze H_0. Víme-li např. jistě, že X je normální náhodná veličina a domníváme se, že $\mu = 0.3$, nemusí mít H nutně tvar $\mu \neq 0.3$. Může se stát, že předem bezpečně víme, že musí platit $\mu \geq 0.3$. Za této dodatečné informace má pak H tvar: $\mu > 0.3$.

Testem statistické hypotézy H_0 rozumíme postup, který na základě realizace náhodného výběru z rozdělení X vede k zamítnutí nebo nezamítnutí hypotézy H_0. Je-li k hypotéze H_0 stanovena alternativní hypotéza H, mluvíme o testu hypotézy H_0 proti (hypotéze) H. Značíme $H_0 \uparrow H$.

Na následujícím příkladu objasníme několik základních pojmů z oblasti testování statistických hypotéz.

Příklad 5.1.1. Mějme dvě „falešné" hrací kostky. Víme, že pravděpodobnost padnutí čísla šest na jedné z nich je 0.1, na druhé 0.5. Na základě pěti hodů jednou kostkou máme rozhodnout, zda jsme si vybrali tu, na které padá šestka častěji, tj. tu kostku, kterou budeme nazývat dále výhodnější.

Řešení. Označme p pravděpodobnost jevu, že při jednom hodu vybranou kostkou padne číslo šest. Za nulovou hypotézu H_0 zvolme hypotézu, že námi vybraná kostka je pro nás nevýhodná, za alternativní hypotézu H pak hypotézu, že tato kostka je výhodná. Tj. budeme testovat jednoduchou hypotézu $H_0 : p = 0.1$ proti jednoduché hypotéze $H : p = 0.5$. Naše rozhodnutí o zamítnutí, resp. nezamítnutí H_0 chceme udělat na základě pěti hodů vybranou kostkou.

Zřejmě $p \in \{0.1, 0.5\}$, protože nevíme, kterou kostku jsme si vybrali. Nazvěme hod vybranou kostkou úspěšný, jestliže padne číslo šest. Nechť X_i je počet úspěchů v i-tém hodu touto kostkou (i=1,2, ... ,5). Potom je zřejmě $(X_1, X_2, \dots, X_5)$ náhodný výběr z rozdělení $X \sim A(p)$, kde $p \in \{0.1, 0.5\}$. Statistika $R = X_1 + X_2 + \dots + X_5$ je pak počet úspěchů v 5-ti hodech vybranou kostkou. Rozhodnutí o zamítnutí, resp. nezamítnutí H_0 založíme tedy na realizaci r statistiky R. Obor hodnot statistiky R je zřejmě množina $\{0, 1, \dots, 5\}$. Je třeba připustit, že všechny tyto hodnoty jsou možné, jak při platnosti H_0, tak při platnosti H. Platí-li však H_0, pak zřejmě velké hodnoty statistiky R dostaneme výjimečně a naopak, platí-li H, dostaneme velké hodnoty statistiky R často. Řekněme tedy např., že H_0 zamítneme (tj. přijmeme H), je-li $r \in \{2, 3, 4, 5\}$ a H_0 nezamítneme, je-li $r \in \{0, 1\}$. Toto rozdělení oboru hodnot statistiky R se jeví rozumné. Můžeme jenom pochybovat, zda hranice mezi oběma částmi oboru hodnot statistiky R byla volena optimálně. Prozkoumejme za tím účelem pravděpodobnost toho, že se při rozhodování o platnosti H_0, resp. H dopustíme omylu.

Náhodná veličina R má zřejmě rozdělení $Bi(5, p)$ (viz poznámka P - 9.1.1 1)), kde $p \in \{0.1, 0.5\}$; tzn. pro pravděpodobnostní funkci q statistiky R platí

$$q \in \left\{ q(r;p);\ q(r;p) = \begin{cases} \binom{5}{r} p^r (1-p)^{5-r} & \text{pro } r = 0, 1, \dots, 5 \\ 0 & \text{jinak} \end{cases},\ p \in \{0.1, 0.5\} \right\}.$$

I když poměrně vzácně, může i při platnosti $H_0 : p = 0.1$ nastat případ, že šestka padne alespoň 2-krát. Protože v tomto případě H_0 zamítneme, bude pravděpodobnost, že $R \in \{2, 3, 4, 5\}$ při platnosti $H_0 : p = 0.1$, tj.

$$P(R \in \{2, 3, 4, 5\}/p = 0.1),$$

představovat riziko mylného zamítnutí H_0. Platí-li $H_0 : p = 0.1$ je $R \sim Bi(5, 0.1)$ a tedy

$$\begin{aligned} P(R \in \{2, 3, 4, 5\}/p = 0.1) &= \sum_{r=2}^{5} \binom{5}{r} 0.1^r 0.9^{5-r} \\ &\doteq 0,0729 + 0,0081 + 0,0005 + 0,0000 = 0,0815. \end{aligned}$$

Riziko mylného zamítnutí H_0 je tedy asi 8%-ní, tj. mýlíme se tímto způsobem v průměru asi v osmi případech ze sta.

Stejně tak může nastat případ, že při platnosti $H : p = 0.5$ dostaneme nízký výsledek, který nás vede k zamítnutí H a přijetí H_0. Riziko takového omylu určíme zcela obdobně:

$$\begin{aligned} P(R \in \{0, 1\}/p = 0.5) &= \sum_{r=0}^{1} \binom{5}{r} 0.5^r 0.5^{5-r} = \sum_{r=0}^{1} \binom{5}{r} 0.5^5 \\ &\doteq 0.0313 + 0.1563 \doteq 0.1876. \end{aligned}$$

Vidíme tedy, že riziko mylného nezamítnutí H_0 je téměř 19%-ní - tj. mýlíme se v průměru v asi 19-ti případech ze sta.

V právě uvedeném příkladu si zejména všimněme, že přicházely v úvahu dva druhy chyb.

♣

Vztah mezi výsledkem testu a realitou a z něj vyplývajících možností chyb můžeme schématicky znázornit tak, jako v tabulce 5.1.1.

Výsledek	skutečnost	
	jde o výhodnou kostku	jde o nevýhodnou kostku
H_0 nezamítáme	mylné nezamítnutí H_0	správné rozhodnutí
H_0 zamítáme	správné rozhodnutí	mylné zamítnutí H_0

Tabulka 5.1.1

Tematicky značně odlehlé situace mají z hlediska rozhodování některé základní rysy stejné. Snadno se sestaví obdobná schémata pro další situace, jako je přijetí dodávky, přechod na nový výrobní program a podobně.

Přejděme nyní k obecnějšímu výkladu principů testování statistických hypotéz. Jak již bylo uvedeno, stojí proti sobě dvě hypotézy týkající se náhodné veličiny X, nulová hypotéza H_0 a alternativní hypotéza H. Své rozhodnutí o zamítnutí H_0 zakládáme na realizaci náhodného výběru $(X_1, X_2, \ldots, X_n)$ z rozdělení X, přesněji řečeno na realizaci určité statistiky $R = r(X_1, X_2, \ldots, X_n)$. Statistiku R nazýváme testovací kritérium. Obor hodnot statistiky R rozdělíme na dvě disjunktní části - jednu z nich označíme W a budeme nazývat kritickým oborem (pro test hypotézy H_0), druhou označíme V a nazveme oborem nezamítnutí (pro test hypotézy H_0). Za W volíme ty hodnoty r testovacího kritéria R, které nesvědčí v prospěch hypotézy H_0 a svědčí v prospěch hypotézy H. Hypotézu H_0 pak zamítneme, jestliže pro realizaci r statistiky R platí $r \in W$; jestliže $r \in V$, tj. $r \notin W$, hypotézu H_0 nezamítneme.

Při zamítání, resp. nezamítání hypotézy H_0 se můžeme dopustit jedné ze dvou chyb. Jestliže zamítneme H_0, ačkoliv je správná, mluvíme o chybě 1. druhu. Stane-li se, že nezamítneme H_0, ačkoliv je nesprávná, mluvíme o chybě 2. druhu. Pravděpodobnost chyby 1. druhu je zřejmě

$$P(R \in W/H_0)$$

a pravděpodobnost chyby 2. druhu je

$$P(R \notin W/H).$$

Uvědomme si, že pro vyjádření pravděpodobnosti chyby 1. druhu musíme znát rozdělení statistiky R za podmínky, že platí hypotéza H_0. Je-li H_0 jednoduchá hypotéza, je rozdělení náhodné veličiny X za podmínky H_0 určeno jednoznačně, a tedy i rozdělení statistiky R za podmínky H_0 je určeno jednoznačně. Tedy pravděpodobnost chyby 1. druhu, tj. $P(R \in W/H_0)$, je určena jednoznačně. Je-li H_0

PDF-Xchange Click to buy NOW! www.docu-track.com

složená hypotéza, není rozdělení náhodné veličiny X za podmínky H_0 určeno jednoznačně (připouštíme alespoň dvě různé možnosti), a tedy i rozdělení statistiky R za podmínky H_0 nemusí být určeno jednoznačně. Tedy pravděpodobnost chyby 1. druhu, tj. $P(R \in W/H_0)$, nemusí být určena jednoznačně (může mít alespoň dvě různé hodnoty).

Číslo α, které udává supremum pravděpodobnosti chyby 1. druhu, nazýváme hladina významnosti testu (hypotézy H_0). Z výše uvedeného plyne, že

$$\alpha = P(R \in W/H_0)$$

v případě jednoduché hypotézy H_0. V případě složené hypotézy H_0 je

$$\alpha \geq P(R \in W/H_0).$$

Podobné úvahy platí pro vyjádření pravděpodobnosti chyby druhého druhu, tj. $P(R \notin W/H)$. Číslo, které udává supremum pravděpodobnosti chyby 2. druhu, značíme β. Číslo $1-\beta$ se nazývá síla testu. V případě složené hypotézy H je tedy

$$\beta \geq P(R \notin W/H).$$

V případě jednoduché alternativní hypotézy H platí

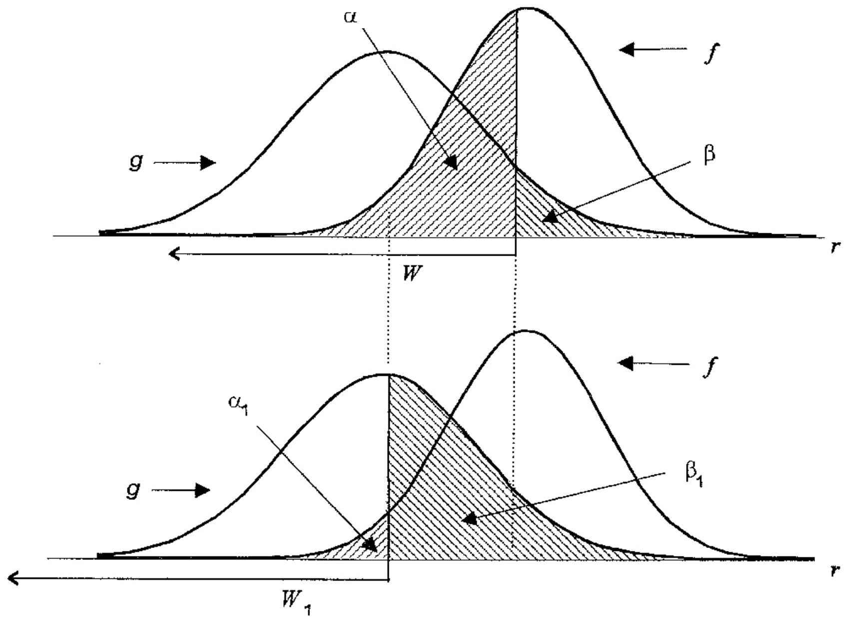
$$\beta = P(R \notin W/H).$$

Hladina významnosti testu α vlastně zaručuje, že riziko mylného zamítnutí H_0 je maximálně $100 \cdot \alpha$ procentní (mýlíme se tímto způsobem v 100α případech ze sta). Podobně definiční požadavek pravděpodobnosti chyby 2. druhu β zaručuje, že riziko mylného přijetí H_0 (v neprospěch H) je maximálně $100 \cdot \beta$ procentní. Je přirozené požadovat, aby čísla α i β byla co nejmenší.

Prozkoumejme nyní některé souvislosti mezi pravděpodobnostmi chyb obou druhů, zejména z hlediska, jak lze tyto pravděpodobnosti ovlivnit.

Pozorujme vztah mezi pravděpodobnostmi obou druhů chyb na obrázku 5.1.1. Z porovnání obou diagramů je ihned patrné, že za jinak stejných podmínek (nestane-li se nic, co by ovlivnilo rozdělení statistiky R při platnosti H_0, resp. H) má snížení pravděpodobnosti chyby 1. druhu za následek zvýšení pravděpodobnosti chyby 2. druhu a naopak.

V praktických úlohách se snažíme rizika chyb obou druhů rozumně vyvážit. Neznamená to ovšem, že chceme dosáhnout toho, aby byla stejná. Spíše se musíme snažit o minimalizaci následků obou druhů chyb.



Obrázek 5.1.1: Vyjádření vztahu mezi pravděpodobnostmi obou druhů chyb při testu jednoduché hypotézy H_0 proti jednoduché hypotéze H. Hustota testovacího kritéria R za podmínky $H_0[H]$ je označena $f[g]$.

Příklad 5.1.2. V příkladu 5.1.1 jsme testovali proti sobě dvě jednoduché hypotézy o parametru p alternativního rozložení X; totiž hypotézu $H_0 : p = 0.1$ proti hypotéze $H : p = 0.5$. K dispozici byl náhodný výběr z X o rozsahu $n = 5$. Za testovací kritérium jsme zvolili statistiku

$$R = \sum_{i=1}^{5} X_i$$

a za kritický obor množinu

$$W = \Big\{r;\ r \in \{2, 3, 4, 5\}\Big\}.$$

Za těchto podmínek jsme zjistili, že

$$\alpha = 0.0815,\ \beta = 0.1876.$$

Sledujme, jak se budou měnit pravděpodobnosti chyb při změněných podmínkách:

a) zmenšíme kritický obor na

$$W_1 = \Big\{r;\ r \in \{3, 4, 5\}\Big\}.$$

PDF-Xchange
Click to buy NOW!
www.docu-track.com

Vzhledem k tomu, že tato změna neovlivnila rozdělení statistiky R, můžeme očekávat zmenšení pravděpodobnosti chyby 1. druhu a zvětšení pravděpodobnosti chyby 2. druhu. Podobně jako v příkladu 5.1.1 dostáváme:

$$\begin{aligned} P\big(R \in W_1/H_0\big) &= P\big(R \in \{3,4,5\}/p = 0.1\big) \\ &= \sum_{i=3}^{5} \binom{5}{r} 0.1^r 0.9^{5-r} \doteq 0.0086, \end{aligned}$$

$$\begin{aligned} P\big(R \notin W_1/H\big) &= P\big(R \in \{0,1,2\}/p = 0.5\big) \\ &= \sum_{r=0}^{2} \binom{5}{r} 0.5^r 0.5^{5-r} \doteq 0.500. \end{aligned}$$

Snížení pravděpodobnosti chyby 1. druhu pod 0.01 má tedy za následek zvýšení pravděpodobnosti chyby 2. druhu na 0.5.

b) zvětšíme rozsah náhodného výběru dvojnásobně a kritický obor W_2 zvolíme proporcionální původnímu, tj.

$$W_2 = \Big\{r;\ r \in \{4,5,\dots,10\}\Big\}.$$

Zvětšení rozsahu náhodného výběru má za následek změnu rozdělení testovacího kritéria

$$R = \sum_{i=1}^{10} X_i$$

Pro R bude platit $R \sim Bi(10,p)$. Dostáváme tedy

$$\begin{aligned} P\big(R \in W_2/H_0\big) &= P\big(R \in \{4,5,\dots,10\}/p = 0.1\big) \\ &= \sum_{i=4}^{10} \binom{10}{r} 0.1^r 0.9^{10-r} \doteq 0.0128, \end{aligned}$$

$$\begin{aligned} P\big(R \notin W_2/H\big) &= P\big(R \in \{0,1,2,3\}/p = 0.5\big) \\ &= \sum_{r=0}^{3} \binom{10}{r} 0.5^r 0.5^{10-r} \doteq 0.1719. \end{aligned}$$

Pozorujeme tedy z hlediska velikosti chyb značné zlepšení testu.

♣

Všimněme si, že jsme v předchozím příkladu snížení pravděpodobnosti chyb obou druhů dosáhli tak, že jsme zvětšili rozsah náhodného výběru. (Tím se změnilo rozdělení testovacího kritéria.) Zvětšení rozsahu náhodného výběru však nemusí být vždycky možné (např. realizace náhodného výběru může být již provedena a není možné v ní pokračovat).

Ukázali jsme, že obecně nelze kritický obor W volit tak, aby byly současně obě chyby tak malé, jak bychom si přáli. Proto se obvykle trvá jenom na tom, aby hladina významnosti testu α bylo nějaké, předem dané (samozřejmě malé) číslo z intervalu (0,1). V technické praxi se nejčastěji volí $\alpha = 0.05$ nebo $\alpha = 0.01$. Z možných testů H_0 na dané hladině významnosti α (tj. z možných voleb testovacího kritéria R a kritického oboru W), pak volíme ten, při kterém je β nejmenší.

Z výše uvedeného vyplývá, že hlavním teoretickým problémem testování hypotéz je vhodná volba testovacího kritéria a následné určení jeho rozdělení. Volbou testovacího kritéria se zabývají rozsáhlejší učebnice matematické statistiky, zde budou pro nejběžnější situace testovací kritéria, zajišťující výhodné vlastnosti testu, uvedena (až na výjimky) bez odvození. Je-li v testu H_0 na dané hladině významnosti α zvoleno testovací kritérium R, je ještě zapotřebí rozhodnout o umístění kritického oboru W. Problém volby kritického oboru však většinou vyřešíme okamžitě, budeme-li mít na zřeteli tvar testovacího kritéria R a znění alternativní hypotézy H. Je totiž většinou snadné určit, které realizace statistiky R svědčí proti H_0 a v prospěch H. Z nich pak vytvoříme kritický obor W.

Rozhodneme-li se ponechat stranou všechny teoretické problémy testování hypotéz naznačené výše, můžeme řešení konkrétních úloh různého druhu shrnout do následujícího obecného postupu.

Na základě všeho, co je nám známo o sledované náhodné veličině X a s uvážením zamýšleného rozsahu realizace náhodného výběru z rozdělení X

1) formulujeme hypotézu H_0 a volíme alternativní hypotézu H k hypotéze H_0, volíme hladinu významnosti testu α
2) stanovíme (nejčastěji vyhledáme) testovací kritérium R pro test hypotézy H_0
3) stanovíme (vyhledáme) kritický obor W pro test hypotézy H_0 proti hypotéze H na hladině významnosti α
4) dosadíme realizaci $(x_1, x_2, \ldots, x_n)$ náhodného výběru do testovacího kritéria R, výsledek označíme r
5) rozhodneme
   - jestliže $r \in W$, potom H_0 zamítneme a přijmeme H (s maximálně $100{\cdot}\alpha\%$ - ním rizikem mylného přijetí H v neprospěch H_0)
   - jestliže $r \notin W$, potom H_0 provedeným testem nezamítáme.

Pokud nemůžeme specifikovat horní hranici pravděpodobnosti chyby 2.druhu, zdržíme se úsudku o platnosti H_0 a test uzavřeme s tím, že se H_0 nepodařilo zamítnout. Pokud je horní hranice pravděpodobnosti chyby 2. druhu β známá a předepsaně malá (např. 0.05), pak H_0 nezamítneme a můžeme ji přijmout (s maximálně $100{\cdot}\beta\%$ - ním rizikem mylného přijetí H_0 v neprospěch H).

Poněvadž se pravděpodobnost chyby 2. druhu někdy určuje dost obtížně (především v případě složené alternativní hypotézy), postupujeme bez její znalosti často tak, že za alternativní hypotézu H volíme ten výrok o rozdělení X, který chceme prokázat, nebo spíše, jehož potvrzení vede k nějakým závažným opatřením. Podle výše uvedeného postupu pak při výsledku testu v kritickém oboru prohlásíme tento výrok za platný s maximálně $100{\cdot}\alpha\%$ - ním rizikem omylu. Jestliže výsledek neleží v kritickém oboru, zdržíme se úsudku. Při výsledku testu $r \notin W$ může někdy pomoci „přehodit“ hypotézy a testovat H proti H_0.

5.2. Testy hypotéz o parametrech rozdělení

Předpokládejme, že $X \sim N(\mu, \sigma^2)$, kde oba nebo některý z parametrů μ, σ^2 neznáme. Zabývejme se testováním hypotéz o skutečných hodnotách těchto parametrů.

Příklad 5.2.1. Nechť X je normální náhodná veličina s neznámou střední hodnotou μ a známým rozptylem σ^2. Buďte α, μ_0 předem daná čísla, $\alpha \in (0,1)$. Navrhněte test statistické hypotézy

1. $H_0 : \mu = \mu_0$ proti hypotéze $H : \mu > \mu_0$
2. $H_0 : \mu = \mu_0$ proti hypotéze $H : \mu < \mu_0$
3. $H_0 : \mu = \mu_0$ proti hypotéze $H : \mu \neq \mu_0$

na hladině významnosti α.

Řešení. Buď $(X_1, X_2, \ldots, X_n)$ náhodný výběr z X. Protože nás zajímají testy o střední hodnotě μ náhodné veličiny X, vyjdeme z nejlepšího nestranného odhadu střední hodnoty, tj. statistiky

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

Jestliže platí hypotéza H_0 (tj. je-li skutečná střední hodnota náhodné veličiny X rovna číslu μ_0) má náhodná veličina X rozdělení $N(\mu_0, \sigma^2)$. Tedy statistika

$$R = \frac{\overline{X} - \mu_0}{\sigma} \sqrt{n}$$

má za podmínky $\mu = \mu_0$ rozdělení $N(0,1)$ (viz věta 4.1 vztah 1).

V prospěch hypotézy H_0 budou svědčit ty realizace $(x_1, x_2, \ldots, x_n)$ náhodného výběru $(X_1, X_2, \ldots, X_n)$ z X, pro které bude realizace $\overline{x}$ statistiky $\overline{X}$ „blízká“ číslu μ_0, tj. pro které bude realizace r statistiky R „blízká“ nule.

Zvolme tedy statistiku R za testovací kritérium pro test hypotézy $H_0 : \mu = \mu_0$. Nyní zbývá určit kritické obory pro jednotlivé možné alternativy.

1. V případě testu 1 budou zřejmě v prospěch hypotézy H svědčit ty realizace $(x_1, x_2, \ldots x_n)$ náhodného výběru $(X_1, X_2, \ldots, X_n)$ z X, pro které bude $\overline{x}$ „podstatně“ větší než číslo μ_0, tj. pro něž $\overline{x} - \mu_0$ bude větší než nějaká kladná konstanta c, tj. pro které bude $r = \frac{\overline{x} - \mu_0}{\sigma} \sqrt{n} > \frac{c\sqrt{n}}{\sigma}$. Označíme-li tedy $\frac{c\sqrt{n}}{\sigma} = k$, pak lze za kritický obor W pro test 1 hypotézy H_0 proti hypotéze H zvolit množinu

$$W = \Big\{ r;\ r > k \Big\},$$

kde $k > 0$. Číslo k vyberme tak, aby hladina významnosti testu byla α, tj. požadujeme, aby

$$\alpha = P\big(R \in W / H_0\big) = P\big(R > k\ / \mu = \mu_0\big).$$

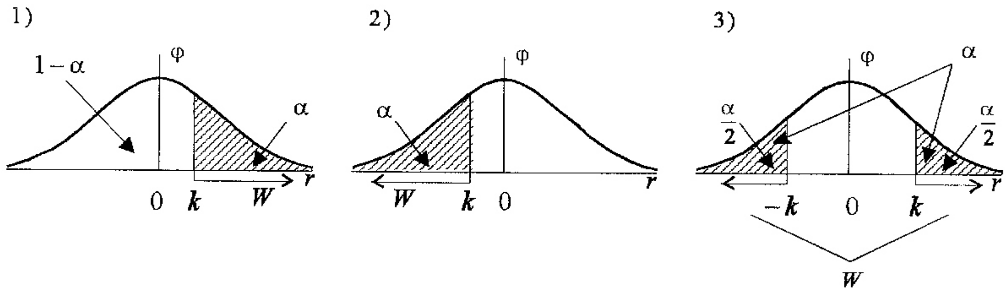
PDF-Xchange Click to buy NOW! www.docu-track.com

Odtud (viz obrázek 5.2.1 1))

$$P(R \leq k\ /\mu = \mu_0) = 1 - \alpha \Rightarrow k\ = u(1-\alpha).$$

Kritickým oborem W pro test hypotézy H_0 proti hypotéze H na hladině významnosti α je tedy množina

$$W = \Big\{r;\ r > u(1-\alpha)\Big\}.$$



Obrázek 5.2.1: Volba kritického oboru W pro test hypotézy $H_0 : \mu = \mu_0$ rozdělení $N(\mu, \sigma^2)$ se známým rozptylem σ^2 proti hypotéze 1) $H : \mu > \mu_0$, 2) $H : \mu < \mu_0$, 3) $H : \mu \neq \mu_0$ na hladině významnosti α.

2. V prospěch hypotézy H v případě testu 2 svědčí ty realizace $\overline{x}$ statistiky $\overline{X}$, pro které je $\overline{x}$ „podstatně“ menší než číslo μ_0, tj. ty realizace r testovacího kritéria R, které jsou menší než nějaká záporná konstanta k. Kritický obor W bude mít tedy tvar

$$W = \Big\{r;\ r < k\Big\}.$$

Číslo k opět vybereme na základě požadavku kladeného na hladinu významnosti testu α. Požadujeme, aby

$$\alpha = P(R \in W/H_0) = P(R < k\ /\mu = \mu_0).$$

Odtud (viz obrázek 5.2.1 2)) $k = u(\alpha)$. Vzhledem k tomu, že prakticky volíme α malé číslo (0.01 nebo 0.05) je $u(\alpha) < 0$ a pro práci s tabulkami potřebujeme „převodní“ vztah $u(\alpha) = -u(1-\alpha)$. Tedy kritický obor W pro test 2 na hladině významnosti α je

$$W = \Big\{r;\ r < -u(1-\alpha)\Big\}.$$

3. V testu 3 lze zřejmě volit za kritický obor W pro test hypotézy H_0 proti hypotéze H množinu

$$W = \Big\{r;\ |r| > k\Big\},$$

kde k je kladná konstanta. Konstantu k opět určíme na základě požadavku kladeného na hladinu významnosti α. Dostaneme (viz obrázek 5.2.1 3))

$$W = \Big\{r;\ |r| > u(1 - \frac{\alpha}{2}\Big\}.$$

♣

Poznámka 5.2.1. V testu 1, resp. 2 z příkladu 5.2.1 bychom obecně mohli uvažovat nulové hypotézy H_0 opačné k alternativním hypotézám H. Tj. uvažovat testy

$$T_1: \quad \text{hypotézy } H_0: \ \mu \leq \mu_0 \ \text{ proti hypotéze } H: \ \mu > \mu_0$$
$$T_2: \quad \text{hypotézy } H_0: \ \mu \geq \mu_0 \ \text{ proti hypotéze } H: \ \mu < \mu_0$$

na hladině významnosti α.

V těchto případech jsou nulové hypotézy složené. Lze ukázat, že za testovací kritérium R lze volit stejnou statistiku jako v příkladu 5.2.1 a kritické obory jsou taktéž stejné, tj. kritický obor pro test $T_1[T_2]$ je stejný jako pro test 1[2]. Pouze riziko mylného přijetí hypotézy H není 100α procentní, ale je maximálně $100\alpha\%$. Platí totiž např., že

$$P(R > k/\mu \leq \mu_0) \leq P(R > k/\mu = \mu_0).$$

Dostáváme tedy následující větu:

Věta 5.2.1. Buď $(X_1, X_2, \ldots, X_n)$ náhodný výběr z rozdělení $X \sim N(\mu, \sigma^2)$ s neznámou střední hodnotou μ a známým rozptylem σ^2. Buďte μ_0, α předem daná čísla, $\alpha \in (0,1)$. Potom pro testy

$$T_1: \quad \text{hypotézy } H_0: \ \mu \leq \mu_0 \ \text{ proti hypotéze } H: \ \mu > \mu_0,$$
$$T_2: \quad \text{hypotézy } H_0: \ \mu \geq \mu_0 \ \text{ proti hypotéze } H: \ \mu < \mu_0,$$
$$T_3: \quad \text{hypotézy } H_0: \ \mu = \mu_0 \ \text{ proti hypotéze } H: \ \mu \neq \mu_0$$

lze za testovací kritérium volit stejnou statistiku

$$R = \frac{\overline{X} - \mu_0}{\sigma}\sqrt{n},$$

která má za podmínky $\mu = \mu_0$ rozdělení $N(0,1)$.

Za kritické obory W_i $(i = 1,2,3)$ pro testy T_i $(i = 1,2,3)$ na hladině významnosti α lze postupně volit množiny

$$W_1 = \Big\{r;\ r > u(1-\alpha)\Big\},$$
$$W_2 = \Big\{r;\ r < -u(1-\alpha)\Big\},$$
$$W_3 = \Big\{r;\ r < -u(1-\frac{\alpha}{2}) \vee r > u(1-\frac{\alpha}{2})\Big\}.$$

Jsou-li nulové hypotézy v testech T_1 a T_2 tvaru $\mu = \mu_0$ zůstává tvrzení věty v platnosti.

Příklad 5.2.2. Nechť X je normální náhodná veličina se známou střední hodnotou μ a neznámým rozptylem σ^2. Buďte α, σ_0^2 předem daná čísla, $\alpha \in (0,1)$. Navrhněte test statistické hypotézy

a) $H_0: \ \sigma^2 = \sigma_0^2$ proti hypotéze $H: \ \sigma^2 < \sigma_0^2$

b) $H_0: \ \sigma^2 \geq \sigma_0^2$ proti hypotéze $H: \ \sigma^2 < \sigma_0^2$

na hladině významnosti α.

Řešení.

a) Buď $(X_1, X_2, \ldots, X_n)$ náhodný výběr z X. Protože nás zajímají testy o rozptylu σ^2 náhodné veličiny X, vyjdeme z nestranného odhadu roztylu, když známe střední hodnotu μ, tj. statistiky

$$S_0^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2.$$

Jestliže je skutečný rozptyl náhodné veličiny X roven číslu σ_0^2, má náhodná veličina X rozdělení $N(\mu, \sigma_0^2)$. Tedy statistika

$$R = \frac{nS_0^2}{\sigma_0^2}$$

má za podmínky $\sigma^2 = \sigma_0^2$ rozdělení $\chi^2(n)$ (viz věta 4.1).

V prospěch hypotézy H_0 budou zřejmě svědčit ty realizace $(x_1, \ldots, x_n)$ náhodného výběru $(X_1, \ldots, X_n)$ z rozdělení X, pro které bude realizace s_0^2 statistiky S_0^2 „blízká" číslu σ_0^2, tj. pro které bude $\frac{s_0^2}{\sigma_0^2}$ blízké číslu jedna, tj. pro které bude realizace r statistiky R blízká číslu n (rozsahu výběru). Zvolme tedy R za testovací kritérium pro test hypotézy H_0.

Nyní zbývá určit kritický obor pro test hypotézy H_0 proti hypotéze H na hladině významnosti α. V prospěch hypotézy H budou zřejmě svědčit ty realizace s_0^2, které budou „dostatečně" menší než číslo σ_0^2, tj. pro které bude $\frac{s_0^2}{\sigma_0^2}$ „dostatečně" menší než číslo jedna, tj. ty realizace r testovacího kritéria R, které budou „dostatečně" menší než číslo n. Zvolme tedy za kritický obor W pro test hypotézy H_0 proti hypotéze H množinu

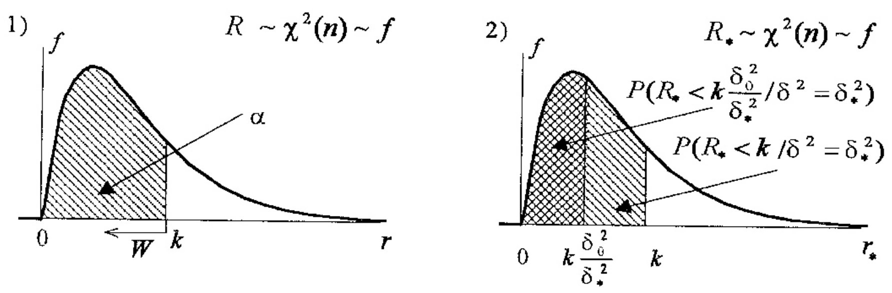
$$W = \left\{r;\ r < k\right\},$$

kde k je nějaká kladná konstanta menší než n. Číslo k vybereme tak, aby hladina významnosti testu byla α, tj. aby platilo

$$\alpha = P\left(R \in W/H_0\right) = P\left(R < k/\sigma^2 = \sigma_0^2\right).$$

Odtud (viz obrázek 5.2.2 1))

$$k = \chi^2(n;, \alpha).$$



Obrázek 5.2.2

PDF-XChange Click to buy NOW! www.docu-track.com

Kritickým oborem W pro test hypotézy H_0 proti hypotéze H na hladině významnosti α je tedy množina

$$W = \Big\{r;\ r < \chi^2(n;\alpha)\Big\}.$$

b) Test b) se liší od testu a) pouze tím, že nulová hypotéza H_0 je opačná k hypotéze H a jedná se o složenou hypotézu. Zvolme tedy opět statistiku

$$R = \frac{nS_0^2}{\sigma_0^2}$$

za testovací kritérium.

Protože hypotéza H je stejná jako v testu a), lze za bude kritický obor W pro test b) zvolit opět množinu

$$W = \Big\{r;\ r < k\Big\},$$

kde k je nějaká kladná konstanta menší než číslo n. Číslo k vybereme tak, aby hladina významnosti testu byla α. Vzhledem k tomu, že H_0 je složená hypotéza, požadujeme, aby nejmenší horní hranice pravděpodobnosti chyby 1. druhu byla α. Ukážeme, že pro pravděpodobnost chyby 1. druhu platí

$$P\big(R \in W/H_0\big) = P\big(R < k/\sigma^2 \geq \sigma_0^2\big) \leq P\big(R < k/\sigma^2 = \sigma_0^2\big), \tag{5.2.1}$$

tj. že nejmenší horní hranice pravděpodobnosti chyby 1. druhu je

$$P\big(R < k/\sigma^2 = \sigma_0^2\big).$$

Je-li skutečný rozptyl σ_*^2, nemá rozdělení $\chi^2(n)$ statistika R, ale statistika

$$R_* = \frac{nS_0^2}{\sigma_*^2}.$$

Zřejmě

$$R = R_* \cdot \frac{\sigma_*^2}{\sigma_0^2}.$$

Vyjádřeme nyní pravděpodobnost chyby 1. druhu ve vztahu (5.2.1) pomocí náhodné veličiny R_*, dostaneme

$$\begin{aligned} P\big(R < k/\sigma^2 = \sigma_*^2\big) &= P\Big(R_* \cdot \frac{\sigma_*^2}{\sigma_0^2} < k/\sigma^2 = \sigma_*^2\Big) \\ &= P\Big(R_* < k \cdot \frac{\sigma_0^2}{\sigma_*^2}/\sigma^2 = \sigma_*^2\Big). \end{aligned} \tag{5.2.2}$$

Pro každé $\sigma_*^2 \geq \sigma_0^2$ je $\dfrac{\sigma_0^2}{\sigma_*^2} \leq 1$ a tedy $k \cdot \dfrac{\sigma_0^2}{\sigma_*^2} \leq k$. Tedy pro každé $\sigma_*^2 \geq \sigma_0^2$ je (viz obrázek 5.2.2 2))

$$P\Big(R_* < k \cdot \frac{\sigma_0^2}{\sigma_*^2}/\sigma^2 = \sigma_*^2\Big) \leq P\big(R_* < k/\sigma^2 = \sigma_*^2\big).$$

Odtud a ze vztahu (5.2.2) dostáváme, že

$$P(R < k / \sigma^2 = \sigma_*^2) \leq P(R_* < k / \sigma^2 = \sigma_*^2) \quad \text{pro každé} \quad \sigma_*^2 \geq \sigma_0^2. \tag{5.2.3}$$

Uvědomme si nyní, že veličina R_* má za podmínky $\sigma^2 = \sigma_*^2$ rozdělení $\chi^2(n)$ a veličina R má za podmínky $\sigma^2 = \sigma_0^2$ také rozdělení $\chi^2(n)$. Platí tedy

$$P(R_* < k / \sigma^2 = \sigma_*^2) = P(R < k / \sigma^2 = \sigma_0^2) \quad \text{pro každé} \quad \sigma_*^2 \geq \sigma_0^2.$$

Ze vztahu (5.2.3) pak dostaneme

$$P(R < k / \sigma^2 = \sigma_*^2) \leq P(R < k / \sigma^2 = \sigma_0^2) \quad \text{pro každé} \quad \sigma_*^2 \geq \sigma_0^2,$$

čímž je vztah (5.2.1) dokázán. Nejmenší horní hranice pravděpodobnosti chyby 1. druhu je tedy

$$P(R < k / \sigma^2 = \sigma_0^2).$$

Protože požadujeme, aby hladina významnosti testu byla α, musí platit

$$P(R < k / \sigma^2 = \sigma_0^2) = \alpha.$$

Odtud plyne

$$k = \chi^2(n; \alpha).$$

Kritickým oborem W pro test b) je tedy stejně jako v testu a) množina

$$W = \left\{ r;\ r < \chi^2(n; \alpha) \right\}.$$

♣

Příklad 5.2.3. Vraťte se k příkladu 5.2.1 a poznámce 5.2.1 a dokažte platnost vztahu

$$P(R > k / \mu \leq \mu_0) \leq P(R > k / \mu = \mu_0).$$

♣

Analogicky jako v příkladu 5.2.2 lze odvodit testy dalších hypotéz o rozptylu normální náhodné veličiny, jejíž střední hodnota je známá. Výsledky shrnuje následující věta.

Věta 5.2.2. Buď $(X_1, X_2, \ldots, X_n)$ náhodný výběr z rozdělení $X \sim N(\mu, \sigma^2)$ s neznámým rozptylem σ^2 a známou střední hodnotou μ. Buďte σ_0^2, α předem daná čísla, $\alpha \in (0, 1)$. Potom pro testy

T_4 : hypotézy $H_0 : \sigma^2 \leq \sigma_0^2$ proti hypotéze $H : \sigma^2 > \sigma_0^2$,
T_5 : hypotézy $H_0 : \sigma^2 \geq \sigma_0^2$ proti hypotéze $H : \sigma^2 < \sigma_0^2$,
T_6 : hypotézy $H_0 : \sigma^2 = \sigma_0^2$ proti hypotéze $H : \sigma^2 \neq \sigma_0^2$

lze za testovací kritérium volit stejnou statistiku

$$R = \frac{n S_0^2}{\sigma_0^2},$$

která má za podmínky $\sigma^2 = \sigma_0^2$ rozdělení $\chi^2(n)$.

Za kritické obory W_i $(i = 4, 5, 6)$ pro testy T_i $(i = 4, 5, 6)$ na hladině významnosti α lze postupně volit množiny

$$W_4 = \Big\{r;\ r > \chi^2(n; 1-\alpha)\Big\},$$
$$W_5 = \Big\{r;\ r < \chi^2(n; \alpha)\Big\},$$
$$W_6 = \Big\{r;\ r < \chi^2(n; \frac{\alpha}{2}) \vee r > \chi^2(n; 1-\frac{\alpha}{2})\Big\}.$$

Jsou-li nulové hypotézy v testech T_4 a T_5 tvaru $\sigma^2 = \sigma_0^2$, zůstává tvrzení věty v platnosti.

Poznámka 5.2.2. Obecně se může stát, že neznáme ani jeden z parametrů μ, σ^2 normálního rozdělení.

Pro testy o střední hodnotě μ normálního rozdělení nelze v tomto případě samozřejmě použít větu 5.2.1. Uvědomme si ale, že nestranným odhadem skutečné střední hodnoty μ je opět statistika $\overline{X}$. Odhadem rozptylu σ^2 je statistika S^2. Statistika $R = \frac{\overline{X}-\mu_0}{S}\sqrt{n}$ má za podmínky $\mu = \mu_0$ rozdělení $t(n-1)$ (viz věta 4.1). Zvolíme-li tedy R za testovací kritérium, kritické obory pro jednotlivé testy budou analogické kritickým oborům ve větě 5.2.1. Budou se lišit pouze tím, že se v nich budou vyskytovat kvantily rozdělení $t(n)$ místo kvantilů rozdělení N(0,1).

Pro testy o hodnotě rozptylu σ^2 normálního rozdělení nelze opět obecně použít větu 5.2.2. Odhadem σ^2 je v tomto případě statistika S^2. Statistika $R = \frac{(n-1)S^2}{\sigma_0^2}$ má za podmínky $\sigma^2 = \sigma_0^2$ rozdělení $\chi^2(n-1)$ (viz věta 4.1). Když zvolíme R za testovací kritérium, pak kritické obory pro jednotlivé testy dostaneme zcela analogicky jako v příkladu 5.2.2. Budeme pracovat s rozdělením $\chi^2(n-1)$ místo s rozdělením $\chi^2(n)$. O výsledcích hovoří následující věta.

Věta 5.2.3. Buď $(X_1, X_2, \ldots, X_n)$ $(n \geq 2)$ náhodný výběr z rozdělení $X \sim N(\mu, \sigma^2)$ s neznámou střední hodnotou μ a neznámým rozptylem σ^2. Buďte μ_0, σ_0^2, α předem daná čísla, $\alpha \in (0, 1)$. Potom

a) pro testy

T_7: hypotézy $H_0 : \mu \leq \mu_0$ proti hypotéze $H : \mu > \mu_0$,
T_8: hypotézy $H_0 : \mu \geq \mu_0$ proti hypotéze $H : \mu < \mu_0$,
T_9: hypotézy $H_0 : \mu = \mu_0$ proti hypotéze $H : \mu \neq \mu_0$

lze za testovací kritérium volit stejnou statistiku

$$R = \frac{\overline{X}-\mu_0}{S}\sqrt{n},$$

která má za podmínky $\mu = \mu_0$ rozdělení $t(n-1)$.

Za kritické obory W_i $(i = 7, 8, 9)$ pro testy T_i $(i = 7, 8, 9)$ na hladině významnosti α lze postupně volit množiny

$$W_7 = \Big\{r;\ r > t(n-1; 1-\alpha)\Big\},$$
$$W_8 = \Big\{r;\ r < -t(n-1; 1-\alpha)\Big\},$$
$$W_9 = \Big\{r;\ r < -t(n-1; 1-\frac{\alpha}{2}) \vee r > t(n-1; 1-\frac{\alpha}{2})\Big\}.$$

b) pro testy

T_{10} : hypotézy $H_0 : \sigma^2 \leq \sigma_0^2$ proti hypotéze $H : \sigma^2 > \sigma_0^2$,
T_{11} : hypotézy $H_0 : \sigma^2 \geq \sigma_0^2$ proti hypotéze $H : \sigma^2 < \sigma_0^2$,
T_{12} : hypotézy $H_0 : \sigma^2 = \sigma_0^2$ proti hypotéze $H : \sigma^2 \neq \sigma_0^2$.

lze za testovací kritérium volit stejnou statistiku

$$R = \frac{(n-1)S^2}{\sigma_0^2},$$

která má za podmínky $\sigma^2 = \sigma_0^2$ rozdělení $\chi^2(n-1)$.

Za kritické obory W_i $(i = 10, 11, 12)$ pro testy T_i $(i = 10, 11, 12)$ na hladině významnosti α lze postupně volit množiny

$$W_{10} = \Big\{r;\ r > \chi^2(n-1; 1-\alpha)\Big\},$$
$$W_{11} = \Big\{r;\ r < \chi^2(n-1; \alpha)\Big\},$$
$$W_{12} = \Big\{r;\ r < \chi^2(n-1; \frac{\alpha}{2}) \vee r > \chi^2(n-1; 1-\frac{\alpha}{2})\Big\}.$$

Jsou-li všechny nulové hypotézy H_0 jednoduché, zůstává tvrzení věty v platnosti.

Poznámka 5.2.3.

1) Testy hypotéz o střední hodnotě normálního rozdělení lze použít i v případě, že se nejedná o normální rozdělení (viz poznámka P - 9.2.1 3)), je-li rozsah výběru větší než 30. Testy hypotéz o rozptylu normálního rozdělení nelze použít v případě, kdy se nejedná o normální rozdělení.

2) Testy o střední hodnotě normálního rozdělení při neznámém rozptylu se nazývají jednovýběrové t-testy podle rozdělení testovacího kritéria.

3) S rozvojem výpočetní techniky se stále častěji při testování hypotéz používá postup, při kterém se určí nejmenší hladina, při které bychom ještě hypotézu zamítli. Tato dosažená hladina testu se v anglicky psaných výstupech označuje jako P-value, Sig. level, P a podobně. Vyjadřuje nejmenší horní hranici pravděpodobnosti počítané za platnosti nulové hypotézy, že dostaneme právě naši realizaci r testovacího kritéia R nebo realizaci ještě více odporující nulové hypotéze. Tak se např. při provedení testu T_7 z věty 5.2.3 vypočítá realizace r testovacího kritéria

$$R = \frac{\overline{X} - \mu_0}{S}\sqrt{n}$$

a následně P-value, tj. nejmenší horní hranice následující pravděpodobnosti

$$P(R > r / \mu \leq \mu_0),$$

což je

$$P(R > r / \mu = \mu_0).$$

Když zamítneme hypotézu H_0 a přijmeme hypotézu H, dopustíme se maximálně $100P$ procentního omylu. Je-li tedy např. P=0.04, zamítáme H_0 s rizikem omylu maximálně 4%. Tedy hypotézu H_0 na hladině významnosti 0.05 zamítáme a na hladině významnosti 0.01 nezamítáme.

4) Pro provedení testů o střední hodnotě a rozptylu normální náhodné veličiny s oběma neznámými parametry můžeme využít např. STATGRAPHICS a to nabídku STATS $\longmapsto$ Estimation and Testing $\longmapsto$ One-Sample Analysis.

Příklad 5.2.4. Při zkoušce šestnácti náhodně vybraných ocelových tyčí byly zjištěny následující meze kluzu v MPa:

249, 250, 248, 255, 246, 250, 246, 245,
243, 245, 251, 248, 246, 245, 250, 249.

Požaduje se, aby střední hodnota meze kluzu byla 250 MPa a variabilita meze kluzu vyjádřená směrodatnou odchylkou byla maximálně 5 MPa. Ověřte, zda jsou oba předpoklady splněny - v prvním případě se připouští riziko omylu maximálně 5%, ve druhém maximálně 1%. Předpokládejte, že mez kluzu má přibližně normální rozdělení.

Řešení. Náhodnou veličinou X je zde mez kluzu oceli. Předpokládáme, že $X \sim N(\mu, \sigma^2)$ a ani jeden z paramterů μ a σ^2 není znám. V prvním případě budeme testovat hypotézu

$$H_0 : \mu = 250 \text{ MPa} \text{ proti hypotéze } H : \mu \neq 250 \text{ MPa}$$

na hladině významnosti 0.05. Jedná se o test T_9 z věty 5.2.3, kde $\mu_0 = 250$ MPa. Za testovací kritérium lze tedy zvolit statistiku

$$R = \frac{\overline{X} - \mu_0}{S} \sqrt{n}.$$

Kritický obor W pro test H_0 proti H na hladině významnosti α je pak množina

$$W = W_9 = \Big\{ r;\ r < -t(n-1; 1-\frac{\alpha}{2}) \vee r > t(n-1; 1-\frac{\alpha}{2}) \Big\}.$$

Výpočtem dostaneme

$$\widehat{\mu} = \overline{x} = 247.875 \text{ MPa},$$
$$\widehat{\sigma} = s \doteq 3.008 \text{ MPa} \Rightarrow \widehat{\sigma^2} \doteq 9.050 \text{ MPa}^2.$$

Odtud realizace r testovacího kritéria R je

$$r \doteq \frac{247.875 - 250}{3.008} \sqrt{16} \doteq -2.826.$$

Dále

$$t(n-1; 1-\frac{\alpha}{2}) = t(15, 0.975) = 2.131 \Rightarrow$$
$$W = \Big\{ r;\ r < -2.131 \vee r > 2.131 \Big\}$$

Protože $r \in W$, zamítáme hypotézu H_0 a přijímáme hypotézu H. Tedy střední hodnota meze kluzu oceli není rovna 250 MPa, riziko omylu je 5%.

Dále máme ověřit, zda $\sigma < 5$ MPa, tj. budeme testovat hypotézu

$$H_0 : \sigma^2 \geq 25 \text{ MPa}^2 \text{ proti hypotéze } H : \sigma^2 < 25 \text{ MPa}^2$$

na hladině významnosti 0.01. Jedná se o test T_{11} (viz věta 5.2.3), kde $\sigma_0^2 = 25 \text{ MPa}^2$. Testovací kritérium je statistika

$$R = \frac{(n-1)S^2}{\sigma_0^2}$$

a kritický obor W je množina

$$W = W_{11} = \Big\{r;\ r < \chi^2(n-1;\alpha)\Big\}$$

Výpočtem a pomocí tabulek dostaneme

$$r \doteq \frac{15 \cdot 9.050}{25} \doteq 5.430$$

$$W = \Big\{r;\ r < \chi^2(15;0.01)\Big\} = \Big\{r;\ r < 5.229\Big\}.$$

Protože $r \notin W$, nezamítáme hypotézu H_0, ale nemůžeme ji ani přijmout, protože neznáme riziko mylného přijetí.
Budeme ještě testovat hypotézu

$$H_0 : \sigma^2 \leq 25 \text{ MPa}^2 \text{ proti hypotéze } H : \sigma^2 > 25 \text{ MPa}^2$$

na stejné hladině významnosti. Testovací kritérium a tedy i jeho realizace zůstávají stejné, mění se pouze kritický obor, který je v tomto případě množina

$$W = W_{10} = \Big\{r;\ r > \chi^2(n-1, 1-\alpha)\Big\}$$

Pomocí tabulek dostaneme

$$W = \Big\{r;\ r > \chi^2(15;0.99)\Big\} = \Big\{r;\ r > 30.58\Big\}.$$

I v tomto případě $r \notin W$, nezamítáme tedy hypotézu $H_0 : \sigma^2 \leq 25 \text{ MPa}^2$.

Naměřená data a použité testy neumožňují rozhodnout, zda je směrodatná odchylka meze průtažnosti oceli menší než 5 MPa.

♣

Příklad 5.2.5. Byla provedena realizace náhodného výběru z $X \sim N(\mu, 4)$ o rozsahu $n = 9$ a vypočten $\overline{x} = 5.5$.

a) Ověřte, že na základě naměřených dat nelze na hladině významnosti 0.05 rozhodnout o zamítnutí hypotézy $H_0 : \mu \leq 6$.

b) Určete riziko mylného přijetí hypotézy H_0 v neprospěch hypotézy $\mu = 7$.

c) Určete rozsah n náhodného výběru z X tak, aby riziko mylného přijetí hypotézy H_0 v neprospěch hypotézy $\mu = 7$ bylo maximálně 5%.

Řešení.

a) Budeme testovat hypotézu

$$H_0 : \mu \leq 6 \text{ proti hypotéze } H : \mu > 6$$

na hladině významnosti 0.05. Testovací kritérium je podle věty 5.2.1 statistika

$$R = \frac{\overline{X} - \mu_0}{\sigma}\sqrt{n},$$

kde $\mu_0 = 6$, $\sigma = 2$, $n = 9$ a kritický obor je množina

$$W = \Big\{r;\ r > u(1-\alpha)\Big\}$$

Realizace testovacího kritéria je

$$r = -0.75,$$

z tabulek dostaneme

$$W = \Big\{r;\ r > u(0.95)\Big\} = \Big\{r;\ r > 1.645\Big\}.$$

Protože $r \notin W$, nezamítáme hypotézu H_0. Testujme ještě na stejné hladině významnosti hypotézu

$$H_0 : \mu \geq 6 \text{ proti hypotéze } H : \mu < 6.$$

Testovací kritérium a jeho realizace zůstávají stejné, kritický obor je tentokráte množina

$$W = \Big\{r;\ r < -u(1-\alpha)\Big\} = \{r;\ r < -u(0.95)\Big\} = \Big\{r;\ r < -1.645\Big\}.$$

Protože $r \notin W$, nezamítáme hypotézu $H_0 : \mu \geq 6$.

Naměřená data tedy neumožňují rozhodnout o zamítnutí, resp. nezamítnutí hypotézy H_0 v původním testu, tj. hypotézy $\mu \leq 6$.

b) Máme vypočítat

$$P(R \notin W/\mu = 7) = P(R \leq 1.645/\mu = 7), \tag{5.2.4}$$

kde

$$R = \frac{\overline{X} - 6}{2} \cdot 3.$$

Z věty 5.2.1 víme, že R má za podmínky $\mu = 6$ rozdělení $N(0,1)$. Ve vztahu (5.2.4) počítáme pravděpodobnost za podmínky, že skutečná střední hodnota je 7. Za této podmínky nemá rozdělení $N(0,1)$ statistika R, ale statistika

$$R^* = \frac{\overline{X} - 7}{2} \cdot 3.$$

Zřejmě
$$R = R^* + \frac{3}{2}.$$
Vyjádřeme tedy pravděpodobnost ve vztahu (5.2.4) pomocí statistiky R^*, dostaneme

$$\begin{aligned} P(R \le 1.645/\mu = 7) &= P\Big(R^* + \frac{3}{2} \le 1.645/\mu = 7\Big) \\ &= P(R^* \le 0.145/\mu = 7) = \Phi(0.145) \doteq 0.60. \end{aligned}$$

Riziko mylného přijetí hypotézy $\mu \le 6$ v neprospěch hypotézy $\mu = 7$ je přibližně 60%.

c) Máme určit rozsah n výběru z X tak, aby

$$P(R \notin W/\mu = 7) = P(R \le 1.645/\mu = 7) \le 0.05 \tag{5.2.5}$$

kde
$$R = \frac{\overline{X} - 6}{2} \cdot \sqrt{n}.$$
Rozdělení $N(0,1)$ má za podmínky $\mu = 7$ statistika

$$R' = \frac{\overline{X} - 7}{2} \cdot \sqrt{n}.$$

Zřejmě
$$R = R' + \frac{1}{2}\sqrt{n}.$$
Vyjádřeme tedy pravděpodobnost ve vztahu (5.2.5) pomocí statistiky R', dostaneme
$$\begin{aligned} 0.05 &\ge P\Big(R' + \frac{1}{2}\sqrt{n} \le 1.645/\mu = 7\Big) \\ 0.05 &\ge P\Big(R' \le 1.645 - \frac{1}{2}\sqrt{n}/\mu = 7\Big) \\ 0.05 &\ge \Phi\Big(1.645 - \frac{1}{2}\sqrt{n}\Big) \\ \Phi\big(u(0.05)\big) &\ge \Phi\Big(1.645 - \frac{1}{2}\sqrt{n}\Big) \\ u(0.05) &\ge 1.645 - \frac{1}{2}\sqrt{n} \\ -u(0.95) &\ge 1.645 - \frac{1}{2}\sqrt{n}. \end{aligned}$$
Odtud
$$-1.645 \ge -1.645 - \frac{1}{2}\sqrt{n} \Rightarrow \sqrt{n} \ge 6.58 \Rightarrow n \ge 43.2964.$$
Riziko mylného přijetí $H_0 : \mu \le 6$ v neprospěch $\mu = 7$ bude maximálně 5% v případě rozsahu výběru $n \ge 44$.

♣

Příklad 5.2.6. Krychlová pevnost betonu třídy 40 se zkoušela na deseti vzorcích jednak určitou nedestruktivní metodou, jednak destruktivně. Výsledné hodnoty v MPa jsou zapsány v následující tabulce:

Pevnost (x_{1i}) nedestr. m.	51.9	49.1	50.5	49.8	50.1	50.3	52.0	50.0	49.9	50.5
Pevnost (x_{2i}) destr. m.	49.1	48.2	50.5	47.1	50.9	50.5	49.0	51.6	49.5	51.0

Zjistěte, zda se výsledky podle nedestruktivní metody v průměru jen náhodně liší od destruktivních zkoušek. Přípustné riziko omylu je 5%.

Řešení. Nechť X_1 [X_2] je pevnost betonu měřená nedestruktivní [destruktivní] metodou. Máme testovat hypotézu

$$E(X_1) = E(X_2) \text{ proti hypotéze } E(X_1) \neq E(X_2)$$

na hladině významnosti 0.05. Označme $X = X_1 - X_2$, potom $E(X_1) = E(X_2)$ právě tehdy, když $E(X) = 0$. Můžeme tedy testovat hypotézu

$$H_0 : \ E(X) = 0 \text{ proti hypotéze } H : \ E(X) \neq 0$$

na hladině významnosti 0.05. Realizace náhodného výběru z X je

2.4, 0.9, 0.0, 2.7, -0.8, -0.2, 3.0, -1.6, 0.4, -0.5

Má-li náhodná veličina X normální rozdělení (což lze v našem případu přepokládat), lze za zestovací kritérium pro test hypotézy H_0 proti hypotéze H použít statistiku (viz věta 5.2.3)

$$R = \frac{\overline{X} - \mu_0}{S}\sqrt{n},$$

kde $\mu_0 = 0$ a $n = 10$.

Za kritický obor W lze zvolit množinu

$$W = \left\{r;\ r < -t(n-1; 1-\frac{\alpha}{2}) \vee r > t(n-1; 1-\frac{\alpha}{2})\right\}.$$

Postupně dostaneme

$$\overline{x} = 0.630,$$
$$s = \sigma_{n-1} \doteq 1.583,$$
$$t(n-1; 1-\frac{\alpha}{2}) = t(9; 0.975) = 2.262$$

Odtud

$$r \doteq \frac{0.630 - 0}{1.583}\sqrt{10} \doteq 1.259,$$
$$W = \left\{r;\ r < -2.262 \vee r > 2.262\right\}.$$

Protože $r \notin W$, nebyla na hladině výnamnosti 0.05 prokázána statistická významnost v rozdílech obou metod.

Poznámka 5.2.4. Jednovýběrový t-test aplikovaný na porovnávání středních hodnot dvou veličin měřených na stejných objektech se nazývá párový t-test.

Příklad 5.2.7. Při sledování doby bezporuchového chodu určitého zařízení byly zjištěny následující výsledky v hodinách

210, 65, 51, 174, 63, 263, 61, 72.

Předpokládá se, že bezporuchový chod zařízení má přibližne exponenciální rozdělení s parametry 0 a δ. Ověřte na hladině významnosti 0.01, resp. 0.05, zda je střední doba bezporuhového chodu maximálně 70 hodin.

Řešení. Náhodnou veličinou X je zde doba bezporuchového chodu zařízení. Je známo, že $X \sim E(0, \delta)$, tj. náhodná veličina X má hustotu

$$g(x) = \frac{1}{\delta}\,\mathrm{e}^{-\frac{x}{\delta}} \quad \text{pro } x \geq 0$$

(viz poznámka P - 2.2). Zajímají nás testy o střední hodnotě náhodné veličiny X. Protože veličina X nemá normální rozdělení a rozsah výběru je menší než 30, nemůžeme použít testy o střední hodnotě normální náhodné veličiny. V literatuře tedy musíme najít testy o střední hodnotě rozdělení $E(0, \delta)$. Běžně jsou uváděny testy o hodnotě parametru δ rozdělení $E(0, \delta)$ (viz např [14]). Potřebujeme tedy zjistit vztah mezi střední hodnotou veličiny X a parametrem δ. Z teorie pravděpodobnosti víme, že

$$E(X) = \int_{-\infty}^{\infty} x\, g(x)\, dx = \int_{0}^{\infty} x \cdot \frac{1}{\delta}\,\mathrm{e}^{-\frac{x}{\delta}}\, dx.$$

Výpočtem posledního integrálu dostaneme

$$E(X) = \delta.$$

Testy o střední hodnotě rozdělení $E(0, \delta)$ jsou tedy totožné s testy o hodnotách parametru δ tohoto rozdělení. Budeme tedy testovat hypotézu

$$H_0 : \sigma \geq 70 \text{ proti hypotéze } H : \sigma < 70$$

na hladině významnosti 0.01, resp. 0.05. Za testovací kritérum lze pro tento test použít statistiku

$$R = \frac{2n\overline{X}}{\delta_0},$$

která má za podmínky $\delta = \delta_0$ rozdělení $\chi^2(2n)$ (v našem případě je $\delta_0 = 70$). Z kapitoly 4 víme, že statistika $\overline{X}$ je vhodným odhadem střední hodnoty, tedy je vhodným odhadem parametru δ. V prospěch hypotézy H budou tedy svědčit ty realizace $\overline{x}$ statistiky $\overline{X}$, které budou „podstatně“ menší než číslo δ_0, tj. ty realizace $\overline{x}$, pro které bude $\frac{\overline{x}}{\delta_0}$ „podstatně“ menší než číslo jedna, tj. pro které bude realizace r testovacího kritéria R podstatně menší než číslo $2n$, tj. pro které bude r menší než nějaká kladná konstatna menší než $2n$. Kritický obor W pro test hypotézy H_0 proti H na hladině významnosti α je tedy množina

$$W = \Big\{r; r < \chi^2(2n;\alpha)\Big\}.$$

Výpočtem a pomocí tabulek dostaneme

$$\begin{aligned}\overline{x} &= 119.875 \Rightarrow r = 27.4,\\ \chi^2(2n;\alpha) &= \chi^2(16;0.01) = 5.812,.\\ \chi^2(2n;\alpha) &= \chi^2(16;0.05) = 7.962.\end{aligned}$$

Označme $W_{0.01}[W_{0.05}]$ kritický obor pro test H_0 proti H na hladině významnosti 0.01 [0.05]. Potom

$$\begin{aligned}W_{0.01} &= \Big\{r;\ r < 5.812\Big\},\\ W_{0.05} &= \Big\{r;\ r < 7.962\Big\}.\end{aligned}$$

Tedy hypotézu $H_0 : \delta \geq 70$ nezamítáme na hladině významnosti 0.01 ani na hladině významnosti 0.05.

Budeme ještě na obou hladinách významnosti testovat hypotézu

$$H_0 : \delta \leq 70 \text{ proti hypotéze } H : \delta > 70.$$

Testovací kritérium R zůstává stejné, tedy opět

$$r = 27.4.$$

Kritický obor W pro test H_0 proti H na hladině významnosti α bude tentokrát množina

$$W = \Big\{r;\ r > \chi^2(2n; 1-\alpha)\Big\}.$$

Označme opět $W_{0.01}[W_{0.05}]$ kritický obor pro poslední test H_0 proti H na hladině významnosti 0.01 [0.05]. Pomocí tabulek dostaneme

$$\begin{aligned}W_{0.01} &= \Big\{r;\ r > 32.0\Big\},\\ W_{0.05} &= \Big\{r;\ r > 26.3\Big\}.\end{aligned}$$

Tedy hypotézu $H_0 : \delta \leq 70$ nezamítáme na hladině významnosti 0.01 a zamítáme na hladině významnosti 0.05.

Na hladině významnosti 0.01 neumožňují použité testy a zjištěná data rozhodnout o platnosti hypotézy $\delta = E(X) \leq 70$. Na hladině významnosti 0.05 zamítáme hypotézu, že $\delta = E(X) \leq 70$, tedy střední doba bezporuchového chodu je větší než 70 hodin s rizikem omylu maximálně 5%.

5.3. Testy dobré shody

V teorii pravděpodobnosti jsme předpokládali, že známe zákon rozdělení náhodné veličiny X, tj. známe její distribuční funkci, resp. rozdělovací funkci. V matematické statistice jsme pak dosud předpokládali, že známe tvar rozdělení až na nějaké neznámé parametry $\Theta_1, \Theta_2, \ldots, \Theta_m$. K úplné znalosti rozdělení náhodné veličiny X pak v této situaci stačilo odhadnout tyto neznámé parametry. Pokud si však rozdělením, resp. tvarem rozdělení nejsme jisti, je zapotřebí testovat, zda není předpoklad o tvaru rozdělení chybný.

Připomeňme v této souvislosti histogram, resp. histogram relativních četností z kapitoly P - 1. Podle jejich tvaru lze usuzovat na tvar rozdělovací funkce náhodné veličiny X. K tomu je samozřejmě vhodné znát grafy některých důležitých rozdělovacích funkcí, abychom měli histogram s čím srovnávat. Tak např. z histogramu váhy betonové kostky (viz obr. P - 1.2 a obr. P - 1.3) se lze domnívat, že váha betonu bude mít normální a ne např. exponenciální rozdělení. Podobně z histogramu počtu ok při jednom hodu hrací kostkou (viz obr. P - 1.1 a)) lze usuzovat, zda má počet ok klasické rozdělení, nebo nemá, tj. zda je hrací kostka regulérní nebo falešná. Tyto tzv. grafické metody však slouží pouze k vytvoření domněnky o tvaru rozdělení. Tuto domněnku pak samozřejmě musíme testovat, abychom zaručili požadované riziko případného omylu. Právě k testování hypotéz o rozdělení, resp. tvaru rozdělení používáme testy dobré shody.

Testů dobré shody můžeme v literatuře najít celou řadu. My zde uvedeme pouze dva nejpoužívanější testy dobré shody - a to Pearsonův nebo χ^2 test dobré shody a Kolmogorovův test dobré shody. Pro použití obou testů je zapotřebí znát jejich podstatu, vyspělejší uživatel statistiky pak může využít statistický software STATGRAPHICS, kde má možnost použít testy dobré shody v případě porovnávání s osmnácti základními typy rozdělení (viz nabídka STATS $\longmapsto$ Distribution Functions $\longmapsto$ Distribution Fitting). Pokud naše domněnka není souhlasná s některou z osmnácti nabídek rozdělení, můžeme využít nabídku STATS $\longmapsto$ Nonparametrics $\longmapsto$ Kolmogorov-Smirnov One-Sample nebo MODELS $\longmapsto$ Categorial Data Analysis $\longmapsto$ Chi-Square Test, kde ale musíme dopočítat tzv. teoretické četnosti. Přejděme nyní k jednotlivým testům dobré shody.

Pearsonův $[\chi^2]$ test dobré shody

Domníváme se, že náhodná veličina X má rozdělovací funkci $g(x; \Theta_1, \ldots, \Theta_m)$, kde $\Theta_1, \ldots, \Theta_m$ jsou neznámé parametry. Připouštíme i $m = 0$, tj. že rozdělovací funkce je $g(x)$ a nejsou v ní obsaženy žádné neznámé parametry. Je-li $m \geq 1$, použijeme realizaci $(x_1, x_2, \ldots, x_n)$ náhodného výběru $(X_1, X_2, \ldots, X_n)$ z X pro výpočet realizací odhadů parametrů $\Theta_1, \Theta_2, \ldots, \Theta_m$ - označme je $\widehat{\Theta}_1, \widehat{\Theta}_2, \ldots, \widehat{\Theta}_m$. Odhady parametrů $\Theta_1, \Theta_2, \ldots, \Theta_m$ hledáme tzv. metodou minimálního χ^2 (viz např. [1]). Realizace odhadů dosadíme za parametry do rozdělovací funkce g, dostaneme rozdělovací funkci $g(x; \widehat{\Theta}_1, \widehat{\Theta}_2, \ldots, \widehat{\Theta}_m)$, která již neobsahuje neznámé parametry. Pearsonův test dobré shody je pak test hypotézy

$$H_0\colon X \text{ má rozdělovací funkci } g(x; \widehat{\Theta}_1, \widehat{\Theta}_2, \ldots, \widehat{\Theta}_m)$$

proti hypotéze

$$H\colon X \text{ nemá rozdělovací funkci } g(x; \Theta_1, \Theta_2, \ldots, \Theta_m)$$
$$\text{pro žádný z možných parametrů } \Theta_1, \Theta_2, \ldots, \Theta_m$$

na hladině významnosti α.

Příklad 5.3.1. Domníváme se, že náhodná veličina X má rozdělení 1) $N(0,1)$, 2) $N(\mu, 1)$, 3) $N(\mu, \sigma^2)$. V situaci 1 nejsou v rozdělení žádné neznáme parametry, tedy $m = 0$. V situaci 2 máme jeden neznámý parametr, tedy $m = 1$. Musíme vypočítat realizaci $\widehat{\mu}$ odhadu μ. V situaci 3 máme $m = 2$ a musíme vypočítat realizace $\widehat{\mu}$ a $\widehat{\sigma^2}$ odhadů parametrů μ a σ^2. Potom bychom v jednotlivých situacích testovali hypotézy

1) $H_0 : X \sim N(0,1)$ proti $H : X \not\sim N(0,1)$,
2) $H_0 : X \sim N(\widehat{\mu}, 1)$ proti $H : X \not\sim N(\mu, 1)$ pro žádné $\mu \in \mathbb{R}$,
3) $H_0 : X \sim N(\widehat{\mu}, \widehat{\sigma^2})$ proti $H : X \not\sim N(\mu, \sigma^2)$ pro žádné $(\mu, \sigma^2) \in \mathbb{R} \times (0, \infty)$.

V případě zamítnutí hypotézy H_0 v testu 1 nemá veličina X normální rozdělení s parametry 0 a 1, ale může mít normální rozdělení s jinými parametry. V případě zamítnutí hypotézy H_0 v testu 2 nemá veličina X normální rozdělení s rozptylem 1, ale může mít normální rozdělení s rozptylem různým od čísla 1. A v případě zamítnutí hypotézy H_0 v testu 3 nemá veličina X normální rozdělení.

♣

Postup při samotném testu je pak následující:

Předpokládáme, že platí hypotéza H_0, tj. že náhodná veličina X má rozdělovací funkci $g(x; \widehat{\Theta}_1, \widehat{\Theta}_2, \ldots, \widehat{\Theta}_m)$. Obor hodnot náhodné veličiny X rozdělíme do k disjunktních tříd Ω_j $(j = 1, 2, \ldots, k)$ a to zpravidla následovně:

- intervalů v případě spojité náhodné veličiny
- bodů, resp. množin obsahujících body oboru hodnot náhodné veličiny X v případě diskrétní náhodné veličiny .

Označme N_j absolutní četnost třídy Ω_j, tj. počet hodnot náhodného výběru $(X_1, X_2, \ldots, X_n)$, které padnou do třídy Ω_j pro $j = 1, 2, \ldots, k$. Uvědomme si, že N_j je pro každé j náhodnou veličinou. Zřejmě

$$N_1 + N_2 + \cdots + N_k = n.$$

Označme dále

$$p_j = P(X \in \Omega_j / H_0), \; j = 1, 2, \ldots, k. \tag{5.3.1}$$

Zdůrazněme, že pravděpodobnosti p_j počítáme za předpokladu platnosti hypotézy H_0. Protože

$$\Omega_1 \cup \Omega_2 \cup \cdots \cup \Omega_k = \Omega,$$

platí

$$p_1 + p_2 + \cdots + p_k = 1.$$

Za testovací kritérium zvolil K. Pearson statistiku

$$R = \sum_{j=1}^{k} \frac{(N_j - np_j)^2}{np_j}, \tag{5.3.2}$$

která má za platnosti hypotézy H_0 při „dostatečně“ velkém rozsahu výběru n přibližně rozdělení $\chi^2(k-m-1)$. „Dostatečná“ velikost rozsahu n je zaručena tím, že lze realizaci náhodného výběru z X roztřídit do tříd Ω_j tak, že

$$\begin{aligned} np_j &\geq 1 \quad \text{pro každé } j,\\ np_j &\geq 5 \quad \text{pro alespoň } 80\%\ j. \end{aligned} \tag{5.3.3}$$

Pearsonův test dobré shody by se měl tedy používat v případě splnění podmínek (5.3.3).

Zbývá určit kritický obor pro test hypotézy H_0 proti hypotéze H. Kdyby platila hypotéza H_0, pak by pravděpodobnost, že výsledek pokusu X nabude hodnoty ze třídy Ω_j, byla p_j (viz vztah 5.3.1). Odhad pravděpodobnosti, že výsledek pokusu padne do třídy Ω_j, je zřejmě počet příznivých výsledků ku celkovému počtu možných výsledků - tedy náhodná veličina N_j/n. Takže v prospěch hypotézy H_0 svědčí ty realizace n_j/n veličiny N_j/n , které jsou „dostatečně“ blízké číslu p_j, tj. pro které „$n_j/n \doteq p_j$“, tj. pro které „$n_j \doteq np_j$“ pro každé $j = 1, \ldots, k$. Číslo n_j se někdy v této souvislosti nazývá empirická četnost třídy Ω_j a číslo np_j pak teoretická četnost třídy Ω_j. V prospěch hypotézy H_0 svědčí tedy ty realizace r testovacího kritéria R, které jsou „bízké“ číslu nula. Odtud dostáváme, že kritický obor W pro test hypotézy H_0 proti hypotéze H na hladině významnosti α je množina

$$W = \left\{r;\ r > \chi^2(k-m-1; 1-\alpha\right\} \tag{5.3.4}$$

Vraťme se nyní k výpočtu pravděpodobností p_j $(j = 1, 2, \ldots, k)$ v případě spojité náhodné veličiny X. V této situaci je třída Ω_j interval, označme jeho krajní body jako t_{j-1}, t_j pro $j = 1, \ldots, k$. Potom

$$\begin{aligned} p_j &= P\big(X \in \Omega_j/H_0\big) = P\big(t_{j-1} < X < t_j/H_0\big)\\ &= \int_{t_{j-1}}^{t_j} g(x; \widehat{\Theta}_1, \widehat{\Theta}_2, \ldots, \widehat{\Theta}_m)\, dx \quad \text{pro } j = 1, 2, \ldots, k. \end{aligned}$$

Výhodnější pro výpočty se často jeví vyjádřit nejprve v hypotéze H_0 z hustoty g distribuční funkci G náhodné veličiny X a pak teprve počítat p_j. Zřejmě

$$G(x; \widehat{\Theta}_1, \widehat{\Theta}_2, \ldots, \widehat{\Theta}_m) = \int_{-\infty}^{x} g(t; \widehat{\Theta}_1, \widehat{\Theta}_2, \ldots, \widehat{\Theta}_m)\, dt,$$

potom

$$\begin{aligned} p_j &= P\big(X \in \Omega_j/H_0\big) = P\big(t_{j-1} < X < t_j/H_0\big)\\ &= G(t_j; \widehat{\Theta}_1, \widehat{\Theta}_2, \ldots, \widehat{\Theta}_m) - G(t_{j-1}; \widehat{\Theta}_1, \widehat{\Theta}_2, \ldots, \widehat{\Theta}_m) \quad \text{pro } j = 1, 2, \ldots, k. \end{aligned}$$

Příklad 5.3.2. Házeli jsme 120-krát hrací kostkou a získali následující výsledky

Počet ok	1	2	3	4	5	6
Počet výsledků	6	20	20	21	23	30

Ověřte na hladině významnosti 0.05, resp. 0.01, zda je hrací kostka regulérní.

Řešení. Náhodnou veličinou X je zde počet ok, které padnou při jednom hodu hrací kostkou. Obor hodnot Ω náhodné veličiny X je množina $\Omega = \{1, 2, 3, 4, 5, 6\}$, jedná se tedy o diskrétní náhodnou veličinu. Kdyby kostka byla regulární, pak by čísla 1 až 6 padala se stejnou pravděpodobností, tj. náhodná veličina X by měla pravděpodobnostní funkci

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

Budeme tedy testovat hypotézu

$$H_0 : X \sim g(x) \text{ proti hypotéze } H : X \not\sim g(x)$$

na hladině významnosti 0.01, resp. 0.05. K dispozici máme realizaci náhodného výběru z X o rozsahu $n = 120$, která byla roztříděna do šesti tříd

$$\Omega_j = j \quad \text{pro } j = 1, 2, 3, 4, 5, 6.$$

Ptáme se na rozdělení náhodné veličiny X, takže použijeme Pearsonův test dobré shody, pokud budou splněny podmínky jeho použitelnosti, tj. podmínky (5.3.3). K ověření těchto podmínek i k výpočtu realizace r testovacího kritéria R (viz (5.3.2)), musíme vypočítat všechny pravděpodobnosti (5.3.1). Tedy

$$p_j = P(X \in \Omega_j / H_0) = P(X = j / H_0) \quad \text{pro } j = 1, 2, 3, 4, 5, 6.$$

Platí-li hypotéza H_0, má náhodná veličina X rozdělovací funkci g a tedy

$$p_j = g(j) = \frac{1}{6} \quad \text{pro } j = 1, 2, 3, 4, 5, 6.$$

Odtud

$$np_j = 20 \quad \text{pro } j = 1, 2, 3, 4, 5, 6$$

a jsou tedy splněny podmínky použitelnosti Pearsonova testu dobré shody. Přejdeme tedy k jeho provedení. Vypočítáme realizaci r testovacího kritéria (5.3.2), tj. statistiky

$$R = \sum_{j=1}^{k} \frac{(N_j - np_j)^2}{np_j}$$

- výpočty jsou v tabulce 5.3.1. V posledním řádku tabulky jsou prováděny součty (tam, kde to má smysl) jednak pro kontrolu výpočtů (sloupce 3 - 5) a jednak pro samotný výpočet realizace r testovacího kritéria R. Tedy

$$r = \sum_{j=1}^{6} \frac{(n_j - np_j)^2}{np_j} = \frac{(6-20)^2}{20} + \cdots + \frac{(30-20)^2}{20}$$
$$= 9.80 + \cdots + 5.00 = 15.30.$$

Kritický obor W pro Perasonův test dobré shody na hladině významnosti α je (5.3.4), tj.

$$W = \Big\{r;\ r > \chi^2(k-m-1; 1-\alpha)\Big\}.$$

Vzhledem k tomu, že počet tříd je šest (tj. $k = 6$) a v rozdělení g nejsou neznámé parametry (tj. $m = 0$), dostaneme

$$W = \Big\{r;\ r > \chi^2(5; 1-\alpha)\Big\}.$$

Označme nyní $W_{0.05}[W_{0.01}]$ kritické obory pro tento test na hladině významnosti 0.05 [0.01], potom

$$W_{0.05} = \Big\{r;\ r > \chi^2(5; 0.95)\Big\} = \Big\{r;\ r > 11.07\Big\},$$
$$W_{0.01} = \Big\{r;\ r > \chi^2(5; 0.99)\Big\} = \Big\{r;\ r > 15.09\Big\}.$$

Protože $r \in W_{0.05}$, na hladině významnosti 0.05 zamítáme hypotézu H_0 a přijímáme hypotézu H, tedy hrací kostka je falešná, riziko omylu je maximálně 5%. Protože $r \notin W_{0.01}$, na hladině významnosti 0.01 nezamítáme hypotézu H_0.

Třída	j	n_j	p_j	np_j	$\frac{(n_j-np_j)^2}{np_j}$
Ω_1	1	6	$\frac{1}{6}$	20	9.80
Ω_2	2	20	$\frac{1}{6}$	20	0.00
Ω_3	3	20	$\frac{1}{6}$	20	0.00
Ω_4	4	21	$\frac{1}{6}$	20	0.05
Ω_5	5	23	$\frac{1}{6}$	20	0.45
Ω_6	6	30	$\frac{1}{6}$	20	5.00
Součet	×	120	1	120	15.30

Tabulka 5.3.1

Vraťme se ještě k tabulce 5.3.1. Ve sloupci 5 jsou uvedena čísla np_j, tj. teoretické četnosti třídy Ω_j, ve sloupci 3 jsou pak uvedeny empirické četnosti n_j. Kdyby platila hypotéza H_0, pak by v každé třídě mělo být přibližně 20 výsledků. Největší rozdíly jsou tedy v první a poslední třídě.

♣

PDF-XChange Click to buy NOW! www.docu-track.com

Příklad 5.3.3. Prohlídkou pěti set tabulí skla byly zjištěny tyto počty bublin v jednotlivých tabulích

Počet bublin	0	1	2	3	4 a více
Počet tabulí	246	181	64	7	2

Ověřte na hladině významnosti 0.05, zda má počet bublin v tabuli skla Poissonovo rozdělení.

Řešení. Náhodnou veličinou X je zde počet bublin v tabuli skla. Máme ověřit, zda má náhodná veličina Poisonovo rozdělení. Budeme chtít použít χ^2 test dobré shody. K dispozici máme realizaci náhodného výběru z X o rozsahu $n = 500$, která byla roztříděna do pěti tříd. V první řadě si uvědomme, že četnost v poslední třídě je velmi malá. Dá se očekávat, že by i teoretická četnost odpovídající této třídě byla velmi malá, a proto sloučíme poslední dvě třídy. Východiskem pro výpočty jsou tedy první tři sloupce tabulky 5.3.2.

Třída	Počet bublin	n_j	p_j	np_j	$\frac{(n_j-np_j)^2}{np_j}$
Ω_1	0	246	0.511	255.5	0.353
Ω_2	1	181	0.343	171.5	0.526
Ω_3	2	64	0.115	57.5	0.735
Ω_4	3 a více	9	0.031	15.5	2.726
Součet	$\times$	500	1.000	500.0	4.340

Tabulka 5.3.2

Domníváme-li se, že má náhodná veličina X Poissonovo rozdělení, tj. $X \sim Po(\lambda)$, neznáme v rozdělovací funkci veličiny X parametr λ, proto λ musíme nejprve odhadnout. Z věty P - 9.1.2 víme, že $E(X) = \lambda$, z věty 3.2 víme, že $\overline{X}$ je nejlepším nestranným odhadem prametru λ. Vypočítejme tedy z dat v tabulce 5.3.2 realizaci $\widehat{\lambda} = \overline{x}$ tohoto odhadu parametru λ. Dostaneme

$$\widehat{\lambda} = \overline{x} = 0.672.$$

(Správněji bychom měli použít pro λ jiný odhad - $\overline{X}$ je pouze počáteční aproximace odhadu λ v metodě minimálního χ^2.) Budeme testovat hypotézu

$$H_0 : X \sim Po(0.672) \quad \text{proti hypotéze} \quad H : X \not\sim Po(\lambda)$$

na hladině významnosti 0.05. Předpokládejme tedy, že platí hypotéza H_0, tj. náhodná veličina X má pravděpodobnostní funkci

$$g(x) = \begin{cases} \dfrac{0.672^x \, e^{-0.672}}{x!} & \text{pro } \ x = 0, 1, 2, \ldots \\ 0 & \text{jinak} \end{cases}$$

PDF-XChange Click to buy NOW! www.docu-track.com

(viz definice P - 9.1.2). Potřebujeme vypočítat teoretické pravděpodobnosti (5.3.1). Postupně dostaneme

$$\begin{aligned}
p_1 &= P(X \in \Omega_1/H_0) = P(X = 0/H_0) = g(0) = \frac{0.672^0\, \mathrm{e}^{-0.672}}{0!} \doteq 0.511,\\
p_2 &= P(X \in \Omega_2/H_0) = P(X = 1/H_0) = g(1) = \frac{0.672^1\, \mathrm{e}^{-0.672}}{1!} \doteq 0.343,\\
p_3 &= P(X \in \Omega_3/H_0) = P(X = 2/H_0) = g(2) = \frac{0.672^2\, \mathrm{e}^{-0.672}}{2!} \doteq 0.115,\\
p_4 &= P(X \in \Omega_3/H_0) = P(X \geq 3/H_0) = 1 - P(X < 3/H_0)\\
&= 1 - \left[P(X = 0/H_0) + P(X = 1/H_0) + P(X = 2/H_0)\right]\\
&= 1 - \left[g(0) + g(1) + g(2)\right] \doteq 0.031.
\end{aligned}$$

Výsledky jsou uvedeny ve sloupci 4 tabulky 5.3.2. Ze sloupce 5 téže tabulky plyne, že jsou splněny podmínky použitelnosti χ^2 testu (5.3.3). Přejděme tedy k výpočtu realizace r testovacího kritéria (5.3.2), dostaneme

$$r = \sum_{j=1}^{4} \frac{(n_j - np_j)^2}{np_j} \doteq \frac{(246 - 255.5)^2}{255.5} + \cdots + \frac{(9 - 15.5)^2}{15.5} \doteq 4.340$$

(viz součet v posledním sloupci tabulky 5.3.2). Kritický obor W pro χ^2 test na hladině významnosti α je (5.3.4), tj.

$$W = \left\{r;\ r > \chi^2(k - m - 1; 1 - \alpha)\right\}.$$

Protože počet stupňů volnosti je $k - m - 1$, kde $k = 4$ je počet tříd a $m = 1$ je počet ohadnutých prametrů a požadovaná hladina významnosti testu je 0.05, dostaneme

$$W = \left\{r;\ r > \chi^2(2; 0.95)\right\} = \left\{r;\ r > 5.992)\right\}.$$

Protože $r \notin W$, nezamítáme hypotézu H_0 na hladině významnosti 0.05, tj. nezamítáme hypotézu o shodě s Poissonovým rozdělením.

♣

Příklad 5.3.4. Ověřte na hladině významnosti 0.01, resp. 0.05, zda má náhodná veličina X rozdělovací funkci

$$g(x) = \begin{cases} x - 1 & \text{pro } x \in \langle 1, 2) \\ 3 - x & \text{pro } x \in (2, 3\rangle \\ 0 & \text{jinak} \end{cases}.$$

Realizace náhodného výběru z X byla roztříděna následovně.

Interval	1.0 – 1.4	1.4 – 1.8	1.8 – 2.2	2.2 – 2.6	2.6 – 3.0
Počet výsledků	1	4	18	4	3

PDF-XChange
Click to buy NOW!
www.docu-track.com

Řešení. Budeme testovat hypotézu

$$H_0 : X \sim g(x) \quad \text{proti hypotéze} \quad H_0 : X \not\sim g(x)$$

na hladině významnosti 0.01, resp. 0.05. Budeme chtít použít χ^2 test dobré shody. Předpokládejme tedy, že platí hypotéza H_0, tj. náhodná veličina X má hustotu g. K dispozici máme realizaci náhodného výběru z X o rozsahu $n = 30$, která je roztříděna do pěti intervalů - označme je Ω_j ($j = 1, \ldots, 5$). Nejprve vypočítáme pravděpodobnosti (5.3.1), tj.

$$p_j = P\big(X \in \Omega_j / H_0\big) \quad \text{pro} \;\; j = 1, \ldots, 5.$$

Označme t_{j-1}, t_j krajní body intervalu Ω_j pro $j = 1, \ldots, 5$. Potom

$$p_j = P\big(X \in \Omega_j / H_0\big) = P\big(t_{j-1} < X < t_j / H_0\big) \quad \text{pro} \;\; j = 1, \ldots, 5.$$

Pravděpodobnosti p_j budeme počítat pomocí distribuční funkce G náhodné veličiny X. Postupně dostaneme

$$G(x) = \int_{-\infty}^{x} g(t)\,dt = \begin{cases} 0 & \text{pro} \;\; x \in (-\infty, 1) \\ \displaystyle\int_1^x (t-1)\,dt = \frac{x^2}{2} - x + 0.5 & \text{pro} \;\; x \in \langle 1, 2) \\ 0.5 + \displaystyle\int_2^x (3-t)\,dt = -\frac{x^2}{2} + 3x - 3.5 & \text{pro} \;\; x \in \langle 2, 3\rangle \\ 1 & \text{pro} \;\; x \in (3, \infty) \end{cases}$$

Potom

$$p_j = G(t_j) - G(t_{j-1}) \quad \text{pro} \;\; j = 1, \ldots, 5.$$

Tedy

$$\begin{aligned} p_1 &= G(t_1) - G(t_0) = G(1.4) - G(1.0) = 0.08 - 0.00 = 0.08, \\ p_2 &= G(t_2) - G(t_1) = G(1.8) - G(1.4) = 0.32 - 0.08 = 0.24, \\ &\;\vdots \\ p_5 &= G(t_5) - G(t_4) = G(3.0) - G(2.6) = 1.00 - 0.92 = 0.08. \end{aligned}$$

Vypočítané hodnoty $G(t_j)$ v pravých krajních bodech intervalů Ω_j jsou ve sloupci 4, hodnoty p_j jsou ve sloupci 5 tabulky 5.3.3. Ze sloupce 6 téže tabulky je patrné, že není splněna druhá z podmínek (5.3.3) použitelnosti χ^2 testu. Tento nedostatek lze v naší situaci odstranit tak, že sloučíme první třídu s druhou a předposlední s poslední, dostaneme tak prvních šest sloupců v tabulce 5.3.4. Obecně se to však podařit nemusí. V praxi se pokládá za dostačující, aby byl rozsah výběru větší než 50. Nyní můžeme přistoupit k samotnému provedení χ^2 testu.

Třída	$t_{j-1} - t_j$	n_j	$G(t_j)$	p_j	np_j
Ω_1	$1.0 - 1.4$	1	0.08	0.08	2.4
Ω_2	$1.4 - 1.8$	4	0.32	0.24	7.2
Ω_3	$1.8 - 2.2$	18	0.68	0.36	10.8
Ω_4	$2.2 - 2.6$	4	0.92	0.24	7.2
Ω_5	$2.6 - 3.0$	3	1.00	0.08	2.4
Součet	$\times$	30	$\times$	1.00	30

Tabulka 5.3.3

Třída	$t_{j-1} - t_j$	n_j	$G(t_j)$	p_j	np_j	$\frac{(n_j - np_j)^2}{np_j}$
Ω_1	$1.0 - 1.8$	5	0.32	0.32	9.6	2.204
Ω_2	$1.8 - 2.2$	18	0.68	0.36	10.8	4.800
Ω_3	$2.2 - 3.0$	7	1.00	0.32	9.6	0.704
Součet	$\times$	30	$\times$	1.00	30.0	7.708

Tabulka 5.3.4

Realizace r testovacího kritéria (5.3.2) je

$$r = \sum_{j=1}^{3} \frac{(n_j - np_j)^2}{np_j} = \frac{(5-9.6)^2}{9.6} + \frac{(18-10.8)^2}{10.8} + \frac{(7-9.6)^2}{9.6}$$
$$= 2.204 + 4.800 + 2.204 = 7.708$$

Kritický obor W pro χ^2 test na hladině významnosti α je (5.3.4), tj.

$$W = \left\{r;\ r > \chi^2(k - m - 1; 1 - \alpha)\right\}.$$

V hustotě g v hypotéze H_0 nebyly žádné neznámé parametry, tj. $m = 0$. Realizaci náhodného výběru z X máme přerozdělenou do tří tříd, tj. $k = 3$. Tedy

$$W = \left\{r;\ r > \chi^2(2; 1 - \alpha)\right\}.$$

Označme $W_{0.01}$ [$W_{0.05}$] kritický obor pro náš test na hladině významnosti 0.01 [0.05]. Potom

$$W_{0.01} = \left\{r;\ r > \chi^2(2; 0.99)\right\} = \left\{r;\ r > 9.210\right\},$$

$$W_{0.05} = \Big\{r;\ r > \chi^2(2; 0.95)\Big\} = \Big\{r;\ r > 5.992)\Big\}.$$

Protože $r \notin W_{0.01}$, nezamítáme hypotézu H_0 na hladině významnosti 0.01. Protože $r \in W_{0.05}$, zamítáme hypotézu H_0 na hladině významnosti 0.05, tedy na této hladině významnosti přijímáme hypotézu H, tj. realizace náhodného výběru nepochází z rozdělení s hustotou g, riziko omylu je 5%.

♣

Příklad 5.3.5. Vraťte se k příkladu P - 1.7 a zjistětě, zda je váha betonové kostky normální náhodná veličina. Přípustné riziko omylu je 5%.

Řešení. Náhodnou veličinou X je váha betonové kostky. Domníváme-li se, že X je normální náhodná veličina, tj. $X \sim N(\mu, \sigma^2)$, neznáme v rozdělovací funkci náhodné veličiny X dva parametry μ a σ^2, které musíme nejprve odhadnout. K dispozici máme realizaci náhodného výběru z X o rozsahu $n = 100$, která byla v příkladu P - 1.7 roztříděna do pěti tříd - označme je Ω_j $j = 1, \dots, 5$. Realizace $\widehat{\mu}$ a $\widehat{\sigma}$ odhadů μ a σ vypočítáme z takto roztříděných dat. Dostaneme (viz poznámka 3.5 7))

$$\widehat{\mu} = \overline{x} = 2.174 \text{ [kg]}$$
$$\widehat{\sigma} = s \doteq 0.172 \text{ [kg]}.$$

(Přesněji bychom měli pro χ^2 test volit $\widehat{\sigma} = \sqrt{m}$, ale při velkém rozsahu je $\sqrt{m} \doteq s$ - viz poznámka 3.5 5).)

Budeme testovat hypotézu

$$H_0 : X \sim N(2.174, 0.172^2) \quad \text{proti hypotéze} \quad H : X \nsim N(\mu, \sigma^2), \mu > 0, \sigma^2 > 0$$

na hladině významnosti 0.05. Předpokládejme tedy, že $X \sim N(2.174, 0.172^2)$ K ověření podmínek použitelnosti Pearsonova testu a provedení samotného testu potřebujeme vypočítat pravděpodobnosti (5.3.1), tj.

$$p_j = P(X \in \Omega_j / H_0).$$

Označme opět t_{j-1}, t_j krajní body intervalu Ω_j pro $j = 1, \dots, 5$. Potom

$$p_j = P(t_{j-1} < X < t_j / H_0).$$

Platí-li hypotéza H_0 má veličina X rozdělení $N(2.174, 0.172^2)$. Označme její distribuční funkci F. Potom (viz věta P - 9.2.3)

$$X \sim N(2.174, 0.172^2) \sim F(x) = \Phi\Big(\frac{x - 2.174}{0.172}\Big).$$

Tedy pro $j = 1, 2, \dots, 5$ platí

$$p_j = P(t_{j-1} < X < t_j / H_0) = F(t_j) - F(t_{j-1})$$
$$= \Phi\Big(\frac{t_j - 2.174}{0.172}\Big) - \Phi\Big(\frac{t_{j-1} - 2.174}{0.172}\Big).$$

Stačí tedy pomocí tabulek distribuční funkce Φ rozdělení $N(0,1)$ určit hodnoty distribuční funkce $\Phi(\frac{t_j - 2.174}{0.172})$. Výsledky jsou uvedeny ve sloupci 6 tabulky 5.3.5. Potom

$$
\begin{aligned}
p_1 &= \Phi\Big(\frac{t_1 - 2.174}{0.172}\Big) - \Phi(-\infty) = \Phi\Big(\frac{t_1 - 2.174}{0.172}\Big) = \Phi(-1.59) = 0.0559,\\
p_2 &= \Phi\Big(\frac{t_2 - 2.174}{0.172}\Big) - \Phi\Big(\frac{t_1 - 2.174}{0.172}\Big) = \Phi(-0.43) - \Phi(-1.59)\\
&= 0.3336 - 0.0559 = 0.2777,\\
&\;\vdots\\
p_5 &= \Phi\Big(\frac{t_5 - 2.174}{0.172}\Big) - \Phi\Big(\frac{t_4 - 2.174}{0.172}\Big) = \Phi(\infty) - \Phi(2.6)\\
&= 1.000 - 0.9712 = 0.0288.
\end{aligned}
$$

Hodnoty p_j jsou uvedeny ve sloupci 7 téže tabulky. Ze sloupce 8 je patrné, že jsou splněny podmínky požitelnosti χ^2 testu, tj. podmínky (5.3.3).

Třída	$t_{j-1} - t_j$	n_j	$\bar{x}_j$	$\frac{t_j - \bar{x}}{s}$	$\Phi(\frac{t_j - \bar{x}}{s})$	p_j	np_j	$\frac{(n_j - np_j)^2}{np_j}$
Ω_1	$-\ 1.9$	5	1.8	-1.59	0.0559	0.0559	5.59	0.062
Ω_2	$1.9 - 2.1$	27	2.0	-0.43	0.3336	0.2777	27.77	0.021
Ω_3	$2.1 - 2.3$	46	2.2	0.73	0.7673	0.4337	43.37	0.159
Ω_4	$2.3 - 2.5$	20	2.4	1.90	0.9712	0.2039	20.39	0.007
Ω_5	$2.5 -$	2	2.6	∞	1.0000	0.0288	2.88	0.269
Součet	$\times$	100	$\times$	$\times$	$\times$	1.0000	100.00	0.518

Tabulka 5.3.5

Ze sloupce 9 dostáváme realizaci r testovacího kritéria χ^2 testu (5.3.3), tj.

$$r = \sum_{j=1}^{5} \frac{(n_j - np_j)^2}{np_j} \doteq 0.518.$$

Kritický obor W pro χ^2 test na hladině významnosti α je (5.3.4), tj.

$$W = \Big\{r;\ r > \chi^2(k - m - 1; 1 - \alpha)\Big\}.$$

Protože počet stupňů volnosti je $k - m - 1$, kde $k = 5$ je počet tříd a $m = 2$ je počet odhadnutých parametrů a požadovaná hladina významnosti testu je 0.05, dostaneme

$$W = \Big\{r;\ r > \chi^2(2; 0.95)\Big\} = \Big\{r;\ r > 5.992)\Big\}.$$

Protože $r \notin W$, nezamítáme hypotézu H_0 na hladině významnosti 0.05, tj. nezamítáme hypotézu o shodě s normálním rozdělením.

Příklad 5.3.6. Nakreslete v příkladech 5.3.2 - 5.3.5 histogramy relativních četností, teoretické rozdělovací funkce v nulových hypotézách a porovnejte je.

♣

Na závěr kapitoly se budeme věnovat dalšímu testu dobré shody.

Kolmogorovův test dobré shody

Pomocí Kolmogorovova testu shody se ověřuje hypotéza H_0 : Náhodná veličina X má rozdělení s distribuční funkcí $F(x)$, kde $F(x)$ je daná distribuční funkce spojitého typu. Při tomto testu se porovnává teoretická distribuční funkce $F(x)$ s tzv. empirickou distribuční funkcí, která je odhadem distribuční funkce veličiny X. Připomeňme, že

$$F(x) = P\big(X \le x\big).$$

Buď tedy $(X_1, X_2, \dots, X_n)$ náhodný výběr z X. <u>Empirickou distibuční funkci</u> (příslušnou náhodnému výběru $(X_1, X_2, \dots, X_n)$ z X) definujeme jako funkci $F_n(x)$ přiřazující každému $x \in \mathbb{R}$ relativní četnost pozorování menších nebo rovných x. Značí-li tedy M_x počet veličin X_i splňujících nerovnost $X_i \le x$, je

$$F_n(x) = \frac{M_x}{n}.$$

Zdůrazněme, že $F_n(x)$ je pro každé $x \in \mathbb{R}$ náhodnou veličinou.

Buď $(x_1, x_2, \dots, x_n)$ realizace náhodného výběru $(X_1, X_2, \dots, X_n)$ z X. Potom realizaci empirické distribuční funkce dostaneme následovně:

Uspořádejme hodnoty $x_1, x_2, \dots, x_n$ vzestupně podle velikosti. Nejmenší z nich označme $x_{(1)}$, druhou nejmenší $x_{(2)}$, až největší $x_{(n)}$. Platí tedy

$$x_{(1)} \le x_{(2)} \le \dots \le x_{(n)}.$$

Pokud jsou všechny hodnoty $x_1, \dots, x_n$ navzájem různé (tj. $x_1 < \dots < x_n$), bude realizace funkce $F_n(x)$ schodovitá fukce rovná číslu nula pro $x < x_{(1)}$ (protože ve výběru není žádné pozorování menší než $x_{(1)}$), rovná $1/n$ pro všechna $x \in \big\langle x_{(1)}, x_{(2)}\big)$ atd., tj. je konstantní v intervalech $\big\langle x_{(i)}, x_{(i+1)}\big)$ se skokem velikosti $1/n$ v bodě $x_{(i)}$. To znamená, že lze psát

$$F_n(x) = \begin{cases} 0 & \text{pro } -\infty < x < x_{(1)} \\ \dfrac{i}{n} & \text{pro } x_{(i)} \le x < x_{(i+1)},\ i = 1, 2, \dots, n-1 \\ 1 & \text{pro } x_{(n)} \le x < \infty \end{cases} \tag{5.3.5}$$

Pokud se však hodnota $x_{(i)}$ vyskytuje v realizaci $(x_1, x_2, \dots, x_n)$ k-krát, zůstává sice realizace funkce $F_n(x)$ schodovitá, ale v bodě $x_{(i)}$ má skok o velikosti k/n.

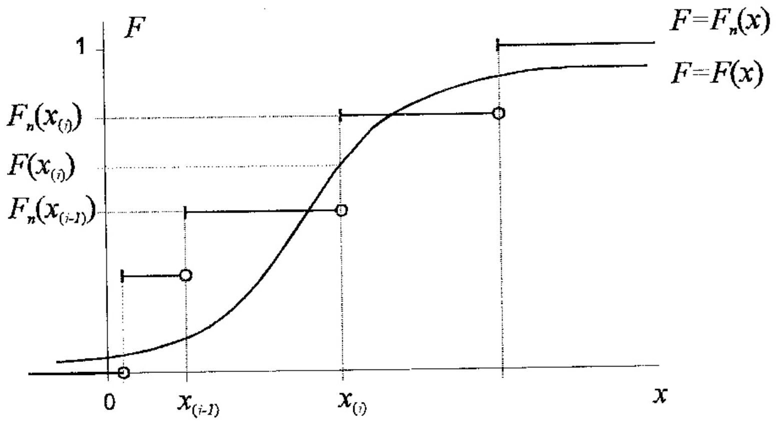
Za testovací kritérium pro test hypotézy H_0 použil A. N. Kolmogorov statistiku

$$D = \sup_x \Big| F_n(x) - F(x) \Big|. \tag{5.3.6}$$

Protože $F(x)$ jakožto distribuční funkce je neklesající a realizace $F_n(x)$ je funkce po částech konstantní a mění svou hodnotu skokem v bodech $x_{(1)}, x_{(2)}, \dots, x_{(n)}$, lze realizaci d statistiky D zapsat jako

$$d = \max_{1 \le i \le n} \left\{ \max \left[\left| F_n(x_{(i)}) - F(x_{(i)}) \right|, \left| F_n(x_{(i-1)}) - F(x_{(i)}) \right| \right] \right\} \tag{5.3.7}$$

(viz obrázek 5.3.1).



Obrázek 5.3.1

Tabelovány jsou např. tzv. kritické hodnoty $D(n;\alpha)$ s vlastností

$$P\big(D > D(n;\alpha)/H_0\big) = \alpha$$

(viz Tab. 5 nebo [13]). Při testu hypotézy

$$H_0 : X \sim F(x) \quad \text{proti hypotéze} \quad H : X \not\sim F(x)$$

svědčí v prospěch hypotézy H_0 ty realizace d testovacího kritéria D, které jsou blízké číslu nula. Kritickým oborem W pro test H_0 proti opačné hypotéze H na hladině významnosti α je tedy množina

$$W = \Big\{ d;\ d > D(n;\alpha) \Big\}. \tag{5.3.8}$$

Přísně vzato, má se Kolmogorovova testu používat jen k ověřování hypotéz, které stanoví distribuční funkci $F(x)$ jednoznačně, tj. v případě, kdy nejsou v distribuční funkci obsaženy neznámé parametry $\Theta_1, \Theta_2, \dots, \Theta_m$. Kdyby totiž nulová hypotéza byla $X \sim F(x, \Theta_1, \Theta_2, \dots, \Theta_m)$, museli bychom najít odhady $\widehat{\Theta}_1, \widehat{\Theta}_2, \dots, \widehat{\Theta}_m$ neznámých parametrů $\Theta_1, \Theta_2, \dots, \Theta_m$ a ve statistice D pracovat místo s funkcí $F(x)$ s funkcí $F(x, \widehat{\Theta}_1, \widehat{\Theta}_2, \dots, \widehat{\Theta}_m)$. Tím by se však změnilo rozdělení statistiky D, a v důsledku toho i rizika chyb.

Testu se má dále používat jen při znalosti všech hodnot z realizace $(x_1, \dots, x_n)$, nikoliv při realizaci roztříděné do tříd. Pokud je totiž realizace roztříděna do tříd, nelze určit celý průběh empirické distribuční funkce $F_n(x)$. Přesto se Kolmogorovův test používá i u roztříděného výběru. Postupuje se následovně. Předpokládejme, že bude výběr roztříděn do tříd $\Omega_j = (t_{j-1}, t_j\rangle$ z absolutními četnostmi N_j $(j = 1, 2, \ldots, k)$. Potom jsme schopni určit empirickou distribuční funkci $F_n(x)$ pouze v bodech $t_j = 1, 2, \ldots, k$. Zřejmě

$$(5.3.9) \qquad F_n(t_j) = \frac{1}{n} \sum_{s=1}^{j} N_s \quad \text{pro } j = 1, 2, \ldots, k.$$

Označme

$$(5.3.10) \qquad D^* = \max_{1 \le j \le k} \left| F_n(t_j) - F(t_j) \right|.$$

Potom zřejmě

$$\max_{1 \le j \le k} \left| F_n(t_j) - F(t_j) \right| \le \sup_x \left| F_n(x) - F(x) \right|.$$

A tedy

$$D^* \le D.$$

Zvolme pro test hypotézy H_0 proti H testovací kritérium D^* a za kritický obor W^* množinu

$$(5.3.11) \qquad W^* = \left\{ d^*;\ d^* > D(n; \alpha) \right\}.$$

Prozkoumejme nyní pravděpodobnosti obou druhů chyb. Zřejmě

$$\begin{aligned} P(D^* \in W^*/H_0) &= P(D^* > D(n;\alpha)/H_0) \le P(D > D(n;\alpha)/H_0) \\ &= P(D \in W/H_0) = \alpha. \end{aligned}$$

Tedy hladina významnosti testu se oproti původní může snížit. Tj. riziko mylného zamítnutí hypotézy H_0 v neprospěch H je maximálně $100\alpha\%$.
Dále platí

$$\begin{aligned} P(D \notin W/H) &= P(D \le D(n;\alpha)/H) \le P(D^* \le D(n;\alpha)/H) \\ &= P(D^* \notin W^*/H). \end{aligned}$$

Tedy pravděpodobnost chyby druhého druhu se oproti původní může zvětšit. Tj. použití Kolmogorovova testu v případě roztříděných dat může mít za následek snížení síly testu.

Příklad 5.3.7. Byly zjištěny následující doby čekání na trolejbus (v minutách)

7.25, 9.00, 1.00, 7.50, 8.75, 1.25, 8.50, 9.50.

Ověřte na hladině významnosti 0.05, zda má doba čekání rovnoměrné rozdělení s parametry 0, 10.

Řešení. Náhodnou veličinou X je zde doba čekání na trolejbus. Náhodná veličina X má rovnoměrné rozdělení s parametry 0, 10, jestliže má hustotu

$$f(x) = \begin{cases} \dfrac{1}{10} & \text{pro } x \in \langle 0, 10 \rangle \\ 0 & \text{jinak} \end{cases} .$$

(viz poznámka P - 3.3). Potom pro distribuční funkci F náhodné veličiny X platí

$$F(x) = \int_{-\infty}^{x} f(t)\,dt = \begin{cases} 0 & \text{pro } x \in (-\infty, 0) \\ \dfrac{x}{10} & \text{pro } x \in \langle 0, 10 \rangle \\ 1 & \text{pro } x \in (10, \infty) \end{cases} .$$

Budeme tedy testovat hypotézu

$$H_0 : X \sim F(x) \quad \text{proti hypotéze} \quad H : X \not\sim F(x)$$

na hladině významnosti 0.05. Protože rozsah výběru je malý, nemůžeme použít Perasonův test dobré shody. Použijeme tedy Kolmogorovův test. K dispozici máme celou realizaci náhodného výběru z X o rozsahu $n = 8$. Vypočítáme tedy realizaci d testovacího kritéria (5.3.6). Pro výpočty použijeme vzorec (5.3.7). Musíme tedy určit realizaci empirické distribuční funkce $F_n(x)$, k tomu je zapotřebí seřadit hodnoty realizace vzestupně. Dostaneme sloupec 2 v tabulce 5.3.6. Protože jsou hodnoty realizace náhodného výběru z X navzájem různé, dostaneme realizaci $F_n(x)$ podle vztahu (5.3.5), kde $n = 8$. Hodnoty $F_8(x)$ v bodech $x_{(i)}$ pro $i = 1, 2, \ldots, 8$ jsou ve sloupci 3 téže tabulky. (Nakreslete grafy funkcí $F_n(x)$ a $F(x)$ a porovnejte je!) Dále vypočítáme hodnoty $F(x_{(i)})$ pro $i = 1, 2, \ldots, 8$ - výsledky jsou ve sloupci 4 téže tabulky. A konečně pro výpočet realizace d statistiky (5.3.6) vypočítáme

$$\left| F_8(x_{(i)}) - F(x_{(i)}) \right| \quad \text{pro } i = 1, 2, \ldots, 8,$$

$$\left| F_8(x_{(i-1)}) - F(x_{(i)}) \right| \quad \text{pro } i = 1, 2, \ldots, 8.$$

($F_8(x_{(0)})$ je zřejmě nula.)

Z posledních dvou sloupců tabulky 5.3.6, pak s využitím vztahu (5.3.7) dostaneme

$$d = 0.475.$$

Kritický obor W pro test hypotézy H_0 proti hypotéze H je množina (5.3.8). Protože $n = 8$ a $\alpha = 0.05$ z tabulek dostaneme

$$W = \left\{ d;\ d > D(n; \alpha) \right\} = \left\{ d;\ d > D(8; 0.05) \right\} = \left\{ d;\ d > 0.45427 \right\}.$$

Protože $d \in W$, zamítáme hypotézu H_0. Tedy doba čekání na trolejbus nemá rovnoměrné rozdělení s parametry 0,10, riziko omylu je 5%. Poznamenejme ale, že může mít rovnoměrné rozdělení s jinými parametry.

PDF-Xchange
Click to buy NOW!
www.docu-track.com

i	$x_{(i)}$	$F_8(x_{(i)})$	$F(x_{(i)})$	$\lvert F_8(x_{(i)}) - F(x_{(i)})\rvert$	$\lvert F_8(x_{(i-1)}) - F(x_{(i)})\rvert$
1	1.00	0.125	0.100	0.025	0.100
2	1.25	0.250	0.125	0.125	0.000
3	7.25	0.375	0.725	0.335	0.475
4	7.50	0.500	0.750	0.025	0.375
5	8.50	0.625	0.850	0.225	0.350
6	8.75	0.750	0.875	0.125	0.250
7	9.00	0.875	0.900	0.025	0.150
8	9.50	1.000	0.950	0.050	0.075

Tabulka 5.3.6

♣

Příklad 5.3.8. Řešte příklad 5.3.4 pomocí Kolmogorovova testu.

Řešení. Máme testovat hypotézu

$$H_0 : X \sim G(x) \quad \text{proti hypotéze} \quad H : X \sim G(x)$$

na hladině významnosti 0.01, resp. 0.05, kde je distribuční funkce G vypočítána v příkladu 5.3.4. Vzhledem k tomu, že je realizace náhodného výběru roztříděna do tříd, použijeme testovací kritérium (5.3.10). Vypočtené hodnoty teoretické distribuční funkce $G(x)$ v bodech t_j $(j = 1, 2, \ldots, 5)$ jsou ve sloupci 4 tabulky 5.3.7. Hodnoty empirické distribuční funkce $G_n(x)$ v bodech t_j $(j = 1, 2, \ldots, 5)$ vypočítáme podle vztahu (5.3.9), kde $n = 30$, $k = 5$. Potom

$$F_{30}(t_j) = \frac{1}{30} \sum_{s=1}^{j} n_s \quad \text{pro} \quad j = 1, 2, \ldots, 5.$$

Tedy

$$F_{30}(t_1) = \frac{1}{30} n_1 \doteq 0.033,$$
$$F_{30}(t_2) = \frac{1}{30} (n_1 + n_2) \doteq 0.167,$$
$$\vdots$$
$$F_{30}(t_5) = \frac{1}{30} (n_1 + n_2 + \cdots + n_5) = 1.000.$$

Výsledky jsou ve sloupci 5 tabulky 5.3.7. Ze sloupce 6 téže tabulky je patrné, že

$$d^* \doteq 0.153.$$

Kritický obor pro test hypotézy H_0 proti hypotéze H je množina (5.3.11). Označme $W^*_{0.01}[W^*_{0.05}]$ kritický obor pro náš test na hladině významnosti 0.01 [0.05], potom

$$W^*_{0.01} = \left\{d^*;\ d^* > D(30; 0.01)\right\} = \left\{d^*;\ d^* > 0.28987\right\},$$

$$W^*_{0.05} = \left\{d^*;\ d^* > D(30; 0.05)\right\} = \left\{d^*;\ d^* > 0.24170\right\}.$$

Protože $d^* \notin W^*_{0.01}$ a $d^* \notin W^*_{0.05}$, nezamítáme hypotézu H_0 ani na jedné z požadovaných hladin významnosti.

Třída	$t_{j-1} - t_j$	n_j	$G(t_j)$	$G_{30}(t_j)$	$\lvert G_{30}(t_j) - G(t_j)\rvert$
Ω_1	1.0 – 1.4	1	0.08	0.033	0.047
Ω_2	1.4 – 1.8	4	0.32	0.167	0.153
Ω_3	1.8 – 2.2	18	0.68	0.767	0.087
Ω_4	2.2 – 2.6	4	0.92	0.900	0.020
Ω_5	2.6 – 3.0	3	1.00	1.000	0.000
Součet	×	30	×	×	×

Tabulka 5.3.7

Poznamenejme, že jsme v příkladu 5.3.4 hypotézu H_0 na hladině významnosti 0.05 zamítli. Vidíme tedy, že síla Kolmogorovova testu je v tomto případě malá.

Tab. 1a Hodnoty distribuční funkce $\Phi(x)$ normované normální náhodné veličiny

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0.00	0.5000	0.40	0.6554	0.80	0.7881	1.20	0.8849	1.60	0.9452
0.01	0.5040	0.41	0.6591	0.81	0.7910	1.21	0.8869	1.61	0.9463
0.02	0.5080	0.42	0.6628	0.82	0.7939	1.22	0.8888	1.62	0.9474
0.03	0.5120	0.43	0.6664	0.83	0.7967	1.23	0.8907	1.63	0.9484
0.04	0.5160	0.44	0.6700	0.84	0.7995	1.24	0.8925	1.64	0.9495
0.05	0.5199	0.45	0.6736	0.85	0.8023	1.25	0.8944	1.65	0.9505
0.06	0.5239	0.46	0.6772	0.86	0.8051	1.26	0.8962	1.66	0.9515
0.07	0.5279	0.47	0.6808	0.87	0.8079	1.27	0.8980	1.67	0.9525
0.08	0.5319	0.48	0.6844	0.88	0.8106	1.28	0.8997	1.68	0.9535
0.09	0.5359	0.49	0.6879	0.89	0.8133	1.29	0.9015	1.69	0.9545
0.10	0.5398	0.50	0.6915	0.90	0.8159	1.30	0.9032	1.70	0.9554
0.11	0.5438	0.51	0.6950	0.91	0.8186	1.31	0.9049	1.71	0.9564
0.12	0.5478	0.52	0.6985	0.92	0.8212	1.32	0.9066	1.72	0.9573
0.13	0.5517	0.53	0.7019	0.93	0.8238	1.33	0.9082	1.73	0.9582
0.14	0.5557	0.54	0.7054	0.94	0.8264	1.34	0.9099	1.74	0.9591
0.15	0.5596	0.55	0.7088	0.95	0.8289	1.35	0.9115	1.75	0.9599
0.16	0.5636	0.56	0.7123	0.96	0.8315	1.36	0.9131	1.76	0.9608
0.17	0.5675	0.57	0.7157	0.97	0.8340	1.37	0.9147	1.77	0.9616
0.18	0.5714	0.58	0.7190	0.98	0.8365	1.38	0.9162	1.78	0.9625
0.19	0.5753	0.59	0.7224	0.99	0.8389	1.39	0.9177	1.79	0.9633
0.20	0.5793	0.60	0.7257	1.00	0.8413	1.40	0.9192	1.80	0.9641
0.21	0.5832	0.61	0.7291	1.01	0.8438	1.41	0.9207	1.81	0.9649
0.22	0.5871	0.62	0.7324	1.02	0.8461	1.42	0.9222	1.82	0.9656
0.23	0.5910	0.63	0.7357	1.03	0.8485	1.43	0.9236	1.83	0.9664
0.24	0.5948	0.64	0.7389	1.04	0.8508	1.44	0.9251	1.84	0.9671
0.25	0.5987	0.65	0.7422	1.05	0.8531	1.45	0.9265	1.85	0.9678
0.26	0.6026	0.66	0.7454	1.06	0.8554	1.46	0.9279	1.86	0.9686
0.27	0.6064	0.67	0.7486	1.07	0.8577	1.47	0.9292	1.87	0.9693
0.28	0.6103	0.68	0.7517	1.08	0.8599	1.48	0.9306	1.88	0.9699
0.29	0.6141	0.69	0.7549	1.09	0.8621	1.49	0.9319	1.89	0.9706
0.30	0.6179	0.70	0.7580	1.10	0.8643	1.50	0.9332	1.90	0.9712
0.31	0.6217	0.71	0.7611	1.11	0.8665	1.51	0.9345	1.91	0.9719
0.32	0.6255	0.72	0.7642	1.12	0.8686	1.52	0.9357	1.92	0.9726
0.33	0.6293	0.73	0.7673	1.13	0.8708	1.53	0.9370	1.93	0.9732
0.34	0.6331	0.74	0.7704	1.14	0.8729	1.54	0.9382	1.94	0.9738
0.35	0.6368	0.75	0.7734	1.15	0.8749	1.55	0.9394	1.95	0.9744
0.36	0.6406	0.76	0.7764	1.16	0.8770	1.56	0.9406	1.96	0.9750
0.37	0.6443	0.77	0.7794	1.17	0.8790	1.57	0.9418	1.97	0.9756
0.38	0.6480	0.78	0.7823	1.18	0.8810	1.58	0.9429	1.98	0.9761
0.39	0.6517	0.79	0.7852	1.19	0.8830	1.59	0.9441	1.99	0.9767

Tab. 1b Hodnoty distribuční funkce $\Phi(x)$ normované normální náhodné veličiny

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
2.00	0.9773	2.40	0.9918	2.80	0.9974	3.20	0.9993	3.60	0.9998
2.01	0.9778	2.41	0.9920	2.81	0.9975	3.21	0.9993	3.61	0.9998
2.02	0.9783	2.42	0.9922	2.82	0.9976	3.22	0.9994	3.62	0.9999
2.03	0.9788	2.43	0.9925	2.83	0.9977	3.23	0.9994	3.63	0.9999
2.04	0.9793	2.44	0.9927	2.84	0.9977	3.24	0.9994	3.64	0.9999
2.05	0.9798	2.45	0.9929	2.85	0.9978	3.25	0.9994	3.65	0.9999
2.06	0.9803	2.46	0.9931	2.86	0.9979	3.26	0.9994	3.66	0.9999
2.07	0.9808	2.47	0.9932	2.87	0.9979	3.27	0.9995	3.67	0.9999
2.08	0.9812	2.48	0.9934	2.88	0.9980	3.28	0.9995	3.68	0.9999
2.09	0.9817	2.49	0.9936	2.89	0.9981	3.29	0.9995	3.69	0.9999
2.10	0.9821	2.50	0.9938	2.90	0.9981	3.30	0.9995	3.70	0.9999
2.11	0.9826	2.51	0.9940	2.91	0.9982	3.31	0.9995	3.71	0.9999
2.12	0.9830	2.52	0.9941	2.92	0.9983	3.32	0.9996	3.72	0.9999
2.13	0.9834	2.53	0.9943	2.93	0.9983	3.33	0.9996	3.73	0.9999
2.14	0.9838	2.54	0.9945	2.94	0.9984	3.34	0.9996	3.74	0.9999
2.15	0.9842	2.55	0.9946	2.95	0.9984	3.35	0.9996	3.75	0.9999
2.16	0.9846	2.56	0.9948	2.96	0.9985	3.36	0.9996	3.76	0.9999
2.17	0.9850	2.57	0.9949	2.97	0.9985	3.37	0.9996	3.77	0.9999
2.18	0.9854	2.58	0.9951	2.98	0.9986	3.38	0.9996	3.78	0.9999
2.19	0.9857	2.59	0.9952	2.99	0.9986	3.39	0.9997	3.79	0.9999
2.20	0.9861	2.60	0.9953	3.00	0.9987	3.40	0.9997	3.80	0.9999
2.21	0.9864	2.61	0.9955	3.01	0.9987	3.41	0.9997	3.81	0.9999
2.22	0.9868	2.62	0.9956	3.02	0.9987	3.42	0.9997	3.82	0.9999
2.23	0.9871	2.63	0.9957	3.03	0.9988	3.43	0.9997	3.83	0.9999
2.24	0.9875	2.64	0.9959	3.04	0.9988	3.44	0.9997	3.84	0.9999
2.25	0.9878	2.65	0.9960	3.05	0.9989	3.45	0.9997	3.85	0.9999
2.26	0.9881	2.66	0.9961	3.06	0.9989	3.46	0.9997	3.86	0.9999
2.27	0.9884	2.67	0.9962	3.07	0.9989	3.47	0.9997	3.87	0.9999
2.28	0.9887	2.68	0.9963	3.08	0.9990	3.48	0.9997	3.88	0.9999
2.29	0.9890	2.69	0.9964	3.09	0.9990	3.49	0.9998	3.89	1.0000
2.30	0.9893	2.70	0.9965	3.10	0.9990	3.50	0.9998	3.90	1.0000
2.31	0.9896	2.71	0.9966	3.11	0.9991	3.51	0.9998	3.91	1.0000
2.32	0.9898	2.72	0.9967	3.12	0.9991	3.52	0.9998	3.92	1.0000
2.33	0.9901	2.73	0.9968	3.13	0.9991	3.53	0.9998	3.93	1.0000
2.34	0.9904	2.74	0.9969	3.14	0.9992	3.54	0.9998	3.94	1.0000
2.35	0.9906	2.75	0.9970	3.15	0.9992	3.55	0.9998	3.95	1.0000
2.36	0.9909	2.76	0.9971	3.16	0.9992	3.56	0.9998	3.96	1.0000
2.37	0.9911	2.77	0.9972	3.17	0.9992	3.57	0.9998	3.97	1.0000
2.38	0.9913	2.78	0.9973	3.18	0.9993	3.58	0.9998	3.98	1.0000
2.39	0.9916	2.79	0.9974	3.19	0.9993	3.59	0.9998	3.99	1.0000

PDF-XChange
Click to buy NOW!
www.docu-track.com

Tab. 2 Kvantily $u(\alpha)$ normované normální náhodné veličiny

α	0.900	0.950	0.975	0.990	0.995
$u(\alpha)$	1.282	1.645	1.960	2.326	2.576

Tab. 3a Kvantily $\chi^2(n;\alpha)$ chí-kvadrát rozdělení

$n \backslash \alpha$	0.005	0.010	0.025	0.050	0.100
1	0.000	0.000	0.001	0.004	0.016
2	0.010	0.020	0.051	0.103	0.211
3	0.072	0.115	0.216	0.352	0.584
4	0.207	0.292	0.484	0.711	1.064
5	0.412	0.554	0.831	1.146	1.610
6	0.676	0.872	1.237	1.635	2.204
7	0.989	1.239	1.690	2.167	2.833
8	1.344	1.647	2.180	2.733	3.490
9	1.735	2.088	2.700	3.325	4.168
10	2.156	2.558	3.249	3.940	4.865
11	2.603	3.054	3.816	4.575	5.578
12	3.074	3.571	4.404	5.226	6.304
13	3.565	4.107	5.009	5.892	7.042
14	4.075	4.660	5.629	6.571	7.790
15	4.601	5.229	6.262	7.261	8.547
16	5.142	5.812	6.608	7.962	9.312
17	5.697	6.408	7.564	8.672	10.09
18	6.265	7.015	8.231	9.391	10.87
19	6.844	7.633	8.907	10.12	11.65
20	7.434	8.260	9.591	10.85	12.44
21	8.034	8.897	10.28	11.59	13.34
22	8.643	9.543	10.98	12.34	14.04
23	9.260	10.20	11.69	13.09	14.85
24	9.887	10.86	12.40	13.85	15.66
25	10.52	11.52	13.12	14.61	16.47
26	11.16	12.20	13.84	15.38	17.29
27	11.81	12.88	14.57	16.15	18.11
28	12.46	13.57	15.31	16.93	18.94
29	13.12	14.26	16.05	17.71	19.77
30	13.79	14.95	16.79	18.49	20.60

PDF-Xchange
Click to buy NOW!
www.docu-track.com

Tab. 3b Kvantily $\chi^2(n;\alpha)$ chí-kvadrát rozdělení

$n\backslash\alpha$	0.900	0.950	0.975	0.990	0.995
1	2.706	3.842	5.024	6.635	7.879
2	4.605	5.992	7.378	9.210	10.60
3	6.251	7.815	9.348	11.35	12.84
4	7.779	9.488	11.14	13.28	14.86
5	9.236	11.07	12.83	15.09	16.75
6	10.65	12.59	14.45	16.81	18.55
7	12.02	14.07	16.01	18.48	20.28
8	13.36	15.51	17.54	20.09	21.96
9	14.68	16.92	19.02	21.67	23.59
10	15.99	18.31	20.48	23.21	25.19
11	17.28	19.68	21.92	24.73	26.76
12	18.55	21.03	23.34	26.22	28.30
13	19.81	22.36	24.74	27.69	29.82
14	21.06	23.69	26.12	29.14	31.32
15	22.31	25.00	27.49	30.58	32.80
16	23.54	26.30	28.85	32.00	34.27
17	24.77	27.59	30.19	33.41	35.72
18	25.99	28.87	31.53	34.81	37.16
19	27.20	30.14	32.85	36.19	38.58
20	28.41	31.41	34.17	37.57	40.00
21	29.62	32.67	35.48	38.93	41.40
22	30.81	33.92	36.78	40.29	42.80
23	32.01	35.17	38.08	41.64	44.18
24	33.20	36.42	39.36	42.98	45.56
25	34.38	37.65	40.65	44.31	46.93
26	35.56	38.89	41.92	45.64	48.29
27	36.74	40.11	43.20	46.96	49.65
28	37.92	41.34	44.46	48.28	50.99
29	39.09	42.56	45.72	49.59	52.34
30	40.26	43.77	46.98	50.89	53.67

Tab. 4 Kvantily $t(n;\alpha)$ t-rozdělení

$n\backslash\alpha$	0.900	0.950	0.975	0.990	0.995
1	3.078	6.314	12.71	31.82	63.66
2	1.886	2.920	4.303	6.964	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	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	3.500
8	1.397	1.860	2.306	2.897	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.813	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	2.625	2.977
15	1.341	1.753	2.131	2.603	2.947
16	1.337	1.746	2.120	2.584	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	2.540	2.861
20	1.325	1.725	2.086	2.520	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	1.320	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	1.705	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
∞	1.282	1.645	1.960	2.326	2.576

Tab. 5 **Kritické hodnoty $D(n; \alpha)$ pro Kolmogorovův test dobré shody**

n	$\alpha = 0.05$	$\alpha = 0.01$	n	$\alpha = 0.05$	$\alpha = 0.01$	n	$\alpha = 0.05$	$\alpha = 0.01$
1	0.97500	0.99500	31	0.23788	0.28530	61	0.17091	0.20506
2	0.84189	0.92929	32	0.23424	0.28094	62	0.16959	0.20343
3	0.70760	0.82900	33	0.23076	0.27677	63	0.16823	0.20184
4	0.62394	0.73424	34	0.22743	0.27279	64	0.16993	0.20029
5	0.56328	0.66853	35	0.22425	0.26897	65	0.16567	0.19877
6	0.51926	0.61661	36	0.22119	0.26532	66	0.16443	0.19729
7	0.48342	0.51581	37	0.21826	0.26180	67	0.16322	0.19854
8	0.45427	0.54179	38	0.21544	0.25843	68	0.16204	0.19442
9	0.43001	0.51332	39	0.21273	0.25518	69	0.16088	0.19303
10	0.40925	0.48893	40	0.21012	0.25205	70	0.15975	0.19167
11	0.39122	0.46770	41	0.20760	0.24904	71	0.15864	0.19034
12	0.37543	0.44905	42	0.20517	0.24613	72	0.15755	0.18903
13	0.36143	0.43247	43	0.20283	0.24332	73	0.15649	0.18776
14	0.34890	0.41762	44	0.20056	0.24060	74	0.15544	0.18650
15	0.33760	0.40420	45	0.19837	0.23798	75	0.15442	0.18528
16	0.32733	0.39201	46	0.19625	0.23544	76	0.15343	0.18408
17	0.31796	0.38086	47	0.19420	0.23298	77	0.15244	0.18290
18	0.30936	0.37062	48	0.19221	0.23059	78	0.15147	0.18174
19	0.30143	0.36117	49	0.19028	0.22828	79	0.15052	0.18060
20	0.29408	0.35241	50	0.18841	0.22604	80	0.14960	0.17949
21	0.28724	0.34427	51	0.18659	0.22386	81	0.14868	0.17840
22	0.28087	0.33666	52	0.18482	0.22174	82	0.14779	0.17732
23	0.27490	0.32954	53	0.18311	0.21968	83	0.14691	0.17627
24	0.26931	0.32286	54	0.18144	0.21768	84	0.14605	0.17523
25	0.26404	0.31687	55	0.17981	0.21574	85	0.14520	0.17421
26	0.25907	0.31064	56	0.17823	0.21384	86	0.14437	0.17321
27	0.25438	0.30502	57	0.17669	0.21199	87	0.14355	0.17223
28	0.24993	0.29971	58	0.17519	0.21019	90	0.14117	0.16938
29	0.24571	0.29466	59	0.17373	0.20844	95	0.13746	0.16493
30	0.24170	0.28987	60	0.17231	0.20673	100	0.13403	0.16081