- **Rozdělení náhodné veličiny**
- **Charakteristiky náhodné veličiny**
- **Významná rozdělení náhodné veličiny**

4. Náhodná veličina

Cíl kapitoly

Pro popis rozsáhlejších statistických souborů je významný pojem náhodného rozdělení. Jedná se o pravidlo chování tzv. náhodné veličiny, která velmi často dobře interpretuje statistické soubory. Následující kapitola Vás proto seznámí s pojmem a definicí náhodné veličiny a některými pravidly podle nichž se chová. Analogii mezi pravděpodobností ve smyslu statistické definice a popisem statistických souborů pomocí četností využijeme i v případě náhodné veličiny. K základním popisným veličinám, jako jsou aritmetický průměr a rozptyl, přibudou jejich protějšky v podobě střední hodnoty a rozptylu náhodné veličiny. Poslední část kapitoly Vás seznámí s některými standardními rozděleními náhodné kapitoly.

Časová zátěž

4 hodiny (1. týden v listopadu)

Náhodná veličina

Výsledkem náhodných pokusů jsou obvykle reálná čísla reprezentující daný náhodný jev (například počet bodů na hrací kostce, počet poruch daného stroje, výsledek zkoušky na vysoké škole). Takový výsledek náhodného pokusu, který je možno vyjádřit reálným číslem označujeme pojmem **náhodná veličina**. Náhodná veličina je tedy veličina, jejíž hodnota je jednoznačně určena výsledkem náhodného pokusu.

Náhodnou veličinu obvykle označujeme velkými písmeny z konce abecedy – tedy například X, Y, Z. Konkrétní realizace této veličiny poté označujeme malými písmeny x, y, z. Zápis je poté například následující: $X = x$.

Například průměrná mzda v českém průmyslovém podniku obecně bude náhodnou veličinou, kterou označíme například X. Konkrétní výše průměrné mzdy v konkrétním podniku ABC již bude konkrétní realizace této náhodné veličiny, například 15000 Kč.

Danou skutečnost tedy můžeme zapsat jako $X = 15000$ Kč.

Z hlediska pravidel počítání a analýz je vhodné náhodné veličiny rozdělit do dvou skupin na náhodné veličiny

- **diskrétní** (nespojité)
  Podobně jako u statistického znaku jsou diskrétní náhodné veličiny takové, které mohou nabývat jen konečný počet určitých hodnot.
- **spojité**
  Spojitá náhodná veličina může nabývat všech hodnot z konečného či nekonečného intervalu.

4.1 Rozdělení náhodné veličiny

Přestože je náhodná veličina jako výsledek náhodného pokusu spojená s nejistotou, není zcela nezkoumatelná. Lze ji charakterizovat například tak, že informaci o jednotlivých realizacích náhodné veličiny doplníme informací o pravděpodobnosti, že tato konkrétní realizace nastane. Náhodnou veličinu poté považujeme za danou, známe-li všechny její možné hodnoty a pravděpodobnosti výskytu každé z nich.

Zákon rozdělení náhodné veličiny

Pravidlo (tabulka, předpis, graf), které každé hodnotě (nebo množině hodnot z daného intervalu) přiřazuje pravděpodobnost, že náhodná veličina nabude této hodnoty (nebo hodnoty z určitého intervalu) je nazýváno **zákonem rozdělení náhodné veličiny**.

Zákon rozdělení náhodné veličiny je možno zadat několika způsoby:

- tabulkou,
- grafem,
- distribuční funkcí,
- pravděpodobnostní funkcí.

4.1.1 Zadání tabulkou

Nejjednodušší formou zadání zákona rozdělení náhodné veličiny je pomocí **tabulky rozdělení náhodné veličiny**. Získáme ji tak, že do prvního řádku umístíme jednotlivé realizace (možné hodnoty) náhodné veličiny a do druhého řádku jim přiřadíme příslušnou pravděpodobnost. Musí platit, že sečteme-li pravděpodobnost pro všechny možné hodnoty, musí být rovna jedné. Tabulka má tedy následující podobu:

x_i	x_1	x_2	x_3	x_4	x_5	x_6	$\sum$
$P(x_i)$	$P(x_1)$	$P(x_2)$	$P(x_3)$	$P(x_4)$	$P(x_5)$	$P(x_6)$	1

4.1.2 Zadání statistickým grafem

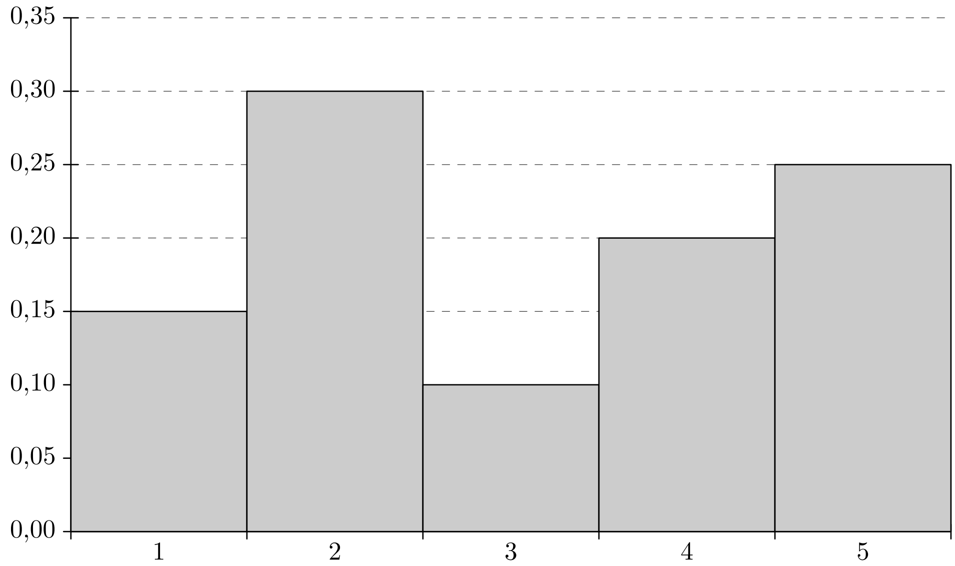
Jinou názornou možností, jak zachytit rozdělení náhodné veličiny, je využití **statistického grafu**. Obvykle jej konstruujeme ve formě polygonu nebo histogramu, kdy na horizontální osu nanášíme jednotlivé realizace náhodné veličiny a na vertikální osu příslušné pravděpodobnosti jejich nastoupení. Graf má například následující podobu:

Všimněte si, že pro studium náhodné veličiny existují zajímavé analogie se studiem statistických souborů. Analogií náhodné veličiny je v popisné statistice do jisté míry statistický znak. Analogií příslušné pravděpodobnosti je relativní četnost jednotlivých variant tohoto znaku. Proto je také konstrukce a podoba tabulek a grafů rozdělení četností velmi podobné tabulkám a grafům rozdělení náhodné veličiny.

Důvody těchto analogií si lze vysvětlit pomocí statistické definice pravděpodobnosti, která ztotožňuje pravděpodobnost s relativní četností.

4.1.3 Distribuční funkce

Přestože výše uvedené příklady zápisu jsou poměrně názorné, je základní formou zápisu zákona rozdělení náhodné veličiny **distribuční funkce**. Dis-



Obrázek 4.1: Rozdělení náhodné veličiny zobrazené pomocí histogramu.

tribuční funkce udává pravděpodobnost, že náhodná veličina nabude hodnoty menší nebo rovné než je pevně zvolená realizace náhodné veličiny x. Značíme ji $F(x)$ a definujeme jako

$$F(x) = P(X \leq x).$$

Jelikož se jedná o pravděpodobnost, pohybují se hodnoty distribuční funkce v intervalu $\langle 0; 1\rangle$.

4.1.4 Pravděpodobnostní funkce

Jiným funkčním předpisem, který je možno užít pro zachycení zákona rozdělení náhodné veličiny je pravděpodobnostní funkce. Pravděpodobnostní funkce každé realizaci náhodné veličiny x přiřazuje příslušnou pravděpodobnost.

$$P(x) = P(X = x)$$

Pro pravděpodobnostní funkci diskrétní veličiny platí, že

$$\sum P(x) = 1.$$

Pro spojitou náhodnou veličinu nahrazujeme pravděpodobnostní funkci hustotou pravděpodobnosti, pro niž platí podobné vztahy (místo výrazu $\sum$ však musíme použít integraci).

Ve smyslu výše uvedené analogie odpovídá grafu distribuční funkce $F(x)$ v popisné statistice graf kumulativních relativních četností. Pravděpodobnostní funkci $P(x)$ odpovídají v popisné statistice relativní četnosti.

Všimněte si také, že z definičního vztahu pro distribuční a pravděpodobnostní funkci vyplývá, že pravděpodobnostní funkci lze vypočítat pomocí distribuční a obráceně.

Příklad 4.1

Student si u zkoušky může vybrat jednu z pěti otázek. Ke zkoušce přijde pět studentů, přičemž žádná z otázek se nesmí opakovat. Poslední student tedy bude odpovídat na otázku, kterou si nevybere žádný z jeho kolegů (tedy tu „která na něj zbude“). Jelikož je pořadí studentů předem stanoveno, může poslední student odhadovat, na kterou z otázek bude odpovídat (která bude ta „nejméně žádaná“). Podle odhadu preferencí kolegů, kteří jdou před ním, odhaduje pravděpodobnost, že na něj vyjde určitá otázka, následovně:

1. otázka 15%
2. otázka 30%
3. otázka 10%
4. otázka 20%
5. otázka 25%

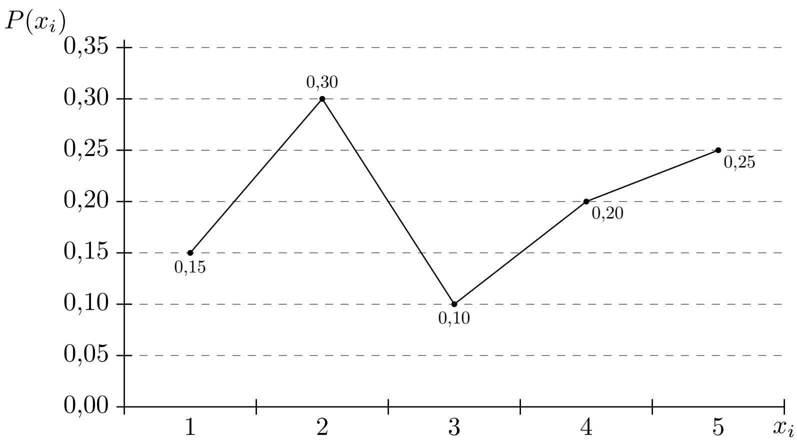
Považujte volbu otázky za náhodný pokus a číslo otázky za náhodnou veličinu. Zobrazte tuto náhodnou veličinu pomocí zákona rozdělení, a to tabulkou, grafem, distribuční funkcí a pravděpodobnostní funkcí.

Řešení

a) tabulkou

Číslo otázky	x_i	1	2	3	4	5
Pravděpodobnost	$P(x_i)$	0,15	0,30	0,10	0,20	0,25

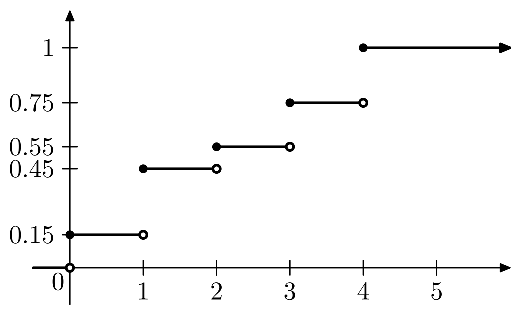
b) grafem (pravděpodobnostní polygon)



c) pomocí distribuční funkce

Distribuční funkce $F(x)$ bude nabývat pěti hodnot:

pro $x = 1$ $F(x) = 0{,}15$
pro $x = 2$ $F(x) = 0{,}45$
pro $x = 3$ $F(x) = 0{,}55$
pro $x = 4$ $F(x) = 0{,}75$
pro $x = 5$ $F(x) = 1{,}00$



d) pomocí pravděpodobnostní funkce
Stejně jako distribuční bude i pravděpodobnostní $P(x)$ nabývat pěti hodnot:
pro $x = 1$ $P(x) = 0{,}15$
pro $x = 2$ $P(x) = 0{,}30$
pro $x = 3$ $P(x) = 0{,}10$
pro $x = 4$ $P(x) = 0{,}20$
pro $x = 5$ $P(x) = 0{,}25$

4.2 Charakteristiky náhodné veličiny

Vlastnosti náhodné veličiny

Některá z výše uvedených forem zápisu plně postačuje k popisu náhodné veličiny. Užívá se zejména distribuční funkce, pomocí níž jsme schopni určit, jakých hodnot může uvažovaná veličina nabývat a jaké jsou pravděpodobnosti, které těmto hodnotám odpovídají. V praktických úlohách však obvykle tato poměrně obecná (a pro více možných realizací náhodné veličiny i nepřehledná) informace nestačí. Obvykle je proto nutno použít nějakého přehlednějšího vyjádření. K těmto účelům používáme popisné veličiny, souhrnně označované jako charakteristiky náhodné veličiny. Mezi nejčastěji používané charakteristiky patří **střední hodnota**, která popisuje úroveň (polohu) náhodné veličiny a **rozptyl**, popisující variabilitu náhodné veličiny.

Jak je již zřejmé z označení těchto charakteristik, jsou opět analogií nejvýznamnějších popisně statistických veličin – průměru a rozptylu. K charakteristikám náhodné veličiny patří také (opět stejně jako u popisné statistiky) kvantil, jehož vlastnosti i použití jsou obdobné jako v popisné statistice (tou jsme se zabývali v kapitole 1).

4.2.1 Střední hodnota rozdělení náhodné veličiny

Střední hodnota je definována jako vážený průměr možných hodnot náhodné veličiny. Úlohu vah zde přebírají hodnoty pravděpodobnosti příslušející každé z možných hodnot náhodné veličiny.

Pro diskrétní náhodné veličiny je střední hodnota náhodné veličiny X definována jako

$$E(X) = \sum_{i=1}^{n} x_i \cdot p(x_i).$$

V případě, že počítáme střední hodnotu spojité náhodné veličiny, je nutno nahradit součet pomocí sumy integrací a místo pravděpodobnosti počítat s hustotou pravděpodobnosti.

$$E(x) = \int_{-\infty}^{\infty} x \cdot f(x)\,dx.$$

Vlastnosti střední hodnoty

a) $E(c) = c$
Střední hodnota konstanty je rovna konstantě.

b) $E(cX) = cE(X)$
Střední hodnota náhodné veličiny, kterou je možno vyjádřit jako součin konstanty a jiné náhodné veličiny, je rovna konstantě násobené střední hodnotou této veličiny .

c) $E(X_1 + X_2 + \cdots + X_k) = E(X_1) + E(X_1) + \cdots + E(X_k)$
Střední hodnota součtu náhodných veličin je rovna součtu jejich středních hodnot.

d) $E(X_1 \cdot X_2 \cdot \ldots \cdot X_k) = E(X_1) \cdot E(X_2) \cdot \cdots \cdot E(X_k)$
Střední hodnota součinu náhodných veličin je rovna součinu jejich středních hodnot, pokud jsou tyto náhodné veličiny nezávislé.

4.2.2 Rozptyl rozdělení náhodné veličiny

Rozptyl náhodné veličiny měří její variabilitu. Je definován jako střední hodnota druhých mocnin odchylek hodnot náhodné veličiny od její střední hodnoty. Rozptyl

Pro diskrétní náhodné veličiny je rozptyl náhodné veličiny X definována jako

$$D(X) = E[X - E(X)]^2 = E(X^2) - [E(X)]^2.$$

Výpočtový tvar rozptylu má pak následující podobu:

$$\begin{aligned} D(X) &= \sum_{i=1}^{n} [x_i - E(X)]^2 \cdot p(x_i) = \sum_{i=1}^{n} [x_i^2 - 2 \cdot x_i \cdot E(X) + E^2(X)] \cdot p(x_i) = \\ &= \sum_{i=1}^{n} x_i^2 \cdot p(x_i) - 2 \cdot E(X) \cdot \sum_{i=1}^{n} x_i \cdot p(x_i) + E^2(X) \sum_{i=1}^{n} p(x_i) = \\ &= \boldsymbol{\sum_{i=1}^{n} x_i^2 \cdot p(x_i) - E^2(X)}. \end{aligned}$$

Výraz $2 \cdot E(X) \cdot \sum_{i=1}^{n} x_i \cdot p(x_i) = 2 \cdot E(X) \cdot E(X) = 2 \cdot E^2(X)$, neboť jak bylo uvedeno dříve, výraz $\sum_{i=1}^{n} x_i \cdot p(x_i) = E(X)$ a výraz $\sum_{i=1}^{n} p(x_i) = 1$. Potom $-2 \cdot E^2(X) + E^2(X) = -E^{(}X)$.

V případě, že počítáme rozptyl spojité náhodné veličiny, je nutno nahradit součet pomocí sumy integrací a místo pravděpodobnosti počítat s hustotou pravděpodobnosti.

$$D(X) = \int_{-\infty}^{\infty} [x - f(x)]^2 \, dx.$$

Vlastnosti rozptylu

a) $D(c) = 0$
Rozptyl (variabilita) konstanty je roven nule.

b) $D(cX) = c^2 D(X)$
Rozptyl náhodné veličiny, kterou je možno vyjádřit jako součin konstanty a jiné náhodné veličiny, je roven druhé mocnině konstanty násobené rozptylem této veličiny.

c) $D(X_1 + X_2 + \cdots + X_k) = D(X_1) + D(X_1) + \cdots + D(X_k)$
Rozptyl součtu náhodných veličin je roven součtu rozptylů těchto náhodných veličin, pokud jsou náhodné veličiny nezávislé.

Příklad 4.2

Vypočítejte rozptyl a střední hodnotu rozdělení náhodné veličiny „vytažení otázky u zkoušky“, která má rozdělení definované tabulkou uvedenou v příkladu 4.1.

Řešení

Máme spočítat rozptyl a střední hodnotu následujícího rozdělení náhodné veličiny:

Číslo otázky	x_i	1	2	3	4	5
Pravděpodobnost	$P(x_i)$	0,15	0,30	0,10	0,20	0,25

a) **Střední hodnotu** vypočítáme podle výše uvedeného definičního vztahu

$$E(X) = \sum_{i=1}^{n} x_i \cdot p(x_i) = 1 \cdot 0{,}15 + 2 \cdot 0{,}30 + 3 \cdot 0{,}10 + 4 \cdot 0{,}20 + 5 \cdot 0{,}25 =$$
$$= 0{,}15 + 0{,}60 + 0{,}30 + 0{,}80 + 1{,}25 = 3{,}10.$$

Nejpravděpodobněji vytaženou otázkou tedy bude otázka číslo tři.

b) **Rozptyl** vypočítáme podle uvedeného výpočtového vztahu. Nejprve však musíme do tabulky doplnit další řádek, ve kterém budou druhé mocniny jednotlivých hodnot náhodné veličiny.

Číslo otázky	x_i	1	2	3	4	5
Pravděpodobnost	$P(x_i)$	0,15	0,30	0,10	0,20	0,25
	x_i^2	1	4	9	16	25

$$D(X) = \sum_{i=1}^{n} x_i^2 \cdot p(x_i) - \left[\sum_{i=1}^{n} x_i \cdot p(x_i)\right]^2 =$$
$$= (1 \cdot 0{,}15 + 4 \cdot 0{,}30 + 9 \cdot 0{,}10 + 16 \cdot 0{,}20 + 25 \cdot 0{,}25) - 3{,}10^2 =$$
$$= (0{,}15 + 1{,}20 + 0{,}90 + 3{,}20 + 6{,}25) - (9{,}61) = 2{,}09.$$

Nejpravděpodobnějším číslem vytažené otázky bude 3. Výsledek však není příliš významný, neboť hodnota rozptylu hovoří o velké variabilitě možných výsledků (vytažených otázek).

Pro výpočet charakteristik spojité náhodné veličiny je nutno vycházet z metod integrálního počtu. Úlohy jsou však ve své podstatě naprosto identické jako v případě nespojité veličiny. Stejně tak je možno určit další charakteristiky náhodné veličiny, jako jsou např. kvantil, či kovariance. Jejich definice i výpočtové vztahy naleznete například v učebnici Seger, Hindls, Hronová: *Statistika v hospodářství* na stranách 93–98.

4.2.3 Směrodatná odchylka

Stejně jako tomu u veličin popisné statistiky i u charakteristik náhodné veličiny je namísto rozptylu obvykle užívána jeho druhá odmocnina, která je označována jako **směrodatná odchylka**.

Je tedy možno ji definovat jako

$$S(X) = \sqrt{D(X)} = \sqrt{\sum_{i=1}^{n} [x_i - E(X)]^2 \cdot p(x_i)} = \sqrt{\sum_{i=1}^{n} x_i^2 \cdot p(x_i) - E^2(X)}.$$

4.3 Významná rozdělení náhodné veličiny

Jak jsme již uvedli, k popisu náhodné veličiny postačuje, abychom znali její rozdělení (ať už zadané tabulkou, grafem, či některou z funkcí). Ve statistice je obvyklé používat pro nejrůznější náhodné veličiny několik standardních modelů, jejichž charakteristiky jsou obecně známé. Popis náhodné veličiny se poté v první fázi zaměřuje na zjištění, zda lze tuto veličinu popsat některým z těchto standardních modelů. V tom případě již máme její popis prakticky hotov, neboť charakteristiky tohoto modelu jsou součástí všech klasických statistických učebnic.

Uvedené modely je poté možno využít pro nejrůznější pokročilejší statistické úlohy. Nejtypičtějšími z nich jsou úlohy konstruující bodové a intervalové odhady a intervaly spolehlivosti (viz kapitola 5 tohoto dílu učebnice), úlohy sloužící k testování statistických hypotéz (kapitoly 6 a 7 tohoto dílu), či aplikace těchto úloh pro regresní a korelační počet (seznámíme se s nimi ve druhém díle učebnice).

V následujícím přehledu se zmíníme pouze o těch nejvýznamnějších standardních rozděleních náhodné veličiny. Patří sem dvě rozdělení nespojité veličiny (alternativní a binomické) a především čtyři rozdělení spojité náhodné veličiny (normální, Studentovo, Fisher-Snedecorovo a chí-kvadrát).

Pro každé z dále uvedených rozdělení jsou prezentovány pouze základní vlastnosti a charakteristiky. Podrobnější popis standardních modelů rozdělení náhodné veličiny je možno nalézt například v učebnici Seger, Hindls, Hronová: *Statistika v hospodářství* na stranách 98–119.

4.3.1 Diskrétní rozdělení

1. Alternativní rozdělení $A(p)$

Alternativní rozdělení

Tento typ rozdělení diskrétní náhodné veličiny se používá pro veličinu, která může nabývat pouze dvou hodnot $x = 0$ a $x = 1$. Parametr p tohoto rozdělení udává jaká je pravděpodobnost, že daný jev nastane ($x = 1$). Alternativního rozdělení lze použít všude tam, kde mohou nastat pouze dva výsledky, přičemž známe pravděpodobnost těchto výsledků. Náhodná veličina s rozdělením $A(p)$ pak vyjadřuje, kolikrát jev A v pokusu nastane.

Pravděpodobnostní funkce tohoto rozdělení má následující podobu:

$$P(x) = \begin{cases} p^x(1-p)^{1-x} & \text{je-li } x = 0, \text{ nebo } x = 1, \\ P(x) = 0 & \text{pro ostatní } x. \end{cases}$$

Střední hodnota rozdělení je $E(X) = p$ a rozptyl $D(X) = p \cdot q = p \cdot (1-p)$.

Binomické rozdělení

2. Binomické rozdělení $\boldsymbol{Bi(n,p)}$

Binomické rozdělení je rozšířením alternativního rozdělení pro n pokusů. Lze jej využít všude tam, kde provádíme několik nezávislých pokusů (pozorování), při nichž mohou nastat dva výsledky. Pravděpodobnost, že v sérii n pokusů jev A nastane právě x-krát je definována funkcí:

$$P(X = x) = P(x) = \binom{n}{x} p^x q^{n-x},$$

kde x může nabývat hodnot od 1 do n a $q = 1 - p$.

Charakteristiky rozdělení $Bi(n,p)$:

střední hodnota	$E(x) = n \cdot p$
rozptyl	$D(x) = n \cdot p \cdot q$

Binomické rozdělení přechází pro rostoucí počet pokusů ($n > 30$) v rozdělení Poissonovo. Poissonovo rozdělení je charakterizováno jedním parametrem λ, který je roven součinu $n \cdot p$.

Poissonova rozdělení je užíváno zejména v průmyslových aplikacích v souvislosti s tzv. poissonovským proudem jevů, kdy sledujeme pravděpodobnost výskytu nějakého jevu v určitém časovém intervalu (například poruchy stroje).

Poissonovo rozdělení

Poissonovo rozdělení lze zapsat jako

$$P(X = x) = P(x) = \begin{cases} \frac{\lambda^x}{x!} e^{-\lambda} & \text{pro } x = 0, 1, 2, 3, \ldots, \\ P(x) = 0 & \text{jinak.} \end{cases}$$

V praxi aproximujeme binomické rozdělení Poissonovým rozdělením v případech kdy $p \leq 0{,}1$ a $n > 30$.

Pro základní číselné charakteristiky Poissonova rozdělení platí:

$$E(X) = D(X) = \lambda.$$

Příklad 4.3

Na telefonní ústřednu je napojeno 300 účastníků. Každý z nich bude volat ústřednu s pravděpodobností 0,01. Jaká je pravděpodobnost, že během hodiny zavolají ústřednu 4 účastníci.

Řešení

Náhodná proměnná X (počet účastníků volajících během hodiny ústřednu) má:

a) *binomické rozdělení* s parametry: $n = 300$, $p = 0{,}01$. To znamená, že pravděpodobnost

$$P(X = 4) = \binom{300}{4} \cdot 0{,}01^4 \cdot 0{,}99^2 96 =$$
$$= 330791117 \cdot 0{,}00000001 \cdot 0{,}0510525 = 0{,}16887739.$$

b) *Poissonovo rozdělení* s parametry $\lambda = 300 \cdot 0{,}01 = 3$, $x = 4$:

$$P(X = 4) = \frac{e^{-3} \cdot 3^4}{4!} = \frac{0{,}04978708 \cdot 81}{24} = 0{,}16803136.$$

4.3.2 Spojitá rozdělení

3. Normální rozdělení $N(\mu, \sigma^2)$

Normální rozdělení

Jedná se o nejklasičtější a nejčastější rozdělení v přírodě. Slouží jako řídící nástroj při výzkumech v oblasti přírodních a společenských věd, v medicíně, zemědělství a stavebnictví. Je nepostradatelným nástrojem pro analýzu a interpretaci základních informací získaných při pozorování a různých experimentech.

Náhodná veličina X má normální rozdělení tehdy, je-li možno náhodný výsledek interpretovat jako součet nekonečně mnoha malých nezávislých vlivů.

Jelikož se jedná o rozdělení spojité náhodné veličiny, je definováno pomocí hustoty pravděpodobnosti $f(x)$. Její tvar je následující

$$f(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

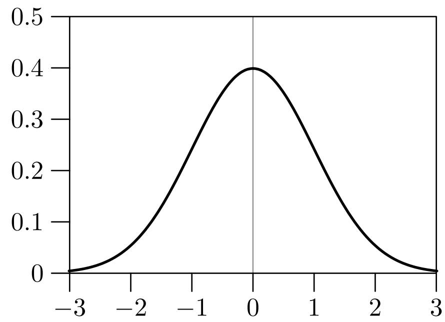
Charakteristiky rozdělení $N(\mu, \sigma^2)$:

střední hodnota	$E(x) = \mu$
rozptyl	$D(x) = \sigma^2$

Normální rozdělení má tvar zvonovité křivky, která nabývá maxima v bodě $x = \mu$. Pro jednodušší výpočet i interpretaci je často obecný tvar normálního rozdělení převáděn na tzv. normované normální rozdělení $N(0, 1)$, jehož graf má maximum ve své střední hodnotě (bodě 0) a je kolem této hodnoty symetrický. Příklad normovaného normálního rozdělení je na následujícím obrázku:

Vlastnosti normálního rozdělení

- Symetrické okolo průměru
- Asymptoticky se přibližuje k osám
- Má zvonovitý tvar
- Plocha pod křivkou = 1
- více než 99% plochy pod křivkou se rozprostírá ve vzdálenosti $\pm 3\sigma$ od průměru



Obrázek 4.2: Normované normální rozdělení náhodné veličiny.

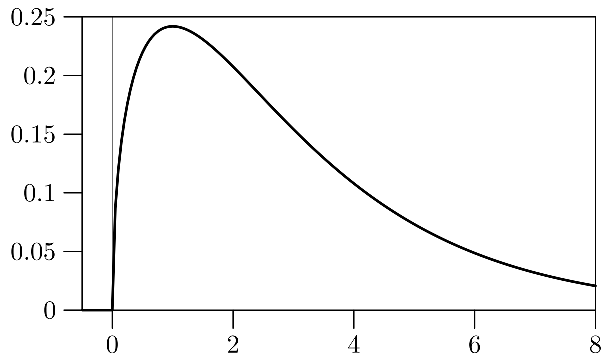
Klasickým příkladem normálního rozdělení je rozdělení náhodných chyb vzniklých při měření nějaké veličiny. Při opakovaném měření způsobují náhodné vlivy odchylky od skutečné hodnoty měřené veličiny. Tyto odchylky jsou (pro velký počet pozorování) rozvrstveny symetricky kolem této skutečné hodnoty.

4. Rozdělení chí-kvadrát $\chi^2(\nu)$

Náhodná veličina $\chi^2(\nu) = U_1^2 + U_2^2 + U_3^2 + \cdots + U_n^2$, kde U_1, U_2, U_3,…,U_n jsou nezávislé náhodné veličiny s rozdělením $N(0,1)$, se řídí rozdělením $\chi^2(\nu)$, kde ν je počet stupňů volnosti. Počtem stupňů volnosti ν rozumíme počet nezávislých sčítanců v součtu. Máme-li n sčítanců, potom $\nu = n - 1$.

Charakteristiky rozdělení $\chi^2(\nu)$:

střední hodnota	$E(x) = \nu$
rozptyl	$D(x) = 2\nu$

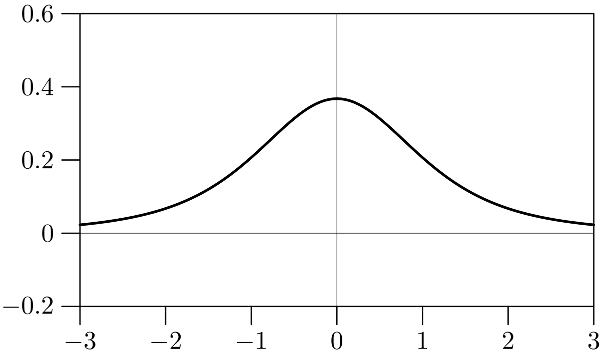


Obrázek 4.3: Chí-kvadrát rozdělení ($\chi^2(3)$).

χ^2 rozdělení má rozsáhlé použití zejména v teorii odhadu a testování hypotéz, zejména při ověřování předpokladu, zda empirické rozdělení četnosti má či nemá pravděpodobnostní rozdělení určitého druhu (tzv. neparametrické statistické testy – viz kapitola 7), dále při ověřování nezávislosti kvalitativních znaků apod.

5. Studentovo t rozdělení $t(\nu)$

Studentovo rozdělení má náhodná veličina, kterou lze vyjádřit jako podíl veličiny s normovaným normálním rozdělením a veličiny s rozdělením χ^2. Má-li náhodná veličina U rozdělení $N(0,1)$ a veličina χ^2 má rozdělení $\chi^2(\nu)$, pak náhodná veličina $t = U : \sqrt{\frac{\chi^2}{\nu}}$ se řídí t rozdělením o $\nu = n - 1$ stupních volnosti. Je charakterizováno parametrem ν, který opět označuje označují stupně volnosti. Pro vysoký počet stupňů volnosti ($\nu > 30$) přechází t-rozdělení v normální rozdělení.



Obrázek 4.4: Studentovo t-rozdělení náhodné veličiny ($t(3)$).

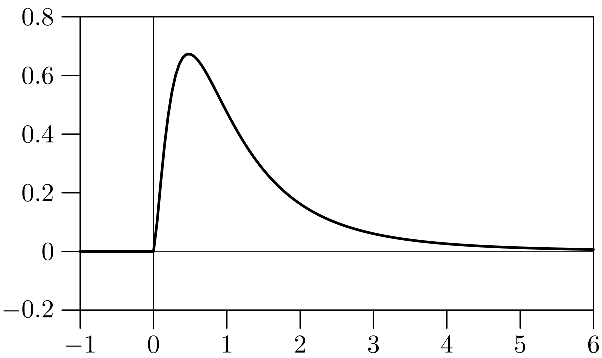
Studentovo rozdělení má (stejně jako výše uvedená rozdělení i následující F-rozdělení) rozsáhlé použití zejména v teorii odhadu a testování hypotéz. Seznámíme se s nimi v následujících kapitolách.

6. Fisher-Snedecorovo F-rozdělení $F(\nu_1, \nu_2)$

Má-li veličina χ_1^2 rozdělení $\chi^2(\nu_1)$ a veličina χ_2^2 rozdělení $\chi^2(\nu_2)$, pak náhodná veličina

$$F = \frac{\chi_1^2}{\nu_1} : \frac{\chi_2^2}{\nu_2}$$

má rozdělení F o ν_1 a ν_2 stupních volnosti.



Obrázek 4.5: Fisher-Snedecorovo F-rozdělení náhodné veličiny ($F(5,8)$).

Shrnutí kapitoly

Náhodná veličina je definována jako výsledek náhodného pokusu. Je pro ni tedy charakteristická velká nejistota o konkrétní hodnotě, které nabude. Pokud však pro jednotlivé možné hodnoty náhodné veličiny známe odpovídající pravděpodobnost, že veličina této hodnoty nabude, získáme poměrně dobrou představu o jejím chování. Přiřazení pravděpodobnosti příslušným hodnotám náhodné veličiny se nazývá zákonem rozdělení náhodné veličiny a je jedním z klíčových pojmů teorie pravděpodobnosti.

V přírodě, technických vědách, sociologických či ekonomických průzkumech, kde jsou prováděna výběrová šetření, jejichž výsledkem jsou soubory náhodných veličin, je obvyklé, že hodnoty v těchto souborech se chovají podle určitých standardních pravidel. Ve statistice jsou proto používány některé standardní modely rozdělení náhodné veličiny. Znalost jejich vlastností umožňuje jejich využití pro vyslovení závěrů o chování a vlastnostech takových výběrových souborů.

Kapitola Vám nabídla pohled na některá nejznámější rozdělení náhodné veličiny. Výlučné místo mezi nimi má tzv. normální rozdělení, které velmi dobře charakterizuje chování velkého množství jevů v přírodě, či technice. Vlastnosti uvedených rozdělení náhodné veličiny využijeme v následujících kapitolách.

Otázky k zamyšlení

1. Vysvětlete pojem náhodné veličiny a uveďte některé příklady náhodné veličiny z hospodářské praxe.
2. Na základě analogie mezi pravděpodobností a relativní četností vysvětlete vztah mezi rozptylem ve smyslu popisné statistiky a rozptylem náhodné veličiny.
3. Vypočítejte rozptyl a střední hodnotu rozdělení náhodné veličiny, které vzniklo jako výsledek sledování spolehlivosti výrobní linky.

Počet selhání za den	1	4	5	7	9	11	12	14
Pravděpodobnost	0,05	0,1	0,05	0,2	0,05	0,2	0,15	0,2

Příloha kapitoly 4

Využití programu MS EXCEL pro analýzu náhodné veličiny

Prostředí programu EXCEL je možno využít pro určení některých vlastností náhodné veličiny. Zejména je ho možno využít pro výpočet **kvantilů** standardních rozdělení náhodné veličiny. Program také umožňuje určit hodnoty **distribuční funkce** těchto standardních rozdělení. Je samozřejmé, že programu EXCEL je v široké míře možno užít i pro zobrazení rozdělení náhodné veličiny pomocí **statistických grafů** (polygonů i histogramů).

Kvantily rozdělení náhodné veličiny

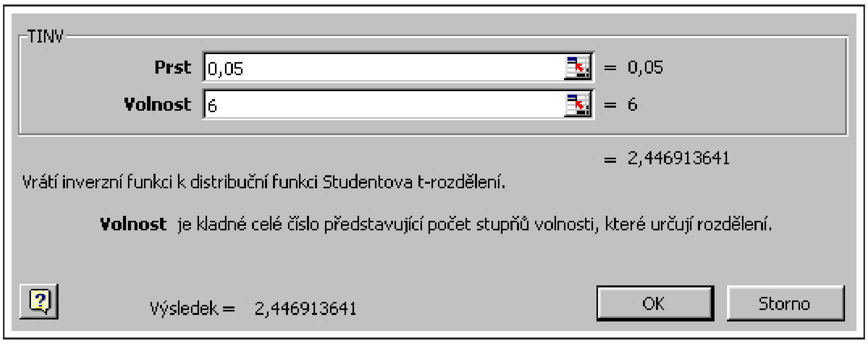
V prostředí programu EXCEL lze spočítat libovolné kvantily vybraných standardních rozdělení náhodné veličiny. Tyto hodnoty jsou obvyklou součástí statistických učebnic, kde jsou zpravidla tabelovány pouze pro vybrané procentní hodnoty kvantilu (viz např. příloha učebnice Seger, Hindls, Hronová: *Statistika v hospodářství*, str. 605–621).

V programu EXCEL lze určit kvantily všech spojitých rozdělení jež jsme zmínili v kapitole:

- **Normální rozdělení $N(\mu; \sigma^2)$**
  funkce NORMINV
- **Normované normální rozdělení $N(0; 1)$**
  funkce NORSMINV
- **Chí-kvadrát rozdělení $\chi^2(\nu)$**
  funkce CHIINV
- **Studentovo t-rozdělení $t(\nu)$**
  funkce TINV
- **Fisher-Snedecorovo F-rozdělení $F(\nu_1, \nu_2)$**
  funkce FINV

Například kvantil t-rozdělení pro 6 stupňů volnosti získáme pomocí funkce TINV v následujícím tvaru:

TINV(prst; volnost)



Obrázek 4.6: Pětiprocentní kvantil t-rozdělení pro 6 stupňů volnosti a pomocí funkce TINV.

Jako první se zadá procentní vyjádření pořadí kvantilu (**Prst**), proměnná **volnost** označuje počet stupňů volnosti (zde 6).

Hodnoty distribuční funkce rozdělení náhodné veličiny

V prostředí programu EXCEL lze také snadno určit hodnoty distribuční funkce libovolné kvantily vybraných standardních rozdělení náhodné veličiny. Také tyto hodnoty jsou součástí statistických učebnic, i když méně obvyklou.

Pomocí programu EXCEL lze určit prakticky jakoukoliv hodnotu distribuční funkce všech výše zmíněných standardních rozdělení. Jejich přehled včetně příslušného názvu funkce uvádí následující tabulka.

Typ rozdělní	**Funkce v EXCELu**
Binomické rozdělení $Bi(n, p)$	BINOMDIST
Poissonovo rozdělení	POISSON
Exponenciální rozdělení	EXPONDIST
Hypergeometrické rozdělení	HYPGEOMDIST
Normální rozdělení $N(\mu; \sigma^2)$	NORMDIST
Normované normální rozdělení $N(0; 1)$	NORMSDIST
Logaritmicko-normální rozdělení	LOGNORMDIST
Chí-kvadrát rozdělení $\chi^2(\nu)$	CHIDIST
Studentovo t-rozdělení $t(\nu)$	TDIST
Fisher-Snedecorovo F-rozdělení $F(\nu_1, \nu_2)$	FDIST