3 Důležitá rozdělení náhodné veličiny

Co to znamená, když prohlásíme, že jsou nějaká důležitá rozdělení? Rozdělení náhodné veličiny je její popis. A náhodná veličina představuje určitý náhodný pokus (kde výsledky prezentujeme jako čísla). Jedná se tedy o typické náhodné pokusy u nichž je znám jejich pravděpodobnostní popis. Několik takových známých rozdělení jak pro diskrétní, tak i pro spojitou náhodnou veličinu si dále uvedeme.

3.1 Diskrétní rozdělení

Diskrétní rozdělení má konečný nebo spočetný počet realizací. Distribuční funkce je skoková

PŘÍKLADY: hod mincí, hod kostkou, tažení králku ze 3 modrých a 5 bílých, počet nehod na křižovatce za 1 rok, počet aut zaznamenaných detektorem za 1 hodinu.

3.1.1 Alternativní rozdělení Alt (π)

Alternativní rozdělení Alt (π) je diskrétní rozdělení, které popisuje situaci, kdy při jednom pokusu náhodná veličina X může nabývat pouze dvou hodnot: 1 (jestliže jev nastoupí - úspěch) a 0 (jestliže jev nenastoupí - neúspěch).

Pravděpodobnostní funkce

$$f(x) = \pi^x (1-\pi)^{1-x}, \; x = \{0,1\}$$

kde π je parametr - pravděpodobnost úspěšného pokusu.
Uvedenou pravděpodobnostní funkci lze také vyjádřit tabulkou

x	0	1
$f(x)$	$1-\pi$	π

Z tabulky je přímo vidět, že pravděpodobnost úspěchu $(x = 1)$ je π a pravděpodobnost neúspěchu $(x = 0)$ je $1-\pi$. Totéž dostaneme i z analytického vyjádření, když dosadíme za x příslušnou hodnotu.
Distribuční funkce

$$F(x) = \begin{cases} 0 & x < 0 \\ 1-\pi & 0 \le x < 1 \\ 1 & x \ge 1 \end{cases}$$

$X \sim \text{Alt}(\pi)$ (náhodná veličina s alternativním rozdělením) má:

- střední hodnotu $E[X] = \pi$,
- rozptyl $D[X] = \pi(1-\pi)$.

Příklad

Máme sklad s výrobky mezi nimiž je 15 % zmetků. Vybereme náhodně jeden výrobek. Úspěch je, když je výrobek dobrý, neúspěch, je-li výrobek zmetek. Určete, s (i) jakou pravděpodobností vybereme zmetek a (ii) s jakou dobrý výrobek. Dále určete (iii) střední hodnotu a (iv) rozptyl a (v) nakreslete pravděpodobnostní a distribuční funkci.

Řešení:

(i) $(1-\pi) = 0.15$ ... pravděpodobnost, že vybereme vadný výrobek

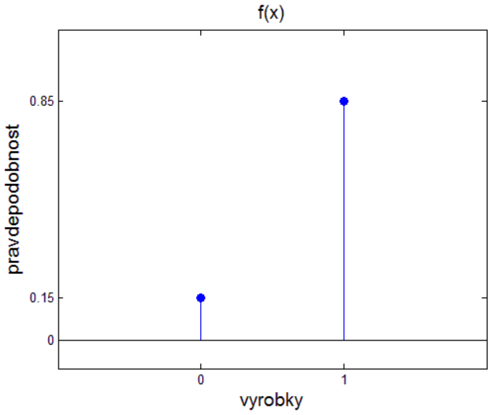
(ii) $\pi = 0.85$ ... pravděpodobnost, že vybereme bezvadný výrobek

(iii) střední hodnota $E[X] = \pi = 0.85$

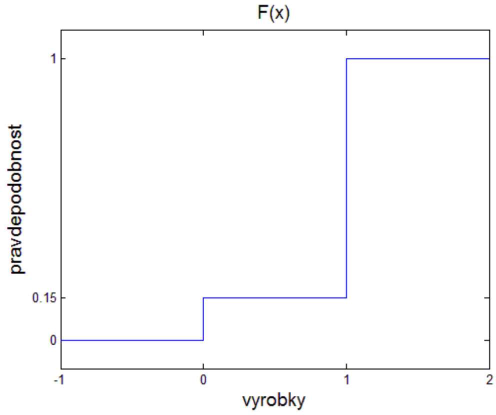
(iv) rozptyl $D[X] = \pi(1-\pi) = 0.85 \times 0.05 = 0.1275$

(v) grafické zobrazení pravděpodobnostní a distribuční funkce

Pravděpodobnostní funkce



Distribuční funkce



3.1.2 Binomické rozdělení Bi (π, n)

Provedeme-li n nezávislých pokusů s alternativním rozdělením, pak součtem výsledků pokusů (tj. za výsledek vezmeme počet dosažených úspěchů) získáme binomické rozdělení. Platí tedy, že pro nezávislé náhodné veličiny $Y_i \sim \mathrm{Alt}(\pi)$, $i = 1, 2, ..., n$ můžeme psát

$$X = \sum_i Y_i \quad \sim \quad Bi\,(\pi, n)$$

Pravděpodobnostní funkce

$$f(x) = \binom{n}{x} \pi^x (1-\pi)^{n-x}, \; x = 0, 1, 2, \ldots, n$$

kde π je pravděpodobnost úspěchu v odpovídajícím alternativní pokusu a n je počet provedených alternativních pokusů. π a n jsou parametry rozdělení.

Vzorec, kterým jsme vyjádřili pravděpodobnostní funkci binomického rozdělení se nazývá vzorec pro binomickou pravděpodobnost. Jeho odvození je poměrně snadné. Člen π^x představuje pravděpodobnost požadovaných x úspěchů (při nezávislých pokusech), člen $1 - \pi^{n-x}$ je pravděpodobnost zbylých pokusů, kdy výsledkem byl neúspěch a kombinační číslo $\binom{n}{x}$ představuje počet způsobů, kterými lze z provedených n pokusů vybrat x úspěchů.

$X \sim \mathrm{Bi}(\pi, n)$ má:

- střední hodnotu $E[X] = n\pi$,
- rozptyl $D[X] = n\pi(1-\pi)$.

Příklad 1

Ve velikém skladu jsou uloženy výrobky[4]. Mezi nimi jich je 85% dobrých výrobků a 15% zmetků. Jak veliký je třeba udělat výběr, aby v něm bylo s pravděpodobností 99% alespoň 5 dobrých výrobků?

Řešení:

Jedná se o binomické rozdělení s velikostí výběru n a pravděpodobností úspěchu $\pi = 0.85$. Hledáme první n takové, že platí

$$P = \sum_{i=5}^{n} \binom{n}{5} \pi^5 (1-\pi)^{n-5} \geq 0.99$$

To je ale implicitní nerovnice, která by se obecně musela řešit nějakou numerickou metodou. V našem případě, kdy n je diskrétní, stačí počítat hodnoty P postupně pro $n = 5, 6, \cdots$ a za výsledek vezmeme první hodnotu n pro kterou sledovaná nerovnice platí. Výpočty uspořádáme do tabulky

n	5	6	7	8	9
P	0.4356	0.7765	0.9262	0.9786	0.9944

Hledaný výběr tedy bude o rozsahu $n = 9$.

[4]Uvažovaný sklad musí být hodně velký. Základním předpokladem pro binomické rozdělení je, že uvažované pokusy jsou nezávislé. Jestliže ale vybíráme výrobky a nevracíme je zpět, jedná se závislé pokusy, protože počet a tím i pravděpodobnost dobrých a špatných výrobků se mění. Jestliže je ale sklad hodně veliký a v něm velké množství výrobků, bude relativní vliv vybraných výrobků na celkový počet zanedbatelný a pravděpodobnosti můžeme pokládat za konstantní.

Příklad 2

Za jednu hodinu projede křižovatkou v průměru 50 aut. Jaká je (i) pravděpodobnost, že 5 z nich odbočí, je-li pravděpodobnost přímé jízdy 0.9? Dále určete (ii) střední hodnotu, (iii) rozptyl a (iv) nakreslete pravděpodobnostní a distribuční funkci.

Řešení:

(i) V tomto příkladě považujeme za úspěch odbočení a jeho pravděpodobnost je $\pi = 1 - 0.9 = 0.1$. Pravděpodobnost odbočení pěti aut z 50 je

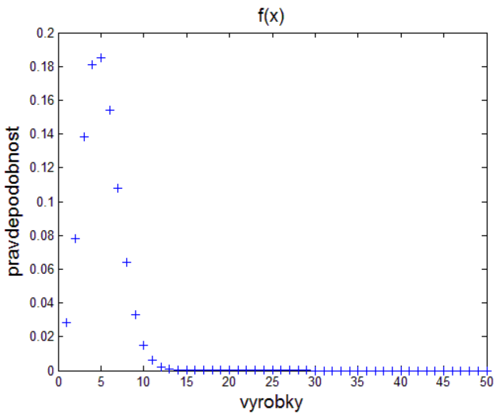
$$f\left(5\right) = \binom{50}{5} 0.1^5 0.9^{45} = 0.185,$$

(ii) střední hodnota $E\left[X\right] = n\pi = 50 \times 0.1 = 5$,

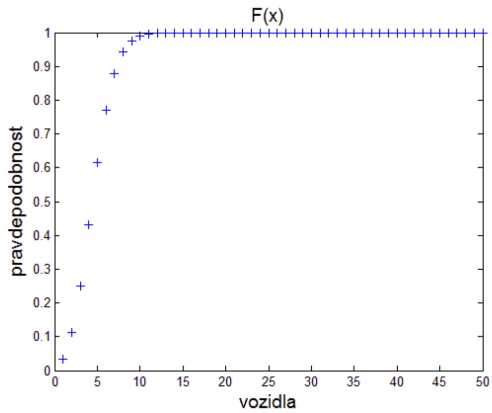
(iii) rozptyl $D\left[X\right] = n\pi\left(1 - \pi\right) = 50 \times 0.1 \times 0.9 = 4.5$,

(iv) grafické zobrazení pravděpodobnostní a distribuční funkce

Pravděpodobnostní funkce



Distribuční funkce



3.1.3 Hypergeometrické rozdělení Hg (N, M, n)

Hypergeometrické rozdělení je rozšířením binomického, resp. binomické rozdělení je limitním případem hypergeometrického. Jedná se o pokus, kdy ze souboru o N prvcích, z nichž M má danou vlastnost, vybereme n prvků, aniž bychom vybrané prvky do souboru vraceli. Náhodná veličina $X \sim \mathrm{Hg}\left(N, M, n\right)$ nabývá hodnot z množiny přirozených čísel $\mathbb{N}$ nebo 0.

Binomické rozdělení je tedy limitním případem hypergeometrického, a sice pro $n \to \infty$ a $\frac{M}{N} \to 0$. Poměr $\frac{M}{N}$ je u binomického rozdělení definován jako pravděpodobnost úspěchu π.

Pravděpodobnostní funkce

$$f\left(x\right) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

pro $x = 0, 1, 2, \ldots, min(n, M)$. Dále musí platit $n < N,\ M < n$.

$X \sim \mathrm{Hg}\left(N, M, n\right)$ má:

- střední hodnotu $E\left[X\right] = n\frac{M}{N}$,
- rozptyl $D\left[X\right] = n\frac{M}{N}\left(1 - \frac{M}{N}\right)\frac{N-n}{N-1}$.

Příklad

Ve skladu je uloženo 40 výrobků. Mezi nimi jich je 85% dobrých výrobků a 15% zmetků. (i) Jak veliký je třeba udělat výběr, aby v něm bylo s pravděpodobností 99% alespoň 5 dobrých výrobků? (ii) Nakreslete pravděpodobnosti a distribuční funkci pro případ, že ze uděláme výběr 10 výrobků a sledujeme pravděpodobnost výskytu dobrých výrobků ve výběru. K tomuto případu určete (iii) střední hodnotu a (iv) rozptyl.

Řešení:

(i) V tomto případě se nejedná o binomické rozdělení, ačkoliv zadání vypadá téměř shodně. Rozdíl je v tom, že v tomto případě každý výběr ovlivní další výběr a tudíž již nelze hovořit o nezávislých náhodných veličinách. Pro toto zadání se tedy použije hypergeometrické rozdělení $\mathrm{Hg}\,(N, M, n)$.

Jedná se tedy o hypergeometrické rozdělení s velikostí souboru $N = 40$, z nichž právě $M = 34$ je dobrých výrobků a vybíráme n prvků. Hledáme první n takové, že platí

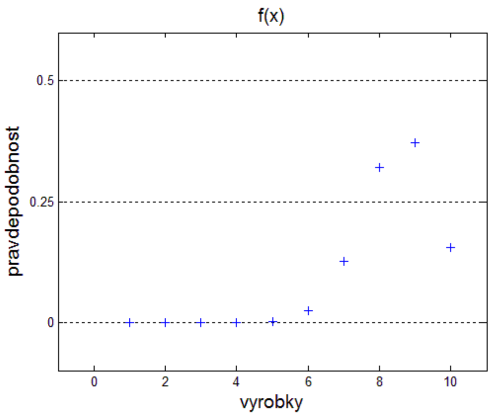
$$P\,(x = 5) \quad = \quad \sum_{i=5}^{n} \frac{\binom{M}{5}\binom{N-M}{n-5}}{\binom{N}{n}} \geq 0.99$$

Stejně jako u binomického rozdělení i tady by se musela tato implicitní nerovnice řešit numerickou metodou. I v tomto případě je možno použít postupný výpočet pravděpodobnosti P, a proto budeme postupovat stejně a pro $n = 5, 6, \ldots$ jí dopočteme a budeme hledat první n pro které bude nerovnice platit.

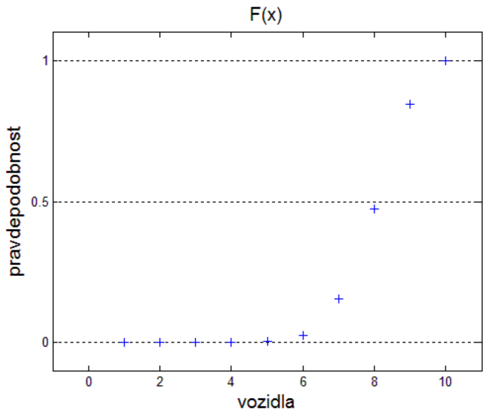
n	5	6	7	8
P	0.4229	0.7853	0.9453	0.9905

Hledaný výběr tedy bude o rozsahu $n = 8$.
(ii) grafické zobrazení pravděpodobnostní a distribuční funkce

Pravděpodobnostní funkce



Distribuční funkce



(iii) střední hodnota $E\,[X] = n\frac{M}{N} = 10\frac{34}{40} = 8.5$,

(iv) rozptyl $D\,[X] = n\frac{M}{N}\left(1 - \frac{M}{N}\right)\frac{N-n}{N-1} = 0.29$.

3.1.4 Poissonovo rozdělení Po (λ)

Při velkých hodnotách n a malých hodnotách π je binomické rozdělení pro výpočet klasickou numerickou metodou nevhodné. Z tohoto důvodu lze pravděpodobnostní a distribuční funkce Binomické rozdělení $\mathrm{Bi}\,(\pi, n)$ aproximovat Poissonovým rozdělením $\mathrm{Po}\,(\lambda)$. Poissonovo rozdělení je tedy limitním případem binomického, a sice pro $n \to \infty$ a $\pi \to 0$, a to tak, že limita součinu $n \cdot \pi = \lambda$ je konstantní.

Toto rozdělení má sice nekonečný počet realizací, ale nekonečně spočetný (realizacím lze jednoznačně přiřadit přirozená čísla). Proto patří mezi diskrétní rozdělení. Pravděpodobnostní funkci Poissonova rozdělení lze odvodit jako limitu binomického rozdělení. Odvození je uvedeno v dodatku **??**

Pravděpodobnostní funkce

$$f\,(x_i) = \mathrm{e}^{-\lambda}\frac{\lambda^{x_i}}{x_i!},\ x = 0, 1, 2, \cdots$$

Podle základních vlastností musí být součet všech členů roven jedné, tj.

$$\sum_{i=0}^{\infty} \mathrm{e}^{-\lambda}\frac{\lambda^{x_i}}{x_i!} = \mathrm{e}^{-\lambda}\sum_{i=0}^{\infty}\frac{\lambda^{x_i}}{x_i!} = 1$$

což skutečně je, protože suma v prostředním výrazu není nic jiného než Taylorův rozvoj funkce $\mathrm{e}^{-\lambda}$ [CITACE].

$X \sim \mathrm{Po}\,(\lambda)$ má:

- střední hodnotu $E\left[X\right]=\lambda$,
- rozptyl $D\left[X\right]=\lambda$.

Poissonovo rozdělení se velmi často využívá v Teorii hromadné obsluhy, protože se vlastně jedná o rozdělení, které je vhodné na popis náhodné veličiny, kde úspěch nastane s malou pravděpodobností na delším intervalu. Příkladem může být vstup zákazníka od obchodu (předpokládáme obchod s menší frekvencí zákazníků, nikoliv nákupní centra).

Příklad 1

Uvažujme podobný příklad jako byl u binomického rozdělení ale s extrémnějšími hodnotami: Za jeden den projede křižovatkou v průměru 3500 aut. Jaká je (i) pravděpodobnost, že 25 z nich odbočí, je-li pravděpodobnost přímé jízdy 0.999? Dále určete (ii) střední hodnotu, (iii) rozptyl a (iv) nakreslete pravděpodobnostní a distribuční funkci.

Řešení:

(i) Pokud bychom chtěli úlohu řešit pomocí binomického rozdělení (protože se skutečně o binomické rozdělení jedná), dospějeme k výrazu

$$f\left(5\right)=\binom{350}{5}0.001^{5}0.999^{3495}$$

což není moc hezké. Proto aproximujeme Poissonovým rozdělením pro $\lambda=350\cdot 0.01=35$. Dostaneme

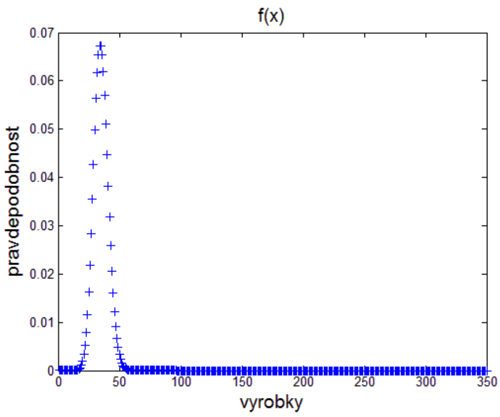
$$f_P\left(5\right)=\mathrm{e}^{-35}\frac{35^5}{25!}=0.132,$$

(ii) střední hodnota $E\left[X\right]=\lambda=3.5$,

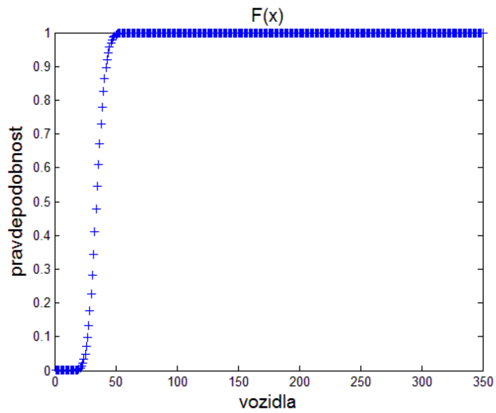
(iii) rozptyl $D\left[X\right]=\lambda=3.5$

(iv) grafické zobrazení pravděpodobnostní a distribuční funkce

pravděpodobnostní funkce



distribuční funkce



Příklad 2

Sledujeme silnici s malým provozem. Po dobu jedné minuty zaznamenáváme každou vteřinu přítomnost auta. Po hodině sledování jsme zjistili, že v průměru za jednu minutu projedou 3 auta. Za předpokladu Poissonova rozdělení počtu projetých aut během jedné minuty určete pravděpodobnost, že během jedné minuty projede více než 5 automobilů.

Řešení:

Pro určení λ známe $n=60,\ \pi=3/60$. Odtud $\lambda=n\pi=3$.

$$P\left(X>5\right)=1-P\left(X\leq 5\right)=1-\mathrm{e}^{-3}\left(\frac{3^0}{0!}+\frac{3^1}{1!}+\frac{3^2}{2!}+\frac{3^3}{3!}+\frac{3^4}{4!}+\frac{3^5}{5!}\right)=0.084.$$

Pravděpodobnost projetí více než pěti aut za jednu minutu je 0.084.

3.1.5 Geometrické rozdělení G $\left(\pi\right)$

Základem pro geometrické rozdělení je opět náhodný pokus s alternativním rozdělením Alt $\left(\pi\right)$, tedy pokus se dvěma výsledky (úspěch a neúspěch) a s pevnou pravděpodobností úspěchu, označenou π. Toto rozdělení počítá pravděpodobnost toho, že při opakovaných nezávislých pokusech bude první úspěch předcházet x neúspěchů.

Pravděpodobnostní funkce

$$f(x) = \pi (1-\pi)^x, \; x = 0, 1, 2, \cdots$$

Pro zpřesnění je lépe uvést, že při výpočtu pravděpodobnostní funkce geometrické náhodné veličiny $X \sim \mathrm{G}(\pi)$ se do výpočtu pravděpodobnosti uvažují všechny neúspěšné pokusy i pokus úspěšný.

$X \sim \mathrm{G}(\pi)$ má:

- střední hodnotu $E[X] = \frac{1}{\pi}$,
- rozptyl $D[X] = \frac{1-\pi}{\pi^2}$.

Příklad 1

Sledujeme silniční síť s řízenými křižovatkami. Sítí projíždí automobil a pravděpodobnost, že ho každý jednotlivý semafor zastaví je stejná a stálá, rovna 0.62. Jaká je pravděpodobnost, že automobil bude na dvou křižovatkách zastaven a teprve na třetí projede bez zastavení? Uvažujeme, že dopravní síť je volná.

Řešení

V tomto příkladě je úspěch projetí automobilu bez zastavení. Pravděpodobnost úspěchu je $\pi = 1 - 0.62 = 0.38$. Sledovaná realizace je $x = 2$ (dvě zastavení před volným průjezdem). Dosazením do pravděpodobnostní funkce dostáváme

$$f(2) = 0.38 (1 - 0.38)^2 = 0.146$$

Pravděpodobnost, že automobil bude dvakrát zastaven než se mu podaří projet bez zastavení je tedy 0.146.

Příklad 2

Máme sklad s výrobky mezi nimiž je 15 % zmetků. Vybereme náhodně jeden výrobek. Úspěch je, když je výrobek dobrý, neúspěch, je-li výrobek zmetek. Určete, s (i) jakou pravděpodobností vybereme zmetek, pokud můžeme vybírat až 5x. Dále určete (ii) střední hodnotu a (iii) rozptyl a (iv) nakreslete pravděpodobnostní a distribuční funkci.

Řešení

V tomto příkladě je úspěch výběr zmetku, tzn. $\pi = 0.15$.

(i)

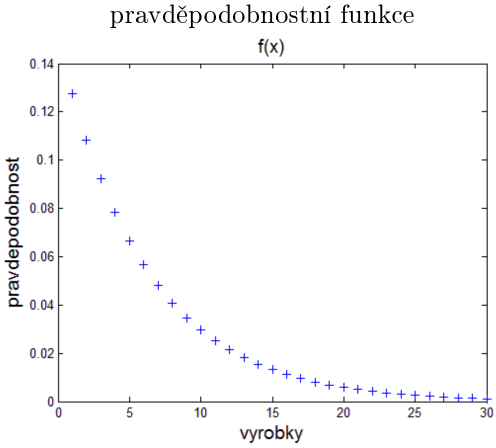
$$\begin{aligned}
f(0) &= 0.15 (1 - 0.15)^0 = 0.150 \\
f(1) &= 0.15 (1 - 0.15)^1 = 0.128 \\
f(2) &= 0.15 (1 - 0.15)^2 = 0.108 \\
f(3) &= 0.15 (1 - 0.15)^3 = 0.092 \\
f(4) &= 0.15 (1 - 0.15)^4 = 0.078
\end{aligned}$$

(ii) střední hodnota $E[X] = \frac{1}{\pi} = \frac{1}{0.15} = 6.667$,

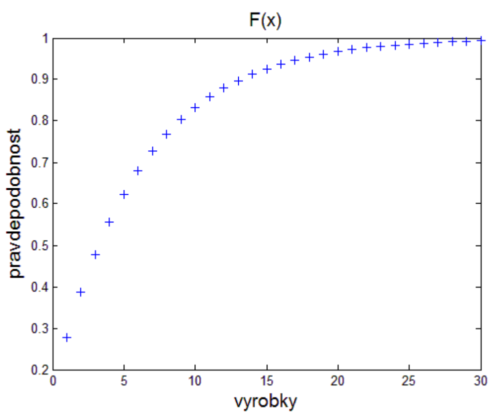
(iii) rozptyl $D[X] = \frac{1-\pi}{\pi^2} = \frac{1-0.15}{0.15^2} = 37.778$,

(iv) grafické zobrazení pravděpodobnostní a distribuční funkce

pravděpodobnostní funkce



distribuční funkce



3.1.6 Negativně binomické rozdělení NBi (n, π)

Negativně binomické rozdělení je rozšířením geometrického rozdělení. Vychází také z alternativního pokusu a podobně jako geometrické rozdělení určuje pravděpodobnost toho, že n-tý úspěch předchází x neúspěchů. Přitom předchozích $n-1$ úspěchů mohlo nastat kdykoli od začátku provádění pokusů.

Zjednodušeně lze říci, že náhodná veličina $X \sim \text{NBi}(n, \pi)$ popisuje počet x neúspěšných pokusů než nastoupí n-tý úspěšný pokus. Pokud $n = 1$, pak hovoříme o geometrickém rozdělení.

Pravděpodobnostní funkce

$$f(x) = \binom{x+n-1}{n-1} \pi^n (1-\pi)^x, \ x = 0, 1, 2, \cdots$$

$X \sim \text{NBi}(n, \pi)$ má:

- střední hodnotu $E[X] = \frac{n(1-\pi)}{\pi}$,
- rozptyl $D[X] = \frac{n(1-\pi)}{\pi^2}$.

Příklad

Sledujeme silniční síť s řízenými křižovatkami. Sítí projíždí automobil a pravděpodobnost, že ho každý jednotlivý semafor zastaví je stejná a stálá, rovna 0.62. (i) Jaká je pravděpodobnost, že automobil bude na prvních dvou křižovatkách zastaven a až na třetí projede bez zastavení? (ii) Jaká je pravděpodobnost, že při projíždění křižovatek bude potřetí zastaven právě na 5. křižovatce. Pro tuto situaci určete (iii) střední hodnotu a (iv) rozptyl. (v) Nakreslete pravděpodobnostní a distribuční funkci pro případ, že n-tý úspěšný pokus bude 3. zastavení a počet neúspěšných pokusů $x = 0, 1, 2, 3, \ldots, 20$.

Řešení

(i) Zadání je $x = 2$ (počet neúspěchů před prvním úspěchem), $\pi = 0.38$ (pravděpodobnost, že nastoupí úspěch - automobil projede) a $n = 1$ (počet úspěchů). Dosazením dostaneme

$$f(2) = \binom{2+1-1}{2} 0.38^1 (1-0.38)^2 = 0.146.$$

V tomto případě se podle definice pravděpodobnostní funkce vlastně nejedná o negativně binomické rozdělení, ale o geometrické rozdělení. Protože je ovšem negativně binomické rozdělení pouze rozšířením, z principu musí být výsledek stejný jako při použití geometrického. V tomto případě lze porovnat výsledek s výsledkem u příkladu č. 1 u geometrického rozdělení (řešené příklady).

(ii) Zadání je $x = 3$, $\pi = 0.38$ a $n = 2$ (až druhý úspěch bude ten „správný“). První úspěch tedy v daném případě může nastoupit při projetí 1. nebo 2. nebo 3. nebo 4. křižovatkou.

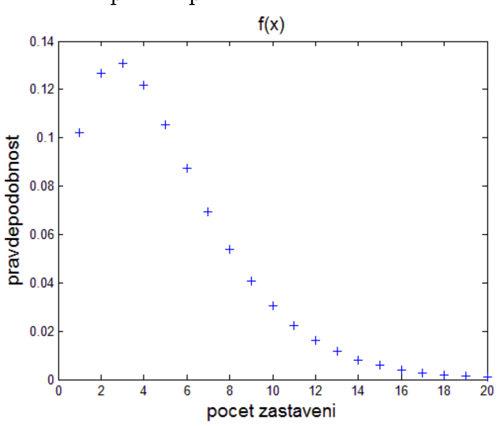
$$f(3) = \binom{3+2-1}{2-1} 0.38^2 (1-0.38)^3 = 0.1377,$$

(iii) střední hodnota $E[X] = \frac{2(1-0.38)}{0.38} = 3.26$,

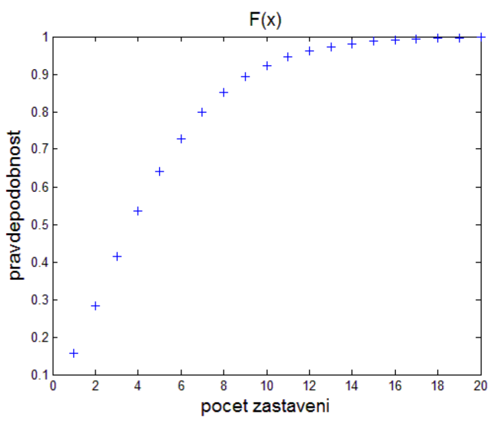
(iv) rozptyl $D[X] = \frac{2(1-0.38)}{0.38^2} = 8.58$,

(v) grafické zobrazení pravděpodobnostní a distribuční funkce

pravděpodobnostní funkce



distribuční funkce



3.1.7 Rovnoměrné rozdělení $\mathrm{U}(n)$

Diskrétní rovnoměrné rozdělení patří k základním a nejjednodušším typům rozdělení. Na něm jsou založeny nejznámější „školní“ příklady o házení mincí nebo kostkou. Diskrétní rozdělení popisuje náhodnou veličinu s konečným počtem n realizací které nastávají se stejnou pravděpodobností $\frac{1}{n}$. Rovnoměrnost rozdělení vyjadřuje skutečnost, že v realizacích nemáme žádné preference nebo, že o nich nic nevíme.

Rovnoměrné rozdělení má dvě důležité charakteristiky

- všechny realizace jsou rovnocenné (žádný výsledek není preferován),
- ostré hranice pro definiční obor (hranice nelze překročit).

Pravděpodobnostní funkce

$$f(x) = \frac{1}{n}, \quad x = 1, 2, \cdots, n$$

Rovnoměrné rozdělení se často používá při generování dat. Jeho využití je důležité například při určení chyby přístroje, resp. rozptylu chyby přístroje.

$X \sim \mathrm{U}(n)$ má:

- střední hodnotu $E[X] = \frac{1}{n}\sum x_i$,
- rozptyl $D[X] = \frac{1}{n}\sum (x_i - E(X))^2$.

Příklad

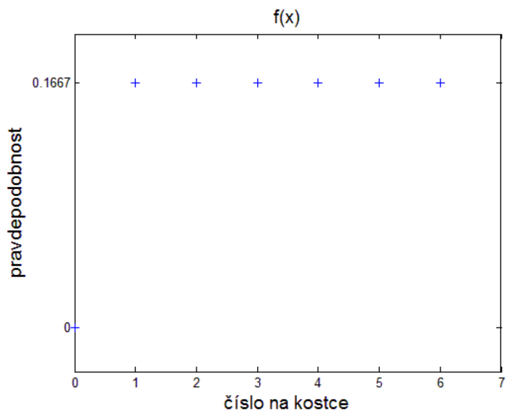
Házíme šestistrannou hrací kostkou. Náhodná veličina je číslo, které padne. Graficky znázorněte (i) pravděpodobnostní a distribuční funkci. Dále určete (ii) střední hodnotu a (iii) rozptyl.

Řešení:

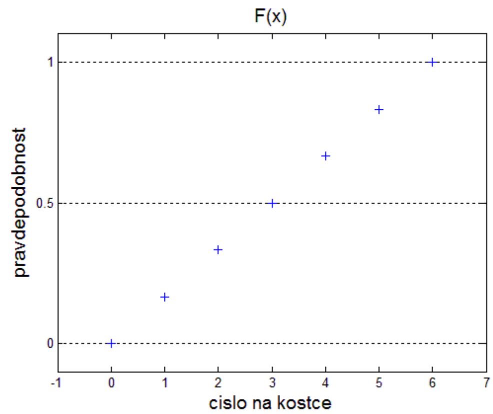
Pokus má šest realizací, tj. na kostce může padnout 6 různých variant a každá má pravděpodobnosti $\frac{1}{6}$.

(i) grafické zobrazení pravděpodobnostní a distribuční funkce

pravděpodobnostní funkce



distribuční funkce



(ii) střední hodnota $E[X] = \frac{1}{6}(1+2+3+4+5+6) = 3.5000$,

(iii) rozptyl $D[X] = \frac{1}{6}\left((1-3.5)^2 + (2-3.5)^2 + \ldots + (6-3.5)^2\right) = 2.9167$,

3.2 Spojité rozdělení

Spojité rozdělení má nekonečný nespočetný počet realizací, tj. jeho realizace jsou z reálné osy. Pravděpodobnost výskytu v jednom daném okamžiku je téměř rovna nule, ale zároveň je distribuční funkce spojitého rozdělení absolutně spojitá.

PŘÍKLADY: chyby v měření rozměru součástky, doba čekání na tramvaj, bezporuchová doba fungování přístroje, intenzita dopravního proudu (je vlastně diskrétní, ale lze ji aproximovat spojitým - půlky aut)

3.2.1 Rovnoměrné rozdělení $\mathrm{U}(a, b)$

O rovnoměrném rozdělení jsme mluvili už v diskrétních rozděleních. Nyní se dostáváme k jeho spojité variantě. I u této varianty platí, že má ve všech bodech na intervalu (a, b) konstantní hustotu pravděpodobnosti. Mimo tento interval je hustota pravděpodobnosti rovna 0.

Hustota pravděpodobnosti

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{pro } x \in (a,b) \\ 0 & \text{pro } x \notin (a,b) \end{cases},$$

kde $a, b \in \mathbb{R}$.

Distribuční funkce

$$F(x) = \begin{cases} 0 & \text{pro } x \leq a \\ \frac{x-a}{b-a} & \text{pro } a < x < b \\ 1 & \text{pro } x \geq b \end{cases}.$$

$X \sim \mathrm{U}(a,b)$ má:

- střední hodnotu $E[X] = \frac{a+b}{2}$,
- rozptyl $D[X] = \frac{(b-a)^2}{12}$.

Příklad

Sledujeme dobu čekání na tramvaj, která má pevný a stálý interval 8 minut, pro náhodně přicházející osoby. Dolní hranice je 0 (osoba tramvaj zrovna doběhla) a horní mez je rovna intervalu (dané osobě tramvaj právě ujela a ona musela čekat na další). Určete (i) střední hodnotu, (ii) rozptyl a (iii) graficky znázorněte hustotu pravděpodobnosti a distribuční funkci.

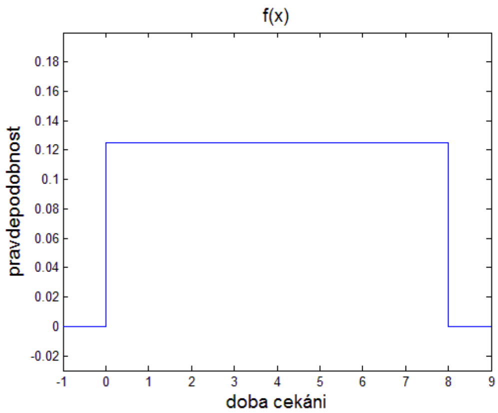
Řešení:

(i) Střední hodnota $E[X] = \frac{0+8}{2} = 4$,

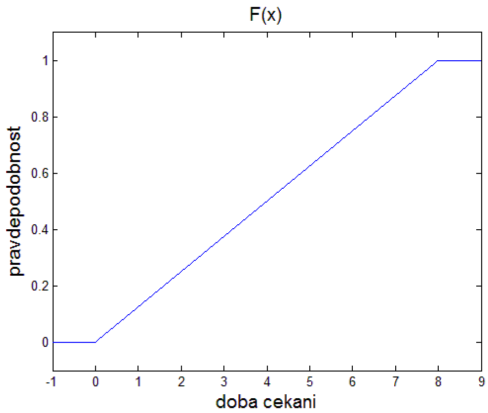
(ii) rozptyl $D[X] = \frac{(8-0)^2}{12} = 5.33$,

(iii) grafické zobrazení pravděpodobnostní a distribuční funkce

pravděpodobnostní funkce



distribuční funkce



3.2.2 Normální rozdělení N (μ, σ^2)

Normální rozdělení (někdy nazývané Gaussovo rozdělení) je nejčastěji používaným rozdělením spojité náhodné veličiny. Důvody k tomu jsou dva:

1. normální rozdělení vzniká při spolupůsobení velkého množství nezávislých malých náhod, takže řada ne přesně definovaných neurčitostí se mu značně podobají - např. když se sype suchý písek, zrníčka na sebe navzájem mnohokráte narazí a výsledkem je hromada, která dosti připomíná „dvourozměrnou gaussovku",

2. normální rozdělení je jedním z mála, pro které se dají analyticky dopočítat i složitější statistické úlohy - jinak je často třeba dělat určité aproximace.

Hustota pravděpodobnosti

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \; x \in R$$

kde parametry rozdělení jsou μ - střední hodnota a σ^2- rozptyl.

Normální rozdělení s nulovou střední hodnotou μ=0 a jednotkovým rozptylem $\sigma^2 = 1$ nazýváme normovaným normálním rozdělením. Budeme-li normálně rozdělenou náhodnou veličinu X normovat tak, aby měla střední hodnotu nula a rozptyl jedna, dostaneme náhodnou veličinu Z se standardním normálním rozdělením. Normovací vztah je $z = (x - \mu)/\sigma$ a odpovídající hustota pravděpodobnosti

$$f(z) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}z^2}$$

Distribuční funkce

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{1}{2\sigma^2}(t-\mu)^2} dt$$

$X \sim \mathrm{N}(\mu, \sigma^2)$ má:

- střední hodnotu $E[X] = \mu$,
- rozptyl $D[X] = \sigma^2$.

Ačkoli je normální rozdělení tak časté a pro svoje dobré výpočetní vlastnosti oblíbené, je s ním trochu potíž. K hustotě pravděpodobnosti totiž neexistuje primitivní funkce. Pravděpodobnosti intervalů, což jsou integrály z hustoty pravděpodobnosti přes tyto intervaly, a stejně tak hodnoty distribuční funkce nelze analyticky počítat. Hodnoty distribuční funkce pro standardní normální rozdělení je možné vypočíst numericky pomocí různých matematických programů nebo lze použít statistické tabulky (po znormování).

Příklad 1

Z vojenského tábora má do sousedního stanoviště dojet nákladní automobil. Po cestě se nachází most o výšce 4 metry. Výška vojenských nákladních automobilů lze popsat normálním rozdělením se střední hodnotou $\mu = 3.8$ m a směrodatnou odchylkou $\sigma = 0.3$. (i) Jaká je pravděpodobnost, že vyslané nákladní auto pod mostem projede? Určete (ii) střední hodnotu a (iii) rozptyl. (iv) nakreslete hustotou pravděpodobnosti a distribuční funkci.

Řešení:

(i) Ze zadání tedy plyne, že výška auta je náhodná veličina $X \sim N(3.8, 0.3^2)$, tj. s normálním rozdělením se střední hodnotou $\mu = 3.8$ a rozptylem $\sigma^2 = 0.3^2 = 0.09$. Ptáme se, na pravděpodobnost $P(X < 4)$. Pro určení této pravděpodobnosti můžeme použít tabulky (ukážeme jen pro přehled, dále budeme uvažovat pouze s programy) nebo nějaký statistický nástroj (Matlab, Statistica...).

1. Použití tabulek:

   při práci s tabulkami se předpokládá, že náhodná veličina je ve formě normovaného normálního rozdělení, tedy $X \sim N(0, 1)$. Musíme tedy úlohu převést do normované oblasti, tj. náhodnou veličinu X a s ní související údaje musíme přepočítat na náhodnou veličinu $Z \sim N(0, 1)$. Vzorec pro přepočet je

   $$Z = \frac{X - \mu}{\sigma} \tag{7}$$

   Pro hledanou pravděpodobnost nyní platí

   $$P(X < 4) = P\left(\frac{X - \mu}{\sigma} < \frac{4 - 3.8}{0.3}\right) = P(Z < 0.67) = 0.749$$

2. Použití programu Matlab[5]

   Hledáme pravděpodobnost, kdy náhodná veličina X bude menší než 4, tzn. hledáme distribuční funkci nikoliv hustotu pravděpodobnosti (tu bychom použili v případě, že by tam bylo znaménko " = "). Použijeme tedy funkci

   normal_cdf(4, 3.8, 0.09)   % normal_cdf(X,m,v)

Tedy pravděpodobnost, že vojenský automobil projede pod mostem je 0.749.

Příklad 2

Na městské komunikaci nedaleko od řízené křižovatky měříme rychlosti projíždějících automobilů a uvažujeme o tom, zda tyto rychlosti můžeme popsat pomocí normálního rozdělení.

Ověření typu rozdělení je předmětem testování, ke kterému se dostaneme až později ve statistice. Nyní se budeme držet jen základních rysů, které by normální rozdělení mělo mít.

[5] Při práci s programem MATLAB se pracuje s verzí 7.5 (R2007b). Dále se předpokládá použití Statistického balíčku, který je vytvořen přímo pro potřeby předmětu. Případně v některých případech lze použít Statistic toolbox, který je rozšířením programu MATLAB. Vzhledem k tomu, že se jedná o placené rozšíření, řešené příklady budou řešeny pomocí statistického balíčku, který je možno stáhnout na stránkách předmětu.

Budeme-li situaci sledovat, ihned zjistíme, že rychlosti automobilů se významně liší v situaci, kdy do měřeného místa zasahuje z křižovatky kolona, a jestliže žádná kolona není nebo je tak malá, že na měřené místo nemá vliv. Rozhodneme-li se popisovat rychlosti bez ohledu na kolonu, dostaneme velmi hrubý popis, který nepopisuje dobře ani jednu ze zmíněných situací. Bude proto lépe situaci rozdělit na dva případy: kolona má velký vliv a kolona má malý vliv.

Všimněme si nejprve prvního stavu, kdy kolona prakticky zasahuje na měřené místo a způsobuje, že často naměříme rychlost nula. V důsledku toho, bude střední hodnota blízká nule. Vzhledem k tomu, že normální rozdělení je symetrické, bude jeho hustota pravděpodobnosti svou levou částí zasahovat do záporné poloosy realizací a tím tvrdíme, že mohou nastat i záporné rychlosti (mají nenulovou pravděpodobnost). V tomto případě je tedy normální rozdělení nevhodné.

V druhém případě jsou rychlosti aut větší (zřejmě někde kolem maximální povolení rychlosti). Průměrná rychlost bude taky velká a odchylky od ní budou na obě strany, jak kladné (auta překračující povolenou rychlost), tak i záporné (ti, co telefonují, hledají neznámou adresu atd.). V tomto případě lze dobře předpokládat, že nulové rychlosti se nebudou vyskytovat. Gaussova křivka, která popisuje hustotu pravděpodobnosti normálního rozdělení bude svým vrcholem nad střední hodnotou a levou částí bude zasahovat do záporných hodnot až v místě, kdy je prakticky nulová, a to tak, že i hodnota integrálu přes záporné hodnoty bude přibližně nula. V tomto případě je pro popis rychlostí projíždějících aut možno použít normální rozdělení.

3.2.3 Log-normální rozdělení $LN(\mu, \sigma)$

Normální rozdělaní, kterým jsme se právě zabývali, má jednu nevýhodu. Je symetrické, definované na celé reálné ose a tedy, připouští i záporné realizace náhodné veličiny. Přitom nezápornost je přirozenou vlastností celé řady reálných veličin, např. délka, počet (tj. intenzita i hustota dopravního proudu), rychlost a mnoho dalších. S touto nevýhodou se částečně vypořádává log-normální rozdělení. Jeho realizace dostaneme transformací normálního rozdělení $U \sim N\left(\mu, \sigma^2\right)$ podle rovnice $X = \mathrm{e}^U$. Toto rozdělení se pro velká μ chová přibližně jako normální, pro malá μ je nesymetrické. Obor hodnot, na kterém je jeho hustota pravděpodobnosti nenulová je množina nezáporných reálných čísel. Za toto vylepšení se ale platí složitějším popisem a jednoduchost při úpravách se většinou ztrácí.

Hustota pravděpodobnosti

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2\sigma^2}(\ln x - \mu)^2} \quad \text{pro } x > 0$$

$X \sim LN(\mu, \sigma)$ má:

- střední hodnotu $E[X] = e^{\mu + \sigma^2/2}$,
- rozptyl $D[X] = \left(e^{\sigma^2} - 1\right) e^{2\mu + \sigma^2}$.

Toto rozdělení se hodí např. pro modelování dopravního proudu. Pro velké intenzity mohou být změny kladné i záporné (blízké normálnímu rozdělení), při nulové intenzitě mohou být změny jen kladné.

3.2.4 Exponenciální rozdělení $Exp(\delta)$

Exponenciální rozdělení je spojitou obdobou geometrického, které popisuje počet neúspěchů, které je třeba absolvovat, než dojde ke změně stavu a obdržíme úspěch. Podobně exponenciální rozdělení vyjadřuje čas mezi náhodně se vyskytujícími událostmi např. dobu fungování určitého přístroje, než dojde ke změně stavu - přístroj se porouchá. Tomuto jevu se běžně říká bezporuchová doba fungování přístroje. Zde se ale jedná o velmi speciální případ, kdy porucha je způsobena nikoli stárnutím, ale nějakou vnější příčinou, která přístroj v čase ohrožuje se stále stejnou intenzitou. Z této skutečnosti plyne vlastnost exponenciálního rozdělení, která říká, že toto rozdělení nemá paměť. To lze vyjádřit vztahem $P(X > t + s | X > t) = P(X > s)$, který říká: jestliže přístroj fungoval až do času t, pak pravděpodobnost, že bude ještě dále fungovat po dobu s je stejná, jako pravděpodobnost, že bude fungovat po dobu s, aniž byl před tím v provozu.

Hustota pravděpodobnosti

$$f(x, \delta) = \frac{1}{\delta}\mathrm{e}^{-\frac{x}{\delta}}, \ x \geq 0$$

$X \sim Exp(\delta)$ má:

- střední hodnotu $E[X] = \delta$,
- rozptyl $D[X] = \delta^2$.

Příklad

Životnost závodního automobilu: Jaká je pravděpodobnost, že závodní automobil dokončí tříhodinový závod, jestliže jeho střední životnost na plný výkon je 2 hodiny a 15 minut?

Hp je $f(x) = \frac{1}{2.25}\mathrm{e}^{-x/2.25}$.

$$
\begin{aligned}
P(X<3) &= \int_0^3 f(x)\,dx = \\
&= \int_0^3 \mathrm{e}^{-x/2.25}dx = \frac{1}{2.25}(-2.25)\left[\mathrm{e}^{-x/2.25}\right]_0^3 = 1-\mathrm{e}^{-3/2.25} = 0.736
\end{aligned}
$$

3.2.5 χ^2rozdělení $\chi^2(n)$

χ^2 rozdělení je odvozeno ze součtu kvadrátu nezávislých náhodných veličin s normovaným normálním rozdělením[6], tedy

$$\chi^2(n) = \sum_{i=1}^{n} (N_i(0,1))^2$$

kde n je počet stupňů volnosti. Se zvyšujícím se počtem stupňů volnosti se hustota χ^2 blíží k normovanému normálnímu rozdělení $N(0,1)$.

Hustota χ^2 rozdělení je pro hodnoty $x \leq 0$ je nulová. Zápis hustoty pravděpodobnosti pro hodnoty $x > 0$ je komplikovaný a využívá se pro něj mmj. Γ (gamma) funkce.

$X \sim \chi^2(n)$ má:

- střední hodnotu $E[X] = n$,
- rozptyl $D[X] = 2n$.

3.2.6 Studentovo t-rozdělení $t(n)$

Další spojité rozdělení, které je odvozeno od normovaného normálního rozdělení. t-rozdělení získáme jako podíl dvou veličin, jedné s normovaným normálním rozdělením a druhé s χ^2rozdělením, tedy

$$t(n) = \frac{N(0,1)}{\chi^2(n)/n}$$

Důležitá vlastnost t-rozdělení je ta, že se vrůstajícím počtem stupňů volnosti n se studentovo t-rozdělení blíží normovanému rozdělení (souvisí to s chováním χ^2 rozdělení). Při nižším počtu stupňů volnosti se podobá normovanému normálními rozdělení, ale s těžšími konci, tzn. $D[X] > 1$.

$X \sim t(n)$ má:

- střední hodnotu $E[X] = 0$,
- rozptyl $D[X] = \frac{n}{n-2}$.

3.2.7 F rozdělení $F(m,n)$

F rozdělení neboli Fisherovo rozdělení je spojité rozdělení pro podíl dvou nezáporných veličin. U obou veličin se předpokládá χ^2 rozdělení se stupni volnosti m a n a platí tedy

$$F(m,n) = \frac{\chi_1^2(m)/m}{\chi_2^2(n)/n}$$

$X \sim F(m,n)$ má:

- střední hodnotu $E[X] = \nu$,
- rozptyl $D[X] = 2\nu$.

F rozdělení popisuje veličinu ve tvaru podílu dvou rozptylů. Typicky se jedná o rozptyl vysvětlený a nevysvětlený při zkoumání závislosti dvou nebo více veličin.

[6] Toto rozdělení doprovází veličiny, které jsou tvořeny kvadráty - rozptyl, kriterium ve tvaru součtu čtverců ...