ZÁPADOČESKÁ UNIVERZITA V PLZNI
FAKULTA PEDAGOGICKÁ

KATEDRA MATEMATIKY, FYZIKY A TECHNICKÉ VÝCHOVY

NORMÁLNÍ ROZDĚLENÍ A ROZDĚLENÍ Z NĚHO ODVOZENÁ

BAKALÁŘSKÁ PRÁCE

Pavel Staněk

Přírodovědná studia, Matematická Studia

Vedoucí práce: RNDr. Václav Kohout

Plzeň, 2013

Prohlašuji, že jsem bakalářskou práci vypracoval samostatně s použitím uvedené literatury a zdrojů informací.

Plzeň, 28. 6. 2013

……………………………………..

vlastnoruční podpis

Poděkování

Tímto bych chtěl poděkovat svému vedoucímu bakalářské práce RNDr. Václavu Kohoutovi za odborné vedení a cenné rady. Dále bych chtěl poděkovat své rodině za podporu a trpělivost.

Obsah

1 ÚVOD

Jako téma své bakalářské práce jsem si vybral Normální rozdělení a rozdělení z něho odvozená, s kterými jsem se setkával během celého svého studia, a možností jejich aplikace.

Normální rozdělení patří mezi jedno z nejdůležitějších a nejvíce používaných spojitých rozdělení. Někdy se též nazývá Laplaceovo-Gaussovo rozdělení podle německého vědce C. F. Gausse a francouzského vědce Pierre-Simon Laplaceho.

Normální rozdělení hraje ze spojitých rozdělení největší roli v teorii pravděpodobnosti a statistice a řídí se jím (alespoň "přibližně") více náhodných veličin (proměnné, jejichž hodnoty jsou jednoznačně určeny výsledkem náhodného pokusu). Jako příklad můžeme uvést velikost náhodných chyb vznikajících při jakékoliv činnosti (např. při výrobě součástek odchylky v rozměrech od předem stanovené normy), dále u většiny měřitelných charakteristik biologických statistických jednotek (u hospodářských zvířat, při pokusných buněčných kulturách, ale také u lidí, aj.). Vezmeme-li to úplně konkrétně, normální rozdělení nalezneme např. u hmotnosti, IQ, výšky, aj.

Velmi dobrým příkladem tohoto rozdělení je rozdělení náhodných chyb, které vznikají při měření veličin. Budeme-li opakovat měření stejné veličiny při shodných podmínkách jako napoprvé, tak způsobují náhodné vlivy odchylky od skutečné hodnoty naměřené veličiny. Proto můžeme říci, že normální rozdělení je dobrým pravděpodobnostním modelem, když na kolísání náhodné veličiny působí větší počet malých na sobě nezávislých vlivů. Díky tomu se normálnímu rozdělení může občas říkat rozdělení chyb. Normální rozdělení se může využívat k aproximaci některých rozdělení (Binomického, Poissonova, aj.). Normální rozdělení může být také velice dobře využito ve fyzikálních a technických veličinách.

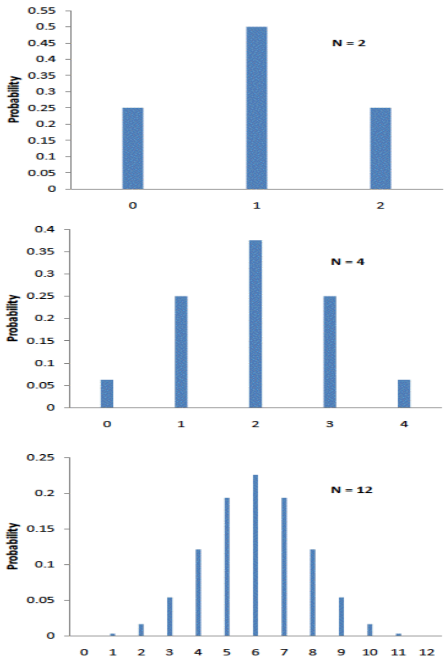
Vzhledem k tématu práce jsem její první část věnoval historii vzniku normálního rozdělení a významných matematiků.

2 HISTORIE NORMÁLNÍHO ROZDĚLENÍ A VÝZNAMNÍ MATEMATICI

V této kapitole si rozebereme historii normálního rozdělení a významných matematiků, kteří se normálním rozdělením během svého života zabývali. Informace byly čerpány z literatury [7], [8], [9], [10].

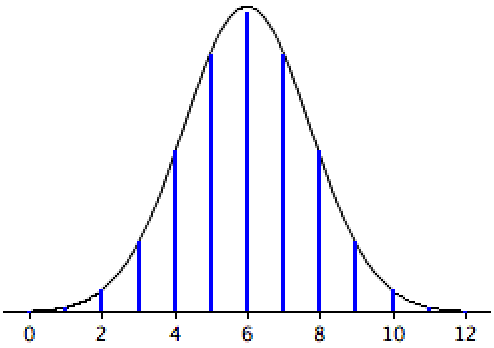
2.1 HISTORIE

Abraham de Moivre byl v 18. století často nabádán, zda by nebyl schopen zkrátit dlouhé výpočty daných matematických úkolů své doby. Uvedeme si zde příklad výpočtu pravděpodobnosti - pokud 100 krát hodíme mincí, jaká je pravděpodobnost, že padne 40 krát nebo vícekrát hlava. Daný úkol by se dal vypočítat pomocí následujícího vzorce $P(x) = \frac{N!}{x!(N-x)}\pi^x(1-\pi)^{N-x}$ kde x je počet hozených hlav (40), N je počet hodů (100), a π je pravděpodobnost, že padne hlava (0,5). Takto bychom museli vždy spočítat pravděpodobnost, že padne hlava 40 krát, poté 41 krát, atd. a sečíst všechny tyto pravděpodobnosti dohromady. Dříve možnost počítačů a kalkulaček nebyla, a proto se obraceli právě na Moivra. Ten poznamenal, že jestli zvýší číslo hodu, tvar binomického rozdělení se přiblíží k hladké křivce.



Obrázek 1: Zobrazení binomického rozdělení pro hodnoty 2, 4, 12

Jakmile byl Moivre schopen najít matematické vyjádření pro křivku, začal se domnívat, že bude schopen vyřešit, jaká je pravděpodobnost, že padne 40 krát nebo vícekrát hlava ze 100 hodů, což se mu povedlo.



Obrázek 2: Aproximace binomického rozdělení normálním rozdělením

Na obrázku 2 si můžeme všimnout aproximace binomického rozdělení rozdělením normálním pro 12 hodů mincí. Zde je patrné, že aproximované binomické rozdělení v modré barvě je srovnatelné s křivkou. Důležitost právě vyvozené křivky spočívá v tom, že mnoho přírodních jevů (viz. výše) se řídí pomocí normálního rozdělení.

Jedno z prvních využití normálního rozdělení byla analýza chyb vzniklých při astronomických měřeních. Tyto chyby mohly vzniknout pomocí nedokonalých přístrojů nebo špatných pozorovatelů. V 17. století Galileo Galilei poznamenal, že tyto chyby jsou symetrické, a navíc se malé chyby vyskytovali častěji než velké chyby. Tato poznámka vedla k několika hypotézám šíření chyb v 19. století kdy, došlo k objevu, že tyto chyby můžeme chápat jako normálního rozdělení. Poté nezávisle na sobě vědci a matematici vytvořili vzorec pro normální rozdělení. Adrian vzorec vytvořil v roce 1808 a Gauss v roce 1809. Díky tomuto vzorci poukázali na dané chyby (viz. následující kapitoly) zapadající do normálního rozdělení. Stejné rozdělení bylo popsáno už v roce 1778 Laplacem, kdy odvodil také velmi důležitou centrální limitní větu. Laplace v podstatě dokázal, že i když není rozdělení normálně distribuované (nejde nahradit normálním rozdělením), opakováním se přiblíží k normálnímu rozdělení, a že čím větší počet opakování, tím více by se to mělo blížit normálnímu rozdělení. Byl také první, kdo aplikoval normální rozdělení na lidské vlastnosti. Poznamenal, že charakteristiky jako výška, váha a síla se řídí normálním rozdělením.

2.2 VÝZNAMNÍ MATEMATICI

V této části kapitoly jsou v krátkosti popsány životy matematiků, kteří se normálním rozdělením zabývali během svého života.

2.2.1 ABRAHAM DE MOIVRE

Abraham de Moivre (1667-1754) se narodil ve Vitře, která se nachází mezi Paříží a Nancy. Jeho otec pracoval jako chirurg v Nancy. Jejich rodina nepatřila mezi bohaté, ale díky stálému příjmu vedli spokojený život. Jejich celá rodina byla protestantského vyznání, přesto šel Abraham nejdříve na katolickou školu, která byla tolerantní vůči jinému vyznání zejména kvůli náboženským problémům ve Francii. V 11 letech byl poslán na protestantskou školu v Sedanu, kde se učil řecky.

Tato škola byla pro náboženské problémy uzavřena, proto musel jít Moivre studovat na Saumur, kde studoval do roku 1684. Na této škole nebyla matematika součástí oboru, který studoval. Přesto svůj veškerý volný čas trávil čtením matematických článků, matematika byla jeho zálibou. V té době se jeho rodiče rozhodli přestěhovat do Paříže. Rozhodl se odejít s nimi a v jeho studiích pokračoval na Harcourtu. Zde poprvé studoval samostatnou matematiku.

Poté co Ludvík XIV. zrušil edikt nantský, byli protestanti pronásledováni, což vedlo k vyhnání Hugenotů a Moivre byl vězněn za jeho náboženské vyznání. Po propuštění odcestoval do Anglie, kde se stal soukromým učitelem.

Základní matematické texty si velmi dobře osvojil, ale poté co četl Newtonovo *Principia* si uvědomil, že bude muset matematiku studovat mnohem hlouběji. Rozhodl se, že se bude tomuto mistrovskému dílu snažit porozumět, což se mu také nakonec podařilo. Mezitím se ucházel o místo matematika v Anglii, ale díky svému vyznání byl stále diskriminován. Později se setkal s Newtonem, a stali se z nich přátelé. Jeho první dokument byl vydán v roce 1695 pod názvem *Method of fluxions*. Moivre byl průkopníkem v analytické geometrii a teorii pravděpodobnosti. Publikoval *The Doctrine of Chance: A methodof calculating the probabilities of events in play* v roce 1718, její latinská verze byla publikovaná v nejvyšších vědeckých kruzích. Francis Robartes navrhl, aby Moivre představil širší pojetí teorie pravděpodobnosti, než kterou uvedl Montmort v *Essay d´analyse sur les jeux de hazard*. Montmort byl ve sporu s Moivrem kvůli napadení svého díla v Moivrově díle. Později se *The Doctrine of Chance* objevila v rozšířených verzích, a to v letech 1718, 1738 a v roce 1756. Vydání v roce 1756 je velice důležité, protože v tomto vydání Moivre přibližuje binomické rozdělení k normálnímu rozdělení v případě velkého počtu pokusů. Moivre také pracoval na statistice úmrtnosti a základech teorie rent. V *Miscellanea Analytica*(1730) popsal Stirlingovu formuli, která je chybně připisovaná Stirlingovi. Formuli v roce 1733 použije k popsání normální křivky aproximací binomického rozdělení při velkém počtu hodnot. V druhém vydání chválí Stirlinga, který vylepšil jeho formuli. Moivre je také připomínán za sepsání formule $(cosx + isinx)^n$, která převádí trigonometrii do analýzy a je také důležitá k brzkému vývoji komplexních čísel. Navzdory jeho dobrému vědeckému postavení a snaze získat křeslo profesora na Cambridge, jeho hlavní příjem byl ze soukromých matematických hodin. Díky tomu také později zemřel v chudobě. Ke křeslu profesora mu nepomohli ani jeho nejlepší přátelé jako Newton a Halley, kteří se o to snažili. Velice zajímavá je jeho smrt. Moivre si ji sám předpověděl. Zjistil, že každý den spí o 15 minut více a sečtením aritmetických progresí a jejich vypočtením zjistil, že zemře v den, kdy bude spát 24 hodin a tak se i stalo.

2.2.2 PIERRE-SIMON LAPLACE

Pierre-Simon Laplace (1749-1827) pocházel z bohaté rodiny. Jeho otec byl zámožný podnikatel a matka byla ze zemědělské rodiny, která vlastnila pozemky okolo Tourgéville. Laplace nejdříve navštěvoval benediktínskou školu Beaumont-en-Auge, protože jeho otec byl toho názoru, že udělá kariéru v církvi. V 16 letech vstoupil na univerzitu v Caen. Zde si zapsal teologii, ale během prvních 2 let si oblíbil matematiku, na kterou měl talent. Zásluhu na tom zjištění měli jeho dva učitelé na univerzitě C. Gadbled a P. Le Canu, kteří si uvědomili jeho matematický potenciál.

Jakmile si uvědomil, že chce dále pracovat s matematikou, odchází v 19 letech do Paříže, aniž by ukončil studium. Zde se setkal s d´Alembertem, kterého hned zaujal, a který se mu snažil pomoci sehnat práci. Krátce poté nastupuje jako učitel matematiky na École Militaire.

Velice brzy začal vydávat zajímavé matematické dokumenty. První z těchto dokumentů byl předložen Akademii věd v Paříži 28. března 1770, popisoval v něm minima a maxima křivek, a zlepšil metody vypracované Lagrangem. Jeho další dokumenty pro Akademii věd na sebe nenechaly dlouho čekat. Například 18. července 1770 předložil dokument o diferenčních rovnicích. Důležité je podotknout, že tyto dokumenty byly určené jen pro čtení a ne pro tisk.

Jeho první dokument pro tisk se představuje v roce 1771, byl přeložen do latiny a vydán v Lipsku pod názvem *Nova acta eruditorum*. O 6 let později publikuje opravenou verzi a za chyby se omlouvá, ale říká, že za ně může tiskárna.

V roce 1771 se také poprvé snaží dostat do Akademie věd, ale upřednostněn byl Vandermonde. O rok později to zkouší znovu, ale tentokrát byl upřednostněn Cousin. Zklamán byl Laplace i jeho velký příznivce D´Alembert. Ten proto 1. ledna 1773 napsal Lagrangovi, zda nemůže být zvolen Laplace v Berlíně v Akademii věd a zda by tam také nemohl pracovat. Předtím než mohl Lagrange zareagovat a odpovědět, naskytla se pro Laplaceho další příležitost a 31. března 1773 byl zvolen do Akademie věd.

Laplace se v průběhu života zabýval 2 hlavními tématy (matematická astronomie a teorie pravděpodobnosti). Mezitím jeho pověst dále a dále rostla a v 80 letech byl považován za největšího vědce, kterého svět viděl. Jak ale rostla pověst, zhoršovaly se vztahy s jeho kolegy, protože neuznával jejich práci.

V roce 1784 byl jmenován zkoušejícím Royal Artillery Corps, kde roku 1785 zkoušel dokonce Napoleona Bonaparte. Tato práce ho velmi zaměstnávala, ale díky ní se znal s lidmi na důležitých místech.

V roce 1785 Lagrange opustil Berlín a přidal se k Laplaceovi v Paříži, jenž byl mezitím povýšen na hlavní pozici v Akademii věd. 15. Května 1788 se oženil s Marií Charlotte de Courty de Romanges. Společně měli 2 děti. Z důvodu revoluce roku 1793 opustil Paříž, kam se nechtěl vracet, aby se vyhnul gilotině.

V roce 1796 představil svoji slavnou hypotézu mlhovin v knize *Exposition du systeme du monde.* Tato kniha je rozdělena na 5 částí. V první části píše o pohybu nebeských těles, pohybu moře a také o atmosférické refrakci, druhá část popisuje skutečný pohyb nebeských těles, třetí část o síle a hybnosti, čtvrtá část o teorii univerzální gravitace a v páté části se objevuje jeho slavná hypotéza mlhovin.

Napoleon ve svých pamětech napsal, že Laplace byl členem senátu, kancléřem a poté obdržel i Řád čestné legie, ale i tak ho z úřadu ministra vnitra odstranil po 6 týdnech. V roce 1806 se stal hrabětem a v roce 1817 byl jmenován markýzem po obnovení Bourbonů.

Hlavním jeho dílem bylo *Théorie analytique des Probabilités.* První vydání vyšlo v roce 1812. Práce se skládala ze dvou částí. První část obsahuje generování funkcí a aproximaci různých výrazů v teorii pravděpodobnosti. Druhá část obsahuje definici teorie pravděpodobnosti a také Bayesovo pravidlo. Pozdější vydání knihy bylo doplněno o aplikování pravděpodobnosti, chyb v měření aj. Na tomto díle pracoval hlavně mezi lety 1817-1822.

Postupně jeho vliv upadá a opouští ho i jeho přátelé. Když roku 1826 odmítl podepsat dokument Francouzské akademie věd, přišel také o své poslední přátele, které měl v politice.

Laplace zemřel ráno 5. 3. 1827. V tento den Akademie věd zrušila zasedání z důvodu úcty k němu. Všichni ho totiž považovali za jednoho z největších vědců všech dob.

2.2.3 JOHANN CARL FRIEDRICH GAUSS

Johann Carl Friedrich Gauss (1777-1855) začal v 7 letech chodit na základní školu, kde jeho nadání objevili učitelé Büttner a asistent Martin Bartels . Velmi je zaujalo, jak rychle Johann spočetl součet čísel od 1 do 100. Využil toho, že vytvořil 50 dvojic. Poté byl výpočet velmi jednoduchý, protože věděl, že součet dvou čísel je vždy 101 (1+100,2+99,…). V roce 1788 začal studovat na gymnáziu. Později obdržel studijní stipendium a v roce 1792 nastoupil na Brunswick Collegium Carolinum. V roce 1795 odešel studovat na Göttingen University, kterou opustil v roce 1798 bez diplomu. Zde i tak přišel na jeden ze svých největších objevů, čímž je konstrukce 17-ti úhelníka pouze pomocí pravítka a kompasu. V této oblasti to byl největší pokrok od dob řecké matematiky. Poté se vrátil zpět na Brunswick, kde v roce 1799 obdržel titul. Svojí disertační práci předložil na Univerzitě v Helmstedtu. Téma jeho práce bylo Základní věty algebry.

Díky svému stipendiu nemusel hledat práci a věnoval se výzkumu. Olbers mu nabídl, aby se stal ředitelem nové hvězdárny v Göttingenu, ale nedošlo k tomu. Mezitím si začal dopisovat s Besselem a Sophií Germain. Roku 1805 se oženil s Johannou Ostoff. V roce 1807 opouští Brunswick a nastupuje jako ředitel hvězdárny v Göttingenu. Roku 1808 mu umírá otec a později tohoto roku i jeho žena při porodu jejich druhého dítěte. O rok později se Gauss oženil podruhé s Minnou. Společně měli 3 děti.

V roce 1809 publikuje svoji druhou knihu *Theoria motus corporum coelestium in sectionibus conicis Solem ambientium*. Tato kniha je rozdělena na dvě hlavní části. První část pojednává o diferenciálních rovnicích, řezech kuželem a eliptických orbitách. V druhé části ukázal, jak odhadnout a poté i upřesnit odhad dané oběžné dráhy planety. Příspěvky k astronomii přestal psát v roce 1817.

Mnoho času Gauss vynaložil do výstavby nové hvězdárny, která byla s jeho pomocí dokončena v roce 1816. Během této doby přesto našel čas k publikování. Vydal například *Disquisitiones generales circa seriem infinitum,* ve které představuje hypergeometrické funkce. Dále *Methodus nova integralium valores per approximationem inveniendi,* která je praktickou esejí o přibližné integraci, či *Bestimmung der Genauigkeit der Beobachtungen* (diskuze o statistických odhadech).

Později byl inspirován geodetickými problémy a zaujatý teorií potenciálu. Začal se tedy více zajímat o geodézii. V této době vynalezl Gauss heliotrop, který pracoval na principu odrážení slunečních paprsků pomocí zrcadel a malých dalekohledům. Mezi lety 1820-1830 vydává kolem 70 dílčích prací.

Od počátku roku 1800 se začal zajímat o otázku existující ne-Eukleidovské geometrie. O tomto tématu si psal na dálku s Farkasem Bolyaiem, v dopisech se Schumacherem a Gerlingem. V recenzi knihy v roce 1816 diskutuje nad důkazy, které poukazují na existenci různých axiomů od Eukleidovských axiomů. Z toho důvodu věřil v existenci ne-Eukleidovské geometrie.

Později se staral o svou nemocnou matku, u které byl po celou dobu její nemoci. V těchto letech se začal velmi zajímat o fyziku, kde spolupracoval s Weberem a společně vydali mnoho článků. Během 6 let společné spolupráce dosáhli mnoha úspěchů. Objevili Kirchhoffovy zákony, vynalezli telegraf, který byl schopen posílat zprávy na vzdálenost 5000 ft (1,524 kilometrů). Toto vše bylo pro Gausse jen zábavou. Hlavním Gaussovým úkolem bylo vytvořit celosvětovou síť magnetických pozorovacích bodů. V roce 1837 byl Weber nucen opustit Göttingen, protože byl zapleten do politických sporů a od té doby Gauss svoji činnost snižoval. Mezi lety 1845-1851 strávil Gauss spravováním a vylepšováním vdovského fondu na Göttingské univerzitě. Tato profese mu přidala zkušenosti ve finanční matematice. Od roku 1850 byly jeho práce pouze jen praktické. Dne 23. 2. 1855 zemřel.

3 TEORIE CHYB

Vypracováno s použitím literatury [12], [14].

Teorie chyb má za sebou poměrně značnou historii. Celkově se za moderního zakladatele teorie chyb považuje Jacob Bernoulli. Obecně totiž nemůžeme říct o žádné naměřené hodnotě, že byla změřena bez chyby s absolutní přesností. Bez chyby nemůžou zůstat ani přístroje. Chyby jsou způsobené nedokonalostí našich smyslů, nepřesností přístrojů, kterými měříme, nebo nemožností splnit úplně přesné podmínky pro dané měření (např. vlhkost, tlak). Dalšími příčinami jsou vnější vlivy (otřesy budov, apod.). Abychom snížili pravděpodobnost výskytu chyb, je důležité, vybrat správnou metodu a měřící přístroje.

O existenci chyb se můžeme přesvědčit, že stejné měření budeme opakovat vícekrát. Po větším počtu měření odhalíme, že jsme získali odlišné výsledky a nikoliv vždy stejné. Zjištěné výsledky se od sebe ve většině měření liší minimálně.

Chyby můžeme rozdělit do tří skupin podle původu jejich vzniku:

1) hrubé chyby (omyly) – u těchto chyb je hlavním faktorem nepozornost lidí při měření, zápisu výsledků nebo použitím špatně fungujících přístrojů aj. Výsledek, který vznikne tímto měřením, poznáme snadněji, protože se více liší od ostatních výsledků. K dosažení výsledků se hrubé chyby nepočítají, místo toho se udělá doplňující měření.
2) systematické chyby – vznikají při pokusu na používaných zařízeních nebo vnějšími vlivy. Vzniklé chyby můžou být způsobeny neúplností, nepřesností nebo nedokonalou metodou. Patří sem i chyby osobní, které jsou způsobeny schopnostmi pozorovatele (patří sem například zaokrouhlování, aj.). Můžeme je omezit, pokud použijeme různé metody k měření a různé přístroje.
3) náhodné chyby – vznikají z blíže neurčených příčin. Zjistíme je, pokud se po odstranění chyb hrubých a systematických budou po měření jednotlivé veličiny od sebe lišit a my nebudeme schopni určit příčinu, z jakého důvodu k tomu dochází. Jejich hodnoty mohou být kladné i záporné a různě veliké. Nejvíce oscilují kolem nuly. K těmto chybám může dojít po kolísání tlaku, teploty, aj. Přestože jejich příčinu nejsme schopni určit, pomocí různých matematických metod jsme schopni dané chyby zpracovat,

protože známe jejich charakteristické vlastnosti (velké chyby nejsou tak časté jako malé, při nekonečně velkém počtu měření je u nahodilých chyb součet roven nule).

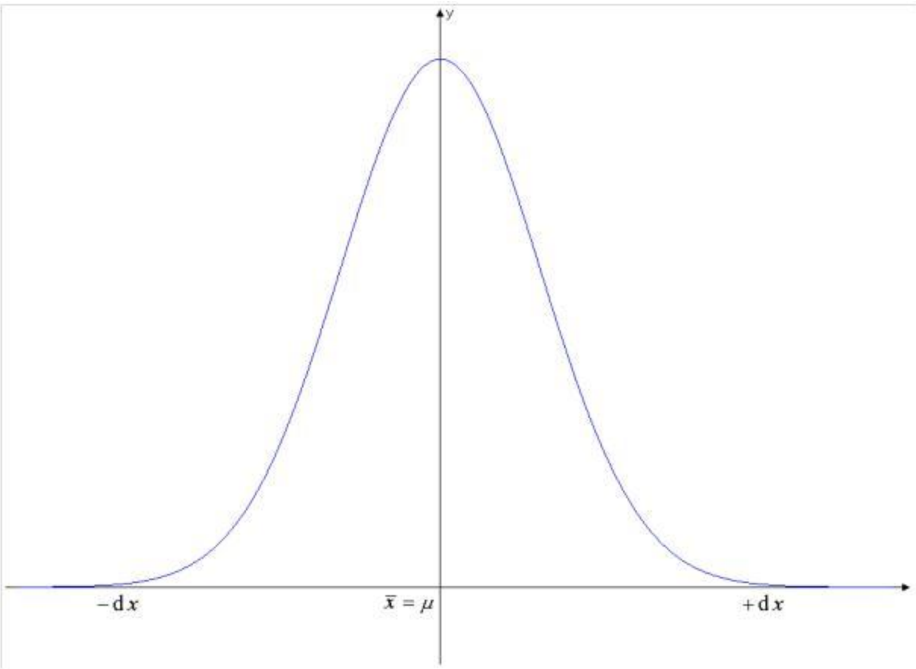
Abychom mohli zpracovat výsledky měření pomocí matematické statistiky, je důležité mít statistický soubor naměřených hodnot. V daném případě je to počet měření a lépe je, pokud jich je co nejvíce.

Velmi dobře můžeme výsledky měření ukázat na grafu. Na vodorovné ose budeme umisťovat úsečky, které odpovídají danému výsledku měření. Na svislé ose budeme nanášet, kolikrát se daný výsledek v měřeních objevil. Výsledek specifického n měření dané veličiny X bude vždy jiný, ale celkově se budou výsledky nejvíce uskupovat kolem střední hodnoty $\bar{x}$. Poté jestli x_n je n-tá naměřená hodnota, tak bude pro $\bar{x}$ platit

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (1),$$

číslo $\bar{x}$ se bude poté moci nazývat aritmetický průměr ze všech hodnot, které byly naměřeny.

Velice podstatné je, že takto určený aritmetický průměr (viz. 1) není úplně přesný. Dáno tím, že každé další měření nám pozmění výsledek vypočteného $\bar{x}$. Při nekonečně mnoho měření bude aritmetický průměr roven skutečné hodnotě $\bar{x} = \mu$. Při velkém počtu pokusů a jejich měření dostaneme danou křivku (viz. Obrázek 3)



Obrázek 3: Gaussova křivka

Obrázek 3 zobrazuje křivku $p(x)$, kde na vodorovnou osu nanášíme naměřené hodnoty x na svislé ose y nanášíme počet měření, kde bylo vždy výsledkem x. Vyjádřením této křivky se zabýval Gauss a došel ke vztahu:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \qquad (2)$$

Střední hodnota (μ) náhodné veličiny X určuje vrchol Gaussovi křivky. Rozptyl (σ^2) určuje, jaký tvar bude křivka mít. Problémem může být, jak nejlépe proložit Gaussovu křivku naměřenými hodnotami. K tomuto úkolu je podstatné znát parametry μ a σ^2, které zjistíme pomocí metody maximální věrohodnosti. Pro jejich zjištění hledáme rozdělení $p(x)$, pro které zjištěné hodnoty $\{x_1, \dots, x_n\}$ je nejvíce pravděpodobné. Určíme si L jako míru pravděpodobnosti, pro kterou bude platit:

$$L = p(x_1).\,p(x_2).\dots.\,p(x_n) = \prod_{i=1}^{n} p(x_i) \qquad (3)$$

V tomto konkrétním případě bude $p(x_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x_i-\mu)^2}{2\sigma^2}\right\}$. Proto můžeme napsat, že $L = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$. Funkce L je tedy funkcí dvou proměnných. Určení μ a σ^2, aby funkce L mohla být maximální, provedeme pomocí matematické analýzy.

$$\frac{\partial L(\mu,\sigma)}{\partial \mu} = 0, \frac{\partial L(\mu,\sigma)}{\partial \sigma^2} = 0 \qquad (4)$$

Jelikož L je funkcí neznámých parametrů, které jsou odhadovány, metoda maximální věrohodnosti je založena na získání takových hodnot, která maximalizují L. V praxi maximalizujeme $\ln L$ místo L, jelikož obě operace jsou ekvivalentní a dávají stejné výsledky. Funkce L a $\ln L$ jsou monotónní, z toho plyne, že mají i stejné extrémy.

$$\frac{\partial L}{\partial \mu} = \frac{\partial}{\partial \mu}[\ln L], \frac{\partial L}{\partial \sigma^2} n = \frac{\partial}{\partial \sigma^2}[\ln L] \qquad (5)$$

Pomocí úprav bychom poté dostali, že $L = -\frac{n}{2} ln[2\pi] - \frac{n}{2} ln[\sigma^2]$. Nyní budeme derivovat podle μ:

$$\frac{\partial L}{\partial \mu} = \frac{\partial}{\partial \mu}\left[-\frac{n}{2} ln[2\pi\sigma^2] - \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2\sigma^2}\right] = -\sum_{i=1}^{n} \frac{-(x_i - \mu)}{2\sigma^2} = \sum_{i=1}^{n} \frac{x_i - \mu}{2\sigma^2}$$

$$\frac{\partial L}{\partial \mu} = 0 \Rightarrow \sum_{i=1}^{n} \frac{x_i - \mu}{2\sigma^2} \Rightarrow \sum_{i=1}^{n} (x_i - \mu) = 0 \Rightarrow \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} \mu \Rightarrow n\mu = \sum_{i=1}^{n} x_i \Rightarrow$$

$$\mu = \frac{\sum_{i=1}^{n} x_i}{n}$$

Dále budeme derivovat podle σ^2 :

$$\frac{\partial L}{\partial \sigma^2} = \frac{\partial L}{\partial \sigma^2}\left[-\frac{n}{2} ln[2\pi] - \frac{n}{2} ln[\sigma^2] - \sum_{i=1}^{n} \frac{x_i - \mu}{2\sigma^2}\right] = 0 - \frac{n}{2} \times \frac{1}{2\sigma^2} + \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2\sigma^4}$$

$$= \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2\sigma^4} - \frac{n}{2\sigma^2}$$

$$\frac{\partial L}{\partial \sigma^2} = 0 \Rightarrow \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2\sigma^4} - \frac{n}{2\sigma^2} = 0 \Rightarrow \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2\sigma^2} - n = 0$$

$$\Rightarrow \sum_{i=1}^{n} (x_i - \mu)^2 = n\sigma^2 \Rightarrow \sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n}}$$

V tomto výrazu pro σ neznáme hodnotu μ . Proto naším dalším krokem bude vyjádření σ takové, že nebude funkcí μ. K tomuto úkolu si zavedeme daná čísla:

$$\varepsilon_k = x_k = \mu \qquad (6)$$

$$\Delta_k = x_k - \bar{x} \qquad (7)$$

$$\varepsilon_k - \Delta_k = \bar{x} - \mu \qquad (8)$$

Rovnost (8) vznikla pomocí porovnání výrazů (6) a (7). Nyní jsme schopni pokračovat a vyjádříme σ^2.

$$\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n} = \frac{\sum_{i=1}^{n} \varepsilon_i^2}{n} \qquad (9)$$

Poté úpravami sumy $\sum_{i=1}^{n} \varepsilon_i^2$ bychom dospěli postupně k výrazu $\sum_{i=1}^{n} \varepsilon_i^2 = \frac{n}{n-1} \sum_{i=1}^{n} \Delta_i^2$. Jelikož jsme si zavedli vztah (9), který platí, tak budeme moci pokračovat:

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \varepsilon_i^2} = \sqrt{\frac{1}{n} \frac{n}{n-1} \sum_{i=1}^{n} \Delta_i^2} = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}$$

Postupně jsme tedy dospěli k tomu, že hodnotám μ a σ, které jsou podstatné pro určení Gaussovy křivky, odpovídá aritmetický průměr a střední kvadratická chyba jednoho měření:

$$\mu = \frac{\sum_{i=1}^{n} x_i}{n}, \sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n}} \qquad (10)$$

Abychom popsali danou křivku co nejlépe, využijeme některé její charakteristiky. Mezi nejdůležitější patří poloha, šikmost, špičatost, a proměnlivost. Tyto charakteristiky nyní rozebereme:

1) charakteristika polohy – určuje střední hodnotu náhodných veličin. Zjištěné hodnoty kolem střední hodnoty kolísají. V tomto případě je to odhad skutečné hodnoty. Polohu charakterizuje medián (specifická hodnota, která rozdělí obor náhodné veličiny na dvě poloviny se stejnou pravděpodobnostní hodnotou). Další, co může charakterizovat polohu, je modus (veličina s největší četností výskytu) a aritmetický průměr, který je úhrnem hodnot veličiny v souboru dělený rozsahem souboru.
2) charakteristika špičatosti a šikmosti – tyto dvě charakteristiky ukazují, zda na měření neměl vliv faktor, který by způsobil asymetrii a zda se velké chyby neobjevují na úkor malých a opačně (vliv na špičatost). Zde je nejvíce používán koeficient šikmosti (tzv. třetí normovaný moment) $A = \mu_3(t) = \frac{\mu_3(x)}{\sigma^3}$. Tato charakteristika pro rozdělení, která jsou symetrická, je rovna nule. Koeficient špičatosti (čtvrtý normovaný moment) můžeme použít pro charakteristiku špičatosti $\mu_4(t) - 3 = \frac{\mu_4(x)}{\sigma^4} - 3$. U normálního rozdělení je roven nule. Pokud je koeficient kladný, dané rozdělení je oproti normálnímu rozdělení špičatější a naopak pokud je záporné, je více ploché.
3) charakteristika variability – ukazuje, jak se hromadí naměřené hodnoty kolem střední hodnoty (jestli jsou více nebo méně rozptýlené). Proměnlivost se charakterizuje pomocí střední kvadratické odchylky od dané střední hodnoty. Proměnlivost můžeme také

charakterizovat pomocí průměrné odchylky nebo také pravděpodobné odchylky od střední hodnoty.

4 NORMÁLNÍ ROZDĚLENÍ A ROZDĚLENÍ Z NĚHO ODVOZENÁ

Vypracováno s využitím literatury [1], [2], definice přejaty z literatury [21].

Pro pokračování si nejdříve zavedeme pojmy, které nyní budeme používat. Budeme uvažovat, že máme pravděpodobnostní prostor $\{\Omega, \mathcal{A}, P\}$. Ω je neprázdná množina všech výsledků náhodného pokusu, výsledky označujeme ω. $\mathcal{A}$ je σ-algebra sestrojená na Ω. $P: \mathcal{A} \rightarrow \langle 0,1 \rangle \subset \mathbb{R}$ je funkce přiřazující každé množině $A \in \mathcal{A}$ její pravděpodobnost.

Definice

Náhodnou veličinou rozumíme každé měřitelné zobrazení X z $\{\Omega, \mathcal{A}, P\}$ do $\mathbb{R}$.

Důležité si je uvědomit, že existují dva typy náhodných veličin. Jestliže množina možných výsledků náhodné veličiny obsahuje konečně mnoho hodnot, pak mluvíme o diskrétní náhodné veličině. Pokud množina možných výsledků náhodné veličiny obsahuje nespočetně mnoho hodnot (je interval), pak hovoříme o spojité náhodné veličině.

Definice

Funkci $F(x) : \mathbb{R} \rightarrow \mathbb{R}$ definovanou vztahem $F_X(x) = P(X \leq x)$ nazýváme **distribuční funkcí** náhodné veličiny X.

Důležité si je uvědomit, že existují dva typy náhodných veličin. Jestliže množina možných výsledků náhodné veličiny obsahuje konečně mnoho hodnot, pak mluvíme o diskrétní náhodné veličině. Pokud množina možných výsledků náhodné veličiny obsahuje nespočetně mnoho hodnot (je interval), pak hovoříme o spojité náhodné veličině.

Definice

Pro náhodnou veličinu X definujeme **hustotu pravděpodobnosti** $f(x)$ pomocí distribuční funkce $F_X(x)$ vztahem$f_X(x) = \frac{dF_X(x)}{dx}$ nebo v integrálním tvaru $F_X(x) = \int_{-\infty}^{x} f(t)\, dt$.

Distribuční funkce tedy přiřazuje každému reálnému číslu pravděpodobnost, že náhodná veličina nabude hodnoty, která je menší nebo rovna tomu číslu.

Než se zaměříme na normální rozdělení, popíšeme si zde pár základních rozdělení diskrétních náhodných veličin, s kterými budeme v následující části také pracovat:

Alternativní rozdělení $A(p)$ představuje úspěch nebo neúspěch pokusů s pravděpodobností $0 < p < 1$. Alternativní rozdělení nabývá pouze dvou hodnot (úspěch – 1, neúspěch – 0).

$$P(X = 1) = p, P(X = 0) = 1 - p.$$

$$E(X) = p, VAR(X) = p(1 - p)$$

Binomické rozdělení $\mathrm{Bi}(n, p)$ představuje počet úspěchů v n nezávislých pokusech, kdy pravděpodobnost úspěchu je $0 < p < 1$.

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, k = 1,2,3, \dots, n$$

$$E(X) = np, VAR(X) = np(1 - p)$$

Poissonovo rozdělení $Po(\lambda)\lambda > 0$ představuje počet událostí, která nastanou za určitý čas.

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

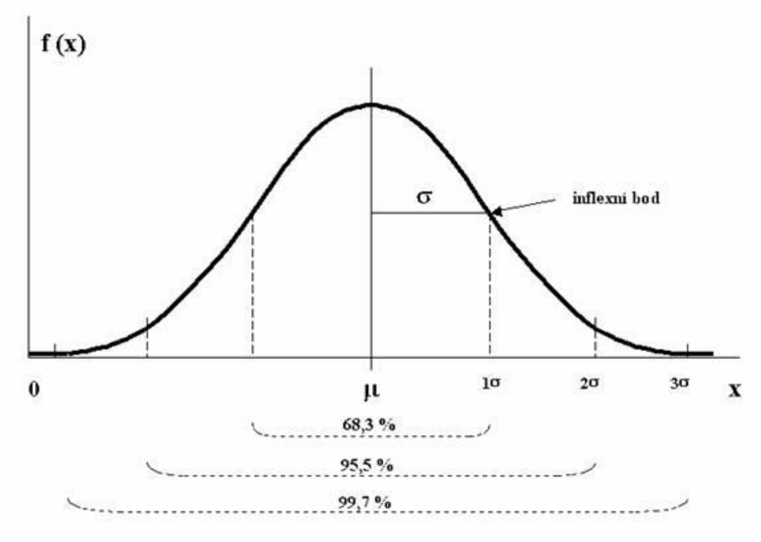
$$E(X) = \lambda, VAR(X) = \lambda$$

4.1 NORMÁLNÍ ROZDĚLENÍ

Normální rozdělení je určeno dvěma parametry. Prvním z nich je střední hodnota, která charakterizuje polohu (u tohoto konkrétního budeme značit μ a ne $E(X)$). Druhým parametrem je směrodatná odchylka (píšeme σ), která charakterizuje variabilitu náhodné veličiny X. Máme-li určeny dané parametry, můžeme náhodnou veličinu X, která má normální rozdělení, zapsat takto: $X \sim \mathrm{N}\,(\mu, \sigma^2), \mu \in \mathrm{R}, \sigma^2 > 0$. Potom může nabývat veličina X hodnot od $-\infty$ do $+\infty$.

Definice

Náhodná veličina X má normální rozdělení $\mathrm{N}\,(\mu, \sigma^2)$ právě tehdy, když má hustota pravděpodobnosti tvar: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}}$ pro $x \in (-\infty, +\infty)$



Obrázek 4: Graf hustoty normálního rozdělení

Grafem hustoty je Gaussova křivka (obr. 4). Z obrázku 4 si můžeme všimnout, že střední hodnota určuje, kde křivka dosahuje svého maxima. Směrodatná odchylka určuje, jak je roztažena do šířky. Hustota pravděpodobnosti nám však nic neříká o případné pravděpodobnosti. Abychom určili pravděpodobnost dané náhodné veličiny, je potřeba distribuční funkce, kterou získáme integrací hustoty pravděpodobnosti:

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} . \int_{-\infty}^{x} e^{\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}} dt, x \in (-\infty, +\infty) \qquad (11)$$

Normální rozdělení je velice důležité, protože se vyskytuje nejčastěji. Tímto rozdělením se řídí velké množství náhodných veličin (výšky, objemy hrudníku, váhy, chyby v měření). Často budeme moci aproximovat (nahradit) nespojitá i spojitá rozdělení pomocí N (μ, σ^2). Důležitá je symetrie kolem $x = \mu$, protože rozdělíme-li zadanou křivku přesně v bodě μ, získáme dvě totožné poloviny. Normální rozdělení i s rozděleními z něj odvozených se využívají při testování hypotéz, bodového odhadu, aj.

Výpočet E(X), VAR(X) :

$$E(X) = \int_{-\infty}^{+\infty} x.f(x)dx = \int_{-\infty}^{+\infty} x.\frac{1}{\sigma\sqrt{2\pi}} e^{\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}} dx = \left(substituce\ \ \frac{x-\mu}{\sigma} = t\right)$$

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

$$= \sigma.\int_{-\infty}^{+\infty} t.\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{t^2}{2}}.\sigma dt + \mu.\int_{-\infty}^{+\infty} .\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{t^2}{2}}.\sigma dt = 0 + \mu.1 = \mu$$

$$VAR(X) = \int_{-\infty}^{+\infty} (x-\mu)^2.f(x)dx = \int_{-\infty}^{+\infty} (x-\mu)^2.\frac{1}{\sigma\sqrt{2\pi}} e^{\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}} dx$$

$$= \left(substituce\ \frac{x-\mu}{\sigma} = t\right)$$

$$= \int_{-\infty}^{+\infty} \sigma.t^2.\frac{1}{\sqrt{2\pi}}.e^{-\frac{t^2}{2}}.\sigma dt = \sigma^2.\frac{1}{\sqrt{2\pi}}.\left(\lim_{n\to\infty}\left[-t.e^{-\frac{t^2}{2}}\right] + \int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} dt\right)$$

$$= \sigma^2.\frac{1}{\sqrt{2\pi}}.\left(0+\sqrt{2\pi}\right) = \sigma^2$$

Příklad 4.1.1

Máme danou náhodnou veličinu X, která je určena rozdělením $N(12{,}9)$. Jaká je pravděpodobnost, že X bude nabývat hodnoty a) menší než 13, b) větší jak 19, c) 11<X<17?

Řešení:

a) $$P(X < 13) = P(-\infty < X < 13) = F(13) - F(-) = F(13)$$

Určit, čemu se rovná distribuční funkce pro hodnotu 13 lze více způsoby. Prvním z nich je převést si normální rozdělení na normované normální rozdělení (viz. níže). Druhým způsobem je výpočet pomocí Excelu, kde je předdefinována funkce NORMDIST. Nyní použijeme druhou možnost.

$$P(X < 13) = F(13) = NORMDIST(13; 12; 3; 1) = 0{,}630559$$

Prvním číslem v závorce je hodnota distribuční funkce, kterou počítáme. Druhým parametrem je střední hodnota normálního rozdělení, třetím směrodatná odchylka určeného rozdělení a posledním je pravdivostní hodnota 1 (zadává se vždy, když chceme určit hodnotu distribuční funkce). Náhodná veličina X bude tedy menší než 13 s 63,0559% pravděpodobností.

b) $P(X > 19) = P(19 < X < \infty) = 1 - F(19) = 1 - NORMDIST(19; 12; 3; 1) = 0{,}009185$

Výsledná pravděpodobnost, že bude větší jak 19, je 0,9185%.

c) $P(11 < X < 17) = NORMDIST(17; 12; 3; 1) - NORMDIST(11; 12; 3; 1) = 0{,}582769$

Náhodná veličina bude v rozmezí 11 až 17 s pravděpodobností 58,2769%.

4.2 NORMOVANÉ NORMÁLNÍ ROZDĚLENÍ

Jedná se o modifikovaný případ normálního rozdělení, kdy bude $\mu = 0, \sigma^2 = 1$. Značí se U. Hodnoty normovaných veličin počítáme pomocí předpisu: $U = \frac{X-\mu}{\sigma}$.
Hustota pravděpodobnosti pro normované normální rozdělení:

$$f(u) = \frac{1}{\sqrt{2\pi}} . e^{-\frac{u^2}{2}} \qquad (12)$$

Pro výpočet distribuční funkce provedeme integraci hustoty:

$$\Phi = \frac{1}{\sqrt{2\pi}} . \int_{-\infty}^{u} e^{-\frac{t^2}{2}} dt \qquad (13)$$

Protože dané rozdělení je tabelováno, využívá se při výpočtu příkladů u normálního rozdělení. Postup při výpočtu, máme-li náhodnou veličinu, která má normální rozdělení je následující. Zjišťujeme její distribuční funkci $F(x) = P(X < x)$. Poté náhodnou veličinu X převedeme podle předpisu na normovanou normální veličinu:

$$F(x) = P(X < x) = P\left(\frac{X-\mu}{\sigma} < \frac{x-\mu}{\sigma}\right) = P(U < u) = F(u).$$

Důležitou vlastností, která vyplývá ze symetrie, pro normální a normované normální rozdělení je následující: hustota pravděpodobnosti se rovná hustotě pravděpodobnosti ze záporných hodnot hustoty dané veličiny.

$$f(x) = f(-x) \qquad (14)$$

Potom platí:

$$F(-x) = 1 - F(x) \qquad (15)$$

Pokud známe tento vztah (15), už není důležité znát všechny hodnoty pro hustotu a distribuční funkci, a budou stačit pouze kladné hodnoty. Konkrétně budeme moci psát:

$$P(X < -1) = 1 - P(X < 1) \qquad (16)$$

$$P(X < -1) = P(X > 1) \qquad (17)$$

Pro distribuční funkci:

$$F(-1) = 1 - F(1) \qquad (18)$$

Příklad 4.2.1

Máme danou náhodnou veličinu X s normálním rozdělením $N(12{,}16)$. Jaká je pravděpodobnost, že X bude nabývat hodnoty a) menší než 4, b) větší jak 13, c) 8<X<11?

Řešení:

a) Převedeme náhodnou veličinu X na náhodnou veličinu U, která bude mít normované normální rozdělení.

$$P(X < 4) = P\left(\frac{X - 12}{4} < \frac{4 - 12}{4}\right) = P(U < -2)$$

Nyní využijeme symetrie, kde platí vztah (15).

$$P(U < -2) = 1 - F(2)$$

Najdeme hodnotu distribuční funkce pro normované normální rozdělení v tabulkách.

$$1 - 0{,}97725 = 0{,}02275$$

Pravděpodobnost, že náhodná veličina bude menší než 4 je 2,275%.

b) $$P(X > 13) = 1 - P(X < 13) = 1 - F(13)$$

Zde musíme provést transformaci na normované normální rozdělení, protože hodnotu distribuční funkce pro normální rozdělení v tabulkách nenalezneme.

$$1 - P(X < 13) = 1 - P\left(\frac{X - 12}{4} < \frac{13 - 12}{4}\right) = 1 - P(U < \frac{1}{4})$$

Nyní najdeme hodnotu distribuční funkce pro $\mu = 0{,}25$ v tabulkách a vyjde nám:

$$1 - 0{,}59871 = 0{,}40129$$

Pravděpodobnost, že náhodná veličina bude větší než 13, je 40,129%.

c) Transformujeme opět na normované normální rozdělení:

$$P(8 < X < 11) = P\left(\frac{8-12}{4} < \frac{X-12}{4} < \frac{11-12}{4}\right) = P(-1 < X < -\frac{1}{4})$$

Nyní odečteme od sebe hodnoty distribučních funkcí a využijeme i vlastnosti, že se jedná o symetrickou funkci.

$$F\left(-\frac{1}{4}\right) - F(-1) = \left[1 - F\left(\frac{1}{4}\right) - (1 - F(1))\right] = F(1) - F\left(\frac{1}{4}\right) = 0{,}84134 - 0{,}59871$$
$$= 0{,}24263$$

Pravděpodobnost, že náhodná veličina X náleží do intervalu od 8 do 11 je 24,263%.

4.3 PEARSONOVO χ^2-ROZDĚLENÍ

χ^2 (čteme chí kvadrát) rozdělení je prvním z odvozených od normálního rozdělení. Má pouze jeden parametr v (počet stupňů volnosti). Máme-li v náhodných veličin $U_1, U_2, U_3, \ldots, U_v$, které jsou navzájem nezávislé (viz. pojmy výše) a každá z nich je normovaným normálním rozdělení $N(0,1)$, potom součet čtverců označujeme χ^2-rozdělení se stupněm volnosti v.

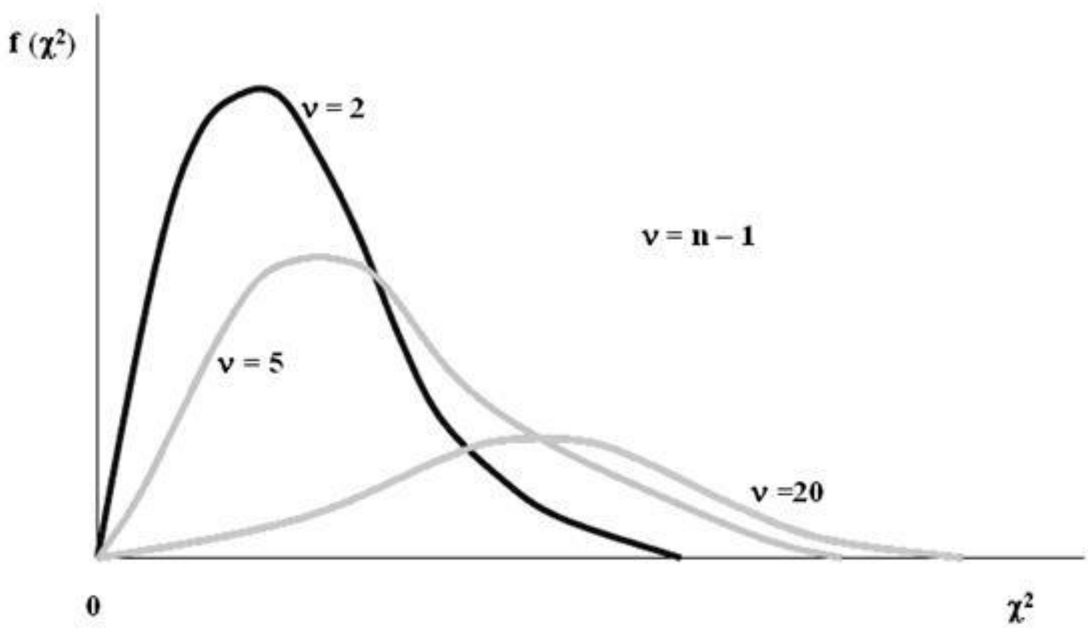
$$\chi^2 = \sum_{i=1}^{v} U_i^2 \qquad (19)$$

Zde si můžeme všimnout, že počet stupňů volnosti je dán počtem náhodných veličin, které sčítáme. Střední hodnota je rovna počtu stupňů volnosti a rozptyl je roven dvojnásobku počtu stupňů volnosti.

$$E(\chi^2) = v, D(\chi^2) = 2.v \qquad (20)$$

Hustota pravděpodobnosti pro χ^2-rozdělení:

$$f(x) = \frac{1}{2^{\frac{v}{2}}\Gamma\left(\frac{v}{2}\right)} e^{-\frac{x}{2}}(x)^{\frac{v}{2}-1} \qquad (21)$$



Obrázek 5: Graf hustoty χ^2 se stupni volnosti v.

Z obrázku 5 je vidět, že křivky hustot χ^2 jsou asymetrické. U malých výběrů je asymetrie značná. Při $v > 100$ budeme dané rozdělení považovat už za normované normální rozdělení. Hodnoty χ^2 jsou tabelovány v statistických tabulkách. Funkce Γ (Eulerův integrál) bude mít pro $p > 1$ dána předpis:

$$\Gamma(p) = \int_0^\infty x^{p-1}.e^{-x}dx \qquad (22)$$

Distribuční funkce:

$$F_v(x) = \frac{1}{2^{\frac{v}{2}}\Gamma\left(\frac{v}{2}\right)}\int_0^x e^{-\frac{x}{2}}(x)^{\frac{v}{2}-1}\,dx \qquad (23)$$

je tabelována. Místo ní mohou být tabelovány jen kvantily nebo kritické hodnoty. Kvantily χ_p^2 budeme chápat jako hodnoty, které splňují pro určené v vztah:

$$P\left(\chi^2 < \chi_p^2\right) = p, \qquad (24)$$

Kritické hodnoty χ^2_α naproti tomu budou splňovat vztah:

$$P(\chi^2 > \chi^2_\alpha) = \alpha = 1 - p, \qquad (25)$$

kdy α popisuje riziko nebo hladinu významnosti.

Větší využití má χ^2-rozdělení při statistických aplikacích:

a) pokud provádíme test dobré shody (určování náhodných veličin zda nepochází z některého konkrétního rozdělení).
b) k ověření nezávislosti náhodných veličin.
c) ke kontrole a ověření (na základě dat), zda dva a více výběrů budou homogenní k určité veličině. Příkladem mohou být náboženské názory a jejich určení, jestli jsou různé v odlišných oblastech.
d) máme-li dány náhodné veličiny $X_1, X_2, X_3, \ldots, X_v$, které budou mít rozdělení $N(\mu, \sigma^2)$, potom výběrový rozptyl (viz. 26) vydělený σ^2 a souběžně vynásobený $(v-1)$ bude mít rozdělení $\chi^2(v-1)$. Velice často je to potom využíváno v testování hypotéz.

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{x})^2}{n-1} \qquad (26)$$

4.4 STUDENTOVO ROZDĚLENÍ

Často se označuje jako t-rozdělení. William Sealy Gosset odvodil t-rozdělení, aby mohl s malým počtem vzorků získávat použitelné závěry příkladů a pokusů. Má jeden parametr v, který udává počet stupňů volnosti.

Máme danou náhodnou veličinu U s rozdělením $N(0,1)$ a náhodnou veličinu Z, která má rozdělení $\chi^2(v)$. Jestli jsou dané náhodné veličiny nezávisle, potom platí, že náhodná veličina

$$T = \frac{U}{\sqrt{\frac{\chi^2}{v}}} \qquad (27)$$

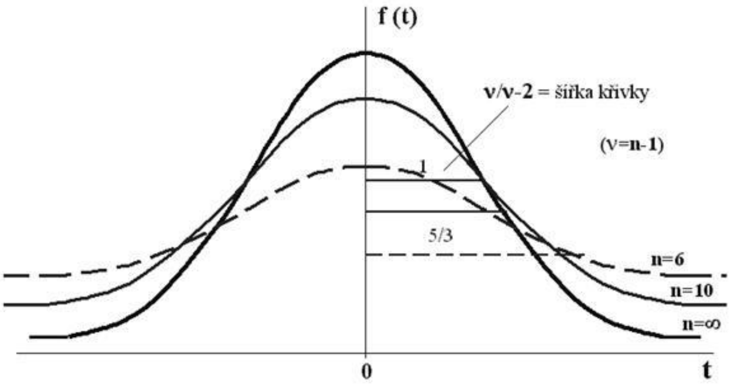
má t-rozdělení o v stupních volnosti. Stručně se značí t_v.

Střední hodnota a rozptyl jsou rovny:

$$E(X) = 0, D(X) = \frac{v}{v-2} \qquad (28)$$

Hustota pravděpodobnosti pro t-rozdělení:

$$f_v(t) = \frac{\Gamma(\frac{v+1)}{2}}{\Gamma\left(\frac{v}{2}\right)\sqrt{\pi k}}\left(1 + \frac{t^2}{v}\right)^{-\frac{v+1}{2}}, -\infty < t < \infty, v = 1, 2, \dots \qquad (29)$$



Obrázek 6: hustoty t-rozdělení

Na obrázku 6 je zobrazeno t-rozdělení pro různě velké soubory. Rozsah výběru určuje n, v udává počet stupňů volnosti. Z obrázku je patrné, že Studentovo rozdělení je symetrické s vrcholem v nule. Při zvětšování výběru se křivka zvyšuje a zužuje. Pro $v > 100$ lze nahradit rozdělením $N(0,1)$.

Distribuční funkce:

$$F_v(t) = \int_{-\infty}^{t} f_v(t)dt \qquad (30)$$

Využití t-rozdělení je mnoho. Uvedeme zde několik příkladů (více popsáno v kapitole s využitím):

a) Při intervalu spolehlivosti pro μ, když neznáme σ^2. Nechť $X_1, \dots, X_n$ je výběr z $N(\mu, \sigma^2)$, $n > 1, \sigma^2 > 0$. Naším úkolem je sestrojit oboustranný interval spolehlivosti pro μ na hladině $1 - \alpha$. Nyní využijeme věty: Nechť $X_1, \dots, X_n$ je výběr z $N(\mu, \sigma^2)$, kde

$n \geq 2$ a $\sigma^2 > 0$, náhodná veličina $T = \frac{\bar{X}-\mu}{S}\sqrt{n}$ má rozdělení t_{n-1}. Následně dostaneme $\mathrm{P}\{|\bar{X} - \mu|\sqrt{n}/S < t_{n-1}(\alpha)\} = 1 - \alpha$. Z rovnice dostaneme $\mathrm{P}\{\bar{X} - St_{n-1}(C)/\sqrt{n} < \mu < \bar{X} + St_{n-1}(\alpha)/\sqrt{n}\} = 1 - \alpha$. Námi hledaný interval spolehlivosti pro μ je tedy $\bar{X} - St_{n-1}(\alpha)/\sqrt{n}$, $\bar{X} + St_{n-1}(\alpha)/\sqrt{n}$

b) Jednovýběrový t test. Budiž $X_1, \dots, X_n$ výběr z $N(\mu, \sigma^2)$, kde $n > 1$. *Parametr* $\sigma^2 > 0$ je neznámý. Potřeba testovat hypotézu H_0: $\mu = \mu_0$ proti $H_1 : \mu \neq \mu_0$, kde μ_0 je dané číslo. Hypotézu H_0 budeme moci zamítnout, když bude X hodně vzdálené od čísla μ_0. H_0 tedy zamítneme na hladině α, když bude platit $|\bar{X} - \mu_0|\sqrt{n}/S \geq t_{n-1}(\alpha)$. Neplatí-li nerovnost, tak H_0 nezamítáme. S zjišťujeme z $S^2 = \frac{1}{n-1}(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2)$.

Příklad 4.4.1

Máme automat, který plní solí balíčky o hmotnosti 1kg. Při kontrole byly při přesném převážení 5 balíčků zjištěny odchylky (v gramech) od požadované hodnoty. Dané odchylky byly $-3, 2, -2, 0, -1$. Úkolem je zjistit, zda daný automat nemá systematickou odchylku od hodnoty, kterou vyžadujeme?

Řešení: jednotlivé odchylky budeme považovat za realizaci nezávislých náhodných veličin s rozdělením $N(\mu, \sigma^2)$, kde σ^2 je neznámá. Musíme otestovat hypotézu H_0: $\mu = 0$.

$$n = 5;\ \mu_0 = 0$$

$$\bar{X} = \frac{1}{n}\sum_{i=0}^{n} X_i = \frac{1}{5}(-3 + 2 - 2 + 0 - 1) = -0{,}8;$$

$$S^2 = \frac{1}{n-1}(\sum_{i=0}^{n} X_i^2 - n.X^2) =$$

$$= \frac{1}{4}\{[(-3)^2 + 2^2 + (-2)^2 + 0^2 + (-1)^2] - 5.(-0{,}8)^2\} = 3{,}7;$$

$$S = 1{,}9235;$$

$$T = \frac{\bar{X} - \mu_0}{S}\sqrt{n} = \frac{-0{,}8}{1{,}9235}\sqrt{5} = -0{,}9300$$

Pak lze porovnat $0{,}9300 < t_4(0{,}05) = 2{,}776$. Z toho můžeme vidět, že zjištěná data neodporují předpokladu, že daný automat nemá systematickou odchylku.

c) Párový t test. Budeme mít výběr$(Y_1, Z_1), \dots, (Y_n, Z_n)$, který bude náhodný, z daného dvojrozměrného rozdělení, a který bude mít vektor středních hodnot (μ_1, μ_2). Naším úkolem bude testovat hypotézu H_o: $\mu_1 - \mu_2 = \Delta$, kde bude Δ nějaké dané číslo. Potom máme nezávislé náhodné veličiny $X_1 = Y_1 - Z_1, \dots, X_n = Y_n - Z_n$. Budeme předpokládat, že dané veličiny mají normální rozdělení, pak je zřejmé, že $\mu_1 = \mu_2$. Splníme-li tyto předpoklady, bude se jednat o test $H_0: \mu = \Delta$ proti $H_1: \mu \neq \Delta$. Z toho se opět dostáváme k předchozímu využití, kterým je jednovýběrový t test. Hypotéza bude moci být zamítnuta na hladině (α), když$|\bar{X} - \Delta|\sqrt{n}/S \geq t_{n-1}(\alpha)$. Tento test je využit v případech, kdy jsou na každém z n objektů měřeny pouze dvě veličiny. Příkladem může být práh slyšitelnosti u lidí na různé frekvenci zvuku. Zjišťujeme, zda levé a pravé ucho mají stejnou citlivost.

Příklad 4.4.2

Naším úkolem bude určit, zda se obě přední pneumatiky u aut sjíždějí stejnou rychlostí. Máme 6 nových osobních automobilů, které mají správně seřízenou geometrii vozu. Po delší době se u nich zjistilo, o kolik mm se sjely jejich přední pneumatiky.

Číslo automobilu	1	2	3	4	5	6
Sjetí pravé pneumatiky	1,8	1,0	2,2	0,9	1,5	1,6
Sjetí levé pneumatiky	1,5	1,1	2,0	1,1	1,4	1,4
Rozdíl	0,3	-0,1	0,2	-0,2	0,1	0,2

Řešení: Rozdílem budeme rozumět realizaci nezávislých veličin s rozdělením $N(\mu, \sigma^2)$. Jestliže se u aut sjíždějí přední pneumatiky stejně, bude hypotéza $H_0: \mu = 0$.

$$n = 6;\ \Delta = 0;$$

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{6}(0{,}3 - 0{,}1 + 0{,}2 - 0{,}2 + 0{,}1 + 0{,}2) = 0{,}0833;$$

$$S^2 = \frac{1}{n-1}\left(\sum_{i=0}^{n} X_i^2 - n.\bar{X}^2\right) =$$

$$= \frac{1}{5}(0{,}3^2 + 0{,}1^2 + 0{,}2^2 + 0{,}2^2 + 0{,}1^2 + 0{,}2^2 - 6.\,0{,}0833^2) = 0.0377;$$

$$S = 0{,}1941;$$

$$T = \frac{\bar{X} - \mu_0}{S}\sqrt{n} = \frac{0.0833}{0{,}1941}\sqrt{6} = 1{,}0512.$$

Ze získaných dat nebudeme moci zamítnout hypotézu, že se u aut sjíždějí obě přední pneumatiky stejně, protože $1{,}051 < t_5(0{,}05) = 2{,}571$. K lepšímu a přesnějšímu určení bychom potřebovali větší výběr.

4.5 FISHEROVO-SNEDECOROVO ROZDĚLENÍ

Toto rozdělení se jinak také nazývá F-rozdělení. Máme dvě nezávislé veličiny $y_{1,}y_2$. Veličina y_1 má rozdělení χ^2 o v_1 stupních volnosti, veličina y_2 má χ^2 rozdělení o v_2 stupních volnosti, potom náhodná veličina

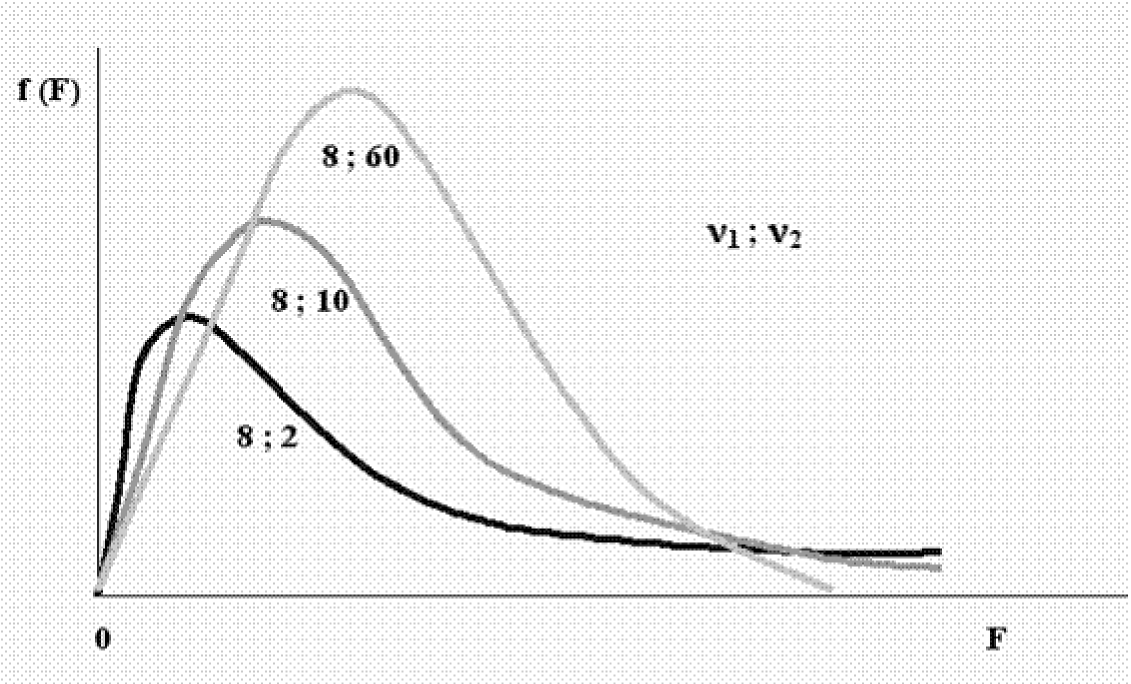
$$Z = \frac{\frac{y_{1,}}{v_1}}{\frac{y_2}{v_2}} \qquad (31)$$

má F-rozdělení, jehož hustota je rovna:

$$f_{v_1,v_2}(z) = \frac{\Gamma(\frac{v_1 + v_2)}{2}}{\Gamma\left(\frac{v_1}{2}\right).\Gamma\left(\frac{v_2}{2}\right)}.\left(\frac{v_1}{v_2}\right)^{\frac{v_1}{2}}.z^{\frac{v_1}{2}-1}.(1 + \frac{v_1}{v_2}.z)^{\frac{-(v_1+v_2)}{2}} \qquad (32)$$

Distribuční funkce:

$$F_{(v_1,v_2)}(z) = \int_0^z f_{v_1,v_2}(z)dz \qquad (33)$$



Obrázek 7: hustoty F-rozdělení o různých stupních volnosti a různém výběru

Je patrné (viz. obrázek 7), že křivka je asymetrická. Při menších výběrech je asymetrie větší.
Střední hodnota a rozptyl náhodné veličiny Z s F-rozdělením jsou rovny:

$$E(Z) = \frac{v_2}{v_2 - 2}, D(Z) = \frac{2v_2^2.(v_1 + v_2 - 2)}{v_1.(v_2 - 2)^2.(v_2 - 4)^2} \qquad (34)$$

Uplatnění F-rozdělení je velice široké. Základní využití test o shodnosti rozptylu dvou náhodných veličin, testy v regresní analýze a test o shodě středních hodnot pro více náhodných veličin.

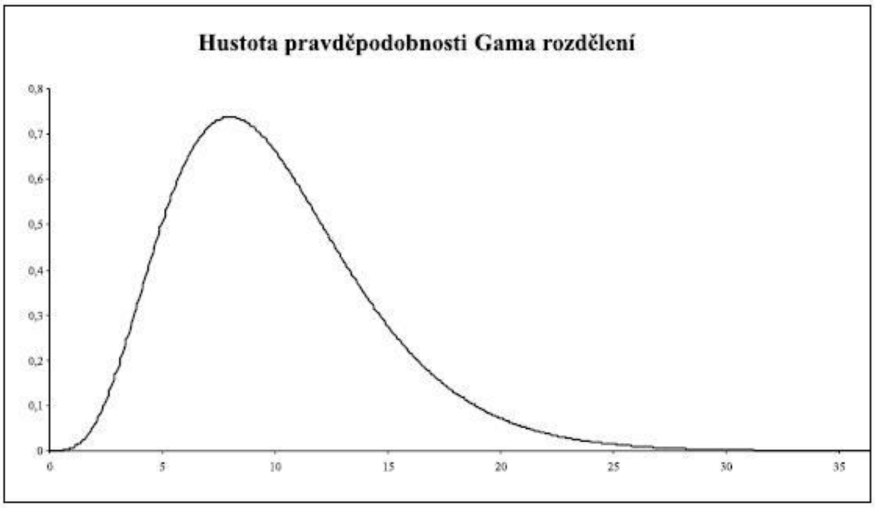
4.6 GAMA ROZDĚLENÍ

Gama rozdělení je závislé na dvou parametrech m a δ. Náhodná veličina X má gama rozdělení, když $m > 0,\ \delta > 0$ a její hustota pravděpodobnosti je rovna

$$f(x) = \frac{1}{\delta^m \Gamma(m)}.x^{m-1}e^{-\frac{x}{\delta}} \qquad x > 0, \qquad (35)$$

pak označíme veličinu, která má gama rozdělení:

$$X \sim \Gamma(m, \delta) \qquad (36)$$



Obrázek 8: Hustota pravděpodobnosti Gama rozdělení s parametry m=5,δ=2

Na obrázku 8 je zobrazena hustota pravděpodobnosti pro parametry $m = 5, \delta = 2$. Rozdělení $X \sim \Gamma(m, \delta)$ je asymetrické. Je-li $m < 1$, je hustota klesající funkcí. Využívá se v teorii spolehlivosti, při popsání velikosti částic disperzních fází kovových materiálů, aj. Distribuční funkci vyjádříme:

$$F(x) = \int_0^x f(t)dt \qquad (37)$$

Střední hodnota a rozptyl jsou rovny:

$$E(X) = m.\delta, \qquad VAR(X) = m.\delta^2 \qquad (38)$$

Zvláštním případem $\Gamma(m, \delta)$ je Erlangovo rozdělení, kdy se m rovná přirozeným číslům. Tento typ se využívá v teorii hromadné obsluhy. Příkladem uvedeme obsluhu při opravě nějakého stroje. Oprava se bude moci rozložit na n fází, která na sebe budou navazovat. Jednotlivé doby obsluhy v každé fázi jsou na sobě nezávislé náhodné veličiny, které mají Erlangovo rozdělení $Erlang(0, \delta)$, kdy $m = 1$ a parametr δ je stejný. Dalším příkladem je životnost n složek kde pracuje jen jedna a ostatní jsou v rezervě. Přestane-li tato jedna konkrétní pracovat, je nahrazena jinou, atd. Jestli jsou na sobě nezávislé a každá má rozdělení $Erlang(0, \delta)$se stejným parametrem δ, potom se jedná o Erlangovo rozdělení.

4.7 BETA ROZDĚLENÍ

Závislé na dvou parametrech m_1 a m_2. Náhodná veličina X má beta rozdělení s parametry $m_1 > 0$ a $m_2 > 0$, jestliže je její hustota pravděpodobnosti $f(x)$ rovna

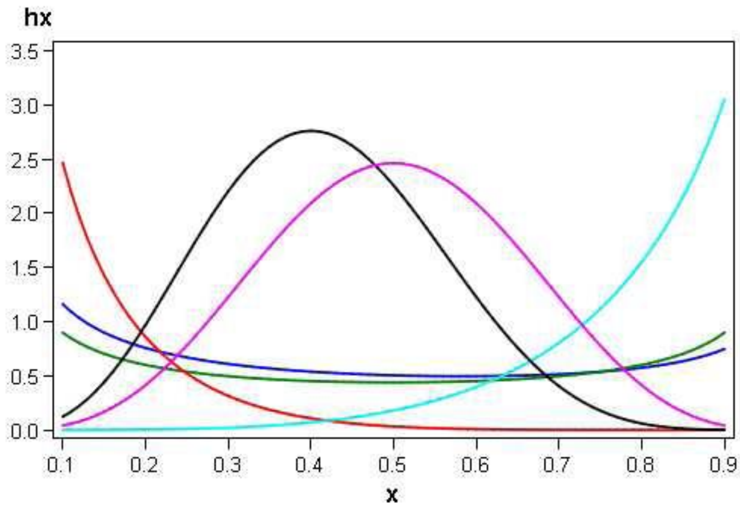
$$f(x) = \frac{1}{B(m_1, m_2)} . x^{m_1 - 1} . (1 - x)^{m_2 - 1}, \quad x \in (0,1), \qquad (39)$$

potom náhodnou veličinu X, která se řídí beta rozdělením, značíme:

$$X \sim B(m_1, m_2) \qquad (40)$$

Tvar hustoty pravděpodobnosti je závislý (viz Obrázek 9) na daných parametrech:

Hustota pravděpodobnosti	m_1	m_2	Tvar funkce
hx1	< 1	< 1	Tvar podobný písmenu U
hx2	$< 1 (m_1 = m_2)$	$< 1 (m_1 = m_2)$	Symetrie kolem bodu 0,5
hx3	< 1	> 1	Klesající
hx4	> 1	< 1	Rostoucí
hx5	> 1	> 1	Unimodální křivka
hx6	$> 1 (m_1 = m_2)$	$> 1 (m_1 = m_2)$	Symetrie kolem bodu 0,5



Obrázek 9: Různé typy grafů hustoty beta rozdělení

Obrázek 9 zobrazuje graf pro různě veliké hodnoty parametru, které jsou uvedeny v tabulce výše.

Střední hodnota a rozptyl jsou rovny:

$$E(X) = \frac{m_1}{m_1 + m_2}, \quad D(X) = \frac{m_1 . m_2}{(m_1 + m_2)^2 . (m_1 + m_2 + 1)} \qquad (41)$$

Beta rozdělení se využívá převážně v ekonomii.

4.7 KRITICKÉ HODNOTY

Kritické hodnoty vyjadřují hranici, kterou náhodná veličina překročí s pravděpodobností α. Může se jim říkat v některých případech také kvantil. Kritické hodnoty jsou nadefinované pro každé rozdělení jiným způsobem. Jejich hodnoty nalezneme v tabulkách.

Kritické hodnoty pro normální rozdělení $u(\alpha)$:

$$X \sim N(0,1), \quad P(X \geq u(\alpha) = 1 - \alpha$$

Kritické hodnoty pro Pearsonovo rozdělení $\chi_k^2(\alpha)$:

$$X \sim \chi_k^2, \quad P\left(X \geq \chi_k^2(\alpha)\right) = \alpha$$

Kritické hodnoty Studentova rozdělení$t_k(\alpha)$:

$$X \sim t_k(\alpha), \quad P\left(-t_k(\alpha) \leq X \leq t_k(\alpha)\right) = 1 - \alpha$$

Kritické hodnoty pro Fisherovo-Snedecorovo rozdělení $F_{v_1,v_2}(\alpha)$:

$$X \sim F_{v_1,v_2}(\alpha), \quad P\left(X \geq F_{v_1,v_2}(\alpha)\right) = \alpha$$

Kritické hodnoty pro Fisherovo-Snedecorovo rozdělení $F_{v_1,v_2}(\alpha)$ jsou tabelovány pro $0 < \alpha \leq 0{,}5$, když $0{,}5 < \alpha \leq 1$ počítáme kritické hodnoty ze vztahu:

$$F_{v_1,v_2}(\alpha) = \frac{1}{F_{v_1,v_2}(1-\alpha)}$$

5 VYUŽITÍ NORMÁLNÍHO ROZDĚLENÍ

V této kapitole rozebereme některé možnosti využití normálního rozdělení a rozdělení z něho odvozených. Vypsání všech možností využití, by bylo na velice obsáhlou knihu, a proto zde rozebereme jen některé.

5.1 TESTOVÁNÍ HYPOTÉZ

Vypracováno s využitím literatury [22], [24].

5.1.1 ÚVOD DO TESTOVÁNÍ HYPOTÉZ

U hypotéz se dělají určité závěry, které se týkají základního souboru. Závěrem nebude číslo, ale výrok (např. výsledkem nebude, že průměr leží s danou pravděpodobností v intervalu od 14 do 19, ale to zda je průměr základního souboru roven 16,5 s jistou pravděpodobností). Výrok má určitou spolehlivost (obsahuje určitou možnost chyby). Abychom mohli vyslovit nějakou hypotézu, musíme udělat náhodný výběr (např. 50 chlapců na základní škole) ze základního souboru (všichni chlapci na škole). Jakmile máme náhodný výběr určen, budeme moci z tohoto náhodného výběru dělat závěry, které se týkají parametrů základního souboru.

Budeme-li vyšetřovat nějaké normální rozdělení $N(\mu, \sigma^2)$, pak můžeme říci, že o samotné normalitě nejsou žádné pochybnosti. Aby se nám co nejjednodušeji pracovalo, můžeme předpokládat, že hodnota parametru $\sigma^2 > 0$ je nám známa a o parametru μ se domníváme, že by mohl být roven μ_0. Proto tvrzení $H_0: \mu = \mu_0$ nazveme nulovou hypotézou (nazývá se nulová, protože vyjadřuje žádný nebo nulový rozdíl mezi testovanými soubory), jelikož si tím nejsme jisti. Druhou možností je alternativní hypotéza (H_1), která může být oboustranná nebo jednostranná:

$$H_1: \mu \neq \mu_0 \qquad (42)$$

$$H_1: \mu > \mu_0 \qquad (43)$$

$$H_1: \mu < \mu_0 \qquad (44).$$

Dvoustrannou hypotézu (42) můžeme přijmout nebo nepřijmout, a to bude náš výsledek, který hledáme. Pravostranná hypotéza (43) nám řekne, že průměr základního souboru je větší než H_0. Oproti tomu levostranná hypotéza nám říká, že průměr základního souboru je menší než H_0. My budeme daný test tvořit, abychom dokázali H_1 a zamítli H_0.

5.1.2 HLADINA VÝZNAMNOSTI A DRUHY CHYB

Velice důležité je určit hladinu významnosti testu (pravděpodobnost, že zamítneme nulovou hypotézu, ale přitom platí). My na základě výsledku náhodného výběru hypotézu můžeme přijmout nebo zamítnout, a proto přijetí nebo zamítnutí H_0 může být správné ale i špatné. Můžeme se dopustit jedné ze dvou chyb. První je chyba 1. druhu (značíme α). Při této chybě zamítáme H_0, když platí. Druhou je chyba 2. druhu (značíme β). Při této chybě přijímáme hypotézu H_0, když neplatí. Test, který by minimalizoval α a β neexistuje, protože chyby spolu souvisí (čím menší je α, tak tím větší je β a naopak). My tedy budeme pokračovat zvolením α, která pro nás bude tedy hladinou významnosti. α budeme volit co nejmenší $(1\%, 5\%, 10\%)$.

5.1.3 POSTUP SESTROJENÍ HYPOTÉZY

Máme-li určenou hladinu významnosti, vybereme testové kritérium (T), do kterého budeme dosazovat konkrétní hodnoty. Výsledkem budou určité hodnoty. Množina všech hodnot, kterých může testovací kritérium nabýt, se bude nazývat výběrový prostor (S), který rozdělíme na 2 části. První je obor přijetí (V). Hodnoty náležící V říkají, že lze přijmout H_0. Druhá část je kritický obor (W). Spadají-li hodnoty do W, přijímáme H_1. Hranice bodů, které oddělují V od W, nazveme kritické body. Velikost W záleží na chybě prvního druhu (α), kterou volíme. Je potřeba splnit

$$\alpha = P(T \in W/H_0) \qquad (45)$$

Rovnost (45) ukazuje, jestli výsledek z testového kritéria bude náležet do kritického oboru a bude platit H_0, tak je potom roven α (1%, 5%, 10%). Při obrácení $(1-\alpha)$ získáváme, s jakou pravděpodobností bude platit H_1.

Budeme vždy určovat H_1. Když pro přijetí alternativní hypotézy H_1 svědčí nízké hodnoty testového kritéria, potom kritický obor určíme intervalem:

$$W = (T_{min}, T_{\alpha}). \qquad (46)$$

T_{min}určuje, jaké nejmenší hodnoty může testové kritérium nabývat. T_{α} je alfa procentní kvantil pro rozdělení testového kritéria při platnosti H_0, protože každé testové kritérium bude určeno pomocí nějakého rozdělení. Tento kvantil budeme hledat v tabulkách. Při vysokých hodnotách testového kritéria bude kritický obor roven:

$$W = (T_{1-\alpha}, T_{max}) \qquad (47)$$

Při extrémně nízkých, nebo extrémně vysokých hodnotách lze kritický obor vyjádřit:

$$W = \left(T_{min}, T_{\frac{\alpha}{2}}\right) \cup \left(T_{1-\frac{\alpha}{2}}, T_{max}\right) \qquad (48)$$

Předem zvolené α nám určuje, jak velkou chybu 1. druhu jsme ochotni nést. Vyjde-li v závěru testu, že $T \in W$, pak je prokázána platnost H_1 a my přijímáme $100\alpha\%$ možnost, že daný výrok je špatně. Při $T \in V$ řekneme, že H_1 nebyla prokázána.

5.1.4 PARAMETRICKÉ TESTY

Zde využíváme předpokladu, že známe rozdělení základního souboru. Z větší části se jedná o normální rozdělení. Hlavním cílem je vytvářet výroky o charakteristikách (parametrech rozdělení) daného základního souboru, z kterého uděláme náhodný výběr a sestavíme hypotézu H_0: $\mu = \mu_0$.

Před samotným testováním je nutné brát v úvahu rozsah výběru, jestliže je malý nebo velký (nad 100). Potom využijeme předpis pro konkrétní testové kritérium. Musíme také znát rozptyl základního souboru. Jestliže ho znát nebudeme, musíme určit výběrový rozptyl, který se vypočte ze vzorce:

$$S_x^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} \qquad (49)$$

Je-li rozsah výběru větší než 100 a známe rozptyl, potom platí:

$$U = \frac{x - \mu}{\sigma} . \sqrt{n} \qquad (50)$$

Pro rozsah větší než 100 a neznámý rozptyl platí:

$$U = \frac{x - \mu}{s_x} . \sqrt{n} \qquad (51)$$

Při rozsahu menším než 100 a neznámém rozptylu platí:

$$t = \frac{x - \mu}{s_x} . \sqrt{n} \qquad (52)$$

U rovnosti (52) při rozsahu menším jak 100 a neznámém rozptylu budeme předpokládat, že místo normálního rozdělení platí Studentovo rozdělení. Posledním krokem je určení formulace alternativního rozdělení (jestli bude větší, menší nebo není roven), protože na jeho základě se bude odvíjet volba kritického oboru.

Příklad 5.1.4.1

Společnost vyrábí auta a garantuje zákazníkům, že průměrná spotřeba je šest litrů na sto kilometrů. Hladina významnosti $\alpha = 0{,}1$. Vybrali jsme 20 aut, u kterých se spočetlo, že výběrový průměr je roven 6,2 litrů a výběrový rozptyl 0,9. Máme zjistit, zda společnost nevydává klamnou reklamu?

Řešení:

$H_0: \mu = 6$

$H_1: \mu > 6$

Jestli bude platit H_1, společnost se dopouští klamné reklamy. Protože máme výběr jen z 20 aut, použijeme testové kritérium $t = \frac{x-\mu}{\sigma} . \sqrt{n}$, které má při platnosti H_0 Studentovo rozdělení.

$$t = \frac{x - \mu}{s_x} . \sqrt{n} = \frac{6,2 - 6}{\sqrt{0,3}} . \sqrt{20} = 1,6329$$

Zjistíme $t_{0,9}(19) = 1,328$ a už vidíme, že $1,6329 > 1,328$. Můžeme hypotézu H_0 zamítnout a řekneme, že společnost využívá klamnou reklamu.

Příklad 5.1.4.2

Vyrábíme auta, kde jejich průměrná spotřeba je 5 litrů na 100 kilometrů. Vybrali jsme 10 aut a změřili jsme jejich průměrnou spotřebu (5,7; 4,3; 4,8; 6; 5,1; 5,3; 4,7; 5; 4,9; 5,1). $\alpha = 0,1$. Máme zjistit, zda skutečná spotřeba je uváděných 5 litrů na 100 kilometrů?

Řešení:

Nejprve zjistíme výběrový průměr.

$$\bar{x} = \frac{\sum_{i=1}^{10} x_i}{10} = \frac{50,9}{10} = 5,09$$

Z důvodu, že neznáme rozptyl, musíme spočítat výběrový rozptyl.

$$S_x^2 = \frac{\sum_{i=1}^{10} (x_i - 5,09)^2}{10 - 1} = 0,23877$$

$H_0: \mu = 5$

$H_1: \mu \neq 5$

Řešíme dvoustrannou hypotézu. Máme pouze 10 výběrů, a proto využijeme testové kritérium, které při platnosti H_0 bude mít Studentovo rozdělení.

$$t = \frac{x - \mu}{s_x} . \sqrt{n} = \frac{5,09 - 5}{\sqrt{0,23877}} . \sqrt{10} = 0,5824$$

Zjistíme si $t_{1-\frac{0,1}{2}}(9) \rightarrow t_{0,95}(9) = 1,833$, a protože se jedná o symetrické rozdělení, $t_{0,05}(9) = -1,833$. Hodnota testového kritéria spadá do oboru přijetí a my nemůžeme H_0 zamítnout.

5.1.5 HYPOTÉZY ROZPTYLU

Logika výpočtu je úplně stejná jako u testování průměru. Máme testové kritérium, které spočteme. Pak rozlišujeme, jestli spadá do kritického oboru. Potom zamítáme nebo přijímáme H_1. Uděláme si výběr a sestavíme hypotézu

$$H_0\colon \sigma^2 = \sigma_0^2, \qquad (53)$$

proti této nulové hypotéze vytvoříme alternativní hypotézu (jednostrannou nebo dvoustrannou)

$$H_1\colon \sigma^2 \neq \sigma_0^2 - dvoustranná\ hypotéza$$
$$H_1\colon \sigma^2 > \sigma_0^2 - \ pravostranná\ hypotéza$$
$$H_1\colon \sigma^2 < \sigma_0^2 - levostranná\ hypotéza.$$

Testové kritérium má tvar:

$$\chi^2 = \frac{(n-1).S_x^2}{\sigma_0^2} \qquad (54)$$

Při platnosti H_0 má statistika χ^2-rozdělení s $v = n - 1$. Proto budeme u těchto testů hledat χ^2 kvantily rozdělení. Vyjde-li nám při oboustranné hypotéze hodnota testového kritéria menší než kvantil s $n - 1$ stupni volnosti

$$X^2 \leq X_{\frac{\alpha}{2}}^2 \qquad (55)$$

nebo hodnota testového kritéria větší než daný kvantil s $n - 1$ stupni volnosti

$$X^2 \leq X_{1-\frac{\alpha}{2}}^2, \qquad (56)$$

potom přijmeme H_1 a zamítneme H_0.

U jednostranných hypotéz získáme kritické obory následovně:

$$X^2 \leq X_\alpha^2, X^2 \leq X_{1-\alpha}^2 \qquad (57)$$

U levostranné hypotézy budou svědčit nízké hodnoty, budou-li menší jak kvantil, můžeme zamítnout H_0 a přijmeme H_1 a naopak. U pravostranné hypotézy budou svědčit pro H_1 vysoké hodnoty.

Příklad 5.1.5.1

Máme přístroj, který plní balíčky. Průměrná hmotnost balíčku je 250g. Víme, že směrodatná odchylka daného stroje je 10g. Chceme zjistit, když daný stroj vyměníme za novou technologií, jestli je lepší a přesnější? Bylo vybráno 10 balíčků, u kterých byla zjištěna následující hmotnost (238; 250; 253; 246; 260; 251; 249; 252; 254; 248).

Řešení:

Vytvoříme si nejdříve danou hypotézu.

$$H_0: \sigma = 10 \rightarrow \sigma^2 = 100$$

$$H_1: \sigma^2 < 10 \rightarrow \sigma^2 < 100$$

Nyní spočteme výběrový průměr a následně výběrový rozptyl.

$$\bar{x} = \frac{\sum_{i=1}^{10} x_i}{10} = 250{,}1$$

$$S_x^2 = \frac{\sum_{i=1}^{10}(x_i - 250{,}1)^2}{10 - 1} = 32{,}7$$

Dosadíme do testového kritéria.

$$\chi^2 = \frac{(n-1).\,S_x^2}{\sigma_0^2} = \frac{(10-1).\,32{,}7}{100} = 2{,}949$$

Zjistíme z tabulek $\chi^2_{0{,}05}(9) = 3{,}33$. Řešíme levostrannou hypotézu, bude muset platit $\chi^2 \leq \chi^2_\alpha \rightarrow 2{,}949 \leq 3{,}33$. Vidíme, že hodnota testového kritéria je menší než kvantil a patří do kritického oboru. Můžeme zamítnout hypotézu H_0 a přijmeme H_1(nově dosazená technologie je účinnější než starší, která byla vyměněna).

5.1.6 DVOUVÝBĚROVÝ T TEST O ROVNOSTI PRŮMĚRU

Máme 2 základní soubory a děláme dva na sobě nezávislé výběry. Při těchto testech rozlišujeme 3 možnosti, které by mohly nastat. První možnost je, že budeme znát rozptyly základních souboru. Druhou možností je, že neznáme rozptyly, ale předpokládáme jejich rovnost. Třetí eventualita je, že neznámé rozptyly a ty jsou navíc různé. Podle toho, o kterou variantu se jedná, se bude muset upravit testové kritérium a daný test.

Budeme-li znát rozptyly, provedeme náhodný výběr a vytvoříme hypotézu, kdy se budou dané výběry rovnat:

$$H_0: \mu_1 = \mu_2 \qquad (58)$$

Proti tomu vytvoříme alternativní hypotézu:

$$H_0: \mu_1 \neq \mu_2 - dvoustranná\ hypotéza$$
$$H_0: \mu_1 > \mu_2 - pravostranná\ hypotéza$$
$$H_0: \mu_1 < \mu_2 - levostranná\ hypotéza.$$

Testové kritérium bude rovno:

$$U = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \qquad (59)$$

Při platnosti H_0 se jedná o $N(0,1)$. Z toho vyvozujeme, že kritické hodnoty jsou brány jako kvantily $N(0,1)$.

5.2 ODHADY PARAMETRŮ

Na statistický soubor $(x_1, \dots, x_n)$, který získáme statistickým šetřením, můžeme nahlížet jako na výběrový soubor získaný realizací náhodného výběru z náhodné veličiny X. Nyní si položíme otázku: Jestli pomocí parametru statistického souboru můžeme určit přesně parametry náhodné veličiny X? My na to odpovídáme ne. Parametry nelze určit přesně, ale budeme je moci odhadnout.

5.2.1 BODOVÉ ODHADY

Statistický soubor lze vyjádřit pomocí charakteristik (střední hodnota, rozptyl, četnost, …). Pro zjednodušení vytváříme výběrové soubory, u kterých zjišťujeme výběrové charakteristiky (statistiky). Důležité je si uvědomit, že základní soubor je pevně stanoven, je neměnný a statistiky (g) se měnit mohou. Statistiky mají charakter náhodné veličiny. Rozptyl a průměr budou kolísat. Budeme-li chtít zjistit charakteristiky základního souboru, nebudeme počítat pouze jen průměr (rozptyl,…) vybraného souboru, ale budeme muset znát pravděpodobnostní rozdělení vhodné výběrové charakteristiky. Provedeme odhad neznámé charakteristiky ze základního souboru (G). Z výběrového souboru spočteme jedno číslo, které se bude nazývat bodový odhad průměru (rozptylu, mediánu,…) základního souboru. Abychom mohli vytvořit daný odhad, je důležité, aby statistika nevedla k systematickým chybám (střední hodnota výběrové charakteristiky má být rovna charakteristice základního souboru). Při rovnosti jsou statistiky nezkresleným odhadem základního souboru. Další možností je asymptotický bodový odhad:

$$\lim_{n \to \infty} (E(g) - G) = 0 \qquad (60)$$

Rovnost (60) říká, že bude-li se rozsah výběru blížit nekonečnu, bude platit rovnost, že při odečítání statistiky základního souboru od střední hodnoty získáme výsledek 0. Jestliže bude nezkreslenost splňovat více statistik, vybíráme statistiku, která má nejmenší rozptyl (nejméně kolísá). Takto zjištěný odhad nazýváme vydatný odhad. Při získání zkresleného odhadu je důležité, aby byl konzistentní.

$$\lim_{n \to \infty} P(|g - G| < \varepsilon) = 1, \;\; \varepsilon > 0 \qquad (61)$$

Určená rovnost (61) říká, že bude-li se rozsah výběru blížit k nekonečnu, pravděpodobnost, že rozdíl $g - G$ je menší než ε, je jedna (rozdíl se nebude zvětšovat s rostoucím výběrem).

Bodové odhady základních číselných charakteristik náhodné veličiny X na základě výběrových charakteristik:

1) $\bar{X}$ je nestranný konzistentní odhad střední hodnoty $E(X)$;
2) $\frac{n}{n-1} . S^2$ je nestranný konzistentní odhad rozptylu $D(X)$.

Tyto dva odhady jsou nejlepší pro normální rozdělení.
Bodové odhady základních číselných charakteristik náhodné veličiny X na základě empirických charakteristik statistického souboru:

1) $\bar{x}$ je bodovým odhadem střední hodnoty $E(X)$;
2) $\frac{n}{n-1} . s^2$ je bodovým odhadem rozptylu $D(X)$;
3) $\sqrt{\frac{n}{n-1}} . s$ je bodovým odhadem směrodatné odchylky $\sigma(X)$;

kde $\bar{x}, s^2, s$ jsou získané empirické charakteristiky.

5.2.2 INTERVALOVÉ ODHADY

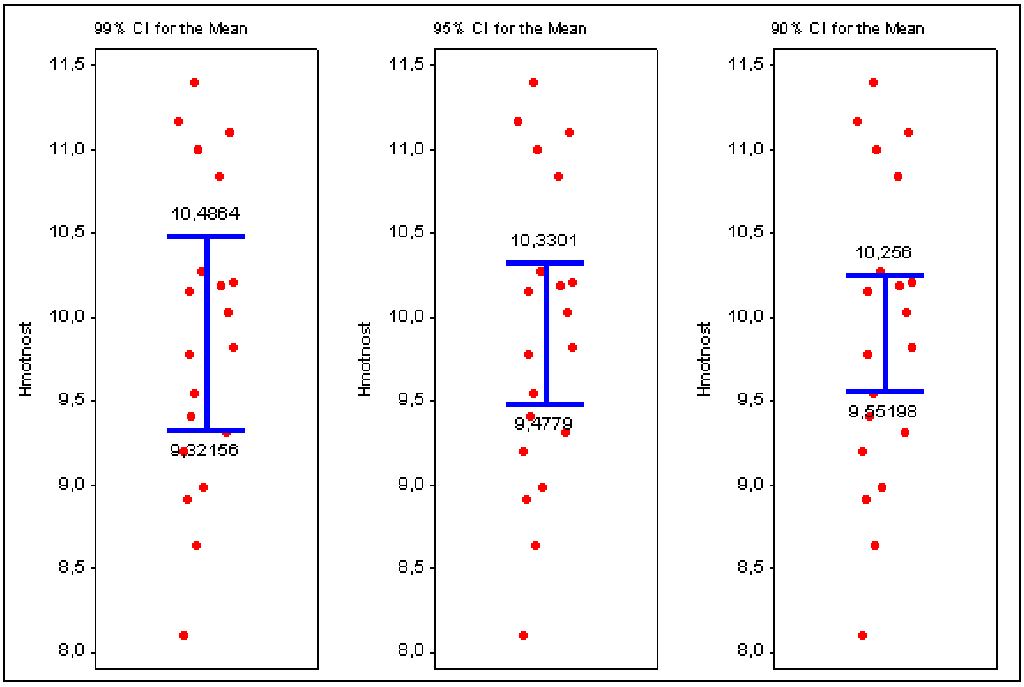
Bodové odhady mají velkou nevýhodu, že jejich spolehlivost je nulová (pravděpodobnost, že parametr určíme přesně) a proto si zavedeme intervalové odhady. Interval spolehlivosti pro parametr ϑ se spolehlivostí $1-\alpha$ je interval, kde $\alpha \in \langle 0,1 \rangle$, je dvojice takových statistik $(T_1; T_2)$, že

$$P(T_1 \leq \vartheta \leq T_2) = 1 - \alpha \qquad (62)$$

pro libovolnou hodnotu ϑ. Pak píšeme, že intervalový odhad parametru ϑ se spolehlivostí $1-\alpha$ je interval $\langle t_1; t_2 \rangle$ a $\vartheta \in \langle t_1; t_2 \rangle$, kde $t_1; t_2$ jsou hodnoty statistik na určeném statistickém souboru.

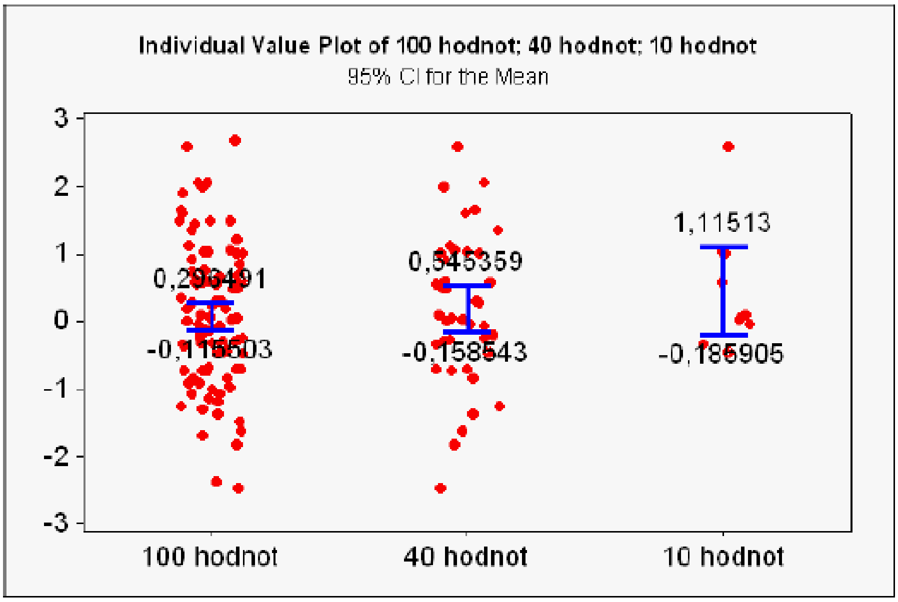
Spolehlivost $1-\alpha$ volíme blízké jedné. Spolehlivost $1-\alpha$ udává, že při velkém počtu opakovaných výběrů s konstantním rozsahem n z daného základního souboru zhruba $(1-\alpha)100\%$ všech intervalových odhadů obsahuje skutečnou hodnotu parametru ϑ a naopak $\alpha 100\%$ ji neobsahuje.

S rostoucí spolehlivostí roste také rozpětí intervalového odhadu. Z toho nám plyne, že chceme-li snížit intervalový odhad, tak snížíme spolehlivost (viz. obrázek 10).



Obrázek 10: Intervalové odhady střední hodnoty se spolehlivostí 0,99, 0,95, 0,90 a pro stejný statistický soubor

Postup, kdy snižujeme interval spolehlivosti, se nedoporučuje, protože chceme udržet co nejvyšší spolehlivost. Lepším způsobem je naopak zvětšení souboru n (důležité si uvědomit, že velikost intervalového odhadu se zmenší úměrně $\sqrt{n}$ (viz. obrázek 11)).



Obrázek 11: Intervalové odhady střední hodnoty pro náhodné výběry s různým rozsahem

5.2.3 ODHADY PARAMETRŮ NORMÁLNÍHO ROZDĚLENÍ

Předpokládáme, že pozorovaná náhodná veličina X má normální rozdělení pravděpodobnosti s parametry μ, σ^2, tak bodové odhady jsou (resp. i ϑ):

$$\mu = \bar{x}, \sigma^2 = \frac{n}{n-1} . s^2, \sigma = \sqrt{\frac{n}{n-1}} . s, \vartheta = r \qquad (63)$$

Intervalový odhad střední hodnoty μ se spolehlivostí $1 - \alpha$ při neznámém rozptylu σ^2 je roven:

$$\langle \bar{x} - t_{1-\frac{\alpha}{2}} . \frac{s}{\sqrt{n-1}} ; \bar{x} + t_{1-\frac{\alpha}{2}} . \frac{s}{\sqrt{n-1}} \rangle, \qquad (64)$$

kde $t_{1-\frac{\alpha}{2}}$ je kvantil Studentova rozdělení s $n - 1$ stupni volnosti.

Intervalový odhad rozptylu σ^2 se spolehlivostí $1 - \alpha$ je roven:

$$\langle \frac{n . s^2}{\chi^2_{1-\frac{\alpha}{2}}} ; \frac{n . s^2}{\chi^2_{\frac{\alpha}{2}}} \rangle, \qquad (65)$$

kde χ^2_p je P-kvantil Pearsonova rozdělení s $n - 1$ stupni volnosti.

5.2.4 INTERVALOVÝ ODHAD STŘEDNÍ HODNOTY POMOCÍ CLV

Nastane-li případ, že náhodné veličiny nemají normální rozdělení, nejsme schopni použít předchozí odhady, a proto využijeme jinou možnost. Máme-li více náhodných veličin, využijeme centrální limitní věty (viz. níže).

Centrální limitní věta: nechť $X_1, \dots, X_n$ je posloupnost nezávislých, stejně rozdělených náhodných veličin se střední hodnotou μ a konečným rozptylem σ^2. Pak náhodná veličina

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n.\sigma^2}} \qquad (66)$$

má při $n \to \infty$ asymptoticky rozdělení $N(0,1)$.

Zjednodušeně řečeno, že součet většího počtu náhodných veličin se chová jako normální rozdělení (závisí to také na tom, jaké rozdělení má veličina X_i). Pro aproximaci budeme využívat rozsah náhodného výběru $n \geq 20$.

Máme náhodný výběr $X_1, \dots, X_n$ z rozdělení s konečnou střední hodnotou μ a konečným rozptylem σ^2. Potom podle centrální limitní věty má

$$\frac{\bar{X} - \mu}{S}\sqrt{n} \to_{n \to \infty} N(0,1) \qquad (67)$$

asymptoticky normované normální rozdělení. Z definice kritické hodnoty normovaného normálního rozdělení následně plyne:

$$P\left(-u\left(1 - \frac{\alpha}{2}\right) \leq \frac{\bar{X} - \mu}{S}\sqrt{n} \leq u\left(1 - \frac{\alpha}{2}\right)\right) = 1 - \alpha \qquad (68)$$

Upravením dostáváme oboustranný intervalový odhad pro střední hodnotu μ o spolehlivosti $1 - \alpha$:

$$\langle \bar{X} - u\left(1 - \frac{\alpha}{2}\right)\frac{S}{\sqrt{n}}, \bar{X} + u\left(1 - \frac{\alpha}{2}\right)\frac{S}{\sqrt{n}} \rangle \qquad (69)$$

Příklad 5.2.4.1

Byl proveden pokus, při kterém jsme 600 krát házeli kostkou, kdy 75 krát padla šestka. Našim úkolem bude zjistit odhad pravděpodobnosti hodu šest na této kostce, jestli je spravedlivá (zda každé číslo padne s pravděpodobností $\frac{1}{6} = 0{,}166$)?

Řešení:

My zavedeme náhodné veličiny $X_1, \dots, X_{600}$s alternativním rozdělením $A(p)$, kde úspěch ($X = 1$) vzniká, když padne 6 a neúspěch ($X = 0$), když padne 1 až 5. Bodový odhad bude:

$$\bar{X} = \frac{75}{600} = 0{,}125$$

Výběrový rozptyl bude roven:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{599}[75(1-0{,}125)^2 + 525(1-0{,}125)^2]$$
$$= 0{,}109$$

Pro výpočet o spolehlivosti 95% potřebujeme ještě zjistit $u(0{,}975)$. Z tabulek zjistíme, že $u(0{,}975) = 1{,}96$. Nyní už dosadíme a vychází interval:

$$\langle \bar{X} - u(0{,}975)\frac{S}{\sqrt{n}}, \bar{X} + u(0{,}975)\frac{S}{\sqrt{n}} \rangle = \langle 0{,}098; 0{,}151 \rangle$$

Z výsledného intervalu vidíme, že s pravděpodobností 95% leží hod šestky v daném intervalu na této kostce. Pravděpodobnost, že padne šestka, je v rozmezí od 9,8% do 15,1%. Z toho také vidíme, že kostka spravedlivá není, protože 0,166 neleží ve vypočteném intervalovém odhadu o spolehlivosti 95%.

5.3 APROXIMACE NORMÁLNÍM ROZDĚLENÍM

Abychom mohli pokračovat, budeme si muset zavést dvě dílčí formulace centrální limitní věty: Linderbergovu-Lévyho větu a Moivreovo-Laplaceovu větu.

Linderbergova-Lévyho věta

Jestliže $X_1, \dots, X_n$ jsou nezávislé náhodné veličiny se stejným (libovolným) rozdělením, stejnými středními hodnotami a se stejnými rozptyly, pak platí:

$$Y_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \Rightarrow \lim_{n\to\infty} P(Y_n < u) = \Phi(u) \qquad (70)$$

(Y_n konverguje v distribuci k rozdělení $N(0,1)$). Vyplývá nám z toho, že pro dostatečně velká n:

$$X = \sum_{i=1}^{n} X_i \Rightarrow E(X) = n.\mu, D(X) = n.\sigma^2 \qquad (71)$$

$$\bar{X} = \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{X}{n} \Rightarrow E(\bar{X}) = \frac{1}{n}.n\mu = \mu, D(\bar{X}) = \left(\frac{1}{n}\right)^2.n\sigma^2 = \frac{\sigma^2}{n} \qquad (72)$$

Je patrné, že rozdělení náhodné veličiny X, můžeme aproximovat rozdělením $N(n\mu; n\sigma^2)$. Obdobně $\bar{X}$ můžeme aproximovat rozdělením $N(\mu; \frac{\sigma^2}{n})$.

Příklad 5.3.1

Dlouhodobým průzkumem bylo zjištěno, že doba potřebná k objevení a odstranění poruchy stroje má střední hodnotu 40 minut a směrodatnou odchylku 30 minut. Jaká je pravděpodobnost, že doba potřebná k objevení a opravení 100 poruch nepřekročí 70 hodin?

Řešení:

X_i - doba, která je zapotřebí k objevení a odstranění i-té poruchy

$E(X_i) = 40, D(X_i) = 30^2, X$ - celková doba k objevení všech 100 chyb

Důležité je, že součet n náhodných veličin, která mají stejné rozdělení (jedno jaké), stejnou střední hodnoty a stejný rozptyl můžeme aproximovat normálním rozdělením (viz. věty výše) s parametry $\mu = n.E(X_i), \sigma^2 = n.D(X_i)$. Vznikne nám

$$X \to N(100.40; 10.30^2)$$

Na závěr spočítáme pravděpodobnost:

$$P(X < 4200) = F(4200) = \Phi\left(\frac{4200 - 4000}{\sqrt{90000}}\right) = \Phi(0,67) = 0,749$$

Pravděpodobnost, že nepřekročíme 70 hodin při hledání 100 chyb je přibližně 75%.

Moivreova- Laplaceova věta

Nechť máme náhodnou veličinu X, která má binomické rozdělení, $E(X) = np, D(X) = np(1-p)$, pak pro velká n platí:

$$U = \frac{X - np}{\sqrt{np(1-p)}} \rightarrow N(0,1) \qquad (73)$$

Můžeme také zapisovat $X \rightarrow N(np; np(1-p))$.

Nyní se zaměříme na samotnou aproximaci normálním rozdělením. Normálním rozdělením lze aproximovat některé diskrétní rozdělení (Binomické rozdělení, Poissonovo rozdělení, Hypergeometrické rozdělení).

My zde vypracujeme pro ukázku aproximaci Poissonova rozdělení normálním rozdělením:

Podmínkou je, aby časový interval $(0, t)$ byl velký, a proto je dostatečně velká i střední hodnota λt:

$$X \rightarrow Po(\lambda t), E(X) = \lambda t, D(X) = \lambda t,$$

pak pro velké t bude platit, že X lze aproximovat normálním rozdělením, které bude mít parametry $\mu = \lambda t, \sigma^2 = \lambda t$.

$$X \rightarrow N(\lambda t, \lambda t)$$

Důkaz: Budeme uvažovat Poissonův proces, který pozorujeme během času t. My předpokládáme, že rychlost výskytu událostí je λ. Pak pravděpodobnost, že se události vyskytují během intervalu $(0, t)$, je úměrná hodnotě λt . Dalším krokem bude si rozdělit interval délky t na n subintervalů stejné délky $\left(\frac{t}{n}\right)$. Výskyt událostí bude v každém intervalu nezávislý a pravděpodobnost výskytu události během jednoho subintervalu bude úměrná hodnotě $(\lambda.\left(\frac{t}{n}\right))$. Je-li n dostatečně velké číslo, tak vyplývá, že délka intervalu bude dostatečně malá, že pravděpodobnost výskytu více než jedné události v tomto intervalu je skoro nulová a pravděpodobnost výskytu jedné události je úměrná $(\lambda.\left(\frac{t}{n}\right))$. Potom, bude-li X_i vyjadřovat počet výskytů událostí v i-tém subintervalu, je zřejmé, že X_i bude mít alternativní rozdělení s parametrem $p = \lambda.\left(\frac{t}{n}\right)$.

$$X_i \rightarrow A\left(\lambda.\frac{t}{n}\right), E(X_i) = \lambda.\frac{t}{n}, D(X_i) = \lambda.\frac{t}{n}\left(1 - \lambda.\frac{t}{n}\right) \qquad (74)$$

Je-li náhodná veličina X definována jako počet výskytů událostí během daného časového intervalu $(0, t)$, pak má Poissonovo rozdělení s parametrem λt.

$$X \rightarrow Po(\lambda t) \qquad (75)$$

Náhodnou veličinu X lze vyjádřit jako součet n náhodných veličin X_i, a proto ji můžeme podle CLV aproximovat normálním rozdělením.

$$X = \sum_{i=1}^{n} X_i \Rightarrow E(X) = n.EX_i = \lambda t; D(X) = n.DX_i$$

$$= \lim_{n\to\infty}\left[\lambda t(1 - \frac{\lambda t}{n})\right] = \lim_{n\to\infty}\left[\lambda t - \frac{(\lambda t)^2}{n})\right] = \lambda t$$

$$X \to N(\lambda t; \lambda t)$$

6 ZÁVĚR

Cílem mé bakalářské práce bylo shrnutí normálního rozdělení a rozdělení z něho odvozených. Na začátku jsem zmínil nejdříve historii, kde jsem popsal vývoj normálního rozdělení. Uvedl jsem významné matematiky, kteří se tímto rozdělením zabývali.

Normální rozdělení a rozdělení z něj odvozená jsou velice důležitá a zajímavá, protože se využívají v mnoha odvětvích. Snažil jsem se je tedy popsat, ukázat jejich využití a spočítat názorné příklady, kde se s těmito rozděleními pracuje (viz testování hypotéz, bodové odhady, atd.). Je důležité poznamenat, že vypsání všech příkladů využití, by bylo na velmi obsáhlou knihu, a proto jsou zde jen některé možnosti využití.

Doufám, že bude moje bakalářská práce srozumitelná, bude se líbit a případně objasní některé z rozdělení, které je zde rozebráno, s jeho využitím.

7 SEZNAM LITERATURY A DALŠÍCH STUDIJNÍCH ZDROJŮ

[1]. ANDĚL, Jiří. *Matematická statistika.* Praha: Nakladatelství technické literatury SNTL, 1978. 352 s

[2]. ZVÁRA, K., ŠTĚPÁN, J. *Pravděpodobnost a matematická statistika.* 2. vydání Praha: nakladatelství Matfyzpress, 2002

[3]. ŠŤĚPÁN, Josef. *Teorie pravděpodobnosti : Matematické základy : Vysokošk. učebnice pro stud. matematicko-fyz. fakult*. Praha : Academia, 1987

[4]. HEBÁK, P., KAHOUNOVÁ, J. *Počet pravděpodobnosti v příkladech.* Praha: Nakladatelství technické literatury SNTL, 1978. 312 s

[5]. LIKEŠ, J., MACHEK, J. *Počet pravděpodobnosti.* Praha: Nakladatelství technické literatury SNTL, 1987. 160 s

[6]. KAHOUNOVÁ, Jana. *Praktikum k výuce matematické statiky I : Odhady*. Praha: Vysoká škola ekonomická, 2000. 97s

INTERNETOVÉ ZDROJE:

[7]. ONLINE STATISTICS EDUCATION. NORMAL DISTRIBUTION [online]. 2008. [cit. 2013-03-25]. Dostupné z WWW: <http://onlinestatbook.com/2/normal_distribution/normal_distribution.html>

[8]. TURNBULL. ABRAHAM DE MOIVRE [online]. 2000. [cit. 2013-03-25]. Dostupné z WWW: <http://www-gap.dcs.st-and.ac.uk/~history/Biographies/De_Moivre.html>

[9]. TURNBULL. JOHANN CARL FRIEDRICH GAUSS [online]. 2000. [cit. 2013-03-25]. Dostupné z WWW: <http://www-gap.dcs.st-and.ac.uk/~history/Biographies/Gauss.html >

[10]. TURNBULL. PIERRE-SIMON LAPLACE [online]. 2000. [cit. 2013-03-25]. Dostupné z WWW: <http://www-gap.dcs.st-and.ac.uk/~history/Biographies/Laplace.html>

[11]. BUSINESS KNOWLEDGE CENTER. NORMAL DISTRIBUTION [online]. 2002-2010. [cit. 2013-04-16]. Dostupné z WWW: <http://www.netmba.com/statistics/distribution/normal/>

[12]. FAKULTA JADERNÁ A FYZIKÁLNĚ INŽENÝRSKÁ. PŘÍMÉ MĚŘENÍ [online]. 2001-2004. [cit. 2013-04-16]. Dostupné z WWW: <http://www.fjfi.cvut.cz/files/k402/pers_hpgs/skoda/prime_mereni1.pdf>

[13]. FCC PUBLIC. ODBORNÉ ČASOPISY – ELEKTRO [online]. 2013. [cit. 2013-04-16]. Dostupné z WWW: <http://www.odbornecasopisy.cz/index.php?id_document=42353>

[14]. ZDROJ. TEORIE CHYB [online]. 2010. [cit. 2013-04-16]. Dostupné z WWW: <http://zdroj.mysteria.cz/files/GNR.pdf>

[15]. INGGEO – PORTÁL INŽENÝRSKÉ GEODÉZIE. TEORIE CHYB [online]. 2012. [cit. 2013-04-16]. Dostupné z WWW: <http://inggeo.fsv.cvut.cz/wiki/doku.php?id=04_teorie_chyb>

[16]. FAKULTA VETERINÁRNÍHO LÉKAŘSTVÍ. BIOSTATISTIKA - NORMÁLNÍ ROZDĚLENÍ [online]. 2013. [cit. 2013-04-16]. Dostupné z WWW: <http://cit.vfu.cz/stat/FVL/Teorie/Predn2/rozdelZS.htm#Gaussovo>

[17]. FAKULTA VETERINÁRNÍHO LÉKAŘSTVÍ. BIOSTATISTIKA – PRAVDĚPODOBNOSTNÍ ROZDĚLENÍ [online]. 2006. [cit. 2013-04-16]. Dostupné z WWW:<http://cit.vfu.cz/stat/FVL/Teorie/Predn2/rozdelVS.htm>

[18]. INTERAKTIVNÍ UČEBNICE STATISTIKY. CHÍ-KVADRÁT ROZDĚLENÍ [online]. 2001. [cit. 2013-05-17]. Dostupné z WWW: <http://iastat.vse.cz/CHikvadr.htm>

[19]. STANFORD UNIVERSITY. NORMAL DISTRIBUTION [online]. 2004. [cit. 2013-05-17]. Dostupné z WWW: <http://www-stat.stanford.edu/~naras/jsm/NormalDensity/NormalDensity.html >

[20]. ČVUT PRAHA - FAKULTA JADERNÁ A FYZIKÁLNĚ INŽENÝRSKÁ. CHÍ-KVADRÁT ROZDĚLENÍ [online]. 2012. [cit. 2013-05-28]. Dostupné z WWW: <http://kfe.fjfi.cvut.cz/~limpouch/sigdat/statodn/node7.html >

[21]. JIHOČESKÁ UNIVERZITA V ČESKÝCH BUDĚJOVICÝCH. ÚVOD DO STATISTIKY [online]. 2006. [cit. 2013-06-14]. Dostupné z WWW: <http://www.pf.jcu.cz/stru/katedry/m/petrasekstat.pdf>

[22]. IOWA STATE UNIVERSITY. LARGE SAMPLE THEORY [online]. 2003-2013. [cit. 2013-06-14]. Dostupné z WWW: <http://www2.econ.iastate.edu/classes/econ671/hallam/documents/LargeSampleTheory_000.pdf>

[23]. TEORIE PRAVDĚPODOBNOSTI. PRAVDĚPODOBNOSTNÍ ROZDĚLENÍ A SAS [online]. 2012. [cit. 2013-06-17] Dostupné z WWW: <http://www.osvobodtepi.cz/wp-content/uploads/4st201/Pravd_SAS.pdf>

[24]. ÚSTAV MATEMATIKY FSI VUT BRNO. MATEMATIKA ONLINE [online]. 2005. [cit. 2013-06-08]. Dostupné z WWW: <http://mathonline.fme.vutbr.cz/Odhady-parametru/sc-1157-sr-1-a-149/default.aspx>

[25]. TECHNICKÁ UNIVERZITA OSTRAVA. ODHADY PARAMETRŮ ZÁKLADNÍHO SOUBORU [online]. 2011. [cit. 2013-06-19]. Dostupné z WWW: <http://homel.vsb.cz/~dom033/predmety/statistika/ucebni_text/12Odhady.pdf >

SEZNAM INTERNETOVÝCH ZDROJŮ OBRÁZKŮ

- **Obrázek 1. -** ONLINE STATISTICS EDUCATION. HISTORY OF NORMAL DISTRIBUTION- EXAMPLES OF BINOMIAL DISTRIBUTION [online]. 2008. [cit. 2013-03-25]. Dostupné z WWW: <http://onlinestatbook.com/2/normal_distribution/history_normal.html>
- **Obrázek 2. -** ONLINE STATISTICS EDUCATION. HISTORY OF NORMAL DISTRIBUTION [online]. 2008. [cit. 2013-03-25]. Dostupné z WWW: <http://onlinestatbook.com/2/normal_distribution/history_normal.html>
- **Obrázek 3. -** ZDROJ. TEORIE CHYB [online]. 2010. [cit. 2013-04-16]. Dostupné z WWW: <http://zdroj.mysteria.cz/files/GNR.pdf>
- **Obrázek 4. –** FAKULTA VETERINÁRNÍHO LÉKAŘSTVÍ. BIOSTATISTIKA– PRAVDĚPODOBNOSTNÍ ROZDĚLENÍ [online]. 2006. [cit. 2013-04-16]. Dostupné z WWW: <http://cit.vfu.cz/stat/FVL/Teorie/Predn2/rozdelZS.htm#NNR>
- **Obrázek 5. –** FAKULTA VETERINÁRNÍHO LÉKAŘSTVÍ. BIOSTATISTIKA – PRAVDĚPODOBNOSTNÍ ROZDĚLENÍ [online]. 2006. [cit. 2013-04-16]. Dostupné z WWW: <http://cit.vfu.cz/stat/FVL/Teorie/Predn2/rozdelVS.htm>
- **Obrázek 6. –** FAKULTA VETERINÁRNÍHO LÉKAŘSTVÍ. BIOSTATISTIKA – PRAVDĚPODOBNOSTNÍ ROZDĚLENÍ [online]. 2006. [cit. 2013-04-16]. Dostupné z WWW: <http://cit.vfu.cz/stat/FVL/Teorie/Predn2/rozdelVS.htm>
- **Obrázek 7. –** FAKULTA VETERINÁRNÍHO LÉKAŘSTVÍ. BIOSTATISTIKA – PRAVDĚPODOBNOSTNÍ ROZDĚLENÍ [online]. 2006. [cit. 2013-04-16]. Dostupné z WWW: <http://cit.vfu.cz/stat/FVL/Teorie/Predn2/rozdelVS.htm>

- **Obrázek 8. –** PRAVDĚPODOBNOSTNÍ ROZDĚLENÍ. GAMA ROZDĚLENÍ [online]. 2006. [cit. 2013-06-17]. Dostupné z WWW: <http://panda.hyperlink.cz/cestapdf/pdf06c6/mares.pdf>
- **Obrázek 9. –** TEORIE PRAVDĚPODOBNOSTI. PRAVDĚPODOBNOSTNÍ ROZDĚLENÍ A SAS [online]. 2012. [cit. 2013-06-17]. Dostupné z WWW: <http://www.osvobodtepi.cz/wp-content/uploads/4st201/Pravd_SAS.pdf>
- **Obrázek 10. –** VYSOKÁ ŠKOLA CHEMICKO-TECHNOLOGICKÁ V PRAZE. APLIKOVANÁ STATISTIKA [online]. 2005. [cit. 2013-06-20]. Dostupné z WWW: <http://sofe2-files.pepiino.cz/ast/ast-vscht.pdf >
- **Obrázek 11. –** ÚSTAV MATEMATIKY FSI VUT BRNO. INTERVALOVÉ ODHADY [online]. 2005. [cit. 2013-06-08]. Dostupné z WWW: <http://mathonline.fme.vutbr.cz/Odhady-parametru/sc-1157-sr-1-a-149/default.aspx>
- **Obrázek 12. –** ÚSTAV MATEMATIKY FSI VUT BRNO. INTERVALOVÉ ODHADY [online]. 2005. [cit. 2013-06-08]. Dostupné z WWW: <http://mathonline.fme.vutbr.cz/Odhady-parametru/sc-1157-sr-1-a-149/default.aspx>

8 SEZNAM OBRÁZKŮ

9 RESUMÉ

The bachelor thesis deals with the topic Normal distribution and distributions derived from the normal distribution. This work consists of four main parts. The first part describes the history of the normal distribution and important mathematicians. Second part is about theory of errors. In this part is described Gaussian curve. The third part is about theory of normal distribution and distributions derived from the normal distribution. This part contains examples of distribution with theirs solution and use. Last part contains use of distributions (statistical hypothesis testing, point estimations and use the normal distribution to approximate the distribution of other).