Matematika pro radiologické asistenty

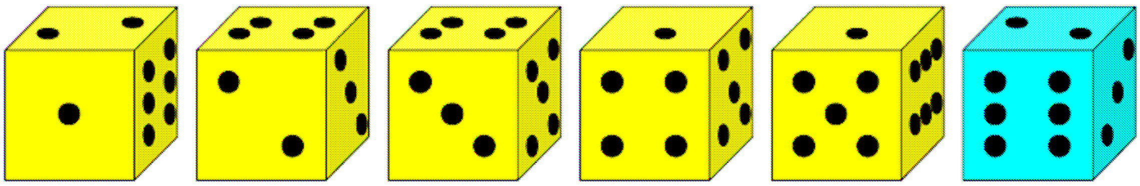
9. Pravděpodobnost

9.1 Náhodné jevy a jejich pravděpodobnost

Začneme dvěma velmi známými příklady ze života:

Příklad 1. Házení mincí – náhodný pokus: jsou celkem dvě možnosti – náhodné jevy (orel, hlava). Sledujeme jev A: padne orel – jedna z možností je příznivá, $p = 1/2$.

Příklad 2. Házení kostkou – náhodný pokus: je celkem šest možností – náhodných jevů (1, 2, 3, 4, 5, 6). Sledujeme jev A: padne šestka – jedna z možností je příznivá, $p = 1/6$.



Definice pravděpodobnosti: Pravděpodobnost nastoupení jevu A je podílem počtu případů *M*, v nichž jev A nastal (čitatel), a počtu *N* všech možných případů (jmenovatel).

Příklad s jednou kostkou (všech možných případů je $N = 6$):

a) Jaká je pravděpodobnost, že při hodu kostkou padne sudé číslo? Příznivé jsou ty případy, kdy padne dvojka nebo čtyřka nebo šestka, tedy $M = 3$. Pravděpodobnost je $p = 3/6 = 1/2$.

b) Jaká je pravděpodobnost, že padne číslo dělitelné třemi? Příznivé jsou ty případy, kdy padne trojka nebo šestka, tedy $M = 2$. Pravděpodobnost je $p = 2/6 = 1/3$.

c) Jaká je pravděpodobnost, že padne číslo menší než tři? Příznivé jsou ty případy, kdy padne jednička nebo dvojka, tedy $M = 2$. Pravděpodobnost je $p = 2/6 = 1/3$.

Příklad se dvěma kostkami (všech možných případů je $N = 36$, každá ze 6 možností, které mohou padnout na první kostce, se nezávisle kombinuje s každou ze 6 možností na druhé kostce):

a) Jaká je pravděpodobnost, že při hodu dvěma kostkami (současně) padne součet sedm? Příznivé jsou ty případy, kdy padne jednička na první a šestka na druhé kostce, nebo naopak, dvojka na první a pětka na druhé kostce, nebo naopak, trojka na první a čtyřka na druhé kostce, nebo naopak), tedy $M = 6$. Pravděpodobnost je $p = 6/36 = 1/6$.

b) Jaká je pravděpodobnost, že součin padnuvších čísel bude lichý? Příznivé jsou ty případy, kdy padne jednička na první kostce a jednička nebo trojka nebo pětka na druhé kostce, trojka na první kostce a jednička nebo trojka nebo pětka na druhé kostce, pětka na první kostce a jednička nebo trojka nebo pětka na druhé kostce, tedy $M=9$. Pravděpodobnost je $p=9/36=1/4$.

Jakých hodnot může pravděpodobnost v principu nabývat? Z definice je zřejmé, že hodnoty budou v intervalu mezi nulou a jedničkou

$$\boxed{0 \le p \le 1}$$

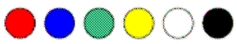
Krajním hodnotám odpovídají jev nemožný – jev s nulovou pravděpodobností a jev jistý – jev s jednotkovou pravděpodobností.

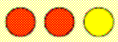
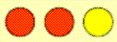
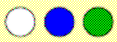
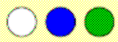
9.2 Kombinatorika

V mnoha aplikacích provádíme výběry *k* prvků z množiny obsahující *n* prvků podle jistých pravidel. Typy pravidel jsou uvedeny v tabulce

	pořadí je podstatné	**pořadí je nepodstatné**
prvky lze opakovat	*variace s opakováním*	*kombinace s opakováním*
prvky nelze opakovat	*variace bez opakování*	*kombinace bez opakování*

Příkladem může být vytváření barevných signálů, kdy šest barev ($n=6$) zaplňuje tři místa ($k=3$):



	variace			**kombinace**		
s opakováním		$\neq$			$\equiv$	
bez opakování		$\neq$			$\equiv$	

Počet možných výběrů je v následující tabulce:

	variace	**kombinace**
s opakováním	$V'_k(n) = n^k$	$C'_k(n) = \dfrac{(n+k-1)!}{(n-1)!k!}$
bez opakování	$V_k(n) = \dfrac{n!}{(n-k)!}$	$C_k(n) = \dfrac{n!}{k!(n-k)!}$

Pro triviální případ $n=k=1$ máme vždy jedinou možnost (připomeňme si, že z definice $0!=1$, $1!=1$, $2!=2$, $3!=6$, $4!=24$, $5!=120$, $6!=720$, $\dots$).

Určení vzorce pro variace opakováním:

Kolika způsoby lze z *n* prvků vybrat první? Je to *n* způsobů. Kolika způsoby lze z *n* prvků vybrat druhý? Opět je *n* způsobů. Kolika způsoby lze tedy vybrat první dva prvky? Je to $n \cdot n = n^2$ způsobů. Pokračujeme až k otázce kolika způsoby lze vybrat *k* prvků? Je to $\underbrace{n \cdot n \cdots n}_{k} = n^k$ způsobů.

Určení vzorce pro variace bez opakování:

Kolika způsoby lze z *n* prvků vybrat první? Je to *n* způsobů. Kolika způsoby lze ze zbývajících $n-1$ prvků vybrat druhý? Je to $n-1$ způsobů. Kolika způsoby lze tedy vybrat první dva prvky? Je to $n \cdot (n-1)$ způsobů. Pokračujeme stejně až nakonec, kdy se ptáme kolika způsoby lze ze zbývajících $n-(k-1)$ prvků vybrat *k*-tý? Je to $n-k+1$ způsobů. Kolika způsoby lze tedy vybrat *k* prvků? Je to $n \cdot (n-1) \cdot (n-2) \cdots (n-k+1)$ neboli $n!/(n-k)!$ prvků.

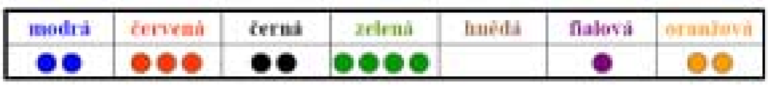
Určení vzorce pro kombinace bez opakování:

Zatímco pro variace bylo pořadí vybraných prvků podstatné, pro kombinace je podstatný pouze výběr těchto prvků. Musíme proto počet způsobů pro variace podělit tím, kolika způsoby lze uspořádat *k* neopakujících se prvků, což je $k!$ Proto je

$$C_k(n) = \frac{V_k(n)}{k!} = \frac{n!}{k!(n-k)!}$$

Určení vzorce pro kombinace s opakováním:

Jak vypadají kombinace s opakováním *k* barevných kuliček při výběru z *n* možných barev? Obrázek ilustruje jednu z možností pro případ $n=7$, $k=14$. Obecně dostaneme počet



kombinací takto: máme *n* přihrádek, tj. $n-1$ přepážek mez nimi a *k* kuliček, celkem tedy $m=n+k-1$ prvků (pozic), z nich vybíráme *k* pozic pro kuličky, tj.

$$C_k'(n) = C_k(m) = \frac{m!}{k!(m-k)!} = \frac{(n+k-1)!}{k!(n-1)!}$$

Při variacích nebo kombinacích bez opakování musí být přirozeně $k \leq n$ („vybraný prvek nevracíme do losování“). Pro $k=1$ jsou samozřejmě počty možností „s opakováním“ – žádné totiž není a „bez opakování“ stejné, pro $n \geq k > 1$ je $V_k(n) < V_k^/(n)$ a $C_k(n) < C_k^/(n)$.

Příklad 1. Jaká je pravděpodobnost hlavní výhry ve Sportce? Tah sportky představuje výběr šesti ze čtyřiceti devíti čísel.

$$N = C_6(49) = \frac{49!}{6!43!} \quad , \quad M = 1 \quad , \quad p = \frac{M}{N} \doteq 7 \cdot 10^{-8}$$

Příklad 2. Jaká je pravděpodobnost páté ceny ve Sportce? Ze šesti tažených je třeba uhodnout tři čísla.

$$N = C_6(49) = \frac{49!}{6!43!} \quad , \quad M = C_3(6)C_3(43) = \frac{6!}{3!3!}\frac{43!}{3!40!} \quad , \quad p = \frac{M}{N} \doteq 0{,}018$$

Příklad 3. Jaká je pravděpodobnost, že ve hře typu Šance milion uhodnete správně taženou skupinu cifer? (Z každého ze šesti bubnů obsahujících cifry 0, 1, 2, . , 9 se náhodně vybere jedna.)

$$N = V_6^/(10) = 10^6 \quad , \quad M = 1 \quad , \quad p = \frac{M}{N} = 10^{-6}$$

9.3 Pravděpodobnosti složených jevů

Někdy je třeba určit pravděpodobnosti jevů, které jsou nějakým způsobem „složeny“ z jevů jednodušších. Uvažujme o dvou jevech ***A*** a ***B***, jejichž pravděpodobnosti jsou známy, $p(A)$ a $p(B)$. Definujme nové jevy ***C*** a ***D*** jako: jev ***C*** je jev, kdy jevy ***A*** a ***B*** nastanou současně, jev ***D*** je jev, kdy nastane jev ***A*** nebo ***B*** (v principu zahrnuje i možnost, že nastanou oba). Za určitých podmínek lze pravděpodobnosti jevů ***C*** a ***D*** určit pomocí pravděpodobností $p(A)$ a $p(B)$.

Nezávislé jevy

Jaká je pravděpodobnost, že při současném hodu dvěma kostkami padne na obou šestka? Uvědomíme si, že to, co padne na jedné kostce, je nezávislé na výsledku druhé kostky. Máme tedy jev ***A***, kdy na první kostce padne šestka – $p(A)=1/6$ a jev ***B***, kdy na druhé kostce padne šestka – $p(B)=1/6$. Jev ***C*** pak odpovídá tomu, že jevy ***A*** a ***B*** nastanou současně. Máme $N = 6 \cdot 6 = 36$ možností a právě jeden příznivý případ $M=1$. Je tedy

$$p(C) = \frac{M}{N} = \frac{1}{36} = \frac{1}{6}\cdot\frac{1}{6} = p(A)\cdot p(B)$$

Pravděpodobnost současného nástupu nezávislých jevů je rovna součinu pravděpodobností těchto jevů.

Neslučitelné jevy

Jaká je pravděpodobnost, že při hodu kostkou padne některé ze dvou nejvyšších čísel, tj. padne šestka nebo pětka? Uvědomíme si, že skutečnosti, že šestka i pětka padnou při stejném hodu, jsou neslučitelné**.** Máme tedy jev ***A***, kdy na kostce padne šestka – $p(A)=1/6$ a jev ***B***, kdy na kostce padne pětka – $p(B)=1/6$. Jev ***C*** pak odpovídá tomu, že nastane buď jev ***A*** nebo jev ***B***. Máme $N=6$ možností a dva příznivé případy $M=2$. Je tedy

$$p(C)=\frac{M}{N}=\frac{2}{6}=\frac{1}{6}+\frac{1}{6}=p(A)+p(B)$$

Pravděpodobnost nástupu některého z jevů, z nichž každé dva jsou neslučitelné, je rovna součtu pravděpodobností těchto jevů.

Označíme-li pravděpodobnost jevu ***A*** $p(A)$, je pravděpodobnost jevu opačného $\overline{\boldsymbol{A}}$ (tj. jev A nenastane) $p(\overline{A})=1-p(A)$. Pro součet pravděpodobností jevu ***A*** a opačného jevu $\overline{\boldsymbol{A}}$ platí

$$p(A)+p(\overline{A})=1$$

Příklad 4. Jaká je pravděpodobnost, že ve skupině *k* osob mají alespoň dvě narozeniny ve stejný den? Rok má $n=365$ dní. Určíme nejprve pravděpodobnost jevu ***A***, že každá z osob má narozeniny *v* jiný den. Počet případů možných pro tento jev je $V_k^{\prime}(n)$ (variace s opakováním – v principu může mít každá z osob narozeniny v kterýkoli den). Počet případů příznivých je $M=V_k(n)$ (variace bez opakování – nechceme, aby se narozeninový den zopakoval u více osob). Jev, který nás zajímá, je opačným jevem k jevu ***A***, jeho pravděpodobnost je tedy

$$p(\overline{A})=1-p(A)=1-\frac{V_k(n)}{V_k^{\prime}(n)}=1-\frac{n!}{(n-k)!n^k}$$

Je-li ve skupině pouze jeden člověk ($k=1$), nemohou připadnout dvoje narozeniny na stejný den – skutečně dostáváme v tomto případě $p(\overline{A})=1-1=0$. Pro skupinu například 40 osob je pak $p(\overline{A})\doteq 1-0{,}11=0{,}89$, tedy skoro 90%.

Bernoulliův pokus

Nejprve zavedeme pojmy a značení. Nastane-li předem definovaný jev ***A*** (například „padne šestka“), nazveme to *zdarem*, v opačném případě *nezdarem*. Pravděpodobnost zdaru označíme p (v příkladu s kostkou p = 1/6), pravděpodobnost nezdaru je $1-p$ (v příkladu s kostkou tedy 5/6). n-krát nezávisle provedeme pokus (například hod kostkou). Bude nás zajímat jev $\boldsymbol{B}$, kdy právě při x provedeních z celkového počtu n provedení pokusu nastane zdar. S touto základní situací souvisí další jevy – jev $\boldsymbol{A_j}$, kdy pro $j=1,2,\ldots,x$ nastane zdar při j-tém provedení pokusu, jev $\boldsymbol{B_k}$, kdy pro $k=x+1,x+2,\ldots,n$ nastane při k-tém provedení pokusu nezdar a konečně jev $\boldsymbol{C}$, kdy jevy $\boldsymbol{A_1}$ až $\boldsymbol{A_x}$ a jevy $\boldsymbol{B_{x+1}}$ až $\boldsymbol{B_n}$ nastoupí současně, tj. právě při prvních x opakováních pokusu nastane zdar, při zbývajících nezdar. Pravděpodobnost jevu $\boldsymbol{C}$ spočteme snadno jako součin pravděpodobností nezávislých jevů (například hodů kostkou)

$$p(C)=p(A_1)\cdots p(A_x)\cdot p(B_{x+1})\cdots p(B_n)=p^x(1-p)^{n-x}$$

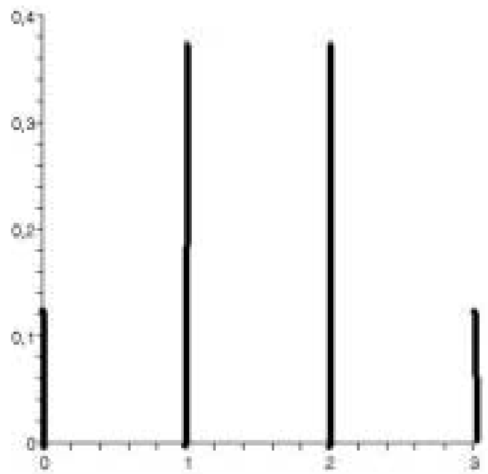
Pro pravděpodobnost jevu $\boldsymbol{B}$ nám ale nezáleží na tom, v jakém pořadí došlo k požadovaným x zdarům ze všech n opakování pokusu. Možností, kdy zdar nastal právě při x ze všech n opakování pokusu, je (kombinace)

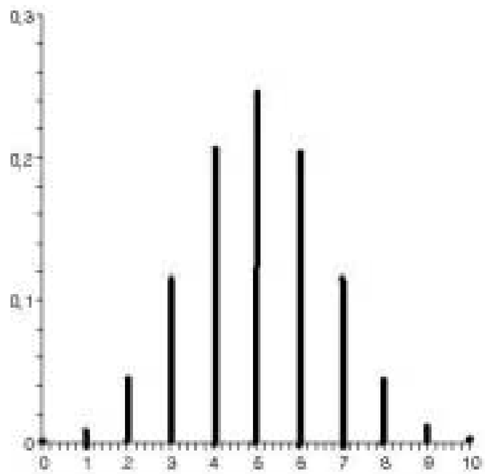
$$C_x(n)=\frac{n!}{x!(n-x)!}\equiv\binom{n}{x}$$

Výsledná pravděpodobnost jevu $\boldsymbol{B}$ je tedy $C_x(n)$-krát větší než pravděpodobnost jevu $\boldsymbol{C}$, tj.

$$p(B)=C_x(n)\,p(C)=\frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}$$

Na obrázcích je závislost pravděpodobnosti jevu $\boldsymbol{B}$ na x při n hodech mincí pro $n=3$ a $n=10$.





Někoho možná překvapí, že při zvyšujícím se n se pravděpodobnost stejného počtu zdarů a nezdarů (pro $n=10$ je to při $x=5$) neblíží jedné polovině, ale klesá. Je to proto, že počítáme pravděpodobnost, kdy je počet zdarů a nezdarů *přesně* stejný. Při velkých hodnotách n pak rozdíl třeba o jedničku nebo dvojku v počtech zdarů a nezdarů má téměř stejnou pravděpodobnost jako když jsou počty stejné, jak vidíme z obrázku už při poměrně malém $n=10$ je pravděpodobnost pro $x=4$ a $x=6$ jen o málo menší než maximální pro $x=5$.

10. Měření a zpracování dat

10.1 Měřené hodnoty veličin jsou „náhodné“

Uvozovky v nadpisu jsou upozorněním na to, že při měření (ať už je jeho podstatou cokoliv) závislosti hodnoty měřené veličiny (závisle proměnná) na hodnotách jiné veličiny (nezávisle proměnná) se mohou uplatňovat také náhodné vlivy, nikoliv snad to, že „naměříme cokoliv“. Tedy za stejných podmínek (tj. při stejných hodnotách nezávisle proměnných veličin) můžeme dostat při opakovaných měřeních různé hodnoty měřené veličiny (závisle proměnné). Základem pro matematický popis této situace je pojem náhodné veličiny.

10.2 Náhodná veličina s diskrétním rozdělením

Náhodná veličina X a její (diskrétní) rozdělení je veličina, která nabývá hodnot $\left(x_0, x_1, \ldots, x_n\right)$ s pravděpodobnostmi $\left(p_0, p_1, \ldots, p_n\right)$. Protože jevy, kdy veličina X nabývá dvou různých hodnot x_j současně jsou neslučitelné, musí být $p_0 + p_1 + \cdots + p_n = 1$. *Rozdělením náhodné veličiny* rozumíme soubor všech dvojic $\left(x_j, p_j\right)$ pro $j=0,1,\ldots,n$. V části o pravděpodobnosti jsme se setkali s výrazem pro *Bernoulliovo rozdělení*

$$x_j = j \quad , \quad p_j = \frac{n!}{j!(n-j)!} q^j (1-q)^{n-j} \quad , \quad 0 \le q \le 1 \quad , \quad j=0,1,\ldots,n$$

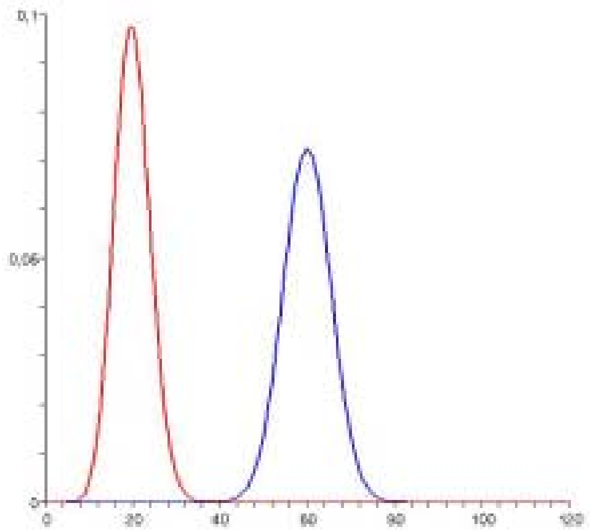
Samozřejmě platí pro součet pravděpodobností

$$\sum_{j=0}^{n} p_j = \sum_{j=0}^{n} \binom{n}{j} q^j (1-q)^{n-j} = \left[(1-q)+q\right]^n = 1$$

V případě, že pravděpodobnost zdaru a nezdaru v jednom pokusu je stejná ($q=1/2$) dostáváme *binomické rozdělení*

$$x_j = j \quad , \quad p_j = \frac{1}{2^n} \frac{n!}{j!(n-j)!} \quad , \quad j=0,1,\ldots,n$$

Na obrázku je Bernoulliovo rozdělení pro $n=120$ a dvě hodnoty q: $q=1/6$ (házení kostkou, červeně) a $q=1/2$ (házení mincí, modře) – pro lepší viditelnost jsou body spojeny plnou křivkou. Maximum je v obou případech u střední hodnoty $\langle j \rangle = q\,n$ (střední hodnotu definujeme dále), tj. $\langle j \rangle = 20$ pro házení kostkou a $\langle j \rangle = 50$ pro házení mincí. Na příkladu



střelce, který vystřelí n-krát na terč si ukážeme, jak zjistit rozdělení experimentálně. Dosažené počty bodů při jednotlivých výstřelech představují hodnoty náhodné veličiny. Jaké jsou pravděpodobnosti jednotlivých hodnot? Pro n = 50 například:

hodnoty	0	1	2	3	4	5	6	7	8	9	10
četnosti	0	0	1	1	2	2	3	3	8	20	10
pravděp	**0,00**	**0,00**	**0,02**	**0,02**	**0,04**	**0,04**	**0,06**	**0,06**	**0,16**	**0,40**	**0,20**

Při různých počtech výstřelů n se pravděpodobnosti budou obecně měnit. Pro rostoucí n budeme pozorovat jejich „ustalování“. Která hodnota nejlépe reprezentuje rozdělení? Náhodnou veličinu samozřejmě nejdokonaleji reprezentuje zadání jejího rozdělení. To je ovšem poněkud nepraktické. U střelce jsme viděli, že jeho kvalita je reprezentována hodnotou blízkou devítce. Realizovala se nejčastěji, má největší váhu. Vhodnější reprezentativní hodnotou bude aritmetický průměr všech hodnot včetně „násobnosti“

$$\langle j \rangle = \frac{n_0 \cdot 0 + n_1 \cdot 1 + \cdots + n_{10} \cdot 10}{n_0 + n_1 + \cdots + n_{10}} = \sum_{j=0}^{10} \frac{n_j}{50}\, j = \sum_{j=0}^{10} p_j\, j = 8{,}12$$

Obecně pak

$$\langle x \rangle = \sum_{j=0}^{n} p_j\, x_j \quad , \quad \sum_{j=0}^{n} p_j = 1$$

Takto definovaná hodnota $\langle x \rangle$ je váženým průměrem jednotlivých hodnot a nazývá se *střední hodnotou náhodné veličiny*. Výpočet střední hodnoty náhodné veličiny, která má Bernoulliovo rozdělení není komplikovaný:

$$\langle j \rangle = \sum_{j=0}^{n} j \frac{n!}{j!(n-j)!} q^j (1-q)^{n-j} = n q \sum_{j=1}^{n} \frac{(n-1)!}{(j-1)!\left[n-1-(j-1)\right]!} q^{j-1} (1-q)^{n-1-(j-1)}$$

a s označením $i = j-1$, $m = n-1$ pak

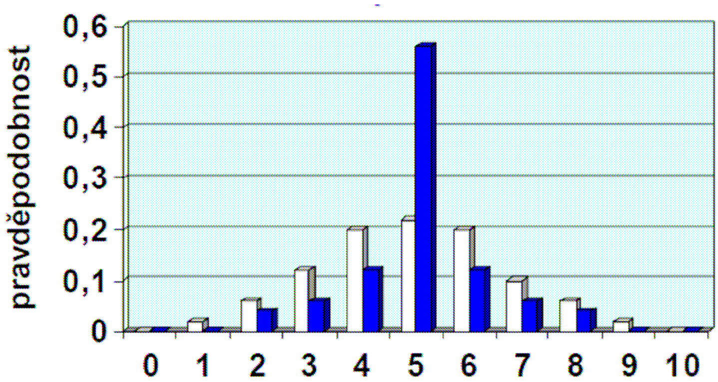
$$\langle j \rangle = n q \sum_{i=0}^{m} \frac{m!}{i!(m-i)!} q^i (1-q)^{m-i} = n q (1-q+q)^m = n q$$

Před zavedením další charakteristiky rozdělení uvedeme další příklad, tentokrát se dvěma střelci, z nichž každý vystřelí *n*-krát na terč. Zvolme opět $n = 50$:

hodnoty	0	1	2	3	4	5	6	7	8	9	10
četnosti	0	1	3	6	10	11	10	5	3	1	0
pravděp	0,00	0,02	0,06	0,12	0,20	0,22	0,20	0,10	0,06	0,02	0,00

hodnoty	0	1	2	3	4	5	6	7	8	9	10
četnosti	0	0	2	3	6	28	6	3	2	0	0
pravděp	0,00	0,00	0,04	0,06	0,12	0,56	0,12	0,06	0,04	0,00	0,00

Na obrázku jsou zobrazena rozdělení obou střelců (pro prvého bílými, pro druhého modrými sloupečky). Výpočtem zjistíme, že oba dosáhli střední hodnoty 5. Čím se tedy jejich výsledky



liší? Je to rozptyl nebo z něho vypočtená *směrodatná odchylka*. Jak se k této charakteristice dostaneme? Uvažujme náhodnou veličinu $Y = f(X)$. Má-li veličina X rozdělení (x_j, p_j), má veličina Y rozdělení $(y_j, p_j) = (f(x_j), p_j)$. Jaká veličina by tedy mohla charakterizovat „odchýlení“ hodnot od střední hodnoty? Nejjednodušší možnost $Y = X - \langle x \rangle$ není vhodná, protože pro libovolné rozdělení je $\langle y \rangle = 0$:

$$\langle y\rangle=\sum_{j=0}^{n} y_j\,p_j=\sum_{j=0}^{n}\left(x_j-\langle x\rangle\right)p_j=\sum_{j=0}^{n} x_j\,p_j-\langle x\rangle\sum_{j=0}^{n} p_j=\langle x\rangle-\langle x\rangle=0$$

Takže musíme vzít náhodnou veličinu $Y=\left(X-\langle x\rangle\right)^2$. Střední hodnotu této veličiny nazýváme *rozptylem* $D(X)=\left\langle\left(X-\langle x\rangle\right)^2\right\rangle$. Provedeme-li umocnění, máme pro rozptyl

$$D(X)=\left\langle X^2-2X\langle x\rangle+\langle x\rangle^2\right\rangle=\left\langle X^2\right\rangle-\langle x\rangle^2$$

Odmocnina z rozptylu (jak vidět z definice rozptyl je vždy kladný) je tzv. *směrodatná odchylka*

$$\sigma(X)=\sqrt{D(X)}=\sqrt{\left\langle X^2\right\rangle-\langle x\rangle^2}$$

Pro úplnost příkladu dodejme, že směrodatná odchylka vyjde 1,7 pro prvého a 1,2 pro druhého střelce. Pro Bernoulliovo rozdělení je

$$\left\langle j^2\right\rangle=\sum_{j=0}^{n} j^2\frac{n!}{j!(n-j)!}q^j(1-q)^{n-j}=n\,q\sum_{j=1}^{n}\frac{j(n-1)!}{(j-1)!\left[n-1-(j-1)\right]!}q^{j-1}(1-q)^{n-1-(j-1)}=$$
$$n\,q\sum_{i=0}^{n-1}\frac{(1+i)(n-1)!}{i!(n-1-i)!}q^i(1-q)^{n-1-i}=n\,q+n(n-1)q^2\sum_{i=1}^{n-1}\frac{(n-2)!}{(i-1)!\left[n-2-(i-1)\right]!}q^{i-1}(1-q)^{n-2-(i-1)}$$
$$=n\,q+n(n-1)q^2\sum_{k=o}^{n-2}\frac{(n-2)!}{k!(n-2-k)!}q^k(1-q)^{n-2-k}=n\,q+n(n-1)q^2=n^2q^2+n\,q(1-q)$$

Rozptyl Bernoulliova rozdělení je tedy

$$D(j)=\left\langle j^2\right\rangle-\langle j\rangle^2=n\,p(1-p)$$

Distribuční funkci $F(x)$ definujeme na reálné ose $\mathbb{R}$ součtem pravděpodobností $p_0+\cdots+p_s$ odpovídajících hodnotám menším než x_{s+1}

$$F(x)=\begin{cases} 0 & x<x_0\\ p_0 & x_0\le x<x_1\\ \vdots & \vdots\\ \sum_{j=0}^{s} p_j & x_s\le x<x_{s+1}\\ \vdots & \vdots\\ \sum_{j=0}^{n} p_j=1 & x\ge x_n\end{cases}$$

Naopak pravděpodobnosti dostaneme jako

$$p_0=F(x_0)\quad,\quad p_j=F\left(x_j\right)-F\left(x_{j-1}\right)\quad 1\le j\le n$$

Medián rozdělení je taková hodnota x_s, pro kterou $F(x_s) < 0,5$ a $F(x_{s+1}) \geq 0,5$.

Pro $n \to \infty$ a j malé přejde Bernoulliovo rozdělení na Poissonovo rozdělení. Upravujeme tedy

$$p_j = \lim_{n\to\infty} \frac{n!}{j!(n-j)!} q^j (1-q)^{n-j} = \frac{\langle j\rangle^j}{j!} \lim_{n\to\infty} \frac{n!}{n^j (n-j)! \left(1-\frac{\langle j\rangle}{n}\right)^j} \lim_{n\to\infty}\left(1-\frac{\langle j\rangle}{n}\right)^n$$

Využijeme vztahů platných pro $n \to \infty$ a j malé: $n! \approx n^j (n-1)!$, $\left(1-\langle j\rangle/n\right)^j \approx 1$ a definice $\exp\left(-\langle j\rangle\right) = \lim_{n\to\infty}\left(1-\langle j\rangle/n\right)^n$. Dostáváme tak Poissonovo rozdělení

$$p_j = \frac{\langle j\rangle^j}{j!} \exp\left(-\langle j\rangle\right)$$

Poissonovo rozdělení se projeví třeba tehdy, sledujeme-li časový průběh počtu produktů radioaktivního rozpadu: počítáme částice detekované v nějakých časových intervalech, ale ty tvoří jen nepatrnou část částic v těchto intervalech vzniklých při rozpadu radioaktivních jader zdroje.

10.3 Náhodná veličina se spojitým rozdělením

Náhodná veličina *X*, která nabývá všech reálných hodnot x z intervalu $[x_m, x_M]$ má (spojité) rozdělení dané funkcí $w(x)$ na tomto intervalu s vlastnostmi

$$w(x) \geq 0 \quad x_m \leq x \leq x_M \quad , \quad \int_{x_m}^{x_M} w(x)\, dx = 1$$

Funkci $w(x)$ se také říká hustota pravděpodobnosti, neboť elementární pravděpodobnost (tj. pravděpodobnost, že veličina *X* má hodnotu v intervalu $[x, x+dx]$) je $dp = w(x)\, dx$. Také pro náhodnou veličinu se spojitým rozdělením můžeme definovat střední hodnotu, směrodatnou odchylku, distribuční funkci a medián:

$$\langle x\rangle = \int_{x_m}^{x_M} x\, w(x)\, dx \quad , \quad \sigma = \sqrt{\int_{x_m}^{x_M} \left(x - \langle x\rangle\right)^2 w(x)\, dx}$$

$$F(x) = \int_{x_m}^{x} \xi\, w(\xi)\, d\xi \quad , \quad F(x_{MED}) = 0,5$$

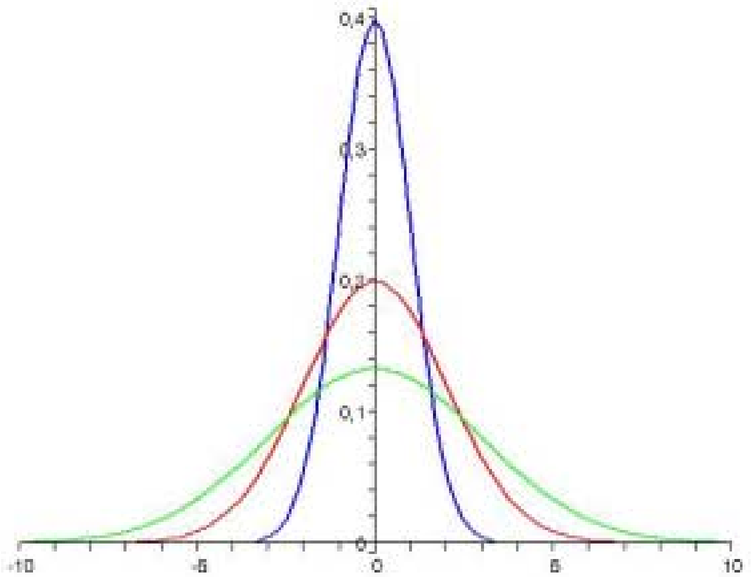
Důležitým příkladem je normální neboli Gaussovo rozdělení na celé reálné ose, tj. $x \in (-\infty, \infty)$:

$$w(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{x^2}{2\sigma^2}\right]$$

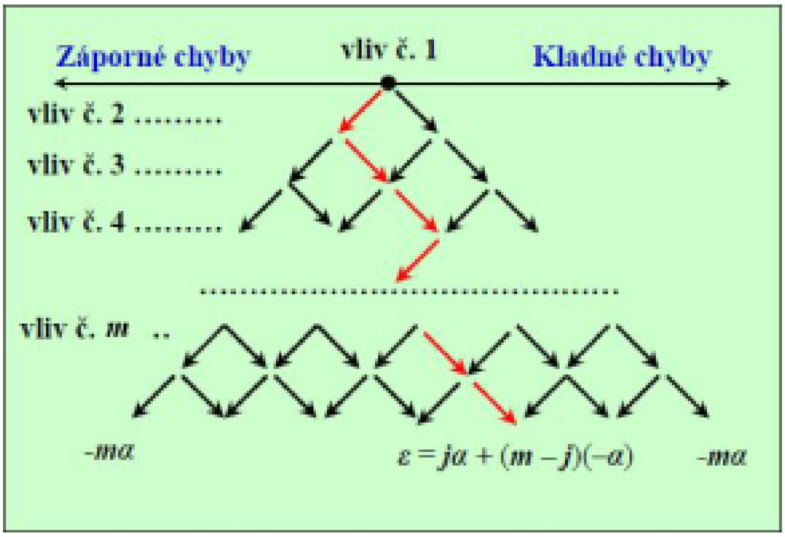
Veličina σ skutečně vyjadřuje standardní odchylku, platí

$$\langle x\rangle=0 \quad , \quad D(X)=\sigma^2$$

Na obrázku je Gaussovo rozdělení pro $\sigma=1$ (modrá křivka), $\sigma=2$ (červená křivka) a $\sigma=3$ (zelená křivka).



Předpokládejme, že existuje nějaká „správná“ hodnota měřené veličiny *x* a že při opakovaném měření získáme hodnoty $(x_0, x_1, \ldots, x_n)$, některé mohou být i stejné. Odchylky od (zatím neznámé) správné hodnoty označme $(\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_n)$. Tyto hodnoty jsou hodnotami náhodné veličiny *E*. Její hustotu pravděpodobnosti označme $w(E)$. Za jistých podmínek je rozdělením veličiny *E* rozdělení normální. Předpokládejme, že odchylky jsou způsobeny *m* nezávislými vlivy, každý z nich odchýlí měřenou hodnotu od *x* o stejnou hodnotu α, kladnou nebo zápornou, s pravděpodobností $1/2$. Kladnou odchylku ($+\alpha$) nazveme zdarem, zápornou ($-\alpha$) nezdarem. Výsledná odchylka naměřené hodnoty ε od *x* leží v intervalu $(-m\alpha, m\alpha)$ a může nabývat pouze celých násobků α.



Pravděpodobnost odchýlení o j kladných a $m-j$ záporných vlivů (j zdarů a $m-j$ nezdarů), tj. pravděpodobnost vzniku odchylky $\varepsilon = j\alpha + (m-j)(-\alpha)$, tj. $\varepsilon = (2j-m)\alpha$ je dána binomickým rozdělením (Bernoulliovým pro $q = 1/2$)

$$p_j = \frac{1}{2^m}\frac{m!}{j!(m-j)!}$$

Pro velká m je lze nahradit rozdělením normálním (Gaussovým) – důkaz naznačíme na konci kapitoly. Předpoklad o normální rozdělení umožňuje následující zpracování výsledků: reprezentativní hodnota měření představuje *aritmetický průměr* všech naměřených hodnot (střední hodnota veličiny)

$$\overline{x} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$$

(číslujeme teď měření od jedničky, nikoliv od nuly). Tento aritmetický průměr neurčuje správnou hodnotu veličiny. Lze však tvrdit, že s pravděpodobností 68,3% tato správná hodnota leží v intervalu určeném aritmetickým průměrem a *směrodatnou odchylkou* příslušnou aritmetickému průměru

$$x \in (\overline{x} - \overline{\sigma}, \overline{x} + \overline{\sigma})$$

což zapisujeme také takto

$$x = \overline{x} \pm \overline{\sigma}$$

Směrodatnou odchylku aritmetického průměru počítáme pomocí odchylek od aritmetického průměru jako

$$\overline{\sigma}^2 = \frac{1}{n(n-1)}\left[\left(x_1 - \langle x\rangle\right)^2 + \left(x_2 - \langle x\rangle\right)^2 + \ldots + \left(x_n - \langle x\rangle\right)^2\right]$$

Jen pro zajímavost uveďme, že pokud bychom chtěli psát výraz pomocí odchylek od skutečné hodnoty (to však z praktického hlediska nemá smysl, skutečnou hodnotu neznáme), měl by výraz tvar

$$\bar{\sigma}^2 = \frac{1}{n^2}\left[\left(x_1 - x\right)^2 + \left(x_2 - x\right)^2 + \ldots + \left(x_n - x\right)^2\right]$$

Vezmeme-li interval daný tzv. krajní chybou (trojnásobek směrodatné odchylky), pak skutečná hodnota měřené veličiny se v tomto intervalu nachází s pravděpodobností 97%.

10.4 Jaký průměr?

Příklad 1. Student měl ze tří matematických písemek v semestru tři hodnocení B a jedno C. U dvou závěrečných písemek měl A a D, u ústní zkoušky E. Jaká je jeho průměrná známka, jestliže všechny známky mají stejnou váhu? Označíme z_j hodnoty, n_j jejich četnosti a w_j váhy. Obecný vztah pro vážený průměr je

$$\langle z \rangle = \frac{\sum_{j=1}^{n} z_j\, n_j\, w_j}{\sum_{j=1}^{n} n_j\, w_j}$$

V našem případě jsou všechny váhy stejné, potom se výraz zjednoduší na

$$\langle z \rangle = \frac{\sum_{j=1}^{n} z_j\, n_j}{\sum_{j=1}^{n} n_j}$$

Máme tedy

$$\langle z \rangle = \frac{1{,}5\cdot 3 + 2\cdot 1 + 1\cdot 1 + 2{,}5\cdot 1 + 3\cdot 1}{3+1+1+1+1} = \frac{13}{7} = 1{,}86\ldots\ldots\mathrm{C}$$

Příklad 2. Řešte předchozí příklad za předpokladu, že závěrečná písemka má dvakrát větší váhu než průběžná a ústní zkouška má dvakrát větší váhu ne. závěrečná písemka:

$$\langle z \rangle = \frac{1{,}5\cdot 3\cdot 1 + 2\cdot 1\cdot 1 + 1\cdot 1\cdot 2 + 2{,}5\cdot 1\cdot 2 + 3\cdot 1\cdot 4}{3\cdot 1 + 1\cdot 1 + 1\cdot 2 + 1\cdot 2 + 1\cdot 4} = \frac{25{,}5}{12} = 2{,}12\ldots\ldots\mathrm{C}$$

Příklad 3. Automobil jel z A do B první úsek rychlostí 130 km/h a stejnou dobu druhý úsek průměrnou rychlostí 70 km/h. Jaká byla jeho průměrná rychlost na celé trase? Je odpověď dána aritmetickým průměrem obou hodnot? Je to věc definice. Průměrná je definována jako podíl celkové dráhy a celkové doby jízdy. Dráhu ale neznáme. Víme však, že oba úseky trvaly stejně času. V takovém případě by bylo

$$\langle v\rangle = \frac{s_1+s_2}{t_1+t_2} = \frac{v_1 t + v_2 t}{2t} = \frac{v_1+v_2}{2} = 100\,\text{km/h}$$

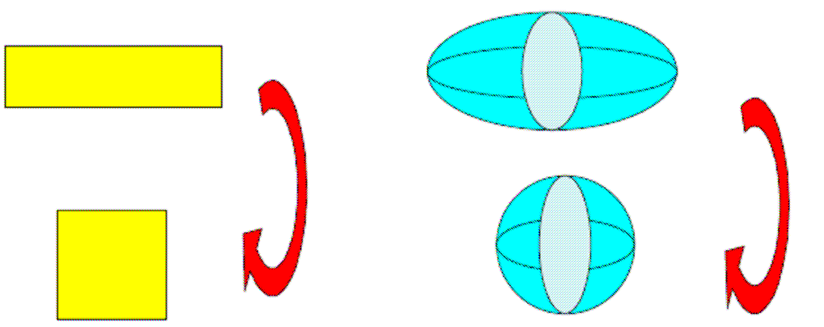
Takže přece jen aritmetický průměr? Zkusme úlohu obměnit.

Příklad 4. Automobil jel z A do B první úsek rychlostí 130 km/h a druhý úsek rychlostí 70 km/h. Oba úseky byly stejně dlouhé. Jaká byla nyní průměrná rychlost?

$$\langle v\rangle = \frac{s_1+s_2}{t_1+t_2} = \frac{2s}{\frac{s}{v_1}+\frac{s}{v_2}} = \frac{2}{\frac{1}{v_1}+\frac{1}{v_2}} = \frac{2v_1 v_2}{v_1+v_2} = 91\,\text{km/h}$$

Jedná se o tzv. harmonický průměr.

Příklad 5. „Průměrný“ obdélník je čtverec (určete stranu čtverce, který má stejný obsah jako obdélník o stranách a a b), „průměrný“ elipsoid je koule (určete poloměr koule, která má stejný objem jako elipsoid o poloosách a , b , c):



V prvním případě máme pro stranu čtverce

$$P = a\,b = x^2 \quad \Rightarrow \quad x = \sqrt{a\,b}$$

a v druhém případě pro poloměr koule

$$V = \frac{4}{3}\pi\,a\,b\,c = \frac{4}{3}\pi\,x^3 \quad \Rightarrow \quad x = \sqrt[3]{a\,b\,c}$$

V obou případech se jedná o tzv. geometrický průměr.

Pro obecný počet n hodnot máme následující výrazy pro výpočet průměru:

aritmetický průměr

$$\langle x\rangle_a = \frac{x_1+x_2+\cdots+x_n}{n}$$

geometrický průměr

$$\langle x\rangle_g = \sqrt[n]{x_1\,x_2\ldots x_n}$$

harmonický průměr

$$\langle x_h\rangle = \frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\cdots+\frac{1}{x_n}}$$

Platí nerovnost

$$\langle x\rangle_a \geq \langle x\rangle_g \geq \langle x\rangle_h$$

Pro $n=2$ je to hned vidět (napišme nerovnosti pro druhé mocniny průměrů)

$$\langle x\rangle_a^2 \geq \langle x\rangle_g^2 \quad \Leftarrow \quad \frac{\left(x_1+x_2\right)^2}{4} \geq x_1 x_2 \quad \Leftarrow \quad \frac{1}{4}\left(x_1+x_2\right)^2 - x_1 x_2 = \left(x_1 - x_2\right)^2 \geq 0$$

$$\langle x\rangle_g^2 \geq \langle x\rangle_h^2 \quad \Leftarrow \quad x_1 x_2 \geq \frac{4 x_1^2 x_2^2}{x_1^2 + 2 x_1 x_2 + x_2^2} \quad \Leftarrow \quad x_1 x_2 \frac{\left(x_1+x_2\right)^2}{4} \geq x_1^2 x_2^2$$

10.5 Přechod od Bernoulliova ke Gaussovu rozdělení

Bernoulliovo rozdělení je

$$x_j = j \quad , \quad p_j = \frac{n!}{j!\left(n-j\right)!} q^j \left(1-q\right)^{n-j} \quad , \quad 0 \leq q \leq 1 \quad , \quad j = 0,1,\ldots,n$$

Logaritmus pravděpodobnosti je

$$\ln\left(p_j\right) = \ln\left(n!\right) - \ln\left(j!\right) - \ln\left(\left[n-j\right]!\right) + j\ln\left(q\right) + \left(n-j\right)\ln\left(1-q\right)$$

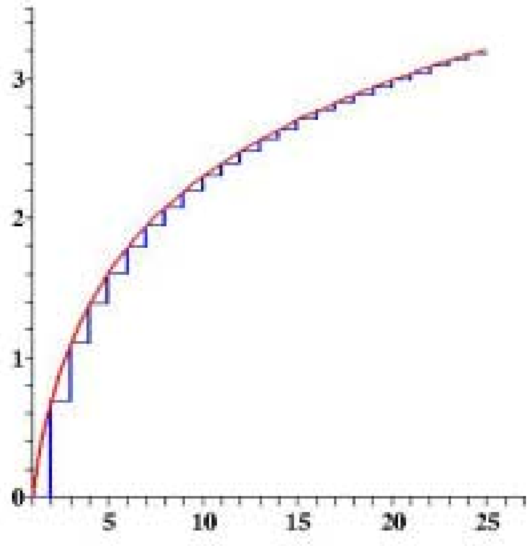
Stejně jako při přechodu k Poissonovu rozdělení předpokládáme $n \to \infty$, ale teď budeme předpokládat, že počet zdarů a nezdarů se nebude příliš lišit, tj. také j a $n-j$ jsou velká čísla. Základem pro výpočet bude přibližné vyjádření $\ln\left(k!\right) = \sum_{r=1}^{k} \ln\left(r\right)$ (logaritmus součinu je roven součtu logaritmů součinitelů) pro velké hodnoty k. Protože logaritmus je monotónně rostoucí funkce, můžeme napsat nerovnost, kdy integrál z logaritmu na intervalu jednotkové délky je větší než hodnota logaritmu ve spodní mezi a menší než hodnota logaritmu v horní mezi

$$\int_{r-1}^{r} \ln\left(\xi\right) d\xi < \ln\left(r\right) < \int_{r}^{r+1} \ln\left(\xi\right) d\xi$$

Sečtením přes r od $r=1$ do $r=k$ dostáváme

$$\int_{0}^{k} \ln\left(\xi\right) d\xi < k\ln\left(k\right) - k < \ln\left(k!\right) < \int_{1}^{k+1} \ln\left(\xi\right) d\xi = \left(k+1\right)\ln\left(k+1\right) - k$$

neboť $\int \ln(x)\,dx = x\ln(x) - x$. Na obrázku vidíme, proč už pro relativně malé



hodnoty můžeme sumaci $\sum_j \ln(j)$ nahradit integrací $\int \ln(x)\,dx$ – tedy diskretní proměnnou j spojitou proměnnou x. Za hodnotu $\ln(k!)$ bychom mohli vzít průměr z obou krajních hodnot v nerovnosti, my vezmeme ještě o něco lepší aproximaci, zvanou Stirlingova formule (její odvození už je však trochu komplikovanější)

$$\ln(k!) \doteq \left(k + \frac{1}{2}\right)\ln(k) - k - \frac{1}{2}\ln(2\pi)$$

Pro logaritmus pravděpodobnosti budeme teď mít

$$P(x) = \ln\left[p(x)\right] =$$
$$-\frac{1}{2}\ln(2\pi) + \left(n + \frac{1}{2}\right)\ln(n) - \left(x + \frac{1}{2}\right)\ln(x) - \left(n - x + \frac{1}{2}\right)\ln(n-x) + x\ln(q) + (n-x)\ln(1-q)$$

Maximum hustoty pravděpodobnosti bude u takové hodnoty x, pro kterou je derivace funkce $p(x)$ (a tedy i jejího logaritmu) rovna nule. Dostáváme tak rovnici

$$\frac{dP(x)}{dx} = \ln\frac{q(n-x)}{(1-q)x} + \frac{1}{2}\left(\frac{1}{n-x} - \frac{1}{x}\right) = 0$$

Pro druhou derivaci máme vztah

$$\frac{d^2P(x)}{dx^2} = -\left(\frac{1}{n-x} + \frac{1}{x}\right) + \frac{1}{2}\left[\frac{1}{(n-x)^2} + \frac{1}{x^2}\right]$$

První derivace je rovna nula pro hodnotu x, která je přibližně (pro $q = 1/2$ přesné) rovna střední hodnotě $x = \langle x \rangle = nq$. Ve stejném přiblížení máme

$$P(\langle x\rangle) = -\frac{1}{2}\ln(2\pi\sigma^2) \quad , \quad \frac{d^2P(\langle x\rangle)}{dx^2} = -\frac{1}{\sigma^2} \quad ,$$

kde σ je směrodatná odchylka $\sigma^2 = nq(1-q)$. Vezmeme-li teď první členy Taylorova rozvoje $P(x)$ kolem $x=\langle x\rangle$, dostáváme

$$P(x) = -\frac{1}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}(x-\langle x\rangle)^2$$

a po odlogaritmování dostáváme skutečně Gaussovo rozdělení se středem v $x=\langle x\rangle$

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

Na obrázcích je porovnáno binomické rozdělení s normálním rozdělením pro $n=25$ (vlevo) a $n=50$ (vpravo).