Catalanova čísla

Antonín Slavík
Matematicko-fyzikální fakulta UK

Catalanova čísla jsou po Fibonacciho číslech patrně druhou nejznámější posloupností čísel, která se vyskytuje v mnoha kombinatorických úlohách. Začneme dvěma úlohami, které nás dovedou k definici Catalanových čísel.

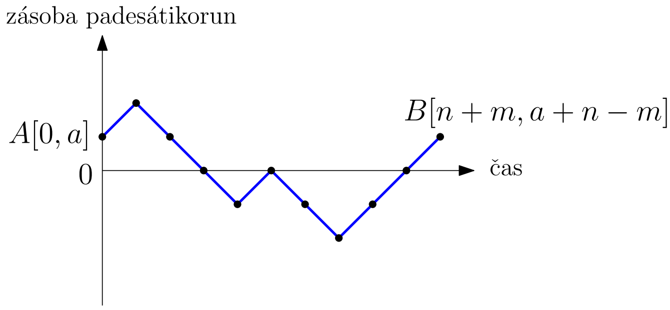
1 Úloha o frontě před pokladnou

U pokladny, kde se prodávají vstupenky v hodnotě 50 Kč, stojí ve frontě $n+m$ osob; n z nich má padesátikorunu, zatímco zbylých m má pouze stokorunu. Jestliže pořadí osob je náhodné a pokladník má počáteční zásobu a padesátikorun, jaká je pravděpodobnost, že bude moci vrátit peníze každé osobě se stokorunou?

Pokud pokladník bude moci vrátit peníze všem osobám se stokorunou, budeme zkráceně říkat, že fronta projde. Stačí se omezit na případ, kdy $a+n \geq m$, jinak je jasné, že fronta neprojde (protože celkový počet padesátikorun bude menší než počet stokorun).

Ve frontě není podstatné pořadí osob, záleží pouze na tom, kdo má padesátikorunu a kdo stokorunu. Frontu si tedy můžeme představit jako uspořádanou $(n+m)$-tici sestavenou ze symbolů 50 a 100. Počet všech front je $\binom{n+m}{n}$, neboť fronta je jednoznačně určena výběrem n pozic pro padesátikoruny;[1] toto číslo bude ve jmenovateli hledané pravděpodobnosti.

Zbývá vypočítat počet příznivých případů, tj. front, které projdou. Ve skutečnosti je jednodušší určit počet front, které neprojdou. Každou frontu (bez ohledu na to, zda projde či neprojde) můžeme znázornit lomenou čarou tak, jak ukazuje obr. 1. Lomená čára znázorňuje pokladníkovu aktuální zásobu padesátikorun. Začíná tedy v bodě $A[0,a]$, každá osoba s padesátikorunou způsobí zvýšení zásoby o 1 a naopak každá osoba se stokorunou způsobí snížení zásoby o 1. Lomená čára tedy končí v bodě $B[n+m, a+n-m]$.

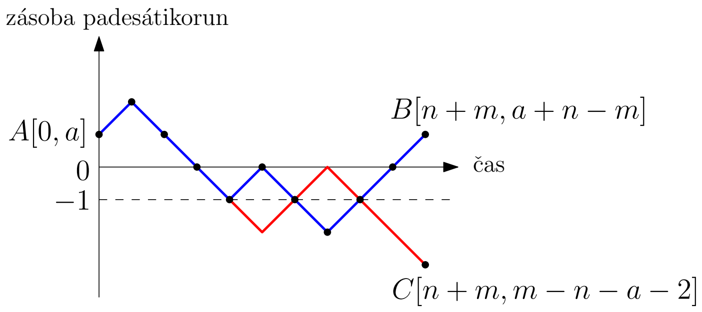


Obrázek 1: Grafické znázornění fronty 50, 100, 100, 100, 50, 100, 100, 50, 50, 50. Ve frontě je $m=5$ osob se stokorunou a $n=5$ osob s padesátikorunou, pokladník má na začátku $a=1$ padesátikorunu.

Na obrázku je příklad fronty, která neprojde. Počet všech takových front je roven počtu lomených čar z A do B, které v nějakém okamžiku klesnou pod osu x. Nyní přichází klíčová myšlenka řešení: Najdeme první

[1] Lze též použít vzorec pro permutace s opakováním: Počet permutací n padesátikorun a m stokorun je $\frac{(n+m)!}{n!\cdot m!} = \binom{n+m}{n}$.

okamžik, kdy čára klesne pod osu x, tj. dostane se do bodu s y-ovou souřadnicí -1. Od tohoto okamžiku začneme cestu zrcadlit podél přímky $y = -1$, viz obr. 2.



Obrázek 2: V okamžiku, kdy lomená čára poprvé protne přímku $y = -1$ (čárkovaně), ji podél této přímky začínáme zrcadlit (červený úsek).

Takto získaná cesta skončí v bodě C, který je zrcadlovým obrazem B vzhledem k přímce $y = -1$. Jelikož bod B leží ve výšce $a+n-m+1$ nad touto přímkou, bod C má y-ovou souřadnici $-1-(a+n-m+1) = m-n-a-2$. Naopak je zřejmé, že každá lomená čára z A do C někde protíná přímku $y = -1$ a zrcadlením příslušného úseku získáme cestu z A do B. Tudíž počet všech lomených čar z A do B, které klesnou pod osu x, je roven počtu všech lomených čar z A do C. Abychom zjistili počet těchto cest, potřebujeme znát počty stoupání a klesání na cestě z A do C.

Uvažujme cestu z A do B, která v nějakém okamžiku protne přímku $y = -1$. Část cesty před tímto okamžikem nazývejme 1. úsek a další část pak 2. úsek. Nechť ℓ je počet stoupání na 1. úseku cesty. Protože 1. úsek začíná ve výšce a a končí ve výšce -1, musí být počet klesání roven číslu $a + \ell + 1$. Jelikož celkové počty stoupání a klesání jsou n a m, musí být na 2. úseku $n - \ell$ klesání a $m - a - \ell - 1$ stoupání. Získané počty shrnuje následující tabulka.

	stoupání	klesání
1. úsek	ℓ	$a + \ell + 1$
2. úsek	$n - \ell$	$m - a - \ell - 1$

Odpovídající cesta z A do C vznikne zrcadlením 2. úseku, počty stoupání a klesání ve druhém řádku tabulky se tedy vymění:

	stoupání	klesání
1. úsek	ℓ	$a + \ell + 1$
2. úsek	$m - a - \ell - 1$	$n - \ell$

Celkový počet stoupání na cestě z A do C je proto $\ell + m - a - \ell - 1 = m - a - 1$ a celkový počet klesání je $a + \ell + 1 + n - \ell = a + n + 1$. To znamená, že počet všech cest z A do C je roven

$$\binom{n+m}{a+n+1}.$$

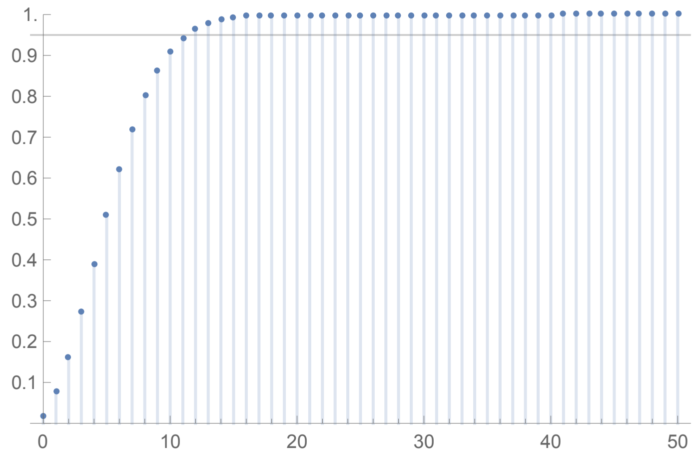
Toto číslo udává počet všech front, které neprojdou. Počet front, které projdou, je

$$\binom{n+m}{n} - \binom{n+m}{a+n+1}$$

a hledaná pravděpodobnost je

$$P = \frac{\binom{n+m}{n} - \binom{n+m}{a+n+1}}{\binom{n+m}{n}}. \tag{1.1}$$

Uvažujme pro ilustraci frontu se 100 osobami, z nichž polovina má padesátikorunu a polovina stokorunu, tj. $m = n = 50$. Pokud by pokladník chtěl mít jistotu, že fronta projde, musí si připravit 50 padesátikorun. Kolik padesátikorun by mu stačilo, kdy se spokojil s 95% pravděpodobností, že fronta projde? Hledáme nejmenší $a \in \mathbb{N}_0$ takové, aby platilo $P \geq 0{,}95$. Postupným dosazováním $a = 0, 1, 2, \ldots, 50$ do vztahu (1.1) získáme hodnoty znázorněné na obr. 3, ze kterého je vidět, že k dosažení pravděpodobnosti 95 % stačí vzít $a \geq 12$ padesátikorun.

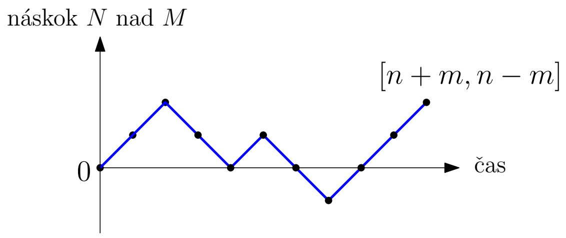


Obrázek 3: Pravděpodobnost, že fronta pro $m = n = 50$ projde v závislosti na volbě $a \in \{0, \ldots, 50\}$

2 Hlasovací problém

Ve volbách soupeří dva kandidáti M a N. Kandidát M získal m hlasů a kandidát N obdržel n hlasů, přičemž $n \geq m$. Jestliže pořadí sčítání hlasů bylo náhodné, jaká je pravděpodobnost, že v průběhu sčítání měl N vždy aspoň tolik hlasů jako M?

Každý průběh sčítání můžeme znázornit lomenou čarou tak, jak ukazuje obr. 4. Lomená čára znázorňuje aktuální náskok kandidáta N nad jeho soupeřem M. Začíná tedy v bodě $[0, 0]$, každý hlas pro N zvýší náskok o 1 a naopak každý hlas pro M sníží náskok o 1. Lomená čára tedy končí v bodě $[n + m, n - m]$.



Obrázek 4: Grafické znázornění sčítání pro posloupnost hlasů N, N, M, M, N, M, M, N, N, N

Skutečnost, že v průběhu sčítání měl N vždy aspoň tolik hlasů jako M, poznáme tak, že lomená čára nikdy neklesne pod osu x.

Vidíme, že hlasovací problém je totožný se speciálním případem úlohy o frontě před pokladnou, kde pokladník nemá na začátku žádné padesátikoruny. Hledanou pravděpodobnost tedy získáme dosazením $a = 0$ do vztahu (1.1):

$$P = \frac{\binom{n+m}{n} - \binom{n+m}{n+1}}{\binom{n+m}{n}}.$$

Rozepsáním kombinačních čísel lze výsledek zjednodušit:

$$P = 1 - \frac{\frac{(n+m)\cdots m}{(n+1)!}}{\frac{(n+m)\cdots(m+1)}{n!}} = 1 - \frac{m}{n+1} = \frac{n+1-m}{n+1}.$$

Pro kontrolu prozkoumejme, co nastane ve dvou speciálních případech:

- Pokud hlasování skončilo remízou, tj. $m = n$, pak $P = \frac{1}{n+1}$. Pro velká n je tedy pravděpodobnost, že jeden z kandidátů měl vždy aspoň tolik hlasů jako druhý, velmi malá.
- Pokud kandidát N zvítězil s velkou převahou, tj. $n \gg m$, pak pravděpodobnost $P = 1 - \frac{m}{n+1}$ je blízká k jedné. Je tudíž velmi pravděpodobné, že vítěz měl v průběhu sčítání vždy aspoň tolik hlasů jako poražený.

3 Catalanova čísla

Z předchozích oddílů víme, že

- počet front s n padesátikorunami a n stokorunami, které projdou, pokud pokladník nemá na začátku žádnou padesátikorunu,
- nebo počet pořadí sčítání hlasů ve volbách, kde oba kandidáti získali shodně n hlasů, přičemž kandidát N měl v průběhu sčítání vždy aspoň tolik hlasů jako kandidát M,

je roven číslu

$$\binom{2n}{n} - \binom{2n}{n+1}.$$

Rozepsáním kombinačních čísel obdržíme

$$\binom{2n}{n} - \binom{2n}{n+1} = \frac{2n\cdots(n+1)}{n!} - \frac{2n\cdots n}{(n+1)!} = \frac{2n\cdots(n+1)}{n!}\left(1 - \frac{n}{n+1}\right) = \binom{2n}{n}\frac{1}{n+1}.$$

Definice 3.1. Čísla $C_n = \frac{1}{n+1}\binom{2n}{n}$, $n \in \mathbb{N}_0$, se nazývají Catalanova čísla.

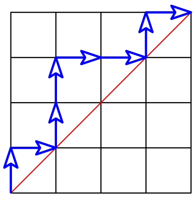
Následující tabulka ukazuje Catalanova čísla C_n pro nízké hodnoty n.

n	0	1	2	3	4	5	6	7	8	9
C_n	1	1	2	5	14	42	132	429	1430	4862

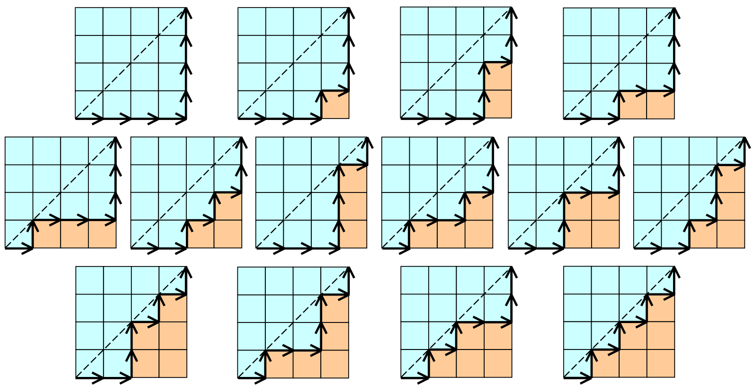
Poznámka 3.2. Catalanova čísla jsou pojmenována na počest belgického matematika Eugèna Charlese Catalana (1814–1894), který řešil následující úlohu: *Kolika způsoby lze uzávorkovat součin $a_0a_1\cdots a_n$ tak, aby pořadí násobení bylo jednoznačně určeno?* (Můžeme si představovat, že operace násobení není asociativní, a hodnota součinu tedy závisí na uzávorkování.) Např. pro $n = 3$ existuje pět způsobů: $(a_0a_1)(a_2a_3)$, $((a_0a_1)a_2)a_3$, $a_0(a_1(a_2a_3))$, $a_0((a_1a_2)a_3)$, $(a_0(a_1a_2))a_3$. Lze ukázat, že počet způsobů pro obecné n je C_n.

Patrně nejjednodušší kombinatorická interpretace Catalanových je následující: Mějme čtverec $n\times n$ rozdělený úsečkami na n^2 jednotkových čtverečků. Uvažujme cesty vedoucí po hranách této mříže, a to z levého dolního do pravého horního rohu, přičemž každý krok vede buď vpravo, nebo nahoru. Pak počet cest, které nikdy nesestoupí pod diagonálu vedoucí z levého dolního do pravého horního rohu, je C_n.

Proč platí toto tvrzení? Obrázek 5 ukazuje příklad přípustné cesty. Otočíme-li tento obrázek o 45 stupňů ve směru pohybu hodinových ručiček, dostáváme lomenou čáru stejného typu, jako v hlasovacím problému, kde oba kandidáti získali n hlasů. Obě úlohy jsou tedy ekvivalentní a počet přípustných cest je skutečně C_n.



Obrázek 5: Příklad cesty ve čtverci 4×4, která nikdy nesestoupí pod diagonálu



Obrázek 6: Cesty ve čtverci 4×4, které nikdy nevystoupí nad diagonálu (Wikimedia Commons, `en.wikipedia.org/wiki/Catalan_number#/media/File:Catalan_number_4x4_grid_example.svg`)

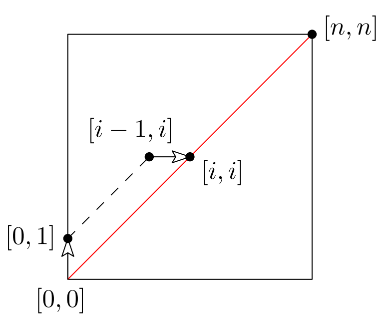
Ze symetrie je zřejmé, že číslo C_n udává rovněž počet cest, které nikdy nevystoupí nad diagonálu, viz obr. 6.

Interpretace Catalanových čísel pomocí cest v čtvercové síti nám umožní odvodit následující rekurentní vzorec pro C_n.

Věta 3.3. *Pro každé $n \in \mathbb{N}$ platí*

$$C_n = C_0C_{n-1} + C_1C_{n-2} + C_2C_{n-3} + \cdots + C_{n-1}C_0 = \sum_{i=1}^{n} C_{i-1}C_{n-i}. \tag{3.1}$$

Důkaz. Víme, že C_n udává počet cest ve čtverci $n \times n$ vedoucích z levého dolního rohu $[0, 0]$ do pravého horního rohu $[n, n]$, které nikdy nesestoupí pod diagonálu. Pokud si odmyslíme počáteční bod $[0, 0]$, musí vždy nastat ještě další okamžik, kdy se cesta dotkne diagonály (nejpozději v bodě $[n, n]$). Nechť první takový okamžik nastane v bodě $[i, i]$, kde $i \in \{1, \ldots, n\}$, viz obr. 7.



Obrázek 7: Cesta z bodu $[0, 0]$ do bodu $[n, n]$, která se poprvé dotkne diagonály v bodě $[i, i]$

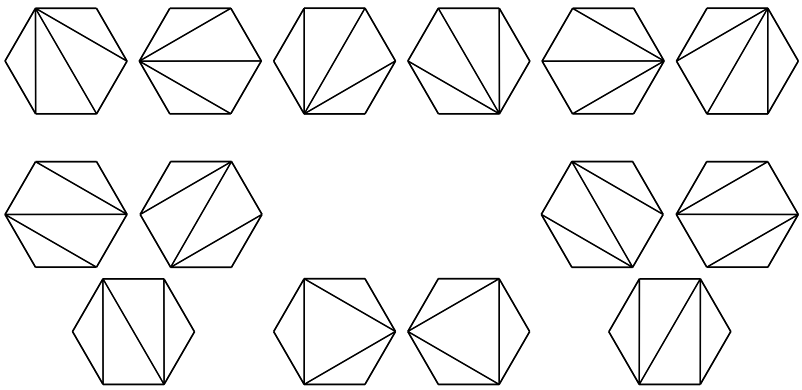
První krok cesty musel vést z bodu $[0, 0]$ do bodu $[0, 1]$ a do bodu $[i, i]$ se cesta nutně musela dostat z bodu

$[i-1, i]$. Mezi body $[0, 1]$ a $[i-1, i]$ nemohla sestoupit pod úsečku spojující tyto dva body (jinak by se dotkla diagonály dříve než v $[i, i]$). Počet způsobů, jak se cesta mohla dostat z $[0, 1]$ do $[i-1, i]$, je C_{i-1}, neboť jde vlastně o cestu ve čtverci o straně $i-1$, která nesmí sestoupit pod diagonálu. Podobně počet způsobů, jak se cesta mohla dostat z bodu $[i, i]$ do bodu $[n, n]$, je C_{n-i}. Vidíme, že počet cest z bodu $[0, 0]$ do bodu $[n, n]$, které se poprvé dotknou diagonály v bodě $[i, i]$, je roven $C_{i-1}C_{n-i}$. Nasčítáním přes všechna možná $i \in \{1, \ldots, n\}$ dostaneme celkový počet cest C_n, čímž je věta dokázána. $\square$

Vztah (3.1) umožňuje vypočítat C_n ze znalosti $C_0, \ldots, C_{n-1}$. Společně s podmínkou $C_0 = 1$ tedy jednoznačně určuje Catalanova čísla. K výpočtu C_n se lépe hodí explicitní vzorec z definice 3.1, přesto je však rekurentní vzorec užitečný. V úlohách vedoucích na Catalanova čísla může být obtížné přímo ověřit, že uvažovaná posloupnost čísel je popsána vzorcem $\frac{1}{n+1}\binom{2n}{n}$; často je mnohem jednodušší ukázat, že posloupnost splňuje stejný rekurentní vztah jako C_n. Následující příklad ilustruje tuto skutečnost.

Kolika způsoby lze rozdělit konvexní mnohoúhelník pomocí neprotínajících se úhlopříček na trojúhelníky?

Připomeňme, že úhlopříčkou se rozumí úsečka spojující nesousední vrcholy. Pro trojúhelník je pouze 1 možnost (nekreslíme žádnou úhlopříčku), pro čtyřúhelník jsou 2 možnosti, pro pětiúhelník 5 možností (nakreslete si obrázek) a pro šestiúhelník 14 možností (viz obr. 8).



Obrázek 8: Dělení šestiúhelníku pomocí neprotínajících se úhlopříček na trojúhelníky (Wikimedia Commons, `en.wikipedia.org/wiki/Catalan_number#/media/File:Catalan-Hexagons-example.svg`)

Zdá se tedy, že pro každé $n \in \mathbb{N}$ udává C_n počet triangulací pro $(n+2)$-úhelník. Označíme tento počet T_n, dodefinujeme $T_0 = 1$ a ukážeme, že $T_n = C_n$ pro každé $n \in \mathbb{N}_0$.

Uvažujme libovolný konvexní $(n+2)$-úhelník a vyberme si jeho libovolnou stranu XY. Bez újmy na obecnosti můžeme předpokládat, že vrcholy X, Y jdou po sobě proti směru pohybu hodinových ručiček (jinak zaměníme jejich označení). V každé triangulaci musí být strana XY součástí nějakého trojúhelníku XYZ. Jdeme-li po obvodu mnohoúhelníku z bodu Y do bodu Z proti směru pohybu hodinových ručiček, projdeme přitom i stran mnohoúhelníku, kde $i \in \{1, \ldots, n\}$. Tato část obvodu společně s úsečkou ZY ohraničuje $(i+1)$-úhelník, který je možné triangulovat T_{i-1} způsoby. Cesta z bodu Z do bodu X (opět proti směru pohybu hodinových ručiček) obsahuje $n+1-i$ stran mnohoúhelníku. Společně s úsečkou XZ ohraničuje $(n+2-i)$-úhelník, který je možné triangulovat T_{n-i} způsoby. Pro pevné Z tedy existuje celkem $T_{i-1}T_{n-i}$ triangulací obsahujících trojúhelník XYZ.[2] Nasčítáme-li tyto počty přes všechny možné volby Z, tj. přes všechna možná $i \in \{1, \ldots, n\}$, dostaneme celkový počet triangulací $(n+2)$-úhelníku:

$$T_n = \sum_{i=1}^{n} T_{i-1}T_{n-i}.$$

Posloupnost $\{T_n\}_{n=0}^{\infty}$ tedy splňuje stejnou rekurentní rovnici jako Catalanova čísla, a protože $T_0 = C_0 = 1$, musí nutně platit $T_n = C_n$ pro každé $n \in \mathbb{N}_0$.

[2] Díky volbě $T_0 = 1$ je toto tvrzení pravdivé i pro $i = 1$ a $i = n$.