Binomická distribuce

Při zjišťování p je nutné znát:

- a) celkový počet možných jednoduchých jevů
- b) počet jednoduchých jevů který spadá do jevu/třídy jevů, jejichž pravděpodobnost zjišťujeme

Otázka: Kolika různými cestami může jev nastat?

Někdy jednoduché, někdy lepší použít početní pravidla = sekvence, permutace, uspořádané (variace) a neuspořádané kombinace

sekvence

- Jednoduchý jev může mít podobu série či sekvence – každý pokus náhodného experimentu rezultuje v právě jednu z K vzájemně vylučujících se a vyčerpávajících jevů a tento pokus je opakován N krát – výsledkem je sekvence
  - Př. Tahám kuličku a po každém pokusu ji vrátím zpět, opakuji čtyřikrát výsledek: Červená, Bílá, Černá, Bílá nebo možná Bílá, Bílá, Černá, Černá = dvě různé sekvence

- **Početní pravidlo 1 : počet možných sekvencí pro N pokusů**

  Pokud může nastat jakýkoli z K vzájemně vylučujících se a vyčerpávajících jevů při každém z N pokusů, pak existuje K^n různých sekvencí které mohou nastat

  - Př. 5 krát (N) házím mincí, přičemž každý hod může padnout jen pana nebo orel (K=2),

    pak existuje $K^n = 2^5 = 32$ sekvencí

- Někdy se počet jevů, které mohou nastat v různých pokusech experimentu liší,

  pak platí Pravidlo 2 pro sekvence: **Pokud K_1, K_2, ….K_n je počet různých jevů které mohou nastat při pokusech 1, 2,….,N v řadě, počet různých sekvencí N jevů je $K_1 * K_2 * \ldots * K_n$.**

- Př. Nejprve házím mincí (K=2), pak kostkou (K=6), pak počet sekvencí je

  2*6 = 12

permutace

- =Počet způsobů jakými může být sada různých objektů seřazena (platí i pro sekvence jevů)
- **Pravidlo 3 permutace:**

  Počet různých způsobů jakými může být seřazeno N různých předmětů je N! = 1*2*3….(N-1)*N, seřazení se nazývá permutace takže celkový počet permutací N objektů je N! (nazýváme N faktoriál)

- Př. Ve třídě je 10 židliček pro 10 studentů. Kolika způsoby můžou být studenti na židle rozděleni?
  - Postup: Kterýkoli ze studentů může být umístěn na židli 1 (10 možností pro židli 1). Pro židli 2 je ale už jen 9 možností protože jeden student už sedí na židli 1 atd. ……tedy 10! = 10*9*8*7…..*1 = 3 628 800 způsobů

- Př. 2: Učitel má v klobouku jména 5 studentů, tahá postupně jedno jméno bez vracení, kolik sekvencí je možné vytáhnout? 5! = 120
  - Jaká je p že jméno jakéhokoli dítěte bude vytaženo jako první? Postup: Pokud bude jakékoli dítě na první pozici, zbývá N-1 různých sekvencí (tj. (N-1)! = 4! = 24) v jakých mohou být seřazeni ostatní děti, tedy p jakéhokoli dítěte že bude vytaženo první je 24/120 = .20
  - Předpokládejme že v klobouku jsou dvě dívky, jaká je p že první dvě jména tažená budou dívky?
    - Postup: Pokud jsou první dvě tažená jména dívčí, pak zbývá 3!=6 způsobů jak vytáhnout zbývající 3 chlapce. Jména dívek mohou být vytažena dvěma způsoby. Proto p = (2*6) / 120 = .10

Uspořádané kombinace (variace)

- Někdy je nezbytné spočítat počet způsobů jakým r počet objektů může být vybrán z N objektů (přičemž r <= N)

- **Platí pravidlo: Počet způsobů vybrání a seřazení r objektů z N různých objektů je N! / (N – r)!**

- Př. 1. Ve třídě je 10 studentů ale jen 5 židlí. Kolika způsoby může učitel vybrat a umístit studenty na židle?

  Postup: Na počátku první židle která může být obsazena 10 možnými způsoby, pak druhá židle 9 způsoby….až pátá židle 6 způsoby, tedy 10*9*8*7*6 = 10! / (10-5!) = 30 240

- Př. 2. V loterii bylo rozdáno 40 losů různým lidem a třem vybraným byly určeny první, druhá a třetí výhra. Kolik může existovat různých sekvencí vítězů?

  40! / (40-3)! = 40! / 37! = 40 * 39 * 38 = 59 280

  - V kolika z těchto sekvencí by John Doe byl první, druhý nebo třetí?

    Pokud by byl první pak by počet možností na druhé a třetí místo byl 39! / 37! = 1 482. Podobně by existovalo 1482 sekvencí kde by byl vytažen jako druhý a stejný počet sekvencí kde by byl tažen jako třetí. Proto p(první, druhý nebo třetí) = (3*1482) / 59 280 = .075

Neuspořádané kombinace

- Při počtech pravděpodobnosti nás však často nezajímá pořadí jevů/událostí, ale jen počet způsobů jakým r věcí může být vybráno z N věcí. Předcházející pravidlo 4 říká, že celkový počet způsobu výběru r věcí z N a jejich seřazení je N! / (N-r)!. Každá sada r objektů má přitom r! možných seřazení podle pravidla 3. Kombinací těchto faktů získáme pravidlo 5:

- **Početní pravidlo 5: Celkový počet způsobů jak vybrat r různých kombinací z N objektů bez ohledu na pořadí je N! / (N-r)!r! (neboli N nad r, neboli binomický koeficient)**

- Př. Máme 33 kandidátů na pozici manažera, a 3 místa. Kolika různými způsoby můžeme vybrat 3 lidi do manažerských pozic?

  – Postup: r = 3, N = 33, takže 33 nad 3 = 33! / (33-3)!3! = (33*32*31) / 6 = 5 456, pokud jsou všechny trojice stejně pravděpodobné, pak je pravděpodobnost výběru jakékoli trojice 1/5456.

Bernoulliho proces

- Nejjednodušší pravděpodobnostní distribuce je ta s pouze dvěma třídami jevů
  - Př. Hod mincí – pana nebo orel
  - Př. Člověk je náhodně vybrán – muž nebo žena
- experiment/proces který může skončit pouze jedním z dvou výsledků = Bernoulliho pokus
- Dvě třídy jevů a jejich pravděpodobnosti = Bernoulliho proces
- Jedna událost = tzv. „úspěch“, druhá = tzv. „neúspěch“
- p = pravděpodnost „úspěchu“, q = 1 – p = pravděpodnost „neúspěchu“
  - Př. Hod mincí: pana=úspěch=p=1/2

    orel=neúspěch=q=1-p=1/2

výběr vzorku jako Bernoulliho proces

- Máme-li S v souladu s Bernoulihho procesem a vybíráme vzorek nezávisle s vracením anebo existuje nekonečný počet elementárních jevů tak, že p a q se nemění.
- Př. Provádíme N=5 pokusů. Kolik různých sekvencí můžeme dostat?

  - Pravidlo 1: $2^{5} = 32$, přičemž pravděpodobnost každé sekvence záleží na p a q. Protože pokusy jsou nezávislé, lze aplikovat rovnici o sdružených jevech $A \cap B = p(A)*p(B)$
  - Jaká je pravděpodobnost 3 úspěchů = panen?

    p (P,P,O,O,P)? Použitím rovnice o sdružených jevech dostáváme $p^2q^2p= p^3q^2$

- Stejná pravděpodobnost je i pro sekvenci (O,O,P,P,P), nezáleží tedy na pořadí úspěchů, jen na jejich počtu a pravděpodobnosti úspěchu

- **Pravděpodobnost jakékoli sekvence N nezávislých Bernoulliho pokusech záleží pouze na počtu úspěchů a jejich pravděpodobnosti: $p^r q^{N-r}$, $= p^r(1-p)^{N-r}$,**

  kde r=počet úspěchů a N-r=počet neúspěchů

- Sekvence 10 pokusů, nastanou 4 úspěchy: $p^4q^6 = p^4(1-p)^6$
  - Pokud např. p=2/3 pak $(2/3)^4(1/3)^6 = .00027$

- Př.2: Pokud hodíme 6krát mincí, jaká je pravděpodobnost že padnou 3 pany následované 3 orly (P,P,P,O,O,O)?
  - $p^3q^3 = (1/2)^3\,(1/2)^3 = 1/64$
  - Stejná pravděpodobnost je i pro sekvence (P,O,P,O,P,O) nebo (P,O,P,P,O,O) nebo jakoukoli jinou sekvenci obsahující 3 pany

- Většinou nás však nezajímají konkrétní sekvence s různým pořadím, ale pravděpodobnost určitého počtu „úspěchů“ bez ohledu na pořadí
  - př. pokud hodíme 5krát mincí, existuje podle pravidla 5 (5 nad 3) = 10 různých sekvencí s 3 úspěchy (pany), přičemž každá má stejnou pravděpodobnost, protože obsahují stejné množství úspěchů
  - Chceme znát pravděpodobnost padnutí 3 úspěchů bez ohledu na pořadí, tedy pravděpodobnost sekvence (P,P,P,O,O) nebo sekvence (P,O,P,O,P) nebo jakékoli jiné obsahující 3 pany:
    - Užitím „nebo“ pravidla pravděpodobnosti dostáváme p(3 úspěchy v 5 pokusech) = $p^3q^2 + p^3q^2 + \ldots + p^3q^2$ = (5 nad 3) p^3q^2 protože každá sekvence má stejnou pravděpodobnost a jejich celkem (5 nad 3) = 10

- **Obecně: Při výběru vzorku z Bernoulliho procesu, s pravděpodobností úspěchu = p, pravděpodobnost dosažení přesně r úspěchů v N nezávislých pokusech je p (r successes, N,p) = (N nad r) p^rq^{N-r}**

=

Binomický výběr

- Př. Binomického výběru

  - V populaci zvířat máme je 80% normálně zbarvených a 20% albínů bez pigmentace. Biolog vybírá náhodný vzorek o 3 zvířatech, jaká je pravděpodobnost že vybere jednoho albína?
    - p (1 albín ze 3 zvířat) = (3 nad 1) $.20^1 .80^2$ = .384

    Biolog má 38 příležitostí ze 100 najít 1 albína ve vzorku 3 náhodně vybraných zvířat

Počet úspěchů jako náhodná diskrétní (nespojitá) proměnná: binomická distribuce

- Proměnná má rozsah od 0 do N
- Distribuce páruje každý počet úspěchů s jejich pravděpodobnostmi
- **Každá náhodná proměnná X s pravděpodobnostní funkcí danou**
  p(x=r, N, p) = (N nad r) $p^r q^{N-r,}$ pro r = 1, 2, …., N,
  má binomickou distribuci s parametry N a p
- změníme-li parametry N a p, dostaneme jinou binomickou distribuci

- Př. Házíme 5 krát kostkou. Zajímá nás počet úspěchů (panen) = hodnota proměnné v každé sekvenci.
  - Postup: Začínáme nejvyšší hodnotou X=5.
    - Použitím pravidla 5 dostáváme:

      (5 nad 5) = 1 možných sekvencí v kterých jsou samé úspěchy,

      přičemž pravděpodobnost sekvence = p^5q^0

    Tedy p (X=5, N=5, p=0.5) = (5 nad 5) p^5 = p^5 = 1/32

  Stejně postupujeme pro X=4, 3, …, 0 a dostáváme:

  p (X=4, N=5, p)= (5 nad 4) p^4q^1 = 5 p^4q^1 = 5/32

  p (X=3, N=5, p)= (5 nad 3) p^3q^2= 10 p^3q^2 = 10/32

  p (X=2, N=5, p)= (5 nad 2) p^2q^3= 10 p^2q^3 = 10/32

  p (X=1, N=5, p)= (5 nad 1) p^1q^4= 5 p^1q^4 = 5/32

  p (X=0, N=5, p)= (5 nad 0) p^0q^5 = q^5 = 1/32

X	P(x)	
5	$(1/2)^{5}$	$=1/32$
4	$5\ (1/2)^{4(}1/2)^{1}$	$=5/32$
3	$10\ (1/2)^{3}(1/2)^{2}$	$=10/32$
2	$10\ (1/2)^{2}(1/2)^{3}$	$=10/32$
1	$5\ (1/2)^{1}(1/2)^{4}$	$=5/32$
0	$(1/2)^{5}$	$=1/32$
		$= 32/32$

Grafické znázornění distribuce

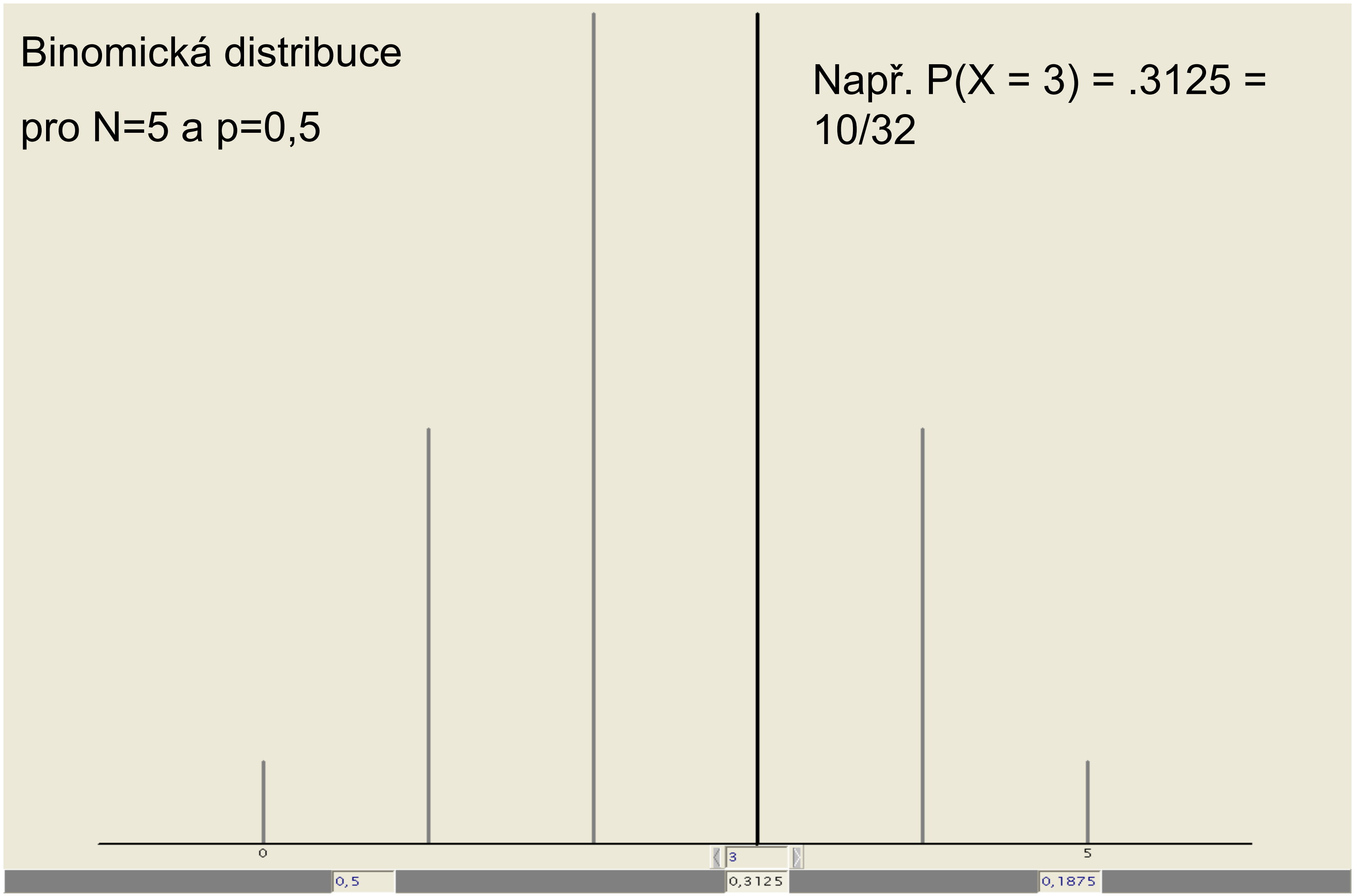
aneb

alternativní zjištění pravděpodobností konkrétního počtu úspěchů v N pokusech nebo intervalu proměnné (např. 3=>x>=2) přes PQRS

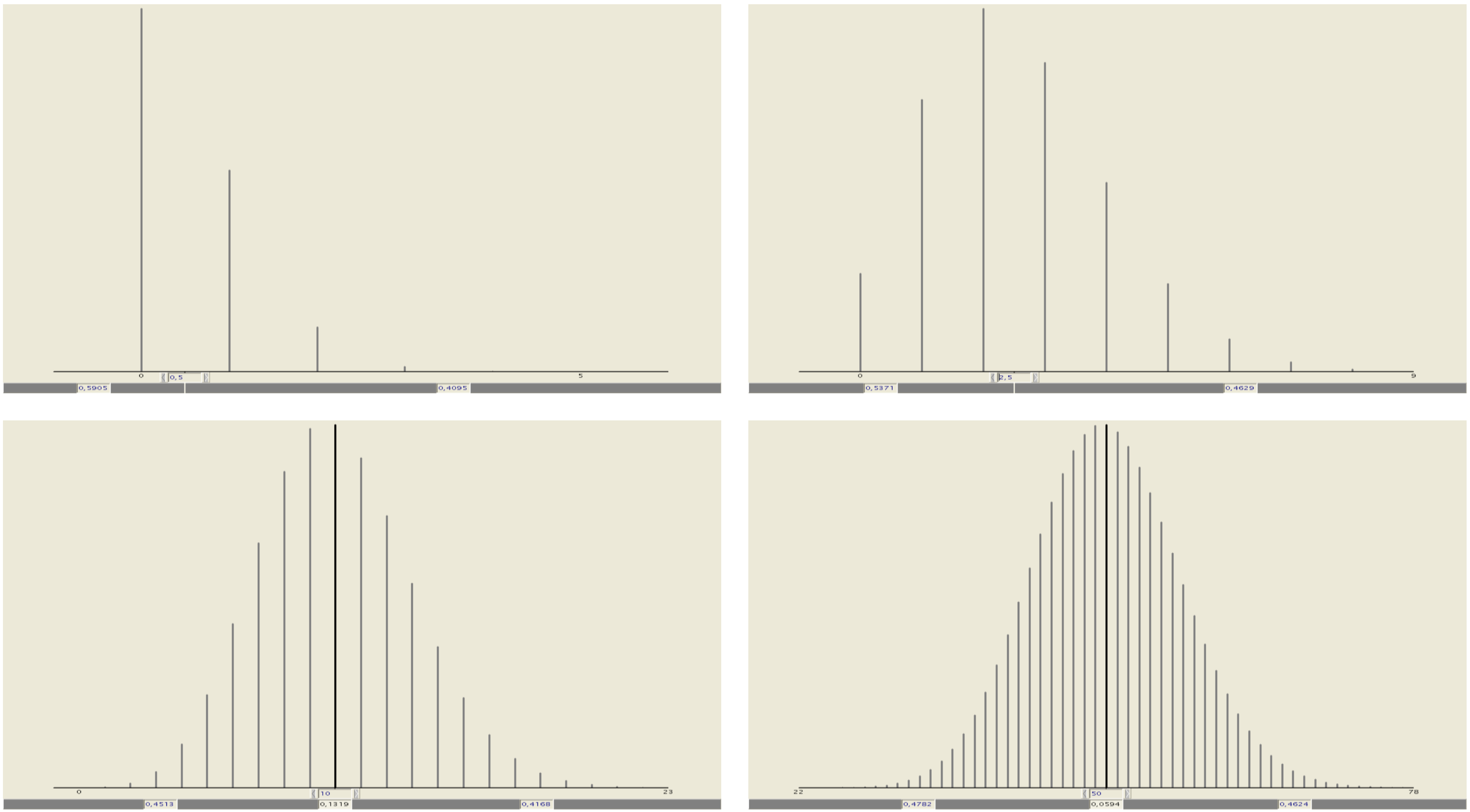
Binomická distribuce

pro N=5 a p=0,5

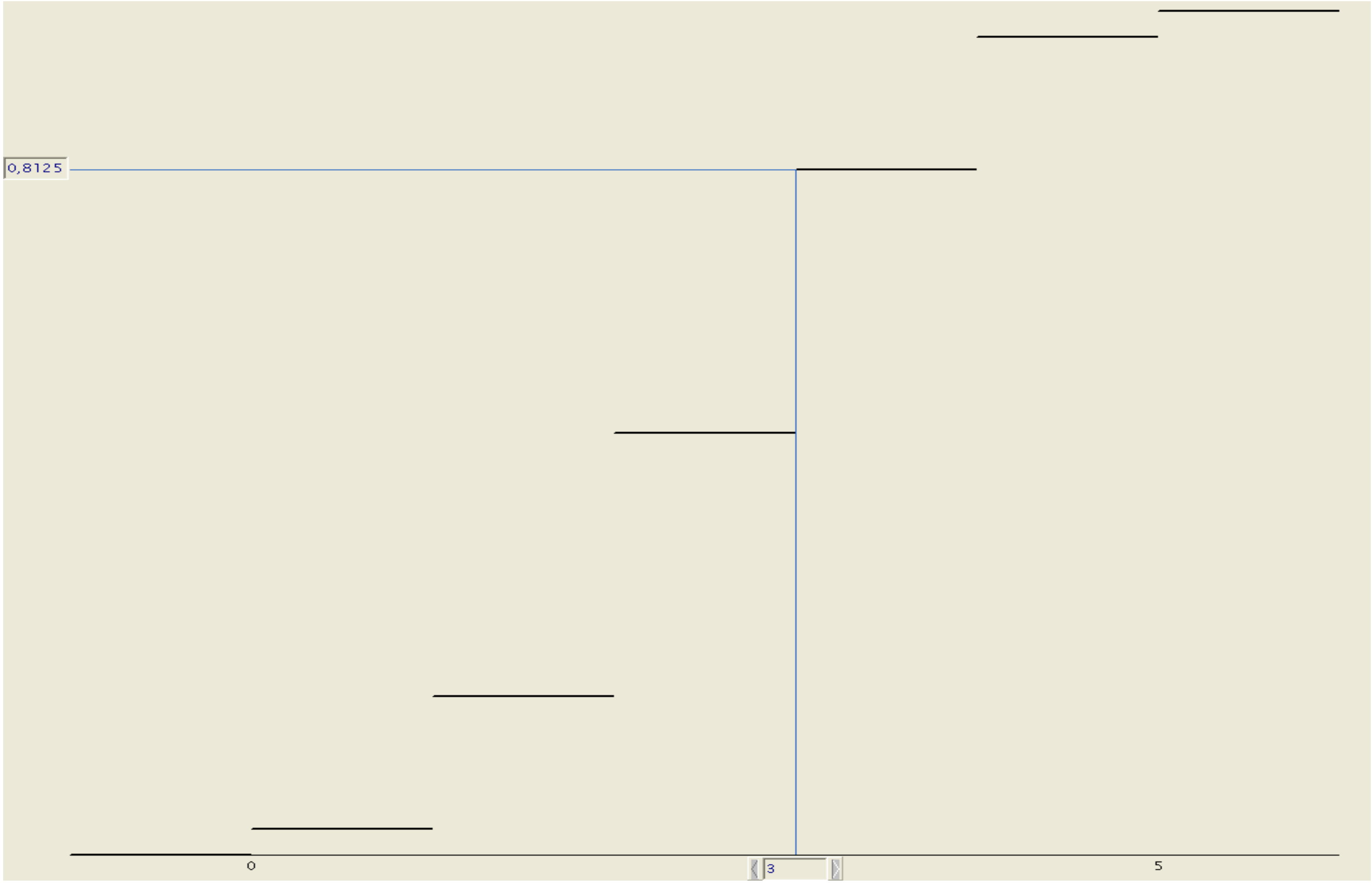
Např. P(X = 3) = .3125 = 10/32



Srovnání binomických distribucí pro N = 5, 25,100 a 500 (zleva nahoře) – se zvětšujícím se N se blíží normální distribuci



Kumulativní funkce hustoty pravděpodobnosti pro N=5 a p=0.5



Druhé alternativní zjištění počtu kombinací použitím magie ☺ - Pascalův trojúhelník

1
1 1
1 2 1
1 3 3 1
1 4 6 4 1
1 5 10 10 5 1
1 6 15 20 15 6 1

Např. při N = 5 (šestá řada)

X = 0 1 kombinace

X = 1 5 kombinací

X = 2 10 .

X = 3 10 .

X = 4 5 .

X = 5 1 kombinace

1

1 1

1 2 1

1 3 3 1

1 4 6 4 1

1 5 10 10 5 1

1 6 15 20 15 6 1

1 7 21 35 35 21 7 1

1 8 28 56 70 56 28 8 1

1 9 36 84 126 126 84 36 9 1

1 10 45 120 210 252 210 120 45 10 1

1 11 55 165 330 462 462 330 165 55 11 1

1 12 66 220 495 792 924 792 495 220 66 12 1

1 13 78 286 715 1287 1716 1716 1287 715 286 78 13 1

1 14 91 364 1001 2002 3003 3432 3003 2002 1001 364 91 14 1

1 15 105 455 1365 3003 5005 6435 6435 5005 3003 1365 455 105 15 1

1 16 120 560 1820 4368 8008 11 440 12 870 11 440 8008 4368 1820 560 120 16 1