Cvičenie 5 – Špeciálne diskrétne rozdelenia

Rovnomerné (uniform) R(n), Ro(n), U(n)
Ktorákoľvek z možností je rovnako pravdepodobná.

X	x_1	x_2	...	x_n
P	1/n	1/n	...	1/n

Alternatívne (Bernoulliho) $\mathcal{A}$(p)
Buď sa „to“ stane (1) alebo nie (0).
Buď sa stane x_1 alebo x_2, tretia možnosť nie je.

X	0	1
P	1-p	p

X	x_1	x_2
P	p	1-p

Binomické Bi(n,p)
„Pokus“ s alternatívnym rozdelením opakujeme n-krát. S akou pravdepodobnosťou (pri n pokusoch) sa „to“ stane 0, 1, 2, ..., n-krát?

X	0	1	2	...	n
P	p_0	p_1	p_2	...	p_n

Pre k=0,1, ..., n platí $p_k = C(n,k)*p^k*(1-p)^{n-k}$

Prečo práve takto?
Pravdepodobnosť, že z n pokusov to vyjde k-krát, je súčet pravdepodobností všetkých možných konkrétnych úspešných k-tic:

– Počet k vieme vybrať z počtu n práve C(n,k) spôsobmi (tj. konkrétne ktorých k pokusov z tých n bude úspešných)
– Ak sme vybrali konkrétnych k pokusov, pravdepodobnosť k-násobného úspechu je $p^k*(1-p)^{n-k}$, lebo k-krát to musí vyjsť (pravdepodnosť p) a (n-k)-krát (v tých pokusoch, ktoré sme nevybrali do vyvolenej k-tice) nevyjsť (pravd. 1-p).

Hypergeometrické H(N,M,n), Hg(N,M,n)
Máme na výber N možností a vieme, že M z nich má vlastnosť, na ktorej nám záleží (sú „prívlastkové“).
Náhodne z celej ponuky (N) vyberieme n kusov. Pýtame sa na pravdepodobnosť, že práve x kusov z tých n bude „prívlastkových“. Aby sme neriešili príliš okrajové situácie, dohodneme sa, že M nie je nula a vyberať budeme aspoň 1 kus (nenulové n).

Najprv trochu hraníc – musí platiť (vysvetlite pri každej, prečo je to tak):

$1 \leq M \leq N$
$1 \leq n \leq N$
$x \leq n, x \leq M$
$0 \leq x$
$M+n-N \leq x$

K poslednej nerovnosti dve nápovedy:

na lepšie pochopenie dôvodov to treba zapísať ako n - (N-M)

Ak máte 10 pív, z toho 7 ležiakov, vo výbere 5 kusov nemôžete mať menej ako 2 ležiaky

Skúmaná pravdepodobnosť pre x „prívlastkových“:

Menovateľ sú všetky možnosti, ako z N vybrať n.

Čitateľ sú všetky možnosti, ako z M vybrať x, násobené možnosťami, ako zo zvyšku neprívlastkových (N-M) vybrať to, čo chýba do počtu n, teda (n-x).

<u>Zhrnuté:</u> $P(x) = C(M,x) * C(N-M, n-x) \ / \ C(N, n)$

Geometrické G(p), Geo(p)

Opakujeme základný pokus s alternatívnym rozdelením dovtedy, kým sa „to“ stane. Napr. dovtedy chodíme s krčahom po vodu, kým nám nenalejú niečo lepšie.

Pýtame sa na pravdepodobnosť, že práve na k-ty pokus „to“ vyjde. Tá sa rovná súčinu pravdepodobnosti, že (k-1)-krát to nevyjde a následne to vyjde:

$P(k) = (1-p)^{k-1} * p, \qquad k=1,2,3,...$

Distribučná funkcia (kumulatívny súčet pravdepodobností) je súčet geometrického radu, odtiaľ pochádza názov rozdelenia.

$F(k) = p + (1-p)^1 * p+(1-p)^2 * p+ ... +(1-p)^{k-1} * p = 1 - (1-p)^k$

Poissonovo Po (λ), Pois (λ)

Týka sa spravidla náhodných udalostí v určitom časovom intervale (ale nie nutne). Za predpokladu, že udalosť zvykne nastávať v priemere n-krát za čas T, budeme skúmať pravdepodobnosť P, že udalosť v časovom intervale I nastane k-krát.

Ak sa „to“ stáva priemerne n-krát za T, môžeme očakavať, že za I sa „to“ stane λ-krát,

kde $\lambda=n*I/T$ (stredná hodnota našich očakávaní).

Hľadaná pravdepodobnosť je potom podľa Poissona

$P(k) = e^{-\lambda} * \lambda^k / k! \quad , \qquad k=0,1,2,3, ... , \infty$

Poissonovo rozdelenie sa používa tiež na aproximáciu binomického rozdelenia (pre veľké n, malé p), keďže

$$\lim_{n\to\infty,\ p\to 0} Bi(n,p) = Pois(\lambda=np)$$

Riešené príklady

Príklad 1:
Hodíme 8 hracích kociek na stôl. Aká je pravdepodobnosť, že na k kockách padne 6-ka?

Riešenie:

Je to prvoplánový príklad na binomické rozdelenie. Odpoveďou je priame dosadenie do vzorca, v ktorom bude p=1/6:

$$p_k = C(8,k) * (1/6)^k * (5/6)^{8-k} \qquad k = 0,1, ..., 8$$

Vyčíslenie:

```
>> for k=0:8, nchoosek(8,k) * (1/6)^k * (5/6)^(8-k), end
ans =  0.23257
ans =  0.37211
ans =  0.26048
ans =  0.10419
ans =  0.026048
ans =  0.0041676
ans =   4.1676e-004
ans =   2.3815e-005
ans =   5.9537e-007
```

Príklad 2:
Študent FEI mal pred skúškou „náročnú noc“, v dôsledku čoho ráno prišiel omylom na inú školu a zasadol si k testu z čínštiny. Na teste bolo 10 otázok, pri každej 4 možné odpovede. Keďže ani zadanie prečítať nevedel, náhodne označil pri každej otázke nejakú odpoveď.

Aká je pravdepodobnosť, že pri takomto slepom hádaní urobí skúšku

a) na A alebo B (aspoň 9 správnych odpovedí)
b) aspoň na E (aspoň 6 správnych odpovedí)
c) za FX

Riešenie:

Opäť príklad na binomické rozdelenie. Tu však bude trochu viac práce.

a) 9 alebo 10 správnych odpovedí: $P(9)+P(10) = (1/4)^{10} + 10 * (1/4)^9 * (3/4) = 2.9564e-005$
Čísla nepustia, je to takmer nemožné.

b) 6 až 10 správnych odpovedí: $P(6) + ... + P(10) = 0.019728$

Postup s pomocou počítača:

```
>> z=0; for k=10:-1:6, p=nchoosek(10,k)*(1/4)^k*(3/4)^(10-k); z=z+p; end, z
        z =  0.019728
```

c) Neurobiť skúšku znamená presný doplnok ku „aspoň na E“:

```
>> 1-z
        ans =  0.98027
```

Príklad 3:

Vinár Jano ponúka burčiak a rôzne iné nápoje. Pravdepodobnosť, že zákazník si kúpi práve burčiak (predpokladajme, že každý si kupuje len jednu vec), je 0.2.

Aká je pravdepodobnosť, že v sledovaný deň po otvorení predajného pultu

a) prvý zákazník si kúpi burčiak
b) až tretí zákazník si kúpi burčiak
c) jeden (!) z prvých 3 zákazníkov si kúpi burčiak

Riešenie:

Ide o jednoduchý príklad na geometrické rozdelenie.

a) 0.2
b) 0.8*0.8*0.2 = 0.128
c) 0.2 + 0.8*0.2 + 0.8*0.8*0.2 = 0.488

Príklad 4:

V prípade nádejného influencera XY je pravdepodobnosť preklepu pri písaní statusu na facebook 0.1 pri každom písmene alebo znaku. Náhodná veličina X bude označovať počet správne napísaných písmen do momentu, keď nastane chyba (potom sa počítadlo vynuluje). Aká je pravdepodobnosť, že XY napíše správne

a) 5 písmen
b) aspoň 5 písmen
c) najviac 4 písmená
d) niečo medzi 4 až 9 písmen

Riešenie:

Ide o geometrické rozdelenie, len si treba dať pozor na to, „čo je čo“, keďže zadanie je trochu inak formulované než bežný vzorec.

Pravdepodobnosť, že napíše presne k písmen správne a potom sa pomýli, je tu

$P(k) = 0.9^k * 0.1$

a) P(5) = 0.059049

c) P(4)+P(3)+P(2)+P(1)+P(0)

```
>> z=0; for k=4:-1:0, p=0.9^k * 0.1; z=z+p; end, z
        z =  0.40951
```

b)
```
>> 1-z
        ans =  0.59049
```

d) P(4)+...+P(9)

```
>> z=0; for k=4:9, p=0.9^k * 0.1; z=z+p; end, z
        z =  0.30742
```

Príklad 5:
Z 24 jabĺk sú 3 červivé (nevidno to na nich, len pod ultrazvukom sa to vyšetrilo). Náhodne vyberieme z nich 6. Aká je pravdepodobnosť, že v našom výbere bude x červivých jabĺk a aké hodnoty x sú možné?

Riešenie:

Hodnota x môže byť od 0 po 3. Ide tu o typický príklad na hypergeometrické rozdelenie. Stačí dosadiť do vzorca.

$$P(x) = C(3,x) * C(21, 6-x) \ / \ C(24, 6)$$

```
>> A=nchoosek(24,6); for x=0:3, x, p=nchoosek(3,x)*nchoosek(21,6-x)/A, end

x = 0
        p =  0.40316
x =  1
        p =  0.45356
x =  2
        p =  0.13340
x =  3
        p =  0.0098814
```

Príklad 6:
Na maturite je 20 otázok, nádejný maturant sa naučil 11 otázok. Z „klobúka“ si vyberá tri otázky. Aká je pravdepodobnosť, že si vytiahne x otázok, ktoré vie?

Riešenie:

Hodnota x môže byť od 0 po 3. Opäť stačí dosadiť do vzorca

$$P(x) = C(11,x) * C(9, 3-x) \ / \ C(20, 3)$$

```
>> A=nchoosek(20,3); for x=0:3, x, p=nchoosek(11,x)*nchoosek(9,3-x)/A, end

x = 0
        p =  0.073684
x =  1
        p =  0.34737
x =  2
        p =  0.43421
x =  3
        p =  0.14474
```

Príklad 7:

Stará telefónna ústredňa je obsluhovaná ručne „tetou telefonistkou“.
V stránkových hodinách úradu prichádza v priemere 15 telefonátov za hodinu, ktoré treba prepojiť. Počas desiatovej prestávky teta odchádza na 20 minút do bufetu a zastupuje ju mladý brigádnik, ktorý je však povestný rozličnými nemiestnymi vtipmi. Preto teta v bufete celý čas trpne, akú hanbu mladý narobí. V záujme odbúrania stresu a upokojenia nervov teta zvykne uvažovať nad tým a počítať, koľko telefonátov a s akou pravdepodobnosťou asi príde počas jej neprítomnosti.

Možno za ten čas nikto nezavolá... Možno zavolá len jeden, bude to musieť prežiť... Možno aj dvaja... ... .. a čo ak just zavolajú desiati?

Vypočítajte pravdepodobnosti jednotlivých reálnejších možností.

Riešenie:

Najprv si určíme lambdu:

$$\lambda = 20*15/60=5$$

A teraz pravdepodobnosti:

$$P(k) = e^{-\lambda} * \lambda^{k} / k! \qquad \text{pre } k = 0,1,2,3, \ldots$$

Len dosadíme:

k = 0	p = 0.0067379
k = 1	p = 0.033690
k = 2	p = 0.084224
k = 3	p = 0.14037
k = 4	p = 0.17547
k = 5	p = 0.17547
k = 6	p = 0.14622
k = 7	p = 0.10444
k = 8	p = 0.065278
k = 9	p = 0.036266
k = 10	p = 0.018133
k = 11	p = 0.0082422
k = 12	p = 0.0034342

...

Všimnime si, že najvyššia pravdepodobnosť je pre 5, čo je pochopiteľné, ale tiež rovnako pre 4, čo môže trochu prekvapovať, ale má to svoju rafinovanú logiku. Ešte sa k tomu (možno) vrátime v ďalších úlohách.

Príklad 8a:

Biofazuľa z farmy XYZ sa predáva „neprebratá“ a obsahuje v priemere 1 kameň na 100 fazuliek (tj. vo výbere 100 kusov je 1 kameň a iba 99 fazuliek). Lojzo sa tiež neunúva to pred varením preberať, tak odsype len tak 60 kusov „materiálu“ a nachystá si poctivú fazuľovicu. Aká je pravdepodobnosť, že vo svojej porcii

a) nebude mať kameň
b) bude mať aspoň 1 kameň
c) bude mať práve 1 kameň
d) bude mať práve 2 / práve 3 kamene

Riešenie:

Použijeme binomické rozdelenie, ktoré systémovo presne zodpovedá situácii.
Pravdepodobnosť, že vybraná „fazuľka“ je v skutočnosti kameň, je podľa zadania p = 0.01 .
Vyberieme n=60 kúskov a už uvažujeme len o tejto vzorke. Na dané otázky odpovedáme priamo dosadením do vzorca.

a) $P(0) = C(60,0)*0.01^{0}*(0.99)^{60} = 0.99^{\wedge}60 = 0.54716$

b) $P(x\geq 1) = 1 - 0.54716 = 0.45284$

c) $P(1) = C(60,1)*0.01^{1}*(0.99)^{59} = 0.33161$

d) $P(2) = C(60,2)*0.01^{2}*(0.99)^{58} = 0.098813$
$P(3) = C(60,3)*0.01^{3}*(0.99)^{57} = 0.019297$

Príklad 8b:

Biofazuľa z farmy XYZ sa predáva „neprebratá“ a obsahuje v priemere 1 kameň na 100 fazuliek (tj. vo výbere 100 kusov je 1 kameň a iba 99 fazuliek). Lojzo sa tiež neunúva to pred varením preberať, tak odsype len tak 1000 kusov „materiálu“ a nachystá celej rodine na týždeň dopredu poctivú fazuľovicu.

Aká je pravdepodobnosť, že v hrnci

a) bude mať najviac 2 kamene
b) bude mať najviac 10 kameňov
c) bude mať aspoň 10 kameňov
d) bude mať presne 0,1,2, ... 15 kameňov

e) Ak odsype 10000 kusov snáď-fazuliek, aká je pravdepodobnosť, že tam bude presne 100 kameňov?

Riešenie:

I. Binomické rozdelenie systémovo presne zodpovedá situácii.

a) Počítame P(0) + P(1) + P(2) .

$C(1000,0)*0.01^0*(0.99)^{1000} + C(1000,1)*0.01*(0.99)^{999} + C(1000,2)*0.01^2*(0.99)^{998} =$

= 0.99^1000 + 1000*0.01*0.99^999 + 1000*999/2*0.01^2*0.99^998 = 0.0026794

b) Počítame P(0) + P(1) + ... + P(10) .

Kalkulačka nás pri týchto číslach asi rýchlo pošle do nevábnych destinácií, a počítač napriek ochote k práci varuje, že vôbec neručí za presnosť výpočtu.
Našťastie, dá sa to trochu uhrať, len treba počítať rozumne, s minimálnym počtom operácií a s vyhýbaním sa veľkým číslam.

```
>> p=0.99^1000; z=p; for k=1:10, p=p/0.99*0.01*(1001-k)/k; z=z+p; end, z

z =  0.58304
```

c) Počítame P(0) + P(1) + ... + P(9) a výsledok odrátame od 1.

```
>> p=0.99^1000; z=p; for k=1:9, p=p/0.99*0.01*(1001-k)/k; z=z+p; end, 1-z

ans =  0.54270
```

d) Aplikujeme postup z b) a necháme si vypísať jednotlivé pravdepodobnosti.

```
0        p =  4.3171e-005
1        p =  4.3607e-004
2        p =  0.0022002
3        p =  0.0073932
4        p =  0.018614
5        p =  0.037453
6        p =  0.062737
7        p =  0.089987
8        p =  0.11282
9        p =  0.12561
```

```
10        p =  0.12574
11        p =  0.11431
12        p =  0.095162
13        p =  0.073053
14        p =  0.052023
15        p =  0.034542
```

e) V tomto prípade priamy pokus o výpočet zlyhá (príliš veľké čísla) a postupný výpočet (podľa bodu b)) už je príliš zdĺhavý na tak banálnu otázku.

II. Vzhľadom na veľké čísla stojí za úvahu použiť miesto binomického rozdelenia aproximáciu Poissonovým rozdelením.

Ako sme spomenuli vyššie v teoretickom prehľade, volíme $\lambda = np = 1000*0.01 = 10$.

Urobíme výpočet pre úlohu d) – tam je možné najlepšie porovnať, nakoľko si výsledky budú blízke. (zároveň z údajov v úlohe d) vyplývajú priamo aj výsledky v a), b), c)).

$P(k) = e^{-10} * 10^k / k!$

```
>>        for k=0:15, exp(-10)*10^k/factorial(k), end
```

výsledky – Binomické rozdelenie	výsledky – Poissonovo rozdelenie
0 p = 4.3171e-005	ans = 4.5400e-005
1 p = 4.3607e-004	ans = 4.5400e-004
2 p = 0.0022002	ans = 0.0022700
3 p = 0.0073932	ans = 0.0075667
4 p = 0.018614	ans = 0.018917
5 p = 0.037453	ans = 0.037833
6 p = 0.062737	ans = 0.063055
7 p = 0.089987	ans = 0.090079
8 p = 0.11282	ans = 0.11260
9 p = 0.12561	ans = 0.12511
10 p = 0.12574	ans = 0.12511
11 p = 0.11431	ans = 0.11374
12 p = 0.095162	ans = 0.094780
13 p = 0.073053	ans = 0.072908
14 p = 0.052023	ans = 0.052077
15 p = 0.034542	ans = 0.034718

Porovnanie ukazuje, že aproximácia je naozaj slušná a tie drobné odchýlky nepredstavujú fatálne skreslenie skutočnosti.

e) S použitím Poissonovho rozdelenia vieme riešiť aj úlohu e) . Dáme $\lambda = np = 10000*0.01 = 100$

```
>> k=100; exp(-100)*100^k/factorial(k)

                    ans =  0.039861
```

Stabilnejší postup dáva rovnaký výsledok, pre dnešný softvér už sú hranice presnosti kdesi inde:

```
>> z=exp(-100); for k=1:100, z=z*100/k; end, z

                    z =  0.039861
```

Neriešené príklady

(na niektoré výpočty sa odporúča využiť programovateľný softvér)

1. Do školy v spádovej obci Y chodí 40 žiakov, prevažne z okolitých lazov. Počas dní so snežením je pravdepodobnosť absencie žiaka 0,25.
a) Aký počet neprítomných sa dá v deň so snežením s najvyššou pravdepodobnosťou (akou? – vyčísliť) očakávať? (10, 0.14436)
b) Aká je pravdepodobnosť, že bude chýbať najviac 7 žiakov? (0.18195)
c) Aká je pravdepodobnosť, že bude chýbať aspoň 8 žiakov? (0.81805)

2. Smartfón značky XYZ sa v priemere počas 10 000 hodín prevádzky zasekne 15-krát. Aká je pravdepodobnosť, že počas určených 500 hodín príde raz/aspoň raz k takejto udalosti? (0.35427, 0.52763)

3. Na prijímačkách je test s 15 otázkami, každá s voľbou jednej z odpovedí A, B, C. Záujemca o štúdium sa vôbec neučil a test vypĺňa náhodne.
a) Aká je pravdepodobnosť, že trafí správne aspoň 5 otázok? (0.59594)
b) Koľko by mal mať test otázok, aby pravdepodobnosť správneho zodpovedania (pri náhodnom krúžkovaní) aspoň polovice otázok bola menej než 0,05 ? (23 pre nepárny počet otázok, 30 pre párny počet)

4. Vo fazuli od výrobcu ZYX sa nachádza priemerne 1 zablúdený hrášok na 1000 fazuliek. Na prípravu polievky sa použije 2000 snáď-fazuliek. Jedlo zjedia piati ľudia (rovnakým dielom). Aká je pravdepodobnosť, že
a) aspoň jeden zo stravníkov nájde u seba aspoň 1 hrášok? (0.86480)
b) práve jeden zo stravníkov nájde u seba aspoň 1 hrášok? (0.33268)

5. Krajský zdravotnícky úrad zhromažďuje údaje o novonarodených deťoch v kraji. Priemerne každé dve hodiny sa narodí ďalšie dieťa. Určte
a) priemerný počet narodených detí za rok (nepriestupný). (4380)
b) pravdepodobnosť, že v danom dni sa narodí najviac 5 detí. (0.020341)
d) pravdepodobnost, že za 4 hodiny sa narodí aspoň 5 detí. (0.052653)

6. Po rannom príboji sa na pláž dĺžky 2 km vyplaví priemerne 200 ulít (ulity zo predošlého dňa sú vyzbierané).
a) Aká je pravdepodobnosť, že prvý ranný nadšenec po prehľadaní 50 m úseku nájde práve 4 ulity / práve 6 ulít? (0.17547 , 0.14622)
b) Aká je pravdepodobnosť, že nájde aspoň 5 ulít? (0.55951)
c) Aká je pravdepodobnosť, že po prehľadaní 20 metrov nenájde nič? (0.13534)
d)* Ako dlhý úsek musí prehľadať, aby s pravdepodobnosťou aspoň 0,9 našiel aspoň 5 ulít? (80)