

## (07) Náhodné vektory I.

Plán prednášok o náhodných vektoroch:

- všeobecné vlastnosti
- korelácia zložiek náhodného vektora
- diskrétné náhodné vektory
- spojité náhodné vektory

### 1. Úvod

Ak  $X_1, X_2, \dots, X_n$  sú náhodné premenné, tak vektor  $(X_1, X_2, \dots, X_n)$  nazývame **náhodným vektorom**.

S náhodnými vektormi sme už pracovali aj bez toho, aby sme ich takto pomenovali. Napríklad keď sme sledovali čísla, ktoré padli na dvoch kockách, mali sme vlastne náhodný vektor  $(X, Y)$ , kde  $X$  je počet bodov na prvej kocke a  $Y$  je počet bodov na druhej kocke.

Dôležité typy náhodných vektorov (podobne ako pri náhodných premenných):

- **Diskrétné náhodné vektory**
  - nadobúdajú konečne alebo spočítateľne veľa hodnôt
  - to je ekvivalentné s tým, že ich zložky sú diskrétné náhodné premenné
- **Spojité náhodné vektory**
  - majú hustotu, bude to analógia hustoty náhodnej premennej – *bude tomu venovaná samostatná téma, zatiaľ sa tým nebudeme zaoberať*
  - každá zložka je spojitou náhodnou premennou
  - uvidíme, že naopak to neplatí – ak dáme do vektora spojité náhodné premenné, nemusí vzniknúť spojitý náhodný vektor (ale platí to, ak sú tieto náhodné premenné nezávislé) – *toto bude dávať, samozrejme, väčší zmysel po prebraní danej témy*

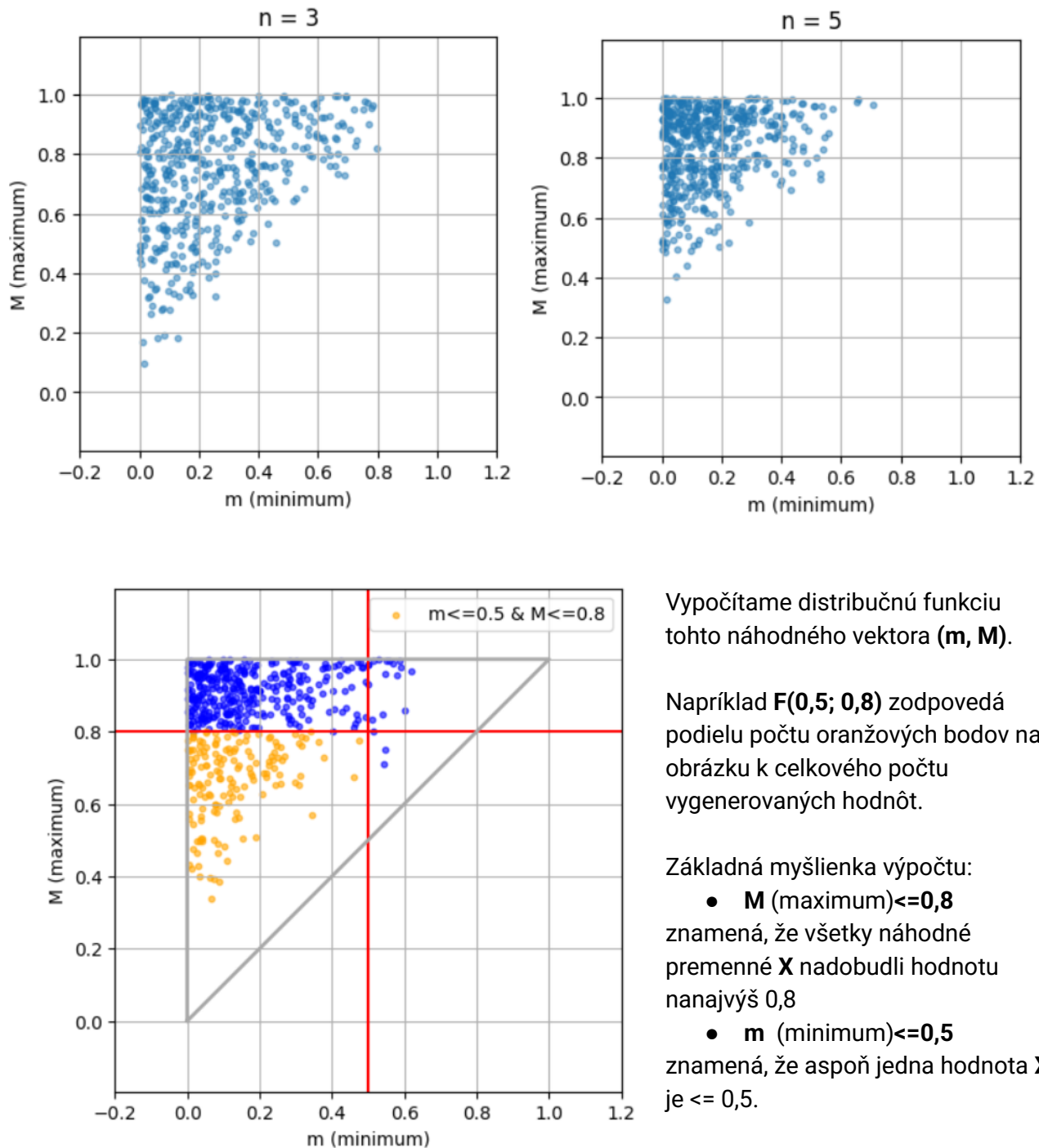
Spoločné pre všetky náhodné vektory:

- **Stredná hodnota náhodného vektora** – vektor, ktorého zložky sú  $E(X_i)$
- **Distribučná funkcia náhodného vektora** je funkcia  $n$  premenných definovaná prepisom  $F(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n)$ 
  - Vidíme, že ak sú **zložky náhodného vektora nezávislé**, tak **distribučná funkcia náhodného vektora je súčinom distribučných funkcií jeho zložiek**.
  - Platí to ak naopak, ak je distribučná funkcia náhodného vektora je súčinom distribučných funkcií jeho zložiek, tak tieto náhodné premenné sú nezávislé.

## 2. Distribučná funkcia náhodného vektora: definícia

### Príklad 1

Majme  $n$  nezávislých náhodných premenných s rovnomerným rozdelením na intervale  $(0, 1)$ . Definujme náhodný vektor  $(\mathbf{m}, \mathbf{M})$ , kde  $\mathbf{m}$  je minimum týchto náhodných premenných a  $\mathbf{M}$  je ich maximum.



Vypočítame distribučnú funkciu tohto náhodného vektora  $(\mathbf{m}, \mathbf{M})$ .

Napríklad  $F(0,5; 0,8)$  zodpovedá podielu počtu oranžových bodov na obrázku k celkovému počtu vygenerovaných hodnôt.

Základná myšlienka výpočtu:

- $\mathbf{M}(\text{maximum}) \leq 0,8$  znamená, že všetky náhodné premenné  $\mathbf{X}$  nadobudli hodnotu najvyšš 0,8
- $\mathbf{m}(\text{minimum}) \leq 0,5$  znamená, že aspoň jedna hodnota  $\mathbf{X}$  je  $\leq 0,5$ .

Niektoré hodnoty sú zrejmé, keďže náhodné premenné  $X_i$ , z ktorých počítame maximum nadobúdajú hodnoty z intervalu  $(0, 1)$ :

- Ak je  $x$  aj  $y$  väčšie alebo rovné ako 1, tak podmienka "minimum je  $\leq x$  a maximum je  $\leq y$ " je určite splnená, takže má pravdepodobnosť rovnú 1. Distribučná funkcia sa v takýchto bodoch  $(x, y)$  rovná 1.
- Ak je niektorá z hodnôt  $x, y$  menšia alebo rovná ako nula, tak podmienka "minimum je  $\leq x$  a maximum je  $\leq y$ " nemôže byť splnená, takže má pravdepodobnosť rovnú 0. Distribučná funkcia sa v takýchto bodoch  $(x, y)$  rovná 0.

Vypočítame teraz  $F(x, y)$  pre  $(x, y)$  vnútri sivého trojuholníka:

$$F(x, y) = P(\underbrace{m \leq x}_{\text{minimum} \leq x \text{ znamená, že aspoň jedno } X_i \text{ je } \leq x}, \underbrace{M \leq y}_{\text{maximum} \leq y \text{ znamená, že všetky } X_i \text{ sú } \leq y}) \equiv$$

$$\text{Ozn. udalosti } A = \{\text{aspoň jedno } X_i \text{ je } \leq x\} \rightarrow A^c = \{\text{všetky } X_i \text{ sú } > x\}$$

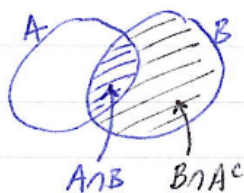
$$B = \{\text{všetky } X_i \text{ sú } \leq y\}$$

Potrebujeeme  $A \cap B$

Vieme vypočítať:

$$P(B) = P(X_1 \leq y, \dots, X_n \leq y) \stackrel{\text{nezávislosť}}{=} \underbrace{P(X_1 \leq y)}_y \dots \underbrace{P(X_n \leq y)}_y = y^n$$

lebo  $X_i$  má rovnomerné rozdelenie na  $(0, 1)$  a  $y \in (0, 1)$

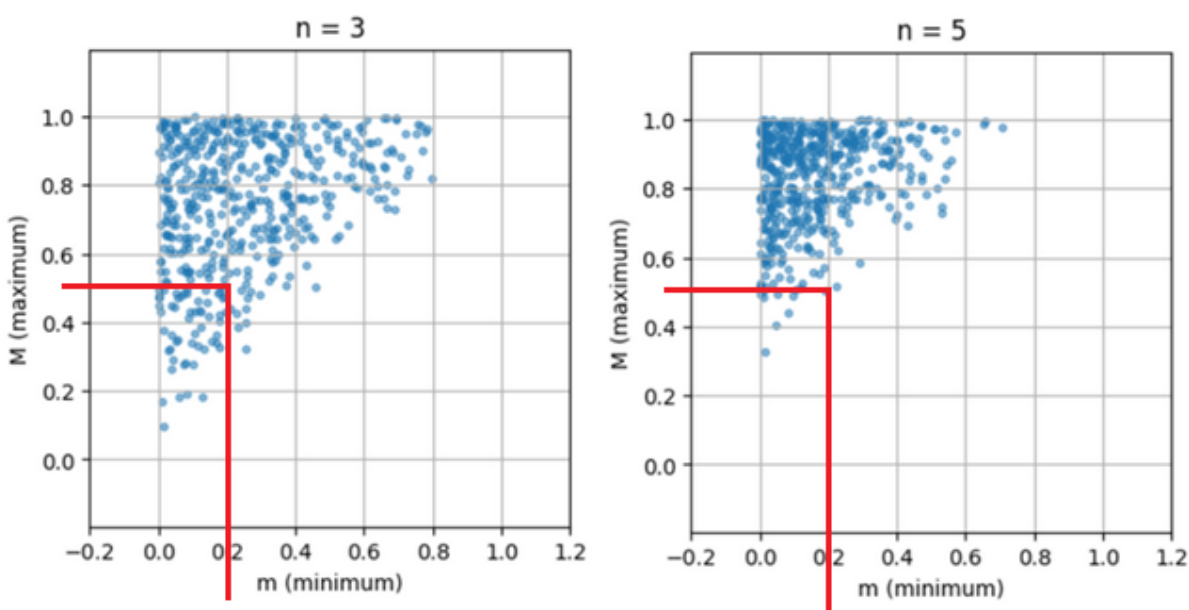


$$\begin{aligned} P(B \cap A^c) &= P(\text{všetky } X_i \text{ sú } \leq y \text{ a súčasne } > x) = \\ &= P(\text{všetky } X_i \text{ sú z intervalu } (x, y]) = \\ &= P(X_1 \in (x, y], \dots, X_n \in (x, y]) \stackrel{\text{nezávislosť}}{=} \\ &= \underbrace{P(X_1 \in (x, y])}_{y-x} \dots \underbrace{P(X_n \in (x, y])}_{y-x} = (y-x)^n \end{aligned}$$

lebo  $X_i \sim R(0, 1)$   
 $x, y \in (0, 1), x < y$

$$\equiv P(A \cap B) = P(B) - P(B \cap A^c) = y^n - (y-x)^n$$

Na základe simulácií očakávame, že napríklad hodnota  $F(0,2; 0,5)$  bude väčšia v prípade distribučnej funkcie počítanej s  $n=3$  (minimum a maximum z troch náhodných čísel) ako v prípade  $n=5$ .

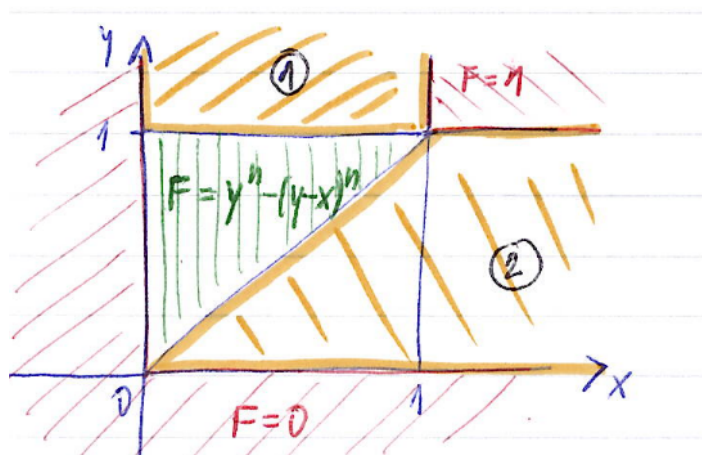


Vyčíslime tieto hodnoty:

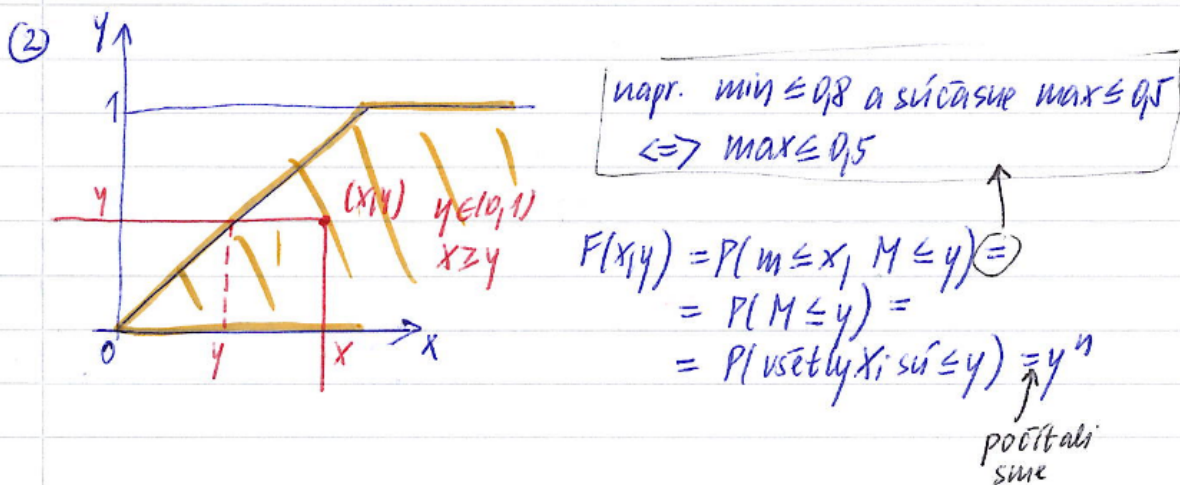
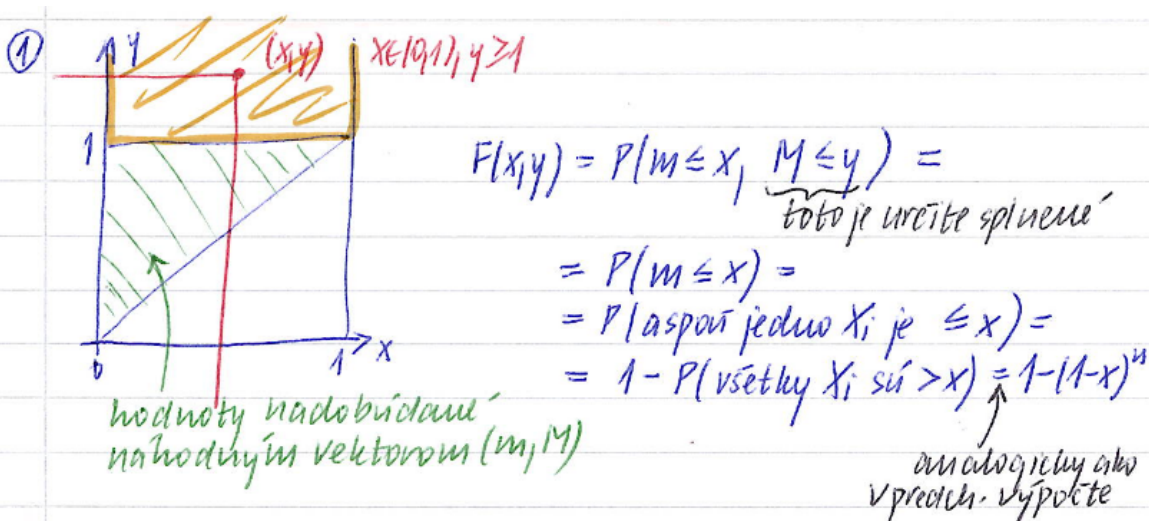
```
F = lambda x, y, n: y**n - (y - x)**n # pre 0 <= x < y <= 1
n_vektor = range(2, 8)
x, y = 0.2, 0.5
for n in n_vektor:
    print(f"Pre n={n} je F({x}, {y}) = {F(x, y, n):.5f}")
```

```
Pre n=2 je F(0.2, 0.5) = 0.16000
Pre n=3 je F(0.2, 0.5) = 0.09800
Pre n=4 je F(0.2, 0.5) = 0.05440
Pre n=5 je F(0.2, 0.5) = 0.02882
Pre n=6 je F(0.2, 0.5) = 0.01490
Pre n=7 je F(0.2, 0.5) = 0.00759
```

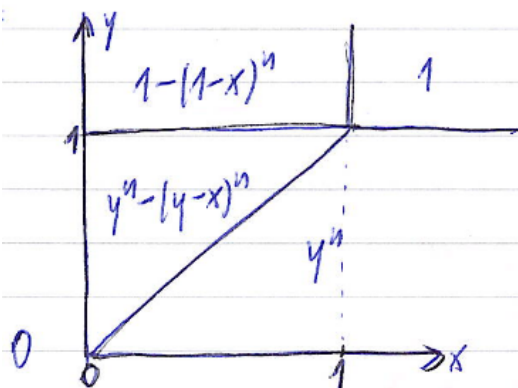
Čo už máme a čo potrebujeme dopočítať:



Zostávajúce dve oblasti:



Záver:



Všimnime si, že táto distribučná funkcia je spojitá, v hraničných bodoch medzi oblasťami, na ktorých sme  $F$  počítali zvlášť, sa hodnoty na seba spojitاً napájajú.

### 3. Distribučná funkcia náhodného vektora: výpočty, vlastnosti

#### (1) Výpočet pravdepodobnosti hodnôt z obdĺžnika pomocou $F(x, y)$

##### Príklad 2

Aká je pravdepodobnosť, že náhodný vektor  $(\mathbf{m}, \mathbf{M})$  nadobudne hodnotu z vyznačeného štvorca? Zdá sa, že s rastúcim  $n$  táto pravdepodobnosť rastie a intuitívne to tiež vyzerá logicky.

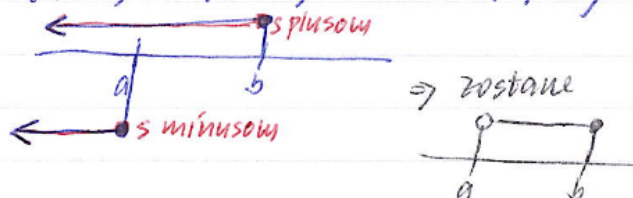


Ako to vyzeralo v 1D:

$F(x)$  - distribučná funkcia náhodnej premennej  $X$

z definície:  $F(x) = P(X \leq x)$

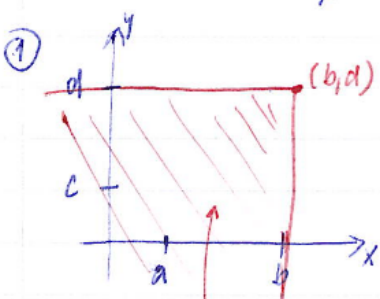
$$\Rightarrow F(b) - F(a) = P(X \leq b) - P(X \leq a) = P(X \in [a, b])$$



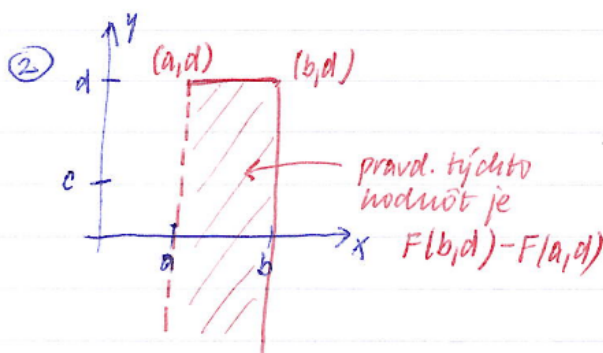
Všeobecné odvedenie pre 2D:

$F(x, y)$  - distribučná funkcia náhodného vektora  $(X, Y)$

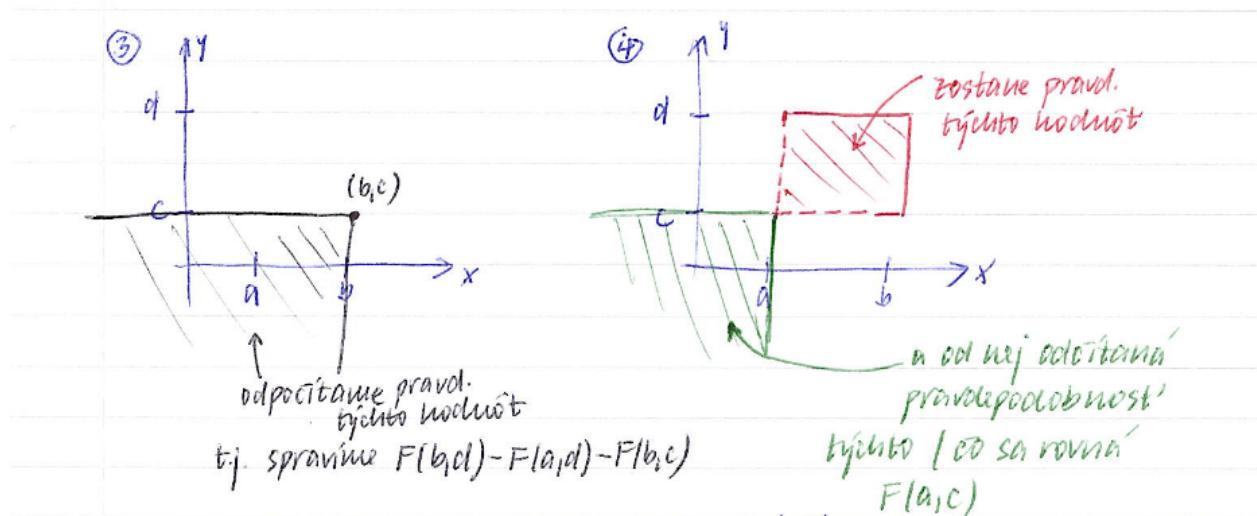
z definície:  $F(x, y) = P(X \leq x, Y \leq y)$



$F(b, d) = \text{pravd. týchto hodnôt}$



pravd. týchto hodnôt je  $F(b, d) - F(a, d)$



Teda pravdepodobnosť, že sa nadobudne hodnota z červeného obdĺžnika na obrázku 4 je  
 $P(a < X \leq b, c < Y \leq d) = F(b, d) - F(a, d) - F(b, c) + F(a, c)$

Numericky pre náš príklad (m, M), pričom predpokladáme, že celý obdĺžnik, pre ktorý sa počíta pravdepodobnosť, leží v trojuholníku, v ktorom sa nadobúdajú hodnoty náhodného vektora. (Ak to neplatí, treba definovať funkciu **F** poriadne, aby vrátila správnu hodnotu pre ľubovoľný argument). Náš štvorec to spĺňa.

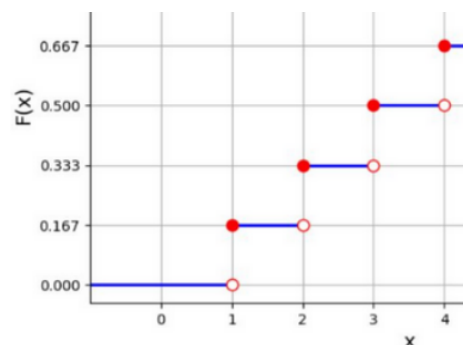
```
p_obdlznika = lambda a, b, c, d, n: F(b, d, n) - F(a, d, n) - F(b, c, n) + F(a, c, n)
n_vektor = range(2, 8)
a, b, c, d = 0, 0.2, 0.8, 1
for n in n_vektor:
    print(f"Pre n={n} je pravdepodobnosť = {p_obdlznika(a, b, c, d, n):.5f}")
```

```
Pre n=2 je pravdepodobnosť = 0.08000
Pre n=3 je pravdepodobnosť = 0.19200
Pre n=4 je pravdepodobnosť = 0.31040
Pre n=5 je pravdepodobnosť = 0.42240
Pre n=6 je pravdepodobnosť = 0.52237
Pre n=7 je pravdepodobnosť = 0.60856
```

## (2) Priebeh distribučnej funkcie v 2D v bodoch nespojitosti

Ako to vyzeralo v 1D:

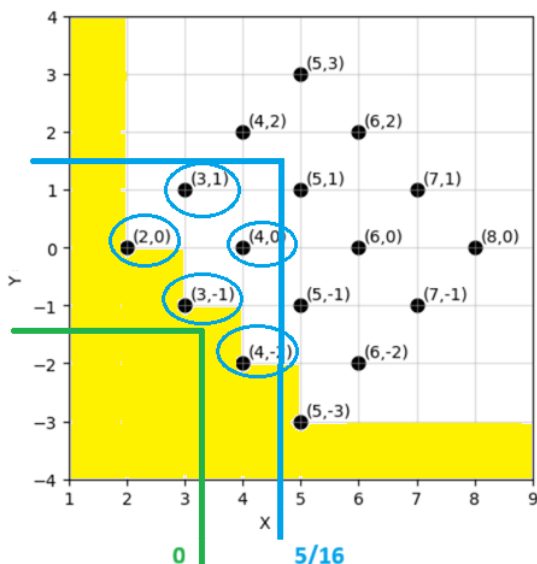
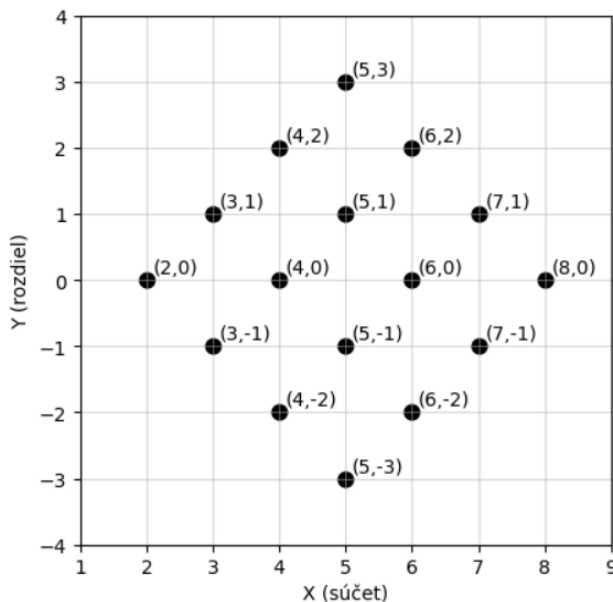
Na obrázku vidíme časť nespojitej distribučnej funkcie náhodnej premennej. Je taká, že v bode nespojitosti sa nadobúda tá vyššia hodnota. Vyplýva to z toho, že v definícii distribučnej funkcie je neostrá nerovnosť  $F(x)=P(X \leq x)$ .



### Príklad 3

Hádzeme dvoma kockami v tvare štvorstenu, na ich stenách sú čísla 1-4. Vytvoríme náhodný vektor  $(X, Y)$ , ktorého prvá zložka je súčet padnutých čísel a druhá zložka je ich rozdiel (prvá kocka mínus druhá kocka). Zaujímá nás distribučná funkcia tohto náhodného vektora.

Ak si vypíšeme dvojice čísel, ktoré padnú na kockách (každá má pravdepodobnosť  $1/16$ ) a príslušné hodnoty  $X, Y$ , dostaneme, že v každom zo 16 prípadov sa nadobúda iná dvojica hodnôt  $(x, y)$ . Máme ich znázornené na obrázku vpravo.



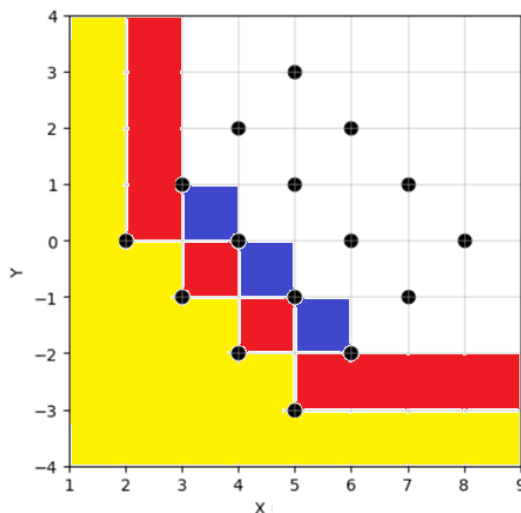
Ak hľadáme hodnotu distribučnej funkcie v nejakom bode, spočítame pravdepodobnosti tých možností, ktoré sú vľavo a dolu od daného bodu (vrátane polpriamok vychádzajúcich z tohto bodu).

Takže napríklad:

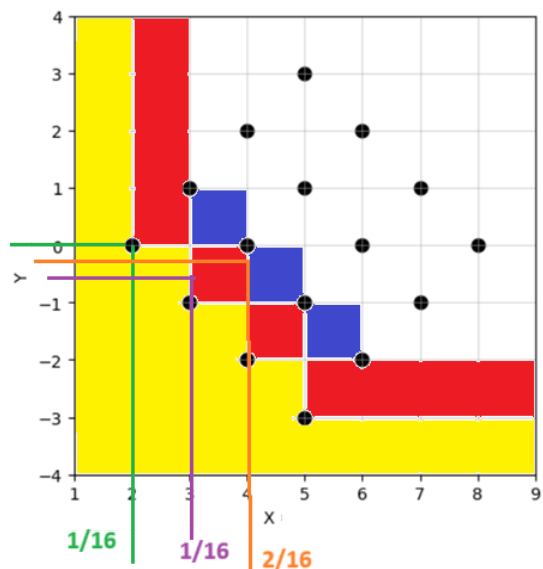
- $F(3,3; -1,4) = 0$  (zelená čiara vyznačuje oblasť, v ktorej sledujeme výsledky pokusu a sčítavame ich pravdepodobnosti)
- $F(4,6; 1,5) = 5/16$  (modrá čiara)
- V žltej oblasti je distribučná funkcia nulová (bez hraničných bodov).

Takto môžeme pokračovať, začiatok výpočtu je na obrázku. Bez hraničných bodov, vo vnútri oblasti:

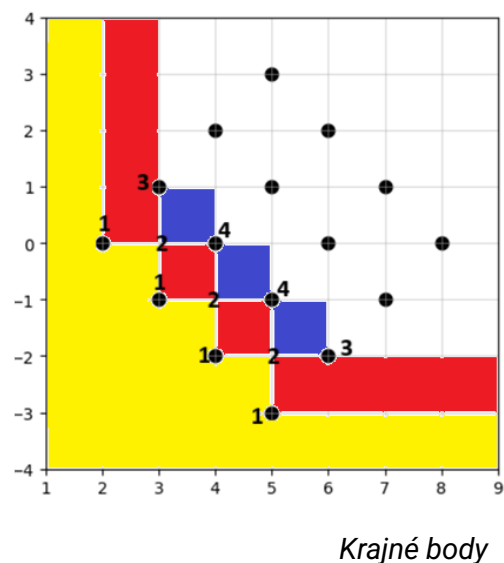
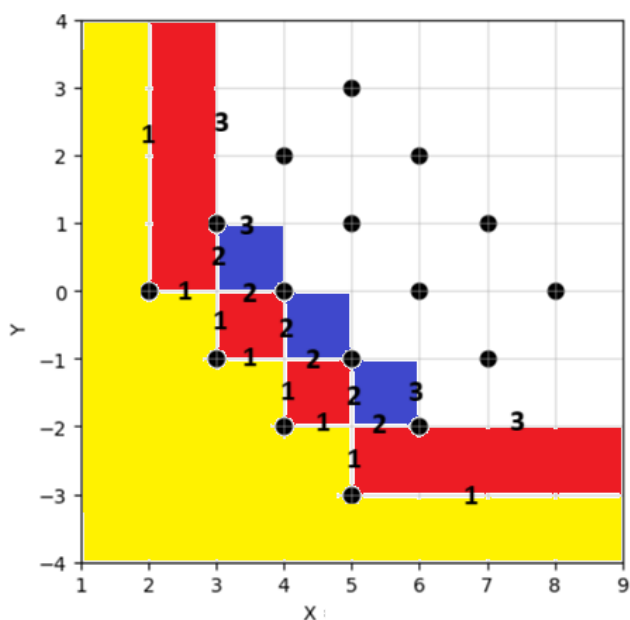
- Žltá farba zodpovedá  $F(x,y)=0$
- Červená farba zodpovedá  $F(x,y)=1/16$
- Modrá farba zodpovedá  $F(x,y)=2/16$



Pozrime sa teraz na body nespojitosti, niektoré sú vyznačené na obrázku. Počítame z definície, stačí si uvedomiť, že body ležiace na hraničnej čiare patria medzi tie, ktoré sa zarátavajú ako "body vľavo dole od sledovaného bodu".



Úsečky medzi konštantnými oblasťami (bez krajných bodov, hodnota čitateľa, menovateľ je 16):



Ak spravíme rez, vidíme, že pri prechode od menšej k väčšej hodnote premennej sa funkcia správa ako jednorozmerná distribučná funkcia – v prípade skoku sa v bode nespojitosti nadobúda

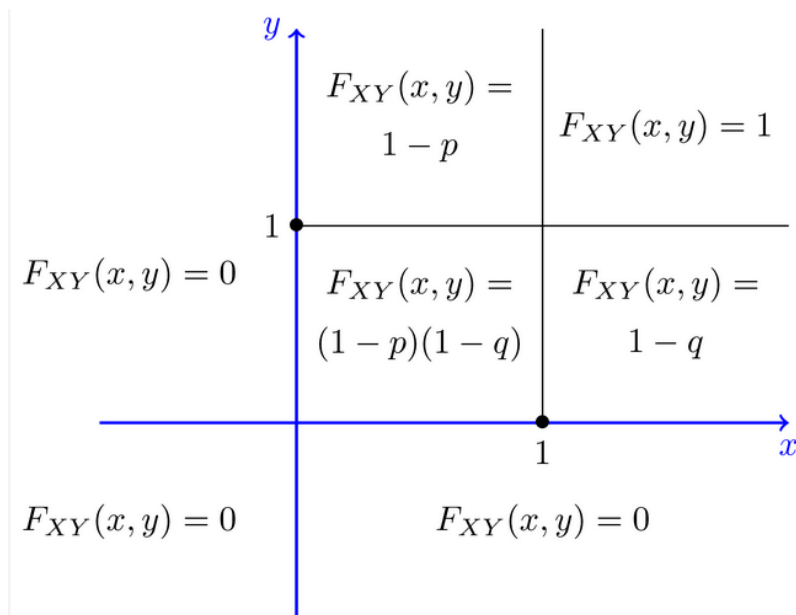
väčšia hodnota. Napríklad na tomto reze: najskôr nula zo žltej oblasti, potom skok na  $1/16$  z červenej, pričom v bode skoku je hodnota  $1/16$ . Potom skok na hodnotu  $2/16$  z modrej oblasti, pričom v bode skoku je hodnota  $2/16$ .



#### Príklad 4

Nech  $\mathbf{X}$  je náhodná premenná s Bernoulliho rozdelením s parametrom  $p$  a  $\mathbf{Y}$  je náhodná premenná s Bernoulliho rozdelením s parametrom  $q$ . Predpokladajme, že sú nezávislé. Chceme nájsť distribučnú funkciu náhodného vektora  $(\mathbf{X}, \mathbf{Y})$ .

Tento príklad sa na stránke [https://www.probabilitycourse.com/chapter5/5\\_1\\_2\\_joint\\_cdf.php](https://www.probabilitycourse.com/chapter5/5_1_2_joint_cdf.php) (online učebnica pravdepodobnosti, odporúčam 😊) je graficky znázorný výsledok. Doplňte doňho hodnoty v bodoch nespojivosti. (V texte učebnice to spravené je.)



#### 4. Marginálne rozdelenia

**Marginálnym rozdeleniami** náhodného vektora sa nazývajú rozdelenia jeho zložiek.

#### Príklad 5

Uvažujme náhodný vektor  $(\mathbf{X}, \mathbf{Y})$  z príkladu 3, teda súčet a rozdiel čísel na kockách v tvare štvorstenu. Nájdeme marginálne rozdelenie jeho druhej zložky.

Riešenie:

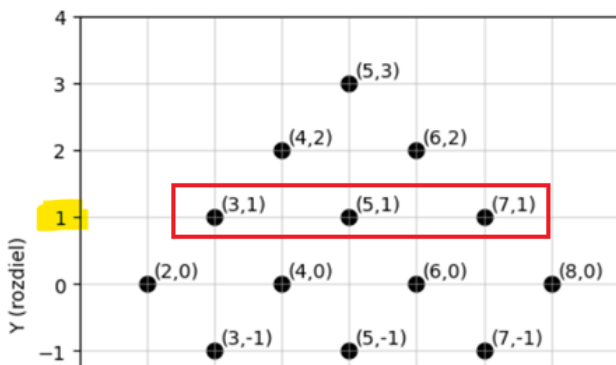
| $Y$        | -3             | -2             | -1             | 0              | 1              | 2              | 3              |
|------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| $P(Y = y)$ | $\frac{1}{16}$ | $\frac{2}{16}$ | $\frac{3}{16}$ | $\frac{4}{16}$ | $\frac{3}{16}$ | $\frac{2}{16}$ | $\frac{1}{16}$ |

## 5. Podmienené rozdelenia

### Príklad 6

Uvažujme náhodný vektor  $(X, Y)$  z príkladu 3, teda súčet a rozdiel čísel na kockách v tvare štvorstenu. Nájdeme podmienené rozdelenie súčtu  $X$ , ak vieme, že rozdiel  $Y$  sa rovná 1.

Riešenie:



Možné hodnoty  $X$  pri tejto podmienke sú 3, 5 a 11 a ich pravdepodobnosti sú podmienené pravdepodobnosti:

$$P(X = x \mid Y = 1) = \frac{P(X = x, Y = 1)}{P(Y = 1)}$$

V tomto prípade sú všetky rovnaké, rovné  $\frac{1}{3}$ .