Příklad 1 (Náhodná veličina)

Uvažujeme experiment: házení mincí. Výsledkem pokusu je "rub" nebo "líc", že padne hrana neuvažujeme. Pokud hovoříme o náhodné veličině, musíme přepsat výsledky pokusu do množiny reálných čísel. Strany mince tedy označíme 0 a 1 a hledáme jejich pravděpodobnosti $f(0) = p_0$ a $f(1) = p_1$[1].

Náhodná veličina je "hod férovou korunou" s pravděpodobnostní funkcí

x	0	1
$f(x)$	0.5	0.5

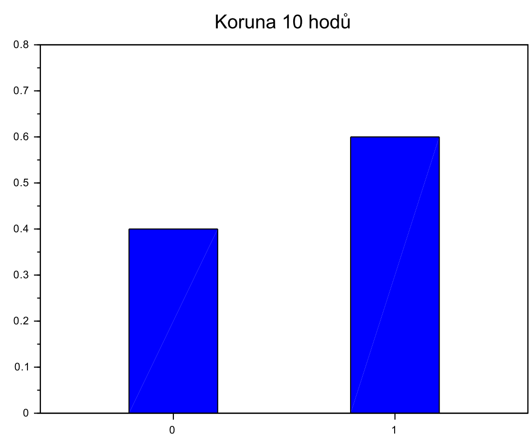
Hody mincí simulujeme ve Scilabu příkazem `x=fix(rand(1,nd,'u')+.5)`[2]; který definuje férovou minci se stejnými pravděpodobnostmi obou stran (rozmyslete). Tento příkaz je tedy generátor s danou pravděpodobnostní funkcí.

Datová množina se liší podle počtu hodů (i při stejném počtu hodů mohou a nemusí vycházet jiné výsledky). Například:

- s 10 daty (provedli jsme 10 hodů) je

  x = {0. 1. 0. 1. 1. 0. 1. 1. 0. 1. }

  a histogram je



- se 100 daty máme

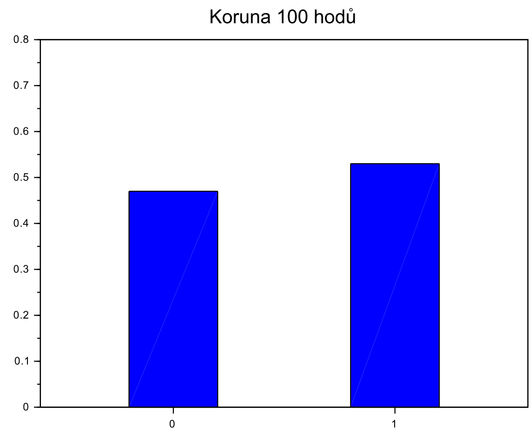
  x = {0. 0. 1. 1. 1. 0. 0. 0. 1. 0. 1. 1. 1. 0. 1. 0. 1. 1. 1. 1. 1. 1. 1. 1. 1. 0. 0. 0. 0. 0. 0. 1.

  1. 1. 0. 1. 1. 1. 0. 0. 1. 0. 1. 1. 1. 0. 1. 0. 0. 1. 0. 1. 0. 1. 0. 1. 1. 0. 0. 0. 0. 1. 0. 0. 0. 1.

  1. 1. 0. 0. 0. 1. 1. 0. 0. 0. 1. 0. 1. 0. 1. 1. 0. 0. 1. 0. 1. 1. 1. 0. 1. 1. 0. 1. 0. 0. 1. 1. 0. 1. }

  a histogram

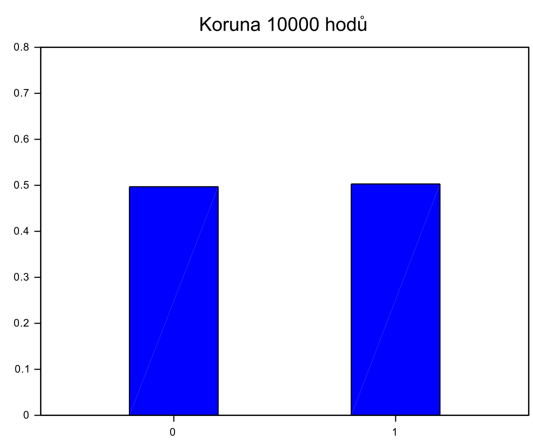
[1] p_0 je tedy pravděpodobnost, že padne rub a p_1 je pravděpodobnost, že padne líc.
[2] funkce fix znamená zaokrouhli, funkce rand(....,'u') generuje hodnoty z intervalu <0,1)



- s 10 000 daty dostaneme mraky čísel, u kterých nemá smysl jejich výpis. Lze je uvést ve tříděném tvaru (který již smysl má - je pochopitelný)

x	0	1
četnosti	4970	5030

s histogramem



Závěry

1. Předpis, generující hody je náhodná veličina. Jeho základní vlastností je
   (a) generuje dvě hodnoty 0 a 1
   (b) obě hodnoty mají stejnou pravděpodobnost 0.5.
2. Pravděpodobnostní funkce náhodné veličiny je

x	0	1
$f(x)$	0.5	0.5

3. Data, generovaná náhodnou veličinou sice ukazují na to, že by pravděpodobnosti mohly být stejné, ale nepřesně.

4. Čím více dat naměříme, tím jsou odhady pravděpodobností lepší.

Příklad 2 (Náhodná veličina)

Uvažujme předchozí příklad s hoden mincí. Ukážeme si rozdíl mezi výpočtem pro charakteristiku náhodné veličiny a charakteristiku pro 10 generovaných hodů mincí.

- Charakteristiky náhodné veličiny

  střední hodnota
  $$E[X] = 0 \cdot 0.5 + 1 \cdot 0.5 = 0.5$$
  rozptyl
  $$D[X] = (0-0.5)^2\, 0.5 + (1-0.5)^2\, 0.5 = 0.25$$
  medián
  $$\tilde{x} = 0.5$$
  (prostředek je mezi 0 a 1 - bere se průměr)

- Charakteristiky datové množiny $\{0,\ 1,\ 0,\ 1,\ 1,\ 0,\ 1,\ 1,\ 0,\ 1\}$[3]

  průměr
  $$\bar{x} = \frac{1}{10}(0+1+0+1+1+0+1+1+0+1) = \frac{3}{5}$$
  rozptyl
  $$\sigma^2 = \frac{1}{10}\left[\left(0-\frac{3}{5}\right)^2 + \left(1-\frac{3}{5}\right)^2 + \left(0-\frac{3}{5}\right)^2 + \left(1-\frac{3}{5}\right)^2 + \left(1-\frac{3}{5}\right)^2 + \right.$$
  $$\left. + \left(0-\frac{3}{5}\right)^2 + \left(1-\frac{3}{5}\right)^2 + \left(1-\frac{3}{5}\right)^2 + \left(0-\frac{3}{5}\right)^2 + \left(1-\frac{3}{5}\right)^2\right] = 0.24$$
  medián

  uspořádaná data jsou
  $$0,\ 0,\ 0,\ 0,\ 1, \quad 1,\ 1,\ 1,\ 1,\ 1$$
  a prostředek je vyznačen mezerou. Jsou tam dva prostřední a jejich průměr je 1. Tedy je
  $$\tilde{x} = 1$$

[3] Tady by byl přehlednější a rychlejší postup, kdy si data nejdříve utřídíme. Ale pro tento ilustrační případ budeme uvažovat klasický zápis.

Pozn.

1. Pokud počítáme charakteristiku náhodné veličiny nelze pracovat jen s datovou množinou, a zjištěné výsledky označit za výsledky celé náhodné veličiny.
2. Zkuste si utřídit data a spočítat charakteristiky datové množiny pomocí metody výpočtu přes tříděná data. Převeďte četnosti u tříděných dat na pravděpodobnosti a spočtěte. Co je nejrychlejší?

Příklad 3 (Náhodný vektor)

Máme 2 náhodné veličiny X s hodnotami 1,2 a Y s hodnotami 1,2,3[4]. Sdružená pravděpodobnostní funkce je dána tabulkou[5]

$X \setminus Y$	1	2	3
1	0.1	0.08	0.2
2	0.05	0.07	0.5

Určete stření hodnotu a rozptyl a kovarianci.

Nejprve bude potřeba určit marginály. To je znázorněno v následující tabulce[6]

$X \setminus Y$	1	2	3	$f(x)$
1	0.1	0.08	0.2	0.38
2	0.05	0.07	0.5	0.62
$f(y)$	0.15	0.15	0.7	

Střední hodnoty[7]

$$E[X] = \sum_x x f(x) = 1.62 \qquad E[Y] = \sum_y y f(y) = 1 \cdot 0.15 + 2 \cdot 0.15 + 3 \cdot 0.7 = 2.55$$

Rozptyly

$$D[X] = (1 - 1.62)^2 \cdot 0.38 + (2 - 1.62)^2 \cdot 0.62 = 0.85393$$

$$D[Y] \quad \text{podobně}$$

Kovariance[8]

$$C[X,Y] = \sum_{x=1,2} \sum_{y=1,2,3} (x - EX)(y - EY) \cdot f(x,y) =$$

$$= (1 - 1.62)(1 - 2.55) \cdot 0.1 + (1 - 1.62)(2 - 2.55) \cdot 0.08 + (1 - 1.62)(3 - 2.55) \cdot 0.2 +$$

$$+ (2 - 1.62)(1 - 2.55) \cdot 0.05 + (2 - 1.62)(2 - 2.55) \cdot 0.15 + (2 - 1.62)(3 - 2.55) \cdot 0.7 = 0.12648$$

[4] Pro lepší představu si představte, že náhodná veličina X je dopravní prostředek (1 - motocykl, 2 - automobil) a Y je obuv (1 - teniska, 2 - sandále, 3 - kotníčková obuv).

[5] Pravděpodobnost, že pojedete na motorce a vezmete si sandále je 0.08, prp. že pojedete v automobilu v kotníčkové obuvi je 0.5.

[6] Pokud se na to podíváte z pohledu výše definované úlohy, tak $f(x)$ je pravděpodobnost, že pojedu na motocyklu, resp. automobil. $f(y)$ je pak pravděpodobnost výběru určitého druhu obuvi.

[7] Pozor, jsou dvě pro každnou náhodnou veličinu zvlášť. Stejně je to i u rozptylů.

[8] Počítáme vliv X na Y, proto je kovariance jen jedna.

Pozn.

1. Zkuste zjistit, co vznikne, pokud budu chtít spočítat $C[X, X]$.
2. Pozor, kovariance může být i záporná.

Příklad 4 (náhodný vektor)

Dejme tomu, že jsme na základě sdružené pravděpodobnosti v příkladu 3 generovali data

x	1	2	1	2	1	2	2	2	1	1	2	1	1	1	2
y	1	3	2	3	1	2	1	3	1	3	3	2	3	3	3

Máme určit průměr, rozptyl a kovarianci.

Průměr

$$\bar{x} = \frac{1}{15}(1+2+1+2+\cdots) = 1.4666$$

$$\bar{y} = \frac{1}{15}(1+3+2+3+\cdots) = 2.2666$$

Rozptyl

$$s_x^2 = \frac{1}{15}\left((1-1.46)^2 + (2-1.46)^2 + (1-1.46)^2 + \cdots\right) = 0.2666$$

$$s_y^2 = \frac{1}{15}\left((1-2.26)^2 + (3-2.26)^2 + (2-2.26)^2 + \cdots\right) = 0.7809$$

Kovariance

$$C[X,Y] = \frac{1}{15}\left[(1-1.46)(1-2.26) + (2-1.46)(3-2.26) + (1-1.46)(2-2.26) + \cdots\right] = 0.1422$$

Příklad 5 (náhodný vektor)

Je dán náhodný vektor $[X, Y]'$ s hodnotami $x \in \{1,2,3\}$, $y \in \{1,2,3,4\}$ a sdruženou pravděpodobnostní funkcí[9]

$X \backslash Y$	1	2	3	4
1	0.1	0.05	0.02	0.2
2	0.08	0.1	0.05	0.2
3	0.01	0.03	0.06	0.1

Určete podmíněnou pravděpodobnostní funkci $f(x|y)$ a $f(y|x)$.[10]

- Nejprve určíme marginály $f(x)$ a $f(y)$

[9] K předchozímu příkladu o botách a dopravních prostředcích nám přibude jeden dopravní prostředek (například 3 - kolo) a jedna obuv (4 - pantofle).

[10] Neboli s jakou pravděpodobností pojedu autem, na kole..., když vím, že jsem si vzala sandále $f(x|y)$, případně s jakou pravděpodobností si vezmu sandále, tenisky...., když vím, že jsem jela autem $f(y|x)$. Všimněte si, že se jedná o dva různé problémy.

$X \backslash Y$	1	2	3	4	$f(x)$
1	0.1	0.05	0.02	0.2	**0.37**
2	0.08	0.1	0.05	0.2	**0.42**
3	0.01	0.03	0.06	0.1	**0.2**
$f(y)$	**0.19**	**0.18**	**0.13**	**0.5**	

- Dále určíme $f(x|y)$ pro všechny možné podmínky y:

  první sloupec bude první sloupec sdružené dělený součtem prvního sloupce (1. hodnotou marginály pro y - viz definice)

  druhý sloupec podmíněné bude druhý sloupec sdružené dělený jeho součtem (příslušnou marginálou)

  atd.

$f(x\|y)$	1	2	3	4
1	0.526	0.278	0.154	0.4
2	0.421	0.556	0.385	0.4
3	0.053	0.167	0.462	0.2

  Pozn. Všimněte si, že součty ve sloupcích dají vždy 1. Když je pevně zvolena hodnota y, vyčerpávají hodnoty $x = 1, 2, 3$ všechny možnosti[11].

- Nakonec budeme konstruovat $f(y|x)$

  Tentokrát budeme brát řádky a normovat na jedničku, tj. dělit součtem řádků (v tabulce bude stále x svisle a y vodorovně)

$f(y\|x)$	1	2	3	4
1	0.27	0.135	0.054	0.541
2	0.186	0.233	0.116	0.465
3	0.05	0.15	0.3	0.5

  Pozn. Tady zase budou součty v řádcích rovny jedné.

Pozn.

1. Je důležité se uvědomit rozdíl mezi $f(x|y)$ a $f(y|x)$.
2. Nezapomínejte na definici pravděpodobnosti při kontrole výpočtu!

[11] Pokud vím, že mám například tenisky, tak si mohu vybírat pouze ze 3 dopravních prostředků, dle definice pravděpodobnosti musí být součet, prp. že pojedu autem, na motocyklu či kole právě 1