



BI-PST 2018

# Domácí Úkol

**Autoři:**

Pavel Jahoda a Jan Lidák

## 1 Úkol 1

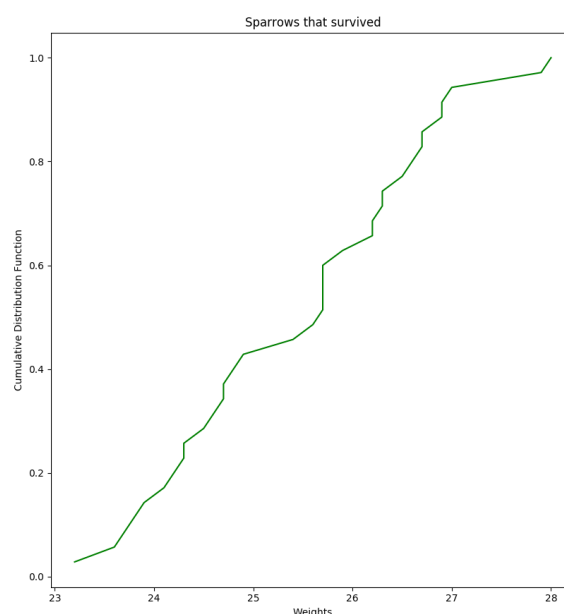
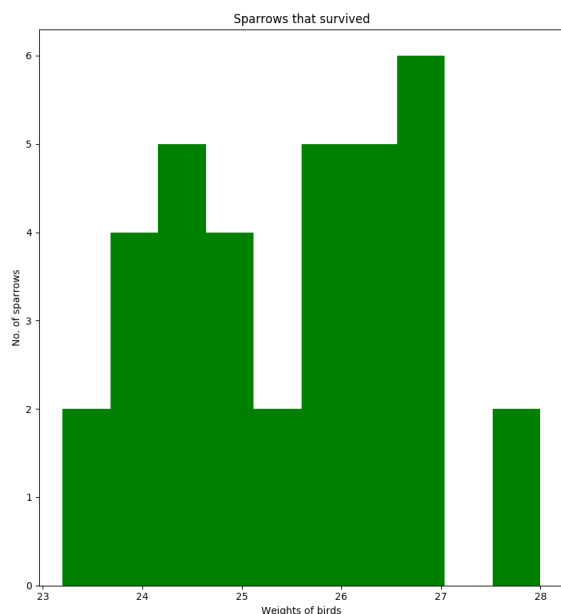
Data jsou z pozorování 59 vrabců během zimy. První veličina  $\mathbf{X}$  reprezentuje váhy vrabců v gramech. Druhá veličina  $\mathbf{Y}$  nabývá dvou hodnot 'survived', pokud vrabec přežil a 'perished' pokud nepřežil. Sledovanou proměnnou  $X$  jsme rozdělili na dvě pozorované skupiny takzvaných *independent and identically distributed random variables*. Tedy v každé skupině jsou náhodné veličiny reprezentující výsledky pokusu prováděného za stejných podmínek.  $\mathbf{X1}$  jsou vrabci co přežili a je jich 35.  $\mathbf{X2}$  jsou vrabci co nepřežili a je jich 24.

$EX1=25.463$ ,  $\text{var}(X1)=1.539$  a medián je 25.7.

$EX2=26.275$ ,  $\text{var}(X2)=2.078$  a medián je 26.

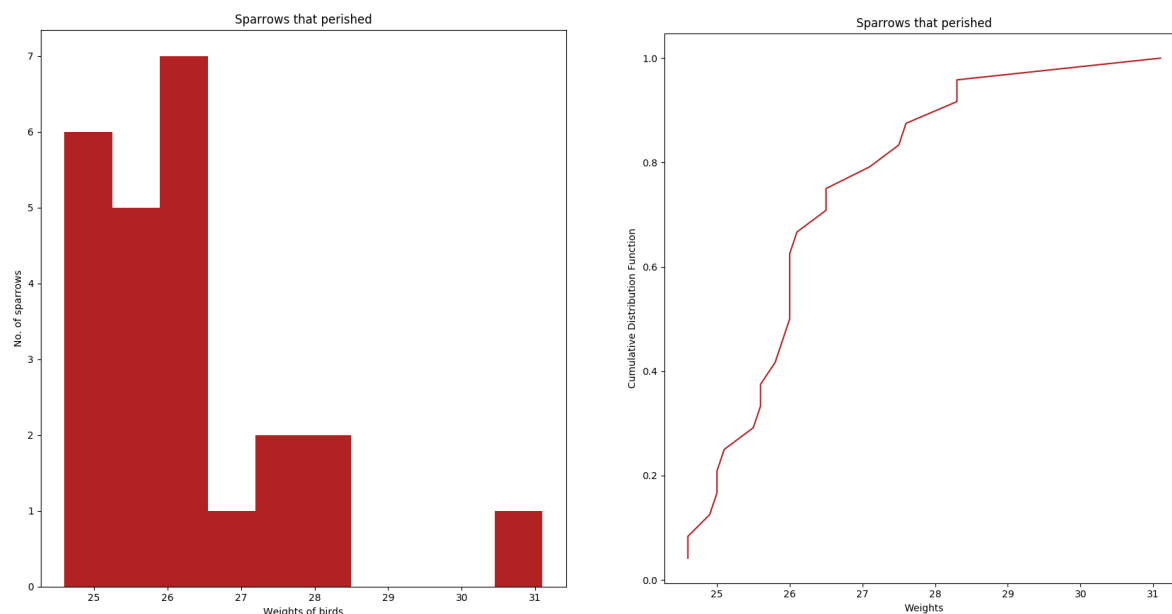
## 2 Úkol 2

Nejprve vykreslíme histogram a graf empirické distribuční funkce pro vrabce kteří přežili.



TODO.

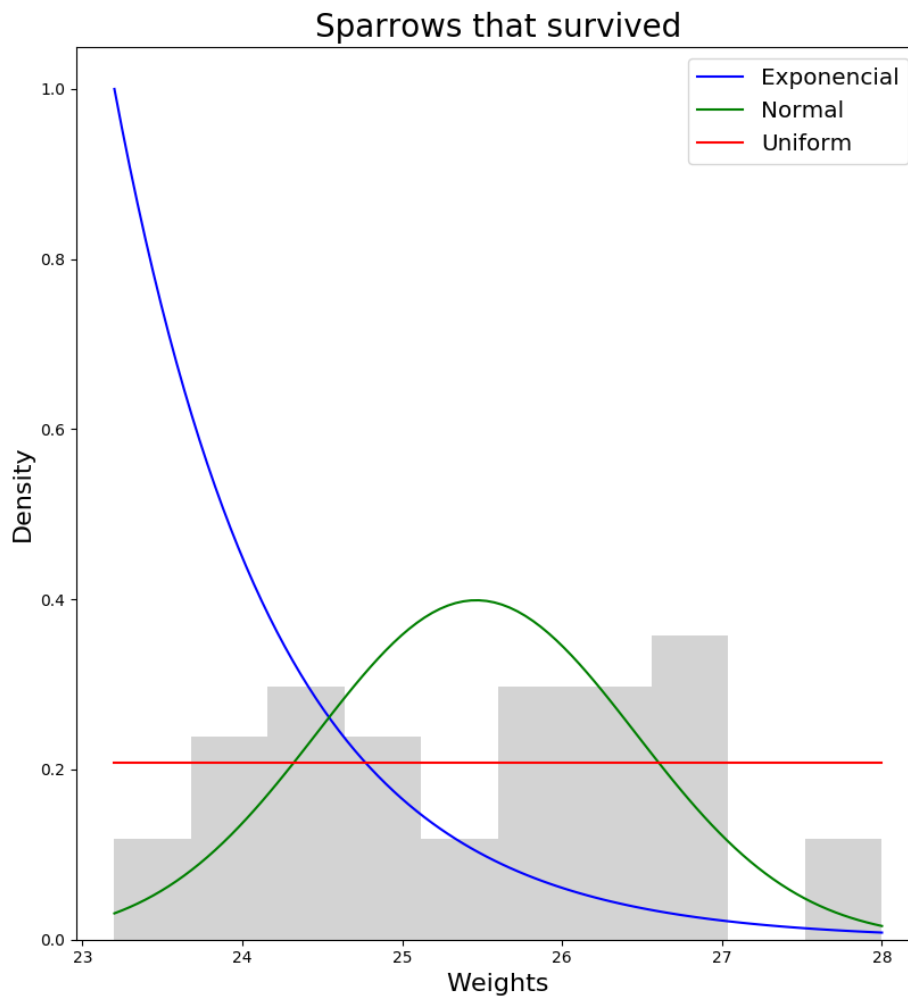
Poté vykreslíme histogram a graf empirické distribuční funkce pro vrabce kteří nepřežili.



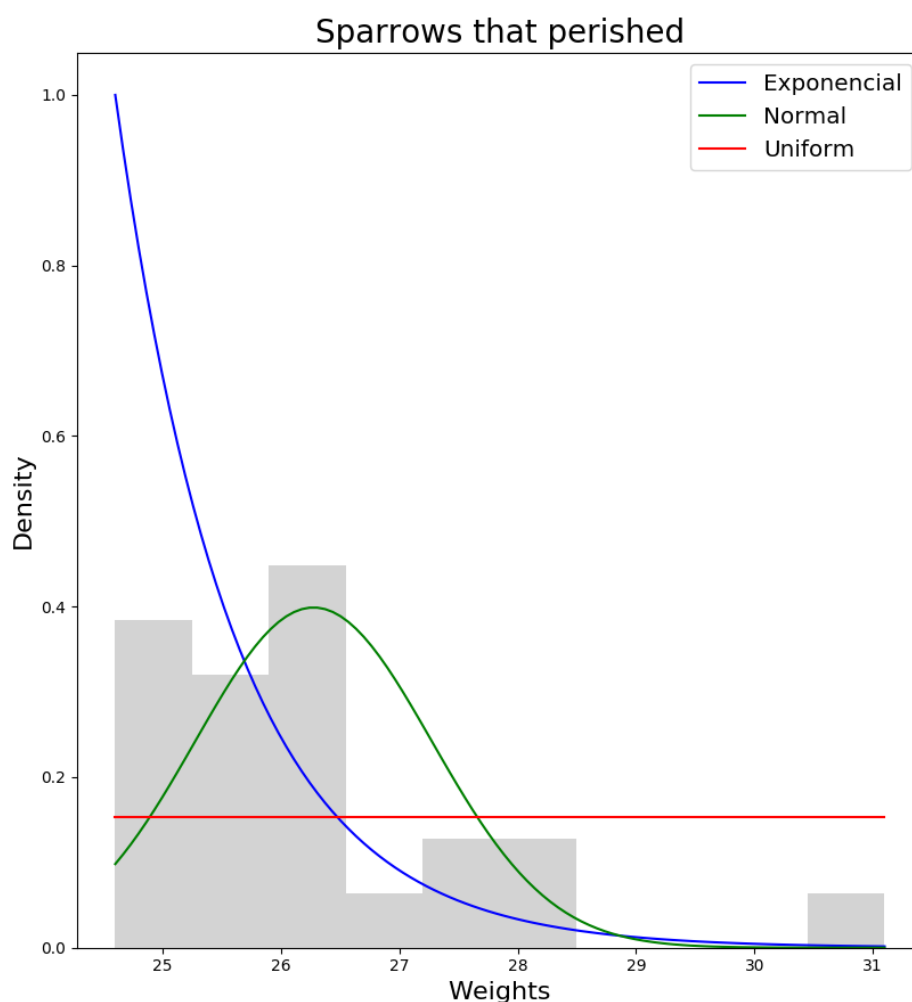
Graf empirické distribuční funkce se podobá grafu exponenciálního rozdělení s parametrem  $\lambda = 1$ .

### 3 Úkol 3

Histogram přeživších vrabců byl nejprve znormován, tak aby obsah histogramu byl roven 1 pro lepší porovnání s grafy rozdělení jejich obsah pod křivkou je také roven 1. Po zanesení normálního, exponenciálního a rovnoměrného rozdělení s odhadnutými parametry do grafů histogramu vidíme, že histogram nejvíce odpovídá normálnímu rozdělení.



Určení, které rozdělení odpovídá nejlepě grafu vrabců, kteří zimu nepřežili je obtížnější, jelikož to není vizuálně patrné. První způsob, který jsem použil na zjištění nejbližšího rozdělení bylo spočítat součet rozdílů mezi výškami sloupců histogramů a hodnotami funkcí pro  $x$  rovno středu daného sloupce. V tomto způsobu vyšlo normální rozdělení jako rozdělení které histogramu více odpovídá ( 0.73 normalání a 0.94 exponenciální). Dále mi co se zjištění podobnosti přišlo přirozené penalizovat extrémní rozdíly mezi výškami sloupců histogramů a hodnotami funkcí. Umocnění rozdílu zapříčiní požadovanou penalizaci. V tomto případě vyšla výsledná suma normálního rozdělení 0.13 oproti 0.20 u exponenciálního rozdělení, tudíž si myslíme, že normální rozdělení je nejvíce podobné histogramu zemřelých vrabců. Statistická významnost tohoto tvrzení se stejně jako u vrabců kteří přežili odvíjí od počtu náhodných veličin, kterých je 24 (nepřežili) a 35 (přežili).

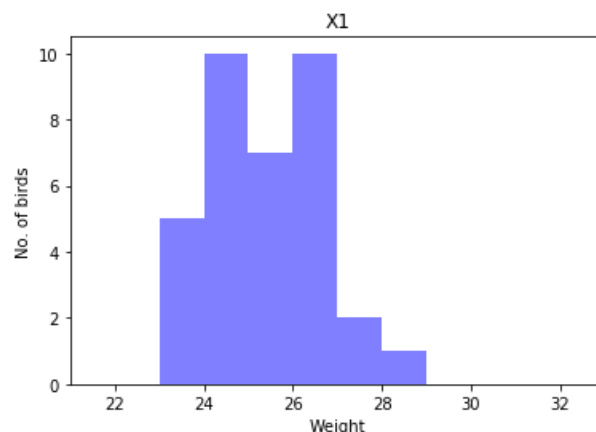
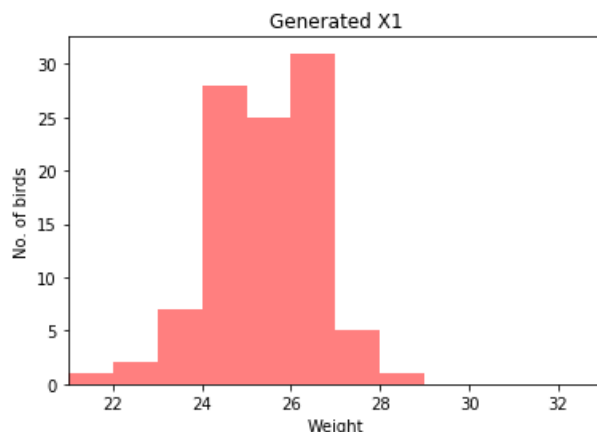


## 4 Úkol 4

Na následujících grafech je zobrazeno 100 vygenerovaných hodnot spolu s načtenými daty z datasetu. Histogram přeživších opeřenců vypadá jako normálního rozdělení s parametry  $\lambda = EX = 25.793$  a  $\sigma^2 = \text{var}X = 1.918$ , hodnoty byly tedy generovány s těmito parametry.

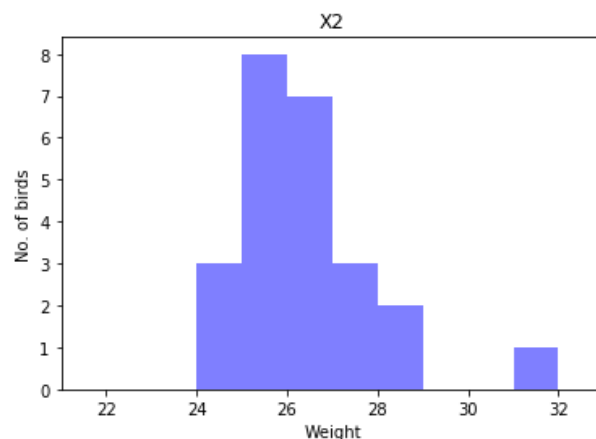
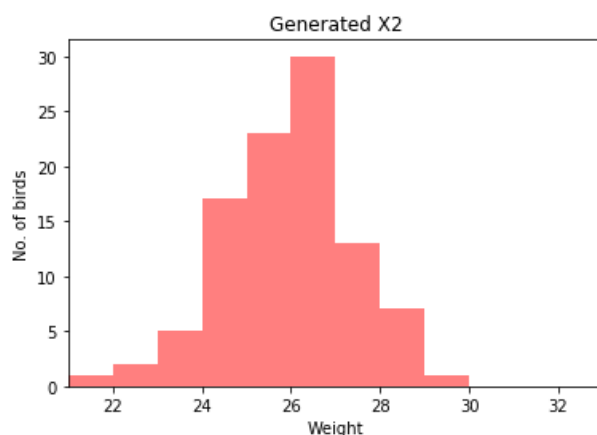
První graf vygenerovaných dat poměrně silně připomíná získaná data, věříme tedy že

toto rozdělení bylo zvoleno správně.



Data vygenerovaná pro mrtvé operece už na tom nejsou tak dobře, část dat "utíká" do strany, zde by se hodilo mít větší dataset aby se dala lépe odhadnout distribuční funkce, popřípadě lépe vystihnout parametry rozdělení.

Pro operece jež nepřežili byla data generována s normální distribuční funkcí s parametry  $\lambda = EX = 26.275$  a  $\sigma^2 = \text{var}X = 2.078$ .



## 5 Úkol 5

$\bar{X}_n$  je výběrový průměr veličiny,  $\sigma$  je směrodatná odchylka,  $n$  je počet prvků v rozdělení. Spolehlivost má být 95%, tedy

$$\alpha = 0.05 \Rightarrow \alpha/2 = 0.025$$

$$z_{\alpha/2} = z_{0.025} = \Phi^{-1}(0.975) \doteq 1.96$$

Oboustranný 95% konfidenční interval spočteme jako:

$$(S_{X_1}, U_{X_1}) = (\bar{X}_1 - z_{\alpha/2} \cdot \frac{\sigma_1}{\sqrt{n_1}}, \bar{X}_1 + z_{\alpha/2} \cdot \frac{\sigma_1}{\sqrt{n_1}})$$

$$= (25,463 - 1,96 \cdot \frac{1,24}{5,9}, 25,463 + 1,96 \cdot \frac{1,24}{5,9}) = (25.052, 25.874) \quad (1)$$

$$\begin{aligned}(S_{X_2}, U_{X_2}) &= (\bar{X}_{2n} - z_{a/2} \cdot \frac{\sigma_2}{\sqrt{n_2}}, \bar{X}_{2n} + z_{a/2} \cdot \frac{\sigma_2}{\sqrt{n_2}}) \\ &= (26.275 - 1,96 \cdot \frac{1,44}{4,9}, 26.275 + 1,96 \cdot \frac{1,44}{4,9}) = (25.698, 26.852)\end{aligned}\quad (2)$$