Kurzy

- **Kurz na jev *A***

$$\boxed{O(A) = \frac{P(A)}{P(\overline{A})} = \frac{P(A)}{1 - P(A)}}$$

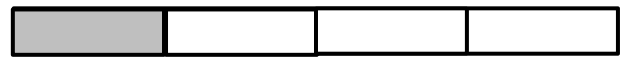
- **Pravděpodobnost jevu *A***

$$P(A) = \frac{O(A)}{1 + O(A)}$$

- **Kurz na to, že jev *A* nenastane** $\quad O(\overline{A}) = \dfrac{1}{O(A)}$

např.

$O(A) = 1{:}3$



$P(A) = 0.25$

Kurzy bookmakerů

- kurz 4 : 1 vypsaný na jev *A*

znamená, že pokud jev *A* nastane, tak při vkladu 100 kč dostanu 100 + 400 = 500 kč

tj. kurz vypsaný bookmakery je kurz na to že jev nenastane $O(\overline{A}) = \dfrac{1}{O(A)}$

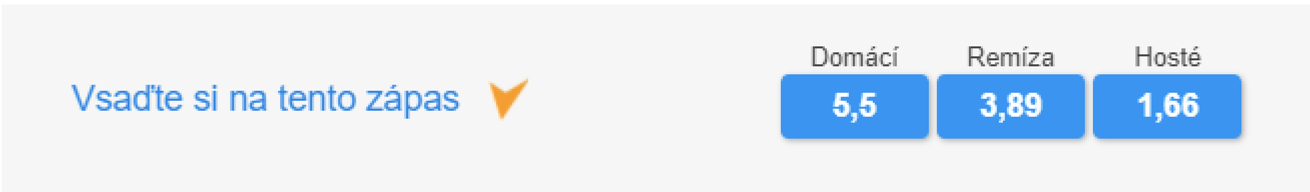
- desetinné kurzy: $O(\overline{A}) + 1 = \dfrac{1}{P(A)}$

- např. kurz 4 : 1 odpovídá desetinnému kurzu 5.0

to znamená, že pokud jev *A* nastane, tak při vkladu 100 kč dostanu 100 × 5.0 = 500 kč

Kurzy bookmakerů

- Mladá Boleslav vs. Slavia Praha, 14.3. 2021

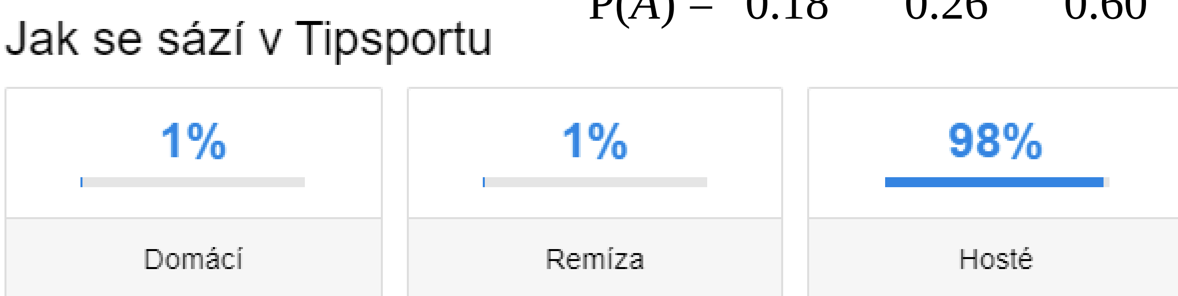


$P(A) =$ 0.18 0.26 0.60

Součet:

$0.18 + 0.26 + 0.60 = 1.04$

To je proto aby bookmaker vydělal.



Kurzy bookmakerů

	domácí	remíza	hosté	součet
$P(A)$	0.50	0.30	0.20	1.0
zvýšíme pravděpodobnosti	0.60	0.40	0.30	1.3
vypsané kurzy	1.7	2.5	3.3	
vklady	50	30	20	100

	vyplatíme	náš zisk
výhra domácích	85	15
remíza	75	25
výhra hostů	66	34

Kurzy

- **Bayesův teorém pro kurzy**

$$\boxed{O(A|B) = \frac{P(A|B)}{P(\overline{A}|B)} = LR(A,B)O(A)}$$

- **Vyjádření v dB**

$$I\,[\mathrm{dB}] = 10 \log I$$

$$\boxed{O(A|B)\,[\mathrm{dB}] = LR(A,B)\,[\mathrm{dB}] + O(A)\,[\mathrm{dB}]}$$

Podmíněná pravděpodobnost: příklad – Máte děťátko?

prostor událostí: Ω = {(t,+) (t,-) (n,+) (n,-)}

index selhání

10 žen za 1 rok ze 100 žen

$\Rightarrow P(T \text{ za 1 rok}) = 0.1$

$\Rightarrow P(T \text{ za 1 měsíc}) = 0.1/12 \cong 0.008$

T = {(t,+) (t,-)} – je těhotná $\quad P(T) = 0.008$

N = {(n,+) (n,-)} – není těhotná $\quad P(N) = 0.992$

R_+ = {(t,+) (n,+)} – test říká ano $\quad P(R_+ \mid T) = 0.995$

R_- = {(t,-) (n,-)} – test říká ne $\quad P(R_+ \mid N) = 0.01$

$$P(T \mid R_+) = \frac{P(R_+ \mid T)P(T)}{P(R_+ \mid T)P(T) + P(R_+ \mid N)P(N)}$$

$$P(T \mid R_+) = 0.45$$

$$LR(T, R_+) = \frac{P(R_+ \mid T)}{P(R_+ \mid N)} = \frac{0.995}{0.01} = 99.5 = 20\,\mathrm{dB}$$

$$O(T) = \frac{P(T)}{P(N)} = \frac{0.008}{0.992} = -20.9\,\mathrm{dB}$$

$$O(T \mid R_+) = LR(T, R_+)O(T)$$

$$O(T \mid R_+) = -0.9\,\mathrm{dB} = 0.81$$

$$P(T \mid R_+) = \frac{81}{181} = 0.45$$

Podmíněná pravděpodobnost: příklad – Máte děťátko?

prostor událostí: Ω = {(t,+) (t,-) (n,+) (n,-)}

index selhání

10 žen za 1 rok ze 100 žen

$$\Rightarrow P(T \text{ za 1 rok}) = 0.1$$

$$\Rightarrow P(T \text{ za 1 měsíc}) = 0.1/12 \cong 0.008$$

T = {(t,+) (t,-)} – je těhotná $\quad P(T) = 0.008$

N = {(n,+) (n,-)} – není těhotná $\quad P(N) = 0.992$

R_+ = {(t,+) (n,+)} – test říká ano $\quad P(R_+ \mid T) = 0.995$

R_- = {(t,-) (n,-)} – test říká ne $\quad P(R_+ \mid N) = 0.01$

$$P(T \mid R_+) = \frac{P(R_+ \mid T)P(T)}{P(R_+ \mid T)P(T) + P(R_+ \mid N)P(N)}$$

$$P(T \mid R_+) = 0.45$$

$$LR(T, R_+) = \frac{P(R_+ \mid T)}{P(R_+ \mid N)} = \frac{0.995}{0.01} = 99.5 = 20\,\text{dB}$$

dva pozitivní testy

$$O(T \mid R_+^2) = 19.1\,\text{dB} = 81.3$$

$$O(T \mid R_+^2) = LR(T, R_+)\,O(T \mid R_+) = 19.1\,\text{dB}$$

$$P(T \mid R_+^2) = \frac{813}{823} = 0.988$$

Kostky

`kostky.py`

- kostky

- počet stran: 4 6 8 12 20

- Někdo náhodně vytáhne jednu kostku a hodí s ní. Padne 6.

- Jaká je pravděpodobnost, že se jedná o kostku s 4, 6, 8, 12 nebo 20 stranami?

- $\Omega = \{4, 6, 8, 12, 20\}$

  hypotézy h_i

- prior $P(h_i) = \frac{1}{5}$

- data: padlo $k = 6$

- posterior $P(h_i|k) = \dfrac{P(k|h_i)P(h_i)}{P(k)}$

- věrohodnost: $P(k|h_i) = \begin{cases} 0 & k > h_i \\ \frac{1}{h_i} & k \leq h_i \end{cases}$

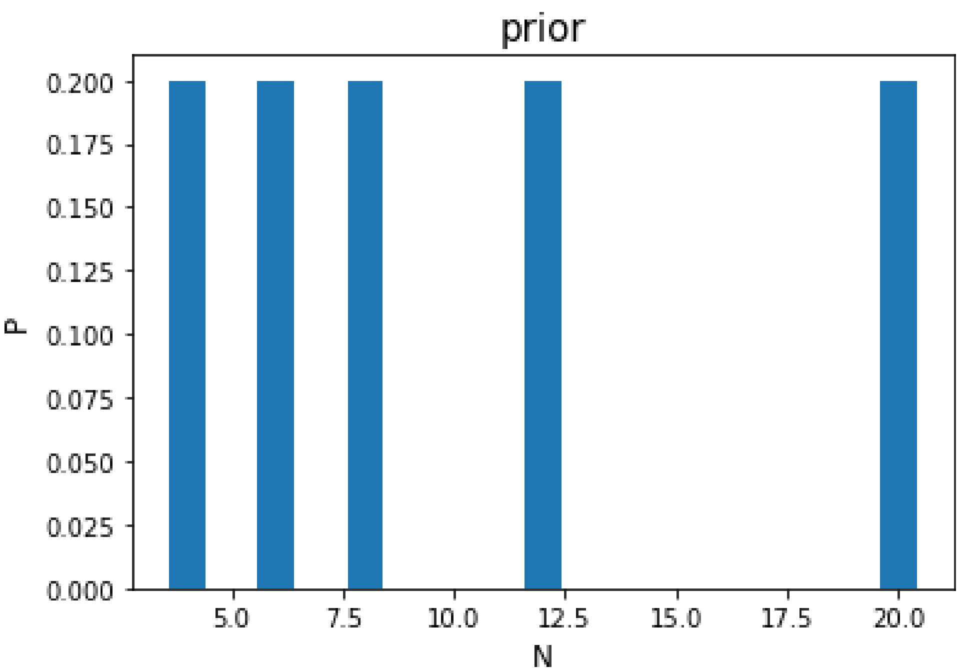
- zákon celkové pravděpodobnosti: $P(k) = \sum_{i=1}^{5} P(k|h_i)P(h_i)$

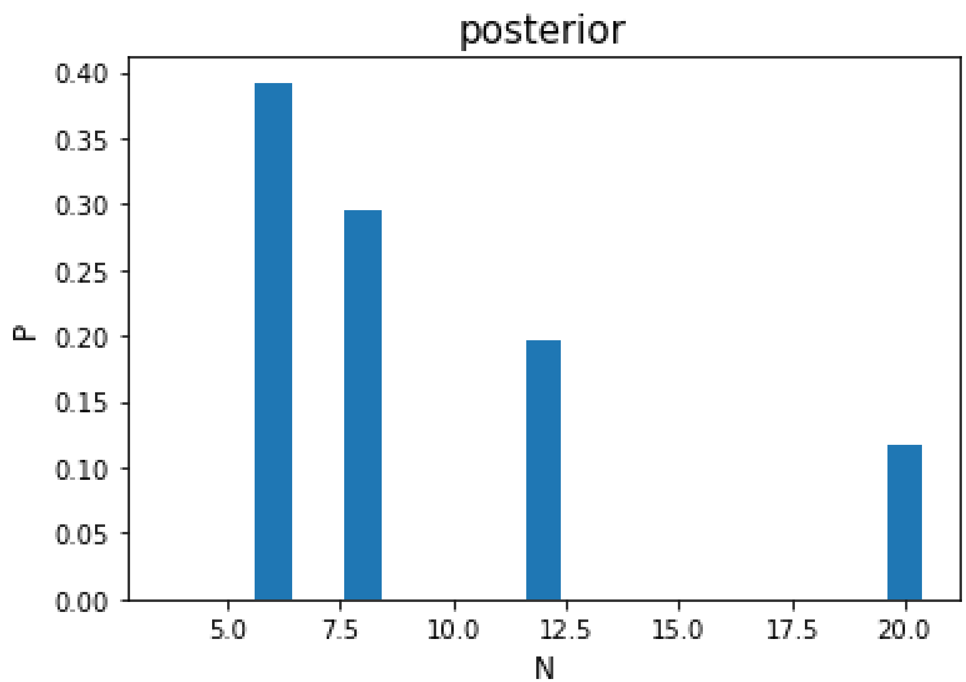
Kostky

`kostky.py`

- kostky

• počet stran:	4	6	8	12	20
• prior:	0.2	0.2	0.2	0.2	0.2
• posterior:	0	0.39	0.29	0.20	0.12



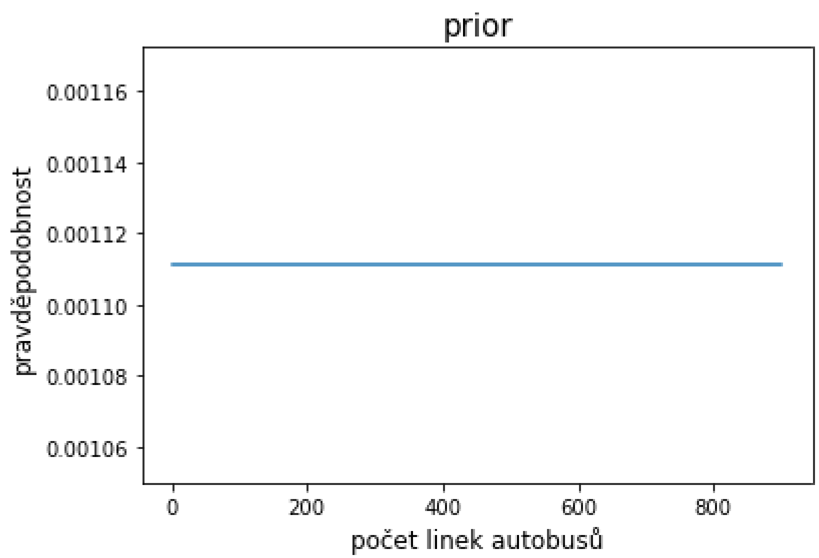


Autobusy

`autobusy.py`

Linky autobusů v Praze nají 3 číslice. Při náhodné procházce Prahou jsme viděli autobusy číslo 177, 108, 159, 124, 223, 185, 107, 218, 149, 175. Kolik je v Praze linek autobusu?

- Jestliže je maximální číslo linky autobusu N_{max}, tak je počet autobusů $N_a = N_{max} - 99$
- prior: jakýkoliv počet autobusů mezi 1 a 900 je stejně pravděpodobný $$P(N_a) = \frac{1}{900}$$



- věrohodnost: $$p(k|N_a) = \begin{cases} 0 & k > N_a + 99 \\ \frac{1}{N_a} & k \leq N_a + 99 \end{cases}$$
- posterior: $$P(N_a|k) = \frac{P(k|N_a)P(N_a)}{P(k)}$$
- zákon celkové pravděpodobnosti: $$P(k) = \sum_{N_a=1}^{900} P(k|N_a)P(N_a)$$

Autobusy

`autobusy.py`

Linky autobusů v Praze nají 3 číslice. Při náhodné procházce Prahou jsme viděli autobusy číslo 177, 108, 159, 124, 223, 185, 107, 218, 149, 175. Kolik je v Praze linek autobusu?

- Update informace:

- prior: $P(N_a) = \frac{1}{900}$

$\downarrow$ první pozorování k_1

- posterior: $P(N_a|k_1) = \frac{P(k_1|N_a)P(N_a)}{P(k_1)}$

$\downarrow$

- vezmi posterior $P(N_a|k_1)$ jako nový prior

$\downarrow$ druhé pozorování k_2

- posterior: $P(N_a|k_1, k_2) = \frac{P(k_2|N_a)P(N_a|k_1)}{P(k_2)}$

$\downarrow$ ….

- věrohodnost: $p(k_i|N_a) = \begin{cases} 0 & k_i > N_a + 99 \\ \frac{1}{N_a} & k_i \leq N_a + 99 \end{cases}$

- zákon celkové pravděpodobnosti: $P(k_i) = \sum_{N_a=1}^{900} P(k_i|N_a)P(N_a)$

Autobusy

`autobusy.py`

Linky autobusů v Praze nají 3 číslice. Při náhodné procházce Prahou jsme viděli autobusy číslo 177, 108, 159, 124, 223, 185, 107, 218, 149, 175. Kolik je v Praze linek autobusu?

- Update informace:
- prior: $P(N_a) = \dfrac{1}{900}$

$\downarrow$ první pozorování k_1

- posterior: $P(N_a|k_1) = \dfrac{P(k_1|N_a)P(N_a)}{P(k_1)}$

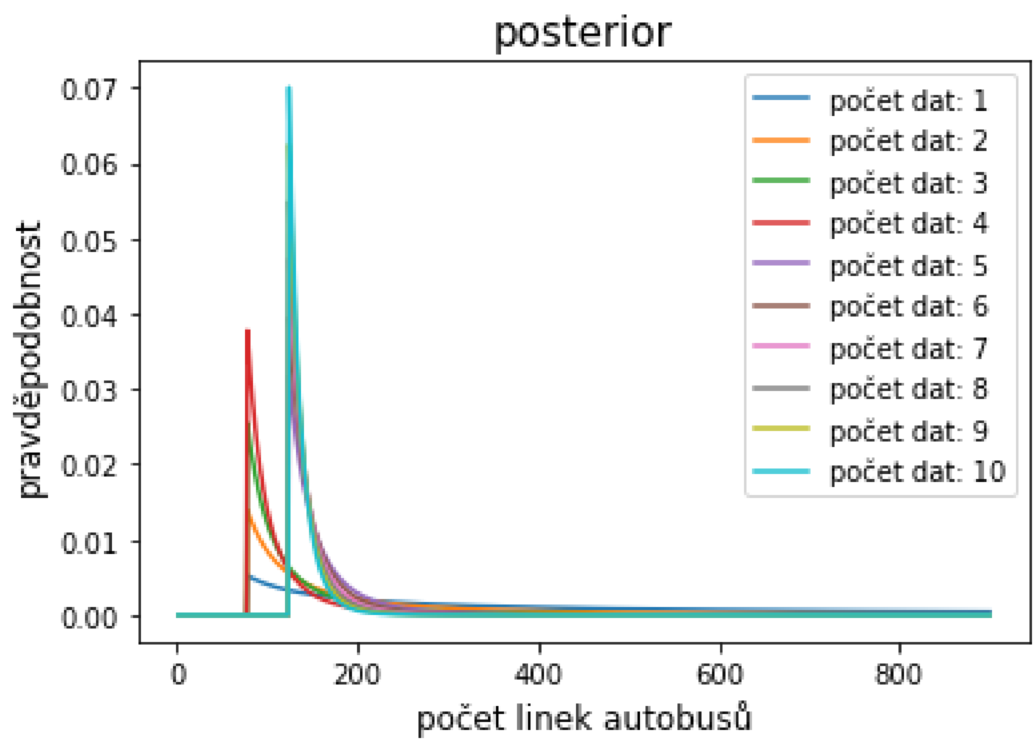
$\downarrow$

- vezmi posterior $P(N_a|k_1)$ jako nový prior

$\downarrow$ druhé pozorování k_2

- posterior: $P(N_a|k_1, k_2) = \dfrac{P(k_2|N_a)P(N_a|k_1)}{P(k_2)}$

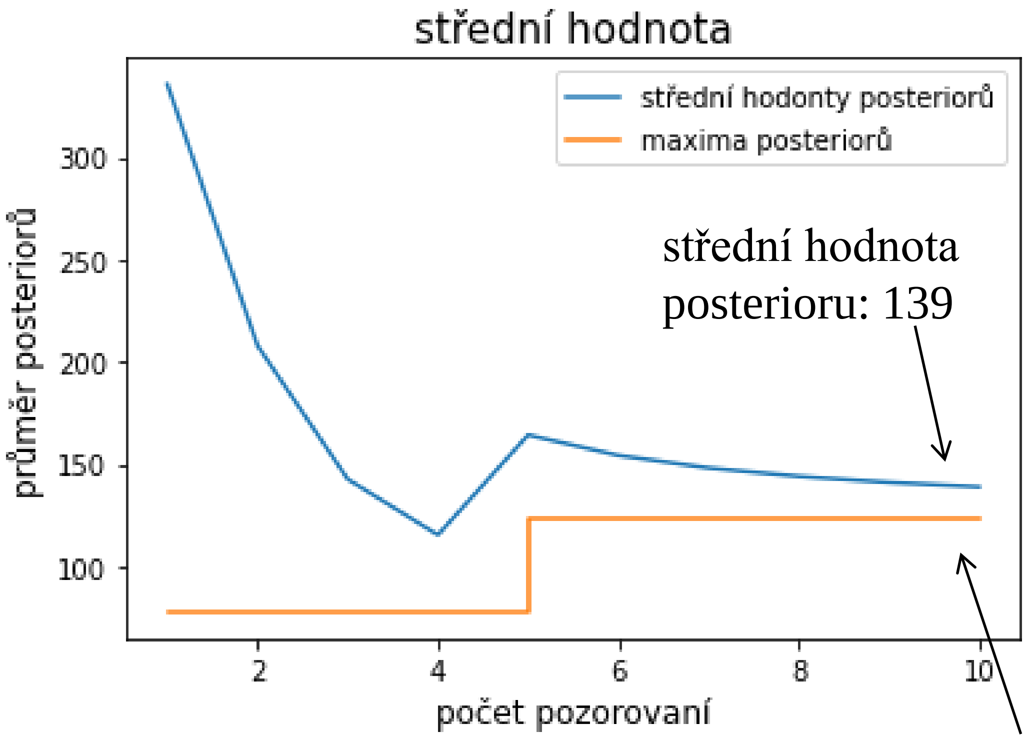
$\downarrow$ ….



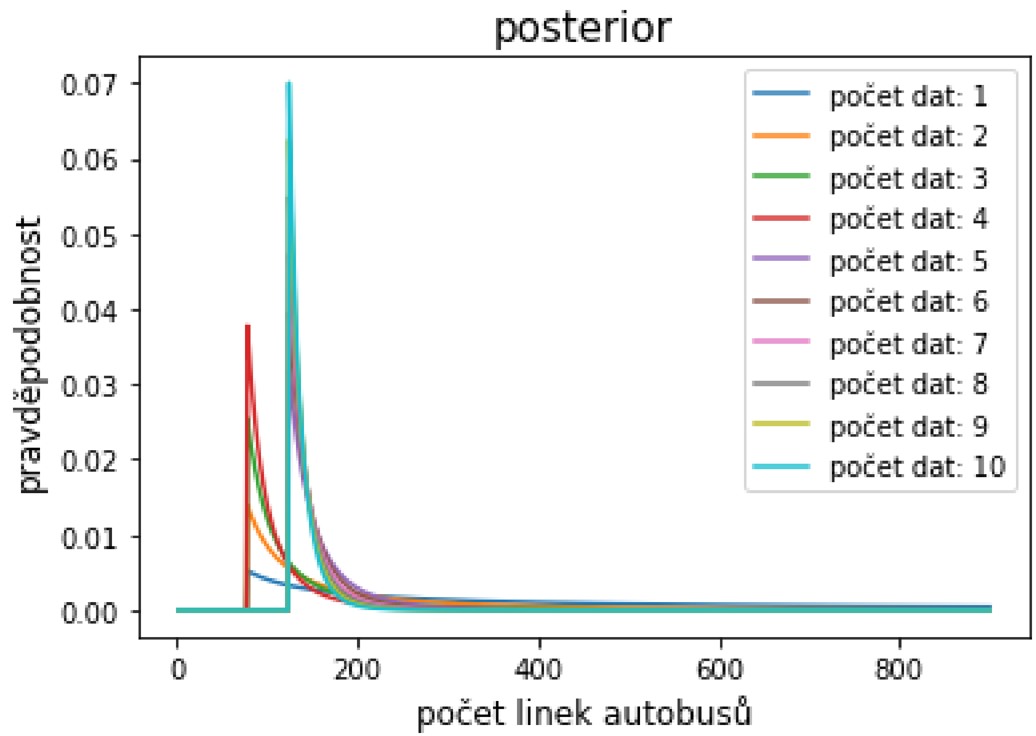
Autobusy

`autobusy.py`

Linky autobusů v Praze nají 3 číslice. Při náhodné procházce Prahou jsme viděli autobusy číslo 177, 108, 159, 124, 223, 185, 107, 218, 149, 175. Kolik je v Praze linek autobusu?



střední hodnota posterioru: 139



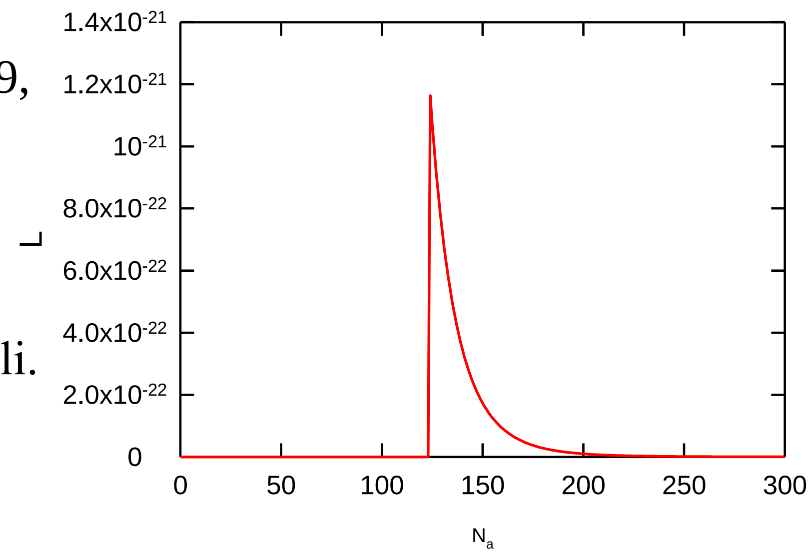
nejpravděpodobnější hodnota (maximum posterioru): 223 – 99 = 124

Autobusy (klasický přístup)

Linky autobusů v Praze nají 3 číslice. Při náhodné procházce Prahou jsme viděli autobusy číslo 177, 108, 159, 124, 223, 185, 107, 218, 149, 175. Kolik je v Praze linek autobusu?
Pomocí metody maximální věrohodnosti najděte odhad počtu linek a spočítejte jeho předpojatost.

- Jestliže je maximální číslo linky autobusu N_{max}, tak je počet autobusů $N_a = N_{max} - 99$
- Potom jsou čísla linek 100, 101, 102, .. $N_a + 99$
- Pravděpodobnost, že uvidíme linku číslo k_i je $1/N_a$ pro každé $100 \leq k_i \leq N_a + 99$
- Jestliže jsme viděli linky číslo $k_1, k_2, \ldots, k_N$, je pravděpodobnost, že se toto stane $(N_a)^{-N}$
- Věrohodnostní funkce je tedy

$L(N_a|\{k\}) = (N_a)^{-N}$ pro $N_a \geq k_{\mathrm{max}}$-99 a nulová pro $N_a < k_{\mathrm{max}}$-99, kde k_{max} je největší z čísel autobusů, které jsme viděli.



- Takže z metody maximální věrohodnosti dostáváme jako nejlepší odhad počtu linek $\hat{N}_a = k_{\max} - 99$, kde k_{max} je nejvyšší číslo autobusové linky, které jsme viděli.
- V našem konkrétním případě je $\hat{N}_a = 223 - 99 = 124$

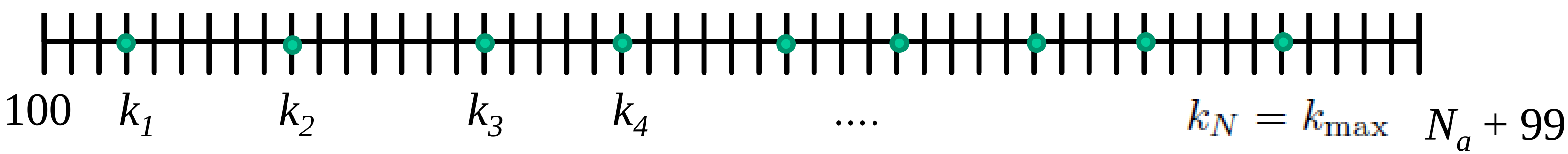
Autobusy (klasický přístup)

Linky autobusů v Praze nají 3 číslice. Při náhodné procházce Prahou jsme viděli autobusy číslo 177, 108, 159, 124, 223, 185, 107, 218, 149, 175. Kolik je v Praze linek autobusu?
Pomocí metody maximální věrohodnosti najděte odhad počtu linek a spočítejte jeho předpojatost.

- Předpojatost: $b = E[\hat{N}_a] - N_a$

- Jaká je předpojatost odhadu: $\hat{N}_a = k_{\max} - 99$
- Je zřejmé, že $\hat{N}_a \leq N_a \;\; \Rightarrow b < 0$ tedy předpojatost je záporná

$$b = -\left[\frac{k_{\max} - 99}{N}\right] + 1$$



- střední počet čísel, které jsme neviděli $k_{\max} - 99 - N$
- jednotlivá pozorovaná čísla jsou rozdělena na intervalu $<100, N_a+99>$ rovnoměrně
- střední počet čísel, která jsme neviděli na intervalu $<k_{\max}, N_a+99>$: $\left[\frac{k_{\max} - 99 - N}{N}\right] = \left[\frac{k_{\max} - 99}{N}\right] - 1$
- nepředpojatý odhad N_a je tedy: $\hat{N}_a = k_{\max} - 99 + \left[\frac{k_{\max} - 99}{N}\right] - 1$
- V našem konkrétním případě: $\hat{N}_a = 135$ (maximální číslo linky: 135 + 99 = 234)