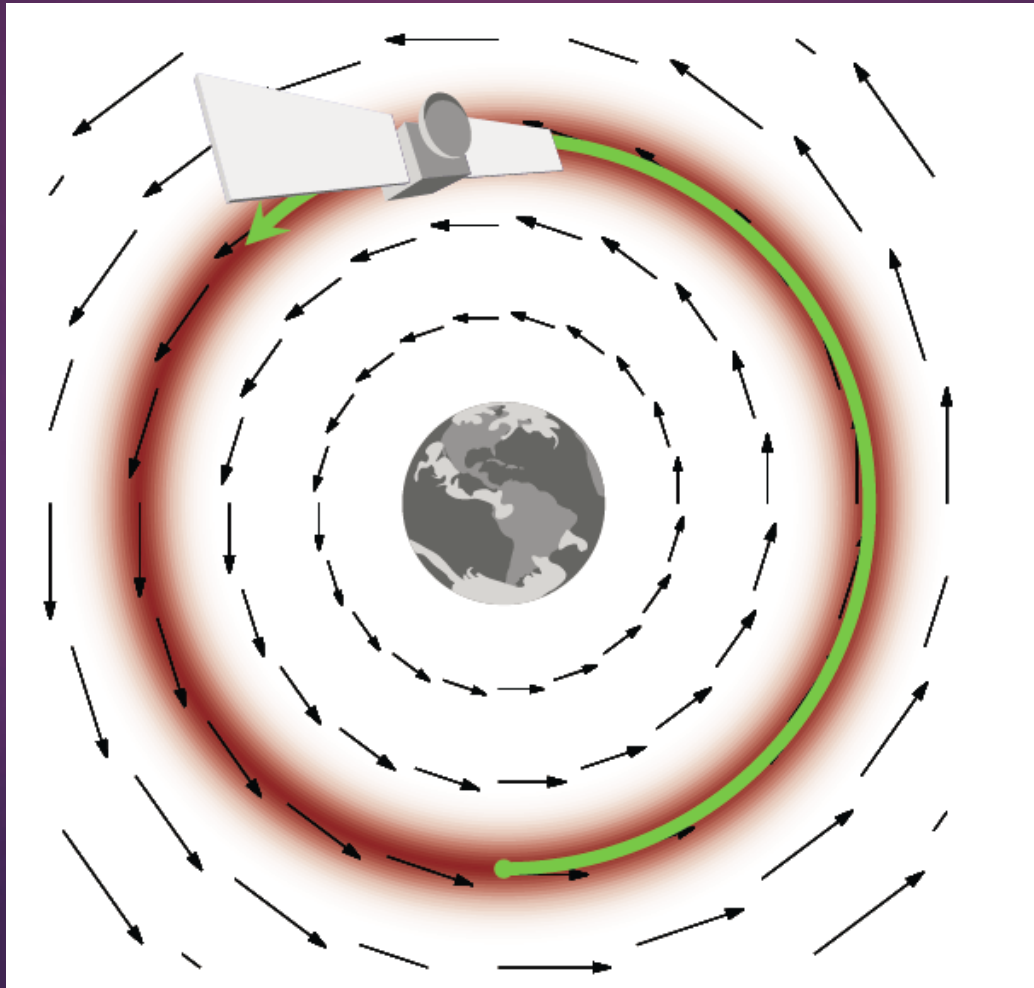


Hamiltonovské Monte Carlo



Filip Kostka

Úvod – o čem to bude

- ▶ *Hamiltonovské Monte Carlo* (HMC), dříve také *Hybridní Monte Carlo*, je gradientová Markov Chain Monte Carlo (MCMC) metoda.
- ▶ K procházení modelového prostoru využívá Hamiltonova formalismu klasické mechaniky.
- ▶ Efektivní alternativa k random-walk algoritmům.
- ▶ První použití Duane et al. (1987) (lattice QCD), dále rozpracoval a zpopularizoval Neal (2012, <https://arxiv.org/pdf/1206.1901.pdf>), teoreticky podložil Betancourt (2014 rigorózní článek: <https://arxiv.org/pdf/1701.02434.pdf>, 2017 laskavý článek: <https://arxiv.org/pdf/1410.5110.pdf>, přednáška: https://www.youtube.com/watch?v=pHsuaPbNbY&ab_channel=MLSSiceland2014)
- ▶ Aplikace v geofyzice: Muir & Tkalčić (2015); Sen & Biswas, (2017), **Fichtner & Simutè** (2017, <http://geo.mff.cuni.cz/seismosoft/clanky/FichtnerSimute.JGR2018.pdf>).

Úvod – o co jde

1. Chceme pracovat s *posteriorní* PDF, například abychom získali optimální modely z modelového prostoru, abychom ji vizualizovali, abychom spočítali očekávané hodnoty nějakých funkcí, abychom mohli získat marginální či podmíněné PDF, atd. $\Rightarrow$ potřebujeme počítat integrály.
2. Pokud je PDF složitá, nebo ji ani nedokážeme analyticky vyjádřit, tak to nepůjde deterministicky. Místo PDF budeme tedy pomocí náhodných metod hledat její *vzorky* a integrály aproximujeme $\Rightarrow$ metody *Monte Carlo*.
3. Pokud budeme *vzorky* hledat *postupným procházením* modelového prostoru pomocí nějakého procesu *nezávislého na historii*, dostaneme metody *Markov Chain Monte Carlo* (MCMC).
4. Díky *fundamentální větě o Markovových řetězcích* se dají najít takové procesy, které nám navzorkují právě naši PDF. Oblíbenou metodou je využití náhodné procházky - *Random Walk MCMC*
5. Ale to prochází modelový prostor *difuzivně*, je neefektivní, obzvlášť ve vysokých dimenzích
6. Lepší je procházet prostor *koherentně*.
7. Metodu, jak to udělat, nám poskytuje hamiltonovská dynamika, které vhodně využívá *Hamiltonovské Monte Carlo*.

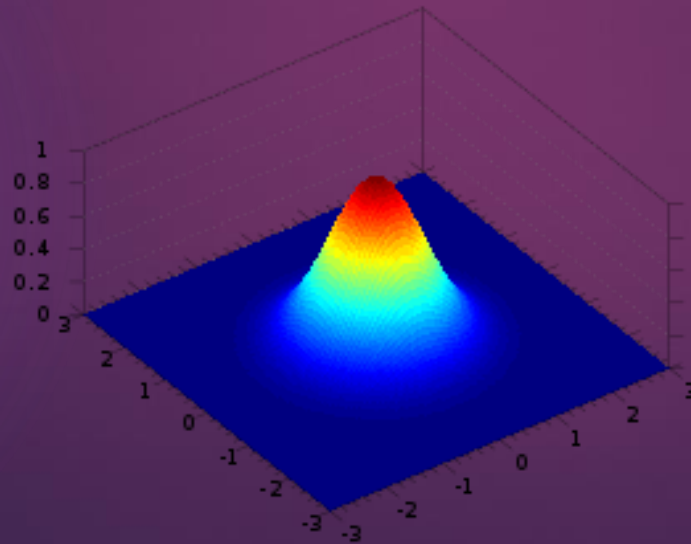
Úvod - osnova

1. Opakování – jak je definovaná pravděpodobnost? Co je to hustota pravděpodobnosti a proč na ní dávat pozor?
2. Monte Carlo, Markov Chain Monte Carlo
3. Fundamentální věta o Markovových řetězcích, Metropolis-Hastings
4. Liouvilleův teorém klasické mechaniky
5. Hamiltonovské Monte Carlo
6. Aplikace HMC: Fichtner 2017

Hustoty pravděpodobnosti – varovný příklad na začátek

- Mějme *standardní* gaussovskou hustotu pravděpodobnosti ve dvou dimenzích:

$$f(x, y|0,1) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}(x^2 + y^2)\right)$$

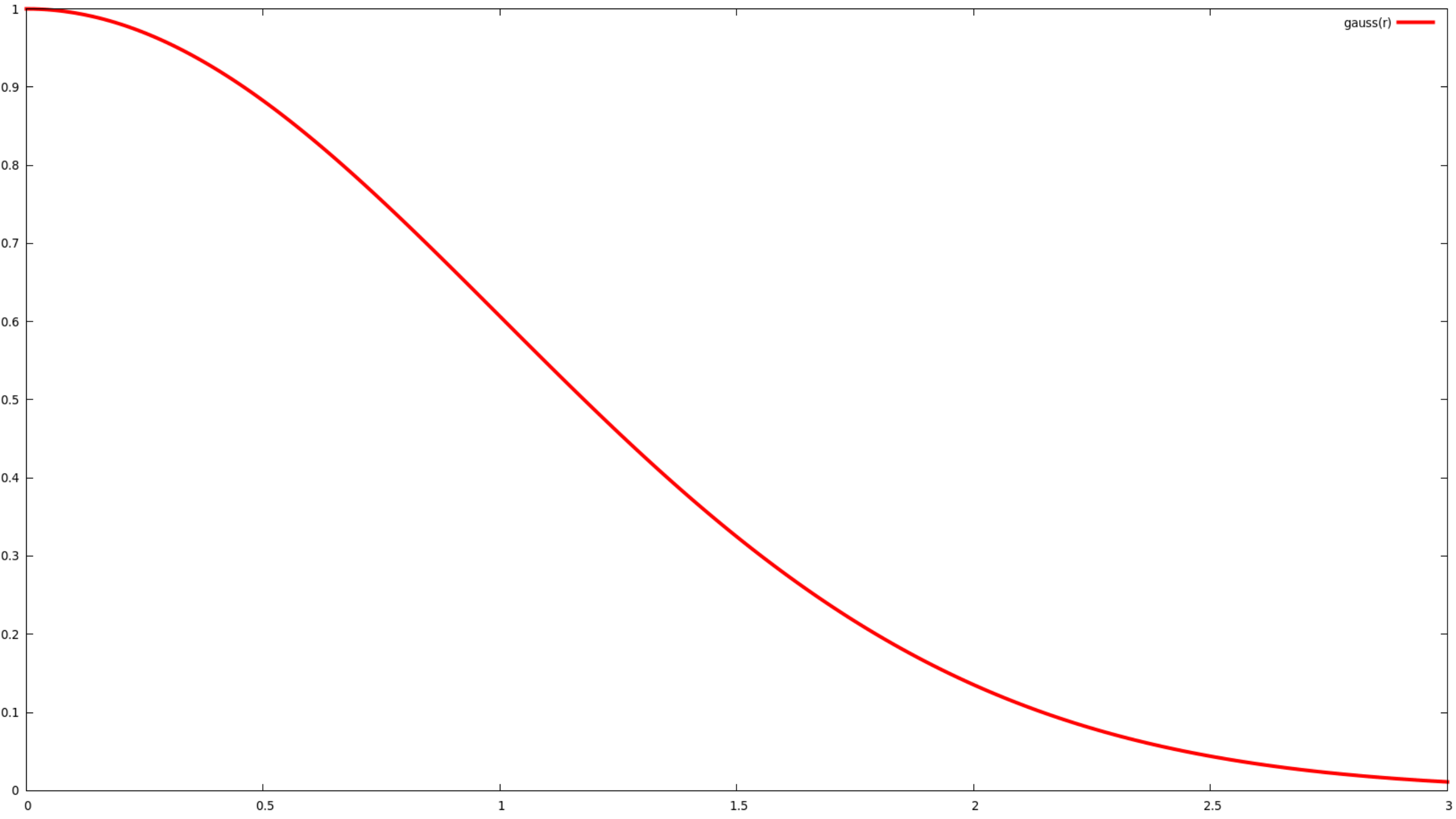


Hustoty pravděpodobnosti – varovný příklad na začátek

- ▶ Mějme *standardní* gaussovskou hustotu pravděpodobnosti ve dvou dimenzích:

$$f(x, y|0,1) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}(x^2 + y^2)\right)$$

- ▶ Nakresleme tuto funkci v závislosti na radiální souřadnici, $r = \sqrt{x^2 + y^2}$:



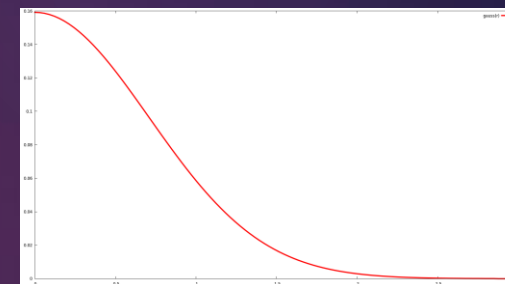
Hustoty pravděpodobnosti – varovný příklad na začátek

- ▶ Mějme *standardní* gaussovskou hustotu pravděpodobnosti ve dvou dimenzích:

$$f(x, y|0,1) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}(x^2 + y^2)\right)$$

- ▶ Nakresleme tuto funkci v závislosti na radiální souřadnici, $r = \sqrt{x^2 + y^2}$:

- ▶ Rozdělme rovinu na mezikruží o tloušťce Δr .



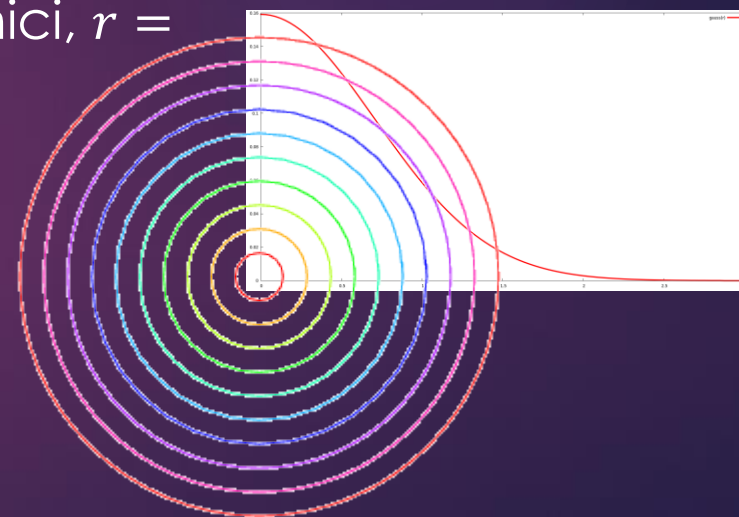
Hustoty pravděpodobnosti – varovný příklad na začátek

- ▶ Mějme *standardní* gaussovskou hustotu pravděpodobnosti ve dvou dimenzích:

$$f(x, y|0,1) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}(x^2 + y^2)\right)$$

- ▶ Nakresleme tuto funkci v závislosti na radiální souřadnici, $r = \sqrt{x^2 + y^2}$:

- ▶ Rozdělme rovinu na mezikruží o tloušťce Δr .



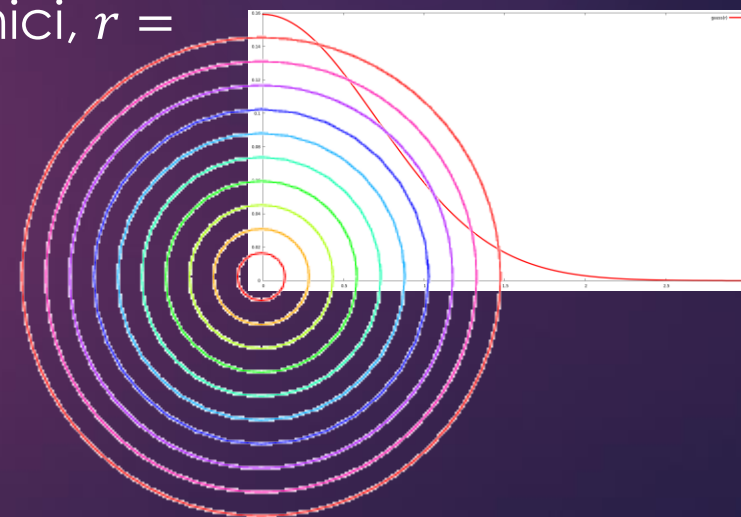
Hustoty pravděpodobnosti – varovný příklad na začátek

- ▶ Mějme *standardní* gaussovskou hustotu pravděpodobnosti ve dvou dimenzích:

$$f(x, y|0,1) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}(x^2 + y^2)\right)$$

- ▶ Nakresleme tuto funkci v závislosti na radiální souřadnici, $r = \sqrt{x^2 + y^2}$:

- ▶ Rozdělme rovinu na mezikruží o tloušťce Δr .
- ▶ V jakém mezikruží bude největší pravděpodobnost?



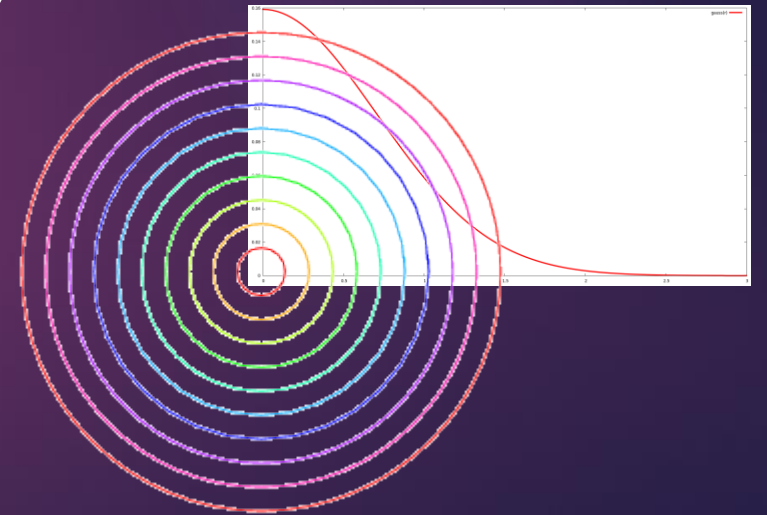
Hustoty pravděpodobnosti – varovný příklad na začátek

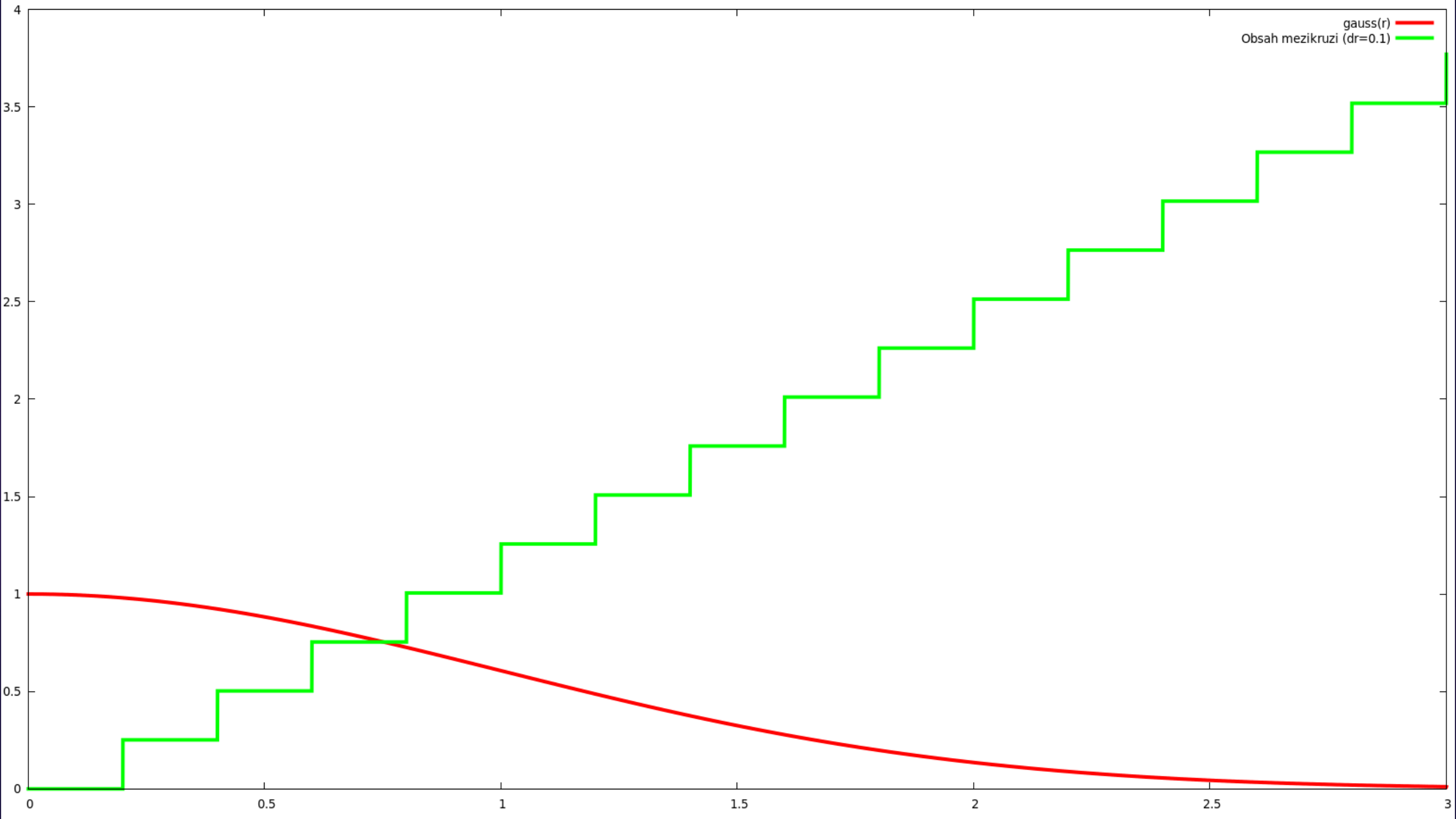
- ▶ Mějme *standardní* gaussovskou hustotu pravděpodobnosti ve dvou dimenzích:

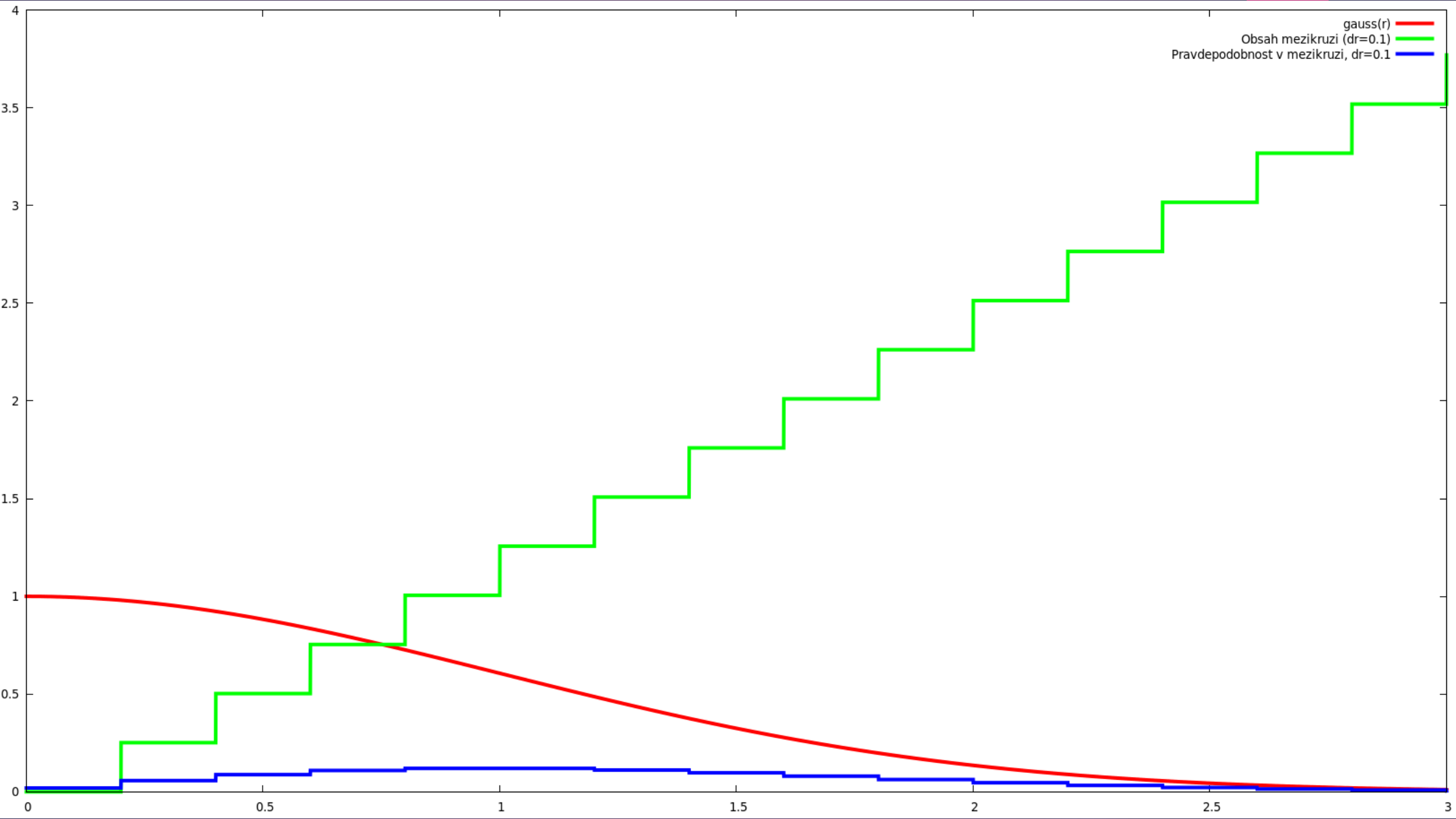
$$f(x, y|0,1) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}(x^2 + y^2)\right)$$

- ▶ Nakresleme tuto funkci v závislosti na radiální souřadnici,
 $r = \sqrt{x^2 + y^2}$:

- ▶ Rozdělme rovinu na mezikruží o tloušťce Δr .
- ▶ V jakém mezikruží bude největší pravděpodobnost?
- ▶ $P(r' \in [r, r + \Delta r]) \approx f(r)\Delta S(r) \approx f(r)2\pi r\Delta r$







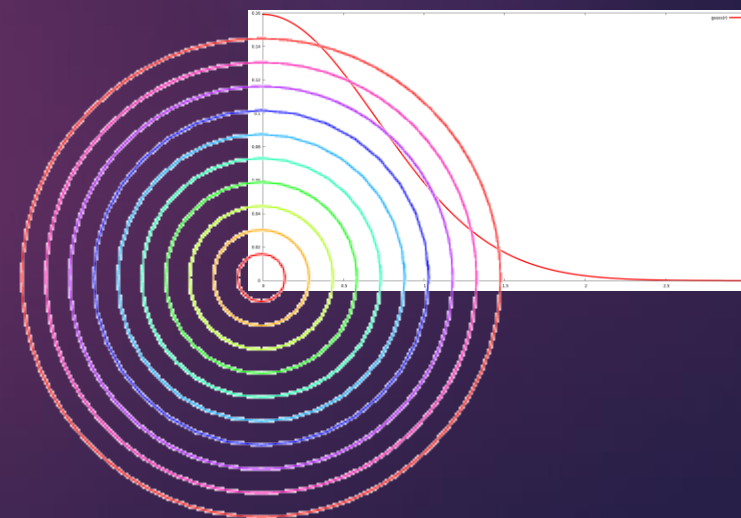
Hustoty pravděpodobnosti – varovný příklad na začátek

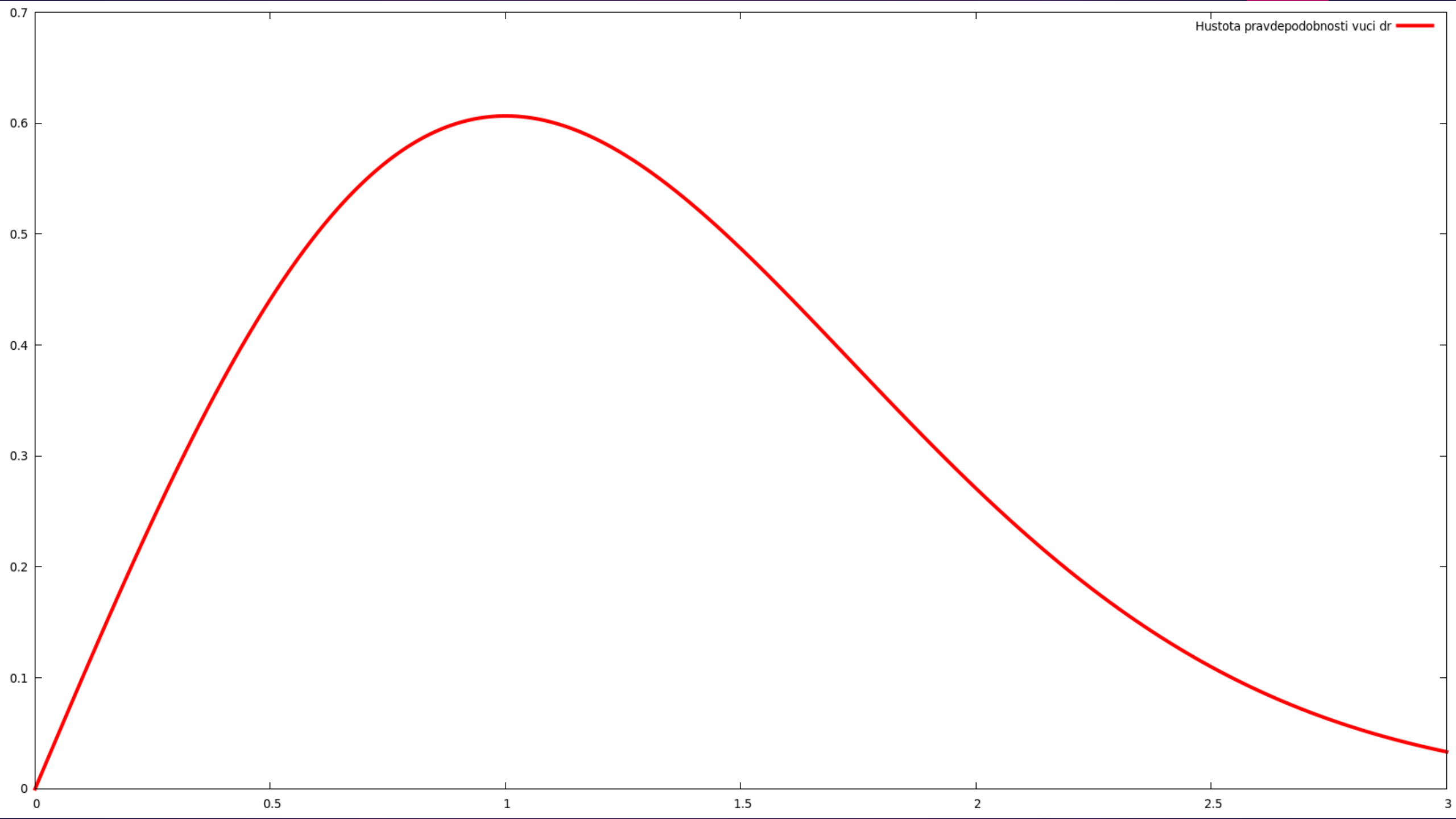
- ▶ Mějme *standardní* gaussovskou hustotu pravděpodobnosti ve dvou dimenzích:

$$f(x, y|0,1) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}(x^2 + y^2)\right)$$

- ▶ Nakresleme tuto funkci v závislosti na radiální souřadnici, $r = \sqrt{x^2 + y^2}$:

- ▶ Rozdělme rovinu na mezikruží o tloušťce Δr .
- ▶ V jakém mezikruží bude největší pravděpodobnost?
- ▶ $P(r' \in [r, r + \Delta r]) \approx f(r)\Delta S(r) \approx f(r)2\pi r\Delta r$
- ▶ Největší P tedy není kolem nuly, ale kolem 1.





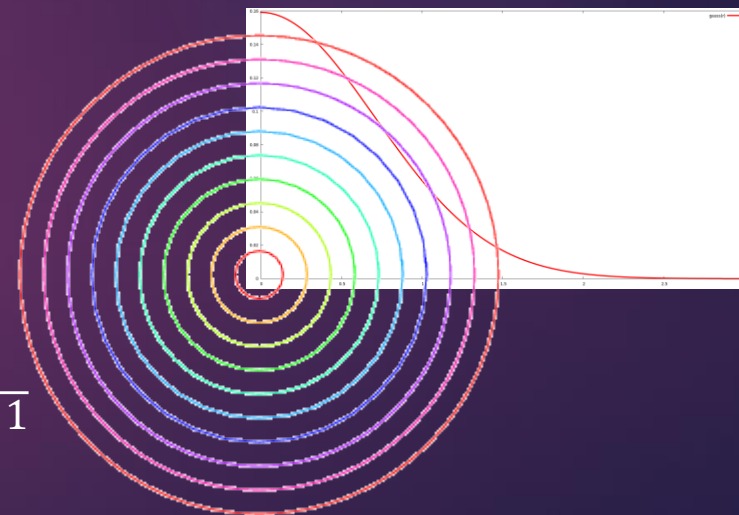
Hustoty pravděpodobnosti – varovný příklad na začátek

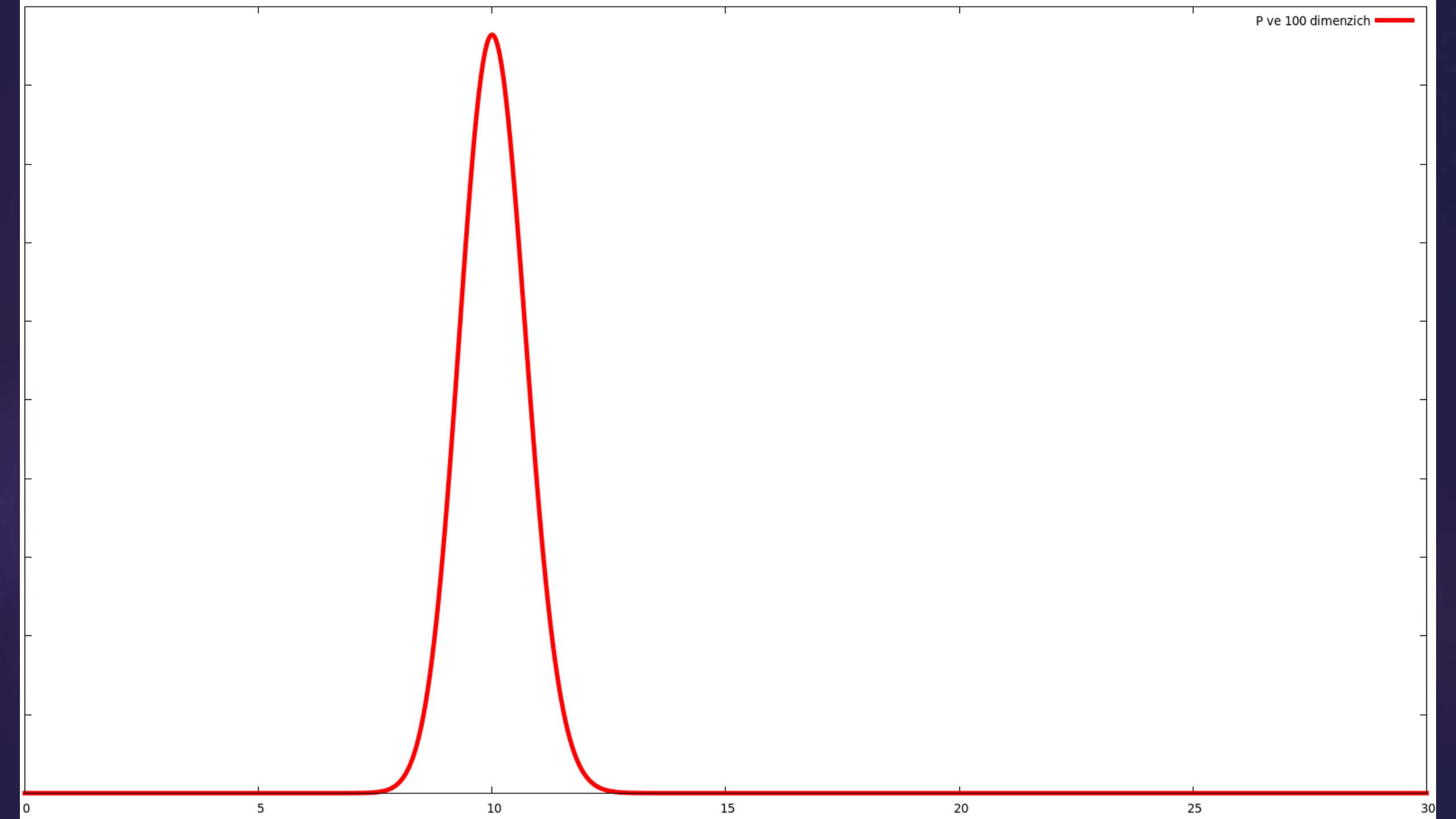
- ▶ Mějme *standardní* gaussovskou hustotu pravděpodobnosti ve dvou dimenzích:

$$f(x, y|0,1) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}(x^2 + y^2)\right)$$

- ▶ Nakresleme tuto funkci v závislosti na radiální souřadnici, $r = \sqrt{x^2 + y^2}$:

- ▶ Rozdělme rovinu na mezikruží o tloušťce Δr .
- ▶ V jakém mezikruží bude největší pravděpodobnost?
- ▶ $P(r' \in [r, r + \Delta r]) \approx f(r)\Delta S(r) \approx f(r)2\pi r\Delta r$
- ▶ Největší tedy není kolem nuly, ale kolem 1.
- ▶ S rostoucí dimenzí n bude tento efekt čím dál tím silnější –
prakticky veškerá pravděpodobnost se bude soustředit kolem $\sqrt{n-1}$





Hustoty pravděpodobnosti – varovný příklad na začátek

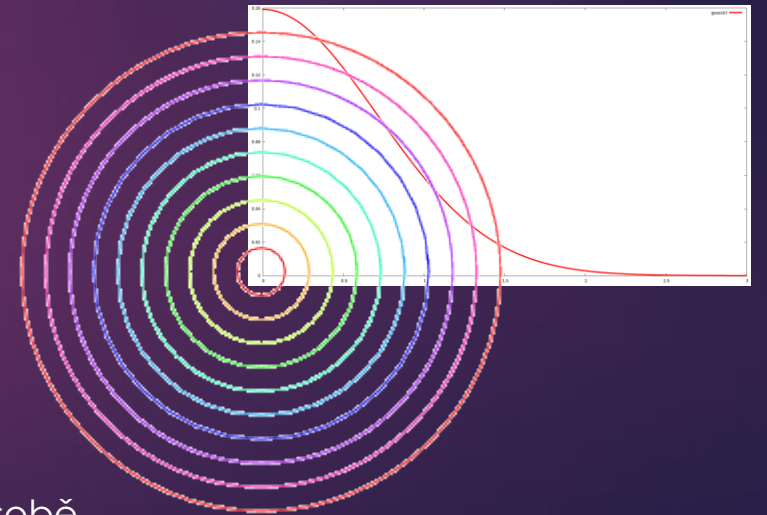
- ▶ Mějme *standardní* gaussovskou hustotu pravděpodobnosti ve dvou dimenzích:

$$f(x, y|0,1) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}(x^2 + y^2)\right)$$

- ▶ Nakresleme tuto funkci v závislosti na radiální souřadnici, $r = \sqrt{x^2 + y^2}$:

- ▶ Rozdělme rovinu na mezikruží o tloušťce Δr .
- ▶ V jakém mezikruží bude největší pravděpodobnost?
- ▶ $P(r' \in [r, r + \Delta r]) \approx f(r)\Delta S(r) \approx f(r)2\pi r\Delta r$
- ▶ Největší tedy není kolem nuly, ale kolem 1
- ▶ S rostoucí dimenzí n bude tento efekt čím dál tím silnější –
prakticky veškerá pravděpodobnost se bude soustředit kolem $\sqrt{n-1}$
Ostatní oblasti dají při integraci téměř nulový příspěvek!

- ▶ **Poučení:** Důležitá je pravděpodobnost, hustota pravděpodobnosti sama o sobě nám nic neříká.



Hustoty pravděpodobnosti – varovný příklad na začátek

- ▶ Mějme *standardní* gaussovskou hustotu pravděpodobnosti ve dvou dimenzích:

$$f(x, y|0,1) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}(x^2 + y^2)\right)$$

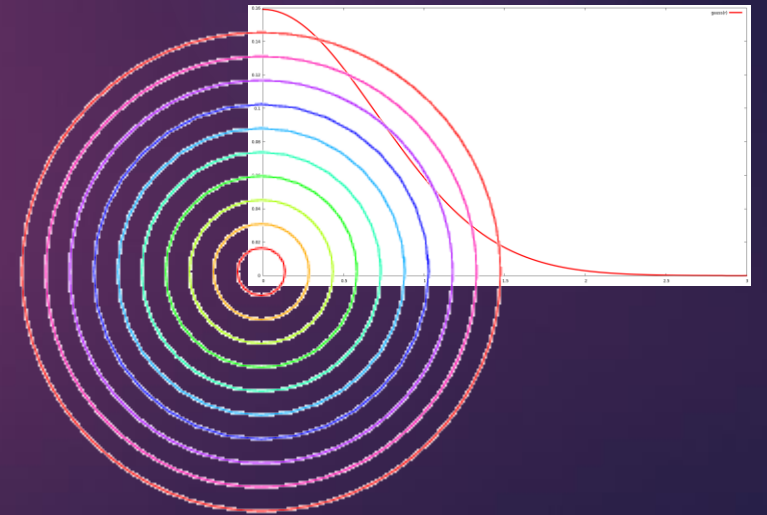
- ▶ Nakresleme tuto funkci v závislosti na radiální souřadnici, $r = \sqrt{x^2 + y^2}$:

- ▶ Rozdělme rovinu na mezikruží o tloušťce Δr .
- ▶ V jakém mezikruží bude největší pravděpodobnost?
- ▶ $P(r' \in [r, r + \Delta r]) \approx f(r)\Delta S(r) \approx f(r)2\pi r\Delta r$
- ▶ Největší tedy není kolem nuly, ale kolem 1
- ▶ S rostoucí dimenzí n bude tento efekt čím dál tím silnější –

prakticky veškerá pravděpodobnost se bude soustředit kolem $\sqrt{n-1}$

Ostatní oblasti dají při integraci téměř nulový příspěvek!

- ▶ **Poučení:** Důležitá je pravděpodobnost, hustota pravděpodobnosti sama o sobě nám nic neříká.
- ▶ Ale co to vlastně je pravděpodobnost?

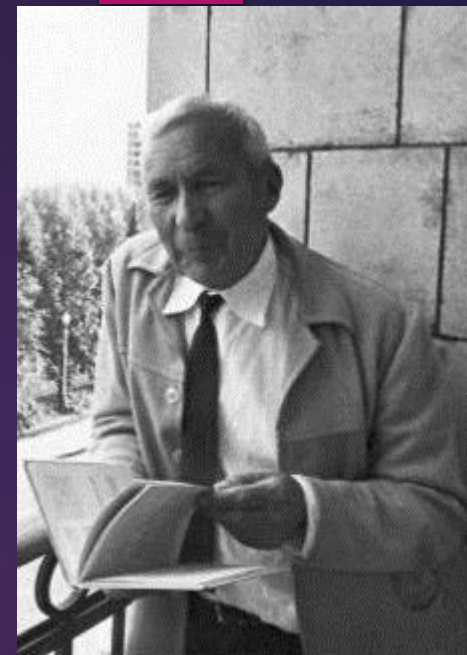


Definice pravděpodobnosti (Kolmogorov, 1933)

- ▶ Nechť X je libovolná neprázdná množina a Σ je σ -algebra definovaná na X . Potom pravděpodobnost P je libovolná míra na Σ , pro kterou platí $P(X)=1$.

Trojice (X, Σ, P) se nazývá pravděpodobnostní prostor.

- ▶ X – Prostor *elementárních jevů* (*sample space*), prvky $x \in X$ nazýváme elementární jevy.
- ▶ Σ - Prostor *náhodných jevů*, prvky $A \in \Sigma$ jsou (*náhodné*) jevy (*events*).
- ▶ P - pravděpodobnost

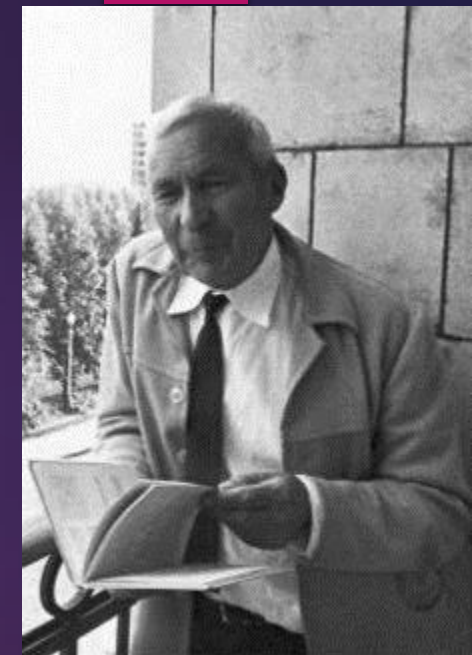


Definice pravděpodobnosti (Kolmogorov, 1933)

- ▶ Nechť X je libovolná neprázdná množina a Σ je σ -algebra definovaná na X . Potom pravděpodobnost P je libovolná míra na Σ , pro kterou platí $P(X)=1$.

Trojice (X, Σ, P) se nazývá pravděpodobnostní prostor.

- ▶ X – Prostor elementárních jevů (*sample space*), prvky $x \in X$ nazýváme elementární jevy.
- ▶ Σ - Prostor náhodných jevů, prvky $A \in \Sigma$ jsou (náhodné) jevy (*events*).
- ▶ P - pravděpodobnost



Připomenutí: σ -algebra

► Definice: Nechť X je libovolná množina. Pak σ -algebra Σ na X je libovolná kolekce podmnožin X , pro kterou platí:

1 Σ obsahuje X : $X \in \Sigma$

2 Uzavřenost pod doplňkem: $A \in \Sigma \Rightarrow X \setminus A \in \Sigma$

3 Uzavřenost pod spočetným sjednocením: Jestliže platí, že $A_k \in \Sigma$ pro $\forall k \in \mathbb{N}$, tak i $\bigcup_{k=1}^{\infty} A_k \in \Sigma$

Připomenutí: σ -algebra

► Definice: Nechť X je libovolná množina. Pak σ -algebra Σ na X je libovolná kolekce podmnožin X , pro kterou platí:

1 Σ obsahuje X : $X \in \Sigma$

2 Uzavřenost pod doplňkem: $A \in \Sigma \Rightarrow X \setminus A \in \Sigma$

3 Uzavřenost pod spočetným sjednocením: Jestliže platí, že $A_k \in \Sigma$ pro $\forall k \in \mathbf{N}$, tak i $\bigcup_{k=1}^{\infty} A_k \in \Sigma$

Z těchto axiomů s pomocí De Morganových zákonů hned dostaneme, že také:

a. Σ musí obsahovat i prázdnou množinu: $\emptyset \in \Sigma$ ($\emptyset = X \setminus X$)

b. platí i uzavřenost pod spočetným průnikem ($\bigcap_{i=1}^{\infty} A_i = X \setminus (\bigcup_{i=1}^{\infty} X \setminus A_i)$)

Připomenutí: σ -algebra

► Definice: Nechť X je libovolná množina. Pak σ -algebra Σ na X je libovolná kolekce podmnožin X , pro kterou platí:

1 Σ obsahuje X : $X \in \Sigma$

2 Uzavřenost pod doplňkem: $A \in \Sigma \Rightarrow X \setminus A \in \Sigma$

3 Uzavřenost pod spočetným sjednocením: Jestliže platí, že $A_k \in \Sigma$ pro $\forall k \in \mathbb{N}$, tak i $\bigcup_{k=1}^{\infty} A_k \in \Sigma$

Z těchto axiomů s pomocí De Morganových zákonů hned dostaneme, že také:

a. Σ musí obsahovat i prázdnou množinu: $\emptyset \in \Sigma$ ($\emptyset = X \setminus X$)

b. platí i uzavřenost pod spočetným průnikem ($\bigcap_{i=1}^{\infty} A_i = X \setminus (\bigcup_{i=1}^{\infty} X \setminus A_i)$)

Dvojice (X, Σ) se nazývá *měřitelný prostor*. Prvky $A \in \Sigma$ se nazývají *měřitelné množiny*.

Příklady σ -algeber

Příklady:

Množina $X = \{\text{koláč}, \text{tofu}, \text{tempeh}\}$:

- ▶ $\Sigma = \{\emptyset, X\} = \{\emptyset, X\}$ (*triviální* σ -algebra)
- ▶ $\Sigma = 2^X = \{\emptyset, \{\text{Koláč}\}, \{\text{tofu}\}, \{\text{tempeh}\}, \{\text{koláč}, \text{tofu}\}, \{\text{tofu}, \text{tempeh}\}, \{\text{koláč}, \text{tempeh}\}, \{\text{koláč}, \text{tofu}, \text{tempeh}\}\}$
(potenční množina X - *diskrétní* σ -algebra)
- ▶ $\Sigma = \{\emptyset, \{\text{koláč}\}, \{\text{tofu}, \text{tempeh}\}, X\}$ – σ -algebra generovaná *{koláčem}*.

Množina $X = \text{Věky lidí v celých letech od 0 do 100}$

- ▶ $\Sigma = \{\emptyset, [0,10], [11,20], \dots [90,100], [0,10] \cup [11,20], \dots, X\}$

Množina $X = \mathbf{N}$:

- ▶ $\Sigma = \{\emptyset, \mathbf{N}\}$
- ▶ $\Sigma = \{\emptyset, \{269\}, \mathbf{N} \setminus \{269\}, \mathbf{N}\}$
- ▶ $\Sigma = \{\emptyset, \text{lichá čísla}, \text{sudá čísla}, \mathbf{N}\}$
- ▶ $\Sigma = 2^{\mathbf{N}}$ (už nespočetná množina, stejná mohutnost jako $\mathbf{R}$).

Definice: Nechť X je libovolná množina. Pak σ -algebra Σ na X je libovolná kolekce podmnožin X , pro kterou platí:

1 Σ obsahuje X : $X \in \Sigma$

2 Uzavřenost pod doplňkem: $A \in \Sigma \Rightarrow X \setminus A \in \Sigma$

3 Uzavřenost pod spočetným sjednocením: Jestliže platí, že $A_k \in \Sigma$ pro $\forall k \in \mathbf{N}$, tak i $\bigcup_{k=1}^{\infty} A_k \in \Sigma$

Připomenutí: σ -algebra

Příklady:

Množina $X = \mathbf{R}^n$

- ▶ $\Sigma = \{\emptyset, \mathbf{R}^n\}$
- ▶ $\Sigma = 2^{\mathbf{R}^n}$ množina všech podmnožin $\mathbf{R}^n$
- ▶ $\Sigma = \{\emptyset, \{\pi\}, [0,1], \pi \cup [0,1], \mathbf{R} \setminus \{\pi\}, \mathbf{R} \setminus [0,1], \mathbf{R} \setminus (\{\pi\} \cup [0,1]), \mathbf{R}\}$
- ▶ Σ = Borelova algebra - nejmenší σ -algebra, která obsahuje všechny otevřené množiny.
- ▶ Σ = Lebesgueova σ -algebra - σ -algebra všech lebesgueovsky měřitelných podmnožin $\mathbf{R}^n$ (pomocí ní definujeme Lebesgueův integrál). Axiom výběru $\Leftrightarrow$ ne všechny podmnožiny $\mathbf{R}^n$ jsou lebesgueovsky měřitelné !
- ▶ X může být libovolná podmnožina $\mathbf{R}^n$, libovolná varieta, atd.

Definice: Nechť X je libovolná množina. Pak σ -algebra Σ na X je libovolná kolekce podmnožin X , pro kterou platí:

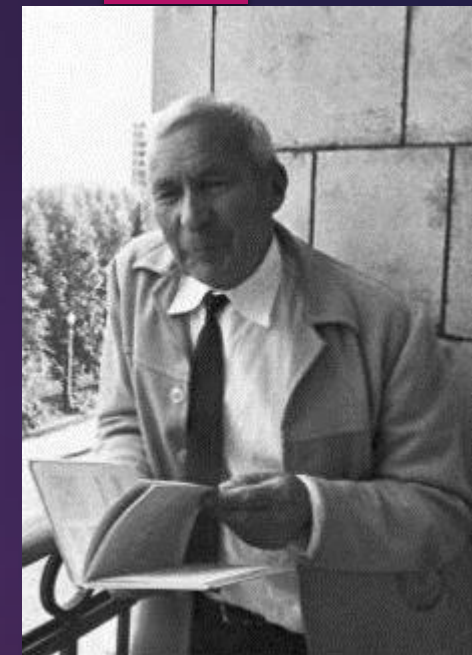
- 1 Σ obsahuje X : $X \in \Sigma$
- 2 Uzavřenost pod doplňkem: $A \in \Sigma \Rightarrow X \setminus A \in \Sigma$
- 3 Uzavřenost pod spočetným sjednocením: Jestliže platí, že $A_k \in \Sigma$ pro $\forall k \in \mathbf{N}$, tak i $\bigcup_{k=1}^{\infty} A_k \in \Sigma$

Definice pravděpodobnosti (Kolmogorov, 1933)

- ▶ Nechť X je libovolná neprázdná množina a Σ je σ -algebra definovaná na X . Potom pravděpodobnost P je libovolná **míra** na Σ , pro kterou platí $P(X)=1$.

Trojice (X, Σ, P) se nazývá pravděpodobnostní *prostor*.

- ▶ X – Prostor *elementárních jevů* (*sample space*), prvky $x \in X$ nazýváme elementární jevy.
- ▶ Σ - Prostor *náhodných jevů*, prvky $A \in \Sigma$ jsou (*náhodné*) jevy (*events*).
- ▶ P - pravděpodobnost



Připomenutí: Míra

- Definice: Nechť (X, Σ) je měřitelný prostor. Míra μ je nezáporná σ -aditivní funkce na prvcích σ -algebry, $\mu: \Sigma \rightarrow [0, \infty) \cup \{\infty\}$:

1 $\mu(\emptyset) = 0$

2 Nechť $A_1, A_2, A_3 \dots \in \Sigma$ jsou po dvou disjunktní: $A_i \cap A_j = \emptyset \forall i \neq j \in \mathbb{N}, A_k \in \Sigma$. Potom $\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$

Pravděpodobnostní míra (pravděpodobnost) je míra, pro kterou navíc

3 $\mu(X) = 1$

Připomenutí: Míra

- Definice: Nechť (X, Σ) je měřitelný prostor. Míra μ je nezáporná σ -aditivní funkce na prvcích σ -algebry, $\mu: \Sigma \rightarrow [0, \infty) \cup \{\infty\}$:

1 $\mu(\emptyset) = 0$

2 Nechť $A_1, A_2, A_3 \dots \in \Sigma$ jsou po dvou disjunktní: $A_i \cap A_j = \emptyset \forall i \neq j \in \mathbb{N}, A_k \in \Sigma$. Potom $\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$

Pravděpodobnostní míra (pravděpodobnost) je míra, pro kterou navíc

3 $\mu(X) = 1$

- Z 2 plyne:

$$\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B) \quad \forall A, B \in \Sigma$$

- Z 2 a nezápornosti plyne:

$$B \subset A \Rightarrow \mu(B) \leq \mu(A) \quad \forall A, B \in \Sigma$$

- Pro pravděpodobnosti z 2 a 3 plyne:

$$\mu(X/A) = 1 - \mu(A) \quad \forall A \in \Sigma$$

$$\mu(A) \leq 1 \quad \forall A \in \Sigma$$

Připomenutí: Míra

- Definice: Nechť (X, Σ) je měřitelný prostor. Míra μ je nezáporná σ -aditivní funkce na prvcích σ -algebry, $\mu: \Sigma \rightarrow [0, \infty) \cup \{\infty\}$:

1 $\mu(\emptyset) = 0$

2 Nechť $A_1, A_2, A_3 \dots \in \Sigma$ jsou po dvou disjunktní: $A_i \cap A_j = \emptyset \forall i \neq j \in \mathbb{N}, A_k \in \Sigma$. Potom $\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$

Pravděpodobnostní míra (pravděpodobnost) je míra, pro kterou navíc

3 $\mu(X) = 1$

- Z 2 plyne:

$$\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B) \quad \forall A, B \in \Sigma$$

- Z 2 a nezápornosti plyne:

$$B \subset A \Rightarrow \mu(B) \leq \mu(A) \quad \forall A, B \in \Sigma$$

- Pro pravděpodobnosti z 2 a 3 plyne:

$$\mu(X/A) = 1 - \mu(A) \quad \forall A \in \Sigma \quad (\Rightarrow \mu(\emptyset) = 0)$$

$$\mu(A) \leq 1 \quad \forall A \in \Sigma$$

Míra – příklady

► $X = \{\text{koláč}, \text{tofu}, \text{tempeh}\}$, $\Sigma = \{\emptyset, \{\text{koláč}\}, \{\text{tofu}, \text{tempeh}\}, X\}$.

$\mu(X) = 1$, $\mu(\emptyset) = 0$, $\mu(\{\text{koláč}\}) = 1/4$, $\mu(\{\text{tofu}, \text{tempeh}\}) = 3/4$
(Nedokážeme říct, jestli si dáme tofu nebo tempeh)

► $X = \mathbf{N}$, $\Sigma = \{\emptyset, \text{sudá}, \text{lichá}, \mathbf{N}\}$

$\mu(\mathbf{N}) = 1$, $\mu(\emptyset) = 0$, $\mu(\text{sudá}) = 1/2$, $\mu(\text{lichá}) = 1/2$

► $X = \mathbf{N}$, $\Sigma = 2^{\mathbf{N}}$

$\mu(\mathbf{N}) = 1$, $\mu(\emptyset) = 0$, $\mu(\{n\}) = \frac{6}{\pi^2} \frac{1}{n^2}$

Definice: Nechť (X, Σ) je měřitelný prostor. Míra μ je nezáporná σ -aditivní funkce na prvcích σ -algebry, $\mu: \Sigma \rightarrow [0, \infty) \cup \{\infty\}$:

- 1 $\mu(\emptyset) = 0$
- 2 Nechť $A_1, A_2, A_3 \dots \in \Sigma$ jsou po dvou disjunktní: $A_i \cap A_j = \emptyset \forall i \neq j \in \mathbf{N}$, $A_k \in \Sigma$. Potom $\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$

Pravděpodobnostní míra
(pravděpodobnost) je míra, pro kterou navíc

- 3 $\mu(X) = 1$

Míra – příklady

- ▶ $X = \{\text{koláč}, \text{tofu}, \text{tempeh}\}$, $\Sigma = \{\emptyset, \{\text{koláč}\}, \{\text{tofu}, \text{tempeh}\}, X\}$.

$$\mu(X) = 1, \mu(\emptyset) = 0, \mu(\{\text{koláč}\}) = 1/4, \mu(\{\text{tofu}, \text{tempeh}\}) = 3/4$$

(Nedokážeme říct, jestli si dáme tofu nebo tempeh)

- ▶ $X = \mathbf{N}$, $\Sigma = \{\emptyset, \text{sudá}, \text{lichá}, \mathbf{N}\}$

$$\mu(\mathbf{N}) = 1, \mu(\emptyset) = 0, \mu(\text{sudá}) = 1/2, \mu(\text{lichá}) = 1/2$$

- ▶ $X = \mathbf{N}$, $\Sigma = 2^{\mathbf{N}}$

$$\mu(\mathbf{N}) = 1, \mu(\emptyset) = 0, \mu(\{n\}) = \frac{6}{\pi^2} \frac{1}{n^2}$$

(rovnoměrná pravděpodobnost na $2^{\mathbf{N}}$ definovat nejde – ale na některých σ -algebrách ano)

Definice: Nechť (X, Σ) je měřitelný prostor. Míra μ je nezáporná σ -aditivní funkce na prvcích σ -algebry, $\mu: \Sigma \rightarrow [0, \infty) \cup \{\infty\}$:

- 1 $\mu(\emptyset) = 0$
- 2 Nechť $A_1, A_2, A_3 \dots \in \Sigma$ jsou po dvou disjunktní: $A_i \cap A_j = \emptyset \forall i \neq j \in \mathbf{N}$. Potom $\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$

Pravděpodobnostní míra (pravděpodobnost) je míra, pro kterou navíc

- 3 $\mu(X) = 1$

Míra – příklady

► $X = \mathbf{R}$, $\Sigma = \{\emptyset, \{\pi\}, [0,1], \pi \cup [0,1], \mathbf{R} \setminus \{\pi\}, \mathbf{R} \setminus [0,1], \mathbf{R} \setminus (\{\pi\} \cup [0,1]), \mathbf{R}\}$

$$\mu(\{\pi\}) = \frac{1}{2}, \quad \mu([0,1]) = 1/4, \quad \mu(\mathbf{R} \setminus (\{\pi\} \cup [0,1])) = 1/4$$

► $X = \mathbf{R}$, $\Sigma =$ Lebegueova σ -algebra

► $\mu = \delta_{x_0}(A) \dots$ Diracova míra: $\delta_{x_0}(A) = \begin{cases} 1, & x_0 \in A \\ 0, & x_0 \notin A. \end{cases}$

► $\mu \dots$ Lebegueova míra („ $\mu = dx$ “)

$$\text{► } \nu(B) = \frac{\int_B f(x) \mu(dx)}{\int_X f(x) \mu(dx)}$$

Pro malá okolí x , dx : $\nu(dx) = f \mu(dx) \Leftrightarrow d\nu = f d\mu$

tj. integrováním f vzhledem k míře μ dostaneme novou míru ν

Míra – příklady

- ▶ $X = \mathbf{R}$, $\Sigma = \{\emptyset, \{\pi\}, [0,1], \pi \cup [0,1], \mathbf{R} \setminus \{\pi\}, \mathbf{R} \setminus [0,1], \mathbf{R} \setminus (\{\pi\} \cup [0,1]), \mathbf{R}\}$

$$\mu(\{\pi\}) = \frac{1}{2}, \quad \mu([0,1]) = 1/4, \quad \mu(\mathbf{R} \setminus (\{\pi\} \cup [0,1])) = 1/4$$

- ▶ $X = \mathbf{R}$, $\Sigma =$ Lebegueova σ -algebra

- ▶ $\mu = \delta_{x_0}(A) \dots$ Diracova míra: $\delta_{x_0}(A) = \begin{cases} 1, & x_0 \in A \\ 0, & x_0 \notin A. \end{cases}$

- ▶ $\mu \dots$ Lebegueova míra („ $\mu = dx$ “)

- ▶ $\nu(B) = \frac{\int_B f(x) \mu(dx)}{\int_X f(x) \mu(dx)}$

Pro malá okolí x , dx : $\nu(dx) = f \mu(dx) \Leftrightarrow d\nu = f d\mu$

tj. integrováním f vzhledem k míře μ dostaneme novou míru ν

- ▶ Platí to i naopak?

Radonova-Nikodymova věta, hustota pravděpodobnosti

Věta: Nechť (X, Σ, μ) je prostor s mírou a nechť je na měřitelném prostoru (X, Σ) definovaná ještě další míra, ν . Nechť jsou obě míry konečné, tj. $\mu(X) < \infty, \nu(X) < \infty$. **Pak platí:**

Míra ν je absolutně spojitá vzhledem k dominantní míře μ (tj. $\mu(A) = 0 \Rightarrow \nu(A) = 0 \forall A \in \Sigma$) právě tehdy, když existuje nezáporná měřitelná funkce $f: X \rightarrow \mathbb{R}$ s vlastností

$$\nu(A) = \int_A f \mu(dx) \quad \forall A \in \Sigma.$$

Funkce $f =: \frac{d\nu}{d\mu}$ se nazývá Radonova-Nikodymova derivace nebo taky *hustota (pravděpodobnosti) vzhledem k μ (PDF)*. Je určena μ -skoro všude jednoznačně (tj. $\exists N \in \Sigma: \mu(N) = 0$, taková, že f je jednoznačná na X/N).

Míra, která všem prvkům $A \in \Sigma$ přiřazuje jejich normalizovaný objem, $\frac{\nu(A)}{\nu(X)}$, se nazývá *homogenní pravděpodobnost*.

Její hustota pravděpodobnosti vzhledem k μ se nazývá *homogenní hustota pravděpodobnosti (vzhledem k μ)*.

Pravděpodobnost vs PDF

- Je třeba brát v úvahu celou pravděpodobnostní míru, ne jen PDF (ta je definována vždy relativně k nějaké míře!)

Pravděpodobnost vs PDF

- Je třeba brát v úvahu celou pravděpodobnostní míru, ne jen PDF (ta je definována vždy relativně k nějaké míře!)

Příklad 1: jednotková sféra

- Uvažujeme úhlové souřadnice θ, ϕ
- Homogenní pravděpodobnost je $H(d\Omega) = \frac{d\Omega}{4\pi}$
- Homogenní hustota pravděpodobnosti vzhledem k míře $d\Omega$: $h_{d\Omega}(\theta, \phi) = \frac{1}{4\pi}$
- Jak vypadá homogenní hustota pravděpodobnosti $h_{d\theta d\phi}(\theta, \phi)$ vzhledem k míře $d\theta d\phi$?

$$\begin{aligned} H(\theta' \in [\theta, \theta + d\theta], \phi' \in [\phi, \phi + d\phi]) &= \frac{1}{4\pi} \int_{\theta}^{\theta+d\theta} \int_{\phi}^{\phi+d\phi} 1 d\Omega = \\ &= \frac{1}{4\pi} \int_{\theta}^{\theta+d\theta} \int_{\phi}^{\phi+d\phi} \sin\theta d\theta d\phi = \frac{1}{4\pi} \sin\theta d\theta d\phi = h_{d\theta d\phi}(\theta, \phi) d\theta d\phi \\ &\Rightarrow h_{d\theta d\phi}(\theta, \phi) = \frac{1}{4\pi} \sin(\theta) \end{aligned}$$

$\Rightarrow$ Pravděpodobnost je stejná pro všechna $d\Omega$, ale $h_{d\theta d\phi}$ na rovníku je větší než na pólech.

Pravděpodobnost vs PDF

Příklad 2 (změna souřadnic):

- Mějme na intervalu $(0,1]$ pravděpodobnost $P([x_1, x_2]) = \int_{x_1}^{x_2} \mu(dx)$, kde $\mu(dx)$ je (třeba) klasická Lebesgueova míra, „ $\mu = dx$ “.

Pravděpodobnost vs PDF

Příklad 2 (změna souřadnic):

- Mějme na intervalu $[0,1]$ pravděpodobnost $P([x_1, x_2]) = \int_{x_1}^{x_2} \mu(dx)$, kde $\mu(dx)$ je (třeba) klasická Lebesgueova míra, „ $\mu = dx$ “.
- Tj. $P(dx) = \mu(dx)$
- Hustota pravděpodobnosti vůči $\mu(dx)$ je $f(x) \equiv 1$.

Pravděpodobnost vs PDF

Příklad 2 (změna souřadnic):

- Mějme na intervalu $(0,1]$ pravděpodobnost $P([x_1, x_2]) = \int_{x_1}^{x_2} \mu(dx)$, kde $\mu(dx)$ je (třeba) klasická Lebesgueova míra, „ $\mu = dx$ “.
- Tj. $P(dx) = \mu(dx)$
- Hustota pravděpodobnosti vůči $\mu(dx)$ je $f(x) \equiv 1$.
- Mějme změnu souřadnic $x \rightarrow y(x) = \ln(x)$. Tato transformace indukuje novou míru ν : $\nu(A) = \mu(y(A))$ („ $\nu = dy$ “). V mat. žargónu je to push-forward μ přes y^{-1} : $\nu(A) = y_*^{-1}\mu(A) := \mu(y(A))$.
- Například $\nu([e^{-1}, 1]) = \mu([-1, 0]) = 1$, ale také $\nu([e^{-2}, e^{-1}]) = \mu([-2, -1]) = 1$ (tj. ν vlastně měří intervaly v log škále).

Pravděpodobnost vs PDF

Příklad 2 (změna souřadnic):

- Mějme na intervalu $(0,1]$ pravděpodobnost $P([x_1, x_2]) = \int_{x_1}^{x_2} \mu(dx)$, kde $\mu(dx)$ je (třeba) klasická Lebesgueova míra, „ $\mu = dx$ “.
- Tj. $P(dx) = \mu(dx)$
- Hustota pravděpodobnosti vůči $\mu(dx)$ je $f(x) \equiv 1$.
- Mějme změnu souřadnic $x \rightarrow y(x) = \ln(x)$. Tato transformace indukuje novou míru ν : $\nu(A) = \mu(y(A))$ („ $\nu = dy$ “). V mat. žargónu je to *push-forward* μ přes y^{-1} : $\nu(A) = y_*^{-1}\mu(A) := \mu(y(A))$.
- Například $\nu([e^{-1}, 1]) = \mu([-1, 0]) = 1$, ale také $\nu([e^{-2}, e^{-1}]) = \mu([-2, -1]) = 1$ (tj. ν vlastně měří intervaly v log škále).
- Pro malá okolí dy platí „ $dy = dy/dx dx$ “: $\nu(dx) = \mu(y(dx)) = \left| \frac{dy}{dx} \right| \mu(dx) = \frac{1}{\ln(x)} \mu(dx) = \frac{1}{y(x)} \mu(dx)$
- $P(dx)$ se tedy pomocí této indukované míry dá vyjádřit jako
$$P(dx) = f_\mu(x) \mu(dx) = 1 \cdot \mu(dx) = \left| \frac{dx}{dy} \right| \nu(dx) = y(x) \nu(dx) = f_\nu(y(x)) \nu(dx) = P(dx)$$

Pravděpodobnost vs PDF

Příklad 2 (změna souřadnic):

- Mějme na intervalu $(0,1]$ pravděpodobnost $P([x_1, x_2]) = \int_{x_1}^{x_2} \mu(dx)$, kde $\mu(dx)$ je (třeba) klasická Lebesgueova míra, „ $\mu = dx$ “.
 - Tj. $P(dx) = \mu(dx)$
 - Hustota pravděpodobnosti vůči $\mu(dx)$ je $f(x) \equiv 1$.
 - Mějme změnu souřadnic $x \rightarrow y(x) = \ln(x)$. Tato transformace indukuje novou míru ν : $\nu(A) = \mu(y(A))$ („ $\nu = dy$ “). V mat. žargónu je to *push-forward* μ přes y^{-1} : $\nu(A) = y_*^{-1}\mu(A) := \mu(y(A))$.
 - Například $\nu([e^{-1}, 1]) = \mu([-1, 0]) = 1$, ale také $\nu([e^{-2}, e^{-1}]) = \mu([-2, -1]) = 1$ (tj. ν vlastně měří intervaly v log škále).
 - Pro malá okolí dy platí „ $dy = dy/dx dx$ “: $\nu(dx) = \mu(y(dx)) = \left| \frac{dy}{dx} \right| \mu(dx) = \frac{1}{\ln(x)} \mu(dx) = \frac{1}{y(x)} \mu(dx)$
 - $P(dx)$ se tedy pomocí této indukované míry dá vyjádřit jako
$$P(dx) = f_\mu(x) \mu(dx) = 1 \cdot \mu(dx) = \left| \frac{dx}{dy} \right| \nu(dx) = y(x) \nu(dx) = f_\nu(y(x)) \nu(dx) = P(dx)$$
- $\Rightarrow$ Hustota pravděpodobnosti $P(\cdot)$ vůči míře $\nu(\cdot)$ je tedy $f_\nu(y) = y$
- $f_\nu(y)$ s y roste, protože $\nu(dy)$ s y naopak klesá $\Rightarrow$ PDF je na intervalu $(0, 1/2)$ větší než na intervalu $[1/2, 1]$, oba ale mají stejnou pravděpodobnost.

Pravděpodobnost vs PDF

Příklad 2 (změna souřadnic):

- Mějme na intervalu $(0,1]$ pravděpodobnost $P([x_1, x_2]) = \int_{x_1}^{x_2} \mu(dx)$, kde $\mu(dx)$ je (třeba) klasická Lebesgueova míra, „ $\mu = dx$ “.
 - Tj. $P(dx) = \mu(dx)$
 - Hustota pravděpodobnosti vůči $\mu(dx)$ je $f(x) \equiv 1$.
 - Mějme změnu souřadnic $x \rightarrow y(x) = \ln(x)$. Tato transformace indukuje novou míru ν : $\nu(A) = \mu(y(A))$ („ $\nu = dy$ “). V mat. žargónu je to *push-forward* μ přes y^{-1} : $\nu(A) = y_*^{-1}\mu(A) := \mu(y(A))$.
 - Například $\nu([e^{-1}, 1]) = \mu([-1, 0]) = 1$, ale také $\nu([e^{-2}, e^{-1}]) = \mu([-2, -1]) = 1$ (tj. ν vlastně měří intervaly v log škále).
 - Pro malá okolí dy platí „ $dy = dy/dx dx$ “: $\nu(dx) = \mu(y(dx)) = \left| \frac{dy}{dx} \right| \mu(dx) = \frac{1}{\ln(x)} \mu(dx) = \frac{1}{y(x)} \mu(dx)$
 - $P(dx)$ se tedy pomocí této indukované míry dá vyjádřit jako
$$P(dx) = f_\mu(x) \mu(dx) = 1 \cdot \mu(dx) = \left| \frac{dx}{dy} \right| \nu(dx) = \mathbf{y(x)} \nu(dx) = \mathbf{f_\nu(y(x))} \nu(dx) = P(dx)$$
- $\Rightarrow$ Hustota pravděpodobnosti $P(\cdot)$ vůči míře $\nu(\cdot)$ je tedy $f_\nu(y) = y$
- $f_\nu(y)$ s y roste, protože $\nu(dy)$ s y naopak klesá $\Rightarrow$ PDF je na intervalu $(0, 1/2)$ větší než na intervalu $(1/2, 1]$, oba ale mají stejnou pravděpodobnost
 - Byla by chyba jen přepočítat $f_\nu(y) = f_\mu(x(y)) = 1$!

Pravděpodobnost vs PDF

Příklad 2 (změna souřadnic):

- Mějme na intervalu $(0,1]$ pravděpodobnost $P([x_1, x_2]) = \int_{x_1}^{x_2} \mu(dx)$, kde $\mu(dx)$ je (třeba) klasická Lebesgueova míra, „ $\mu = dx$ “.
- Tj. $P(dx) = \mu(dx)$
- Hustota pravděpodobnosti vůči $\mu(dx)$ je $f(x) \equiv 1$.
- Mějme změnu souřadnic $x \rightarrow y(x) = \ln(x)$. Tato transformace indukuje novou míru ν : $\nu(A) = \mu(y(A))$ („ $\nu = dy$ “). V mat. žargónu je to push-forward μ přes y^{-1} : $\nu(A) = y_*^{-1}\mu(A) := \mu(y(A))$.
- Například $\nu([e^{-1}, 1]) = \mu([-1, 0]) = 1$, ale také $\nu([e^{-2}, e^{-1}]) = \mu([-2, -1]) = 1$ (tj. ν vlastně měří intervaly v log škále).
- Pro malá okolí dy platí „ $dy = dy/dx dx$ “: $\nu(dx) = \mu(y(dx)) = \left| \frac{dy}{dx} \right| \mu(dx) = \frac{1}{\ln(x)} \mu(dx) = \frac{1}{y(x)} \mu(dx)$

- $P(dx)$ se tedy pomocí této indukované míry dá vyjádřit jako

$$P(dx) = f_\mu(x)\mu(dx) = 1 \cdot \mu(dx) = \left| \frac{dx}{dy} \right| \nu(dx) = y(x)\nu(dx) = f_\nu(y(x))\nu(dx) = P(dx)$$

$\Rightarrow$ Hustota pravděpodobnosti $P(\cdot)$ vůči míře $\nu(\cdot)$ je tedy $f_\nu(y) = y$

- $f_\nu(y)$ s y roste, protože $\nu(dy)$ s y naopak klesá $\Rightarrow$ PDF je na intervalu $(0, 1/2)$ větší než na intervalu $(1/2, 1]$, oba ale mají stejnou pravděpodobnost
- Byla by chyba jen přepočítat $f_\nu(y) = f_\mu(x(y)) = 1$
- Obecně máme při změně souřadnic:

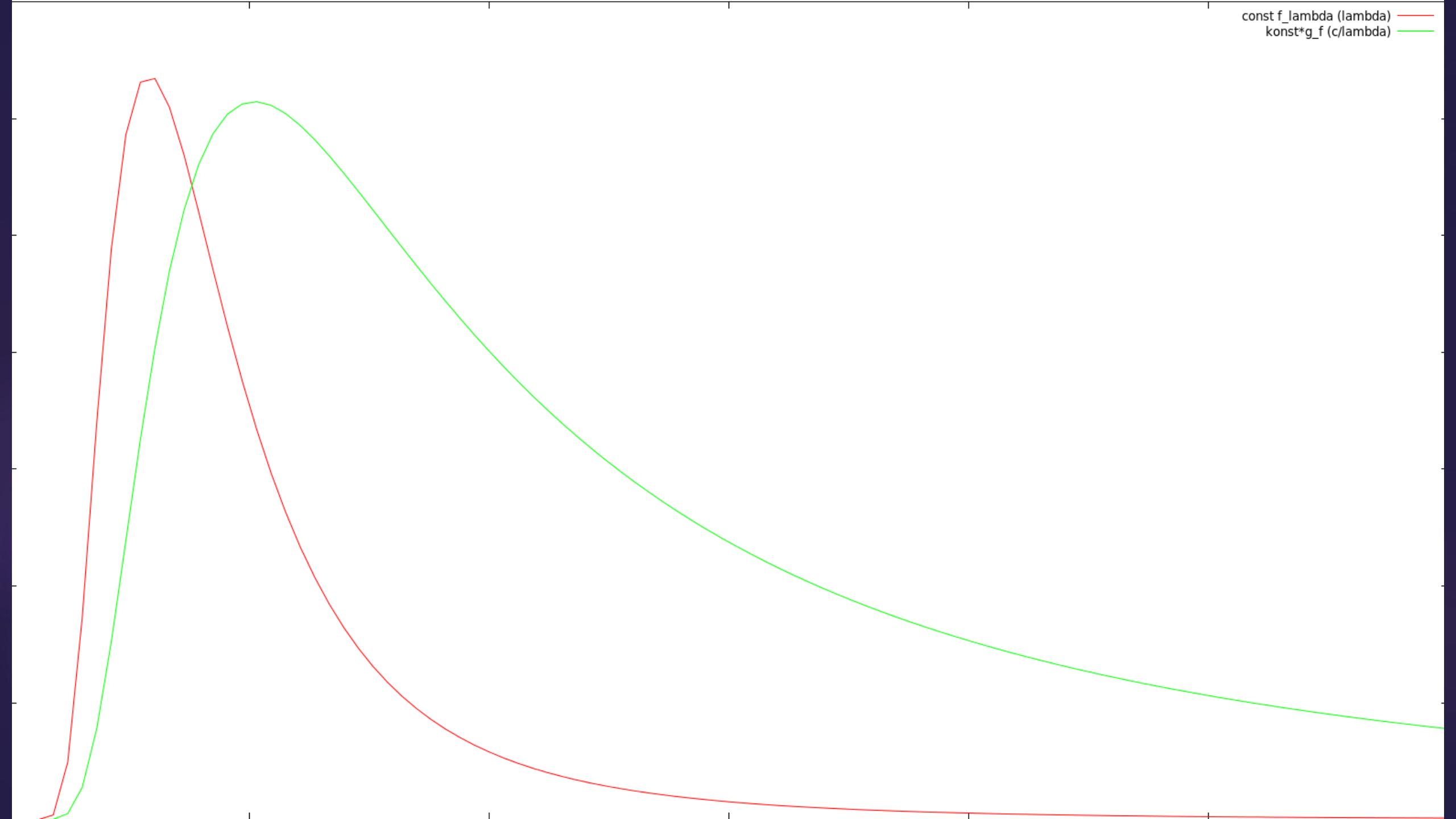
$$f_\nu(y) = \left| \frac{D(x_1, \dots, x_n)}{D(y_1, \dots, y_n)} \right| f_\mu(x(y))$$

Pravděpodobnost vs PDF

► **Příklad 3** „Planckův zákon“:

- $g_\lambda(\lambda, T) = \frac{k_\lambda}{\lambda^5} \frac{1}{e^{hc/\lambda kT} - 1}$, kde k_λ je normalizační konstanta
- Položením derivace g_λ podle λ rovnou nule dostaneme maximum g_λ , jeho závislost na T vyjadřuje „Wienův posunovací zákon“: $\lambda_{max} = \frac{b}{T}$
- Kdybychom udělali transformaci $\lambda \rightarrow f(\lambda) = \frac{c}{f}$; $g_\lambda \rightarrow g_f = g_\lambda \left| \frac{d\lambda}{df} \right| = k_f f^3 \frac{1}{e^{hf/kT} - 1}$, mohli bychom spočítat maximum g_f : $f_{max} = \frac{c}{a} T$.
- $a \neq b \Rightarrow \lambda(f_{max}) = \frac{c}{f_{max}} = \frac{a}{T} = \frac{a}{b} \frac{b}{T} = \frac{a}{b} \lambda_{max} \neq \lambda_{max}$
- Neplatí tedy, že $\lambda_{max} = \frac{c}{f_{max}}$!

const f_lambda (lambda) —
konst*g_f (c/lambda) —



Pravděpodobnost vs PDF

► Příklad 3 „Planckův zákon“:

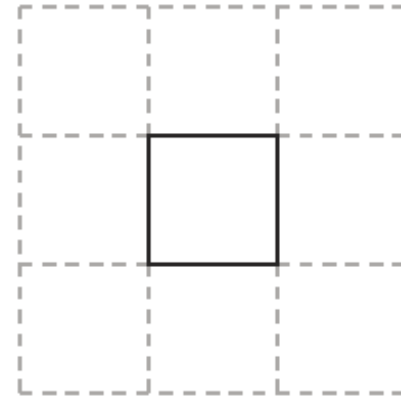
- $g_\lambda(\lambda, T) = \frac{k_\lambda}{\lambda^5} \frac{1}{e^{hc/\lambda kT} - 1}$, kde k_λ je normalizační konstanta
- Položením derivace g_λ podle λ rovnou nule dostaneme maximum g_λ , jeho závislost na T vyjadřuje „Wienův posunovací zákon“: $\lambda_{max} = \frac{b}{T}$
- Kdybychom udělali transformaci $\lambda \rightarrow f(\lambda) = \frac{c}{f}$; $g_\lambda \rightarrow g_f = g_\lambda \left| \frac{d\lambda}{df} \right| = k_f f^3 \frac{1}{e^{hf/kT} - 1}$, mohli bychom spočítat maximum g_f : $f_{max} = \frac{c}{a} T$.
- $a \neq b \Rightarrow \lambda(f_{max}) = \frac{c}{f_{max}} = \frac{a}{T} = \frac{a}{b} \frac{b}{T} = \frac{a}{b} \lambda_{max} \neq \lambda_{max}$
- Neplatí tedy, že $\lambda_{max} = \frac{c}{f_{max}}$!
- Pravděpodobnost je pořád stejná $P(df) = g_f df = g_\lambda d\lambda$.
- Ale g_f je PDF vůči df , zatímco g_λ je PDF vůči $d\lambda = \left| \frac{d\lambda}{df} \right| df = \frac{c}{f^2} df = \frac{\lambda(f)^2}{c} df$, což je velký rozdíl.

Prokletí dimenzí

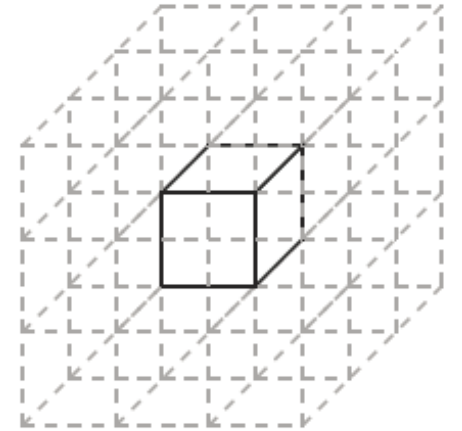
- I když je kolem nějakého bodu vysoká hodnota PDF, s rostoucí dimenzí bude objem okolí maxima relativně čím dál menší, takže pravděpodobnost okolí bude malá.
- A naopak v oblastech s velkým objemem bude typicky malá PDF.
- S rostoucí dimenzí bude tento efekt čím dál silnější.



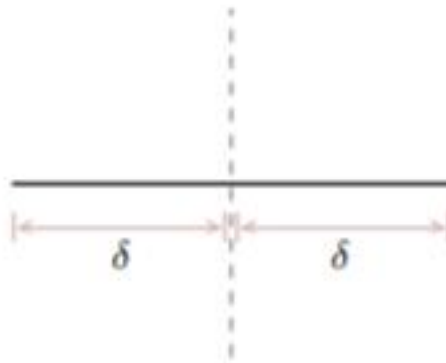
(a)



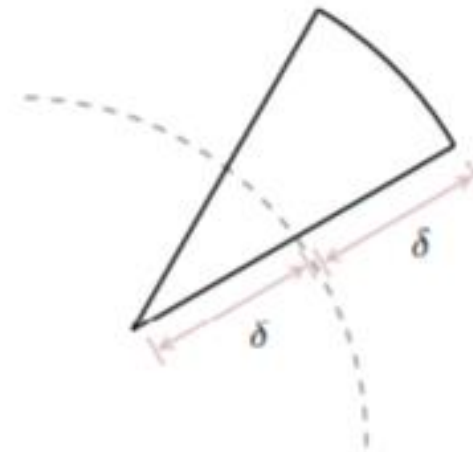
(b)



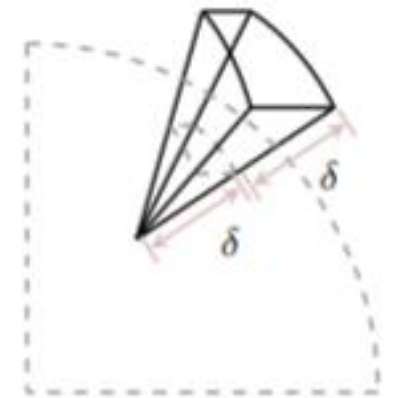
(c)



(a)



(b)



(c)

Typická množina

- Pravděpodobnost bude nezanedbatelná jen někde mezi – *typická množina* (typical set)
- Při výpočtu integrálů dají oblasti mimo typickou množinu zanedbatelný výsledek.

8

BETANCOURT

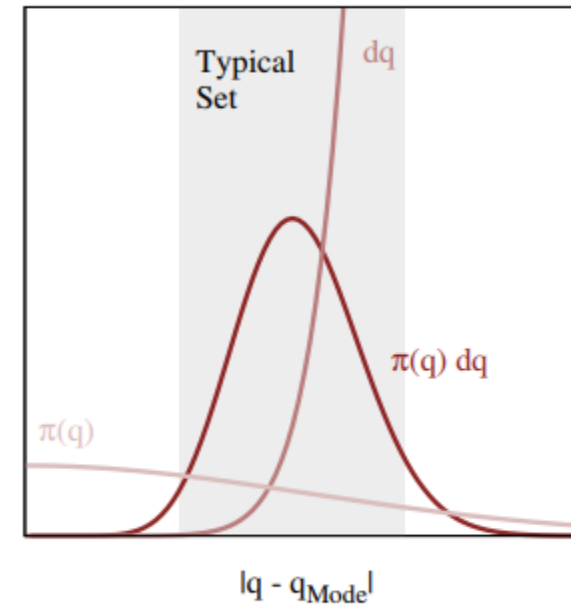
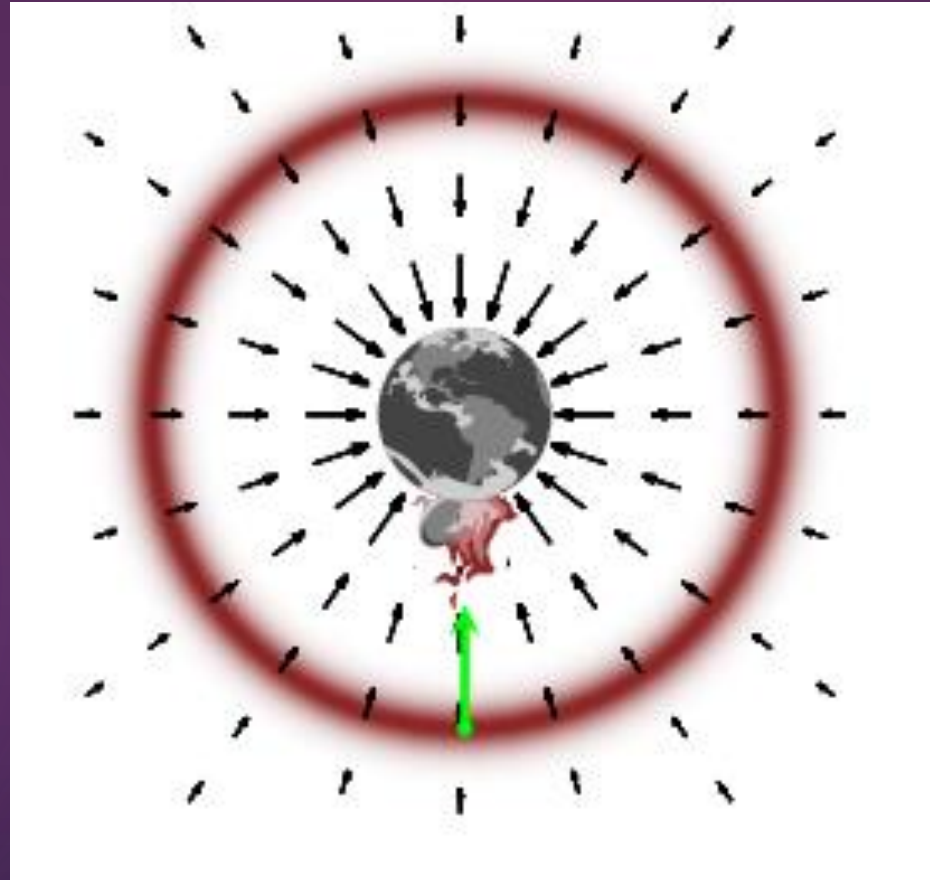


FIG 3. In high dimensions a probability density, $\pi(q)$, will concentrate around its mode, but the volume over which we integrate that density, dq , is much larger away from the mode. Contributions to any expectation are determined by the product of density and volume, $\pi(q) dq$, which then concentrates in a nearly-singular neighborhood called the typical set (grey).



Sledování gradientu PDF $\Rightarrow$ katastrofa

Sledování gradientu PDF $\Rightarrow$ katastrofa



Dostaneme se do oblasti maxima PDF, která je ale malá a má malou pravděpodobnost.

Markov Chain Monte Carlo

- ▶ Mějme pravděpodobnostní prostor $(X, \mathcal{A}, P)$. Chceme provádět nějaké operace s $P(\cdot)$: například vizualizaci, odhalení korelací, zjistit marginální či podmíněné pravděpodobnosti nějakých náhodných veličin, odhadnout střední hodnoty funkcí:

$$E_P[f] = \int_X f(x)P(dx) \quad \text{atd.}$$

- ▶ Příslušné integrály jsou analyticky řešitelné jen pro malou třídu problémů (často nemáme analytický tvar ani pro hustotu pravděpodobnosti P), s rostoucí dimenzí rychle klesá efektivita systematického prohledávání.

Markov Chain Monte Carlo

- ▶ Mějme pravděpodobnostní prostor $(X, \mathcal{A}, P)$. Chceme provádět nějaké operace s $P(\cdot)$: například vizualizaci, odhalení korelací, zjistit marginální či podmíněné pravděpodobnosti nějakých náhodných veličin, odhadnout střední hodnoty funkcí:

$$E_P[f] = \int_X f(x)P(dx) \quad \text{atd.}$$

- ▶ Příslušné integrály jsou analyticky řešitelné jen pro malou třídu problémů (často nemáme analytický tvar ani pro hustotu pravděpodobnosti P), s rostoucí dimenzí rychle klesá efektivita systematického prohledávání.

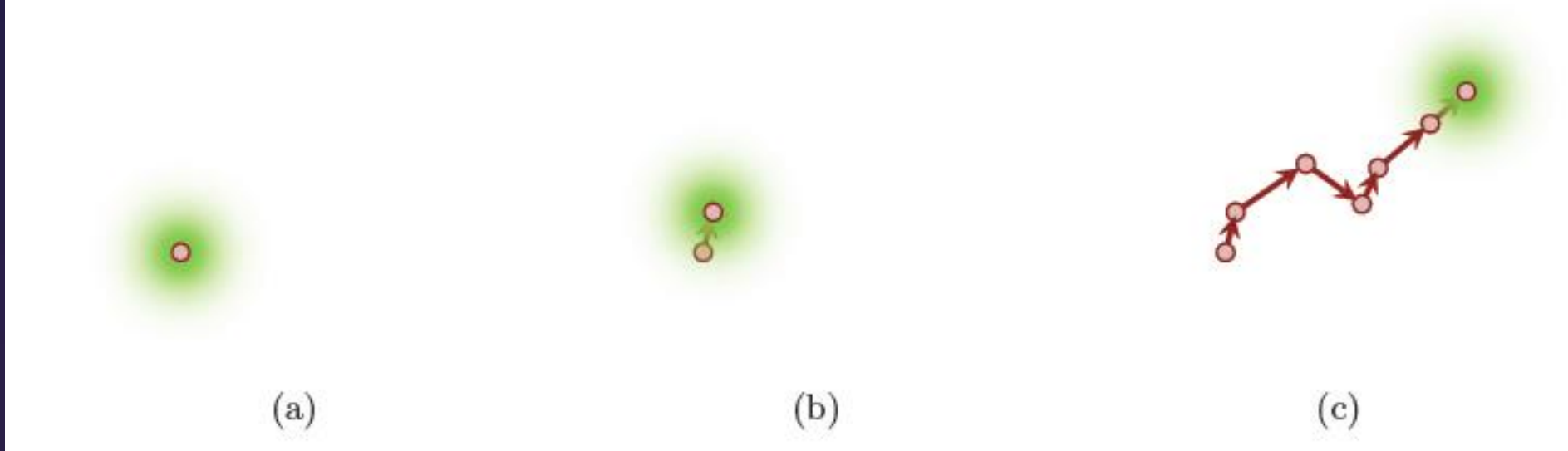
⇒

- ▶ Metoda Monte Carlo: Generování vzorků Z_i z $P(\cdot)$. Když jich budeme mít hodně, můžeme místo PDF kreslit histogramy Z_i , očekávané hodnoty f můžeme odhadnout přes aritmetický průměr:

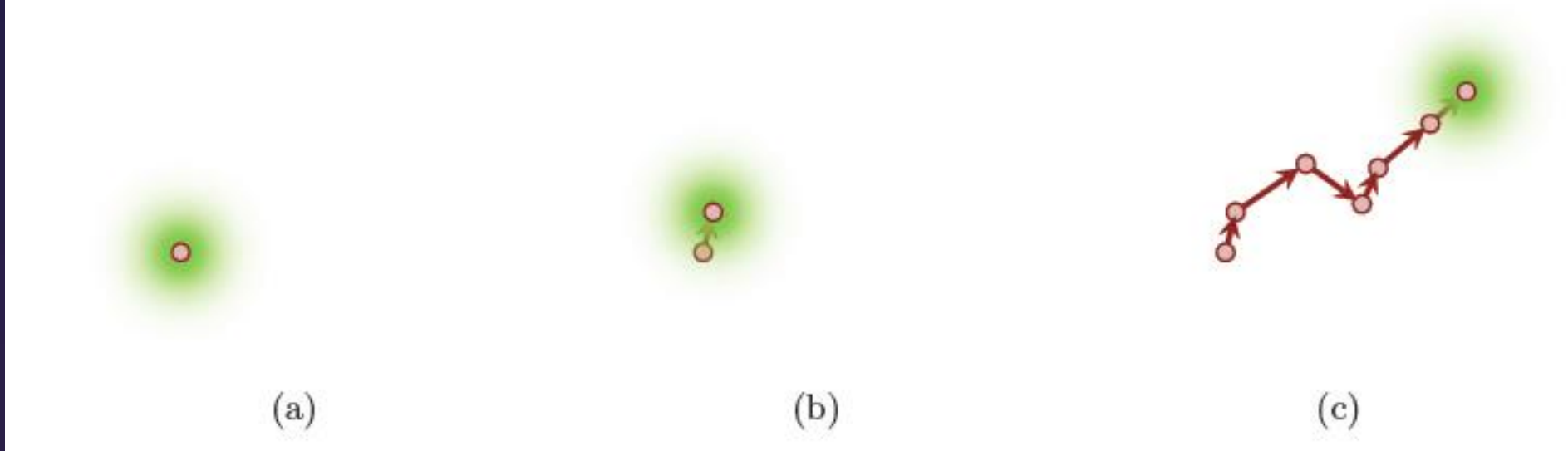
$$E_P[f] = \int_X f(x)P(dx) \rightarrow \hat{E}_P[f] = \frac{1}{N} \sum_{i=1}^N f(Z_i)$$

Markov Chain Monte Carlo

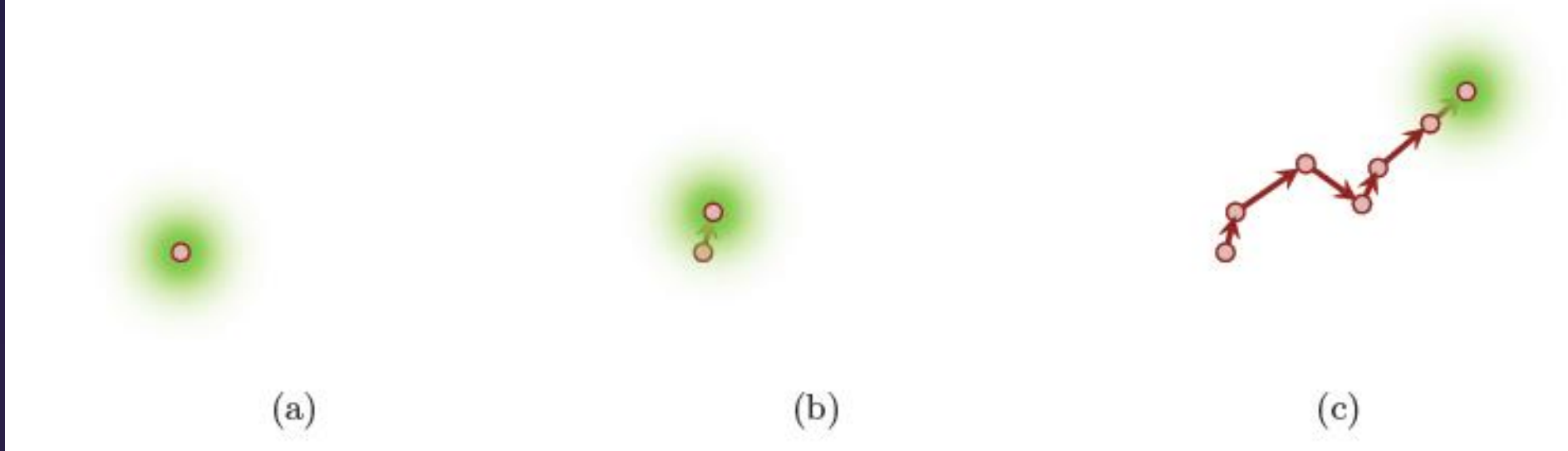
- ▶ Jak ale vzorky najít? Existuje víc metod (viz Tarantola, kap. 2): Inversion method, rejection method, sequential realization... ale fungují v málo dimenzích, nebo je třeba analytické vyjádření, což často nemáme k dispozici $\Rightarrow$ Markov Chain Monte Carlo



- Markov Chain Monte Carlo na pravděpodobnostním prostoru (X, Σ, P) . Cílem je získat vzorky z P .
 1. Začni v arbitrárním bodě $x_0 \in X$. Předpokládejme, že x je vzorkem nějaké pravděpodobnostní míry $P_0(\cdot)$
 2. Definuj Markovovo jádro (kernel) $\tau: X \times \Sigma \rightarrow [0,1], (x, A) \rightarrow \tau((x, A))$, které každému bodu $x \in X$ přiřadí pravděpodobnost, že se přesuneme do $A \in \Sigma$.
 3. Vygeneruj vzorek $x_1 \sim \tau(x_0, \cdot)$.
 4. Jaká je nyní pravděpodobnost, že x je v $A \in \Sigma$? Je to: $P_1(A) = (P_0 T)(A) = \int \tau(x, A) P_0(dx)$. Zobrazení $T: P_i \rightarrow P_{i+1}$ se říká Markovův přechod (Markov transition).
 5. Opakuj kroky 3. a 4. pořád dokola



- Markov Chain Monte Carlo na pravděpodobnostním prostoru (X, Σ, P) . Cílem je získat vzorky z P .
- 1. Začni v arbitrárním bodě $x_0 \in X$. Předpokládejme, že x je vzorkem nějaké pravděpodobnostní míry $P_0(\cdot)$
- 2. Definuj Markovovo jádro (kernel) $\tau: X \times \Sigma \rightarrow [0,1], (x, A) \rightarrow \tau((x, A))$, které každému bodu $x \in X$ přiřadí pravděpodobnost, že se přesuneme do $A \in \Sigma$.
- 3. Vygeneruj vzorek $x_1 \sim \tau(x_0, \cdot)$.
- 4. Jaká je nyní pravděpodobnost, že x je v $A \in \Sigma$? Je to: $P_1(A) = (P_0 T)(A) = \int \tau(x, A) P_0(dx)$. Zobrazení $T: P_i \rightarrow P_{i+1}$ se říká Markovův přechod (Markov transition).
- 5. Opakuj kroky 3. a 4. pořád dokola
- 6. Modli se, že dostaneš vzorky z cílové pravděpodobnosti.



- Markov Chain Monte Carlo na pravděpodobnostním prostoru (X, Σ, P) . Cílem je získat vzorky z P .
- 1. Začni v arbitrárním bodě $x_0 \in X$. Předpokládejme, že x je vzorkem nějaké pravděpodobnostní míry $P_0(\cdot)$
- 2. Definuj Markovovo jádro (kernel) $\tau: X \times \Sigma \rightarrow [0,1], (x, A) \rightarrow \tau((x, A))$, které každému bodu $x \in X$ přiřadí pravděpodobnost, že se přesuneme do $A \in \Sigma$.
- 3. Vygeneruj vzorek $x_1 \sim \tau(x_0, \cdot)$.
- 4. Jaká je nyní pravděpodobnost, že x je v $A \in \Sigma$? Je to: $P_1(A) = (P_0 T)(A) = \int \tau(x, A) P_0(dx)$. Zobrazení $T: P_i \rightarrow P_{i+1}$ se říká Markovův přechod (Markov transition).
- 5. Opakuj kroky 3. a 4. pořád dokola
- 6. *Modli se, že dostaneš vzorky z cílové pravděpodobnosti?*

Markov Chain Monte Carlo

- ▶ *Modli se, že dostaneš vzorky z cílové pravděpodobnosti?*
- ▶ *Naštěstí to není potřeba.*

Fundamentální věta o Markovových řetězcích: Nechť je Markovův řetězec indukovaný jádrem $\tau(x, A)$ definovaný na měřitelném prostoru (X, Σ) se spočetně generovanou σ -algebrou **aperiodický a ireducibilní** a nechť má stacionární pravděpodobnost $\pi(\cdot)$, tj. pravděpodobnost invariantní vůči Markovově přechodu T :

Viz Wiki a protipříklady dále

$$\pi(A) = (\pi T)(A) = \int_X \tau(x, A) \pi(dx)$$

Potom pro π - skoro všechna $x \in X$:

$$\lim_{n \rightarrow \infty} \|T^n(x, \cdot) - \pi(\cdot)\| = 0$$

Markov Chain Monte Carlo

- *Modli se, že dostaneš vzorky z cílové pravděpodobnosti?*
- *Naštěstí to není potřeba.*

Fundamentální věta o Markovových řetězcích: Nechť je Markovův řetězec indukovaný jádrem $\tau(x, A)$ definovaný na měřitelném prostoru (X, Σ) se spočetně generovanou σ -algebrou aperiodický a ireducibilní a nechť má stacionární pravděpodobnost $\pi(\cdot)$, tj. pravděpodobnost invariantní vůči Markovově přechodu T :

$$\pi(A) = (\pi T)(A) = \int_X \tau(x, A) \pi(dx)$$

Potom pro π - skoro všechna $x \in X$:

$$\lim_{n \rightarrow \infty} \|T^n(x, \cdot) - \pi(\cdot)\| = 0$$

$$\begin{aligned} & \|f_1(\cdot) - f_2(\cdot)\| \\ &= \sup_{A \in \Sigma} |f_1(A) - f_2(A)| \end{aligned}$$

Markov Chain Monte Carlo

- *Modli se, že dostaneš vzorky z cílové pravděpodobnosti?*
- *Naštěstí to není potřeba.*

Fundamentální věta o Markovových řetězcích: Nechť je Markovův řetězec indukovaný jádrem $\tau(x, A)$ definovaný na měřitelném prostoru (X, Σ) se spočetně generovanou σ -algebrou aperiodický a ireducibilní a nechť má stacionární pravděpodobnost $\pi(\cdot)$, tj. pravděpodobnost invariantní vůči Markovově přechodu T :

$$\pi(A) = (\pi T)(A) = \int_X \tau(x, A) \pi(dx)$$

Potom pro π - skoro všechna $x \in X$:

$$\lim_{n \rightarrow \infty} \|T^n(x, \cdot) - \pi(\cdot)\| = 0$$

$$\begin{aligned} & \|f_1(\cdot) - f_2(\cdot)\| \\ &= \sup_{A \in \Sigma} |f_1(A) - f_2(A)| \end{aligned}$$

Speciálně $\lim_{n \rightarrow \infty} T^n(x, A) = \pi(A) \forall A \in \Sigma$

Markov Chain Monte Carlo

- *Modli se, že dostaneš vzorky z cílové pravděpodobnosti?*
- Naštěstí to není potřeba.

Fundamentální věta o Markovových řetězcích: Nechť je Markovův řetězec indukovaný jádrem $\tau(x, A)$ definovaný na měřitelném prostoru (X, Σ) se spočetně generovanou σ -algebrou aperiodický a ireducibilní a nechť má stacionární pravděpodobnost $\pi(\cdot)$, tj. pravděpodobnost invariantní vůči Markovově přechodu T :

$$\pi(A) = (\pi T)(A) = \int_X \tau(x, A) \pi(dx)$$

Potom pro π - skoro všechna $x \in X$:

$$\lim_{n \rightarrow \infty} \|T^n(x, \cdot) - \pi(\cdot)\| = 0$$

$$\begin{aligned} & \|f_1(\cdot) - f_2(\cdot)\| \\ &= \sup_{A \in \Sigma} |f_1(A) - f_2(A)| \end{aligned}$$

Speciálně $\lim_{n \rightarrow \infty} T^n(x, A) = \pi(A) \forall A \in \Sigma$

*Navíc platí **silná verze zákona velkých čísel**:*

$$\lim_{n \rightarrow \infty} \sum_{i=1}^{\infty} f(Z_i) = E_{\pi}(f)$$

Příklad: prostor s konečným počtem prvků

► $X = \{\text{koláč}, \text{tofu}, \text{tempeh}\}$, $\Sigma = 2^X$.

$$P_0(\{\text{koláč}\}) = P_0(\{\text{tofu}\}) = P_0(\{\text{tempeh}\}) = \frac{1}{3} \Rightarrow$$

Příklad: prostor s konečným počtem prvků

► $X=\{\text{koláč}, \text{tofu}, \text{tempeh}\}$, $\Sigma = 2^X$.

$$P_0(\{\text{koláč}\}) = P_0(\{\text{tofu}\}) = P_0(\{\text{tempeh}\}) = \frac{1}{3} \Rightarrow \overrightarrow{P_0} = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$$

1/3



1/3



1/3



Příklad: prostor s konečným počtem prvků

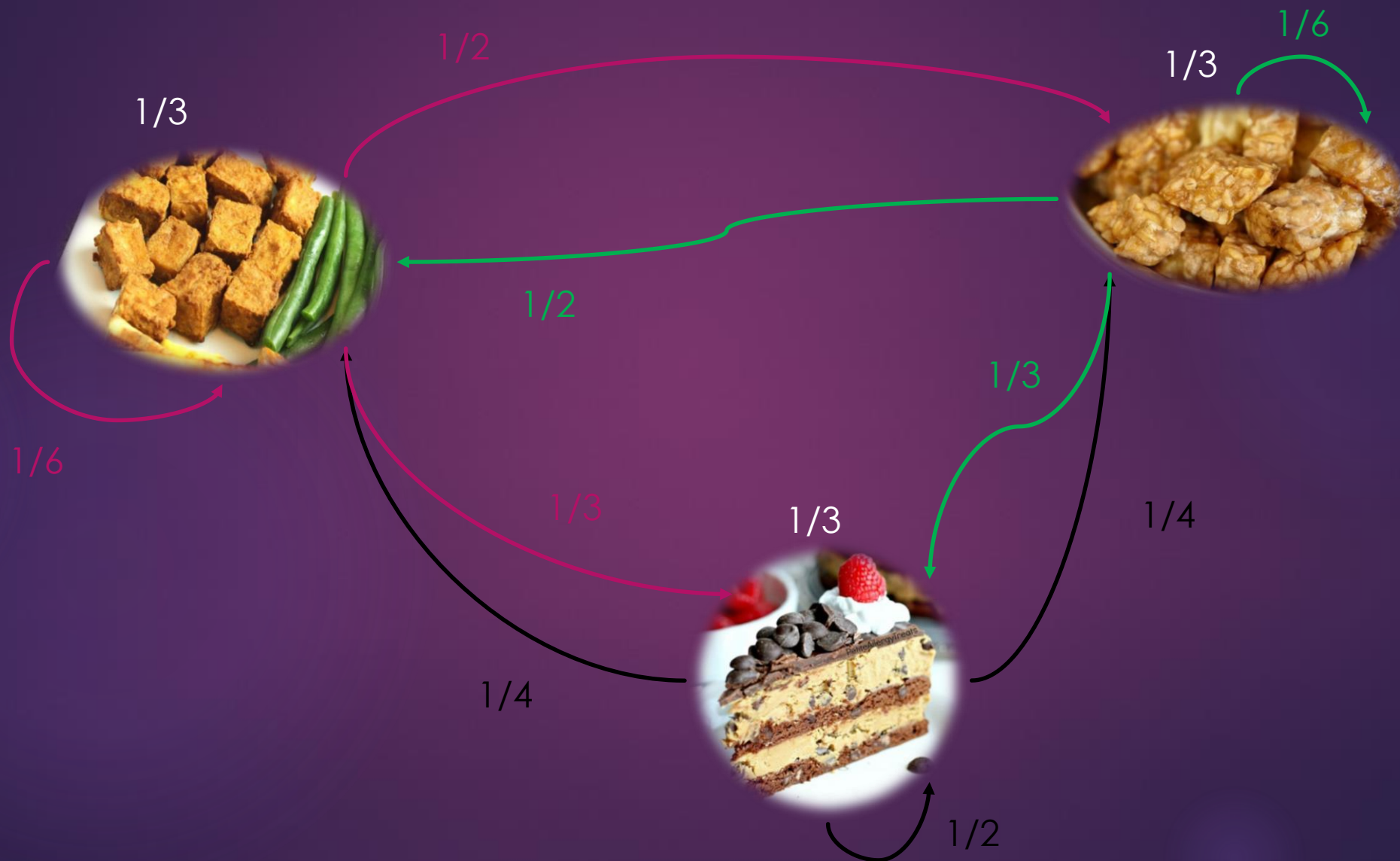
- ▶ $X = \{\text{koláč}, \text{tofu}, \text{tempeh}\}$, $\Sigma = 2^X$.

$$P_0(\{\text{koláč}\}) = P_0(\{\text{tofu}\}) = P_0(\{\text{tempeh}\}) = \frac{1}{3} \Rightarrow \overrightarrow{P_0} = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$$

- ▶ $\tau(\text{koláč}, \{\text{koláč}\}) = \frac{1}{2}$, $\tau(\text{koláč}, \{\text{tofu}\}) = \frac{1}{4}$, $\tau(\text{koláč}, \{\text{tempeh}\}) = \frac{1}{4}$

$$\tau(\text{tofu}, \{\text{koláč}\}) = \frac{1}{3}, \tau(\text{tofu}, \{\text{tofu}\}) = \frac{1}{6}, \tau(\text{tofu}, \{\text{tempeh}\}) = \frac{1}{2}$$

$$\tau(\text{tempeh}, \{\text{koláč}\}) = \frac{1}{3}, \tau(\text{tempeh}, \{\text{tofu}\}) = \frac{1}{2}, \tau(\text{tempeh}, \{\text{tempeh}\}) = \frac{1}{6}$$



Příklad: prostor s konečným počtem prvků

► $X = \{\text{koláč}, \text{tofu}, \text{tempeh}\}$, $\Sigma = 2^X$.

$$P_0(\{\text{koláč}\}) = P_0(\{\text{tofu}\}) = P_0(\{\text{tempeh}\}) = \frac{1}{3} \Rightarrow \overrightarrow{P_0} = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$$

► $\tau(\text{koláč}, \{\text{koláč}\}) = \frac{1}{2}$, $\tau(\text{koláč}, \{\text{tofu}\}) = \frac{1}{4}$, $\tau(\text{koláč}, \{\text{tempeh}\}) = \frac{1}{4}$

$$\tau(\text{tofu}, \{\text{koláč}\}) = \frac{1}{3}, \tau(\text{tofu}, \{\text{tofu}\}) = \frac{1}{6}, \tau(\text{tofu}, \{\text{tempeh}\}) = \frac{1}{6}$$

$$\tau(\text{tempeh}, \{\text{koláč}\}) = \frac{1}{3}, \tau(\text{tempeh}, \{\text{tofu}\}) = \frac{1}{2}, \tau(\text{tempeh}, \{\text{tempeh}\}) = \frac{1}{6}$$

$$\Rightarrow T = \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/3 & 1/6 & 1/2 \\ 1/3 & 1/2 & 1/6 \end{pmatrix}$$

Příklad: prostor s konečným počtem prvků

► $X = \{\text{koláč}, \text{tofu}, \text{tempeh}\}$, $\Sigma = 2^X$.

$$P_0(\{\text{koláč}\}) = P_0(\{\text{tofu}\}) = P_0(\{\text{tempeh}\}) = \frac{1}{3} \Rightarrow \overrightarrow{P_0} = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$$

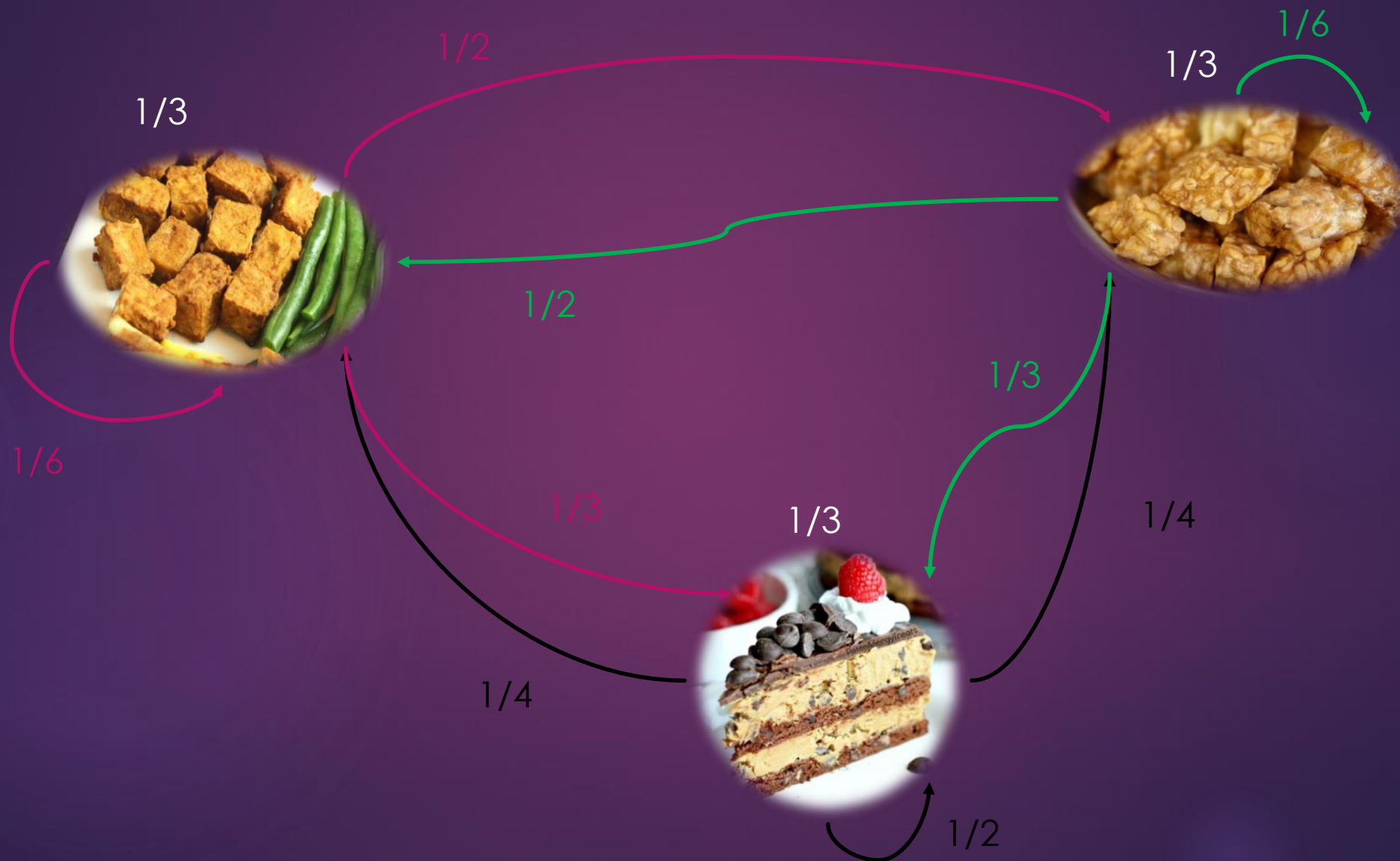
► $\tau(\text{koláč}, \{\text{koláč}\}) = \frac{1}{2}$, $\tau(\text{koláč}, \{\text{tofu}\}) = \frac{1}{4}$, $\tau(\text{koláč}, \{\text{tempeh}\}) = \frac{1}{4}$

$$\tau(\text{tofu}, \{\text{koláč}\}) = \frac{1}{3}$$
$$\tau(\text{tofu}, \{\text{tofu}\}) = \frac{1}{6}$$
$$\tau(\text{tofu}, \{\text{tempeh}\}) = \frac{1}{2}$$

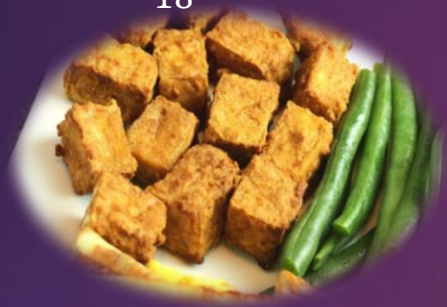
$$\tau(\text{tempeh}, \{\text{koláč}\}) = \frac{1}{3}$$
$$\tau(\text{tempeh}, \{\text{tofu}\}) = \frac{1}{2}$$
$$\tau(\text{tempeh}, \{\text{tempeh}\}) = \frac{1}{6}$$

$$\Rightarrow T = \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/3 & 1/6 & 1/2 \\ 1/3 & 1/2 & 1/6 \end{pmatrix}$$

► $\overrightarrow{P_1} = \overrightarrow{P_0}T = \left(\frac{7}{18}, \frac{11}{36}, \frac{11}{36}\right)$



$$\frac{7}{18}$$



$$\frac{11}{36}$$



$$\frac{11}{36}$$



Příklad: prostor s konečným počtem prvků

- ▶ $X = \{\text{koláč}, \text{tofu}, \text{tempeh}\}$, $\Sigma = 2^X$.

$$P_0(\{\text{koláč}\}) = P_0(\{\text{tofu}\}) = P_0(\{\text{tempeh}\}) = \frac{1}{3} \Rightarrow \overrightarrow{P_0} = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$$

- ▶ $\tau(\text{koláč}, \{\text{koláč}\}) = \frac{1}{2}$, $\tau(\text{koláč}, \{\text{tofu}\}) = \frac{1}{4}$, $\tau(\text{koláč}, \{\text{tempeh}\}) = \frac{1}{4}$

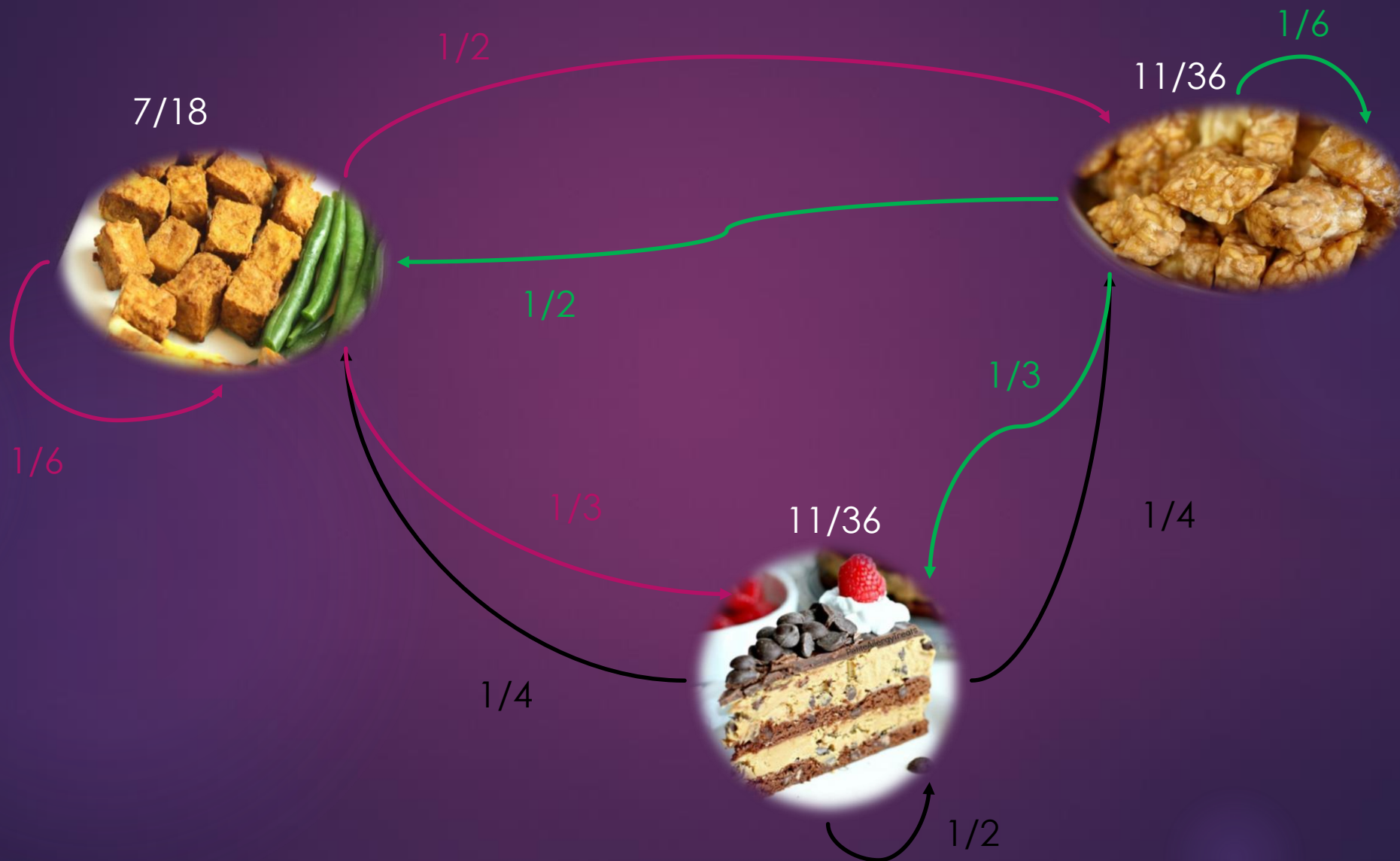
$$\tau(\text{tofu}, \{\text{koláč}\}) = \frac{1}{3}, \tau(\text{tofu}, \{\text{tofu}\}) = \frac{1}{6}, \tau(\text{tofu}, \{\text{tempeh}\}) = \frac{1}{6}$$

$$\tau(\text{tempeh}, \{\text{koláč}\}) = \frac{1}{3}, \tau(\text{tempeh}, \{\text{tofu}\}) = \frac{1}{2}, \tau(\text{tempeh}, \{\text{tempeh}\}) = \frac{1}{6}$$

$$\Rightarrow T = \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/3 & 1/6 & 1/2 \\ 1/3 & 1/2 & 1/6 \end{pmatrix}$$

- ▶ $\overrightarrow{P_1} = \overrightarrow{P_0}T = \left(\frac{7}{18}, \frac{11}{36}, \frac{11}{36}\right)$

- ▶ $\overrightarrow{P_2} = \overrightarrow{P_1}T = \overrightarrow{P_0}T^2$



Příklad: prostor s konečným počtem prvků

► $X = \{\text{koláč}, \text{tofu}, \text{tempeh}\}$, $\Sigma = 2^X$.

$$P_0(\{\text{koláč}\}) = P_0(\{\text{tofu}\}) = P_0(\{\text{tempeh}\}) = \frac{1}{3} \Rightarrow \overrightarrow{P_0} = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$$

$$\text{► } \tau(\text{koláč}, \{\text{koláč}\}) = \frac{1}{2}, \tau(\text{koláč}, \{\text{tofu}\}) = \frac{1}{4}, \tau(\text{koláč}, \{\text{tempeh}\}) = \frac{1}{4}$$

$$\tau(\text{tofu}, \{\text{koláč}\}) = \frac{1}{3}, \tau(\text{tofu}, \{\text{tofu}\}) = \frac{1}{6}, \tau(\text{tofu}, \{\text{tempeh}\}) = \frac{1}{6}$$

$$\tau(\text{tempeh}, \{\text{koláč}\}) = \frac{1}{3}, \tau(\text{tempeh}, \{\text{tofu}\}) = \frac{1}{2}, \tau(\text{tempeh}, \{\text{tempeh}\}) = \frac{1}{6}$$

$$\Rightarrow T = \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/3 & 1/6 & 1/2 \\ 1/3 & 1/2 & 1/6 \end{pmatrix}$$

$$\text{► } \overrightarrow{P_1} = \overrightarrow{P_0}T = \left(\frac{7}{18}, \frac{11}{36}, \frac{11}{36}\right)$$

$$\text{► } \overrightarrow{P_2} = \overrightarrow{P_1}T = \overrightarrow{P_0}T^2$$

$$\text{► } \overrightarrow{P_n} = \overrightarrow{P_0}T^n, \lim_{n \rightarrow \infty} \overrightarrow{P_n} = \overrightarrow{P_0} \lim_{n \rightarrow \infty} T^n =: \overrightarrow{P_0}T^\infty$$

Příklad: prostor s konečným počtem prvků

- ▶ $X = \{\text{koláč}, \text{tofu}, \text{tempeh}\}$, $\Sigma = 2^X$.

$$P_0(\{\text{koláč}\}) = P_0(\{\text{tofu}\}) = P_0(\{\text{tempeh}\}) = \frac{1}{3} \Rightarrow \overrightarrow{P_0} = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$$

- ▶ $\tau(\text{koláč}, \{\text{koláč}\}) = \frac{1}{2}$, $\tau(\text{koláč}, \{\text{tofu}\}) = \frac{1}{4}$, $\tau(\text{koláč}, \{\text{tempeh}\}) = \frac{1}{4}$

$$\tau(\text{tofu}, \{\text{koláč}\}) = \frac{1}{3}, \tau(\text{tofu}, \{\text{tofu}\}) = \frac{1}{6}, \tau(\text{tofu}, \{\text{tempeh}\}) = \frac{1}{6}$$

$$\tau(\text{tempeh}, \{\text{koláč}\}) = \frac{1}{3}, \tau(\text{tempeh}, \{\text{tofu}\}) = \frac{1}{6}, \tau(\text{tempeh}, \{\text{tempeh}\}) = \frac{1}{6}$$

$$\Rightarrow T = \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/3 & 1/6 & 1/2 \\ 1/3 & 1/2 & 1/6 \end{pmatrix}$$

- ▶ $\overrightarrow{P_1} = \overrightarrow{P_0}T = \left(\frac{7}{18}, \frac{11}{36}, \frac{11}{36}\right)$

- ▶ $\overrightarrow{P_2} = \overrightarrow{P_1}T = \overrightarrow{P_0}T^2$

- ▶ $\overrightarrow{P_n} = \overrightarrow{P_0}T^n$, $\lim_{n \rightarrow \infty} \overrightarrow{P_n} = \overrightarrow{P_0} \lim_{n \rightarrow \infty} T^n =: \overrightarrow{P_0}T^\infty$

$$T^\infty = ?$$

Příklad: prostor s konečným počtem prvků

$$\Rightarrow T = \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/3 & 1/6 & 1/2 \\ 1/3 & 1/2 & 1/6 \end{pmatrix}$$

► $T^\infty = ?$

► Matice T má vlastní čísla $(1, -1/3, 1/6)$ $\Rightarrow T^n = V^{-1} \text{diag}(1^n, (\frac{1}{3})^n, (\frac{1}{6})^n) V$, kde V je matice vl. vektorů. V limitě „přežije jen 1 “ $\Rightarrow$

$$T^\infty = \vec{1}\vec{\pi} = \begin{pmatrix} \vec{\pi} \\ \vec{\pi} \\ \vec{\pi} \end{pmatrix} = \begin{pmatrix} 4/10 & 3/10 & 3/10 \\ 4/10 & 3/10 & 3/10 \\ 4/10 & 3/10 & 3/10 \end{pmatrix},$$

kde $\vec{1} = (1 \ 1 \ 1)^T$ a $\vec{\pi} = (\frac{4}{10} \ \frac{3}{10} \ \frac{3}{10})$ jsou levé a pravé vektory příslušné vl. číslu 1.

Příklad: prostor s konečným počtem prvků

$$\Rightarrow T = \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/3 & 1/6 & 1/2 \\ 1/3 & 1/2 & 1/6 \end{pmatrix}$$

► $T^\infty = ?$

► Matice T má vlastní čísla $(1, -1/3, 1/6)$ $\Rightarrow T^n = V^{-1} \text{diag}(1^n, (\frac{1}{3})^n, (\frac{1}{6})^n) V$,
kde V je matice vl. vektorů. V limitě „přežije jen $\mathbf{1}$ “ $\Rightarrow$

$$T^\infty = \vec{1}\vec{\pi} = \begin{pmatrix} \vec{\pi} \\ \vec{\pi} \\ \vec{\pi} \end{pmatrix} = \begin{pmatrix} 4/10 & 3/10 & 3/10 \\ 4/10 & 3/10 & 3/10 \\ 4/10 & 3/10 & 3/10 \end{pmatrix},$$

kde $\vec{1} = (1 \ 1 \ 1)^T$ a $\vec{\pi} = (\frac{4}{10} \ \frac{3}{10} \ \frac{3}{10})$ jsou levé a pravé vektory příslušné vl. číslu 1.

► Po aplikaci na libovolný náhodný vektor $\vec{P}$ dostaneme $\vec{P}T^\infty = (\vec{P} \cdot \vec{1}) \vec{\pi} = 1\vec{\pi} = \vec{\pi}$

Příklad: prostor s konečným počtem prvků

$$\Rightarrow T = \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/3 & 1/6 & 1/2 \\ 1/3 & 1/2 & 1/6 \end{pmatrix}$$

► $T^\infty = ?$

► Matice T má vlastní čísla $(1, -1/3, 1/6) \Rightarrow T^n = V^{-1} \text{diag}(1^n, (\frac{1}{3})^n, (\frac{1}{6})^n) V$,
kde V je matice vl. vektorů. V limitě „přežije jen $\mathbf{1}$ “ $\Rightarrow$

$$T^\infty = \vec{1} \vec{\pi} = \begin{pmatrix} \vec{\pi} \\ \vec{\pi} \\ \vec{\pi} \end{pmatrix} = \begin{pmatrix} 4/10 & 3/10 & 3/10 \\ 4/10 & 3/10 & 3/10 \\ 4/10 & 3/10 & 3/10 \end{pmatrix}, \quad \begin{pmatrix} 2/3 & 1/3 & 0 \end{pmatrix} \begin{pmatrix} \frac{4}{10} & \frac{3}{10} & \frac{3}{10} \\ \frac{4}{10} & \frac{3}{10} & \frac{3}{10} \\ \frac{4}{10} & \frac{3}{10} & \frac{3}{10} \end{pmatrix} = \left(\frac{4}{10}, \frac{3}{10}, \frac{3}{10} \right)$$

kde $\vec{1} = (1 \ 1 \ 1)^T$ a $\vec{\pi} = (\frac{4}{10} \ \frac{3}{10} \ \frac{3}{10})$ jsou levé a pravé vektory příslušné vl. číslu 1.

► Po aplikaci na libovolný náhodný vektor $\vec{P}$ dostaneme $\vec{P} T^\infty = (\vec{P} \cdot \vec{1}) \vec{\pi} = 1 \vec{\pi} = \vec{\pi}$

Příklad: prostor s konečným počtem prvků

$$\Rightarrow T = \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/3 & 1/6 & 1/2 \\ 1/3 & 1/2 & 1/6 \end{pmatrix}$$

- ▶ $\vec{P}_n = \vec{P}_0 T^n, \lim_{n \rightarrow \infty} \vec{P}_n = \vec{P}_0 \lim_{n \rightarrow \infty} T^n =: \vec{P}_0 T^\infty$
- ▶ Matice T má vlastní čísla $(1, -1/3, 1/6) \Rightarrow T^n = V^{-1} \text{diag}(1^n, (\frac{1}{3})^n, (\frac{1}{6})^n) V$, kde V je matice vl. vektorů. V limitě „přežije jen **1**“ $\Rightarrow$

$$T^\infty = \vec{1} \vec{\pi} = \begin{pmatrix} \vec{\pi} \\ \vec{\pi} \\ \vec{\pi} \end{pmatrix} = \begin{pmatrix} 4/10 & 3/10 & 3/10 \\ 4/10 & 3/10 & 3/10 \\ 4/10 & 3/10 & 3/10 \end{pmatrix},$$

kde $\vec{1} = (1 \ 1 \ 1)^T$ a $\vec{\pi} = (\frac{4}{10} \ \frac{3}{10} \ \frac{3}{10})$ jsou levé a pravé vektory příslušné vl. číslu 1.

- ▶ Po aplikaci na libovolný náhodný vektor $\vec{P}$ dostaneme $\vec{P} T^\infty = (\vec{P} \cdot \vec{1}) \vec{\pi} = 1 \vec{\pi} = \vec{\pi}$
- ▶ Co vlastně představuje $\vec{\pi}$?

Příklad: prostor s konečným počtem prvků

$$\Rightarrow T = \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/3 & 1/6 & 1/2 \\ 1/3 & 1/2 & 1/6 \end{pmatrix}$$

- ▶ $\vec{P}_n = \vec{P}_0 T^n, \lim_{n \rightarrow \infty} \vec{P}_n = \vec{P}_0 \lim_{n \rightarrow \infty} T^n =: \vec{P}_0 T^\infty$
- ▶ Matice T má vlastní čísla $(1, -1/3, 1/6)$ $\Rightarrow T^n = V^{-1} \text{diag}(1^n, \left(\frac{1}{3}\right)^n, \left(\frac{1}{6}\right)^n) V$, kde V je matice vl. vektorů. V limitě „přežije jen 1“ $\Rightarrow$

$$T^\infty = \vec{1} \vec{\pi} = \begin{pmatrix} \vec{\pi} \\ \vec{\pi} \\ \vec{\pi} \end{pmatrix} = \begin{pmatrix} 4/10 & 3/10 & 3/10 \\ 4/10 & 3/10 & 3/10 \\ 4/10 & 3/10 & 3/10 \end{pmatrix},$$

kde $\vec{1} = (1 \ 1 \ 1)^T$ a $\vec{\pi} = (\frac{4}{10} \ \frac{3}{10} \ \frac{3}{10})$ jsou levé a pravé vektory příslušné vl. číslu 1.

- ▶ Po aplikaci na libovolný náhodný vektor $\vec{P}$ dostaneme $\vec{P} T^\infty = (\vec{P} \cdot \vec{1}) \vec{\pi} = 1 \vec{\pi} = \vec{\pi}$
- ▶ Co vlastně představuje $\vec{\pi}$? Je to stacionární distribuce vůči T !:

$$\vec{\pi} T = \vec{\pi}$$

Příklad: prostor s konečným počtem prvků

$$\Rightarrow T = \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/3 & 1/6 & 1/2 \\ 1/3 & 1/2 & 1/6 \end{pmatrix}$$

- ▶ $\vec{P}_n = \vec{P}_0 T^n, \lim_{n \rightarrow \infty} \vec{P}_n = \vec{P}_0 \lim_{n \rightarrow \infty} T^n =: \vec{P}_0 T^\infty$
- ▶ Matice T má vlastní čísla $(1, -1/3, 1/6)$ $\Rightarrow T^n = V^{-1} \text{diag}(1^n, (-1/3)^n, (1/6)^n) V$, kde V je matice vl. vektorů. V limitě „přežije jen 1“ $\Rightarrow$

$$T^\infty = \vec{1} \vec{\pi} = \begin{pmatrix} \vec{\pi} \\ \vec{\pi} \\ \vec{\pi} \end{pmatrix} = \begin{pmatrix} 4/10 & 3/10 & 3/10 \\ 4/10 & 3/10 & 3/10 \\ 4/10 & 3/10 & 3/10 \end{pmatrix},$$

kde $\vec{1} = (1 \ 1 \ 1)^T$ a $\vec{\pi} = (\frac{4}{10} \ \frac{3}{10} \ \frac{3}{10})$ jsou levé a pravé vektory příslušné vl. číslu 1.

- ▶ Po aplikaci na libovolný náhodný vektor $\vec{P}$ dostaneme $\vec{P} T^\infty = (\vec{P} \cdot \vec{1}) \vec{\pi} = 1 \vec{\pi} = \vec{\pi}$
- ▶ Co vlastně představuje $\vec{\pi}$? Je to stacionární distribuce vůči T !:

$$\vec{\pi} T = \vec{\pi}$$

Pro konečný počet prvků tedy stacionární distribuce odpovídá levému vl. vektoru T s vl. číslem 1. Pokud ho najdeme, můžeme okamžitě spočítat T^∞ .

Protipříklady

- 1) $T = \begin{pmatrix} 3/4 & 1/4 & 0 \\ 1/3 & 2/3 & 0 \\ 0 & 0 & 1 \end{pmatrix} \Rightarrow$ vl. čísla jsou $\{1, 1, \frac{5}{12}\}$. Existuje více stacionárních vektorů: $(0,0,1)$ a $(4/7, 3/7, 0)$. Začneme-li s $(0,0,1)$, nikdy nedokonvergujeme k $(4/7, 3/7, 0)$. Tato matice je blokově diagonální, působí na dva „oddělené světy“ – přechod je *reducibilní*.
- 2) $T = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \Rightarrow$ vl. čísla jsou $\{1, e^{\frac{2\pi i}{3}}, e^{-\frac{2\pi i}{3}}\}$, stacionární vektor je $(1/3, 1/3, 1/3)$. Limita kvůli výskytu komplexních jednotek neexistuje. Tato matice iteruje $(1,0,0) \rightarrow (0,1,0) \rightarrow (0,0,1) \rightarrow (1,0,0)$. Přechod je *periodický*.

Markov Chain Monte Carlo

Fundamentální věta o Markovových řetězcích: Nechť je Markovův řetězec indukovaný jádrem $\tau(x, A)$ definovaný na měřitelném prostoru (X, Σ) se spočetně generovanou σ -algebrou **aperiodický a ireducibilní** a nechť má stacionární pravděpodobnost $\pi(\cdot)$, tj. pravděpodobnost invariantní vůči Markovově přechodu T :

$$\pi(A) = (\pi T)(A) = \int_X \tau(x, A) \pi(dx)$$

Potom pro π - skoro všechna $x \in X$:

$$\lim_{n \rightarrow \infty} \|T^n(x, \cdot) - \pi(\cdot)\| = 0$$

Reversibilita – podmínka detailní rovnováhy

- ▶ Ale jak naopak najít Markovovo jádro, pro který je naše cílová pr. distribuce $P(\cdot)$ stacionární?
- ▶ Postačující (ale ne nutná) podmínka je, že Markovův řetězec je reverzibilní - Markovovo jádro splňuje podmínku *detailní rovnováhy*:

$$P(dx)\tau(x, dy) = P(dy)\tau(y, dx) \quad \forall x, y$$

Reversibilita – podmínka detailní rovnováhy

- ▶ Ale jak naopak najít Markovovo jádro, pro který je naše cílová pr. distribuce $P(\cdot)$ stacionární?
- ▶ Postačující (ale ne nutná) podmínka je, že Markovův řetězec je reverzibilní - Markovovo jádro splňuje podmínku *detailní rovnováhy*:

$$P(dx)\tau(x, dy) = P(dy)\tau(y, dx) \quad \forall x, y$$

- ▶ Integrací přes stavový prostor totiž dostaneme:

$$(PT)(dy) = \int_{x \in X} P(dx)\tau(x, dy) = \int_{x \in X} P(dy)\tau(y, dx) = P(dy) \int_{x \in X} \tau(y, dx) = P(dy),$$

což je přesně podmínka stacionarity P vůči T .

Reversibilita – podmínka detailní rovnováhy

- ▶ Ale jak naopak najít Markovovo jádro, pro který je naše cílová pr. distribuce $P(\cdot)$ stacionární?
- ▶ Postačující (ale ne nutná) podmínka je, že Markovův řetězec je reverzibilní - Markovovo jádro splňuje podmínku *detailní rovnováhy*:

$$P(dx)\tau(x, dy) = P(dy)\tau(y, dx) \quad \forall x, y$$

- ▶ Integrací přes stavový prostor totiž dostaneme:

$$(PT)(dy) = \int_{x \in X} P(dx)\tau(x, dy) = \int_{x \in X} P(dy)\tau(y, dx) = P(dy) \int_{x \in X} \tau(y, dx) = P(dy),$$

což je přesně podmínka stacionarity P vůči T .

- ▶ Pro konečný prostor to odpovídá podmínce

$$P_i \tau_{ij} = P_j \tau_{ji} \quad \forall i, j$$

A důkaz součtu přes i :

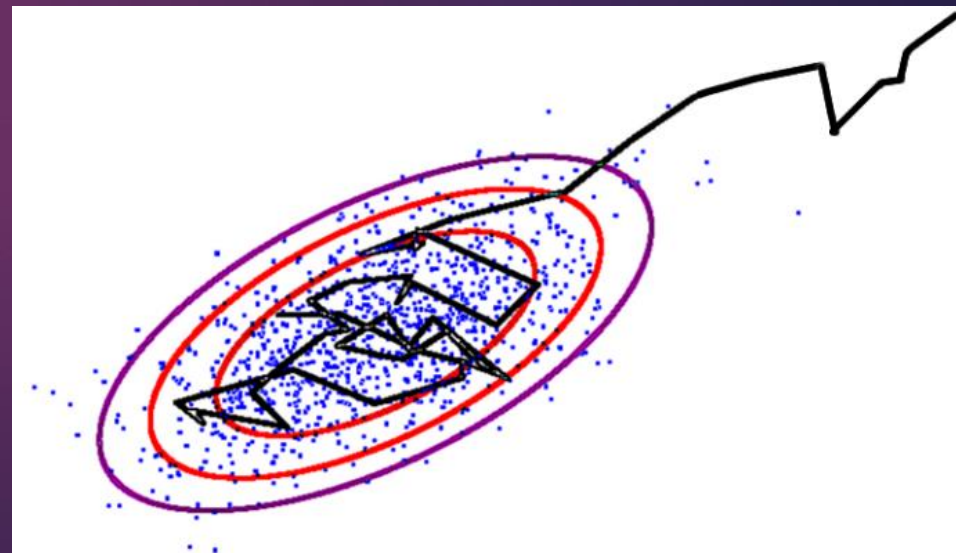
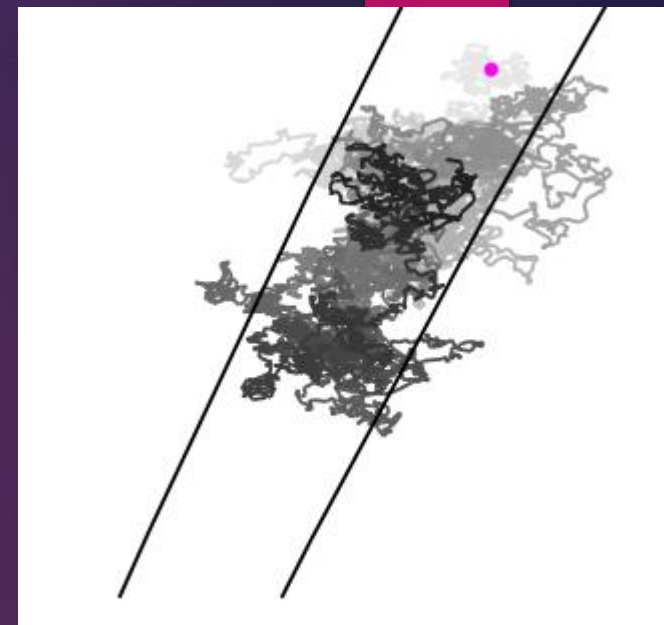
$$\sum_i P_i \tau_{ij} = \sum_i P_j \tau_{ji} = P_j \Rightarrow \vec{P} \text{ je vl. vektor } \tau.$$

Příklad –Metropolis-Hastings

1. Začni v arbitrárním bodě $x \in X$, $x \sim P_0(\cdot)$. Předpokládejme, že chceme navzorkovat pravděpodobnost $P(\cdot)$, $P(dx) = f(x)dx$
 2. Definuj si libovolné Markovovo jádro $Q(x, dy) \propto q(x, y)dy$ (často $q(x, y) = q(|x - y|)$ – symetrický random-walk MH. Typicky $q(x, y) = N(x - y|0, \sigma^2)$).
 3. Vygeneruj návrh $n, n \sim Q(x_0, \cdot)$
 4. Přijmi návrh $x_i \rightarrow x_{i+1} = n$ s pravděpodobností $\alpha(x_i, n) = \min[1, \frac{f(n)q(n, x_i)}{f(x_i)q(x_i, n)}]$
 5. Opakuj kroky 3. a 4. pořád dokola
- Složení kroků 3 a 4 dohromady vytvoří nové jádro $\tau(x, dy) = q(x, y)\alpha(x, y)dy$, o kterém se dá ukázat, že splňuje podmínku detailní rovnováhy vůči $P(\cdot)$. $P(\cdot)$ je tedy vůči přechodům indukovaným náhodnou procházkou stacionární a proto nám náhodná procházka nakonec bude dávat vzorky $P(\cdot)$.

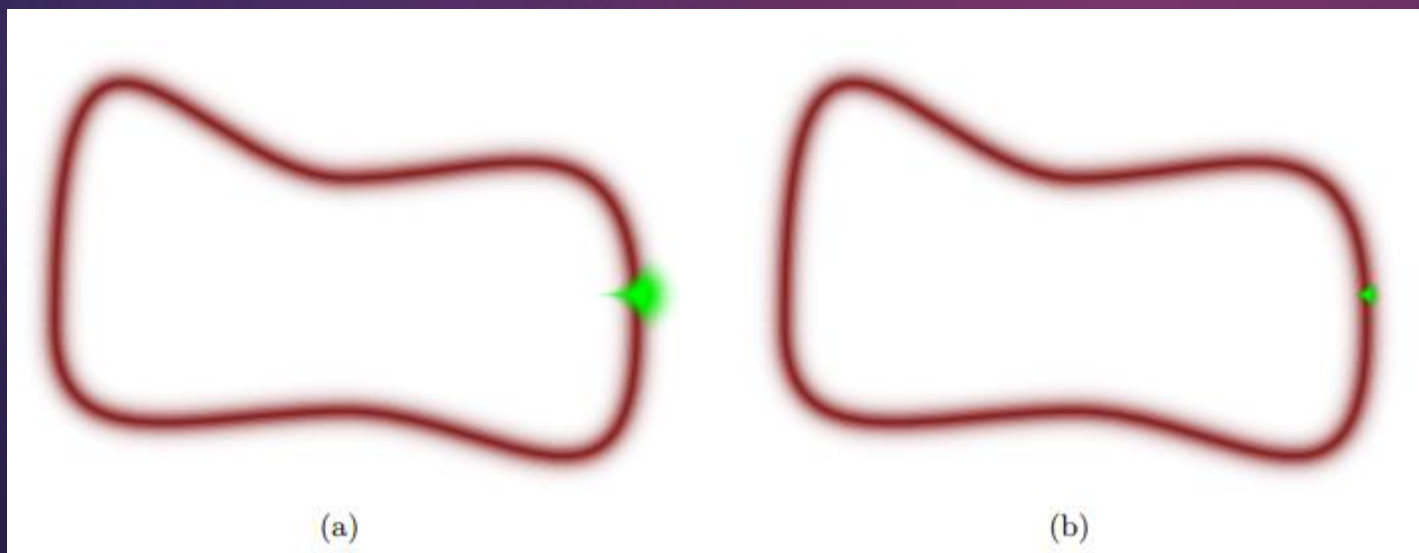
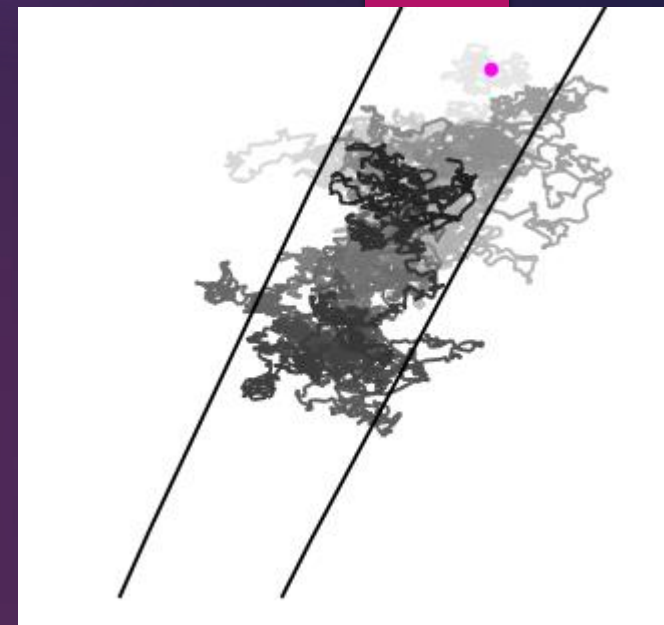
Příklad –Metropolis-Hastings

- ▶ Ale jak rychlá bude konvergence? Například pro symetrický random walk máme difuzní pohyb, často budeme dělat neefektivní kroky zpátky (uražená vzdálenost bude růst zhruba jako $\sqrt{n}$).
- ▶ $P(\cdot)$ se sice díky přechodům zachová, ale body mají tendenci jít mimo typickou množinu, kde je (hlavně ve vysokých dimenzích extrémně nízké PDF $\Rightarrow$ vysoká míra zamítání).



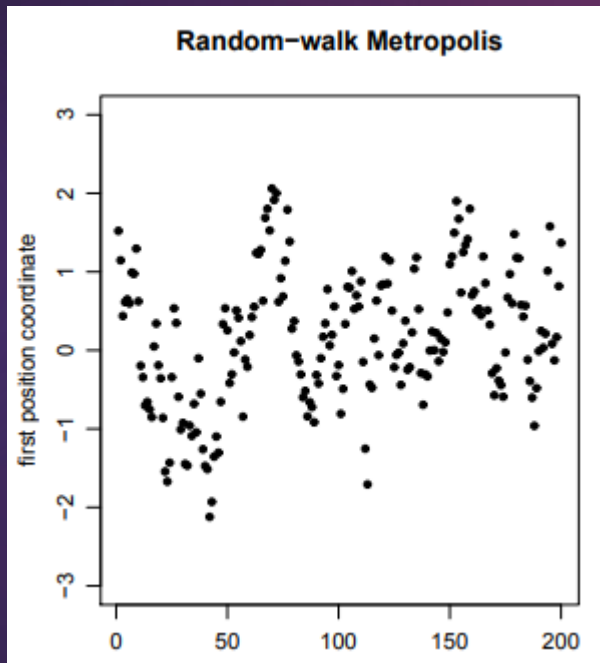
Příklad –Metropolis-Hastings

- ▶ Ale jak rychlá bude konvergence? Například pro symetrický random walk máme difuzní pohyb, často budeme dělat neefektivní kroky zpátky (uražená vzdálenost bude růst zhruba jako $\sqrt{n}$).
- ▶ $P(\cdot)$ se sice díky přechodům zachová, ale body mají tendenci jít mimo typickou množinu, kde je (hlavně ve vysokých dimenzích extrémně nízké PDF $\Rightarrow$ vysoká míra zamítání).
- ▶ To se dá řešit zmenšením kroku, ale pak se budem soustředit jen kolem počátečního bodu.

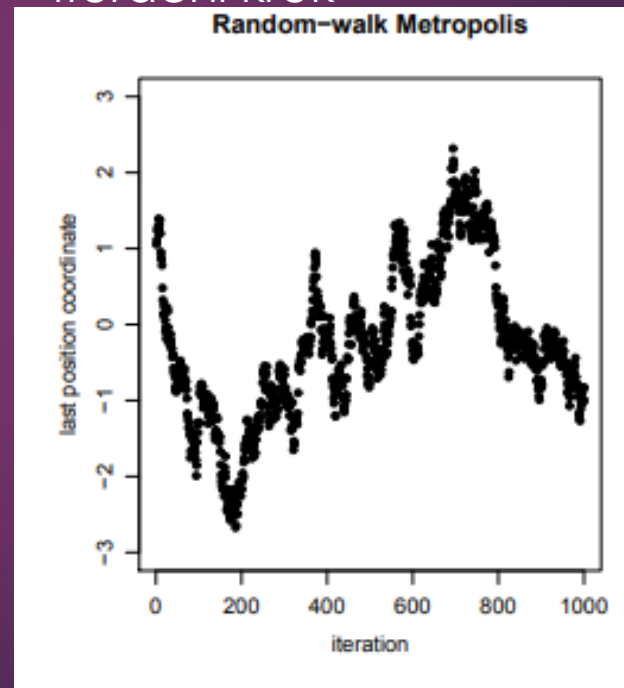


Příklad –Metropolis-Hastings Samplování 2D gaussovky

Poloha první souřadnice
vs iterační krok



Aritmetický průměr vs
iterační krok



Toky

- ▶ Lepší by bylo, kdybychom mohli modelový prostor procházet nějak koherentněji.

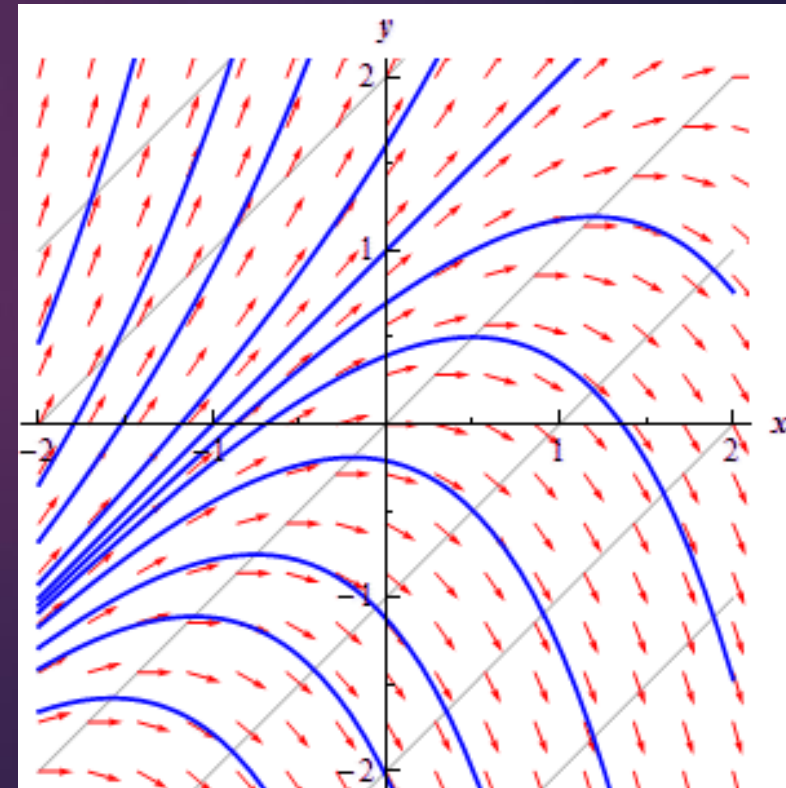
Toky

- ▶ Lepší by bylo, kdybychom mohli modelový prostor procházet nějak koherentněji. Ideální kandidátem jsou *toky* (flows), tj. zobrazení

$\phi: X \times \mathbb{R} \rightarrow X : (x, t) \rightarrow y = \phi(x, t) = \phi_t(x)$, parametrizovaná časem t , pro která

- ▶ $\phi_0 = Id_X$
- ▶ $\phi_t^{-1} = \phi_{-t}$
- ▶ $\phi_{t_1} \circ \phi_{t_2} = \phi_{t_1+t_2}$

Například $\phi_t(x, y) = (x + t, y + t)$, $\phi_t(x, y) = (x + \cos t, y + \sin t)$



Toky

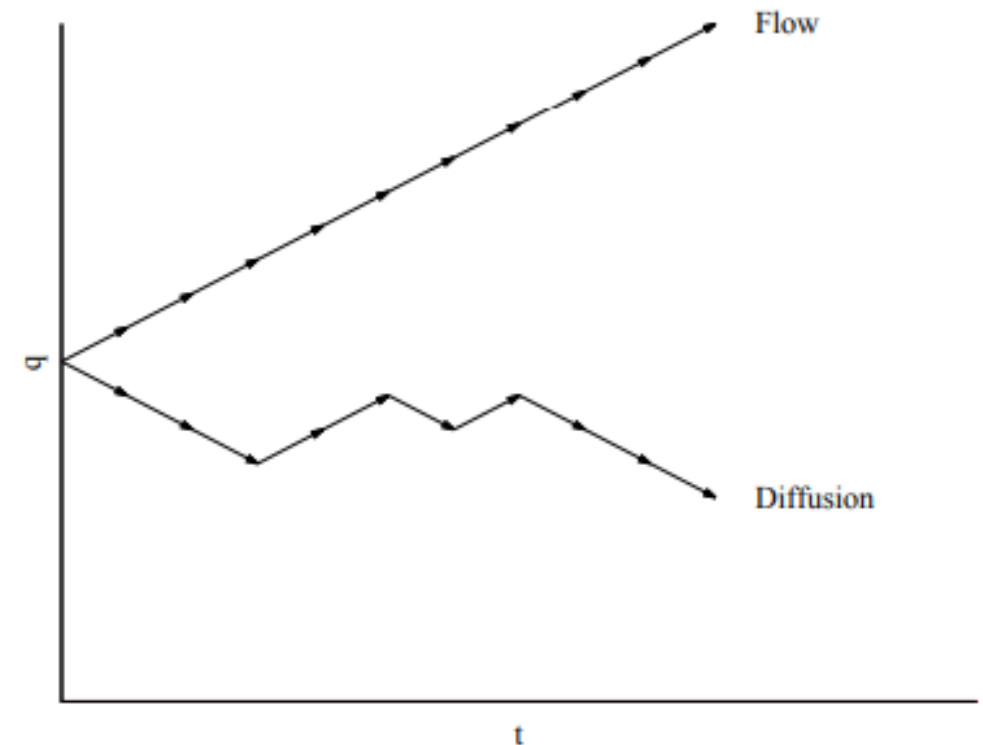
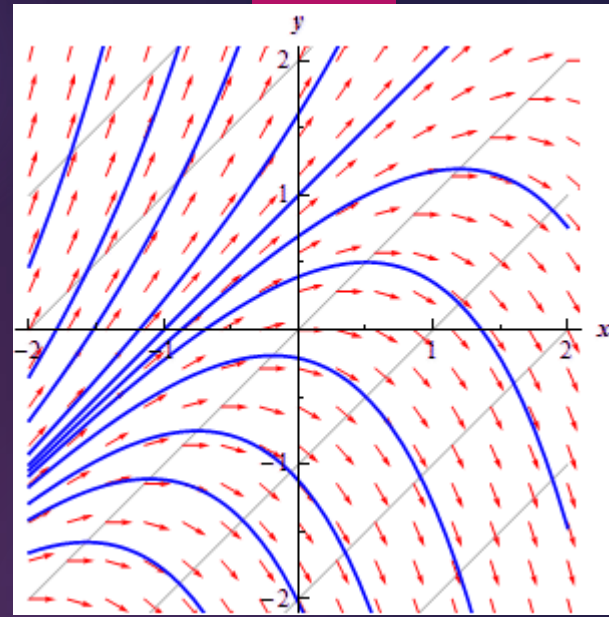
- Lepší by bylo, kdybychom mohli modelový prostor procházet nějak koherentněji. Ideální kandidátem jsou *toky* (flows), tj. zobrazení

$\phi: X \times \mathbb{R} \rightarrow X: (x, t) \rightarrow y = \phi(x, t) = \phi_t(x)$, parametrizovaná časem t , pro která:

1. $\phi_0 = Id_X$
2. $\phi_t^{-1} = \phi_{-t}$
3. $\phi_{t_1} \circ \phi_{t_2} = \phi_{t_1+t_2}$

Například $\phi_t(x, y) = (x + t, y + t)$, $\phi_t(x, y) = (x + \cos t, y + \sin t)$

- Toky mají tu vlastnost, že se nemohou jen tak vracet zpátky.



Toky

- Lepší by bylo, kdybychom mohli modelový prostor procházet nějak koherentněji. Ideální kandidátem jsou *toky* (flows), tj. zobrazení

$\phi: X \times \mathbb{R} \rightarrow X: (x, t) \rightarrow y = \phi(x, t) = \phi_t(x)$, parametrizovaná časem t , pro která:

1. $\phi_0 = Id_X$
2. $\phi_t^{-1} = \phi_{-t}$
3. $\phi_{t_1} \circ \phi_{t_2} = \phi_{t_1+t_2}$

Například $\phi_t(x, y) = (x + t, y + t)$, $\phi_t(x, y) = (x + \cos t, y + \sin t)$

- Toky mají tu vlastnost, že se nemohou jen tak vracet zpátky.

Pokud by se totiž tok vrátil sám na sebe, $\exists t_1, t_2 \in \mathbb{R}: \phi_{t_1} = \phi_{t_2}$, tak:

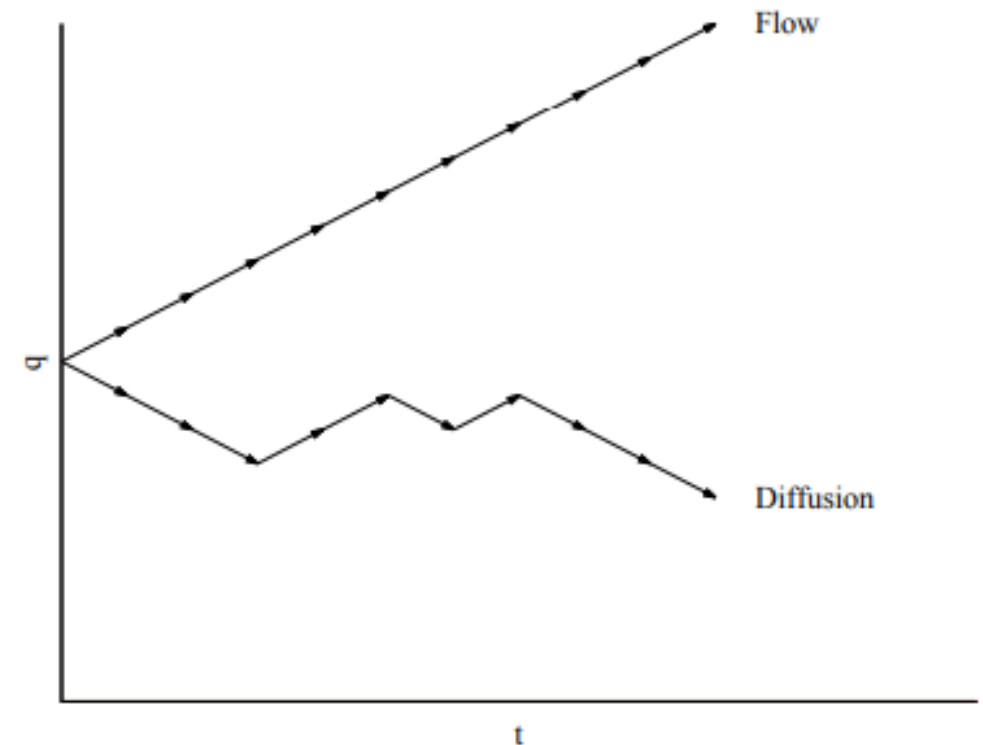
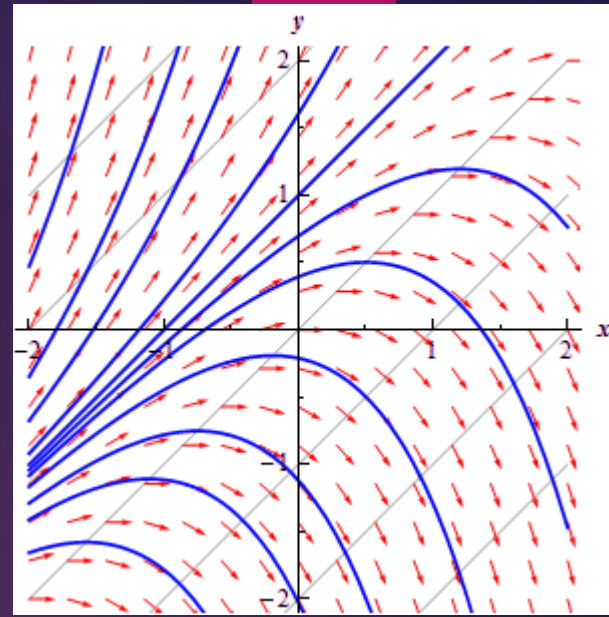
$$\phi_{t_1} = \phi_{t_2}$$

$$\phi_{t_1} \circ \phi_{t_2}^{-1} = \phi_{t_2} \circ \phi_{t_2}^{-1}$$

$$\phi_{t_1} \circ \phi_{-t_2} = \phi_{t_2} \circ \phi_{-t_2}$$

$$\phi_{t_1-t_2} = Id_X$$

$\Rightarrow T = t_1 - t_2$ je perioda toku, $\phi_t = \phi_{t+T} \forall t \in \mathbb{R}$.



Toky zachovávající pravděpodobnost

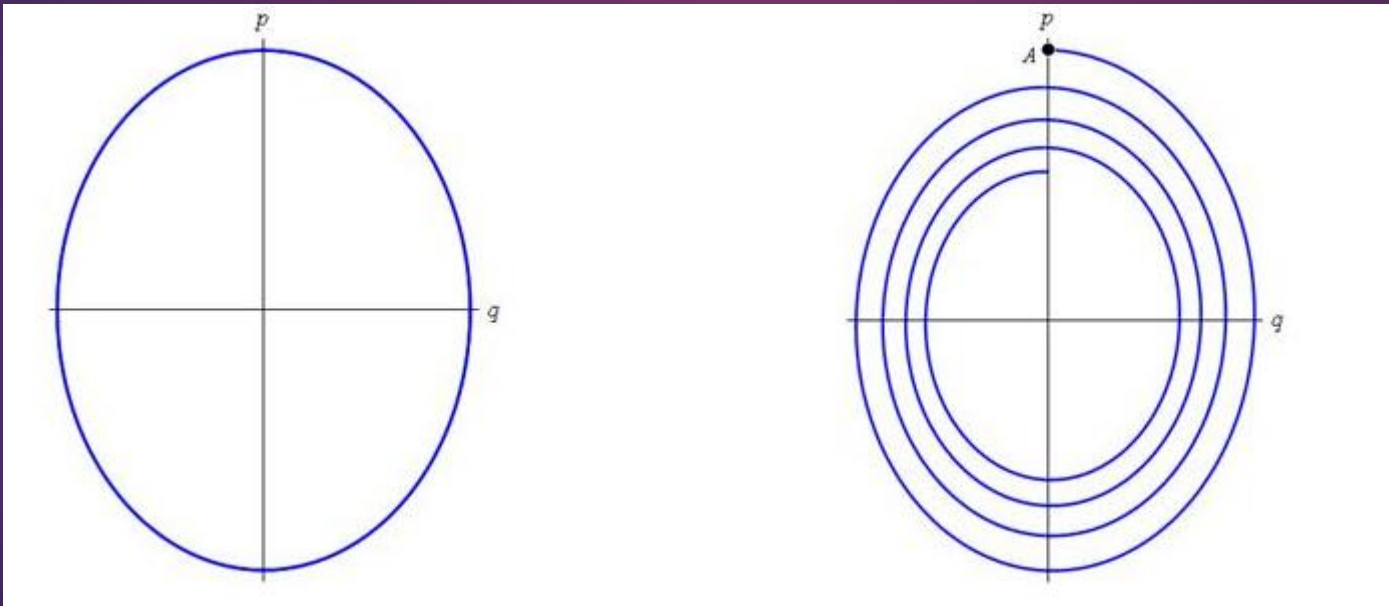
- ▶ Nám ale nestačí jen rychle procházet modelový prostor, musíme ho procházet tak, aby nám přechody zachovávaly pravděpodobnostní míru, jinak nepůjde použít fundamentální věta o Markovových řetězcích.
- ▶ Známe nějaké toky, které zachovávají míru?

Toky zachovávající pravděpodobnost

- ▶ Nám ale nestačí jen rychle procházet modelový prostor, musíme ho procházet tak, aby nám přechody zachovávaly pravděpodobnostní míru, jinak nepůjde použít fundamentální věta o Markovových řetězcích.
- ▶ Známe nějaké toky, které zachovávají míru?
- ▶ Ano! Toky odpovídající časovému vývoji systémů klasické fyziky! *Liouvilleova věta*

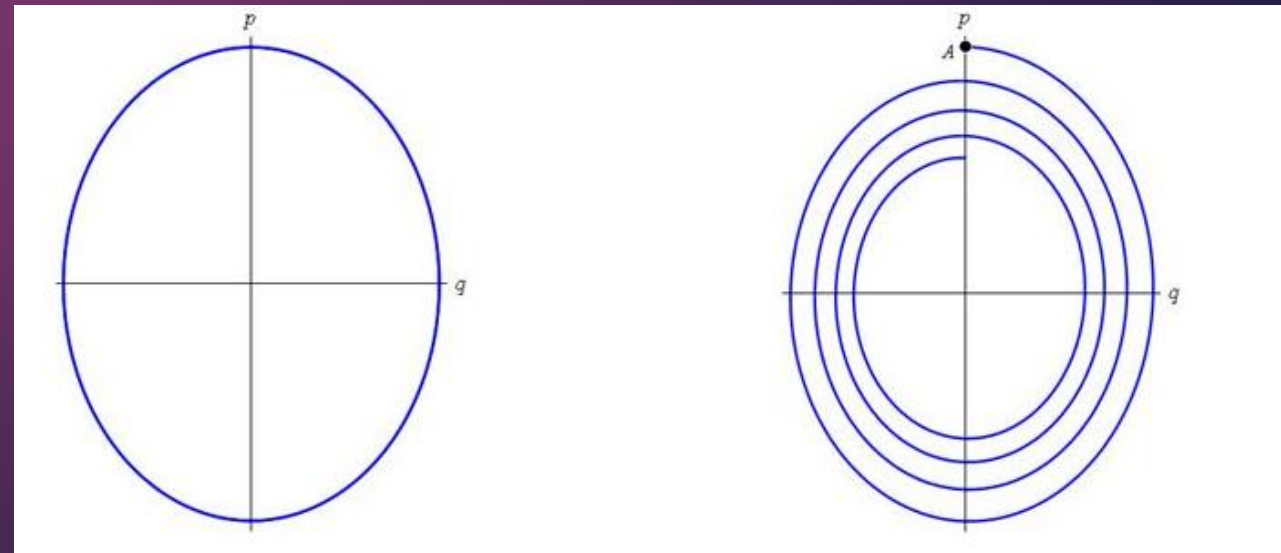
Připomenutí: fázový prostor

- Prostor, na kterém „žijí“ stavy klasické mechaniky. Definujeme na něm *kanonické souřadnice*: zobecněné pozice q^i a zobecněné hybnosti p_i . Jeho dimenze je vždy sudá. Dynamika stavů fázového prostoru se řídí Hamiltonovými rovnicemi.



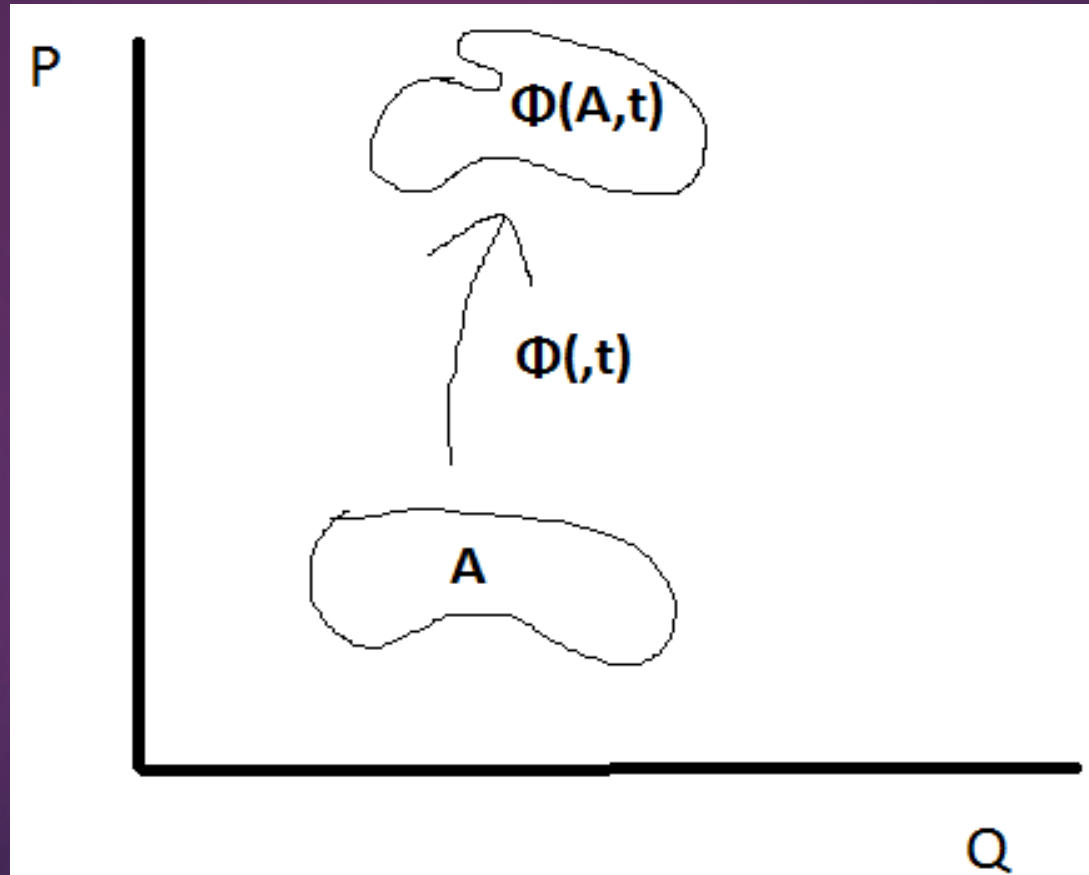
Připomenutí: fázový prostor

- ▶ Prostor, na kterém „žijí“ stavy klasické mechaniky. Definujeme na něm *kanonické souřadnice*: zobecněné pozice q^i a zobecněné hybnosti p_i . Jeho dimenze je vždy sudá. Dynamika stavů fázového prostoru se řídí Hamiltonovými rovnicemi.
- ▶ Z hlediska dif. geometrie se jedná o tzv. kotečný bandl T^*Q konfigurační variety Q . Tato geometrická struktura (fibrovaného bandlu) je z matematického hlediska pro metodu důležitá, protože umožňuje dobře definovat podmíněné pravděpodobnosti (viz Betancourt, 2014).



Liouvilleova věta

- Objem oblasti ve fázovém prostoru se při hamiltonovském vývoji nemění.



Liouvilleova věta

- Objem oblasti ve fázovém prostoru se při hamiltonovském vývoji nemění.
- Přesněji: Nechť M je $2n$ -dimenzionální fázový prostor s kanonickými souřadnicemi $q^i, p_i, i = 1, 2, \dots, n$. Nechť $H: M \times R \rightarrow R$ je hamiltonián (obecně závislý na čase) a nechť $\Phi: M \times R \rightarrow M : ((\mathbf{q}_0, \mathbf{p}_0), t) \rightarrow \Phi((\mathbf{q}_0, \mathbf{p}_0), t) = (\mathbf{q}(t), \mathbf{p}(t))$ je řešením Hamiltonových rovnic:

$$\frac{dq^i}{dt} = \frac{\partial H}{\partial p_i}, \frac{dp_i}{dt} = -\frac{\partial H}{\partial q^i}$$

s počátečními podmínkami $\mathbf{q}(0) = \mathbf{q}_0, \mathbf{p}(0) = \mathbf{p}_0$ v čase t . Nechť A je nějaká (borelovská) podmnožina M a $\mu(A)$ její objem.

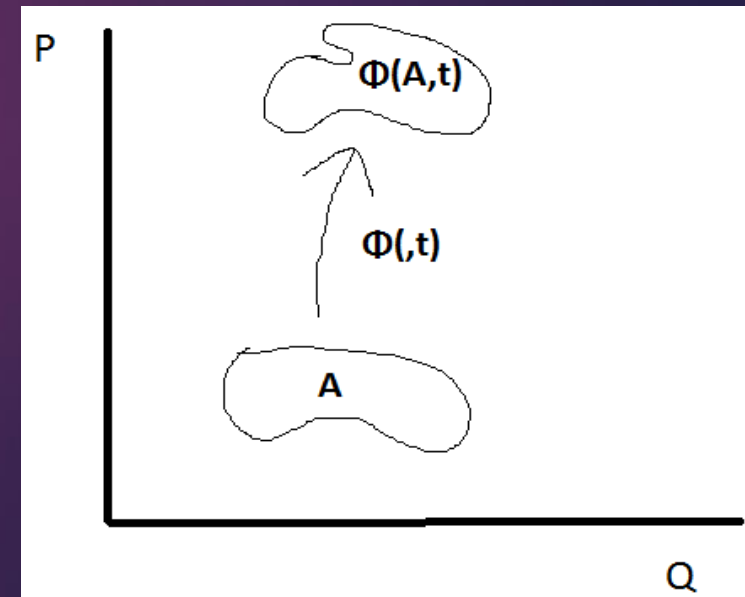
$$\text{Potom } \mu(\Phi(A, t)) = \mu(A) \quad \forall t \in R$$

Infinitesimálně:

$$dq^1 \dots dq^n dp^1 \dots dp^n = dq(t)^1 \dots dq(t)^n dp(t)^1 \dots dp(t)^n$$

Zkráceně

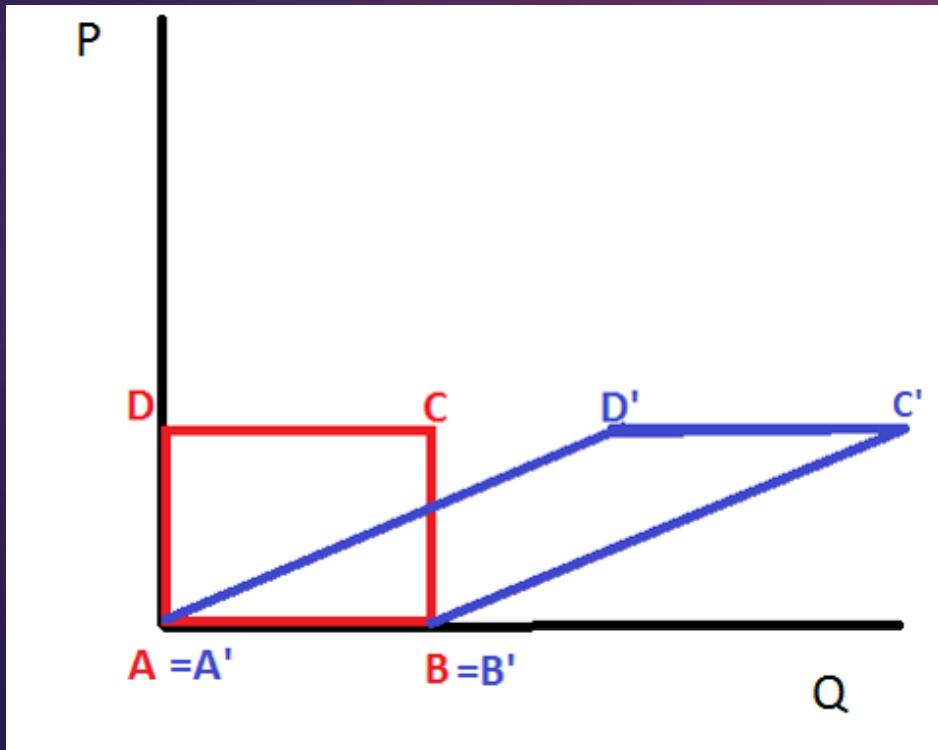
$$dqdp = dq(t)dp(t)$$



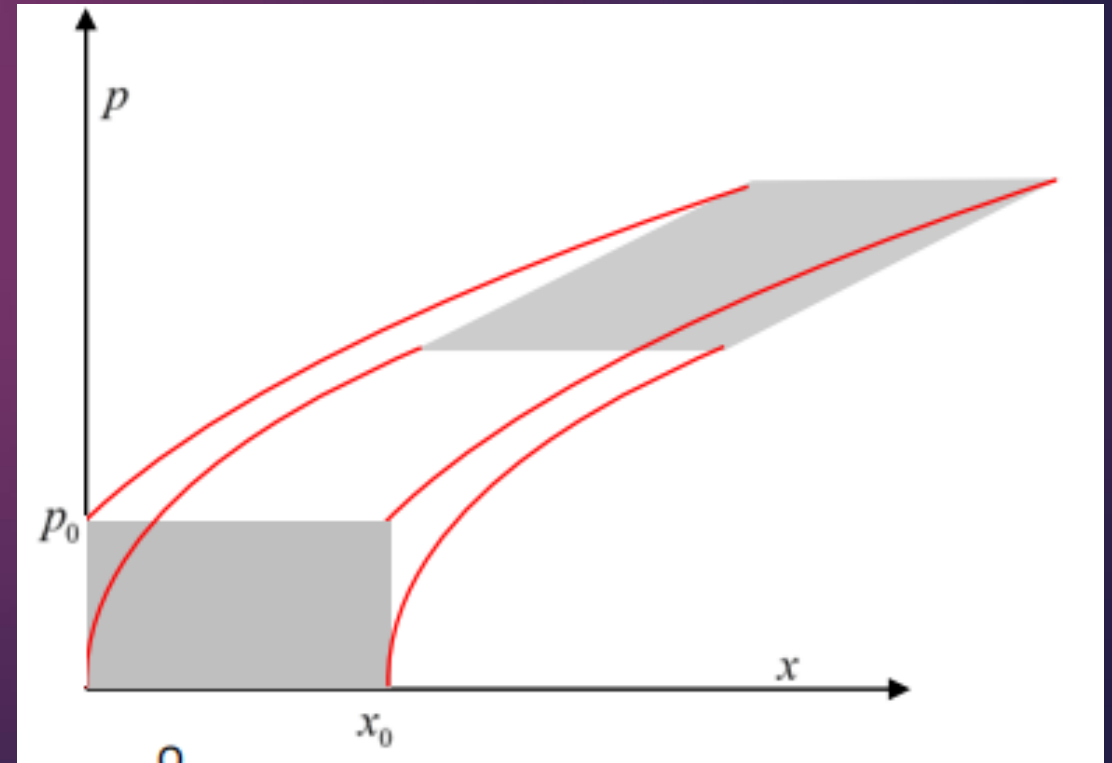
Liouvilleova věta - příklady

- ▶ Objem oblasti ve fázovém prostoru se při vývoji hamiltonovského systému nemění.

- ▶ Příklad: volná částice, $H = \frac{p^2}{m}$



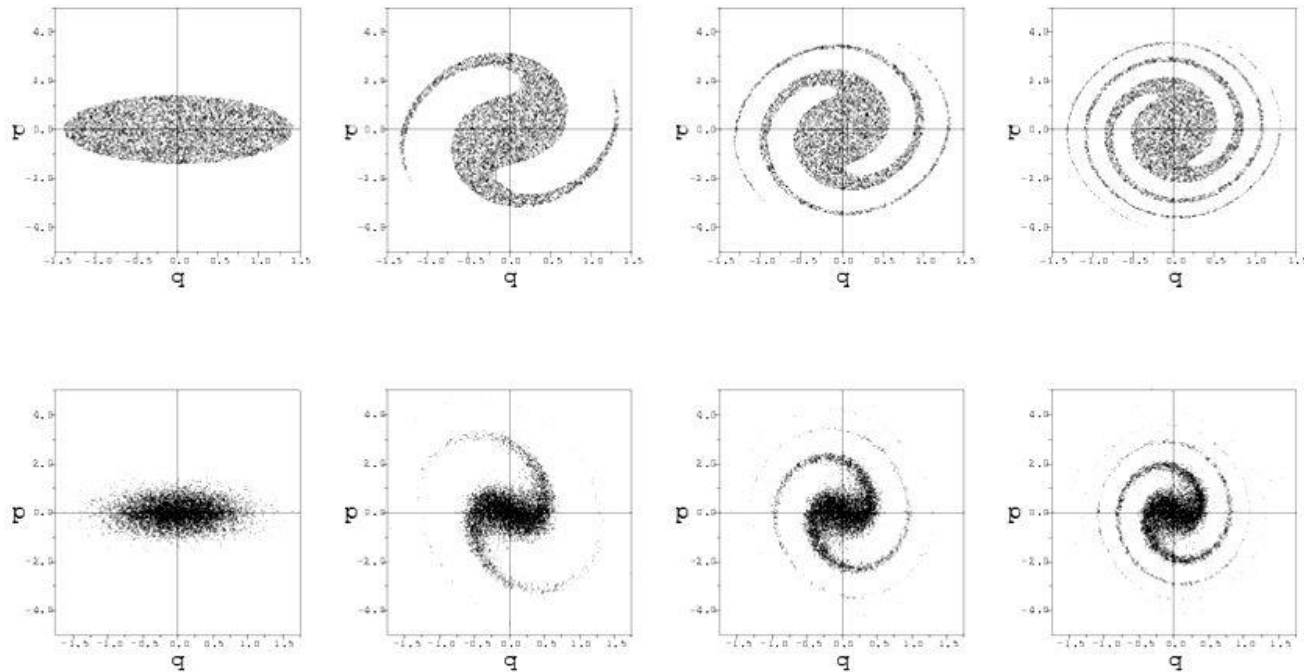
- ▶ Příklad: volný pád, $H = \frac{p^2}{m} + mgq$



Liouvilleova věta - příklady



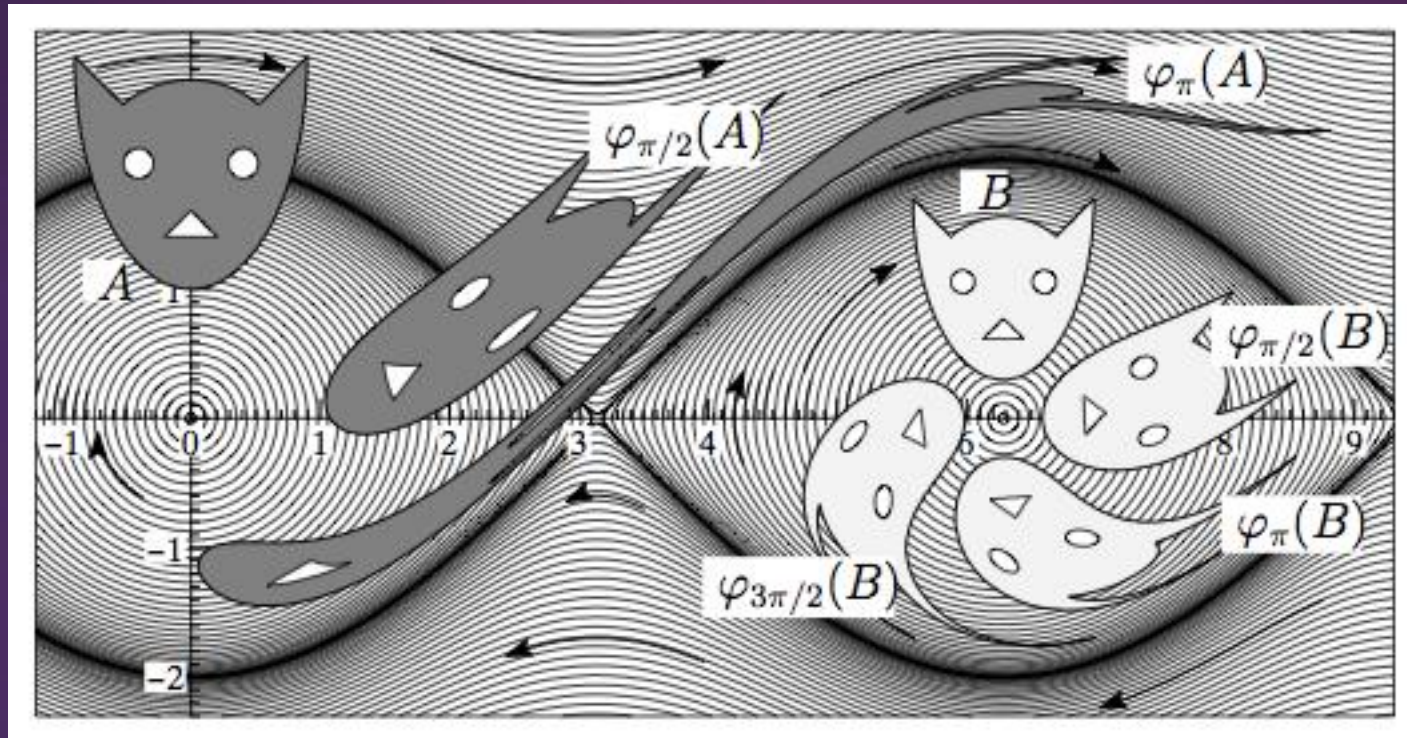
1. Review of Hamiltonian dynamics 1.5. Liouville's theorem



Time evolution of a uniform (top) and a Gaussian (bottom) phase-space distribution of initial conditions for the mathematical pendulum.

Liouvilleova věta - příklady

I objem kočiček je invariant.



Toky zachovávající pravděpodobnost

- ▶ Nám ale nestačí jen rychle procházet modelový prostor, musíme ho procházet tak, aby nám přechody zachovávaly pravděpodobnostní míru, jinak nepůjde použít fundamentální věta o Markovových řetězcích.
- ▶ Známe nějaké toky, které zachovávají míru?
- ▶ Ano! Toky odpovídající časovému vývoji systémů klasické fyziky! $d\mathbf{q}d\mathbf{p} = d\mathbf{q}(t)d\mathbf{p}(t)$
- ▶ Navíc se při vývoji systému zachovává hodnota hamiltoniánu: $H(\mathbf{q}(t), \mathbf{p}(t)) = H(\Phi^i((\mathbf{q}_0, \mathbf{p}_0), t)) = H((\mathbf{q}_0, \mathbf{p}_0))$

Toky zachovávající pravděpodobnost

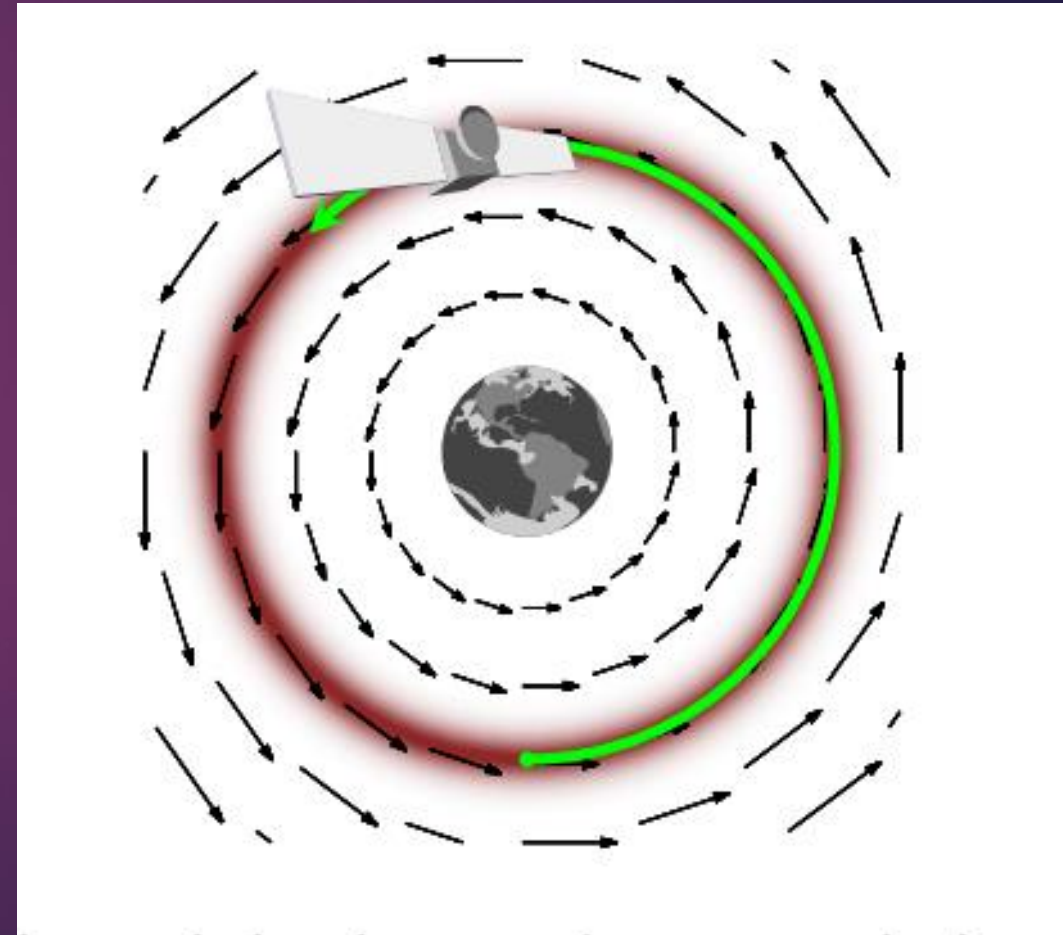
- ▶ Nám ale nestačí jen rychle procházet modelový prostor, musíme ho procházet tak, aby nám přechody zachovávaly pravděpodobnostní míru, jinak nepůjde použít fundamentální věta o Markovových řetězcích.
- ▶ Známe nějaké toky, které zachovávají míru?
- ▶ Ano! Toky odpovídající časovému vývoji systémů klasické fyziky! $d\mathbf{q}d\mathbf{p} = d\mathbf{q}(t)d\mathbf{p}(t)$
- ▶ Navíc se při vývoji systému zachovává hodnota hamiltoniánu: $H(\mathbf{q}(t), \mathbf{p}(t)) = H(\Phi^i((\mathbf{q}_0, \mathbf{p}_0), t)) = H((\mathbf{q}_0, \mathbf{p}_0))$
- ▶ Zachovávají se tedy i všechny míry $P(\cdot)$ na fázovém prostoru ve tvaru $P(d\mathbf{q}d\mathbf{p}) = \exp(-H(\mathbf{q}, \mathbf{p})) d\mathbf{q}d\mathbf{p}$

Toky zachovávající pravděpodobnost

- ▶ Nám ale nestačí jen rychle procházet modelový prostor, musíme ho procházet tak, aby nám přechody zachovávaly pravděpodobnostní míru, jinak nepůjde použít fundamentální věta o Markovových řetězcích.
- ▶ Známe nějaké toky, které zachovávají míru?
- ▶ Ano! Toky odpovídající časovému vývoji systémů klasické fyziky! $d\mathbf{q}d\mathbf{p} = d\mathbf{q}(t)d\mathbf{p}(t)$
- ▶ Navíc se při vývoji systému zachovává hodnota hamiltoniánu: $H(\mathbf{q}(t), \mathbf{p}(t)) = H(\Phi^i((\mathbf{q}_0, \mathbf{p}_0), t)) = H((\mathbf{q}_0, \mathbf{p}_0))$
- ▶ Zachovávají se tedy i všechny míry $P(\cdot)$ na fázovém prostoru ve tvaru $P(d\mathbf{q}d\mathbf{p}) = \exp(-H(\mathbf{q}, \mathbf{p})) d\mathbf{q}d\mathbf{p}$
- ▶ Po normalizaci se z těchto měr stanou pravděpodobnosti.

Hamiltonovské Monte Carlo

- ▶ Chceme navzorkovat pravděpodobnost $P(\cdot) = \pi(q)dq$ na modelovém prostoru Q .
- ▶ Idea: Začneme s nějakým bodem $q \in Q$ a přidejme mu nějakou náhodnou hybnost p . Tím vyzdihneme Q na fázový prostor M (a nespadne nám do planety!). Definujme si tam nějaký hamiltonián, který bude šikovně souviset s cílovou pravděpodobností a nechme (q, p) nějakou dobu vyvíjet pomocí hamiltonovského toku, pak promítněme zpátky. Tím získáme Markovův přechod, který efektivně prochází M a zároveň je vůči němu cílová pravděpodobnost invariantní. Opakujme tuto posloupnost – vyzdvih, tok, projekce – pořád dokola. Díky fundamentální větě Markovových řetězců tak dostaneme vzorky požadované pravděpodobnosti.

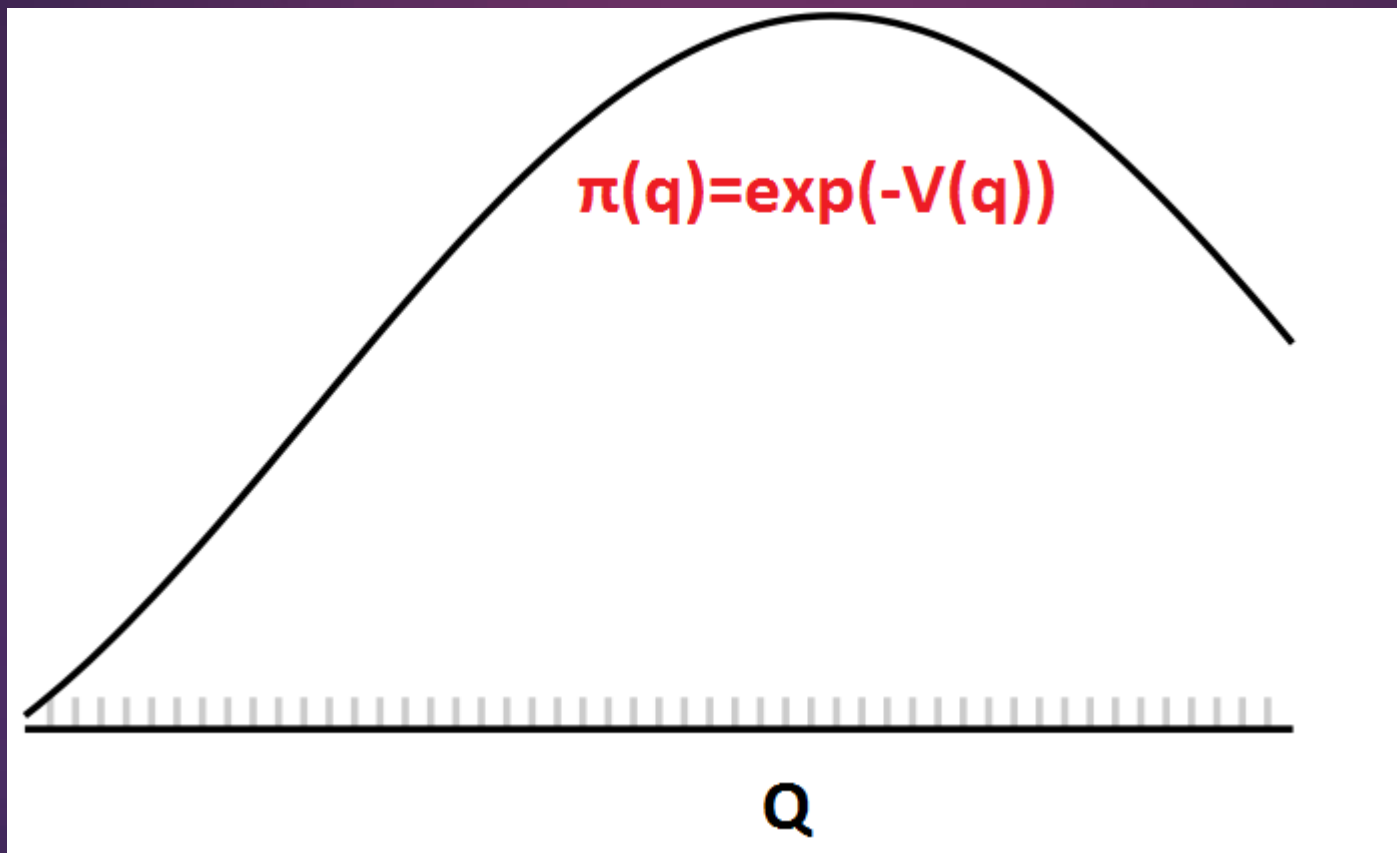


Hamiltonovské Monte Carlo

- ▶ Mějme n -dimenzionální modelový prostor Q se souřadnicemi q^i , chceme navzorkovat $P(\cdot)$: $P(d\mathbf{q}) = \exp(-V(\mathbf{q}))d\mathbf{q} \Rightarrow$ PDF: $\pi(\mathbf{q}) = \exp(-V(\mathbf{q}))$. Funkci $V(\mathbf{q})$ budeme říkat *potenciální* energie.

Hamiltonovské Monte Carlo

- Mějme n -dimenzionální modelový prostor Q se souřadnicemi q^i , chceme navzorkovat $P(\cdot)$: $P(dq) = \exp(-V(q))dq \Rightarrow$ PDF: $\pi(q) = \exp(-V(q))$. Funkci $V(q)$ budeme říkat *potenciální* energie.



Hamiltonovské Monte Carlo

- ▶ Mějme n -dimenzionální modelový prostor Q se souřadnicemi q^i , chceme navzorkovat $P(\cdot)$: $P(d\mathbf{q}) = \exp(-V(\mathbf{q}))d\mathbf{q} \Rightarrow$ PDF: $\pi(\mathbf{q}) = \exp(-V(\mathbf{q}))$. Funkci $V(\mathbf{q})$ budeme říkat *potenciální* energie.
- ▶ Rozšířme Q na $2n$ -dimenzionální fázový prostor M přidáním n hybností p_i , $i = 1, 2, \dots, n$:

$$(q^i) \rightarrow (q^i, p_i) \text{ tj. } (\mathbf{q}) \rightarrow (\mathbf{q}, \mathbf{p})$$

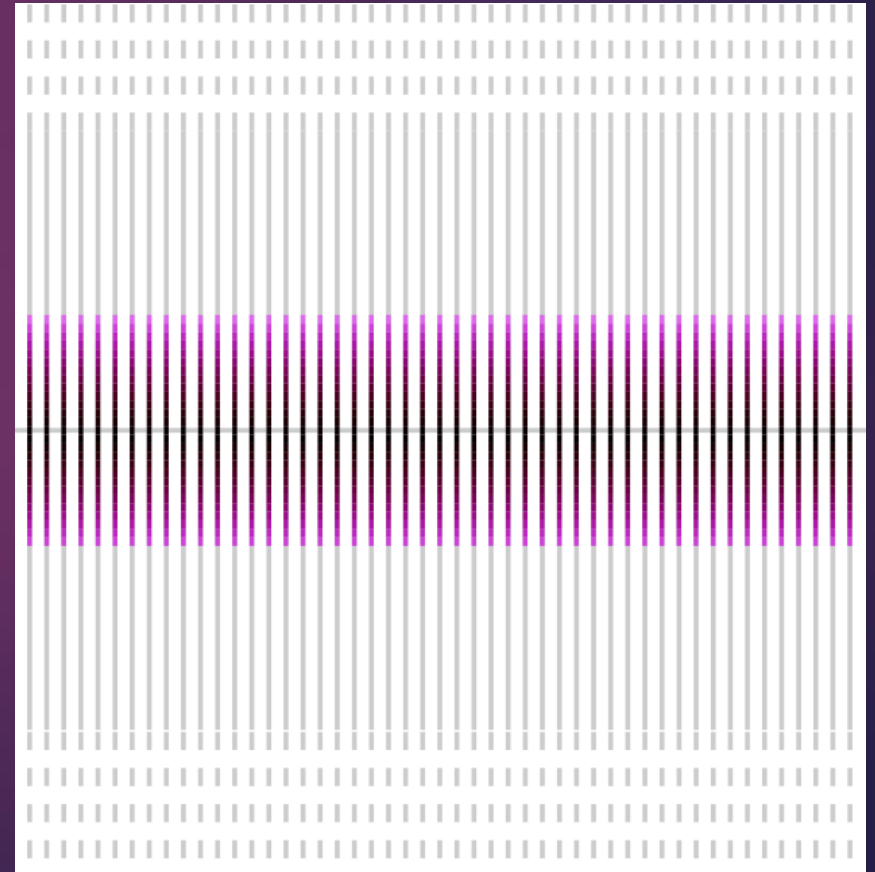
- ▶ $\mathbf{p}$ dostaneme jako $\mathbf{p} \sim T(\cdot)$, kde $T(d\mathbf{p}) = \exp(-K(\mathbf{q}, \mathbf{p}))d\mathbf{p} \Rightarrow$ PDF: $\pi(\mathbf{p}|\mathbf{q}) = \exp(-K(\mathbf{q}, \mathbf{p}))$. Funkci $K(\mathbf{q}, \mathbf{p})$, kterou můžeme zvolit libovolně, budeme říkat *kinetická* energie.

Hamiltonovské Monte Carlo

- ▶ Mějme n -dimenzionální modelový prostor Q se souřadnicemi q^i , chceme navzorkovat $P(\cdot)$: $P(d\mathbf{q}) = \exp(-V(\mathbf{q}))d\mathbf{q} \Rightarrow$ PDF: $\pi(\mathbf{q}) = \exp(-V(\mathbf{q}))$. Funkci $V(\mathbf{q})$ budeme říkat *potenciální* energie.
- ▶ Rozšířme Q na $2n$ -dimenzionální fázový prostor M přidáním n hybností p_i , $i = 1, 2, \dots, n$:

$$(q^i) \rightarrow (q^i, p_i) \text{ tj. } (\mathbf{q}) \rightarrow (\mathbf{q}, \mathbf{p})$$

- ▶ $\mathbf{p}$ dostaneme jako $\mathbf{p} \sim T(\cdot)$, kde $T(d\mathbf{p}) = \exp(-K(\mathbf{q}, \mathbf{p}))d\mathbf{p} \Rightarrow$ PDF: $\pi(\mathbf{p}|\mathbf{q}) = \exp(-K(\mathbf{q}, \mathbf{p}))$. Funkci $K(\mathbf{q}, \mathbf{p})$, kterou můžeme zvolit libovolně, budeme říkat *kinetická* energie.



Hamiltonovské Monte Carlo

- ▶ Mějme n -dimenzionální modelový prostor Q se souřadnicemi q^i , chceme navzorkovat $P(\cdot)$: $P(d\mathbf{q}) = \exp(-V(\mathbf{q}))d\mathbf{q} \Rightarrow$ PDF: $\pi(\mathbf{q}) = \exp(-V(\mathbf{q}))$. Funkci $V(\mathbf{q})$ budeme říkat *potenciální* energie.
- ▶ Rozšířme Q na $2n$ -dimenzionální fázový prostor M přidáním n hybností p_i , $i = 1, 2, \dots, n$:

$$(q^i) \rightarrow (q^i, p_i) \text{ tj. } (\mathbf{q}) \rightarrow (\mathbf{q}, \mathbf{p})$$

- ▶ $\mathbf{p}$ dostaneme jako $\mathbf{p} \sim T(\cdot)$, kde $T(d\mathbf{p}) = \exp(-K(\mathbf{q}, \mathbf{p}))d\mathbf{p} \Rightarrow$ PDF: $\pi(\mathbf{p}|\mathbf{q}) = \exp(-K(\mathbf{q}, \mathbf{p}))$. Funkci $K(\mathbf{q}, \mathbf{p})$, kterou můžeme zvolit libovolně, budeme říkat *kinetická* energie.
- ▶ Kombinací $P(\cdot)$ a $T(\cdot)$ dostaneme pravděpodobnost na celém fázovém prostoru:
$$\pi(\mathbf{q}, \mathbf{p})d\mathbf{q}d\mathbf{p} = \pi(\mathbf{p}|\mathbf{q})\pi(\mathbf{q})d\mathbf{q}d\mathbf{p}$$
- ▶ $\pi(\mathbf{q}, \mathbf{p})$ je hustota pravděpodobnosti na celém fázovém prostoru, $\pi(\mathbf{q}, \mathbf{p}) = \exp(-(V(\mathbf{q}) + K(\mathbf{q}, \mathbf{p}))) = \exp(-H(\mathbf{q}, \mathbf{p}))$. Funkci $H(\mathbf{q}, \mathbf{p})$ budeme nazývat *hamiltoniánem*.

Hamiltonovské Monte Carlo

- ▶ Mějme n -dimenzionální modelový prostor Q se souřadnicemi q^i , chceme navzorkovat $P(\cdot)$: $P(d\mathbf{q}) = \exp(-V(\mathbf{q}))d\mathbf{q} \Rightarrow$ PDF: $\pi(\mathbf{q}) = \exp(-V(\mathbf{q}))$. Funkci $V(\mathbf{q})$ budeme říkat *potenciální* energie.
- ▶ Rozšířme Q na $2n$ -dimenzionální fázový prostor M přidáním n hybností p_i , $i = 1, 2, \dots, n$:

$$(q^i) \rightarrow (q^i, p_i) \text{ tj. } (\mathbf{q}) \rightarrow (\mathbf{q}, \mathbf{p})$$

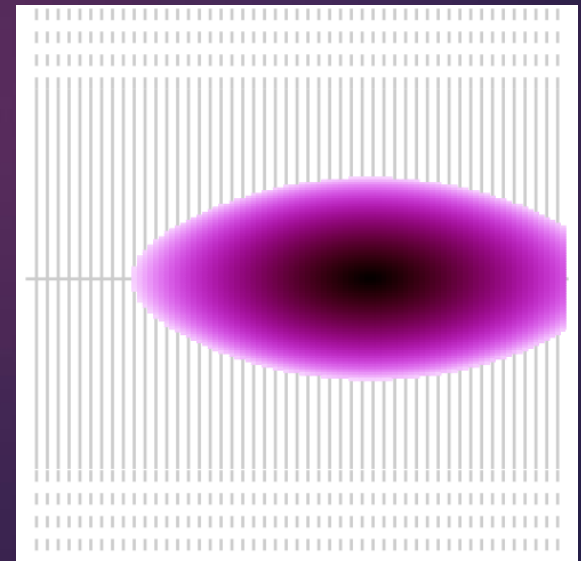
- ▶ $\mathbf{p}$ dostaneme jako $\mathbf{p} \sim T(\cdot)$, kde $T(d\mathbf{p}) = \exp(-K(\mathbf{q}, \mathbf{p}))d\mathbf{p} \Rightarrow$ PDF: $\pi(\mathbf{p}|\mathbf{q}) = \exp(-K(\mathbf{q}, \mathbf{p}))$. Funkci $K(\mathbf{q}, \mathbf{p})$, kterou můžeme zvolit libovolně, budeme říkat *kinetická* energie.
- ▶ Kombinací $P(\cdot)$ a $T(\cdot)$ dostaneme pravděpodobnost na celém fázovém prostoru: *Kanonický rozklad*
 $\pi(\mathbf{q}, \mathbf{p})d\mathbf{q}d\mathbf{p} = \pi(\mathbf{p}|\mathbf{q})\pi(\mathbf{q})d\mathbf{q}d\mathbf{p}$
- ▶ $\pi(\mathbf{q}, \mathbf{p})$ je hustota pravděpodobnosti na celém fázovém prostoru, $\pi(\mathbf{q}, \mathbf{p}) = \exp(-(V(\mathbf{q}) + K(\mathbf{q}, \mathbf{p}))) = \exp(-H(\mathbf{q}, \mathbf{p}))$. Funkci $H(\mathbf{q}, \mathbf{p})$ budeme nazývat *hamiltoniánem*.

Hamiltonovské Monte Carlo

- ▶ Mějme n -dimenzionální modelový prostor Q se souřadnicemi q^i , chceme navzorkovat $P(\cdot)$: $P(d\mathbf{q}) = \exp(-V(\mathbf{q}))d\mathbf{q} \Rightarrow$ PDF: $\pi(\mathbf{q}) = \exp(-V(\mathbf{q}))$. Funkci $V(\mathbf{q})$ budeme říkat *potenciální* energie.
- ▶ Rozšířme Q na $2n$ -dimenzionální fázový prostor M přidáním n hybností p_i , $i = 1, 2, \dots, n$:

$$(q^i) \rightarrow (q^i, p_i) \text{ tj. } (\mathbf{q}) \rightarrow (\mathbf{q}, \mathbf{p})$$

- ▶ $\mathbf{p}$ dostaneme jako $\mathbf{p} \sim T(\cdot)$, kde $T(d\mathbf{p}) = \exp(-K(\mathbf{q}, \mathbf{p}))d\mathbf{p} \Rightarrow$ PDF: $\pi(\mathbf{p}|\mathbf{q}) = \exp(-K(\mathbf{q}, \mathbf{p}))$. Funkci $K(\mathbf{q}, \mathbf{p})$, kterou můžeme zvolit libovolně, budeme říkat *kinetická* energie.
- ▶ Kombinací $P(\cdot)$ a $T(\cdot)$ dostaneme pravděpodobnost na celém fázovém prostoru: Kanonický rozklad
 $\pi(\mathbf{q}, \mathbf{p})d\mathbf{q}d\mathbf{p} = \pi(\mathbf{p}|\mathbf{q})\pi(\mathbf{q})d\mathbf{q}d\mathbf{p}$
- ▶ $\pi(\mathbf{q}, \mathbf{p})$ je hustota pravděpodobnosti na celém fázovém prostoru, $\pi(\mathbf{q}, \mathbf{p}) = \exp(-(V(\mathbf{q}) + K(\mathbf{q}, \mathbf{p}))) = \exp(-H(\mathbf{q}, \mathbf{p}))$. Funkci $H(\mathbf{q}, \mathbf{p})$ budeme nazývat *hamiltoniánem*.



Hamiltonovské Monte Carlo

- ▶ Mějme n -dimenzionální modelový prostor Q se souřadnicemi q^i , chceme navzorkovat $P(\cdot)$: $P(d\mathbf{q}) = \exp(-V(\mathbf{q}))d\mathbf{q} \Rightarrow$ PDF: $\pi(\mathbf{q}) = \exp(-V(\mathbf{q}))$. Funkci $V(\mathbf{q})$ budeme říkat *potenciální* energie.
- ▶ Rozšířme Q na $2n$ -dimenzionální fázový prostor M přidáním n hybností p_i , $i = 1, 2, \dots, n$:

$$(q^i) \rightarrow (q^i, p_i) \text{ tj. } (\mathbf{q}) \rightarrow (\mathbf{q}, \mathbf{p})$$

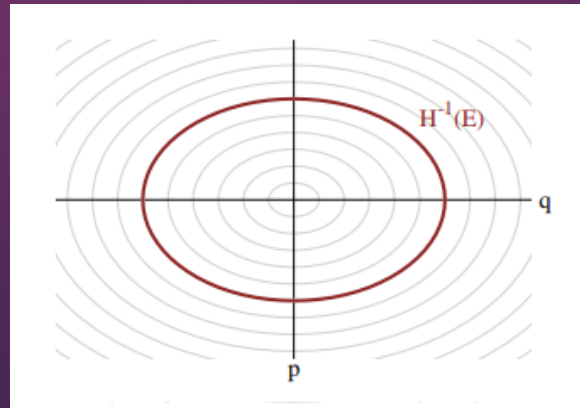
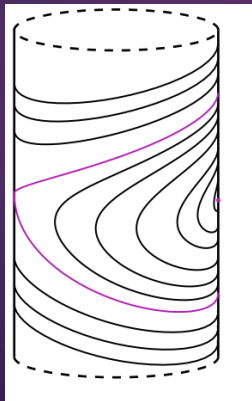
- ▶ $\mathbf{p}$ dostaneme jako $\mathbf{p} \sim T(\cdot)$, kde $T(d\mathbf{p}) = \exp(-K(\mathbf{q}, \mathbf{p}))d\mathbf{p} \Rightarrow$ PDF: $\pi(\mathbf{p}|\mathbf{q}) = \exp(-K(\mathbf{q}, \mathbf{p}))$. Funkci $K(\mathbf{q}, \mathbf{p})$, kterou můžeme zvolit libovolně, budeme říkat *kinetická* energie.
- ▶ Kombinací $P(\cdot)$ a $T(\cdot)$ dostaneme pravděpodobnost na celém fázovém prostoru: *Kanonický rozklad*
$$\pi(\mathbf{q}, \mathbf{p})d\mathbf{q}d\mathbf{p} = \pi(\mathbf{p}|\mathbf{q})\pi(\mathbf{q})d\mathbf{q}d\mathbf{p}$$
- ▶ $\pi(\mathbf{q}, \mathbf{p})$ je hustota pravděpodobnosti na celém fázovém prostoru, $\pi(\mathbf{q}, \mathbf{p}) = \exp(-(V(\mathbf{q}) + K(\mathbf{q}, \mathbf{p}))) = \exp(-H(\mathbf{q}, \mathbf{p}))$. Funkci $H(\mathbf{q}, \mathbf{p})$ budeme nazývat *hamiltoniánem*.
- 1. Po vyzdvihnutí $(\mathbf{q}) \rightarrow (\mathbf{q}, \mathbf{p})$ necháme $(\mathbf{q}, \mathbf{p})$ se vyvíjet pomocí Hamiltonových rovnic s hamiltoniánem H . Získáme tak vzorky $(\mathbf{q}_k, \mathbf{p}_k)$. Protože se při toku zachovává H , budeme kroužit na level setech $\{\mathbf{q}, \mathbf{p}: H(\mathbf{q}, \mathbf{p}) = E\}$.

Hamiltonovské Monte Carlo

- ▶ Mějme n -dimenzionální modelový prostor Q se souřadnicemi q^i , chceme navzorkovat $P(\cdot)$: $P(d\mathbf{q}) = \exp(-V(\mathbf{q}))d\mathbf{q} \Rightarrow$ PDF: $\pi(\mathbf{q}) = \exp(-V(\mathbf{q}))$. Funkci $V(\mathbf{q})$ budeme říkat *potenciální* energie.
- ▶ Rozšířme Q na $2n$ -dimenzionální fázový prostor M přidáním n hybností p_i , $i = 1, 2, \dots, n$:

$$(q^i) \rightarrow (q^i, p_i) \text{ tj. } (\mathbf{q}) \rightarrow (\mathbf{q}, \mathbf{p})$$

- ▶ $\mathbf{p}$ dostaneme jako $\mathbf{p} \sim T(\cdot)$, kde $T(d\mathbf{p}) = \exp(-K(\mathbf{q}, \mathbf{p}))d\mathbf{p} \Rightarrow$ PDF: $\pi(\mathbf{p}|\mathbf{q}) = \exp(-K(\mathbf{q}, \mathbf{p}))$. Funkci $K(\mathbf{q}, \mathbf{p})$, kterou můžeme zvolit libovolně, budeme říkat *kinetická* energie.
- ▶ Kombinací $P(\cdot)$ a $T(\cdot)$ dostaneme pravděpodobnost na celém fázovém prostoru: $\pi(\mathbf{q}, \mathbf{p})d\mathbf{q}d\mathbf{p} = \pi(\mathbf{p}|\mathbf{q})\pi(\mathbf{q})d\mathbf{q}d\mathbf{p}$ Kanonický rozklad
- ▶ $\pi(\mathbf{q}, \mathbf{p})$ je hustota pravděpodobnosti na celém fázovém prostoru, $\pi(\mathbf{q}, \mathbf{p}) = \exp(-(V(\mathbf{q}) + K(\mathbf{q}, \mathbf{p}))) = \exp(-H(\mathbf{q}, \mathbf{p}))$. Funkci $H(\mathbf{q}, \mathbf{p})$ budeme nazývat *hamiltoniánem*.
- 1. Po vyzdvihnutí $(\mathbf{q}) \rightarrow (\mathbf{q}, \mathbf{p})$ necháme $(\mathbf{q}, \mathbf{p})$ se vyvíjet pomocí Hamiltonových rovnic s hamiltoniánem H . Získáme tak vzorky $(\mathbf{q}_k, \mathbf{p}_k)$. Protože se při toku zachovává H , budeme kroužit na level setech $\{\mathbf{q}, \mathbf{p}: H(\mathbf{q}, \mathbf{p}) = E\}$.



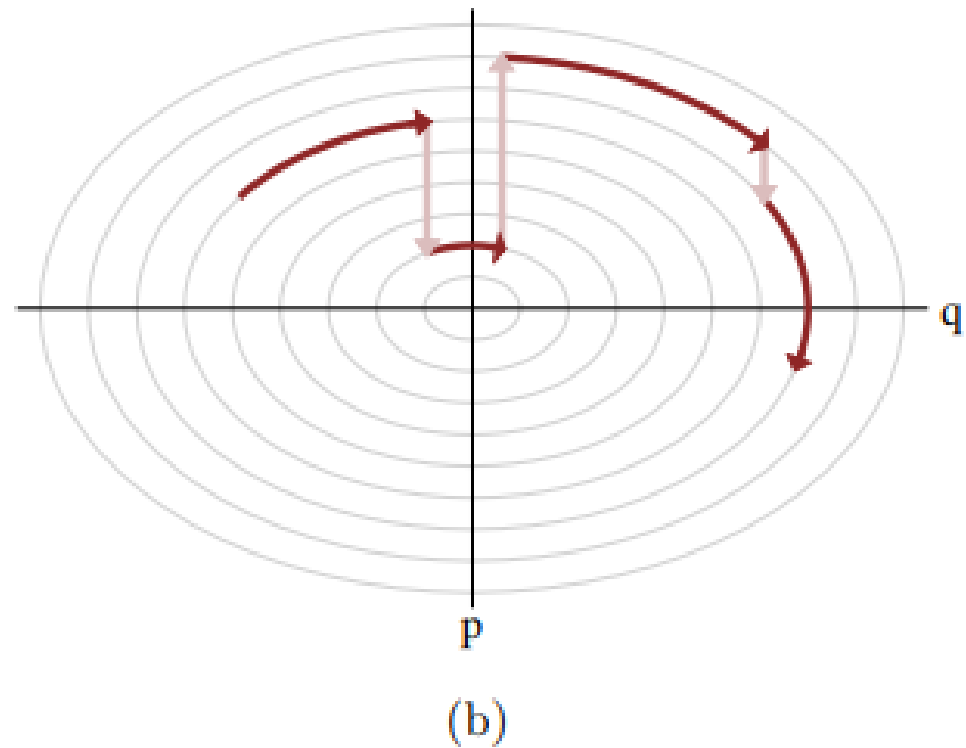
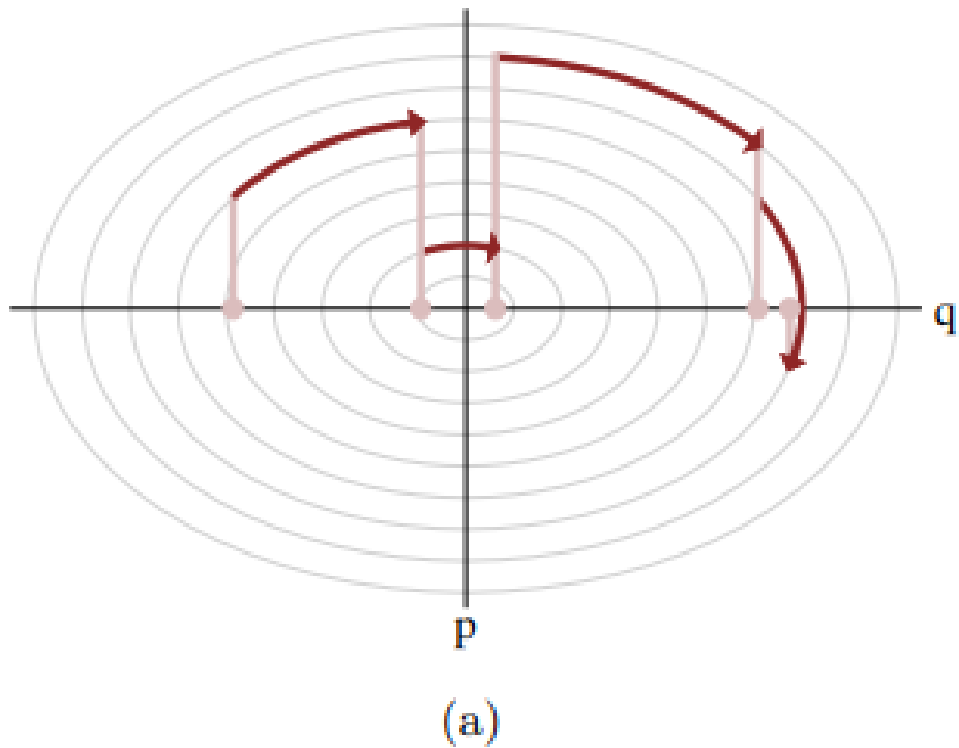
Hamiltonovské Monte Carlo

- ▶ Mějme n -dimenzionální modelový prostor Q se souřadnicemi q^i , chceme navzorkovat $P(\cdot)$: $P(d\mathbf{q}) = \exp(-V(\mathbf{q}))d\mathbf{q} \Rightarrow$ PDF: $\pi(\mathbf{q}) = \exp(-V(\mathbf{q}))$. Funkci $V(\mathbf{q})$ budeme říkat *potenciální* energie.
- ▶ Rozšířme Q na $2n$ -dimenzionální fázový prostor M přidáním n hybností p_i , $i = 1, 2, \dots, n$:

$$(q^i) \rightarrow (q^i, p_i) \text{ tj. } (\mathbf{q}) \rightarrow (\mathbf{q}, \mathbf{p})$$

- ▶ $\mathbf{p}$ dostaneme jako $\mathbf{p} \sim T(\cdot)$, kde $T(d\mathbf{p}) = \exp(-K(\mathbf{q}, \mathbf{p}))d\mathbf{p} \Rightarrow$ PDF: $\pi(\mathbf{p}|\mathbf{q}) = \exp(-K(\mathbf{q}, \mathbf{p}))$. Funkci $K(\mathbf{q}, \mathbf{p})$, kterou můžeme zvolit libovolně, budeme říkat *kinetická* energie.
 - ▶ Kombinací $P(\cdot)$ a $T(\cdot)$ dostaneme pravděpodobnost na celém fázovém prostoru: *Kanonický rozklad*
 $\pi(\mathbf{q}, \mathbf{p})d\mathbf{q}d\mathbf{p} = \pi(\mathbf{p}|\mathbf{q})\pi(\mathbf{q})d\mathbf{q}d\mathbf{p}$
 - ▶ $\pi(\mathbf{q}, \mathbf{p})$ je hustota pravděpodobnosti na celém fázovém prostoru, $\pi(\mathbf{q}, \mathbf{p}) = \exp(-(V(\mathbf{q}) + K(\mathbf{q}, \mathbf{p}))) = \exp(-H(\mathbf{q}, \mathbf{p}))$. Funkci $H(\mathbf{q}, \mathbf{p})$ budeme nazývat *hamiltoniánem*.
1. Po vyzdvihnutí $(\mathbf{q}) \rightarrow (\mathbf{q}, \mathbf{p})$ necháme $(\mathbf{q}, \mathbf{p})$ se vyvíjet pomocí Hamiltonových rovnic s hamiltoniánem H . Získáme tak vzorky $(\mathbf{q}_k, \mathbf{p}_k)$. Protože se při toku zachovává H , budeme kroužit na level setech $\{\mathbf{q}, \mathbf{p}: H(\mathbf{q}, \mathbf{p}) = E\}$.
 2. Po nějaké době promítneme zpátky $(\mathbf{q}, \mathbf{p}) \rightarrow (\mathbf{q})$
 3. $(\mathbf{q})$ vyzdvihneme pomocí kinetické energie na nový bod ve fázovém prostoru.
 4. Opakujeme 1,2,3 dokola. Díky tomu, že míra je vůči tomuto přechodu stacionární, dostaneme vzorky z $\pi(\mathbf{q}, \mathbf{p})d\mathbf{q}d\mathbf{p}$.

- a) Výzdvih $\rightarrow$ tok $\rightarrow$ projekce.
- b) Složení projekce a výzdvihu můžeme interpretovat jako skok mezi jednotlivými level sety $H(\mathbf{q}, \mathbf{p})$.



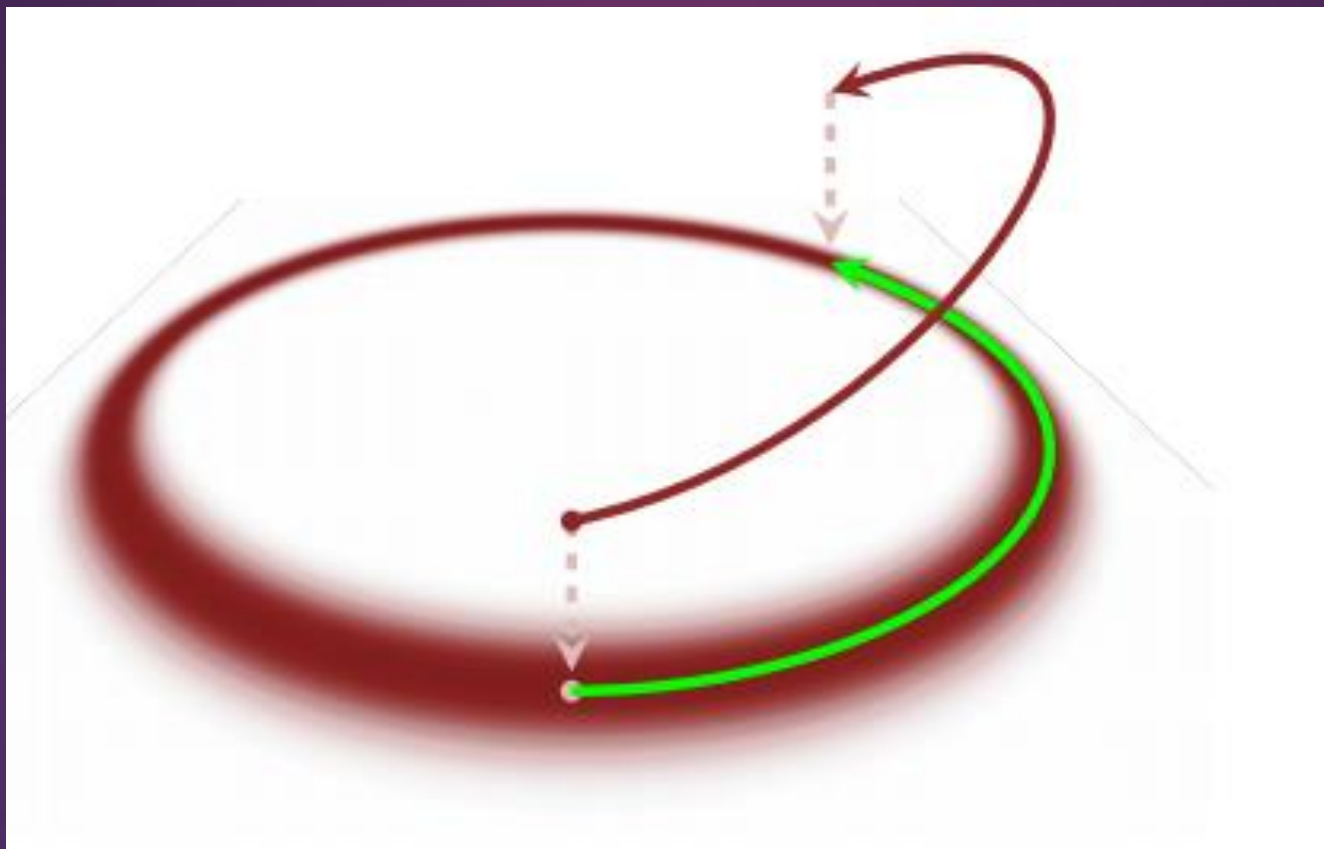
Hamiltonovské Monte Carlo

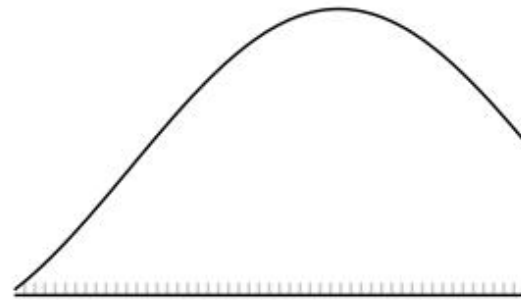
- ▶ Mějme n -dimenzionální modelový prostor Q se souřadnicemi q^i , chceme navzorkovat $P(\cdot)$: $P(d\mathbf{q}) = \exp(-V(\mathbf{q}))d\mathbf{q} \Rightarrow$ PDF: $\pi(\mathbf{q}) = \exp(-V(\mathbf{q}))$. Funkci $V(\mathbf{q})$ budeme říkat *potenciální* energie.
- ▶ Rozšířme Q na $2n$ -dimenzionální fázový prostor M přidáním n hybností p_i , $i = 1, 2, \dots, n$:

$$(q^i) \rightarrow (q^i, p_i) \text{ tj. } (\mathbf{q}) \rightarrow (\mathbf{q}, \mathbf{p})$$

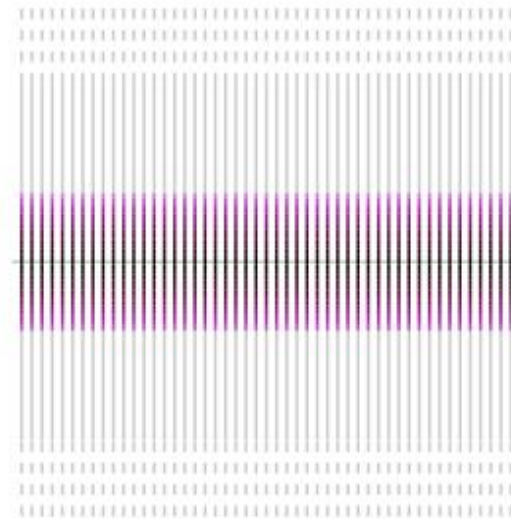
- ▶ $\mathbf{p}$ dostaneme jako $\mathbf{p} \sim T(\cdot)$, kde $T(d\mathbf{p}) = \exp(-K(\mathbf{q}, \mathbf{p}))d\mathbf{p} \Rightarrow$ PDF: $\pi(\mathbf{p}|\mathbf{q}) = \exp(-K(\mathbf{q}, \mathbf{p}))$. Funkci $K(\mathbf{q}, \mathbf{p})$, kterou můžeme zvolit libovolně, budeme říkat *kinetická* energie.
- ▶ Kombinací $P(\cdot)$ a $T(\cdot)$ dostaneme pravděpodobnost na celém fázovém prostoru:
$$\pi(\mathbf{q}, \mathbf{p})d\mathbf{q}d\mathbf{p} = \pi(\mathbf{p}|\mathbf{q})\pi(\mathbf{q})d\mathbf{q}d\mathbf{p}$$
 Kanonický rozklad

- ▶ $\pi(\mathbf{q}, \mathbf{p})$ je hustota pravděpodobnosti na celém fázovém prostoru, $\pi(\mathbf{q}, \mathbf{p}) = \exp(-(V(\mathbf{q}) + K(\mathbf{q}, \mathbf{p}))) = \exp(-H(\mathbf{q}, \mathbf{p}))$. Funkci $H(\mathbf{q}, \mathbf{p})$ budeme nazývat *hamiltoniánem*.
1. Po vyzdvihnutí $(\mathbf{q}) \rightarrow (\mathbf{q}, \mathbf{p})$ necháme $(\mathbf{q}, \mathbf{p})$ se vyvíjet pomocí Hamiltonových rovnic s hamiltoniánem H . Získáme tak vzorky $(\mathbf{q}_k, \mathbf{p}_k)$. Protože se při toku zachovává H , budeme kroužit na level setech $\{\mathbf{q}, \mathbf{p}: H(\mathbf{q}, \mathbf{p}) = E\}$.
 2. Po nějaké době promítneme zpátky $(\mathbf{q}, \mathbf{p}) \rightarrow (\mathbf{q})$
 3. $(\mathbf{q})$ vyzdvihneme pomocí kinetické energie na nový bod ve fázovém prostoru.
 4. Opakujeme 1,2,3 dokola. Díky tomu, že míra je vůči tomuto přechodu stacionární, dostaneme vzorky z $\pi(\mathbf{q}, \mathbf{p})d\mathbf{q}d\mathbf{p}$.
 5. Nakonec ignorováním $\mathbf{p}_k$ dostaneme vzorky marginální pravděpodobnosti $\pi(\mathbf{q})d\mathbf{q}$ - což je naše cílová pravděpodobnost.

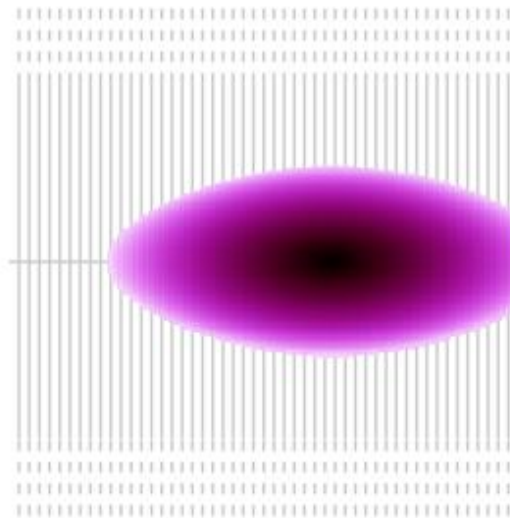




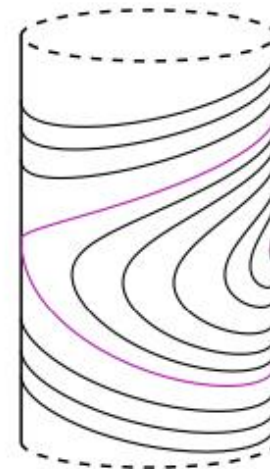
(a)



(b)

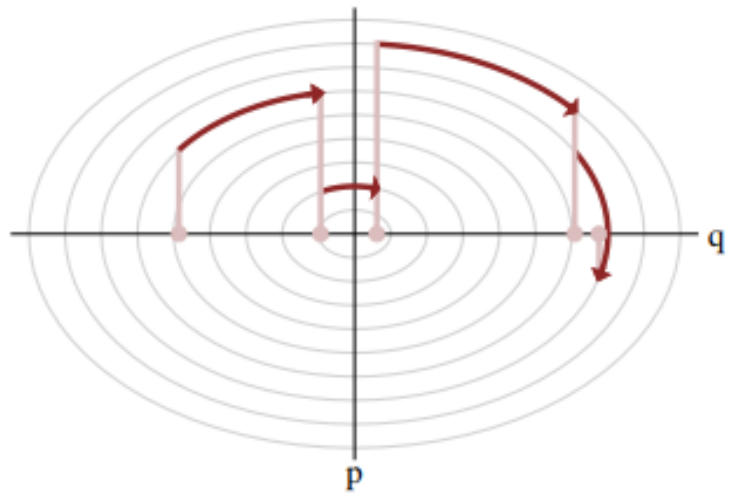


(c)

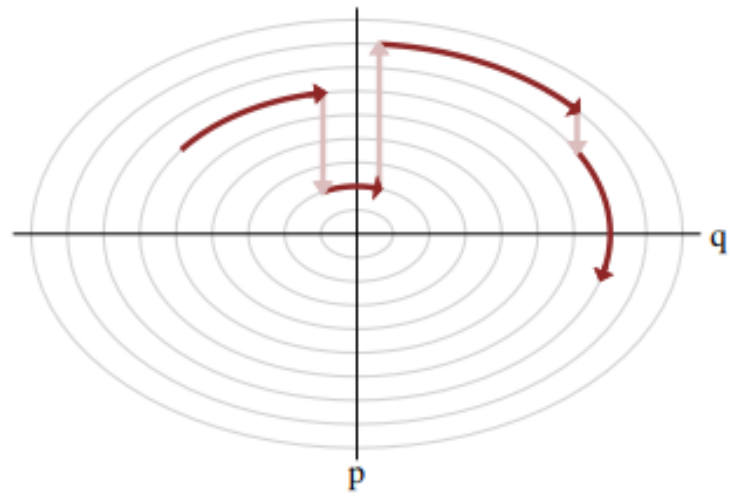


(d)

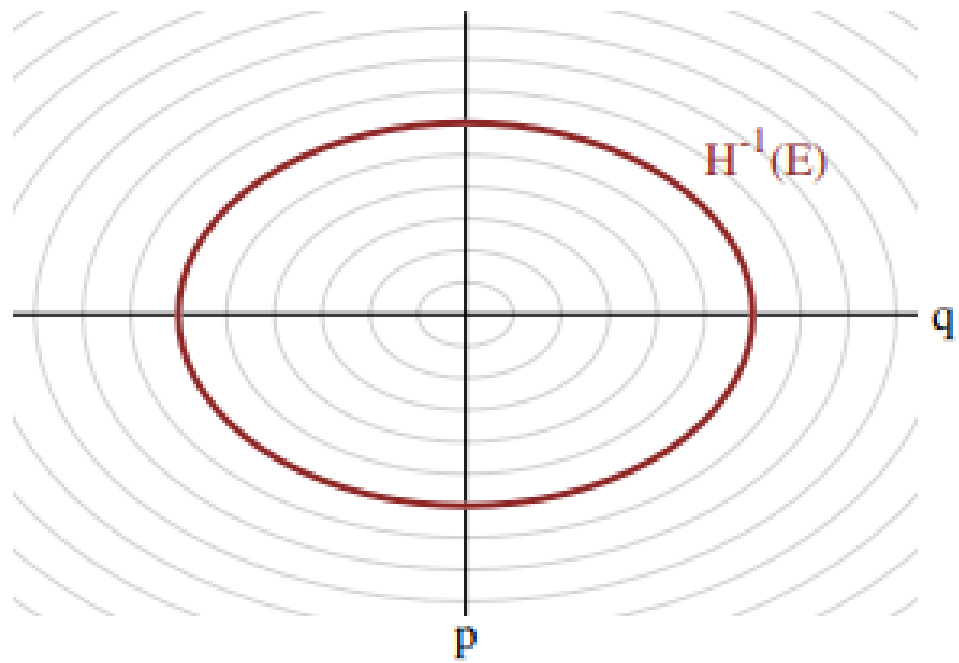
FIG 13. (a) *Hamiltonian Monte Carlo begins with a target measure on the base space, for example $Q = \mathbb{S}^1$.* (b) *The choice of a disintegration on the cotangent bundle, $T^*Q = \mathbb{S}^1 \times \mathbb{R}$, defines (c) a joint measure on the cotangent bundle which immediately defines (d) a Hamiltonian system given the canonical symplectic structure. The Hamiltonian flow of this system is then used to construct an efficient Markov transition.*



(a)



(b)

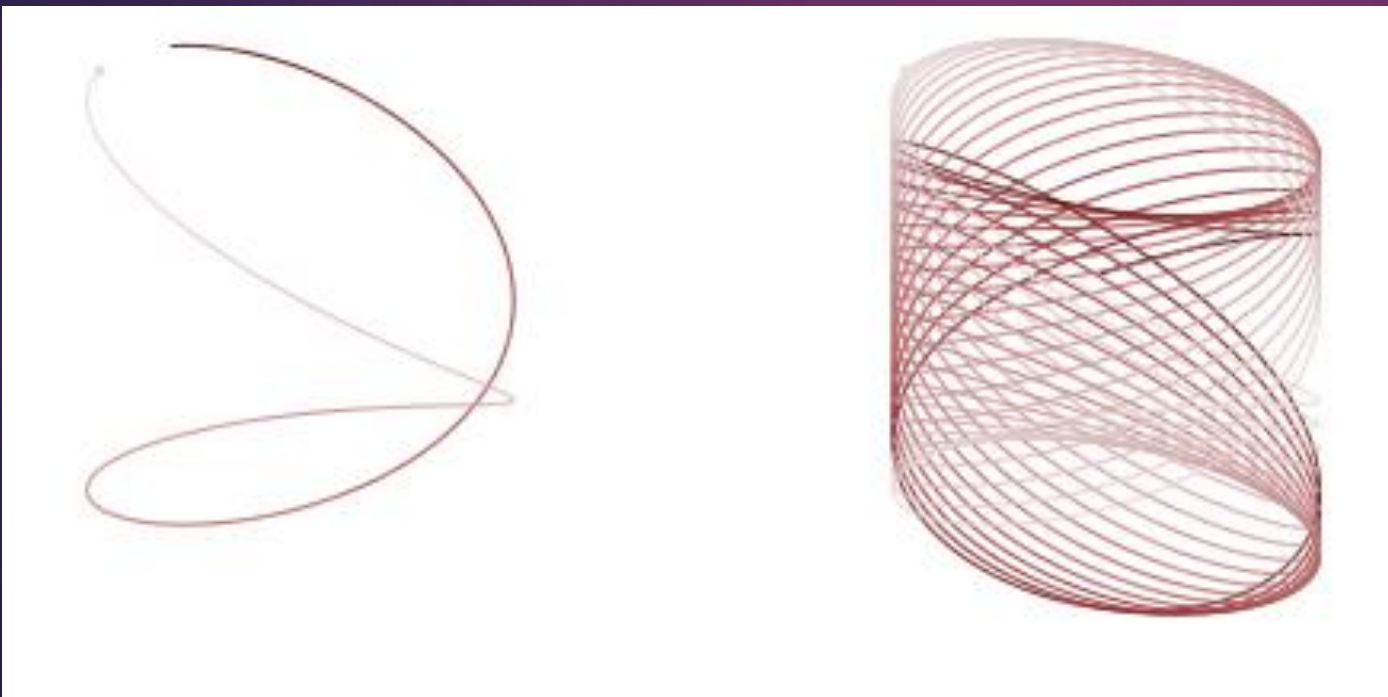


Optimalizace

- ▶ Po každém výzdvihu algoritmus prochází jeden level set hamiltoniánu, tj. celkové pravděpodobnosti. Tyto level sety vyplní celý fázový prostor (až na kritické body).
- ▶ Optimalizace algoritmu se proto dá přirozeně rozložit na optimalizaci deterministického procházení level setu a na optimalizaci skoků mezi hybnostmi.
- ▶ Formálně se joint pravděpodobnost rozloží pomocí *mikronanického rozkladu*: $\pi(q, p) = \pi(\theta_E | E) \pi(E)$, kde θ_E je level set o energii E . Obě distribuce vpravo se pak dají zkoumat zvlášť.

Optimalizace

Optimalizace procházení



Optimalizace skoků

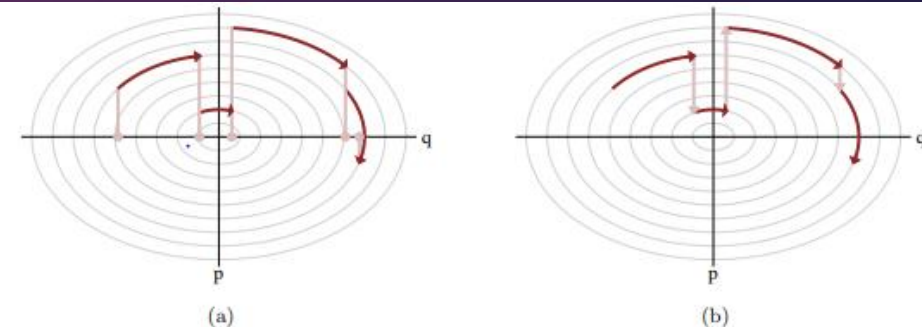


FIG 22. (a) Each Hamiltonian Markov transition lifts the initial state onto a random level set of the Hamiltonian, which can then be explored with a Hamiltonian trajectory before projecting back down to the target parameter space. (b) If we consider the projection and random lift steps as a single momentum resampling step, then the Hamiltonian Markov chain alternates between deterministic trajectories along these level sets (dark red) and a random walk across the level sets (light red).

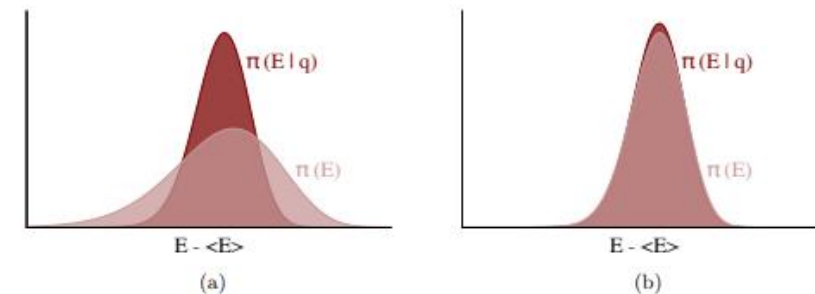


FIG 23. The momentum resampling in a Hamiltonian Markov transition randomly changes the energy, inducing a random walk between level sets. (a) When the energy transition distribution, $\pi(E | q)$ is narrow relative to the marginal energy distribution, $\pi(E)$, this random walk will explore the marginal energy distribution only very slowly, requiring many expensive transitions to survey all of the relevant energies. (b) On the other hand, when the two distributions are well-matched the random walk will explore the marginal energy distribution extremely efficiently.

HcMC korekce

- ▶ Pro úspěch metody je klíčové jak zachování míry při hamiltonovských tocích, tak zachování energie při oběhu.
- ▶ Pokud budeme Hamiltonovy rovnice integrovat nepřesně, tak se metoda zhroutí
- ▶ Explicitní ani implicitní Eulerova metoda není moc dobrá, lepší je tzv. *leap-frog* algoritmus
- ▶ Ale v praxi stejně budeme mít nějakou chybu
- ▶ Proto se používají pravděpodobnostní korekce – pomocí hamiltonovské dynamiky se navrhuje nové kroky a ty se pak podrobí HCMC korekci (viz Fichtner).

Výhody metody

- ▶ Efektivní prohledávání modelového prostoru
- ▶ Hamiltonián přirozeně zakódovává informace o geometrii pravděpodobnostního rozložení
- ▶ Solidní matematický základ, vše se dá jasně přeložit do řeči pravděpodobnosti a diferenciální geometrie.
- ▶ Usnadněná diagnostika

Nevýhody metody

- ▶ Potřeba mít k dispozici derivace
- ▶ V nemodifikované podobě funguje jen na unimodální PDF.
- ▶ Určitá volnost ve volbě kinetické energie (nejčastěji se ale používá nějaká kvadratická forma v $\mathbf{p}$).
- ▶ Ještě není jasné, jak optimálně volit časy integrace.
- ▶ V nemodifikované podobě stejně musíme používat McMC kvůli nepřesnosti v integraci

Aplikace metody na seismickou inverzi

Fichtner & Simutè (2017):

<http://geo.mff.cuni.cz/seismosoft/clanky/FichtnerSimute.JGR2018.pdf>



Zdroje:

- ▶ <https://arxiv.org/pdf/1410.5110.pdf> Betancourt 1
- ▶ <https://arxiv.org/pdf/1701.02434.pdf> Betancourt 2
- ▶ https://www.youtube.com/watch?v=pHsulaPbNbY&ab_channel=MLSSIceland2014 Betancourt přednáška
- ▶ <https://arxiv.org/pdf/1206.1901.pdf> Neal
- ▶ <https://arxiv.org/pdf/math/0404033.pdf> Roberts
- ▶ Radonova-Nikodymova věta – bakalářská práce, Tomáš Kuric (https://is.muni.cz/th/bb1k3/bachelor_thesis.pdf)
- ▶ Inverse Problem Theory, Albert Tarantola:
<http://www.ipgp.fr/~tarantola/Files/Professional/Books/InverseProblemTheory.pdf>
- ▶ Reichl – Encyklopedie fyziky (<http://fyzika.jreichl.com/>)
- ▶ Google Images
- ▶ Wikipedia

Marginální a podmíněná pravděpodobnost

Expektance

$$\mathbb{E}[f(\theta)] = \int d\theta f(\theta) \pi(\theta|\mathcal{D})$$

Marginální

$$\pi(\theta_2, \dots, \theta_n|\mathcal{D}) = \int d\theta_1 \pi(\theta|\mathcal{D})$$

Podmíněná

$$\pi(\theta_2, \dots, \theta_n|\theta_1, \mathcal{D}) = \frac{\pi(\theta|\mathcal{D})}{\int d\theta_1 \pi(\theta|\mathcal{D})}$$

Lebesgueova míra a Lebesgueova σ -algebra

Definition [\[edit \]](#)

Given a subset $E \subseteq \mathbb{R}$, with the length of [interval](#) $I = [a, b]$ (or $I = (a, b)$) given by $\ell(I) = b - a$, the Lebesgue [outer measure](#) [\[3\]](#) $\lambda^*(E)$ is defined as

$$\lambda^*(E) = \inf \left\{ \sum_{k=1}^{\infty} \ell(I_k) : (I_k)_{k \in \mathbb{N}} \text{ is a sequence of open intervals with } E \subseteq \bigcup_{k=1}^{\infty} I_k \right\}.$$

The Lebesgue measure is defined on the Lebesgue [\$\sigma\$ -algebra](#), which is the collection of all sets E which satisfy the condition that, for every $A \subseteq \mathbb{R}$,

$$\lambda^*(A) = \lambda^*(A \cap E) + \lambda^*(A \cap E^c)$$

For any set in the Lebesgue σ -algebra, its Lebesgue measure is given by its Lebesgue outer measure $\lambda(E) = \lambda^*(E)$.

Sets that are not included in the Lebesgue σ -algebra are not Lebesgue-measurable. Such sets do exist, i.e., the Lebesgue σ -algebra is strictly contained in the [power set](#) of $\mathbb{R}$.