Teorie pravděpodobnosti a teorie informace

(opakování s akcentem na potřeby teorie informace)

Ω - neprázdná množina

$\mathcal{P}(\Omega)$ - množina všech podmnožin množiny Ω

<u>Definice</u> Pravděpodobnostní prostor je trojice $(\Omega, \mathcal{A}, P)$, kde

(i) $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ je σ-algebra tj.

(i-1) $\emptyset \in \Omega$ $\qquad \forall A \in \mathcal{A}$

(i-2) $A \in \mathcal{A} \Rightarrow A^c = \Omega \setminus A \in \mathcal{A}$

(i-3) Je-li $A_1, A_2, \ldots$ posloupnost prvků z $\mathcal{A}$, pak

$$\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}.$$

(ii) $P : \mathcal{A} \to \langle 0, 1 \rangle$

(ii-1) $P(\Omega) = 1$

(ii-2) $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$

kdykoliv je $A_1, A_2, \ldots$ disjunktní posloupnost prvků z $\mathcal{A}$.

Příklad: $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$

$\mathcal{A} = \mathcal{P}(\Omega)$

Každá pravděpodobnost P je zadána n-ticí

$p_1 = P(\{\omega_1\})$

$p_2 = P(\{\omega_2\})$

$p_n = P(\{\omega_n\})$,

kde $0 \leq p_i \leq 1$ a $\sum_{i=1}^{n} p_i = 1$

· $\vec{p} = (p_1, \dots, p_n)$ se nazývá pravděpodobnostní vektor.

Pak $P(A) = \sum_{\{i \mid \omega_i \in A\}} p_i$, $A \subseteq \Omega$.

V teorii informace

Ω = konečná množina znaků (abeceda)

$\Omega = \{a, b, c, \dots\}$

Rovnoměrné rozdělení

$\vec{p} = (1/n, \dots, 1/n)$

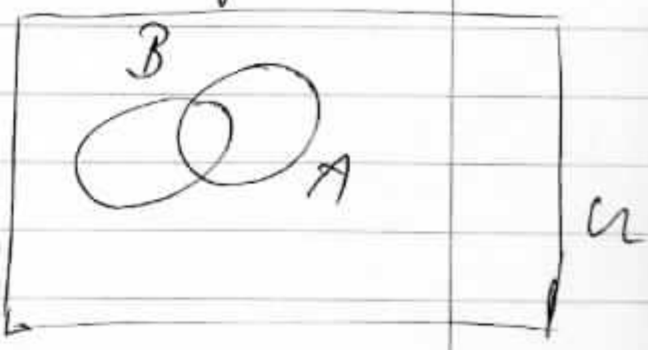
Diracovo rozdělení

$\vec{p} = (0, \dots, 1, 0, \dots)$

Podmíněná pravděpodobnost

$(\Omega, \mathcal{A}, P)$ – pravděpodobnostní prostor

$B \in \mathcal{A}$; $P(B) > 0$.



$$P(A|B) = \frac{P(A \wedge B)}{P(B)}$$

tj. $P(A \wedge B) = P(B) \cdot P(A|B)$

vytváří další pravděpodobnostní prostor s pravděpodobností

$$P_B : \mathcal{A} \to \langle 0, 1 \rangle$$

$$P_B(A) = P(A|B).$$

Nezávislé jevy:

$A, B \in \mathcal{A}$

$$P(A \wedge B) = P(A) \cdot P(B)$$

neboli $P(A) = P(A|B)$

je-li $P(B) > 0$

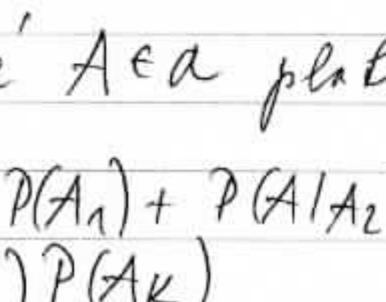
Věta o úplné pravděpodobnosti

Jsou-li $A_1, \dots, A_k$ disjunktní množiny v $\mathcal{A}$ takové, že

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = 1$$

pak pro každé $A \in \mathcal{A}$ platí

$$P(A) = P(A|A_1) \cdot P(A_1) + P(A|A_2)\,P(A_2) + \dots$$
$$\dots + P(A|A_k)\,P(A_k)$$



odvození: $P(A) = P(A \cap A_1) + P(A \cap A_2) + \dots + P(A \cap A_k)$
$= P(A|A_1)P(A_1) + \dots + P(A|A_k)P(A_k)$

řetězce symbolů jako pravděpodobnostní prostor + konzistentní systém pravděpodobností

Δ … konečná abeceda $|\Delta| = n$

$\Delta^2 = \Delta \times \Delta$ … řetězce symbolů z Δ délky 2

(např: ab, eh, …)

$\Delta^k = \underbrace{\Delta \times \dots \times \Delta}_{k \times \text{krát}}$ … řetězce délky k

$\Delta^\infty = \Delta \times \Delta \cdots$ … nekonečné řetězce = posloupnosti symbolů.

např. binární abeceda: $\Delta = \{0, 1\}$

Řetězce též můžeme modelovat stromy

např. 0010

Nějaké zadány pravděpodobnosti
$P_1, P_2, \ldots$
na řetězcích $\Delta, \Delta^2, \Delta^3$

konzistence znamená například
$$P_1\{0\} = P_2\{00\} + P_2\{01\}$$

<u>podmínka konzistence:</u>

$$P_\ell\{a_1 a_2 \ldots a_\ell\} = \sum_{a_{\ell+1} \ldots a_m \in \Delta^{m-\ell}} P_m\{a_1 \ldots a_\ell a_{\ell+1} \ldots a_m\}$$

$$\forall \ell \leq m$$

" pravděp. v uzlu = součet na všech následných větvičkách

Konstrukce pravděpodobnostního prostoru zahrnujícího tento rámec

$\Omega = \Delta^\infty$

Cylindrické množiny: posloupnosti mající fixní prefix

např.

$C(\langle a_1, a_2 \rangle) = \{ (x_i \in \Delta^\infty) \mid x_1 = a_1, x_2 = a_2 \}$

$a_1 a_2$ ztotožníme $C(\langle a_1 a_2 \rangle)$

$\mathcal{A}$ = nejmenší σ-algebra obsahující všechny cylindrické množiny

Dají se nějak pravděpodobnosti $P_1, P_2, \dots$ „slepit", tak abychom dostali pravděpodobnost „na" Ω

<u>Věta</u>: Kolmogorov

Nechť Δ je konečná abeceda a $P_1, P_2, \dots$ je konzistentní posloupnost pravděpodobností na na $\Delta, \Delta^2, \dots$

Pak existuje právě jedna pravděpodobnost P na $\mathcal{A}$ tak, že

$$P_k \langle a_1 a_2 \dots a_k \rangle = P(C \langle a_1 a_2 \dots a_k \rangle)$$

Příklad:

$\Delta = \{0,1\}$; $\Omega = \Delta^\infty$ = posloupnosti nul a jedniček
"binární kód, házení mincí"

$P_1\{1\} = p$

$P_1\{0\} = 1-p \qquad 0 < p < 1$

$P_K\{\omega_1 \dots \omega_K\} = p^{\text{počet jedniček}} (1-p)^{\text{počet nul}}$

Tím dostaneme konzistentní systém pravděpodobností! $\boxed{\text{(Cvičení) Návod: } [p+(1-p)]^K = 1}$

Ω = binární posloupnosti $\equiv [0,1)$

$\omega \in [0,1)$

$$\omega = \omega_1 \cdot \frac{1}{2} + \omega_2 \cdot \frac{1}{4} + \omega_3 \cdot \frac{1}{8} + \dots$$
$$= \sum_{i=1}^{\infty} \frac{\omega_i}{2^i}$$

Jaká je pravděpodobnost ω : $P\{\omega\}$?

Víme si

$$P(\underbrace{C\{\omega_1 \dots \omega_m\}}_{=A_m}) = p^{\text{počet } 1} \cdot (1-p)^{\text{počet } 0.}$$

$A_m \downarrow 0 \qquad \bigcap_{m=1}^{\infty} A_m = \{\omega\}$

$P\{\omega\} = \lim_{m\to\infty} P(A_m)$

$P(A_m) \leq \max\{p, 1-p\}^m \to 0$

$\Rightarrow P\{\omega\} = 0.$

Vynecháme-li posloupnosti s ocasem 11… + periodické máme jednoznačnou korespondenci s $[0,1)$ co vynechává vůči mince 0.

Bod má nulovou pravděpodobnost
Je to paradox?

$C\langle 01 \rangle \quad \alpha = \underbrace{0 \cdot 1/2}_{\text{dolní interval}} + \underbrace{1 \cdot \frac{1}{4}}_{\text{horní interval}} + \underbrace{\ldots}_{\text{cokoliv}}$

$\le \frac{1}{8} + \frac{1}{16} + \ldots = 1/8 \cdot \frac{1}{1 - 1/2} = 1/4$

Tedy $C\langle 01 \rangle = [1/4, 1/2]$

$P[C\langle 01 \rangle] = (1-p)p$

Pro férovou minci $p = 1/2$ je

$P[1/4, 1/2] = \frac{1}{4}$ ... délka intervalu

V tomto případě vede konstrukce k Lebesgueově integrálu.

Zobecnění: Bernoulliovo schéma

Δ = konečná abeceda

P_1 ... pravděpodobnost na Δ

konzistentní systém

$P_K(\omega_1 \ldots \omega_K) = P_1\langle \omega_1 \rangle P_1\langle \omega_2 \rangle \ldots P_1\langle \omega_K \rangle$

- modelují posloupnost nezávislých pokusů.

Náhodná veličina

"Náhodná veličina není ani náhodná ani veličina"

<u>Def</u>: Náhodná veličina je funkce X na pravděpodobnostním prostoru $(\Omega, \mathcal{A}, P)$ taková, že

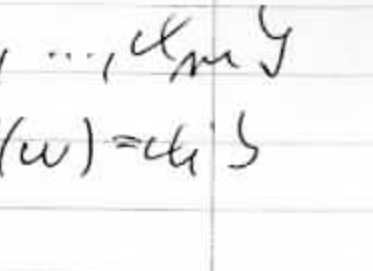
$$\{\omega \in \Omega \mid X(\omega) \in I\} \in \mathcal{A}$$

pro každý interval $I \subseteq \mathbb{R}$.

Zkratky: $P[X = a] = P(a) = p(a)$, atd...

Náhodná veličina s konečně mnoha hodnotami:

$\mathcal{X}$ - množina hodnot $\{a_1, a_2, \dots, a_m\}$



$\Omega_i = \{\omega \in \Omega \mid X(\omega) = a_i\}$

disjunktní rozklad Ω

střední hodnota

$$EX = a_1 P(\Omega_1) + a_2 P(\Omega_2) + \dots + a_m P(\Omega_m)$$

Pravděpodobnostní funkce, tabulka

x_1	x_2	$\dots$	x_m
p_1	p_2		p_m

kde $\vec{p} = (p_1, p_2, \dots, p_m)$ je pravděpodobnostní vektor.

---

náhodný vektor $\vec{X}$ na prostoru $(\Omega, \mathcal{A}, P)$ je n-tice náhodných veličin

$$\vec{X} = (X_1, \dots, X_n)$$

X … n.v. s konečnou množinou hodnot $\mathcal{X}$

Y … n.v. s konečnou množinou hodnot $\mathcal{Y}$

Pak (X, Y) má také konečně mnoho hodnot v $\mathcal{X} \times \mathcal{Y}$.

pravděp. funkce $p(x, y) : \mathcal{X} \times \mathcal{Y}$

$$p(x, y) = P[X = x \wedge Y = y]$$

$x \in \mathcal{X}$; $y \in \mathcal{Y}$.

Matice $(p(x,y))$ je symetrická.
Součet řádkový dá rozdělení X
a sloupcový rozdělení Y.

Náhodné veličiny X a Y s konečnou množinou hodnot jsou <u>nezávislé</u> jestliže

$$p(x,y) = p(x)p(y)$$

$$\forall x \in \mathcal{X}, y \in \mathcal{Y}.$$

<u>Časová řada</u> (diskrétní stochastický proces) je posloupnost

$$X_1, X_2, \ldots$$

náhodných veličin na pravděpodobnostním prostoru $(\Omega, \mathcal{A}, P)$.

<u>Bernoulliovo schéma</u> – opakování!

Δ – konečná abeceda

$X : (\Omega, \mathcal{A}, P) \to \Delta$

$P_K : \Delta^K \to [0,1]$

$P_K \langle x_1 \ldots x_K \rangle = P(x_1)P(x_2) \ldots P(x_K)$

Součinový prostor: $(\Delta^\infty, \mathcal{A}, P)$

X_i – projekce na i-tou souřadnici

$X_i \langle x_1 x_2 \ldots \rangle = x_i.$

Pak $X_1, X_2, \dots$ jsou nezávislé stejně rozdělené náhodné veličiny.

<u>Odvození</u>: $P[X_1 = x_1 \wedge X_2 = x_2] =$
$= P[C\{x_1, x_2\}] = P(x_1) \cdot P(x_2)$.

<u>Matice přechodu</u> (závislé náhodné veličiny)

$(\Omega, \mathcal{A}, P)$ – prav. prostor

$X \dots$ hodnoty $\mathcal{X} = \{x_1, \dots, x_n\}$
$Y \dots$ hodnoty $\mathcal{Y} = \{y_1, \dots, y_m\}$

přechodové pravděpodobnosti:

$$p_{ij} = P[Y = y_j / X = x_i]$$

X	Y
i-tý stav	j-tý stav

(p_{ij}) – matice $n \times m$

řádkové součty jsou 1

$$\sum_j P[Y = y_j / X = x_i] = \sum_j \frac{P[Y = y_j, X = x_i]}{P(x_i)} = \frac{P(x_i)}{P(x_i)} = 1$$

Řekneme, že X má pravděpodobnostní vektor

$$\vec{p} = (p_1, \dots, p_m)$$

tj. $P[X = x_i] = p_i$, $i = 1, \dots, m$.

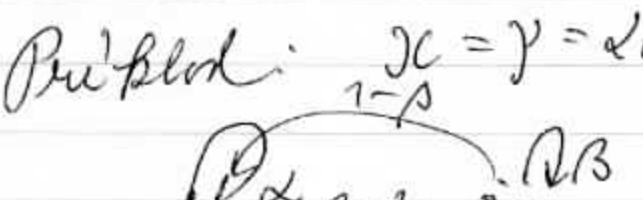
Jak spočítat pravděp. vektor $\vec{q} = (q_1, \dots, q_m)$ pro Y?

$$P[Y = y_j] = \sum_i P[Y = y_j / X = x_i] \, P[X = x_i]$$
$$= \sum_i p_{ij} \, p_i$$

maticově

$$\vec{q} = (q_1, \dots, q_m) = (p_1 \dots p_m) \underbrace{\begin{pmatrix} p_{ij} \end{pmatrix}}_{P}$$
$$\vec{q} = \vec{p} \cdot P$$

Příklad: $\mathcal{X} = \mathcal{Y} = \{0, 1\}$



X	Y
0	0
1	1

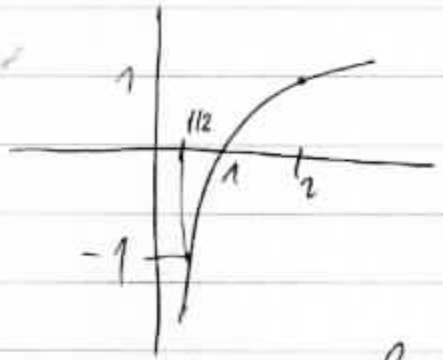
matice přechodu $\begin{pmatrix} \alpha & 1-\alpha \\ 1-\beta & \beta \end{pmatrix}$

$$(q_1, q_2) = (p_1, p_2) \begin{pmatrix} \alpha & 1-\alpha \\ 1-\beta & \beta \end{pmatrix} !$$
$$= (\alpha p_1 + (1-\beta) p_2, (1-\alpha) p_1 + \beta p_2)$$

směs rozdělení $(\alpha, 1-\alpha)$ $(1-\beta, \beta)$

Entropie

konvence: $\log x = \log_2 x$

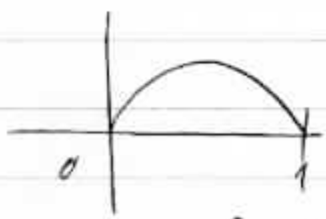


$\log_D x$, $D>0$, obecný základ

$$\boxed{\log_D x = \log x \cdot \log_D 2}$$

Odvození: $2^{\log x} = x \Rightarrow \log x \cdot \log_D 2 = \log_D x$

$-x \log x$:



$$\lim_{x\to 0} x \log x = \lim_{x\to 0} \frac{\log x}{1/x} = \lim_{x\to 0} \frac{\frac{\ln x}{\ln 2}}{1/x}$$

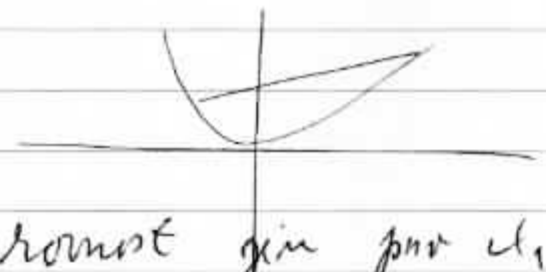
$$= \frac{1}{\ln 2} \cdot \lim_{x\to 0} \frac{\frac{1}{x}}{-1/x^2} = 0.$$

konvence $\boxed{0 \log 0 = 0.}$

Def: Funkce $f: \langle a,b\rangle \to \mathbb{R}$ je **konvexní** jestliže pro každé $\alpha \in \langle 0,1\rangle$, $\forall x_1, x_2 \in \langle a,b\rangle$ platí nerovnost

$$f(\alpha x_1+(1-\alpha)x_2) \leq \alpha f(x_1)+(1-\alpha)f(x_2)$$

$\downarrow$ funkční hodnota $\qquad\qquad$ $\downarrow$ sečna



Platí-li rovnost jen pro $x_1=x_2$ je f **ryze konvexní**.

Tvrzení: $f: \langle a,b\rangle \to \mathbb{R}$ je konvexní $\iff$

$$(*) \quad f(\alpha_1 x_1+\alpha_2 x_2+\dots+\alpha_n x_n) \leq \alpha_1 f(x_1)+\dots+\alpha_n f(x_n)$$

pro všechna $\alpha_1, \alpha_2, \dots \alpha_n \in \langle 0,1\rangle$, $\sum_{i=1}^{n} \alpha_i = 1$ a všechna $x_1, x_2, \dots x_n \in \langle a,b\rangle$.

f je ryze konvexní $\iff$ rovnost v $(*)$ nastává právě tehdy když $\alpha_1=\alpha_2=\dots=\alpha_n$.

Dk: Indukcí dle n. Ať platí pro $n-1$.

$$f(\alpha_1 x_1+\alpha_2 x_2+\dots+\alpha_n x_n) = f\left(\alpha_1 x_1+(1-\alpha_1)\frac{\alpha_2 x_2+\dots+\alpha_n x_n}{1-\alpha_1}\right)$$

$$\leq \alpha_1 f(x_1) + (1-\alpha_1) f\left(\frac{\alpha_2 x_2+\dots+\alpha_n x_n}{1-\alpha_1}\right)$$

$\downarrow$ konvexní kombinace $x_2, \dots x_n$

indukční předpoklad.

$$\leq \alpha_1 f(x_1) + (1-\alpha_1)\left[\frac{\alpha_2}{1-\alpha_1} f(x_2) + \cdots + \frac{\alpha_n}{1-\alpha_1} f(x_n)\right]$$

$$= \alpha_1 f(x_1) + \alpha_2 f(x_2) + \cdots + \alpha_n f(x_n)$$

Striktní konvexita podobně.

- f je konkávní, striktně konkávní $\Leftrightarrow$ $-f$ je konvexní,
- $\log x$ je ryze konkávní: ryze konvexní

$$(\log x)'' = \frac{1}{\ln 2} \cdot \frac{-1}{x^2} < 0$$

<u>Definice</u> Pro pravděpodobnostní vektor $\vec{p} = (p_1, p_2, \ldots, p_n)$ definujeme <u>entropii</u>

$$\boxed{H(\vec{p}) = -p_1 \log p_1 - p_2 \log p_2 - \cdots - p_n \log p_n = -\sum_{i=1}^{n} p_i \log p_i = -\sum_{\{i \mid p_i \neq 0\}} p_i \log p_i}$$

Základní vlastnosti

- $H(\vec{p}) \geq 0$
- $H(\vec{p})$ nezávisí na permutaci složek vektoru $\vec{p}$
- $H(\vec{p}) = 0 \Leftrightarrow \vec{p}$ je Diracovo rozložení

<u>Definice</u> Je-li X náhodná veličina s konečně mnoha hodnotami. Pak <u>entropie X</u>, $H(X)$, je definována

$$H(X) = H(\vec{p}),$$

kde $\vec{p}$ je pravděpodobnostní vektor pro X.

Příklady: 1) $\vec{p} = (1, 0, \dots, 0)$

$$H(\vec{p}) = 0$$

2) $\vec{p} = \left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right)$

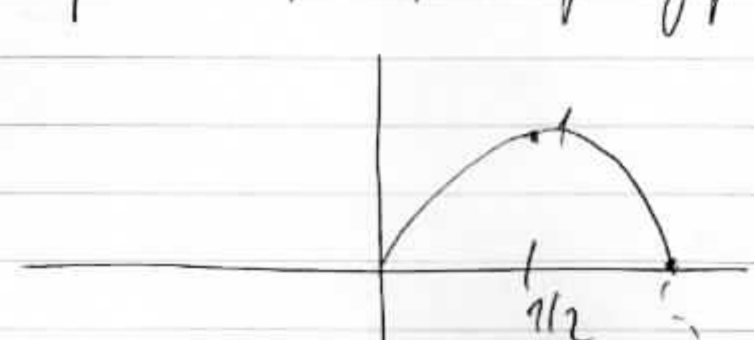
$$H(\vec{p}) = n \cdot \left(-\frac{1}{n} \log \frac{1}{n}\right) = \log n$$

---

3) $p \in (0,1)$

$\vec{p} = (p, 1-p)$

$$H(\vec{p}) = H(p, 1-p) = -p \log p - (1-p) \log(1-p)$$



konvence $H(p) = H(\vec{p})$.

Dirack

maximálně 1 bit.

symetrie díky permutacím

Tvrzení: ať $\vec{p} = (p_1, p_2, \dots, p_n)$ je pravděpodobnostní vektor.
Pak

$$H(\vec{p}) \le \log n$$

Rovnost nastává $\Leftrightarrow$

$$\vec{p} = \left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right)$$

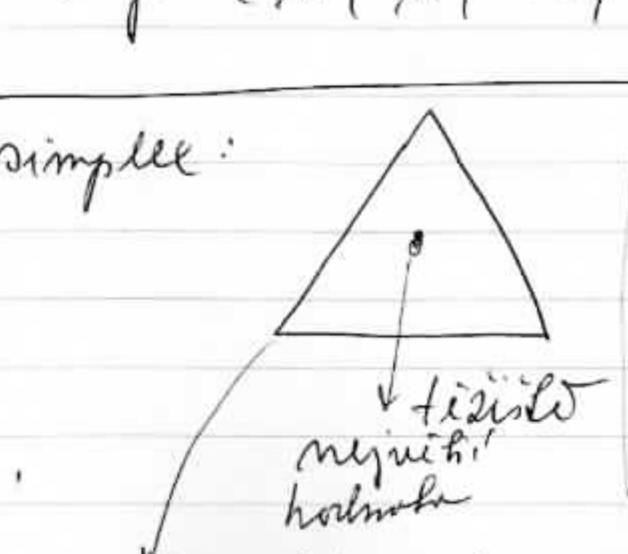
Důkaz: $f(x) = \log x$ je ryze konkávní funkce.

$$H(p) = p_1 f(1/p_1) + p_2 f(1/p_2) + \dots + p_n f(1/p_n)$$
$$\le f\left(p_1 \cdot \frac{1}{p_1} + \dots + p_n \cdot \frac{1}{p_n}\right) =$$
$$= \log n.$$

Rovnost nastává $\Leftrightarrow \frac{1}{p_i} = \frac{1}{p_j} \ \forall i, j$
$\Rightarrow \vec{p} = \left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right)$.

---

simplex:



~~[illegible]~~ rovnostranný trojúhelník/R2
bod $\equiv (p_1, p_2, p_3)$
pravděp. vektor je konvexní kombinace vrcholů

symetrie – obrázek na slidu