*Poznámka*
Toto nejsou úplné zápisky z přednášky, toto je jen moje příprava k zápočtovému testu a později ke zkoušce.

# 1 Markovovy řetězce

**Definice 1.1** (Markovův řetězec)

Nechť $S$ je nejvýše spočetná množina. Posloupnost $(X_t)_{t=0}^{\infty}$ náhodných veličin s oborem hodnot v $S$ je Markovův řetězec (s diskrétním časem, s diskrétním prostorem a časově homogenní) pokud pro každé $t \geqslant 0$ a každé $a_0, \ldots, a_{t+1} \in S$ platí

$$P(X_{t+1} = a_{t+1} | X_t = a_t \wedge \ldots \wedge X_0 = a_0) = P(X_{t+1} = a_{t+1} | X_t = a_t) = p_{a_t, a_{t+1}},$$

pokaždé, když $P(X_t = a_t \wedge \ldots \wedge X_0 = a_0) > 0$.

Množině $S$ se říká stavy, budeme předpokládat, že jsou nějak (pevně) očíslované přirozenými čísly (resp. přirozenými čísly s 0). $p_{a_t, a_{t+1}}$ je pravděpodobnost přechodu ze stavu $a_t$ do stavu $a_{t+1}$

## 1.1 Přechody

**Definice 1.2** (Přechodová matice)

Matice $P$, jejíž prvek $p_{i,j}$ vyjadřuje pravděpodobnost přechodu ze stavu $i$ do stavu $j$.

*Důsledek*
Každý řádek přechodové matice má součet jeho prvků roven 1. Tj. $P \cdot (1, \ldots, 1)^T = (1, \ldots, 1)^T$.

**Definice 1.3** (Přechodový graf/diagram)

Přechodový graf je ohodnocený orientovaný graf se smyčkami, jehož množina vrcholů je $S$. Hrana mezi vrcholy $i, j \in S$ vede právě tehdy, když $p_{i,j} > 0$ a má váhu $p_{i,j}$.

**Definice 1.4** (Pravděpodobnostní rozdělení $X$)

Nechť $(X_t)_{t=0}$ je Markovův řetězec. Pravděpodobnostní rozdělení $X_t$ budeme značit $\pi_i^{(t)} = P(X_t = i)$ pro každý stav $i \in S$, $t \in \mathbb{N}_0$. $\pi^{(t)}$ pak značí řádkový vektor hodnot $\pi_i^{(t)}$.

**Věta 1.1**

*Pro libovolný Markovův řetězec s pravděpodobnostním rozdělením $\pi$ a přechodovou maticí $P$ a libovolné $k \geqslant 0$*

$$\pi^{(k)} = \pi^{(0)} \cdot P^k.$$

*Dokonce obecněji* $\pi^{(t+k)} = \pi^{(t)} P^k$.

*Důkaz*

$$\forall m \in \mathbb{N} : P(X_m = j) = \sum_{i \in S} P(X_{m-1} = i) \cdot P(X_m = j | X_{m-1} = i),$$

$$\pi_j^{(m)} = \sum_{i \in S} \pi_i^{(m-1)} \cdot P_{i,j},$$

$$\pi^{(m)} = \pi^{(m-1)} \cdot P.$$

□

**Definice 1.5** ($k$-krokový přechod)

$$r_{i,j}(k) = P(\text{přechod z } i \text{ do } j \text{ za } k \text{ kroků}) = P(X_k = j | X_0 = i).$$

*Důsledek*

$$r_{i,j}(k) = P(X_{t+k} = j | X_t = i).$$

**Věta 1.2** (Chapman-Kolmogorov)

*Pro libovolný Markovův řetězec a libovolné* $k, l \in \mathbb{N}_0$ *platí*

- $r_{i,j}(k) = \left(P^{(k)}\right)_{i,j}$;
- $r_{i,j}(k+l) = \sum_{u \in S} r_{i,u}(k) r_{u,j}(l)$;
- $r_{i,j}(k+1) = \sum_{u \in S} r_{i,u}(k) p_{u,j}$.

## 1.2 Klasifikace stavů

**Definice 1.6** (Dosažitelný stav)

Pro stavy $i, j$ Markovova řetězce říkáme, že $j$ je dosažitelný z $i$ (píšeme $j \in A(i)$ nebo $i \to j$), pokud je nenulová pravděpodobnost, že začínaje v $i$ dosáhneme $j$ v konečném čase. Tedy

$$j \in A(i) \equiv \exists t \in \mathbb{N}_0 : P(X_t = j | X_0 = i) > 0.$$

*Důsledek*
$j \in A(i)$ odpovídá existenci orientované cesty z $i$ do $j$ v přechodovém grafu.

**Definice 1.7** (Komutující stavy)

Říkáme, že stavy $i, j$ Markovova řetězce komutují, pokud $i \in A(j)$ a $j \in A(i)$. Píšeme $i \leftrightarrow j$.

**Věta 1.3**

*Pro libovolný Markovův řetězec je relace $\leftrightarrow$ (na $S$) ekvivalence.*

**Definice 1.8** (Ireducibilní Markovův řetězec)

Markovův řetězec se nazývá ireducibilní, pokud $\forall i, j \in S : i \leftrightarrow j$.

**Definice 1.9** (Rekurentní stav)

Stav $i \in S$ Markovova řetězce se nazývá rekurentní, pokud $\forall j \in A(i) : i \in A(j)$.

**Definice 1.10** (Transientní stav)

Stav $i \in S$ Markovova řetězce se nazývá transientní (význam: dočasný, přechodný, pomíjivý), pokud není rekurentní.

**Věta 1.4**

*Pro stav $i \in S$ Markovova řetězce označme $f_{ii} = P(\exists t \in \mathbb{N} : X_t = i | X_0 = i)$. Potom, když $f_{ii} = 1$, tak je stav rekurentní, pokud $f_{ii} < 1$, tak je transientní.*

*Důkaz*

Označme $j$ to $j \in A(i)$, pro které $i \notin A(j)$. Potom $P(\exists t \in \mathbb{N} : X_t = j | X_0 = i) \neq 0$ a zřejmě $P(\exists t \in \mathbb{N}\ \forall 0 < t_1 < t : X_t = j \wedge X_{t_1} \neq i | X_0 = i) \neq 0$ a $P(\exists t_2 > t : X_{t_2} = i | X_t = j) = 0$, tedy $f_{ii} \neq 1$.

Naopak pokud $f_{ii} = 1$, tak $\forall j \in A(i)$ musí být $P(\exists t_2 > t : X_{t_2} = i | X_t = j) \neq 0$, tedy $i \in A(j)$. □

**Definice 1.11** (Počet návštěv)

Pro stav $i \in S$ Markovova řetězce označme náhodnou veličinu $V_i$ s oborem hodnot v $\mathbb{N}_0^*$ počet návštěv $i$, tedy $V_i = |\{t | X_t = i\}|$.

**Věta 1.5**

*Stav $i \in S$ Markovova řetězce je rekurentní $\implies P(V_i = \infty | X_0 = i) = 1$. $i$ je transientní, pokud $V_i|_{X_0=i} \sim Geo(1 - f_{ii})$.*

**Definice 1.12** (Stacionární rozložení)

Nechť $\pi$ je pravděpodobnostní rozložení na stavech $S$ Markovova řetězce. Řekneme, že $\pi$ je stacionární rozložení, pokud $\pi \cdot P = \pi$, kde $\pi$ považujeme za řádkový vektor.

*Důsledek*
Pokud $\pi^{(0)}$ je stacionární rozložení, pak $\forall k \in \mathbb{N}_0 : \pi^{(k)} = \pi^{(0)}$.

**Definice 1.13** (Periodický stav, periodický Markovův řetězec, aperiodický ...)

Stav $i \in S$ Markovova řetězce je periodický, pokud $\exists \Delta \in \mathbb{N} \setminus \{1\}$:

$$P(X_t = i | X_0 = i) > 0 \implies \Delta | t.$$

Markovův řetězec se nazývá periodický, pokud jsou všechny jeho stavy periodické.

Stav nebo Markovův řetězec se nazývá aperiodický, pokud není periodický.

**Věta 1.6**

*Buď $(X_t)_{t=0}^{\infty}$ Markovův řetězec, který je ireducibilní, aperiodický a $|S| < \infty$. Potom $\exists \pi$ stacionární rozložení a*

$$\forall j \ \forall i \lim_{k \to \infty} r_{i,j}(k) = \pi_j;$$

*navíc $\pi$ je jednoznačné řešení $\pi \cdot P = \pi$ a $\pi \cdot (1, \ldots, 1)^T = 1$.*

**Definice 1.14** (Absorbující stav)

Stav $a \in S$ Markovova řetězce je absorbující, pokud $p_{a,a} = 1$.

**Definice 1.15** (Čas absorbování)

Předpokládejme $A \subseteq S$ neprázdnou množinu absorbujících stavů Markovova řetězce a BÚNO $0 \in A$. Pro každý stav $i \in S$ definujeme $\mu_i$ jako střední hodnotu času absorbování z $i$, tedy

$$\mu_i = \mathbb{E}(T | X_0 = i), \qquad T = \min \{t : X_t \in A\}.$$

Dále $a_i$ buď pravděpodobnost, že začínaje ve stavu $i$ skončíme v stavu 0.

$$a_i = \sum_{t \in \mathbb{N}_0} P(X_t = 0 | X_0 = i).$$

**Věta 1.7**

*Pravděpodobnosti $a_i$ jsou jednoznačné řešení*

$$a_0 = 1, \qquad a_i = 0, \quad 0 \neq i \in A, \qquad a_i = \sum_{j \in S} p_{i,j} a_j, \quad i \in (S \backslash A) \cup \{0\}.$$

*Důkaz*
TODO? Jednoduchý, větou o úplné pravděpodobnosti. □

**Věta 1.8**

*Střední hodnoty času ($\mu_i$) jsou jednoznačné řešení*

$$\mu_i = 0, \quad i \in A, \qquad \mu_i = 1 + \sum_{j \in S} p_{i,j} \mu_j, \quad i \in S \backslash A.$$

*Důkaz*
TODO? Jednoduchý, větou o úplné střední hodnotě. □

TODO!!! (SAT)

# 2 Bayesovská statistika

## 2.1 Postup

**Definice 2.1** (Parametr hledaného rozdělení)

Hledáme rozdělení s parametrem $\Theta$, který budeme považovat za náhodnou veličinu.

**Definice 2.2** (Apriorní rozdělení)

Nejprve vybereme apriorní rozdělení s pmf (probability mass function) $p_\Theta(\vartheta)$ nebo pdf (probability density function) $f_\Theta(\vartheta)$ náhodné veličiny $\Theta$ nezávisle na datech.

**Definice 2.3** (Statistický model)

Potom zvolíme statistický model $p_{X|\Theta}(x|\vartheta)$ (nebo $f_{X|\Theta}(x|\vartheta)$), který popisuje jak jsou (věříme, že jsou) rozděleny data, pokud je $\Theta$ rovno nějakému konkrétnímu $\vartheta$.

**Definice 2.4** (Posteriorní rozdělení)

Poté, co pozorujeme $X = x$ (více měření považujeme za pozorování jednoho $X = x$ z vícedimenzionálního rozdělení) spočítáme posteriorní rozdělení $f_{\Theta|X}(\vartheta|x)$.

*Poznámka*
Nakonec najdeme, co potřebujeme vědět, například $a, b$ tak, aby $P(a \leqslant \Theta \leqslant b|X = x) = \int_a^b f_{(\Theta|X)}(\vartheta|x)d\vartheta \geqslant 1 - \alpha$.

## 2.2 Bayesova věta

**Věta 2.1** (Bayesova pro obě diskrétní)

*Nechť $X$, $\Theta$ jsou diskrétní náhodné veličiny, pak*

$$p_{\Theta|X}(\vartheta|x) = \frac{p_{X|\Theta}(x|\vartheta)p_\Theta(\vartheta)}{\sum_{\vartheta' \in \operatorname{Im}\Theta \setminus \{p_\Theta(\vartheta')=0\}} p_{X|\Theta}(x|\vartheta')p_\Theta(\vartheta')}.$$

**Věta 2.2** (Bayesova pro obě spojitá)

*Nechť $X$, $\Theta$ jsou spojité náhodné veličiny, pak*

$$f_{\Theta|X}(\vartheta|x) = \frac{f_{X|\Theta}(x|\vartheta)f_\Theta(\vartheta)}{\int_{\vartheta' \in \operatorname{Im}\Theta \setminus \{f_\Theta(\vartheta')=0\}} f_{X|\Theta}(x|\vartheta')f_\Theta(\vartheta')}.$$

**Věta 2.3** (Bayesova pro diskrétní a spojité)

*Nechť $X$ je diskrétní a $\Theta$ spojitá náhodná veličina, pak*

$$f_{\Theta|X}(\vartheta|x) = \frac{p_{X|\Theta}(x|\vartheta)f_\Theta(\vartheta)}{\int_{\vartheta' \in \operatorname{Im}\Theta \setminus \{f_\Theta(\vartheta')=0\}} p_{X|\Theta}(x|\vartheta')f_\Theta(\vartheta')}.$$

## 2.3 Bodové odhady

**Definice 2.5** (MAP – maximum a-posteriori)

Zvolíme modus $\Theta$.

*Poznámka*
Tj. maximum $p_{\Theta|X}(\vartheta|x)$, resp $f_{\Theta|X}(\vartheta|x)$.

**Definice 2.6** (LMS – least mean square)

Zvolíme střední hodnotu $\Theta$, tedy $\mathbb{E}(\Theta|X = x)$.

*Poznámka*
Dostaneme nestranný bodový odhad, který minimalizuje $\mathbb{E}((\Theta - \cdot)^2|X = x)$.

*Poznámka* (Medián)
Obdobně, když vezmeme medián (tj. $m$ tak, že $P(\Theta \leqslant m|X = x) = \frac{1}{2}$), tak minimalizujeme $\mathbb{E}((\Theta - \cdot)|X = x)$, tento přístup však nebudeme dále používat.

TODO? (Zbytek B. statistiky)

# 3 Stochastické procesy

*Poznámka*
I Markovovy řetězce jsou vlastně stochastický proces.

## 3.1 Bernoulliho proces

**Definice 3.1** (Bernoulliho proces)

Bernoulliho proces (s parametrem $p$), píšeme $Bp(p)$, je posloupnost nezávislých náhodných proměnných $(X_t)_{t=1}^{\infty}$, kde $X_t \sim Ber(p)$, tedy $p(X_t = 1) = p$ a $p(X_t = 0) = 1 - p$, $\forall t \in \mathbb{N}$.

*Důsledek*

$$\{X_t\}_{t=1}^{\infty} \sim Bp(p) \implies \{X_t\}_{t=k}^{\infty} \sim Bp(p), \forall k \in \mathbb{N}.$$

$$\{X_t\}_{t=1}^{\infty} \sim Bp(p) \implies \{X_t\}_{t=N}^{\infty} \sim Bp(p),$$

kde $N$ je náhodná veličina závisející pouze na minulosti.

**Definice 3.2** (Čas prvního úspěchu, čas $k$-tého)

$$T := \min\{t|X_t = 1\}, \qquad T_k := \min\left\{t \middle| \sum_{s=1}^{t} X_s = k\right\}.$$

*Důsledek*

$$T \sim Geom(p), \qquad \mathbb{E}[T] = \frac{1}{p}, \qquad \operatorname{var} T = \frac{1-p}{p^2}.$$

**Definice 3.3** (Doba čekání)

$$L_k := T_k - T_{k-1}, \qquad (T_0 = 0).$$

*Důsledek*

$$L_k \sim T \sim Geom(p).$$

*Důkaz*
Restartujeme Bernoulliho proces v $T_{k-1}$. □

*Důsledek*

$$T_k = \sum_{i=1}^{k} L_i.$$

$$\mathbb{E}[T_k] = \sum_{i=1}^{k} \mathbb{E}L_i = \frac{k}{p}, \qquad \operatorname{var} T_k = \sum_{i=1}^{k} \operatorname{var} L_i = k \cdot \frac{1-p}{p^2}.$$

$$p(T_k = t) = \binom{t-1}{k-1} \cdot p^k \cdot (1-p)^{t-k}, \qquad \chi(T_k = t) \sim Pas(p, k),$$

kde $Pas(p, k)$ je tzv. Pascalovo rozdělení (definované právě $p(T_k = t) = \ldots$ výše), také nazývané negativní binomické.

**Věta 3.1** (Spojování Bernoulliho procesů)
*Mějme* $\{X_t\}_{t=1}^{\infty} \sim Bp(p)$ *a* $\{Y_t\}_{t=1}^{\infty} \sim Bp(q)$, *pak* $\{X_t \vee Y_t\}_{t=1}^{\infty} \sim Bp(p + q - pq)$.

**Věta 3.2** (Rozdělování Bernoulliho procesů)
*Mějme* $\{Z_t\}_{t=1}^{\infty} \sim Bp(p)$. *Potom* $\{Z_t \cdot Y_t\}_{t=1}^{\infty} \sim Bp(p \cdot q)$, *kde* $Y_t \sim Ber(q)$.

## 3.2 Poissonův proces

**Definice 3.4** (Poissonův proces)

Definujme časy příchodů jako reálná čísla: $0 < T_1 < T_2 < T_3 < \ldots$. Po Poissonově procesu požadujeme:

1. Pro každou délku intervalu $\tau$ chceme, aby pravděpodobnost $k$ příchodů v tomto intervalu byla stejná, označme ji $p(k, \tau)$.
2. Počet příchodů v intervalu $[a, b]$ je nezávislý na počtu příchodů v $[0, a]$.
3. $p(0, \tau) = 1 - \lambda\tau + o(\tau)$, $p(1, \tau) = \lambda\tau + o(\tau)$ ( $\implies$ $p(k, \tau) = o(\tau)$, $\forall k \geqslant 2$).

Poissonův proces je tedy posloupnost náhodných reálných veličin $0 < T_1 < T_2 < T_3 < \ldots$, která splňuje tyto 3 body.

**Definice 3.5** (Počet příchodů do času $t$)

$$N_t := \max k | T_k \leqslant t$$

**Věta 3.3**

$$N_t \sim Pois(\lambda \cdot t), \qquad p(N_t = k) = e^{-\lambda \cdot t} \frac{(\lambda \cdot t)^k}{k!}.$$

*Důkaz*

Rozdělme si interval $[0, t]$ na $l$ intervalů pro nějaké $l$ velké. Pak délka jednoho intervalu je $\frac{t}{l}$, $p\left(1, \frac{t}{l}\right) = \frac{\lambda \cdot t}{l} + o\left(\frac{t}{l}\right)$ a $p\left(k, \frac{t}{l}\right) = o\left(\frac{t}{l}\right)$. $o\left(\frac{t}{l}\right)$ zanedbáme, tedy máme Binomické rozdělení s parametry $l$ a $\frac{\lambda \cdot t}{l}$, což pro rostoucí $l$ vede k Poissonovu rozdělení s parametrem $\lambda \cdot t$. Tedy

$$p(N_t = k) = e^{-\lambda \cdot t} \frac{(\lambda \cdot t)^k}{k!}.$$

□

**Definice 3.6** (Čekání na další příchod)

$$L_k := T_k - T_{k-1}.$$

*Důsledek*

$$p(L_k \geqslant t) = p(0, t) = e^{-\lambda \cdot t}, \qquad p(L_k \leqslant t) = 1 - p(L_k \geqslant t) = 1 - e^{-\lambda \cdot t}.$$

$$L_k \sim Exp(\lambda).$$

*Důsledek*

$$\mathbb{E}T_k = \sum_{i=1}^{k} \mathbb{E}L_i = k \cdot \frac{1}{\lambda}.$$

$$\operatorname{var} T_k = \sum_{i=1}^{k} \operatorname{var} L_i = k \cdot \frac{1}{\lambda^2}.$$

$$f_{T_k}(t) = \frac{\lambda^k t^{k-1} e^{-\lambda \cdot t}}{(k-1)!}$$

**Věta 3.4** (Rozdělování Poissonových procesů)

*Mějme $0 < T_1 < T_2 < \ldots$ Poissonův proces s parametrem $\lambda$ a každý příchod nezávisle s pravděpodobností $p$ ponechejme. Pak nová $0 < T'_1 < T'_2 < \ldots$ jsou Poissonův proces s parametrem $\lambda \cdot p$. Odstraněné $0 < \tilde{T}_1 < \tilde{T}_2 < \ldots$ jsou Poissonův proces s parametrem $\lambda \cdot (1-p)$. A tyto procesy jsou na sobě nezávislé.*

*Důkaz*

$$p_p(k, \tau) = \sum_{n=k}^{\infty} p(n, \tau) \cdot P(Bin(n, p) = k).$$

Následně se ověří podmínky Poissonova procesu (na přednášce ukázán trochu zjednodušený výpočet).

Nezávislé $\Leftrightarrow P(X = k \wedge Y = l) = P(X = k) \cdot P(Y = l)$. Následně jsme ověřili dosazením. □

**Věta 3.5** (Spojování Poissonových procesů)

*Nechť $0 < T_1 < T_2 < \ldots$ a $0 < S_1 < S_2 < \ldots$ jsou Poissonovy procesy s parametry $\lambda$, $\varkappa$. Potom jejich sjednocením získáme Poissonův proces $0 < R_1 < R_2 < \ldots$ s parametrem $\lambda + \varkappa$. (Případně můžeme spojovat i libovolně mnoho Poissonových procesů do Poissonova procesu s parametrem rovným součtu parametrů původních.)*

*Důkaz*

$$p(R_1 > t) = P(T_1 > t \wedge S_1 > t) = P(T_1 > t) \cdot P(S_1 > t) = e^{-\lambda t} \cdot e^{-\varkappa t} = e^{-(\lambda + \varkappa)t}.$$

Následně restartujeme procesy v $R_1$ a začínáme nanovo :) □

TODO!!! (Balls and bins)

# 4 Neparametrická statistika

**Definice 4.1** (Neparametrická statistika)

Nemáme model (rozdělení závisející na parametru).

**Definice 4.2** (Permutační test)

Mějme data $x_1, \ldots, x_n$ a $y_1, \ldots, y_m$ (např. testovací a kontrolní vzorek). Dále mějme $f$, které rozhoduje, zda dané $z_1, \ldots, z_{m+n}$ splňuje nulovou hypotézu.

$$\mathcal{F} := \{f(\pi(z))\}_{\pi \in S_{n+m}}$$

$p$-hodnota je podíl prvků souboru $\mathcal{F}$, které splňují nulovou hypotézu. Nulovou hypotézu zamítneme, pokud je tento podíl menší než $\alpha$.

(Požadujeme, aby za nulové hypotézy byla pravděpodobnost každého prvku $\mathbb{F}$ stejná.)

**Definice 4.3** (Permutační test ++)

Pokud nemůžeme počítat $f$ pro všechny $\pi \in S_{n+m}$, nasamplujeme $\mathbb{F}^* \subset \mathbb{F}$.

**Definice 4.4** (Znamínkový test)

$X_1, \ldots, X_n$ nezávislé náhodné veličiny z neznámého spojitého rozdělení symetrické podle střední hodnoty. Nulová hypotéza je, že střední hodnota je 0.

Nechť $Y_i = \operatorname{sgn}(X_i) = +1$ nebo 0 (pozor, ne $-1$). Potom při předpokladu nulové hypotézy $Y = \sum_{i=1}^{n} Y_i \sim Binom(n, \frac{1}{2})$. Tedy nulovou hypotézu zamítneme, pokud $Y \leqslant Y_{\alpha/2}$ nebo $Y > Y_{1-\alpha/2}$, kde $P(Binom(n, \frac{1}{2}) < Y_x) = x$.

**Definice 4.5** (Pair test)

Mějme data, která jsou přirozeně v párech (např. hodnota před a po vylepšení algoritmu) a mějme nějakou hypotézu, kterou můžeme testovat po prvcích (např. jestli se průměr nových a starých hodnot shoduje, což můžeme testovat jako „jestli je průměr rozdílů hodnot 0“). Potom se můžeme na pár dívat jako na jeden prvek.

**Definice 4.6** (Wilcoxonův test znamínka hodnosti)

$X_1, \ldots, X_n$ nezávislé náhodné veličiny z neznámého spojitého rozdělení symetrické podle střední hodnoty. Nulová hypotéza je, že střední hodnota je 0.

Hodnost $(\operatorname{rank}, r_i)$ je pořadí v seřazení $|X_i|$ (místo sdíleného pořadí vezmeme průměr sdílených míst, to se ve skutečnosti v spojitém rozdělení nemůže stát). Definujeme

$$T := (W :=) \sum_{i=1}^{n} r_i \cdot \operatorname{sgn}(X_i) = T^+ - T^-.$$

Zamítneme nulovou hypotézu, pokud $T$ je moc velké nebo moc malé, tj. $T < Y_{\alpha/2}$ nebo

$T > Y_{1-\alpha/2}$ ve správném (TODO?) rozdělení.

**Definice 4.7** (Mannův–Whitneyho U-test)

Máme dvě množiny $X_1, \ldots, X_n$ a $Y_1, \ldots, Y_m$.

$$U := \sum_{i=1}^{n} \sum_{j=1}^{m} S(X_i, Y_i), \qquad S(X,Y) := \begin{cases} 0, & X > Y, \\ \frac{1}{2}, & X = Y, \\ 1, & X < Y. \end{cases}$$

Nulová hypotéza je $P(X < Y) = P(Y < X)$.

TODO!!! (Simpson paradox)

TODO!!! (CLV)