



Zadání bakalářské práce

Název:	Predikce časových rozestupů pomocí dat z evakuačních experimentů
Student:	Ondřej Hývnar
Vedoucí:	Ing. Pavel Hrabák, Ph.D.
Studijní program:	Informatika
Obor / specializace:	Znalostní inženýrství
Katedra:	Katedra aplikované matematiky
Platnost zadání:	do konce letního semestru 2023/2024

Pokyny pro vypracování

Časové rozestupy mezi jednotlivými průchody úzkým hrdlem, především jejich rozdělení, jsou důležitou charakteristikou dynamiky pohybu chodců. Autoři v (Bode a Codling, 2016) navrhli statistický model založený na gamma rozdělení, jehož střední hodnota závisí na stavu okolí chodce před průchodem úzkým hrdlem. Cílem práce je reprodukovat výsledky zmíněného článku pro datovou sadu získanou během experimentu E4 (data poskytne školitel), analyzovat naučený model a porovnat jej s výsledky ze zmíněného článku.

1. Proveďte rešerši statistického modelování chování chodců se zaměřením na časové rozestupy mezi průchody úzkým hrdlem.
2. Proveďte rešerši nástrojů souvisejících s odhadem parametrů modelu založený na maximální věrohodnosti (MLE).
3. Zpracujte poskytnutá data z experimentu E4 a vybranými nástroji reprodukujte výsledky z článku (Bode a Codling, 2016). Výstupy porovnejte.

Bode, N., Codling, E. Statistical Models for Pedestrian Behaviour in Front of Bottlenecks. Traffic and Granular Flow '15. Eds. Knoop V.L., Daamen W. Springer, 2016. pp. 81-88.

Bakalářská práce

PREDIKCE ČASOVÝCH ROZESTUPŮ POMOCÍ DAT Z EVAKUAČNÍCH EXPERIMENTŮ

Ondřej Hývnar

Fakulta informačních technologií
Katedra aplikované matematiky
Vedoucí: Ing. Pavel Hrabák, Ph.D.
11. května 2023

České vysoké učení technické v Praze
Fakulta informačních technologií

© 2023 Ondřej Hývnar. Všechna práva vyhrazena.

Tato práce vznikla jako školní dílo na Českém vysokém učení technickém v Praze, Fakultě informačních technologií. Práce je chráněna právními předpisy a mezinárodními úmluvami o právu autorském a právech souvisejících s právem autorským. K jejímu užití, s výjimkou bezúplatných zákonných licencí a nad rámec oprávnění uvedených v Prohlášení, je nezbytný souhlas autora.

Odkaz na tuto práci: Hývnar Ondřej. *Predikce časových rozestupů pomocí dat z evakuačních experimentů*. Bakalářská práce. České vysoké učení technické v Praze, Fakulta informačních technologií, 2023.

Obsah

Poděkování	vi
Prohlášení	vii
Abstrakt	viii
Seznam zkratk	ix
1 Úvod	1
2 Teoretický základ	3
2.1 Základní pojmy	3
2.2 Gamma rozdělení	5
2.3 Odhad parametrů statistického modelu	7
2.3.1 Metoda maximální věrohodnosti	7
2.4 Matematická optimalizace	8
2.5 Porovnávání statistických modelů	11
3 Statistické modelování chování chodců	13
3.1 Studie časových rozestupů	13
3.2 Model Shafiri, Christensen, Chen	17
3.3 Model Bode a Codling	20
4 Praktická část	23
4.1 Experiment	23
4.2 Předzpracování dat	24
4.3 Odhad parametrů	27
5 Zhodnocení výsledků	29
6 Závěr	37
A Výsledky jednotlivých běhů experimentu	39
Obsah přiloženého média	53

Seznam obrázků

2.1	Příklady gamma rozdělení	6
3.1	Snímek z experimentu [12]	14
3.2	Rozdělení časových rozestupů pro šířky chodby a) 0,5 m, b) 0,8 m, c) 1,3 m [12]	15
3.3	Vztah rychlosti a pořadí chodce při průchodu pro různé šířky chodby [12]	16
3.4	Průměrná rychlost skupiny při průchodu pro šířku chodby 1.2m [12]	17
3.5	Rozdělení časových rozestupů pro jednotlivé typy vůdců. Křivka znázorňuje odhad modelu. [13]	19
3.6	Vektory \vec{x} , \vec{y} a úhel θ_{ij}	21
3.7	Snímky z experimentů a provedená simulace[3]	22
3.8	Vypočtené hodnoty AIC pro experimenty a simulaci[3]	22
4.1	Snímek z experimentu E4	23
4.2	Červeně označený střed úzkého hrdla, modře zed'	24
4.3	Označená oblast před vstupem do úzkého hrdla	25
5.1	Informační kritéria výsledných modelů	29
5.2	Porovnání výsledných modelů s modely z reprodukováného článku	30
5.3	Residua gamma modelu	31
5.4	Predikce μ_{p-1} pomocí kvadratické regrese	32
5.5	Residua kvadratické regrese	32
5.6	Řez skrz hodnoty d_i	33
5.7	Řez skrz hodnoty d_{ij}	34
5.8	Porovnání skutečných rozestupů a odhadnutého rozdělení	35
5.9	Rozdělení hodnot d_i a d_{ij}	36
5.10	Porovnání skutečných rozestupů z okolí bodu P	36
A.1	Průchod 2	40
A.2	Průchod 3	41
A.3	Průchod 4	42
A.4	Průchod 5	43
A.5	Průchod 6	44
A.6	Průchod 7	45
A.7	Průchod 8	46
A.8	Průchod 9	47
A.9	Průchod 10	48
A.10	Průchod 11	49
A.11	Průchod 12	50

Seznam tabulek

4.1	Formát poskytnutých dat, soubor <code>data/E4-R_pruchochod_2.txt</code>	24
4.2	Extrahovaná data, soubor <code>regressors/E4-R_pruchochod_3.txt</code>	26
4.3	Nalezené hodnoty parametrů, soubor <code>models/all.txt</code> , hodnoty jsou zaokrouhleny na 4 desetinná místa	28

Seznam algoritmů

1	Detekce časů průchodu	25
2	Filtrace trajektorií	25
3	Extrakce dat pro modely	26
4	Odhad parametru k	27

Chtěl bych poděkovat především vedoucímu panu Ing. Pavlu Hrabákovi, Ph.D. za trpělivost během naší spolupráce na této práci. Dále bych chtěl poděkovat rodičům a kamarádům za podporu při studiu a při psaní této práce.

Prohlášení

Prohlašuji, že jsem předloženou práci vypracoval samostatně a že jsem uvedl veškeré použité informační zdroje v souladu s Metodickým pokynem o dodržování etických principů při přípravě vysokoškolských závěrečných prací. Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona, ve znění pozdějších předpisů, zejména skutečnost, že České vysoké učení technické v Praze má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 citovaného zákona.

V Praze dne 11. května 2023

.....

Abstrakt

Časové rozestupy mezi průchody úzkým hrdlem jsou důležitou charakteristikou pohybu chodců. Tato práce reprodukuje výsledky předchozího výzkumu v oblasti modelování časových rozestupů, ve kterém byly představeny statistické modely pro predikci rozdělení časových rozestupů při průchodu úzkým hrdlem v závislosti na okolních parametrech. Parametry těchto modelů odhaduji na datech z evakuačních experimentů pořádaných ČVUT a diskutuji rozdíly s výsledky předchozího výzkumu.

Klíčová slova časové rozestupy chodců, metoda maximální věrohodnosti, gamma rozdělení, matematická optimalizace

Abstract

Time headway in bottlenecks is an important characteristic of pedestrian behaviour. This thesis reproduces the result of previous research in this field, where several statistical models for prediction of time headway distribution in bottleneck based on walker's surroundings were proposed. I estimate parameters of these models on data from evacuation experiments organized by CTU and discuss the differences with results from previous research.

Keywords pedestrian time headway, maximum likelihood estimate, gamma distribution, mathematical optimization

Seznam zkratk

NV	Náhodná veličina
AIC	Akaikeho informační kritérium
BIC	Bayesovské informační kritérium



Kapitola 1

Úvod

Časové rozestupy mezi průchody úzkým hrdlem jsou důležitou charakteristikou pohybu chodců, která ovlivňuje celkovou dynamiku davu a může být klíčovým faktorem pro plánování a návrh veřejných prostor. V článku *Statistical Models for Pedestrian Behaviour in Front of Bottlenecks* navrhli autoři Nikolai W. F. Bode a Edward A. Codling statistický model založený na gamma rozdělení, který se zaměřuje na vliv okolí chodce na časový rozestup od ostatních chodců při průchodu úzkým hrdlem. Tato práce si klade za cíl reprodukovat výsledky tohoto článku pomocí datové sady získané během experimentu E4, který proběhl v rámci ČVUT. Kromě toho bude provedena analýza naučených modelů a porovnání s výsledky zmíněného článku. Výsledky této práce mohou přispět k lepšímu pochopení pohybu chodců.

Časovým rozestupem při průchodu se rozumí rozdíl v časech průchodu dvou po sobě jdoucích chodců. Cílem modelování těchto rozestupů je porozumět, jak chodci interagují s prostředím a jaké faktory ovlivňují jejich chování. Těmito faktory mohou být například vzdálenost chodců k úzkému hrdlu nebo rychlost pohybu. V této práci studuji vliv vzdálenosti nejbližšího chodce, hustoty chodců a úhlu mezi dvěma nejbližšími chodci.

Teoretický základ

Tato kapitola poskytuje úvod do statistických pojmů, matematické optimalizace a porovnávání statistických modelů.

2.1 Základní pojmy

► **Definice 2.1** (Pravděpodobnostní prostor[1]). *Pravděpodobnostní prostor, nebo také experiment ε je trojice*

$$\varepsilon = (\Omega, F, P)$$

Ω označuje množinu všech možných výsledků experimentu, F množinu všech jevů, pro které lze spočítat pravděpodobnost P .

► **Definice 2.2** (Náhodná veličina[1]). *Bud' $\varepsilon = (\Omega, F, P)$ pravděpodobnostní prostor. Náhodná veličina X na ε je funkce, která každému výsledku experimentu $\omega \in \Omega$ přiřadí hodnotu $X(\omega) \in R$ a pro kterou platí podmínka měřitelnosti:*

$$\{X \leq x\} \in F, \forall x \in R$$

► **Definice 2.3** (Distribuční funkce[1]). *Distribuční funkci F náhodné veličiny X definujeme vztahem:*

$$F_X(x) = P(X \leq x)$$

Pokud náhodná veličina X nabývá hodnot z konečné, nebo spočetně nekonečné množiny, nazýváme X *diskrétní* náhodnou veličinou a distribuční funkci můžeme vyjádřit jako:

$$F_X = \sum_{k: x_k \leq x} P(X = x_k)$$

V případě, že existuje nezáporná funkce f_X , pomocí níž lze distribuční funkci F_X vyjádřit jako:

$$F_X(x) = \int_{-\infty}^x f_X(t) dt, \forall x \in R$$

nazýváme X *spojitou* náhodnou veličinou a funkci f_X nazýváme *hustotou pravděpodobnosti* náhodné veličiny X .

► **Definice 2.4** (Střední hodnota[1]). *Střední hodnotu μ , EX náhodné veličiny X definujeme pro diskrétní NV jako:*

$$EX = \mu = \sum_k x_k P(X = x_k)$$

a pro spojitou NV jako:

$$EX = \mu = \int_{-\infty}^{\infty} x f_X(x) dx$$

► **Definice 2.5** (Rozptyl[1]). *Rozptyl σ^2 , $\text{var}X$ náhodné veličiny X definujeme jako:*

$$\text{var}X = \sigma^2 = E[(X - EX)^2]$$

V části 2.4 budou důležité pojmy *parciální derivace*, *gradient funkce* a *Hessova matice*.

► **Definice 2.6** (Parciální derivace[2]). *Bud' $f : R^n \rightarrow R$ funkce. Derivaci funkce f dle proměnné x_i , $i \in \{1, 2, \dots, n\}$ při ostatních proměnných konstantních označujeme jako parciální derivaci a značíme $\frac{\partial f}{\partial x_i}$.*

► **Definice 2.7** (Gradient funkce[2]). *Bud' $f : R^n \rightarrow R$ funkce. Vektor parciálních derivací funkce f v bodě $p \in R^n$ nazveme gradient funkce f v bodě p :*

$$\nabla f(p) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(p) \\ \frac{\partial f}{\partial x_2}(p) \\ \vdots \\ \frac{\partial f}{\partial x_n}(p) \end{bmatrix}$$

► **Definice 2.8** (Hessova matice[2]). *Bud' $f : R^n \rightarrow R$ funkce. Necht' existují všechny druhé parciální derivace funkce f . Hessovu matici funkce f poté definujeme jako:*

$$H_f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

2.2 Gamma rozdělení

Model z článku [3], který v této práci reprodukuji, je založen na gamma rozdělení, pomocí něhož modeluje střední hodnotu časových rozestupů při průchodu chodců úzkým hrdlem.

Gamma rozdělení[4] je skupina spojitých rozdělení pravděpodobnosti, které můžeme vyjádřit pomocí dvou parametrů. Existuje několik způsobů parametrizace gamma rozdělení, nejčastějšími způsoby jsou:

- Tvarem k a měřítkem θ
- Tvarem α a inverzním měřítkem β
- Tvarem k a střední hodnotou μ
- Střední hodnotou μ a rozptylem σ^2

Pro každou z těchto parametrizací musí být oba parametry kladná reálná čísla. Tyto parametrizace jsou mezi sebou ekvivalentní. V této práci budu využívat parametrizaci pomocí tvaru k a měřítka θ z balíčku SciPy[5].

Pro střední hodnotu μ gamma rozdělení platí:

$$\mu = k\theta = \frac{\alpha}{\beta}$$

a pro rozptyl σ^2 :

$$\sigma^2 = k\theta^2 = \frac{\alpha}{\beta^2}$$

► **Definice 2.9** (Gamma funkce[6]). *Gamma funkce Γ je zobecněním integrálu pro komplexní čísla. Je definována pro všechna komplexní čísla s výjimkou nekladných celých čísel. Pro kladná celá čísla je definována jako*

$$\Gamma(n) = (n - 1)!$$

a pro komplexní čísla s kladnou reálnou částí jako

$$\Gamma(z) = \int_0^{\infty} x^{z-1} e^{-x} dx$$

Dále definujeme horní gamma funkci Γ :

$$\Gamma(s, x) = \int_x^{\infty} t^{s-1} e^{-t} dt$$

a dolní gamma funkci γ :

$$\gamma(s, x) = \int_0^x t^{s-1} e^{-t} dt$$

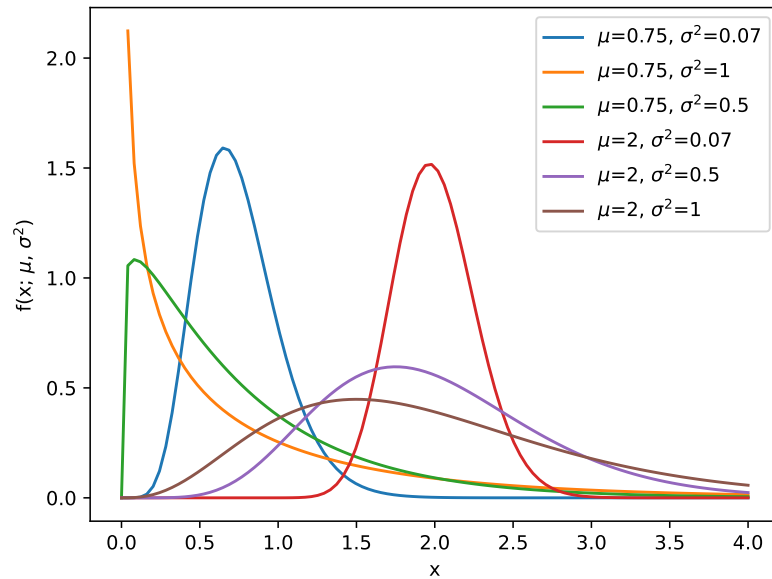
Hustota pravděpodobnosti náhodné veličiny X o gamma rozdělení $\Gamma(\alpha, \beta)$ má tvar:

$$f(x; \alpha, \beta) = \frac{x^{\alpha-1} e^{-\beta x} \beta^{\alpha}}{\Gamma(\alpha)}$$

pro $x, \alpha, \beta > 0$

Distribuční funkce náhodné veličiny X o gamma rozdělení $\Gamma(\alpha, \beta)$ má tvar:

$$F(x; \alpha, \beta) = \int_0^x f(u; \alpha, \beta) du = \frac{\gamma(\alpha, \beta x)}{\Gamma(\alpha)}$$



■ **Obrázek 2.1** Příklady gamma rozdělení

V praktické části budu kromě modelování časových rozestupů pomocí gamma rozdělení využívat i predikci těchto rozestupů pomocí polynomiální regrese stupně dva (kvadratická regrese) pro dvě vysvětlující proměnné, pro jejíž implementaci využiji lineární regresi.

► **Definice 2.10** (Lineární regrese[2]). *Bud' y vysvětlovaná proměnná, bud' (x_1, x_2, \dots, x_n) vektor vysvětlujících proměnných (příznaků), bud' $w = (w_0, w_1, \dots, w_n)$ vektor koeficientů. Model lineární regrese zapíšeme jako:*

$$y = w_0 + \sum_{i=1}^n w_i x_i + \epsilon,$$

kde ϵ je náhodná veličina se střední hodnotou 0.

Hodnota w_0 se označuje jako *intercept* a odpovídá hodnotě vysvětlované proměnné y při nulových hodnotách příznaků. Predikci \hat{y} skutečné hodnoty y získáme jako:

$$\hat{y} = \hat{w}_0 + \sum_{i=1}^n \hat{w}_1 x_i,$$

kde $(\hat{w}_0, \hat{w}_1, \dots, \hat{w}_n)$ je odhad vektoru parametrů w .

► **Definice 2.11** (Kvadratická regrese[7]). *Bud' y vysvětlovaná proměnná, bud'te x_1, x_2 vysvětlující proměnné, bud' (w_1, w_2, \dots, w_5) vektor koeficientů. Model kvadratické regrese zapíšeme jako:*

$$y = w_0 + w_1 x_1 + w_2 x_2 + w_3 x_1^2 + w_4 x_1 x_2 + w_5 x_2^2 + \epsilon$$

Predikci \hat{y} skutečné hodnoty y získáme jako:

$$\hat{y} = \hat{w}_0 + \hat{w}_1 x_1 + \hat{w}_2 x_2 + \hat{w}_3 x_1^2 + \hat{w}_4 x_1 x_2 + \hat{w}_5 x_2^2$$

kde $(\hat{w}_0, \hat{w}_1, \dots, \hat{w}_n)$ je odhad vektoru parametrů w .

2.3 Odhad parametrů statistického modelu

► **Definice 2.12** (Náhodný výběr[1]). n -tici X_1, X_2, \dots, X_n stejně rozdělených nezávislých veličin s distribuční funkcí F nazveme náhodný výběr z rozdělení F .

► **Definice 2.13** (Bodový odhad[1]). Funkci $\hat{\theta}_n(X_1, X_2, \dots, X_n)$ náhodného výběru, která nezávisí na hodnotách θ , nazveme bodovým odhadem parametru θ .

Pro získání bodových odhadů existuje několik metod, např. *momentová metoda* a *metoda maximální věrohodnosti*, která je využívána v této práci. [1]

2.3.1 Metoda maximální věrohodnosti

► **Definice 2.14** (Věrohodnostní funkce[1]). Buď X_1, X_2, \dots, X_n náhodný výběr se sdruženou hustotou

$$f_{\theta}(x) = \prod_{i=1}^n f_{\theta}(x_i)$$

pro spojité rozdělení, nebo sdruženou pravděpodobností

$$p_{\theta}(x) = \prod_{i=1}^n P_{\theta}(X_i = x_i)$$

pro diskrétní rozdělení. Při pevných hodnotách dat $x = (x_1, x_2, \dots, x_n)$ se $f_{\theta}(x)$, nebo $p_{\theta}(x)$ jako funkce θ nazývá věrohodnostní funkcí a značí se $\mathcal{L}(\theta; x)$, nebo zkráceně $\mathcal{L}(\theta)$.

V praxi se často používá logaritmická věrohodnostní funkce:

$$\ell(\theta; x) = \ln \mathcal{L}(\theta; x)$$

čímž se součin změní na součet:

$$\ell(\theta; x) = \sum_{i=1}^n f_{\theta}(x_i), \text{ pro spojité rozdělení}$$

$$\ell(\theta; x) = \sum_{i=1}^n P_{\theta}(X_i = x_i), \text{ pro diskrétní rozdělení}$$

► **Definice 2.15** (Maximálně věrohodný odhad[1]). Nechť Θ označuje podmnožinu Euklidovského prostoru konečné dimenze zvanou parametrický prostor. Hodnotu $\hat{\theta}_n$ parametru θ , pro kterou nabývá věrohodnostní funkce $\mathcal{L}(\theta; x)$ pro pevné hodnoty dat x maxima, nazýváme maximálně věrohodný odhad parametru θ . To znamená, že pro všechna $\theta \in \Theta$:

$$\mathcal{L}(\hat{\theta}_n; x) \geq \mathcal{L}(\theta; x)$$

Protože je logaritmus monotónní funkce, nabývá logaritmická věrohodnostní funkce maxima ve stejném bodě jako věrohodnostní funkce. Pokud je věrohodnostní funkce $\ell(\theta; x)$ diferencovatelná, nutné podmínky pro existenci maxima jsou:

$$\frac{\partial \ell}{\partial \theta_1} = 0, \quad \frac{\partial \ell}{\partial \theta_2} = 0, \quad \dots, \quad \frac{\partial \ell}{\partial \theta_k} = 0$$

Tyto podmínky nazýváme *věrohodnostní rovnice*. Hodnotu $\hat{\theta}_n$ lze získat vyřešením daných rovnic, neexistuje ale obecný způsob, jak toto řešení najít, a tak je pro nalezení $\hat{\theta}_n$ nutno využít *matematické optimalizace*.

2.4 Matematická optimalizace

Matematická optimalizace, také označovaná jako *nelineární programování*, *matematické programování* nebo *numerická optimalizace*, je oblast matematiky zabývající se hledáním minimálních nebo maximálních hodnot funkcí. Analyzovanou funkci nazýváme *objektivní funkce*.

► **Definice 2.16** (Optimalizační problém). *Bud' $f : R^n \rightarrow R$ objektivní funkce. Řešením minimalizačního optimalizačního problému je $x^* \in R^n$ takové, že:*

$$\forall x \in R^n : f(x^*) \leq f(x)$$

Řešením maximalizačního optimalizačního problému je $x^ \in R^n$ takové, že:*

$$\forall x \in R^n : f(x^*) \geq f(x)$$

Optimalizační problém může být doplněn omezujícími podmínkami:

$$g_j(x) \leq 0, \quad j = 1, 2, \dots, m$$

$$h_j(x) = 0, \quad j = 1, 2, \dots, r$$

V takovém případě hovoříme o *omezeném optimalizačním problému*, v opačném případě o *neomezeném optimalizačním problému*.

Pro některé optimalizační problémy lze získat řešení analyticky, čímž získáme přesný popis řešení ve formě matematického vzorce. Pro složitější problémy analytické řešení neexistuje a je nutné jej získat numericky pomocí *optimalizačních algoritmů*. [8]

L-BFGS algoritmus

V této práci využiji pro optimalizaci implementaci *L-BFGS*[9] algoritmu z balíčku *SciPy*[5] pro programovací jazyk Python. Jedná se o optimalizační algoritmus pro hledání minima n -dimenzionální funkce $f : R^n \rightarrow R$ se známým gradientem g a bodu x^* , ve kterém funkce minima nabývá. Algoritmus může být doplněn o vektory l a u představující dolní, resp. horní meze pro jednotlivé složky x :

$$x^* = \arg \min_{x \in R^n} (f(x))$$

$$l \leq x \leq u$$

Na začátku každé iterace algoritmu je znám bod x_k , hodnota funkce f_k , gradient g_k a odhad Hessiany matice B_k . V k -té iteraci je vypočítán odhad objektivní funkce v bodě x_k jako:

$$m_k(x) = f(x_k) + g_k^T(x - x_k) + \frac{1}{2}(x - x_k)^T B_k(x - x_k)$$

Algoritmus iterativně minimalizuje hodnotu $m_k(x)$ vzhledem k daným mezím l a u . Nejprve jsou pomocí gradientní projekce nalezeny aktivní meze, které jsou poté v procesu minimalizace $m_k(x)$ považovány za podmínky rovnosti. Definujme po částech lineární funkci x :

$$x(t) = P(x_k - t g_k, l, u)$$

kde funkce P představuje projekci do mezí l a u :

$$P(x, l, u) = \begin{cases} l_i & x_i < l_i \\ x_i & x_i \in [l_i, u_i] \\ u_i & x_i > u_i \end{cases}$$

Následně spočteme zobecněný Cauchyho bod x^c definovaný jako první bod lokálního minima po částech kvadratické funkce

$$q_k(t) = m_k(x(t))$$

Označme $\mathcal{A}(x^c)$ množinu indexů proměnných, jejichž hodnota je v bodě x^c rovna horní nebo dolní mezi. Na podprostoru volných proměnných poté uvažme následující kvadratický problém:

$$\min\{m_k(x) : x_i = x_i^c, \forall i \in \mathcal{A}(x^c)\}$$

s omezujícími podmínkami

$$l_i \leq x_i \leq u_i, \forall i \notin \mathcal{A}(x^c)$$

Daný problém lze vyřešit dvěma způsoby, Metodou *Lagrangeových multiplikátorů*, nebo iterativně se začátkem v bodu x^c a postupným zkracováním cesty k řešení, dokud nejsou splněny dané omezující podmínky. Řešení tohoto problému označíme \bar{x}_{k+1} .

Následně spočteme bod x_{k+1} metodou *line search* přes $d_k = \bar{x}_{k+1} - x_k$ splňující podmínku dostatečného snížení:

$$f(x_{k+1}) \leq f(x_k) + \alpha \lambda_k g_k^T d_k$$

a podmínku zakřivení:

$$|g_{k+1}^T d_k| \leq \beta |g_k^T d_k|$$

kde λ_k značí délku kroku a $\alpha = 10^{-4}$, $\beta = 0.9$ jsou parametry.

Následně spočteme gradient v bodě x_{k+1} a nový odhad Hessovy matice B_{k+1} a pokračujeme další iterací.

Aktualizace odhadu Hessovy matice

Pro aktualizaci odhadu Hessovy matice B_k je využito m opravných párů $\{s_i, y_i\}$, $i = k-1, k-2, \dots, k-m$:

$$s_k = x_{k+1} - x_k$$

$$y_k = g_{k+1} - g_k$$

Tyto opravné páry vyjadřují změnu aktuálního bodu a gradientu objektivní funkce mezi dvěma iteracemi a můžeme je využít pro sestavení opravných matic Y_k a S_k :

$$Y_k = [y_{k-m}, y_{k-m+1}, \dots, y_{k-1}]$$

$$S_k = [s_{k-m}, s_{k-m+1}, \dots, s_{k-1}]$$

Pokud páry $\{s_i, y_i\}$ splňují $s_i^T y_i > 0$ a θ je kladný parametr, můžeme matici B_k vyjádřit jako

$$B_k = \theta E - W_k M_k W_k^T$$

kde

$$W_k = [Y_k \quad \theta S_k]$$

$$M_k = \begin{bmatrix} -D_k & L_k^T \\ L_k & \theta S_k^T S_k \end{bmatrix}^{-1}$$

a matice D_k a L_k jsou definovány jako

$$(L_k)_{i,j} = \begin{cases} (s_{k-m-1+i})^T (y_{k-m-1+j}) & i > j \\ 0 & \text{jinak} \end{cases}$$

$$D_k = \begin{bmatrix} s_{k-m}^T y_{k-m} & 0 & \cdots & 0 \\ 0 & s_{k-m+1}^T y_{k-m+1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_{k-1}^T y_{k+1} \end{bmatrix}$$

Kvůli daným mezím l a u nemusí metoda *line search* vždy splňovat podmínku zakřivení, což může mít za následek, že není vždy splněna podmínka $s_k^T y_k > 0$. Pro zachování pozitivní definitnosti Hessiany matice jsou zahazeny opravné páry $\{s_k, y_k\}$, pokud podmínka zakřivení $s_k^T y_k > \epsilon \|x\|^2$ pro malou kladnou konstantu ϵ není splněna. Z tohoto důvodu mohou být sloupce matic S_k a Y_k tvořeny vektory s indexy menšími než $k - m$.

Výpočet Cauchyho bodu

V této části nebudeme pro lepší čitelnost psát dolní index označující aktuální iteraci, g, x a B tedy budou označovat g_k, x_k a B_k a dolní index i bude označovat $i - \text{tou}$ složku daného vektoru a horní index j bude značit $j - \text{tou}$ iteraci hledání Cauchyho bodu.

Na začátku nastavíme $x^0 = x_k$ a spočteme

$$t_i = \begin{cases} \frac{(x_i^0 - u_i) g_i}{g_i} & g_i < 0 \\ \frac{x_i^0 - l_i}{g_i} & g_i > 0 \\ \infty & \text{jinak} \end{cases}$$

a soubor $(t_i | i = 1, 2, \dots, n)$ vzestupně seřadíme. Hledání přes po částech lineární funkci $P(x^0 - tg, l, u)$ poté můžeme vyjádřit jako

$$x_i(t) = \begin{cases} x_i^0 - tg_i & t \leq t_i \\ x_i^0 - t_i g_i & \text{jinak} \end{cases}$$

Předpokládejme, že zkoumáme interval $[t^{j-1}, t^j]$. Definujme bod zlomu $j - 1$ jako

$$x^{j-1} = x(t^{j-1})$$

tak, aby na intervalu $[t^{j-1}, t^j]$ platilo

$$x(t) = x^{j-1} + \Delta t d^{j-1}$$

kde

$$\Delta t = t - t^{j-1}$$

$$d_i^{j-1} = \begin{cases} -g_i & t^{j-1} < t_i \\ 0 & \text{jinak} \end{cases}$$

Pomocí tohoto značení můžeme odhad objektivní funkce f pro iteraci k na části přímky $[x(t^{j-1}), x(t^j)]$ vyjádřit jako

$$\begin{aligned} m(x) &= f + g^T (x - x^0) + \frac{1}{2} (x - x^0)^T B (x - x^0) \\ &= f + g^T (z^{j-1} + \Delta t d^{j-1}) + \frac{1}{2} (z^{j-1} + \Delta t d^{j-1})^T B (z^{j-1} + \Delta t d^{j-1}) \end{aligned}$$

kde

$$z^{j-1} = x^{j-1} - x^0$$

V bodě Δt můžeme $m(x)$ pro část přímky $[x(t^{j-1}), x(t^j)]$ vyjádřit jako

$$\begin{aligned} \hat{m}(\Delta t) &= (f + g^T z^{j-1} + \frac{1}{2} z^{j-1T} B z^{j-1}) + (g^T d^{j-1} + d^{j-1T} B z^{j-1}) \Delta t + \frac{1}{2} (d^{j-1T} B d^{j-1}) \Delta t^2 \\ &= f_{j-1} + f'_{j-1} \Delta t + \frac{1}{2} f''_{j-1} \Delta t^2 \end{aligned}$$

což je kvadratická rovnice s neznámou Δt a parametry

$$\begin{aligned} f_{j-1} &= f + g^T z^{j-1} + \frac{1}{2} z^{j-1T} B z^{j-1} \\ f'_{j-1} &= g^T d^{j-1} + d^{j-1T} B z^{j-1} \\ f''_{j-1} &= d^{j-1T} B d^{j-1} \end{aligned}$$

Zderivováním $\hat{m}(x)$ a vyřešením $\hat{m}'(\Delta t) = 0$ získáme

$$\Delta t^* = \frac{-f'_{j-1}}{f''_{j-1}}$$

Pokud $t^{j-1} + \Delta t^*$ leží na $[t^{j-1}, t^j]$, definuje Δt^* bod lokálního minima. Pokud tomu tak není, Cauchyho bod se nazývá v bodě $x(t^{j-1})$, pokud $f'_{j-1} \geq 0$ a v bodě $x(t^j)$, pokud $f'_{j-1} < 0$.

2.5 Porovnávání statistických modelů

Po odhadu parametrů statistického modelu je nutné z množiny kandidátních modelů vybrat ten nejlepší. Komplexní modely s mnoha parametry mohou dosahovat vyšších hodnot věrohodnostní funkce, ale mají riziko přeučení. Naopak jednodušší modely s menším počtem parametrů nemusí dostatečně vystihovat vztahy v datech. Pro ohodnocení modelů se proto zavádí *informační kritéria*, která měří kvalitu modelu vzhledem k počtu parametrů, případně i k velikosti trénovací datové sady.

► **Definice 2.17** (Informační kritérium). *Bud' k počet parametrů modelu a ℓ logaritmičká věrohodnostní funkce modelu. Informační kritérium IC spočteme jako:*

$$IC = 2k - p\ell$$

kde $p \in \mathbb{R}$ značí penalizační faktor.

V této práci jsou využity Akaikeho(AIC) a Bayesovské informační kritérium(BIC).

► **Definice 2.18** (Akaikeho informační kritérium).

$$AIC = 2k - 2\ell$$

AIC bylo formulováno Hirotugu Akaiem v [10]. Modely s nižší hodnotou AIC jsou považovány za lepší. Tato hodnota ale nereflektuje absolutní kvalitu modelu, ale pouze relativní kvalitu vzhledem k ostatním modelům.

► **Definice 2.19** (Bayesovské informační kritérium).

$$BIC = 2k - \log(n)\ell$$

kde $n \in \mathbb{N}$ označuje počet bodů trénovací množiny.

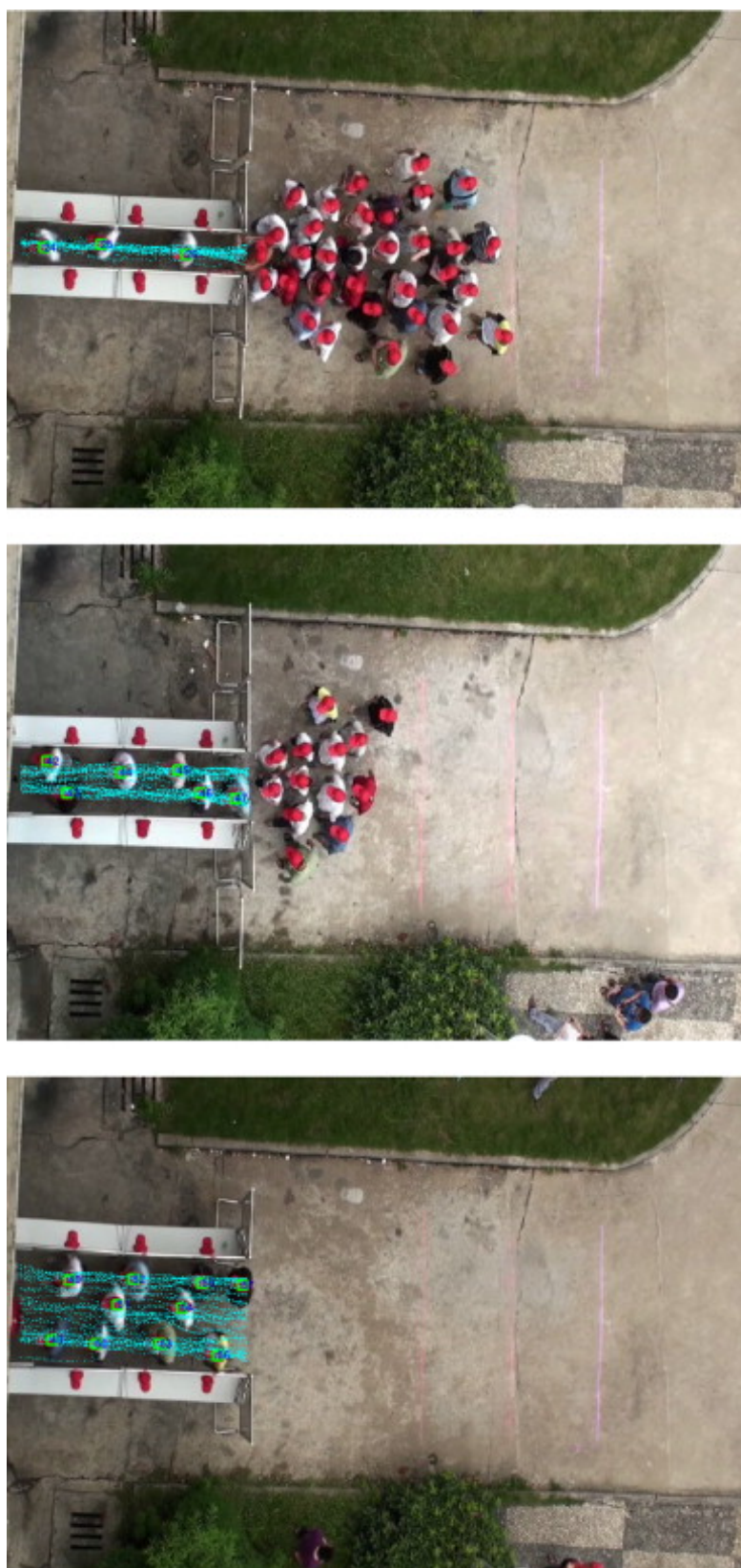
BIC bylo představeno Gideonem Schwartzem v [11]. Pokud je počet bodů trénovací množiny větší než 7, BIC přináší větší penalizační faktor než AIC, čím preferuje jednodušší modely.

Statistické modelování chování chodců

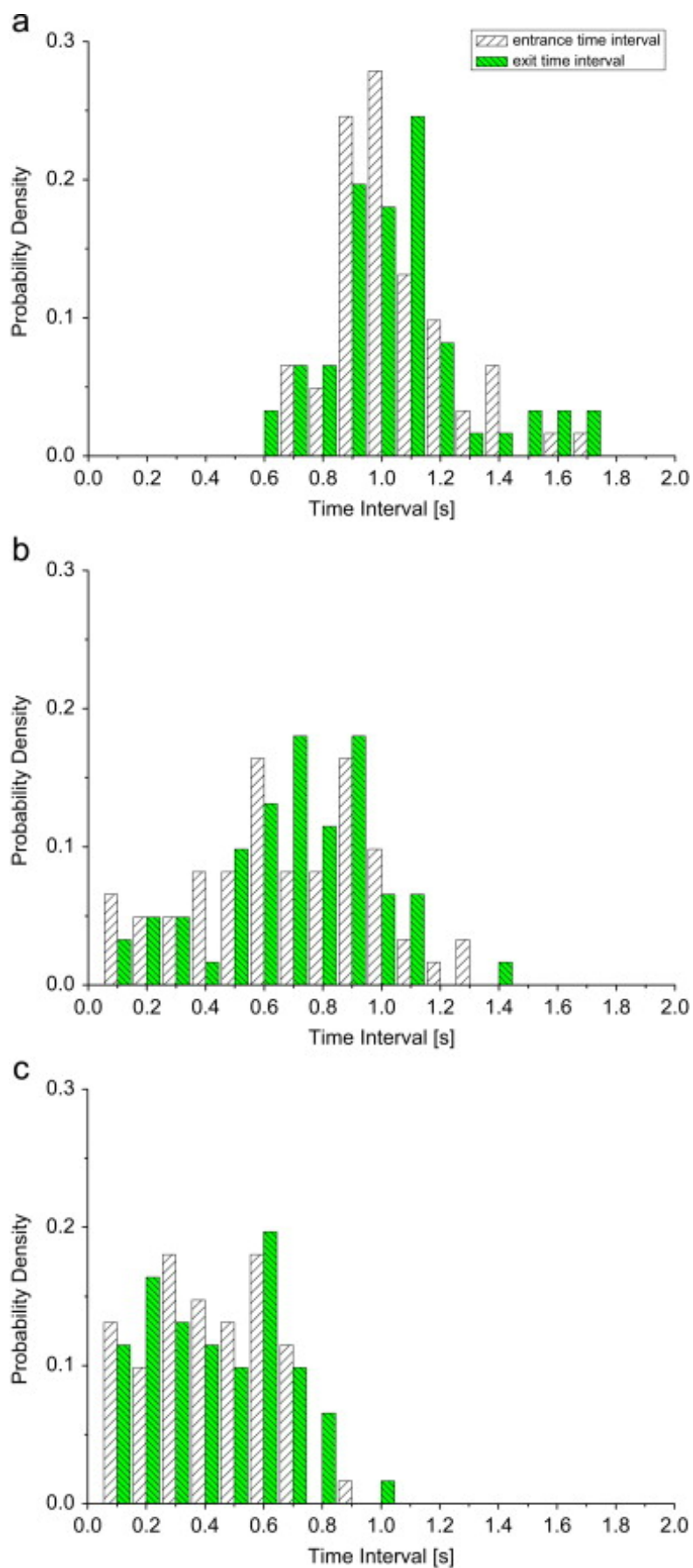
Statistické modelování chování chodců je proces, při kterém se využívají statistické metody k analýze a předpovídání chování chodců v různých situacích, jako je například průchod úzkým hrdlem, čemuž se věnuji v této práci. Data pro modelování mohou být získána například z kamerových záznamů experimentů nebo umělým generováním například pomocí celulárních modelů nebo neuronových sítí. Tato část představuje studie a statistické modely z oblasti modelování časových rozestupů.

3.1 Studie časových rozestupů

V [12] autoři zkoumají vliv počtu chodců na jejich průměrnou rychlost a vliv šířky chodby na časové rozestupy mezi chodci při vstupu a výstupu z této chodby. Šířka chodby (úzkého hrdla) se pro jednotlivé experimenty pohybovala od 0.5 m do 1.4 m, což umožňovalo řazení chodců do jedné až tří řad, jak lze vidět na obrázku 3.1. Počet řad, do kterých se chodci mohli řadit, měl velký vliv na časové rozestupy při vstupu a výstupu z chodby. Dle očekávání měli chodci nejmenší časové rozestupy v situaci se třemi řadami a největší časové rozestupy v situaci pouze s jednou řadou, jak lze vidět na obrázku 3.2.

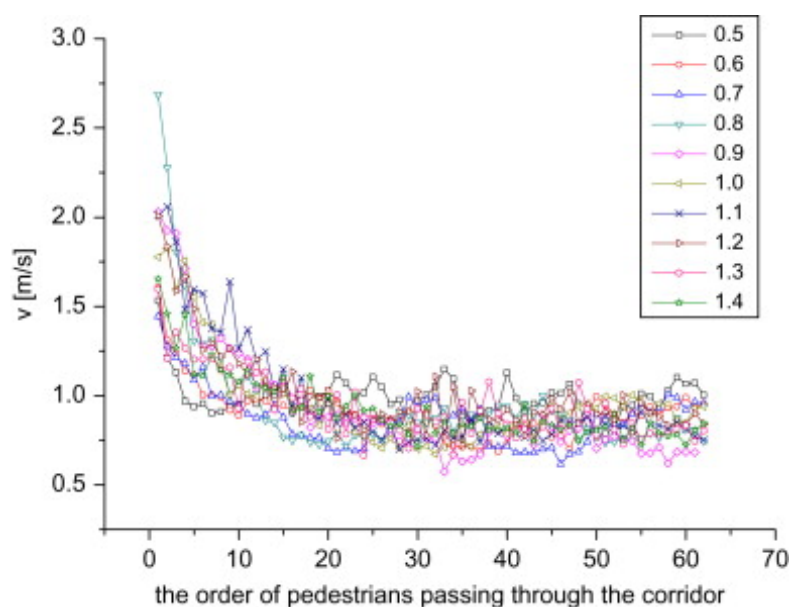


■ Obrázek 3.1 Snímek z experimentu [12]



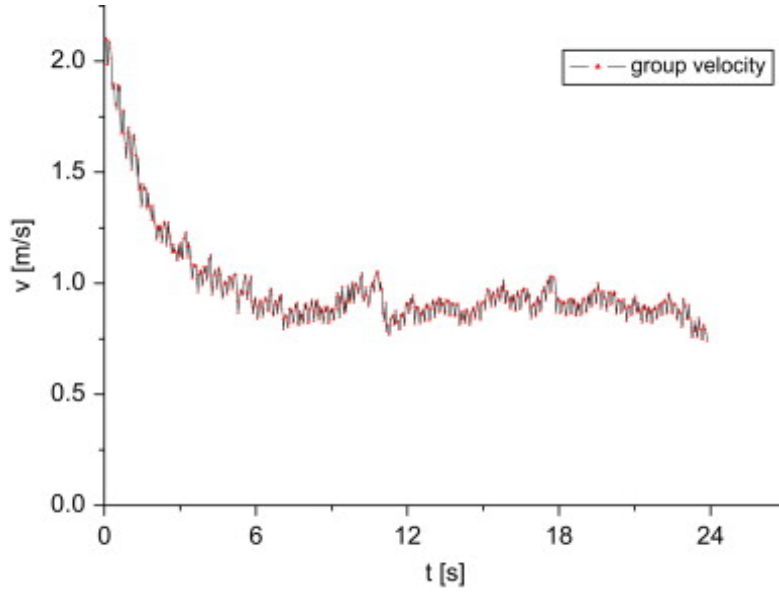
■ **Obrázek 3.2** Rozdělení časových rozestupů pro šířky chodby a) 0,5 m, b) 0,8 m, c) 1,3 m [12]

V grafu *a* na obrázku 3.2 můžeme, že většina hodnot rozestupů je mezi 1 s a 1,2 s, což autoři připisují tomu, že chodci vstupují do chodby po jednom a časové rozestupy se tedy koncentrují kolem jedné hodnoty. Graf *b* můžeme rozdělit na dvě části, první část od 0 s do 0,5 s dle autorů odpovídá situaci, kdy chodci při vstupu do chodby využívají techniku *zípování*. Druhá část grafu *b* od 0,5 s do 1,2 s odpovídá stejné strategii řazení jako v situaci pouze s jednou řadou. Třetí graf *c* lze podle autorů rozdělit na tři části. První dvě odpovídají strategiím řazení jako v situaci s jednou a dvěma řadami a třetí část grafu je situace, kdy chodbou prochází několik osob vedle sebe. Autoři ve studii dále zjistili, že při šířce chodby větší než 1,2 metrů preferovali chodci krajní řady, pravděpodobně z důvodu snížení šance kolize s ostatními chodci. Autoři předpokládají, že při ještě větších šířkách chodby mohou existovat i další strategie řazení vedoucí k menším časovým rozestupům.



■ **Obrázek 3.3** Vztah rychlosti a pořadí chodce při průchodu pro různé šířky chodby [12]

Na obrázku 3.3 můžeme vidět, že chodci, kteří chodbou procházeli mezi prvními, měli vyšší rychlost než chodci, kteří procházeli později. To autoři vysvětlují tím, že se chodci, kteří se dostali ke vchodu do chodby dříve, nezasekli v davu a mohli tedy do chodby vstoupit bez znatelného zpomalení. Dále je také zjevné, že při vstupu do nejšířší chodby měli první chodci nejvyšší rychlost. Autoři také zkoumali průměrnou rychlost skupiny chodců, kteří zrovna procházejí chodbou. Na obrázku 3.4 můžeme vidět, že díky tomu, že první z chodců nemusí zpomalovat, je tato rychlost na začátku experimentu nejvyšší a po přibližně šesti sekundách se ustálí.



■ **Obrázek 3.4** Průměrná rychlost skupiny při průchodu pro šířku chodby 1.2m [12]

3.2 Model Shafiri, Christensen, Chen

V [13] je představen model založený na gamma rozdělení, který rozlišuje dva typy chodců, vůdce a následníky. Představený model pro odhad hustoty pravděpodobnosti časových rozestupů vychází z modelu GQM (Generalized Queing Model) představeného v [14], který časové rozestupy modeluje jako součet dvou nezávislých veličin, *omezený* a *volný* rozestup. Omezené rozestupy jsou vždy menší než volné rozestupy a existuje pravděpodobnostní práh pro rozdělení rozestupů na volné a omezené. GQM má tedy formu:

$$f_{\pi}(t) = \theta_{\pi} g_{\pi}(t) + (1 - \theta_{\pi}) h_{\pi}(t),$$

kde $f_{\pi}(t)$ označuje hustotu pravděpodobnosti vůdce typu π , θ_{π} označuje podíl omezených rozestupů vůdce typu π , $g_{\pi}(t)$ je hustota pravděpodobnosti omezených rozestupů vůdce typu π a $h_{\pi}(t)$ je hustota pravděpodobnosti volných rozestupů vůdce typu π .

Dle [14] lze omezené časové rozestupy modelovat gamma rozdělením a volné časové rozestupy smíšeným gamma-exponenciálním rozdělením. Takovýto model se nazývá Gamma-GQM a pro konkrétní typ vůdce jej lze napsat jako:

$$f(t) = \theta \frac{\beta^{\alpha} t^{\alpha-1}}{\Gamma(\alpha)} e^{-\beta t} + (1 - \theta) \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda e^{-\lambda t} \int_0^t x^{\alpha-1} e^{-x(\beta-\lambda)} dx,$$

kde α označuje tvar gamma rozdělení, β měřítko gamma rozdělení, λ průměrnou intenzitu příchodu při volných podmínkách a Γ označuje gamma funkci. Parametry modelu lze získat pomocí metody maximální věrohodnosti z věrohodnostní funkce:

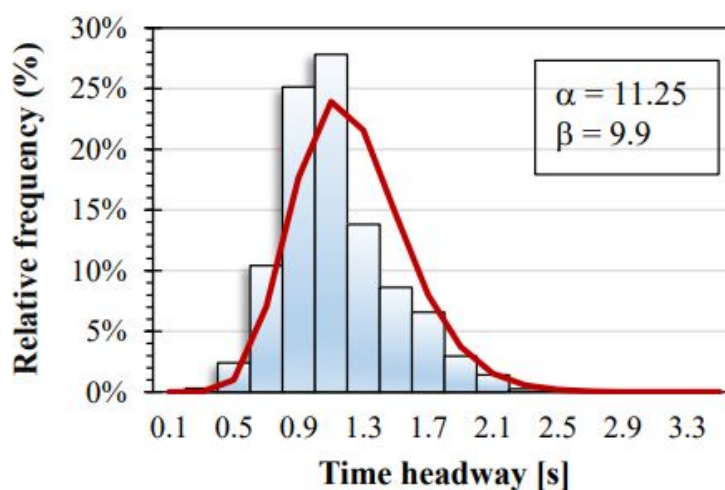
$$LL(t) = -n \ln \Gamma(\alpha) + \sum_{j=1}^n \ln \left[\theta (\beta t_j)^{\alpha-1} \beta e^{\beta t_j} + (1 - \theta) \left(\frac{\beta}{\beta - \lambda} \right)^{\alpha} \frac{\gamma[\alpha, t(\beta - \lambda)]}{\Gamma(\alpha)} \lambda e^{-\lambda t} \right],$$

kde n označuje celkový počet bodů datové sady a γ je dolní gamma funkce.

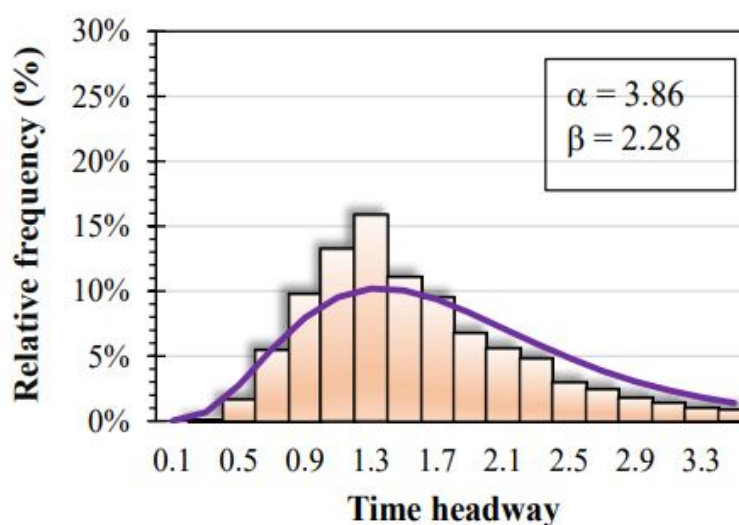
Jako vůdci jsou považováni chodci, kteří jsou následováni ostatními, za následníky autoři považují osoby bez pohybového nebo zrakového postižení. Autoři v této studii rozlišují 5 typů vůdců:

- osoby bez zrakového nebo pohybového postižení
- osoby se zrakovým postižením
- osoby chodící o holi
- osoby využívající pohybové pomůcky bez motoru, např. invalidní vozík, chodítka
- osoby využívající pohybové pomůcky s motorem

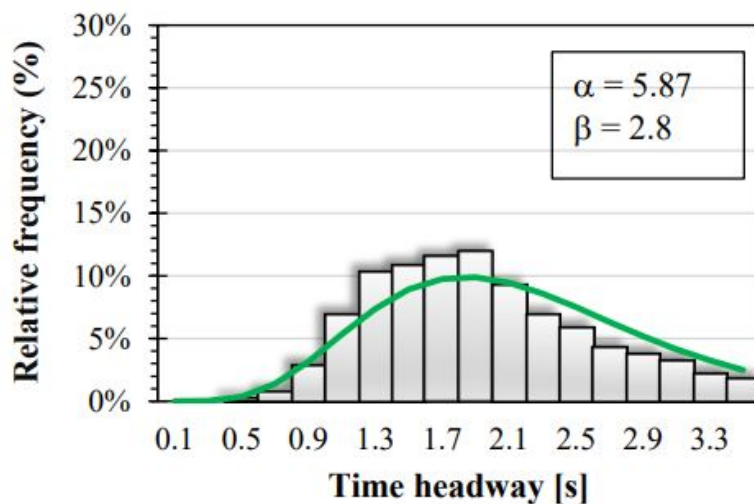
Pro každý z těchto typů vůdců zkoumali autoři rozdělení časových rozestupů od následníků. Z výsledků výzkumu vyplývá, že různé typy zdravotních postižení mohou mít významný vliv na časový rozestup a že je nutné brát tyto faktory v úvahu při návrhu a plánování veřejných prostorů.



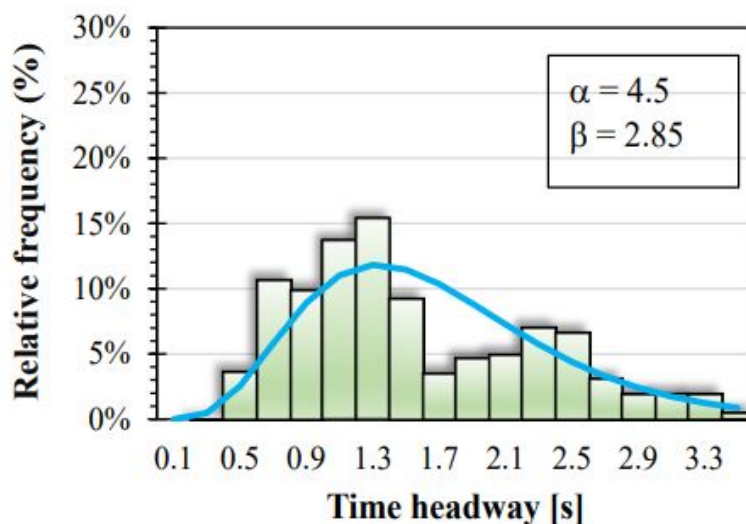
(a) osoby bez postižení



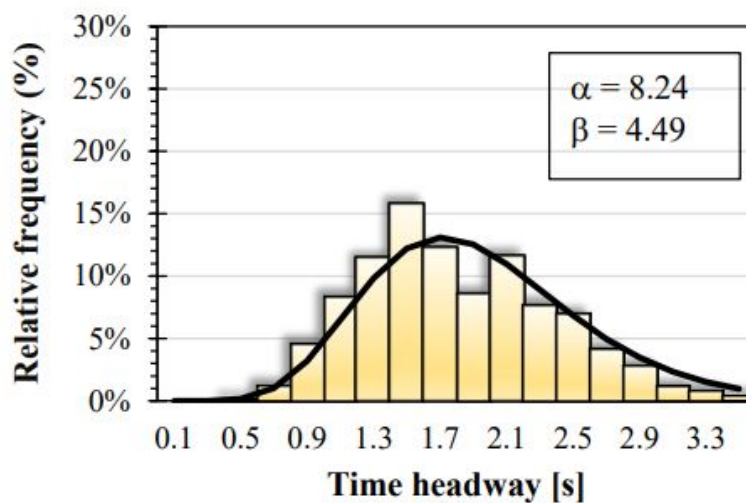
(b) osoby se zrakovým postižením



(c) osoby chodící o holi



(d) osoby využívající pohybové pomůcky bez motoru



(e) osoby využívající pohybové pomůcky s motorem

■ **Obrázek 3.5** Rozdělení časových rozestupů pro jednotlivé typy vůdců. Křivka znázorňuje odhad modelu. [13]

3.3 Model Bode a Codling

V této práci reprodukuji výsledky z [3], kde autoři Bode a Codling představili statistický model založený na gamma rozdělení pro predikci střední hodnoty časových rozestupů při průchodu chodců úzkým hrdlem.

Nechť t_p a t_{p-1} označuje časy průchodu úzkým hrdlem dvou po sobě jdoucích chodců. Dále předpokládejme, že tímto úzkým hrdlem může v jednu chvíli procházet pouze jeden chodec. Označme Δt_p časový rozestup těchto chodců. Nechť T je náhodná veličina o hodnotách Δt_p . Představený model má tvar $T \sim \Gamma(\mu_{p-1}, \sigma^2)$, kde Γ značí gamma rozdělení se střední hodnotou μ_{p-1} a rozptylem σ^2 . Rozptyl je v rámci modelu považován za konstantní. Model popisuje vliv pozic chodců v čase $p-1$ na časový rozestup Δt_p . Pro odhad střední hodnoty náhodné veličiny T model využívá pět vysvětlujících faktorů.

Prvním faktorem je konstanta:

$$m_0 = \alpha_1$$

α_1 je parametrem modelu. Tento faktor představuje očekávaný časový rozestup bez ohledu na pozice chodců.

Druhý faktor m_1 zachycuje hustotu chodců před vstupem do úzkého hrdla. Nechť $\langle d \rangle_k$ označuje průměrnou vzdálenost $k = 2, 3, 4, 5$ chodců od středu průchodu úzkým hrdlem. Faktor m_1 poté definujeme jako:

$$m_1 = (\alpha_2 \langle d \rangle_k - \alpha_3)^2$$

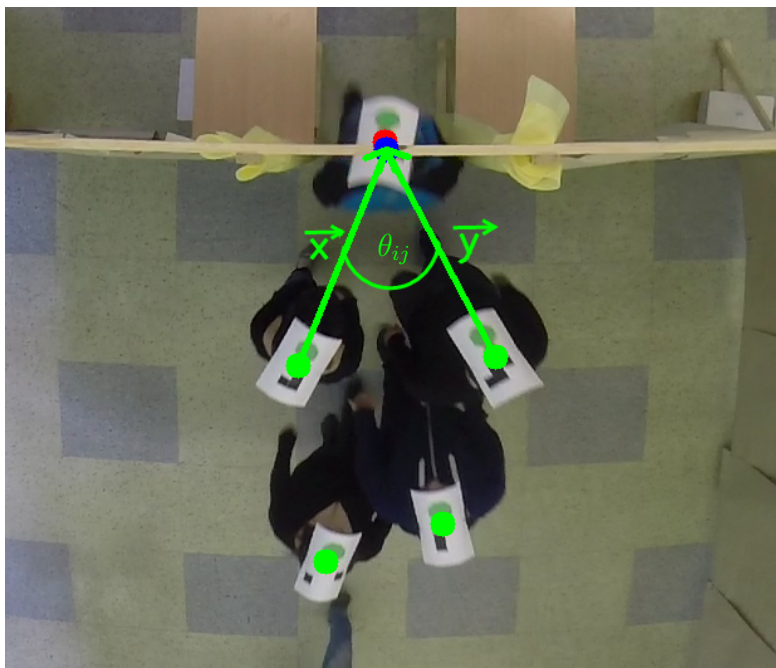
k, α_2, α_3 jsou parametry modelu.

Faktory m_2 a m_3 popisují vliv dvou nejbližších chodců ke vstupu do úzkého hrdla. Označme tyto chodce i a j , označme d_{ij} rozdíl ve vzdálenosti těchto chodců od středu vstupu do úzkého hrdla a θ_{ij} úhel vektorů \vec{x} a \vec{y} , kde \vec{x} je vektor směřující od chodce i ke středu vstupu do úzkého hrdla a \vec{y} je vektor směřující od chodce j ke středu vstupu do úzkého hrdla. Vysvětlující faktory jsou poté definovány jako:

$$m_2 = (\alpha_4 d_{ij} - \alpha_5)^2$$

$$m_3 = (\alpha_6 \theta_{ij} - \alpha_7)^2$$

$\alpha_4, \dots, \alpha_7$ jsou parametry modelu.



■ **Obrázek 3.6** Vektory \vec{x} , \vec{y} a úhel θ_{ij}

Pátý vysvětlující faktor m_4 je definován jako:

$$m_4 = (\alpha_8 d_i - \alpha_9)^2$$

d_i je vzdálenost nejbližšího chodce ke středu vstupu do úzkého hrdla a α_8, α_9 jsou parametry modelu.

Využitím všech možných kombinací faktorů m_1, m_2, m_3 a m_4 je sestaveno 16 různých statistických modelů a faktor m_0 je zahrnut do každého z těchto modelů. Predikci střední hodnoty μ_{p-1} získáme jako součet hodnot faktorů daného modelu. Pokud například model obsahuje faktory m_0, m_1 a m_4 , predikovaná střední hodnota bude $\mu_{p-1} = m_0 + m_1 + m_4$. Porovnáním výsledných modelů lze poté zjistit, které z vysvětlujících faktorů nejlépe popisují časové rozestupy Δt_p .

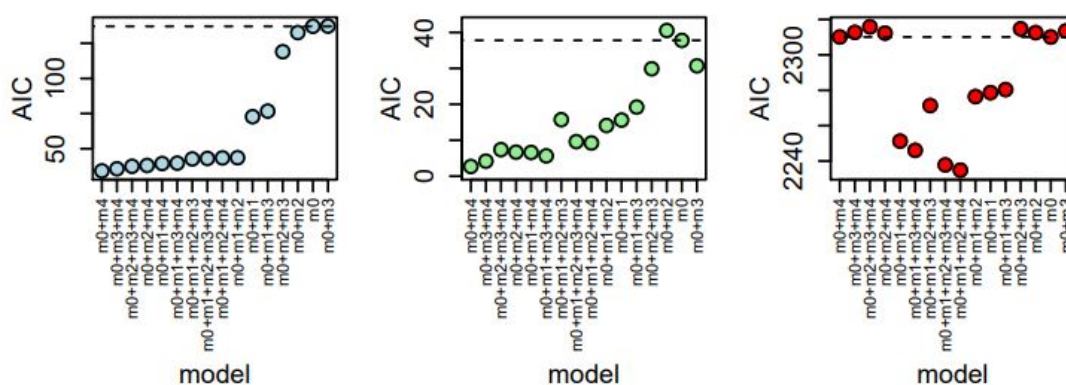
Pro odhad parametrů modelů je využita metoda maximální věrohodnosti. Označme hustotu pravděpodobnosti gamma rozdělení se střední hodnotou μ_{p-1} a rozptylem σ^2 v bodě Δt_p jako $f_\Gamma(\Delta t_p; \mu_{p-1}, \sigma^2)$. Věrohodnostní funkci poté získáme jako:

$$\mathcal{L} = \prod_{p \in T} f_\Gamma(\Delta t_p; \mu_{p-1}, \sigma^2)$$

kde T označuje trénovací množinu. Jednotlivé body trénovací množiny jsou považovány za nezávislé. Pro každý z modelů jsou nalezeny parametry maximalizující věrohodnostní funkci \mathcal{L} a poté jsou porovnány pomocí AIC. Parametry modelů autoři odhadovali na třech datových sadách. První dvě pocházely z experimentů se šířkou úzkého hrdla 0,6 metru, třetí byla získána simulací průchodu davu úzkým hrdlem o šířce 0,5 metru.



■ **Obrázek 3.7** Snímky z experimentů a provedená simulace[3]



■ **Obrázek 3.8** Vypočtené hodnoty AIC pro experimenty a simulaci[3]

Autoři své výsledky považují za výchozí bod pro další výzkumy, které by mohly zkoumat další faktory ovlivňující časové rozestupy. Jedním z faktorů, který autoři navrhuji, jsou interakce mezi různými sociálními skupinami chodců, např. přátelé. Dále autoři navrhuji rozšířit model o analýzu širších úzkých hrdel, které jsou chodci preferovány a bývají frekventovanější. Jako nedostatek modelu autoři považují absenci analýzy situace za úzkým hrdlem, která může být také jedním z faktorů ovlivňující časové rozestupy, například pokud dojde k zácpě za tímto úzkým hrdlem.

Kapitola 4

Praktická část

4.1 Experiment

Data analyzovaná v této práci pocházejí z experimentu E4[15], který proběhl v rámci ČVUT. Experimentu se zúčastnilo kolem 80 studentů, jejichž úkolem bylo na signál vstoupit do místnosti a co nejrychleji ji opustit úzkým hrdlem o šířce 0,6 metru. Běhání a fyzický kontakt s jinými účastníky byl zakázán. Každý z účastníků byl vybaven čepičkou s unikátním obrazovým identifikátorem.



■ **Obrázek 4.1** Snímek z experimentu E4

Z videozáznamů, které byly z experimentu pořízeny, extrahoval Vít Pospíšil v rámci své bakalářské práce [16] trajektorie jednotlivých účastníků, ze kterých budu v této práci vycházet. Pro extrakci těchto trajektorií využil čepičkové identifikátory, které měli účastníci na sobě. Celkem mám k dispozici trajektorie z 11 běhů tohoto experimentu a dále budu používat i datovou sadu obsahující data ze všech jedenácti běhů dohromady. Trajektorií chodce se rozumí seznam pozic x a y chodce seřazený vzestupně dle čísla snímku.

4.2 Předzpracování dat

Poskytnutá data jsou v csv formátu, kde se na každém řádku nachází číslo trajektorie následované seznamem snímků a pozic, kde byla tato trajektorie detekována. Snímky jsou číslovány dle jejich pořadí ve videu, pozice x a y jsou dány v pixelech.

■ **Tabulka 4.1** Formát poskytnutých dat, soubor data/E4.R_pruchod.2.txt

trajektorie	snímek x y	snímek x y	...
1	1776 969 1053	1777 970 1050	...
2	1842 899 1054	1843 899 1052	...
⋮	⋮	⋮	⋮

Tato data jsem si načel do dvou slovníků. První slovník *time_to_points* má jako klíče čísla snímků a hodnoty jsou tvořeny seznamem pozic, které byly v daném čase detekovány. Druhý slovník *trajectories* má jako klíče čísla trajektorií a jako hodnoty seznam pozic pro tuto trajektorii včetně čísla snímku, ve kterém byla daná pozice detekována.

Abych mohl extrahovat data pro modely, bylo nutné zjistit, ve kterých časech došlo k průchodu chodce úzkým hrdlem. Identifikoval jsem tedy oblast před vstupem do úzkého hrdla. Tato oblast je vpravo ohraničena zdí a nahoře samotným úzkým hrdlem.



■ **Obrázek 4.2** Červeně označený střed úzkého hrdla, modře zeď



■ **Obrázek 4.3** Označená oblast před vstupem do úzkého hrdla

Pro každou trajektorii jsem našel dvojici pozic p_{t-1}, p_t , kde t označuje čas pozice a pro které platí, že v čase $t - 1$ se chodec nacházel v identifikované oblasti, ale v čase t již ne. Jako čas průchodu tohoto chodce jsem označil čas t . Pseudokód je v algoritmu 1.

Algoritmus 1: Detekce časů průchodu

```

1 Function in_front_of_bottleneck(point)
2   return point.y < BOTTLENECK.y and point.x < WALL.START.x
3 Function get_passage_times(trajectories{})
4   times ← []
5   for key in trajectories do
6     points ← trajectories[key]
7     N ← length(points)
8     for i ← 1 to N do
9       if in_front_of_bottleneck(points[i - 1]) and not in_front_of_bottleneck(points[i])
10        then
11          times.append(points[i].time.frame)
12   return times

```

Po identifikaci časů průchodu jsem z trajektorií odstranil pozice, které se nenachází v oblasti před úzkým hrdlem a které neodpovídají žádnému z časů průchodu, jelikož tyto pozice nejsou potřeba k extrakci parametrů. Statistické modely, jejichž parametry budu odhadovat, totiž předpokládají, že časový rozestup chodce je ovlivněn situací před úzkým hrdlem v čase průchodu předchozího chodce. Pseudokód je v algoritmu 2.

Algoritmus 2: Filtrace trajektorií

```

1 Function get_passage_times(time_to_points{}, passage_times[])
2   result ← {}
3   for time in passage_times do
4     result[time] ← {}
5     for point in time_to_points[time] do
6       if in_front_of_bottleneck(point) then
7         result[time].append(point)
8   return result

```

4.3 Odhad parametrů

Pro odhad parametrů modelů jsem využil funkci *minimize* z balíčku *SciPy*[5]. Funkce *minimize* má jako parametry objektivní funkci, která se má minimalizovat, vektor počátečních hodnot parametrů x_0 a případně omezení pro tyto parametry. Objektivní funkcí je v mém případě logaritmická věrohodnostní funkce ℓ , kterou jsem počítal jako součet logaritmu:

$$\ell = \ln \prod_{p \in T} f_{\Gamma}(\Delta t_p; \mu_{p-1}, \sigma^2) = \sum_{p \in T} \ln f_{\Gamma}(\Delta t_p; \mu_{p-1}, \sigma^2)$$

a parametry jsou $\alpha_i, \alpha_j, \dots$ příslušející vysvětlujícím faktorům obsaženým v daném modelu a celé číslo $k \in \{2, 3, 4, 5\}$, pokud model obsahuje faktor m_1 . Posledním faktorem je rozptyl σ^2 . Například pro model obsahující faktory m_0 a m_4 budou parametry $\alpha_1, \alpha_8, \alpha_9, \sigma^2$. Protože potřebuji najít hodnoty parametrů, při kterých objektivní funkce ℓ nabývá maxima, budu hledat hodnoty, při kterých nabývá minima funkce $-\ell$.

Funkce *minimize* se všemi parametry pracuje jako s reálným číslem, z toho důvodu nemůže být parametry k součástí vektoru x_0 . Místo toho spouštím pro každou hodnotu parametru k funkci *minimize* zvlášť a vybírám k , pro které je výsledná hodnota objektivní funkce nejnižší. 4

Algoritmus 4: Odhad parametru k

```

1 best_k ← null
2 minimum ← ∞
3 for k ← 2 to 5 do
4   value ← minimize( $-\ell$ )
5   if value < minimum then
6     best_k ← k
7   minimum ← value

```

Počáteční hodnoty vektoru x_0 jsem našel pomocí metody sítí. Pro každý parametr jsem náhodně zvolil spodní mez l , horní mez u a délku kroku s a v tomto intervalu jsem našel hodnoty parametrů, pro které funkce $-\ell$ nabývá minima. Pokud byla nalezená hodnota parametru rovna spodní nebo horní mezi, snížil, resp. zvýšil jsem hodnotu příslušné meze. Pokud byla nalezená hodnota v daných mezích, zmenšil jsem délku kroku s . Tento proces jsem opakoval, dokud docházelo ke značnému snížení hodnoty objektivní funkce $-\ell$ a nalezené hodnoty jsem použil jako počáteční hodnoty vektoru x_0 pro funkci *minimize*.

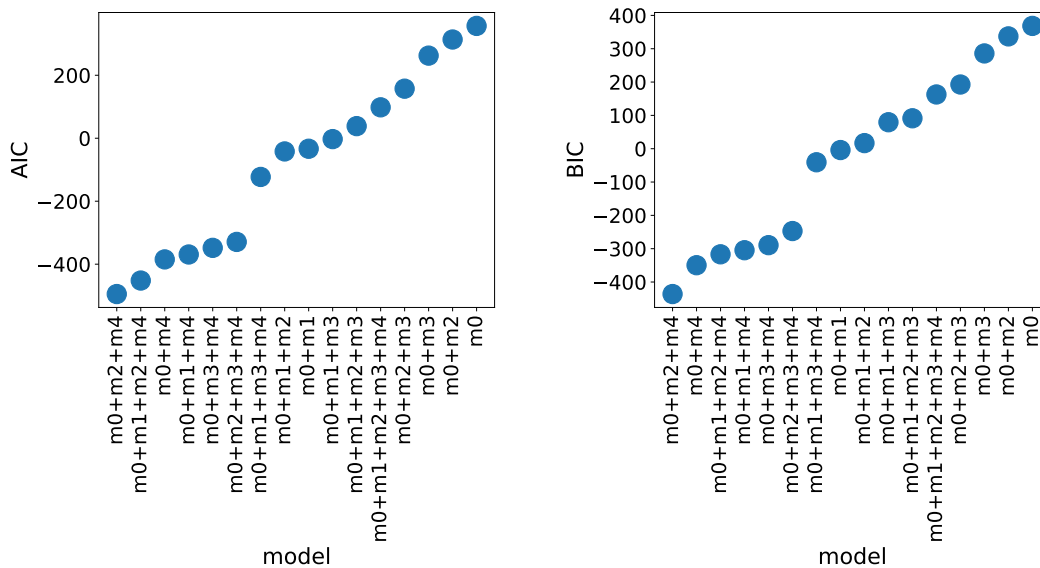
Hodnoty parametrů nalezené funkcí *minimize* jsem uložil do cvs souborů ve složce *models* ve formátu obsažené faktory, $-\ell, \alpha_1, \dots, \sigma^2, k$, kde ... označuje ostatní parametry α ve vstoupném pořadí a hodnota k je vynechána, pokud model neobsahuje faktor m_1 . Obsahem složky *models* je tedy celkem 12 souborů, 11 pro modely trénované na jednotlivých bězích a jeden soubor pro modely trénované na agregovaných datech ze všech běhů. Každý soubor obsahuje 16 řádků, na každém řádku jsou hodnoty parametrů pro jeden konkrétní model.

■ **Tabulka 4.3** Nalezené hodnoty parametrů, soubor *models/all.txt*, hodnoty jsou zaokrouhleny na 4 desetinná místa

faktory	$-\ell$	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	α_9	σ^2	k
0	176.3701	0.7544	-	-	-	-	-	-	-	-	0.0732	-
01	-21.6931	0.7005	0.0013	-0.0008	-	-	-	-	-	-	0.0638	2
02	152.7668	0.6995	-	-	0.0016	0.3049	-	-	-	-	0.0718	-
03	127.1902	0.7253	-	-	-	-	0.1808	0.3122	-	-	0.0728	-
04	-198.4115	0.3063	-	-	-	-	-	-	0.0010	-0.5603	0.0558	-
012	-30.9704	0.6871	0.0013	-0.0099	0.0003	0.1027	-	-	-	-	0.0637	2
013	-15.3845	0.5842	0.0010	-0.1114	-	-	0.1024	0.3927	-	-	0.0660	2
014	-195.6030	0.3595	0.00002	0.0260	-	-	-	-	0.0010	-0.5048	0.0549	4
023	72.6942	0.5311	-	-	0.0016	0.3523	0.0992	0.4638	-	-0.0704	-	-
024	-257.4206	0.0926	-	-	0.0015	0.4037	-	-	0.0009	-0.6376	0.0522	-
034	-184.0740	0.0925	-	-	-	-	0.0988	0.4517	0.0009	-0.6190	0.0557	-
0123	10.2272	0.3213	1.3143 * 10 ⁻⁸	0.0341	0.0020	0.5287	0.1750	0.6364	-	-	0.0706	4
0124	-248.9713	0.1917	0	0.0426	0.0015	0.3781	-	-	0.0010	-0.5697	0.0526	5
0134	-75.4581	0.1668	0.0005	0.0418	-	-	0.1875	0.5866	0.0008	-0.5206	0.0586	5
0234	-178.6377	0.1613	-	-	0.0015	0.3734	0.1945	0.5675	0.0007	-0.4994	0.0604	-
01234	38.0753	0.2155	0.0010	0.0061	0.0002	0.0585	0.1511	0.5810	0	0.5419	0.0723	2

Zhodnocení výsledků

V této části prezentuji výsledky modelů, jejichž parametry byly odhadnuty na datovém souboru agregující data ze všech běhů experimentu. Výsledky modelů na datech z jednotlivých běhů lze nalézt v příloze. Po nalezení parametrů jednotlivých modelů jsem spočítal jejich AIC a BIC, viz obrázek 5.1. Jednotlivé modely jsou seřazeny dle hodnot AIC vzestupně, můžeme ovšem vidět, že pro obě informační kritéria je toto pořadí stejné. U obou informačních kritérií je také patrný nárůst hodnoty daného kritéria mezi modely $m_0 + m_2 + m_3 + m_4$ a $m_0 + m_1 + m_3 + m_4$. Modely bychom tedy mohli rozdělit do dvou skupin (prvních 6 modelů a zbylých 10 modelů), ve kterých dochází k mírnému nárůstu hodnot informačního kritéria a tyto skupiny jsou odděleny zmíněným skokem mezi hodnotami informačního kritéria.



(a) AIC výsledných modelů

(b) BIC výsledných modelů

■ **Obrázek 5.1** Informační kritéria výsledných modelů

Z vypočtených hodnot AIC lze vidět, že nejlepším modelem pro agregovaný soubor je model obsahující faktory m_0 , m_2 a m_4 , které byly definovány jako

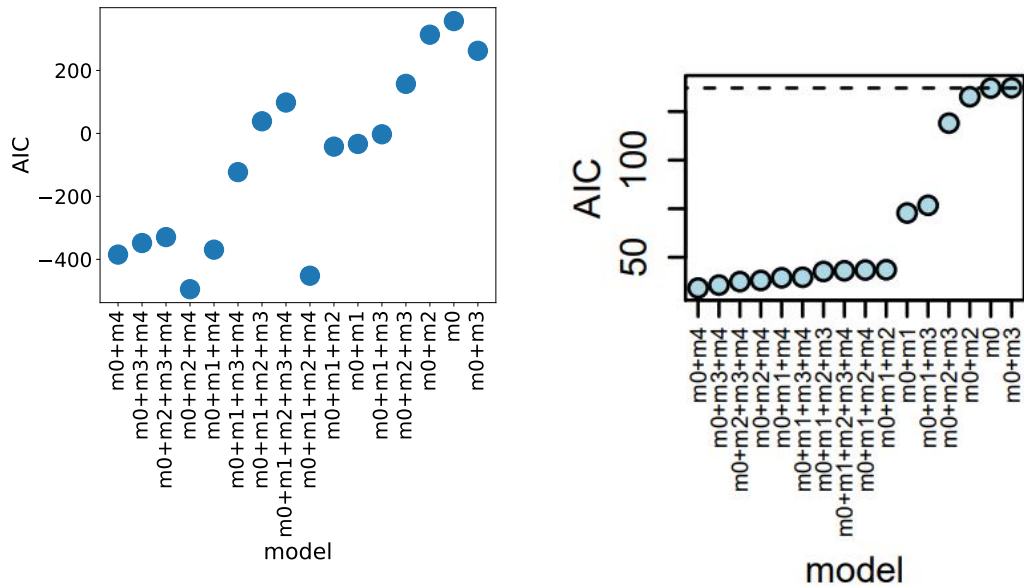
$$\begin{aligned} m_0 &= \alpha_1 \\ m_2 &= (\alpha_4 d_{ij} - \alpha_5)^2 \\ m_4 &= (\alpha_8 d_i - \alpha_9)^2, \end{aligned}$$

kde α_1 , α_4 , α_5 , α_8 a α_9 byly parametry modely, d_i vzdálenost nejbližšího chodce ke středu úzkého hrdla a d_{ij} rozdíl ve vzdálenosti dvou nejbližších chodců ke středu úzkého hrdla. Tím se odlišují od výsledků v [3], kde byl nejlepší model s faktory m_0 a m_4 a model $m_0 + m_2 + m_4$ byl až čtvrtý nejlepší. Nižších hodnot AIC než model $m_0 + m_2 + m_4$ dosáhly v reprodukováném článku kromě nejlepšího $m_0 + m_4$ ještě modely $m_0 + m_3 + m_4$ a $m_0 + m_2 + m_3 + m_4$.

Model $m_0 + m_4$, který byl v reprodukováném článku nejlepší, je v mé práci třetím nejlepším, lépe dopadl kromě již zmíněného nejlepšího $m_0 + m_2 + m_4$ ještě model $m_0 + m_1 + m_2 + m_4$. Parametr m_1 z tohoto modelu je definován jako

$$m_1 = (\alpha_2 \langle d_k \rangle - \alpha_3)^2,$$

kde α_2 , α_3 a $k = 2, 3, 4, 5$ jsou parametry modelu a $\langle d_k \rangle$ je průměrná vzdálenost k nejbližších chodců ke středu úzkého hrdla. Lze tedy říci, že hustota chodců před úzkým hrdlem je v modelech v mé práci důležitým faktorem pro predikci střední hodnoty rozestupu.



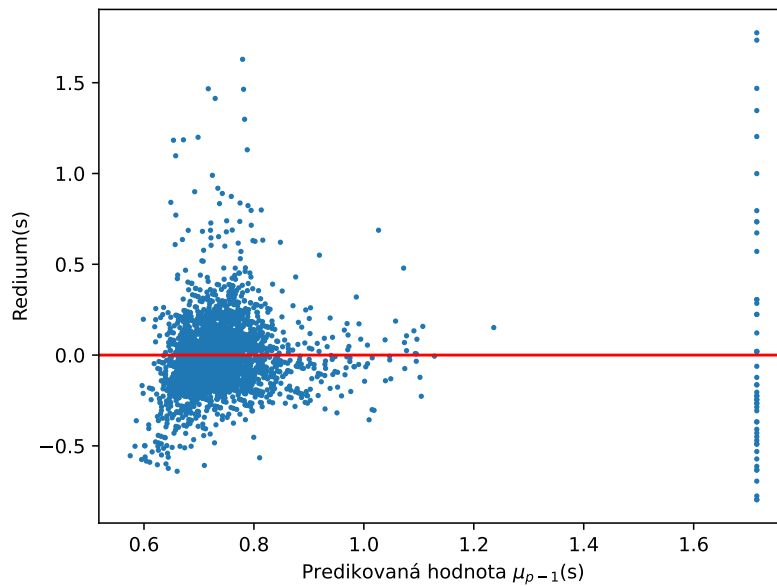
(a) AIC výsledných modelů, seřazeno dle pořadí z reprodukováného článku

(b) AIC modelů z reprodukováného článku[3]

■ **Obrázek 5.2** Porovnání výsledných modelů s modely z reprodukováného článku

Pro model s nejnižším AIC jsem počítal residua r predikovaných hodnot μ_{p-1} a skutečných hodnot Δt .

$$r(\Delta t, \mu_{p-1}) = \Delta t - \mu_{p-1}$$

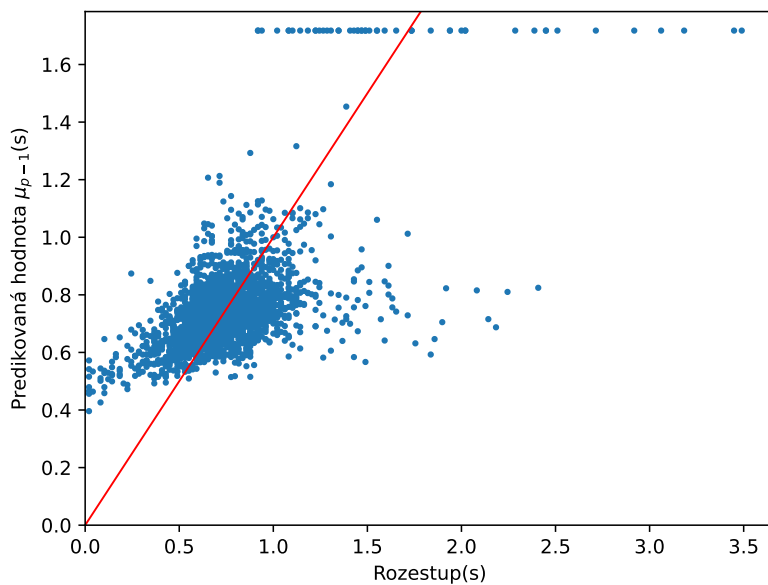


■ **Obrázek 5.3** Residua gamma modelu

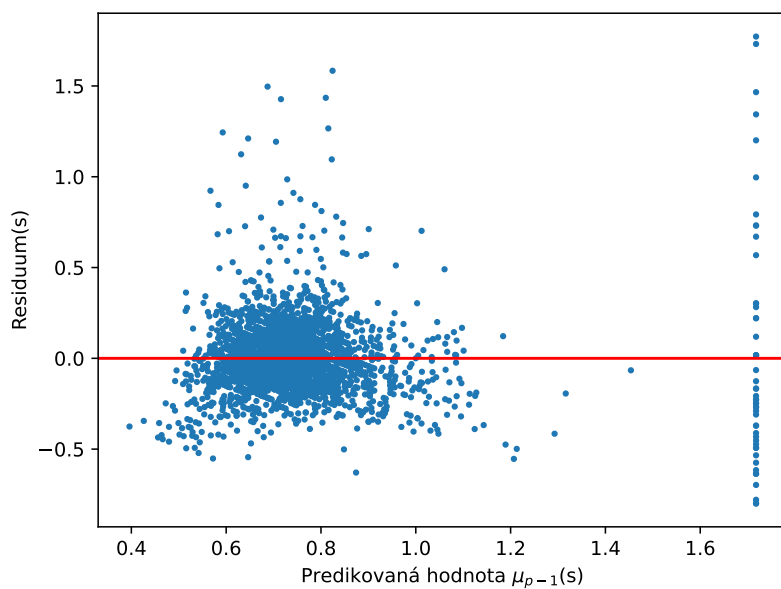
Z grafu lze vidět, že pro některé chodce model predikuje hodnotu ≈ 1.71 sekund, která se výrazně liší od ostatních predikovaných hodnot. Jedná se o situace, kdy se po průchodu chodce žádný další chodec nenacházel ve sledované oblasti a tedy byly hodnoty d_2 , d_3 , d_4 a d_5 , θ_{ij} nahrazeny umělou hodnotou.

Kromě gamma modelu jsem implementoval i predikci pomocí kvadratické regrese. K tomu jsem využil třídy *PolynomialFeatures* pro transformaci dat a *LinearRegression* pro samotnou predikci z balíčku *scikit-learn*[17]. Vysvětlujícími proměnnými jsou vzdálenost nejbližšího chodce d_i a rozdíl ve vzdálenosti dvou nejbližších chodců d_{ij} , což jsou hodnoty obsažené ve vysvětlujících faktorech výsledného gamma modelu s nejnižším AIC. Rozdílem oproti predikci gamma modelem je, že u gamma modelu předpokládáme, že šum v datech má gamma rozdělení, u regresního modelu předpokládáme, že tento šum má normální rozdělení. Natrénovaný model kvadratické regrese má tvar:

$$\mu_{p-1} = 0.0035d_i - 0.0015d_{ij} - 2.5031 * 10^{-6}d_i^2 - 2.4588 * 10^{-6}d_id_{ij} + 3.0915 * 10^{-6}d_{ij}^2$$



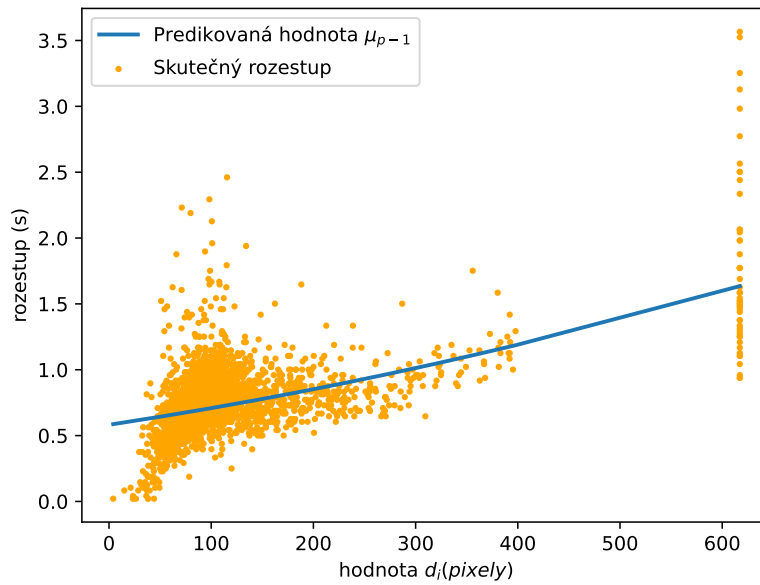
■ **Obrázek 5.4** Predikce μ_{p-1} pomocí kvadratické regrese



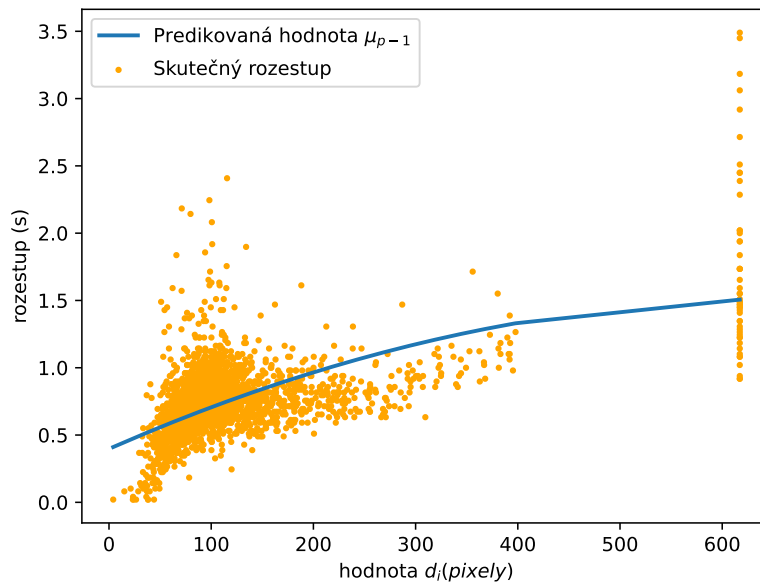
■ **Obrázek 5.5** Residua kvadratické regrese

Pro gamma model i kvadratickou regresi jsem spočítal R^2 statistiku, která vyjadřuje, jakou část rozptylu závislé proměnné dokáže model vysvětlit. Pro gamma model je tato hodnota ≈ 0.35 , pro regresní model ≈ 0.38 . Regresní model tedy vysvětluje větší část rozptylu závislé proměnné μ_{p-1} .

Modely jsem dále analyzoval pomocí řezů skrz hodnoty d_i a d_{ij} . Pro řez skrz hodnoty d_i jsem hodnotu d_{ij} nastavil na průměrnou hodnotu d_{ij} v rámci všech dat, pro řez skrz hodnoty d_{ij} jsem nastavil hodnoty d_i na průměrnou hodnotu d_i v rámci všech dat. Na obrázcích 5.6 a 5.7 můžeme vidět srovnání skutečných hodnot rozestupů a křivky odhadnuté jednotlivými modely.

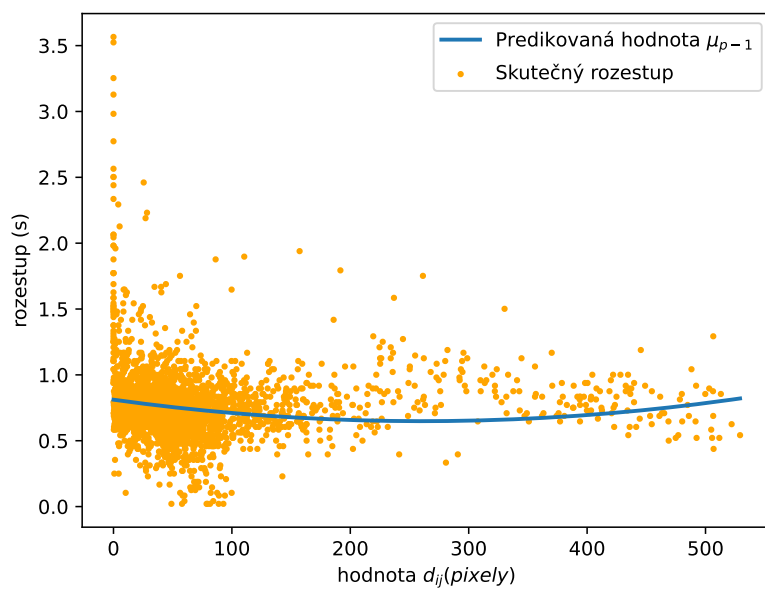


(a) Gamma model

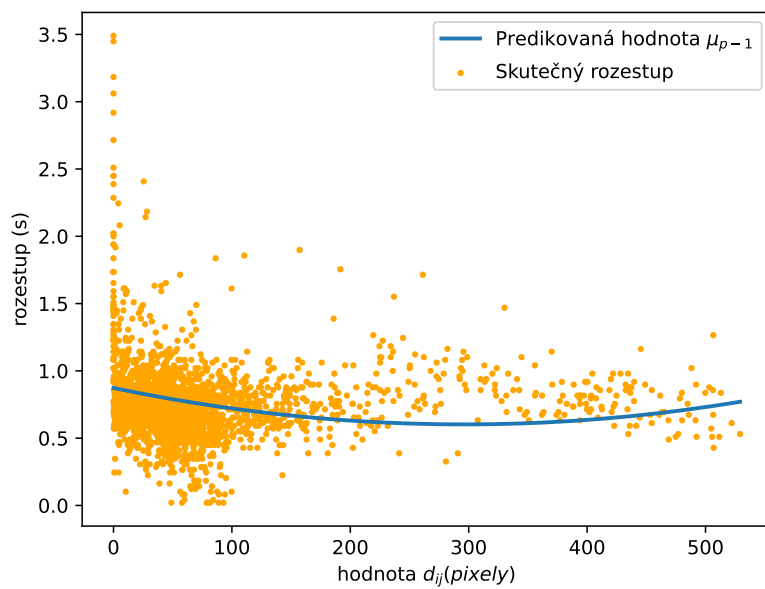


(b) Kvadratická regrese

■ **Obrázek 5.6** Řez skrz hodnoty d_i



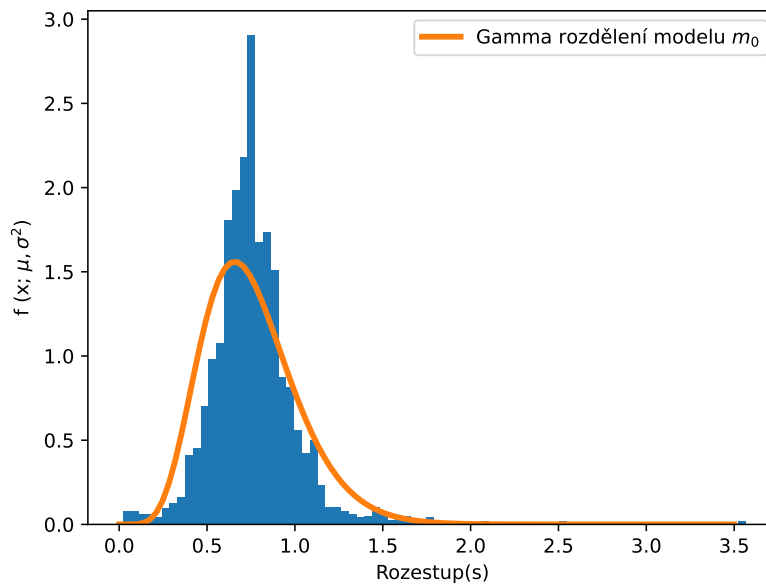
(a) Gamma model



(b) Kvadratická regrese

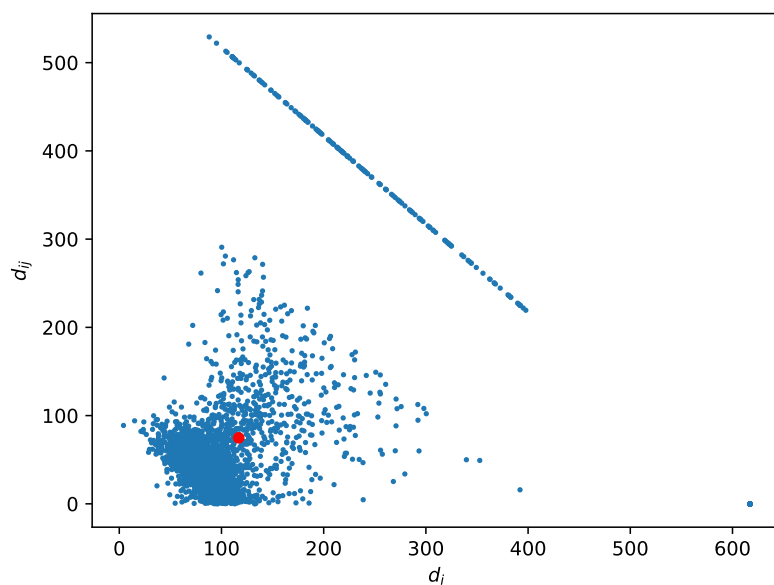
■ **Obrázek 5.7** Řez skrz hodnoty d_{ij}

Gamma rozdělení odhanuté modelem m_0 můžeme přímo porovnat s histogramem skutečných rozestupů. Parametry tohoto rozdělení jsou $\mu \approx 0.75$ a $\sigma^2 \approx 0.07$. Pro toto rozdělení jsem pomocí funkce *goodness-of-fit* z balíčku SciPy[5] provedl test dobré shody, jehož p-hodnota vyšla 0.0001. Statistický test, že hodnoty rozestupů odpovídají tomuto gamma rozdělení, bychom pro hladinu významnosti $\alpha > 0.0001$ zamítli.



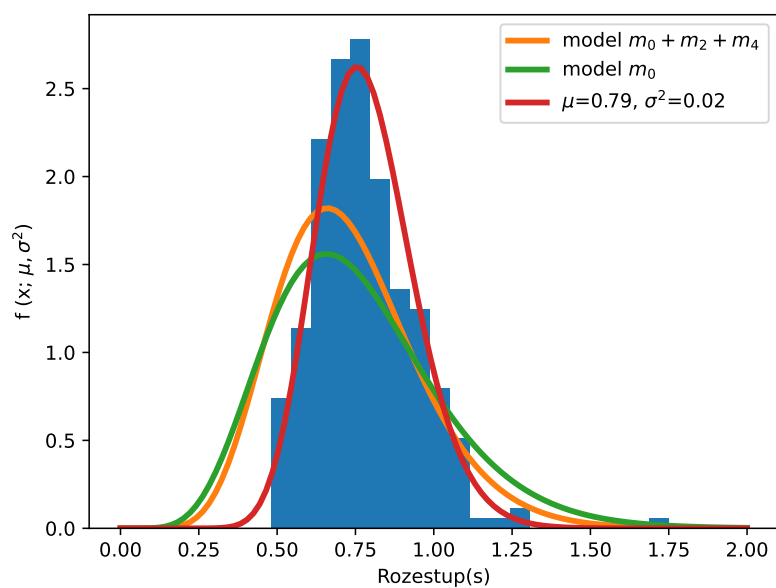
■ **Obrázek 5.8** Porovnání skutečných rozestupů a odhadnutého rozdělení

Na obrázku 5.9 můžeme vidět, jak jsou hodnoty d_i a d_{ij} , které model $m_0 + m_2 + m_4$ využívá, rozprostřeny. V horní části grafu můžeme vidět skupinu bodů, které jsou daleko od ostatních. Jedná se o body, u kterých byla známa skutečná hodnota vzdálenosti nejbližšího chodce, ale vzdálenost druhého chodce již byla doplněna. Červeně označený bod P označuje průměrné hodnoty daných hodnot d_i a d_{ij} .



■ **Obrázek 5.9** Rozdělení hodnot d_i a d_{ij}

Dále jsem skutečné rozestupy bodů, které se nacházejí v intervalu $\pm 30\%$ průměrných hodnot d_i a d_{ij} , porovnal s gamma rozděleními odhadnutými modelem m_0 , modelem $m_0 + m_2 + m_4$ pro bod P označený v obrázku 5.9 a gamma rozdělením odhadnutým pouze pro tyto body metodou maximální věrohodnosti. Můžeme vidět, že data nejlépe vystihuje červená křivka gamma rozdělení odhadnutého pro tato data a data nejméně kopíruje křivka modelu m_0 .



■ **Obrázek 5.10** Porovnání skutečných rozestupů z okolí bodu P

Kapitola 6

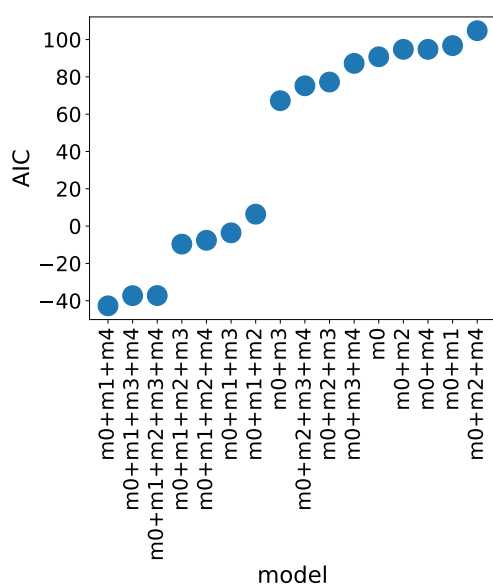
Závěr

Cílem mé bakalářské práce bylo reprodukovat statistické modely z článku [3] pro modelování střední hodnoty časových rozestupů chodců při průchodu úzkým hrdlem a odhadnout jejich parametry. Toho jsem docílil zpracováním dat z evakuačních experimentů pořádaných ČVUT. Tyto data jsem předzpracoval, abych získal potřebná data pro odhad parametrů daných modelů. Po odhadnutí parametrů metodou maximální věrohodnosti jsem poté výsledky srovnal s výsledky z článku [3] a provedl diskuzi rozdílů v těchto výsledcích. Některé rozdíly byly pravděpodobně způsobeny doplňováním chybějících hodnot, jelikož reprodukováný článek nezmiňuje, jak byly chybějící hodnoty řešeny.

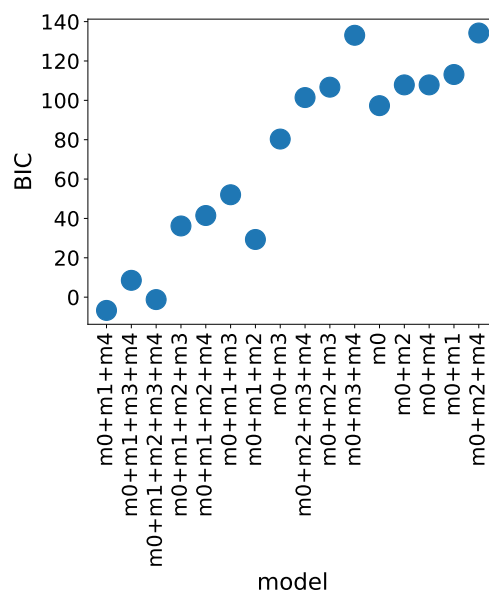
Přínosem mé bakalářské práce jsou informace o tom, jaké faktory ovlivňují časové rozestupy při průchodu úzkým hrdlem. Na její výsledky lze navázat například podrobnější analýzou výsledných modelů.

..... Příloha A

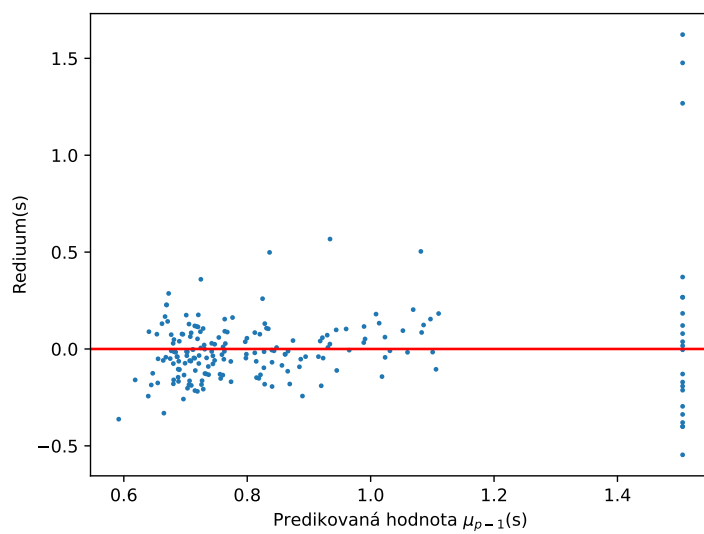
Výsledky jednotlivých běhů experimentu



(a) AIC výsledných modelů

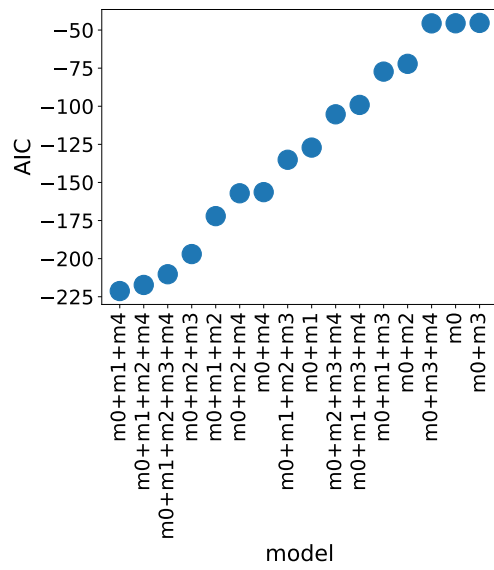


(b) BIC výsledných modelů

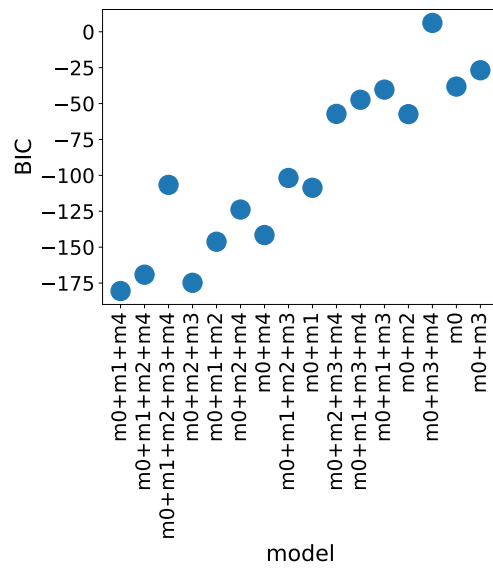


(c) Residua gamma modelu

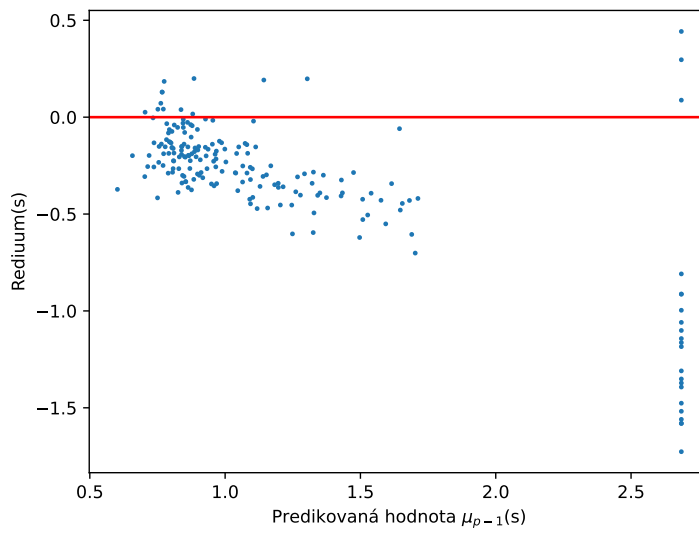
■ Obrázek A.1 Průchod 2



(a) AIC výsledných modelů

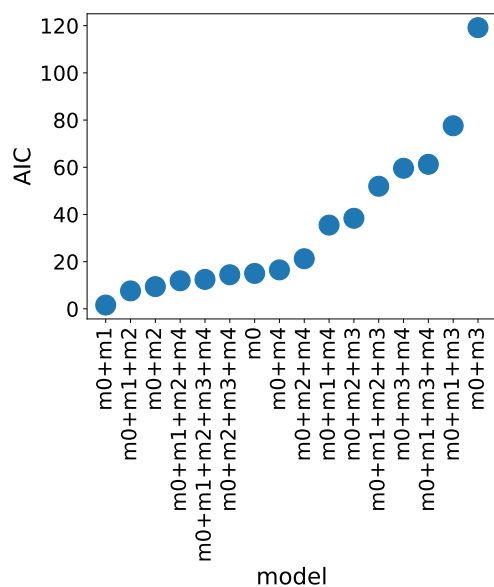


(b) BIC výsledných modelů

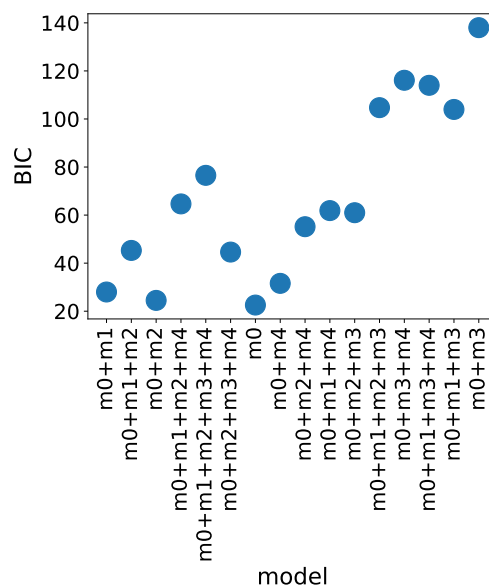


(c) Residua gamma modelu

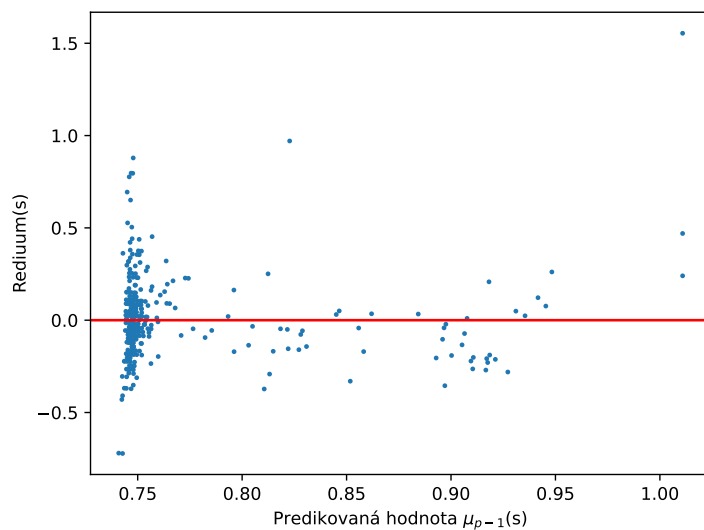
■ **Obrázek A.2** Průchod 3



(a) AIC výsledných modelů

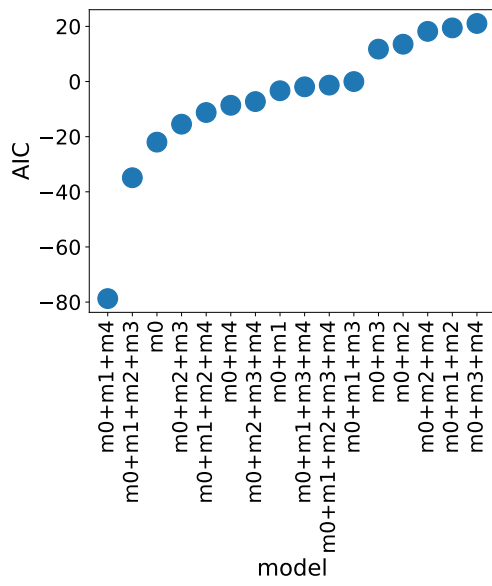


(b) BIC výsledných modelů

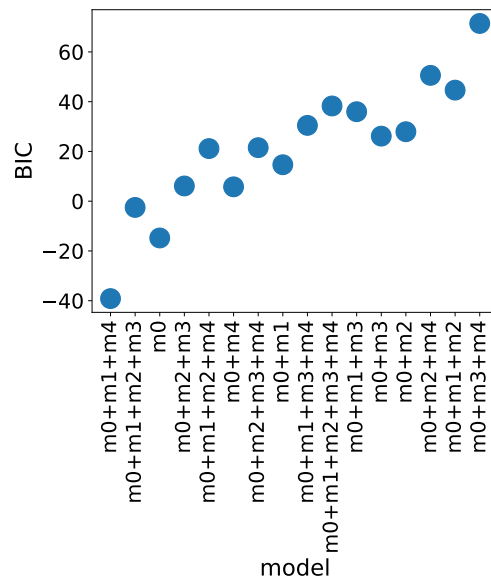


(c) Residua gamma modelu

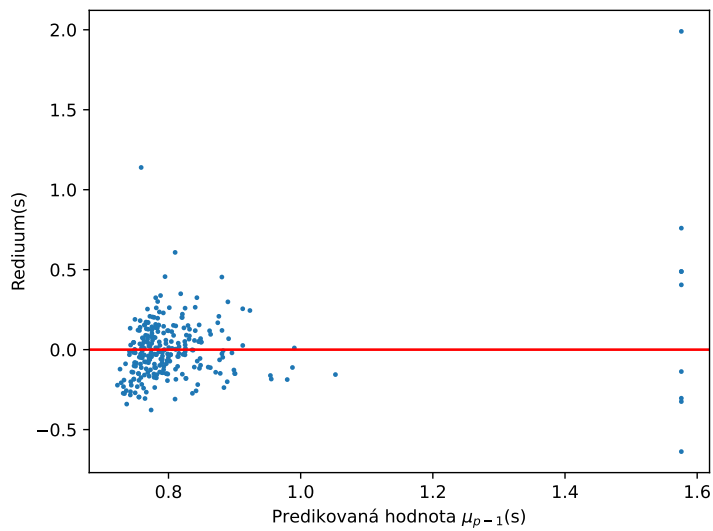
■ **Obrázek A.3** Průchod 4



(a) AIC výsledných modelů

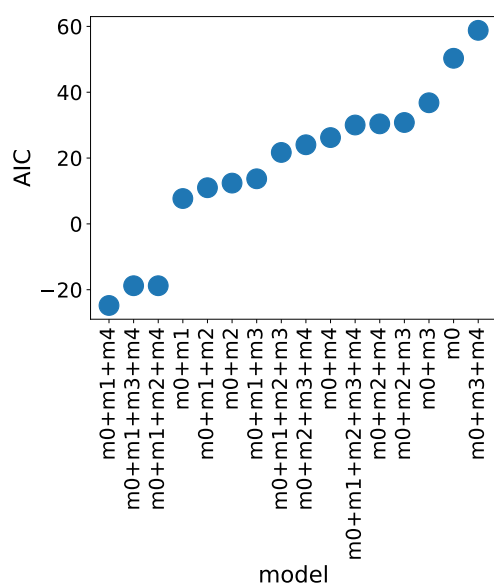


(b) BIC výsledných modelů

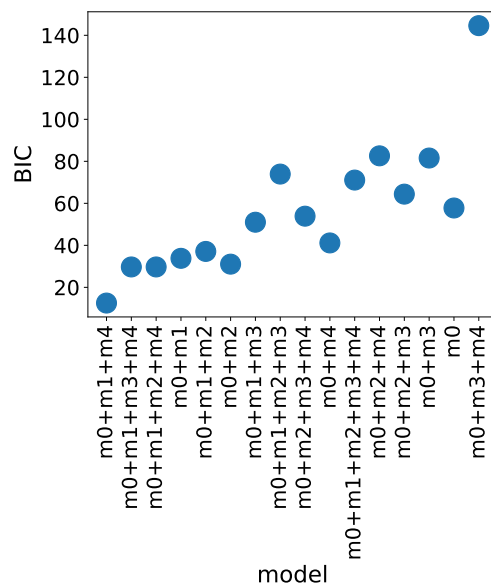


(c) Residua gamma modelu

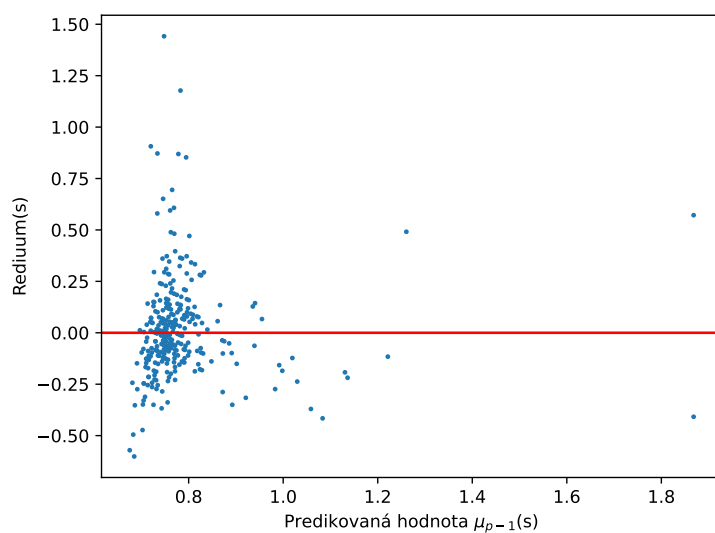
■ **Obrázek A.4** Průchod 5



(a) AIC výsledných modelů

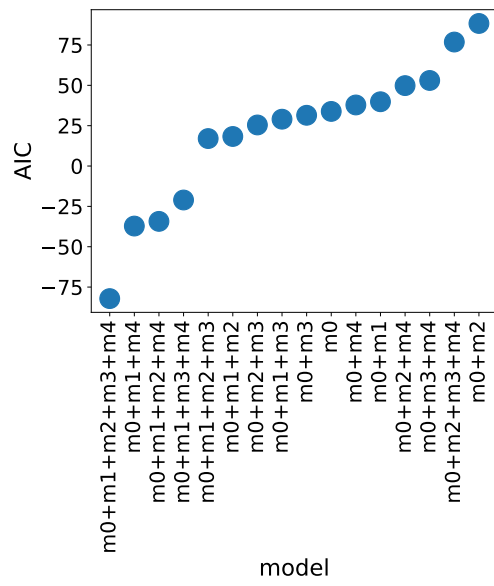


(b) BIC výsledných modelů

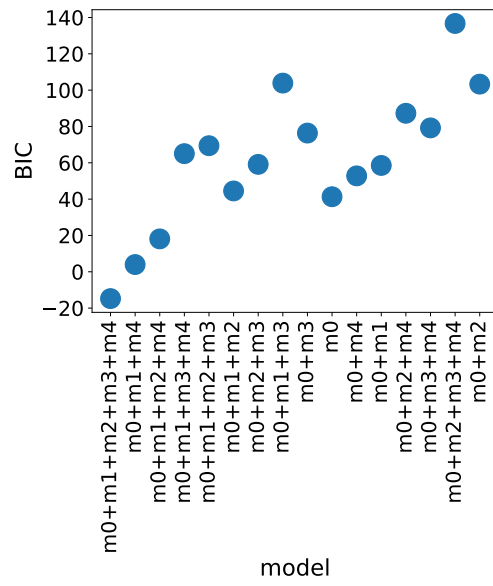


(c) Residua gamma modelu

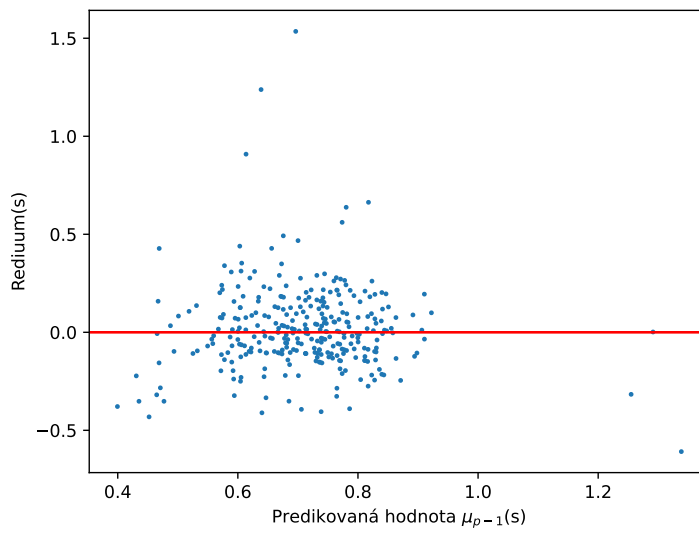
■ **Obrázek A.5** Průchod 6



(a) AIC výsledných modelů

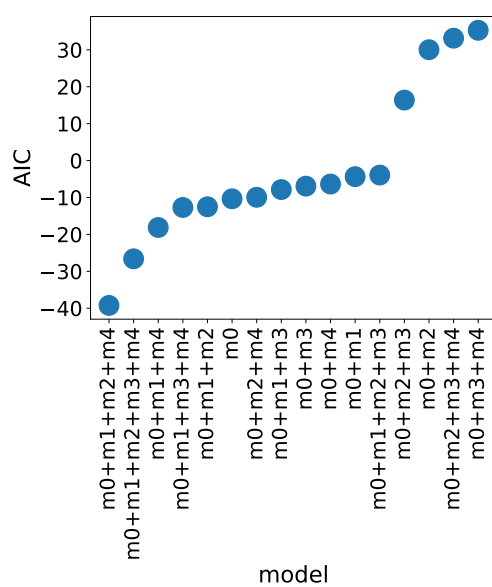


(b) BIC výsledných modelů

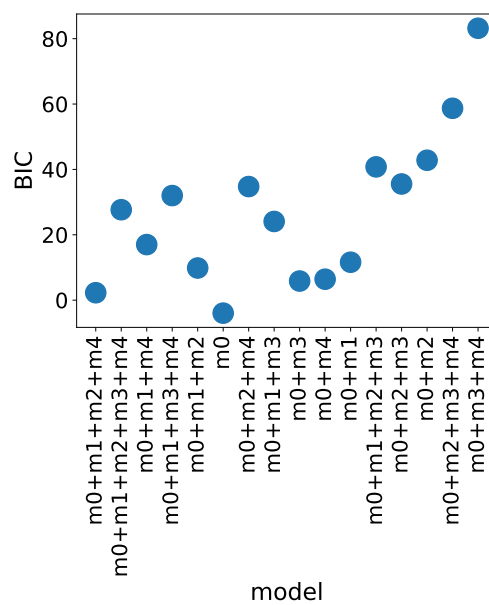


(c) Residua gamma modelu

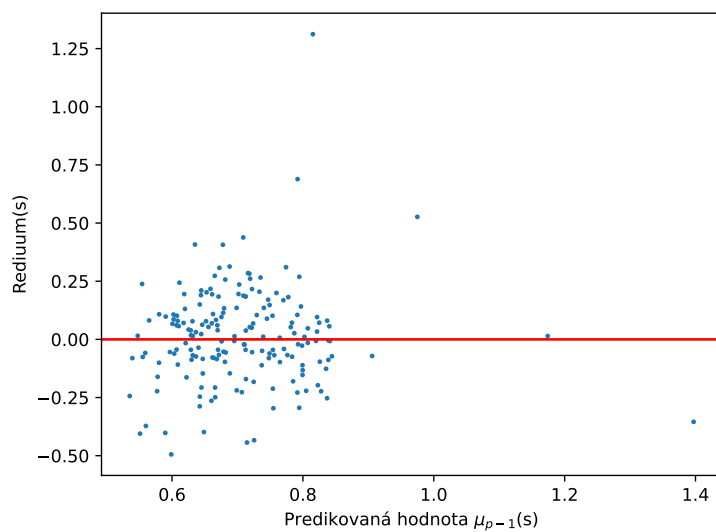
■ **Obrázek A.6** Průchod 7



(a) AIC výsledných modelů

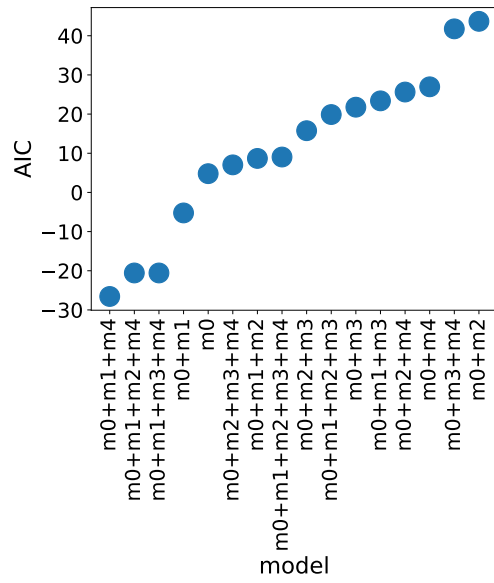


(b) BIC výsledných modelů

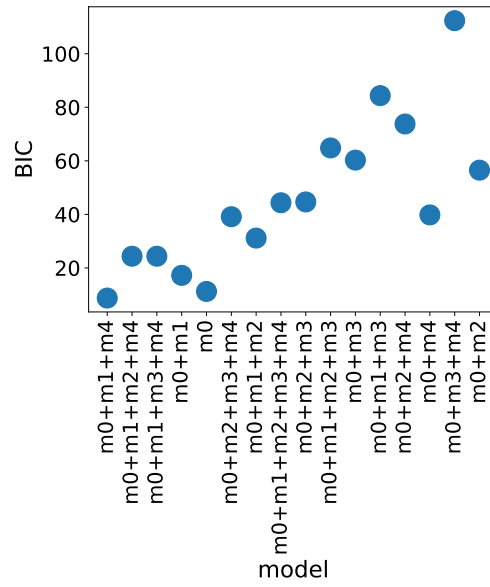


(c) Residua gamma modelu

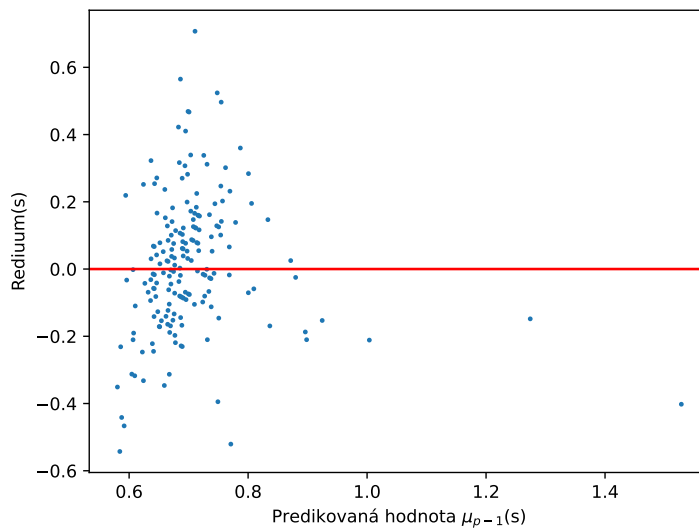
■ **Obrázek A.7** Průchod 8



(a) AIC výsledných modelů

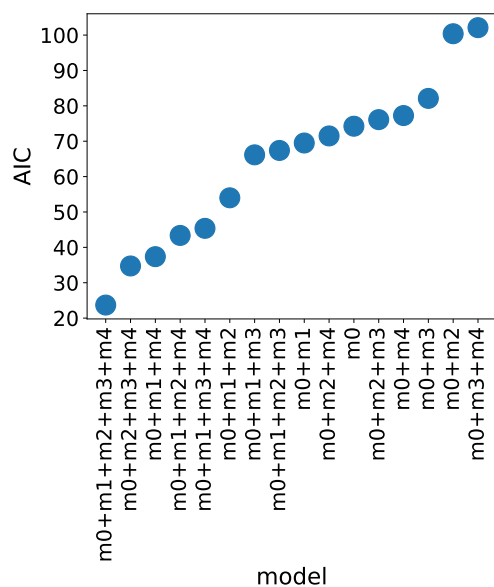


(b) BIC výsledných modelů

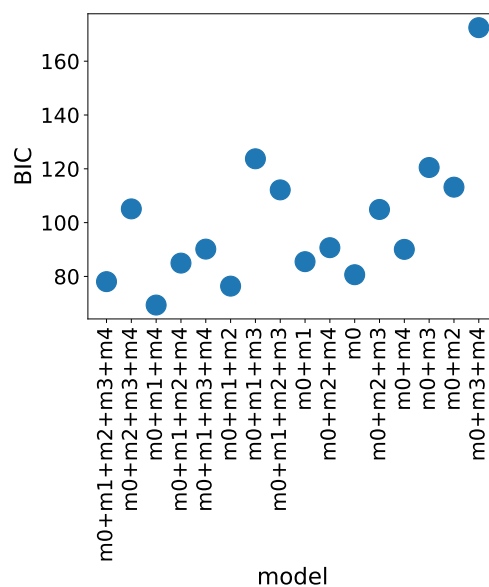


(c) Residua gamma modelu

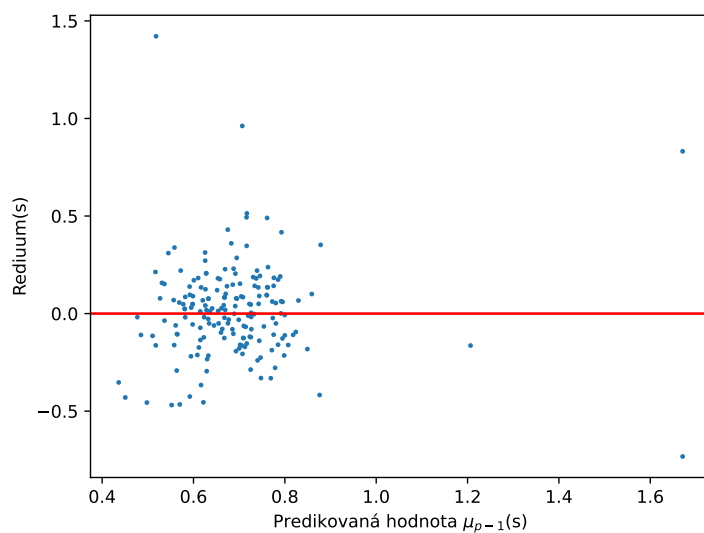
■ **Obrázek A.8** Průchod 9



(a) AIC výsledných modelů

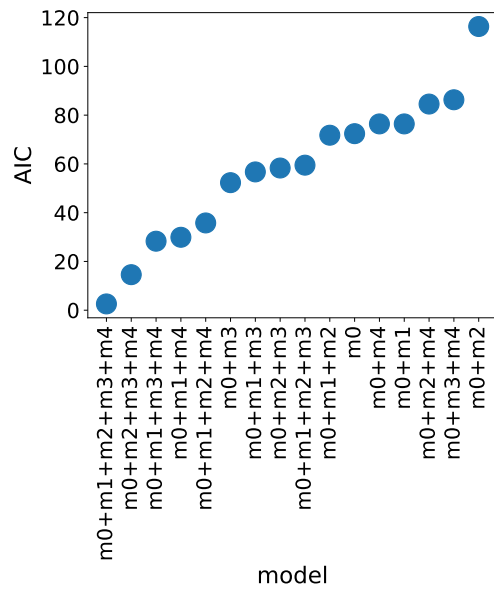


(b) BIC výsledných modelů

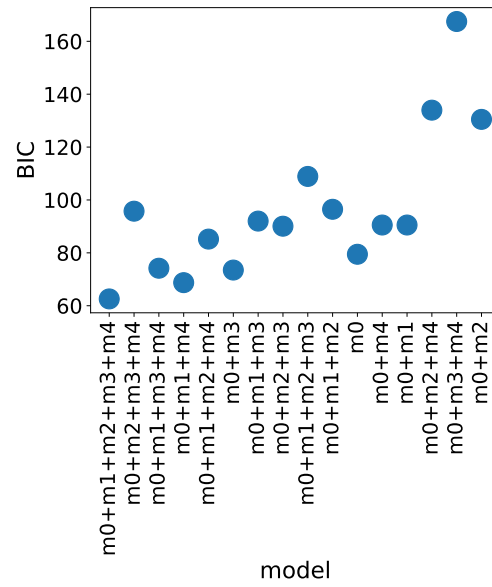


(c) Residua gamma modelu

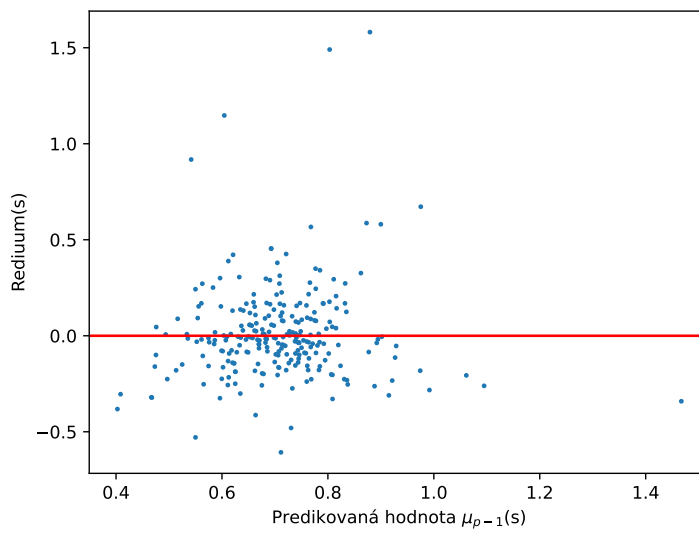
■ Obrázek A.9 Průchod 10



(a) AIC výsledných modelů

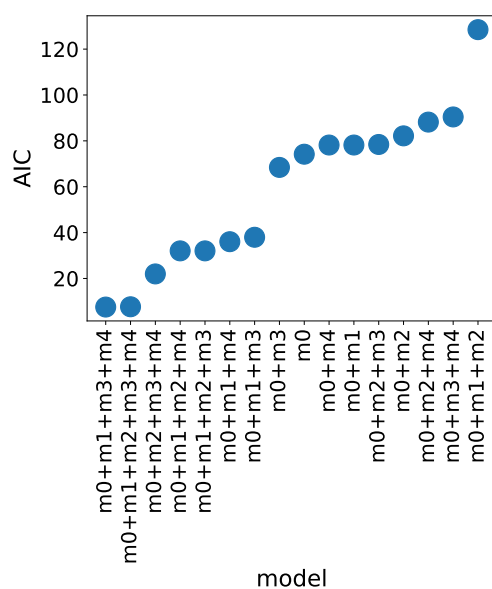


(b) BIC výsledných modelů

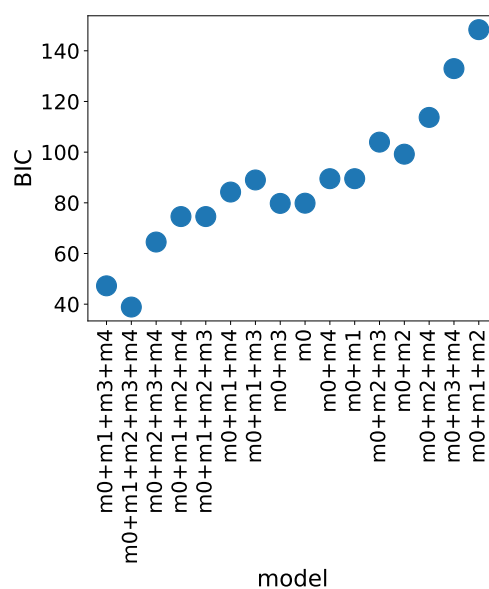


(c) Residua gamma modelu

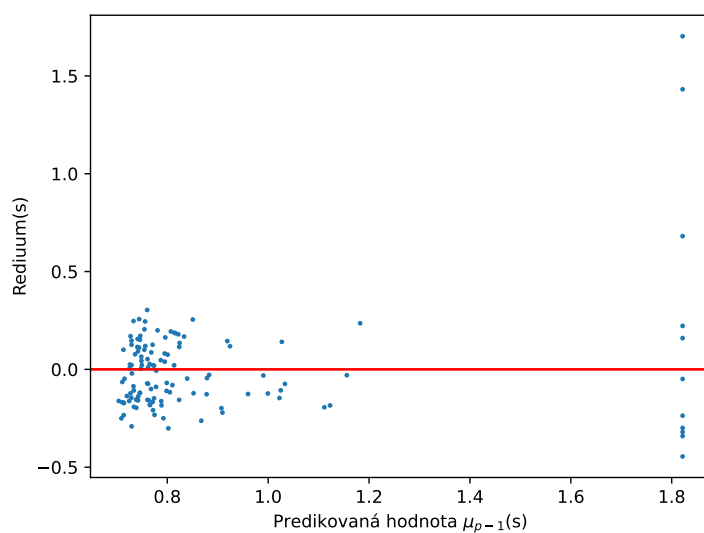
■ **Obrázek A.10** Průchod 11



(a) AIC výsledných modelů



(b) BIC výsledných modelů



(c) Residua gamma modelu

■ **Obrázek A.11** Průchod 12

Bibliografie

1. BLAŽEK, Rudolf B.; KOTECKÝ, Roman; VAŠATA, Daniel; HRABÁKOVÁ, Jitka; NOVÁK, Petr. *BI-PST – Pravděpodobnost a statistika*. FIT ČVUT, 2022.
2. KLOUDA, Karel; LOPEZ, Juan Pablo Maldonado; VAŠATA, Daniel. *BI-VZD přednáška 5*. FIT ČVUT, 2022-06.
3. BODE, Nikolai WF; CODLING, Edward A. Statistical models for pedestrian behaviour in front of bottlenecks. In: *Traffic and Granular Flow'15*. Springer, 2016.
4. THOM, Herbert CS. A note on the gamma distribution. *Monthly weather review*. 1958, roč. 86, č. 4, s. 117–122.
5. VIRTANEN, Pauli; GOMMERS, Ralf; OLIPHANT, Travis E.; HABERLAND, Matt; REDDY, Tyler; COURNAPEAU, David; BUROVSKI, Evgeni; PETERSON, Pearu; WECKESSER, Warren; BRIGHT, Jonathan; VAN DER WALT, Stéfan J.; BRETT, Matthew; WILSON, Joshua; MILLMAN, K. Jarrod; MAYOROV, Nikolay; NELSON, Andrew R. J.; JONES, Eric; KERN, Robert; LARSON, Eric; CAREY, C J; POLAT, İlhan; FENG, Yu; MOORE, Eric W.; VANDERPLAS, Jake; LAXALDE, Denis; PERKTOLD, Josef; CIMRMAN, Robert; HENRIKSEN, Ian; QUINTERO, E. A.; HARRIS, Charles R.; ARCHIBALD, Anne M.; RIBEIRO, Antônio H.; PEDREGOSA, Fabian; VAN MULBREGT, Paul; SCIPY 1.0 CONTRIBUTORS. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*. 2020, roč. 17, s. 261–272. Dostupné z DOI: 10.1038/s41592-019-0686-2.
6. ARTIN, Emil. *The gamma function*. Courier Dover Publications, 2015.
7. SINHA, Priyanka. Multivariate polynomial regression in data mining: methodology, problems and solutions. *Int. J. Sci. Eng. Res.* 2013, roč. 4, č. 12, s. 962–965.
8. SNYMAN, Jan A; WILKE, Daniel N et al. *Practical mathematical optimization*. Springer, 2005.
9. BYRD, Richard H; LU, Peihuang; NOCEDAL, Jorge; ZHU, Ciyou. A limited memory algorithm for bound constrained optimization. *SIAM Journal on scientific computing*. 1995, roč. 16, č. 5, s. 1190–1208.
10. AKAIKE, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*. 1974, roč. 19, č. 6, s. 716–723. Dostupné z DOI: 10.1109/TAC.1974.1100705.
11. SCHWARZ, Gideon. Estimating the Dimension of a Model. *The Annals of Statistics*. 1978, roč. 6, č. 2, s. 461–464. Dostupné z DOI: 10.1214/aos/1176344136.

12. TIAN, Wei; SONG, Weiguo; MA, Jian; FANG, Zhiming; SEYFRIED, Armin; LIDDLE, Jack. Experimental study of pedestrian behaviors in a corridor based on digital image processing. *Fire Safety Journal*. 2012, roč. 47, s. 8–15. ISSN 0379-7112. Dostupné z DOI: <https://doi.org/10.1016/j.firesaf.2011.09.005>.
13. SHARIFI, Sadra; STUART, Dan; CHRISTENSEN, Keith; CHEN, Anthony. Time Headway Modeling and Capacity Analysis of Pedestrian Facilities Involving Individuals with Disabilities. *Transportation Research Board*. 2016, roč. 2553. Dostupné z DOI: 10.3141/2553-05.
14. COWAN, Richard J. Useful headway models. *Transportation Research*. 1975, roč. 9, č. 6, s. 371–375. ISSN 0041-1647. Dostupné z DOI: [https://doi.org/10.1016/0041-1647\(75\)90008-8](https://doi.org/10.1016/0041-1647(75)90008-8).
15. BUKÁČEK, M.; HRABÁK, P.; KRBÁLEK, M. Experimental Study of Phase Transition in Pedestrian Flow. *Transportation Research Procedia*. 2014, roč. 2. ISSN 2352-1465. Dostupné z DOI: <https://doi.org/10.1016/j.trpro.2014.09.014>. The Conference on Pedestrian and Evacuation Dynamics 2014 (PED 2014), 22-24 October 2014, Delft, The Netherlands.
16. POSPÍŠIL, Vít. *Analýza videozáznamů evakuačních experimentů – zúžení*. 2022. Dostupné také z: <http://hdl.handle.net/10467/101918>. Bakalářská práce.
17. PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V.; VANDERPLAS, J.; PASSOS, A.; COURNAPEAU, D.; BRUCHER, M.; PERROT, M.; DUCHESNAY, E. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*. 2011, roč. 12, s. 2825–2830.

Obsah přiloženého média

src	
├─ readme.txt.....	stručný popis zdrojových kódů
├─ impl.....	zdrojové kódy řešení
├─ thesis.....	zdrojová forma práce ve formátu L ^A T _E X
└─ text.....	text práce
├─ thesis.pdf.....	text práce ve formátu PDF