

Západočeská univerzita v Plzni

Fakulta aplikovaných věd

Katedra matematiky

Bakalářská práce

Odhad počtu návštěvníků nákupního centra

ZÁPADOČESKÁ UNIVERZITA V PLZNI

Fakulta aplikovaných věd

Akademický rok: 2021/2022

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

(projektu, uměleckého díla, uměleckého výkonu)

Jméno a příjmení: **Barbora FÜRSTOVÁ**
Osobní číslo: **A20B0011P**
Studijní program: **B0541A170006 Matematika a finanční studia**
Téma práce: **Odhad počtu návštěvníků nákupního centra**
Zadávající katedra: **Katedra matematiky**

Zásady pro vypracování

Studentka se bude věnovat úloze spojené s čítacím procesem v situaci, kdy některé příchody nejsou zachyceny. Popíše příslušný pravděpodobnostní model, navrhne odhady zkoumaných parametrů. Vyšetří (aspoň simulačně) jejich vlastnosti a porovná je s odhady z modelu s úplnou informací.

Rozsah bakalářské práce: **20-50 stan**
Rozsah grafických prací: **dle potřeby**
Forma zpracování bakalářské práce: **tištěná**

Seznam doporučené literatury:

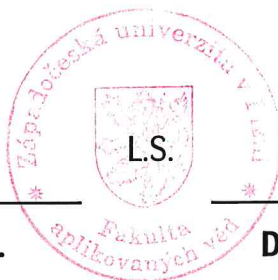
- J. Anděl: Matematická statistika, SNTL, Praha, 1985.
- P. Mandl: Pravděpodobnostní dynamické modely, Academia, Praha, 1985.
- Z. Prášková, P. Lachout: Základy náhodných procesů, Nakladatelství Karolinum, Praha, 2001.

Vedoucí bakalářské práce: **Mgr. Michal Friesl, Ph.D.**
Katedra matematiky

Datum zadání bakalářské práce: **1. října 2021**
Termín odevzdání bakalářské práce: **25. května 2022**



Doc. Ing. Miloš Železný, Ph.D.
děkan



Doc. Ing. Marek Brandner, Ph.D.
vedoucí katedry

Prohlášení

Prohlašuji, že jsem bakalářskou práci vypracovala samostatně a výhradně s použitím citovaných pramenů.

V Plzni dne 22. května 2022

Barbora Fürstová

Poděkování

Na tomto místě bych velmi ráda poděkovala vedoucímu mé práce, panu Mgr. Michalu Frieslovi, PhD., za jeho odborné konzultace a velmi vstřícný přístup.

Abstract

This bachelor thesis explores a probability model for two different counting methods of incoming customers into a shopping centre. The investigated counting devices are a sensor and a camera. The estimators of the model are proposed and evaluated using simulations.

Abstrakt

Tato bakalářská práce se zabývá pravděpodobnostním modelem pro dva různé způsoby sčítání (resp. detekci) příchodu zákazníků do obchodního centra. Uvažované sčítací přístroje jsou čidlo a kamera. Odhady modelu jsou navrženy a vyhodnoceny pomocí simulací.

Obsah

1	Úvod	8
2	Základní pojmy	9
2.1	Alternativní rozdělení	9
2.2	Binomické rozdělení	9
2.3	Poissonovo rozdělení	10
2.4	Podmíněná pravděpodobnost, podmíněná střední hodnota .	11
2.5	Kovariance	12
3	Zavedení proměnných	13
4	Matematický model	14
4.1	Předpoklady modelu	14
4.2	Popis modelu	14
4.2.1	Čidlo	14
4.2.2	Kamera	15
4.2.3	Závislost čidla a kamery	16
4.3	Vyjádření parametrů pomocí momentů	17
5	Odhady	19
5.1	Odvození odhadů	19
5.1.1	Odhad 0	20
5.1.2	Odhad 1	20
5.1.3	Odhad 2	20
5.2	Omezení odhadů	21
5.2.1	Odhad 1	21
5.2.2	Odhad 2	23
6	Simulace	25
6.1	Analýza simulací	27
6.1.1	Závislost průměrné úspěšnosti odhadů pro čidlo na pa- rametru p	28
6.1.2	Závislost průměrné úspěšnosti odhadů pro kameru na parametru q	29
6.1.3	Závislost průměrné úspěšnosti odhadů pro čidlo na pa- rametru λ	30

6.1.4	Závislost průměrné úspěšnosti odhadů pro kameru na parametru λ	30
6.1.5	Závislost průměrné úspěšnosti odhadů pro čidlo na parametru n	31
6.1.6	Závislost průměrné úspěšnosti odhadů pro kameru na parametru n	32
7	Závěr	34
	Literatura	46

1 Úvod

Pro účely této práce uvažujeme existenci obchodního centra, v němž je prováděno sčítání příchozích zákazníků pomocí čidla a kamery. Pro jednotlivá automatická zařízení je pracováno s odhady, které jsou vyhodnocovány pomocí simulací. Simulační výpočty jsou prováděny pomocí programu Matlab. Při následné analýze předpokládáme, že známe skutečný počet příchozích zákazníků (viz Odhad 0), tudíž je možno porovnat úspěšnost záchytu pomocí čidla i kamery ve zkoumaných odhadech.

V rámci této práce si nejprve zdefinujeme základní pojmy z pravděpodobnostní teorie, které jsou v této práci dále využívány; následně zavedeme vlastní značení jednotlivých proměnných. Poté je představen samotný model a jeho dva odhady - jak pro čidlo, tak (alternativně) i pro kameru. Při konstrukci odhadů poté budeme nahrazovat teoretické momenty z modelu momenty výběrovými a provedeme simulace, jejichž grafické výstupy zanalyzujeme.

2 Základní pojmy

Na úvod si zavedeme některé základní pojmy z teorie pravděpodobnosti, se kterými se v rámci této práce setkáme.

2.1 Alternativní rozdělení

Veličina X s alternativním rozdělením pravděpodobnosti je dána parametrem p ($0 < p < 1$), a může nabývat pouze hodnot 0, nebo 1. Hodnoty $X = 1$ nabývá s pravděpodobností p , hodnoty $X = 0$ s pravděpodobností $1 - p$.

Alternativní rozdělení můžeme zapisovat $P(X = x) = p^x \cdot (1 - p)^{1-x}$ pro $x = 0$ a $x = 1$.

Střední hodnotu alternativního rozdělení náhodné veličiny X určujeme jako

$$\mathbf{E}(X) = p. \quad (2.1)$$

Zapíšeme $X \sim A(p)$. [2, str. 19]

Toto rozdělení budeme využívat pro zachycení přicházejících zákazníků. Jak kamera, tak čidlo budto zákazníka detekují (hodnota 1 - odpovídá jednomu zaznamenanému zákazníkovi), nebo nedetekují (hodnota 0).

2.2 Binomické rozdělení

Náhodná veličina X se řídí binomickým rozdělením s parametry $n \in \mathbb{N}$ a $p \in (0, 1)$, pokud X nabývá pouze hodnot $0, 1, \dots, n$ (tj. může nabývat více hodnot než jen $x = 0$ a $x = 1$ (viz podkapitola 2.1; dále budeme značit $x = k$), a platí

$$P(X = k) = \binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k}, \quad (2.2)$$

pro $k = 0, 1, \dots, n$. Pak píšeme $X \sim Bi(n, p)$.

Střední hodnota binomického rozdělení se vypočítá dle vzorce

$$\mathbf{E}(X) = n \cdot p \quad (2.3)$$

a rozptyl

$$\sigma^2(X) = n \cdot p \cdot (1 - p). \quad (2.4)$$

Binomické rozdělení pravděpodobnosti je aplikováno v případech, kdy uvažujeme určitý náhodný jev příslušný nějakému pokusu. Pro náš model je

pravděpodobnost uvažovaného jevu zachycení příchozího zákazníka kamerou rovna p , resp. q pro zachycení čidlem (viz kapitola 4). Počet nezávislých opakování pokusu je n .¹ Pak se počet nastalých jevů z celkem n opakování prováděného pokusu řídí binomickým rozdělením pravděpodobnosti $Bi(n, p)$. [1, str. 32-33]

2.3 Poissonovo rozdělení

Náhodná veličina diskrétního typu X má Poissonovo rozdělení s parametrem $\lambda > 0$, jestliže X je z množiny nezáporných celých čísel, a platí

$$P(X = k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}$$

pro $k = 0, 1, 2, \dots$. Zapisujeme $X \sim Po(\lambda)$. Střední hodnota a rozptyl jsou určeny následovně:

$$\mathbf{E}(X) = \lambda, \quad (2.5)$$

$$\sigma^2(X) = \lambda, \quad (2.6)$$

Distribuční funkce Poissonova rozdělení je určena pomocí vzorce

$$F(x) = \sum_{k=0}^{[x]} e^{-\lambda} \cdot \frac{\lambda^k}{k!},$$

kde $[x]$ je celá část čísla x . [1, str. 37-38]

Hodnoty distribuční funkce Poissonova rozdělení pro nižší λ bývají tabulovány. Pro velké hodnoty λ (např. $\lambda \geq 9$) můžeme aproximovat Poissonovo rozdělení pomocí rozdělení normálního. Poissonovo rozdělení lze použít k aproximaci rozdělení binomického. Pro velká n a malá p ($n \geq 30$, $p \leq 0,1$) můžeme provést aproximaci $Bi(n, p) \approx Po(\lambda)$, kde $\lambda = n \cdot p$. [2, str. 18], [1, str. 38]

Náhodná veličina s Poissonovým rozdělením pravděpodobnosti se používá jako model např. pro počet poruch přístroje, počet pojistných událostí, počet hovorů či počet zákazníků v obchodě za určitou časovou jednotku (např. 1 hodinu). [1, str. 37-38],[4]

Pro součet nezávislých náhodných veličin řídicích se Poissonovým rozdělením pravděpodobnosti platí, že daný součet má rovněž Poissonovo rozdělení s parametrem λ , jenž se rovná součtu parametrů původních rozdělení. [1, str. 38]

¹Nezávislost zde chápeme tak, že výsledek každého jednotlivého pokusu nijak nezávisí na výsledcích z předchozích pokusů.

2.4 Podmíněná pravděpodobnost, podmíněná střední hodnota

Nechť A, B jsou jevy.

Pro podmíněnou pravděpodobnost jevu A za podmínky, že nastal jev B platí

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (2.7)$$

za předpokladu, že $P(B) > 0$.

Nechť X, Y jsou diskrétní náhodné veličiny.

Pokud je splněno, že $P(Y = y) > 0$, je podmíněná pravděpodobnostní funkce pro konkrétní zadané $Y = y$ (tj. náhodný jev) stanovena

$$p_{X|Y}(x|y) = \frac{P(X = x \wedge Y = y)}{P(Y = y)}. \quad (2.8)$$

Ze znalosti marginální (okrajové) pravděpodobnostní funkce $p_Y(y) = \sum_x p(x, y)$ dále dostaneme rovnost

$$p_{X|Y}(x|y) = \frac{p_{XY}(x, y)}{p_Y(y)}, \quad (2.9)$$

pokud $p_Y > 0$ a $x, y = 0, 1, \dots$

Ze zákona celkové pravděpodobnosti (Law of total probability) poté získáme vztah

$$P(X = x) = \sum_{y=0}^x p_{X|Y}(x|y) \cdot p_Y(y). \quad (2.10)$$

[3, str. 57-58]

Pro podmíněnou střední hodnotu, tj. kdy jedna z proměnných nabyla určité hodnoty (např. zde $Y = y$), vycházíme ze vztahu:

$$\mathbf{E}(X|Y = y) = \sum_x x \cdot p_{X|Y}(x|y), \quad (2.11)$$

jestliže $p_Y(y) > 0$. Ze zákona celkové pravděpodobnosti pro podmíněnou střední hodnotu dostaneme rovnost

$$\mathbf{E}(X) = \sum_x \mathbf{E}(X|Y = y) \cdot p_Y(y). \quad (2.12)$$

Tento vztah můžeme upravit do podoby

$$\mathbf{E}(X) = \mathbf{E}(\mathbf{E}(X|Y)). \quad (2.13)$$

[5]

2.5 Kovariance

Nechť X, Y jsou náhodné proměnné. Kovariance mezi těmito dvěma veličinami je dána vzorcem

$$\text{cov}(X, Y) = \mathbf{E}[(X - \mathbf{E}(X)) \cdot (Y - \mathbf{E}(Y))]. \quad (2.14)$$

Nechť X, Y, Z jsou náhodné proměnné. Zákon totální kovariance je potom určen následující rovností

$$\text{cov}(X, Y) = \mathbf{E}(\text{cov}(X, Y|Z)) + \text{cov}(\mathbf{E}(X|Z), \mathbf{E}(Y|Z)). \quad (2.15)$$

3 Zavedení proměnných

V rámci této práce se budeme setkávat s následujícími proměnnými:

- N ... náhodná veličina - počet příchozích zákazníků
- $N^{(p)}$... náhodná veličina - počet zákazníků zachycených kamerou
- $N^{(q)}$... náhodná veličina - počet zákazníků zachycených čidlem
- n_i ... konkrétní (i -té) pozorování náhodné veličiny N
- $n_i^{(p)}$... konkrétní (i -té) pozorování náhodné veličiny $N^{(p)}$
- $n_i^{(q)}$... konkrétní (i -té) pozorování náhodné veličiny $N^{(q)}$
- p ... značení pravděpodobnosti pro detekci zákazníka kamerou
- q ... značení pravděpodobnosti pro detekci zákazníka čidlem
- λ ... intenzita (průměrný počet zákazníků, kteří přijdou za jednu časovou jednotku)
- n ... počet měřených časových úseků (uvažujeme intervaly stejné délky, přičemž délka každého časového úseku je rovna uvažované časové jednotce)

4 Matematický model

Nyní si představíme model detekce zákazníků přicházejících do obchodního centra.

4.1 Předpoklady modelu

Dále představený model se řídí následujícími předpoklady:

- Skutečný počet zákazníků v daném časovém období se řídí Poissonovým rozdělením (viz podkapitola 2.3).
- Každý zákazník je zachycen a započten kamerou s pravděpodobností p .
- Každý zákazník je zachycen a započten automatem, tj. čidlem s pravděpodobností q .
- Započtení zákazníka kamerou a čidlem jsou dva nezávislé jevy.

4.2 Popis modelu

Do fiktivního obchodního centra přicházejí návštěvníci, jejichž vstup je zaznamenáván a sčítán pomocí čidla a kamery. Abychom byli schopni určovat pravděpodobnost detekce danými přístroji, předpokládáme, že známe skutečný počet příchodu zákazníků. Tuto veličinu si označíme N .

Jelikož tato náhodná veličina zaznamenává počet výskytů daného jevu (počet zákazníků) v na sobě nezávislých časových intervalech, budeme ji popisovat pomocí Poissonova rozdělení, tj. $N \sim Po(\lambda)$ (viz podkapitola 2.3).

Parametr $\lambda > 0$ odpovídá průměrnému množství zákazníků, kteří do obchodního centra přijdou za testovanou časovou jednotku.

4.2.1 Čidlo

Nechť tedy máme v daném obchodním centru instalované čidlo. Počet zákazníků zachycených čidlem bude popisován náhodnou veličinou $N^{(q)}$; parametr q si zvolíme jako značení pravděpodobnosti detekce zákazníka pomocí čidla.

Podmíněné rozdělení náhodné veličiny $N^{(q)}$ při daném $N = n$ je binomické: $(N^{(q)}|N = n) \sim Bi(n, q)$ (viz podkapitola 2.2). Parametr q leží

v intervalu $(0; 1)$. Detekce jednoho konkrétního zákazníka v daném časovém úseku se řídí alternativním rozdělením pravděpodobnosti - zákazník je čidlem buďto zachycen ($1 =$ úspěšné zaznamenání příchozího zákazníka), nebo ne ($0 =$ neúspěšné zachycení zákazníka). Posčítáním úspěšných zachytů čidlem z $N = n$ příchozích návštěvníků v rámci jednoho časového úseku získáme hodnotu $N^{(q)}$.

Nyní stanovíme střední hodnotu $\mathbf{E}(N^{(q)})$ pomocí vzorce podmíněné střední hodnoty (viz vzorec (2.13) v podkapitole 2.4):

$$\mathbf{E}(N^{(q)}) = \mathbf{E}(\mathbf{E}(N^{(q)}|N)).$$

Jak již bylo zmíněno, podmíněné rozdělení počtu jednotlivých zákazníků čidlem se řídí binomickým rozdělením:

$$(N^{(q)}|N = n) \sim \sum_{i=1}^n I_i^{(q)} \sim Bi(n, q), \quad (4.1)$$

kde $N = n$ je bráno jako podmínka. A $I_i^{(q)}, i = 1, 2, \dots, n$ nabývají hodnot:

$$I_i^{(q)} = \begin{cases} 1 & \text{s pravděpodobností } q, \text{ tj. } i\text{-tý zákazník je zachycen čidlem,} \\ 0 & \text{s pravděpodobností } 1 - q, \text{ tj. } i\text{-tý zákazník čidlem zachycen nebyl.} \end{cases}$$

Dále víme, že pro binomické rozdělení (viz podkapitola 2.2, resp. vztah (4.1)) je střední hodnota rovna $\mathbf{E}(N^{(q)}|N = n) = n \cdot q$, takže za použití vzorců z podkapitoly 2.4 platí rovnost

$$\mathbf{E}(N^{(q)}) = \mathbf{E}(\mathbf{E}(N^{(q)}|N)) = \mathbf{E}(N \cdot q) = q \cdot \mathbf{E}(N).$$

Poté, ze znalosti střední hodnoty Poissonova rozdělení (tj. $\mathbf{E}(N) = \lambda$), již vyvodíme finální tvar rovnice, tedy:

$$\mathbf{E}(N^{(q)}) = q \cdot \lambda. \quad (4.2)$$

4.2.2 Kamera

Obdobně jako u čidla, i u kamery předpokládáme, že toto zařízení instalované v obchodním centru zachycuje příchod návštěvníků, který budeme popisovat náhodnou veličinou $N^{(p)}$ (parametr p odpovídá pravděpodobnosti zachycení zákazníků pomocí kamery). I zde budeme jednotlivé (i -té) pozorování označovat $n_i^{(p)}$. Náhodná veličina určující počet zákazníků zachycených kamerou se rovněž řídí binomickým rozdělením pravděpodobnosti:

$N^{(p)} \sim Bi(n, p)$. Parametr p uvažujeme z intervalu $(0; 1)$ a parametr n bude i zde vycházet z měření skutečných hodnot pro (i -té pozorování) n_i .

Tedy - jako u čidla, i zde dospějeme k následujícímu vztahu:

$$\mathbf{E}(N^{(p)}) = \mathbf{E}(\mathbf{E}(N^{(p)}|N = n)) = \mathbf{E}(N \cdot p) = p \cdot \mathbf{E}(N) = p \cdot \lambda.$$

Tj. zísaná rovnost je

$$\mathbf{E}(N^{(p)}) = p \cdot \lambda. \quad (4.3)$$

4.2.3 Závislost čidla a kamery

Nyní popíšeme závislost mezi čidlem a kamerou pomocí kovariance.

Kovariance dvou náhodných veličin se počítá dle vzorce (2.14), pro naše veličiny má tedy podobu:

$$cov(N^{(p)}, N^{(q)}) = \mathbf{E}[(N^{(p)} - \mathbf{E}(N^{(p)})) \cdot (N^{(q)} - \mathbf{E}(N^{(q)}))].$$

Dále budeme počítat se vzorcem pro podmíněnou kovarianci. Ze zákona totální kovariance (viz vzorec (2.15)) obdržíme pro naše veličiny následující vztah

$$cov(N^{(p)}, N^{(q)}) = \mathbf{E}(cov(N^{(p)}, N^{(q)}|N)) + cov(\mathbf{E}(N^{(p)}|N), \mathbf{E}(N^{(q)}|N)). \quad (4.4)$$

Pro podmíněnou kovarianci $cov(N^{(p)}, N^{(q)}|N)$ jsou náhodné veličiny $N^{(p)}$ a $N^{(q)}$ díky podmíněnosti N nezávislé, tj. $cov(N^{(p)}, N^{(q)}|N) = 0$. Takže i $\mathbf{E}(cov(N^{(p)}, N^{(q)}|N)) = 0$.

Druhý člen pravé strany rovnice (4.4), v němž mj. vycházíme ze vzorce pro střední hodnotu binomického rozdělení náhodných veličin $N^{(p)}$ a $N^{(q)}$, je roven $cov(\mathbf{E}(N^{(p)}|N), \mathbf{E}(N^{(q)}|N)) = cov(N \cdot p, N \cdot q)$.

Tím získáme následující rovnost

$$cov(N^{(p)}, N^{(q)}) = 0 + cov(N \cdot p, N \cdot q) = p \cdot q \cdot cov(N, N).$$

Z definice rozptylu a jeho vztahu s kovariancí je patrné, že $cov(N, N) = \sigma^2(N)$. A tudíž celková kovariance je rovna

$$p \cdot q \cdot cov(N, N) = p \cdot q \cdot \sigma^2(N).$$

Díky tomu, že pracujeme s Poissonovým rozdělením, jehož rozptyl je roven λ , získáváme

$$cov(\mathbf{E}(N^{(p)}|N), \mathbf{E}(N^{(q)}|N)) = p \cdot q \cdot \lambda.$$

Tím jsme tedy dostali vztah:

$$\text{cov}(N^{(p)}, N^{(q)}) = p \cdot q \cdot \lambda. \quad (4.5)$$

4.3 Vyjádření parametrů pomocí momentů

Z podkapitol 4.2.1, 4.2.2 jsme získali dvě rovnice: $\mathbf{E}(N^{(p)}) = p \cdot \lambda$ a $\mathbf{E}(N^{(q)}) = p \cdot \lambda$.

Jelikož potřebujeme vyjádřit celkem tři proměnné (p, q, λ) , musíme získat soustavu tří rovnic o třech neznámých. Třetí vztah je odvozen ze vzájemné závislosti mezi čidlem a kamerou, tj. z kovariance mezi těmito dvěma zkoumanými náhodnými veličinami $N^{(p)}$ a $N^{(q)}$ (viz podkapitola 4.2.3).

Z předchozích výpočtů jsme tedy dostali celkem tři rovnice o třech neznámých.

$$\mathbf{E}(N^{(p)}) = p \cdot \lambda, \quad (4.6)$$

$$\mathbf{E}(N^{(q)}) = q \cdot \lambda, \quad (4.7)$$

$$\text{cov}(N^{(p)}, N^{(q)}) = p \cdot q \cdot \lambda. \quad (4.8)$$

Vhodnou úpravou tedy získáme i předpisy pro vyjádření těchto samotných parametrů. Pokud rovnici (4.6) podělíme parametrem λ různým od nuly (tato podmínka je splněna, neboť $\lambda > 0$), dostaneme vyjádření pro parametr p v následujícím tvaru

$$p = \frac{\mathbf{E}(N^{(p)})}{\lambda}. \quad (4.9)$$

Obdobně z rovnice (4.7) získáme tvar pro výpočet parametru q :

$$q = \frac{\mathbf{E}(N^{(q)})}{\lambda}. \quad (4.10)$$

A z poslední rovnice (4.8) následně pomocí dosazení získaných vztahů (4.9) a (4.10) za p a q vyjádříme zbývající parametr λ .

$$\text{cov}(N^{(p)}, N^{(q)}) = \frac{\mathbf{E}(N^{(p)})}{\lambda} \cdot \frac{\mathbf{E}(N^{(q)})}{\lambda} \cdot \lambda$$

$$\text{cov}(N^{(p)}, N^{(q)}) = \frac{\mathbf{E}(N^{(p)}) \cdot \mathbf{E}(N^{(q)})}{\lambda}$$

$$\lambda = \frac{\mathbf{E}(N^{(p)}) \cdot \mathbf{E}(N^{(q)})}{cov(N^{(p)}, N^{(q)})}. \quad (4.11)$$

V následující kapitole 5, kde jsou modelu přiřazeny dva odhady, budou teoretické momenty nahrazeny momenty výběrovými.

5 Odhady

V této kapitole budou pro náš model představeny dva odhady.

Pro dané odhady budeme provádět opakovaná měření, a to následujícím způsobem: Uvažujeme počet testovaných časových jednotek (značíme n). Z jednotlivých n^1 pozorování poté zjistíme pro každé jednotlivé měření hodnoty záchyťů pro kameru i čidlo, tj. $(N_1^{(p)}, N_1^{(q)}), (N_2^{(p)}, N_2^{(q)}), \dots, (N_n^{(p)}, N_n^{(q)})$.

Zároveň také platí, že pracujeme s výsledky získanými opakovaným měřením (za podmínek intervalů stejné časové délky a při stejné intenzitě provozu).

Jak již bylo zmíněno, dále také předpokládáme, že celkové počty příchoďů $N^{(p)}$ a $N^{(q)}$ jsou pro jednotlivá (i -tá) pozorování vždy sledovaná ve stejném časovém intervalu. Tato pozorování budeme označovat $n_i^{(p)}$, resp. $n_i^{(q)}$. Provedeme n opakování² těchto pozorování (přičemž se tyto intervaly nijak navzájem nepřekrývají) a následně odhadujeme průměrnou úspěšnost detekce zákazníků automatickým čidlem, resp. kamerou.

Pro odhady budeme využívat výběrovou kovarianci mezi náhodnými veličinami X a Y , která je definována následovně

$$s_{XY} = \frac{1}{n-1} \cdot \sum_{i=1}^n (X_i - \bar{X}) \cdot (Y_i - \bar{Y}), \quad (5.1)$$

přičemž je tento výraz definován pro $n \geq 2$, $n \in N$. Pro naše odhady jen přeznačíme náhodné veličiny, a sice: $X = N^{(p)}$, $Y = N^{(q)}$, tj. získáme rovnost

$$s_{N^{(p)}N^{(q)}} = \frac{1}{n-1} \cdot \sum_{i=1}^n (N_i^{(p)} - \overline{N^{(p)}}) \cdot (N_i^{(q)} - \overline{N^{(q)}}), \quad (5.2)$$

5.1 Odvození odhadů

Nyní budou představeny odhady pravděpodobnosti zachycení zákazníků jednotlivými přístroji. U obou těchto sčítacích mechanismů (čidla a kamery) předpokládáme, že jimi prováděné detekce zákazníků jsou dva navzájem nezávislé jevy (viz 4.1).

¹Proměnná n odpovídá v našem modelu značení parametrem n_i , tj. skutečnému naměřenému počtu zákazníků v daném (i -tém) měření náhodné veličiny N (viz kapitola 3).

²V případě, kdybychom měli k dispozici informace jen z jednoho pozorování, tento postup bychom volit nemohli, neboť výběrová směrodatná odchylka je definována pro alespoň 2 pozorování (viz vzorec (5.1)).

5.1.1 Odhad 0

Pro simulační účely si zavedeme tzv. Odhad 0. Tento odhad odpovídá pravdivým údajům o všech příchozích zákaznících (tj. skutečný počet zákazníků, kteří přišli v daném časovém úseku). Zároveň zaznamenává skutečné záchyty čidlem i kamerou, proto jej využíváme jako zdroj zpětné kontroly údajů v simulacích.

5.1.2 Odhad 1

Pro tento odhad vycházíme ze znalosti celkových počtů zákazníků zachycených jednak kamerou, jednak automatickým čidlem, a dále ze střední hodnoty detekovaných zákazníků pro druhé testované zařízení. Výpočet pravděpodobnosti automatického čidla je potom vyjádřen podílem kovariance (mezi čidlem ($N^{(q)}$) a kamerou ($N^{(p)}$)) a střední hodnotou počtu zákazníků zachycených kamerou,

$$q = \frac{\text{cov}(N^{(p)}, N^{(q)})}{\mathbf{E}(N^{(p)})}. \quad (5.3)$$

Alternativně, pro relativní četnost kamery, stanovíme podíl kovariance (mezi kamerou ($N^{(p)}$) a čidlem ($N^{(q)}$)) a střední hodnoty počtu zákazníků zachycených automatickým čidlem,

$$p = \frac{\text{cov}(N^{(p)}, N^{(q)})}{\mathbf{E}(N^{(q)})}. \quad (5.4)$$

Nyní převedeme vzorec (5.3) do tvaru odhadu

$$\hat{q}_1 = \frac{s_{N^{(p)}N^{(q)}}}{\overline{N^{(p)}}}, \quad (5.5)$$

kde $s_{N^{(p)}N^{(q)}}$ značí výběrovou kovarianci a $\overline{N^{(p)}}$ průměr hodnot z celkových hodnot záchytů zákazníků kamerou (tj. z počtů $N_1^{(p)}, N_2^{(p)}, \dots, N_n^{(p)}$).

Obdobně získáme tento odhad pro kameru ve tvaru

$$\hat{p}_1 = \frac{s_{N^{(p)}N^{(q)}}}{\overline{N^{(q)}}}, \quad (5.6)$$

kde $s_{N^{(p)}N^{(q)}}$ značí výběrovou kovarianci a $\overline{N^{(q)}}$ průměr z celkových hodnot záchytů zákazníků automatickým čidlem (tj. z počtů $N_1^{(q)}, N_2^{(q)}, \dots, N_n^{(q)}$).

5.1.3 Odhad 2

Pro další odhad budeme rovněž vycházet z výpočtu výběrové kovariance spočtené z hodnot $(N_1^{(p)}, N_1^{(q)}), (N_2^{(p)}, N_2^{(q)}), \dots, (N_n^{(p)}, N_n^{(q)})$, kterou podělíme

výběrovým rozptylem kamery (jenž získáme z hodnot $N_1^{(p)}, N_2^{(p)}, \dots, N_n^{(p)}$), resp. automatického čidla (zde bychom pracovali s hodnotami $N_1^{(q)}, N_2^{(q)}, \dots, N_n^{(q)}$).

$$\frac{\sigma_{N^{(p)}N^{(q)}}}{\sigma_{N^{(p)}}^2}, \quad (5.7)$$

Aby byla zajištěna existence výběrové kovariance, je i zde nutné provádět opakovaná měření.

Odhad 2 úspěšnosti záchytu čidlem má tedy tvar

$$\hat{q}_2 = \frac{s_{N^{(p)}N^{(q)}}}{s_{N^{(p)}}^2}, \quad (5.8)$$

kde $s_{N^{(p)}N^{(q)}}$ značí výběrovou kovarianci mezi veličinami "zachycení zákazníka kamerou" ($N^{(p)}$) a "zachycení zákazníka čidlem" ($N^{(q)}$) a $s_{N^{(p)}}^2$ výběrový rozptyl z celkových hodnot záchytů zákazníků kamerou (tj. z počtů $N_1^{(p)}, N_2^{(p)}, \dots, N_n^{(p)}$).

Druhý odhad úspěšnosti záchytu kamerou je obdobně

$$\hat{p}_2 = \frac{s_{N^{(p)}N^{(q)}}}{s_{N^{(q)}}^2}, \quad (5.9)$$

kde $s_{N^{(p)}N^{(q)}}$ značí výběrovou kovarianci a $s_{N^{(q)}}^2$ výběrový rozptyl z celkových hodnot záchytů zákazníků čidlem (tj. z počtů $N_1^{(q)}, N_2^{(q)}, \dots, N_n^{(q)}$).

5.2 Omezení odhadů

Nyní rozebereme omezení modelů, tj. definiční obory odhadů pro kameru i čidlo.

Pro oba dva odhady platí, že výsledná hodnota musí ležet v intervalu $\hat{q} = \langle 0; 1 \rangle$, $\hat{p} = \langle 0; 1 \rangle$. Proto stanovíme, že vyjde-li hodnota $\hat{q} < 0$ upravíme ji na $\hat{q} = 0$, je-li hodnota $\hat{q} > 1$, bude nahrazena $\hat{q} = 1$.

Alternativně upravíme hodnoty pro \hat{p} .

5.2.1 Odhad 1

Čidlo

První odhad je stanoven jako podíl výběrové kovariance mezi veličinami $N^{(p)}$ (kamera) a $N^{(q)}$ (čidlo) a hodnotou průměru z měření kamerou,

$$\hat{q}_1 = \frac{s_{N^{(p)}N^{(q)}}}{N^{(p)}}.$$

Tento odhad není definován v případě, kdy se průměr $\overline{N^{(p)}} = \sum_{i=1}^n \frac{N_i^{(p)}}{n}$ rovná nule. Tato situace nastává v případě, kdy jmenovatel zlomku, tj. suma jednotlivých pozorování, je nula. Průměr roven nule dostaneme tehdy, je-li součet jednotlivých pozorování nulový. K tomuto dochází tehdy, jsou-li jednotlivá pozorování $N_i^{(p)}$ rovna nule,

$$\forall i = 1, 2, \dots, n : N_i^{(p)} = 0.$$

K výše zmíněnému obecně platí, že průměr rovný hodnotě nula, lze získat i součtem reálných čísel, která se v sumě rovnají nule. Z povahy úlohy vycházíme z toho, že počítáme s nezápornými celými čísly (počet zákazníků).

Tento odhad je omezen i v případě, pokud bychom obdrželi neurčitý výraz $\frac{0}{0}$. Nulový výraz ve jmenovateli je rozebrán výše. Zbývá tedy zjistit, v jakém případě se výběrová kovariance rovná nulové hodnotě. Výběrová kovariance je definována pro $n \geq 2, n \in N$ (viz (5.2)). Dále platí, že je-li jmenovatel odhadu $\overline{N^{(p)}} = 0$, hodnota $\sigma_{N^{(p)}N^{(q)}}$ bude rovněž nulová, neboť

$$\sum_{i=1}^n (0 - 0) \cdot (N_i^{(q)} - \overline{N^{(q)}}) = 0,$$

a tedy i výběrová kovariance

$$s_{N^{(p)}N^{(q)}} = \frac{1}{n-1} \cdot 0 = 0 \quad \forall n \geq 2.$$

Aby byly splněny podmínky definičního oboru pro výběrovou kovarianci mezi $N^{(p)}$ a $N^{(q)}$, uvažujme $n = 50$ měření (viz úvod kapitoly 6).

Pro výběrovou kovarianci musíme ošetřit následující případ: Nastane-li situace, kdy je záchyt zákazníků pomocí kamery roven nule (což by způsobilo nulovost výběrové kovariance v čitateli), v rámci simulací provedeme validaci na stejném testovaném vzorku pomocí "zpětné kontroly" hodnot záchyty čidlem (tj. zda-li byl záchyt čidlem rovněž nulový).

Pokud v daném testovaném vzorku existuje $i \in 1, 2, \dots, n$ takové, že $N_i^{(q)} \neq 0$, tedy $\overline{N^{(q)}} \neq 0$, tj. čidlo nějaké zákazníky detekovalo, budeme předpokládat, že kamera nefunguje ($p = 0$).

Jsou-li $N_i^{(q)} = 0 \quad \forall i = 1, 2, \dots, n$ (tj. i průměr $\overline{N^{(q)}} = 0$), tedy pokud čidlo nezachytilo žádného příchozího zákazníka, stanovíme, že úspěšnosti čidla i kamery budou rovny nule (tj. $q = 0, p = 0$).

Kamera

Odhad 1 pro kameru je stanoven jako podíl výběrové kovariance a průměru hodnot získaných z měření čidlem, tj.

$$\hat{p}_1 = \frac{S_{N^{(p)}N^{(q)}}}{N^{(q)}}.$$

Další analýza probíhá analogicky jako v Odhadu 1 pro čidlo.

5.2.2 Odhad 2

Čidlo

Druhý odhad pro čidlo je určen vzorcem

$$\hat{q}_2 = \frac{S_{N^{(p)}N^{(q)}}}{s_{N^{(p)}}^2},$$

kde $s_{N^{(p)}}^2$ značí výběrový rozptyl kamery. Opět nejprve stanovíme, za jakých podmínek má tento zlomek smysl, tzn. provedeme rozbor definiční oboru. Výběrový rozptyl

$$s^2(N^{(p)}) = \frac{1}{n-1} \cdot \sum_{i=1}^n (N_i^{(p)} - \overline{N^{(p)}})^2 \quad (5.10)$$

je definován pro $n \geq 2, n \in N$. Potřebujeme, aby celý výběrový rozptyl byl roven kladné hodnotě, čímž předejdeme nulovému výrazu ve jmenovateli.

Nulový výsledek pro výraz $\sum_{i=1}^n (N_i^{(p)} - \overline{N^{(p)}})^2$, bychom dostali pouze v případě, pokud by byly všechny hodnoty $N_i^{(p)}$ stejné, tj. $N_i^{(p)} = c$, kde $c \geq 0$ je konstantní hodnota. To nám způsobí, že i průměr $\overline{N^{(p)}} = c$.

$$\sum_{i=1}^n (N_i^{(p)} - \overline{N^{(p)}})^2 = \sum_{i=1}^n (c - c)^2 = \sum_{i=1}^n 0 = 0 \quad \forall i = 1, 2, \dots, n.$$

I zde tedy musíme ošetřit případný výskyt neurčitého výrazu $\frac{0}{0}$ při definici \hat{q}_2 . Víme tedy, že není definována situace, kdy je nulová výběrová kovariance - tentokrát v podílu s výběrovou směrodatnou odchylkou rovnou nule. V rámci simulací provedeme "zpětnou kontrolu" hodnot: Pokud je výběrový rozptyl kamery roven nule, zkontrolujeme, zda k tomuto výsledku došlo kvůli tomu, že do obchodního centra v daném testovaném intervalu skutečně nepřišli žádní zákazníci, či zda byl nulový výsledek zapříčiněn součtem rozdílů jiných (tentokrát nenulových) hodnot.

Pro případ, kdy tedy byla nulová hodnota výběrového rozptylu kamery způsobena sčítáním rozdílů shodných nenulových hodnot, zjistíme, zda podobné výstupy zachytilo i čidlo. Je-li tomu tak, poté stanovíme $q = 1$. V opačném případě (tj. kdy příchod zákazníků není v daných pozorováních roven pouze jedné konkrétní hodnotě), nahradíme hodnotu q poměrem počtu přichodících zákazníků obou testovaných zařízení v daném pozorování $\frac{n_i^{(q)}}{n_i^{(p)}}$. U

tohoto výpočtu ovšem musíme stanovit, že v případě, pokud je hodnota tohoto poměru větší než 1, provedeme zaokrouhlení na 1 (maximální rozsah pravděpodobnosti: vycházíme z toho, že čidlo skutečně zachytilo všechny příchozí zákazníky).

Kamera

Druhý odhad pro kameru je stanoven jako podíl výběrové směrodatné odchylky a výběrového rozptylu podle veličiny $N^{(a)}$, tj. čidla:

$$\hat{p}_2 = \frac{s_{N^{(p)}, N^{(a)}}}{s_{N^{(a)}}^2}.$$

Omezení Odhadu 2 pro kameru odvodíme analogicky jako pro čidlo.

6 Simulace

Vlastnosti odhadů jsou následně analyzovány na základě simulací. Samotné simulace byly prováděny v programu Matlab.

Na úvod je tedy nutné vytvořit simulační proces, který generuje náhodný příchod zákazníků v určitém časovém okamžiku. Zde se vychází z Poissonova rozdělení při stanoveném parametru λ , přičemž tento parametr závisí na počtu pozorování za sledované časové období. Pro účely této práce byla horní hranice počtu období n stanovena na 50 pozorování, a to následujícím způsobem: Předpokládáme tzv. non-stop otevřené obchodní centrum (např. v nějaké světové metropoli), kde detekce zákazníků probíhá měřením padesátkrát denně: (Celková otevírací doba $24 \cdot 60 = 1440$ minut; měření probíhá každých $\frac{1440}{50} = 28,8$ minut.)

Následně se v závislosti na vygenerovaném vektoru příchodů provede simulace pomocí binomického rozdělení - jak pro záchyt kamerou, tak pro záchyt čidlem. A pro každého jednotlivého zákazníka je pomocí alternativního rozdělení rozhodnuto, zda došlo k jeho detekci pomocí čidla, resp. kamery:

$$x_i^{(q)} = \begin{cases} 1 & \text{je-li } i\text{-tý zákazník čidlem zachycen,} \\ 0 & \text{pokud } i\text{-tý zákazník čidlem zachycen není.} \end{cases}$$
$$x_i^{(p)} = \begin{cases} 1 & \text{je-li } i\text{-tý zákazník kamerou zachycen,} \\ 0 & \text{pokud } i\text{-tý zákazník kamerou zachycen není.} \end{cases}$$

Pro účely porovnávání si stanovíme, že počet příchozích zákazníků generovaný pomocí Poissonova rozdělení odpovídá skutečnému příchodu zákazníků, tj. bude se jednat o kontrolní údaj pro Odhad 0. Z nagenерованého skutečného příchodu zákazníků i z něj dle binomického rozdělení odvozených záchytů kamerou a čidlem vypočítáme průměr, výběrový rozptyl, výběrovou směrodatnou odchylku a výběrovou kovarianci. Tyto údaje nám nyní slouží k sestavení výpočtového tvaru pro jednotlivé odhady.

Na úvod začneme tzv. Odhadem 0 (viz podkapitola 5.1.1), který vychází z kompletní znalosti údajů o příchodu zákazníků i jejich jednotlivých záchytech kamerou a čidlem. Pravděpodobnost záchytu kamerou a čidlem při celkové znalosti dat je poté stanovena jako podíl počtu zachycených zákazníků kamerou (resp. čidlem) vůči skutečnému počtu zákazníků. Díky vícečetnému měření získáváme vektor hodnot, který dále použijeme pro výpočet hodnoty průměru, rozptylu a směrodatné odchylky pro kameru a čidlo v rámci Odhadu 0.

Dále pokračujeme výpočty pro relativní úspěšnost záchytu kamerou/čidlem dle Odhadu 1, jenž je určen výrazem

$$\frac{s_{N^{(p)}N^{(q)}}}{N^{(q)}}$$

pro kameru a výrazem

$$\frac{s_{N^{(p)}N^{(q)}}}{N^{(p)}}$$

pro automatické čidlo.

Pro další analýzu je potřeba upravit výstupní data, která nesplňují kritérium, že se jedná o hodnotu pravděpodobnosti. Konkrétně je nutné ošetřit výstupy Odhadu 1 (pro kameru i čidlo), která jsou menší než 0 (upravíme na hodnotu 0), větší než 1 (upravíme na hodnotu 1), resp. nejsou definovány (*NaN*). Zde se řídíme náhradou hodnot popsanou v podkapitole 5.2, resp. 5.2.1.

Obdobně stanovíme i "pravidla" pro náhradu *NaN* pro Odhady 2 (viz podkapitola 5.2.2). Druhý odhad záchytu kamerou je určen pomocí vzorce

$$\frac{s_{N^{(p)}N^{(q)}}}{s_{N^{(q)}}^2},$$

obdobně pro záchyt čidlem

$$\frac{s_{N^{(p)}N^{(q)}}}{s_{N^{(p)}}^2}.$$

Nejprve opět hodnoty nižší než 0 zaokrouhlíme na nulu, naopak hodnoty vyšší než 1 zaokrouhlíme na tuto maximální krajní hodnotu, tj. 1. Dále zanalyzujeme jednotlivé příčiny vzniku výrazu *NaN*.

Ke vzniku *NaN* může v rámci simulací dojít z několika důvodů, proto je nutná analýza definičního oboru a obecně jakási zpětná detekce možných příčin:

- Počet opakování jednotlivých měření musí být $n \geq 2$, abychom splnili podmínku $\frac{1}{n-1}$ ze vzorce pro výběrovou kovarianci (viz vzorec (5.2) a výběrový rozptyl (vzorec (5.10))).
- Hodnota průměru \bar{X} , resp. \bar{Y} ve jmenovateli Odhadu 1 musí být různá od nuly. Za jakých okolností může hodnota průměru být rovna nule, viz podkapitola 5.2.1.
- Hodnota výběrového rozptylu $s_{N^{(p)}}^2$, $s_{N^{(q)}}^2$ ve jmenovateli Odhadu 2 musí být číslo různé od nuly. Analýza konkrétních situací, kdy se výběrový rozptyl rovná nule, je rozebrán v podkapitole 5.2.2.

- Vznik neurčitého výrazu $\frac{0}{0}$ je zapříčiněn nulovým jmenovatelem (viz předchozí body) a zároveň nulovým čitatelem, který je v obou odhadech určen výběrovou kovariancí. Za jakých okolností je výběrová kovariance rovna nule je analyzováno v kapitole 5.2.1.

Je patrné, že první zmiňovaný bod je možné ošetřit nastavením parametru n . Vznikl-li výraz NaN díky nulové hodnotě průměru $\overline{N^{(p)}}$, resp. $\overline{N^{(q)}}$ (analýza pomocí zpětné detekce), je potřeba zjistit, zda skutečně do obchodního centra nepřišel žádný zákazník (poté lze nahradit NaN hodnotou 1), nebo nějací zákazníci přišli, ale byli zachyceni pouze čidlem (nebo jen kamerou). Protože v Odhadu 1 používáme ve jmenovateli hodnotu odvozené vždy z druhé veličiny, lze v případě, kdy zjistíme, že námi testovaná veličina jisté záchyty provedla, určit $\hat{q} = 0$ (v případě, kdy kamera i čidlo nezachytili žádného příchozího zákazníka), resp. $\hat{q} = 1$ (v případě, kdy kamera nezachytila žádný příchod, ale čidlo ano). Alternativně se provede úprava hodnot \hat{p} pro Odhad 1.

6.1 Analýza simulací

Nyní provedeme analýzu grafických výstupů testovaných odhadů.

Všechny analyzované grafické výstupy byly v programu Matlab počítány pro 1000 simulací. Provedením těchto simulací získáme náhodný výběr $\hat{p}_1, \hat{p}_2, \dots, \hat{p}_{1000}$ z rozdělení odhadu \hat{p} . Střední hodnotu $\mathbf{E}(\hat{p})$ odhadneme výběrovým průměrem tohoto výběru, tj. $\bar{\hat{p}}$. Rozptyl $\sigma_{\hat{p}}^2$ odhadneme pomocí výběrového rozptylu $s_{\hat{p}}^2$. Směrodatnou odchylku $\sqrt{\sigma_{\hat{p}}^2}$ odhadneme výběrovou směrodatnou odchylkou $s_{\hat{p}}$.

Obdobně získáme náhodný výběr $\hat{q}_1, \hat{q}_2, \dots, \hat{q}_{1000}$ z rozdělení odhadu \hat{q} . A alternativně obdržíme výběrové momenty $\bar{\hat{q}}$, $s_{\hat{q}}^2$ a $s_{\hat{q}}$.

Pro samotná rozdělení \hat{p} i \hat{q} byl proveden test na normalitu pomocí Q-Q grafů (viz Obrázky 7.1, 7.2 a 7.3 pro všechny odhady čidla, resp. Obrázky 7.4, 7.5 a 7.6 pro všechny odhady kamery). Grafické výstupy nají v porovnání s normálním rozdělením těžší konce.

Pro 1000 simulací ($\sqrt{1000} \doteq 31.6228$) se rozmezí pohybuje v intervalu hodnot $+ / - \approx 31.6228$ procentních bodů od hledaného parametru.

V grafických výstupech používáme pravidlo "3-sigma" pro vyjádření variability. Tzn., že u přibližně normálního rozdělení by 99,73% testovaných hodnot mělo náležet do rozmezí trojnásobku směrodatných odchylek od hodnoty výběrového průměru.

6.1.1 Závislost průměrné úspěšnosti odhadů pro čidlo na parametru p

Dané simulace byly provedeny pro hodnoty $n = 50$ a $\lambda = 50$.

Porovnáváme-li závislost průměrné úspěšnosti odhadů pro čidlo na hodnotě parametru p , je z Obrázků 7.7¹, 7.8², 7.9³ a 7.10⁴ patrné, že pro vybrané trojice hodnot se průměrné hodnoty odhadů úspěšnosti s rostoucí hodnotou parametru p blíží k průměrné hodnotě Odhadu 0. Nicméně, analyzujeme-li závislost na rostoucím parametru q , platí opačné tvrzení: S rostoucím parametrem q jsou nasimulované hodnoty v průměru méně přesné (srovnejme např. grafický výstup pro $q = 0.01$ a $q = 0.99$ na Obrázku 7.7). Pro vyšší hodnoty pozorovaného parametru q v Odhadu 1 nepozorujeme ani při rostoucím p zlepšení přesnosti měření (viz srovnání s Odhadem 2)).

U všech simulací je také patrné, že odhad číslo 2 se více blíží hodnotám Odhadu 0, vycházíme-li z grafického porovnávání rozptylů. Odhady 1 pro čidlo pro nižší hodnoty parametru q (např. $q = 0.1$, $q = 0.25$, $q = \frac{1}{3}$) mají obecně větší rozptyl než Odhady 2, tj. Odhady 2 jsou pro rostoucí p bližší hodnotám Odhadu 0. Pro rostoucí hodnoty q je rozdíl mezi naměřenými hodnotami rozptylů mezi oběma odhady stále více patrný. Viz Obrázek 7.7: Téměř identické hodnoty rozptylů získáme pro oba odhady pro hodnotu $q = 0.01$. Pokud je pravděpodobnost příchodu zákazníků rovna 1%, pravděpodobnost jejich detekce pomocí automatického čidla je rovněž velmi malá - zde není rozdíl mezi odhady identifikovatelný. Pro $q = 0.5$ se rozdíl mezi rozptyly obou odhadů zvětšuje, až pro hodnotu $q = 0.99$, kde Odhad 2 nejen dosahuje přesnějších hodnot z hlediska průměrné úspěšnosti, ale je více přesný i z hlediska rozptylu. Obrázek 7.8, jenž zachycuje závislost

¹V dané simulaci bylo pro Odhad 1 pro čidlo 1.60% vyšších hodnot upraveno na hodnotu 1, 1.14% nižších hodnot na hodnotu 0. Pro Odhad 2 bylo upraveno 1.15% vyšších hodnot na hodnotu 1, 1.14% nižších hodnot na hodnotu 0 a pro oba dva odhady bylo celkem ošetřeno 0,01% výskytů *NaN*.

²V této simulaci pro Odhad 1 pro čidlo bylo 0.99% vyšších hodnot upraveno na hodnotu 1, 0.43% nižších hodnot na hodnotu 0. Pro Odhad 2 bylo upraveno 0.75% vyšších hodnot na hodnotu 1, 0.43% nižších hodnot na hodnotu 0 a pro oba dva odhady bylo celkem ošetřeno 0.00009% výskytů *NaN*.

³Pro tuto simulaci pro Odhad 1 pro čidlo bylo 0.56% vyšších hodnot upraveno na hodnotu 1, 0.23% nižších hodnot na hodnotu 0. Pro Odhad 2 bylo upraveno 0.35% vyšších hodnot na hodnotu 1, 0.23% nižších hodnot na hodnotu 0 a pro oba dva odhady bylo celkem ošetřeno 0.0001% výskytů *NaN*.

⁴V dané simulaci pro Odhad 1 pro čidlo bylo 0.37% vyšších hodnot upraveno na hodnotu 1, 0.20% nižších hodnot na hodnotu 0. Pro Odhad 2 bylo upraveno 0.25% vyšších hodnot na hodnotu 1, 0.20% nižších hodnot na hodnotu 0 a pro oba dva odhady bylo celkem ošetřeno 0.0001% výskytů *NaN*.

průměrnou úspěšnost odhadu pro čidlo na konkrétních hodnotách $q = 0.1$, $q = 0.5$ a $q = 0.9$, vykazuje podobné chování. Pro nižší testované hodnoty $q = 0.1$ pozorujeme u obou dvou odhadů téměř identický grafický výstup. Pro nejvyšší testovanou hodnotu $q = 0.9$ opět můžeme pozorovat, že Odhad 2 dosahuje vyšší přesnosti (s rostoucím p až řádově okolo jedné desetiny) než Odhad 1, rozptyl druhého odhadu je rovněž menší v porovnání s odhadem prvním.

V Obrázku 7.9 porovnáváme kromě $q = 0.5$ také vybrané hodnoty $q = 0.25$ a $q = 0.75$. I zde je patrné následující chování - pro nižší hodnotu p jsou odhady téměř identické (porovnáme-li tuto závislost na parametru p s předchozími Obrázky 7.7, resp. 7.8, můžeme pozorovat, že s rostoucí hodnotou q je rozdíl mezi variabilitou jednotlivých odhadů skutečně zjevnější (viz testování pro $q = 0.1$ a $q = 0.25$). I pro druhou krajní hodnotu platí, že naměřená průměrná úspěšnost odhadu pro vyšší vybranou hodnotu (zde $q = 0.75$) je v rámci daného výběru závislosti na testovaných hodnotách nejméně přesná (viz porovnání odhadů 1 a 2 v daném obrázku) - jak pro průměrné hodnoty úspěšnosti, tak pro hodnoty rozptylu.

Poslední testovaný Obrázek 7.10 zachycuje analýzu hodnot $q = \frac{1}{3}$ a $q = \frac{2}{3}$. Tento grafický výstup opět potvrzuje pozorované chování obou odhadů, tj. Odhad 1 je obecně přesnější než Odhad 2 (tato tendence je ovšem méně pozorovatelná než u vyšších hodnot parametru p (např. Obrázek 7.7 a 7.8)). A zároveň s rostoucím parametrem q také roste rozdíl mezi rozptyly obou odhadů.

Z analýzy všech dosavadních grafických výstupů není pro pozorování s vyššími hodnotami q patrná závislost obou odhadů úspěšnosti automatického čidla na parametru p . Jinak řečeno, nelze jednoznačně potvrdit, že zvýší-li se přesnost jedné měřící možnosti (zde kamera), ovlivní to přesnost detekce druhou metodou měření (zde automatické čidlo).

6.1.2 Závislost průměrné úspěšnosti odhadů pro kameru na parametru q

Obdobná simulace byla provedena pro testování závislosti odhadů kamery na parametru q . Grafické výstupy jsou pro různé hodnoty testovaných hodnot parametru q zachyceny na Obrázcích 7.11, 7.12, 7.13 a 7.14. Díky symetrii úlohy jsme dostali obdobné výstupy, z nichž tedy vyplývají i obdobné závěry.⁵

⁵Pro dané grafické výstupy se úprava hodnot nižších než 0, vyšších než 1 a *NaN* rovněž pohybovala mezi 0% - 2%.

6.1.3 Závislost průměrné úspěšnosti odhadů pro čidlo na parametru λ

Veškeré simulace byly provedeny pro hodnoty $n = 50$ a $p = 0.95$.⁶

Dále byla testována závislost průměrné úspěšnosti všech odhadů na parametru λ . Na Obrázku 7.15 je zachycen parametr λ pro hodnotu $q = 0.01$. Odhad 1 a Odhad 2 se pro parametr λ z intervalu $(1; 100)$ příchozích zákazníků za časový úsek chovají téměř identicky. Naměřená data sice dosahují mírně vyšších hodnot než Odhad 0, jedná se ale pouze o jednotky tisícín.

Na Obrázku 7.16 je zachycena závislost parametru $q = 0.5$ na λ . Pro tuto vyšší hodnotu q je patrné, že střední hodnota Odhadu 2 se více blíží střední hodnotě Odhadu 0. Pro rozptyl odhadů platí to samé, tj. variance Odhadu 2 dosahuje nižších hodnot (a je bližší hodnotám nulového odhadu) než rozptyl Odhadu 1.

Obrázek 7.17 zobrazuje závislost čidla na λ pro hodnotu $q = 0.99$. Podobně jako pro testování úspěšnosti záchytu zákazníka čidlem na parametru p , i zde dochází k většímu rozdílu mezi Odhady 1 a 2. Je patrné, že i zde je pro rostoucí parametr q Odhad 1 vykazuje vyšší hodnoty výběrového rozptylu než Odhad 2. Samotná závislost na parametru λ ovšem není z těchto grafických výstupů patrná.

6.1.4 Závislost průměrné úspěšnosti odhadů pro kameru na parametru λ

Alternativně byly provedeny simulace závislosti všech odhadů pro kameru na parametru λ . Tentokrát pro hodnoty parametrů $q = 0.85$ a $n = 50$.⁷

Na Obrázku 7.18 je zachycena závislost na parametru λ pro konkrétní hodnotu $p = 0.25$. Ze zachyceného měření lze vypožorovat, že Odhad 1 a Odhad 2 se chovají velmi podobně. Rozdíl je avšak zřetelný mezi výběrovými rozptyly daných odhadů. A sice, výběrový rozptyl Odhadu 2 je menší než výběrový rozptyl Odhadu 1.

Závislost na parametru λ pro hodnotu $p = 0.5$ je zobrazena na Obrázku 7.19. S rostoucí hodnotou p si můžeme povšimnout i většího přiblížování odhadů k hodnotám Odhadu 0. Ovšem závislost průměrné úspěšnosti odhadů na rostoucím λ až tak zřetelná není. Z tohoto grafického výstupu můžeme potvrdit, že výběrový rozptyl Odhadu 1 dosahuje širšího intervalu než výběrový rozptyl Odhadu 2.

⁶Úprava hodnot nižších než 0, vyšších než 1, resp. *NaN* se týkala 0% - 1.56% hodnot.

⁷Pro tyto simulace se úprava hodnot nižších než 0, vyšších než 1, resp. *NaN* se týkala 0% - 1.62% hodnot.

Poslední grafický výstup, tj. Obrázek 7.20 znázorňuje simulaci závislosti úspěšnosti odhadů pro kameru λ při daném parametru $p = 0.75$. Lze si povšimnout, že i zde s rostoucí hodnotou p je stále patrnější rozdíl v přibližování naměřených hodnot průměrné úspěšnosti i výběrového rozptylu obou odhadů vůči nulovému odhadu. A sice, daný grafický výstup se jeví ve prospěch Odhadu 2. Co se týče samotné závislosti přesnosti odhadů na parametru λ , ani z tohoto grafického výstupu nelze tuto závislost vypožorovat.

6.1.5 Závislost průměrné úspěšnosti odhadů pro čidlo na parametru n

Závislost průměrné úspěšnosti obou odhadů pro čidlo na parametru n byla testována pomocí simulačních výpočtů.

Odhad 1

Výsledkem simulace pro pevné hodnoty parametrů $q = 0.85$, $p = 0.95$ a $\lambda = 20$ jsou následující údaje:

Parametr n	Průměrná úspěšnost čidla	výb. směrodatná odchylka čidla
5	0.8193	0.6460
10	0.8445	0.4201
25	0.8424	0.2251
50	0.8689	0.1954
100	0.8385	0.1202

Ze získaných dat je patrné, že pro $n = 5$ průměrná úspěšnost prvního odhadu pro čidlo odpovídá hodnotě přibližně o 3% nižší, než je hodnota parametru pro Odhad 0. Z vybraných dat je skutečnému parametru nejbližší průměrná úspěšnost odhadu pro $n = 10$. Z hlediska výběrového rozptylu, resp. výběrové směrodatné odchylky se však odhad s rostoucím n zpřesňuje.

Odhad 2

Nyní zanalyzujeme výsledky simulace pro druhý odhad pro čidlo s pevnými parametry $q = 0.85$, $p = 0.95$ a $\lambda = 20$.

Parametr n	Průměrná úspěšnost čidla	výb. směrodatná odchylka čidla
5	0.7928	0.3435
10	0.8520	0.1425
25	0.8450	0.0755
50	0.8543	0.0637
100	0.8510	0.0423

Z těchto dat je patrné, že nejbližší výchozí hodnotě je odhad průměrné úspěšnosti čidla pro nejvyšší testovanou hodnotu n , tj. $n = 100$. Nejnižší testovaná hodnota $n = 5$ vykazuje největší zaznamenaný rozdíl mezi odhadem a $q = 0.85$ (a v porovnání s Odhadem 1 dokonce i větší). Výběrová směrodatná odchylka s rostoucím n klesá.

6.1.6 Závislost průměrné úspěšnosti odhadů pro kameru na parametru n

Obdobně i závislost obou odhadů pro kameru na parametru n byla testována pomocí simulačních výpočtů.

Odhad 1

Výsledkem simulace pro pevné hodnoty parametrů $p = 0.95$, $q = 0.85$ a $\lambda = 20$ jsou následující údaje:

Parametr n	Průměrná úspěšnost kamery	výb. směrodatná odchylka kamery
5	0.9130	0.7102
10	0.9440	0.4621
25	0.9411	0.2562
50	0.9732	0.2174
100	0.9350	0.1344

Ze získaných údajů je patrné, že nejbližší je odhad průměrné úspěšnosti kamery výchozímu parametru pro hodnotu $n = 10$ (obdobně jako v Odhadu 1 pro čidlo). Výběrová směrodatná odchylka se s rostoucím parametrem n zmenšuje.

Odhad 2

I pro druhý odhad pro kameru byly provedeny simulace pro pevné hodnoty parametrů $p = 0.95$, $q = 0.85$ a $\lambda = 20$.

Parametr n	Průměrná úspěšnost kamery	výb. směrodatná odchylka kamery
5	0.9363	0.3737
10	0.9496	0.1691
25	0.9632	0.0817
50	0.9516	0.0665
100	0.9432	0.0508

Odhad průměrné úspěšnosti kamery nejbližší hodnotě $p = 0.95$ byl v provedených simulacích zaznamenán pro $n = 10$. S rostoucím n klesá výběrová směrodatná odchylka.

Ze získaných dat se nedá jednoznačně potvrdit závislost průměrné úspěšnosti odhadů pro čidlo či kameru na parametru n .

7 Závěr

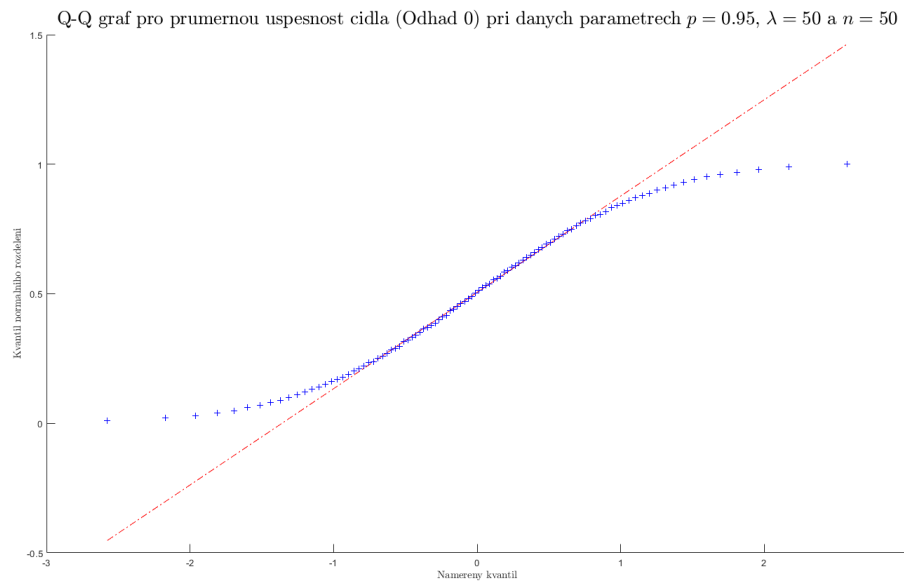
V této práci byly představeny dva odhady zabývající se počtem návštěvníků nákupního centra. Po teoretickém nástihu odhadů byla provedena simulační analýza, která se zabývala otázkami vlastností těchto odhadů.

Z grafické analýzy vyplývá, že chybějící informace o celkovém počtu příchozích návštěvníků nákupního centra má vliv na jednotlivé odhady, zejména na Odhad 1.

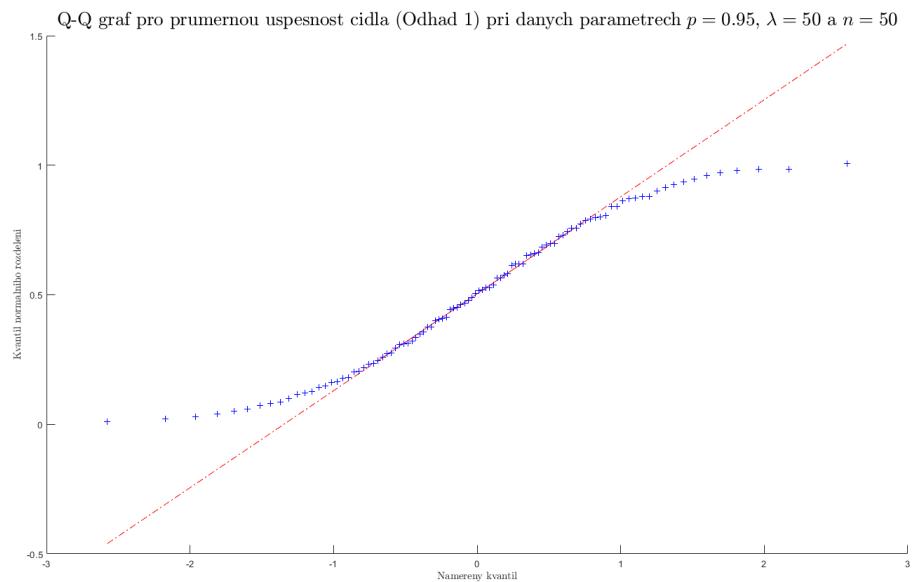
Porovnáme-li Odhad 0 s Odhadem 2, simulační analýza ukázala, že s rostoucími hodnotami parametrů p , resp. q se odhad průměrné úspěšnosti testovaných zařízení blíží průměrné úspěšnosti Odhadu 0. Srovnáme-li nulový odhad i s Odhadem 1, je patrné, že pro vyšší testované hodnoty parametrů $p \approx 0.9$ a $q \approx 0.9$ dochází ve srovnání s Odhadem 2 k větším rozdílům mezi průměrnými úspěšnostmi těchto odhadů s odhadem nulovým. Stejný závěr je patrný i pro výběrový rozptyl daných odhadů.

Ačkoliv se pro nižší hodnoty jeví grafické výstupy Odhadu 1 a Odhadu 2 jako velmi podobné, lepší variantou pro informace o průměrné úspěšnosti a variabilitě je volba Odhadu 2.

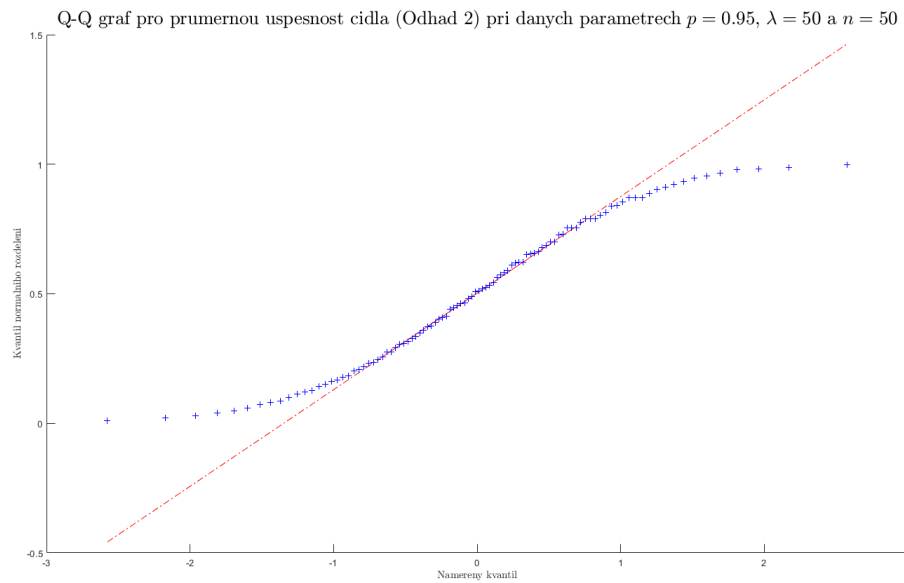
Obrazové přílohy



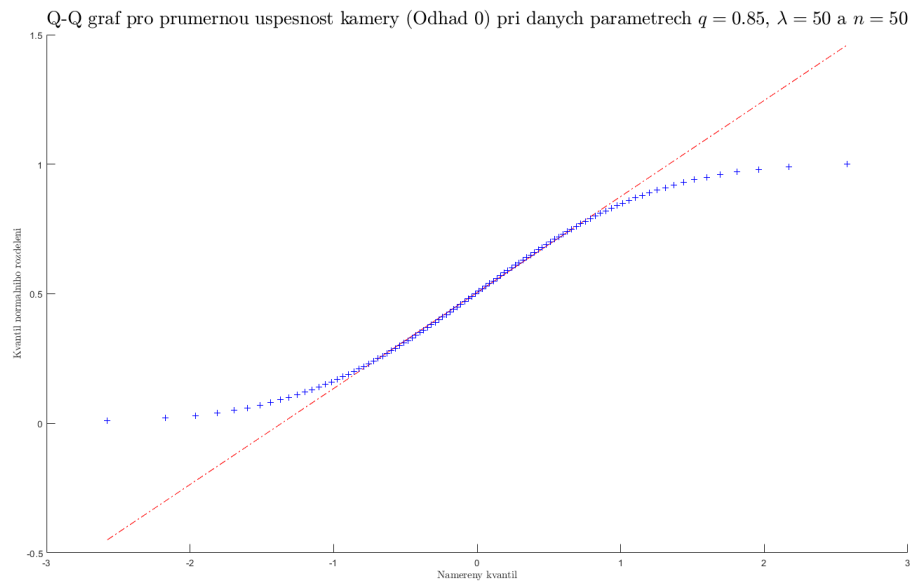
Obrázek 7.1: *Q-Q graf pro čidlo (Odhad 0) při daných parametrech $p = 0.95$, $\lambda = 50$ a $n = 50$.*



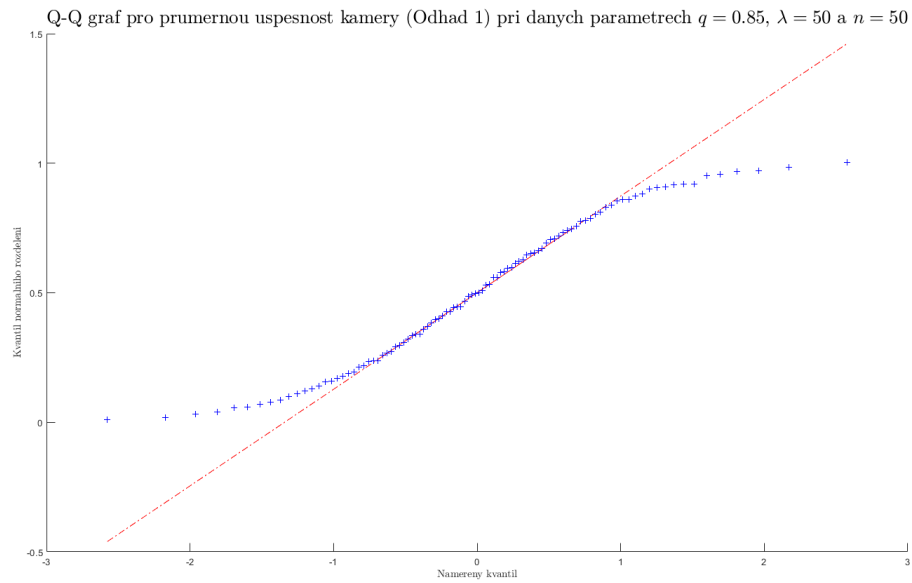
Obrázek 7.2: *Q-Q graf pro čidlo (Odhad 1) při daných parametrech $p = 0.95$, $\lambda = 50$ a $n = 50$.*



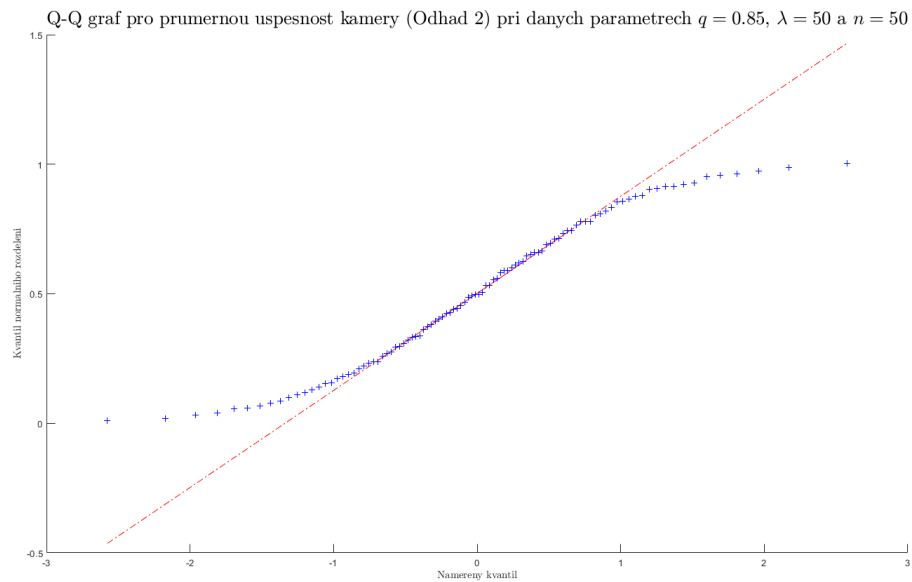
Obrázek 7.3: *Q-Q graf pro čidlo (Odhad 3) při daných parametrech $p = 0.95$, $\lambda = 50$ a $n = 50$.*



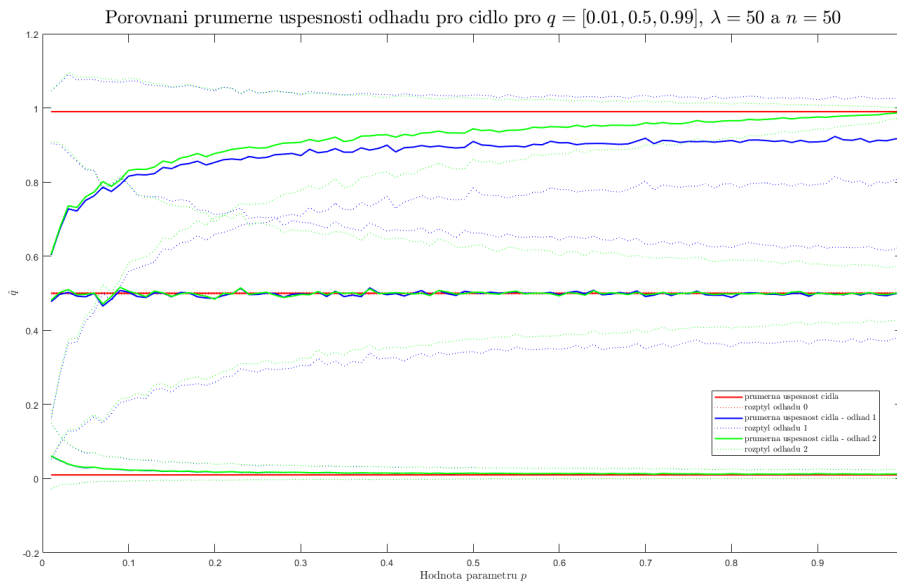
Obrázek 7.4: *Q-Q graf pro kameru (Odhad 0) při daných parametrech $q = 0.85$, $\lambda = 50$ a $n = 50$.*



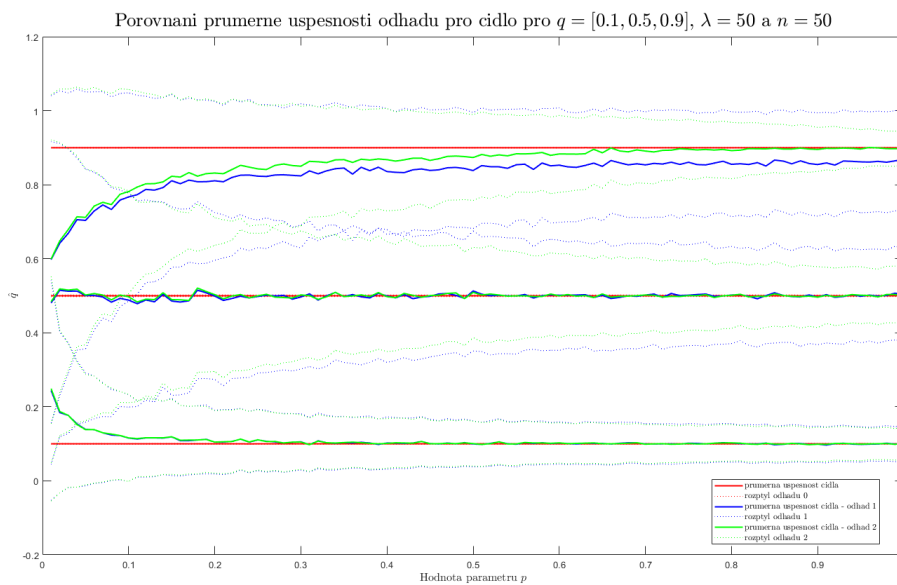
Obrázek 7.5: *Q-Q graf pro kameru (Odhad 1) při daných parametrech $q = 0.85$, $\lambda = 50$ a $n = 50$.*



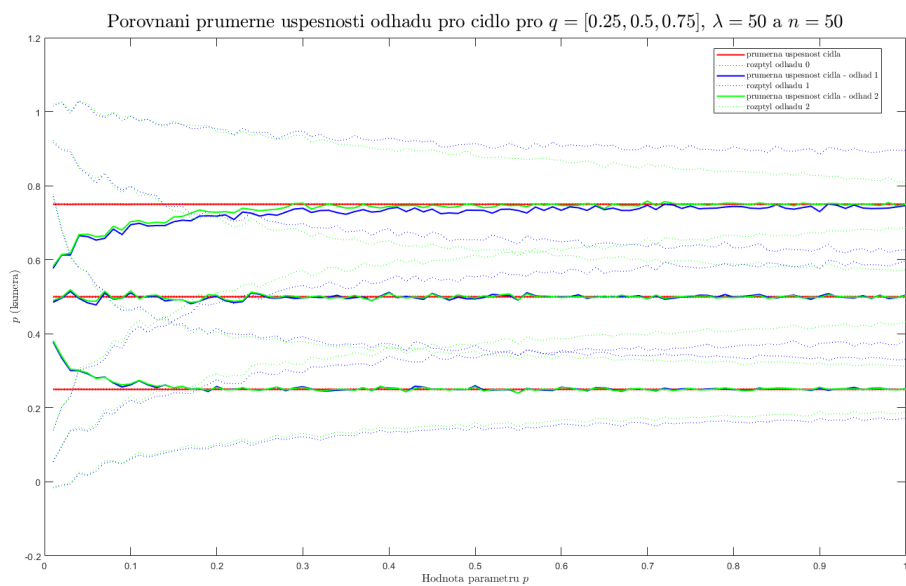
Obrázek 7.6: *Q-Q graf pro kameru (Odhad 2) při daných parametrech $q = 0.85$, $\lambda = 50$ a $n = 50$.*



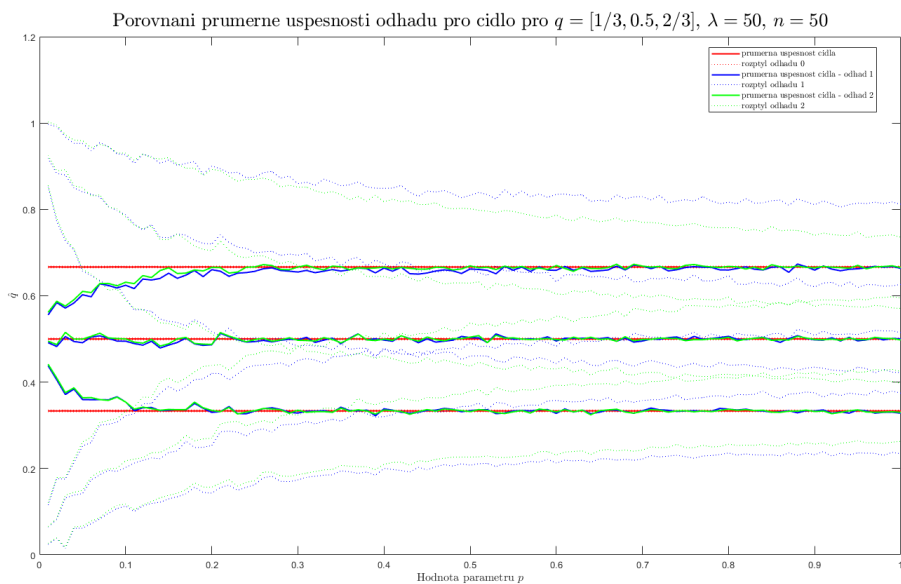
Obrázek 7.7: Porovnání odhadů průměrné úspěšnosti čidla pro konkrétní hodnoty. Závislost na parametru p , zde pro vybrané hodnoty $q = 0.01$, $q = 0.5$, $q = 0.99$ při hodnotách parametrů $\lambda = 50$ a $n = 50$.



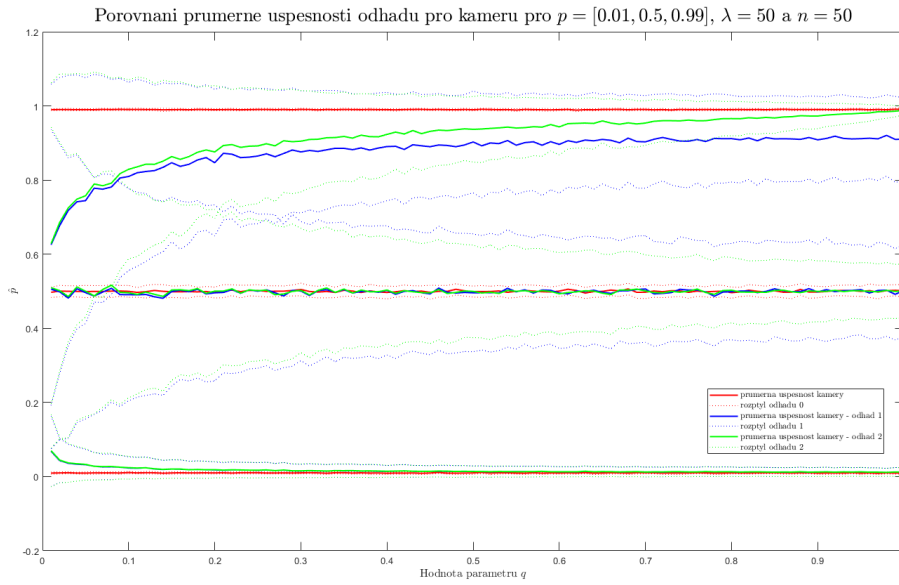
Obrázek 7.8: Porovnání odhadů průměrné úspěšnosti čidla pro konkrétní hodnoty. Závislost na parametru p , zde pro vybrané hodnoty $q = 0.1$, $q = 0.5$, $q = 0.9$ při hodnotách parametrů $\lambda = 50$ a $n = 50$.



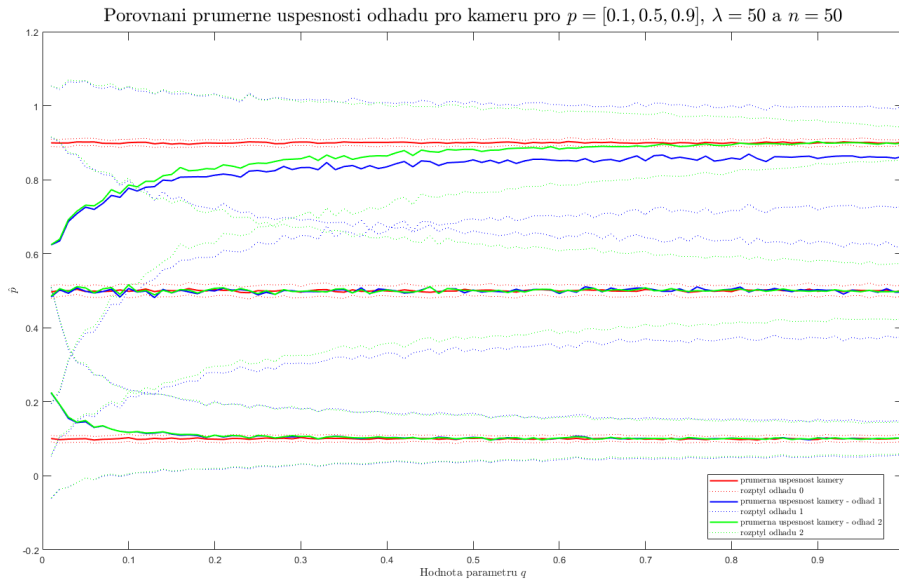
Obrázek 7.9: Porovnání odhadů průměrné úspěšnosti čidla pro konkrétní hodnoty. Závislost na parametru p , zde pro vybrané hodnoty $q = 0.25$, $q = 0.5$, $q = 0.75$ při hodnotách parametrů $\lambda = 50$ a $n = 50$.



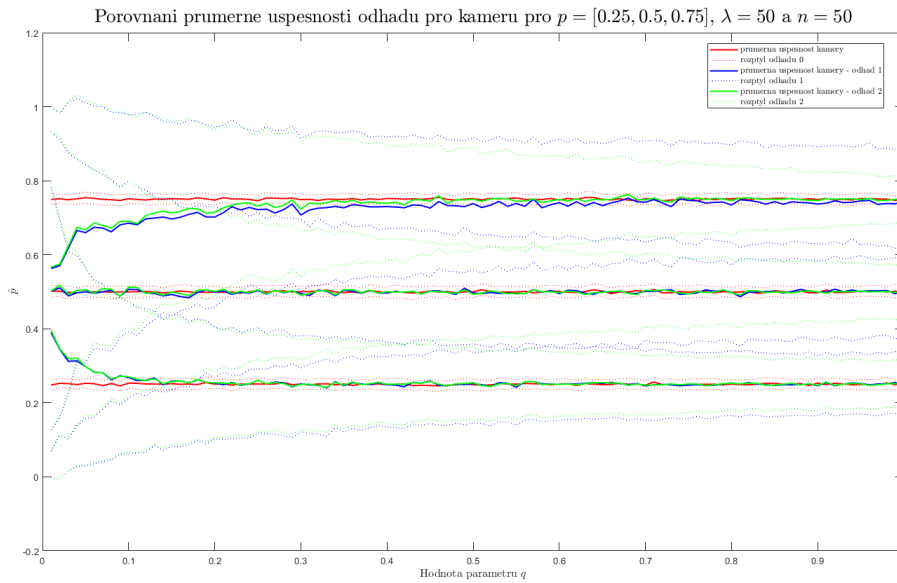
Obrázek 7.10: Porovnání odhadů průměrné úspěšnosti čidla pro konkrétní hodnoty. Závislost na parametru p , zde pro vybrané hodnoty $q = \frac{1}{3}$, $q = 0.5$, $q = \frac{2}{3}$ při hodnotách parametrů $\lambda = 50$ a $n = 50$.



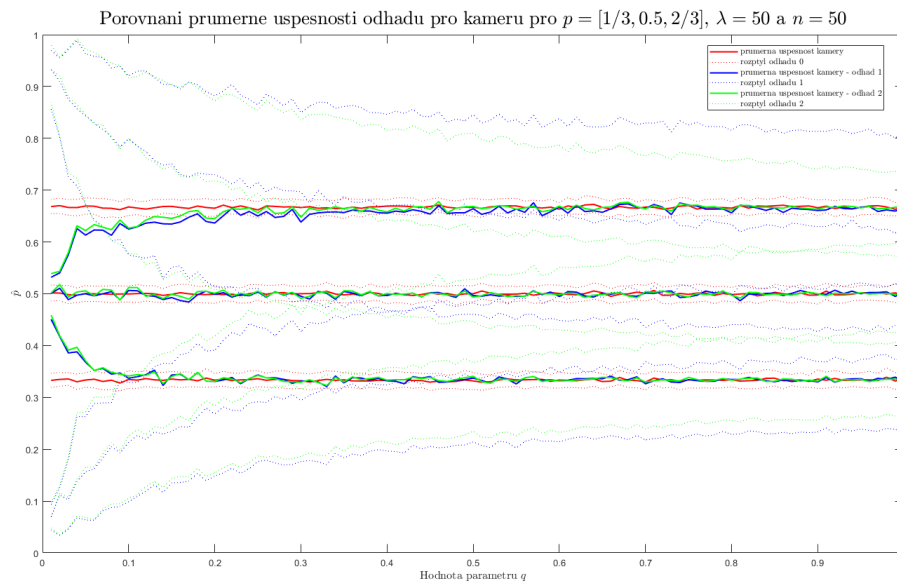
Obrázek 7.11: Porovnání odhadů průměrné úspěšnosti kamery pro konkrétní hodnoty. Závislost na parametru q , zde pro vybrané hodnoty $p = 0.01$, $p = 0.5$, $p = 0.99$ při hodnotách parametrů $\lambda = 50$ a $n = 50$.



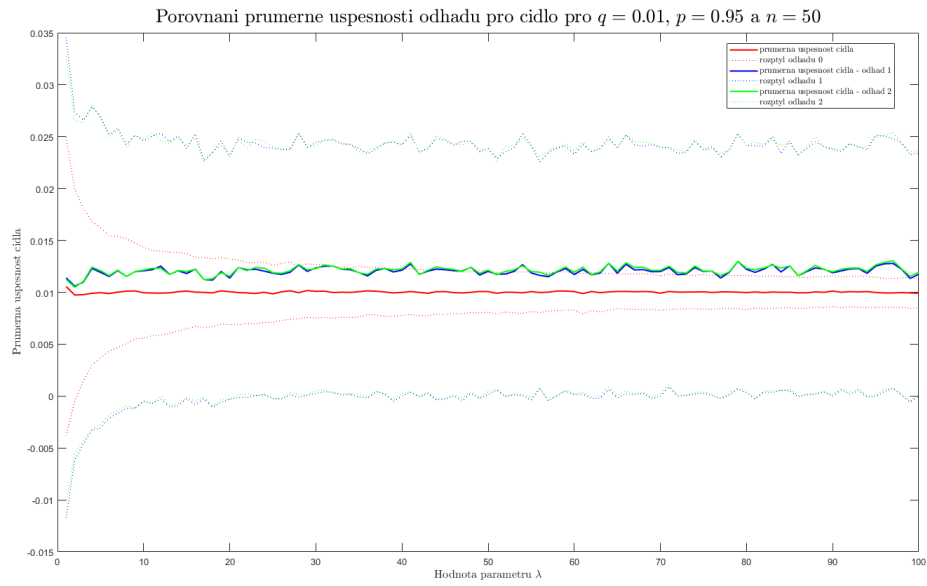
Obrázek 7.12: Porovnání odhadů průměrné úspěšnosti kamery pro konkrétní hodnoty. Závislost na parametru q , zde pro vybrané hodnoty $p = 0.1$, $p = 0.5$, $p = 0.9$ při hodnotách parametrů $\lambda = 50$ a $n = 50$.



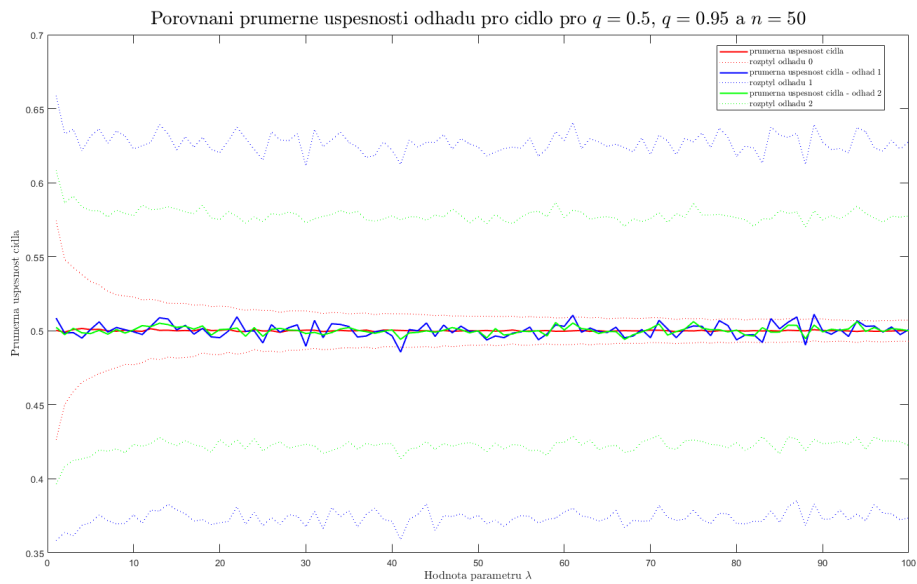
Obrázek 7.13: Porovnání odhadů průměrné úspěšnosti kamery pro konkrétní hodnoty. Závislost na parametru q , zde pro vybrané hodnoty $p = 0.25$, $p = 0.5$, $p = 0.75$ při hodnotách parametrů $\lambda = 50$ a $n = 50$.



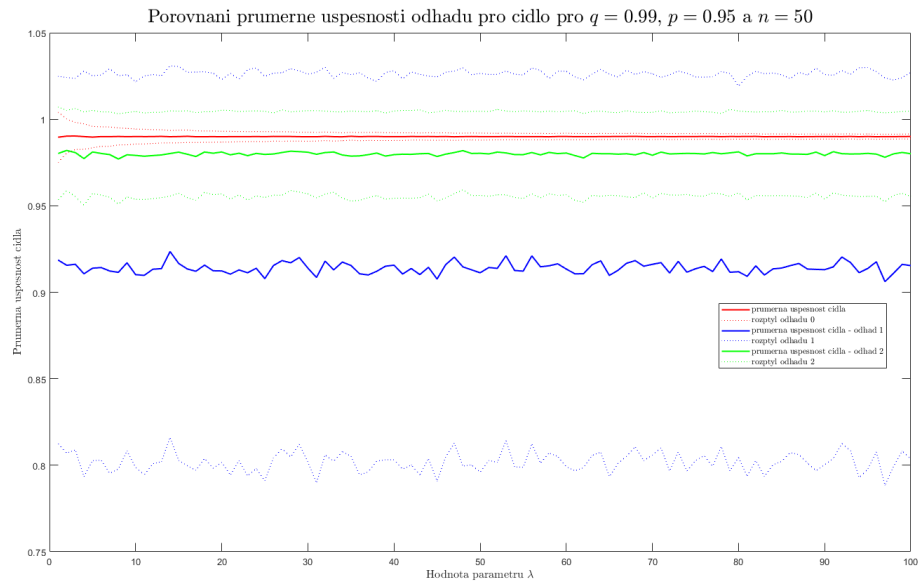
Obrázek 7.14: Porovnání odhadů průměrné úspěšnosti kamery pro konkrétní hodnoty. Závislost na parametru q , zde pro vybrané hodnoty $p = \frac{1}{3}$, $p = 0.5$, $p = \frac{2}{3}$ při hodnotách parametrů $\lambda = 50$ a $n = 50$.



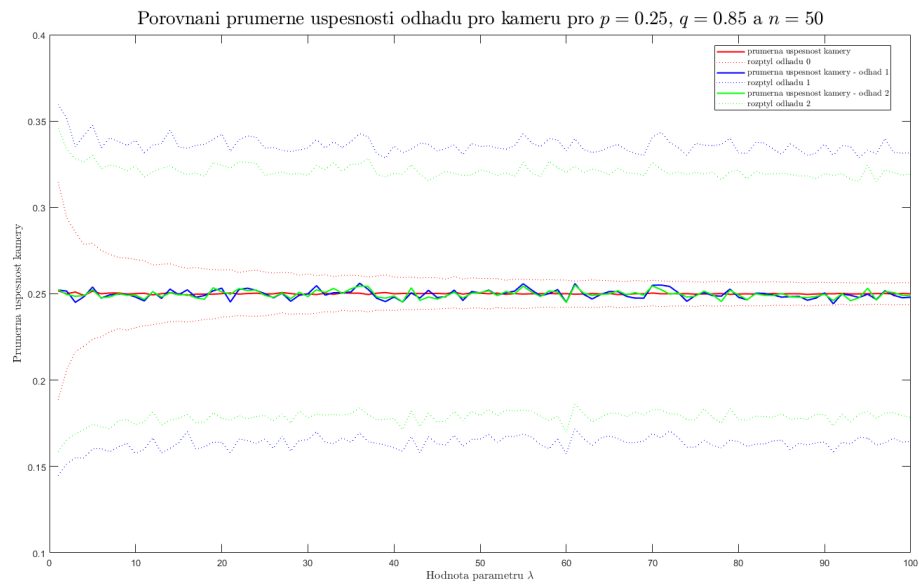
Obrázek 7.15: Porovnání odhadů průměrné úspěšnosti čidla pro hodnotu $q = 0.01$ při hodnotách parametrů $p = 0.95$ a $n = 50$ v závislosti na parametru λ .



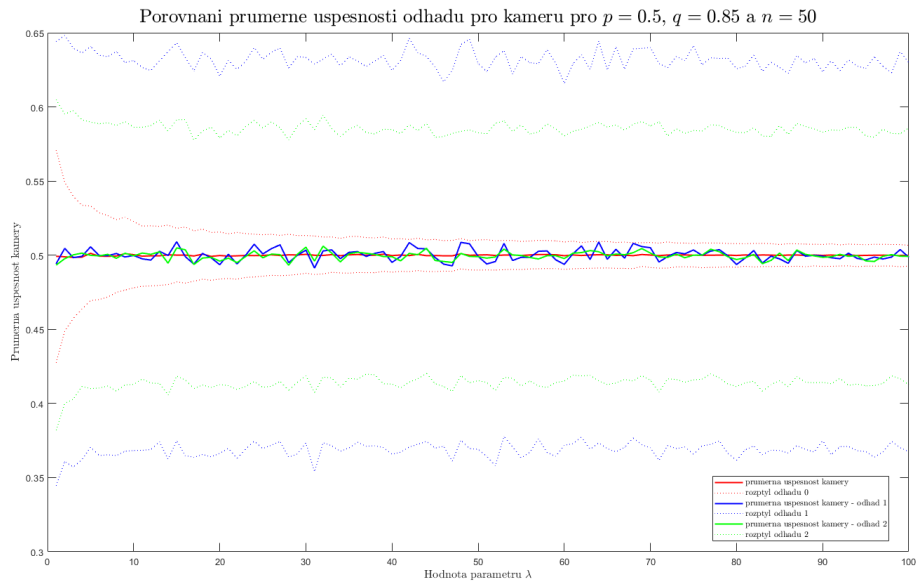
Obrázek 7.16: Porovnání odhadů průměrné úspěšnosti čidla pro hodnotu $q = 0.5$ při hodnotách parametrů $p = 0.95$ a $n = 50$ v závislosti na parametru λ .



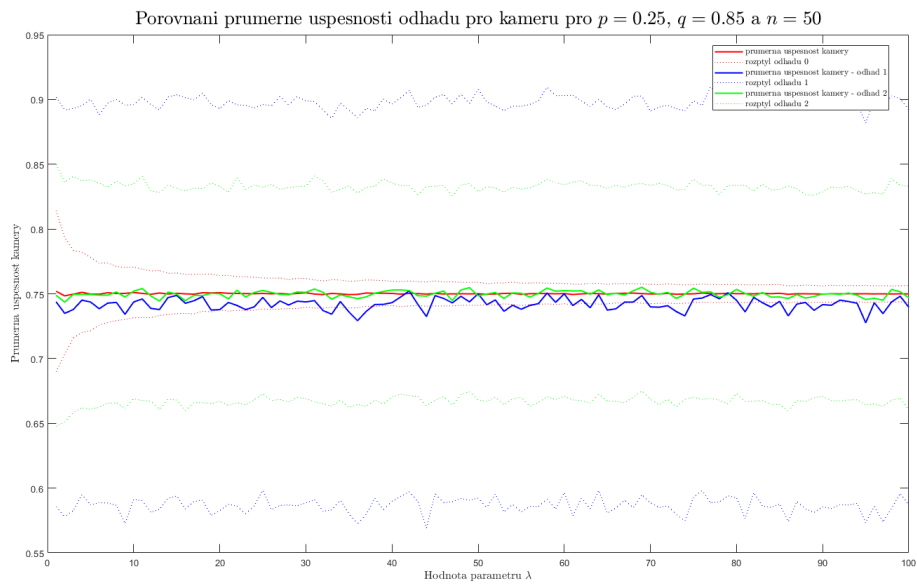
Obrázek 7.17: Porovnání odhadů průměrné úspěšnosti čidla pro hodnotu $q = 0.99$ při hodnotách parametrů $p = 0.95$ a $n = 50$ v závislosti na parametru λ .



Obrázek 7.18: Porovnání odhadů průměrné úspěšnosti kamery pro hodnotu $p = 0.25$ při hodnotách parametrů $q = 0.85$ a $n = 50$ v závislosti na parametru λ .



Obrázek 7.19: Porovnání odhadů průměrné úspěšnosti kamery pro hodnotu $p = 0.5$ při hodnotách parametrů $q = 0.85$ a $n = 50$ v závislosti na parametru λ .



Obrázek 7.20: Porovnání odhadů průměrné úspěšnosti kamery pro hodnotu $p = 0.75$ při hodnotách parametrů $q = 0.85$ a $n = 50$ v závislosti na parametru λ .

Literatura

- [1] REIF, Jiří, KOBEDA, Zdeněk. *Úvod do pravděpodobnosti a spolehlivosti*. Západočeská univerzita v Plzni, Plzeň, 2004. ISBN 80-7043-333-7.
- [2] REIF, Jiří. *Metody matematické statistiky*. Západočeská univerzita v Plzni, Plzeň, 2004. ISBN 80-743-302-7.
- [3] TAYLOR, Howard M., KARLIN, Samuel, *An Introducing To Stochastic Modeling - 3rd Edition*. Academic press, San Diego, California, USA, 1998. ISBN 978-0-12-684887-8. Dostupné také z: https://appliedmath.arizona.edu/sites/default/files/0f04d86a836182cbf608dfc86c7a70f5e5f6_0.pdf
- [4] NEUBAUER, Jiří. *Modely diskrétní náhodné veličiny* [online]. [cit. 2021-09-10]. Dostupné z: https://k101.unob.cz/neubauer/pdf/diskretni_modely.pdf
- [5] OTIPKA, Petr, ŠMAJSTRLA, Vladislav. *Pravděpodobnost a statistika: Náhodný vektor* [online]. [cit. 2022-05-16]. Dostupné z: <https://homel.vsb.cz/oti73/cdpast1/KAP06/PRAV6.HTM>