Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

BAKALÁŘSKÁ PRÁCE

Vojtěch Skubanič

Odhady rozptylů pro závislá pozorování

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: RNDr. Zbyněk Pawlas, Ph.D.
Studijní program: Matematika
Studijní obor: Obecná matematika

2007

Na tomto místě bych rád poděkoval vedoucímu práce RNDr. Zbyňku Pawlasovi, Ph.D. za poskytnutí velkého množství času při konzultacích a za připomínky k obsahu a koncepci této práce.

Prohlašuji, že jsem svou bakalářskou práci napsal samostatně a výhradně s použitím citovaných pramenů. Souhlasím se zapůjčováním práce a jejím zveřejňováním.

V Praze dne 06.07.2007 Vojtěch Skubanič

Obsah

Název práce: Odhady rozptylů pro závislá pozorování
Autor: Vojtěch Skubanič
Katedra (ústav): Katedra pravděpodobnosti a matematické statistiky
Vedoucí bakalářské práce: RNDr. Zbyněk Pawlas, Ph.D.
e-mail vedoucího: Zbynek.Pawlas@mff.cuni.cz

Abstrakt: V předložené práci studujeme odhady rozptylů z náhodných veličin, které nemusí být nezávislé. Cílem je nalézt takové odhady, které budou co nejméně vychýlené s přihlédnutím ke střední čtvercové chybě. Stanovit jejich střední hodnotu, rozptyl a střední čtvercovou chybu. Odhady odvodíme z výběrového rozptylu pro náhodný výběr. Dále popisujeme některé tvary a průběhy autokovariančních funkcí. Provádíme teoretické výpočty středních hodnot a středních čtvercových chyb odhadů rozptylu ze závislých pozorování z normálního rozdělení pro konkrétní volby autokovarianční funkce.

Klíčová slova: odhad rozptylu, závislé pozorování, normální rozdělení

Title: Variance estimation for dependent observations
Author: Vojtěch Skubanič
Department: Department of Probability and Mathematical Statistics
Supervisor: RNDr. Zbyněk Pawlas, Ph.D.
Supervisor's e-mail address: Zbynek.Pawlas@mff.cuni.cz

Abstract: In the present work we study estimators of variance of chance quantities, which may not be independent. Our goal is to find such estimators, which are low biased. Determinate their mean, variation and mean square error. Estimators will be deduce from standard estimator of variance. Hereafter we will describe some shapes and course of autocovariance functions and we will make theoretical calculations of estimators of variance for dependent observations from normal distribution with specific choices of autocovariance function.

Keywords: estimator of variance, dependent observation, normal distribution

Úvod

Tato práce pojednává o odhadech rozptylů ze stejně rozdělených náhodných veličin, které ale nemusí být navzájem nezávislé. V 1. kapitole se bude pojednávat o takovýchto odhadech. Základním odhadem, ze kterého se vyjde, bude výběrový rozptyl S^2. Bude vypočtena jeho střední hodnota a speciálně pro odhad z náhodných veličin s normálním rozdělením jeho rozptyl a střední čtvercová chyba. Tento odhad je vychýlený a proto se ho pokusíme upravit odstraněním některých porovnání náhodných veličin v odhadu S^2. Takto sestavíme odhady, které budou označeny U_h. Opět se vyjádří jejich střední hodnota a rozptyl společně se střední čtvercovou chybou pro případ normálního rozdělení. Další odhad bude založen na vhodném odhadnutí vychýlení u odhadu S^2, tímto způsobem sestavíme odhady T_m. Takový odhad ale bude pro náhodný výběr vychýlený a proto odhad T_m vhodně přenásobíme a tak získáme odhady nové, které budou pro nezávislé náhodné veličiny nestranné. Tyto odhady označíme $\widehat{T}_m$. Ve 2. kapitole budeme pojednávat o tvarech autokovariančních funkcí a pro exponenciální autokovarianční funkci vyjádříme jednoduchým způsobem střední hodnotu odhadů S^2 a U_h. Dále provedeme výpočet středních hodnot a středních čtvercových chyb z normálního rozdělení pro konkrétní volby autokovarianční funkce.

Kapitola 1

Odhady rozptylů

Cílem této kapitoly bude zavést některé odhady rozptylů, nalézt jejich střední hodnotu a rozptyl. Jako další charakteristiku určíme střední čtvercovou chybu, která je často označována jako MSE.

Definice 1.1. Nechť $X_1, \ldots, X_n$ jsou stejně rozdělené náhodné veličiny, U je odhad parametru θ, pak střední čtvercovou chybu MSE definuji přepisem:

$$\mathsf{MSE}\,(U) = \mathsf{E}(U - \theta)^2.$$

Pro výpočet střední čtvercové chyby se bude používat vztah mezi střední čtvercovou chybou, střední hodnotou a rozptylem příslušného odhadu v následujícím jednoduchém lemmatu.

Lemma 1.2. Nechť U je odhad parametru θ, pak lze střední čtvercová chyba vyjádřit vztahem:

$$\mathsf{MSE}\,(U) = \mathsf{var}\, U + (\mathsf{E}\, U - \theta)^2. \tag{1.1}$$

Důkaz.

$$\begin{aligned}
\mathsf{E}\,(U - \theta)^2 &= \mathsf{E}\, U^2 - 2\theta \mathsf{E}\, U + \theta^2 + \\
&\quad + (\mathsf{E}\, U - \theta)^2 - (\mathsf{E}\, U)^2 + 2\theta \mathsf{E}\, U - (\theta)^2 \\
&= \mathsf{E}\, U^2 - (\mathsf{E}\, U)^2 + (\mathsf{E}\, U - \theta)^2 \\
&= \mathsf{var}\, U + (\mathsf{E}\, U - \theta)^2.
\end{aligned}$$

□

Pro výpočet rozptylu odhadů z normálního rozdělení budeme využívat nasledující formuli.

Lemma 1.3. (Isserlitzova formule) Nechť Y_1, Y_2, Y_3, Y_4 jsou náhodné veličiny z normálního rozdělení o střední hodnotě 0, pak platí

$$\mathsf{E}\, Y_1 Y_2 Y_3 Y_4 = \mathsf{E}\, Y_1 Y_2 \mathsf{E}\, Y_3 Y_4 + \mathsf{E}\, Y_2 Y_3 \mathsf{E}\, Y_1 Y_4 + \mathsf{E}\, Y_1 Y_4 \mathsf{E}\, Y_2 Y_3. \tag{1.2}$$

Důkaz. Důkaz uvedl Isserlitz v [2]. □

1.1 Odhad rozptylu pomocí S^2

Definice 1.4. Mějme $n \geq 2$, nechť $X_1, \ldots, X_n$ jsou stejně rozdělené náhodné veličiny, $\overline{X}$ je jejich výběrový průměr, pak odhad S^2 definuji předpisem.

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2. \tag{1.3}$$

Poznámka. Tento odhad se obvykle nazývá výběrový rozptyl.

Lemma 1.5. Nechť $n \geq 2$ a $X_1, \ldots, X_n$ jsou stejně rozdělené náhodné veličiny z rozdělení o střední hodnotě μ, pak pro S^2 platí následující vztah:

$$S^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2 - \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} (X_i - \mu)(X_j - \mu). \tag{1.4}$$

Důkaz.

$$\begin{aligned} S^2 &= \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2 \\ &= \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \mu - \overline{X} + \mu)^2 \\ &= \frac{1}{n-1} \sum_{i=1}^{n} \left((X_i - \mu)^2 + (\overline{X} - \mu)^2 + 2(X_i - \mu)(\mu - \overline{X}) \right). \end{aligned}$$

Z definice výběrového průměru jako $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ lze rovnice upravit následujícím způsobem:

$$\begin{aligned}
S^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\mu)^2 - \frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1}^{n}(X_i-\mu)(X_j-\mu) \\
&\quad + \frac{1}{n^2(n-1)}\sum_{i=1}^{n}\left(\sum_{j=1}^{n}(X_j-\mu)\right)^2 \\
&= \frac{n-2}{n(n-1)}\sum_{i=1}^{n}(X_i-\mu)^2 - \frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}(X_i-\mu)(X_j-\mu) \\
&\quad + \frac{1}{n(n-1)}\sum_{j=1}^{n}(X_j-\mu)^2 + \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}(X_i-\mu)(X_j-\mu) \\
S^2 &= \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 - \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}(X_i-\mu)(X_j-\mu).
\end{aligned}$$

□

Věta 1.6. Nechť $n \geq 2$ a $X_1, \ldots, X_n$ jsou stejně rozdělené náhodné veličiny s rozptylem σ^2, pak lze střední hodnota odhadu S^2 vyjádřit explicitně podle vztahu

$$\mathsf{E}\, S^2 = \sigma^2 - \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\mathsf{cov}\,(X_i, X_j). \tag{1.5}$$

Důkaz. Označme μ střední hodnotu rozdělení ze kterého je daný závislý výběr. Pak z lemmatu 1.5 a z linearity operátoru E platí:

$$\begin{aligned}
\mathsf{E}\, S^2 &= \frac{1}{n}\sum_{i=1}^{n}\mathsf{E}\,(X_i-\mu)^2 - \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\mathsf{E}\,(X_i-\mu)(X_j-\mu) \\
\mathsf{E}\, S^2 &= \sigma^2 - \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\mathsf{cov}\,(X_i, X_j).
\end{aligned}$$

□

Z předcházející věty je patrné, že odhad rozptylu pomocí S^2 je nestranný pro náhodný výběr.

Následující úvahy budou mít za účel vyjádřit $\mathsf{var}\, S^2$ a posléze střední čtvercovou chybu pomocí vztahu (1.1). Mějme stejně rozdělené náhodné veličiny $X_1, \ldots, X_n$ ($n \geq 2$) z rozdělení o střední hodnotě μ a označme vektor náhodných veličin $\boldsymbol{X} = (X_1, \ldots, X_n)$. Vektorem $\boldsymbol{X} - \boldsymbol{\mu}$ budu rozumět vektor $(X_1 - \mu, \ldots, X_n - \mu)$. Vztah (1.4) v lemmatu 1.5 nám umožňuje následující vyjádření odhadu S^2. Zaveďme matici A předpisem

$$A = \begin{pmatrix} \frac{1}{n} & -\frac{1}{n(n-1)} & \cdots & a_n \\ -\frac{1}{n(n-1)} & \frac{1}{n} & \cdots & b_n \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{n(n-1)} & -\frac{1}{n(n-1)} & \cdots & \frac{1}{n} \end{pmatrix}. \tag{1.6}$$

Pak lze odhad S^2 vyjádřit jako

$$S^2 = (\boldsymbol{X} - \boldsymbol{\mu})' A (\boldsymbol{X} - \boldsymbol{\mu}). \tag{1.7}$$

Jednoduchým roznásobením vektorů a matice se lze přesvědčit, že vztah platí z rovnosti (1.4).

Věta 1.7. Nechť $X_1, \ldots, X_n$ jsou stejně rozdělené náhodné veličiny s rozptylem σ^2, a_{ij} jsou prvky libovolné matice A o rozměrech $n \times n$. Dále označím r_{ij} jako prvky varianční matice náhodného vektoru $(X_1, \ldots, X_n)$, tedy:

$$r_{ij} = \begin{cases} \sigma^2 & \text{pro } i = j, \\ \mathsf{cov}\,(X_i - X_j) & \text{pro } i \neq j, \end{cases}$$

pak lze střední hodnota náhodné veličiny Y, dané předpisem

$$Y = (\boldsymbol{X} - \boldsymbol{\mu})' A (\boldsymbol{X} - \boldsymbol{\mu}),$$

vyjádřit jako

$$\mathsf{E}\, Y = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} r_{ij}, \tag{1.8}$$

jsou-li navíc veličiny $X_1, \ldots, X_n$ z normálního rozdělení, pak lze rozptyl vyjádřit jako

$$\mathsf{var}\, Y = \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \sum_{l=1}^{n} a_{ij} a_{kl} (r_{ik} r_{jl} + r_{il} r_{jk}). \tag{1.9}$$

Důkaz.

$$\begin{aligned}
\mathsf{E}\,(Y) &= \mathsf{E}\,(\boldsymbol{X}-\boldsymbol{\mu})'A(\boldsymbol{X}-\boldsymbol{\mu}) \\
&= \mathsf{E}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}(X_i-\mu)(X_j-\mu) \\
\mathsf{E}\,(Y) &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}r_{ij}.
\end{aligned}$$

A nyní vypočteme rozptyl

$$\begin{aligned}
\mathsf{var}\,Y &= \mathsf{E}\,Y^2 - (\mathsf{E}\,Y)^2 \\
\mathsf{var}\,Y &= \mathsf{E}\,\left[(\boldsymbol{X}-\boldsymbol{\mu})'A(\boldsymbol{X}-\boldsymbol{\mu})\right] - (\mathsf{E}\,Y)^2 \\
&= \mathsf{E}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}(X_i-\mu)(X_j-\mu)a_{ij}(X_k-\mu)(X_l-\mu)a_{kl}- \\
&\quad -\left(\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}r_{ij}\right)^2.
\end{aligned}$$

Díky linearity operátoru E a Isserlitzovy formule z tvrzení 1.3 dále dostáváme pro $X_1,\ldots,X_n$ z normálního rozdělení,

$$\begin{aligned}
\mathsf{var}\,Y &= \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}(r_{ij}r_{kl}+r_{ik}r_{jl}+r_{il}r_{jk})a_{ij}a_{kl}- \\
&\quad -\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} a_{ij}a_{kl}r_{ij}r_{kl} \\
\mathsf{var}\,Y &= \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} a_{ij}a_{kl}(r_{ik}r_{jl}+r_{il}r_{jk}).
\end{aligned}$$

□

Věta 1.8. Nechť pro $n \geq 2$ jsou $X_1,\ldots,X_n$ stejně rozdělené náhodné veličiny z normálního rozdělení. Nechť r_{ij} jsou prvky varianční matice vektoru $(X_1,\ldots,X_n)$, a_{ij} jsou prvky matice A dané vztahem (1.6), pak lze rozptyl odhadu S^2 vyjádřit jako

$$\mathsf{var}\,S^2 = \frac{2}{(n-1)^2}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} r_{ij}^2 - \frac{2}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} r_{ij}r_{ik} + \frac{1}{n^2}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} r_{ij}\right)^2\right).$$

Důkaz. $X_1, \dots, X_n$ mají normální rozdělení a tedy z věty 1.7 dostáváme

$$
\begin{aligned}
\mathsf{var}\, S^2 &= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n a_{ij} a_{kl} (r_{ik} r_{jl} + r_{il} r_{jk}) \\
&= 2 \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n a_{ij} a_{kl} r_{ik} r_{jl} \\
&= 2 \left(\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n a_{ij} a_{kk} r_{ik} r_{jk} + \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{\substack{l=1 \\ l \neq k}}^n a_{ij} a_{kl} r_{ik} r_{jl} \right) \\
\frac{1}{2} \mathsf{var}\, S^2 &= \sum_{i=1}^n \sum_{k=1}^n a_{ii} a_{kk} r_{ik} r_{ik} + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \sum_{k=1}^n a_{ij} a_{kk} r_{ik} r_{jk} + \\
&+ \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \sum_{k=1}^n \sum_{\substack{l=1 \\ l \neq k}}^n a_{ij} a_{kl} r_{ik} r_{jl} + \sum_{i=1}^n \sum_{k=1}^n \sum_{\substack{l=1 \\ l \neq k}}^n a_{ii} a_{kl} r_{ik} r_{il},
\end{aligned}
$$

vyjádřením prvků matice A a rozepsáním jednotlivých sum dostáváme

$$
\begin{aligned}
\frac{1}{2} \mathsf{var}\, S^2 &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n (r_{ij})^2 - \frac{1}{n^2(n-1)} \left(\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n r_{ik} r_{jk} - \sum_{i=1}^n \sum_{j=1}^n r_{ij}^2 \right) + \\
&+ \frac{1}{n^2(n-1)^2} \left(\sum_{i=1}^n \sum_{j=1}^n r_{ij} \right)^2 - \frac{1}{n^2(n-1)^2} \sum_{i=1}^n \sum_{k=1}^n \sum_{\substack{l=1 \\ l \neq k}}^n r_{ik} r_{il} - \\
&- \frac{1}{n^2(n-1)^2} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \sum_{k=1}^n r_{ik} r_{jk} - \frac{1}{n^2(n-1)^2} \sum_{i=1}^n \sum_{k=1}^n r_{ik}^2 - \\
&- \frac{1}{n^2(n-1)} \sum_{i=1}^n \sum_{k=1}^n \sum_{\substack{l=1 \\ l \neq k}} r_{ik} r_{il} \\
\mathsf{var}\, S^2 &= \frac{2}{(n-1)^2} \left(\sum_{i=1}^n \sum_{j=1}^n r_{ij}^2 - \frac{2}{n} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n r_{ij} r_{ik} + \frac{1}{n^2} \left(\sum_{i=1}^n \sum_{j=1}^n r_{ij} \right)^2 \right).
\end{aligned}
$$

□

Poznámka. Tento vztah bude vhodnější pro výpočet rozptylu odhadu S^2 než vztah ve větě 1.7, protože se v něm vyskytuje pouze trojná suma.

Důsledek 1.9. Nechť pro $n \geq 2$ jsou $X_1, \ldots, X_n$ stejně rozdělené náhodné veličiny z normálního rozdělení, pak pro rozptyl odhadu S^2 platí

$$\operatorname{var} S^2 = \frac{2\sigma^4}{n-1}.$$

Důkaz. Vyjádřením ze vztahu v předcházející větě dostáváme

$$\begin{aligned}\operatorname{var} S^2 &= \frac{2}{(n-1)^2}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} r_{ij}^2 - \frac{2}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} r_{ij}r_{ik} + \frac{1}{n^2}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} r_{ij}\right)^2\right) \\ &= \frac{2}{(n-1)^2}\left(n\sigma^4 - 2\sigma^4 + \sigma^4\right) \\ \operatorname{var} S^2 &= \frac{2\sigma^4}{n-1}.\end{aligned}$$

□

Poznámka. Wilks uvádí ve [3] vztah pro rozptyl odhadu S^2 pro náhodný výběr o rozsahu $n \geq 2$ z rozdělení s konečným čtvrtým centrálním momentem μ_4 a rozptylem σ^2.

$$\frac{1}{n}\left(\mu_4 - \frac{n-3}{n-1}\sigma^4\right).$$

Snadno se lze přesvědčit, že z vyjádření pro normální rozdělení $\mu_4 = 3\sigma^2$ plyne vztah v předcházejícím důsledku.

Nyní již mohu vypočíst střední čtvercovou chybu odhadu S^2 pro $X_1, \ldots, X_n$ z normálního rozdělení, ze vztahu (1.1).

$$\operatorname{MSE} S^2 = \operatorname{var} S^2 - (\operatorname{E} S^2 - \sigma^2)^2. \tag{1.10}$$

Nyní mějme funkci $M(n)$ přirozených čísel, která je definována od nějákého n_0 přirozeného. Pro stejně rozdělené náhodné veličiny $X_1, \ldots, X_n$; $n \geq n_0$. Uvažujme odhady v následujícím tvaru:

$$S^2_{M(n)} = M(n)\sum_{i=1}^{n}(X_i - \overline{X})^2 \tag{1.11}$$

a tyto odhady můžeme vyjádřit pomocí S^2 jako

$$S^2_{M(n)} = (n-1)M(n)S^2.$$

Pro střední hodnotu a rozptyl tedy snadno z linearity operátoru E dostáváme

$$\begin{aligned}\mathsf{E}\, S^2_{M(n)} &= (n-1)M(n)\mathsf{E}\, S^2,\\ \mathsf{var}\, S^2_{M(n)} &= (n-1)^2 M^2(n)\mathsf{var}\, S^2.\end{aligned}$$

Často se za funkci $M(n)$ volí $\frac{1}{n}$ a $\frac{1}{n+1}$. Tyto odhady označím $\widehat{\sigma^2}$ a $\widetilde{S^2}$.

$$\begin{aligned}\widehat{\sigma^2} &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2\\ &= \frac{n-1}{n}S^2.\\ \widetilde{S^2} &= \frac{1}{n+1}\sum_{i=1}^{n}(X_i - \overline{X})^2\\ &= \frac{n-1}{n+1}S^2.\end{aligned}$$

Odhady $\widehat{\sigma^2}$ a $\widetilde{S^2}$ mají pro $X_1, \ldots, X_n$ nezávislé stejně rozdělené náhodné veličiny z normálního rozdělení menší střední čtvercovou chybu než odhad S^2, navíc odhad $\widetilde{S^2}$ má nejmenší čtvercovou chybu z odhadů tvaru (1.11), jak uvádí Anděl (2005) v [1].

1.2 Odhad rozptylu pomocí U_h

Pro zavedení dalšího odhadu provedeme následující motivaci. Pro $n \geq 2$ mějme stejně rozdělené náhodné veličiny $X_1, \ldots, X_n$. Předpokládejme, že

$$\forall i, j, k \in \{1, \ldots, n\}\ |i-j| < |i-k|\ : \mathsf{cov}\,(X_i, X_j) \geq \mathsf{cov}\,(X_i, X_k) \geq 0.$$

Tento předpoklad je přirozený, neboť očekáváme, že budeme-li mít posloupnost pozorování, pak vliv na dané pozorování bude u bližších pozorováních větší než u vzdálenějších. Cílem tedy bude vyřazení porovnání blízkých veličin v odhadu S^2 a tím zmenšení vychýlení odhadu S^2. Vyjdeme z vyjádření S^2 vztahem (1.4),

$$
\begin{aligned}
S^2 &= \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 - \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}(X_i-\mu)(X_j-\mu)\\
&= \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 - \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}(X_i-\mu)(X_j-\mu)\\
&= \frac{1}{2n(n-1)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\left((X_i-\mu)-(X_j-\mu)\right)^2\\
&= \frac{1}{2n(n-1)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}(X_i-X_j)^2\\
S^2 &= \frac{1}{n(n-1)}\sum_{i=2}^{n}\sum_{j=1}^{i-1}(X_i-X_j)^2. \qquad (1.12)
\end{aligned}
$$

A nyní mějme h přirozené menší než $n-1$, ze sumy na pravé straně můžeme vypustit členy $(X_i - X_j)$ takové, že $i - j \leq h$. Na základě této úvahy dostaneme následující odhad rozptylu.

Definice 1.10. Pro $n \geq 2$ mějme stejně rozdělené náhodné veličiny $X_1, \ldots, X_n$, h celé takové, že $0 \leq h \leq n-2$, nechť $A_h := \{(i,j) : i,j \in \mathbb{N}, i > j, i-j > h\}$ je podmnožina $\mathbb{N}^2$. Odhad rozptylu U_h zavedeme následujícím předpisem:

$$
U_h = \frac{1}{(n-h)(n-h-1)}\sum_{(i,j)\in A_h}(X_i-X_j)^2. \qquad (1.13)
$$

Poznámka. Lze snadno ukázat, že $(n-h)(n-h-1)$ je dvojnásobkem počtu prvků množiny A_h. Důvodem bude zachovaní nestrannosti u náhodného výběru, jak bude ukázáno za chvíli.

Snadno se lze přesvědčit srovnáním s rovností (1.12), že $U_0 = S^2$.

Věta 1.11. Nechť pro $n \geq 2$ jsou $X_1, \ldots, X_n$ stejně rozdělené náhodné veličiny z rozdělení s rozptylem σ^2, množina A_h je zavedena stejně jako v definici 1.10, pak lze střední hodnotu U_h vyjádřit explicitně podle následujícího vztahu.

$$\mathsf{E}\, U_h = \sigma^2 - \frac{2}{(n-h)(n-h-1)} \sum_{(i,j)\in A_h} \mathsf{cov}\,(X_i, X_j). \qquad (1.14)$$

Důkaz. Označme μ jako střední hodnotu rozdělení ze kterého je daný výběr, pak

$$\begin{aligned}
\mathsf{E}\, U_h &= \frac{1}{(n-1)(n-h-1)} \sum_{(i,j)\in A_h} \mathsf{E}\,(X_i - X_j)^2 \\
&= \frac{1}{(n-1)(n-h-1)} \sum_{(i,j)\in A_h} \mathsf{E}\,(X_i-\mu)^2 + \mathsf{E}\,(X_j-\mu)^2 - 2\mathsf{E}\,(X_i-\mu)(X_j-\mu) \\
&= \sigma^2 - \frac{2}{(n-1)(n-h-1)} \sum_{(i,j)\in A_h} \mathsf{cov}\,(X_i, X_j).
\end{aligned}$$

□

Z předchozí věty je patrné, že odhad rozptylu pomocí U_h je nestranný pro náhodný výběr. Navíc srovnáním s (1.5) za předpokladu, že všechny kovariance jsou kladné, platí, že vychýlení odhadu U_h je menší nebo rovno než vychýlení odhadu S^2.

Nyní provedeme podobnou konstrukci pro určení $\mathsf{var}\, U_h$ a posléze $\mathsf{MSE}\,(U_h)$, jako u S^2. Mějme stejně rozdělené náhodné veličiny $X_1, \ldots, X_n$ $(n \geq 2)$ z rozdělení o střední hodnotě μ, označme vektor náhodných veličin $\boldsymbol{X} = (X_1, \ldots, X_n)$, vektorem $\boldsymbol{X} - \boldsymbol{\mu}$ budu opět rozumět vektor $(X_1 - \mu, \ldots, X_n - \mu)$. Hledáme matici B, takovou, že bude platit rovnice

$$U_h = (\boldsymbol{X} - \boldsymbol{\mu})' B (\boldsymbol{X} - \boldsymbol{\mu}). \qquad (1.15)$$

Vyjdu z definice odhadu U_h,

$$\begin{aligned}
U_h &= \frac{1}{(n-h)(n-h-1)} \sum_{(i,j)\in A_h} (X_i - X_j)^2 \\
&= \frac{1}{(n-h)(n-h-1)} \sum_{i=h+2}^{n} \sum_{j=1}^{i-h-1} \big((X_i-\mu)^2 + (X_j-\mu)^2\big) + \\
&\quad + \frac{2}{(n-h)(n-h-1)} \sum_{i=h+2}^{n} \sum_{j=1}^{i-h-1} (X_i-\mu)(X_j-\mu)
\end{aligned}$$

$$U_h = \frac{1}{(n-h)(n-h-1)} \sum_{i=h+2}^{n} (i-h-1)(X_i-\mu)^2$$
$$+ \frac{1}{(n-h)(n-h-1)} \sum_{i=h+2}^{n} \sum_{j=1}^{i-h-1} (X_j-\mu)^2$$
$$+ \frac{2}{(n-h)(n-h-1)} \sum_{i=h+2}^{n} \sum_{j=1}^{i-h-1} (X_i-\mu)(X_j-\mu)$$

$$U_h = \frac{1}{(n-h)(n-h-1)} \sum_{i=h+2}^{n} (i-h-1)(X_i-\mu)^2$$
$$+ \frac{1}{(n-h)(n-h-1)} \sum_{i=1}^{n-h-1} (n-h-i)(X_i-\mu)^2$$
$$+ \frac{2}{(n-h)(n-h-1)} \sum_{i=h+2}^{n} \sum_{j=1}^{i-h-1} (X_i-\mu)(X_j-\mu).$$

Na základě tohoto vyjádření již mohu sestavit matici B. Nechť b_{ij} jsou prvky matice B, pak pro $i \neq j$

$$b_{ij} = \begin{cases} 0 & \text{pro } i,j \text{ takové, že } i-j \leq h, \\ \frac{2}{(n-h)(n-h-1)} & \text{pro } i,j \text{ takové, že } i-j > h \end{cases}$$

a pro $i = j$, je třeba rozlišit dva případy v závislosti na volbě h.

Pro h takové, že $\frac{n-3}{2} \geq h$

$$b_{ii} = \begin{cases} \frac{n-h-i}{(n-h)(n-h-1)} & \text{pro } i \text{ takové, že } i < h+1, \\ \frac{n-2h-1}{(n-h)(n-h-1)} & \text{pro } i \text{ takové, že } i \in [h+2, n-h-1], \\ \frac{i-h-1}{(n-h)(n-h-1)} & \text{pro } i \text{ takové, že } i > n-h-1. \end{cases}$$

Pro h takové, že $\frac{n-3}{2} < h$

$$b_{ii} = \begin{cases} \frac{n-h-i}{(n-h)(n-h-1)} & \text{pro } i \text{ takové, že } i \leq n-h-1, \\ 0 & \text{pro } i \text{ takové, že } n-h-1 < i < h+2, \\ \frac{i-h-1}{(n-h)(n-h-1)} & \text{pro } i \text{ takové, že } i \geq h+2. \end{cases}$$

Střední hodnotu odhadu U_h ponechám ve tvaru uvedeném v (1.14). Rozptyl U_h ze stejně rozdělených náhodných veličin $X_1, \dots, X_n$ z normálního rozdělení vypočtu pomocí věty 1.7.

$$\operatorname{var} U_h = \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \sum_{l=1}^{n} b_{ij} b_{kl} (r_{ik} r_{jl} + r_{il} r_{jk}). \tag{1.16}$$

A střední čtvercovou chybu vyjádřím pomocí (1.1).

$$\mathsf{MSE}\,(U_h) = \operatorname{var} U_h + (\mathsf{E}\, U_h - \theta)^2. \tag{1.17}$$

1.3 Odhad rozptylu pomocí odhadu T_m a $\widehat{T}_m$

Následující odhad bude odvozen pro *slabě stacionární posloupnosti* náhodných veličin.

Definice 1.12. Nechť $X_1, \dots, X_n$ jsou náhodné veličiny o střední hodnotě μ, pak posloupnost náhodných veličin $X_1, \dots, X_n$ nazývám slabě stacionární, jestliže existuje funkce $r(k)$ pro $k \in \{-n+1, \dots, 0, \dots, n-1\}$ taková, že

$$\forall k \in \{-n+1, \dots, 0, \dots, n-1\}\ \forall i, j \in N; i, j \leq n; i - j = k : r(k) = \operatorname{cov}(X_i, X_j).$$

Takovou funkci budu nazývat autokovarianční funkce.

Poznámka. Z vlastností kovariance snadno plyne, že $r(-k) = r(k)$, navíc platí $\operatorname{var} X_i = r(0)$ a tedy náhodné veličiny ve stacionární posloupnosti mají stejný rozptyl.

Předpokládejme tedy, že máme takovou autokovarianční funkci a vyjdeme ze vztahu (1.5), tj.:

$$\begin{aligned}
\mathsf{E}\, S^2 &= \sigma^2 - \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \operatorname{cov}(X_i, X_j) \\
&= \sigma^2 - \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} r(i-j) \\
&= \sigma^2 - \frac{1}{n(n-1)} \sum_{k=-n+1}^{n-1} (n - |k|) r(k).
\end{aligned}$$

A z předcházející poznámky plyne

$$= \sigma^2 - \frac{2}{n(n-1)} \sum_{k=1}^{n-1} (n - |k|) r(k).$$

Pokusíme se odhadnout některé hodnoty $r(k)$ následujícím odhadem $\widehat{r}(k)$:

$$\widehat{r}(k) = \frac{1}{n-k} \sum_{i=1}^{n-k} (X_i - \overline{X})(X_{i+k} - \overline{X}). \tag{1.18}$$

Poznámka. Tento odhad autokovarianční funkce blíže popsal Wilks v [3].

Definice 1.13. Nechť $n \geq 2$ a $X_1, \ldots, X_n$ je slabě stacionární posloupnost stejně rozdělených náhodných veličin, nechť m je přirozené takové, že $m < n$, nechť S^2 je výběrový rozptyl definován vztahem (1.3) a nechť $\widehat{r}(k)$ je definováno vztahem (1.18), pak odhad T_m definuji předpisem:

$$T_m = S^2 + \frac{2}{n(n-1)} \sum_{k=1}^{m} (n-k) \widehat{r}(k). \tag{1.19}$$

Nyní, stejně jako v předcházejících případech, uvažujme náhodný vektor $\boldsymbol{X} = (X_1, \ldots, X_n)$ a hledejme matici C takovou, že bude platit

$$T_m = (\boldsymbol{X} - \boldsymbol{\mu})' C (\boldsymbol{X} - \boldsymbol{\mu}).$$

Nejprve upravíme odhad T_m.

$$\begin{aligned}
T_m &= S^2 + \frac{2}{n(n-1)} \sum_{k=1}^{m} \sum_{i=1}^{n-k} (X_i - \overline{X})(X_{i+k} - \overline{X}) \\
&= S^2 + \frac{2}{n(n-1)} \sum_{k=1}^{m} \sum_{i=1}^{n-k} \big((X_i - \mu) - (\overline{X} - \mu)((X_i - \mu) - (\overline{X} - \mu)\big) \\
&= S^2 + \frac{2}{n(n-1)} \sum_{k=1}^{m} \sum_{i=1}^{n-k} (X_i - \mu)(X_{i+k} - \mu) - (X_i - \mu)(\overline{X} - \mu) - \\
&\qquad\qquad - (\overline{X} - \mu)(X_{i+k}) + (\overline{X} - \mu)^2.
\end{aligned}$$

$$\begin{aligned}
T_m = S^2 + \frac{2}{n(n-1)} \sum_{k=1}^{m} \sum_{i=1}^{n-k} &(X_i - \mu)(X_{i+k} - \mu) - \frac{1}{n}(X_i - \mu) \left(\sum_{j=1}^{n} (X_j - \mu) \right) - \\
&- \frac{1}{n}(X_{i+k} - \mu) \left(\sum_{j=1}^{n} (X_j - \mu) \right) + (\overline{X} - \mu)^2.
\end{aligned}$$

A budeme hledat matice $P_{i,k}, Q_{i,k}, R_{i,k}, S_{i,k}$, takové, že

$$\begin{aligned}
(\boldsymbol{X} - \boldsymbol{\mu})' P_{i,k} (\boldsymbol{X} - \boldsymbol{\mu}) &= (X_i - \mu)(X_{i+k} - \mu), \\
(\boldsymbol{X} - \boldsymbol{\mu})' Q_{i,k} (\boldsymbol{X} - \boldsymbol{\mu}) &= -\frac{1}{n} \sum_{j=1}^{n} (X_i - \mu)(X_j - \mu), \\
(\boldsymbol{X} - \boldsymbol{\mu})' R_{i,k} (\boldsymbol{X} - \boldsymbol{\mu}) &= -\frac{1}{n} \sum_{j=1}^{n} (X_{i+k} - \mu)(X_j - \mu), \\
(\boldsymbol{X} - \boldsymbol{\mu})' S_{i,k} (\boldsymbol{X} - \boldsymbol{\mu}) &= \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} (X_i - \mu)(X_j - \mu).
\end{aligned}$$

Matice $P_{i,k}$ bude mít pouze na pozici $[i, i+k]$ hodnotu 1 a všude jinde bude mít hodnotu 0. A tedy nechť p_{ab} jsou prvky matice $\sum_{k=1}^{m} \sum_{i=1}^{n-k} P_{i,k}$, pak

$$p_{ab} = \begin{cases} 1 & \text{pro } 1 \leq b - a \leq m, \\ 0 & \text{jinak.} \end{cases}$$

Matice $Q_{i,k}$ bude mít v i-tém řádku hodnotu $-\frac{1}{n}$ a jinde 0 a tedy nechť q_{ab} jsou prvky matice $\sum_{k=1}^{m} \sum_{i=1}^{n-k} Q_{i,k}$, pak

$$q_{ab} = \begin{cases} -\frac{m}{n} & \text{pro } a \leq n-m, \\ -\frac{n-a}{n} & \text{pro } a \geq n-m+1. \end{cases}$$

Matice $R_{i,k}$ bude mít v $(i+k)$-tém řádku hodnotu $-\frac{1}{n}$ a jinde 0 a tedy nechť r_{ab} jsou prvky matice $\sum_{k=1}^{m}\sum_{i=1}^{n-k} R_{i,k}$, pak

$$r_{ab} = \begin{cases} -\frac{m}{n} & \text{pro } a \geq m+1, \\ -\frac{a-1}{n} & \text{pro } a \leq m. \end{cases}$$

Matice $S_{i,k}$ bude mít na všech pozicích hodnotu $\frac{1}{n^2}$ a tedy nechť s_{ab} jsou prvky matice $\sum_{k=1}^{m}\sum_{i=1}^{n-k} S_{i,k}$, pak

$$s_{ab} = \frac{m(2n-m-1)}{2n^2}.$$

Pro zjednodušení výpočtu ještě vyjadřím prvky matice

$$T := \sum_{k=1}^{m}\sum_{i=1}^{n-k}(Q_{i,k} + R_{i,k}).$$

Tyto prvky označím t_{ab}. Pro jejich vyjádření je třeba rozdělit situaci na dva případy v závislosti na m.

Pro $m \leq \frac{n-1}{2}$ dostáváme

$$t_{ab} = \begin{cases} -\frac{m+a-1}{n} & \text{pro } a < m, \\ -\frac{2m}{n} & \text{pro } m \leq a \leq n-m, \\ -\frac{n+m-a}{n} & \text{pro } n-m < a. \end{cases}$$

Pro $m > n-m$ dostáváme

$$t_{ab} = \begin{cases} -\frac{m+a-1}{n} & \text{pro } a \leq n-m, \\ -\frac{n-1}{n} & \text{pro } n-1 < a < m+1, \\ -\frac{m+n-a}{n} & \text{pro } a \geq m+1. \end{cases}$$

Poznámka. Sčítat ostatní matice nemá význam, neboť výpočet nijak neulehčí.

Mějme matici A zavedenou pro S^2 pomocí (1.6). Potom lze matice C vyjádřit součtem matic A, P, T, S jako

$$C = A + \frac{2}{n(n-1)}(P + T + S). \tag{1.20}$$

Následující věta shrnuje výpočet střední hodnoty a rozptylu pro odhad T_m.

Věta 1.14. Nechť pro $n \geq 2$ je $X_1, \dots, X_n$ slabě stacionární posloupnost stejně rozdělených náhodných veličin, c_{ij} jsou prvky matice C z rovnice (1.20), r_{ij} prvky varianční matice vektoru $(X_1, \dots, X_n)$, pak střední hodnota odhadu T_m z definice 1.13 lze vyjádřit jako

$$ET_m = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} r_{ij}.$$

Jsou-li navíc $X_1, \dots, X_n$ náhodné veličiny s normálním rozdělením, pak

$$\operatorname{\mathsf{var}} T_m = \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \sum_{l=1}^{n} c_{ij} c_{kl} (r_{ik} r_{jl} + r_{il} r_{jk}).$$

Důkaz. Vyjádření střední hodnoty plyne aplikací vztahu (1.8) ve větě 1.7 na odhad T_m. Pro určení rozptylu odhadu T_m z $X_1, \dots, X_n$ s normálním rozdělením, aplikujeme vztah (1.9). □

Výpočet střední čtvercové chyby pro $X_1, \dots, X_n$ z normálního rozdělení provedeme stejně jako u předcházejících odhadů ze vztahu (1.1), tj.:

$$\operatorname{\mathsf{MSE}}(T_m) = \operatorname{\mathsf{var}} T_m + (\mathsf{E}\, T_m - \theta)^2. \tag{1.21}$$

Nyní se podívejme, jak je to s nestranností pro $X_1, \dots, X_n$ nezávislé stejně rozdělené veličiny.

Věta 1.15. Nechť pro $n \geq 2$ je $X_1, \dots, X_n$ slabě stacionární posloupnost náhodných veličin s rozptylem σ^2, pak pro střední hodnotu odhadu T_m platí

$$\mathsf{E}\, T_m = \sigma^2 \left(1 + \frac{m^2 - 2mn + m}{n^2(n-1)}\right).$$

Důkaz. Z nezávislosti dostáváme pro prvky r_{ij} varianční matice $(X_1, \dots, X_n)$

$$r_{ij} = \begin{cases} \sigma^2 & \text{pro } i = j, \\ 0 & \text{pro } i \neq j. \end{cases}$$

A tedy z maticového vyjádření střední hodnoty T_m (1.14) dostáváme

$$\mathsf{E}\, T_m = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} r_{ij} = \sigma^2 \sum_{i=1}^{n} c_{ii}.$$

Z (1.20) víme, že $c_{ii} = a_{ii} + \frac{2}{n(n-1)}(p_{ii} + t_{ii} + s_{ii})$, snadno se můžeme přesvědčit, že

$$\begin{aligned}
\sum_{i=1}^{n} p_{ii} &= 0 \\
\sum_{i=1}^{n} a_{ii} &= \sigma^2 \\
\sum_{i=1}^{n} s_{ii} &= \frac{m(2n-m-1)}{2n} \\
\sum_{i=1}^{n} t_{ii} &= \frac{2m(m-n) + (1-m)m}{n}
\end{aligned}$$

a tedy pro střední hodnotu po úpravách dostáváme

$$\mathsf{E}\, T_m = \sigma^2 \sum_{i=1}^{n} c_{ii} = \sigma^2 \left(1 + \frac{m^2 - 2mn + m}{n^2(n-1)} \right).$$

□

Proto vhodně odhad přenásobíme, abychom v případě nezávislých $X_1, \ldots, X_n$ stejně rozdělených náhodných veličin získali odhad nestranný, tento odhad označím $\widehat{T}_m$.

Definice 1.16. Nechť pro $n \geq 2$ je $X_1, \ldots, X_n$ slabě stacionární posloupnost, nechť m je přirozené takové, že $m < n$, nechť T_m je odhad rozptylu definován v 1.13, pak odhad $\widehat{T}_m$ definuji předpisem:

$$\widehat{T}_m = \frac{n^2(n-1)}{m^2 - 2mn + m + n^2(n-1)} T_m.$$

Tento odhad je již nestranný pro $X_1, \dots, X_n$ nezávislé stejně rozdělené veličiny. Momenty určím snadno z momentů T_m.

$$\mathsf{E}\,\widehat{T}_m = \frac{n^2(n-1)}{m^2 - 2mn + m + n^2(n-1)}\mathsf{E}\,T_m,$$
$$\mathsf{var}\,\widehat{T}_m = \left(\frac{n^2(n-1)}{m^2 - 2mn + m + n^2(n-1)}\right)^2 \mathsf{var}\,T_m.$$

A střední čtvercovou chybu pomocí vztahu (1.1).

$$\mathsf{MSE}\,(\widehat{T}_m) = \mathsf{var}\,\widehat{T}_m + (\mathsf{E}\,\widehat{T}_m - \sigma^2)^2.$$

Kapitola 2

Odhady rozptylů pro konkrétní případy

Cílem této kapitoly bude zavést několik základních tvarů autokovarianční funkce, vyjádřit střední hodnotu odhadů S^2 a U_h pro exponenciální tvar autokovariančních funkcí. A provést výpočty středních hodnot a středních čtvercových chyb odhadů pro některé konkrétní případy.

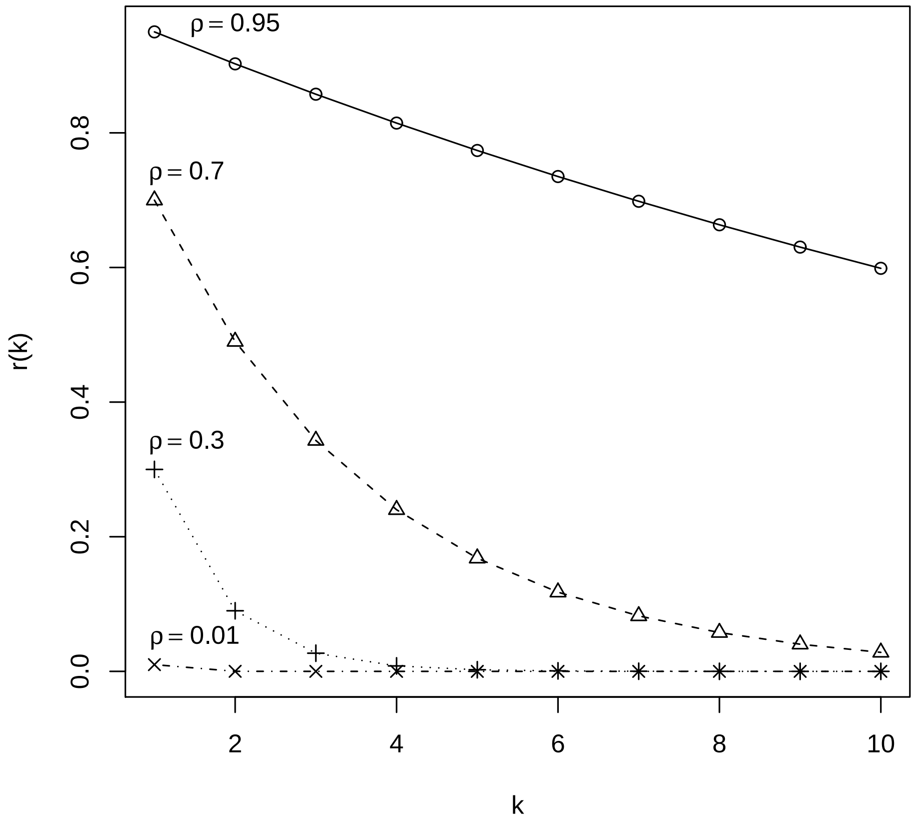
2.1 Autokovarianční funkce

Mějme $X_1, \ldots, X_n$ stejně rozdělené náhodné veličiny s roztylem σ^2 a podívejme se nyní na některé tvary autokovarianční funkce.

Nechť $|\rho| < 1$ je nenulový reálný parametr, pak autokovarianční funkci budu nazývat *exponenciální*, jestliže je dána předpisem

$$r(k) = \sigma^2 \rho^k.$$

Tato autokovarianční funkce bude mít pro $\sigma^2 = 1$ v závislosti na volbě ρ následující tvar podle k.



Podívejme se nyní na tvar střední hodnoty odhadů S^2 a U_h.

Věta 2.1. Nechť pro $n \geq 2$ jsou $X_1, \ldots, X_n$ stejně rozdělené náhodné veličiny s exponenciální autokovarianční funkcí, pak střední hodnoty odhadů rozptylu U_h a S^2 lze vyjádřit jako

$$\mathsf{E}\, S^2 = \sigma^2 \left(1 - \frac{2\rho(\rho^n - 2\rho + 1)}{n(n-1)(\rho-1)^2}\right),$$
$$\mathsf{E}\, U_h = \sigma^2 \left(1 - \frac{2\rho^{h+1}(\rho^{n-h} - 2\rho + 1)}{(n-h)(n-h-1)(\rho-1)^2}\right).$$

Důkaz. Nejprve vyjádřím $\sum_{i=h+2}^{n} \sum_{j=1}^{i-h-1} \mathsf{cov}\,(X_i, X_j)$ jako

$$\begin{aligned}
\sum_{i=h+2}^{n} \sum_{j=1}^{i-h-1} \mathsf{cov}\,(X_i, X_j) &= \sigma^2 \sum_{i=h+2}^{n} \sum_{j=1}^{i-h-1} \rho^{|i-j|} \\
&= \sigma^2 \sum_{i=h+2}^{n} \rho^{h+1} \frac{\rho^{i-h-1} - 1}{\rho - 1} \\
&= \sigma^2 \left(\frac{\rho^{h+1}(\rho^{n-h} - 2\rho + 1)}{(\rho - 1)^2} \right).
\end{aligned}$$

A tedy z (1.14) dostáváme pro $\mathsf{E}\, U_h$

$$\mathsf{E}\, U_h = \sigma^2 \left(1 - \frac{2\rho^{h+1}(\rho^{n-h} - 2\rho + 1)}{n(n-1)(\rho - 1)^2} \right).$$

A ze vztahu $S^2 = U_0$ dostáváme snadnou úpravou

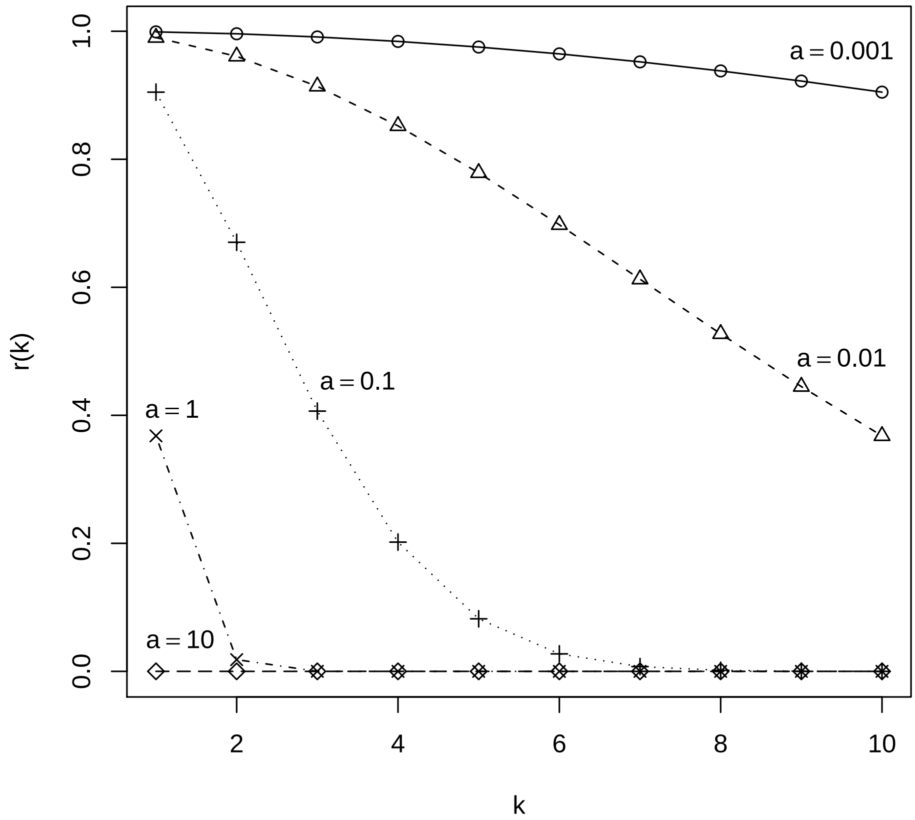
$$\mathsf{E}\, S^2 = \sigma^2 \left(1 - \frac{2\rho(\rho^{n} - 2\rho + 1)}{(n-h)(n-h-1)(\rho - 1)^2} \right).$$

□

Další autokovarianční funkci budu nazývat *gaussovská* a bude dána vztahem

$$r(k) = \sigma^2 e^{-a|k|^2}, \qquad \text{kde } a > 0 \text{ je reálný parametr.} \tag{2.1}$$

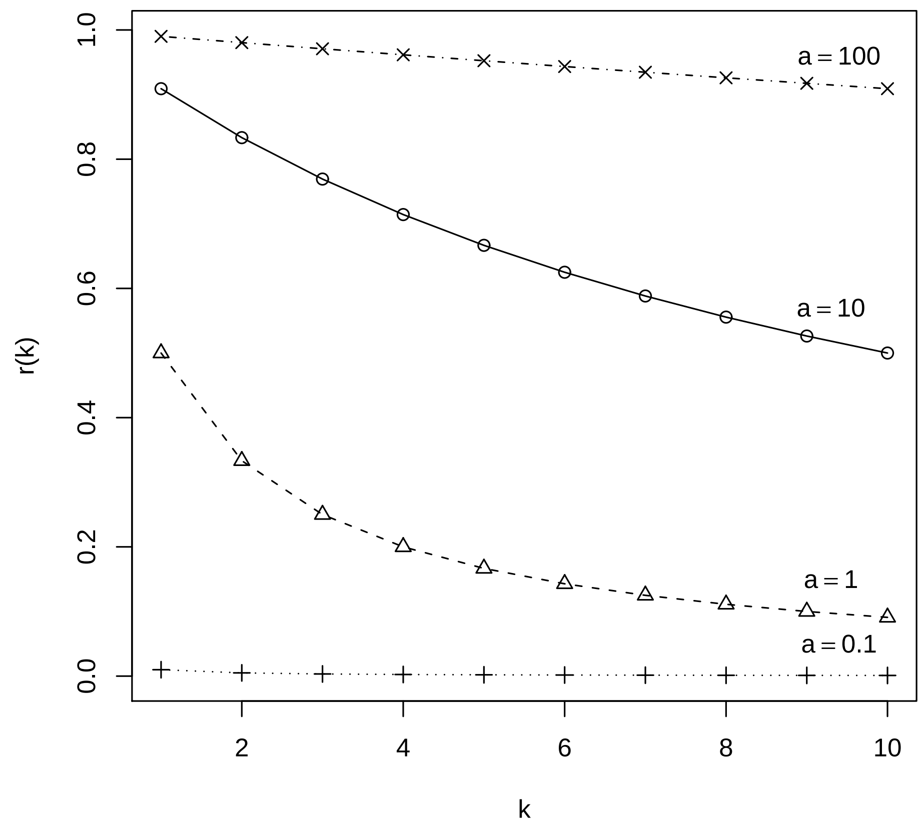
Tato autokovarianční funkce bude mít pro $\sigma^2 = 1$ v závislosti na volbě a následující tvar podle k.



Následující autokovarianční funkci budu nazývat *hyperbolická* a bude dána předpisem

$$r(k) = \sigma^2 \frac{1}{1 + \frac{|k|}{a}}, \qquad \text{kde } a > 0 \text{ je reálný parametr.} \tag{2.2}$$

Tato autokovarianční funkce bude mít pro $\sigma^2 = 1$ v závislosti na volbě a následující tvar podle k.

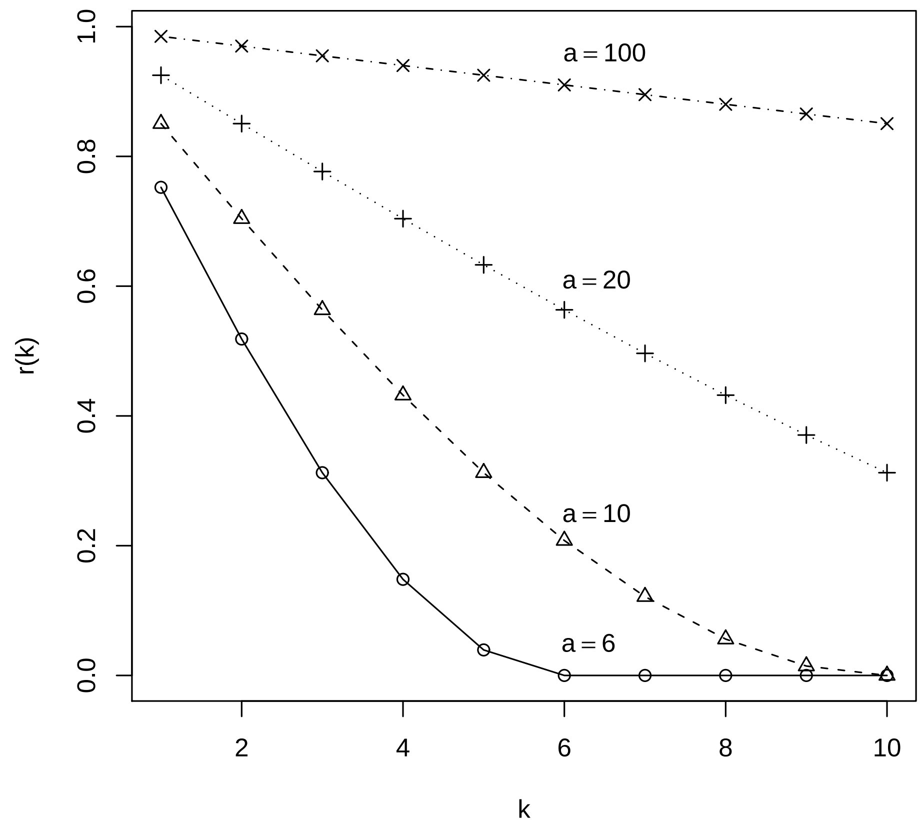


Nechť $a \geq 0$ je reálný parametr. Další autokovarianční funkci budu nazývat *sférická* a bude dána předpisem

$$r(k) \begin{cases} \sigma^2 \left(1 - \frac{3|k|}{2a} + \frac{|k|^3}{a^3}\right) & \text{pro } |k| \leq a, \\ 0 & \text{jinak.} \end{cases} \tag{2.3}$$

Poznámka. Význam omezení na $|k| \leq a$ je zachování monotonie pro k kladná, resp. záporná.

Tato autokovarianční funkce bude mít pro $\sigma^2 = 1$ v závislosti na volbě a následující tvar podle k.



2.2 Odhady rozptylu pro pozorování z normálního rozdělení

Nyní budeme předpokládat, že máme pro $n \geq 2$ slabě stacionární posloupnost náhodných veličin $X_1, \ldots, X_n$ ze stejného normálního rozdělením a porovnáme výsledky výpočtů teoretických středních hodnot a středních čtvercových chyb pro různé volby autokovariančních funkcí a jejich parametrů. Výpočty jsem provedl v programu R.2.2.0.

Zvolme parametry následovně

$$\begin{aligned} \sigma^2 &= 1 & &\text{rozptyl náhodných veličin} \\ n &= 20 & &\text{počet náhodných veličin.} \end{aligned}$$

Pro volbu exponenciální autokovarianční funkce s parametrem ρ, jsem získal následující výsledky,

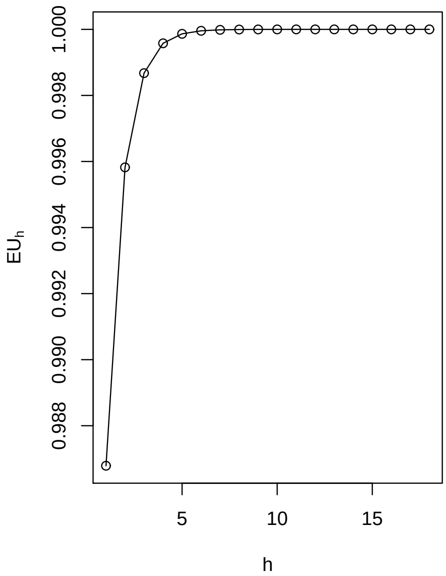
Pro volbu $\rho = 0.30$

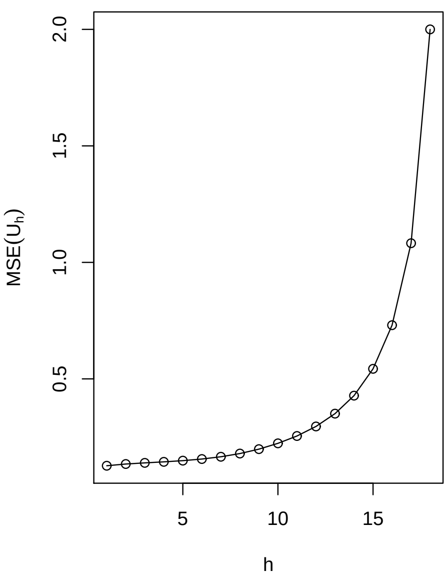
$\rho = 0.30$	S^2	$\widehat{\sigma^2}$	U_1	T_2	$\widehat{T_2}$
Střední hodnota	0.958	0.910	0.987	0.979	0.988
Střední čtvercová chyba	0.115	0.110	0.127	0.128	0.130

Poznámka. Volby dolních indexů u odhadů U_h, T_m a $\widehat{T}$ jsou pouze orientační, tyto odhady budou podrobněji popsány dále.

Nyní porovnáme střední hodnoty a střední čtvercové chyby pro různé hodnoty h v odhadu U_h pro volbu $\rho = 0.3$.

h	1	2	3	4	10	15	18
$\mathsf{E}\, U_h$	0.9867	0.9958	0.9986	0.9995	1.0000	1.0000	1.0000
$\mathsf{MSE}\, U_h$	0.1271	0.1345	0.1394	0.1438	0.2233	0.5432	2.0000

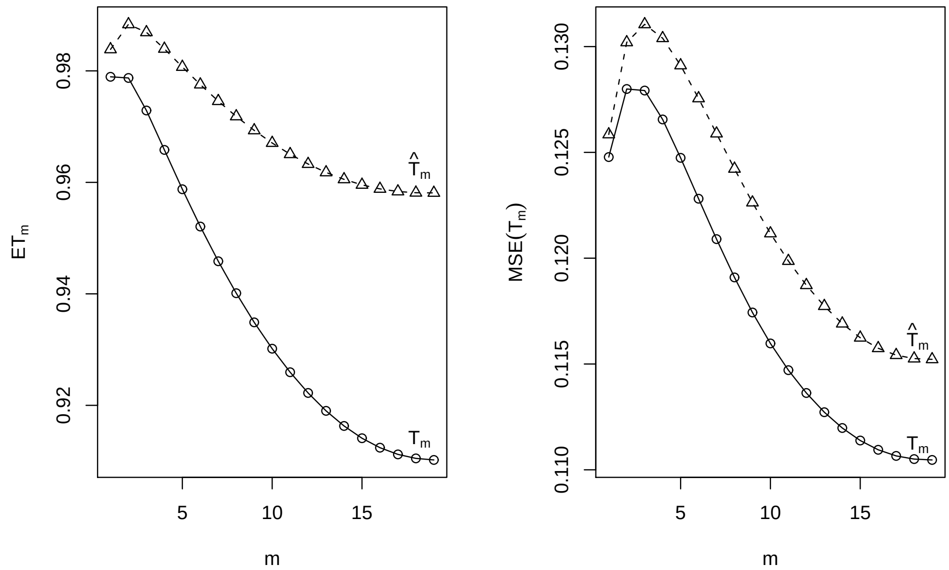




Z grafu střední čtvercové chyby je patrné, že volba velkého h je zcela nevhodná.

Nyní porovnáme střední hodnoty a střední čtvercové chyby pro různé hodnoty m v odhadu T_m a $\widehat{T}_m$ pro volbu $\rho = 0.3$.

m	1	2	4	5	6	15	19
$\mathsf{E}\, T_m$	0.9789	0.9787	0.9658	0.9587	0.9520	0.9141	0.9102
$\mathsf{E}\, \widehat{T}_m$	0.9838	0.9883	0.9839	0.9807	0.9775	0.9595	0.9581
$\mathsf{MSE}\, T_m$	0.1247	0.1279	0.1265	0.1247	0.1228	0.1113	0.1104
$\mathsf{MSE}\, \widehat{T}_m$	0.1258	0.1301	0.1303	0.1291	0.1275	0.1162	0.1152



Nejmenší střední čtvercové chyby pro volbu $\rho = 0.30$ vykazuje odhad $\widehat{\sigma^2}$. Nejmenší vychýlení pak nabízí odhad U_{18}, ten je však nevhodný kvůli nepřiměřeně velké střední čtvercové chybě, proto bych doporučil použít odhad U_1, nebo $\widehat{T_2}$, je-li vyžadováno malé vychýlení.

Výsledky pro malou hodnotu ρ. Kovariance budou tedy velmi blízko nule.

$\rho = 0.01$	S^2	$\widehat{\sigma^2}$	U_2	T_5	$\widehat{T_5}$
Střední hodnota	0.999	0.949	1.000	0.977	1.000
Střední čtvercová chyba	0.105	0.097	0.107	0.102	0.107

Výsledky pro vysokou hodnotu ρ. Kovariance blízké hodnotě rozptylu, tj. korelační koeficienty budou blízké jedné.

$\rho = 0.95$	S^2	$\widehat{\sigma^2}$	U_2	T_5	$\widehat{T_5}$
Střední hodnota	0.283	0.269	0.334	0.334	0.342
Střední čtvercová chyba	0.565	0.581	0.521	0.528	0.522

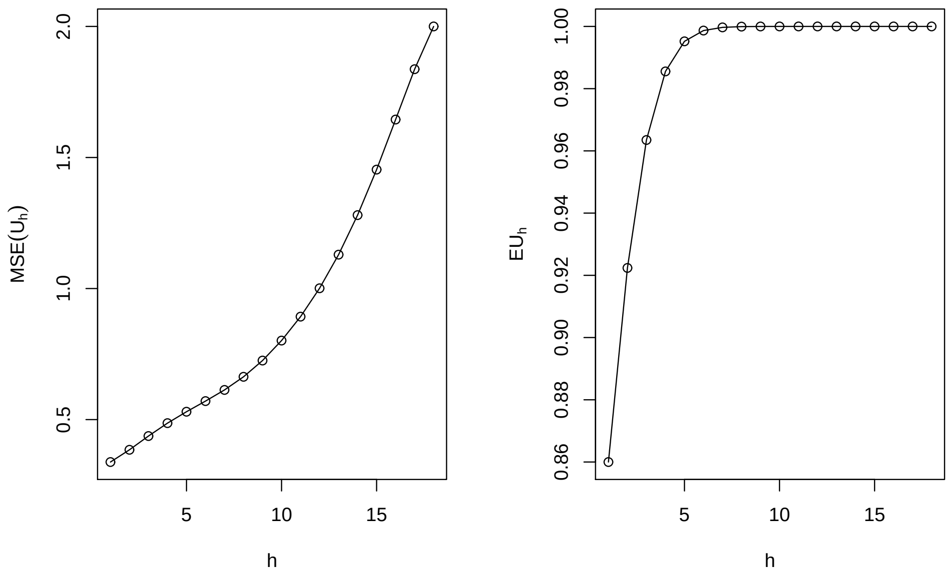
Za povšimnutí stojí výsledek u odhadu $\widehat{\sigma^2}$, ukazuje se, že při vysokých závislostech v pozorování se stává tento odhad jak po stránce střední hodnoty, tak po stránce střední čtvercové chyby nejhorším, ale ani ostatní odhady nejsou příliš dobré.

Podívejme se nyní na výsledky při volbě gaussovské autokovarianční funkce s parametrem $a = 0.1$.

$a = 0.1$	S^2	$\widehat{\sigma^2}$	U_2	T_3	$\widehat{T}_3$
Střední hodnota	0.784	0.743	0.922	0.897	0.910
Střední čtvercová chyba	0.308	0.301	0.385	0.382	0.391

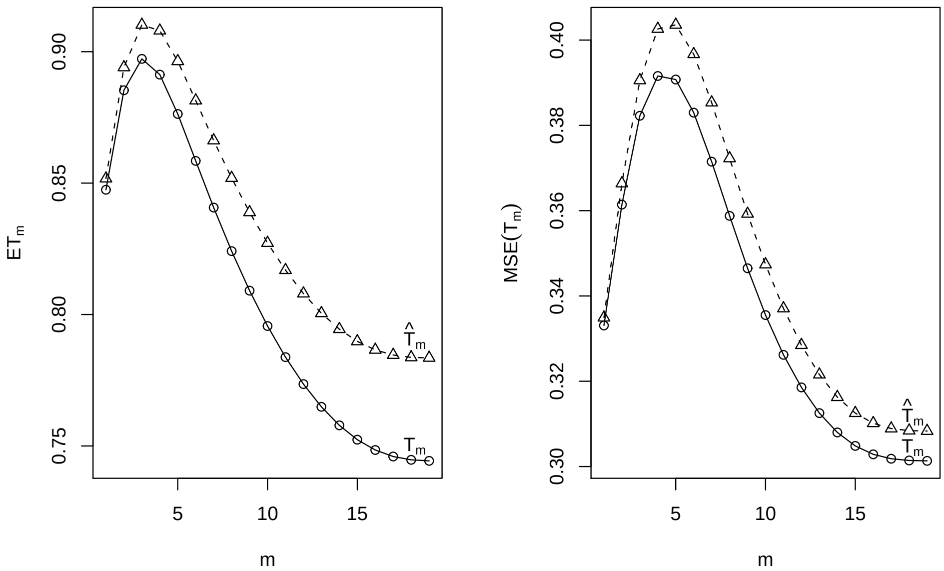
Speciálně pro odhady U_h

h	1	2	3	4	10	15	18
$\mathsf{E}\, U_h$	0.8599	0.9223	0.9634	0.9855	0.9999	1.0000	1.0000
$\mathsf{MSE}\, U_h$	0.3380	0.3847	0.4372	0.4864	0.8015	1.4537	2.0000



A pro odhady T_m a $\widetilde{T}_m$

m	1	2	3	4	5	15	19
$\mathsf{E}\, T_m$	0.8474	0.8853	0.8972	0.8912	0.8763	0.7523	0.7443
$\mathsf{E}\, \widehat{T}_m$	0.8517	0.8940	0.9102	0.9080	0.8963	0.7897	0.7835
$\mathsf{MSE}\, T_m$	0.3330	0.3614	0.3822	0.3915	0.3907	0.3048	0.3013
$\mathsf{MSE}\, \widehat{T}_m$	0.3349	0.3663	0.3905	0.4026	0.4035	0.3125	0.3083



Stejně jako pro exponenciální autokovarianční funkci s parametrem $\rho = 0.3$ dostáváme nejmenší čtvercovou chybu při použití $\widehat{\sigma^2}$, ale za cenu nepříjemného vychýlení. Naproti tomu nejmenšího vychýlení s přijatelnou střední čtvercovou chybou dosahují odhady U_2 a $\widehat{T}_3$.
Výsledky pro volbu malého parametru a, kdy budou náhodné veličiny mít kovariance blízké 0, a pro volbu velkého parametru a, kdy budou kovariance blízké hodnotě rozptylu, neuvádím neboť jsou téměř totožné s výsledky pro exponenciální autokovarianční funkci s příslušným parametrem ρ.

Literatura

[1] Anděl J. (2005): Základy matematické statistiky, Matfyzpress, Praha.

[2] Isserlitz. L. : On certain probable errors and correlation coefficients of multiple frequency distributions with skew regression, *Biometrika* **11**, 185–190.

[3] Wilks S.S. (1962): Mathematical Statistics. Wiley, New York.