MATEMATICKO-FYZIKÁLNÍ FAKULTA
Univerzita Karlova

BAKALÁŘSKÁ PRÁCE

Martin Farda

Intervaly spolehlivosti pro korelační koeficient

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: doc. Mgr. Michal Kulich, Ph.D.
Studijní program: Matematika
Studijní obor: Finanční matematika

Praha 2021

Prohlašuji, že jsem tuto bakalářskou práci vypracoval(a) samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů. Tato práce nebyla využita k získání jiného nebo stejného titulu.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V .... dne ....

Podpis autora

Rád bych tímto poděkoval panu docentu Kulichovi za jeho vedení, věnovaný čas, užitečné rady a doporučení k bakalářské práci.

Název práce: Intervaly spolehlivosti pro korelační koeficient

Autor: Martin Farda

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: doc. Mgr. Michal Kulich, Ph.D., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: Cílem práce je podrobně představit metody používané pro konstrukci intervalu spolehlivosti pro korelační koeficient a porovnat je na různých příkladech. První kapitola práce se věnuje úvodu o korelačním koeficientu a jeho vlastnostech a stručnému představení Fisherovy z-transformace. Druhá kapitola se věnuje metodě založené na zobecněných pivotech. Vysvětluje také proč je pro tuto metodu potřebný předpoklad dvojrozměrného normálního rozdělení. Třetí kapitola popisuje dvě metody založené na empirické věrohodnosti. Tyto metody jsou vhodné i pro jiná dvojrozměrná rozdělení než normální. V závěrečné kapitole jsou všechny metody aplikovány na několik příkladů a navzájem porovnávány.

Klíčová slova: korelační koeficient, interval spolehlivosti, normální rozdělení

Title: Confidence intervals for the correlation coefficient

Author: Martin Farda

Department: Department of Probability and Mathematical Statistics

Supervisor: doc. Mgr. Michal Kulich, Ph.D., Department of Probability and Mathematical Statistics

Abstract: The main goal of the thesis is to introduce methods used for the construction of confidence intervals for correlation coefficient in detail and to show their performances on various examples. In the first chapter is an introduction of basic properties of correlation coefficient and Fisher's z-transformation. The second chapter is about a method based on generalized pivotal quantities. It also contains an explanation why is an assumption of bivariate normal distribution necessary for this method. In the third chapter there is a description of two methods based on empirical likelihood. These methods are approptiate also for non-normal bivariate distributions. In the last chapter are all mentioned methods applied on several examples and compared with each other.

Keywords: correlation coefficient, confidence interval, normal distribution

Obsah

1. Úvod a základní pojmy

V první kapitole si zavedeme základní pojmy, nezbytné pro naši práci. Také se seznámíme s běžně používanou metodou pro testování nezávislosti náhodných veličin a metodou používanou pro konstrukci intervalu spolehlivosti pro korelační koeficient za předpokladu normality.

1.1 Základní pojmy a věty

Definice 1 (korelační koeficient)**.** *Nechť jsou X, Y náhodné veličiny s konečnými druhými momenty, s rozptyly $var X > 0$ a $var Y > 0$ a kovariancí $cov(X, Y)$. Pak definujeme jejich* korelační koeficient *ρ_{XY} vztahem*

$$\rho_{XY} = \frac{cov(X, Y)}{\sqrt{var(X)\, var(Y)}}.$$

Podívejme se na některé vlastnosti korelačního koeficientu. Ve zbytku kapitoly předpokládejme, že náhodné veličiny X a Y splňují podmínky z definice korelačního koeficientu. Již z definice a vlastností kovariance je zřejmé, že platí $\rho_{XY} = \rho_{YX}$.

Věta 1. *Platí*

$$-1 \leq \rho_{XY} \leq 1,$$

přičemž platí, že

$$|\rho_{XY}| = 1 \iff Y = a + bX \text{ s pravděpodobností 1 a } b \neq 0.$$

Důkaz můžeme najít v knize Anděl (2011, str. 39).

Věta 2 (Anděl, 2011,str. 39)**.** *pro reálná čísla a, b, c, d splňující $ac \neq 0$ platí*

$$\rho_{aX+b,cY+d} = \begin{cases} -\rho_{XY}, & pro\ ac < 0, \\ \rho_{XY}, & pro\ ac > 0. \end{cases}$$

Tedy lineární transformací se případně mění znaménko.

Věta 3. *Jsou-li náhodné veličiny X, Y nezávislé, pak*

$$\rho_{XY} = 0$$

Důkaz plyne z Anděl (2011, str. 34, věta 2.9) a definice korelačního koeficientu. Stejně jako v případě kovariance, nulový korelační koeficient obecně neimplikuje nezávislost náhodných veličin. Nicméně, je tomu tak v případě normálního rozdělení, jak plyne následující věty.

Věta 4. *Nechť má náhodný vektor $\begin{pmatrix} X \\ Y \end{pmatrix}$ dvojrozměrné normální rozdělení. Je-li $cov(X, Y) = 0$, pak jsou náhodné veličiny X, Y nezávislé.*

Důkaz najdeme v knize Anděl (2011, str. 66, věta 4.11). Dohromady s větou 3 dostáváme, že pro normální rozdělení platí ekvivalence mezi nulovým korelačním koeficientem a nezávislostí náhodných veličin.

Zaveďme značení pro výběrové průměry, výběrové rozptyly a výběrovou kovarianci náhodného výběru $\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix}, \ldots, \begin{pmatrix} X_n \\ Y_n \end{pmatrix}$.

$$\overline{X} = \frac{1}{n-1}\sum_{i=1}^{n} X_i, \quad \overline{Y} = \frac{1}{n-1}\sum_{i=1}^{n} Y_i,$$
$$S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2, \quad S_Y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2,$$
$$S_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y}).$$

Definice 2 (výběrový korelační koeficient)**.** *Nechť je* $\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix}, \ldots, \begin{pmatrix} X_n \\ Y_n \end{pmatrix}$ *náhodný výběr z nějakého dvojrozměrného rozdělení a nechť platí* $S_X^2 > 0$ *a* $S_Y^2 > 0$. *Pak definujeme* výběrový korelační koeficient r_{XY} *vztahem*

$$r_{XY} = \frac{S_{XY}}{\sqrt{S_X^2 S_Y^2}}.$$

Jak nám ukáže následující věta, výběrový korelační koeficient má podobné vlastnosti jako korelační koeficient.

Věta 5 (Anděl, 2011,str. 93)**.** *Platí*

$$-1 \leq r_{XY} \leq 1$$

a pro reálná čísla a, b, c, d *splňující* $ac \neq 0$ *platí*

$$r_{aX+b,cY+d} = \begin{cases} -r_{XY}, & \text{pro } ac < 0, \\ r_{XY}, & \text{pro } ac > 0. \end{cases}$$

Věta 6. *Pro náhodný výběr o rozsahu alespoň 2 pocházející ze spojitého rozdělení je výběrový korelační koeficient definován s pravděpodobností jedna.*

Důkaz můžeme nalézt znovu v knize Anděl (2011, str. 93, věta 6.1). Věta, kterou nyní položíme, je často využívána při testech nezávislosti dvou náhodných veličin.

Věta 7. *Nechť je* $\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix}, \ldots, \begin{pmatrix} X_n \\ Y_n \end{pmatrix}$ *náhodný výběr z dvojrozměrného normálního rozdělení s korelačním koeficientem* $\rho = 0$ *a nechť platí* $S_X^2 > 0$, $S_Y^2 > 0$ *a* $n \geq 3$. *Pak*

$$\sqrt{n-2}\frac{r_{XY}}{\sqrt{1-r^2}} \sim t_{n-2}.$$

Důkaz nalezneme opět v knize Anděl (2011, str.94).

1.2 Fisherova z-transformace

Konstrukce intervalů spolehlivosti pomocí Fisherovy z-transformace využívá její asymptotické normality za předpokladu, že náhodný výběr pochází z dvojrozměrného normálního rozdělení. Fisher definoval z-transformaci vztahem

$$Z(r) = \frac{1}{2} \ln \frac{1+r}{1-r}. \tag{1.1}$$

Odvození Fisherovy z-transformace lze nalézt v práci Winterbottom (1979).

Věta 8 (Anděl, 2011, str. 95)**.** *Nechť náhodný výběr* $\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix}, \ldots, \begin{pmatrix} X_n \\ Y_n \end{pmatrix}$ *pochází z dvojrozměrného normálního rozdělení s korelačním koeficientem* ρ *a nechť* r *je jeho výběrový korelační koeficient. Pak*

$$\frac{\sqrt{n-3}}{2} \ln \frac{(1+r)(1-\rho)}{(1-r)(1+\rho)} \xrightarrow{D} \mathcal{N}(0,1). \tag{1.2}$$

Užitím vztahu (1.2) a pomocí inverze $Z^{-1}(x) = \frac{e^{2x}-1}{e^{2x}+1}$ jsme schopni sestavit interval spolehlivosti I s pravděpodobností pokrytí $1-\alpha$ pro korelační koeficient ρ ve tvaru:

$$\begin{aligned} I &= \left(Z^{-1}\left[\left(\frac{1}{2} \ln \frac{1-r}{1+r} \right) - \frac{u_{1-\alpha/2}}{\sqrt{n-3}} \right], Z^{-1}\left[\left(\frac{1}{2} \ln \frac{1-r}{1+r} \right) + \frac{u_{1-\alpha/2}}{\sqrt{n-3}} \right] \right) \\ &= \left(\frac{\frac{1-r}{1+r} \exp\left(-\frac{2}{\sqrt{n-3}} u_{1-\alpha/2}\right) - 1}{\frac{1-r}{1+r} \exp\left(-\frac{2}{\sqrt{n-3}} u_{1-\alpha/2}\right) + 1}, \frac{\frac{1-r}{1+r} \exp\left(\frac{2}{\sqrt{n-3}} u_{1-\alpha/2}\right) - 1}{\frac{1-r}{1+r} \exp\left(\frac{2}{\sqrt{n-3}} u_{1-\alpha/2}\right) + 1} \right), \end{aligned} \tag{1.3}$$

kde $u_{1-\alpha/2}$ značí $(1-\alpha/2)$-tý kvantil normovaného normálního rozdělení.

2. Konstrukce pomocí zobecněných pivotů

V této kapitole se budeme věnovat konstrukci intervalů spolehlivosti pro korelační koeficient pomocí postupu se zobecněnými pivoty, který zavedli ve své práci Hu, Jung a Qin (2020).

Nejprve si připravme půdu pro aplikaci metody založené na zobecněných pivotech. K zobecněným pivotů samotným a jejich užití při konstrukci intervalu spolehlivosti pro korelační koeficient se dostaneme v druhé části kapitoly.

2.1 Příprava na použití metody

Korelační koeficient je vyjádřen vztahem

$$\rho = \frac{\sigma_{XY}}{\sqrt{\sigma_X^2 \sigma_Y^2}},$$

pro náhodné veličiny X, Y, jejich směrodatné odchylky a kovarianci. Korelační koeficient je tedy funkcí parametrů σ_{XY}, σ_X^2 a σ_Y^2. V této části kapitoly chceme najít takové parametry z nichž lze σ_{XY}, σ_X^2 a σ_Y^2 sestavit a které budeme schopni vyjádřit pomocí na nich založených statistik a náhodných veličin se známým rozdělením. Tím připravíme data pro druhou části kapitoly.

Nechť $\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix}, \ldots, \begin{pmatrix} X_n \\ Y_n \end{pmatrix}$ je náhodný výběr z rozdělení s konečnými druhými momenty, se střední hodnotou $\boldsymbol{\mu} = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}$, kovarianční maticí $\Sigma = \begin{pmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{pmatrix}$, kde oba rozptyly jsou kladné a nechť $n > 2$.

Budeme používat značení $\overline{X}, \overline{Y}$ pro výběrové průměry a S_X^2, S_Y^2 pro výběrové rozptyly náhodných veličin X a Y. Jejich výběrovou kovarianci budeme značit S_{XY}. Písmenem ρ budeme vyjadřovat jejich korelační koeficient.

Z vlastností výběrového průměru, výběrového rozptylu a výběrové kovariance víme, že

$$\hat{\boldsymbol{\mu}} = \begin{pmatrix} \overline{X} \\ \overline{Y} \end{pmatrix}$$

je nestranným a konzistentním odhadem střední hodnoty $\boldsymbol{\mu}$ a

$$\hat{\Sigma} = \begin{pmatrix} S_X^2 & S_{XY} \\ S_{XY} & S_Y^2 \end{pmatrix}$$

je konzistentním odhadem kovarianční matice Σ.

Dále označme

$$\begin{aligned}\boldsymbol{X} &= (X_1, \dots, X_n)^\top,\\ \boldsymbol{Y} &= (Y_1, \dots, Y_n)^\top,\\ \boldsymbol{\beta} &= (\beta_1, \beta_2)^\top,\\ \mathbb{Z} &= \begin{pmatrix} 1 & X_1 \\ \vdots & \vdots \\ 1 & X_n \end{pmatrix}.\end{aligned}$$

Při uvedeném značení zaveďme lineární regresní model

$$\boldsymbol{Y} = \beta_1 + \beta_2 \boldsymbol{X} + \boldsymbol{\epsilon}, \tag{2.1}$$

který rovněž lze psát ve tvaru

$$\boldsymbol{Y} = \mathbb{Z}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{2.2}$$

kde $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_n)^\top, \mathsf{E}\,\boldsymbol{\epsilon} = \boldsymbol{0}$ a $\operatorname{var}\boldsymbol{\epsilon} = \operatorname{var}(Y_i|X_i)I_n$ a kde I_n je jednotková matice o rozměrech $n \times n$.

Označme $\boldsymbol{B} = (B_1, B_2)^\top$ odhad parametru $\boldsymbol{\beta}$ metodou nejmenších čtverců.

Věta 9. *Odhad $\boldsymbol{B}$ parametru $\boldsymbol{\beta}$ metodu nejmenších čtverců splňuje*

$$\boldsymbol{B} = (\mathbb{Z}^\top\mathbb{Z})^{-1}\mathbb{Z}^\top\boldsymbol{Y}.$$

Věta 10. *Platí*

$$\begin{aligned}\mathsf{E}\,\boldsymbol{B} &= \boldsymbol{\beta},\\ var\,\boldsymbol{B} &= var(Y_i|X_i)(\mathbb{Z}^\top\mathbb{Z})^{-1}.\end{aligned}$$

Důkaz věty 9 najdeme v knize Anděl (2011, str 81, Věta 5.1), důkaz věty 10 najdeme rovněž v knize Anděl (2011, str.82, Věta 5.2).

Zjistíme $(\mathbb{Z}^\top\mathbb{Z})^{-1}$:

$$(\mathbb{Z}^\top\mathbb{Z})^{-1} = \begin{pmatrix} \frac{1}{n} + \frac{\overline{X}^2}{(n-1)S_X^2} & -\frac{\overline{X}}{(n-1)S_X^2} \\ -\frac{\overline{X}}{(n-1)S_X^2} & \frac{1}{(n-1)S_X^2} \end{pmatrix}$$

a $\mathbb{Z}^\top\boldsymbol{Y}$:

$$\mathbb{Z}^\top\boldsymbol{Y} = \left(n\overline{Y}, (n-1)S_{XY} + \overline{X}n\overline{Y}\right)^\top.$$

Z věty 9 pak můžeme zjistit tvar $\boldsymbol{B}$.

$$\boldsymbol{B} = (B_1, B_2)^\top = (\mathbb{Z}^\top\mathbb{Z})^{-1}\mathbb{Z}^\top\boldsymbol{Y} = \left(\overline{Y} - \frac{S_{XY}}{S_X^2}\overline{X}, \frac{S_{XY}}{S_X^2}\right)^\top,$$

Tento odhad je podle věty 10 nestranným odhadem parametru $\boldsymbol{\beta}$.

Věta 11 (Anděl, 2011, str. 105)**.** *Nechť $\mathsf{E}\,\hat{\theta}_n < \infty$ pro každé $n \in \mathbb{N}$. Platí-li $\mathsf{E}\,\hat{\theta}_n \to \theta$ a $var\hat{\theta}_n \to 0$, pak je $\hat{\theta}_n$ konzistentním odhadem parametru θ.*

Důkaz najdeme v knize Anděl (2011, str. 105). Z věty 10 známe vyjádření $\operatorname{var} \boldsymbol{B} = \operatorname{var}(Y_i|X_i)(\mathbb{Z}^\top\mathbb{Z})^{-1}$ a z tvaru $(\mathbb{Z}^\top\mathbb{Z})^{-1}$ vidíme, že $\operatorname{var} B_1 \to 0$ a $\operatorname{var} B_2 \to 0$ pro $n \to \infty$. Jsou tedy splněny předpoklady pro použití věty 11. Z ní plyne, že B_1, B_2 jsou konzistentními odhady parametrů β_1, β_2. Také víme, že $\hat{\Sigma}$ je konzistentním odhadem kovarianční matice Σ. Nyní již můžeme najít vyjádření parametrů β_1, β_2 :

$$\begin{aligned} \beta_1 &= \mu_Y - \frac{\sigma_{XY}}{\sigma_X^2}\mu_X, \\ \beta_2 &= \frac{\sigma_{XY}}{\sigma_X^2}. \end{aligned} \tag{2.3}$$

Při známém vyjádření parametru β_2 jsme schopni získat následující vyjádření pro reziduální rozptyl $\sigma^2_{Y|X} = \operatorname{var}(Y_i|X_i)$ (Anděl, 2011, str.38):

$$\sigma^2_{Y|X} = \sigma^2_Y - \beta_2^2\sigma_X^2 = \sigma_Y^2 - \frac{\sigma^2_{XY}}{\sigma^2_X}.$$

Označme reziduální součet čtverců modelu (2.1) jako SS_e. Pak platí následující věta.

Věta 12. $\mathsf{E}\,\frac{SS_e}{n-2} = \sigma^2_{Y|X}$.

Důkaz můžeme najít v knize Anděl (2011, str. 82, Věta 5.4) pro volbu $k = 2$. Nyní zjistíme tvar reziduálního součtu čtverců SS_e v našem modelu. Platí

$$SS_e = \sum_{i=1}^n (Y_i - B_1 - B_2X_i)^2.$$

Výraz upravíme

$$\begin{aligned} \sum_{i=1}^n (Y_i - B_1 - B_2X_i)^2 &= \sum_{i=1}^n (Y_i - \overline{Y} + B_2\overline{X} - B_2X_i)^2 \\ &= \sum_{i=1}^n \Big[(Y_i - \overline{Y}) - B_2(X_i - \overline{X})\Big]^2 \\ &= \sum_{i=1}^n (Y_i - \overline{Y})^2 - 2\frac{S_{XY}}{S_X^2}\sum_{i=1}^n (Y_i - \overline{Y})(X_i - \overline{X}) \\ &\quad + \frac{S^2_{XY}}{S^4_X}\sum_{i=1}^n (X_i - \overline{X})^2 \\ &= (n-1)S_Y^2 - 2\frac{S_{XY}}{S_X^2}(n-1)S_{XY} + \frac{S^2_{XY}}{S^4_X}(n-1)S_X^2 \\ &= (n-1)\left(S_Y^2 - \frac{S^2_{XY}}{S^2_X}\right). \end{aligned}$$

Známe tedy S_X^2, nestranný a konzistentní odhad σ_X^2, B_2, nestranný a konzistentní odhad parametru β_2 a z věty 12 víme, že $\frac{SS_e}{n-2}$ je nestranný odhad parametru $\sigma^2_{Y|X}$.

Poznámka. Za předpokladu normality rozdělení z kterého náhodný výběr pochází bychom mohli dojít k vyjádření parametrů $\beta_1, \beta_2, \sigma^2_{Y|X}$ přímou cestou z vyjádření lineárního regresního modelu. Předpokládáme-li, že náhodný vektor $\begin{pmatrix} X \\ Y \end{pmatrix}$ má rozdělení $\mathcal{N}_2(\boldsymbol{\mu}, \Sigma)$, můžeme si uvědomit, že podmíněné rozdělení náhodné veličiny Y při daném $X = x$ má následující tvar (Anděl, 2011, str. 67):

$$(Y|X = x) \sim \mathcal{N}\left(\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X), \sigma_Y^2(1 - \rho^2)\right) \tag{2.4}$$

a že z rovnosti (2.1) a vlastností $\boldsymbol{\epsilon}$ plyne:

$$\begin{aligned} \beta_1 &= \mathsf{E}\,(Y|X = 0), \\ \beta_2 &= \mathsf{E}\,(Y|X = x + 1) - \mathsf{E}\,(Y|X = x). \end{aligned}$$

Získáváme tak vyjádření parametrů β_1, β_2:

$$\begin{aligned} \beta_1 &= \mu_Y - \rho\frac{\sigma_Y}{\sigma_X}\mu_X = \mu_Y - \frac{\sigma_{XY}}{\sigma_X^2}\mu_X, \\ \beta_2 &= \rho\frac{\sigma_Y}{\sigma_X} = \frac{\sigma_{XY}}{\sigma_X^2}. \end{aligned}$$

Reziduální rozptyl $\sigma_{Y|X}$ bychom pak zjistili ze vztahu (2.4):

$$\sigma_{Y|X}^2 = \sigma_Y^2(1 - \rho^2) = \sigma_Y^2 - \frac{\sigma_{XY}^2}{\sigma_X^2}.$$

Všimněme si, že pomocí σ_X a výše zavedených parametrů $\beta_2, \sigma_{Y|X}^2$ lze vyjádřit σ_{XY} a σ_Y^2 následujícím způsobem:

$$\sigma_{XY} = \beta_2\sigma_X^2, \quad \sigma_Y^2 = \beta_2^2\sigma_X^2 + \sigma_{Y|X}^2. \tag{2.5}$$

Tohoto zjištění využijeme při konstrukci zobecněných pivotů v další části kapitoly.

Až doposud jsme byli schopni se obejít bez předpokladu, že náhodný výběr pochází z dvojrozměrného normálního rozdělení. Předpokládejme nyní, že rozdělení ze kterého náhodný výběr pochází skutečně je dvojrozměrné normální rozdělení. Pak můžeme položit následující větu, která právě předpoklad normality vyžaduje.

Věta 13. *Při zavedeném značení platí:*

$$\begin{aligned} &a) \quad U = \frac{(n-1)S_X^2}{\sigma_X^2} \sim \chi_{n-1}^2, \\ &b) \quad U_{Y|X} = \frac{SS_e}{\sigma_{Y|X}^2} \sim \chi_{n-2}^2, \\ &c) \quad Z = (B_2 - \beta_2)\sqrt{\frac{(n-1)S_X^2}{\sigma_{Y|X}^2}} \sim \mathcal{N}(0,1). \end{aligned}$$

Důkaz. Viz Anděl (2011, Věta 4.21) pro a), Anděl (2011, Věta 5.6) pro b) a Anděl (2011, Věta 5.5) pro c).

□

Poznámka. Pokud náhodný výběr z normálního rozdělení nepochází, platí část c) věty 13 alespoň asymptoticky (Flynn, 1999, str. 6), tedy

$$(B_2 - \beta_2)\sqrt{\frac{(n-1)S_X^2}{\sigma_{Y|X}^2}} \to \mathcal{N}(0,1).$$

Části a), b) věty 13 však bez předpokladu normality neplatí ani asymptoticky. je nám známé asymptotické rozdělení výběrového rozptylu

$$\sqrt{n}(S_X^2 - \sigma_X^2) \xrightarrow{D} \mathcal{N}(0, \sigma_X^4(\gamma_2 - 1)),$$

kde γ_2 značí špičatost. Normální rozdělení má špičatost $\gamma_2 = 3$, nicméně pro data pocházející obecně z nějakého jiného neznámého rozdělení zůstává parametr špičatosti neznámý a proto není ani toto asymptotické rozdělení výběrového rozptylu pro metodu zobecněných pivotů vhodné. Abychom mohli metodu rozšířit i na data nepocházející z normálního rozdělení, bylo by třeba najít alespoň asymptotická rozdělení funkce statistiky $(n-1)S_X^2$ a parametru σ_X^2 a funkce statistiky SS_e a parametru $\sigma_{Y|X}^2$.

2.2 Zobecněné pivoty

Nejprve si obecně zavedeme značení, příslušné pojmy a popíšeme metodu. Její použití si ukážeme na jednoduchém příkladu a poté se přesuneme k aplikaci na situaci připravenou v první části kapitoly. Na závěr si krátce shrneme celý postup konstrukce intervalu spolehlivosti pro korelační koeficient.

Nechť $\boldsymbol{X}$ je náhodný vektor, $\boldsymbol{x}$ je pozorovaná hodnota náhodného vektoru $\boldsymbol{X}$ a $\boldsymbol{\xi}$ je neznámý vektorový parametr. Předpokládejme, že parametr θ je funkcí parametru $\boldsymbol{\xi}$. Je-li $Q = Q(\boldsymbol{X}, \boldsymbol{\xi})$ funkcí $\boldsymbol{X}$ a $\boldsymbol{\xi}$, řekneme, že Q je pivotem parametru θ, pokud rozdělení Q nezávisí na $\boldsymbol{\xi}$ a pokud její pozorovaná hodnota $q = Q(\boldsymbol{x}, \boldsymbol{\xi})$ závisí pouze na parametru θ.

Pojem zobecněný pivot[1] zavedl ve své práci Weerahandi (1993). Definujme si nyní zobecněné pivoty se zavedeným značením $\boldsymbol{X}, \boldsymbol{x}$ a $\boldsymbol{\xi}$.

Definice 3. *Nechť je $R = r(\boldsymbol{X}; \boldsymbol{x}, \boldsymbol{\xi})$ je funkcí $\boldsymbol{X}, \boldsymbol{x}$ a $\boldsymbol{\xi}$. Pak R nazveme* zobecněným pivotem*, pokud splňuje následující dvě podmínky:*
Podmínka A: R má pravděpodobnostní rozdělení nezávislé na $\boldsymbol{\xi}$,
Podmínka B: $r_{obs} = r(\boldsymbol{x}; \boldsymbol{x}, \boldsymbol{\xi})$ závisí na $\boldsymbol{\xi}$ pouze skrze parametr θ.

Následně Hanning, Iyer a Patterson (2006) definovali pojem fiduciální zobecněný pivot[2]. Fiduciální zobecněné pivoty tvoří důležitou podskupinu zobecněných pivotů. Vyznačují se tím, že pro ně platí v podmínce B dokonce rovnost $r_{obs} = \theta$.

Popišme si nyní postup, jakým se fiduciální zobecněné pivoty a na nich založené intervaly spolehlivosti konstruují. Tento postup odpovídá metodě používané v práci Hanning a kol. (2006) a je přehledně popsán v článku Wang a Iyer (2005).

Nejprve najdeme pivoty pro minimální postačující statistiky zadaných parametrů. Připomeňme, že statistiku T nazveme postačující pro parametr θ, pokud

[1] Generalized pivotal quantity (GPQ)
[2] Fiducial generalized pivotal quantity (FGPQ)

podmíněné rozdělení náhodného vektoru $\boldsymbol{X}$ při daném T nezávisí na θ. Postačující statistiku nazýváme minimální, jestliže je funkcí kterékoli jiné postačující statistiky (Anděl, 2011, str. 124-126).

V druhém kroku vyjádříme parametry jako funkce minimálních postačujících statistik a zjištěných pivotů.

Fiduciální zobecněné pivoty poté získáme, dosadíme-li za minimální postačující statistiky jejich pozorované hodnoty. Pokud hledáme fiduciální zobecněný pivot pro funkci parametrů, dosadíme do vyjádření této funkce za parametry jejich fiduciální zobecněné pivoty.

Pro zjištěný fiduciální zobecněný pivot $R = r(\boldsymbol{X}, \boldsymbol{x}, \boldsymbol{\xi})$ parametru θ platí

$$P\left[R \in C\right] = 1 - \alpha,$$

kde

$$C = \left(R_{\alpha/2}, R_{1-\alpha/2}\right),$$

a kde R_α je α-kvantil rozdělení $R = r(\boldsymbol{X}, \boldsymbol{x}, \boldsymbol{\xi})$.

Je-li $\boldsymbol{X}^*$ nezávislá kopie vektoru $\boldsymbol{X}$ a $\boldsymbol{x}^*$ její pozorovaná hodnota, pak dosazením pozorované hodnoty $\boldsymbol{x}^*$ za $\boldsymbol{X}$ dostáváme, že C je $100(1-\alpha)\%$ interval spolehlivosti pro $r(\boldsymbol{x}^*, \boldsymbol{x}, \boldsymbol{\xi})$. Vektor $\boldsymbol{x}$ se skládá ze statistik, které odhadují zadané parametry a které k těmto parametrům konvergují v pravděpodobnosti (Hanning a kol., 2006, str. 256). Tedy i složky nezávislé kopie $\boldsymbol{x}^*$ takového vektoru konvergují k zadaným parametrům. Pro dostatečně velký rozsah výběru n tak jsou hodnoty složek obou vektorů libovolně blízko hodnotám zadaných parametrů a tedy i hodnoty složek $\boldsymbol{x}^*$ a $\boldsymbol{x}$ si budou pro dostatečně velký rozsah výběru n blízké. Dosadíme-li $\boldsymbol{x}$ namísto $\boldsymbol{x}^*$ a uvědomíme-li si, že $r(\boldsymbol{x}, \boldsymbol{x}, \boldsymbol{\xi}) = \theta$, dostáváme C jako asymptotický interval spolehlivosti pro parametr θ (Hanning a kol., 2006, str. 257).

Popsaný postup si předvedeme na jednoduchém příkladu.

Příklad. Uvažujme náhodný výběr $W_1, \ldots, W_m$ z normálního rozdělení s parametry μ_W, σ_W^2. Budeme hledat fiduciální zobecněné pivoty pro oba zadané parametry pomocí výše popsaného postupu. Nejprve najdeme pivoty pro statistiky $\overline{W}, S_W^2$.

$$U_W = \frac{(m-1)S_W^2}{\sigma_W^2} \sim \chi^2_{m-1},$$
$$Z_W = \sqrt{m}\frac{\overline{W} - \mu_W}{\sigma_W} \sim \mathcal{N}(0{,}1).$$

Parametry si vyjádříme jako funkce statistik $\overline{W}, S_W^2$ a pivotů U_W, Z_W.

$$\sigma_W^2 = \frac{(m-1)S_W^2}{U_W},$$
$$\mu_W = \overline{W} - \frac{Z_W}{\sqrt{m}}\sqrt{\frac{(m-1)S_W^2}{U_W}}.$$

Dosadíme do vyjádření pro μ_W, σ_W^2 pozorované hodnoty $\overline{w}, s_W^2$ výběrového průměru a výběrového rozptylu. U_W, Z_W pak ve vyjádření fiduciálních zobecněných pivotů chápeme již jako náhodné veličiny se zjištěnými rozděleními a nikoli jako funkce zadaných parametrů dat. Fiduciální zobecněné pivoty tak jsou funkcemi pozorovaných hodnot statistik $\overline{w}, s_W^2$ a náhodných veličin U_W, Z_W. Získáváme tak fiduciální zobecněné pivoty R_{μ_W} a $R_{\sigma_W^2}$ ve tvaru:

$$R_{\sigma_W^2} = \frac{(m-1)s_W^2}{U_W},$$
$$R_{\mu_W} = \overline{w} - \frac{Z_W}{\sqrt{m}}\sqrt{\frac{(m-1)s_W^2}{U_W}}.$$

Můžeme si uvědomit, že $\frac{Z_W}{U_W/(m-1)}$ má t-rozdělení s $m-1$ stupni volnosti. R_{μ_W} tak můžeme psát také ve tvaru

$$R_{\mu_W} = \overline{w} - \frac{s_W}{\sqrt{m}}T_{m-1},$$

kde T_{m-1} má t-rozdělení s $m-1$ stupni volnosti. Našli jsme tedy fiduciální zobecněné pivoty k parametrům μ_W, σ_W^2. Na závěr, již snadno, ze zjištěných tvarů fiduciálních zobecněných pivotů, sestavíme pro parametry μ_W a σ_W^2 $100(1-\alpha)\%$ intervaly spolehlivosti I_{μ_W} a $I_{\sigma_W^2}$:

$$I_{\mu_W} = \left(\overline{w} - \frac{s_W}{\sqrt{m}}t_{1-\alpha/2}, \overline{w} - \frac{s_W}{\sqrt{m}}t_{\alpha/2}\right),$$
$$I_{\sigma_W^2} = \left(\frac{(m-1)s_W^2}{q_{1-\alpha/2}}, \frac{(m-1)s_W^2}{q_{\alpha/2}}\right),$$

kde t_α je α-kvantil t-rozdělení s $m-1$ stupni volnosti a q_α je α-kvantil rozdělení χ^2_{m-1}. Na závěr příkladu poznamenejme, že výsledné intervaly spolehlivosti jsou stejné jako známé intervaly spolehlivosti pro tyto parametry, k nimž se dojde pomocí pivotu $\sqrt{m}\frac{\overline{W}-\mu_W}{S_W}$ v případě střední hodnoty a pomocí stejného pivotu U_W v případě rozptylu.

Ukázali jsme si na příkladu jak zjistit fiduciální zobecněné pivoty. Nyní navažme na první část kapitoly a aplikujme postup na problém zjišťování fiduciálního zobecněného pivotu pro korelační koeficient. V první části kapitoly jsme zjistili, že korelační koeficient ρ je funkcí parametrů σ_X^2, σ_Y^2 a σ_{XY} a v bodě (2.5) jsme si ukázali jak vyjádřit σ_Y^2 a σ_{XY} pomocí parametrů $\sigma_X^2, \sigma_{Y|X}^2$ a β_2. Korelační koeficient tak můžeme chápat jak funkci právě těchto tří parametrů. Proto budeme chtít najít jejich fiduciální zobecněné pivoty $R_{\sigma_X^2}, R_{\sigma_{Y|X}^2}$ a R_{β_2}.

V první kapitole jsme ve větě 13 našli pivoty se známými rozděleními. Pomocí ve větě 13 zavedených náhodných veličin U,$U_{Y|X}$ a Z jsme schopni snadno vyjádřit parametry $\sigma_X^2, \sigma_{Y|X}^2$ a β_2 následujícím způsobem:

$$\sigma_X^2 = \frac{(n-1)S_X^2}{U},$$
$$\sigma_{Y|X}^2 = \frac{SS_e}{U_{Y|X}},$$
$$\beta_2 = B_2 - Z\sqrt{\frac{SS_e}{U_{Y|X}}\frac{1}{(n-1)S_X^2}}.$$

Podle popsaného postupu získáme fiduciální zobecněné pivoty zmíněných parametrů dosazením s_X^2, ss_e, b_2, pozorovaných hodnot statistik S_X^2, SS_e, B_2.

$$
\begin{aligned}
R_{\sigma_X^2} &= \frac{(n-1)s_X^2}{U}, \\
R_{\sigma_{Y|X}^2} &= \frac{ss_e}{U_{Y|X}}, \\
R_{\beta_2} &= b_2 - Z\sqrt{\frac{ss_e}{U_{Y|X}}\frac{1}{(n-1)s_X^2}},
\end{aligned}
$$

kde náhodné veličiny $U, U_{Y|X}$ a Z generujeme z rozdělení zjištěných ve větě 13 nezávisle na sobě (Hu a kol., 2020, str. 30), tedy $U \sim \chi^2_{n-1}$, $U_{Y|X} \sim \chi^2_{n-2}$ a $Z \sim \mathcal{N}(0,1)$.

Připomeňme, že ze vztahu (2.5) víme, jak z parametrů $\sigma_X^2, \sigma_{Y|X}^2$ a β_2 složit σ_Y^2 a σ_{XY}. Fiduciální zobecněné pivoty pro σ_Y^2, σ_{XY} pak jsou podle metody popsané výše:

$$
\begin{aligned}
R_{\sigma_Y^2} &= R_{\beta_2}^2 R_{\sigma_X^2} + R_{\sigma_{Y|X}^2} \\
R_{\sigma_{XY}} &= R_{\beta_2} R_{\sigma_X^2}
\end{aligned}
$$

Fiduciální zobecněný pivot pro korelační koeficient ρ je tedy

$$
R_\rho = \frac{R_{\sigma_{XY}}}{R_{\sigma_X} R_{\sigma_Y}}. \tag{2.6}
$$

Z vlastností fiduciálních zobecněných pivotů plyne, že dosazením pozorovaných hodnot s_X^2, ss_e, b_2 do $U, U_{Y|X}, Z$ ve vyjádřeních fiduciálních zobecněných pivotů $R_{\sigma_X^2}, R_{\sigma_{Y|X}^2}$ a R_{β_2} dostaneme samotné parametry $\sigma_X^2, \sigma_{Y|X}^2$ a β_2. Proto platí, že tímto dosazením pozorovaných hodnot do vyjádření R_ρ dostaneme samotný korelační koeficient ρ. Tedy $100(1-\alpha)\%$ oboustranný asymptotický interval spolehlivosti pro ρ založený na fiduciálních zobecněných pivotech je roven

$$
\left(R_{\alpha/2}, R_{1-\alpha/2}\right),
$$

kde R_α je α-kvantil rozdělení R_ρ. Rozdělení R_ρ v praxi neznáme, vygenerujeme dostatek realizací náhodných veličin $U, U_{Y|X}, Z$, pro každou vygenerovanou trojici hodnot spočteme R_ρ a tímto způsob odhadneme jeho rozdělení. Za R_α pak vezmeme α-kvantil vzniklého empirického rozdělení.

Na závěr kapitoly si krátce shrňme popsaný postup pro zjišťování intervalu spolehlivosti pro korelační koeficient ρ. Nejprve ze zadaného náhodného výběru $\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix}, \dots, \begin{pmatrix} X_n \\ Y_n \end{pmatrix}$ spočteme pozorované hodnoty výběrových průměrů, výběrových rozptylů a výběrové kovariance. Pomocí vztahů

$$
B_2 = \frac{S_{XY}}{S_X^2}, \quad SS_e = (n-1)\left(S_Y^2 - \frac{S_{XY}^2}{S_X^2}\right).
$$

spočteme pozorované hodnoty b_2 a ss_e. Poté vygenerujeme hodnoty náhodných veličin $U, U_{Y|X}, Z$ z příslušných rozdělení:

$$
U \sim \chi^2_{n-1}, \quad U_{Y|X} \sim \chi^2_{n-2}, \quad Z \sim \mathcal{N}(0,1).
$$

Následně zjistíme hodnoty fiduciálních zobecněných pivotu pro $\sigma_X^2, \sigma_{Y|X}^2$ a β_2 :

$$\begin{aligned}
R_{\sigma_X^2} &= \frac{(n-1)s_X^2}{U},\\
R_{\sigma_{Y|X}^2} &= \frac{ss_e}{U_{Y|X}},\\
R_{\beta_2} &= b_2 - Z\sqrt{\frac{ss_e}{U_{Y|X}}\frac{1}{(n-1)s_X^2}}.
\end{aligned}$$

Z nich spočteme fiduciální zobecněné pivoty pro σ_{XY}, σ_Y :

$$\begin{aligned}
R_{\sigma_Y^2} &= R_{\beta_2}^2 R_{\sigma_X^2} + R_{\sigma_{Y|X}^2},\\
R_{\sigma_{XY}} &= R_{\beta_2} R_{\sigma_X^2}.
\end{aligned}$$

Nakonec zjistíme hodnotu fiduciálního zobecněného pivotu pro ρ :

$$R_\rho = \frac{R_{\sigma_{XY}}}{R_{\sigma_X} R_{\sigma_Y}}.$$

Tento postup opakujeme dostatečněkrát abychom mohli odhadovat rozdělení R_ρ. Hu a kol. (2020) ve své práci doporučují 10 000 opakování. Nakonec určíme $100\,(1-\alpha)\,\%$ oboustranný asymptotický interval spolehlivosti pro ρ jako interval

$$\left(R_{\alpha/2}, R_{1-\alpha/2}\right),$$

kde R_α značí α-kvantil empirického rozdělení spočtených hodnot R_ρ.

3. Empirická věrohodnost

Pojem empirická věrohodnost[1] zavedl Owen (1988). Definujme si ji nejprve obecně právě podle práce Owena (Owen, 1988) a poté převeďme na případ korelačního koeficientu podle článku Hu a kol. (2020).

Nechť $X_1, \dots, X_n$ je náhodný výběr z rozdělení s distribuční funkcí F_0. Empirická distribuční funkce F_n se často považuje za neparametrický maximálně věrohodný odhad distribuční funkce F_0, protože maximalizuje

$$L(F) = \prod_{i=1}^{n} [F(X_i) - F(X_i-)]$$

přes všechny distribuční funkce F. Se zavedeným vyjádřením $L(F)$ definujme empirický věrohodnostní poměr.

Definice 4. *Za platnosti výše zavedeného značení pro $L(F)$ definujeme* empirický věrohodnostní poměr *vztahem*

$$R(F) = \frac{L(F)}{L(F_n)}.$$

Předpokládejme, že

$$w_i \geq 0, \qquad \sum_{j: X_j = X_i} w_j = F(X_i) - F(X_i-), \quad i = 1, \dots, n. \tag{3.1}$$

Pak w_i mají podobu pravděpodobností přiřazených jednotlivým pozorováním. Platí proto, že součet prvků vektoru $\boldsymbol{w} = (w_1, \dots, w_n)^\top$ je roven 1.

Definice 5. *Definujme*

$$\widetilde{L}(F,\boldsymbol{w}) = \prod_{i=1}^{n} w_i, \tag{3.2}$$

kde $\boldsymbol{w} = (w_1, \dots, w_n)^\top$ a nazvěme $\widetilde{L}(F, \boldsymbol{w})$ věrohodnostní funkcí založenou na pozorování*.*

Protože součet složek vektoru vektoru $\boldsymbol{w}$ je roven 1, je zřejmé, že funkce $\widetilde{L}$ nabývá maxima pro $F = F_n$ a $w_1 = \cdots = w_n = \frac{1}{n}$. Můžeme vyjádřit také empirický věrohodnostní poměr pomocí $\boldsymbol{w}$.

Definice 6. *Definujme* věrohodnostní poměr založený na pozorování *vztahem*

$$\widetilde{R}(F,\boldsymbol{w}) = \frac{\widetilde{L}(F,\boldsymbol{w})}{\widetilde{L}(F_n,\boldsymbol{w})} = \prod_{i=1}^{n} n w_i. \tag{3.3}$$

[1] Empirical likelihood (EL)

3.1 Intervaly spolehlivosti pomocí empirické věrohodnosti

Uvažujme náhodný výběr $\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix}, \dots, \begin{pmatrix} X_n \\ Y_n \end{pmatrix}$ z rozdělení s distribuční funkcí F_0 a konečnými druhými momenty. Předpokládejme také, že oba rozptyly jsou kladné. Označme $W_i = \begin{pmatrix} X_i \\ Y_i \end{pmatrix}$ pro $i = 1, \dots, n$ a $W = \begin{pmatrix} X \\ Y \end{pmatrix}$. Víme, že pro náhodné veličiny X a Y se středními hodnotami μ_X a μ_Y, s rozptyly $\sigma_X^2 > 0$ a $\sigma_Y^2 > 0$ a s korelačním koeficientem ρ platí:

$$\mathsf{E}\left(\frac{X - \mu_X}{\sigma_X}\frac{Y - \mu_Y}{\sigma_Y}\right) = \rho.$$

Jednoduchou úpravou dostáváme:

$$\mathsf{E}\left(\frac{X - \mu_X}{\sigma_X}\frac{Y - \mu_Y}{\sigma_Y} - \rho\right) = 0. \tag{3.4}$$

Zaveďme funkci V vztahem $V(W_i) = \left(\frac{X_i - \mu_X}{\sigma_X}\frac{Y_i - \mu_Y}{\sigma_Y}\right)$ pro $i = 1, \dots, n$. Nyní definujme empirickou věrohodnost pro korelační koeficient ρ užitím rovnosti (3.2), vlastností (3.1) pravděpodobnostního vektoru $\boldsymbol{w}$, rovnosti (3.4) a právě zavedné funkce V.

Definice 7. *Označme*

$$L_0(\rho) = \sup_{\boldsymbol{w}}\Big\{\prod_{i=1}^n w_i : w_i \geq 0, \sum_{i=1}^n w_i = 1, \sum_{i=1}^n w_i\,(V(W_i) - \rho) = 0\Big\}. \tag{3.5}$$

Funkci $L_0(\rho)$ budeme nazývat empirickou věrohodností pro korelační koeficient ρ*.*

Střední hodnoty μ_X a μ_Y a směrodatné odchylky σ_X a σ_Y neznáme, nicméně je jsme schopni odhadnout pomocí výběrových průměrů $\overline{X}$ a $\overline{Y}$ a výběrových směrodatných odchylek S_X a S_Y. Tyto odhady dosadíme za střední hodnoty a směrodatné odchylky do vztahu (3.5). Označíme-li $\widehat{V}(W_i) = \left(\frac{X_i - \overline{X}}{S_X}\frac{Y_i - \overline{Y}}{S_Y}\right)$ pro $i = 1, \dots, n$, dostáváme

$$\widehat{L_0}(\rho) = \sup_{\boldsymbol{w}}\Big\{\prod_{i=1}^n w_i : w_i \geq 0, \sum_{i=1}^n w_i = 1, \sum_{i=1}^n w_i\left(\widehat{V}(W_i) - \rho\right) = 0\Big\}. \tag{3.6}$$

$\widehat{L_0}(\rho)$ jsme získali dosazením za střední hodnoty a za směrodatné odchylky do vyjádření $L_0(\rho)$. Můžeme proto nazvat $\widehat{L_0}(\rho)$ *dosazovanou empirickou věrohodností pro korelační koeficient ρ.*

Z důvodu jednoduššího postupu budeme hledat logaritmickou dosazovanou empirickou věrohodnost pro korelační koeficient. To znamená, že místo suprema funkce $\prod_{i=1}^n w_i$ budeme za podmínek ze vztahu (3.10) hledat supremum funkce $\log \prod_{i=1}^n w_i$. Pro získání požadovaného tvaru vyjádření w_i rovněž vynásobíme hodnotou $-n$ obě strany poslední podmínky. Dostáváme tak poslední podmínku ve tvaru

$$-n\sum_{i=1}^n w_i\left(\widehat{V}(W_i) - \rho\right) = 0.$$

Nyní aplikujme metodu Lagrangeových multiplikátorů za účelem získání vyjádření pro w_i, $i = 1, \ldots, n$. Zaveďme Lagrangeovu funkci

$$\mathcal{L}(w_1, \ldots, w_n, \lambda_1, \lambda_2) = \log \prod_{i=1}^{n} w_i + \lambda_1 (1 - \sum_{i=1}^{n} w_i) - \lambda_2 n \sum_{i=1}^{n} w_i \left(\widehat{V}(W_i) - \rho \right).$$

Zderivujme ji podle jednotlivých proměnných:

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial w_i} &= \frac{1}{w_i} - \lambda_1 - \lambda_2 n \left(\widehat{V}(W_i) - \rho \right), \quad i = 1, \ldots, n, \\
\frac{\partial \mathcal{L}}{\partial \lambda_1} &= 1 - \sum_{i=1}^{n} w_i, \\
\frac{\partial \mathcal{L}}{\partial \lambda_2} &= -n \sum_{i=1}^{n} w_i \left(\widehat{V}(W_i) - \rho \right).
\end{aligned}$$

Derivace položíme rovny nule:

$$1 - \lambda_1 w_i - \lambda_2 n w_i \left(\widehat{V}(W_i) - \rho \right) = 0, \quad i = 1, \ldots, n, \tag{3.7}$$

$$1 - \sum_{i=1}^{n} w_i = 0,$$

$$-n \sum_{i=1}^{n} w_i \left(\widehat{V}(W_i) - \rho \right) = 0. \tag{3.8}$$

Pokud sečteme rovnosti pro všechna $i = 1, \ldots, n$, dostaneme rovnost

$$n - \lambda_1 \sum_{i=1}^{n} w_i - \lambda_2 n \sum_{i=1}^{n} w_i \left(\widehat{V}(W_i) - \rho \right) = 0.$$

Uvědomíme-li si, že $\sum_{i=1}^{n} w_i = 1$ a $\sum_{i=1}^{n} w_i \left(\widehat{V}(W_i) - \rho \right) = 0$, získáme vyjádření pro λ_1 :

$$\lambda_1 = n.$$

Získaný tvar λ_1 dosadíme do rovnosti (3.7) a obdržíme tak rovnost

$$1 - n w_i - \lambda_2 n w_i \left(\widehat{V}(W_i) - \rho \right) = 0, \quad i = 1, \ldots, n.$$

Z této rovnosti už jsme jednoduchou úpravou schopni odvodit požadované vyjádření pro w_i :

$$w_i = \frac{1}{n \left[1 + \lambda_2 \left(\widehat{V}(W_i) - \rho \right) \right]}, \quad i = 1, \ldots, n,$$

kde λ_2 je řešením rovnice, kterou získáme z (3.8) dosazením za w_i:

$$\sum_{i=1}^{n} \frac{\widehat{V}(W_i) - \rho}{1 + \lambda_2 (\widehat{V}(W_i) - \rho)} = 0. \tag{3.9}$$

Funkce $\prod_{i=1}^{n} w_i$ nabývá svého maxima pro $w_1 = \cdots = w_n = \frac{1}{n}$ za podmínky, že $\sum_{i=1}^{n} w_i = 1$. Zaveďme tak, podobným způsobem jako v (3.3), *dosazovaný empirický věrohodnostní poměr pro korelační koeficient* ρ vztahem:

$$\widehat{R}(\rho) = \prod_{i=1}^{n} n w_i = \prod_{i=1}^{n} \frac{1}{1 + \lambda_2 \left(\widehat{V}(W_i) - \rho \right)} \tag{3.10}$$

Úpravou výrazu (3.10) můžeme získat *dosazovaný empirický logaritmický věrohodnostní poměr pro korelační koeficient* ρ:

$$\widehat{l}(\rho) = -2\log \widehat{R}(\rho) = 2\sum_{i=1}^{n} \log\left[1 + \lambda_2\left(\widehat{V}(W_i) - \rho\right)\right]. \tag{3.11}$$

V následující větě zjistíme, k jakému rozdělení konverguje dosazovaný empirický logaritmický věrohodnostní poměr $\widehat{l}(\rho)$. Nejprve však zavedeme pomocnou konstantu A a následně lemma, které využijeme v důkazu věty.

Převezměme následující značení z článku Hu a kol. (2020)

$$\begin{aligned}
\sigma_0^2 &= \operatorname{var}\left(\frac{X-\mu_X}{\sigma_X}\frac{Y-\mu_Y}{\sigma_Y}\right),\\
\sigma_v^2 &= \operatorname{var}\left(\frac{X-\mu_X}{\sigma_X}\frac{Y-\mu_Y}{\sigma_Y} - \frac{\rho}{2}\left[\left(\frac{X-\mu_X}{\sigma_X}\right)^2 + \left(\frac{Y-\mu_Y}{\sigma_Y}\right)^2\right]\right),\\
A &= \frac{\sigma_0^2}{\sigma_v^2}.
\end{aligned}$$

Lemma 14. *Při zavedeném značení pro* σ_0^2, σ_v^2 *platí:*

$$\begin{aligned}
&a)\quad \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\widehat{V}(W_i) - \rho\right) \xrightarrow{D} \mathcal{N}(0,\sigma_v^2),\\
&b)\quad \frac{1}{n}\sum_{i=1}^{n}\left(\widehat{V}(W_i) - \rho\right)^2 \xrightarrow{P} \sigma_0^2.
\end{aligned}$$

Důkaz lemmatu 14 najdeme v článku Hu a kol. (2020, str. 35). Z definice σ_0^2 vidíme, že σ_0^2 je rozptylem náhodné veličiny $V(W)$ se střední hodnotou ρ. Tvar části a) lemma 14 vychází z centrální limitní věty a říká nám, že σ_v^2 je vlastně rozptyl náhodné veličiny $\widehat{V}(W)$, která má také střední hodnotu ρ. Z toho také plyne, že musíme zesílit předpoklad položený za začátku této sekce kapitoly. Nyní již musíme předpokládat existenci konečných druhých momentů náhodných veličin $V(W), \widehat{V}(W)$. To znamená, že pro náhodné veličiny X, Y je třeba předpokládat konečnost dokonce čtvrtých momentů.

Věta 15. *Platí*

$$A \cdot \widehat{l}(\rho) \xrightarrow{D} \chi_1^2,$$

Před samotným důkazem věty poznamenejme, že důkaz věty je uveden také v článku Hu a kol. (2020). Důkaz v této práci je proveden podrobněji, některé části jsou vysvětleny, chybějící kroky doplněny a některé pozměněny. Dále poznamenejme, že v důkazu využíváme vztah $\lambda_2 = \mathcal{O}_p(\frac{1}{\sqrt{n}})$, který je převzat z článku Owen (1988, str. 242). V tomto článku se při zjišťování vztahu $\lambda_2 = \mathcal{O}_p(\frac{1}{\sqrt{n}})$ pracuje pouze s jednou náhodnou veličinou X a její střední hodnotu, místo situace s náhodnou veličinou $\widehat{V}(W)$ a korelačním koeficientem v naší práci. V článku je použit předpoklad konečnosti třetích momentů pro onu náhodnou veličinu X. V našem případě je tak postačující předpoklad konečnosti třetích momentů náhodné veličiny $\widehat{V}(W)$, tedy předpoklad konečnosti až šestých momentů náhodných veličin X, Y. Shrnutím, z definice σ_0^2, σ_v^2 a následně A vidíme, že konečnost čtvrtých momentů náhodných veličin X, Y je předpokladem nutným, z předchozí úvahy pak, že konečnost jejich šestých momentů je předpokladem postačujícím. Je možné, že lze ukázat platnost vztahu $\lambda_2 = \mathcal{O}_p(\frac{1}{\sqrt{n}})$ pomocí konečnosti pouze pátých nebo dokonce oněch čtvrtých momentů náhodných veličin X, Y, to však není součástí této práce. Předpokládejme proto dále, že náhodný výběr $\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix}, \ldots, \begin{pmatrix} X_n \\ Y_n \end{pmatrix}$ pochází z rozdělení s konečnými šestými momenty. Nyní se přesuňme k avizovanému důkazu věty 15.

Důkaz. Zaveďme substituci $M_i = \lambda_2(\widehat{V}(W_i) - \rho)$. Maclaurinův rozvoj (Taylorův rozvoj se středem v bodě 0) funkce $\log(1 + M_i)$ je

$$\log(1 + M_i) = M_i - \frac{M_i^2}{2} + \frac{M_i^3}{3} - \ldots.$$

Aplikujeme-li Maclaurinův rozvoj stupně 2 na funkci $2\sum_{i=1}^n \log(1 + M_i)$, dostáváme

$$2\sum_{i=1}^n \log(1 + M_i) = 2\sum_{i=1}^n (M_i - \frac{M_i^2}{2}) + r_2,$$

kde r_2 je druhý zbytek Maclaurinova rozvoje funkce $2\sum_{i=1}^n \log(1 + M_i)$. Zpětnou substitucí dostáváme

$$\widehat{l}(\rho) = 2\sum_{i=1}^n \left(\lambda_2(\widehat{V}(W_i) - \rho) - \frac{[\lambda_2(\widehat{V}(W_i) - \rho)]^2}{2} \right) + r_2.$$

Pro zbytek r_2 existuje konstanta $K \in \mathbb{R}$ taková, že

$$|r_2| \leq K \sum_{i=1}^n \left| \lambda_2(\widehat{V}(W_i) - \rho) \right|^3 \leq K \; |\lambda_2|^3 \, n \frac{1}{n} \sum_{i=1}^n \left| \widehat{V}(W_i) - \rho \right|^3 = o_p(1).$$

Využili jsme, že $\lambda_2 = \mathcal{O}_p(\frac{1}{\sqrt{n}})$ z článku Owen (1988, str. 242) a Owen (1990, str. 98, lemma 3), které říká, že $\frac{1}{n}\sum_{i=1}^n \left| \widehat{V}(W_i) - \rho \right|^3 = o(\sqrt{n})$. Máme tedy nyní vztah

$$\widehat{l}(\rho) = 2\lambda_2 \sum_{i=1}^n (\widehat{V}(W_i) - \rho) - \lambda_2^2 \sum_{i=1}^n (\widehat{V}(W_i) - \rho)^2 + o_p(1). \tag{3.12}$$

Připomeňme si nyní rovnost (3.9):

$$\sum_{i=1}^n \frac{\widehat{V}(W_i) - \rho}{1 + \lambda_2(\widehat{V}(W_i) - \rho)} = 0.$$

Zaveďme substituci $M_i = \widehat{V}(W_i) - \rho$. Maclaurinův rozvoj funkce $\frac{M_i}{1+\lambda_2 M_i}$ je

$$\frac{M_i}{1 + \lambda_2 M_i} = M_i - \lambda_2 M_i^2 + \lambda_2^2 M_i^3 - \ldots.$$

Aplikujme Maclaurinův rozvoj stupně 2 na funkci $\sum_{i=1}^n \frac{M_i}{1+\lambda_2 M_i}$, tedy na levou stranu rovnosti vzniklé substitucí s M_i.

$$\sum_{i=1}^n M_i - \sum_{i=1}^n \lambda_2 M_i^2 + r_2 = 0.$$

Zpětnou substitucí za M_i dostáváme

$$\sum_{i=1}^{n}(\widehat{V}(W_i)-\rho)-\sum_{i=1}^{n}\lambda_2(\widehat{V}(W_i)-\rho)^2+r_2=0.$$

I tentokrát jsme schopni omezit zbytek r_2 :

$$|r_2|\le K\ \sum_{i=1}^{n}\lambda_2^2\left|\widehat{V}(W_i)-\rho\right|^3=K\ \lambda_2^2 n\frac{1}{n}\sum_{i=1}^{n}\left|\widehat{V}(W_i)-\rho\right|^3=o_p(\sqrt{n}),$$

přičemž jsme opět využili vztah $\lambda_2=\mathcal{O}_p(\frac{1}{\sqrt{n}})$ a článek Owen (1990, str. 98, lemma 3). Úpravou předchozí rovnosti a dosazením za r_2 dostáváme vyjádření pro λ_2 :

$$\lambda_2=\frac{\sum_{i=1}^{n}(\widehat{V}(W_i)-\rho)}{\sum_{i=1}^{n}(\widehat{V}(W_i)-\rho)^2}+\frac{o_p(\sqrt{n})}{\sum_{i=1}^{n}(\widehat{V}(W_i)-\rho)^2}. \tag{3.13}$$

Dosadíme pomocí vyjádření (3.13) za λ_2 ve vztahu (3.12). Dostáváme

$$\begin{aligned}\widehat{l}(\rho)&=2\frac{[\sum_{i=1}^{n}(\widehat{V}(W_i)-\rho)]^2}{\sum_{i=1}^{n}(\widehat{V}(W_i)-\rho)^2}+o_p(1)-\frac{[\sum_{i=1}^{n}(\widehat{V}(W_i)-\rho)]^2}{\sum_{i=1}^{n}(\widehat{V}(W_i)-\rho)^2}-o_p(1/n)+o_p(1)\\&=\frac{[\sum_{i=1}^{n}(\widehat{V}(W_i)-\rho)]^2}{\sum_{i=1}^{n}(\widehat{V}(W_i)-\rho)^2}+o_p(1).\end{aligned}$$

Využili jsme vztahu $\max_{1\le i\le n}\left|\widehat{V}(W_i)-\rho\right|=o(\sqrt{n})$ z článku Owen (1990, str. 98, lemma 3).

Našli jsme příhodné vyjádření pro $\widehat{l}(\rho)$. Nyní toto vyjádření pro $\widehat{l}(\rho)$ a vztah $A=\sigma_0^2/\sigma_v^2$ dosadíme do součinu $A\cdot\widehat{l}(\rho)$. Dostáváme

$$\begin{aligned}A\ \cdot\widehat{l}(\rho)&=\frac{\sigma_0^2}{\sigma_v^2}\frac{[\sum_{i=1}^{n}(\widehat{V}(W_i)-\rho)]^2}{\sum_{i=1}^{n}(\widehat{V}(W_i)-\rho)^2}+o_p(1)\\&=\frac{\sigma_0^2}{\frac{1}{n}\sum_{i=1}^{n}(\widehat{V}(W_i)-\rho)^2}\left[\frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\widehat{V}(W_i)-\rho)}{\sigma_v}\right]^2+o_p(1)\end{aligned}$$

Aplikací lemmatu 14 na tento tvar zjišťujeme, že první zlomek konverguje k 1 v pravděpodobnosti, čitatel zlomku s druhou mocninou konverguje v distribuci k normálnímu rozdělení s nulovou střední hodnotou a rozptylem σ_v^2 a $o_p(1)$ značí konvergenci k nule v pravděpodobnosti. Z posledního tvaru proto plyne, že

$$A\ \cdot\widehat{l}(\rho)\xrightarrow{D}\chi_1^2.$$

□

V praxi jsou nám parametry středních hodnot a směrodatných odchylek neznámé. Pro konstrukci intervalu spolehlivosti použitím věty 15 je třeba tyto neznámé parametry ve vyjádření pomocné konstanty A nahradit konzistentními odhady, tak aby asymptotické rozdělení součinu $A\cdot\widehat{l}(\rho)$ zůstalo zachováno. Výběrové průměry, výběrové rozptyly a výběrový korelační koeficient jsou konzistentními odhady parametrů středních hodnot, rozptylů a korelačního koeficientu, proto dosadíme $\overline{X},\overline{Y},S_X^2,S_Y^2,r$ do vyjádření parametrů σ_0^2,σ_v^2 (Hu a kol., 2020, str. 32). Dále použijeme vztah $varZ = \mathsf{E}\,(Z - \mathsf{E}\ Z)^2$ pro nějakou náhodnou veličinu Z. Označíme-li

$$A_{0,i} = \frac{X_i - \overline{X}}{S_X}\frac{Y_i - \overline{Y}}{S_Y}, \tag{3.14}$$

$$A_{v,i} = \frac{X_i - \overline{X}}{S_X}\frac{Y_i - \overline{Y}}{S_Y} - \frac{r}{2}\left[\left(\frac{X_i - \overline{X}}{S_X}\right)^2 + \left(\frac{Y_i - \overline{Y}}{S_Y}\right)^2\right], \tag{3.15}$$

dostáváme $\hat{\sigma}_0^2,\hat{\sigma}_v^2$, konzistentní odhady parametrů σ_0^2,σ_v^2 :

$$\hat{\sigma}_0^2 = \frac{1}{n}\sum_{i=1}^{n}\left(A_{0,i} - \frac{1}{n}\sum_{i=1}^{n}A_{0,i}\right)^2, \tag{3.16}$$

$$\hat{\sigma}_v^2 = \frac{1}{n}\sum_{i=1}^{n}\left(A_{v,i} - \frac{1}{n}\sum_{i=1}^{n}A_{v,i}\right)^2. \tag{3.17}$$

Z toho plyne, že

$$\widehat{A} = \frac{\hat{\sigma}_0^2}{\hat{\sigma}_v^2} \tag{3.18}$$

je konzistentním odhadem A. Můžeme tedy vyjádřit $100(1-\alpha)\%$ asymptotický interval spolehlivosti pro korelační koeficient ρ jako množinu

$$\left\{\rho : \widehat{A}\cdot\widehat{l}(\rho) \leq q_{1-\alpha}\right\},$$

kde $q_{1-\alpha}$ je $(1-\alpha)$-kvantil rozdělení χ_1^2. Řešíme soustavu rovnic $\widehat{A}\cdot\widehat{l}(\rho) = q_{1-\alpha}$ a (3.9) pro neznámé λ_2 a ρ. Podle článku Hall a La Scala (1990) má tato soustava dvě reálná řešení a ρ_1,ρ_2 která jsou součástí řešení soustavy tvoří krajní body našeho intervalu spolehlivosti. Soustavu řešíme numericky např. pomocí Newtonovy iterační metody. Jako počáteční aproximaci pro ρ lze použít výběrový korelační koeficient a jako počáteční aproximaci pro λ_2 lze využít první výraz pravé strany vztahu (3.13) z důkazu věty 15. Další počáteční aproximace získáme vhodnou úpravou těchto výrazů, pro účely příkladů v kapitole 4 měníme hodnotu počáteční aproximace pro ρ v krocích o velikosti 0.2 a počáteční aproximaci pro lambda měníme o 0.1 a to v obou případech kladným i záporným směrem. Pro příklady v této práci jsou tyto změny počátečních aproximací postačující pro nalezení obou krajních bodů intervalů spolehlivosti. Pokud by byla metoda použita na příkladu, kde by takové změny počátečních aproximací nevedly k nalezení obou krajních bodů, bylo by třeba velikost kroku pro změny počátečních aproximací vhodně korigovat.

Shrňme krátce popsaný postup pro zjišťování intervalu spolehlivosti pro korelační koeficient pomocí empirické věrohodnosti. Nejprve spočteme z dat výběrové průměry $\overline{X},\overline{Y}$, výběrové rozptyly S_X^2,S_Y^2 a výběrový korelační koeficient r. Pomocí těchto statistik spočteme hodnoty funkce $\widehat{V}(W_i) = \left(\frac{X_i-\overline{X}}{S_X}\frac{Y_i-\overline{Y}}{S_Y}\right)$ pro $i = 1,\ldots,n$, které potřebujeme pro výpočet $\widehat{l}(\rho)$ i pro výpočet $\widehat{A}$. Spočteme odhad $\widehat{A}$ pomocné konstanty A podle vztahů (3.14) až (3.18). Pak $100(1-\alpha)\%$ asymptotický interval spolehlivosti tvoří množina takových ρ, která splňují nerovnost $\widehat{A}\cdot\widehat{l}(\rho) \leq q_{1-\alpha}$, kde $q_{1-\alpha}$ je $(1-\alpha)$-kvantil rozdělení χ_1^2 a přičemž $\widehat{l}(\rho)$

splňuje (3.11) a λ_2 je řešením rovnice (3.9). Tuto soustavu dvou rovnic o dvou neznámých řešíme numericky např. pomocí Newtonovy iterační metody a měníme počáteční aproximace dokud nenajdeme dvě reálná řešení této soustavy. Nakonec, ρ_1, ρ_2 z takových řešení tvoří krajní body hledaného intervalu spolehlivosti.

3.2 Konstrukce bez pomocné konstanty

V první části kapitoly jsme ukázali, jak konstruovat intervaly spolehlivosti pro korelační koeficient pomocí empirické věrohodnosti. Zjistili jsme, že je třeba pro určení intervalu spolehlivosti spočíst $\widehat{A}$, odhad hodnoty pomocné konstanty A. Ponechme i pro druhou část kapitoly stejné předpoklady jako pro část první. V této části kapitoly si ukážeme úpravu, s jejíž pomocí získáme obdobným postupem interval spolehlivosti pro korelační koeficient, ovšem bez nutnosti počítat odhad $\widehat{A}$.

Zaveďme *funkci vlivu*

$$V_I(W_i, \rho) = (V(W_i) - \rho) - \frac{\rho}{2}\left[\left(\frac{X_i - \mu_X}{\sigma_X}\right)^2 - 1 + \left(\frac{Y_i - \mu_Y}{\sigma_Y}\right)^2 - 1\right].$$

Tato úprava funkce tvaru $(V(W_i) - \rho)$ ve vyjádření $L_0(\rho)$ zajistí, že nebude třeba použít pomocnou konstantu. Postup je dále podobný jako v první části.

Definice 8. *Označme*

$$L_I(\rho) = \sup_{\boldsymbol{w}}\Big\{\prod_{i=1}^n w_i : w_i \geq 0, \sum_{i=1}^n w_i = 1, \sum_{i=1}^n w_i V_I(W_i, \rho) = 0\Big\}. \tag{3.19}$$

Funkci $L_I(\rho)$ budeme nazývat empirickou věrohodností s funkcí vlivu pro korelační koeficient *ρ.*

Stejně jako v předchozí části kapitoly dosadíme výběrové průměry a výběrové rozptyly za příslušné parametry ve vyjádření funkce $V_I(W_i, \rho)$. Dostáváme vyjádření

$$\widehat{V}_I(W_i, \rho) = (\widehat{V}(W_i) - \rho) - \frac{\rho}{2}\left[\left(\frac{X_i - \overline{X}}{S_X}\right)^2 - 1 + \left(\frac{Y_i - \overline{Y}}{S_Y}\right)^2 - 1\right].$$

Dosazením za $V_I(W_i, \rho)$ ve vyjádření $L_I(\rho)$ získáme

$$\widehat{L}_I(\rho) = \sup_{\boldsymbol{w}}\Big\{\prod_{i=1}^n w_i : w_i \geq 0, \sum_{i=1}^n w_i = 1, \sum_{i=1}^n w_i \widehat{V}_I(W_i, \rho) = 0\Big\}.$$

$\widehat{L}_I(\rho)$ budeme nazývat *dosazovanou empirickou věrohodností s funkcí vlivu pro korelační koeficient ρ.* Dále identickým postupem jako v první části kapitoly zjistíme metodou Lagrangeových multiplikátorů, že

$$w_i = \frac{1}{n\left[1 + \lambda_I \widehat{V}_I(W_i, \rho)\right]}, \quad i = 1, \ldots, n,$$

kde λ_I je řešením rovnice

$$\sum_{i=1}^{n} \frac{\widehat{V}_I(W_i,\rho)}{1+\lambda_I \widehat{V}_I(W_i,\rho)} = 0. \tag{3.20}$$

Opět, podobně jako v první části kapitoly, zavedeme vztah

$$\widehat{l}_I(\rho) = 2\sum_{i=1}^{n} \log\left[1+\lambda_I \widehat{V}_I(W_i,\rho)\right]. \tag{3.21}$$

Funkci $\widehat{l}_I(\rho)$ budeme říkat *dosazovaný empirický logaritmický věrohodnostní poměr s funkcí vlivu* pro ρ.

Věta 16 (Hu a kol. (2020), str.32)**.** *Při zavedeném značení platí*

$$\widehat{l}_I(\rho) \xrightarrow{D} \chi_1^2.$$

Jak jsme avizovali na začátku této sekce, není třeba odhadovat hodnotu pomocné konstanty. Proto přímo z věty získáváme předpis pro $100(1-\alpha)\%$ asymptotický interval spolehlivosti pro korelační koeficient:

$$\left\{\rho : \widehat{l}_I(\rho) \leq q_{1-\alpha}\right\}, \tag{3.22}$$

kde $q_{1-\alpha}$ opět značí $(1-\alpha)$-kvantil rozdělení χ_1^2.

Postup pro konstrukci je tedy velmi podobný jako v první sekci kapitoly. Nejprve spočteme výběrové průměry a výběrové rozptyly a dosadíme vypočtené hodnoty do vyjádření funkce $\widehat{V}_I(W_i,\rho)$. $100(1-\alpha)\%$ asymptotický interval spolehlivosti pak je množinou z (3.22), kde $\widehat{l}_I(\rho)$ splňuje (3.21) a kde λ_I splňuje rovnost (3.20). Soustavu rovnic $\widehat{l}_I(\rho) = q_{1-\alpha}$ a (3.20) pro neznámé ρ a λ_I opět řešíme numericky, například pomocí Newtonovy iterační metody, obdobně jako v případě empirické věrohodnosti v první části kapitoly.

4. Simulace

V této kapitole vyzkoušíme metody popsané v práci na několika příkladech a porovnáme jak si jednotlivé metody vedly. K výpočtům byl použit program Wolfram Mathematica 12.0.

Nejprve popišme jakým způsobem bude simulace probíhat. Ve všech příkladech volíme hodnotu $\alpha = 0.05$, tedy budeme konstruovat 95% intervaly spolehlivosti pro korelační koeficient, ve zbytku kapitoly proto vždy, když budeme zmiňovat intervaly spolehlivosti, budeme tím myslet právě 95% intervaly spolehlivosti pro korelační koeficient. Na začátku každého příkladu bude specifikováno s jakými rozsahy n náhodného výběru, s jakým pravděpodobnostním rozdělením a jakými hodnotami korelačního koeficientu ρ budeme pracovat. V každém příkladu a pro každou těchto volbu parametrů vygenerujeme 1000 náhodných výběrů. Z nich pomocí každé metody spočteme 1000 intervalů spolehlivosti pro korelační koeficient. Pro tyto intervaly spočteme jejich průměrnou délku $\overline{d}$:

$$\overline{d} = \frac{\sum_{i=1}^{n}(\rho_{U,i} - \rho_{L,i})}{1000},$$

kde $\rho_{U,i}$ značí horní a $\rho_{L,i}$ značí spodní hranici i-tého intervalu spolehlivosti. A spočteme *opp*, odhad pravděpodobnosti pokrytí skutečné hodnoty korelačního koeficientu:

$$opp = \frac{\sum_{i=1}^{n} \mathbf{1}\left\{\rho_{L,i} \leq \rho \leq \rho_{U,i}\right\}}{1000},$$

kde ρ je skutečná hodnota korelačního koeficientu. Výsledné hodnoty pro jednotlivé metody zapíšeme do tabulek, metody porovnáme a budeme pozorovat zda a jak moc se odhady pravděpodobností pokrytí liší od hodnoty 0.95. Poznamenejme, že v této kapitole budeme metodu Fisherovy z-transformace označovat zkratkou FZ, metodu zobecněných pivotů ZP, metodu empirické věrohodnosti EV a metodu empirické věrohodnosti s funkcí vlivu zkratkou EVFV.

4.1 Příklad 1

V prvním příkladu pochází náhodný výběr z dvojrozměrného normálního rozdělení s nulovým vektorem středních hodnot a kovarianční maticí $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. Tedy

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}_2 \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right].$$

Volíme hodnoty rozsahu náhodného výběru $n = 30, 90, 150$ a hodnoty korelačního koeficientu $\rho = 0.05, -0.5, 0.9$. Následují dvě tabulky. Tabulka 4.1 obsahuje odhady pravděpodobností pokrytí skutečného korelačního koeficientu pro jednotlivé metody a různé volby parametrů n a ρ. Druhá tabulka obsahuje hodnoty průměrných délek intervalů spolehlivosti opět pro jednotlivé metody a pro různé volby parametrů n a ρ.

Z tabulky 4.1 vidíme, že metoda Fisherovy z-transformace a metoda zobecněných pivotů mají hodnoty odhadů *opp* velice podobné pro zvolená data. Dokonce jsou výsledky metody Fisherovy z-transformace nepatrně lepší. Hodnoty

ρ		0.05			-0.5			0.9	
n	30	90	150	30	90	150	30	90	150
FZ	0.944	0.960	0.943	0.960	0.939	0.956	0.961	0.948	0.953
ZP	0.945	0.961	0.939	0.959	0.941	0.955	0.961	0.949	0.954
EV	0.875	0.936	0.923	0.912	0.919	0.949	0.947	0.946	0.959
EVFV	0.912	0.949	0.933	0.935	0.922	0.956	0.920	0.942	0.951

Tabulka 4.1: Odhady pravděpodobností pokrytí, příklad 1

ρ		0.05			-0.5			0.9	
n	30	90	150	30	90	150	30	90	150
FZ	0.697	0.409	0.318	0.550	0.311	0.241	0.153	0.081	0.062
ZP	0.688	0.407	0.317	0.547	0.311	0.241	0.156	0.082	0.062
EV	0.669	0.414	0.320	0.500	0.303	0.239	0.125	0.076	0.059
EVFV	0.629	0.398	0.312	0.496	0.303	0.239	0.137	0.079	0.061

Tabulka 4.2: Průměrné délky intervalů spolehlivosti, příklad 1

odhadů pravděpodobností pokrytí se u obou metod se poměrně stabilně pohybují okolo požadovaných 0.95. Odhady pravděpodobností pokrytí metodou empirické věrohodnosti jsou pro volbu korelačního koeficientu $\rho = 0.05$ nižší. Pro rostoucí rozsah výběru n se sice hodnota odhadu postupně zvyšuje, nicméně i pro $n = 150$ je stále nižší než u zbylých metody. Při výraznější absolutní hodnotě korelačního koeficientu se tato metoda postupně zlepšuje a pro $\rho = 0.9$ už můžeme označit výsledky dosažené metodou empirické věrohodnosti za velmi dobré. Metoda empirické věrohodnosti s funkcí vlivu vykazuje pro $\rho = 0.05$ a $\rho = -0.5$ vyšší hodnoty než metoda EV. Pro rozsah výběru $n = 150$ se dostává na srovnatelnou úroveň s metodami FZ a ZP.

Tabulka 4.2 znázorňuje průměrné délky intervalů spolehlivosti. Vidíme, že metoda Fisherovy z-transformace a metoda zobecněných pivotů opět vykazují velmi podobné výsledky. Intervaly spolehlivosti spočtené metodou empirické věrohodnosti mají pro 7 z 9 voleb ρ a n průměrnou délku menší než je tomu u předchozích dvou metod. Metoda empirické věrohodnosti s funkcí vlivu má průměrné délky nižší pro volbu $\rho = 0.05$, ale naopak vyšší pro $\rho = 0.9$ než metoda empirické věrohodnosti. Pro $\rho = -0.5$ měly obě tyto metody průměrné délky intervalů spolehlivosti téměř identické. Můžeme si všimnout, že pro zvyšující se rozsah výběru a zvyšující se absolutní hodnotu korelačního koeficientu jsou si odhady i průměrné délky pro jednotlivé metody obecně stále bližší. Rovněž můžeme pozorovat, že průměrná délka intervalu spolehlivosti se zkracuje při zvyšující absolutní hodnotě korelačního koeficientu a pro zvyšující se rozsah výběru.

4.2 Příklad 2

Také v příkladu 2 vyzkoušíme jednotlivé metody na náhodném výběru pocházejícím z dvojrozměrného normálního rozdělení, tentokrát s nenulovým vektorem středních hodnot a různými rozptyly. Náhodný vektor $(X,Y)^\top$ v příkladu 2 má

ρ		0.05			-0.5			0.9	
n	30	90	150	30	90	150	30	90	150
FZ	0.944	0.949	0.940	0.945	0.955	0.954	0.948	0.954	0.949
ZP	0.945	0.950	0.940	0.944	0.954	0.954	0.947	0.957	0.948
EV	0.862	0.916	0.937	0.896	0.946	0.952	0.934	0.962	0.948
EVFV	0.908	0.936	0.943	0.910	0.933	0.951	0.906	0.944	0.947

Tabulka 4.3: Odhady pravděpodobností pokrytí, příklad 2

ρ		0.05			-0.5			0.9	
n	30	90	150	30	90	150	30	90	150
FZ	0.697	0.409	0.318	0.551	0.313	0.241	0.157	0.083	0.062
ZP	0.687	0.407	0.317	0.547	0.313	0.241	0.160	0.083	0.062
EV	0.677	0.415	0.323	0.503	0.305	0.238	0.129	0.077	0.060
EVFV	0.634	0.399	0.314	0.499	0.304	0.238	0.141	0.081	0.062

Tabulka 4.4: Průměrné délky intervalů spolehlivosti, příklad 2

rozdělení:

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}_2 \left[\begin{pmatrix} 2 \\ 10 \end{pmatrix}, \begin{pmatrix} 4 & 6\rho \\ 6\rho & 9 \end{pmatrix} \right].$$

Volby parametrů rozsahu náhodného výběru n a korelačního koeficientu ρ ponechme stejné jako v příkladu 1.

Na základě tabulek 4.3 a 4.4 můžeme metody hodnotit velmi podobně jako v příkladu 1. Metody Fisherovy z-transformace a zobecněných pivotů mají opět velmi podobné výsledky, s tím, že odhad pravděpodobnosti pokrytí je u intervalů pomocí metody Fisherovy z-transformace nepatrně vyšší. Na druhou stranu mají intervaly metodou zobecněných pivotů nepatrně menší průměrnou délku pro $\rho = 0.05$ i pro $\rho = -0.5$. Metoda empirické věrohodnosti má opět většinou menší průměrnou délku intervalů spolehlivosti, ale znovu má také pro hodnoty $\rho = 0.05$ a $\rho = -0.5$ znatelně nižší hodnoty odhadů v tabulce 4.3. Pro $\rho = 0.9$ už jsou její hodnoty odhadu pravděpodobnosti pokrytí opět podobné jako u zbylých metod. Na základě tabulek z příkladu 1 a 2 lze říci, že pro náhodný výběr pocházející z normálního rozdělení vypadají metoda Fisherovy z-transformace a metoda zobecněných pivotů spolehlivěji než metoda empirické věrohodnosti. Zvláště pak v případě menšího rozsahu výběru a nižší absolutní hodnoty korelačního koeficientu. Metoda empirické věrohodnosti s funkcí vlivu vykazuje znatelně lepší výsledky pro volbu $\rho = 0.05$ než metoda empirické věrohodnosti. Při rozsahu výběru $n = 150$ jsou výsledky všech čtyř metod srovnatelné pro všechny volby ρ.

4.3 Příklad 3

Nyní vyzkoušíme metody na náhodném výběru z dvojrozměrného Poissonova rozdělení s parametry $\lambda_1, \lambda_2, \lambda_{12}$. Víme, že pro metodu Fisherovy z-transformace a metodu zobecněných pivotů již není splněn předpoklad normálního rozdělení. Přesto je použijme také v tomto a dalších příkladech s nenormálními daty a

ρ		0.1			0.5			0.9	
n	30	90	150	30	90	150	30	90	150
FZ	0.946	0.934	0.941	0.934	0.930	0.932	0.930	0.937	0.944
ZP	0.947	0.934	0.942	0.933	0.929	0.933	0.929	0.936	0.944
EV	0.875	0.901	0.930	0.886	0.918	0.931	0.934	0.956	0.954
EVFV	0.907	0.926	0.938	0.910	0.926	0.926	0.907	0.942	0.946

Tabulka 4.5: Odhady pravděpodobností pokrytí, příklad 3

ρ		0.1			0.5			0.9	
n	30	90	150	30	90	150	30	90	150
FZ	0.692	0.406	0.315	0.547	0.312	0.241	0.158	0.083	0.062
ZP	0.683	0.404	0.315	0.544	0.312	0.241	0.161	0.083	0.062
EV	0.681	0.421	0.331	0.523	0.326	0.256	0.136	0.083	0.063
EVFV	0.633	0.399	0.318	0.514	0.321	0.253	0.149	0.086	0.065

Tabulka 4.6: Průměrné délky intervalů spolehlivosti, příklad 3

pozorujme jak si tyto metody pro různá rozdělení vedou. Zvolme $\lambda_1 = \lambda_2 = 1$. Pak korelační koeficient je roven $\rho = \frac{\lambda_{12}}{1+\lambda_{12}}$. Opět zvolme tři různé hodnoty korelačního koeficientu $\rho = 0.1, 0.5, 0.9$. Jim příslušné volby parametru λ_{12} jsou $\lambda_{12} = \frac{1}{9}, 1, 9$. Volby rozsahu výběru n ponechme stejné jako v předchozích příkladech.

Výsledky simulace příkladu 3 jsou obsaženy v tabulkách 4.5 a 4.6. Jak jsme již zmínili, předpoklad normálního rozdělení pro metody FZ a ZP je porušen. Přesto si obě tyto metody vedli na datech s Poissonovým rozdělením velmi dobře. V případě $\rho = 0.1$ a $\rho = 0.5$ dokonce lépe než obě metody založené na empirické věrohodnosti. V případě $\rho = 0.9$ vykazuje nejlepší výsledky metoda EV. Její výsledky jsou ovšem pro $\rho = 0.1$ naopak nejslabší. Opět můžeme pozorovat, že průměrná délka intervalů spolehlivosti klesá při zvyšující se hodnotě korelačního koeficientu a zvyšujícím se rozsahu výběru.

4.4 Příklad 4

V příkladu 4 opět vyzkoušíme metody na datech z jiného než normálního rozdělení. Tentokrát půjde o dvojrozměrné rovnoměrné rozdělení složené ze dvou spojitých rovnoměrných rozdělení $R(0, 1)$ a s korelačním koeficientem ρ. Hodnoty korelačního koeficientu ρ a rozsahu výběru n ponechme stejné jako v příkladu 3. Každý prvek náhodného výběru generujeme pomocí algoritmu, který je podrobněji odvozen v článku Demirtas (2014, str. 3576). Nejprve spočteme

$$a = -\frac{5}{2} + \frac{1}{2}\sqrt{\frac{\rho + 49}{\rho + 1}}.$$

Poté generujeme náhodné veličiny

$$X \sim R(0, 1), \quad V \sim R(0, 1), \quad U \sim Beta(a, 1).$$

ρ		0.1			0.5			0.9	
n	30	90	150	30	90	150	30	90	150
FZ	0.942	0.933	0.945	0.906	0.891	0.892	0.574	0.591	0.612
ZP	0.944	0.935	0.945	0.907	0.890	0.894	0.564	0.592	0.609
EV	0.901	0.925	0.945	0.907	0.934	0.944	0.776	0.872	0.910
EVFV	0.932	0.928	0.948	0.920	0.942	0.950	0.740	0.887	0.916

Tabulka 4.7: Odhady pravděpodobností pokrytí, příklad 4

ρ		0.1			0.5			0.9	
n	30	90	150	30	90	150	30	90	150
FZ	0.691	0.405	0.315	0.540	0.310	0.240	0.141	0.080	0.062
ZP	0.682	0.403	0.314	0.536	0.309	0.240	0.143	0.080	0.062
EV	0.682	0.411	0.320	0.590	0.357	0.281	0.225	0.162	0.132
EVFV	0.661	0.405	0.318	0.583	0.357	0.280	0.241	0.171	0.138

Tabulka 4.8: Průměrné délky intervalů spolehlivosti, příklad 4

Nakonec, pokud $V < 0.5$, pak položíme

$$Y = |U - X|,$$

jinak položíme

$$Y = 1 - |1 - U - X|.$$

Spočtené odhady pravděpodobností pokrytí a průměrné délky intervalů spolehlivosti jsou obsaženy v tabulkách 4.7 a 4.8. Pro metody FZ a ZP ani tentokrát není splněn předpoklad dvojrozměrného normálního rozdělení. Přesto si obě vedou pro volbu $\rho = 0.1$ velmi dobře. Při této volbě ρ a rozsahu výběru $n = 90$ a $n = 150$ jsou výsledky všech čtyř metod velmi podobné. Pro hodnotu $\rho = 0.5$ vykazují metody EV a EVFV velmi dobré výsledky opět pro rozsah výběru $n = 90$, ale především pro rozsah výběru $n = 150$ při kterém metoda EVFV dosahuje hranice 0.95. Naopak odhady pravděpodobností pokrytí metodami FZ a ZP se, pro všechny tři volby n, pohybují pouze kolem hodnoty 0.9. Pro volbu $\rho = 0.9$ již naplno vidíme selhání metod FZ a ZP, jejichž odhady pravděpodobností pokrytí se pohybují kolem hodnoty 0.6. I metody EV a EVFV produkují slabší výsledky než pro předchozí volby ρ, vývoj hodnot odhadů pravděpodobností pokrytí však má výrazně stoupavou tendenci pro rostoucí n a pro $n = 150$ je tato hodnota u EVFV již 0.916. Z tabulky 4.8 lze pozorovat, že průměrné délky intervalů spolehlivosti jsou pro jednotlivé metody podobné při volbě $\rho = 0.1$. Pro rostoucí ρ se průměrné délky intervalů spolehlivosti pomocí metod FZ a ZP výrazně zkracují oproti intervalům spolehlivosti metod EV a EVFV. I to jistě přispívá ke stále výraznějšímu selhání metod FZ a ZP, jejichž intervaly spolehlivosti vykazují velmi podobnou průměrnou délku jako v předchozích příkladech a nepřizpůsobují svou délku rozdělení dat. Na příkladu 4 jsme se tak ukázali situaci, kdy metoda Fisherovy z-transformace a metoda zobecněných pivotů vykazují velmi slabé výsledky. Docházíme tak k závěru, že pro data z nenormálního rozdělení je skutečně vhodnější použití metody empirické věrohodnosti s funkcí vlivu, případně metody empirické věrohodnosti.

	$R(-1{,}1)$ a $R(0,2)$			χ_1^2 a χ_2^2			$\mathcal{N}(0,1)$ a $Po(1)$		
n	30	90	150	30	90	150	30	90	150
FZ	0.954	0.955	0.957	0.944	0.945	0.950	0.938	0.940	0.951
ZP	0.953	0.955	0.957	0.948	0.946	0.951	0.939	0.941	0.953
EV	0.912	0.948	0.954	0.767	0.857	0.904	0.858	0.913	0.929
EVFV	0.939	0.953	0.959	0.870	0.902	0.927	0.902	0.930	0.937

Tabulka 4.9: Odhady pravděpodobností pokrytí, příklad 5

	$R(-1,1)$ a $R(0,2)$			χ_1^2 a χ_2^2			$\mathcal{N}(0,1)$ a $Po(1)$		
n	30	90	150	30	90	150	30	90	150
FZ	0.698	0.410	0.319	0.697	0.410	0.319	0.698	0.409	0.318
ZP	0.689	0.408	0.318	0.688	0.408	0.318	0.688	0.408	0.318
EV	0.679	0.409	0.318	0.599	0.396	0.314	0.674	0.413	0.326
EVFV	0.658	0.404	0.315	0.546	0.362	0.292	0.629	0.395	0.315

Tabulka 4.10: Průměrné délky intervalů spolehlivosti, příklad 5

4.5 Příklad 5

V příkladu 5 budeme uvažovat tři různé dvojice nezávislých náhodných veličin. První jsou dvě nezávislé náhodné veličiny obě se spojitým rovnoměrným rozdělením, konkrétně $R(-1,1)$ a $R(0,2)$. Druhou dvojici tvoří náhodná veličina s rozdělením χ_1^2 a náhodná veličina s rozdělením χ_2^2. Poslední dvojicí je náhodná veličina s normovaným normálním rozdělením $\mathcal{N}(0,1)$ a náhodná veličina s Poissonovým rozdělením $Po(1)$. V každé dvojici jsou náhodné veličiny nezávislé, korelační koeficienty jsou tedy nulové. Pro každou dvojici náhodných veličin vygenerujeme tři náhodné výběry opět o rozsazích $n = 30, 90, 150$.

Tabulka 4.9 obsahuje hodnoty odhadů pravděpodobnosti pokrytí korelačního koeficientu $\rho = 0$ pro jednotlivé dvojice nezávislých náhodných výběrů a tabulka 4.10 obsahuje průměrné délky spočtených intervalů spolehlivosti zjištěných čtyřmi námi zkoumanými metodami. Z tabulky 4.10 můžeme pozorovat, že průměrné délky intervalů jsou pro metody FZ a ZP opět velmi podobné jako v předchozích příkladech. Intervalu spolehlivosti pomocí metod EV a EVFV vykazují průměrné délky kratší pro nezávislé náhodné výběry z rozdělení χ_1^2 a χ_2^2. Je však rovněž zřejmé, že odhady pravděpodobností pokrytí jsou v případě metody empirické věrohodnosti právě u této druhé kombinace nižší než bychom požadovali a to především pro menší rozsahy výběru. Ve shrnutí, vidíme, že metody FZ a ZP si i přes porušení předpokladu normality vedly velmi dobře. Metody EV a EVFV vykazují velmi dobré výsledky pro kombinaci dvou nezávislých rovnoměrných rozdělení, stále dobře si vedly pro normální a Poissonovo rozdělení. V případě dvou χ^2 rozdělení by bylo vhodné zvolit, obzvlášť v případě metody empirické věrohodnosti, rozsah výběru ještě větší než $n = 150$, pokud chceme aby se odhad pravděpodobnosti pokrytí přiblížil hodnotě 0.95. Závěrem poznamenejme, že metoda empirické věrohodnosti s funkcí vlivu vykázala pro všechny tři kombinace lepší výsledky, než metoda empirické věrohodnosti a to i přes kratší průměrné délky intervalů spolehlivosti.

Závěr

V této bakalářské práci jsme se zabývali metodami pro konstrukci intervalů spolehlivosti pro korelační koeficient. V první kapitole jsme si připomenuli základní vlastnosti korelačního koeficientu a stručně jsme si představili metodu Fisherovy z-transformace, která je metodou nejběžněji používanou.

V druhé kapitole jsme se zabývali metodou založenou na zobecněných pivotech. Nejprve jsme si zavedli lineární regresní model, zjistili jsme pomocí jakých parametrů tohoto lineární regresního modelu lze složit korelační koeficient a našli jsme pravděpodobnostní rozdělení pro funkce těchto parametrů a jejich odhadů. V druhé části kapitoly jsme zavedli pojem zobecněný pivot. Popsali jsme obecný postup pro konstrukci intervalů spolehlivosti pomocí zobecněných pivotů a tento postup jsme ukázali na jednoduchém příkladu. Poté jsme jej převedli, i díky zjištěním z první části kapitoly, na problém intervalu spolehlivosti pro korelační koeficient. Na závěr jsme postup, jakým se interval spolehlivosti metodou zobecněných pivotů počítá, stručně shrnuli.

Ve třetí kapitole jsme se věnovali metodám založeným na empirické věrohodnosti. Nejprve jsme zavedli pojem empirická věrohodnost a další pojmy obecně, poté jsme definovali empirickou věrohodnost pro korelační koeficient. Pomocí metody Lagrangeových multiplikátorů jsme získali nové rovnice a následně po zadefinování pojmu dosazovaný empirický logaritmický věrohodnostní poměr jsme dospěli ke klíčové větě o konvergenci v distribuci k chi-kvadrát rozdělení s jedním stupněm volnosti. Větu jsme si v práci také dokázali. Na základě této věty jsme sestavili předpis množiny, která je intervalem spolehlivosti pro korelační koeficient. Následoval popis jak z předpisu dojít přímo k vyjádření intervalu spolehlivosti pro korelační koeficient ve standardním tvaru. V druhé části kapitoly jsme, již stručněji, prošli celý postup znovu, tentokrát s úpravou, díky níž jsme se zbavili nutnosti odhadovat pomocnou konstantu při konstrukci intervalu spolehlivosti. Tuto upravenou metodu jsme dále v práci nazývali metoda empirické věrohodnosti s funkcí vlivu.

V závěrečné kapitole jsme se věnovali simulacím a porovnávání v práci probraných metod. První dva příklady pracovaly s daty z dvojrozměrného normálního rozdělení. Z hodnot obsažených v tabulkách prvních dvou příkladů bylo patrné, že metoda Fisherovy z-transformace a metoda zobecněných pivotů se jeví vhodnější na použití v případě, že náhodný výběr pochází z dvojrozměrného normálního rozdělení, než metody založené na empirické věrohodnosti. I jejich výsledky ovšem byly velmi dobré. V příkladu 3 jsme aplikovali metody na náhodné výběry z dvojrozměrného Poissonova rozdělení. Pozorovali jsme, že metody pracují dobře i v případě diskrétního rozdělení a také, že, i přes porušení předpokladu normality, si Fisherova z-transformace a metoda zobecněných pivotů vedly velmi dobře. V příkladu 4, s daty z dvojrozměrného spojitého rovnoměrného rozdělení, jsme pozorovali výrazné selhání metody Fisherovy z-transformace a metody zobecněných pivotů pro vyšší hodnoty korelačního koeficientu. Obě metody založené na empirické věrohodnosti vykazovaly v případě korelačního koeficientu 0.9 a rozsahu výběru 30 také slabších výsledků, ale pro větší rozsahy výběru se, na rozdíl od zbylých dvou metod, výrazně zlepšily. V příkladu 5 jsme si na závěr ukázali, jak metody pracují pro dvojice nezávislých náhodných výběrů z různých rozdělení.

V průběhu jsme pozorovali, že ve většině příkladů se metoda empirické věrohodnosti s funkcí vlivu jevila lepší než metoda empirické věrohodnosti. Celkově proto docházíme, na základě výsledků těchto simulací, k závěru, že na náhodné výběry nepocházející z normálního rozdělení, nebo pokud rozdělení neznáme, je vhodné použít právě metodu empirické věrohodnosti s funkcí vlivu.

Seznam použité literatury

Anděl, J. (2011). *Základy matematické statistiky.* Třetí vydání. Matfyzpress, Praha. ISBN 978-80-7378-162-0.

Demirtas, H. (2014). Generating bivariate uniform data with a full range of correlations and connections to bivariate binary data. *Communications in Statistics - Theory and Methods*, **43**(17), 3574–3579.

Flynn (1999). Asymptotic results for the linear regression model.

Hall, P. a La Scala, B. (1990). Methodology and algorithms of empirical likelihood. *International Statistical Review*, **58**(2), 109–127.

Hanning, J., Iyer, H. K. a Patterson, P. (2006). Fiducial generalized confidence intervals. *Journal of the American Statistical Association*, **101**, 254–269.

Hu, X., Jung, A. a Qin, G. (2020). Interval estimation for the correlation coefficient. *The American Statistician*, **74**, 29–36.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, **75**, 237–249.

Owen, A. B. (1990). Empirical likelihood ratio confidence regions. *Biometrika*, **18**, 90–120.

Wang, C.-M. J. a Iyer, H. (2005). Propagation of uncertainty in measurement using generalized inference. *Metrologia*, **42**, 145–153.

Weerahandi, S. (1993). Generalized confidence intervals. *Journal of the American Statistical Association*, **88**, 899–905.

Winterbottom, A. (1979). A note on the derivation of fisher's transformation of the correlation coefficient. *The American Statistician*, **33**, 142—-143.

Seznam tabulek