



**MATEMATICKO-FYZIKÁLNÍ
FAKULTA**
Univerzita Karlova

DIPLOMOVÁ PRÁCE

Šárka Horská

Testy nezávislosti pro funkcionální data

Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: doc. RNDr. Zdeněk Hlávka, Ph.D.

Studijní program: Matematika

Studijní obor: Pravděpodobnost, matematická
statistika a ekonometrie

Praha 2023

Prohlašuji, že jsem tuto diplomovou 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

Ráda bych poděkovala svému vedoucímu doc. RNDr. Zdeňkovi Hlávkovi, Ph.D. za cenné připomínky a rady při vedení diplomové práce a za zasvěcení do práce se školním výpočetním clusterem. Dále bych ráda poděkovala prof. RNDr. Marii Huškové, DrSc. za přínosné konzultace. Nakonec děkuji svým rodičům, kteří mě po celou dobu studia podporovali.

Název práce: Testy nezávislosti pro funkcionální data

Autor: Šárka Horská

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: doc. RNDr. Zdeněk Hlávka, Ph.D., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: Tato práce se zabývá testy sériové nezávislosti pro časové řady funkcionálních dat. V první části práce je představena problematika sériové nezávislosti na časových řadách náhodných veličin. Druhá část už je zaměřena na testy sériové nezávislosti funkcionálních pozorování. Věnuje se testu založenému na autokorelaci a především testu, jehož testová statistika je odvozena přímo z definice nezávislosti. Tento test je modifikován na test se slabší alternativou sub-závislosti. V závěru práce jsou tyto tři testy sériové nezávislosti porovnány na základě simulací.

Klíčová slova: test nezávislosti, sériová závislost, funkcionální data, časové řady

Title: Tests of independence for functional data

Author: Šárka Horská

Department: Department of Probability and Mathematical Statistics

Supervisor: doc. RNDr. Zdeněk Hlávka, Ph.D., Department of Probability and Mathematical Statistics

Abstract: This thesis deals with tests of serial independence for functional time series. The first part of the thesis introduces the issue of serial independence in time series of random variables. The second part focuses on tests of serial independence for functional observations. It examines a test based on autocorrelation and, in particular, a test whose test statistic is derived directly from the definition of independence. This test is modified to a test with a weaker alternative of sub-dependence. The thesis concludes with a comparison of these three tests, which is based on a simulation.

Keywords: independence test, serial dependence, functional data, time series

Obsah

Úvod	2
1 Nezávislost náhodných veličin	3
1.1 Zavedení pojmů	3
1.2 Testy nezávislosti složek vektorů	6
1.2.1 Test nezávislosti odvozený z definice nezávislosti	7
1.2.2 Test nezávislosti odvozený z definice sub-nezávislosti	8
1.3 Testy sériové nezávislosti jednorozměrné časové řady	10
1.3.1 Test založený na autokorelaci	10
1.3.2 Test odvozený z definice nezávislosti	12
1.3.3 Test odvozený z definice sub-nezávislosti	13
2 Nezávislost funkcionálních dat	14
2.1 Testy sériové nezávislosti funkcionálních dat	15
2.1.1 Test založený na autokorelaci	15
2.1.2 Test odvozený z definice nezávislosti	16
2.1.3 Test odvozený z definice sub-nezávislosti	21
2.2 Porovnání testů sériové nezávislosti funkcionálních dat	36
2.2.1 Výpočetní složitost testů	36
2.2.2 Empirická hladina testů	37
2.2.3 Síla testu	38
Závěr	42
Seznam použité literatury	43
A Přílohy	45
A.1 Generování dat	45
A.2 Výpočet testové statistiky	46

Úvod

V posledních 20 letech zaznamenala analýza funkcionálních dat velký vzestup díky neustálému vývoji výpočetních technologií. Své uplatnění nalézá v oborech jako je finančníctví, fyzika nebo například medicína. Stále jsou vyvíjeny nové metody testování na funkcionálních datech. Většina metod testování uvedených v knize Ramsay a Silverman (1997) využívá předpokladu, že funkcionální pozorování jsou nezávislá a stejně rozdělená. V mnohých případech však předpoklad nezávislosti není zajištěn z povahy způsobu sběru dat. Zanedbáním závislosti pozorovaných dat můžeme dojít k zavádějícím výsledkům (viz Horváth, Kokoszka a Reeder, 2013b).

Testování nezávislosti posloupnosti funkcionálních náhodných veličin se věnují články Gabrys a Kokoszka (2007), Horváth, Hušková a Rice (2013a) a 7. kapitola knihy Horváth a Kokoszka (2012). Tyto testy sériové nezávislosti zkoumají míru korelace náhodných funkcí dané posloupnosti. Jiný přístup byl představen v článku Hlávka, Hušková a Meintanis (2021), kde je test sériové nezávislosti odvozený z definice nezávislosti formulované charakteristickými funkcemi. V této práci se budeme věnovat jednomu z testů založeném na nutné podmínce neko-relovanosti, kterým je test z článku Horváth a kol. (2013a). Především se ale zaměříme na test nezávislosti z článku Hlávka a kol. (2021). Tento test budeme dále modifikovat na test vycházející ze slabší podmínky sub-nezávislosti, která byla představena v knize Hamedani (2015).

Celá práce je rozdělena na dvě kapitoly. V kapitole 1 si představíme problematiku sériové nezávislosti na jednorozměrných náhodných veličinách. Uvedeme zde testy, které se dají modifikovat i pro funkcionální případ. Kapitola 2 se již věnuje testům sériové nezávislosti posloupnosti funkcionálních náhodných veličin. Součástí této kapitoly je porovnání testů na základě simulace.

1. Nezávislost náhodných veličin

Typů nezávislostí pro spojité náhodné veličiny je mnoho, vzájemně se prolínají a většinou ke každému najdeme velké množství testů. Nezávislost můžeme zkoumat mezi dvěma náhodnými veličinami, n -ticí náhodných veličin, složkami vektorů, dvěma vektory, n -ticí vektorů nebo můžeme testovat nezávislost náhodných veličin či vektorů uspořádaných do časové řady. Například na test nezávislosti dvou náhodných veličin můžeme zároveň nahlížet jako na test nezávislosti složek dvourozměrného vektoru, pokud k oběma náhodným veličinám máme stejně velký náhodný výběr a máme k dispozici klíč, pomocí kterého je spárujeme do vektorů. Způsobů, jak nezávislost testovat, je také mnoho. Obvykle testujeme nezávislost přímo z její definice nebo testujeme nekorelovanost či sub-nezávislost, které jsou nutnou podmínkou nezávislosti.

Cílem této kapitoly je seznámit se s pojmy a testy, které budeme v další kapitole dále rozvíjet pro funkcionální pozorování. Budeme se tedy věnovat především těm testům, jejichž modifikace jsou vhodné pro testování nezávislosti funkcionálních pozorování.

Nejprve si zavedeme základní pojmy, které budeme dále využívat. K tomu, abychom mohli testovat nezávislost po sobě jdoucích členů časové řady, je dobré se nejdříve seznámit s testy nezávislosti složek náhodného vektoru. Těmto testům se věnujeme v sekci 1.2. V sekci 1.3 přejdeme k testům nezávislosti členů časové řady, které budeme říkat sériová nezávislost. Tato sekce obsahuje test nezávislosti založený na nutné podmínce nekorelovanosti, především se ale věnujeme testu nezávislosti přímo z její definice formulované pomocí distribuční nebo charakteristické funkce. Právě pro test sériové nezávislosti založený na charakteristických funkcích nalezneme využití i ve funkcionálním případě. Tento test ze sekce 1.3.2 modifikujeme v sekci 1.3.3 na test, jehož testová statistika je odvozena z definice sub-nezávislosti.

1.1 Zavedení pojmů

Definice 1. *Reálné náhodné veličiny X a Y jsou nezávislé, pokud pro $\forall x, y \in \mathbb{R}$ platí*

$$F_{X,Y}(x,y) = F_X(x) F_Y(y),$$

kde F_X , F_Y jsou marginální distribuční funkce příslušné náhodným veličinám X, Y a F_{XY} je sdružená distribuční funkce náhodného vektoru $(X, Y)^T$.

Charakteristická funkce dle (Feller, 1968, str. 508) jednoznačně určuje distribuční funkci. Tudíž charakteristická funkce zcela popisuje rozdělení náhodné veličiny a můžeme s její pomocí definovat nezávislost. Pro reálnou náhodnou veličinu X definujeme charakteristickou funkci následovně:

$$\varphi_X(t) = \mathbb{E}e^{itX} \text{ pro každé } t \in \mathbb{R}.$$

Podmínku nezávislosti pak můžeme ekvivalentně definovat v následujícím tvaru:

$$\varphi_{X,Y}(t, s) = \varphi_X(t) \varphi_Y(s) \text{ pro každé } t, s \in \mathbb{R},$$

kde $\varphi_{X,Y}$ je tvaru $\varphi_{X,Y}(t, s) = \mathbb{E}e^{i(tX+sY)}$ a nazýváme ji sdružená charakteristická funkce náhodného vektoru $(X, Y)^T$.

Slabším předpokladem než nezávislost je nekorelovanost, která je definována za předpokladu konečných druhých momentů na základě korelačního koeficientu:

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

který nabývá hodnot v intervalu $[-1,1]$.

Definice 2. *Nechť pro náhodné veličiny X a Y platí $\mathbb{E}X^2 < \infty$, $\mathbb{E}Y^2 < \infty$. Náhodné veličiny X a Y nazýváme nekorelované, pokud platí*

$$\rho_{X,Y} = 0.$$

Jsou-li dvě veličiny nezávislé, pak jsou také nekorelované, protože pro X, Y nezávislé platí

$$\text{Cov}(X,Y) = E[X - EX][Y - EY] = E[X - EX] E[Y - EY] = 0.$$

Opačná implikace ovšem neplatí. Jako protipříklad uvažujme náhodnou veličinu $X \sim U(-a,a)$, kde $a \in \mathbb{R}$, a náhodnou veličinu $Y = X^2$. Tyto dvě náhodné veličiny jsou nekorelované:

$$\text{Cov}(X,Y) = E X Y - EX EY = E X Y = EX^3 = 0.$$

Druhá a čtvrtá rovnost vychází z nulovosti lichých momentů rovnoměrného rozdělení na symetrickém intervalu $[-a,a]$. V tomto případě jsou tedy veličiny X, Y nekorelované, ale nejsou nezávislé. Pokud však navíc předpokládáme, že sdružené rozdělení vektoru $(X, Y)^T$ je dvourozměrné normální, potom jsou pojmy nekorelovanost a nezávislost ekvivalentní (viz Melnick a Tenenbein, 1982, str. 372).

Dalším vztahem náhodných veličin, který je slabší než nezávislost, je sub-nezávislost.

Definice 3. *Náhodné veličiny X a Y nazýváme sub-nezávislé, pokud platí*

$$\varphi_{X+Y}(t) = \varphi_X(t) \varphi_Y(t) \text{ pro každé } t \in \mathbb{R}.$$

Z nezávislosti plyne sub-nezávislost snadno:

$$\varphi_{X+Y}(t) = \mathbb{E} e^{it(X+Y)} \stackrel{\text{nez.}}{=} \mathbb{E} e^{itX} \mathbb{E} e^{itY} = \varphi_X(t) \varphi_Y(t).$$

Sub-nezávislost je však silnější předpoklad než nekorelovanost. Tento vztah již neplyne triviálně.

Lemma 1. *Nechť jsou náhodné veličiny X a Y sub-nezávislé a mají konečné druhé momenty. Potom jsou náhodné veličiny X, Y nekorelované.*

Důkaz. Budeme pracovat s rozepsanou definicí sub-nezávislosti:

$$\mathbb{E} e^{it(X+Y)} = \mathbb{E} e^{itX} \mathbb{E} e^{itY}.$$

Celou rovnici za pomoci Fubiniho věty (viz Roussas, 2005, kapitola 5, věta 12) zderivujeme dle t :

$$\mathbb{E}[i(X+Y) e^{it(X+Y)}] = \mathbb{E}[iX e^{itX}] \mathbb{E} e^{itY} + \mathbb{E} e^{itX} \mathbb{E}[iY e^{itY}].$$

Rovnici pokrátíme o i a zderivujeme podruhé:

$$\mathbb{E}[i(X+Y)^2 e^{it(X+Y)}] = \mathbb{E}[iX^2 e^{itX}] \mathbb{E} e^{itY} + 2i \mathbb{E}[X e^{itX}] \mathbb{E}[Y e^{itY}] + \mathbb{E} e^{itX} \mathbb{E}[iY^2 e^{itY}]$$

Opět pokrátíme rovnici o i a položíme $t = 0$:

$$\mathbb{E}[(X+Y)^2] = \mathbb{E}X^2 + 2\mathbb{E}X\mathbb{E}Y + \mathbb{E}Y^2.$$

Po umocnění součtu a převedení všech členů na pravou stranu získáváme:

$$\mathbb{E}[XY] - \mathbb{E}X\mathbb{E}Y = 0.$$

Tedy $Cov(X,Y) = 0$ a dokázali jsme tak nekorelovanost náhodných veličin X a Y . □

V následujícím lemmatu si uvedeme zajímavé vlastnosti sub-nezávislosti, které jsou uvedeny v článku Schennach (2019).

Lemma 2 (Schennach, 2019, věta 5).

1. Jsou-li náhodné veličiny X, Y sub-nezávislé, potom jsou také náhodné veličiny $(a + cX), (b + cY)$ sub-nezávislé pro $\forall a, b, c \in \mathbb{R}$.
2. Ačkoliv jsou náhodné veličiny X, Y sub-nezávislé, náhodné veličiny $(a + cX), (b + dY)$ nemusí být sub-nezávislé, je-li $c \neq d$.
3. Platí-li, že náhodné veličiny X, aY jsou sub-nezávislé pro $\forall a \in \mathbb{R}$, potom jsou náhodné veličiny X, Y nezávislé.

Bohužel tedy ani všechny lineární transformace nezachovávají sub-nezávislost mezi náhodnými veličinami.

Pojmy nezávislosti a sub-nezávislosti, které jsme zavedli pro dvojici náhodných veličin, nyní zobecníme pro posloupnost náhodných veličin X_1, \dots, X_n .

Definice 4. Reálné náhodné veličiny X_1, \dots, X_n nazýváme

- nezávislé, pokud platí $\varphi_{X_1, \dots, X_n}(\mathbf{t}) = \varphi_{X_1}(t_1) \cdots \varphi_{X_n}(t_n)$ pro každé $\mathbf{t} \in \mathbb{R}^n$,
- sub-nezávislé, pokud platí $\varphi_{X_1 + \dots + X_n}(t) = \varphi_{X_1}(t) \cdots \varphi_{X_n}(t)$ pro každé $t \in \mathbb{R}$.

Ekvivalentně lze podmínku nezávislosti formulovat pomocí distribučních funkcí. Reálné náhodné veličiny X_1, \dots, X_n nazýváme nezávislé, pokud pro každé $\mathbf{x} \in \mathbb{R}^n$ platí

$$F_{X_1, \dots, X_n}(\mathbf{x}) = F_{X_1}(x_1) \cdots F_{X_n}(x_n).$$

Vztah mezi sub-nezávislostí posloupnosti náhodných veličin a sub-nezávislostí dvou náhodných veličin vyplývá z následujícího lemmatu.

Lemma 3 (Schennach, 2019, věta 7).

1. Necht jsou náhodné veličiny X_1, X_2 sub-nezávislé a náhodné veličiny $(X_1 + X_2), X_3$ jsou také sub-nezávislé. Potom je posloupnost náhodných veličin X_1, X_2, X_3 sub-nezávislá.
2. Ačkoliv jsou náhodné veličiny X_1, X_2, X_3 po dvojicích sub-nezávislé, tato posloupnost nemusí být sub-nezávislá.

Na posloupnost náhodných veličin X_1, \dots, X_n , které jsou v posloupnosti smysluplně uspořádány, nyní začněme nahlížet jako na časovou řadu $\{X_j, j \in \mathbb{N}\}$. Definujme si pojmy autokovarianční a autokorelační funkce.

Definice 5. Pro reálnou časovou řadu $\{X_j, j \in \mathbb{N}\}$ splňující $\mathbb{E}|X_j|^2 < \infty$ definujme funkci:

- autokovarianční: $R(j, k) = \mathbb{E}[(X_j - \mu_j)(X_k - \mu_k)]$ pro $j, k \in \mathbb{Z}$, $\mu_j := \mathbb{E}[X_j]$,
- autokorelační:

$$r(j, k) = \frac{R(j, k)}{\sqrt{R(j, j)}\sqrt{R(k, k)}} \text{ pro } j, k \in \mathbb{Z}.$$

O časové řadě $\{X_j, j \in \mathbb{Z}\}$ budeme předpokládat, že je reálná a stacionární. Stacionární časová řada splňuje vztah:

$$F_{t_1, \dots, t_n}(x_1, \dots, x_n) = F_{t_1+h, \dots, t_n+h}(x_1, \dots, x_n) \text{ pro } \forall h \in \mathbb{Z}.$$

Náhodné veličiny takové časové řady mají tedy stejné rozdělení. Tudíž funkce střední hodnoty μ je konstantní. U stacionární časové řady je pro autokovarianční funkce postačující jeden parametr, platí totiž vztah $R(j, k) = R(j - k, 0)$. Značme tedy $R_j := R(j, 0)$, $r_j := r(j, 0)$. Autokorelační funkce stacionární řady se zjednoduší na tvar

$$r_j = \frac{R_j}{R_0},$$

kde R_0 je rozptylem členů časové řady $\{X_j, j \in \mathbb{Z}\}$.

Důležitým příkladem časových řad, se kterým se budeme dále setkávat, je bílý šum. Bílý šum je definován následujícími vlastnostmi: jeho složky jsou vzájemně nekorelované, střední hodnota nulová a rozptyl konečný.

1.2 Testy nezávislosti složek vektorů

Abychom lépe porozuměli testům sériové nezávislosti členů časové řady, seznámíme se nejprve s testy nezávislosti složek náhodného vektoru. V sekci 1.2.1 se zaměříme na test nezávislosti přímo z její definice, který budeme moci dále využít pro časové řady. Dále v sekci 1.2.1 odvodíme test nezávislosti založený na nutné podmínce sub-nezávislosti. Jelikož odvození kritických hodnot těchto testů je poměrně složité a pro dimenze vektorů $m \geq 2$ i výpočetně náročné, v praxi se pak často napočítávají kritické hodnoty těchto testů za pomoci permutačních testů.

1.2.1 Test nezávislosti odvozený z definice nezávislosti

Předpokládejme náhodný výběr $\mathbf{X}_1, \dots, \mathbf{X}_n$ vektorů $\mathbf{X}_j = (X_{j1}, \dots, X_{jm})^T$ dimenze m se spojitou sdruženou distribuční funkcí F_{X_1, \dots, X_m} a marginálními distribučními funkcemi F_{X_1}, \dots, F_{X_m} . Položme následující hypotézu a alternativu:

$$H_0 : \text{pro } \forall \mathbf{x} \in \mathbb{R}^m \text{ platí, že } F_{X_1, \dots, X_m}(\mathbf{x}) = F_{X_1}(x_1) \cdots F_{X_m}(x_m),$$

$$H_1 : \exists \mathbf{x} \in \mathbb{R}^m \text{ takové, že platí } F_{X_1, \dots, X_m}(\mathbf{x}) \neq F_{X_1}(x_1) \cdots F_{X_m}(x_m).$$

První test nezávislosti složek vektorů byl představen v článku Hoeffding (1948) a předpokládal pouze dvousložkové vektory. Rozšíření na vícesložkové vektory bylo později odvozeno v článku Blum a kol. (1961). Oba testy pracují se vzdáleností empirické sdružené distribuční funkce $\hat{F}_{n,m}$ od součinu empirických marginálních distribučních funkcí $\hat{F}_{n,X_1}, \dots, \hat{F}_{n,X_m}$:

$$D_n(\mathbf{x}) = \sqrt{n} (\hat{F}_{n,m}(\mathbf{x}) - \prod_{k=1}^m \hat{F}_{n,X_k}(x_k)) \quad \text{pro } \forall \mathbf{x} \in \mathbb{R}^m.$$

Empirická marginální \hat{F}_{n,X_k} a sdružená distribuční funkce $\hat{F}_{n,m}$ jsou tvaru:

$$\hat{F}_{n,k}(x) = \frac{1}{n} \sum_{j=1}^n I\{X_{jk} \leq x\} \quad \text{pro } \forall x \in \mathbb{R},$$

$$\hat{F}_{n,m}(\mathbf{x}) = \frac{1}{n} \sum_{j=1}^n \prod_{k=1}^m I\{X_{jk} \leq x_k\} \quad \text{pro } \forall \mathbf{x} \in \mathbb{R}^m.$$

Testová statistika pak může být tvaru Kolmogorov-Smirnovovy testové statistiky $A_n = \sup_{\mathbf{x} \in \mathbb{R}^m} |D_n(\mathbf{x})|$ nebo tvaru Cramér-von Misesovy statistiky:

$$K_n = \int_{\mathbb{R}^m} D_n^2(\mathbf{x}) d\hat{F}_{n,m}(\mathbf{x}).$$

Jelikož budeme později používat modifikaci testu pro funkcionální pozorování a hledání suprema je výpočetně obtížné, zaměříme naši pozornost na statistiku K_n . Dle článku Blum a kol. (1961) má za platnosti hypotézy H_0 testová statistika K_n stejné asymptotické rozdělení jako Cramér-von Misesova statistika centrovaného Gaussovského procesu D , jelikož empirický proces D_n konverguje slabě k tomuto Gaussovskému procesu D . Kovarianční funkce procesu D je v článku Blum a kol. (1961) odvozena a s odkazem na práci Dugue (1969) je zde odvozena i charakteristická funkce pro Cramér-von Misesovu statistiku Gaussovského procesu D . Abychom mohli napočítat kritické hodnoty pro asymptotické rozdělení testové statistiky K_n , potřebujeme získat inverzní tvar zmiňované charakteristické funkce, což bývá problematické zejména pro vyšší dimenze náhodného výběru. Různé přístupy k napočítání inverzního tvaru jsou uvedeny v 7. kapitole článku Blum a kol. (1961) nebo v článku Cotterill a Csorgo (1982). Pro případ $m = 2$ jsou kritické hodnoty k nalezení v článku Cotterill a Csorgo (1982). V tomto článku nalezneme i kritické hodnoty napočítané pomocí Cornish-Fisherova rozvoje limitní testové statistiky pro různá m .

Hypotézu nezávislosti můžeme ekvivalentně definovat pomocí charakteristických funkcí následovně:

$$H_0 : \varphi_{X_1, \dots, X_m}(\mathbf{t}) = \varphi_{X_1}(t_1) \cdots \varphi_{X_m}(t_m) \quad \text{pro } \forall \mathbf{t} \in \mathbb{R}^m.$$

Definujme proces C_n , který je obdobou procesu D_n , následovně:

$$C_n(\mathbf{t}) = \sqrt{n}(\hat{\varphi}_{n,m}(\mathbf{t}) - \prod_{k=1}^m \hat{\varphi}_{n,X_k}(t_k)) \quad \text{pro } \forall \mathbf{t} \in \mathbb{R}^m.$$

Rozepišme si empirickou marginální $\hat{\varphi}_{n,X_k}$ a sdruženou charakteristickou funkci $\hat{\varphi}_{n,m}$:

$$\begin{aligned} \hat{\varphi}_{n,X_k}(t) &= \frac{1}{n} \sum_{j=1}^n e^{itX_{jk}} \quad \text{pro } \forall t \in \mathbb{R}, \\ \hat{\varphi}_{n,m}(\mathbf{t}) &= \frac{1}{n} \sum_{j=1}^n e^{it^T \mathbf{X}_j} = \int_{\mathbb{R}^m} e^{it^T \mathbf{x}} d\hat{F}_{n,m}(\mathbf{x}) \quad \text{pro } \forall \mathbf{t} \in \mathbb{R}^m. \end{aligned}$$

Uvažujme testovou statistiku opět ve formě Cramér-von Misesovy statistiky:

$$S_n = \int_{\mathbb{R}^m} C_n^2(\mathbf{t}) w(\mathbf{t}) d\mathbf{t},$$

kde $w(\cdot)$ je vhodně zvolená váhová funkce.

Dle článku Csörgő (1985) konverguje empirický proces C_n slabě k centrovanému Gaussovskému procesu C . Konvergence rozdělení testové statistiky S_n k rozdělení Cramér-von Mises statistiky Gaussovského procesu C a napočítání kritických hodnot tohoto asymptotického testu se již článek Csörgő (1985) nevěnuje. Zdá se, že odvození bude velmi podobné jako v článku Blum a kol. (1961). Nicméně tento článek se odkazuje na netriviální výsledky pro distribuční funkce, které pro charakteristické funkce nejsou všechny odvozeny. Hledání charakteristické funkce Cramér-von Mises statistiky Gaussovského procesu C , kterou potřebujeme k odvození asymptotického rozdělení statistiky S_n , se věnují v článcích Durbin (1970) a Dugue (1969).

1.2.2 Test nezávislosti odvozený z definice sub-nezávislosti

V této kapitole se pokusíme odvodit tvar testové statistiky pro test nezávislosti založený na slabší podmínce sub-nezávislosti. Pokud jsou složky náhodného vektoru sub-závislé, zamítáme pak hypotézu o jejich nezávislosti. Předpokládejme náhodný výběr $\mathbf{X}_1, \dots, \mathbf{X}_n$ m -složkových vektorů $\mathbf{X}_i = (X_{i1}, \dots, X_{im})^T$ se spojitou sdruženou distribuční funkcí. Položme hypotézu a alternativu v následujícím tvaru:

$$\begin{aligned} H_0 &: \varphi_{X_1, \dots, X_m}(\mathbf{t}) = \varphi_{X_1}(t_1) \cdots \varphi_{X_m}(t_m) \quad \text{pro } \forall \mathbf{t} \in \mathbb{R}^m, \\ H_1 &: \exists t \in \mathbb{R} \text{ takové, že platí } \varphi_{X_1 + \dots + X_m}(t) \neq \varphi_{X_1}(t) \cdots \varphi_{X_m}(t). \end{aligned}$$

Nyní se pokusíme postupně odvodit testovou statistiku založenou na charakteristických funkcích podobně jako v článku Csörgő (1985). Zajímá nás vzdálenost empirické charakteristické funkce součtu náhodných veličin $\hat{\varphi}_{n,\Sigma}$ od součinu empirických marginálních charakteristických funkcí $\hat{\varphi}_{n,k}$. Z definice sub-nezávislosti by měla být tato vzdálenost za platnosti hypotézy H_0 nulová pro $\forall t \in \mathbb{R}$. Definujme tedy empirický proces C_n^* :

$$C_n^*(t) = \sqrt{n}(\hat{\varphi}_{n,\Sigma}(t) - \prod_{k=1}^m \hat{\varphi}_{n,X_k}(t)) \quad \text{pro } \forall t \in \mathbb{R}.$$

Empirická charakteristická funkce součtu náhodných veličin X_1, \dots, X_m je tvaru:

$$\hat{\varphi}_{n,\Sigma}(t) = \frac{1}{n} \sum_{j=1}^n e^{it \sum_{k=1}^m X_{jk}} = \int_{\mathbb{R}^m} e^{it \sum_{k=1}^m x_{kk}} d\hat{F}_{n,m}(\mathbf{x}) \quad \text{pro } \forall t \in \mathbb{R}.$$

Pokud si prohlédneme tvar empirické charakteristické funkce součtu $\hat{\varphi}_{n,\Sigma}(t)$ vidíme, že je speciálním případem empirické marginální charakteristické funkce $\hat{\varphi}_{n,m}(\mathbf{t})$. Pro $\forall \mathbf{t} \in \mathbb{R}^m$ takové, že splňuje $t_1 = \dots = t_m$, platí:

$$\hat{\varphi}_{n,m}(\mathbf{t}) = \hat{\varphi}_{n,\Sigma}(t_1), \quad C_n(\mathbf{t}) = C_n^*(t_1).$$

Jelikož dle článku Csörgő (1985) platí, že $C_n(\mathbf{t})$ slabě konverguje k $C(\mathbf{t})$ pro $\forall \mathbf{t} \in \mathbb{R}^m$ a $n \rightarrow \infty$, pak tato konvergence platí také pro $t_1 = \dots = t_m$. Tedy $C_n^*(t_1)$ konverguje slabě k $C^*(t_1)$, kde $C^*(t) = C(t, \dots, t)$ pro $\forall t \in \mathbb{R}$ je centrováný Gaussovský proces.

Testovou statistiku ponecháme v Cramér-von Misesově tvaru:

$$S_n^* = \int_{\mathbb{R}} [C_n^*(t)]^2 w(t) dt.$$

Rozdělení testové statistiky S_n^* neplyne z rozdělení testové statistiky S_n . Postup pro odvození kritických hodnot testu však bude velmi podobný.

Nabízí se otázka, proč se rovnou nezabývat testem slabší hypotézy sub-nezávislosti. Předpoklad nezávislosti často neplatí a slabší předpoklad sub-nezávislosti by mohl být v mnoha případech dostačující (viz Hamedani, 2015).

$$\begin{aligned} H_0 : & \text{ pro } \forall t \in \mathbb{R} \text{ platí, že } \varphi_{X_1+\dots+X_n}(t) = \varphi_{X_1}(t) \cdot \dots \cdot \varphi_{X_n}(t), \\ H_1 : & \exists t \in \mathbb{R} \text{ takové, že platí } \varphi_{X_1+\dots+X_m}(t) \neq \varphi_{X_1}(t) \cdot \dots \cdot \varphi_{X_n}(t). \end{aligned}$$

Je problematické, že za předpokladu hypotézy – tedy sub-nezávislosti nemůžeme odvodit kritické hodnoty podobným způsobem jako v předchozí sekci 1.2.1. Ačkoliv máme pro případ sub-nezávislých stejně rozdělených náhodných veličin k dispozici centrální limitní větu (viz Hamedani a Walter, 1984, tvrzení 5), transformace exponenciální funkcí nezachovává sub-nezávislost, jako je tomu v případě nezávislosti.

Nicméně odvození rozdělení testové statistiky je náročné teoreticky i výpočetně, a proto je výhodnější napočítat kritické hodnoty testové statistiky permutačním testem. V takovém případě není problémem testovat přímo hypotézu sub-nezávislosti s použitím testové statistiky S_n^* .

1.3 Testy sériové nezávislosti jednorozměrné časové řady

V této kapitole si ukážeme, jak zkonstruovat testové statistiky pro testy nezávislosti členů časové řady $\{X_t, t = 1, \dots, n\}$. Budeme se věnovat těm testům, jejichž tvar testové statistiky bude možné modifikovat pro časovou řadu funkcionálních pozorování. Nejprve se zaměříme na hojně využívané testování nezávislosti pomocí nutné podmínky nekorelovanosti (viz sekce 1.3.1). Především se ale budeme věnovat testování nezávislosti přímo z její definice, k čemuž využijeme poznatků ze sekce 1.2 o testování nezávislosti složek vektoru. V případě jednorozměrných časových řad články o testech nezávislosti (viz Skaug a Tjøstheim (1993), Ghoudi, Kulperger a Rémillard (2001)), jejichž testové statistiky jsou založeny na definici nezávislosti, využívají k testování nezávislosti distribuční funkce. Takový test si představíme v sekci 1.3.2. Pro použití na funkcionálních datech odvodíme ještě v téže sekci testovou statistiku, která využívá charakteristických funkcí. Na závěr se pokusíme v sekci 1.3.3 odvodit testovou statistiku pro testování nezávislosti za pomoci nutné podmínky sub-nezávislosti. Bohužel pro testy nezávislosti využívající charakteristické funkce nemáme k dispozici žádné podpůrné materiály a ani se nebudeme pokoušet odvodit asymptotické rozdělení testových statistik. Odvození jejich rozdělení by nebylo analogické k funkcionálnímu případu.

Chceme testovat hypotézu nezávislosti členů časové řady proti alternativě, že členy časové řady nesplňují definici nezávislosti nebo některou z nutných podmínek nezávislosti. Předpokládáme, že závislost členů časové řady se s rostoucí vzdáleností mezi členy postupně vytrácí. Proto namísto testování nezávislosti všech členů časové řady budeme testovat nezávislost pouze H po sobě jdoucích členů dané časové řady. Testujeme tedy sériovou nezávislost. Test s takto oslabenou alternativou by měl mít dostatečnou sílu detekovat závislost v časové řadě, pokud za alternativy předpokládáme data z modelu ARIMA.

1.3.1 Test založený na autokorelaci

Uvažujme reálnou stacionární časovou řadu $\{X_j, j = 1, \dots, n\}$. Určeme rozmezí $H \in \{1, \dots, n\}$, ve kterém budeme testovat nezávislost členů časové řady. Položme hypotézu a alternativu následujícího tvaru:

$$\begin{aligned} H_0 &: X_1, \dots, X_n \text{ jsou nezávislé,} \\ H_1 &: \text{existuje } h = 1, \dots, H \text{ takové, že} \\ &\quad \text{cov}(X_1, X_{1+h}) \neq 0. \end{aligned}$$

Pokud prokážeme korelovanost členů časové řady, zamítáme potom hypotézu o jejich nezávislosti.

Nejprve si představíme Boxův-Piercův test publikovaný v článku Box a Pierce (1970). Autoři zkonstruovali tento test, aby testovali korelace mezi chybovými členy ARIMA modelů, které se při vhodně zvoleném modelu chovají jako bílý šum. Předpokládali tedy centrovanou časovou řadu a testovali nekorelovanost jejich členů. Testová statistika Box-Piercova testu je potom součtem korelací chyb do vzdálenosti H . Rozepišme si definici autokorelační funkce:

$$r_h = \frac{\text{Cov}(X_1, X_{1+h})}{\text{Var}(X_1)}, \quad \text{pro } \forall h = 1, \dots, H.$$

Korelace je funkcí rozptylu a kovariance, na které centrování nemá vliv. Rozdělení testové statistiky odvozené v článku Anderson (1942), na které se článek Box a Pierce (1970) odvolává, lze tedy aplikovat i pro necentrované veličiny.

Uvažujme za odhad autokorelací r_s výběrovou autokorelací:

$$\hat{r}_h = \frac{\frac{1}{n} \sum_{j=1}^{n-h} X_j X_{j+h} - \hat{\mu}^2}{\frac{1}{n} \sum_{j=1}^n X_j^2 - \hat{\mu}^2}, \quad \text{kde } \hat{\mu} = \frac{1}{n} \sum_{j=1}^n X_j.$$

Odhad v tomto tvaru zavedl již H. Hotelling a je pro něj v článku Anderson (1942) odvozeno za platnosti hypotézy asymptotické normální rozdělení. Box-Piercův test využívá následující aproximace:

$$\text{var}(\hat{r}_h) = \frac{n-h}{n(n+2)} \approx \frac{1}{n}.$$

Získáváme tak asymptotické rozdělení autokorelace: $\sqrt{n} \hat{r}_h \stackrel{as.}{\sim} \mathcal{N}(0,1)$. Potom testová statistika Box-Piercova testu má za platnosti hypotézy asymptoticky χ_H^2 -rozdělení:

$$n \sum_{h=1}^H \hat{r}_h^2 \stackrel{as.}{\sim} \chi_H^2.$$

Hypotézu o nezávislosti zamítáme, je-li $n \sum_{h=1}^H \hat{r}_h^2 > \chi_H^2(1-\alpha)$, kde $\chi_H^2(1-\alpha)$ je $(1-\alpha)$ -kvantilem χ^2 -rozdělení s H stupni volnosti.

Ljung-Boxův test, který byl představen v článku Ljung a Box (1978), je jednoduchou modifikací Box-Piercova testu. Autoři poukazují na několik nedostatků Box-Piercova testu. Odhad rozptylu autokorelací, který je v testové statistice použit, je příliš hrubý a test celkově slabý. Ljung-Boxova testová statistika vychází za platnosti hypotézy z asymptotického rozdělení:

$$\sqrt{\frac{n(n+2)}{n-h}} \hat{r}_h \stackrel{as.}{\sim} \mathcal{N}(0,1).$$

Testová statistika Ljung-Boxova testu je tedy tvaru:

$$n(n+2) \sum_{h=1}^H (n-h)^{-1} \hat{r}_h^2 \stackrel{as.}{\sim} \chi_H^2.$$

Jelikož mají oba testy za platnosti hypotézy asymptotické χ_H^2 rozdělení, měla by být jejich střední hodnota blízká hodnotě H a jejich rozptyl blízký hodnotě $2H$. Článek Kan a Wang (2010) zkoumá přesné rozdělení autokorelací \hat{r}_h a napočítává skutečné momenty Box-Piercovy i Ljung-Boxovy testové statistiky. Rozptyl Box-Piercovy statistiky je ve skutečnosti menší než rozptyl jeho asymptotického χ_H^2 -rozdělení, test má tedy malou sílu. Navíc je jeho střední hodnota signifikantně odlišná, pokud rozmezí H není opravdu malé. Naopak rozptyl Ljung-Boxovy statistiky je signifikantně větší než $2H$, což vede k vyšší pravděpodobnosti zamítnutí platné hypotézy. Rozptyl obou testových statistik se pro H rostoucí dále vzdaluje od rozptylu χ_H^2 -rozdělení. Tyto testy jsou tedy vhodné pro případy, kdy máme H poměrově malé k n . Pro opačný případ najdeme složitější modifikace těchto testů v článcích Dufour a Roy (1986), Kwan a Sim (1996) a Kan a Wang (2010).

1.3.2 Test odvozený z definice nezávislosti

Stejně jako v předchozí kapitole uvažujeme reálnou stacionární časovou řadu $\{X_j, j = 1, \dots, n\}$. Určíme rozmezí $H = 1, \dots, n$, ve kterém budeme testovat nezávislost členů časové řady. Položme hypotézu a alternativu následujícího tvaru:

$$\begin{aligned} H_0 &: X_1, \dots, X_n \text{ jsou nezávislé,} \\ H_1^* &: \text{existují } \mathbf{x} \in \mathbb{R}^{H+1} \text{ takové, že} \\ &F_{X_1, \dots, X_{1+H}}(\mathbf{x}) \neq F_{X_1}(x_1) \cdots F_{X_{1+H}}(x_{1+H}). \end{aligned}$$

Mohli bychom uspořádat H po sobě jdoucích náhodných veličin do náhodného vektoru $(X_j, \dots, X_{j+H})^T$ pro každé $j = 1, \dots, n - H$ a testovat nezávislost složek tohoto vektoru. Dle článku Skaug a Tjøstheim (1993) je ale takto koncipovaný test poměrně slabý.

Přeformulujeme alternativu do následujícího tvaru:

$$\begin{aligned} H_1 &: \text{existují } \mathbf{x} \in \mathbb{R}^2 \text{ a } h = 1, \dots, H \text{ takové, že} \\ &F_{X_1, X_{1+h}}(\mathbf{x}) \neq F_{X_1}(x_1) \cdot F_{X_{1+h}}(x_2). \end{aligned}$$

Pozorujeme závislost dvojic náhodných veličin X_j, X_{j+h} vzdálených vždy o $h = 1, \dots, H$, což můžeme nahlédnout také jako testování nezávislosti složek náhodného vektoru $(X_j, X_{j+h})^T$ a využít poznatků sekce 1.2.1. Položme:

$$D_{n,h}(\mathbf{x}) = \sqrt{n-h} \left(\hat{F}_{n;X_1, X_{1+h}}(\mathbf{x}) - \hat{F}_{n;X_1}(x_1) \hat{F}_{n;X_{1+h}}(x_2) \right) \quad \text{pro } \forall \mathbf{x} \in \mathbb{R}^2.$$

Empirická marginální \hat{F}_{X_1} a sdružená distribuční funkce $\hat{F}_{n;X_1, X_2}$ jsou tvaru:

$$\begin{aligned} \hat{F}_{n;X_1}(x) &= \frac{1}{n-h} \sum_{j=1}^{n-h} I\{X_j \leq x\} \quad \text{pro } \forall x \in \mathbb{R}, \\ \hat{F}_{n;X_1, X_{1+h}}(\mathbf{x}) &= \frac{1}{n-h} \sum_{j=1}^n I\{X_j \leq x_1\} I\{X_{j+h} \leq x_2\} \quad \text{pro } \forall \mathbf{x} \in \mathbb{R}^2. \end{aligned}$$

Testovou statistiku definujeme ve tvaru Cramér-von Misesovy statistiky:

$$K_n = \sum_{h=1}^H \int_{\mathbb{R}^2} D_{n,h}^2(\mathbf{x}) d\hat{F}_{n;X_1, X_{1+h}}(\mathbf{x}).$$

Za platnosti hypotézy konverguje testová statistika v distribuci k náhodné veličině, která je lineární kombinací nezávislých χ_H^2 -rozdělených náhodných veličin (viz Skaug a Tjøstheim, 1993, věta 3). V článku jsou k nalezení také kritické hodnoty $u_{H,\alpha}$ testové statistiky pro různé vzdálenosti H a různé hladiny spolehlivosti (viz Skaug a Tjøstheim, 1993, tabulka 2). Hypotézu o nezávislosti zamítáme, pokud $K_n > u_{H,\alpha}$.

Nyní se pokusíme odvodit tvar testové statistiky z definice nezávislosti, ve které jsou použity charakteristické funkce namísto distribučních. Přeformulujeme alternativu H_1 do ekvivalentní podoby:

$$\begin{aligned} H_1 &: \text{existují } \mathbf{t} \in \mathbb{R}^2 \text{ a } h = 1, \dots, H \text{ takové, že} \\ &\varphi_{X_1, X_{1+h}}(\mathbf{t}) \neq \varphi_{X_1}(t_1) \cdot \varphi_{X_{1+h}}(t_2). \end{aligned}$$

Testová statistika bude potom tvaru:

$$S_n = \sum_{h=1}^H \int_{\mathbb{R}^2} C_{n,h}^2(\mathbf{t}) w(\mathbf{t}) d\mathbf{t},$$

kde $C_{n,h}(\mathbf{t}) = \sqrt{n-h} \left(\hat{\varphi}_{n;X_1, X_{1+h}}(\mathbf{t}) - \hat{\varphi}_{n;X_1}(t_1) \hat{\varphi}_{n;X_{1+h}}(t_2) \right)$ pro $\forall \mathbf{t} \in \mathbb{R}^2$,

$$\hat{\varphi}_{n;X_1}(t) = \frac{1}{n-h} \sum_{j=1}^{n-h} e^{itX_j} \quad \text{pro } \forall t \in \mathbb{R},$$

$$\hat{\varphi}_{n;X_1, X_{1+h}}(\mathbf{t}) = \frac{1}{n-h} \sum_{j=1}^n e^{i(t_1 X_j + t_2 X_{j+h})} \quad \text{pro } \forall \mathbf{t} \in \mathbb{R}^2$$

a $w(\cdot)$ je nezáporná funkce na \mathbb{R}^2 , která splňuje:

$$0 < \int_{\mathbb{R}^2} w(\mathbf{t}) d\mathbf{t} < \infty, \quad w(\mathbf{t}) = w(-\mathbf{t}) \quad \text{pro každé } \mathbf{t} \in \mathbb{R}^2.$$

1.3.3 Test odvozený z definice sub-nezávislosti

V této sekci sestavíme test sériové nezávislosti vycházející z nutné podmínky sub-nezávislosti. Tentokrát vyjdeme z definice sub-nezávislosti, ve které využijeme charakteristické funkce. Alternativu z předchozí sekce 1.3.2 oslabíme. Testujme tedy:

$H_0 : X_1, \dots, X_n$ jsou nezávislé,

$H_1 : existují t \in \mathbb{R}$ a $h = 1, \dots, H$ takové, že

$$\varphi_{X_1+X_{1+h}}(t) \neq \varphi_{X_1}(t) \cdot \varphi_{X_{1+h}}(t).$$

Podobně jako v předchozí sekci definujme testovou statistiku ve tvaru:

$$S_n^* = \sum_{h=1}^H \int_{\mathbb{R}^2} (C_{n,h}^*)^2(\mathbf{t}) w(\mathbf{t}) d\mathbf{t},$$

kde $C_{n,h}^*(\mathbf{t}) = \sqrt{n-h} \left(\hat{\varphi}_{n;X_1+X_{1+h}}(\mathbf{t}) - \hat{\varphi}_{n;X_1}(t_1) \hat{\varphi}_{n;X_{1+h}}(t_2) \right)$ pro $\forall \mathbf{t} \in \mathbb{R}^2$,

$$\hat{\varphi}_{n;X_1}(t) = \frac{1}{n-h} \sum_{j=1}^{n-h} e^{itX_j} \quad \text{pro } \forall t \in \mathbb{R},$$

$$\hat{\varphi}_{n;X_1+X_{1+h}}(\mathbf{t}) = \frac{1}{n-h} \sum_{j=1}^n e^{it(X_j+X_{j+h})} \quad \text{pro } \forall \mathbf{t} \in \mathbb{R}^2$$

a $w(\cdot)$ je nezáporná funkce na \mathbb{R}^2 , která splňuje:

$$0 < \int_{\mathbb{R}^2} w(\mathbf{t}) d\mathbf{t} < \infty, \quad w(\mathbf{t}) = w(-\mathbf{t}) \quad \text{pro každé } \mathbf{t} \in \mathbb{R}^2.$$

Výhodou takto koncipovaného testu by mohlo být výpočetní zjednodušení testové statistiky S_n^* , protože oproti testové statistice S_n zde integrujeme pouze přes \mathbb{R}^2 namísto \mathbb{R}^2 . Kritické hodnoty můžeme napočítat pomocí permutačního testu. Namísto nezávislosti by bylo možné položit za hypotézu sub-nezávislost. Jak již bylo zmíněno na konci sekce 1.2.2, v takovém případě je vhodné napočítat kritické hodnoty permutačním testem. Pokud bychom chtěli však odvozovat asymptotické rozdělení testové statistiky S_n^* , předpoklad nezávislosti je žádaný, ne-li nezbytný.

2. Nezávislost funkcionálních dat

Nyní přejdeme k testům nezávislosti funkcionálních veličin časové řady. Pro seznámení s tematikou funkcionálních dat jsou vhodné knihy Horváth a Kokoszka (2012) a Ramsay a Silverman (1997). Funkcionální veličiny můžeme chápat jako náhodné prvky z prostoru čtvercově integrovatelných funkcí L^2 . Pro jednoduchost budeme uvažovat prostor $L^2[0,1]$. Jsou to tedy čtvercově integrovatelné náhodné funkce na $[0,1]$, tzn.

$$\mathbb{E} \int_0^1 X^2(t) dt < \infty.$$

Ačkoliv problematika funkcionálních dat bývá zpracována pro náhodné funkce, v praxi často pracujeme spíše s mnoharozměrnými vektory. Problematické je totiž sbírání dat ve spojitě formě a také jejich zaznamenávání. Pokud bychom chtěli pracovat se spojitou formou funkcionálních veličin, je potřeba diskrétně pozorovaná data vyhlazovat (viz Ramsay a Silverman, 1997, kapitola 3). Uvažujme časovou řadu náhodných funkcí $\{X_j(\cdot), j = 1, \dots, n\}$ z prostoru $L^2[0,1]$ a předpokládejme, že jsme se rozhodli pracovat s diskretizovanou formou funkcionálních dat. Náhodnou funkci $X_j(\cdot)$ pozorujeme v nějakých časech t_{j1}, \dots, t_{jp_j} . Pozorovací časy i jejich počet se může lišit pro různá j . Pro jednoduchost značení předpokládejme shodné pozorovací časy funkcí a také shodné rozestupy mezi časy pozorování, tzn. $t_m = m/p$ pro $m = 1, \dots, p$. Pozorujeme pak náhodné vektory \mathbf{X}_j dimenze p , kde hodnota X_{jm} by odpovídala hodnotě $X_j(t_m)$ veličiny ve funkcionálním tvaru.

At už pracujeme s funkcionálními daty v jakékoliv podobě, je důležité na každé pozorování nahlížet jako na pozorování, které se vyvíjí v čase, a nikoliv jako na p jednotlivých hodnot. I v případě, kdy pracujeme s funkcionálními veličinami v diskretizované podobě, není často vhodné využívat metod určených pro běžné náhodné vektory z důvodu vysoké dimenze.

Nyní si definujeme pojmy týkající se nezávislosti funkcionálních dat. Předpokládejme časovou řadu náhodných funkcí $\{X_j(\cdot), j = 1, \dots, n\}$ z prostoru $L^2[0,1]$. Řekneme, že náhodné funkce $X_1(\cdot), \dots, X_n(\cdot)$ jsou nezávislé, pokud pro $\forall u_1, \dots, u_n \in L^2[0,1]$ platí:

$$\Phi_{X_1, \dots, X_n}(u_1, \dots, u_n) = \varphi_{X_1}(u_1) \cdots \varphi_{X_n}(u_n),$$

kde $\varphi_{X_1}(u_1) = \mathbb{E} \exp\{i \int_0^1 u_1(t) X_1(t) dt\}$ nazýváme marginální charakteristickou funkcí funkce $X_1(\cdot)$ a $\Phi_{X_1, \dots, X_n}(u_1, \dots, u_n) = \mathbb{E} \exp\{i \int_0^1 \sum_{j=1}^n u_j(t) X_j(t) dt\}$ nazýváme sdruženou charakteristickou funkcí pro funkce $X_1(\cdot), \dots, X_n(\cdot)$. Řekneme, že náhodné funkce $X_1(\cdot), \dots, X_n(\cdot)$ jsou sub-nezávislé, pokud pro $\forall u \in L^2[0,1]$ platí:

$$\Phi_{X_1 + \dots + X_n}^*(u) = \varphi_{X_1}(u) \cdots \varphi_{X_n}(u),$$

kde $\Phi_{X_1 + \dots + X_n}^*(u) = \mathbb{E} \exp\{i \int_0^1 u(t) \sum_{j=1}^n X_j(t) dt\}$.

Pro úplnost si přepíšeme následující definice také pro případ diskretizovaných funkcionálních veličin. Řekneme, že diskretizovaná funkcionální pozorování $\mathbf{X}_1, \dots, \mathbf{X}_n$ dimenze p jsou nezávislá, pokud pro $\forall \mathbf{u}_1, \dots, \mathbf{u}_n \in \mathbb{R}^p$ platí:

$$\Phi_{\mathbf{X}_1, \dots, \mathbf{X}_n}(\mathbf{u}_1, \dots, \mathbf{u}_n) = \varphi_{\mathbf{X}_1}(\mathbf{u}_1) \cdots \varphi_{\mathbf{X}_n}(\mathbf{u}_n),$$

$\varphi_{\mathbf{X}_1}(\mathbf{u}_1) = \mathbb{E} \exp\{i\mathbf{u}_1^T \mathbf{X}_1\}$ nazýváme marginální charakteristickou funkcí vektoru \mathbf{X}_1 a $\Phi_{\mathbf{X}_1, \dots, \mathbf{X}_n}(\mathbf{u}_1, \dots, \mathbf{u}_n) = \mathbb{E} \exp\{i \sum_{j=1}^n \mathbf{u}_j^T \mathbf{X}_j\}$ nazýváme sdruženou charakteristickou funkcí vektorů $\mathbf{X}_1, \dots, \mathbf{X}_n$. Řekneme, že diskretizovaná funkcionální pozorování $\mathbf{X}_1, \dots, \mathbf{X}_n$ jsou sub-nezávislá, pokud pro $\forall \mathbf{u} \in \mathbb{R}^p$ platí:

$$\Phi_{\mathbf{X}_1 + \dots + \mathbf{X}_n}^*(\mathbf{u}) = \varphi_{\mathbf{X}_1}(\mathbf{u}) \cdots \varphi_{\mathbf{X}_n}(\mathbf{u}),$$

kde $\Phi_{\mathbf{X}_1 + \dots + \mathbf{X}_n}^*(\mathbf{u}) = \mathbb{E} \exp\{i \sum_{j=1}^n \mathbf{u}^T \mathbf{X}_j\}$.

2.1 Testy sériové nezávislosti funkcionálních dat

Navážeme na sekci 1.3, která pojednává o testech sériové nezávislosti časové řady jednorozměrných náhodných veličin, a představíme si je tentokrát pro funkcionální náhodné veličiny. Zaměříme se na testy publikované v článcích Horváth, Hušková a Rice (2013a) a Hlávka, Hušková a Meintanis (2021). V sekci 2.1.1 si představíme test z článku Horváth a kol. (2013a), který testuje nezávislost za pomoci nutné podmínky nekorelovanosti. V sekci 2.1.2 se budeme věnovat testu z článku Hlávka a kol. (2021), který vychází přímo z definice nezávislosti. Článek Horváth a kol. (2013a) pracuje s funkcionálními veličinami ve formě funkcí z prostoru $L^2[0,1]$, zatímco článek Hlávka a kol. (2021) pracuje s funkcionálními veličinami v diskretizované podobě. Formu funkcionálních veličin stanovenou články jsem se rozhodla pro nerozporuplnost platnosti výsledků článků zachovat. Test nezávislosti vycházející z definice nezávislosti budeme v sekci 2.1.3 modifikovat na test nezávislosti založený na nutné podmínce sub-nezávislosti a odvodíme některé vlastnosti jeho testové statistiky.

2.1.1 Test založený na autokorelaci

Test představený v článku Horváth, Hušková a Rice (2013a) testuje nutnou podmínku nekorelovanosti členů časové řady. Určíme rozmezí H , ve kterém budeme testovat nekorelovanost členů časové řady. Předpokládejme reálnou stacionární ergodickou časovou řadu funkcionálních pozorování $\{X_j(\cdot), j = 1, \dots, n\}$ z prostoru $L^2[0,1]$. Položme následující hypotézu a alternativu:

$$\begin{aligned} H_0 &: X_1(\cdot), \dots, X_n(\cdot) \text{ jsou nezávislé,} \\ H_1 &: \text{existuje } h = 1, \dots, H \text{ takové, že } \int_0^1 \int_0^1 C_h^2(t,s) dt ds > 0, \end{aligned}$$

kde $C_h(t,s) = \text{cov}(X_1(t), X_{1+h}(s))$ je autokovarianční funkcí. Testová statistika pak bude součtem odhadů autokovariancí dvou funkcionálních veličin vzdálených o h pro $h = 1, \dots, H$. Autokovarianční funkce $C_h(t,s)$, empirická autokovarianční funkce $\hat{C}_{n,h}(t,s)$, kovarianční funkce $C_0(t,s)$ a empirická kovarianční funkce $\hat{C}_{n,0}(t,s)$ jsou následujícího tvaru:

$$\begin{aligned} C_h(t,s) &= \mathbb{E}[X_1(t) - \mu(t)][X_{1+h}(s) - \mu(s)], \\ \hat{C}_{n,h}(t,s) &= \frac{1}{n-h} \sum_{j=1}^{n-h} [X_j(t) - \hat{\mu}(t)][X_{j+h}(s) - \hat{\mu}(s)], \\ C_0(t,s) &= \mathbb{E}[X_1(t) - \mu(t)][X_1(s) - \mu(s)], \end{aligned}$$

$$\hat{C}_{n,0}(t,s) = \frac{1}{n} \sum_{j=1}^n [X_j(t) - \hat{\mu}(t)][X_j(s) - \hat{\mu}(s)],$$

kde $\mu(t) = \mathbb{E}X_1(t)$ je funkcí střední hodnoty a $\hat{\mu}(t) = 1/n \sum_{j=1}^n X_j$ jejím odhadem. Položme:

$$V_{n,H} = \sum_{h=1}^H \int_0^1 \int_0^1 \hat{C}_{n,h}^2(t,s) dt ds.$$

Jelikož za platnosti hypotézy je $C_h^2(t,s) = 0$ pro $h \geq 1$ a skoro všechna $t,s \in [0,1]$, v neprospěch hypotézy budou svědčit velké hodnoty $V_{n,H}$. Následující věta určuje asymptotické rozdělení testové statistiky.

Věta 4 (Horváth a kol., 2013a, str. 102). *Nechť $\{X_j(\cdot), j = 1, \dots, n\}$ je reálnou stacionární časovou řadou funkcionálních pozorování splňující $\mathbb{E}\|X_1\|^4 < \infty$. Dále necht pro $H = H_n$ při $n \rightarrow \infty$ platí:*

$$H_n \rightarrow \infty, \quad H = O((\log n)^\alpha) \text{ pro některé } \alpha > 0.$$

Potom za předpokladu hypotézy pro $n \rightarrow \infty$ platí:

$$\frac{1}{\sqrt{2H\hat{\sigma}_n^2}} (nV_{n,h} - H\hat{\nu}_n) \xrightarrow{D} \mathcal{N}(0,1),$$

kde

$$\hat{\nu}_n = \left(\int_0^1 \hat{C}_{n,0}(t,t) dt \right)^2, \quad \hat{\sigma}_n^2 = \left(\int_0^1 \int_0^1 \hat{C}_{n,0}^2(t,s) dt ds \right)^2.$$

Hypotézu o nezávislosti zamítáme, pokud

$$\frac{1}{\sqrt{2H\hat{\sigma}_n^2}} (nV_{n,h} - H\hat{\nu}_n) > u_{1-\alpha},$$

kde $u_{1-\alpha}$ je $(1 - \alpha)$ -kvantil normovaného normálního rozdělení.

2.1.2 Test odvozený z definice nezávislosti

Test sériové nezávislosti založený na charakteristických funkcích pro funkcionální data byl představen v článku Hlávka, Hušková a Meintanis (2021). Namísto testování slabší podmínky nekorelovanosti veličin testujeme přímo nezávislost definovanou charakteristickými funkcemi. Zvolme rozmezí H , ve kterém budeme zkoumat závislost členů časové řady. Předpokládejme reálnou stacionární časovou řadu funkcionálních pozorování v diskretizované podobě náhodných vektorů $\{\mathbf{X}_j, j = 1, \dots, n\}$, které jsou dimenze p . Položme následující hypotézu a alternativu:

$$H_0 : \mathbf{X}_1, \dots, \mathbf{X}_n \text{ jsou nezávislé,}$$

$$H_1 : \text{existuje } \mathbf{u}, \mathbf{v} \in \mathbb{R}^p \text{ a } h = 1, \dots, H \text{ takové, že}$$

$$\Phi_{\mathbf{X}_1, \mathbf{X}_{1+h}}(\mathbf{u}, \mathbf{v}) \neq \varphi_{\mathbf{X}_1}(\mathbf{u}) \cdot \varphi_{\mathbf{X}_{1+h}}(\mathbf{v}),$$

kde $\Phi_{\mathbf{X}_1, \mathbf{X}_{1+h}}(\mathbf{u}, \mathbf{v}) = \mathbb{E}e^{i(\mathbf{u}^T \mathbf{X}_1 + \mathbf{v}^T \mathbf{X}_{1+h})}$. Zajímá nás vzdálenost sdružené empirické charakteristické funkce od součinu marginálních empirických charakteristických

funkcí. Tuto vzdálenost dvojic náhodných pozorování vzdálených o h napočítáme pro všechny $h = 1, \dots, H$:

$$D_{n,h;p}(\mathbf{u}, \mathbf{v}) = \hat{\Phi}_{1,n-h;p}(\mathbf{u}, \mathbf{v}) - \hat{\varphi}_{1,n-h;p}(\mathbf{u}) \hat{\varphi}_{h+1,n;p}(\mathbf{v}) \text{ pro každé } \mathbf{u}, \mathbf{v} \in \mathbb{R}^p,$$

kde $\hat{\Phi}_{1,n-h;p}(\mathbf{u}, \mathbf{v}) = (n-h)^{-1} \sum_{j=1}^{n-h} e^{i(\mathbf{u}^T \mathbf{X}_j + \mathbf{v}^T \mathbf{X}_{j+h})}$ je empirickou sdruženou charakteristickou funkcí a funkce $\hat{\varphi}_{r,s;p}(\mathbf{u}) = (s-r+1)^{-1} \sum_{j=r}^s e^{i\mathbf{u}^T \mathbf{X}_j}$ pro $1 \leq r \leq s \leq n$ je empirickou marginální charakteristickou funkcí. Pro konkrétní vzdálenost h pak napočítáme celkovou odchylku od nezávislosti integrací přes $\mathbf{u}, \mathbf{v} \in \mathbb{R}^p$. Testová statistika tak bude v duchu Cramér-von Mises testové statistiky. Mohli bychom také napočítat supremum přes $\mathbf{u}, \mathbf{v} \in \mathbb{R}^p$, což by odpovídalo Kolmogorov-Smirnovově typu testové statistiky, ale hledání suprema by bylo výpočetně náročnější. Položme tedy testovou statistiku následovně:

$$\Delta_{n,H;p} = \sum_{h=1}^H (n-h) \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} |D_{n,h;p}(\mathbf{u}, \mathbf{v})|^2 w(\mathbf{u})w(\mathbf{v}) d\mathbf{u}d\mathbf{v},$$

kde $w(\cdot)$ je vhodně zvolená váhová funkce. Je zapotřebí, aby váhová funkce byla nezáporná, konečně integrovatelná a symetrická. Symetričnost vyžadujeme pro zjednodušení výpočtů momentů testové statistiky a usnadnění ověření předpokladů asymptotické normality.

Věta 5 (Hlávka a kol., 2021, str. 610). *Nechť $\{\mathbf{X}_j, j = 1, \dots, n\}$ je reálnou stacionární časovou řadou p -rozměrných náhodných vektorů pro p pevné a $w(\cdot)$ je nezáporná funkce na \mathbb{R}^p , která splňuje:*

$$0 < \int_{\mathbb{R}^p} w(\mathbf{u})d\mathbf{u} < \infty, \quad w(\mathbf{u}) = w(-\mathbf{u}) \text{ pro každé } \mathbf{u} \in \mathbb{R}^p.$$

Dále necht' pro $H = H_n$ při $n \rightarrow \infty$ platí:

$$H_n \rightarrow \infty, \quad H_n^3/n \rightarrow 0.$$

Potom za předpokladu hypotézy pro $n \rightarrow \infty$ platí:

$$\frac{\Delta_{n,H;p} - H\gamma_p}{\sqrt{H\nu_p}} \xrightarrow{D} \mathcal{N}(0,1),$$

$$\begin{aligned} \text{kde } \gamma_p &= (1 - \mathbb{E}I_{w,p}(\mathbf{X}_1 - \mathbf{X}_2))^2, \\ \nu_p &= (\mathbb{E}\{I_{w,p}(\mathbf{X}_1 - \mathbf{X}_2) - \mathbb{E}[I_{w,p}(\mathbf{X}_1 - \mathbf{X}_2)|\mathbf{X}_1] \\ &\quad - \mathbb{E}[I_{w,p}(\mathbf{X}_1 - \mathbf{X}_2)|\mathbf{X}_2] + \mathbb{E}I_{w,p}(\mathbf{X}_1 - \mathbf{X}_2)\}^2)^2 > 0, \\ I_{w,p}(\mathbf{x}) &= \int_{\mathbb{R}^p} \cos(\mathbf{u}^T \mathbf{x})w(\mathbf{u})d\mathbf{u}. \end{aligned}$$

Abychom mohli testovat hypotézu o nezávislosti, potřebujeme odhadnout neznámé parametry γ_p a ν_p . Nestranným odhadem $\mathbb{E}I_{w,p}(\mathbf{X}_1 - \mathbf{X}_2)$ je za platnosti H_0 výraz:

$$\hat{I}_{w,p} = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j=1}^n I\{i \neq j\} I_{w,p}(\mathbf{X}_i - \mathbf{X}_j).$$

Nestranným odhadem $\mathbb{E}[I_{w,p}(\mathbf{X}_1 - \mathbf{X}_2)|\mathbf{X}_1]$ je za platnosti H_0 výraz:

$$\hat{I}_{w,p,1} = \frac{1}{n-1} \sum_{j=2}^n I_{w,p}(\mathbf{X}_1 - \mathbf{X}_j).$$

S jejich pomocí definujeme odhady:

$$\begin{aligned} \hat{\gamma}_p &= (1 - \hat{I}_{w,p})^2, \\ \hat{\nu}_p &= \left(\frac{1}{n(n-1)} \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n (I_{w,p}(\mathbf{X}_i - \mathbf{X}_j) - \hat{I}_{w,p,i} - \hat{I}_{w,p,j} + \hat{I}_{w,p})^2 \right)^2. \end{aligned}$$

Hypotézu H_0 zamítáme, pokud

$$\frac{\Delta_{n,H;p} - H\hat{\gamma}_p}{\sqrt{H\hat{\nu}_p}} > u_{1-\alpha},$$

kde $u_{1-\alpha}$ je $(1 - \alpha)$ -kvantil normovaného normálního rozdělení. Kritický obor je potom tvaru $(H\hat{\gamma}_p + u_{1-\alpha}\sqrt{H\hat{\nu}_p}, \infty)$.

Stanovili jsme si testovou statistiku $\Delta_{n,H;p}$ a nyní ji potřebujeme upravit do výpočetně jednoduššího tvaru. Začneme úpravou výrazu $D_{n,h;p}$. Přepočítaný tvar $D_{n,h;p}$ je v článku Hlávka a kol. (2021) uveden, chybí zde však jeho odvození, které provedeme v důkazu lemmatu 6. Připomeňme si nejprve důležité vztahy mezi goniometrickými funkcemi, které mimo tento důkaz využijeme také v důkazech v sekci 2.1.3. Pro $\forall x, y \in \mathbb{R}$ platí:

$$\begin{aligned} 1 &= \sin^2(x) + \cos^2(x) \\ \sin(2x) &= 2 \sin(x) \cos(x) \\ \cos(2x) &= \cos^2(x) - \sin^2(x) \\ \sin(x + y) &= \sin(x) \cos(y) + \cos(x) \sin(y) \\ \sin(x - y) &= \sin(x) \cos(y) - \cos(x) \sin(y) \\ \cos(x + y) &= \cos(x) \cos(y) - \sin(x) \sin(y) \\ \cos(x - y) &= \cos(x) \cos(y) + \sin(x) \sin(y) \end{aligned}$$

Tyto vztahy jsou uvedeny v knize Jarník (1963).

Lemma 6. *Nechť $\{\mathbf{X}_j, j = 1, \dots, n\}$ je reálnou časovou řadou p -rozměrných náhodných vektorů pro p pevné. Potom platí:*

$$\begin{aligned} |D_{n,h;p}(\mathbf{u}, \mathbf{v})|^2 &= \frac{1}{(n-h)^2} \sum_{j,k=1}^{n-h} \cos(\mathbf{u}^T \mathbf{X}_{j,k} + \mathbf{v}^T \mathbf{X}_{j+h,k+h}) \\ &\quad - \frac{2}{(n-h)^3} \sum_{j,k,l=1}^{n-h} \cos(\mathbf{u}^T \mathbf{X}_{j,k} + \mathbf{v}^T \mathbf{X}_{j+h,l+h}) \\ &\quad + \frac{1}{(n-h)^4} \sum_{j,k,l,m=1}^{n-h} \cos(\mathbf{u}^T \mathbf{X}_{j,k} + \mathbf{v}^T \mathbf{X}_{l+h,m+h}), \end{aligned}$$

kde značíme zkráceně $\mathbf{X}_{j,k} = \mathbf{X}_j - \mathbf{X}_k$.

Důkaz. Výraz $D_{n,h;p}(\mathbf{u}, \mathbf{v})$ nejprve rozdělíme na reálnou a imaginární část.

$$\begin{aligned}
D_{n,h;p}(\mathbf{u}, \mathbf{v}) &= \\
&= \hat{\Phi}_{1,n-h;p}(\mathbf{u}, \mathbf{v}) - \hat{\varphi}_{1,n-h;p}(\mathbf{u}) \hat{\varphi}_{h+1,n;p}(\mathbf{v}) \\
&= \frac{1}{n-h} \sum_{j=1}^{n-h} \exp\{i(\mathbf{u}^T \mathbf{X}_j + \mathbf{v}^T \mathbf{X}_{j+h})\} \\
&\quad - \frac{1}{n-h} \sum_{j=1}^{n-h} \exp\{i\mathbf{u}^T \mathbf{X}_j\} \frac{1}{n-h} \sum_{j=1}^{n-h} \exp\{i\mathbf{v}^T \mathbf{X}_{j+h}\} \\
&= \frac{1}{n-h} \sum_{j=1}^{n-h} \cos(\mathbf{u}^T \mathbf{X}_j + \mathbf{v}^T \mathbf{X}_{j+h}) + i \sin(\mathbf{u}^T \mathbf{X}_j + \mathbf{v}^T \mathbf{X}_{j+h}) \\
&\quad - \frac{1}{n-h} \sum_{j=1}^{n-h} \cos(\mathbf{u}^T \mathbf{X}_j) + i \sin(\mathbf{u}^T \mathbf{X}_j) \\
&\quad \times \frac{1}{n-h} \sum_{j=1}^{n-h} \cos(\mathbf{v}^T \mathbf{X}_{j+h}) + i \sin(\mathbf{v}^T \mathbf{X}_{j+h}) \\
&= \frac{1}{n-h} \sum_{j=1}^{n-h} \cos(\mathbf{u}^T \mathbf{X}_j + \mathbf{v}^T \mathbf{X}_{j+h}) - \frac{1}{(n-h)^2} \sum_{j=1}^{n-h} \sum_{k=1}^{n-h} \cos(\mathbf{u}^T \mathbf{X}_j) \cos(\mathbf{v}^T \mathbf{X}_{k+h}) \\
&\quad + \frac{1}{(n-h)^2} \sum_{j=1}^{n-h} \sum_{k=1}^{n-h} \sin(\mathbf{u}^T \mathbf{X}_j) \sin(\mathbf{v}^T \mathbf{X}_{k+h}) + \frac{i}{n-h} \sum_{j=1}^{n-h} \sin(\mathbf{u}^T \mathbf{X}_j + \mathbf{v}^T \mathbf{X}_{j+h}) \\
&\quad - \frac{i}{(n-h)^2} \sum_{j=1}^{n-h} \sum_{k=1}^{n-h} \cos(\mathbf{u}^T \mathbf{X}_j) \sin(\mathbf{v}^T \mathbf{X}_{k+h}) \\
&\quad - \frac{i}{(n-h)^2} \sum_{j=1}^{n-h} \sum_{k=1}^{n-h} \sin(\mathbf{u}^T \mathbf{X}_j) \cos(\mathbf{v}^T \mathbf{X}_{k+h}) \\
&= \left[\frac{1}{n-h} \sum_{j=1}^{n-h} \cos(\mathbf{u}^T \mathbf{X}_j + \mathbf{v}^T \mathbf{X}_{j+h}) - \frac{1}{(n-h)^2} \sum_{j=1}^{n-h} \sum_{k=1}^{n-h} \cos(\mathbf{u}^T \mathbf{X}_j + \mathbf{v}^T \mathbf{X}_{k+h}) \right] \\
&\quad + i \left[\frac{1}{n-h} \sum_{j=1}^{n-h} \sin(\mathbf{u}^T \mathbf{X}_j + \mathbf{v}^T \mathbf{X}_{j+h}) \right. \\
&\quad \quad \left. - \frac{1}{(n-h)^2} \sum_{j=1}^{n-h} \sum_{k=1}^{n-h} \sin(\mathbf{u}^T \mathbf{X}_j + \mathbf{v}^T \mathbf{X}_{k+h}) \right]
\end{aligned}$$

Nyní pro napočítání absolutní hodnoty $|D_{n,h;p}(\mathbf{u}, \mathbf{v})|^2$ umocníme reálnou a imaginární část každou zvlášť.

$$\begin{aligned}
|D_{n,h;p}(\mathbf{u}, \mathbf{v})|^2 &= \\
&= \frac{1}{(n-h)^2} \sum_{j=1}^{n-h} \sum_{k=1}^{n-h} \cos(\mathbf{u}^T \mathbf{X}_j + \mathbf{v}^T \mathbf{X}_{j+h}) \cos(\mathbf{u}^T \mathbf{X}_k + \mathbf{v}^T \mathbf{X}_{k+h}) \\
&\quad - \frac{2}{(n-h)^3} \sum_{j=1}^{n-h} \sum_{k=1}^{n-h} \sum_{l=1}^{n-h} \cos(\mathbf{u}^T \mathbf{X}_j + \mathbf{v}^T \mathbf{X}_{j+h}) \cos(\mathbf{u}^T \mathbf{X}_k + \mathbf{v}^T \mathbf{X}_{l+h}) \\
&\quad + \frac{1}{(n-h)^4} \sum_{j=1}^{n-h} \sum_{k=1}^{n-h} \sum_{l=1}^{n-h} \sum_{m=1}^{n-h} \cos(\mathbf{u}^T \mathbf{X}_j + \mathbf{v}^T \mathbf{X}_{l+h}) \cos(\mathbf{u}^T \mathbf{X}_k + \mathbf{v}^T \mathbf{X}_{m+h})
\end{aligned}$$

$$\begin{aligned}
& + \frac{1}{(n-h)^2} \sum_{j=1}^{n-h} \sum_{k=1}^{n-h} \sin(\mathbf{u}^T \mathbf{X}_j + \mathbf{v}^T \mathbf{X}_{j+h}) \sin(\mathbf{u}^T \mathbf{X}_k + \mathbf{v}^T \mathbf{X}_{k+h}) \\
& - \frac{2}{(n-h)^3} \sum_{j=1}^{n-h} \sum_{k=1}^{n-h} \sum_{l=1}^{n-h} \sin(\mathbf{u}^T \mathbf{X}_j + \mathbf{v}^T \mathbf{X}_{j+h}) \sin(\mathbf{u}^T \mathbf{X}_k + \mathbf{v}^T \mathbf{X}_{l+h}) \\
& + \frac{1}{(n-h)^4} \sum_{j=1}^{n-h} \sum_{k=1}^{n-h} \sum_{l=1}^{n-h} \sum_{m=1}^{n-h} \sin(\mathbf{u}^T \mathbf{X}_j + \mathbf{v}^T \mathbf{X}_{l+h}) \sin(\mathbf{u}^T \mathbf{X}_k + \mathbf{v}^T \mathbf{X}_{m+h}) \\
& = \frac{1}{(n-h)^2} \sum_{j=1}^{n-h} \sum_{k=1}^{n-h} \cos(\mathbf{u}^T (\mathbf{X}_j - \mathbf{X}_k) + \mathbf{v}^T (\mathbf{X}_{j+h} - \mathbf{X}_{k+h})) \\
& - \frac{2}{(n-h)^3} \sum_{j=1}^{n-h} \sum_{k=1}^{n-h} \sum_{l=1}^{n-h} \cos(\mathbf{u}^T (\mathbf{X}_j - \mathbf{X}_k) + \mathbf{v}^T (\mathbf{X}_{j+h} - \mathbf{X}_{l+h})) \\
& + \frac{1}{(n-h)^4} \sum_{j=1}^{n-h} \sum_{k=1}^{n-h} \sum_{l=1}^{n-h} \sum_{m=1}^{n-h} \cos(\mathbf{u}^T (\mathbf{X}_j - \mathbf{X}_k) + \mathbf{v}^T (\mathbf{X}_{l+h} - \mathbf{X}_{m+h}))
\end{aligned}$$

□

Pokračujme v úpravách testové statistiky $\Delta_{n,H;p}$:

$$\begin{aligned}
& \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} |D_{n,h;p}(\mathbf{u}, \mathbf{v})|^2 w(\mathbf{u}) w(\mathbf{v}) d\mathbf{u} d\mathbf{v} \\
& = \frac{1}{(n-h)^2} \sum_{j,k=1}^{n-h} \left[\int_{\mathbb{R}^p} \cos(\mathbf{u}^T \mathbf{X}_{j,k}) w(\mathbf{u}) d\mathbf{u} \int_{\mathbb{R}^p} \cos(\mathbf{v}^T \mathbf{X}_{j+h,k+h}) w(\mathbf{v}) d\mathbf{v} \right. \\
& \quad \left. - \int_{\mathbb{R}^p} \sin(\mathbf{u}^T \mathbf{X}_{j,k}) w(\mathbf{u}) d\mathbf{u} \int_{\mathbb{R}^p} \sin(\mathbf{v}^T \mathbf{X}_{j+h,k+h}) w(\mathbf{v}) d\mathbf{v} \right] \\
& - \frac{2}{(n-h)^3} \sum_{j,k,l=1}^{n-h} \left[\int_{\mathbb{R}^p} \cos(\mathbf{u}^T \mathbf{X}_{j,k}) w(\mathbf{u}) d\mathbf{u} \int_{\mathbb{R}^p} \cos(\mathbf{v}^T \mathbf{X}_{j+h,l+h}) w(\mathbf{v}) d\mathbf{v} \right. \\
& \quad \left. - \int_{\mathbb{R}^p} \sin(\mathbf{u}^T \mathbf{X}_{j,k}) w(\mathbf{u}) d\mathbf{u} \int_{\mathbb{R}^p} \sin(\mathbf{v}^T \mathbf{X}_{j+h,l+h}) w(\mathbf{v}) d\mathbf{v} \right] \\
& + \frac{1}{(n-h)^4} \sum_{j,k,l,m=1}^{n-h} \left[\int_{\mathbb{R}^p} \cos(\mathbf{u}^T \mathbf{X}_{j,k}) w(\mathbf{u}) d\mathbf{u} \int_{\mathbb{R}^p} \cos(\mathbf{v}^T \mathbf{X}_{l+h,m+h}) w(\mathbf{v}) d\mathbf{v} \right. \\
& \quad \left. - \int_{\mathbb{R}^p} \sin(\mathbf{u}^T \mathbf{X}_{j,k}) w(\mathbf{u}) d\mathbf{u} \int_{\mathbb{R}^p} \sin(\mathbf{v}^T \mathbf{X}_{l+h,m+h}) w(\mathbf{v}) d\mathbf{v} \right] \\
& = \frac{1}{(n-h)^2} \sum_{j,k=1}^{n-h} I_{w,p}(\mathbf{X}_{j,k}) I_{w,p}(\mathbf{X}_{j+h,k+h}) \\
& - \frac{2}{(n-h)^3} \sum_{j,k,l=1}^{n-h} I_{w,p}(\mathbf{X}_{j,k}) I_{w,p}(\mathbf{X}_{j+h,l+h}) \\
& + \frac{1}{(n-h)^4} \left(\sum_{j,k=1}^{n-h} I_{w,p}(\mathbf{X}_{j,k}) \right) \left(\sum_{j,k=1}^{n-h} I_{w,p}(\mathbf{X}_{j+h,k+h}) \right),
\end{aligned}$$

kde $\mathbf{X}_{j,k} = \mathbf{X}_j - \mathbf{X}_k$ a $I_{w,p}(x) = \int_{\mathbb{R}^p} \cos(\mathbf{u}^T x) w(\mathbf{u}) d\mathbf{u}$. Všimněme si, že funkce $I_{w,p}(x)$ je charakteristickou funkcí centrovaného náhodného vektoru U určeného

hustotou $w(\cdot)$:

$$\begin{aligned} I_{w,p}(\mathbf{x}) &= \int_{\mathbb{R}^p} \cos(\mathbf{u}^T \mathbf{x}) + i \sin(\mathbf{u}^T \mathbf{x}) w(\mathbf{u}) d\mathbf{u} \\ &= \int_{\mathbb{R}^p} \exp(i\mathbf{u}^T \mathbf{x}) w(\mathbf{u}) d\mathbf{u} \\ &= \varphi_U(\mathbf{x}). \end{aligned}$$

Jako váhovou funkci $w(\cdot)$ zvolme hustotu p -rozměrného normálního rozdělení $\mathcal{N}_p(\mathbf{0}, \mathbf{V})$. Potom využijeme znalosti charakteristické funkce normálního rozdělení a získáme následující vztah (násobení konstantou $\exp(1/2)$ zanedbáme):

$$I_{w,p}(\mathbf{x}) = \exp\{-\mathbf{x}^T \mathbf{V} \mathbf{x}\}.$$

Máme tedy zkompletovaný výpočetní tvar testové statistiky $\Delta_{n,H;p}$, který je závislý na volbě kovarianční matice \mathbf{V} .

2.1.3 Test odvozený z definice sub-nezávislosti

Podobně jako jsme v sekci 1.3.3 odvozovali tvar testové statistiky detekující sub-závislost z testové statistiky v sekci 1.3.2, pokusíme se nyní o totéž ve funkcionálním případě. Vyjdeme tedy z testu, který byl představen v článku (Hlávka a kol., 2021) a kterému jsme se věnovali v sekci 2.1.2. Stejně jako v předešlé sekci zvolme rozmezí H a předpokládejme reálnou stacionární časovou řadu diskretizovaných funkcionálních pozorování $\{\mathbf{X}_j, j = 1, \dots, n\}$ dimenze p . Položme následující hypotézu a alternativu:

$$\begin{aligned} H_0 &: \mathbf{X}_1, \dots, \mathbf{X}_n \text{ jsou nezávislé,} \\ H_1 &: \text{existuje } \mathbf{u} \in \mathbb{R}^p \text{ a } h = 1, \dots, H \text{ takové, že} \\ &\quad \Phi_{\mathbf{X}_1 + \mathbf{X}_{1+h}}^*(\mathbf{u}) \neq \varphi_{\mathbf{X}_1}(\mathbf{u}) \cdot \varphi_{\mathbf{X}_{1+h}}(\mathbf{u}), \end{aligned}$$

kde $\Phi_{\mathbf{X}_1 + \mathbf{X}_{1+h}}^*(\mathbf{u}) = \mathbb{E} e^{i\mathbf{u}^T(\mathbf{X}_1 + \mathbf{X}_{1+h})}$ je empirická charakteristická funkce součtu.

Definujme vzdálenost empirické charakteristické funkce součtu od součinu marginálních empirických charakteristických funkcí:

$$D_{n,h;p}^*(\mathbf{u}) = \hat{\Phi}_{1,n-h;p}^*(\mathbf{u}) - \hat{\varphi}_{1,n-h;p}(\mathbf{u}) \hat{\varphi}_{h+1,n;p}(\mathbf{u}) \text{ pro každé } \mathbf{u} \in \mathbb{R}^p,$$

$$\text{kde } \hat{\Phi}_{1,n-h;p}^*(\mathbf{u}) = (n-h)^{-1} \sum_{j=1}^{n-h} e^{i\mathbf{u}^T(\mathbf{X}_j + \mathbf{X}_{j+h})},$$

$$\hat{\varphi}_{r,s;p}(\mathbf{u}) = (s-r+1)^{-1} \sum_{j=r}^s e^{i\mathbf{u}^T \mathbf{X}_j} \text{ pro každé } 1 \leq r \leq s \leq n.$$

Testovou statistiku pro test nezávislosti definujme následovně:

$$\Delta_{n,H;p}^* = \sum_{h=1}^H (n-h) \int_{\mathbb{R}^p} |D_{n,h;p}^*(\mathbf{u})|^2 w(\mathbf{u}) d\mathbf{u},$$

kde $w(\cdot)$ je nezáporná funkce na \mathbb{R}^p , která splňuje:

$$0 < \int_{\mathbb{R}^p} w(\mathbf{u}) d\mathbf{u} < \infty, \quad w(\mathbf{u}) = w(-\mathbf{u}) \text{ pro každé } \mathbf{u} \in \mathbb{R}^p.$$

V neprospěch hypotézy svědčí velké hodnoty testové statistiky $\Delta_{n,H;p}^*$, která detekuje sub-závislost časové řady. S využitím vztahu $\Phi_{\mathbf{X}_1+\mathbf{X}_2}^*(\mathbf{u}) = \Phi_{\mathbf{X}_1,\mathbf{X}_2}(\mathbf{u}, \mathbf{u})$ mezi sub-nezávislostí a nezávislostí můžeme využít některých výsledků odvozených pro test se silnější alternativou závislosti, například výpočetní tvar $D_{n,h;p}^*$. V rámci testové statistiky $\Delta_{n,H;p}^*$ pak ale integrujeme odchylky od nezávislosti pouze jednou přes $\mathbf{u} \in \mathbb{R}^p$ a nikoliv dvakrát přes $\mathbf{u}, \mathbf{v} \in \mathbb{R}^p$ jako u testové statistiky $\Delta_{n,H;p}$. Nemůžeme tak využít asymptotických výsledků věty 5. Alespoň do bodu, než začneme manipulovat s integrály, se můžeme držet postupu z článku Hlávka a kol. (2021).

Nyní se zaměříme na odvození střední hodnoty testové statistiky $\Delta_{n,H;p}^*$, k čemuž je zapotřebí upravit tvar testové statistiky. Položme $T_n^* = \Delta_{n,H;p}^*/\sqrt{H}$ a rozepíšme jej následovně:

$$T_n^* = \frac{1}{\sqrt{H}} \sum_{h=1}^H (n-h) \int_{\mathbb{R}^p} \left| \frac{1}{n-h} \sum_{j=1}^{n-h} \exp\{i\mathbf{u}^T(\mathbf{X}_j + \mathbf{X}_{j+h})\} - \frac{1}{(n-h)^2} \sum_{j=1}^{n-h} \exp(i\mathbf{u}^T\mathbf{X}_j) \sum_{r=1}^{n-h} \exp(i\mathbf{u}^T\mathbf{X}_{r+h}) \right|^2 w(\mathbf{u}) d\mathbf{u}.$$

Nyní T_n^* rozšíříme o charakteristické funkce:

$$T_n^* = \frac{1}{\sqrt{H}} \sum_{h=1}^H (n-h) \int_{\mathbb{R}^p} \left| \frac{1}{n-h} \sum_{j=1}^{n-h} [e^{i\mathbf{u}^T\mathbf{X}_j} - \varphi_{\mathbf{X}}(\mathbf{u})] [e^{i\mathbf{u}^T\mathbf{X}_{j+h}} - \varphi_{\mathbf{X}}(\mathbf{u})] - \frac{1}{(n-h)^2} \sum_{j=1}^{n-h} [e^{i\mathbf{u}^T\mathbf{X}_j} - \varphi_{\mathbf{X}}(\mathbf{u})] \sum_{r=1}^{n-h} [e^{i\mathbf{u}^T\mathbf{X}_{r+h}} - \varphi_{\mathbf{X}}(\mathbf{u})] \right|^2 w(\mathbf{u}) d\mathbf{u}.$$

Pokud bychom si roznásobili členy z první sumy, zjistíme, že sčítance obsahující $\varphi_{\mathbf{X}}(\mathbf{u})$ se odečtou se sčítanci ze součinu dvou sum. Tento krok je tedy legitimní. Výraz T_n^* rozdělíme na jednodušší výraz $T_{n,1}^*$ a na součet výrazů, které jsou zanedbatelné. Tento krok zformulujeme do následujícího lemmatu 7.

Lemma 7. *Nechť $\{\mathbf{X}_j, j = 1, \dots, n\}$ je reálnou stacionární časovou řadou p -rozměrných náhodných vektorů pro p pevné a $w(\cdot)$ je nezáporná funkce na \mathbb{R}^p , která splňuje:*

$$0 < \int_{\mathbb{R}^p} w(\mathbf{u}) d\mathbf{u} < \infty, \quad w(\mathbf{u}) = w(-\mathbf{u}) \text{ pro každé } \mathbf{u} \in \mathbb{R}^p.$$

Dále necht pro $H = H_n$ při $n \rightarrow \infty$ platí:

$$H_n \rightarrow \infty, \quad H_n^3/n \rightarrow 0.$$

Potom za předpokladu hypotézy H_0 platí:

$$T_n^* = T_{n,1}^* + O_p(H^{1/2}n^{-1/2}),$$

$$T_{n,1}^* = \frac{1}{\sqrt{H}} \sum_{h=1}^H (n-h) \int_{\mathbb{R}^p} \left| \frac{1}{n-h} \sum_{j=1}^{n-h} [e^{i\mathbf{u}^T\mathbf{X}_j} - \varphi_{\mathbf{X}}(\mathbf{u})] [e^{i\mathbf{u}^T\mathbf{X}_{j+h}} - \varphi_{\mathbf{X}}(\mathbf{u})] \right|^2 \times w(\mathbf{u}) d\mathbf{u}.$$

Důkaz. Zavedme zkrácené značení:

$$S_{1;n,h}(\mathbf{u}) = \frac{1}{n-h} \sum_{j=1}^{n-h} [e^{i\mathbf{u}^T \mathbf{X}_j} - \varphi_{\mathbf{X}}(\mathbf{u})][e^{i\mathbf{u}^T \mathbf{X}_{j+h}} - \varphi_{\mathbf{X}}(\mathbf{u})],$$

$$S_{2;n,h}(\mathbf{u}) = \frac{1}{(n-h)^2} \sum_{j=1}^{n-h} [e^{i\mathbf{u}^T \mathbf{X}_j} - \varphi_{\mathbf{X}}(\mathbf{u})] \sum_{r=1}^{n-h} [e^{i\mathbf{u}^T \mathbf{X}_{r+h}} - \varphi_{\mathbf{X}}(\mathbf{u})].$$

Rozepišme si T_n^* následovně:

$$\begin{aligned} T_n^* &= \frac{1}{\sqrt{H}} \sum_{h=1}^H (n-h) \int_{\mathbb{R}^p} |S_{1;n,h}(\mathbf{u}) - S_{2;n,h}(\mathbf{u})|^2 w(\mathbf{u}) d\mathbf{u} \\ &= \frac{1}{\sqrt{H}} \sum_{h=1}^H (n-h) \int_{\mathbb{R}^p} [\Re(S_{1;n,h}(\mathbf{u})) - \Re(S_{2;n,h}(\mathbf{u}))]^2 \\ &\quad + [\Im(S_{1;n,h}(\mathbf{u})) - \Im(S_{2;n,h}(\mathbf{u}))]^2 w(\mathbf{u}) d\mathbf{u} \\ &= \frac{1}{\sqrt{H}} \sum_{h=1}^H (n-h) \int_{\mathbb{R}^p} [\Re(S_{1;n,h}(\mathbf{u}))]^2 + [\Im(S_{1;n,h}(\mathbf{u}))]^2 \\ &\quad - 2\Re(S_{1;n,h}(\mathbf{u}))\Re(S_{2;n,h}(\mathbf{u})) \\ &\quad - 2\Im(S_{1;n,h}(\mathbf{u}))\Im(S_{2;n,h}(\mathbf{u})) \\ &\quad + [\Re(S_{2;n,h}(\mathbf{u}))]^2 + [\Im(S_{2;n,h}(\mathbf{u}))]^2 w(\mathbf{u}) d\mathbf{u} \\ &= T_{n,1}^* + \frac{1}{\sqrt{H}} \sum_{h=1}^H (n-h) \int_{\mathbb{R}^p} -2\Re(S_{1;n,h}(\mathbf{u}))\Re(S_{2;n,h}(\mathbf{u})) \\ &\quad - 2\Im(S_{1;n,h}(\mathbf{u}))\Im(S_{2;n,h}(\mathbf{u})) \\ &\quad + |S_{2;n,h}(\mathbf{u})|^2 w(\mathbf{u}) d\mathbf{u} \end{aligned}$$

Nyní si rozdělme $S_{1;n,h}(\mathbf{u})$ a $S_{2;n,h}(\mathbf{u})$ na reálnou a imaginární část:

$$\begin{aligned} S_{1;n,h}(\mathbf{u}) &= \frac{1}{n-h} \sum_{j=1}^{n-h} [\cos(\mathbf{u}^T \mathbf{X}_j) + i \sin(\mathbf{u}^T \mathbf{X}_j) - \mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_1) - i \mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_1)] \\ &\quad \times [\cos(\mathbf{u}^T \mathbf{X}_{j+h}) + i \sin(\mathbf{u}^T \mathbf{X}_{j+h}) - \mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_1) - i \mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_1)] \\ &= \frac{1}{n-h} \sum_{j=1}^{n-h} [\cos(\mathbf{u}^T \mathbf{X}_j) - \mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_1)][\cos(\mathbf{u}^T \mathbf{X}_{j+h}) - \mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_1)] \\ &\quad - [\sin(\mathbf{u}^T \mathbf{X}_j) - \mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_1)][\sin(\mathbf{u}^T \mathbf{X}_{j+h}) - \mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_1)] \\ &+ \frac{i}{n-h} \sum_{j=1}^{n-h} [\cos(\mathbf{u}^T \mathbf{X}_j) - \mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_1)][\sin(\mathbf{u}^T \mathbf{X}_{j+h}) - \mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_1)] \\ &\quad + [\sin(\mathbf{u}^T \mathbf{X}_j) - \mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_1)][\cos(\mathbf{u}^T \mathbf{X}_{j+h}) - \mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_1)] \\ &=: \Re(S_{1;n,h}(\mathbf{u})) + i \Im(S_{1;n,h}(\mathbf{u})) \end{aligned}$$

$$\begin{aligned} S_{2;n,h}(\mathbf{u}) &= \\ &= \frac{1}{(n-h)^2} \sum_{j=1}^{n-h} \sum_{r=1}^{n-h} [\cos(\mathbf{u}^T \mathbf{X}_j) + i \sin(\mathbf{u}^T \mathbf{X}_j) - \mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_1) - i \mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_1)] \end{aligned}$$

$$\begin{aligned}
& \times [\cos(\mathbf{u}^T \mathbf{X}_{r+h}) + i \sin(\mathbf{u}^T \mathbf{X}_{r+h}) - \mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_1) - i \mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_1)] \\
= & \frac{1}{(n-h)^2} \sum_{j=1}^{n-h} \sum_{r=1}^{n-h} [\cos(\mathbf{u}^T \mathbf{X}_j) - \mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_1)] [\cos(\mathbf{u}^T \mathbf{X}_{r+h}) - \mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_1)] \\
& - [\sin(\mathbf{u}^T \mathbf{X}_j) - \mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_1)] [\sin(\mathbf{u}^T \mathbf{X}_{r+h}) - \mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_1)] \\
& + \frac{i}{(n-h)^2} \sum_{j=1}^{n-h} \sum_{r=1}^{n-h} [\cos(\mathbf{u}^T \mathbf{X}_j) - \mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_1)] [\sin(\mathbf{u}^T \mathbf{X}_{r+h}) - \mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_1)] \\
& + [\sin(\mathbf{u}^T \mathbf{X}_j) - \mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_1)] [\cos(\mathbf{u}^T \mathbf{X}_{r+h}) - \mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_1)] \\
= & \Re(S_{2;n,h}(\mathbf{u})) + i \Im(S_{2;n,h}(\mathbf{u}))
\end{aligned}$$

Chceme ukázat omezenost v pravděpodobnosti následujícího výrazu:

$$\begin{aligned}
& \frac{1}{\sqrt{H}} \sum_{h=1}^H (n-h) \int_{\mathbb{R}^p} -2\Re(S_{1;n,h}(\mathbf{u}))\Re(S_{2;n,h}(\mathbf{u})) - 2\Im(S_{1;n,h}(\mathbf{u}))\Im(S_{2;n,h}(\mathbf{u})) \\
& + |S_{2;n,h}(\mathbf{u})|^2 w(\mathbf{u}) d\mathbf{u}
\end{aligned}$$

Abychom odvodili omezenost v pravděpodobnosti členu $|S_{2;n,h}(\mathbf{u})|^2$, ukážeme omezenost jeho střední hodnoty:

$$\begin{aligned}
& \mathbb{E}|S_{2;n,h}(\mathbf{u})|^2 = \\
& = \mathbb{E} \left| \frac{1}{n-h} \sum_{j=1}^{n-h} [e^{i\mathbf{u}^T \mathbf{X}_j} - \varphi_{\mathbf{X}}(\mathbf{u})] \right|^2 \left| \frac{1}{n-h} \sum_{j=1}^{n-h} [e^{i\mathbf{u}^T \mathbf{X}_{j+h}} - \varphi_{\mathbf{X}}(\mathbf{u})] \right|^2 \\
& \leq \left[\mathbb{E} \left| \frac{1}{n-h} \sum_{j=1}^{n-h} [e^{i\mathbf{u}^T \mathbf{X}_j} - \varphi_{\mathbf{X}}(\mathbf{u})] \right|^4 \right]^{\frac{1}{2}} \left[\mathbb{E} \left| \frac{1}{n-h} \sum_{j=1}^{n-h} [e^{i\mathbf{u}^T \mathbf{X}_{j+h}} - \varphi_{\mathbf{X}}(\mathbf{u})] \right|^4 \right]^{\frac{1}{2}}
\end{aligned}$$

Rozepišme si střední hodnotu čtvrté mocniny:

$$\begin{aligned}
& \mathbb{E} \left| \frac{1}{n-h} \sum_{j=1}^{n-h} [e^{i\mathbf{u}^T \mathbf{X}_j} - \varphi_{\mathbf{X}}(\mathbf{u})] \right|^4 \\
& = \frac{1}{(n-h)^4} \mathbb{E} \left(\left| \sum_{j=1}^{n-h} [e^{i\mathbf{u}^T \mathbf{X}_j} - \varphi_{\mathbf{X}}(\mathbf{u})] \right|^2 \right)^2 \\
& = \frac{1}{(n-h)^4} \mathbb{E} \left(\left| \sum_{j=1}^{n-h} [\cos(\mathbf{u}^T \mathbf{X}_j) - \mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_j) + i \sin(\mathbf{u}^T \mathbf{X}_j) - i \mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_j)] \right|^2 \right)^2 \\
& = \frac{1}{(n-h)^4} \mathbb{E} \left(\left| \sum_{j=1}^{n-h} [\cos(\mathbf{u}^T \mathbf{X}_j) - \mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_j)] \right. \right. \\
& \quad \left. \left. + i \sum_{j=1}^{n-h} [\sin(\mathbf{u}^T \mathbf{X}_j) - \mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_j)] \right|^2 \right)^2 \\
& = \frac{1}{(n-h)^4} \mathbb{E} \left(\left(\sum_{j=1}^{n-h} [\cos(\mathbf{u}^T \mathbf{X}_j) - \mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_j)] \right)^2 \right. \\
& \quad \left. + \left(\sum_{j=1}^{n-h} [\sin(\mathbf{u}^T \mathbf{X}_j) - \mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_j)] \right)^2 \right)^2
\end{aligned}$$

Zavedme si zkrácené značení:

$$\begin{aligned} A(\mathbf{u}, j) &= \cos(\mathbf{u}^T \mathbf{X}_j) - \mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_j), \\ B(\mathbf{u}, j) &= \sin(\mathbf{u}^T \mathbf{X}_j) - \mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_j). \end{aligned}$$

Platí, že $A(\mathbf{u}, 1), \dots, A(\mathbf{u}, n)$ jsou nezávislé, centrované a $B(\mathbf{u}, 1), \dots, B(\mathbf{u}, n)$ jsou nezávislé, centrované. Pro $j_1 \neq j_2$ platí:

$$\begin{aligned} \mathbb{E}A(\mathbf{u}, j_1)A(\mathbf{u}, j_2) &= 0, \\ \mathbb{E}A(\mathbf{u}, j_1)B(\mathbf{u}, j_2) &= 0, \\ \mathbb{E}B(\mathbf{u}, j_1)B(\mathbf{u}, j_2) &= 0. \end{aligned}$$

S nově zavedeným značením pokračujeme ve výpočtu:

$$\begin{aligned} & \mathbb{E} \left| \frac{1}{n-h} \sum_{j=1}^{n-h} [e^{i\mathbf{u}^T \mathbf{X}_j} - \varphi_{\mathbf{X}}(\mathbf{u})] \right|^4 \\ &= \frac{1}{(n-h)^4} \mathbb{E} \left(\sum_{j_1=1}^{n-h} \sum_{j_2=1}^{n-h} A(\mathbf{u}, j_1)A(\mathbf{u}, j_2) + \sum_{j_1=1}^{n-h} \sum_{j_2=1}^{n-h} B(\mathbf{u}, j_1)B(\mathbf{u}, j_2) \right)^2 \\ &= \frac{1}{(n-h)^4} \mathbb{E} \left(\sum_{j_1=1}^{n-h} \sum_{j_2=1}^{n-h} \sum_{j_3=1}^{n-h} \sum_{j_4=1}^{n-h} A(\mathbf{u}, j_1)A(\mathbf{u}, j_2)A(\mathbf{u}, j_3)A(\mathbf{u}, j_4) \right. \\ & \quad + \sum_{j_1=1}^{n-h} \sum_{j_2=1}^{n-h} \sum_{j_3=1}^{n-h} \sum_{j_4=1}^{n-h} B(\mathbf{u}, j_1)B(\mathbf{u}, j_2)B(\mathbf{u}, j_3)B(\mathbf{u}, j_4) \\ & \quad \left. + 2 \sum_{j_1=1}^{n-h} \sum_{j_2=1}^{n-h} \sum_{j_3=1}^{n-h} \sum_{j_4=1}^{n-h} A(\mathbf{u}, j_1)A(\mathbf{u}, j_2)B(\mathbf{u}, j_3)B(\mathbf{u}, j_4) \right) \\ &= \frac{1}{(n-h)^4} \left(\sum_{j_1=1}^{n-h} \mathbb{E}A^4(\mathbf{u}, j_1) + 3 \sum_{j_1=1}^{n-h} \sum_{j_2=1}^{n-h} I\{j_1 \neq j_2\} \mathbb{E}A^2(\mathbf{u}, j_1) \mathbb{E}A^2(\mathbf{u}, j_2) \right. \\ & \quad + \sum_{j_1=1}^{n-h} \mathbb{E}B^4(\mathbf{u}, j_1) + 3 \sum_{j_1=1}^{n-h} \sum_{j_2=1}^{n-h} I\{j_1 \neq j_2\} \mathbb{E}B^2(\mathbf{u}, j_1) \mathbb{E}B^2(\mathbf{u}, j_2) \\ & \quad + 2 \sum_{j_1=1}^{n-h} \mathbb{E}A^2(\mathbf{u}, j_1) \mathbb{E}B^2(\mathbf{u}, j_1) \\ & \quad + 2 \sum_{j_1=1}^{n-h} \sum_{j_2=1}^{n-h} I\{j_1 \neq j_2\} \mathbb{E}A^2(\mathbf{u}, j_1) \mathbb{E}B^2(\mathbf{u}, j_2) \\ & \quad \left. + 4 \sum_{j_1=1}^{n-h} \sum_{j_2=1}^{n-h} I\{j_1 \neq j_2\} \mathbb{E}[A(\mathbf{u}, j_1)B(\mathbf{u}, j_1)] \mathbb{E}[A(\mathbf{u}, j_2)B(\mathbf{u}, j_2)] \right) \end{aligned}$$

Díky centrovanosti $A(\mathbf{u}, j), B(\mathbf{u}, j)$ nám vypadly všechny sčítance, které měly všechny indexy různé nebo měly alespoň jeden index různý od všech ostatních. Jelikož $A(\mathbf{u}, j)$ a $B(\mathbf{u}, j)$ mají konečné čtvrté momenty, existuje taková konstanta K , která omezí všechny momenty a součiny momentů uvnitř sum. Dostali jsme tak, že:

$$\mathbb{E} \left| \frac{1}{n-h} \sum_{j=1}^{n-h} [e^{i\mathbf{u}^T \mathbf{X}_j} - \varphi_{\mathbf{X}}(\mathbf{u})] \right|^4 \leq \frac{K}{(n-h)^4} [(n-h) + 3(n-h)(n-h-1)]$$

$$\begin{aligned}
& + (n-h) + 3(n-h)(n-h-1) \\
& + 2(n-h) \\
& + 2(n-h)(n-h-1) \\
& + 4(n-h)(n-h-1)].
\end{aligned}$$

Zvolme tedy konstantu C takovou, že platí:

$$\mathbb{E} \left| \frac{1}{n-h} \sum_{j=1}^{n-h} [e^{i\mathbf{u}^T \mathbf{x}_j} - \varphi_{\mathbf{x}}(\mathbf{u})] \right|^4 \leq C \frac{(n-h)^2}{(n-h)^4}.$$

Získali jsme tak:

$$\mathbb{E} |S_{2;n,h}(\mathbf{u})|^2 \leq C \frac{1}{(n-h)^2}.$$

Jelikož $\mathbb{E} |S_{2;n,h}(\mathbf{u})|^2 = \mathbb{E} [\Re(S_{2;n,h}(\mathbf{u}))]^2 + \mathbb{E} [\Im(S_{2;n,h}(\mathbf{u}))]^2$, platí také následující nerovnost:

$$\begin{aligned}
\mathbb{E} [\Re(S_{2;n,h}(\mathbf{u}))]^2 & \leq C \frac{1}{(n-h)^2}, \\
\mathbb{E} [\Im(S_{2;n,h}(\mathbf{u}))]^2 & \leq C \frac{1}{(n-h)^2}.
\end{aligned}$$

Z Markovovy nerovnosti (viz van der Vaart, 1998, příklad 2.6) tak pro každé $h = 1, \dots, H$ a $\mathbf{u} \in \mathbb{R}^p$ získáme:

$$\begin{aligned}
|S_{2;n,h}(\mathbf{u})|^2 & = O_p(n^{-2}), \\
\Re(S_{2;n,h}(\mathbf{u})) & = O_p(n^{-1}), \\
\Im(S_{2;n,h}(\mathbf{u})) & = O_p(n^{-1}).
\end{aligned}$$

Abychom omezili v pravděpodobnosti výraz $[\Re(S_{1;n,h}(\mathbf{u}))]^2$, ukážeme omezenost jeho střední hodnoty, kterou si rovnou rozepíšeme ve zkrácené formě:

$$\begin{aligned}
& \mathbb{E} [\Re(S_{1;n,h}(\mathbf{u}))]^2 \\
& = \mathbb{E} \left[\frac{1}{n-h} \sum_{j=1}^{n-h} (A(\mathbf{u}, j)A(\mathbf{u}, j+h) - B(\mathbf{u}, j)B(\mathbf{u}, j+h)) \right]^2 \\
& = \frac{1}{(n-h)^2} \mathbb{E} \sum_{j_1=1}^{n-h} \sum_{j_2=1}^{n-h} (A(\mathbf{u}, j_1)A(\mathbf{u}, j_1+h) - B(\mathbf{u}, j_1)B(\mathbf{u}, j_1+h)) \\
& \quad \times (A(\mathbf{u}, j_2)A(\mathbf{u}, j_2+h) - B(\mathbf{u}, j_2)B(\mathbf{u}, j_2+h)) \\
& = \frac{1}{(n-h)^2} \mathbb{E} \sum_{j_1=1}^{n-h} \sum_{j_2=1}^{n-h} A(\mathbf{u}, j_1)A(\mathbf{u}, j_1+h)A(\mathbf{u}, j_2)A(\mathbf{u}, j_2+h) \\
& \quad - A(\mathbf{u}, j_1)A(\mathbf{u}, j_1+h)B(\mathbf{u}, j_2)B(\mathbf{u}, j_2+h) \\
& \quad - B(\mathbf{u}, j_1)B(\mathbf{u}, j_1+h)A(\mathbf{u}, j_2)A(\mathbf{u}, j_2+h) \\
& \quad + B(\mathbf{u}, j_1)B(\mathbf{u}, j_1+h)B(\mathbf{u}, j_2)B(\mathbf{u}, j_2+h) \\
& = \frac{1}{(n-h)^2} \sum_{j_1=1}^{n-h} \mathbb{E} A^2(\mathbf{u}, j_1) \mathbb{E} A^2(\mathbf{u}, j_1+h)
\end{aligned}$$

$$\begin{aligned}
& - 2\mathbb{E}[A(\mathbf{u}, j_1)B(\mathbf{u}, j_1)] \mathbb{E}[A(\mathbf{u}, j_{1+h})B(\mathbf{u}, j_{1+h})] \\
& + \mathbb{E}B^2(\mathbf{u}, j_1) \mathbb{E}B^2(\mathbf{u}, j_{1+h}) \\
& + \frac{1}{(n-h)^2} \mathbb{E} \sum_{j_1=1}^{n-h} \sum_{j_2=1}^{n-h} I\{j_1 \neq j_2, |j_1 - j_2| \leq H\} \\
& \quad \times \left(A(\mathbf{u}, j_1)A(\mathbf{u}, j_{1+h})A(\mathbf{u}, j_2)A(\mathbf{u}, j_{2+h}) \right. \\
& \quad - A(\mathbf{u}, j_1)A(\mathbf{u}, j_{1+h})B(\mathbf{u}, j_2)B(\mathbf{u}, j_{2+h}) \\
& \quad - B(\mathbf{u}, j_1)B(\mathbf{u}, j_{1+h})A(\mathbf{u}, j_2)A(\mathbf{u}, j_{2+h}) \\
& \quad \left. + B(\mathbf{u}, j_1)B(\mathbf{u}, j_{1+h})B(\mathbf{u}, j_2)B(\mathbf{u}, j_{2+h}) \right) \\
& + \frac{1}{(n-h)^2} \mathbb{E} \sum_{j_1=1}^{n-h} \sum_{j_2=1}^{n-h} I\{|j_1 - j_2| > H\} \\
& \quad \times \left(A(\mathbf{u}, j_1)A(\mathbf{u}, j_{1+h})A(\mathbf{u}, j_2)A(\mathbf{u}, j_{2+h}) \right. \\
& \quad - A(\mathbf{u}, j_1)A(\mathbf{u}, j_{1+h})B(\mathbf{u}, j_2)B(\mathbf{u}, j_{2+h}) \\
& \quad - B(\mathbf{u}, j_1)B(\mathbf{u}, j_{1+h})A(\mathbf{u}, j_2)A(\mathbf{u}, j_{2+h}) \\
& \quad \left. + B(\mathbf{u}, j_1)B(\mathbf{u}, j_{1+h})B(\mathbf{u}, j_2)B(\mathbf{u}, j_{2+h}) \right)
\end{aligned}$$

Za podmínky $|j_1 - j_2| > H$ jsou všechny indexy $j_1, j_{1+h}, j_2, j_{2+h}$ různé, a proto $L_{3;n,h} = 0$. V případě podmínky $|j_1 - j_2| \leq H$ pro $j_1 \neq j_2$ může nastat situace $j_1 = j_{2+h}$, další dva indexy jsou ale stále různé. Proto platí:

$$\mathbb{E}[A(\mathbf{u}, j_1)A(\mathbf{u}, j_{1+h})A(\mathbf{u}, j_2)A(\mathbf{u}, j_{2+h})] = \mathbb{E}A^2(\mathbf{u}, j_1) \mathbb{E}A(\mathbf{u}, j_{1+h}) \mathbb{E}A(\mathbf{u}, j_2),$$

kde $\mathbb{E}A(\mathbf{u}, j_{1+h}) = 0, \mathbb{E}A(\mathbf{u}, j_2) = 0$. Získali jsme tak $L_{2;n,h} = 0$ a zbývá nám pouze člen $L_{1;n,h}$:

$$\begin{aligned}
& \mathbb{E}[\mathfrak{R}(S_{1;n,h}(\mathbf{u}))]^2 \\
& = \frac{1}{(n-h)^2} \sum_{j=1}^{n-h} \left(\mathbb{E}A^2(\mathbf{u}, j) \right)^2 - 2 \left(\mathbb{E}[A(\mathbf{u}, j)B(\mathbf{u}, j)] \right)^2 + \left(\mathbb{E}B^2(\mathbf{u}, j) \right)^2 \\
& \leq K \frac{n-h}{(n-h)^2}.
\end{aligned}$$

Jelikož $A(\mathbf{u}, j)$ a $B(\mathbf{u}, j)$ mají konečné druhé momenty, existuje konstanta K , kterou jsme mohli celý výraz v sumě omezit. Analogicky lze odvodit omezenost pro výraz $\mathbb{E}[\mathfrak{R}(S_{1;n,h}(\mathbf{u}))]^2$:

$$\begin{aligned}
& \mathbb{E}[\mathfrak{R}(S_{1;n,h}(\mathbf{u}))]^2 \\
& = \mathbb{E} \left[\frac{1}{n-h} \sum_{j=1}^{n-h} \left(A(\mathbf{u}, j)B(\mathbf{u}, j+h) - B(\mathbf{u}, j)A(\mathbf{u}, j+h) \right) \right]^2 \\
& \leq K \frac{n-h}{(n-h)^2}.
\end{aligned}$$

Opět využijeme Markovovy nerovnosti (viz van der Vaart, 1998, příklad 2.6). Pro každé $h = 1, \dots, H$ a $\mathbf{u} \in \mathbb{R}^p$ potom platí:

$$\mathfrak{R}(S_{1;n,h}(\mathbf{u})) = O_p(n^{-1/2}),$$

$$\Im(S_{1;n,h}(\mathbf{u})) = O_p(n^{-1/2}).$$

Celkem tedy máme:

$$\begin{aligned} T_n^* &= T_{n,1}^* + \frac{1}{\sqrt{H}} \sum_{h=1}^H (n-h) \int_{\mathbb{R}^p} -2\Re(S_{1;n,h}(\mathbf{u}))\Re(S_{2;n,h}(\mathbf{u})) \\ &\quad - 2\Im(S_{1;n,h}(\mathbf{u}))\Im(S_{2;n,h}(\mathbf{u})) \\ &\quad + |S_{2;n,h}(\mathbf{u})|^2 w(\mathbf{u}) d\mathbf{u} \\ &= T_{n,1}^* + \frac{1}{\sqrt{H}} \sum_{h=1}^H (n-h) [O_p(n^{-1/2})O_p(n^{-1}) + O_p(n^{-2})] \\ &= T_{n,1}^* + \frac{1}{\sqrt{H}} \sum_{h=1}^H (n-h) [O_p(n^{-3/2}) + O_p(n^{-2})] \\ &= T_{n,1}^* + \frac{1}{\sqrt{H}} \sum_{h=1}^H (n-h) O_p(n^{-3/2}) \\ &= T_{n,1}^* + \frac{\sqrt{H}}{H} \sum_{h=1}^H O_p(n^{-1/2}) \\ &= T_{n,1}^* + O_p(H^{1/2}n^{-\frac{1}{2}}). \end{aligned}$$

□

Dále tedy budeme pracovat pouze s $T_{n,1}^*$, které přepíšeme za pomoci přímých ale obsáhlých úprav do následujícího tvaru (viz Hlávka a kol., 2021, str. 624):

$$\begin{aligned} T_{n,1}^* &= \frac{1}{\sqrt{H}} \sum_{h=1}^H (n-h) \int_{\mathbb{R}^p} \left[\frac{1}{n-h} \sum_{j=1}^{n-h} \tilde{g}(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h}) \right]^2 w(\mathbf{u}) d\mathbf{u}, \\ \tilde{g}(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h}) &= g(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h}) - \mathbb{E}[g(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h}) | \mathbf{X}_j] \\ &\quad - \mathbb{E}[g(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h}) | \mathbf{X}_{j+h}] + \mathbb{E}g(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h}), \\ g(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h}) &= \cos(\mathbf{u}^T(\mathbf{X}_j + \mathbf{X}_{j+h})) + \sin(\mathbf{u}^T(\mathbf{X}_j + \mathbf{X}_{j+h})). \end{aligned}$$

Nyní rozložme $T_{n,1}^*$ na 3 sčítance $T_{n,11}^*$, $T_{n,12}^*$, $T_{n,13}^*$:

$$\begin{aligned} T_{n,1}^* &= \frac{1}{\sqrt{H}} \sum_{h=1}^H \frac{1}{n-h} \int_{\mathbb{R}^p} \sum_{j=1}^{n-h} \tilde{g}^2(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h}) w(\mathbf{u}) d\mathbf{u} \\ &\quad + \frac{1}{\sqrt{H}} \sum_{h=1}^H \frac{1}{n-h} \int_{\mathbb{R}^p} \sum_{j_1=1}^{n-h} \sum_{j_2=1}^{n-h} I\{|j_1 - j_2| \geq H\} \tilde{g}(\mathbf{u}, \mathbf{X}_{j_1}, \mathbf{X}_{j_1+h}) \\ &\quad \quad \quad \times \tilde{g}(\mathbf{u}, \mathbf{X}_{j_2}, \mathbf{X}_{j_2+h}) w(\mathbf{u}) d\mathbf{u} \\ &\quad + \frac{1}{\sqrt{H}} \sum_{h=1}^H \frac{1}{n-h} \int_{\mathbb{R}^p} \sum_{j_1=1}^{n-h} \sum_{j_2=1}^{n-h} I\{j_1 \neq j_2; |j_1 - j_2| < H\} \tilde{g}(\mathbf{u}, \mathbf{X}_{j_1}, \mathbf{X}_{j_1+h}) \\ &\quad \quad \quad \times \tilde{g}(\mathbf{u}, \mathbf{X}_{j_2}, \mathbf{X}_{j_2+h}) w(\mathbf{u}) d\mathbf{u}. \end{aligned}$$

Z následujícího lemmatu 8 plyne, že $\mathbb{E}T_{n,12}^* = 0$ a $\mathbb{E}T_{n,13}^* = 0$. Abychom odvodili střední hodnotu testové statistiky $\Delta_{n,H;p}^*$, zbývá napočítat $\mathbb{E}T_{n,11}^*$. Výpočet $\mathbb{E}T_{n,11}^*$ provedeme v důkazu věty 9.

Lemma 8 (Hlávka a kol., 2021, str. 624). *Nechť $\{\mathbf{X}_j, j = 1, \dots, n\}$ je reálnou stacionární časovou řadou p -rozměrných náhodných vektorů pro p pevné a $w(\cdot)$ je nezáporná funkce na \mathbb{R}^p , která splňuje:*

$$0 < \int_{\mathbb{R}^p} w(\mathbf{u}) d\mathbf{u} < \infty, \quad w(\mathbf{u}) = w(-\mathbf{u}) \text{ pro každé } \mathbf{u} \in \mathbb{R}^p.$$

Potom za předpokladu platnosti H_0 pro $h = 1, \dots, H$ a $j_1, j_2 = 1, \dots, n-h$ takové, že $j_1 \neq j_2$ platí:

$$\begin{aligned} \mathbb{E}\tilde{g}(\mathbf{u}; \mathbf{X}_{j_1}, \mathbf{X}_{j_1+h}) &= 0, \\ \mathbb{E}[\tilde{g}(\mathbf{u}; \mathbf{X}_{j_1}, \mathbf{X}_{j_1+h})\tilde{g}(\mathbf{u}, \mathbf{X}_{j_2}, \mathbf{X}_{j_2+h})] &= 0. \end{aligned}$$

Věta 9. *Nechť $\{\mathbf{X}_j, j = 1, \dots, n\}$ je reálnou stacionární časovou řadou p -rozměrných náhodných vektorů pro p pevné a $w(\cdot)$ je nezáporná funkce na \mathbb{R}^p , která splňuje:*

$$0 < \int_{\mathbb{R}^p} w(\mathbf{u}) d\mathbf{u} < \infty, \quad w(\mathbf{u}) = w(-\mathbf{u}) \text{ pro každé } \mathbf{u} \in \mathbb{R}^p.$$

Potom za předpokladu platnosti H_0 platí:

$$\mathbb{E}T_{n,11}^* = \sqrt{H} \left[1 - \int_{\mathbb{R}^p} L_p(\mathbf{u})(2 - L_p(\mathbf{u})) w(\mathbf{u}) d\mathbf{u} \right],$$

kde $L_p(\mathbf{u}) = (\mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_1))^2 + (\mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_1))^2$.

Důkaz.

$$\mathbb{E}T_{n,11}^* = \frac{1}{\sqrt{H}} \sum_{h=1}^H \frac{1}{n-h} \sum_{j=1}^{n-h} \mathbb{E} \int_{\mathbb{R}^p} \tilde{g}^2(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h}) w(\mathbf{u}) d\mathbf{u}$$

Z Fubiniho věty (viz Roussas, 2005, kapitola 5, věta 12) jsme oprávněni zaměnit pořadí střední hodnoty a integrálu, jelikož výsledná střední hodnota je konečná. Potřebujeme umocnit součet následujících výrazů:

$$\begin{aligned} \tilde{g}(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h}) &= g(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h}) - \mathbb{E}[g(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h}) | \mathbf{X}_j] \\ &\quad - \mathbb{E}[g(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h}) | \mathbf{X}_{j+h}] + \mathbb{E}g(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h}) \end{aligned}$$

$$\begin{aligned} \tilde{g}^2(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h}) &= g^2(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h}) + [\mathbb{E}[g(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h}) | \mathbf{X}_j]]^2 \\ &\quad + [\mathbb{E}[g(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h}) | \mathbf{X}_{j+h}]]^2 + [\mathbb{E}g(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h})]^2 \\ &\quad - 2g(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h}) \mathbb{E}[g(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h}) | \mathbf{X}_j] \\ &\quad - 2g(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h}) \mathbb{E}[g(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h}) | \mathbf{X}_{j+h}] \\ &\quad + g(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h}) \mathbb{E}g(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h}) \\ &\quad + \mathbb{E}[g(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h}) | \mathbf{X}_j] \mathbb{E}[g(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h}) | \mathbf{X}_{j+h}] \\ &\quad - \mathbb{E}[g(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h}) | \mathbf{X}_j] \mathbb{E}g(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h}) \\ &\quad - \mathbb{E}[g(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h}) | \mathbf{X}_{j+h}] \mathbb{E}g(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h}) \end{aligned}$$

Uvažujeme-li střední hodnotu z $\tilde{g}^2(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h})$, některé sčítance jsou si potom rovny. Snadno nahlédneme rovnost následujících výrazů:

$$[\mathbb{E}g(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h})]^2 = \mathbb{E}[\mathbb{E}g(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h})]^2,$$

$$\begin{aligned}
& - \sin^2(\mathbf{u}^T \mathbf{X}_j) \sin(\mathbf{u}^T \mathbf{X}_{j+h}) \mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_{j+h}) \\
& - \sin^2(\mathbf{u}^T \mathbf{X}_j) \mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_{j+h}) \cos(\mathbf{u}^T \mathbf{X}_{j+h}) \\
& - 2 \sin(\mathbf{u}^T \mathbf{X}_j) \cos(\mathbf{u}^T \mathbf{X}_j) \sin(\mathbf{u}^T \mathbf{X}_{j+h}) \mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_{j+h}) \\
& + \sin(\mathbf{u}^T \mathbf{X}_j) \cos(\mathbf{u}^T \mathbf{X}_j) \sin(\mathbf{u}^T \mathbf{X}_{j+h}) \mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_{j+h}) \\
& + \sin(\mathbf{u}^T \mathbf{X}_j) \cos(\mathbf{u}^T \mathbf{X}_j) \mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_{j+h}) \cos(\mathbf{u}^T \mathbf{X}_{j+h}) \Big] \\
& = \mathbb{E}[\mathbb{E}[g(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h}) | \mathbf{X}_j]]^2
\end{aligned}$$

Analogicky se ověří platnost rovnosti:

$$\mathbb{E}[\mathbb{E}[g(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h}) | \mathbf{X}_{j+h}]]^2 = \mathbb{E}[g(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h}) \mathbb{E}[g(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h}) | \mathbf{X}_{j+h}]].$$

Zjednodušili jsme tak výpočet střední hodnoty na součet čtyř výrazů a za pomoci předpokladu stacionarity dostaneme součet tří výrazů:

$$\begin{aligned}
\mathbb{E}g^2(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h}) & = \mathbb{E}g^2(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h}) - \mathbb{E}[\mathbb{E}[g(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h}) | \mathbf{X}_j]]^2 \\
& \quad - \mathbb{E}[\mathbb{E}[g(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h}) | \mathbf{X}_{j+h}]]^2 + [\mathbb{E}g(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h})]^2 \\
& = \mathbb{E}g^2(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h}) - 2\mathbb{E}[\mathbb{E}[g(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h}) | \mathbf{X}_j]]^2 \\
& \quad + [\mathbb{E}g(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h})]^2
\end{aligned}$$

Napočítáme požadovaný integrál pro každý sčítanec zvlášť. Nejsnáze napočítáme integrál prvního sčítance $\mathbb{E}g^2(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h})$:

$$\begin{aligned}
\mathbb{E}g^2(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h}) & = \mathbb{E}[\cos(\mathbf{u}^T (\mathbf{X}_j + \mathbf{X}_{j+h})) + \sin(\mathbf{u}^T (\mathbf{X}_j + \mathbf{X}_{j+h}))]^2 \\
& = \mathbb{E}[1 + 2 \sin(\mathbf{u}^T (\mathbf{X}_j + \mathbf{X}_{j+h})) \cos(\mathbf{u}^T (\mathbf{X}_j + \mathbf{X}_{j+h}))] \\
& = 1 + \mathbb{E} \sin(2\mathbf{u}^T (\mathbf{X}_j + \mathbf{X}_{j+h}))
\end{aligned}$$

Potom pro $\forall j = 1, \dots, n-h$ a $\forall h = 1, \dots, H$ platí:

$$\begin{aligned}
\int_{\mathbb{R}^p} \mathbb{E}g^2(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h}) w(\mathbf{u}) d\mathbf{u} & = 1 + \int_{\mathbb{R}^p} \mathbb{E} \sin(2\mathbf{u}^T (\mathbf{X}_j + \mathbf{X}_{j+h})) w(\mathbf{u}) d\mathbf{u} \\
& = 1
\end{aligned}$$

Využili jsme Fubiniho věty (viz Roussas, 2005, kapitola 5, věta 12) pro záměnu pořadí střední hodnoty a integrálu. Integrovali jsme tak přes lichý součin liché funkce $\sin(2\mathbf{u}^T (\mathbf{X}_j + \mathbf{X}_{j+h}))$ a sudé funkce $w(\mathbf{u})$. Tudíž integrál součinu je nulový.

Nyní se zaměříme na další sčítanec:

$$\begin{aligned}
& 2[\mathbb{E}[g(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h}) | \mathbf{X}_j]]^2 \\
& = 2 \left[\cos(\mathbf{u}^T \mathbf{X}_j) \mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_{j+h}) - \sin(\mathbf{u}^T \mathbf{X}_j) \mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_{j+h}) \right. \\
& \quad \left. + \sin(\mathbf{u}^T \mathbf{X}_j) \mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_{j+h}) + \cos(\mathbf{u}^T \mathbf{X}_j) \mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_{j+h}) \right]^2 \\
& = 2 \left[\cos(\mathbf{u}^T \mathbf{X}_j) (\mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_{j+h}) + \mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_{j+h})) \right. \\
& \quad \left. + \sin(\mathbf{u}^T \mathbf{X}_j) (\mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_{j+h}) - \mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_{j+h})) \right]^2 \\
& = 2 \left[\cos^2(\mathbf{u}^T \mathbf{X}_j) (\mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_{j+h}) + \mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_{j+h}))^2 \right. \\
& \quad \left. + \sin^2(\mathbf{u}^T \mathbf{X}_j) (\mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_{j+h}) - \mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_{j+h}))^2 \right. \\
& \quad \left. + 2 \sin(\mathbf{u}^T \mathbf{X}_j) \cos(\mathbf{u}^T \mathbf{X}_j) [(\mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_{j+h}))^2 - (\mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_{j+h}))^2] \right]
\end{aligned}$$

$$\begin{aligned}
&= \cos^2(\mathbf{u}^T \mathbf{X}_j) (\mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_{j+h}) + \mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_{j+h}))^2 \\
&\quad + (1 - \cos^2(\mathbf{u}^T \mathbf{X}_j)) (\mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_{j+h}) - \mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_{j+h}))^2 \\
&\quad + (1 - \sin^2(\mathbf{u}^T \mathbf{X}_j)) (\mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_{j+h}) + \mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_{j+h}))^2 \\
&\quad + \sin^2(\mathbf{u}^T \mathbf{X}_j) (\mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_{j+h}) - \mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_{j+h}))^2 \\
&\quad + 2 \sin(2\mathbf{u}^T \mathbf{X}_j) [(\mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_{j+h}))^2 - (\mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_{j+h}))^2] \\
&= (\mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_{j+h}) - \mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_{j+h}))^2 \\
&\quad + 4 \cos^2(\mathbf{u}^T \mathbf{X}_j) \mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_{j+h}) \mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_{j+h}) \\
&\quad + (\mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_{j+h}) + \mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_{j+h}))^2 \\
&\quad - 4 \sin^2(\mathbf{u}^T \mathbf{X}_j) \mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_{j+h}) \mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_{j+h}) \\
&\quad + 2 \sin(2\mathbf{u}^T \mathbf{X}_j) [(\mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_{j+h}))^2 - (\mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_{j+h}))^2] \\
&= 2(\mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_{j+h}))^2 + 2(\mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_{j+h}))^2 \\
&\quad + 4 \cos(2\mathbf{u}^T \mathbf{X}_j) \mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_{j+h}) \mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_{j+h}) \\
&\quad + 2 \sin(2\mathbf{u}^T \mathbf{X}_j) [(\mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_{j+h}))^2 - (\mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_{j+h}))^2]
\end{aligned}$$

Napočítejme ještě střední hodnotu tohoto výrazu:

$$\begin{aligned}
&2\mathbb{E}[\mathbb{E}[g(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h}) | \mathbf{X}_j]]^2 \\
&= 2[\mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_{j+h})]^2 + 2[\mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_{j+h})]^2 \\
&\quad + 4\mathbb{E} \cos(2\mathbf{u}^T \mathbf{X}_j) \mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_{j+h}) \mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_{j+h}) \\
&\quad + 2\mathbb{E} \sin(2\mathbf{u}^T \mathbf{X}_j) [(\mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_{j+h}))^2 - (\mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_{j+h}))^2]
\end{aligned}$$

Potom pro $\forall j = 1, \dots, n-h$ a $\forall h = 1, \dots, H$ platí:

$$\begin{aligned}
&\int_{\mathbb{R}^p} 2\mathbb{E}[\mathbb{E}[g(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h}) | \mathbf{X}_j]]^2 w(\mathbf{u}) d\mathbf{u} \\
&= \int_{\mathbb{R}^p} 2[\mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_{j+h})]^2 + 2[\mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_{j+h})]^2 w(\mathbf{u}) d\mathbf{u} \\
&\quad + 4 \int_{\mathbb{R}^p} \mathbb{E} \cos(2\mathbf{u}^T \mathbf{X}_j) \mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_{j+h}) \mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_{j+h}) w(\mathbf{u}) d\mathbf{u} \\
&\quad + 2 \int_{\mathbb{R}^p} \mathbb{E} \sin(2\mathbf{u}^T \mathbf{X}_j) [(\mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_{j+h}))^2 - (\mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_{j+h}))^2] w(\mathbf{u}) d\mathbf{u} \\
&= \int_{\mathbb{R}^p} 2[\mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_{j+h})]^2 + 2[\mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_{j+h})]^2 w(\mathbf{u}) d\mathbf{u}
\end{aligned}$$

Nulovost druhého integrálu v součtu můžeme nahlédnout s pomocí Fubiniho věty následovně:

$$\begin{aligned}
&\int_{\mathbb{R}^p} \mathbb{E} \cos(2\mathbf{u}^T \mathbf{X}_j) \mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_{j+h}) \mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_{j+h}) w(\mathbf{u}) d\mathbf{u} \\
&= \int_{\mathbb{R}^p} \left[\int_{\mathbb{R}^p} \cos(2\mathbf{u}^T \mathbf{x}) dF_{\mathbf{X}_1}(\mathbf{x}) \int_{\mathbb{R}^p} \sin(\mathbf{u}^T \mathbf{x}) dF_{\mathbf{X}_1}(\mathbf{x}) \int_{\mathbb{R}^p} \cos(\mathbf{u}^T \mathbf{x}) dF_{\mathbf{X}_1}(\mathbf{x}) \right] \\
&\quad \times w(\mathbf{u}) d\mathbf{u} \\
&= \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} \cos(2\mathbf{u}^T \mathbf{x}_1) \sin(\mathbf{u}^T \mathbf{x}_2) \cos(\mathbf{u}^T \mathbf{x}_3) dF_{\mathbf{X}_1}(\mathbf{x}_1) dF_{\mathbf{X}_1}(\mathbf{x}_2) dF_{\mathbf{X}_1}(\mathbf{x}_3) \\
&\quad \times w(\mathbf{u}) d\mathbf{u}
\end{aligned}$$

$$\begin{aligned}
&= \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} \cos(2\mathbf{u}^T \mathbf{x}_1) \sin(\mathbf{u}^T \mathbf{x}_2) \cos(\mathbf{u}^T \mathbf{x}_3) w(\mathbf{u}) d\mathbf{u} \\
&\quad \times dF_{\mathbf{X}_1}(\mathbf{x}_1) dF_{\mathbf{X}_1}(\mathbf{x}_2) dF_{\mathbf{X}_1}(\mathbf{x}_3) \\
&= 0
\end{aligned}$$

Využili jsme toho, že integrujeme lichý součin tří sudých a jedné liché funkce přes $\mathbf{u} \in \mathbb{R}^p$. Analogicky odvodíme, že:

$$\int_{\mathbb{R}^p} \mathbb{E} \sin(2\mathbf{u}^T \mathbf{X}_j) [(\mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_{j+h}))^2 - (\mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_{j+h}))^2] w(\mathbf{u}) d\mathbf{u} = 0.$$

Zbývá nám napočítat poslední ze třech sčítanců výrazu $\mathbb{E} \tilde{g}^2(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h})$:

$$\begin{aligned}
&[\mathbb{E} g(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h})]^2 \\
&= [\mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_j) (\mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_{j+h}) + \mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_{j+h})) \\
&\quad + \mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_j) (\mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_{j+h}) - \mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_{j+h}))]^2 \\
&= (\mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_j))^2 (\mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_{j+h}) + \mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_{j+h}))^2 \\
&\quad + (\mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_j))^2 (\mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_{j+h}) - \mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_{j+h}))^2 \\
&\quad + 2\mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_j) \mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_j) ((\mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_{j+h}))^2 - (\mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_{j+h}))^2) \\
&= (\mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_j))^4 + 2\mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_j) (\mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_j))^3 \\
&\quad + (\mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_j))^2 (\mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_j))^2 \\
&\quad + (\mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_j))^2 (\mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_j))^2 - 2(\mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_j))^3 \mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_j) \\
&\quad + (\mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_j))^4 \\
&\quad + 2\mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_j) (\mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_j))^3 - 2(\mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_j))^3 \mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_j) \\
&= (\mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_j))^4 + 2(\mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_j))^2 (\mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_j))^2 + (\mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_j))^4 \\
&\quad + 4\mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_j) (\mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_j))^3 - 4(\mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_j))^3 \mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_j) \\
&= [(\mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_j))^2 + (\mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_j))^2]^2 \\
&\quad + 4\mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_j) \mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_j) [\{\mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_j)\}^2 - (\mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_j))^2]
\end{aligned}$$

Potom pro $\forall j = 1, \dots, n-h$ a $\forall h = 1, \dots, H$ platí:

$$\begin{aligned}
&\int_{\mathbb{R}^p} [\mathbb{E} g(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h})]^2 w(\mathbf{u}) d\mathbf{u} \\
&= \int_{\mathbb{R}^p} [(\mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_j))^2 + (\mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_j))^2]^2 w(\mathbf{u}) d\mathbf{u} \\
&\quad + 4 \int_{\mathbb{R}^p} \mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_j) \mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_j) [(\mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_j))^2 - (\mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_j))^2] w(\mathbf{u}) d\mathbf{u} \\
&= \int_{\mathbb{R}^p} [(\mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_j))^2 + (\mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_j))^2]^2 w(\mathbf{u}) d\mathbf{u}
\end{aligned}$$

Druhý integrál součtu je opět nulový z Fubiniho věty a lichosti funkce uvnitř integrálu.

Dopracovali jsme se tak ke střední hodnotě v následujícím tvaru:

$$\mathbb{E} T_{n,11}^* = \frac{1}{\sqrt{H}} \sum_{h=1}^H \frac{1}{n-h} \sum_{j=1}^{n-h} \int_{\mathbb{R}^p} \mathbb{E} \tilde{g}^2(\mathbf{u}, \mathbf{X}_j, \mathbf{X}_{j+h}) w(\mathbf{u}) d\mathbf{u}$$

$$\begin{aligned}
&= \frac{1}{\sqrt{H}} \sum_{h=1}^H \frac{1}{n-h} \sum_{j=1}^{n-h} \int_{\mathbb{R}^p} 1 - 2[\mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_{j+h})]^2 - 2[\mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_{j+h})]^2 \\
&\quad + [(\mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_j))^2 + (\mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_j))^2]^2 w(\mathbf{u}) d\mathbf{u} \\
&= \sqrt{H} \int_{\mathbb{R}^p} 1 - 2[\mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_1)]^2 - 2[\mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_1)]^2 \\
&\quad + [(\mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_1))^2 + (\mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_1))^2]^2 w(\mathbf{u}) d\mathbf{u} \\
&= \sqrt{H} \left[1 - \int_{\mathbb{R}^p} L_p(\mathbf{u})(2 - L_p(\mathbf{u}))w(\mathbf{u}) d\mathbf{u} \right],
\end{aligned}$$

kde $L_p(\mathbf{u}) = (\mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_1))^2 + (\mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_1))^2$.

□

Odvodili jsme tak střední hodnotu testové statistiky $\Delta_{n,H;p}^*$:

$$\begin{aligned}
\mathbb{E} \Delta_{n,H;p}^* &= \mathbb{E}[\sqrt{H}(T_{n,1}^* + O_p(H^{1/2}n^{-1/2}))] \\
&= \sqrt{H} \mathbb{E} T_{n,1}^* + O(Hn^{-1/2}) \\
&= H \left[1 - \int_{\mathbb{R}^p} L_p(\mathbf{u})(2 - L_p(\mathbf{u}))w(\mathbf{u}) d\mathbf{u} \right] + O(Hn^{-1/2}),
\end{aligned}$$

kde $L_p(\mathbf{u}) = (\mathbb{E} \cos(\mathbf{u}^T \mathbf{X}_1))^2 + (\mathbb{E} \sin(\mathbf{u}^T \mathbf{X}_1))^2$.

V článku Hlávka a kol. (2021) je pro test odvozený z definice nezávislosti odvozen i rozptyl a asymptotické normální rozdělení $\Delta_{n,H;p}$ (viz Věta 5). Tyto výpočty jsou však mnohonásobně delší a výrazně by přesáhly rozsah diplomové práce. Abychom mohli využít testu odvozeného v této sekci, p-hodnoty budeme napočítávat permutačním testem. Přejdeme nyní k odvození výpočetního tvaru testové statistiky $\Delta_{n,H;p}^*$. Budeme postupovat podobně jako v sekci 2.1.2. V následujícím lemmatu 10 upravíme výraz $|D_{n,h;p}^*(\mathbf{u}, \mathbf{v})|^2$.

Lemma 10. *Nechť $\{\mathbf{X}_j, j = 1, \dots, n\}$ je časovou řadou p -rozměrných náhodných vektorů pro p pevné. Potom platí:*

$$\begin{aligned}
|D_{n,h;p}^*(\mathbf{u}, \mathbf{v})|^2 &= \frac{1}{(n-h)^2} \sum_{j,k=1}^{n-h} \cos(\mathbf{u}^T \mathbf{X}_{j,k,j+h,k+h}) \\
&\quad - \frac{2}{(n-h)^3} \sum_{j,k,l=1}^{n-h} \cos(\mathbf{u}^T \mathbf{X}_{j,k,j+h,l+h}) \\
&\quad + \frac{1}{(n-h)^4} \sum_{j,k,l,m=1}^{n-h} \cos(\mathbf{u}^T \mathbf{X}_{j,k,l+h,m+h}),
\end{aligned}$$

kde značíme zkráceně $\mathbf{X}_{j,k,l,m} = \mathbf{X}_j - \mathbf{X}_k + \mathbf{X}_l - \mathbf{X}_m$.

Důkaz. Lemma plyne triviálně z lemmatu 6, položíme-li $\mathbf{u} = \mathbf{v}$.

□

Výpočetní tvar $\int_{\mathbb{R}^p} |D_{n,h;p}^*(\mathbf{u})|^2 w(\mathbf{u}) d\mathbf{u}$ je tedy následující:

$$\int_{\mathbb{R}^p} |D_{n,h;p}^*(\mathbf{u})|^2 w(\mathbf{u}) d\mathbf{u} = \frac{1}{(n-h)^2} \sum_{j,k=1}^{n-h} I_{w,p}(\mathbf{X}_{j,k,j+h,k+h})$$

$$\begin{aligned}
& - \frac{2}{(n-h)^3} \sum_{j,k,l=1}^{n-h} I_{w,p}(\mathbf{X}_{j,k,j+h,l+h}) \\
& + \frac{1}{(n-h)^4} \sum_{j,k,l,m=1}^{n-h} I_{w,p}(\mathbf{X}_{j,k,l+h,m+h}),
\end{aligned}$$

kde $\mathbf{X}_{j,k,l,m} = \mathbf{X}_j - \mathbf{X}_k + \mathbf{X}_l - \mathbf{X}_m$ a $I_{w,p}(x) = \exp\{-\mathbf{x}^T \mathbf{V} \mathbf{x}\}$, zvolíme-li za váhovou funkci $w(\cdot)$ hustotu p -rozměrného normálního rozdělení $\mathcal{N}_p(\mathbf{0}, \mathbf{V})$. Tvar funkce $I_{w,p}(\cdot)$ byl odvozen na konci sekce 2.1.2 z tvaru charakteristické funkce rozdělení $\mathcal{N}_p(\mathbf{0}, \mathbf{V})$. Umíme tak napočítat hodnotu testové statistiky $\Delta_{n,H;p}^*$.

2.2 Porovnání testů sériové nezávislosti funkcionálních dat

V sekci 2.1 jsme si představili tři přístupy testování sériové nezávislosti funkcionálních členů časové řady. Tyto testy porovnáme na základě vypočetní složitosti, síly a pravděpodobnosti chyby prvního druhu. Test založený na korelacích (Horváth, Hušková a Rice, 2013a) budeme podle autorů zkráceně označovat test HHR, test odvozený z definice nezávislosti (Hlávka, Hušková a Meintanis, 2021) označme jako HHM a test odvozený z definice sub-nezávislosti jako HHMsub.

P-hodnoty pro výpočet odhadu síly a pravděpodobnosti chyby prvního druhu testu HHMsub napočítáme permutačním testem. Pozorujeme časovou řadu funkcionálních pozorování $\{X_j(\cdot), j = 1, \dots, n\}$, pro kterou napočítáme hodnotu testové statistiky $\Delta_{n,H;p}^*$, ozn. zkráceně $\Delta^*(X_1, \dots, X_n)$. Dále pozorování původní časové řady permutujeme, čímž získáme časové řady $\{X_j^{(b)}(\cdot), j = 1, \dots, n\}$ pro $b = 1, \dots, B$. Pro tyto časové řady napočítáme testové statistiky $\Delta_{n,H;p}^*$, ozn. $\Delta^*(X_1^{(b)}, \dots, X_n^{(b)})$. Požadovaná p-hodnota je potom následujícího tvaru:

$$\frac{1}{B} \sum_{b=1}^B I\{\Delta^*(X_1, \dots, X_n) < \Delta^*(X_1^{(b)}, \dots, X_n^{(b)})\}.$$

Jelikož z permutovaných časových řad sériová závislost zmizí, napočítávané hodnoty $\Delta^*(X_1^{(b)}, \dots, X_n^{(b)})$ tak vlastně odhadují rozdělení testové statistiky $\Delta_{n,H;p}^*$ za platnosti hypotézy o nezávislosti členů časové řady.

V článku Hlávka a kol. (2021) byla provedena simulační studie, která porovnávala testy HHM využívající asymptotického rozdělení testové statistiky $\Delta_{n,H;p}$ za platnosti hypotézy (viz věta 5) a HHM, jehož rozdělení za platnosti hypotézy je odhadnuto permutačním přístupem. Zatímco odhad pravděpodobnosti chyby prvního druhu vycházel v rozmezí 4.0% až 7.8% u testu HHM založeného na permutacích, u testu HHM využívajícího asymptotického rozdělení byl tento odhad v rozmezí 0.4% až 13.5% (viz Hlávka a kol., 2021, Tabulka 1). Dále byl zkoumán vliv volby varianční matice \mathbf{V} (viz Hlávka a kol., 2021, Tabulky 2 až 7). Na základě simulací z tohoto článku budeme porovnávat test HHMsub s testem HHM, který je také založen na permutacích, a pro oba testy určíme za kovarianční matici \mathbf{V} inverz kovarianční matice odpovídající Wienerovu procesu.

2.2.1 Výpočetní složitost testů

Celkově nejrychlejší je bezesporu test HHR. Nejenže kritickou hodnotu nepočítá permutačním testem, vychází z asymptotického rozdělení testové statistiky $V_{n,H}$, ale i výpočet samotné testové statistiky $V_{n,H}$ je velmi rychlý. Připomeňme si nyní výpočetní tvary testových statistik HHM a HHMsub:

$$\Delta_{n,H;p} = \sum_{h=1}^H (n-h) \left[\frac{1}{(n-h)^2} \sum_{j,k=1}^{n-h} I_{w,p}(\mathbf{X}_{j,k}) I_{w,p}(\mathbf{X}_{j+h,k+h}) - \frac{2}{(n-h)^3} \sum_{j,k,l=1}^{n-h} I_{w,p}(\mathbf{X}_{j,k}) I_{w,p}(\mathbf{X}_{j+h,l+h}) \right]$$

$$+ \frac{1}{(n-h)^4} \left(\sum_{j,k=1}^{n-h} I_{w,p}(\mathbf{X}_{j,k}) \right) \left(\sum_{j,k=1}^{n-h} I_{w,p}(\mathbf{X}_{j+h,k+h}) \right),$$

kde $\mathbf{X}_{j,k} = \mathbf{X}_j - \mathbf{X}_k$ a $I_{w,p}(x) = \exp\{-\mathbf{x}^T \mathbf{V} \mathbf{x}\}$.

$$\begin{aligned} \Delta_{n,H;p}^* = \sum_{h=1}^H (n-h) & \left[\frac{1}{(n-h)^2} \sum_{j,k=1}^{n-h} I_{w,p}(\mathbf{X}_{j,k,j+h,k+h}) \right. \\ & - \frac{2}{(n-h)^3} \sum_{j,k,l=1}^{n-h} I_{w,p}(\mathbf{X}_{j,k,j+h,l+h}) \\ & \left. + \frac{1}{(n-h)^4} \sum_{j,k,l,m=1}^{n-h} I_{w,p}(\mathbf{X}_{j,k,l+h,m+h}) \right], \end{aligned}$$

kde $\mathbf{X}_{j,k,l,m} = \mathbf{X}_j - \mathbf{X}_k + \mathbf{X}_l - \mathbf{X}_m$ a $I_{w,p}(x) = \exp\{-\mathbf{x}^T \mathbf{V} \mathbf{x}\}$.

Na první pohled se může jevit výpočet hodnoty $\Delta_{n,H;p}^*$ jako snazší oproti výpočtu hodnoty $\Delta_{n,H;p}$. Nicméně v posledním sčítanci je zapotřebí napočítat hodnoty $I_{w,p}$ pro všechny kombinace čtyř vektorů, tedy přibližně n^4 hodnot $I_{w,p}(\mathbf{X}_{j,k,l,m})$. Při výpočtu testové statistiky u testu HHM napočítáváme pouze n^2 hodnot $I_{w,p}(\mathbf{X}_{j,k})$. Sice s využitím symetrií $I_{w,p}(\mathbf{X}_{j,k,l,m})$ pro některé indexy j, k, l, m snížíme počet výpočtů několikanásobně v případě testu HHMsub, při vysokém n je to ale stále diametrální rozdíl. Ačkoliv byly simulace napočítány v programu R, napočítání testové statistiky bylo pro urychlení výpočtů provedeno v programu C pro oba testy HHM a HHMsub.

Při rozsahu dat $n = 30$ a volbě parametrů $B = 400, p = 50, H = 2$ je čas jedné simulace pro test HHM 0.98 sekundy a pro test HHMsub 1 minuta a 22 sekund. Při rozsahu dat $n = 50$ a stejné volbě ostatních parametrů je čas jedné simulace pro test HHM 1.25 sekund a pro test HHMsub 11 minut a 12 sekund. Aby bylo reálné provést simulace v sekcích 2.2.2 a 2.2.3 využili jsme školního výpočetního clusteru Sněhurka, kde jsme si rozdělili výpočty na menší skupinky simulací, které jsme pouštěli souběžně.

2.2.2 Empirická hladina testů

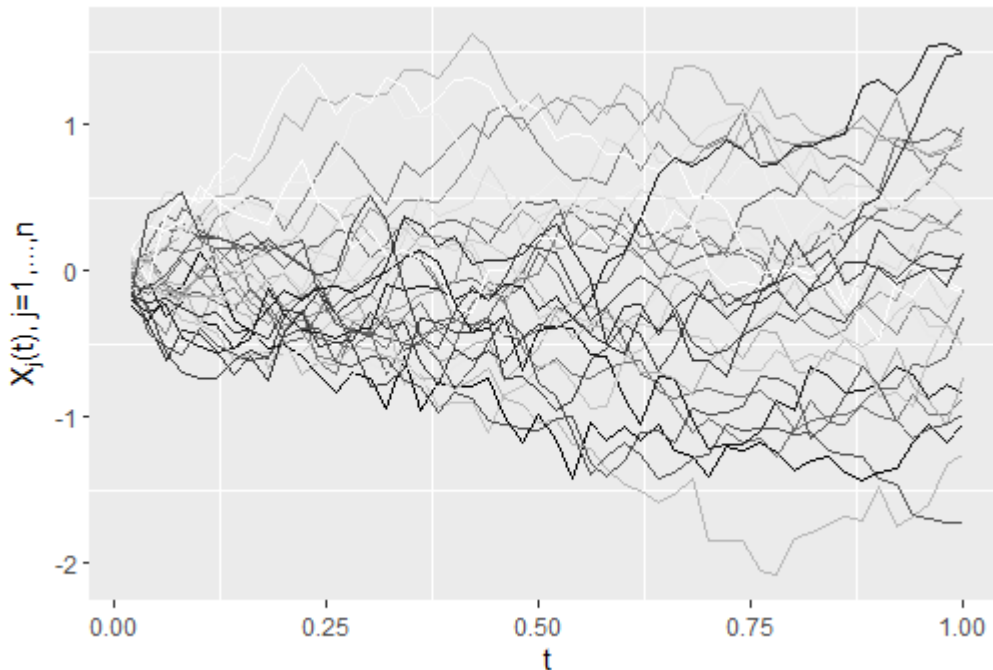
Zajímá nás, jestli testy HHR, HHM a HHMsub splňují stanovenou hladinu významnosti $\alpha = 0.05$. Zjišťujeme tedy, jaké procento platných hypotéz testy zamítnou. Budeme generovat n nezávislých stejně rozdělených funkcionálních pozorování Wienerova procesu. Pro ilustraci jsme si na Obrázku 2.1 vykreslili průběhy 30 nezávislých Wienerových procesů. Odhady hladin testů uvedené v Tabulce 2.1 byly napočítány na 200 simulacích při počtu 400 permutací v rámci jedné simulace.

Z hodnot uvedených v Tabulce 2.1 vidíme, že všechny testy poměrně dobře dodržují stanovenou hladinu $\alpha = 0.05$. Empirické hladiny se pohybují v rozmezí 3% až 6%.

Aplikujeme-li naše tři testy na konkrétní realizaci 30 nezávislých Wienerových procesů z Obrázku 2.1, dostaneme následující p-hodnoty: 0.63 pro test HHR, 0.88 pro test HHM a 0.24 pro test HHMsub. Žádný z testů tedy na těchto datech nezamítl hypotézu o nezávislosti.

n	p	H	HHR	HHM	HHMsub	
50	10	1	3.5	4.5	3.5	
		2	5.5	5.5	6.0	
	20	1	5.0	6.0	5.0	
		2	3.5	4.5	5.0	
	50	1	5.0	5.0	3.0	3.0
		2	4.5	4.5	4.5	3.0

Tabulka 2.1: Empirické hladiny (v %) testů HHR, HHM a HHMsub pro n simulovaných iid Wienerových procesů napočítané na základě 200 simulací s $B = 400$ permutacemi.



Obrázek 2.1: Realizace 30 nezávislých náhodně generovaných Wienerových procesů, $p = 50$.

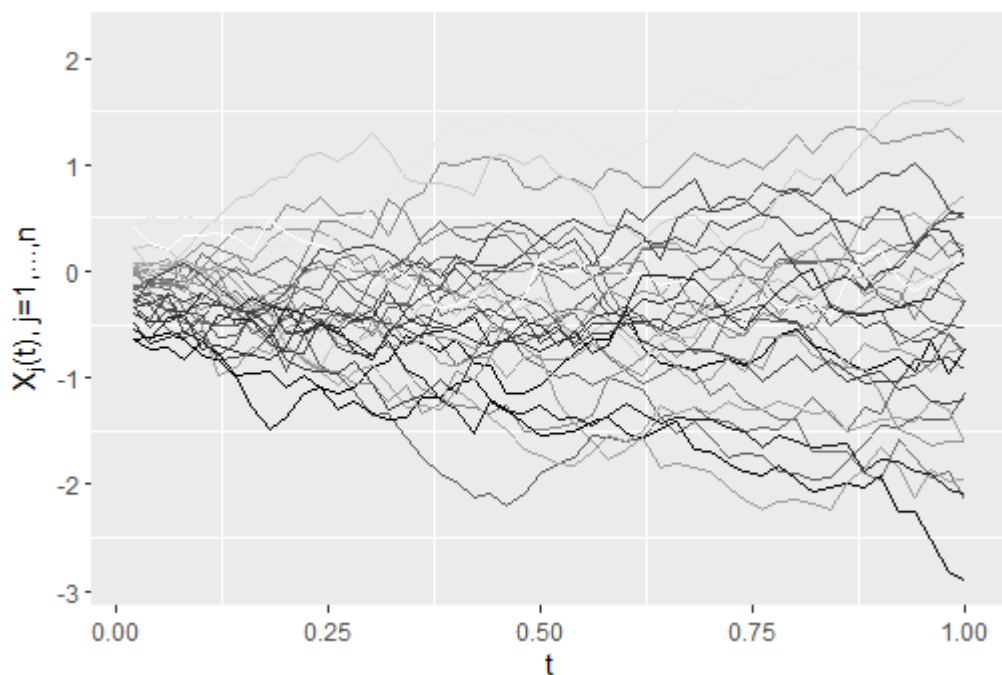
2.2.3 Síla testu

Abychom odhadli sílu testů, budeme generovat data za platnosti alternativy a zkoumat poměr zamítnutých hypotéz. Nejprve uvažujme následující funkcionální AR(1) model časové řady $\{X_j(\cdot), j = 1, \dots, n\}$, který byl použit v simulacích obou článků Horváth a kol. (2013a) a Hlávka a kol. (2021):

$$X_i(t) = \int_0^1 \Psi(s, t) X_{i-1}(s) ds + \epsilon_i(t),$$

kde $\epsilon_i(t)$ jsou iid Wienerovy procesy a $\Psi(t, s) = \pi \exp(s^2 + t^2)/2$. Na Obrázku 2.2 jsme takovou časovou řadu vykreslili pro $n = 30$.

V Tabulce 2.2 jsou uvedeny odhady síly testů HHR, HHM a HHMsub pro rozsah dat $n = 30$ a různé hodnoty parametrů p a H . Empirická síla testu se pohybuje v rozmezí 39.5% až 74.0% pro HHR test, 22.0% až 56.0% pro HHM



Obrázek 2.2: Realizace 30 pozorování z modelu FAR(1), $p = 50$.

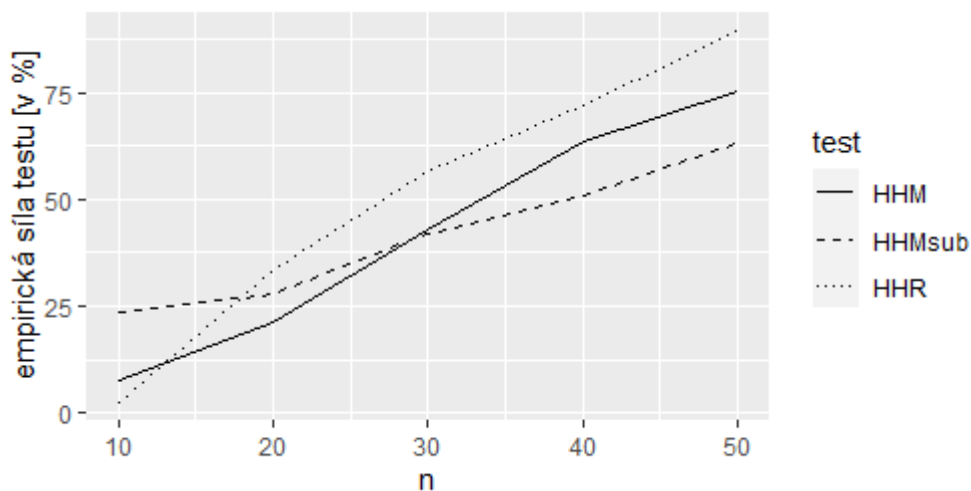
n	p	H	HHR	HHM	HHMsub
30	10	1	74.0	51.5	39.5
		2	66.5	50.0	37.5
		3	55.5	32.5	44.5
		5	45.0	24.5	40.5
	20	1	70.5	55.5	29.5
		2	60.0	39.5	37.5
		3	54.5	35.0	47.0
		5	45.5	23.5	39.5
	50	1	66.5	56.0	32.5
		2	52.0	39.5	42.0
		3	53.0	35.5	49.0
		5	39.5	22.0	41.0

Tabulka 2.2: Empirické síly (v %) testů HHR, HHM a HHMsub pro n simulovaných procesů z modelu FAR(1) napočítané na základě 200 simulací s $B = 400$ permutacemi.

test a 29.5% až 49.0% pro HHMsub test. Zdá se, že hustota mřížky, na která data pozorujeme (tzn. parametr p), nemá na sílu testu přílišný vliv. Naopak výběr rozmezí H , ve kterém sledujeme závislost členů časové řady, je pro testy HHR a HHM důležitý. Tyto testy jsou výrazně silnější, zvolíme-li H co nejmenší. Pro $H = 1$ je nejsilnější test HHR s empirickou silou 74%. U testu HHMsub se zdá, že volba parametru H nemá tak zásadní vliv. Pro $H = 3$ je tento test nejsilnější, ale pokud H dále poroste, měla by jeho síla klesat.

Na Obrázku 2.3 jsou vykresleny empirické síly testů napočítané na datech

z modelu FAR(1) pro $n = 10, 20, 30, 40, 50$ a pevnou volbu parametrů $H = 2$, $p = 50$. V souladu s očekáváním spolu s rostoucím rozsahem výběru, roste také empirická hladina všech testů. Empirická síla testu HHMsub sice také roste spolu s n , avšak rozdíl pro malé a větší rozsahy n není tak velký jako u ostatních testů. Empirická síla testu HHMsub je pro $n = 10$ výrazně lepší než u testů HHR a HHM.

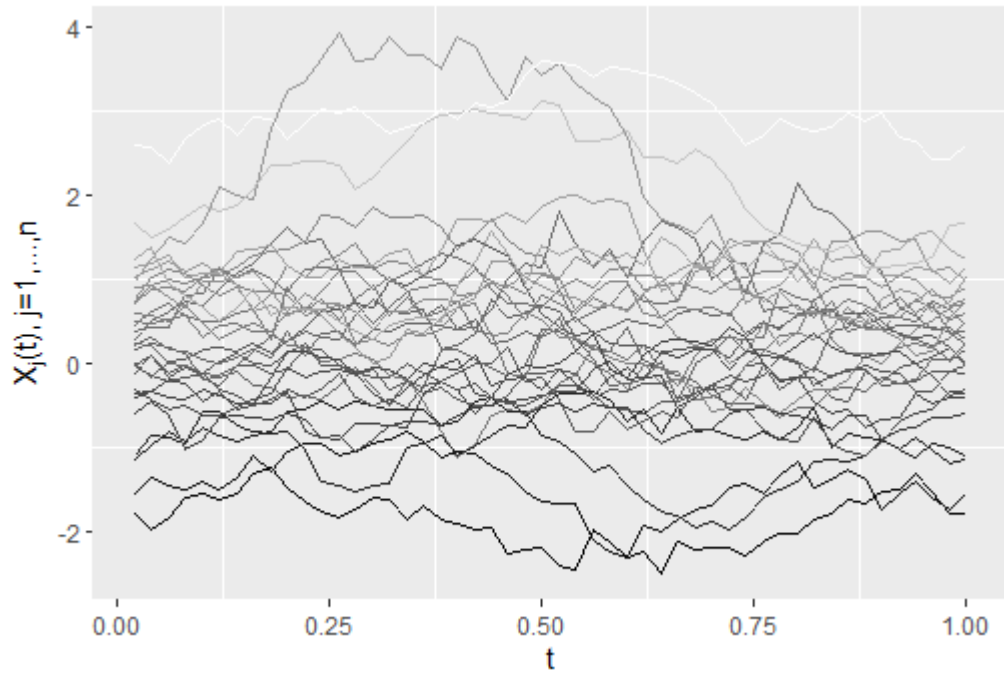


Obrázek 2.3: Empirické síly (v %) testů HHR, HHM a HHMsub pro n simulovaných procesů z modelu FAR(1) s parametry $H = 2, p = 50$ napočítané na základě 200 simulací s $B = 400$ permutacemi.

Nyní uvažujme pro časovou řadu $\{X_j(\cdot), j = 1, \dots, n\}$ funkcionální ARCH(1) model, který byl jedním z modelů použitých v simulacích článku Hlávka a kol. (2021). Závislost členů časové řady ARCH(1) modelu je definována následovně: $X_i(t) = 10\epsilon_i(t)\sigma_i(t)$, kde $\sigma_i^2(t) = \delta(t) \int_0^1 \beta(s,t) X_{i-1}^2(s) ds$ pro $\delta(t) = 0.01$, $\beta(s,t) = 16s(1-s)t(1-t)$ a $\epsilon_i(t) = B_i(t) + U_i \sqrt{1-t(1-t)}$ pro $B_i(\cdot)$ iid standardizované Brownovské můstky a U_i iid náhodné veličiny s rozdělením $\mathcal{N}(0,1)$. Tuto časovou řadu jsme si vykreslili pro $n = 30$ na Obrázku 2.4.

Vzhledem k tomu, že závislosti členů časové řady se projevují pouze na jejich rozptylu, nemůžeme očekávat velkou sílu testů. Podívejme se na empirické síly testů v Tabulce 2.3. Jako nejsilnější se jeví test HHM, jehož empirická síla se pohybuje v rozmezí 10.5% až 24.0%. Naproti tomu oba testy HHR a HHMsub jsou poměrně slabé a nestabilní.

Nevýhodou HHR a HHMsub testů je, že mohou odhalit závislost členů určité časové řady, ale již neodhalí závislost transformovaných členů stejné časové řady. Test HHR navíc není vhodný, očekáváme-li rozdělení dat s těžkými chvosty.



Obrázek 2.4: Realizace 30 pozorování z modelu FARCH(1), $p = 50$.

n	p	H	HHR	HHM	HHMsub
30	10	1	15.0	24.0	19.5
		2	6.5	20	5.0
		3	9.5	13.5	2.0
		5	6.5	10.5	7.0
	20	1	14.5	20.0	7.5
		2	5.5	19.0	6.5
		3	5.5	13.0	8.5
		5	10.5	12.5	2.5
	50	1	14.5	21.0	11.5
		2	7.5	18.5	4.0
		3	5.5	23.5	5.0
		5	10.5	16.0	6.0

Tabulka 2.3: Empirické síly (v %) testů HHR, HHM a HHMsub pro n simulovaných procesů z modelu FARCH(1) napočítané na základě 200 simulací s $B = 400$ permutacemi.

Závěr

Tématem této práce byly testy sériové nezávislosti posloupnosti funkcionálních náhodných veličin. V první části práce jsme si představili problematiku sériové nezávislosti na časových řadách náhodných veličin. Řešili jsme nejprve testy vzájemné nezávislosti složek iid vektorů, z jejichž dvourozměrného případu jsme pak vycházeli při odvozování testu sériové nezávislosti časové řady náhodných veličin. Věnovali jsme se těm testům, jejichž tvar testové statistiky je možné modifikovat pro časovou řadu funkcionálních pozorování. Zaměřili jsme se tak na testování nezávislosti pomocí nutné podmínky nekorelovanosti a na testování nezávislosti přímo z její definice. Jelikož ve funkcionálním případě není distribuční funkce dobře definována, věnovali jsme se i odvození tvaru testové statistiky z definice nezávislosti charakteristickými funkcemi. Tento test jsme modifikovali na test slabší podmínky sub-nezávislosti.

Druhá část práce už je zaměřena na testy sériové nezávislosti časové řady funkcionálních pozorování. Zde jsme se věnovali opět testu založeném na nutné podmínce nekorelovanosti, sub-nezávislosti a testu vycházejícího přímo z definice nezávislosti. Pro test sub-nezávislosti jsme odvodili střední hodnotu jeho testové statistiky. Odvození rozptylu a asymptotického rozdělení, které máme k dispozici pro test vycházející z definice nezávislosti, by však přesáhlo vhodný rozsah diplomové práce. Pro tento test jsme také odvodili výpočetní tvar testové statistiky.

Nakonec jsme tyto tři testy nezávislosti porovnali z hlediska výpočetní složitosti, empirické hladiny testů a empirické síly testů. Test založený na korelacích je výpočetně velmi snadný a rychlý i pro vysoký počet členů časové řady. Test z definice nezávislosti je již výpočetně složitější a doba napočítávání roste kvadraticky v závislosti na počtu pozorování. Ačkoliv se využití definice sub-nezávislosti zdálo jako usnadnění výpočtu testové statistiky, ukázalo se nakonec, že tento test je nejpomalejší. Napočítáváme zde totiž všechny kombinace funkce 4 funkcionálních pozorování. Pro porovnání vlastností testů jsme provedli simulaci, k jejíž realizaci jsme využili školního výpočetního clusteru Sněhurka. U testu založeném na korelacích jsme využili jeho asymptotické rozdělení. U testů odvozených z definice nezávislosti a sub-nezávislosti jsme k napočítávání kritických hodnot použili permutačního přístupu. Všechny tři testy mají empirické hladiny blízké požadované hladině 0.05. V síle testů se však již liší. Test založený na sub-nezávislosti má největší sílu, zkoumáme-li závislost členů vzdálených v časové řadě do vzdálenosti 3. Ostatní testy mají největší sílu, zkoumají-li závislost pouze 2 po sobě jdoucích pozorování, pak jejich síla výrazně klesá.

Test nezávislosti odvozený z nutné podmínky sub-nezávislosti, který jsme v této práci odvodili, je tedy použitelný v případě, že testujeme časovou řadu s malým počtem pozorování. Lze ho také použít k testu slabší hypotézy sub-nezávislosti, pokud nechceme odvozovat žádné teoretické vlastnosti tohoto testu. Použijeme-li k napočítání kritických hodnot permutační přístup, předpoklad sub-nezávislosti je dostačující.

Využití těchto tří testů nezávislosti nalezneme například v regresních modelech s funkcionální odezvou, kde je zapotřebí ověřit nezávislost chybových členů. Zajímavá by mohla být také aplikace konceptu sub-nezávislosti na problematiku testování dvojic funkcionálních pozorování.

Seznam použité literatury

- ANDERSON, R. L. (1942). Distribution of the serial correlation coefficient. *The Annals of Mathematical Statistics*, **13**(1), 1–13.
- BLUM, J. R., KIEFER, J. a ROSENBLATT, M. (1961). Distribution free tests of independence based on the sample distribution function. *The Annals of Mathematical Statistics*, **32**(2), 485–498.
- BOX, G. E. P. a PIERCE, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, **65**, 1509–1526.
- COTTERILL, D. S. a CSORGO, M. (1982). On the limiting distribution of and critical values for the multivariate Cramer-von Mises statistic. *The Annals of Statistics*, **10**(1), 233–244.
- CSÖRGŐ, S. (1985). Testing for independence by the empirical characteristic function. *Journal of Multivariate Analysis*, **16**(3), 290–299.
- DUFOUR, J. a ROY, R. (1986). Generalized portmanteau statistics and tests of randomness. *Communications in Statistics - Theory and Methods*, **15**(10), 2953–2972.
- DUGUE, D. (1969). Characteristic functions of random variables connected with Brownian motion and of the von Mises multidimensional ω_n^2 . **2**, 289–301.
- DURBIN, J. (1970). Asymptotic distributions of some statistics based on the bivariate sample distribution function. In *Nonparametric Techniques in Statistical Inference*, pages 435–451. Cambridge Univ. Press London.
- FELLER, W. (1968). *An Introduction to Probability Theory and Its Applications*, volume 1. John Wiley and Sons, New York. ISBN 0471257087.
- GABRYS, R. a KOKOSZKA, P. (2007). Portmanteau test of independence for functional observations. *Journal of the American Statistical Association*, **102** (480), 1338–1348.
- GHOUDI, K., KULPERGER, R. J. a RÉMILLARD, B. (2001). A nonparametric test of serial independence for time series and residuals. *Journal of Multivariate Analysis*, **79**(2), 191–218.
- HAMEDANI, G. G. (2015). *Sub-independence : a useful concept*. Mathematics Research Developments. Novinka, New York. ISBN 1-63463-555-8.
- HAMEDANI, G. a WALTER, G. (1984). A fixed point theorem and its application to the central limit theorem. *Archiv der Mathematik*, **43**, 258–264.
- HLÁVKA, Z., HUŠKOVÁ, M. a MEINTANIS, S. G. (2021). Testing serial independence with functional data. *Test (Madrid, Spain)*, **30**(3), 603–629.
- HOEFFDING, W. (1948). A non-parametric test of independence. *The Annals of Mathematical Statistics*, **19**(4), 546–557.

- HORVÁTH, L. a KOKOSZKA, P. (2012). *Inference for functional data with applications*. Springer series in statistics. Springer, New York. ISBN 978-1-4614-3654-6.
- HORVÁTH, L., HUŠKOVÁ, M. a RICE, G. (2013a). Test of independence for functional data. *Journal of Multivariate Analysis*, **117**, 100–119.
- HORVÁTH, L., KOKOSZKA, P. a REEDER, R. (2013b). Estimation of the mean of functional time series and a two-sample problem. *Journal of the Royal Statistical Society. Series B, Statistical methodology*, **75**(1), 103–122.
- JARNÍK, V. (1963). *Diferenciální počet I*. Československá akademie věd, Praha.
- KAN, R. a WANG, X. (2010). On the distribution of the sample autocorrelation coefficients. *Journal of Econometrics*, **154**(2), 101–121.
- KWAN, A. C. C. a SIM, A.-B. (1996). On the finite-sample distribution of modified Portmanteau tests for randomness of a Gaussian time series. *Biometrika*, **83**(4), 938–943.
- LJUNG, G. M. a BOX, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, **65**(2), 297–303.
- MELNICK, E. L. a TENENBEIN, A. (1982). Misspecifications of the normal distribution. *The American Statistician*, **36**(4), 372–373.
- RAMSAY, J. O. a SILVERMAN, B. W. (1997). *Functional data analysis*. Springer series in statistics. Springer, New York. ISBN 0-387-94956-9.
- ROUSSAS, G. G. (2005). *An introduction to measure-theoretic probability*. Elsevier Academic Press, Amsterdam. ISBN 1-280-58200-6.
- SCHENNACH, S. M. (2019). Convolution without independence. *Journal of Econometrics*, **211**(1), 308–318.
- SKAUG, H. J. a TJØSTHEIM, D. (1993). A nonparametric test of serial independence based on the empirical distribution function. *Biometrika*, **80**(3), 591–602.
- VAN DER VAART, A. W. (1998). *Asymptotic statistics*. Cambridge series in statistical and probabilistic mathematics. Cambridge University Press, Cambridge. ISBN 0-521-78450-6.

A. Přílohy

Zde nalezneme stěžejní části programů, které byly použity k napočítání testové statistiky ze sekce 2.1.3 a využity v simulacích sekce 2.2. V příloze A.1 nalezneme část skriptu, kterou jsme generovali nezávislé Wienerovy procesy a data z modelů FAR(1) a FARCH(1). Příloha A.2 obsahuje výpočet testové statistiky.

A.1 Generování dat

Tento kód byl napsán v programu R. Vstupní hodnotou je název modelu, ze kterého chceme data generovat. Výstupem jsou funkcionální data.

```
if (alt=="H0w") {
  data=matrix(NA, ncol=n.grid,nrow=n)
  for (i in 1:n) {
    data[i,]=rwiener(frequency=n.grid)
  }
}
if (alt=="H1har") {
  PSI=0.3416*outer(xx,xx,function(t,s)exp((t^2+s^2)/2))/n.grid
  data=matrix(NA, ncol=n.grid,nrow=n+1)
  data[1,]=rwiener(frequency=n.grid)
  for (i in 2:(n+1)) {
    data[i,]=PSI%%data[i-1,]+rwiener(frequency=n.grid)
  }
  data=data[2:(n+1),]
}
if (alt=="H1arhr") { #model FARCH(1)
  xxx=(0:(n.grid-1))/(n.grid-1)
  data=t(BB(N =n.grid-1,M=n+10))+
    (rnorm(n+10)%*%t(sqrt(1-xxx*(1-xxx))))
  delta=0.01
  beta=outer(xxx,xxx,function(s,t){16*s*(1-s)*t*(1-t)})
  for (i in 2:(n+10)) {
    sigi2=delta+beta%%((data[i-1,])^2)/n.grid
    data[i,]=data[i,]*sqrt(sigi2)
    # print(sigi2[10])
  }
  data=10*data[11:(n+10),]
}
```

A.2 Výpočet testové statistiky

Funkce napočítávající testovou statistiku detekující sub-závislost byla pro urychlení výpočtu napsána v programu C. Na vstupu je vyžadován rozsah dat, vzdálenost mezi náhodnými funkcemi, do které budeme zkoumat jejich závislost, a matice Jw , jejíž výpočet byl proveden v programu R a je zde také uveden. Funkce napočítávající matici Jw vyžaduje na vstupu data, kovarianční matici \mathbf{V} a parametr a , který jsme volili vždy $a = 2$.

```
Jwfun = function(data, V, a){
  n=nrow(data)
  Jw = matrix(NA, nrow=(n*(n-1))/2, ncol=(n*(n-1))/2)
  for (j in 2:n) {
    for (k in 1:(j-1)) {
      for (l in j:n) {
        for (m in max(1,sum(0:(j-2))-sum(0:(l-2))+k):(l-1)) {
          Jw[sum(0:(j-2))+k, sum(0:(l-2))+m] =
            exp((- (data[j,]-data[k,])^T
                  %% V %% (data[l,]-data[m,]))/(2*a))
          Jw[sum(0:(l-2))+m, sum(0:(j-2))+k] =
            Jw[sum(0:(j-2))+k, sum(0:(l-2))+m]
        }
      }
    }
  }
  return(Jw)
}
```

```
#include <math.h>
#define PI 3.141592653589793238462643

void subindep(double *jw, int *n, int *hh, double *stat)
{
  int j, k, l, m, h;
  double delta, s1, s2, s2_jk, s2_lm, s2_jklm,
         s3, s3_jk, s3_jl, s3_jkl;

  *stat = 0.0 ;
  delta = 0.0;
  s1 = 1.0 ;
  s2 = 0.0 ;
  s2_jk = 0.0 ;
  s2_lm = 0.0 ;
  s2_jklm = 0.0 ;
  s3 = 0.0 ;
  s3_jk = 0.0 ;
  s3_jl = 0.0 ;
  s3_jkl = 0.0 ;

  for(h = 1; h < (*hh + 1) ; h++) {
```

```

//výpočet prvního sčítance z výpočetního tvaru integral |D*|^2
for(j = 2; j < (*n - h + 1); j++) {
  s1 += 1;
  for(k = 1; k < j; k++) {
    s1 += 2.0 * jw[(((j - 2) * (j - 1) / 2) + k - 1) *
      ((*n * (*n - 1) / 2) + 1)] *
      jw[(((j + h - 2) * (j + h - 1) / 2) + k + h - 1) *
      ((*n * (*n - 1) / 2) + 1)] *
      (jw[(((j + h - 2) * (j + h - 1) / 2) + k + h - 1) *
      (*n * (*n - 1) / 2)) +
      ((j - 2) * (j - 1) / 2) + k - 1]) *
      (jw[(((j + h - 2) * (j + h - 1) / 2) + k + h - 1) *
      (*n * (*n - 1) / 2)) +
      ((j - 2) * (j - 1) / 2) + k - 1]);
  }
}
s1 = s1 / ((*n - h) * (*n - h));

//////////////////////////////////////
//výpočet třetího sčítance z výpočetního tvaru integral |D*|^2
for (j = 2; j < (*n - h + 1); j++) {
  for (k = 1; k < j; k++) {
    s2_jk += jw[(((j - 2) * (j - 1) / 2) + k - 1) *
      ((*n * (*n - 1) / 2) + 1)];
  }
}

for (l = 2; l < (*n - h + 1); l++) {
  for (m = 1; m < l; m++) {
    s2_lm += jw[(((l + h - 2) * (l + h - 1) / 2) + m + h - 1) *
      ((*n * (*n - 1) / 2) + 1)];
  }
}

for (j = 2; j < (*n - h + 1); j++) {
  for (k = 1; k < j; k++) {
    for (l = 2; l < (*n - h + 1); l++) {
      for (m = 1; m < l; m++) {
        s2_jklm += jw[(((j - 2) * (j - 1) / 2) + k - 1) *
          ((*n * (*n - 1) / 2) + 1)] *
          jw[(((l + h - 2) * (l + h - 1) / 2) + m + h - 1) *
          ((*n * (*n - 1) / 2) + 1)] *
          (((jw[(((l + h - 2) * (l + h - 1) / 2) +
            m + h - 1) *
            (*n * (*n - 1) / 2)) +
            ((j - 2) * (j - 1) / 2) + k - 1]) *
            (jw[(((l + h - 2) * (l + h - 1) / 2) +
            m + h - 1) *
            (*n * (*n - 1) / 2)) +
            ((j - 2) * (j - 1) / 2) + k - 1]))+

```

```

(1 / ((jw[(((1 + h - 2) * (1 + h - 1) / 2) +
          m + h - 1) *
          (*n * (*n - 1) / 2)) +
        ((j - 2) * (j - 1) / 2) + k - 1]) *
      (jw[(((1 + h - 2) * (1 + h - 1) / 2) +
          m + h - 1) *
          (*n * (*n - 1) / 2)) +
        ((j - 2) * (j - 1) / 2) + k - 1]]));
    }
  }
}

s2 = 1.0 / ((*n - h) * (*n - h)) +
      (2.0 / ((*n - h) * (*n - h) * (*n - h))) * s2_jk +
      (2.0 / ((*n - h) * (*n - h) * (*n - h))) * s2_lm +
      (2.0 / ((*n - h) * (*n - h) * (*n - h) * (*n - h))) * s2_jklm;

//////////////////////////////////////
//výpočet druhého sčítance z výpočetního tvaru integral |D^*|^2
for (j = 2; j < (*n - h + 1); j++) {
  for (k = 1; k < j; k++) {
    s3_jk += 2 * jw[(((j - 2) * (j - 1) / 2) + k - 1) *
                    ((*n * (*n - 1) / 2) + 1)];
  }
}

for (j = 2; j < (*n - h + 1); j++) {
  for (l = 1; l < j; l++) {
    s3_jl += 2 * jw[(((j + h - 2) * (j + h - 1) / 2) + l + h - 1) *
                    ((*n * (*n - 1) / 2) + 1)];
  }
}

for (j = 2; j < (*n - h + 1); j++) {
  for (k = 1; k < j; k++) {
    for (l = 1; l < j; l++) {
      s3_jkl += jw[(((j - 2) * (j - 1) / 2) + k - 1) *
                  ((*n * (*n - 1) / 2) + 1)] *
                jw[(((j + h - 2) * (j + h - 1) / 2) + l + h - 1) *
                  ((*n * (*n - 1) / 2) + 1)] *
                ((jw[(((j + h - 2) * (j + h - 1) / 2) +
                    l + h - 1) *
                    (*n * (*n - 1) / 2)) +
                  ((j - 2) * (j - 1) / 2) + k - 1]) *
                (jw[(((j + h - 2) * (j + h - 1) / 2) +
                    l + h - 1) *
                    (*n * (*n - 1) / 2)) +
                  ((j - 2) * (j - 1) / 2) + k - 1]]);
    }
  }
}

```

```

}
for (j = 2; j < (*n - h); j++) {
  for (k = 1; k < j; k++) {
    for (l = (j + 1); l < (*n - h + 1); l++) {
      s3_jkl += jw[(((j - 2) * (j - 1) / 2) + k - 1) *
        ((*n * (*n - 1) / 2) + 1)] *
        jw[(((l + h - 2) * (l + h - 1) / 2) + j + h - 1) *
        ((*n * (*n - 1) / 2) + 1)] *
        (1 / ((jw[(((l + h - 2) * (l + h - 1) / 2) +
          j + h - 1) * (*n * (*n - 1) / 2)) +
          ((j - 2) * (j - 1) / 2) + k - 1]) *
          (jw[(((l + h - 2) * (l + h - 1) / 2) +
            j + h - 1) *
            (*n * (*n - 1) / 2)) +
            ((j - 2) * (j - 1) / 2) + k - 1]]));
    }
  }
}
for (j = 2; j < (*n - h); j++) {
  for (k = (j+1); k < (*n - h + 1); k++) {
    for (l = 1; l < j; l++) {
      s3_jkl += jw[(((k - 2) * (k - 1) / 2) + j - 1) *
        ((*n * (*n - 1) / 2) + 1)] *
        jw[(((j + h - 2) * (j + h - 1) / 2) +
          l + h - 1) *
          ((*n * (*n - 1) / 2) + 1)] *
        (1 / ((jw[(((j + h - 2) * (j + h - 1) / 2) +
          l + h - 1) *
          (*n * (*n - 1) / 2)) +
          ((k - 2) * (k - 1) / 2) + j - 1]) *
          (jw[(((j + h - 2) * (j + h - 1) / 2) +
            l + h - 1) *
            (*n * (*n - 1) / 2)) +
            ((k - 2) * (k - 1) / 2) + j - 1]]));
    }
  }
}
for (j = 1; j < (*n - h); j++) {
  for (k = (j+1); k < (*n - h + 1); k++) {
    for (l = (j+1); l < (*n - h + 1); l++) {
      s3_jkl += jw[(((k - 2) * (k - 1) / 2) + j - 1) *
        ((*n * (*n - 1) / 2) + 1)] *
        jw[(((l + h - 2) * (l + h - 1) / 2) +
          j + h - 1) *
          ((*n * (*n - 1) / 2) + 1)] *
        ((jw[(((l + h - 2) * (l + h - 1) / 2) +
          j + h - 1) * (*n * (*n - 1) / 2)) +
          ((k - 2) * (k - 1) / 2) + j - 1]) *
          (jw[(((l + h - 2) * (l + h - 1) / 2) +
            j + h - 1) * (*n * (*n - 1) / 2)) +
            ((k - 2) * (k - 1) / 2) + j - 1]));
    }
  }
}

```

```
    }  
  }  
}  
  
s3 = (-2.0 / ((*n - h) * (*n - h) * (*n - h))) *  
      ((*n - h) + s3_jk + s3_jl + s3_jkl);  
  
////////////////////////////////////  
////////////////////////////////////  
  
delta = s1 + s2 + s3;  
*stat += (*n - h) * delta;  
}  
}
```