BRNO UNIVERSITY OF TECHNOLOGY
VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

FACULTY OF MECHANICAL ENGINEERING
FAKULTA STROJNÍHO INŽENÝRSTVÍ

INSTITUTE OF MATHEMATICS
ÚSTAV MATEMATIKY

SELECTED RANDOM VARIABLES TRANSFORMATIONS USED IN CLASSICAL LINEAR REGRESSION
VYBRANÉ TRANSFORMACE NÁHODNÝCH VELIČIN UŽÍVANÉ V KLASICKÉ LINEÁRNÍ REGRESI

MASTER'S THESIS
DIPLOMOVÁ PRÁCE

AUTHOR
AUTOR PRÁCE
Bc. Martin Tejkal

SUPERVISOR
VEDOUCÍ PRÁCE
doc. Mgr. Zuzana Hübnerová, Ph.D.

BRNO 2017

Specification Master's Thesis

Department:	Institute of Mathematics
Student:	**Bc. Martin Tejkal**
Study programme:	Applied Sciences in Engineering
Study field:	Mathematical Engineering
Leader:	**doc. Mgr. Zuzana Hübnerová, Ph.D.**
Academic year:	2016/17

Pursuant to Act no. 111/1998 concerning universities and the BUT study and examination rules, you have been assigned the following topic by the institute director Master's Thesis:

Selected random variables transformations used in classical linear regression

Concise characteristic of the task:

Classical linear regression model is based on an assumption of normally distributed response variables and on an assumption of variance equality. In case, the normality assumption is not fulfilled a transformation of the response variables is often applied. In literature, there are several transformations being suggested for a given distribution of the response variables. The aim of the transformations is either to standardize the variance or to change other characteristics of the distribution. Frequently, the logarithm transformations considered. The problem of calculating logarithm of non–negative values is usually solved by adding a constant 1. The thesis will aim on properties of the parameter estimates of the distribution of a random variable Y depending on the value c when the transformation log(Y+c) is used. Special attention will be paid to random variables with Poisson or negative binomial distribution.

Goals Master's Thesis:

Introduction of the necessary terms regarding transformation of random variables and properties of parameter estimates.

Simulation or theoretical study of the properties of parameter estimates of random variable Y with Poisson or negative binomial distribution in dependence on constant c when the transformation log(y+c) is applied.

Recommendation for the optimal choice of constant c when the transformation log(y+c) is applied.

Comparison of the properties of the logarithm transformation log(y+c) with other available transformation which can be done either by simulation or theoretical approach.

Recommended bibliography:

ANSCOMBE, Francis John. The transformation of Poisson, binomial and negative-binomial data. Biometrika. 1948, 35 (3–4), 246–254.

ANDĚL, Jiří. Základy matematické statistiky. Praha: Matfyzpress, 2011. ISBN 978-80-7378-1620.

DRAPER, Norman Richard and SMITH, Harry. Applied regression analysis. Third edition. Hoboken, New Jersey: John Wiley & Sons, Inc., 1998. Wiley series in probability and statistics. ISBN 978-1-1-8-62562-0.

Deadline for submission Master's Thesis is given by the Schedule of the Academic year 2016/17

In Brno,

L. S.

prof. RNDr. Josef Šlapal, CSc. Director of the Institute	doc. Ing. Jaroslav Katolický, Ph.D. FME dean

Abstract
Classical linear regression model and the respective tests are based on an assumption of normally distributed response variables and on an assumption of variance equality. If the normality assumption is not fulfilled, then the response variables are usually transformed. In the first part of this work variance stabilising transformations are discussed. Great deal of attention is given to random variables of Poisson and negative binomial distribution, for which generalised variance stabilising transformations with addition constants in their arguments are studied. Optimal values of the constants for the generalised transformations are determined. The second part aims to provide a comparison of the transformations introduced in the first part and some other commonly used transformations. The comparison is done within the ANOVA framework by testing the hypothesis of equality of expectations among p random samples via F test. The properties of the distribution of the F test under the assumptions of equal and unequal variances are studied. Finally a comparison of the power functions of the F test applied to p random samples from Poisson distribution transformed via square root, logarithmic and Yeo-Johnson transformation, and to p random sample of negative binomial distribution transformed via argument of hyperbolic sine, logarithmic and the Yeo-Johnson transformation is carried out theoretically and by simulations.

Abstrakt
Klasická lineární regrese a z ní odvozené testy hypotéz jsou založeny na předpokladu normálního rozdělení a shodnosti rozptylu závislých proměnných. V případě že jsou předpoklady normality porušeny, obvykle se užívá transformací závisle proměnných. První část této práce se zabývá transformacemi stabilizujícími rozptyl. Značná pozornost je udělena náhodným veličinám s Poissonovým a negativně binomickým rozdělením, pro které jsou studovány zobecněné transformace stabilizující rozptyl obsahující parametry v argumentu navíc. Pro tyto parametry jsou stanoveny jejich optimální hodnoty. Cílem druhé části práce je provést srovnání transformací uvedených v první části a dalších často užívaných transformací. Srovnání je provedeno v rámci analýzy rozptylu testováním hypotézy shodnosti středních hodnot p nezávislých náhodných výběrů s pomocí F testu. V této části jsou nejprve studovány vlastnosti F testu za předpokladu shodných a neshodných rozptylů napříč výběry. Následně je provedeno srovnání silofunkcí F testu aplikovaného pro p výběrů z Poissonova rozdělení transformovanými odmocninovou, logaritmickou a Yeo Johnsnovou transformací a z negativně binomického rozdělení transformovaného argumentem hyperbolického sinu, logaritmickou a Yeo-Johnsnovou transformací.

Keywords
Poisson distribution, negative binomial distribution, variance stabilising transformation, logarithmic transformation, square root transformation, argument of hyperbolic sine transformation, Yeo-Johnson transformation, Linear regression, ANOVA, F-test, power function

Klíčová slova
Poissonovo rozdělení, negativně binomické rozdělení, transformace stabilizující rozptyl, odmocninová transformace, transformace argument hyperbolického sinu, logaritmická transformace, Yeo-Johnsnova transformace, Lineární regrese, ANOVA, F-test, silofunkce

TEJKAL, M. *Selected random variables transformations used in classical linear regression.* Brno: Vysoké učení technické v Brně, Fakulta strojního inženýrství, 2017. 119 s. Diploma thesis supervisor Doc. Mgr. Zuzana Hübnerová, Ph.D.

I hereby declare that I wrote the master thesis Selected Random Variables Transformations Used in Classical Linear Regression by myself under the supervision of doc. Mgr. Zuzana Hübnerová, Ph.D. with use of the materials listed in Bibliography

Martin Tejkal

I would like to thank to everybody who helped me and supported me during the time of my work on this thesis. I am especially thankful to doc. Mgr. Zuzana Hübnerová, Ph.D. for exemplary supervision of the thesis and many helpful suggestions and comments.

Martin Tejkal

Contents

Preface **13**

1 Theoretical Basis **15**
1.1 Basic Concepts 15
1.1.1 Special Functions 15
1.1.2 Touchard polynomials 15
1.2 Results of Matrix Theory 16
1.2.1 Symmetric Matrices, Positive Semidefinite, and Definite Matrices . 16
1.2.2 Eigenvalues 18
1.2.3 Idempotent matrices 19
1.2.4 Pseudoinverse Matrices 20
1.3 Properties of Selected Probability Distributions 20
1.3.1 Poisson Probability Distribution 20
1.3.2 Negative Binomial Probability Distribution 21
1.3.3 Pearson Chi Square Distribution 22
1.4 Estimation theory 24
1.4.1 Statistics and Unbiased Estimators 24

2 Transformations of Random Variables **25**
2.1 Variance Stabilising Transformation 25
2.2 Box - Cox Transformation 26
2.3 Yeo-Johnson Transformation 28

3 A Study of Selected Transformations **29**
3.1 Variance Stabilising Transformation for Random Variable with Poisson Probability Distribution 29
3.2 Study of Chosen Transformations for Random Variable with Negative Binomial Probability Distribution 35

4 Theoretical Background of Performance Comparison of Selected Transformations Within ANOVA Framework **48**
4.1 Theoretical Results for Power of F-Test 49
4.1.1 The Distribution of Denominator 50
4.1.2 Distribution of the Numerator Under Alternative 52
4.1.3 Power of F-Test 54
4.2 Theoretical Results for Power of F-Test with Unequal Variances 55
4.2.1 Study of Denominator 57
4.2.2 Study of Numerator 60
4.2.3 Approximation of F Statistic 63

5 Application and Results of Comparison of Selected Transformations within ANOVA Framework **67**
5.1 Transformation ln(X + 1) Applied on Sample from Poisson Distribution . . 67
5.2 Transformation ln(X + 1) Applied on Sample from Negative Binomial Distribution 71
5.3 Simulation Study of Used Approximations, Effect of Transformations on Parameter Estimates of Poisson or Negative Binomial Distribution 73
5.3.1 Data Input and Results for Poisson Case 75
5.3.2 Data Input and Results for Negative Binomial Case 79
5.4 Comparison of Power Functions by Simulation 87
5.4.1 Computation of Theoretical Power Function under Asumption of Equal Variances 87
5.4.2 Computation of Theoretical Power Function Under Assumption of Unequal Variances 88
5.4.3 Power Functions for Case of Poisson Distribution 91
5.4.4 Power Functions for Case of Negative Binomial Distribution 97

A Computation of Numerical Characteristics of Selected Distributions **108**
A.0.5 Poisson Probability Distrbution 108
A.0.6 Negative Binomial Probability Disribution 109

Conclusion **112**

Bibliography **116**

Notation Index **118**

Electronic Appendix Index **119**

Preface

Classical linear regression model and the respective tests are based on an assumption of normally distributed response variables and on an assumption of variance equality. If the normality assumption is not fulfilled, then the response variables are usually transformed. In literature, there are several transformations suggested for the frequently occurring distributions of the response variables. Often, the logarithmic transformation is applied. The problem of calculating a logarithm of nonnegative values is usually solved by adding a constant 1. This work aims to study the logarithmic transformation $\ln(X + 1)$ and alternative transformations that can be used instead of $\ln(X+1)$ and provide comparisons of the studied transformations. Great deal of attention is paid to random variables X of Poisson and Negative binomial distribution.

In the first chapter the necessary theoretical background concerning matrix theory and properties of selected probability distributions is introduced. In the last section of the first chapter some basic results of estimation theory are summarised.

In the second chapter the important results concerning transformations of random variables are introduced. Namely the variance stabilising transformation and a possible way of determining it given a random variable of arbitrary probability distribution. In the second and the third section of the second chapter the commonly used variance stabilising transformations used for normality approximation are discussed.

In the third chapter selected generalised variance stabilising transformations with additional general constants added for random variables with Poisson and Negative binomial probability distributions are studied. Great deal of attention is given to finding approximations of numerical characteristics of the transformed variables when the studied transformations are applied, namely the variance. Using the approximations of the numerical characteristic optimal values for the general constants are found.

The fourth and fifth chapter both tackle with the problematic of comparison of the transformations introduced in Chapter 3 of the work. The comparison itself is done within the One-Way Analysis of Variance Framework by testing the hypothesis of equality of expectations of random samples originating from Poisson or Negative binomial distribution, that were transformed either via transformation $\ln(X + 1)$ or via the variance stabilising transformations introduced in Chapter 3, evaluating the power functions of the tests and comparing them.

The fourth chapter provides a theoretical background for the framework of the transformations comparison. It is assumed that some of the transformations might not have the variance stabilising effect, for such cases an approximation of the distribution of the test statistic in case of violated assumption of equality of variances is derived.

The first two sections of the fifth chapter provide some additional information about the transformation $\ln(X + 1)$ applied to either Poisson distributed random variable or Negative binomially distributed random variable. Namely, for both cases the approximations of the moments of $Y = \ln(X + 1)$ are derived. The rest of the chapter describes the used methods and provides the computational and graphical results of the numerical analysis. First the goodness of all the derived numerical characteristics approximations and other approximations is checked. Finally the power functions of the executed tests are computed by two different approaches, one theoretical and one based on simulations and compared.

Chapter 1

Theoretical Basis

Before we start with the topic of interest of this work, we first define some basic tools that will be used in our study. The whole chapter is based on [2], [3], [10] and [13].

1.1 Basic Concepts

In this section, various theoretical results concerning special functions and Touchard polynomials are collected. The content of this section are based on [2] and [3].

1.1.1 Special Functions

Definition 1.1. (**Gamma Function**) Let $a > 0$. We define gamma function $\Gamma(a)$ as

$$\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx \tag{1.1}$$

Definition 1.2. Let $\Gamma(t)$ be a Gamma function given by Definition 1.1, and assume that p is a nonnegative integer. We define a polygamma function of order p by

$$\psi^p(t) = \left(\frac{d}{dt}\right)^{p+1} \ln \Gamma(t). \tag{1.2}$$

Namely for $p = 0$ we obtain

$$\psi(t) = \left(\frac{d}{dt}\right) \ln \Gamma(t) = \frac{\Gamma'(t)}{\Gamma(t)}. \tag{1.3}$$

Function $\psi(t)$ will be called digamma function.

Definition 1.3. Let $a > 0$, and $b > 0$. We define beta function $B(a, b)$ as

$$B(a, b) = \int_0^1 x^{a-1}(1-x)^{b-1} dx \tag{1.4}$$

1.1.2 Touchard polynomials

In this subsection we provide some results for Touchard polynomials which will be used in Section 3.1. This subsection is based on [3].

Definition 1.4. (Touchard polynomials) The collection $\{T_n : n \geq 0\}$ of the Touchard (also called exponential) polynomials in one dimension is defined as $T_0 = 1$ and

$$T_n(x) = \sum_{k=1}^{n} S(n,k)x^k, \quad n \geq 1 \tag{1.5}$$

Where $S(n,k)$ are Stirling numbers of the second kind, as defined in [3].

1.2 Results of Matrix Theory

In this section we will provide some useful properties of particular types of matrices that appear in Classical Linear Regression and Analysis of Variance. This chapter is mainly based on [2], some important results are taken from [13]. Where there can not be any misunderstanding, we will by abuse of notation drop the indexes denoting the type of a matrix.

Proposition 1.5. *Let $\boldsymbol{A}_{n\times n}, \boldsymbol{B}_{n\times n}$ be matrices, then if $\boldsymbol{B}$ is nonsignular, $rank(\boldsymbol{AB}) = rank(\boldsymbol{BA}) = rank(\boldsymbol{A})$*

Proof. See [10]. □

Proposition 1.6. *Let $\boldsymbol{A}_{m\times n}$ be a $m \times n$ matrix, such that $rank(\boldsymbol{A}_{m\times n}) = r \geq 1$, then there exist matrices $\boldsymbol{B}_{m\times r}$ and $\boldsymbol{C}_{r\times n}$, such that $\boldsymbol{A}_{m\times n} = \boldsymbol{B}_{m\times r}\boldsymbol{C}_{r\times n}$, and $rank(\boldsymbol{B}_{m\times r}) = rank(\boldsymbol{C}_{r\times n}) = r$.*

Proof. As columns of $\mathbf{B}$ we take those r linearly independent columns of matrix $\mathbf{A}$, whose existence follows from the assumption $rank(\mathbf{A}) = r$. Then the j-th column of the matrix $\mathbf{A}$ is a linear combination of the columns of matrix $\mathbf{B}$ with some coefficients $c_{1j}, ..., c_{rj}$. Assume that a vector of these coefficients is the j-th column of matrix $\mathbf{C}$.

Since the rank of product of two matrices is at most equal to the rank of any of the matrices entering the product we have that $rank(\mathbf{B}) \geq r$, and $rank(\mathbf{C} \geq r$. But since $\mathbf{B}$ has r columns, and $\mathbf{C}$ r rows, we also get $rank(\mathbf{B}) \leq r$, and $rank(\mathbf{C}) \leq r$ which concludes the proof. □

1.2.1 Symmetric Matrices, Positive Semidefinite, and Definite Matrices

Definition 1.7. A square matrix $\mathbf{A}_{m\times m}$ is called positively semidefinite, and we denote $\mathbf{A}_{m\times m} \geq 0$, if it is symmetric and for every m-dimensional vector $\mathbf{x}_m \neq \mathbf{0}$ we have

$$\mathbf{x}_m^T \mathbf{A}_{m\times m} \mathbf{x}_m \geq 0. \tag{1.6}$$

Definition 1.8. A square matrix $\mathbf{A}_{m\times m}$ is called positively definite, and we denote $\mathbf{A}_{m\times m} > 0$, if it is symmetric, and for every nonzero m-dimensional vector $\mathbf{x}_m$ we have

$$\mathbf{x}_m^T \mathbf{A}_{m\times m} \mathbf{x}_m > 0. \tag{1.7}$$

Definition 1.9. Let $\mathcal{H}$ be a Hilbert space, a subset $\mathcal{M}$ of $\mathcal{H}$ that is closed under the addition of vectors, and scalar multiplication is called a linear manifold.

Proposition 1.10. *Let $\boldsymbol{A}_{p\times p}$ be a symmetric real matrix, then the eigenvectors $\boldsymbol{p}_i$, $\boldsymbol{p}_j$ corresponding to eigenvalues λ_i, λ_j, where $\lambda_i \neq \lambda_j$ are orthogonal.*

Proof. From the definition we have

$$\mathbf{A}\mathbf{p}_i = \lambda_i \mathbf{p}_i, \quad \mathbf{A}\mathbf{p}_j = \lambda_j \mathbf{p}_j. \tag{1.8}$$

Multiplication of the first by $\mathbf{p}_j^T$ and the second by $\mathbf{p}_i^T$ and a subtraction gives $(\lambda_i - \lambda_j)\mathbf{p}_i^T\mathbf{p}_j = 0$, and since the eigenvalues are distinct we have that $\mathbf{p}_i$, and $\mathbf{p}_j$ are orthogonal. □

Proposition 1.11. *Let $\boldsymbol{A}_{p\times p}$ be a real matrix. If $\boldsymbol{x}$ is an arbitrary non-null vector, there exists an eigenvector $\boldsymbol{y}$ belonging to the linear manifold $\mathcal{M}(\boldsymbol{x}, \boldsymbol{A}\boldsymbol{x}, \boldsymbol{A}^2\boldsymbol{x}, ...)$.*

Proof. The vectors $\mathbf{x}, \mathbf{A}\mathbf{x}, ...$ can not all be independent. Let k be the smallest value, such that

$$\mathbf{A}^k\mathbf{x} + b_{k-1}\mathbf{A}^{k-1}\mathbf{x} + ... + b_0\mathbf{x} = \mathbf{0}. \tag{1.9}$$

Factorising (1.9), we see that

$$(\mathbf{A} - \mu_1\mathbf{I}) \cdot ... \cdot (\mathbf{A} - \mu_k\mathbf{I})\mathbf{x} = \mathbf{0}, \tag{1.10}$$

and therefore

$$(\mathbf{A} - \mu_1\mathbf{I})\mathbf{y} = \mathbf{0}, \tag{1.11}$$

where

$$\mathbf{y} = (\mathbf{A} - \mu_2\mathbf{I}) \cdot ... \cdot (\mathbf{A} - \mu_k\mathbf{I}) \neq \mathbf{0}. \tag{1.12}$$

Furthermore, $(\mathbf{A} - \mu_1\mathbf{I})\mathbf{y} = \mathbf{0}$, i. e. $\mathbf{y}$ is the eigenvector associated with the eigenvalue μ_1. Since $\mathbf{A}$ is real, μ_1 is real (see [13], or [2]). Similarly each μ_i is real, and the equation (1.12) shows that $\mathbf{y} \in \mathcal{M}(\mathbf{x}, \mathbf{A}\mathbf{x}, \mathbf{A}^2\mathbf{x}, ...)$. □

Following important result will be provided with complete proof which can also be found in [13]. The proof is of importance, as it is constructive, and some steps of it will come in handy again in Chapter 4 Section 4.2.

Proposition 1.12. *Let $\boldsymbol{A}_{m\times m}$ be a real symmetric matrix. Let us denote $\lambda_1 \geq ... \geq \lambda_m$ the eigenvalues of $\boldsymbol{A}$ including the multiplicities. Put $\boldsymbol{\Lambda}_{m\times m} = diag(\lambda_1, ..., \lambda_m)$. Then there exists an orthogonal matrix $\boldsymbol{P}_{m\times m}$ such that each column vector $\boldsymbol{p}_i$ of $\boldsymbol{P}$ is an eigenvector corresponding to λ_i, and*

$$\boldsymbol{A}_{m\times m} = \boldsymbol{P}_{m\times m}\boldsymbol{\Lambda}_{m\times m}\boldsymbol{P}^T_{m\times m}, \ \ \textit{and} \ \ \boldsymbol{I}_{m\times m} = \boldsymbol{P}_{m\times m}\boldsymbol{P}^T_{m\times m}. \tag{1.13}$$

Proof. Suppose there exist s orthonormal vectors $\mathbf{p}_1, ..., \mathbf{p}_s$ such that

$$\mathbf{A}\mathbf{p}_i = \lambda_i\mathbf{p}_i, \quad i = 1, ..., s. \tag{1.14}$$

The result (1.14) implies, that $\mathbf{A}^2\mathbf{p}_i = \lambda_i\mathbf{A}\mathbf{p}_i = \lambda_i^2\mathbf{p}_i, ..., \mathbf{A}^r\mathbf{p}_i = \lambda^r\mathbf{p}_i, ...$ Choose a vector $\mathbf{x}$ orthogonal to $\mathcal{M}(\mathbf{p}_1, ..., \mathbf{p}_s)$, then

$$\mathbf{x}^T\mathbf{A}^r\mathbf{p}_i = \mathbf{x}^T\lambda_i^r\mathbf{p}_i = 0 \tag{1.15}$$

for all r and $i = 1, ..., s$. Hence, due to symmetry of $\mathbf{A}$ we have $\mathcal{M}(\mathbf{x}, \mathbf{A}\mathbf{x}, \mathbf{A}^2\mathbf{x}, ...)$ is orthogonal to $\mathcal{M}(\mathbf{p}_1, ..., \mathbf{p}_s)$. From Proposition 1.11 we know, that there exists an eigenvector $\mathbf{p}_{s+1} \in \mathcal{M}(\mathbf{x}, \mathbf{A}\mathbf{x}, \mathbf{A}^2\mathbf{x}, ...)$, which in view of (1.15) is orthogonal to $\mathbf{p}_1, ..., \mathbf{p}_s$.

Since $\mathbf{p}_1$ can be chosen corresponding to any latent vector to start with, we have established the existence of m mutually orthogonal latent vectors $\mathbf{p}_1, ..., \mathbf{p}_m$ such that

$$\mathbf{A}\mathbf{p}_i = \lambda_i\mathbf{p}_i, \quad i = 1, ..., m, \tag{1.16}$$

which may be written

$$\mathbf{AP} = \mathbf{P\Lambda}, \quad \mathbf{PP}^T = \mathbf{I}, \tag{1.17}$$

where $\mathbf{P}$ is the orthogonal matrix with $\mathbf{P}_i$ as its columns and $\mathbf{\Lambda}$ is the diagonal matrix with λ_i as its i-th diagonal element.

$\square$

Due to the nonnegativity of all eigenvalues of a positive semidefinite matrix we can define following useful term.

Definition 1.13. Let $\mathbf{A}_{m\times m}$ be a positive semidefinite matrix, let us denote $\mathbf{\Lambda}^{\frac{1}{2}}_{m\times m} = diag(\sqrt{\lambda_1}, ..., \sqrt{\lambda_m})$ an eigenvalue square root matrix.

Remark 1.14. *It is easy to see that* $\mathbf{\Lambda}^{\frac{1}{2}}_{m\times m}\mathbf{\Lambda}^{\frac{1}{2}}_{m\times m} = \mathbf{\Lambda}_{m\times m}$.

Proposition 1.15. *Let* $\boldsymbol{A}_{m\times m}$ *be a positive semidefinite matrix of* $rank(\boldsymbol{A}_{m\times m}) = r \geq 1$. *Then there exists a matrix* $\boldsymbol{B}_{m\times r}$ *such that* $rank(\boldsymbol{B}_{m\times r}) = r$ *and we have*

$$\boldsymbol{A}_{m\times m} = \boldsymbol{B}_{m\times r}\boldsymbol{B}^T_{m\times r}. \tag{1.18}$$

Proof. By Proposition 1.12 we have that $\mathbf{A}_{m\times m} = \mathbf{U}_{m\times m}\mathbf{\Lambda}_{m\times m}\mathbf{U}^T_{m\times m}$. From the assumption $rank(\mathbf{A}_{m\times m}) = r \geq 1$ follows that $rank(\mathbf{\Lambda}_{m\times m}) = r$. Hence, $\mathbf{\Lambda}_{m\times m}$ has to have form $\mathbf{\Lambda}_{m\times m} = diag(\lambda_1, ..., \lambda_r, 0, ..., 0)$. When using the block notation, we can write $\mathbf{\Lambda}^{\frac{1}{2}}_{m\times m} = (\mathbf{L}, \mathbf{0})_{m\times(m-r)}$, where

$$\mathbf{L}_{m\times r} = \begin{pmatrix} \sqrt{\lambda_1} & 0 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & \sqrt{\lambda_r} \\ 0 & 0 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & 0 \end{pmatrix}$$

. We have that

$$\mathbf{A} = \mathbf{U\Lambda}^{\frac{1}{2}}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{U}^T = \mathbf{U}(\mathbf{L}, \mathbf{0})\begin{pmatrix} \mathbf{L}^T \\ \mathbf{0} \end{pmatrix}\mathbf{U}_T = (\mathbf{UL}, \mathbf{0})\begin{pmatrix} \mathbf{L}^T\mathbf{U}^T \\ \mathbf{0} \end{pmatrix} = \mathbf{ULL}^T\mathbf{U}^T = \mathbf{BB}^T,$$

where $\mathbf{B} = \mathbf{UL}$. Since $\mathbf{U}$ is regular, we have that $rank(\mathbf{B}) = r$. $\square$

1.2.2 Eigenvalues

Proposition 1.16. *Let* $\boldsymbol{A}_{n\times n}$ *be a real matrix with eigenvalues* $\lambda_1, ..., \lambda_n$. *Then*

$$Tr(\boldsymbol{A}) = \sum_{i=1}^{n} \lambda_i. \tag{1.19}$$

Proof. The proof is based on the theory of Jordan Canonical Forms that is not developed in this work, for details see [10]. $\square$

Proposition 1.17. *Let $\boldsymbol{A}_{n\times n}$ be a real matrix with eigenvalues λ_i, $i = 1, ...n$, and let $\boldsymbol{B}_{m\times m}$ be a real matrix with eigenvalues μ_i, $i = 1, ...m$. Then the eigenvalues of $\boldsymbol{A} \otimes \boldsymbol{B}$ are*

$$\lambda_1\mu_1, ..., \lambda_1\mu_m, \lambda_2\mu_1, ..., \lambda_2\mu_m, ..., \lambda_n\mu_m. \tag{1.20}$$

Moreover, if $\boldsymbol{x}_1, ..., \boldsymbol{x}_p$ are linearly independent right eigenvectors of $\boldsymbol{A}$ corresponding to $\lambda_1, ..., \lambda_p$ ($p \le n$), and $\boldsymbol{z}_1, ..., \boldsymbol{z}_q$ are linearly independent right eigenvectors of $\boldsymbol{B}$ corresponding to $\mu_1, ..., \mu_q$ ($q \le m$), then $\boldsymbol{x}_i \otimes \boldsymbol{z}_j \in \mathbb{R}^{nm}$ are linearly independent right eigenvectors of $\boldsymbol{A} \otimes \boldsymbol{B}$ corresponding to $\lambda_i\mu_j$, where $i = 1, ..., p$, and $j = 1, ..., q$.

Proof. See [10]. □

1.2.3 Idempotent matrices

Definition 1.18. We say that an m-dimensional matrix is idempotent if $\mathbf{A}^2_{m\times m} = \mathbf{A}_{m\times m}$.

Proposition 1.19. *Eigenvalues of idempotent matrix are only zeroes and ones.*

Proof. We know, that λ is an eigenvalue and $\mathbf{x}_m \neq \mathbf{0}$ a corresponding eigenvector of a matrix $\mathbf{A}_{m\times m}$, if $\mathbf{A}_{m\times m}\mathbf{x}_m = \lambda\mathbf{x}_m$. If we multiply this equation from the left by the matrix $\mathbf{A}_{m\times m}$ we obtain $\mathbf{A}^2_{m\times m}\mathbf{x}_m = \lambda\mathbf{A}_{m\times m}\mathbf{x}_m$. On the left hand side of the equality we have $\mathbf{A}^2_{m\times m}\mathbf{x}_m = \mathbf{A}_{m\times m}\mathbf{x}_m = \lambda\mathbf{x}_m$, and on the right hand side we have $\lambda\mathbf{A}_{m\times m}\mathbf{x}_m = \lambda^2\mathbf{x}_m$. Hence, we get the equality $\lambda\mathbf{x}_m = \lambda^2\mathbf{x}_m$. Since $\mathbf{x}_m \neq \mathbf{0}$ by assumption, obviously $\lambda(1-\lambda) = 0$ must be satisfied, hence, λ has to be either zero, or one. □

Proposition 1.20. *Let $\boldsymbol{A}_{m\times m}$ be an idempotent matrix. The $rank(\boldsymbol{A}_{m\times m})$ of $\boldsymbol{A}_{m\times m}$ is equal to its trace.*

Proof. Let $rank(\mathbf{A}_{m\times m}) = r$. If $r = 0$, then the statement is obviously satisfied, therefore assume that $r \geq 1$. By Proposition 1.6 we have that $\mathbf{A}_{n\times n} = \mathbf{B}_{n\times r}\mathbf{C}_{r\times n}$, where $rank(\mathbf{B}_{n\times r}) = rank(\mathbf{C}_{r\times n}) = r$. Let us denote $\mathbf{L}_{r\times n}$ the left inverse matrix (see [10]) with respect to matrix $\mathbf{B}_{n\times r}$, and $\mathbf{P}_{n\times r}$ the right inverse matrix (see [10]) with respect to $\mathbf{C}_{r\times n}$. The assumption $\mathbf{A}_{n\times n} = \mathbf{A}^2_{n\times n}$ can be written in the form $\mathbf{B}_{n\times r}\mathbf{C}_{r\times n}\mathbf{B}_{n\times r}\mathbf{C}_{r\times n} = \mathbf{B}_{n\times r}\mathbf{C}_{r\times n}$. Now by multiplying by $\mathbf{L}_{r\times n}$ from the left and $\mathbf{P}_{n\times r}$ from the right, we obtain

$$\begin{aligned}\mathbf{L}(\mathbf{BCBC})\mathbf{P} &= \mathbf{L}(\mathbf{BC})\mathbf{P}\\ &= (\mathbf{LB})(\mathbf{CP})\\ &= \mathbf{I}\end{aligned} \tag{1.21}$$

Since also $(\mathbf{LB})\mathbf{CB}(\mathbf{CP}) = \mathbf{ICBI} = \mathbf{CB}$, we have $\mathbf{CB} = \mathbf{I}$. From here follows that

$$Tr(\mathbf{A}) = Tr(\mathbf{BC}) = Tr(\mathbf{CB}) = Tr(\mathbf{I}) = r = rank(\mathbf{A}) \tag{1.22}$$

□

Proposition 1.21. *Symmetric idempotent matrix is positively semidefinite.*

Proof. Let $\mathbf{A}_{n\times n}$ be a symmetric idempotent matrix. Let $\mathbf{x}_n$ be an n-dimensional vector. Since $\mathbf{A} = \mathbf{A}^2$, and $\mathbf{A}^T = \mathbf{A}$ we have

$$\mathbf{x}^T\mathbf{A}\mathbf{x} = \mathbf{x}^T\mathbf{A}^2\mathbf{x} = \mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x} = (\mathbf{A}\mathbf{x})^T(\mathbf{A}\mathbf{x}) \geq 0. \tag{1.23}$$

□

1.2.4 Pseudoinverse Matrices

Definition 1.22. Let $\mathbf{A}_{m\times n}$ be a matrix. A pseudoinverse matrix $\mathbf{A}^-_{n\times m}$ of $\mathbf{A}_{m\times n}$ is such matrix, that satisfies

$$\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}. \tag{1.24}$$

Remark 1.23. *Pseudoinverse matrix $\boldsymbol{A}^-_{n\times m}$ of $\boldsymbol{A}_{m\times n}$ always exists, but is not given uniquely. For more detailed statement and proof the reader is kindly advised to see [2].*

Proposition 1.24. *Let rank$(\boldsymbol{A}_{m\times r}) = r \geq 1$. Then for any pseudoinverses $\boldsymbol{A}^-$, $\boldsymbol{A}^{T-}$, $(\boldsymbol{A}\boldsymbol{A}^T)^-$ we have that*

(i) $\boldsymbol{A}^-\boldsymbol{A} = \boldsymbol{I}$,

(ii) $\boldsymbol{A}^T\boldsymbol{A}^{T-} = \boldsymbol{I}$,

(iii) $\boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{A}^T)^-\boldsymbol{A} = \boldsymbol{I}$.

Proof. (i) From the assumptions of the theorem follows, that the columns of the matrix $\mathbf{A}$ are linearly independent. I. e. for every vector $\mathbf{y} \in \mathbb{R}_r$ the following holds

$$[\mathbf{A}\mathbf{y}] = \mathbf{0} \implies [\mathbf{y} = \mathbf{0}]. \tag{1.25}$$

By the definition of pseudoinverse matrix we have that $\mathbf{A}\mathbf{A}^-\mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{x}$ for every $\mathbf{x} \in \mathbb{R}_r$. Therefore we have $\mathbf{A}(\mathbf{A}^-\mathbf{A}\mathbf{x} - \mathbf{x}) = \mathbf{0}$ and by using (1.25) we obtain, that $\mathbf{A}^-\mathbf{A}\mathbf{x} = \mathbf{x}$, and since this holds for arbitrary $\mathbf{x}$, we have $\mathbf{A}^-\mathbf{A} = \mathbf{I}$.

(ii) The proof is done analogously as (i).

(iii) Let us denote $\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^-\mathbf{A} = \mathbf{C}$. By the definition of pseudoinverse matrix we have that $\mathbf{A}\mathbf{C}\mathbf{A}^T = \mathbf{A}\mathbf{A}^T$. If we multiply by $\mathbf{A}^-$ from the left, and by $\mathbf{A}^{T-}$ from the right, due to $(i), (ii)$ we obtain $\mathbf{C} = \mathbf{I}$. □

1.3 Properties of Selected Probability Distributions

In this section we will provide characteristics of the probability distributions used in the work. This section is based mainly on [2] and [4]. The details of some computations are featured in the appendix A of this work.

1.3.1 Poisson Probability Distribution

Definition 1.25. Let $\lambda \in [0, \infty)$ and let X be a random variable, such that

$$p(x) = \begin{cases} e^{-\lambda}\frac{\lambda^x}{x!} \ \ for \ \forall \ x \in \mathbb{N}_0, \\ 0 \ \text{otherwise} \end{cases} \tag{1.26}$$

then we say that X has a Poisson probability distribution with parameter λ, and we write $X \sim Po(\lambda)$.

Proposition 1.26. *Let X be a random variable with Poisson probability distribution, then the expectation of X is*

$$\mathbf{E}X = \lambda. \tag{1.27}$$

Proof. Comes directly from Lemma A.1. □

Proposition 1.27. *Let X be a random variable with Poisson probability distribution, then the variance of X is*

$$\mathbf{var}X = \lambda. \tag{1.28}$$

Proof. By Lemma A.2 we have that

$$\mathbf{E}X^2 = \lambda^2 + \lambda. \tag{1.29}$$

Variance of X is then given by (see[2]),

$$\mathbf{var}X = \mathbf{E}\big[X - \mathbf{E}X\big]^2 = \mathbf{E}X^2 - (\mathbf{E}X)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda. \tag{1.30}$$

□

1.3.2 Negative Binomial Probability Distribution

Definition 1.28. Let $\kappa \in \mathbb{N}$ and $p \in (0, 1)$, let X be a random variable such that

$$p(x) = \begin{cases} \binom{\kappa+x-1}{\kappa} q^{\kappa}(1-q)^{\kappa} \;\; for \;\; \forall \;\; x \in \mathbb{N}_0, \\ 0 \;\; \text{otherwise} \end{cases} \tag{1.31}$$

then we say that X has a negative binomial probability distribution and we write $X \sim NBi(\kappa, q)$.

Proposition 1.29. *Let X be a random variable with negative binomial probability distribution, then the expectation of X is*

$$\mathbf{E}X = \frac{\kappa(1-q)}{q}. \tag{1.32}$$

Proof. This comes directly from Lemma A.3. □

Proposition 1.30. *Let X be a random variable with negative binomial probability distribution, then the variance of X is*

$$\mathbf{var}X = \frac{\kappa(1-q)}{q^2}. \tag{1.33}$$

Proof. By using the results of Lemmata A.4, and A.3 we may write

$$\begin{aligned} \mathbf{var}X &= \mathbf{E}X^2 - (\mathbf{E}X)^2 \\ &= \frac{\kappa(\kappa+1)(1-q)^2}{q^2} + \frac{\kappa(1-q)q}{q^2} - \frac{\kappa^2(1-q)^2}{q^2} \\ &= \frac{\kappa[1 - 2q + q^2 + q - q^2]}{q^2} \\ &= \frac{\kappa(1-q)}{q^2}. \end{aligned} \tag{1.34}$$

□

Now we will provide a generalisation of the negative binomial distribution for positive real valued parameter κ.

Definition 1.31. Let $\kappa > 0$ and $p \in (0, 1)$, let X be a random variable such that

$$p(x) = \begin{cases} \frac{\Gamma(x+\kappa)}{x!\Gamma(\kappa)} q^\kappa (1-q)^\kappa \ \ for \ \forall \ x \in \mathbb{N}_0, \\ 0 \ \text{otherwise}, \end{cases} \tag{1.35}$$

then we say that X has a negative binomial probability distribution with positive real parameter κ, and we write $X \sim NBi(\kappa, q)$.

Further on we will use a reparametrisation of the probability mass function of negative binomial distribution that will be introduced in the following Proposition.

Proposition 1.32. *Let $\kappa > 0$ and $p \in (0, 1)$, let X be a random variable such that $X \sim NBi(\kappa, q)$. Set $\mu = \mathbf{E}X$, then the probability mass function $p(x)$ of X can be written as*

$$p(x; \mu, \kappa) = \begin{cases} \frac{\Gamma(x+\kappa)}{x!\Gamma(\kappa)} \left(\frac{\mu}{\kappa+\mu}\right)^x \left(\frac{\kappa}{\kappa+\mu}\right)^\kappa, \quad for \ \forall \ x \in \mathbb{N}_0, \\ 0 \ \text{otherwise}. \end{cases} \tag{1.36}$$

Proof. By Lemma 1.29 we have

$$\mathbf{E}X = \mu = \frac{\kappa(1-q)}{q}. \tag{1.37}$$

From here we obtain

$$q = \frac{\kappa}{\kappa + \mu}. \tag{1.38}$$

By plugging (1.38) into (1.41) we obtain (1.41) which concludes the proof. □

Proposition 1.33. *The variance of the negative binomially distributed random variable X under the reparametrisation introduced in Proposition 1.32 is*

$$\mathbf{var}(X) = \mu + \frac{\mu^2}{\kappa}. \tag{1.39}$$

Proof. By plugging (1.38) into (1.33) we obtain

$$\mathbf{var}(X) = \frac{(\kappa+\mu)^2}{\kappa^2}\left(\frac{\mu}{\kappa+\mu}\right)\kappa = \frac{\mu(\kappa+\mu}{\kappa} = \mu + \frac{\mu^2}{\kappa}. \tag{1.40}$$

□

1.3.3 Pearson Chi Square Distribution

Definition 1.34. Let $n \geq 1$. Let X be a random variable such that

$$f(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{\frac{-x}{2}}, \ \ x > 0, \\ 0 \ \text{otherwise}, \end{cases} \tag{1.41}$$

then we say that X has a χ^2 probability distribution with n degrees of freedom denoted by χ^2_n.

Proposition 1.35. *Let $X_1, ..., X_n$ be independent identically distributed random variables, with Standard Gaussian probability distribution $N(0, 1)$. Then the random variable $Y = (X_1^2 + ... + X_n^2) \sim \chi^2_n$.*

Proof. We need to show that the probability density function of Y is

$$f_n(y) = \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} y^{\frac{n}{2}-1} e^{\frac{-y}{2}}, \ \ y > 0. \tag{1.42}$$

Let $n = 1$. Then for $y > 0$ is the distribution function G of a random variable Y given by

$$G(y) = \boldsymbol{P}(Y < y) = \boldsymbol{P}(X_1^2 < y) = \boldsymbol{P}(-\sqrt{y} < X < \sqrt{y}) = \Phi(\sqrt{y}) - \Phi(-\sqrt{y}). \tag{1.43}$$

Since $\Phi'(x) = \phi(x)$, we have that

$$g(y) = G'(y) = \frac{1}{2\sqrt{y}}\phi(\sqrt{y}) + \frac{1}{2\sqrt{y}}\phi(-\sqrt{y}) = \frac{1}{\sqrt{2}\sqrt{\pi}} y^{-\frac{1}{2}} e^{\frac{y}{2}}. \tag{1.44}$$

So we have that the formula (1.42) holds for $n = 1$. We will continue via induction. Let (1.42) holds for some $n \geq 1$. Then $f_{n+1}(y) = \int f_n(y-z) f_1(z) dz$ and after plugging in $z = uy$ and some further computation we obtain density (1.42) corresponding to index $n+1$. □

Let us denote by $\mathbf{X}^0 = (X_1^0, ..., X_r^0)^T$, where X_i^0 are independent identically distributed Gaussian random variables with $\mathbf{E}(X_i^0) = 0$, $\mathbf{var}(X_i^0) = 1$, i. e. $\mathbf{X}^0 \sim N_p(\mathbf{0}, \mathbf{I}_{p\times p})$.

Proposition 1.36. *Let $\boldsymbol{X} = (X_1, ..., X_n)^T \sim N_n(\boldsymbol{\mu}, \boldsymbol{V})$, and let $rank(\boldsymbol{V}) = r \geq 1$. Furthermore let $\boldsymbol{B}_{n\times r}$ be a matrix of rank r, such that $\boldsymbol{V} = \boldsymbol{B}\boldsymbol{B}^T$. Then $\boldsymbol{\mu} + \boldsymbol{B}\boldsymbol{X}^0 \sim N(\boldsymbol{\mu}, \boldsymbol{V})$.*

Proof. For every vector $\mathbf{c}_n$ we have that

$$\mathbf{c}^T(\boldsymbol{\mu} + \mathbf{B}\mathbf{X}^0) = \mathbf{c}^T\boldsymbol{\mu} + \mathbf{c}^T\mathbf{B}\mathbf{X}^0 \sim N(\mathbf{c}^T\boldsymbol{\mu}, \mathbf{c}^T\mathbf{B}\mathbf{B}^T\mathbf{c}), \tag{1.45}$$

where (1.45) comes from a property of multidimensional normal distribution (see [2]). □

Proposition 1.37. *Let $\boldsymbol{X} \sim N_n(\boldsymbol{\mu}, \boldsymbol{V})$, where $rank(\boldsymbol{V}) = r \geq 1$, then the random variable $Y = (\boldsymbol{X} - \boldsymbol{\mu})^T \boldsymbol{V}^-(\boldsymbol{X} - \boldsymbol{\mu})$ has distribution χ_r^2 for arbitrary choice of the pseudoinverse matrix $\boldsymbol{V}^-$.*

Proof. Due to Proposition 1.36 $\mathbf{X}$ has the same distribution as $\boldsymbol{\mu} + \mathbf{B}\mathbf{X}^0$, where $\mathbf{B}\mathbf{B}^T = \mathbf{V}$, and $\mathbf{X}^0 \sim N_r(\mathbf{0}, \mathbf{I})$. Therefore Y has the same distribution as

$$(\mathbf{B}\mathbf{X}^0)^T(\mathbf{B}\mathbf{B}^T)^-(\mathbf{B}\mathbf{X}^0) = \mathbf{X}^{0T}\mathbf{B}^T(\mathbf{B}\mathbf{B}^T)^-\mathbf{B}\mathbf{X}^0 = \mathbf{X}^{0T}\mathbf{X}^0,$$

since due to Proposition 1.24 (*iii*) $\mathbf{B}^T(\mathbf{B}\mathbf{B})^-\mathbf{B} = \mathbf{I}_{r\times r}$. Since we have also $\mathbf{X}^{0T}\mathbf{X}^0 = (X_1^0)^2 + ... + (X_r^0)^2$ the statement follows from Proposition 1.35. □

Proposition 1.38. *Let $\boldsymbol{X} \sim N_n(\boldsymbol{0}, \boldsymbol{V})$, let $\boldsymbol{A}_{n\times n}$ be as symmetric positive semidefinite matrix. If matrix $\boldsymbol{A}\boldsymbol{V}$ is nonzero and idempotent, then the random variable $\boldsymbol{X}^T\boldsymbol{A}\boldsymbol{X}$ has a distribution χ^2 with degrees of freedom $m = Tr(\boldsymbol{A}\boldsymbol{V})$.*

Proof. By assumption $\mathbf{A}\mathbf{V} \neq 0$ we get that $rank(\mathbf{A}) \geq 1$. Then $\mathbf{A}$ has by Proposition 1.15 decomposition, $\mathbf{A} = \mathbf{B}\mathbf{B}^T$, where $\mathbf{B}$ is of type $n \times rank(\mathbf{A})$. Recall that due to Proposition 1.24 (i) we have that $\mathbf{B}^-\mathbf{B} = \mathbf{I}$. Set $\mathbf{Y} = \mathbf{B}^T\mathbf{X}$. Then $\mathbf{Y} \sim N(\mathbf{0}, \mathbf{B}^T\mathbf{V}\mathbf{B})$ and $\mathbf{X}^T\mathbf{A}\mathbf{X} = \mathbf{Y}\mathbf{Y}$. It is enough to see that $\mathbf{B}^T\mathbf{V}\mathbf{B}$ is idempotent. In that case $\mathbf{I}$ is its pseudoinverse and due to Proposition 1.36 we have that $\mathbf{Y}^T\mathbf{Y} \sim \chi_r^2$, where $r = rank(\mathbf{B}^T\mathbf{V}\mathbf{B}) = Tr(\mathbf{B}^T\mathbf{V}\mathbf{B}) = Tr(\mathbf{B}\mathbf{B}^T\mathbf{V}) = Tr(\mathbf{A}\mathbf{V})$. By the assumption $(\mathbf{A}\mathbf{V})(\mathbf{A}\mathbf{V}) = \mathbf{A}\mathbf{V}$. That means $\mathbf{B}\mathbf{B}^T\mathbf{V}\mathbf{B}\mathbf{B}^T\mathbf{V} = \mathbf{B}\mathbf{B}^T\mathbf{V}$. It follows, that $\mathbf{B}^-\mathbf{B}\mathbf{B}^T\mathbf{V}\mathbf{B}\mathbf{B}^T\mathbf{V} = \mathbf{B}^-\mathbf{B}\mathbf{B}^T\mathbf{V}$, i. e. $(\mathbf{B}^T\mathbf{V}\mathbf{B})(\mathbf{B}^T\mathbf{V}\mathbf{B}) = (\mathbf{B}^T\mathbf{V}\mathbf{B})$. □

Proposition 1.39. *Let $X_1, ..., X_n$ are independent random variables with $X_i \sim N(\mu_i, 1)$, $i = 1, ..., n$. Let $\lambda = \sum_{i=1}^n \mu_i^2 \neq 0$. Then the distribution of random variable $Y = \sum_{i=1}^n X_i^2$ depends only on n and λ and is called a noncentral χ^2 distribution with n degrees of freedom, and a parameter of noncentrality λ, and is denoted $\chi^2_{n,\lambda}$.*

Proof. Set $X_i^0 = X_i - \mu_i$, $\mathbf{X}^0 = (X_1^0, ..., X_n^0)^T$, $\boldsymbol{\mu} = (\mu_1, ..., \mu_n)^T$. Then $Y = (\mathbf{X}^0 + \boldsymbol{\mu})^T(\mathbf{X}^0 + \boldsymbol{\mu})$. Since $\lambda \neq 0$, there is such an orthonormal matrix $\mathbf{B}$, such that its first row is equal to $\lambda^{-\frac{1}{2}}\boldsymbol{\mu}^T$. Since $\mathbf{X}^0 \sim N(\mathbf{0}, \mathbf{I})$ we have that $\mathbf{Z} = \mathbf{B}\mathbf{X}^0 \sim N(\mathbf{0}, \mathbf{I})$ (This can be seen as a result of a more general statement, that linear combination of independent identically distributed Gaussian random variables is again Gaussian with corresponding parameters, for details see [2]). Out of properties of $\mathbf{B}$ we have that $\mathbf{a} = \mathbf{B}\boldsymbol{\mu} = (\sqrt{\lambda}, 0, ..., 0)^T$. Therefore we have

$$\begin{aligned} Y &= (\mathbf{X}^0 + \boldsymbol{\mu})^T(\mathbf{X}^0 + \boldsymbol{\mu}) = (\mathbf{X}^0 + \boldsymbol{\mu})^T\mathbf{B}^T\mathbf{B}(\mathbf{X}^0 + \boldsymbol{\mu}) \\ &= (\mathbf{B}\mathbf{X}^0 + \mathbf{B}\boldsymbol{\mu})^T(\mathbf{B}\mathbf{X}^0 + \mathbf{B}\boldsymbol{\mu}) = (\mathbf{Z} + \mathbf{a})^T(\mathbf{Z} + \mathbf{a}) \end{aligned} \tag{1.46}$$

has distribution dependant only on n and λ. $\square$

1.4 Estimation theory

In this section we will provide theory concerning estimators of parameters of random variables. This section is mainly based on [2].

1.4.1 Statistics and Unbiased Estimators

Assume that random vector $\mathbf{X} = (X_1, ..., X_n)^T$ has a probability density $f(\mathbf{x}, \boldsymbol{\theta})$ with respect to some σ-finite measure μ, where $\boldsymbol{\theta} = (\theta_1, ..., \theta_m)^T$ is an unknown parameter. Our goal is to get the best estimate of $\boldsymbol{\theta}$ based on the vector $\mathbf{X}$, while about $\boldsymbol{\theta}$ we know only that it belongs to some parametric space $\Omega \subset \mathbb{R}_m$. When we are doing a *point estimate* our task is to find a measurable mapping $g : (\mathbb{R}_n, \mathcal{B}_n) \longrightarrow (\mathbb{R}_m, \mathcal{B}_m)$, such that the random vector $\mathbf{T} = g(\mathbf{X})$ would be in some sense the best approximation of the value $\boldsymbol{\theta}$.

Definition 1.40. Assume that $\mathbf{T}$ is an estimator of an unknown parameter $\boldsymbol{\theta}$. We say that $\mathbf{T}$ is *unbiased*, if $\mathbf{ET} = \boldsymbol{\theta}$ for $\forall \boldsymbol{\theta} \in \Omega$. If $\mathbf{ET} = \boldsymbol{\theta} + \mathbf{b}(\boldsymbol{\theta})$, where function $\mathbf{b}$ is not identically zero on Ω we call estimator $\mathbf{T}$ *biased*.

Chapter 2

Transformations of Random Variables

In the following text we will provide some theoretical background regarding the most common transformations of random variables. This chapter is based mainly on [1], partially also on [16].

Let X be a random variable with probability density f and set $Y = t(X)$, where t is measurable function with respect to Lebesgue measure.

2.1 Variance Stabilising Transformation

In classical linear regression model, as well as in corresponding tests it is assumed that the response variable are n independent Gaussian random variables X_i, $i = 1, ..., n$. In such case we have

$$X_i \sim N(\theta_i, \sigma^2), \ \ where \ \mathbf{E}X_i = \theta_i, \ \ and \ \mathbf{var}X_i = \sigma^2, \ \ i = 1, ..., n,$$

where variance of X_i is independent of the expectation of X_i, i. e. σ^2 is constant with respect to the parameter θ_i. In practice this might, and quite often is not the case. In the following we will consider a situation when the independence hypothesis is violated and derive a transformation tackling this problem.

Let X be a random variable with a probability distribution that is depending on a parameter θ. Let the parameter be such that $\mathbf{E}X = \theta$. Furthermore assume that the variance $\mathbf{var}X = \sigma^2$ is a function of the parameter θ as well, i. e. $\mathbf{var}X = \sigma^2(\theta)$. Our task is to find a nonconstant function g, such that $Y = g(X)$ would have a variance, that does not depend on θ. In general this problem does not have a solution, we will try to obtain at least a suitable approximation. In the following theorem we will considered random variable Y to be obtained via transformation $Y = g(\theta) + (X - \theta)g'(\theta)$, rather than using $Y = g(X)$. We are hence, approximating $g(X)$ by it's Taylor expansion around the point θ up to a linear term.

Theorem 2.1. *Let X be a random variable with probability distribution dependent on a parameter θ, such that $\mathbf{E}X = \theta$, and $\mathbf{var}X = \sigma^2(\theta)$. Let g be a function that is smooth along with it's first and second derivative. Let Y be a random variable given by $Y = g(\theta) + (X - \theta)g'(\theta)$, then $\mathbf{var}Y$ is constant with respect to θ if*

$$g(\theta) = c \int \frac{d\theta}{\sigma(\theta)}, \ \ c \in \mathbb{R}. \tag{2.1}$$

Proof. By definition of Y we have

$$\begin{aligned}\mathbf{E}Y &= \big[\mathbf{g}(\theta) + \mathbf{g}'(\theta)\mathbf{X} - \mathbf{g}'(\theta)\theta\big] \\ &= g(\theta) + g'(\theta)\theta - g'(\theta)\theta \\ &= g(\theta),\end{aligned} \tag{2.2}$$

and

$$\begin{aligned}\mathbf{var}Y &= \mathbf{E}Y^2 - (\mathbf{E}Y)^2 \\ &= \mathbf{E}\big[g(\theta) + (X-\theta)g'(\theta)\big]^2 - \big(\mathbf{E}Y\big)^2 \\ &= \mathbf{E}\big[g^2(\theta) + 2Xg'(\theta)g(\theta) - 2\theta g'(\theta)g(\theta) + X^2\big(g'(\theta)\big)^2 - 2X\theta\big(g'(\theta)\big)^2 \\ &+ \theta^2\big(g'(\theta)\big)^2\big] - \big(g(\theta)\big)^2.\end{aligned} \tag{2.3}$$

Since we have that $\mathbf{E}X^2 = \sigma^2 + \big(\mathbf{E}X\big)^2 = \sigma^2 + \theta^2$ we can plug it in (2.3) and obtain

$$\mathbf{var}Y = \sigma^2(\theta)\big(g'(\theta)\big)^2. \tag{2.4}$$

In order for $\mathbf{var}Y$ to be independent with respect to θ we need

$$\sigma(\theta)g'(\theta) = c, \tag{2.5}$$

where c is an arbitrary constant. Out of this condition we obtain

$$g(\theta) = c\int \frac{d\theta}{\sigma(\theta)},$$

which concludes the proof. □

Definition 2.2. The function g from Theorem 2.1 that satisfies (2.1) is called variance stabilising transformation.

2.2 Box - Cox Transformation

In this section we will describe the Box-Cox transformation. A transformation from the family of power transformations that is often applied on nonnormal response data in order to achieve stability of variances among the data. This section is based on [16]. Suppose we have data sample $(X_1, ..., X_n)$ of a distribution of a variable X that is strictly positive. We will consider a power transformation

$$Z = X^\lambda \tag{2.6}$$

and try to find the best value of λ to use. It is obvious that a problem occurs for the choice $\lambda = 0$ that would make all the entries of the sample equal to one.

Definition 2.3. Given a random variable Y we define the family of power transformations for varying parameter λ as follows

$$W = \begin{cases} (Y^\lambda - 1)/\lambda \;\; for \;\; \lambda \neq 0, \\ \\ \ln(Y), \;\; for \;\; \lambda = 0 \end{cases}. \tag{2.7}$$

We will call this family of transformations Box-Cox transformations.

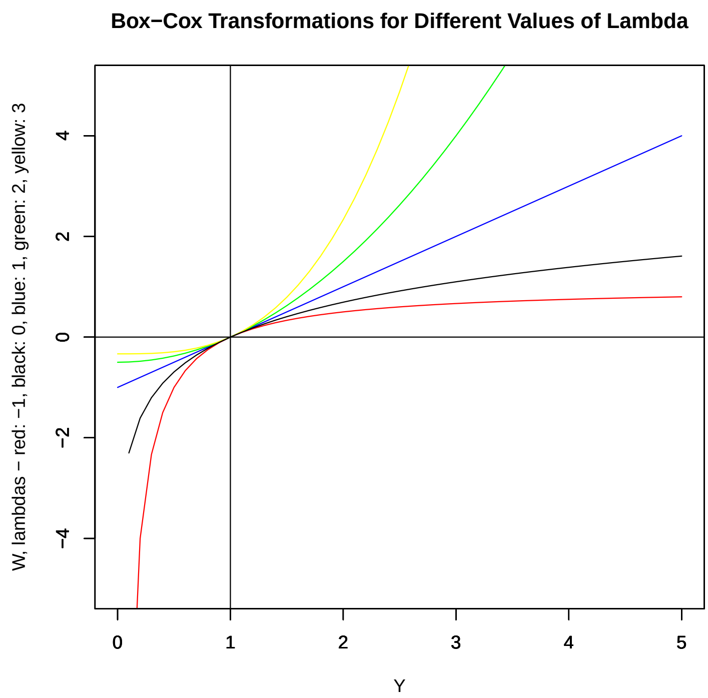


Figure 2.1: Comparison Box-Cox transformations for different values of λ, by colours: $\lambda = -1$ - red, $\lambda = 0$ - black, $\lambda = 1$ - blue, $\lambda = 2$ - green, $\lambda = -3$ - yellow).

The problem in $\lambda = 0$ is overcome, because $\ln(Y)$ is the appropriate limit of $(Y^\lambda - 1)/\lambda$ as $\lambda \longrightarrow 0$. Therefore the family of transformations is now continuous in λ.

The values of W of (2.7) can change greatly as λ varies, which complicates finding the optimal value of λ. For that reason we shall introduce an alternative form of family of transformations W.

Definition 2.4. Given a random variable Y we define the family of power transformations for varying parameter λ as follows

$$V = \begin{cases} (Y^\lambda - 1)/(\lambda \cdot \hat{Y}^{\lambda-1}) \;\; for \;\; \lambda \neq 0, \\ \\ \hat{Y}\ln(Y), \;\; for \;\; \lambda = 0 \end{cases}, \tag{2.8}$$

where the term $\hat{Y}^{\lambda-1}$ is the n-th power of the appropriate Jacobian of the transformation which converts the set Y_i into the set of W_i.

Remark 2.5. *The multiplication with the the n-th power of the Jacobian of the transformation which converts the set Y_i into the set of W_i ensures that the unit volume is preserved in moving from the set of Y_i to W_i.*

The best value of the parameter λ can be determined by using maximum likelihood estimation (see [16]). In [17] the estimation via maximum likelihood and also its Bayesian equivalent is discussed.

It is also possible to relax the assumption of the positiveness of X, if it is negative, but bounded from below, by introducing a shift parameter. In such scenario however the standard asymptotic results of maximum likelihood theory may not apply since the range of the distribution is determined by unknown shift parameter (see [18]).

2.3 Yeo-Johnson Transformation

In the previous section we have discussed the Box-Cox transformations family. Our main limitation was the assumption of positiveness of the random variable X to which the transformation was applied. In this section we will discuss another family of transformations appropriate to approximate normality, the Yeo-Johnson transformations, which unlike the Box-Cox transformations are well defined on the whole real line, so we may drop the assumption on positiveness of X. This section is based on [18].

Definition 2.6. Let X be a real random variable, we define a family of Yeo-Johnson transformations as follows.

$$J = \begin{cases} \frac{(1+X)^{\lambda}-1}{\lambda} & for \ \ X \geq 0, \lambda \neq 0 \\ \ln(1+X) & for \ \ X \geq 0, \lambda = 0, \\ \frac{-((1-X)^{2-\lambda}-1)}{2-\lambda} & for \ \ X < 0, \lambda \neq 2, \\ -\ln(1-X) & for \ \ X < 0\lambda = 2. \end{cases} \tag{2.9}$$

The transformation given by (2.9) is also designed so that it would reduce the skewness parameter of the variable X. The value of the parameter λ can be estimated by maximum likelihood estimation (see [18]).

Chapter 3

A Study of Selected Transformations

3.1 Variance Stabilising Transformation for Random Variable with Poisson Probability Distribution

In this section we will study a behaviour of a variance stabilising transformation of a random variable X with Poisson probability distribution, when a nonnegative constant c is added. We will derive moments of the transformed variable and find an optimal value of c such that the transformation would stabilise the variance of X. The content of this section is based on [1].

First we will provide derivation of the variance stabilising transformation using the formula (2.1) of Theorem 2.1.

Proposition 3.1. *Let $X \sim Po(\lambda)$ be a random variable with Poisson probability distribution. Then the variance stabilising transformation in the sense of the Theorem 2.1 is given by*

$$Y = g(X) = \sqrt{x} \tag{3.1}$$

Proof. By Proposition 1.27 we have that

$$\mathbf{var}(X) = \lambda, \tag{3.2}$$

hence, the standard deviation of X is $\sigma = \sqrt{\lambda}$. The random variable X clearly satisfies the assumptions of Theorem 2.1, hence, the variance stabilising transformation g is given by

$$g(\lambda) = c_0 \int \frac{d\lambda}{\sqrt{\lambda}} = 2c_0\sqrt{\lambda} + c_1. \tag{3.3}$$

By choosing $c_0 = \frac{1}{2}$, and $c_1 = 0$ we get $g(x) = \sqrt{x}$. □

In the rest of this section if not explicitly stated otherwise we will assume that X is a random variable with Poisson probability distribution and parameter λ (see Definition 1.25), and Y a random variable obtained by transformation

$$Y = \sqrt{X + c} \tag{3.4}$$

where c is a positive constant. In order to simplify the following computations, we will first consider the following, let

$$Z = X - \lambda \tag{3.5}$$

be a random variable, and

$$\lambda' = \lambda + c. \tag{3.6}$$

The transformation 3.4 is then

$$Y = \sqrt{Z + \lambda'}. \tag{3.7}$$

By Taylor theorem for any $z \geq -\lambda'$ we obtain an infinite series representation

$$y = \sqrt{\lambda'}\left[1 + a_1\frac{z}{\lambda'} - a_2\left(\frac{z}{\lambda'}\right)^2 + ... + (-1)^s a_{s-1}\left(\frac{z}{\lambda'}\right)^{s-1}\right] + R_s, \tag{3.8}$$

where R_s is a reminder term and coefficients a_s are given by

$$a_s = (-1)^{s+1}\frac{-2s+3}{2^s s!}. \tag{3.9}$$

Lemma 3.2. *For $z > 0$ the term R_s satisfies*

$$|R_s| < \frac{a_s z^s}{(\lambda')^{s-\frac{1}{2}}}. \tag{3.10}$$

Proof. This is a direct result of Lagrange's form of the reminder term (see [6]). □

Now we would like to find a bound $C(s)$ for R_s such that $|R_s| < C(s)$ on a larger interval, namely $z > -\lambda^{'}$.

Lemma 3.3. *For $z > -\lambda^{'}$ the term R_s satisfies*

$$|R_s| \leq G(s)\frac{|z|^s}{(\lambda^{'})^{s-\frac{1}{2}}}.$$

Proof. If we assume that $|z| \leq \lambda^{'}$, we obtain directly from (3.8) the following

$$R_s(\lambda^{'})^{-\frac{1}{2}} = \left(1 + \frac{z}{\lambda'}\right)^{\frac{1}{2}} - \left\{1 + a_1\frac{z}{\lambda'} - a_2\left(\frac{z}{\lambda'}\right)^2 + ... + (-1)^s a_{s-1}\left(\frac{z}{\lambda'}\right)^{s-1}\right\} \tag{3.11}$$

$$= \sum_{n=1}^{\infty}(-1)^{i+1}a_i\left(\frac{z}{\lambda'}\right)^i. \tag{3.12}$$

We notice that the series on the right hand side of the equation (3.11) is convergent and hence, we can write

$$\frac{R_s(\lambda^{'})^{s-\frac{1}{2}}}{z^s} = \sum_{n=1}^{\infty}(-1)^{i+1}a_i\left(\frac{z}{\lambda'}\right)^{i-s}, \tag{3.13}$$

where the right hand side is convergent and bounded. Let us assume $G(s)$ as the bound of the right hand side to its absolute magnitude, then we have

$$|R_s| \leq G(s)\frac{|z|^s}{(\lambda^{'})^{s-\frac{1}{2}}}. \tag{3.14}$$

Due to Lemma 3.2 the inequality (3.10) holds, if we compare (3.10) with (3.14) we see, that (3.14) holds for all $z > -\lambda^{'}$. □

Following propositions will provide us with some tools necessary for deriving the approximations of expectation and variance of Y.

Lemma 3.4. *Let us denote* $\mu'_{Z,k} = \mathbf{E}Z^k$ *the* k*-th moment of* Z *and* $\mu_{X,k} = \mathbf{E}\big[X - \mathbf{E}X\big]^k$. *the* k*-th central moment of* X. *Then we have*

$$\mu'_{Z,k} = \mu_{X,k} \tag{3.15}$$

Proof. By the properties of a random variable with Poisson probability distribution we have $\mathbf{E}X = \lambda$ (see [2]),

$$\begin{aligned}\mu'_{Z,k} &= \mathbf{E}Z^k \\ &= \mathbf{E}\big[X - \lambda\big]^k \\ &= \mathbf{E}\big[X - \mathbf{E}X\big]^k \\ &= \mu_{X,k}.\end{aligned}$$

□

Lemma 3.5. *For every* $n \geq 0$*, one has that*

$$\mathbf{E}_\lambda\big[X\big]^k = T_k(\lambda), \quad \lambda > 0, \tag{3.16}$$

where T_n *is the* n*-th Touchard polynomial, as defined in Definition 1.4.*

Proof. The proof uses results of the theory of Bell polynomials and cumulants, which is not developed in this work, and therefore the proof is not given here, and can be found in [3]. □

Corollary 3.6. *For every* $n \geq 0$*, there exists a polynomial of degree at most* n*, denoted by* $\tilde{T}_n$*, such that*

$$\tilde{T}_n(\lambda) = \mathbf{E}_\lambda\big[(X - \lambda)^n\big], \quad \lambda > 0. \tag{3.17}$$

Proof. This is a direct result of a Lemma 3.5 (see [3]). □

Lemma 3.7. *For every* $n \geq 1$ *we have*

$$\tilde{T}_{n+1}(\lambda) = \lambda \sum_{k=0}^{n-1} \binom{n}{k} \tilde{T}_k(\lambda). \tag{3.18}$$

Proof. The proof of this Lemma is based on a theory that is not developed in this work and can be found in [3]. □

Remark 3.8. *Using the results obtained by Lemmata 3.4, 3.7, and Corollary 3.6 we obtain by direct computation the first few moments of* Z*:*

$$\begin{aligned}\mu'_{Z,1} &= 0, \\ \mu'_{Z,2} &= \lambda, \\ \mu'_{Z,3} &= \lambda, \\ \mu'_{Z,4} &= 3\lambda^2 + \lambda, \\ \mu'_{Z,5} &= 10\lambda^2 + \lambda, \\ \mu'_{Z,6} &= 15\lambda^3 + 25\lambda^2 + \lambda.\end{aligned} \tag{3.19}$$

Let us first derive the approximation of the expected value $\mathbf{E}Y$ for λ large.

Lemma 3.9. *Let Y be the random variable obtained by transformation* (3.4). *Then its expectation may be approximated by*

$$\mathbf{E}Y = \sqrt{\lambda + c} - \frac{1}{8}\frac{1}{\lambda^{\frac{1}{2}}} + \frac{24c - 7}{128\lambda^{\frac{3}{2}}} + O\Big(\frac{1}{\lambda^{\frac{5}{2}}}\Big). \tag{3.20}$$

Proof. By the Corollary 3.6 we may take expected values of the right hand side of (3.8), and its powers, and derive asymptotic expansions for the moments of Y as $\lambda \longrightarrow \infty$.

For the expected values we have

$$\begin{aligned}\mathbf{E}Y = \mathbf{E}\Big\{&\sqrt{\lambda + c} + \frac{1}{2}\frac{Z}{(\lambda + c)^{\frac{1}{2}}} - \frac{1}{8}\frac{Z^2}{(\lambda + c)^{\frac{3}{2}}} + \frac{1}{16}\frac{Z^3}{(\lambda + c)^{\frac{5}{2}}}\\ &- \frac{5}{128}\frac{Z^4}{(\lambda + c)^{\frac{7}{2}}} + \frac{7}{256}\frac{Z^5}{(\lambda + c)^{\frac{9}{2}}} - \frac{21}{1024}\frac{Z^6}{(\lambda + c)^{\frac{11}{2}}} + O\Big(\frac{1}{(\lambda + c)^{\frac{13}{2}}}\Big)\Big\}.\end{aligned} \tag{3.21}$$

By using (3.19), and the linearity property of expectation (see [4]) we obtain

$$\begin{aligned}\mathbf{E}Y = &\sqrt{\lambda + c} + \frac{1}{2}\frac{0}{(\lambda + c)^{\frac{1}{2}}} - \frac{1}{8}\frac{\lambda}{(\lambda + c)^{\frac{3}{2}}} + \frac{1}{16}\frac{\lambda}{(\lambda + c)^{\frac{5}{2}}}\\ &- \frac{5}{128}\frac{3\lambda^2 + \lambda}{(\lambda + c)^{\frac{7}{2}}} + \frac{7}{256}\frac{10\lambda^2 + \lambda}{(\lambda + c)^{\frac{9}{2}}} - \frac{21}{1024}\frac{15\lambda^3 + 25\lambda^2 + \lambda}{(\lambda + c)^{\frac{11}{2}}} + O\Big(\frac{1}{(\lambda + c)^{\frac{13}{2}}}\Big) \ .\end{aligned} \tag{3.22}$$

We derive the asymptotic expansions of all listed fractions

$$\begin{aligned}\mathbf{E}Y = &\sqrt{\lambda + c} + \Big[-\frac{1}{8}\frac{1}{\lambda^{\frac{1}{2}}} + \frac{3}{16}\frac{c}{\lambda^{\frac{3}{2}}} - \frac{15}{64}\frac{c^2}{\lambda^{\frac{5}{2}}} + O\Big(\frac{1}{\lambda^{\frac{7}{2}}}\Big)\Big]\\ &+ \Big[\frac{1}{16}\frac{1}{\lambda^{\frac{3}{2}}} - \frac{5}{32}\frac{c}{\lambda^{\frac{5}{2}}} + O\Big(\frac{1}{\lambda^{\frac{7}{2}}}\Big)\Big] + \Big[-\frac{15}{128}\frac{1}{\lambda^{\frac{3}{2}}} + \frac{1}{256}\frac{105c - 10}{\lambda^{\frac{5}{2}}} + O\Big(\frac{1}{\lambda^{\frac{7}{2}}}\Big)\Big]\\ &+ \Big[\frac{35}{128}\frac{1}{\lambda^{\frac{5}{2}}} + O\Big(\frac{1}{\lambda^{\frac{7}{2}}}\Big)\Big] + \Big[-\frac{315}{1024}\frac{1}{\lambda^{\frac{5}{2}}} + O\Big(\frac{1}{\lambda^{\frac{7}{2}}}\Big)\Big] \ .\end{aligned} \tag{3.23}$$

Hence,

$$\begin{aligned}\mathbf{E}Y &= \sqrt{\lambda + c} - \frac{1}{8}\frac{1}{\lambda^{\frac{1}{2}}} + \frac{3}{16}\frac{c}{\lambda^{\frac{3}{2}}} + \frac{1}{16}\frac{1}{\lambda^{\frac{3}{2}}} - \frac{15}{128}\frac{1}{\lambda^{\frac{3}{2}}}\\ &\quad - \frac{15}{64}\frac{c^2}{\lambda^{\frac{5}{2}}} - \frac{5}{32}\frac{c}{\lambda^{\frac{5}{2}}} + \frac{1}{256}\frac{105c - 10}{\lambda^{\frac{5}{2}}} + \frac{35}{128}\frac{1}{\lambda^{\frac{5}{2}}} - \frac{315}{1024}\frac{1}{\lambda^{\frac{5}{2}}} + O\Big(\frac{1}{\lambda^{\frac{7}{2}}}\Big)\\ &= \sqrt{\lambda + c} - \frac{1}{8}\frac{1}{\lambda^{\frac{1}{2}}} + \frac{24c - 7}{128\lambda^{\frac{3}{2}}} + \frac{-240c^2 + 260c - 75}{\lambda^{\frac{5}{2}}} + O\Big(\frac{1}{\lambda^{\frac{7}{2}}}\Big)\end{aligned} \tag{3.24}$$

And we get the approximation of the expectation of Y

$$\mathbf{E}Y = \sqrt{\lambda + c} - \frac{1}{8}\frac{1}{\lambda^{\frac{1}{2}}} + \frac{24c - 7}{128\lambda^{\frac{3}{2}}} + O\Big(\frac{1}{\lambda^{\frac{5}{2}}}\Big) \tag{3.25}$$

□

Let us now derive the approximation of variance of Y.

Lemma 3.10. *Let Y be the random variable obtained by transformation* (3.4). *Then its variance may be approximated by*

$$\mathbf{var}Y = \frac{1}{4} + \frac{3 - 8c}{32\lambda} + \frac{32c^2 - 52c + 17}{128\lambda^2} + O\Big(\frac{1}{\lambda^3}\Big). \tag{3.26}$$

Proof. In order to obtain the approximation of variance of Y we will use (3.20) to compute $(\mathbf{E}Y)^2$ i. e.

$$\begin{aligned}(\mathbf{E}Y)^2 = \lambda + c + \frac{1}{64\lambda} - 2\cdot\sqrt{\lambda+c}\cdot\frac{1}{8}\cdot\frac{1}{\lambda^{\frac{1}{2}}} + 2\cdot\sqrt{\lambda+c}\cdot\frac{24c-7}{128\lambda^{\frac{3}{2}}}\\ + 2\cdot\sqrt{\lambda+c}\cdot\frac{-240c^2+260c-75}{1024\lambda^{\frac{5}{2}}} - 2\cdot\frac{1}{8\lambda^{\frac{1}{2}}}\cdot\frac{24c-7}{128\lambda^{\frac{3}{2}}} + O\left(\frac{1}{\lambda^3}\right).\end{aligned} \tag{3.27}$$

The following holds for $\lambda \longrightarrow \infty$

$$\sqrt{\lambda+c} = \sqrt{\lambda} + \frac{1}{2}\cdot\frac{c}{\lambda^{\frac{1}{2}}} - \frac{1}{8}\cdot\frac{c^2}{\lambda^{\frac{3}{2}}} + \frac{1}{16}\cdot\frac{c^3}{\lambda^{\frac{5}{2}}} + O\left(\frac{1}{\lambda^{\frac{7}{2}}}\right), \tag{3.28}$$

and by pluging it into (3.27) we obtain

$$\begin{aligned}(\mathbf{E}Y)^2 = \lambda + c + \frac{1}{64\lambda} - \frac{1}{4} - \frac{c}{8\lambda} + \frac{c^2}{32\lambda^2} + \frac{24c-7}{64\lambda} + \frac{24c^2-7c}{128\lambda^2}\\ + \frac{-240c^2+260c-75}{512\lambda^2} + \frac{7-24c}{512\lambda^2} + O\left(\frac{1}{\lambda^3}\right),\end{aligned} \tag{3.29}$$

and after some further computation we finally obtain

$$(\mathbf{E}Y)^2 = -\frac{1}{4} + \lambda + c + \frac{8c-3}{32\lambda} + \frac{-32c^2+52c-17}{128\lambda^2} + O\left(\frac{1}{\lambda^3}\right). \tag{3.30}$$

$\mathbf{E}Y^2$ is obtained as follows

$$\mathbf{E}Y^2 = \mathbf{E}(\sqrt{Z+\lambda'})^2 = \mathbf{E}(Z+\lambda') = \lambda' = \lambda + c. \tag{3.31}$$

And finally we can derive the variance of Y

$$\begin{aligned}\mathbf{var}Y &= \mathbf{E}(Y-\mathbf{E}Y)^2 = \mathbf{E}Y^2 - (\mathbf{E}Y)^2\\ &= \lambda + c + \frac{1}{4} - \lambda - c + \frac{3-8c}{32\lambda} + \frac{32c^2-52c+17}{128\lambda^2} + O\left(\frac{1}{\lambda^3}\right)\\ &= \frac{1}{4} + \frac{3-8c}{32\lambda} + \frac{32c^2-52c+17}{128\lambda^2} + O\left(\frac{1}{\lambda^3}\right).\end{aligned} \tag{3.32}$$

□

We will now show that for the choice of $c = \frac{3}{8}$ Y has most nearly constant variance for large values of parameter λ.

Theorem 3.11. *Let Y be a random variable obtained using transformation* (3.4), *where c is a positive parameter. Let us denote*

$$h(\lambda, c) = \mathbf{var}Y - \frac{1}{4} = \frac{3-8c}{32\lambda} + \frac{32c^2-52c+17}{128\lambda^2} + O\left(\frac{1}{\lambda^3}\right). \tag{3.33}$$

Then $\min_{c>0}\{|h(\lambda,c)|\}$, *given that c is constant with respect to λ is attained for $c = \frac{3}{8}$ for $\lambda \longrightarrow \infty$.*

Proof. From the form of $h(\lambda, c)$ given by (3.33) we observe, that the terms will vanish one after another as $\lambda \longrightarrow \infty$ with the term $\frac{3-8c}{32\lambda}$ vanishing the last for its denominator is of linear order, hence, the minimum will be attained for $\lambda \longrightarrow \infty$ if $\frac{3-8c}{32\lambda} = 0$ from here we obtain that $c = \frac{3}{8}$. □

Lemma 3.12. *Let $Y_1, ..., Y_n$ be a random sample of the distribution identical to the one of Y given by* (3.4)*. Let us denote λ_Y the estimate of λ derived by applying the transformation* (3.4) *in reverse to $\overline{Y} = \frac{1}{n}\sum_{k=1}^{n} Y_k$ of $Y_1, ..., Y_n$ for $n \longrightarrow \infty$. Then*

$$\lambda_Y = \lambda - \frac{1}{4} + \frac{8c-3}{32\lambda} + O\Big(\frac{1}{\lambda^{\frac{3}{2}}}\Big). \tag{3.34}$$

Proof. Set

$$\mathbf{E}(Y) = \sqrt{\lambda_Y + c}. \tag{3.35}$$

From 3.20 we have

$$\sqrt{\lambda_Y + c} = \sqrt{\lambda + c} - \frac{1}{8\lambda^{\frac{1}{2}}} + \frac{24c-7}{128\lambda^{\frac{3}{2}}} + O\Big(\frac{1}{\lambda^{\frac{5}{2}}}\Big). \tag{3.36}$$

By taking a square of (3.36) we obtain the following

$$\lambda_Y + c = \lambda + c + \frac{1}{64\lambda} + -2\sqrt{\lambda + c}\cdot\frac{1}{8\lambda^{\frac{1}{2}}} + 2\sqrt{\lambda + c}\cdot\frac{24c-7}{128\lambda^{\frac{3}{2}}} + O\Big(\frac{1}{\lambda^{\frac{3}{2}}}\Big). \tag{3.37}$$

Now we substitute the term $\sqrt{\lambda + c}$ on the right hand side of (3.37) by (3.28) and obtain

$$\begin{aligned}\lambda_Y &= \frac{1}{64\lambda} - 2\Big[\lambda^{\frac{1}{2}} + \frac{1}{2}\cdot\frac{c}{\lambda^{\frac{1}{2}}} - \frac{1}{8}\cdot\frac{c^2}{\lambda^{\frac{3}{2}}} + O\Big(\frac{1}{\lambda^{\frac{5}{2}}}\Big)\Big]\cdot\frac{1}{8\lambda^{\frac{1}{2}}}\\ &= 2\cdot\Big[\lambda^{\frac{1}{2}} + \frac{1}{2}\cdot\frac{c}{\lambda^{\frac{1}{2}}} - \frac{1}{8}\cdot\frac{c^2}{\lambda^{\frac{3}{2}}} + O\Big(\frac{1}{\lambda^{\frac{5}{2}}}\Big)\Big]\cdot\frac{24c-7}{128\lambda^{\frac{3}{2}}} + O\Big(\frac{1}{\lambda^{\frac{3}{2}}}\Big)\\ &= \lambda + \frac{1}{64\lambda} - \frac{1}{4} - \frac{c}{8\lambda} + \frac{24c-7}{64\lambda} + O\Big(\frac{1}{\lambda^{\frac{3}{2}}}\Big)\\ &= \lambda - \frac{1}{4} + \frac{8c-3}{32\lambda} + O\Big(\frac{1}{\lambda^{\frac{3}{2}}}\Big)\end{aligned} \tag{3.38}$$

as intended. □

Definition 3.13. Regarding the results obtained in Lemma 3.12 we define the bias as

$$b_Y = \lambda_Y - \lambda. \tag{3.39}$$

Corollary 3.14. *The following equation holds*

$$b_Y = -\frac{1}{4} + \frac{8c-3}{32\lambda} + O\Big(\frac{1}{\lambda^{\frac{3}{2}}}\Big). \tag{3.40}$$

Proof. This is a direct result of Lemma 3.12. □

With the Corollary 3.14 we can state the following Theorem.

Theorem 3.15. *Let X be a random variable with Poisson distribution with parameter λ. Let Y be a random variable given by transformation* (3.4)*. Then* $\min_{c\geq 0}\{|b_Y(\lambda, c)|\}$*, given that c is constant with respect to λ is attained $c = \frac{3}{8}$ as $\lambda \longrightarrow \infty$.*

Proof. The term whose denominator is a linear function of λ will vanish the last as $\lambda \longrightarrow \infty$. Therefore we achieve a minimal value of $|b_Y(\lambda, c)|$ by eliminating this term by choosing $c = \frac{3}{8}$ as $\lambda \longrightarrow \infty$. □

3.2 Study of Chosen Transformations for Random Variable with Negative Binomial Probability Distribution

In this section we will study the variance stabilising transformations for the random variable X with a negative binomial probability distribution with expected value μ (see Proposition 1.32) and a known shape parameter κ. Its probability mass function is given by (1.41) of Proposition 1.32, its variance is given by Proposition 1.33. Many of the theoretical results of this section are based on [1].

First we will provide derivation of the variance stabilising transformation using the formula (2.1) of Theorem 2.1.

Proposition 3.16. *Let $X \sim NBi(\mu, \kappa)$ be a random variable with negative binomial probability distribution. Then the variance stabilising transformation in the sense of the Theorem 2.1 is given by*

$$Y = g(X) = 2\sinh^{-1}\left(\sqrt{\frac{X}{\kappa}}\right). \tag{3.41}$$

Proof. By Proposition 1.33 we have that

$$\mathbf{var}(X) = \mu + \frac{\mu^2}{\kappa}, \tag{3.42}$$

hence, the standard deviation σ of X is given by

$$\sigma = \sqrt{\mu + \frac{\mu^2}{\kappa}}. \tag{3.43}$$

The random variable X clearly satisfies the assumptions of Theorem 2.1, hence, the variance stabilising transformation g is given by

$$\begin{aligned} g(\mu) &= c_0 \int \frac{d\mu}{\sqrt{\mu + \frac{\mu^2}{\kappa}}} = c_0\sqrt{\kappa} \int \frac{d\mu}{\sqrt{\kappa\mu + \mu^2}} = c_0\sqrt{\kappa} \int \frac{d\mu}{\sqrt{\kappa\mu + \mu^2 + \frac{\kappa^2}{4} - \frac{\kappa^2}{4}}} \\ &= c_0\sqrt{\kappa} \int \frac{d\mu}{\sqrt{\left(\mu + \frac{\kappa}{2}\right)^2 - \frac{\kappa^2}{4}}}, \end{aligned} \tag{3.44}$$

where c_0 is a constant. In order to compute this integral we will first substitute u for $\mu + \frac{\kappa}{2}$ to obtain

$$g(u) = c_0\sqrt{\kappa} \int \frac{du}{\sqrt{u^2 - \frac{\kappa^2}{4}}} \tag{3.45}$$

There are several possibilities to solve this integral, one might be for example to use a transformation of u into hyperbolometric function and then use the corresponding hyperbolometric identity formula. We will for instance use another possible approach and substitute u with the term $\frac{1}{2}\kappa\sec(s)$ to get

$$g(s) = c_0\sqrt{\kappa} \int \frac{\frac{1}{2}\kappa\tan(s)\sec(s)ds}{\sqrt{\frac{1}{4}\kappa^2\sec^2(s) - \frac{\kappa^2}{4}}} = c_0\sqrt{\kappa} \int \frac{\frac{1}{2}\kappa\tan(s)\sec(s)ds}{\frac{\kappa}{2}\sqrt{\sec^2(s) - 1}} \tag{3.46}$$

Using now the following trigonometric identity (see [5])

$$\sec^2(s) - 1 = \tan^2(s) \tag{3.47}$$

we obtain

$$\begin{aligned} g(s) &= c_0\sqrt{\kappa}\int \sec(s)ds = c_0\sqrt{\kappa}\int \frac{1}{\cos(s)}ds = c_0\sqrt{\kappa}\int \frac{\frac{1+\sin(s)}{\cos(s)}}{\frac{1+\sin(s)}{\cos(s)}} \cdot \frac{ds}{\cos(s)} \\ &= \int \frac{\frac{1+\sin(s)}{\cos^2(s)}}{\frac{1+\sin(s)}{\cos(s)}}ds \end{aligned} \tag{3.48}$$

we notice that $\frac{1+\sin(s)}{\cos(s)} = \sec(s) + \tan(s)$ and $\frac{1+\sin(s)}{\cos^2(s)} = \frac{d}{ds}(\sec(s) + \tan(s))$, and therefore we have

$$g(s) = c_0\sqrt{\kappa}\ln\big[\tan(s) + \sec(s)\big] + c_1. \tag{3.49}$$

Now we return to the original variable μ. First, we have that $s = \sec^{-1}\left(\frac{2u}{\kappa}\right)$, so we obtain

$$g(u) = c_0\sqrt{\kappa}\ln\left[\tan\left(\sec^{-1}\left(\frac{2u}{\kappa}\right)\right) + \sec\left(\sec^{-1}\left(\frac{2u}{\kappa}\right)\right)\right] + c_1. \tag{3.50}$$

Now using that $\sec\sec^{-1}(z) = z$ and $\tan\sec^{-1}(z) = \sqrt{1 - \frac{1}{z^2}} \cdot z$ we obtain that

$$g(u) = c_0\sqrt{\kappa}\ln\left[\frac{2u}{\kappa}\sqrt{1 - \frac{\kappa^2}{4u^2}} + \frac{2u}{\kappa}\right] + c_1, \tag{3.51}$$

which can be written as

$$g(u) = c_0\sqrt{\kappa}\ln\left[\frac{u}{\kappa}\left(2\sqrt{1 - \frac{\kappa^2}{4u^2}} + 2\right)\right] + c_1 \tag{3.52}$$

And finally we have $u = \frac{\kappa}{2} + \mu$ hence, we obtain

$$\begin{aligned} g(\mu) &= c_0\sqrt{\kappa}\ln\left[\frac{\kappa + 2\mu}{2\kappa}\left(2\sqrt{1 - \frac{\kappa^2}{(\kappa+2\mu)^2}} + 2\right)\right] + c_1 \\ &= c_0\sqrt{\kappa}\ln\left[\frac{\kappa + 2\mu}{\kappa}\left(\sqrt{\frac{4\mu\kappa + 4\mu^2}{(\kappa+2\mu)^2}} + 1\right)\right] + c_1 \\ &= c_0\sqrt{\kappa}\ln\left[\frac{1}{\kappa}\left(2\sqrt{\mu\kappa + \mu^2} + (\kappa + 2\mu)\right)\right] + c_1 \end{aligned} \tag{3.53}$$

Now we observe that

$$2\sqrt{\mu\kappa + \mu^2} + (\kappa + 2\mu) = \mu + 2\sqrt{\mu\kappa + \mu^2} + (\mu + \kappa) = (\sqrt{\mu} + \sqrt{\mu + \kappa})^2, \tag{3.54}$$

and therefore we get

$$\begin{aligned} g(\mu) &= 2c_0\sqrt{\kappa}\ln\left[\frac{\sqrt{\kappa + \mu} + \sqrt{\mu}}{\sqrt{\kappa}}\right] + c_1 = 2c_0\sqrt{\kappa}\ln\left[\sqrt{1 + \frac{\mu}{\kappa}} + \sqrt{\frac{\mu}{\kappa}}\right] + c_1 \\ &= 2c_0\sqrt{\kappa}\sinh^{-1}\left(\sqrt{\frac{\mu}{\kappa}}\right) + c_1 \end{aligned} \tag{3.55}$$

Where the last equality is due to definition of argument of hyperbolic sine via natural logarithm (see [5]). By choosing $c_0 = \frac{1}{\sqrt{\kappa}}$, and $c_1 = 0$ we get the desired form of the transformation g, which concludes the proof. $\square$

The case when the ratio $\frac{\kappa}{\mu}$ is constant and μ large allows direct application of asymptotic expansions in order to obtain the approximations of characteristics of the transformed random variable, similarly to the Poisson case seen in the previous section, or a binomial case (see [1]). More details can be found in [1]. It is of more interest, however to consider μ large and κ fixed. The preceding method ceases to work in this case (see [1]), and hence, different strategy is needed.

We will consider the following two transformations, that we will study further

$$Y = 2 \cdot \sinh^{-1}\left[\sqrt{\frac{X+c}{\kappa+d}}\right] \tag{3.56}$$

and

$$Y = \ln(X + A). \tag{3.57}$$

where c and d are positive constants, and term A will be determined later.

The reason for introducing the transformation (3.56) is obvious, it comes as a generalisation of transformation (3.41) obtained via applying the Theorem 2.1. In what follows we will see that the linear term of the asymptotic series expansion of the transformation (3.56) for x large is in fact the transformation (3.57) for a specific choice of A. Hence, the transformation (3.57) might be viewed as an approximation of (3.56).

Lemma 3.17. *A transformation of random variable X given by* (3.56) *differs from a term*

$$2 \cdot \ln\left(\sqrt{x+c} + \sqrt{x+c+\kappa+d}\right) \tag{3.58}$$

by a constant with respect to x.

Proof. By the relation of inverse hyperbolic sine and natural logarithm (see [5]) we have

$$\sinh^{-1}(z) = \ln(z + \sqrt{1+z^2}), \tag{3.59}$$

where in our case $z = \sqrt{\frac{x+c}{\kappa+d}}$. By plugging in we obtain

$$\begin{aligned}\sinh^{-1}(z) &= 2 \cdot \ln\left(\sqrt{\frac{x+c}{\kappa+d}} + \sqrt{1+\frac{x+c}{\kappa+d}}\right) = 2 \cdot \ln\left(\frac{\sqrt{x+c}}{\sqrt{\kappa+d}} + \frac{\sqrt{x+c+\kappa+d}}{\sqrt{\kappa+d}}\right)\\ &= 2 \cdot \ln\left(\sqrt{x+c} + \sqrt{x+c+\kappa+d}\right) - 2 \cdot \ln\left(\sqrt{\kappa+d}\right),\end{aligned}$$

where $r_1 = -2 \cdot \ln\left(\sqrt{\kappa+d}\right)$ is the constant with respect to x. □

For x large enough we can obtain the following result.

Lemma 3.18. *Let Y be the random variable given by transformations* (3.56) *or* (3.57). *Assume that $X \longrightarrow \infty$. Then we have the following approximation*

$$Y = r_1 + r_2 + \ln(X) + \frac{A}{X} - \frac{B^2}{2X^2} + O\left(\frac{1}{X^3}\right). \tag{3.60}$$

where for (3.56) *we have $r_1 = -2\ln(\sqrt{\kappa+d})$ (see Lemma 3.17), $r_2 = 2\ln(2)$ and*

$$A = \frac{1}{2}\left(2c + \kappa + d\right), \tag{3.61}$$

$$B^2 = \frac{1}{8}\left(8c^2 + 8c(\kappa+d) + 3(\kappa+d)^2\right). \tag{3.62}$$

and for (3.57) *$r_1 = r_2 = 0$ and*

$$B = A. \tag{3.63}$$

Proof. Let us consider the transformation given by (3.56), due to Lemma 3.17 we have

$$
\begin{aligned}
y &= 2 \cdot \ln\left(\sqrt{x+c} + \sqrt{x+c+\kappa+d}\right) + r_1 = \ln\left(\left(\sqrt{x+c} + \sqrt{x+c+\kappa+d}\right)^2\right) + r_1 \\
&= \ln\left(2x + 2c + \kappa + d + 2 \cdot \sqrt{x+c} \cdot \sqrt{x+c+\kappa+d}\right) + r_1.
\end{aligned} \tag{3.64}
$$

Now given the $x \longrightarrow \infty$ we can derive the asymptotic series

$$
\sqrt{x+c+\kappa+d} = \sqrt{x+c} + \frac{1}{2} \cdot \frac{\kappa+d}{(x+c)^{\frac{1}{2}}} - \frac{1}{8} \cdot \frac{(\kappa+d)^2}{(x+c)^{\frac{3}{2}}} + O\left(\frac{1}{(x+c)^{\frac{5}{2}}}\right)
$$

and by plugging it in (3.64) we obtain

$$
\begin{aligned}
y &= \ln\Bigg(2x + 2c + \kappa + d + 2 \cdot \sqrt{x+c} \cdot \Bigg[\sqrt{x+c} + \frac{1}{2} \cdot \frac{\kappa+d}{(x+c)^{\frac{1}{2}}} - \frac{1}{8} \cdot \frac{(\kappa+d)^2}{(x+c)^{\frac{3}{2}}} \\
&\quad + O\left(\frac{1}{(x+c)^{\frac{5}{2}}}\right)\Bigg]\Bigg) + r_1 \\
&= \ln\left(4x + 4c + 2\kappa + 2d - \frac{1}{4} \cdot \frac{(\kappa+d)^2}{(x+c)} + O\left(\frac{1}{(x+c)^2}\right)\right) + r_1 \\
&= \ln\left(4x + 4c + 2\kappa + 2d - \frac{1}{4} \cdot \frac{(\kappa+d)^2}{x+c} + O\left(\frac{1}{(x+c)^2}\right)\right) + r_1.
\end{aligned}
$$

Now we substitute the whole logarithm by its asymptotic expansion for $x \longrightarrow \infty$ to obtain

$$
\begin{aligned}
y &= 2 \cdot \ln(2) + \ln(x) + \frac{1}{2}\frac{2c+\kappa+d}{x} + \frac{-\frac{1}{16}(\kappa+d)^2 + \frac{1}{2}(-c - \frac{1}{2}\kappa - \frac{1}{2}d)(c + \frac{1}{2}\kappa + \frac{1}{2}d)}{x^2} \\
&\quad + O\left(\frac{1}{x^3}\right) + r_1 \\
&= 2 \cdot \ln(2) + \ln(x) + \frac{1}{2}\frac{2c+\kappa+d}{x} + \frac{-\frac{1}{16}(\kappa+d)^2 - \frac{1}{16}(8c^2 + 8(\kappa+d) + 2(\kappa+d)^2)}{x^2} \\
&\quad + O\left(\frac{1}{x^3}\right) + r_1 \\
&= 2 \cdot \ln(2) + \ln(x) + \frac{1}{2x} \cdot (2c+\kappa+d) - \frac{1}{16x^2} \cdot (8c^2 + 8(\kappa+d) + 3(\kappa+d)^2) + O\left(\frac{1}{x^3}\right) \\
&\quad + r_1.
\end{aligned}
$$

Let us denote $r_2 = 2 \cdot \ln(2)$. We obtained the approximation (3.60) for (3.56). The second part of the statement concerning the transformation (3.57) is obtained immediately by taking the asymptotic expansion of $y = \ln(x + A)$ for $x \longrightarrow \infty$. □

Definition 3.19. Let us introduce the following notation. By Y^* we denote the random variable, obtained from (3.60) as follows

$$
Y^* = Y - r_1 - r_2 = \ln(X) + \frac{A}{X} - \frac{B^2}{2X^2} + O\left(\frac{1}{X^3}\right). \tag{3.65}
$$

where for (3.56) the constants A and B are given by (3.61) and (3.62), and for (3.57) by (3.63).

We are interested in finding variance approximations of the transformations (3.56) and (3.57). Since the random variable Y^* differs from Y given by (3.60) only by added constant, we have $\mathbf{var}[Y] = \mathbf{var}[Y^*]$, and therefore we may in order to simplify our computations continue with Y^* instead. Now we proceed to find an asymptotic expansion of the moment generating function of the approximation Y^* given by Definition 3.19 as $\mu \longrightarrow \infty$ with κ fixed. The moment generating function of Y^* is given by

$$M^*(t) = \sum_{x=0}^{\infty} e^{y^*(x)t} p(x). \tag{3.66}$$

(see [1]), i.e. if we plug in (1.32) for $p(x)$ we have

$$M^*(t) = \sum_{x=0}^{\infty} e^{y^* t} \frac{\Gamma(x+\kappa)}{x!\Gamma(\kappa)} \left(\frac{\mu}{\kappa+\mu}\right)^x \left(\frac{\kappa}{\kappa+\mu}\right)^\kappa \tag{3.67}$$

Let us introduce a new parameter α.

Proposition 3.20. *Let*

$$\alpha = \ln\left(\frac{\mu+\kappa}{\mu}\right), \tag{3.68}$$

then $\alpha \longrightarrow 0$ as $\mu \longrightarrow \infty$.

Proof. This is seen immediately by taking

$$\lim_{\mu \longrightarrow \infty} \ln\left(\frac{\mu+\kappa}{\mu}\right) = 0. \tag{3.69}$$

□

Definition 3.21. Let us denote $u(\alpha, x, t, \kappa)$ as

$$u(\alpha, x, t, \kappa) = e^{y^* t} \frac{\Gamma(x+\kappa)}{x!\Gamma(\kappa)} e^{-\alpha x}. \tag{3.70}$$

Informally put, the following Lemma will allow us to approximate an infinite sum in the moment generating function $M(t)$ by an integral.

Lemma 3.22. *As $\alpha \longrightarrow 0$*

$$\sum_{x=0}^{\infty} u(\alpha, x, t, \kappa) - \int_0^{\infty} u(\alpha, x, t, \kappa) dx \tag{3.71}$$

tends to a finite limit (depending on κ and t, and on which function y^* of x is chosen, namely (3.56) or (3.57)).

Proof. The proof relies on the use of Euler-Maclaurin expansion, a theory not developed in this work, for the complete proof the reader is kindly advised to see [1]. □

Corollary 3.23. *Let moment generating function be given by (3.66), let $\alpha \longrightarrow 0$, then following holds*

$$M^*(t) = (1 - e^{-\alpha})^\kappa \int_0^{\infty} u(\alpha, x, t, \kappa) dx + O(\alpha^\kappa) \tag{3.72}$$

Proof. Direct result of applying Lemma 3.22 to (3.66), see [1]. □

In general the integral in (3.72) can not be evaluated exactly. An approach is proposed in [1], based on expansion of $u_X(\alpha, x, t, \kappa)$ for x large asymptotically. The error of such expansion is always less than a multiple of the next term (independent of α) for $x \geq 1$ (see [1]). Integrating term by term between the limits 0 and ∞ gives then the following result.

Theorem 3.24. *Let t be confined to a neighbourhood of zero. $M(t)$ can be expanded asymptotically for $\alpha \longrightarrow 0$ in the form*

$$\begin{aligned} M^*(t) = \frac{\Gamma(\kappa+t)}{\alpha^t\Gamma(\kappa)}\Bigg\{&1 + \left(A - \frac{1}{2}\kappa\right)t\frac{\alpha}{\kappa+t-1} + \Bigg[\left(\frac{1}{2}\left(A-\frac{1}{2}\kappa\right)^2 + \frac{1}{24}\kappa\right)t^2 \\ &+ \left(\frac{1}{2}\kappa A - \frac{1}{24}\kappa(\kappa+3) - \frac{1}{2}B^2\right)t\Bigg]\frac{\alpha^2}{(\kappa+t-1)(\kappa+t-2)} + \ldots\Bigg\} + O(\alpha^\kappa) \end{aligned} \tag{3.73}$$

The series in braces is continued as far as the term in α^n, where n is the greatest integer less than κ.

Proof. The proof is a result of the results (1.41), (3.66), and Lemmata 3.18, 3.21. For more detail the reader is kindly advised to see [1]. □

We will now derive the approximations of numerical characteristics of Y^*, as mentioned already, we are namely interested in the variance approximations. For this we will first need to derive the cumulant generating function.

Lemma 3.25. *Consider the moment generating function* (3.73)*, then the corresponding cumulant generating function of Y^* is given by*

$$\begin{aligned} K^*(t) = -t\cdot\ln(\alpha) + \ln\Gamma(\kappa+t) - \ln\Gamma(\kappa) + \Bigg\{&\left(A-\frac{1}{2}\kappa\right)t\cdot\frac{\alpha}{\kappa+t-1} \\ + \Bigg[&\frac{\left(\frac{1}{2}\left(A-\frac{1}{2}\kappa\right)^2 + \frac{1}{24}\kappa\right)\cdot t^2 + \left(\frac{1}{2}\kappa A - \frac{1}{24}\kappa(\kappa+3) - \frac{1}{2}B^2\right)}{(\kappa+t-1)(\kappa+t-2)} \\ &- \frac{1}{4}\frac{\left(2A-\kappa\right)\left(A-\frac{1}{2}\kappa\right)t^2}{(\kappa+t-1)^2}\Bigg]\cdot\alpha^2\Bigg\} + O(\alpha^\kappa) \end{aligned} \tag{3.74}$$

The series in braces is continued as far as the term in α^n, where n is the greatest integer less than κ.

Proof. By [1] the cumulant generating function is found by taking the logarithm of the moment generating function (3.73). By doing so we obtain

$$\begin{aligned} K^*(t) &= \ln\left\{\frac{\Gamma(\kappa+t)}{\alpha^t\Gamma(\kappa)}\Bigg\{1 + \left(A-\frac{1}{2}\kappa\right)t\frac{\alpha}{\kappa+t-1} + \Bigg[\left(\frac{1}{2}\left(A-\frac{1}{2}\kappa\right)^2 + \frac{1}{24}\kappa\right)t^2\right. \\ &\left.+ \left(\frac{1}{2}\kappa A - \frac{1}{24}\kappa(\kappa+3) - \frac{1}{2}B^2\right)t\Bigg]\frac{\alpha^2}{(\kappa+t-1)(\kappa+t-2)} + \ldots\Bigg\} + O(\alpha^\kappa)\right\} \\ &= \ln\left\{\frac{\Gamma(\kappa+t)}{\alpha^t\Gamma(\kappa)}\right\} + \ln\Bigg\{1 + \left(A-\frac{1}{2}\kappa\right)t\frac{\alpha}{\kappa+t-1} + \Bigg[\left(\frac{1}{2}\left(A-\frac{1}{2}\kappa\right)^2 + \frac{1}{24}\kappa\right)t^2 \\ &+ \left(\frac{1}{2}\kappa A - \frac{1}{24}\kappa(\kappa+3) - \frac{1}{2}B^2\right)t\Bigg]\frac{\alpha^2}{(\kappa+t-1)(\kappa+t-2)} + \ldots\Bigg\} + O(\alpha^\kappa)\Bigg\}. \end{aligned} \tag{3.75}$$

Now we will approximate the second logarithm by it's Taylor series in α around the point $\alpha = 0$. The complete form of the Taylor series used to approximate the second logarithm is featured in a maple worksheet on a CD attached as an appendix to this work (file CumulantGeneratingFunction.mw). We obtain

$$\begin{aligned} K^*(t) &= \ln\big(\Gamma(\kappa+t)\big) - t\cdot\ln(\alpha) - \ln\big(\Gamma(\kappa)\big) + \frac{(A-\frac{1}{2}\kappa)\cdot t}{\kappa+t-1}\cdot\alpha \\ &+ \left\{ \frac{\big(\frac{1}{2}(A-\frac{1}{2}\kappa)^2 + \frac{1}{24}\kappa\big)\cdot t^2 + \big(\frac{1}{2}\kappa A - \frac{1}{24}\kappa(\kappa+3) - \frac{1}{2}B^2\big)\cdot t}{(\kappa+t-1)(\kappa+t-2)} \right. \\ &- \left. \frac{1}{4}\frac{(2A-\kappa)\cdot(A-\frac{1}{2}\kappa)\cdot t^2}{(\kappa+t-1)^2} \right\}\cdot\alpha^2 + O(\alpha^\kappa), \end{aligned} \tag{3.76}$$

which concludes the proof. □

Now we will focus on finding the approximations of the variance of Y^* using the cumulant generating function in the form given by (3.74). We will separate different situations based on a value of parameter κ. Further on we will use the notation given by Definitions 1.1 and 1.2. Let us start by formalising following observation.

Lemma 3.26. *Let Y be given by Lemma 3.18, and Y^* be given by Definition 3.19, then for the expectation of Y^* we have*

$$\mathbf{E}[Y^*] = \mathbf{E}[Y] + r_1 + r_2, \tag{3.77}$$

where $r_1 = -2\ln(\sqrt{\kappa+d})$ and $r_2 = 2\ln(2)$ (see Lemmata 3.17, 3.18), and for any k-th central moment of Y^* and Y

$$\mu_{Y^*,k} = \mu_{Y,k} \tag{3.78}$$

Proof. By Definition 3.19 we have that

$$Y^* = Y - r_1 - r_2, \tag{3.79}$$

where r_1, r_2 are deterministic constants. Then the first statement of the Lemma comes directly as a result of the property of expectation

$$\mathbf{E}[a+X] = a + \mathbf{E}[X] \tag{3.80}$$

where X is an arbitrary random variable and a an arbitrary (deterministic) constant (see again [2], or [4]). The second statement is obtained as follows. Let

$$\mu_{Y^*,k} = \mathbf{E}\big[Y + r - \mathbf{E}[Y+r]\big]^k \tag{3.81}$$

be the k-th central moment of Y^*, where $r = r_1 + r_2$, then by property (3.80) we have

$$\mu_{Y^*,k} = \mathbf{E}\big[Y + r - \mathbf{E}[Y+r]\big]^k = \mathbf{E}\big[Y + r - \mathbf{E}[Y] - r\big]^k = \mathbf{E}\big[Y - \mathbf{E}[Y]\big]^k = \mu_{Y,k} \tag{3.82}$$

□

Remark 3.27. *The second result of Lemma 3.26 namely implies that the variances of Y and Y^* are the same.*

Lemma 3.28. *Let $\kappa > 1$. Let the cumulant function $K^*(t)$ be given by Lemma 3.25, then the variance of the random variable Y can be approximated by*

$$\mathbf{var}Y = \psi'(\kappa) + \frac{\kappa - 2A}{(\kappa-1)^2}\alpha + O(\alpha^\kappa). \tag{3.83}$$

Proof. From Lemma 3.25 we have for $\kappa > 1$ the following

$$K^*(t) = -t \cdot \ln(\alpha) + \ln\Gamma(\kappa+t) - \ln\Gamma(\kappa) + \left(A - \frac{1}{2}\kappa\right)t \cdot \frac{\alpha}{\kappa+t-1} + O(\alpha^\kappa). \tag{3.84}$$

The second cumulant is obtained by computing second derivative of cumulant generating function in $t = 0$ (see [3]). The second derivative of (3.84) with respect to t is given by

$$(K^*)''(t) = \psi'(\kappa+t) + (A - \frac{1}{2}\kappa)\frac{-2(\kappa-1)}{(\kappa+t-1)^3} + O(\alpha^\kappa). \tag{3.85}$$

By evaluating in $t = 0$ we obtain

$$k_2 = (K^*)''(0) = \psi'(\kappa) + \frac{\kappa-2A}{(\kappa-1)^2}\alpha + O(\alpha^\kappa), \tag{3.86}$$

where by k_2 we denote the second cumulant of Y^*. The second cumulant of a random variable is equal to its second central moment (see [3]) which altogether with the result of the Lemma 3.26 concludes the proof. □

With the results of Lemma 3.28 we are now able to find an optimal choice for the constant A as will be seen in the following theorem. The idea is, by the right choice of A, eliminate the term dependant on α that for $\alpha \longrightarrow 0$ converges to zero the slowest.

Theorem 3.29. *Let $\kappa > 1$. Let us denote $h(\alpha, A) = \mathbf{var}Y - \psi'(\kappa)$. Then $\min_{A\geq 0}\left[|h(\alpha, A)|\right]$, given that A is constant with respect to α, is attained for $A = \frac{1}{2}\kappa$.*

Proof. Since by assumption $\kappa > 1$, by Theorem 3.28 we have that

$$\mathbf{var}Y = \psi'(\kappa) + \frac{\kappa-2A}{(\kappa-1)^2}\alpha + O(\alpha^\kappa). \tag{3.87}$$

From here follows that

$$h(\alpha, A) = \frac{\kappa-2A}{(\kappa-1)^2}\alpha + O(\alpha^\kappa). \tag{3.88}$$

The term $|h(\alpha, A)|$ will be minimal for such a choice of A that will ensure that

$$\frac{\kappa-2A}{(\kappa-1)^2} = 0, \tag{3.89}$$

hence, $A = \frac{1}{2}\kappa$. □

Corollary 3.30. *Let $\kappa > 1$. Let $A = \frac{1}{2}\kappa$, then $d = -2c$.*

Proof. By Lemma 3.18 we have that

$$A = \frac{1}{2}\big(2c + \kappa + d\big). \tag{3.90}$$

By Theorem 3.29 we have that

$$A = \frac{1}{2}\kappa. \tag{3.91}$$

By subtracting (3.91) from (3.90) and some computation we obtain $d = -2c$. □

Let us for the completeness also derive the approximation of the expectation of Y for $\kappa > 1$.

Lemma 3.31. *Let $\kappa > 1$ and the cumulant function be given by Lemma 3.25, then the expectation of the random variable Y can be approximated by*

$$\boldsymbol{E}[Y] = r_1 + r_2 - \ln(\alpha) + \psi(\kappa) + O(\alpha^\kappa), \tag{3.92}$$

where r_1 and r_2 are the constants introduced in Lemma 3.17 and Definition 3.19 respectively.

Proof. From Lemma 3.25 we have for $\kappa > 1$ the following

$$K^*(t) = -t \cdot \ln(\alpha) + \ln\Gamma(\kappa+t) - \ln\Gamma(\kappa) + \left(A - \frac{1}{2}\kappa\right) t \cdot \frac{\alpha}{\kappa+t-1} + O(\alpha^\kappa). \tag{3.93}$$

The first cumulant is obtained by computing the first derivative of cumulant generating function in $t = 0$ (see [3]). The first derivative of (3.93) with respect to t is

$$K'(t) = -\ln(\alpha) + \psi(\kappa+t) + (A - \frac{1}{2}\kappa)\frac{\kappa-1}{(\kappa+t-1)^2} + O(\alpha^\kappa). \tag{3.94}$$

By evaluating (3.94) in $t = 0$ we obtain

$$k_1 = K'(0) = -\ln(\alpha) + \psi(\kappa) + \frac{A - \frac{1}{2}\kappa}{\kappa-1}\alpha + O(\alpha^\kappa). \tag{3.95}$$

Since $\kappa > 1$ we have by Theorem 3.29 $A = \frac{1}{2}\kappa$ and hence, the summand containing α in the first power is equal to zero. The fact that the first cumulant of a random variable is equal to the first moment (see [3]) altogether with the result introduced in Lemma 3.26 concludes the proof. □

We will now use a more restrictive assumption on the shape parameter κ, which will in turn allow us to find a better approximation of the numerical characteristics of the transformed random variable Y.

Lemma 3.32. *Let $\kappa > 2$. Assume that $A = \frac{1}{2}\kappa$. Let the cumulant function $K(t)$ be given by Theorem 3.25, then the variance of the random variable Y can be approximated by*

$$\mathbf{var}Y = \psi'(\kappa) + \frac{\kappa(\kappa-1)(\kappa-2) - (2\kappa-3)(5\kappa^2 - 3\kappa - 12B^2)}{12(\kappa-1)^2(\kappa-2)^2}\alpha^2 + O(\alpha^\kappa). \tag{3.96}$$

Proof. From Lemma 3.25 we have for $\kappa > 2$ the following

$$\begin{aligned} K^*(t) = {} & \ln\big(\Gamma(\kappa+t)\big) - t \cdot \ln(\alpha) - \ln\big(\Gamma(\kappa)\big) + \frac{(A - \frac{1}{2}\kappa)\cdot t}{\kappa+t-1} \cdot \alpha \\ & + \left\{ \frac{\left(\frac{1}{2}(A - \frac{1}{2}\kappa)^2 + \frac{1}{24}\kappa\right)\cdot t^2 + \left(\frac{1}{2}\kappa A - \frac{1}{24}\kappa(\kappa+3) - \frac{1}{2}B^2\right)\cdot t}{(\kappa+t-1)(\kappa+t-2)} \right. \\ & \left. - \frac{1}{4}\frac{(2A-\kappa)\cdot(A - \frac{1}{2}\kappa)\cdot t^2}{(\kappa+t-1)^2} \right\} \cdot \alpha^2 + O(\alpha^\kappa), \end{aligned} \tag{3.97}$$

The second cumulant is obtained by computing second derivative of cumulant generating function in $t = 0$ (see [3]). Using that $A = \frac{1}{2}\kappa$ the second derivative of (3.97) with respect to t is given by

$$
\begin{aligned}
(K^*)''(t) = \psi'(\kappa+t)\frac{1}{12}\Bigg[&\frac{(-12B^2+3\kappa^2)t^3+(-3\kappa^3+9\kappa^2-6\kappa)t^2}{(\kappa+t-1)^3(\kappa+t-2)^3} \\
&+\frac{(36B^2\kappa^2-15\kappa^4-108B^2\kappa+54\kappa^3+72B^2-57\kappa^2+18\kappa)t}{(\kappa+t-1)^3(\kappa+t-2)^3} \\
&+\frac{24B^2\kappa^3-9\kappa^5-108B^2\kappa^2+45\kappa^4+156B^2\kappa-79\kappa^3-72B^2+57\kappa^2-14\kappa}{(\kappa+t-1)^3(\kappa+t-2)^3}\Bigg]\alpha^2 \\
&+O(\alpha^\kappa)
\end{aligned}
\tag{3.98}
$$

By evaluating (3.98) for $t = 0$ and some further computation we obtain

$$
k_2 = (K^*)''(0) = \psi'(\kappa)+\frac{\kappa(\kappa-1)(\kappa-2)-(2\kappa-3)(5\kappa^2-3\kappa-12B^2)}{12(\kappa-1)^2(\kappa-2)^2}\alpha^2+O(\alpha^\kappa), \tag{3.99}
$$

where by k_2 we denote the second cumulant of Y^*. By [3] the second cumulant of a random variable is equal to its second central moment, which altogether with the result of Lemma 3.26 concludes the proof. □

With the results of Lemma 3.32 and Theorem 3.29 we are now able to find the optimal value of the constant c.

Theorem 3.33. *Let $\kappa > 2$. Let us denote $h(\alpha, c) = \mathbf{var}Y - \psi'(\kappa)$. Then $\min_{c\geq 0}\big[|h(\alpha, c)|\big]$, given that c is constant with respect to α, is attained in*

$$
c = -\frac{1}{6}\cdot\frac{\sqrt{6\kappa\big(6\kappa^3-27\kappa^2+41\kappa-21\big)}-6\kappa^2+9\kappa}{2\kappa-3}. \tag{3.100}
$$

Proof. Since by assumption $\kappa > 2$, by Lemma 3.32 we have that

$$
\mathbf{var}Y = \psi'(\kappa)+\frac{\kappa(\kappa-1)(\kappa-2)-(2\kappa-3)(5\kappa^2-3\kappa-12B^2)}{12(\kappa-1)^2(\kappa-2)^2}\alpha^2+O(\alpha^\kappa). \tag{3.101}
$$

From here by using (3.62) and Corollary 3.30 follows that

$$
\begin{aligned}
h(\alpha, c) = &\frac{\kappa(\kappa-1)(\kappa-2)-(2\kappa-3)(5\kappa^2-3\kappa-\frac{3}{2}[8c^2+8c(\kappa-2c)+3(\kappa-2c)^2])}{12(\kappa-1)^2(\kappa-2)^2}\alpha^2 \\
&+O(\alpha^\kappa).
\end{aligned}
\tag{3.102}
$$

The term $|h(\alpha, c)|$ will be minimal for such a choice of c that will ensure that

$$
\frac{\kappa(\kappa-1)(\kappa-2)-(2\kappa-3)(5\kappa^2-3\kappa-\frac{3}{2}[8c^2+8c(\kappa-2c)+3(\kappa-2c)^2])}{12(\kappa-1)^2(\kappa-2)^2}=0 \tag{3.103}
$$

This in general is a quadratic equation with respect to c, but since we assumed c to be a positive constant, we have to drop one of the solutions of (3.103), which leaves us with

$$
c = -\frac{1}{6}\cdot\frac{\sqrt{6\kappa\big(6\kappa^3-27\kappa^2+41\kappa-21\big)}-6\kappa^2+9\kappa}{2\kappa-3}, \tag{3.104}
$$

which concludes the proof. □

Corollary 3.34. *Let c be optimal in the sense of Theorem 3.33, and $\kappa \longrightarrow \infty$, then we have*

$$c = \frac{3}{8} + \frac{23}{192\kappa} + O\Big(\frac{1}{\kappa^2}\Big). \tag{3.105}$$

Proof. By Theorem 3.33 we have

$$c = -\frac{1}{6} \cdot \frac{\sqrt{6}\sqrt{\kappa\big(6\kappa^3 - 27\kappa^2 + 41\kappa - 21\big)} - 6\kappa^2 + 9\kappa}{2\kappa - 3}. \tag{3.106}$$

Since $\kappa \longrightarrow \infty$ we may approximate (3.106) by its asymptotic expansion up to term κ^{-2} and obtain

$$c = \frac{3}{8} + \frac{23}{192\kappa} + O\Big(\frac{1}{\kappa^2}\Big), \tag{3.107}$$

which concludes the proof. □

To complete the theory of approximations for $\kappa > 2$ let us find the approximation of the expectation of Y.

Lemma 3.35. *Let $\kappa > 2$ and let the cumulant function $K^*(t)$ be given by Lemma 3.25, then the expectation of the random variable Y can be approximated by*

$$\boldsymbol{E}(Y) = r_1 + r_2 - \ln(\alpha) + \psi(\kappa) + \frac{1}{24}\frac{\kappa}{2\kappa - 3}\alpha^2 + O(\alpha^\kappa). \tag{3.108}$$

Proof. From Lemma 3.25 we have for $\kappa > 2$ that the approximation of the cumulant generating function may be given by formula (3.97). The first cumulant is obtained by computing the first derivative of the cumulant generating function in $t = 0$ (see [3]). The first derivative of the approximation of the cumulant generating function $(K^*)'(t)$ is given by

$$(K^*)'(t) = -\ln(\alpha) + \psi(\kappa + t) + (A - \frac{1}{2}\kappa)\frac{\kappa - 1}{(\kappa + t - 1)^2} + G(t, \kappa, c, d) + O(\alpha^\kappa), \tag{3.109}$$

where $G(t, \kappa, c, d)$ represents the first derivative of the coefficient of α^2 with respect to t. Taking this derivative is tedious, yet not particularly technically interesting part of the proof and hence, the detailed form and derivation of the term $G(t, \kappa, c, d)$ is provided in the Maple Document ExpectationQuadraticTermApprox.mw included in the digital appendix of this work. Due to $\kappa > 2$ we may apply the results given by Theorems 3.29, 3.33 and Corollary 3.30, and plug the optimal values of the constants into (3.109), and by evaluating $t = 0$ we obtain

$$k_1 = -\ln(\alpha) + \psi(\kappa) + \frac{1}{24}\frac{\kappa}{2\kappa - 3}\alpha^2 + O(\alpha^\kappa). \tag{3.110}$$

The fact that the first cumulant of a random variable is equal to its first moment (see [3]) altogether with the result of the Lemma 3.26 concludes the proof. □

We have found optimal values of all the constants of the generalised transformations proposed in this section. We will end this section by discussing the behaviour of the shape characteristics of the transformed random variable when μ is large.

Theorem 3.36. *The limiting value of the skewness parameter γ_1 of the random variable Y obtained via transformation* (3.56) *or* (3.57) *for $\mu \longrightarrow \infty$ is*

$$\gamma_1 = \frac{\psi''(\kappa)}{[\psi'(\kappa)]^{\frac{3}{2}}}. \tag{3.111}$$

Proof. The skewness parameter is given by

$$\gamma_1 = \frac{\mu_{Y,3}}{[\mu_{Y,2}]^{\frac{3}{2}}}, \tag{3.112}$$

where $\mu_{Y,3}$ and $\mu_{Y,2}$ are the third and second central moments of Y. If we use the relationship between central moments and cumulants (see [3]) altogether with the second result of Lemma 3.26, we can rewrite (3.112) in the following way

$$\gamma_1 = \frac{k_3}{[k_2]^{\frac{3}{2}}}, \tag{3.113}$$

where k_3 and k_2 are the third and second cumulants of Y^*. We can obtain their asymptotic expansions by evaluating the third and the second derivative of the cumulant generating function asymptotic expansion given by (3.74) in $t = 0$. Let us for the sake of simplicity denote

$$G(t, \kappa) = \frac{(A - \frac{1}{2}\kappa)}{\kappa - 1 + t} \tag{3.114}$$

and

$$H(t, \kappa) = \frac{(\frac{1}{2}(A - \frac{1}{2}\kappa)^2 + \frac{1}{24}\kappa) \cdot t^2 + (\frac{1}{2}\kappa A - \frac{1}{24}\kappa(\kappa + 3) - \frac{1}{2}B^2)}{(\kappa + t - 1)(\kappa + t - 2)} - \frac{1}{4}\frac{(2A - \kappa)(A - \frac{1}{2}\kappa)t^2}{(\kappa + t - 1)^2}. \tag{3.115}$$

The second derivative is given by

$$(K^*)''(t) = \psi'(t + \kappa) + G''(t, \kappa)\alpha + H''(t, \kappa)\alpha^2 \tag{3.116}$$

and by evaluating it in $t = 0$ we obtain

$$k_2 = (K^*)''(0) = \psi'(\kappa) + G''(0, \kappa)\alpha + H''(0, \kappa)\alpha^2. \tag{3.117}$$

Similarly the third derivative

$$k_3 = (K^*)'''(0) = \psi''(\kappa) + G'''(0, \kappa)\alpha + H'''(0, \kappa)\alpha^2. \tag{3.118}$$

By plugging (3.116) and (3.118) into (3.113) we obtain

$$\gamma_1 = \frac{\psi''(\kappa) + G'''(0, \kappa)\alpha + H'''(0, \kappa)\alpha^2}{[\psi'(\kappa) + G''(0, \kappa)\alpha + H''(0, \kappa)\alpha^2]^{\frac{3}{2}}}. \tag{3.119}$$

Due to the reparametrisation introduced in Proposition 3.20 taking limit for $m \longrightarrow \infty$ is equivalent to the limit for $\alpha \longrightarrow 0$ and if we proceed to take this limit of (3.119) we obtain

$$\gamma_1 = \frac{\psi''(\kappa)}{[\psi'(\kappa)]^{\frac{3}{2}}}. \tag{3.120}$$

□

Theorem 3.37. *The limiting value of the kurtosis parameter* γ_2 *of the random variable* Y *obtained via transformation* (3.56) *or* (3.57) *for* $\mu \longrightarrow \infty$ *is*

$$\gamma_1 = \frac{\psi'''(\kappa)}{[\psi'(\kappa)]^2}. \tag{3.121}$$

Proof. The kurtosis parameter is given by

$$\gamma_2 = \frac{\mu_{Y,4}}{[\mu_{Y,2}]^2} - 3, \tag{3.122}$$

where $\mu_{Y,4}$ and $\mu_{Y,2}$ are the forth and second central moments of Y. If we use the relationship between central moments and cumulants (see [3]) altogether with the second result of the Lemma 3.26, we can rewrite (3.122) in the following way

$$\gamma_2 = \frac{k_4 + 3k_2^2}{k_2} - 3, \tag{3.123}$$

where k_4 and k_2 are the fourth and second cumulants of Y^* respectively and we can obtain their asymptotic expansions by evaluating the fourth and the second derivative of the cumulant generating function asymptotic expansion given by (3.74) in $t = 0$. As in Theorem 3.36 let us for the sake of simplicity denote

$$G(t, \kappa) = \frac{(A - \frac{1}{2}\kappa)}{\kappa - 1 + t} \tag{3.124}$$

and

$$H(t, \kappa) = \frac{(\frac{1}{2}(A - \frac{1}{2}\kappa)^2 + \frac{1}{24}\kappa) \cdot t^2 + (\frac{1}{2}\kappa A - \frac{1}{24}\kappa(\kappa + 3) - \frac{1}{2}B^2)}{(\kappa + t - 1)(\kappa + t - 2)} - \frac{1}{4}\frac{(2A - \kappa)(A - \frac{1}{2}\kappa)t^2}{(\kappa + t - 1)^2}. \tag{3.125}$$

The second derivative evaluated in $t = 0$ is given by

$$k_2 = (K^*)''(0) = \psi'(\kappa) + G''(0, \kappa)\alpha + H''(0, \kappa)\alpha^2. \tag{3.126}$$

Similarly for the fourth derivative we have

$$k_4 = (K^*)^{iv}(0) = \psi'''(\kappa) + G^{iv}(0, \kappa)\alpha + H^{iv}(0, \kappa)\alpha^2. \tag{3.127}$$

by plugging (3.126), (3.127) back into (3.123) we obtain

$$\gamma_2 = \frac{\psi'''(\kappa) + G^{iv}(0, \kappa)\alpha + H^{iv}(0, \kappa)\alpha^2 + 3[\psi'(\kappa) + G''(0, \kappa)\alpha + H''(0, \kappa)\alpha^2]^2}{[\psi'(\kappa) + G''(0, \kappa)\alpha + H''(0, \kappa)\alpha^2]^2} - 3 \tag{3.128}$$

Due to the reparametrisation introduced in Proposition 3.20 taking limit for $m \longrightarrow \infty$ is equivalent to the limit for $\alpha \longrightarrow 0$ and if we proceed to take this limit of (3.2) we obtain

$$\gamma_2 = \frac{\psi'''(\kappa)}{[\psi'(\kappa)]^2}. \tag{3.129}$$

□

Remark 3.38. *It is easy to see, that the limiting value of skewness for* $\mu \longrightarrow \infty$ *given by Theorem 3.36 goes to zero, as* κ *goes to infinity, therefore, for large values of* κ *the distribution of* Y *acts as the normal distribution.*

Chapter 4

Theoretical Background of Performance Comparison of Selected Transformations Within ANOVA Framework

In this chapter our goal will be to provide a theoretical tool that will be used to compare performance of the transformations studied in Chapter 3 and some other commonly used transformations, namely $\ln(X + 1)$. The comparison will be done within the One-Way Analysis of Variance Framework (see [2] or [8] for more detail about One-Way ANOVA).

We will now provide a brief description of the assumed model. Let us have p samples $\mathbf{Y}_1 = (Y_{11}, ..., Y_{1n}), ..., \mathbf{Y}_p = (Y_{p1}, ..., Y_{pn})$ of equal size n from independent distributions $\mathcal{L}_1(\theta_1), ..., \mathcal{L}_p(\theta_p)$ in sequence, such that $\mathbf{E}Y_{ij} = \theta_i$ for all $i = 1, ..., p$ and all $j = 1, ..., n$. The task is to test the hypothesis

$$H_0 : \theta_1 = ... = \theta_p \tag{4.1}$$

of equality of expectations among the p samples against the alternative

$$H_1 : \exists i, k \in \{1, ..., p\} \;\; i \neq k \;\; such \;\; that \;\; \theta_i \neq \theta_k \tag{4.2}$$

of inequality of expectations among the samples. The test will be based on the F statistics. The comparison of the transformations will be done by comparing the powers of the F test while testing the hypothesis of equality of expectations (4.1) of p random samples from either Poisson or negative Binomial probability distribution, on which the above mentioned transformations will be applied in order to meet the assumptions of One-Way ANOVA. Furthermore we will assume, that some of the transformations considered will not have the variance stabilising effect, and therefore in order to proceed with the comparison in Section 4.2 we will study the properties of the F statistic for the case of violated equality of variances assumption.

In Theorems and statements as well as in their respective proofs through the whole chapter matrices and vectors of different dimensions will be used. Where needed, we will differentiate between the same kinds of matrices of different type by stating their type as the lower index. E. g. $\mathbf{I}_{n\times n}$ will denote the identity matrix of a type $n \times n$, whereas $\mathbf{I}_{p\times p}$ will denote the identity matrix of a type $p \times p$. In such case the notation will be kept through the whole statement and the respective proof. In cases where no misunderstanding will be possible however, we will write the matrix along with its type only when introducing it (e. g. in the statement of a theorem) and afterwards (e. g. in the respective proof) we will, by abuse of notation, drop the index in order to make the text more compact and legible.

4.1 Theoretical Results for Power of F-Test

We will first provide a derivation of the power of the F-test, when the assumption on the equality of variances is satisfied. This section is based on results from [2], [8], results concerning matrix algebra can be found for example in [10]. Assume the classical One-way Analysis of Variance setting (for more details see [2], [8]). For each $i = 1, ..., p$ set $\mathbf{Y}_i = (Y_{i1}, ..., Y_{in})^T$ the random sample of a size n of $N(\theta_i, \sigma)$ and assume that the p random samples are mutually independent. By stacking the p samples above each other we obtain a random vector $\mathbf{Y} = (Y_{11}, ..., Y_{1n}, ..., Y_{p1}, ..., Y_{pn})^T$, with distribution

$$\mathbf{Y} \sim N_{np}(\boldsymbol{\theta} \otimes \mathbf{1}_n, \sigma^2 \mathbf{I}_{np\times np}), \tag{4.3}$$

where $\boldsymbol{\theta} = \boldsymbol{\theta}_p = (\theta_1, ..., \theta_p)^T$, and σ^2 is the variance. Let us furthermore denote

$$\overline{\mathbf{Y}}_i = \frac{1}{n}\sum_{j=1}^{n} Y_{ij} \;\; for \;\; i = 1, ..., p \tag{4.4}$$

the arithmetic mean of each sample, and by

$$\overline{\mathbf{Y}} = \frac{1}{np}\sum_{i=1}^{p}\sum_{j=1}^{n} Y_{ij} \;\; for \;\; i = 1, ..., p \tag{4.5}$$

the arithmetic mean of all the samples.

The test statistic derived from the likelihood ratio test statistic is of following form (see [2], [8])

$$F = \frac{p(n-1)}{p-1}\frac{\frac{1}{\sigma^2}\sum_{i=1}^{p} n(\overline{Y}_i - \overline{Y})^2}{\frac{1}{\sigma^2}\sum_{i=1}^{p}\sum_{j=1}^{n}(Y_{ij} - \overline{Y}_i)^2} = \frac{p(n-1)}{p-1}\frac{K_1}{K_2} \tag{4.6}$$

For computing the power of the test based on the statistic (4.6) it is necessary to know the distribution of F under the null hypothesis and the alternative. Using the matrix notation, statistics K_1, K_2 can be expressed as follows

$$K_1 = \frac{n}{\sigma^2}\mathbf{Y}_{np}^T\left(\frac{1}{n}\mathbf{I}_{p\times p}\otimes \mathbf{1}_n\right)\left(\mathbf{I}_{p\times p} - \frac{1}{p}\mathbf{1}_p\mathbf{1}_p^T\right)\left(\frac{1}{n}\mathbf{I}_{p\times p}\otimes \mathbf{1}_n^T\right)\mathbf{Y}_{np} = \mathbf{Y}_{np}\mathbf{M}_{np\times np,1}\mathbf{Y}_{np} \tag{4.7}$$

$$K_2 = \frac{1}{\sigma^2}\mathbf{Y}_{np}^T\left(\mathbf{I}_{np\times np} - \frac{1}{n}\mathbf{I}_{p\times p}\otimes(\mathbf{1}_n\mathbf{1}_n^T)\right)\mathbf{Y}_{np} = \mathbf{Y}_{np}^T\mathbf{M}_{np\times np,2}\mathbf{Y}_{np} \tag{4.8}$$

The first important result is given by the following lemma.

Lemma 4.1. *Quadratic forms K_1, K_2 given by* (4.7), (4.8) *are independent both under null hypothesis and alternative.*

Proof. It is enough to see that $\mathbf{M}_{np\times np,1}\sigma^2\mathbf{I}_{np\times np}\mathbf{M}_{np\times np,2} = \mathbf{0}_{np\times np}$ (see [2]), where by $\mathbf{0}$ we denote the matrix, whose entries are only zeros. Let us denote $P_0 = \mathbf{M}_{np\times np,1}\sigma^2\mathbf{I}_{np\times np}\mathbf{M}_{np\times np,2}$, then we have

$$
\begin{aligned}
P_0 = & \frac{n}{\sigma^2}\left(\frac{1}{n}\mathbf{I}_{p\times p}\otimes\mathbf{1}_n\right)\left(\mathbf{I}_{p\times p}-\frac{1}{p}\mathbf{1}_p\mathbf{1}_p^T\right)\left(\frac{1}{n}\mathbf{I}_{p\times p}\otimes\mathbf{1}_n^T\right)\\
& \sigma^2\mathbf{I}_{np\times np}\frac{1}{\sigma^2}\left(\mathbf{I}_{np\times np}-\frac{1}{n}\mathbf{I}_{p\times p}\otimes(\mathbf{1}_n\mathbf{1}_n^T)\right)\\
= & \frac{n}{\sigma^2}\left(\frac{1}{n}\mathbf{I}_{p\times p}\otimes\mathbf{1}_n\right)\left(\mathbf{I}_{p\times p}-\frac{1}{p}\mathbf{1}_n\mathbf{1}_n^T\right)\left(\frac{1}{n}\mathbf{I}_{p\times p}\otimes\mathbf{1}_n^T\right)\left(\mathbf{I}_{np\times np}-\frac{1}{n}\mathbf{I}_{p\times p}\otimes(\mathbf{1}_n\mathbf{1}_n^T)\right)\\
= & \frac{n}{\sigma^2}\left(\frac{1}{n}\mathbf{I}_{p\times p}\otimes\mathbf{1}_n\right)\left(\mathbf{I}_{p\times p}-\frac{1}{p}\mathbf{1}_n\mathbf{1}_n^T\right)\left(\frac{1}{n}\mathbf{I}_{p\times p}\otimes\mathbf{1}_n^T\right)\left((\mathbf{I}_{p\times p}\otimes\mathbf{I}_{n\times n})-\frac{1}{n}\mathbf{I}_{p\times p}\otimes(\mathbf{1}_n\mathbf{1}_n^T)\right)
\end{aligned}
\tag{4.9}
$$

Let us denote

$$
\begin{aligned}
P_1 &= \left(\frac{1}{n}\mathbf{I}_{p\times p}\otimes\mathbf{1}_n^T\right)\left((\mathbf{I}_{p\times p}\otimes\mathbf{I}_{n\times n})-\frac{1}{n}\mathbf{I}_{p\times p}\otimes(\mathbf{1}_n\mathbf{1}_n^T)\right)\\
&= \left(\frac{1}{n}\mathbf{I}_{p\times p}\otimes\mathbf{1}_n^T\right)(\mathbf{I}_{p\times p}\otimes\mathbf{I}_{n\times n})-\left(\frac{1}{n}\mathbf{I}_{p\times p}\otimes\mathbf{1}_n^T\right)\left(\frac{1}{n}\mathbf{I}_{p\times p}\otimes(\mathbf{1}_n\mathbf{1}_n^T)\right)\\
&= \left(\frac{1}{n}\mathbf{I}_{p\times p}\mathbf{I}_{p\times p}\right)\otimes(\mathbf{1}_n^T\mathbf{I}_{n\times n})-\left(\frac{1}{n}\mathbf{I}_{p\times p}\mathbf{I}_{p\times p}\right)\otimes(\frac{1}{n}\mathbf{1}_n^T\mathbf{1}_n\mathbf{1}_n^T)\\
&= \left(\frac{1}{n}\mathbf{I}_{p\times p}\right)\otimes(\mathbf{1}_n^T)-\left(\frac{1}{n}\mathbf{I}_{p\times p}\right)\otimes(\mathbf{1}_n^T)=\mathbf{0}_{np\times np},
\end{aligned}
\tag{4.10}
$$

and hence, also $P_0 = \mathbf{0}_{np\times np}$. □

4.1.1 The Distribution of Denominator

The following Lemma will be given for a quadratic form

$$K_2' = \mathbf{Y}_{np}^T\mathbf{M'}_{np\times np,2}\mathbf{Y}_{np} \tag{4.11}$$

with more general matrix $\mathbf{M'}_{np\times np,2} = \sigma^2\mathbf{M}_{np\times np,2}$. It is obvious, that the same results hold for K_2 with $\mathbf{M}_{np\times np,2}$.

Lemma 4.2. *Let K_2' be a quadratic form given by* (4.11)*, assume that under null hypothesis $\boldsymbol{Y}_{np}\sim N_{np}(\theta\mathbf{1}_{np},\sigma^2\mathbf{I}_{np\times np})$, and under alternative $\boldsymbol{Y}_{np}\sim N_{np}(\boldsymbol{\theta}\otimes\mathbf{1}_n,\sigma^2\mathbf{I}_{np\times np})$, where $\boldsymbol{\theta}=(\theta_1,...,\theta_p)^T$, then we have*

$$K_2' = (\boldsymbol{Y}_{np}-\boldsymbol{\theta}_p\otimes\boldsymbol{1}_n)^T\boldsymbol{M'}_{np\times np,2}(\boldsymbol{Y}_{np}-\boldsymbol{\theta}_p\otimes\boldsymbol{1}_n) = (\boldsymbol{Y}_{np}-\theta\boldsymbol{1}_{np})^T\boldsymbol{M'}_{np\times np,2}(\boldsymbol{Y}_{np}-\theta\boldsymbol{1}_{np}). \tag{4.12}$$

Proof. Let us first show that $(\mathbf{Y}_{np}-\theta\mathbf{1}_{np})^T\mathbf{M'}_{np\times np,2}(\mathbf{Y}_{np}-\theta\mathbf{1}_{np}) = K_2'$. We have that

$$
\begin{aligned}
K_{20}' &= \big(\mathbf{Y}_{np}-\theta\mathbf{1}_{np}\big)^T\mathbf{M'}_{np\times np,2}\big(\mathbf{Y}_{np}-\theta\mathbf{1}_{np}\big)\\
&= \big(\mathbf{Y}_{np}^T\mathbf{M'}_{np\times np,2}-\theta\mathbf{1}_{np}^T\mathbf{M'}_{np\times np,2}\big)\big(\mathbf{Y}_{np}-\theta\mathbf{1}_{np}\big)\\
&= \big(\mathbf{Y}_{np}^T\mathbf{M'}_{np\times np,2}\mathbf{Y}_{np}-\theta\mathbf{1}_{np}^T\mathbf{M'}_{np\times np,2}\mathbf{Y}_{np}-\mathbf{Y}_{np}^T\mathbf{M'}_{np\times np,2}\theta\mathbf{1}_{np}\\
&\quad+\theta\mathbf{1}_{np}^T\mathbf{M'}_{np\times np,2}\theta\mathbf{1}_{np}\big)
\end{aligned}
\tag{4.13}
$$

Observe that

$$\mathbf{1}_{np}^T \mathbf{M'}_{np\times np,2} = \mathbf{1}_{np}^T\left[\mathbf{I}_{np\times np} - \frac{1}{n}\mathbf{I}_{p\times p}\otimes(\mathbf{1}_n\mathbf{1}_n^T)\right] = \mathbf{1}_{np}^T - (\mathbf{1}_p^T\otimes\mathbf{1}_n^T)\left(\frac{1}{n}\mathbf{I}_{p\times p}\otimes(\mathbf{1}_n\mathbf{1}_n^T)\right)$$
$$= \mathbf{1}_{np}^T - (\mathbf{1}_p^T\mathbf{I}_{p\times p})\otimes\left(\frac{1}{n}\mathbf{1}_n^T\mathbf{1}_n\mathbf{1}_n^T\right) = \mathbf{1}_{np}^T - (\mathbf{1}_p^T)\otimes(\mathbf{1}_n^T) = \mathbf{0}_{np\times np} \qquad (4.14)$$

Consequently also $\mathbf{M'}_{np\times np,2}\mathbf{1}_{np} = (\mathbf{1}_{np}^T\mathbf{M'}_{np\times np,2})^T = 0$, where the first equation is due to symmetry of $\mathbf{M'}_{np\times np,2}$, and hence, we have $K'_{20} = \mathbf{Y}_{np}^T\mathbf{M'}_{np\times np,2}\mathbf{Y}_{np} = K'_2$.

Let us now see that $(\mathbf{Y}_{np} - \boldsymbol{\theta}_p\otimes\mathbf{1}_n)^T\mathbf{M'}_{np\times np,2}(\mathbf{Y}_{np} - \boldsymbol{\theta}_p\otimes\mathbf{1}_n) = K'_2$ Similarly as before we have

$$\begin{aligned}K'_{21} &= \big(\mathbf{Y}_{np} - \boldsymbol{\theta}_p\otimes\mathbf{1}_n\big)^T\mathbf{M'}_{np\times np,2}\big(\mathbf{Y}_{np} - \boldsymbol{\theta}_p\otimes\mathbf{1}_n\big)\\
&= \big(\mathbf{Y}_{np}^T\mathbf{M'}_{np\times np,2} - (\boldsymbol{\theta}_p\otimes\mathbf{1}_n)^T\mathbf{M'}_{np\times np,2}\big)\big(\mathbf{Y}_{np} - (\boldsymbol{\theta}_p\otimes\mathbf{1}_n)\big)\\
&= \big(\mathbf{Y}_{np}^T\mathbf{M'}_{np\times np,2}\mathbf{Y}_{np} - (\boldsymbol{\theta}_p\otimes\mathbf{1}_n)^T\mathbf{M'}_{np\times np,2}\mathbf{Y}_{np} - \mathbf{Y}_{np}^T\mathbf{M'}_{np\times np,2}(\boldsymbol{\theta}_p\otimes\mathbf{1}_n)\\
&+ (\boldsymbol{\theta}_p\otimes\mathbf{1}_n)^T\mathbf{M'}_{np\times np,2}\theta\mathbf{1}_{np}\big),\end{aligned} \qquad (4.15)$$

and we have that

$$\begin{aligned}(\boldsymbol{\theta}_p\otimes\mathbf{1}_n)_{np}^T\mathbf{M'}_{np\times np,2} &= (\boldsymbol{\theta}_p\otimes\mathbf{1}_n)^T\left[\mathbf{I}_{np\times np} - \frac{1}{n}\mathbf{I}_{p\times p}\otimes(\mathbf{1}_n\mathbf{1}_n^T)\right]\\
&= (\boldsymbol{\theta}_p^T\otimes\mathbf{1}_n^T) - (\boldsymbol{\theta}_p^T\otimes\mathbf{1}_n^T)\left(\frac{1}{n}\mathbf{I}_{p\times p}\otimes(\mathbf{1}_n\mathbf{1}_n^T)\right)\\
&= (\boldsymbol{\theta}_p^T\otimes\mathbf{1}_n^T) - (\boldsymbol{\theta}_p^T\mathbf{I}_{p\times p})\otimes\left(\frac{1}{n}\mathbf{1}_n^T\mathbf{1}_n\mathbf{1}_n^T\right)\\
&= (\boldsymbol{\theta}_p^T\otimes\mathbf{1}_n^T) - (\boldsymbol{\theta}_p^T)\otimes(\mathbf{1}_n^T) = 0\end{aligned} \qquad (4.16)$$

Again we have also that $\mathbf{M'}_{np\times np,2}(\boldsymbol{\theta}_p^T\otimes\mathbf{1}_n^T) = (\boldsymbol{\theta}_p\otimes\mathbf{1}_n)^T\mathbf{M'}_{np\times np,2})^T = 0$, where the first equation is due to symmetry of $\mathbf{M'}_{np\times np,2}$, and hence, we have $K'_{21} = \mathbf{Y}_{np}^T\mathbf{M'}_{np\times np,2}\mathbf{Y}_{np} = K'_2$, which concludes the proof. □

Definition 4.3. Let us define a matrix $\mathbf{H}$ of a type $p\times p$ as

$$\mathbf{H}_{p\times p} = \mathbf{I}_{p\times p} - \frac{1}{p}\mathbf{1}_p\mathbf{1}_p^T = \begin{pmatrix} 1-\frac{1}{p} & -\frac{1}{p} & \dots & -\frac{1}{p}\\ -\frac{1}{p} & 1-\frac{1}{p} & \dots & -\frac{1}{p}\\ \dots & \dots & \dots & \dots\\ -\frac{1}{p} & -\frac{1}{p} & \dots & 1-\frac{1}{p}\end{pmatrix} \qquad (4.17)$$

this matrix is called the Centering matrix of a type $p\times p$ (see [11]).

Lemma 4.4. *The matrix $\boldsymbol{H}_{p\times p}$ given by Definition* (4.3) *is positively semidefinite and idempotent, and rank*$(\boldsymbol{H}_{p\times p}) = p-1$.

Proof. We will first show, that matrix $\mathbf{H}_{p\times p}$ is idempotent. We have that

$$\begin{aligned}\mathbf{H}_{p\times p}^2 = \mathbf{H}_{p\times p}\mathbf{H}_{p\times p} &= \left(\mathbf{I}_{p\times p} - \frac{1}{p}\mathbf{1}_p\mathbf{1}_p^T\right)\left(\mathbf{I}_{p\times p} - \frac{1}{p}\mathbf{1}_p\mathbf{1}_p^T\right)\\
&= \mathbf{I}_{p\times p} - \frac{2}{p}\mathbf{1}_p\mathbf{1}_p^T + \frac{1}{p^2}\mathbf{1}_p\mathbf{1}_p^T\mathbf{1}_p\mathbf{1}_p^T = \mathbf{I}_{p\times p} - \frac{2}{p}\mathbf{1}_p\mathbf{1}_p^T + \frac{1}{p^2}p\mathbf{1}_p\mathbf{1}_p^T\\
&= \mathbf{I}_{p\times p} - \frac{1}{p}\mathbf{1}_p\mathbf{1}_p^T = \mathbf{H}_{p\times p}\end{aligned} \qquad (4.18)$$

Obviously by its definition the matrix $H_{p\times p}$ is symmetric. Due to Proposition 1.21 we have that a symmetric, idempotent matrix is positively semidefinite. It is also easy to see, that the trace of $H_{p\times p}$ is $p-1$. Due to Proposition 1.20 we have that $rank(H_{p\times p}) = Tr(H_{p\times p}) = p-1$. $\square$

Lemma 4.5. *The matrix* $\boldsymbol{M'}_{np\times np,2}$ *is nonzero, and idempotent, with trace* $p(n-1)$.

Proof. The fact that $\mathbf{M'}_{np\times np,2}$ is nonzero is obvious. Idempotency of $\mathbf{M'}_{np\times np,2}$ follows from the fact that

$$\begin{aligned}\mathbf{M'}_{np\times np,2} &= \mathbf{I}_{np\times np} - \frac{1}{n}\mathbf{I}_{p\times p}\otimes \mathbf{1}_n\mathbf{1}_n^T = \mathbf{I}_{p\times p}\otimes\mathbf{I}_{n\times n} - \frac{1}{n}\mathbf{I}_{p\times p}\otimes \mathbf{1}_n\mathbf{1}_n^T \\ &= \mathbf{I}_{p\times p}\otimes\left[\mathbf{I}_{n\times n} - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T\right] = \mathbf{I}_{p\times p}\otimes\mathbf{H}_{n\times n},\end{aligned} \tag{4.19}$$

where $\mathbf{H}_{n\times n}$ is a Centering matrix from Definition 4.3. Due to Lemma 4.4 we have, that $\mathbf{H}_{n\times n}$ is idempotent, $\mathbf{I}_{p\times p}$ is clearly idempotent as well, and it is easy to see that a Kronecker product of two idempotent matrices is again an idempotent matrix. Due to Proposition 1.20 we have that $rank(\mathbf{M'}_{np\times np,2}) = Tr(\mathbf{M'}_{np\times np,2}) = p(n-1)$. $\square$

With the previous Lemmata we can conclude about the distribution of the quadratic form K_2'. That result is collected in the following theorem.

Theorem 4.6. *The distribution of* K_2' *is* $\chi^2_{p(n-1)}$ *both under zero hypothesis and under the alternative.*

Proof. This comes as a direct result of Lemmata 4.2, and 4.5, and Proposition 1.38. $\square$

4.1.2 Distribution of the Numerator Under Alternative

Let us now examine the quadratic form K_1 of the numerator under the alternative. Let us denote

$$\mathbf{Z}_p = \left(\frac{1}{n}\mathbf{I}_{p\times p}\otimes\mathbf{1}_n^T\right)\mathbf{Y}_n = (\overline{Y}_1, ..., \overline{Y}_p) \tag{4.20}$$

the random vector $\mathbf{Z}_p$

Proposition 4.7. *The random vector* $\boldsymbol{Z}_p$ *has a distribution* $\boldsymbol{Z}_p \sim N_p\big(\boldsymbol{\theta}_p, \frac{\sigma^2}{n}\boldsymbol{I}_{p\times p}\big)$

Proof. This comes as a direct result of the distribution of the arithmetic mean (see [2]). $\square$

With the $\mathbf{Z}_p$ and $\mathbf{H}_{p\times p}$ notation we can express the quadratic form K_1 in the following way

$$K_1 = \frac{n}{\sigma^2}\mathbf{Z}_p^T\mathbf{H}_{p\times p}\mathbf{Z}_p. \tag{4.21}$$

Now let us introduce a new random vector $\mathbf{U}_p$ given by

$$\mathbf{U}_p = \frac{\sqrt{n}}{\sigma}\mathbf{H}_{p\times p}\mathbf{Z}_p. \tag{4.22}$$

Proposition 4.8. *The random vector* $\boldsymbol{U}_p$ *given by equation* (4.22) *has a probability distribution* $\boldsymbol{U}_p \sim N_p(\frac{\sqrt{n}}{\sigma}\boldsymbol{H}_{p\times p}\boldsymbol{\theta}_p, \boldsymbol{H}_{p\times p})$.

Proof. The proof relies on the linearity property of the expectation and the properties of the normal distribution (see [2], [4]). We have that

$$\mathbf{E}(\mathbf{U}) = \mathbf{E}(\frac{\sqrt{n}}{\sigma}\mathbf{HZ}) = \frac{\sqrt{n}}{\sigma}\mathbf{HE}(\mathbf{Z}) = \frac{\sqrt{n}}{\sigma}\mathbf{H}\boldsymbol{\theta}, \tag{4.23}$$

and for the variance we have

$$\mathbf{var}(\mathbf{U}) = \mathbf{var}(\frac{\sqrt{n}}{\sigma}\mathbf{HZ}) = \frac{\sqrt{n}}{\sigma}\mathbf{H}\mathbf{var}(\mathbf{Z})\frac{\sqrt{n}}{\sigma}\mathbf{H}^T = \frac{n}{\sigma^2}\mathbf{H}\frac{\sigma^2}{n}\mathbf{I}\mathbf{H}^T, \tag{4.24}$$

And due to Lemma 4.4 H is idempotent, and we have that

$$\mathbf{var}(\mathbf{U}) = \mathbf{H}. \tag{4.25}$$

□

Proposition 4.9. *Using the transformation given by equation* (4.22) *we can rewrite* K_1 *as follows*

$$K_1 = \boldsymbol{U}_p^T\boldsymbol{U}_p. \tag{4.26}$$

Proof. We have

$$\mathbf{U}^T\mathbf{U} = \frac{\sqrt{n}}{\sigma}(\mathbf{HZ})^T\frac{\sqrt{n}}{\sigma}(\mathbf{HZ}) = \frac{n}{\sigma^2}\mathbf{Z}^T\mathbf{H}^T\mathbf{HZ} = \frac{n}{\sigma^2}\mathbf{Z}^T\mathbf{HZ}, \tag{4.27}$$

where the last equality is due to the idempotency of $\mathbf{H}$. □

Lemma 4.10. *Let* $\boldsymbol{H}_{p\times p}$ *be the centering matrix given by* (4.17). *Then there exists its decomposition*

$$\boldsymbol{H}_{p\times p} = \boldsymbol{B}_{p\times(p-1)}\boldsymbol{B}_{p\times(p-1)}^T, \text{ such that } \boldsymbol{B}_{p\times(p-1)}^T\boldsymbol{B}_{p\times(p-1)} = \boldsymbol{I}_{(p-1)\times(p-1)} \tag{4.28}$$

Proof. Since $\mathbf{H}$ is positive semidefinite, with rank $rank(H) = Tr(H) = p-1$ (see Proposition 1.20), we have due to Proposition 1.15 that $\mathbf{H} = \mathbf{BB}^T$, where $\mathbf{B}$ is a $p\times(p-1)$ matrix and $rank(\mathbf{B}) = p-1$. Due to Lemma 4.4 $\mathbf{H}$ is also idempotent. Let $\mathbf{L}_{(p-1)\times p}$ be the left inverse of $\mathbf{B}$, and $\mathbf{P}_{p\times(p-1)}$ be the right inverse of $\mathbf{B}^T$. We have that

$$\mathbf{L}(\mathbf{BB}^T\mathbf{BB}^T)\mathbf{P} = \mathbf{L}(\mathbf{BB}^T)\mathbf{P} = (\mathbf{LB})(\mathbf{B}^T\mathbf{P}) = \mathbf{I}, \tag{4.29}$$

where the first equality is due to idempotency of $\mathbf{H} = \mathbf{BB}^T$. We also have that

$$\mathbf{L}(\mathbf{BB}^T\mathbf{BB}^T)\mathbf{P} = (\mathbf{LB})\mathbf{B}^T\mathbf{B}(\mathbf{B}^T\mathbf{P}) = \mathbf{IB}^T\mathbf{BI} = \mathbf{B}^T\mathbf{B}, \tag{4.30}$$

and therefore $\mathbf{B}^T\mathbf{B} = \mathbf{I}$. □

Lemma 4.11. *Let* $\boldsymbol{W}_{p-1}$ *be a random vector given by transformation* $\boldsymbol{W}_{p-1} = \boldsymbol{B}_{p\times(p-1)}^T\boldsymbol{U}_p$, *where* $\boldsymbol{B}_{p\times(p-1)}^T$ *is obtained via the decomposition of Centering matrix* $\boldsymbol{H}_{p\times p}$ *as given in Lemma 4.10. Then* $\boldsymbol{W}_{p-1} \sim N_{p-1}\big(\frac{\sqrt{n}}{\sigma}\boldsymbol{B}_{p\times(p-1)}^T\boldsymbol{\theta}_p, \boldsymbol{I}_{(p-1)\times(p-1)}\big)$, *and the quadratic form* K_1 *can be written as*

$$K_1 = \boldsymbol{W}^T\boldsymbol{B}^T\boldsymbol{B}\boldsymbol{W} = \boldsymbol{W}^T\boldsymbol{W}. \tag{4.31}$$

Proof. Let us first show, that the distribution of $\mathbf{W}_{p-1}$ is $N_{p-1}\big(\frac{\sqrt{n}}{\sigma}\mathbf{B}^T_{p\times(p-1)}\boldsymbol{\theta}_p, \mathbf{I}_{(p-1)\times(p-1)}\big)$ we will again make use of the results of linearity of expectation as an operator and properties of the normal distribution (see [2], [4]). We have that

$$\mathbf{E}(\mathbf{W}) = \mathbf{E}(\mathbf{B}^T\mathbf{U}) = \mathbf{B}^T\mathbf{E}(\mathbf{U}) = \frac{\sqrt{n}}{\sigma}\mathbf{B}^T\mathbf{H}\boldsymbol{\theta}_p, \tag{4.32}$$

where the last equality is due to Proposition 4.8. For variance of $\mathbf{W}$ we have

$$\mathbf{var}(\mathbf{W}) = \mathbf{var}(\mathbf{B^T U}) = \mathbf{B}^T\mathbf{var}(\mathbf{U})\mathbf{B} = \mathbf{B}^T\mathbf{H}\mathbf{B} = \mathbf{B}^T\mathbf{B}\mathbf{B}^T\mathbf{B} = \mathbf{I} \tag{4.33}$$

where $\mathbf{var}(\mathbf{U}) = \mathbf{H}$ is given by Proposition 4.8, and the last inequality is due to column orthogonality of $\mathbf{B}$, which is given by Lemma 4.10. Lastly we have $\mathbf{W} = \mathbf{B}^T\mathbf{U}$ and therefore also $\mathbf{U} = \mathbf{B}\mathbf{W}$, and hence, using Proposition 4.9,

$$K_1 = \mathbf{U}^T\mathbf{U} = (\mathbf{B}\mathbf{W})^T(\mathbf{B}\mathbf{W}) = \mathbf{W}^T\mathbf{B}^T\mathbf{B}\mathbf{W} = \mathbf{W}^T\mathbf{W}. \tag{4.34}$$

□

Theorem 4.12. *The distribution of K_1 under the alternative is $K_1 \sim \chi^2_{p-1,\delta}$ where parameter of noncentrality $\delta = \frac{n}{\sigma^2}\boldsymbol{\theta}^T\boldsymbol{B}\boldsymbol{B}^T\boldsymbol{\theta} = \frac{n}{\sigma^2}\boldsymbol{\theta}^T\boldsymbol{H}\boldsymbol{\theta} = \frac{n}{\sigma^2}\sum_{i=1}^p(\theta_i - \overline{\theta})^2$, where $\overline{\theta} = \frac{1}{p}\sum_{j=1}^p \theta_j$.*

Proof. This comes as a result of Lemma 4.11, and Proposition 1.39. □

Corollary 4.13. *The distribution of K_1 under the null hypothesis $\theta_1 = ... = \theta_p$ is central χ^2 with $p-1$ degrees of freedom*

Proof. Given the hypothesis $\theta_1 = ... = \theta_p$, we have that the parameter of noncentrality $\delta = \frac{n}{\sigma^2}\boldsymbol{\theta}^T\mathbf{B}\mathbf{B}^T\boldsymbol{\theta} = \frac{n}{\sigma^2}\boldsymbol{\theta}^T\mathbf{H}\boldsymbol{\theta} = \frac{n}{\sigma^2}\sum_{i=1}^p(\theta_i - \overline{\theta})^2 = 0$ □

4.1.3 Power of F-Test

With the results regarding the distribution and independence of numerator and denominator of the F statistic under null hypothesis and alternative, we are able to provide a result on the distribution of the F statistics.

Theorem 4.14. *The distribution of the F statistic* (4.6) *under the null hypothesis is $F \sim F_{p-1,p(n-1)}$.*

Proof. Due to Corollary 4.13 we have that distribution of K_1 under the null hypothesis is χ^2_{p-1}, and due to Theorem 4.6 the distribution of K_2 is both under the null hypothesis and the alternative $\chi^2_{p(n-1)}$. We have seen in Lemma 4.1 that K_1, and K_2 are independent, and therefore we have $F \sim F_{p-1,p(n-1)}$ (see [8]). □

Theorem 4.15. *The distribution of the F statistic* (4.6) *under the alternative is $F \sim F_{p-1,p(n-1),\delta}$, with the parameter of noncentrality $\delta = \frac{n}{\sigma^2}\sum_{i=1}^p(\theta_i - \overline{\theta})^2$, where $\overline{\theta} = \frac{1}{p}\sum_{j=1}^p \theta_j$.*

Proof. Due to Theorem 4.12 the distribution of K_1 under the alternative is noncentral $\chi^2_{p-1,\delta}$. Due to Theorem 4.6 the distribution of K_2 is both under the null hypothesis and the alternative $\chi^2_{p(n-1)}$. We have seen in Lemma 4.1 that K_1, and K_2 are independent, and hence, F statistic has the noncentral distribution $F_{p-1,p(n-1),\delta}$ (see [8]). □

Now we will provide the formula for the power of the F-test. The power of a test at the significance level α is the conditional probability of rejecting the null hypothesis, given the condition that the alternative holds.

Definition 4.16. Let the test statistic be given by (4.6). We define the power of the F-test at the significance level α as

$$\beta_\alpha(\boldsymbol{\theta}) = \mathbf{P}\{F > Q_F(1-\alpha; p-1, p(n-1))|\boldsymbol{\theta}\}, \tag{4.35}$$

where by $Q_F(1-\alpha; p-1, p(n-1))$ we denote the $1-\alpha$ quantile of F distribution with degrees of freedom $p-1$, $p(n-1)$.

Proposition 4.17. *We may write the power of the F-test $\beta(\boldsymbol{\theta})$ as follows*

$$\beta_\alpha(\boldsymbol{\theta}) = 1 - \mathcal{F}_{F_\delta}(Q_F(1-\alpha; p-1, p(n-1)), p-1, n(p-1), \delta) \tag{4.36}$$

where $\mathcal{F}_{F_\delta}$ is the distribution function of noncentral F distribution $F_{p-1,p(n-1),\delta}$ from the Theorem 4.15.

Proof. If $\mathbf{P}(A) > 0$, then clearly $\mathbf{P}(\cdot|A)$ is also a probability measure on $(\Omega, \mathcal{A})$ (see [4]), hence, it has all the properties of probability measure, namely $\mathbf{P}(\overline{B}|A) = 1 - \mathbf{P}(B|A)$, where $\overline{B}$ is the opposite event to B. Hence, we have

$$\begin{aligned}\beta_\alpha(\boldsymbol{\theta}) &= \mathbf{P}\{F > Q_F(1-\alpha; p-1, p(n-1))|\boldsymbol{\theta}\} &(4.37)\\ &= 1 - \mathbf{P}\{F < Q_F(1-\alpha; p-1, p(n-1))|\boldsymbol{\theta}\}\\ &= 1 - \mathcal{F}_{F_\delta}(Q_F(1-\alpha; p-1, p(n-1)), p-1, p(n-1), \delta). &(4.38)\end{aligned}$$

□

4.2 Theoretical Results for Power of F-Test with Unequal Variances

In this section we will provide an approximation of the distribution of the F-test statistic in case that the assumption of the equality of variances is violated (see the beginning of the Chapter 4 for details about the One-Way Anova and the F test and [2] and [8], compare with Section 4.1). Some results of the matrix algebra not featured in Chapter 1 may be found in [10]. Assume that we have p independent random samples $\mathbf{Y}_1, ..., \mathbf{Y}_p$ of a size n of probability distribution $N(\mu_i, \sigma_i^2)$, where $i = 1, ..., p$. Set $\mathbf{Y}_i = (Y_{i1}, ..., Y_{ip})^T$ for $i = 1, ..., p$ By stacking the samples one above another we obtain a random vector $\mathbf{Y}_{np} = (Y_{11}, ..., Y_{1n}, ..., Y_{p1}, ..., Y_{pn})^T$ with probability distribution

$$\mathbf{Y} \sim N_{np}(\boldsymbol{\theta} \otimes \mathbf{1}_n, diag(\boldsymbol{\sigma^2}) \otimes \mathbf{I}_{n\times n}), \tag{4.39}$$

where $\boldsymbol{\theta} = \boldsymbol{\theta}_p = (\theta_1, ..., \theta_p)^T$, $\boldsymbol{\sigma^2} = \boldsymbol{\sigma^2_p} = (\sigma_1^2, ..., \sigma_p^2)$ the vector of variances.

The test statistic derived from the likelihood ratio test statistic is (see [2], [8])

$$F = \frac{p(n-1)}{p-1} \frac{\sum_{i=1}^p n(\overline{Y}_i - \overline{Y})^2}{\sum_{i=1}^p \sum_{j=1}^n (Y_{ij} - \overline{Y}_i)^2} = \frac{p(n-1)}{p-1} \frac{K_1'}{K_2'}, \tag{4.40}$$

where the terms $\overline{Y}, \overline{Y}_i$ are given by formulas (4.4) and (4.5) applied to the current setting. For computing the power of the test based on the statistic (4.40) it is necessary to know the distribution of F under the null hypothesis and the alternative. Using the matrix notation, statistics K_1, K_2 can be expressed as follows

$$K_1' = n\mathbf{Y}_{np}^T \left(\frac{1}{n}\mathbf{I}_{p\times p} \otimes \mathbf{1}_n\right) \left(\mathbf{I}_{p\times p} - \frac{1}{p}\mathbf{1}_p\mathbf{1}_p^T\right) \left(\frac{1}{n}\mathbf{I}_{p\times p} \otimes \mathbf{1}_n^T\right) \mathbf{Y}_{np} = \mathbf{Y}_{np}\mathbf{M'}_{np\times np,1}\mathbf{Y}_{np} \tag{4.41}$$

$$K_2' = \mathbf{Y}_{np}^T \left(\mathbf{I}_{np\times np} - \frac{1}{n}\mathbf{I}_{p\times p} \otimes (\mathbf{1}_n\mathbf{1}_n^T)\right) \mathbf{Y}_{np} = \mathbf{Y}_{np}^T\mathbf{M'}_{np\times np,2}\mathbf{Y}_{np} \tag{4.42}$$

Lemma 4.18. *Quadratic forms K'_1, K'_2 given by* (4.41), (4.42) *are independent both under null hypothesis and alternative.*

Proof. As previously in Lemma 4.18 we need to show $\mathbf{M'}_{np\times np,1}\mathbf{var}(\mathbf{Y}_{np})\mathbf{M'}_{np\times np,2} = \mathbf{0}_{np\times np}$ (see [2]). Let us denote

$$P'_0 = \mathbf{M'}_{np\times np,1}\mathbf{var}(\mathbf{Y}_{np})\mathbf{M'}_{np\times np,2} = \mathbf{M'}_{np\times np,1}(diag(\boldsymbol{\sigma}^2)\otimes \mathbf{I}_{n\times n})\mathbf{M'}_{np\times np,2}, \tag{4.43}$$

then we have

$$\begin{aligned} P'_0 = n\left(\frac{1}{n}\mathbf{I}_{p\times p}\otimes \mathbf{1}_n\right)\left(\mathbf{I}_{p\times p}-\frac{1}{p}\mathbf{1}_p\mathbf{1}_p^T\right)\left(\frac{1}{n}\mathbf{I}_{p\times p}\otimes \mathbf{1}_n^T\right)& \\ (diag(\boldsymbol{\sigma}^2)\otimes \mathbf{I}_{n\times n})\left(\mathbf{I}_{np\times np}-\frac{1}{n}\mathbf{I}_{p\times p}\otimes(\mathbf{1}_n\mathbf{1}_n^T)\right)& \end{aligned} \tag{4.44}$$

Notice that the matrices $(diag(\boldsymbol{\sigma}^2)\otimes \mathbf{I}_{n\times n})$, and $\left(\mathbf{I}_{np\times np}-\frac{1}{n}\mathbf{I}_{p\times p}\otimes(\mathbf{1}_n\mathbf{1}_n^T)\right)$ commute, indeed we have

$$\begin{aligned} P'_1 &= (diag(\boldsymbol{\sigma}^2)\otimes \mathbf{I}_{n\times n})\left(\mathbf{I}_{np\times np}-\frac{1}{n}\mathbf{I}_{p\times p}\otimes(\mathbf{1}_n\mathbf{1}_n^T)\right)\\ &= (diag(\boldsymbol{\sigma}^2)\otimes \mathbf{I}_{n\times n})\mathbf{I}_{np\times np}-(diag(\boldsymbol{\sigma}^2)\otimes \mathbf{I}_{n\times n})\left(\frac{1}{n}\mathbf{I}_{p\times p}\otimes(\mathbf{1}_n\mathbf{1}_n^T)\right)\\ &= (diag(\boldsymbol{\sigma}^2)\otimes \mathbf{I}_{n\times n})-\left(diag(\boldsymbol{\sigma}^2)\frac{1}{n}\mathbf{I}_{p\times p}\right)\otimes(\mathbf{I}_{n\times n}\mathbf{1}_n\mathbf{1}_n^T)\\ &= (diag(\boldsymbol{\sigma}^2)\otimes \mathbf{I}_{n\times n})-\left(diag(\boldsymbol{\sigma}^2)\frac{1}{n}\right)\otimes(\mathbf{1}_n\mathbf{1}_n^T)\\ &= \mathbf{I}_{np\times np}(diag(\boldsymbol{\sigma}^2)\otimes \mathbf{I}_{n\times n})-\left(\frac{1}{n}\mathbf{I}_{p\times p}diag(\boldsymbol{\sigma}^2)\right)\otimes(\mathbf{1}_n\mathbf{1}_n^T\mathbf{I}_{n\times n})\\ &= \mathbf{I}_{np\times np}(diag(\boldsymbol{\sigma}^2)\otimes \mathbf{I}_{n\times n})-\left(\frac{1}{n}\mathbf{I}_{p\times p}\otimes(\mathbf{1}_n\mathbf{1}_n^T)\right)(diag(\boldsymbol{\sigma}^2)\otimes \mathbf{I}_{n\times n})\\ &= \left(\mathbf{I}_{np\times np}-\frac{1}{n}\mathbf{I}_{p\times p}\otimes(\mathbf{1}_n\mathbf{1}_n^T)\right)(diag(\boldsymbol{\sigma}^2)\otimes \mathbf{I}_{n\times n}), \end{aligned} \tag{4.45}$$

and hence, we can write

$$\begin{aligned} P'_0 = n\left(\frac{1}{n}\mathbf{I}_{p\times p}\otimes \mathbf{1}_n\right)\left(\mathbf{I}_{p\times p}-\frac{1}{p}\mathbf{1}_p\mathbf{1}_p^T\right)\left(\frac{1}{n}\mathbf{I}_{p\times p}\otimes \mathbf{1}_n^T\right)& \\ \left(\mathbf{I}_{np\times np}-\frac{1}{n}\mathbf{I}_{p\times p}\otimes(\mathbf{1}_n\mathbf{1}_n^T)\right)(diag(\boldsymbol{\sigma}^2)\otimes \mathbf{I}_{n\times n}).& \end{aligned} \tag{4.46}$$

We have seen in the proof of Lemma 4.1 that the term

$$\left(\frac{1}{n}\mathbf{I}_{p\times p}\otimes \mathbf{1}_n\right)\left(\mathbf{I}_{p\times p}-\frac{1}{p}\mathbf{1}_p\mathbf{1}_p^T\right)\left(\frac{1}{n}\mathbf{I}_{p\times p}\otimes \mathbf{1}_n^T\right)\left(\mathbf{I}_{np\times np}-\frac{1}{n}\mathbf{I}_{p\times p}\otimes(\mathbf{1}_n\mathbf{1}_n^T)\right)=\mathbf{0}, \tag{4.47}$$

and hence, $P'_0 = \mathbf{0}$. □

4.2.1 Study of Denominator

In the Section 4.1 we saw, that regardless of the assumption on variance of $\mathbf{Y}$ we have

$$\begin{aligned} K_2' &= \mathbf{Y}_{np}^T \mathbf{M'}_{np\times np,2} \mathbf{Y}_{np} \\ &= (\mathbf{Y}_{np} - \boldsymbol{\theta}_p \otimes \mathbf{1}_n)^T \mathbf{M'}_{np\times np,2} (\mathbf{Y}_{np} - \boldsymbol{\theta}_p \otimes \mathbf{1}_n) \\ &= (\mathbf{Y}_{np} - \theta \mathbf{1}_{np})^T \mathbf{M'}_{np\times np,2} (\mathbf{Y}_{np} - \theta \mathbf{1}_{np}). \end{aligned} \tag{4.48}$$

Therefore we may without the loss of generality in the following consider the expression

$$K_2' = (\mathbf{Y}_{np} - \boldsymbol{\theta}_p \otimes \mathbf{1}_n)^T \mathbf{M'}_{np\times np,2} (\mathbf{Y}_{np} - \boldsymbol{\theta}_p \otimes \mathbf{1}_n) \tag{4.49}$$

Proposition 4.19. *Let us denote $\boldsymbol{Y}_0 = (\boldsymbol{Y} - \boldsymbol{\theta}_p \otimes \boldsymbol{1}_n)$, then $\boldsymbol{Y}_0 \sim N_{np}(\boldsymbol{0}, diag(\boldsymbol{\sigma}_p^2) \otimes \boldsymbol{I}_{n\times n})$, and the quadratic form may be written as*

$$K_2' = \boldsymbol{Y}_0^T \boldsymbol{M'}_{np\times np,2} \boldsymbol{Y}_0. \tag{4.50}$$

Proof. Due to linearity of expectation and properties of the normal distribution (see [4]) we have that

$$\mathbf{E}(\mathbf{Y}_0) = \mathbf{E}(\mathbf{Y} - \boldsymbol{\theta}_p \otimes \mathbf{1}_n) = \mathbf{E}(\mathbf{Y}) - \boldsymbol{\theta}_p \otimes \mathbf{1}_n = \boldsymbol{\theta}_p \otimes \mathbf{1}_n - \boldsymbol{\theta}_p \otimes \mathbf{1}_n = \mathbf{0}. \tag{4.51}$$

And for the variance

$$\mathbf{var}(\mathbf{Y}_0) = \mathbf{var}(\mathbf{Y} - \boldsymbol{\theta}_p \otimes \mathbf{1}_n) = \mathbf{var}(\mathbf{Y}) = diag(\boldsymbol{\sigma}_p^2) \otimes \mathbf{I}_{n\times n}. \tag{4.52}$$

The result (4.50) is obvious. □

Proposition 4.20. *Let $\boldsymbol{X}_0 \sim N_{np}(\boldsymbol{0}_{np}, \boldsymbol{I}_{np\times np})$, then there exists a matrix $\boldsymbol{T}_{np\times np}$ such that*

$$\boldsymbol{Y}_0 = \boldsymbol{T}\boldsymbol{X}_0. \tag{4.53}$$

Proof. We will provide a constructive proof by finding the matrix $\mathbf{T}$. In order for $\mathbf{T}$ to satisfy (4.53), the following must hold.

$$\mathbf{0} = \mathbf{E}(\mathbf{Y}_0) = \mathbf{E}(\mathbf{T}\mathbf{X}_0) = \mathbf{T}\mathbf{E}(\mathbf{X}_0) = \mathbf{T}\mathbf{0}, \tag{4.54}$$

and

$$diag(\boldsymbol{\sigma}^2) \otimes \mathbf{I} = \mathbf{var}(\mathbf{Y}_0) = \mathbf{var}(\mathbf{T}\mathbf{X}_0) = \mathbf{T}\mathbf{var}(\mathbf{X}_0)\mathbf{T}^T = \mathbf{T}\mathbf{I}\mathbf{T}^T. \tag{4.55}$$

Since the equation (4.54) is satisfied for any matrix $\mathbf{T}$, by choosing $\mathbf{T} = diag(\boldsymbol{\sigma}_p) \otimes \mathbf{I}_{n\times n}$, where $\boldsymbol{\sigma}_p = (\sigma_1, ..., \sigma_p)^T$ is the vector of standard deviations of $\mathbf{Y}$, also the equation (4.55) is satisfied, which concludes the proof. □

Lemma 4.21. *The quadratic form K_2' may be written in a form*

$$K_2' = \boldsymbol{X}_0^T \boldsymbol{N}_{np\times np} \boldsymbol{X}_0, \tag{4.56}$$

where

$$\boldsymbol{N}_{np\times np} = diag(\boldsymbol{\sigma}^2) \otimes \left[\boldsymbol{I}_{n\times n} - \frac{1}{n} \boldsymbol{1}_n \boldsymbol{1}_n^T \right]. \tag{4.57}$$

Proof. Let us first see, that (4.57) holds. We have

$$\begin{aligned}
\mathbf{N} &= \mathbf{T}^T \mathbf{M'}_{np\times np,2}\mathbf{T} \\
&= (diag(\boldsymbol{\sigma}_p)\otimes \mathbf{I}_{n\times n})^T \left[\mathbf{I}_{np\times np} - \frac{1}{n}\mathbf{I}_{p\times p}\otimes(\mathbf{1}_n\mathbf{1}_n^T)\right](diag(\boldsymbol{\sigma}_p)\otimes \mathbf{I}_{n\times n}) \\
&= (diag(\boldsymbol{\sigma}_p)\otimes \mathbf{I}_{n\times n})\left[(\mathbf{I}_{p\times p}\otimes \mathbf{I}_{n\times n}) - \frac{1}{n}\mathbf{I}_{p\times p}\otimes(\mathbf{1}_n\mathbf{1}_n^T)\right](diag(\boldsymbol{\sigma}_p)\otimes \mathbf{I}_{n\times n}) \\
&= (diag(\boldsymbol{\sigma}_p)\otimes \mathbf{I}_{n\times n}))\left(\mathbf{I}_{p\times p}\otimes\left[(\mathbf{I}_{n\times n}) - \frac{1}{n}(\mathbf{1}_n\mathbf{1}_n^T)\right]\right)(diag(\boldsymbol{\sigma}_p)\otimes \mathbf{I}_{n\times n}) \\
&= (diag(\boldsymbol{\sigma}_p)\mathbf{I}_{p\times p})\otimes\left(\mathbf{I}_{n\times n}\left[(\mathbf{I}_{n\times n}) - \frac{1}{n}(\mathbf{1}_n\mathbf{1}_n^T)\right]\right)(diag(\boldsymbol{\sigma}_p)\otimes \mathbf{I}_{n\times n}) \\
&= (diag(\boldsymbol{\sigma}_p)\mathbf{I}_{p\times p}diag(\boldsymbol{\sigma}_p))\otimes\left(\mathbf{I}_{n\times n}\left[(\mathbf{I}_{n\times n}) - \frac{1}{n}(\mathbf{1}_n\mathbf{1}_n^T)\right]\mathbf{I}_{n\times n}\right) \\
&= diag(\boldsymbol{\sigma}^2)\otimes\left[\mathbf{I}_{n\times n} - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T\right].
\end{aligned} \tag{4.58}$$

Seeing that (4.56) is satisfied is simple. Indeed, we have due to Proposition 4.19

$$K_2' = \mathbf{Y}_0^T\mathbf{M'}_{np\times np,2}\mathbf{Y}_0. \tag{4.59}$$

Since due to Proposition 4.20 $\mathbf{Y}_0 = \mathbf{T}\mathbf{X}_0$, by plugging this into (4.59) we obtain

$$K_2' = (\mathbf{T}\mathbf{X}_0)^T\mathbf{M'}_{np\times np,2}(\mathbf{T}\mathbf{X}_0) = \mathbf{X}_0^T\mathbf{T}^T\mathbf{M'}_{np\times np,2}\mathbf{T}\mathbf{X}_0 = \mathbf{X}_0^T\mathbf{N}\mathbf{X}_0. \tag{4.60}$$

□

Our task now is to find eigenvalues and eigenvectors of the matrix $\mathbf{N}$ so we can determine, what is the distribution of the quadratic form K_2'. Let us introduce the following notation. Let $\mathbf{N}_1 = diag(\boldsymbol{\sigma}^2)$, and $\mathbf{N}_2 = \left[\mathbf{I}_{n\times n} - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T\right]$.

Proposition 4.22. *The matrix $\mathbf{N}_2$ is symmetric, idempotent, and positive semidefinite.*

Proof. Observe that $\mathbf{N}_2 = \mathbf{H}_{n\times n}$ is in fact a Centering matrix of a type $n\times n$ as defined in Section 4.1, Definition 4.3. The proof of symmetry, idempotency, and positive definitness of a Centering matrix is given by Lemma 4.4. □

Lemma 4.23. *$\mathbf{N}_2$ has $n-1$ eigenvalues equal to one, and one eigenvalue equal to zero.*

Proof. In Proposition 4.22 we have seen that N_2 is idempotent. Furthermore we have by Proposition 1.16 that

$$\sum_{i=1}^{n}\lambda_i = Tr(\mathbf{N}_2) = n-1, \tag{4.61}$$

and since due to Proposition 1.19 we have that eigenvalues of Idempotent matrix are either ones or zeroes, we see that $n-1$ eigenvalues have to be ones and exactly one eigenvalue is zero. □

Lemma 4.24. *Eigenvalues of matrix $\mathbf{N}_1$ are $\sigma_1^2,...,\sigma_p^2$.*

Proof. This follows trivially from the form of characteristic polynomial of a diagonal matrix. □

Lemma 4.25. *Eigenvalues of the matrix* $\boldsymbol{N}$ *with their multiplicities are* $\sigma_1^2, ..., \sigma_p^2$ *with multiplicities* $n-1$*, and* 0 *with multiplicity* p*.*

Proof. Due to Proposition 1.17, and Lemma 4.24 we have that the eigenvalues of a Kronecker product of matrices $\mathbf{N}_1, \mathbf{N}_2$ are

$$\sigma_1^2\mu_1, ..., \sigma_1^2\mu_m, \sigma_2^2\mu_1, ..., \sigma_2^2\mu_m, ..., \sigma_n^2\mu_m, \tag{4.62}$$

where σ_i^2 are the eigenvalues of $\mathbf{N}_1$ and μ_i are the eigenvalues of $\mathbf{N}_2$. Since by Lemma 4.23 exactly one eigenvalue of $\mathbf{N}_2$ is $\mu_j = 0$, and the rest are ones, exactly p terms of (4.62) are zeroes and $n-1$ are σ_i^2 for $i = 1, ..., p$. □

Theorem 4.26. *There exists a decomposition of* $\boldsymbol{N}$ *such that*

$$\boldsymbol{N} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^T, \text{ and } \boldsymbol{P}\boldsymbol{P}^T = \boldsymbol{I}, \tag{4.63}$$

where $\boldsymbol{P}$ *is the column orthonormal matrix of eigenvectors of* $\boldsymbol{N}$*. Consequently the quadratic form* K_2' *can be written in the form*

$$K_2' = \boldsymbol{X}_0^T\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^T\boldsymbol{X}_0 = \boldsymbol{V}^T\boldsymbol{\Lambda}\,\boldsymbol{V}, \tag{4.64}$$

where $\boldsymbol{V}$ *is a Gaussian random vector given by* $\boldsymbol{V} = \boldsymbol{P}^T\boldsymbol{X}_0$*, satisfying* $\boldsymbol{E}(\boldsymbol{V}) = \boldsymbol{0}$*, and* $\boldsymbol{var}(\boldsymbol{V}) = \boldsymbol{I}$*.*

Proof. Since $\mathbf{N}$ is clearly a symmetric real matrix the decomposition (4.63) is given by Proposition 1.12. The equation (4.64) is obtained merely by plugging (4.63) into (4.56). $\mathbf{V}$ is clearly Gaussian, since it is obtained as a linear transformation of Gaussian random vector $\mathbf{X}_0$ (see [2]), as for the characteristics we have

$$\mathbf{E}(\mathbf{V}) = \mathbf{E}(\mathbf{P}^T\mathbf{X}_0) = \mathbf{P}^T\mathbf{E}(\mathbf{X}_0) = \mathbf{0}, \tag{4.65}$$

since $\mathbf{E}(\mathbf{X}_0) = \mathbf{0}$, and

$$\mathbf{var}(\mathbf{V}) = \mathbf{var}(\mathbf{P}^T\mathbf{X}_0) = \mathbf{P}^T\mathbf{var}(\mathbf{X}_0)\mathbf{P} = \mathbf{P}^T\mathbf{I}(\mathbf{X}_0)\mathbf{P} = \mathbf{I}, \tag{4.66}$$

where the last equality is due to matrix $\mathbf{P}$ being column orthonormal matrix. □

Corollary 4.27. *Quadratic form* K_2' *can be expressed as*

$$\sum_{i=1}^{p} \sigma_i^2 X_i, \tag{4.67}$$

a linear combination of p *independent identically distributed random variables* $X_i \sim \chi^2_{n-1}$ $i = 1, ..., p$*.*

Proof. Due to Theorem 4.26 we have that $K_2' = \mathbf{V}^T\mathbf{\Lambda}\mathbf{V}$, evaluating the product we obtain that

$$K_2' = \sum_{i=1}^{np} \lambda_i V_i^2, \tag{4.68}$$

where λ_i are the eigenvalues of $\mathbf{N}$ and V_i are the entries of the vector $\mathbf{V} \sim N(\mathbf{0}, \mathbf{I})$. Due to Lemma 4.25 we have that p eigenvalues of $\mathbf{N}$ are zeroes, hence, p terms in the sum of (4.68) are zeroes. Furthermore also due to 4.25 we have that the rest of $p(n-1)$ eigenvalues are $\sigma_1^2, ..., \sigma_p^2$, each with multiplicity $n-1$, hence, we have that

$$\begin{aligned} K_2' &= \sum_{i=1}^{p(n-1)} \lambda_i V_i^2 = \sigma_1^2 V_1^2 + ... + \sigma_1^2 V_{n-1}^2 + \sigma_2^2 V_n^2 + ... + \sigma_2^2 V_{2n-2}^2 + ... + \sigma_p^2 V_{p(n-1)}^2 \\ &= \sigma_1^2 (V_1^2 + ... + V_{n-1}^2) + ... + \sigma_p^2 (V_{p(n-1)-n+2}^2 + ... + V_{p(n-1)}^2). \end{aligned} \tag{4.69}$$

Clearly $V_i \sim N(0,1)$, and are mutually independent ($\mathbf{var}(\mathbf{V}) = \mathbf{I}$), and hence, due to Proposition 1.35 we have that $X_1 = (V_1^2 + ... + V_{n-1}^2) \sim \chi_{n-1}^2, ..., X_p = (V_{p(n-1)-n+2}^2 + ... + V_{p(n-1)}^2) \sim \chi_{n-1}^2$, are p independent identically distributed random variables, and we may write

$$K_2' = \sum_{i=1}^{p} \sigma_i^2 X_i. \tag{4.70}$$

□

4.2.2 Study of Numerator

Let us recall the random vector

$$\mathbf{Z}_p = (\frac{1}{n} \mathbf{I}_{p \times p} \otimes \mathbf{1}_n^T) \mathbf{Y} = (\overline{Y}_1, ..., \overline{Y}_n), \tag{4.71}$$

that we introduced already in Section 4.1 (see (4.20)).

Proposition 4.28. *The distribution of $\boldsymbol{Z}_p$ is $N_p(\boldsymbol{\theta}, \frac{1}{n} diag(\boldsymbol{\sigma}^2))$,*

Proof. This comes directly as a result of distribution of arithmetic mean (see [2]). □

Let us recall that in Section 4.1 we have defined the Centering matrix $\mathbf{H}_{p \times p}$ (see Definition 4.3). With the notion of $\mathbf{Z}$, and the Centering matrix $\mathbf{H}$ we can rewrite K_1' in the following form

$$K_1' = n \mathbf{Z}_p^T \mathbf{H}_{p \times p} \mathbf{Z}_p. \tag{4.72}$$

Now let us introduce a new random vector $\mathbf{U}_p$ by

$$\mathbf{U}_p = \sqrt{n} \mathbf{H}_{p \times p} \mathbf{Z}_p \tag{4.73}$$

Proposition 4.29. *The random vector $\boldsymbol{U}_p$ given by equation 4.73 has a probability distribution $\boldsymbol{U}_p \sim N_p(\sqrt{n} \boldsymbol{H}_{p \times p} \boldsymbol{\theta}, \boldsymbol{H}_{p \times p} diag(\boldsymbol{\sigma}_p^2) \boldsymbol{H}_{p \times p}$, and the quadratic form K_1' may be expressed as follows*

$$K_1' = \boldsymbol{U}_p^T \boldsymbol{U}_p. \tag{4.74}$$

Proof.

$$\mathbf{E}(\mathbf{U}) = \mathbf{E}(\sqrt{n} \mathbf{H} \mathbf{Z}) = \sqrt{n} \mathbf{H} \mathbf{E}(\mathbf{Z}) = \sqrt{n} \mathbf{H} \boldsymbol{\theta}, \tag{4.75}$$

and for the variance we have

$$\mathbf{var}(\mathbf{U}) = \mathbf{var}(\sqrt{n} \mathbf{H} \mathbf{Z}) = \sqrt{n} \mathbf{H} \mathbf{var}(\mathbf{Z}) \sqrt{n} \mathbf{H}^T = n \mathbf{H} \frac{1}{n} diag(\boldsymbol{\sigma}^2) \mathbf{H}^T = \mathbf{H} diag(\boldsymbol{\sigma}^2) \mathbf{H}^T., \tag{4.76}$$

As for the second part, we have

$$\mathbf{U}^T \mathbf{U} = \sqrt{n} (\mathbf{H} \mathbf{Z})^T \sqrt{n} (\mathbf{H} \mathbf{Z}) = n \mathbf{Z}^T \mathbf{H}^T \mathbf{H} \mathbf{Z} = n \mathbf{Z}^T \mathbf{H} \mathbf{Z}, \tag{4.77}$$

where the last equality is due to the idempotency of $\mathbf{H}_{p \times p}$. □

Proposition 4.30. *There exists a random vector $\boldsymbol{X}_0 \sim N_p(\boldsymbol{\mu}_p, \boldsymbol{I}_{p\times p})$, and a matrix $\boldsymbol{S}_{p\times p}$ such that*

$$\boldsymbol{U}_p = \boldsymbol{S}_{p\times p}\boldsymbol{X}_0, \tag{4.78}$$

and consequently the quadratic form K_1' can be written as

$$K_1' = \boldsymbol{X}_0^T \boldsymbol{S}_{p\times p}^T \boldsymbol{S}_{p\times p}\boldsymbol{X}_0. \tag{4.79}$$

Proof. We will provide a constructive proof of (4.78) by finding the matrix $\mathbf{S}$, and the vector $\mathbf{X}_0$. In order for $\mathbf{S}$, and $\mathbf{X}_0$ to satisfy (4.78), the following must hold.

$$\sqrt{n}\mathbf{H}\boldsymbol{\theta} = \mathbf{E}(\mathbf{U}) = \mathbf{E}(\mathbf{S}\mathbf{X}_0) = \mathbf{S}\mathbf{E}(\mathbf{X}_0) = \mathbf{S}\boldsymbol{\mu}, \tag{4.80}$$

and

$$\mathbf{H}diag(\boldsymbol{\sigma}^2)\mathbf{H}^T = \mathbf{var}(\mathbf{U}) = \mathbf{var}(\mathbf{S}\mathbf{X}_0) = \mathbf{S}\mathbf{var}(\mathbf{X}_0)\mathbf{S}^T = \mathbf{S}\mathbf{I}\mathbf{S}^T. \tag{4.81}$$

By choosing $\mathbf{S} = \mathbf{H}_{p\times p}diag(\boldsymbol{\sigma}_p)$, where $\boldsymbol{\sigma}_p = (\sigma_1, ..., \sigma_p)^T$, is the vector of standard deviations of $\mathbf{Y}$, and $\mathbf{H}_{p\times p}$ is a Centering matrix, the equation (4.81) is satisfied. If we now plug $\mathbf{S} = \mathbf{H}_{p\times p}diag(\boldsymbol{\sigma}_p)$ into (4.80) we get that in order for the equality to be satisfied $\boldsymbol{\mu} = \sqrt{n}\big(\frac{\theta_1}{\sigma_1}, ..., \frac{\theta_p}{\sigma_p}\big)^T$.

As for the second part, we have

$$K_1' = \mathbf{U}^T\mathbf{U} = (\mathbf{S}\mathbf{X}_0)^T\mathbf{S}\mathbf{X}_0 = \mathbf{X}^T\mathbf{S}^T\mathbf{S}\mathbf{X}_0. \tag{4.82}$$

□

Proposition 4.31. *The matrix $\boldsymbol{S}_{p\times p}^T\boldsymbol{S}_{p\times p}$ is positively semidefinite of a rank $p-1$.*

Proof. Let us first do a following observation

$$\mathbf{S}^T\mathbf{S} = (\mathbf{H}diag(\boldsymbol{\sigma}))^T\mathbf{H}diag(\boldsymbol{\sigma}) = diag(\boldsymbol{\sigma})\mathbf{H}\mathbf{H}diag(\boldsymbol{\sigma}) = diag(\boldsymbol{\sigma})\mathbf{H}diag(\boldsymbol{\sigma}), \tag{4.83}$$

where the second equality is due to symmetry, and the third equality is due to idempotency of the centering matrix $\mathbf{H}$. Now let $\mathbf{v}_p \in \mathbb{R}^p$ be an arbitrary vector. First we check the symmetry

$$\begin{aligned}\big(diag(\boldsymbol{\sigma})\mathbf{H}diag(\boldsymbol{\sigma})\big)^T &= (diag(\boldsymbol{\sigma})^T(diag(\boldsymbol{\sigma})\mathbf{H})^T \\ &= (diag(\boldsymbol{\sigma}))^T\mathbf{H}^T diag(\boldsymbol{\sigma})^T \\ &= diag(\boldsymbol{\sigma})\mathbf{H}diag(\boldsymbol{\sigma}),\end{aligned} \tag{4.84}$$

where the last equality is due to the symmetry of $\mathbf{H}$ and $diag(\boldsymbol{\sigma})$. Now we check the condition of positive semidefiniteness

$$\mathbf{v}^T diag(\boldsymbol{\sigma})\mathbf{H}diag(\boldsymbol{\sigma})\mathbf{v} = \big[(diag(\boldsymbol{\sigma}))^T(\mathbf{v}^T)^T\big]^T\mathbf{H}diag(\boldsymbol{\sigma})\mathbf{v} = \big(diag(\boldsymbol{\sigma})\mathbf{v}\big)^T\mathbf{H}diag(\boldsymbol{\sigma})\mathbf{v}, \tag{4.85}$$

and if we denote $\mathbf{u} = diag(\boldsymbol{\sigma})\mathbf{v}$, then $\mathbf{u}_p \in \mathbb{R}^p$ is again an arbitrary vector, and we have

$$\mathbf{u}\mathbf{H}\mathbf{u} \geq 0, \tag{4.86}$$

due to positive semidefinitness of the centering matrix $\mathbf{H}$. Regarding the rank, clearly $diag(\boldsymbol{\sigma})$ is of a full rank p, since it is a diagonal matrix. We also know that the rank of the Centering matrix $\mathbf{H}$ is $p-1$ (see Lemma 4.4). Due to the Proposition 1.5 we have that $rank(diag(\boldsymbol{\sigma})\mathbf{H}) = rank(\mathbf{H}) = p-1$, and therefore by applying the proposition again we have $rank(diag(\boldsymbol{\sigma})\mathbf{H}diag(\boldsymbol{\sigma})) = p-1$. □

Theorem 4.32. *There exists a decomposition of matrix $\boldsymbol{S}_{p\times p}^T \boldsymbol{S}_{p\times p}$ such that*

$$\boldsymbol{S}_{p\times p}^T \boldsymbol{S}_{p\times p} = \boldsymbol{P}_{p\times p} \boldsymbol{\Lambda}_{p\times p} \boldsymbol{P}_{p\times p}, \text{ and } \boldsymbol{P}_{p\times p} \boldsymbol{P}_{p\times p}^T = \boldsymbol{I}_{p\times p}, \tag{4.87}$$

where $\boldsymbol{P}_{p\times p}$ is an orthonormal matrix of eigenvectors of $\boldsymbol{S}_{p\times p}^T \boldsymbol{S}_{p\times p}$. Consequently the quadratic form K_1' can be written in the form

$$K_1' = \boldsymbol{X}_0^T \boldsymbol{P}_{p\times p} \boldsymbol{\Lambda}_{p\times p} \boldsymbol{P}_{p\times p}^T \boldsymbol{X}_0 = \boldsymbol{V}_p^T \boldsymbol{\Lambda}_{p\times p} \boldsymbol{V}_p, \tag{4.88}$$

where $\boldsymbol{V}_p$ is a Gaussian random vector given by $\boldsymbol{V}_p = \boldsymbol{P}_{p\times p}^T \boldsymbol{X}_0$, satisfying $\boldsymbol{E}(\boldsymbol{V}_p) = \boldsymbol{P}_{p\times p}^T \boldsymbol{\mu}_p$, and $\mathbf{var}(\boldsymbol{V}_p) = \boldsymbol{I}_{p\times p}$.

Proof. Due to Proposition 4.31 $\mathbf{S}^T\mathbf{S}$ is a real symmetric positive semidefinite matrix, and hence, we can apply Proposition 1.12 to obtain (4.87). The equation (4.88) is obtained merely by plugging (4.87) into (4.79). $\mathbf{V}$ is clearly Gaussian, since it is obtained as a linear transformation of Gaussian random vector $\mathbf{X}_0$ (see [2]), as for the characteristics we have

$$\mathbf{E}(\mathbf{Z}) = \mathbf{E}(\mathbf{P}^T\mathbf{X}_0) = \mathbf{P}^T\mathbf{E}(\mathbf{X}_0) = \mathbf{P}^T\boldsymbol{\mu}, \tag{4.89}$$

since $\mathbf{E}(\mathbf{X}_0) = \boldsymbol{\mu}$, and

$$\mathbf{var}(\mathbf{Z}) = \mathbf{var}(\mathbf{P}^T\mathbf{X}_0) = \mathbf{P}^T\mathbf{var}(\mathbf{X}_0)\mathbf{P} = \mathbf{P}^T\mathbf{I}\mathbf{P} = \mathbf{I}, \tag{4.90}$$

where the last equality is due to matrix $\mathbf{P}$ being an orthonormal matrix. □

Corollary 4.33. *Quadratic form K_1' can be expressed as*

$$K_1' = \sum_{i=1}^{p-1} \lambda_i X_i, \tag{4.91}$$

a linear combination of $p-1$ independent random variables $X_i \sim \chi^2_{1,\delta_i}$ $i = 1, ..., p-1$, where strictly positive numbers λ_i are the nonzero eigenvalues of $\boldsymbol{S}_{p\times p}^T \boldsymbol{S}_{p\times p}$.

Proof. Due to Theorem 4.32 we have that $K_1' = \mathbf{V}^T\boldsymbol{\Lambda}\mathbf{V}$, evaluating the product we obtain that

$$K_1' = \sum_{i=1}^{p} \lambda_i V_i^2, \tag{4.92}$$

where λ_i are the eigenvalues of $\mathbf{S}^T\mathbf{S}$ and V_i are the mutually independent entries of the vector $\mathbf{V} \sim N(\mathbf{P}^T\boldsymbol{\mu}, \mathbf{I})$. If we denote $\mathbf{P} = [\mathbf{p}_1, ..., \mathbf{p}_p]$, where $\mathbf{p}_1, ..., \mathbf{p}_p$ are the corresponding eigenvectors of $\mathbf{S}^T\mathbf{S}$, we have that $\mathbf{P}^T\boldsymbol{\mu} = (\mathbf{p}_1^T\boldsymbol{\mu}, ..., \mathbf{p}_p^T\boldsymbol{\mu})^T$, and hence, we can write the distribution of each entry of $\mathbf{V}$ as $V_i \sim N(\mathbf{p}_i^T\boldsymbol{\mu}, 1)$. By applying the Proposition 1.39 for special case $j = 1$, we obtain that each $V_i^2 \sim \chi^2_{1,\delta_i}$, where $\delta_i = (\mathbf{p}_i^T\boldsymbol{\mu})^2$ for each X_i. Due to Proposition 4.31 we have that $rank(\mathbf{S}^T\mathbf{S}) = p-1$, and hence, exactly one eigenvalue of $\mathbf{S}^T\mathbf{S}$ is equal to zero, and we can write

$$K_1' = \sum_{i=1}^{p-1} \lambda_i X_i, \tag{4.93}$$

where the λ_i, $i = 1, ..., p-1$ are the remaining nonzero eigenvalues of $\mathbf{S}^T\mathbf{S}$. Finally, since $\mathbf{S}^T\mathbf{S}$ is positive semidefinite as seen in Proposition 4.31, all its nonzero eigenvalues are strictly positive. □

4.2.3 Approximation of F Statistic

We have seen in the previous subsections, that the denominator of the F statistic in case of violated variance equality assumption is a linear combination of central χ^2 distributed independent random variables, and the numerator is a linear combination of noncentral χ^2 distributed independent random variables. In order to determine the power of the F-test based on the test statistic (4.40) we will provide a method of approximating linear combinations of independent χ^2 random variable by a single χ^2 random variable. The following two statements are based on [12].

Lemma 4.34. *Let $Q = \sum_{k=1}^{m} \lambda_k X_k$, with $\lambda_k > 0$ and X_k mutually independent random variables with probability distribution $\chi^2_{\nu_k,\delta_k}$, then there exist strictly positive numbers λ^*, ν^*, and nonnegative number δ^*, such that $Q^* = \lambda^* X^*$, with $X^* \sim \chi^2_{\nu^*,\delta_*}$, satisfying $\boldsymbol{E}(Q) = \boldsymbol{E}(Q^*)$, and $\boldsymbol{var}(Q) = \boldsymbol{var}(Q^*)$, and therefore*

$$\boldsymbol{P}\{Q \leq q\} \approx \boldsymbol{P}\{Q^* \leq q\}, \tag{4.94}$$

and the formulas for the parameters are

$$\lambda^* = \frac{R_3 + 2R_4}{R_1 + 2R_2}, \tag{4.95}$$

$$\nu^* = \frac{R_1(R_1 + 2R_2)}{R_3 + 2R_4}, \tag{4.96}$$

$$\delta^* = \frac{R_2(R_1 + 2R_2)}{R_3 + 2R_4}, \tag{4.97}$$

with

$$\sum_{k=1}^{m} \lambda_k \nu_k = R_1 = \lambda^* \nu^*, \tag{4.98}$$

$$\sum_{k=1}^{m} \lambda_k \nu_k + \sum_{k=1}^{m} \lambda_k \delta_k = R_1 + R_2 = \lambda^*(\nu^* + \delta^*), \tag{4.99}$$

$$2\sum_{k=1}^{m} \lambda_k^2 \nu_k + 4\sum_{k=1}^{m} \lambda_k^2 \delta_k = 2R_3 + 4S_4 = 2(\lambda^*)^2(\nu^* + 2\delta^*). \tag{4.100}$$

Proof. See [12] for details. □

Theorem 4.35. *Let $Q_1 = \sum_{k=1}^{m_1} \lambda_k X_k$, and $Q_2 = \sum_{k=m_1+1}^{m_1+m_2} \lambda_k X_k$, where $\lambda_k > 0$, and random variables X_k are all mutually independent with probability distributions $\chi^2_{\nu_k,\delta_k}$, with $\delta_k = 0$ for $k > m_1$, then there exist strictly positive numbers λ_1^*, λ_2^*, ν_1^*, ν_2^* and nonnegative number δ_1^* such that $Q_1^* = \lambda_1^* X_1^* \sim \chi^2_{\nu_1^*,\delta_1^*}$, $Q_2^* = \lambda_2^* X_2^* \sim \chi^2_{\nu_2^*}$, and we have $\boldsymbol{E}(Q_1) = \boldsymbol{E}(Q_1^*)$, $\boldsymbol{var}(Q_1) = \boldsymbol{var}(Q_1^*)$, and $\boldsymbol{E}(Q_2) = \boldsymbol{E}(Q_2^*)$, $\boldsymbol{var}(Q_2) = \boldsymbol{var}(Q_2^*)$. In turn with $r^* = r \cdot \frac{\lambda_2^* \nu_2^*}{\lambda_1^* \nu_1^*}$,*

$$\boldsymbol{P}\left\{\frac{Q_1}{Q_2} \leq r\right\} \approx \boldsymbol{P}\left\{\frac{\lambda_1^* X_1^*}{\lambda_2^* X_2^*} \leq r\right\} = \boldsymbol{P}\left\{\frac{X_1^* \nu_2^*}{X_2^* \nu_1^*} \leq r \cdot \frac{\lambda_2^* \nu_2^*}{\lambda_1^* \nu_1^*}\right\} = \mathcal{F}_{F_\delta}(r^*; \nu_1^*, \nu_2^*, \delta_1^*). \tag{4.101}$$

Proof. This is a result of Lemma 4.34. □

Using Lemma 4.34, and Theorem 4.35 we are able to approximate distribution of the ratio (4.40) of quadratic forms K_1', K_2' expressed as linear combinations of χ^2 distributed random variables with strictly positive coefficients (see Corollaries 4.27, 4.33) by a single random variable with F distribution with nonnegative noncentrality parameter. In fact, it can be seen, that the noncentrality parameter is zero under the hypothesis H_0 of equality of expectations of the samples, and positive under the alternative H_1. This is summarised in the following theorem.

Theorem 4.36. *The random variable F given by* (4.40) *assuming unequal variances of the p samples may be approximated by a random variable F^* where $\frac{\lambda_1^*\nu_1^*}{\lambda_2^*\nu_2^*}F^* \sim F_{\nu_1^*,\nu_2^*,\delta_1^*}$, where the parameters $\lambda_1^*, \lambda_2^*, \nu_1^*, \nu_2^*, \delta_1^*$ are given by Lemma 4.34, and Theorem 4.35. Moreover, under the hypothesis H_0 of equality of expectations of the samples is the parameter $\delta_1^* = 0$, and under the alternative H_1 (see* (4.2)*) δ_1^* is positive.*

Proof. **Step 1.** It is easily seen, that the approximation provided by Lemma 4.34, and consequently by Theorem 4.35 may be applied. Indeed, due to Corollaries 4.27, and 4.33 we have seen, that K_1' and K_2' respectively may be written in the form of linear combination of independent χ^2 distributed random variables, moreover, we have seen, that coefficients of the linear combination for K_1' are strictly positive. It is obvious that the coefficients of the linear combination for K_2' are also strictly positive, since they are precisely the variances $\sigma_1^2, ..., \sigma_p^2$ of the p samples. Neither multiplying K_1' by $p(n-1)$ nor K_2' by $p-1$ has any effect on the parity of the coefficients of the sum i. e. the coefficients $n(p-1)\lambda_i$, $i = 1, ...p-1$ and $(p-1)\sigma_j^2$, $j = 1, ..., p$ are still positive coefficients of linear combinations of χ^2 distributed random variables say $K_1'' = p(n-1)K_1'$ and $K_2'' = (p-1)K_2'$ such that $F = \frac{K_1''}{K_2''}$ and hence, we may apply Lemma 4.34 and consequently Theorem 4.35.

Step 2. We will now show, that the distribution of $\frac{\lambda_1^*\nu_1^*}{\lambda_2^*\nu_2^*}F^*$ under the null hypothesis is central F. In other words we want to show, that the parameter δ_1^* given by (4.97), is equal to zero. This is only possible, if the parameters δ_i of all the random variables V_i, $i = 1, ..., p-1$ appearing in the second power in the linear combination are equal to zero (see proof of 4.33). This is equivalent to asking for expectation of each V_i of the linear combination to be equal to zero. The expectation of V_i is given by the product $\mathbf{p}_i^T\boldsymbol{\mu}$ (see Corollary 4.33), where $\mathbf{p}_i$ is an eigenvector - a column of the matrix $\mathbf{P}$ of eigenvectors of $\mathbf{S}^T\mathbf{S}$, and $\boldsymbol{\mu}$ is a vector defined in Proposition 4.30. In order to see, that the product is indeed zero for each $i = 1, ..., p-1$ it is necessary to know what are the entries of the matrix $\mathbf{S}^T\mathbf{S}$. We have that

$$\mathbf{S}^T\mathbf{S} = diag(\boldsymbol{\sigma})\mathbf{H}_{p\times p}diag(\boldsymbol{\sigma}) = \begin{pmatrix} \sigma_1^2(1-\frac{1}{p}) & -\sigma_1\sigma_2\frac{1}{p} & -\sigma_1\sigma_3\frac{1}{p} & ... & -\sigma_1\sigma_p\frac{1}{p} \\ -\sigma_2\sigma_1\frac{1}{p} & \sigma_2^2(1-\frac{1}{p}) & -\sigma_2\sigma_3\frac{1}{p} & ... & -\sigma_2\sigma_p\frac{1}{p} \\ -\sigma_3\sigma_1\frac{1}{p} & -\sigma_3\sigma_1\frac{1}{p} & \sigma_3^2(1-\frac{1}{p}) & ... & -\sigma_3\sigma_p\frac{1}{p} \\ ... & ... & ... & ... & ... \\ -\sigma_p\sigma_1\frac{1}{p} & -\sigma_p\sigma_2\frac{1}{p} & -\sigma_p\sigma_3\frac{1}{p} & ... & \sigma_p^2(1-\frac{1}{p}) \end{pmatrix}. \tag{4.102}$$

Let us now find an eigenvector $\mathbf{p}_p$ of the matrix $\mathbf{S}^T\mathbf{S}$ corresponding with the single zero eigenvalue of $\mathbf{S}^T\mathbf{S}$. We need to find a nontrivial solution of $(\mathbf{S}^T\mathbf{S} - \lambda_p\mathbf{I})\mathbf{p}_p = \mathbf{0}$ where

$\lambda_p = 0$. I. e.

$$\begin{pmatrix} \sigma_1^2(1-\frac{1}{p}) & -\sigma_1\sigma_2\frac{1}{p} & -\sigma_1\sigma_3\frac{1}{p} & ... & -\sigma_1\sigma_p\frac{1}{p} \\ -\sigma_2\sigma_1\frac{1}{p} & \sigma_2^2(1-\frac{1}{p}) & -\sigma_2\sigma_3\frac{1}{p} & ... & -\sigma_2\sigma_p\frac{1}{p} \\ -\sigma_3\sigma_1\frac{1}{p} & -\sigma_3\sigma_1\frac{1}{p} & \sigma_3^2(1-\frac{1}{p}) & ... & -\sigma_3\sigma_p\frac{1}{p} \\ ... & ... & ... & ... & ... \\ -\sigma_p\sigma_1\frac{1}{p} & -\sigma_p\sigma_2\frac{1}{p} & -\sigma_p\sigma_3\frac{1}{p} & ... & \sigma_p^2(1-\frac{1}{p}) \end{pmatrix} \begin{pmatrix} p_{p,1} \\ p_{p,2} \\ p_{p,3} \\ ... \\ p_{p,p} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ ... \\ 0 \end{pmatrix}. \tag{4.103}$$

Since the standard deviations are by definition positive, we can for $i = 1, ..., p$ multiply i-th row by $\frac{1}{\sigma_i}$ to obtain

$$\begin{pmatrix} \sigma_1(1-\frac{1}{p}) & -\sigma_2\frac{1}{p} & -\sigma_3\frac{1}{p} & ... & -\sigma_p\frac{1}{p} \\ -\sigma_1\frac{1}{p} & \sigma_2(1-\frac{1}{p}) & -\sigma_3\frac{1}{p} & ... & -\sigma_p\frac{1}{p} \\ -\sigma_1\frac{1}{p} & -\sigma_1\frac{1}{p} & \sigma_3(1-\frac{1}{p}) & ... & -\sigma_p\frac{1}{p} \\ ... & ... & ... & ... & ... \\ -\sigma_1\frac{1}{p} & -\sigma_2\frac{1}{p} & -\sigma_3\frac{1}{p} & ... & \sigma_p(1-\frac{1}{p}) \end{pmatrix} \begin{pmatrix} p_{p,1} \\ p_{p,2} \\ p_{p,3} \\ ... \\ p_{p,p} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ ... \\ 0 \end{pmatrix}. \tag{4.104}$$

From the system of the equations (4.104) it is obvious that the nontrivial solution is $\mathbf{p}_p = c \cdot \left(\frac{1}{\sigma_1}, ..., \frac{1}{\sigma_p}\right)$, where c is a real constant. Due to Theorem 4.32, there exists a decomposition of $\mathbf{S}^T\mathbf{S}$ into a product of a diagonal matrix of eigenvalues of $\mathbf{S}^T\mathbf{S}$, and an orthogonal matrix $\mathbf{P}$, whose columns are eigenvectors of $\mathbf{S}^T\mathbf{S}$. This implies that the arithmetic and geometric multiplicity of each eigenvalue of $\mathbf{S}^T\mathbf{S}$ are equal (see for example [13]), namely both the geometric and algebraic multiplicity of $\lambda_p = 0$ is equal to one. We want to see that the vector $\mathbf{p}_p$ is always present among the columns of $\mathbf{P}$. We recall now, how the matrix $\mathbf{P}$ was found in the proof of Proposition 1.12. We have assumed the existence of an orthonormal set $\mathbf{p}_1, ..., \mathbf{p}_s$, choose a vector $\mathbf{x}$ orthogonal to $\mathcal{M}(\mathbf{p}_1, ..., \mathbf{p}_s)$, and it was show that there exists an eigenvector $\mathbf{p}_{s+1} \in \mathcal{M}(\mathbf{x}, \mathbf{A}\mathbf{x}, \mathbf{A}^2\mathbf{x}, ...)$, which is orthogonal to $\mathbf{p}_1, ..., \mathbf{p}_s$. Two possibilities may arise, either $\mathbf{p}_p$ belongs to the orthonormal set $\mathbf{p}_1, ..., \mathbf{p}_s$, and hence, is one of the columns of $\mathbf{P}$, or it does not belong into $\mathbf{p}_1, ..., \mathbf{p}_s$, but then, by choosing $\mathbf{x}$ we must eventually pick $\mathbf{x} = \mathbf{p}_p$, since the multiplicity of λ_p is one and therefore using Proposition 1.10 we have that $\mathbf{p}_p$ is orthogonal to all eigenvectors corresponding to any other eigenvalue of $\mathbf{S}^T\mathbf{S}$. Clearly then $\mathbf{p}_p$ is the eigenvector $\mathbf{p}_{s+1} \in \mathcal{M}(\mathbf{p}_p, \mathbf{S}^T\mathbf{S}\mathbf{p}_p, ...)$, where $(\mathbf{S}^T\mathbf{S})^k\mathbf{p}_p = \mathbf{0}$ for $k > 0$ and hence, is again one of the columns of $\mathbf{P}$.

Finally we observe, that under the hypothesis $H_0 : \theta_1 = ... = \theta_p$ (see (4.1)) of equality of expectations among the random sample the vector $\boldsymbol{\mu} = \left(\frac{\theta}{\sigma_1}, ..., \frac{\theta}{\sigma_p}\right) = \theta\left(\frac{1}{\sigma_1}, ..., \frac{1}{\sigma_p}\right)$, and therefore is obviously orthogonal to all eigenvectors of $\mathbf{S}^T\mathbf{S}$ but $\mathbf{p}_p$. Therefore the expectations of all entries of the vector $\mathbf{V}$ (see Corollary 4.33) apart from one is equal to zero. But the coefficient of the one eigenvalue with a nonzero expectation is $\lambda_p = 0$ and so it is not present in the linear combination K_1'.

Step 3. Lastly, let us see that under the alternative the coefficient δ_1^* is strictly positive, that is equivalent to the claim, that there exists at least one random variable X_i in the linear combination K_1' that has nonzero expectation. The vector $\boldsymbol{\mu}$ under the alternative has the form $\boldsymbol{\mu} = \left(\frac{\theta_1}{\sigma_1}, ..., \frac{\theta_p}{\sigma_p}\right)$. Clearly $\boldsymbol{\mu}$ is not orthogonal with $\mathbf{p}_p$, and $\boldsymbol{\mu} \neq c \cdot \mathbf{p}_p$, where $c \in \mathbb{R}$. Therefore it can not be orthogonal to $\mathbf{p}_1, ...\mathbf{p}_{p-1}$ either since vectors $\mathbf{p}_1, ...\mathbf{p}_p$ form an orthonormal base and hence, the product $\mathbf{p}_i^T\boldsymbol{\mu}$ will be nonzero for all $i = 1, ..., p$ which concludes the proof. □

Remark 4.37. *For the end of this subsection let us remark, that the results for the power functions of the F statistic introduced in the Subsection 4.1.3 hold for the F statistic* *developed under the assumption of unequal variances. The degrees of freedom of* $\frac{p-1}{p(n-1)}F$ *must be of course taken in accordance with Theorem 4.35 under the hypothesis* H_0.

Chapter 5

Application and Results of Comparison of Selected Transformations within ANOVA Framework

In this chapter we will apply the theoretical results obtained in the Chapters 3 and 4 on samples from Poission and negative binomial distribution.

For a random variable with Poisson or negative binomial probability distribution a logarithmic transformation is often used (see [15]). The problem of occurrence of zero observations is solved usually by adding one. The goal of this chapter is to provide computations of both theoretical and simulated powers of the F test (see the first paragraph of Chapter 4 for the description of the model and Sections 4.1 and 4.2, where the F statistic under the assumption of equal and unequal variances is studied, for further reading about One-Way Anova see [2], or [8]) applied to test the hypothesis of equality of expectations of p samples from a size n of either Poisson or negative binomial probability distribution transformed via the logarithmic transformation

$$Y = \ln(X + 1) \tag{5.1}$$

and via the variance stabilising transformations introduced in Chapter 3 and compare them. We will also provide all the necessary theory concerning the transformation (5.1) applied to Poisson and negative binomially distributed random variable.

5.1 Transformation ln(X + 1) Applied on Sample from Poisson Distribution

Through the whole section we will assume, that $X \sim Po(\lambda)$ if not explicitly stated otherwise. The goal of this section is to develop approximation formulae for the numerical characteristics of random variable Y obtained via the transformation (5.1) when applied to X. From these approximations we will see that in the model we are assuming the approximation of variance of Y is a function of the parameter λ of the Poisson distribution (see Proposition 1.25) and therefore it can not be equal among p samples from Poisson distributions $Po(\lambda_i)$, $i = 1, ..., p$ transformed via (5.1).

For the Poisson case, the transformation (5.1) can not be obtained neither as a result of the variance stabilising condition (2.1) nor by any other natural way. In order to develop the approximations of numerical characteristics of random variable Y obtained via the transformation (5.1) when applied to X we will use the same method that was applied in Section 3.1 when dealing with transformation (3.4) which is based on [1]. Since the procedure is almost identical we will omit some details.

Let us as in Section 3.1 consider the following, let

$$Z = X - \lambda \tag{5.2}$$

be a random variable, and

$$\lambda' = \lambda + 1. \tag{5.3}$$

The transformation (5.1) may be rewritten as

$$Y = \ln(Z + \lambda'). \tag{5.4}$$

By Taylor theorem for any $z \geq -\lambda'$ we obtain an infinite series representation

$$y = \ln(\lambda') + \frac{1}{\lambda'}z - \frac{1}{2(\lambda')^2}z^2 + \frac{1}{3(\lambda')^3}z^3 - \frac{1}{4(\lambda')^4}z^4 + ... + \frac{1}{(s-1)(\lambda')^{s-1}}z^{s-1} + R_s, \tag{5.5}$$

where R_s is a reminder term.

Lemma 5.1. *For $z > 0$ the term R_s satisfies*

$$|R_s| < \frac{1}{s(\lambda')^s}z^s \tag{5.6}$$

Proof. This is a direct result of Lagrange's form of the reminder term (see [6]). □

The following Lemma corresponds with the Lemma 3.3.

Lemma 5.2. *For $z > -\lambda'$ the term R_s satisfies*

$$|R_s| < G(s)\frac{1}{(\lambda')^s}|z^s| \tag{5.7}$$

Proof. The idea of the proof is identical with the one of the proof of Lemma 3.3 and hence, will be omitted. □

Remark 5.3. *In this section we set the random variables X and Z to be the same as those introduced in the Section 3.1. Therefore the Lemmata 3.4, 3.5, 3.7, and the Corollary 3.6 hold and the result* (3.19) *given by Remark 3.8 is valid as well.*

Lemma 5.4. *Let Y be the random variable obtained by transformation* (5.1) *applied on X. Then its expectation may be approximated by*

$$\boldsymbol{E}Y = \ln(\lambda + 1) - \frac{1}{2\lambda} + \frac{7}{12\lambda^2} + O\left(\frac{1}{\lambda^3}\right). \tag{5.8}$$

Proof. By Corollary 3.6 we may take the expectation of the right hand side of (5.5) and its powers and derive asymptotic expansions for the moments of Y as $\lambda \longrightarrow \infty$. For the expected values we have

$$\begin{aligned} \mathbf{E}Y = \mathbf{E}\Big\{ & \ln(\lambda+1) + \frac{1}{\lambda+1}Z - \frac{1}{2(\lambda+1)^2}Z^2 + \frac{1}{3(\lambda+1)^3}Z^3 - \frac{1}{4(\lambda+1)^4}Z^4 \\ & + \frac{1}{5(\lambda+1)^5}Z^5 - \frac{1}{6(\lambda+1)^6}Z^6 + O\left(z^5\right)\Big\}. \end{aligned} \tag{5.9}$$

By using (3.19) and the linearity property of the expectation (see [4]) we obtain

$$\mathbf{E}Y = \ln(\lambda+1) - \frac{\lambda}{2(\lambda+1)^2} + \frac{\lambda}{3(\lambda+1)^3} - \frac{3\lambda^2+\lambda}{4(\lambda+1)^4} + \frac{10\lambda^2+\lambda}{5(\lambda+1)^5} - \frac{15\lambda^3+25\lambda^2+\lambda}{6(\lambda+1)^6} + O\Big(\frac{1}{\lambda^4}\Big). \tag{5.10}$$

Now by approximating all listed terms with their asymptotic expansions for $\lambda \longrightarrow \infty$ and some further computation we obtain

$$\mathbf{E}Y = \ln(\lambda+1) - \frac{1}{2\lambda} + \frac{7}{12\lambda^2} - \frac{1}{4\lambda^3} - \frac{22}{15\lambda^4} + O\Big(\frac{1}{\lambda^5}\Big), \tag{5.11}$$

which concludes the proof (for more details on the exact form of the asymptotic expansions used and the additional computations the reader is kindly advised to see file LnPlusOnePoExpectation.mw that can be found on the included CD of the electronic appendix of this work. □

Theorem 5.5. *Let Y be the random variable obtained by transformation* (5.1) *applied on X. Then its variance may be approximated by*

$$\mathbf{var}Y = \frac{1-\ln(\lambda)+\ln(\lambda+1)}{\lambda} - \frac{1}{6}\frac{9-7\ln(\lambda)+7\ln(\lambda+1)}{\lambda^2} \tag{5.12}$$

Proof. The variance of Y may be obtained as $\mathbf{var}Y = \mathbf{E}Y^2 - (\mathbf{E}Y)^2$ (see [2]), hence, we will proceed with finding the approximations for $\mathbf{E}Y^2$ and $(\mathbf{E}Y)^2$. The random variable $Y^2 = \big(\ln(X+1)\big)^2 = \big(\ln(Z+\lambda')\big)^2$ where the last equation is due to reparametrisation given by (5.2) and (5.3) may be expanded into Taylor series as follows

$$y = \ln^2(\lambda') + \frac{2\ln(\lambda')}{\lambda'}z + \Big(-\frac{\ln(\lambda')}{(\lambda')^2} + \frac{1}{(\lambda')^2}\Big)z^2 + \Big(\frac{2}{3}\frac{\ln(\lambda')}{(\lambda')^3} - \frac{1}{(\lambda')^3}\Big)z^3 + \Big(-\frac{1}{2}\frac{\ln(\lambda')}{(\lambda')^4} + \frac{11}{12(\lambda')^4}\Big)z^4 + O(z^5). \tag{5.13}$$

For a Lagrange reminder term R_s of the Taylor series statements similar to Lemmata 5.1, and 5.2 may be shown. We will omit them. Furthermore as already mentioned in Remark 5.3 the Lemmata 3.4, 3.5, 3.7, and the Corollary 3.6 hold and the result (3.19) given by Remark 3.8 is valid as well. Therefore we take expectations of the random equivalent of the right hand side of (5.13) and its powers, and derive asymptotic expansions for the moments of Y as $\lambda \longrightarrow \infty$. We suppose that the idea of this step is rather obvious since it was used on multiple occasions in proofs in Section 3.1 as well as in previous Lemma of this section, hence, we will omit writing all the computations explicitly. Reader interested in the details of this computation is kindly advised to see file LnPlusOnePoVariance.mw included in on the CD of the electronic appendix of this work.

From the computation we get that the expectation of Y^2 may be approximated by

$$\mathbf{E}Y^2 = \ln^2(\lambda+1) + \frac{1-\ln(\lambda)}{\lambda} - \frac{\frac{5}{4}-\frac{7}{6}\ln(\lambda)}{\lambda^2} + O\Big(\frac{1}{\lambda^3}\Big). \tag{5.14}$$

We obtain the term $(\mathbf{E}Y)^2$ by computing the second power of formula (5.8) and neglecting the terms of order $O\big(\frac{1}{\lambda^3}\big)$. Hence,

$$(\mathbf{E}Y)^2 = \ln^2(\lambda+1) - \frac{\ln(\lambda+1)}{\lambda} + \frac{7}{6}\frac{\ln(\lambda+1)}{\lambda^2} + \frac{1}{4\lambda^2} + O\Big(\frac{1}{\lambda^3}\Big). \tag{5.15}$$

The approximation of variance of Y is then obtained by taking

$$\mathbf{var}Y = \mathbf{E}Y^2 - (\mathbf{E}Y)^2. \tag{5.16}$$

□

As an immediate result of the variance approximation formula we have the result about inequality of variance approximations among p independent random samples $\mathbf{X}_1, ..., \mathbf{X}_p$ from distributions $Po(\lambda_i)$, $i = 1, ..., p$ transformed via (5.1), when their expectations are not equal.

Corollary 5.6. *Let $\boldsymbol{X}_1, ..., \boldsymbol{X}_p$ be p independent samples from distributions $Po(\lambda_1), ...,$ $Po(\lambda_p)$. We assume that there exist $i, k \in \{1, ..., p\}$, $i \neq k$ such that $\lambda_i \neq \lambda_k$, i. e. the hypothesis H_1 of inequality of expectations holds (see* (4.2)*). Let $\boldsymbol{Y}_1, ..., \boldsymbol{Y}_p$ be the p random samples obtained from $\boldsymbol{X}_1, ..., \boldsymbol{X}_p$ by applying the transformation* (5.1). *Let $\sigma^2_{11}, ..., \sigma^2_{1n}, ..., \sigma^2_{p1}, ..., \sigma^2_{pn}$ be the approximations of the variance of the random variables Y_{ij}, $i = 1, ..., p$ $j = 1, ..., n$ obtained via Theorem 5.5. Then $\sigma^2_{ij} \neq \sigma^2_{kj}$ for all $i \neq k$, $i, k \in \{1, ..., p\}$.*

Proof. We clearly have that for i fixed we have $\sigma^2_{ij} = \sigma^2_{il}$ for all $j, l \in \{1, ..., p\}$. From equation (5.12) we see, that $\sigma^2_{ij} = \sigma^2(\lambda_{ij})$. Therefore given that there exist i, k such that if $i \neq k$ then $\lambda_i \neq \lambda_k$, the variance approximations $\sigma^2_{ij} = \sigma^2(\lambda_i)$ and $\sigma^2_{kj} = \sigma^2(\lambda_k)$ have to be different. □

Let us also present rather obvious, yet important result covering the case when the assumed random variables that undergo the transformation (5.1) have equal expectations.

Corollary 5.7. *Let $\boldsymbol{X}_1, ..., \boldsymbol{X}_p$ be p independent random samples from distributions $Po(\lambda_i)$, $i = 1, ..., p$. We assume that the expectations λ_i is equal for all $i = 1, ..., p$ (i. e. the hypothesis H_0 of equality of expectations holds, see* (4.1)*). Let $\boldsymbol{Y}_1, ..., \boldsymbol{Y}_p$ be the p random samples obtained from $\boldsymbol{X}_1, ..., \boldsymbol{X}_p$ by applying the transformation* (5.1). *Let $\sigma^2_{11}, ..., \sigma^2_{1n}, ..., \sigma^2_{p1}, ..., \sigma^2_{pn}$ be the approximations of the variance of the random variables Y_{ij}, $i = 1, ..., p$, $j = 1, ..., n$ obtained via Theorem 5.5. Then $\sigma^2_{11} = \sigma^2_{1n} = ... = \sigma^2_{p1} = \sigma^2_{pn} = \sigma^2$.*

Proof. We clearly have that for i fixed we have $\sigma^2_{ij} = \sigma^2_{il}$ for all $j, l \in \{1, ..., p\}$. We have already seen in proof of Corollary 5.6, that the variability in the variance approximations $\sigma^2_{ij} = \sigma^2(\lambda_i)$ of the variance of random variables Y_{ij} is caused by the fact that there exist $i, k \in \{1, ..., p\}$ such that λ_i, λ_k different. If we assume the converse we obviously get $\sigma^2_{11} = \sigma^2_{1n} = ... = \sigma^2_{p1} = \sigma^2_{pn} = \sigma^2$. □

Based on Lemmata 3.9, 3.10 and the proofs of the Corollaries 5.6, 5.7 we will in a form of remark introduce the result on equality of variances when the transformation (3.4) is applied.

Remark 5.8. *Let $\boldsymbol{Y}_1, ..., \boldsymbol{Y}_p$ be p independent random samples of a size obtained by transforming independent random samples $\boldsymbol{X}_1, ..., \boldsymbol{X}_p$ of a size n from distributions $Po(\lambda_i)$ via transformation* (3.4).

The idea of the method for finding the optimal values of the constants introduced in the transformation (3.4) *was to choose the constant in such way, that the term of the variance of Y of the highest order that is dependant on λ would vanish. The higher order terms dependant on λ still survive, but for λ large their input will be not significant and* *hence, we may neglect them. Therefore we may say that the variance approximations of random variables Y_{ij} obtained via the transformation* (3.4) *from the random variables X_{ij} will be equal up to the term of the order we decide to neglect regardless of whether the hypothesis H_0 or H_1 holds.*

5.2 Transformation ln(X + 1) Applied on Sample from Negative Binomial Distribution

The goal of this section is to develop approximation formulae for the numerical characteristics of random variable Y obtained via the transformation (5.1) when applied to $X \sim NBi(\mu, \kappa)$. From these approximations we will see that in the model we are assuming the approximation of variance of Y is a function of the parameter μ of the negative binomial distribution (see Proposition 1.32) and therefore it can not be equal among p samples from distributions $NBi(\mu_i, \kappa)$, $i = 1, ..., p$ transformed by (5.1).

Hereinafter, whenever we assume p samples $\mathbf{X}_1, ..., \mathbf{X}_p$ from negative binomial distributions $NBi(\mu_i, \kappa), ..., NBi(\mu_i, \kappa)$ in sequence, we will always assume that the parameter κ is identical for each distribution.

For the negative binomial case the transformation (5.1) is not obtained directly as a result of the variance stabilising condition (2.1) but may be obtained as a special case of the transformation (3.57), where for the parameter A we choose the value $A = 1$. First let us observe how the expectation and variance of Y transformed via (5.1) can be approximated.

Lemma 5.9. *Let Y be a random variable obtained by transformation* (5.1)*. Let $\kappa > 2$ and let the cumulant function $K^*(t)$ be given by Lemma 3.25, then the expectation of the random variable Y can be approximated by*

$$\boldsymbol{E}(Y) = -\ln(\alpha) + \psi(\kappa) + \frac{1 - \frac{1}{2}\kappa}{\kappa - 1}\alpha - \frac{1}{24}\frac{\kappa^5 - 13\kappa^4 + 53\kappa^3 - 95\kappa^2 + 78\kappa - 24}{(\kappa - 1)^3(\kappa - 2)^2}\alpha^2 + O(\alpha^\kappa) \tag{5.17}$$

Proof. The transformation (5.1) may be obtained from (3.57) by taking $A = 1$, therefore the moment generating function approximation given by Theorem 3.73 and consequently also the cumulant generating function approximation given by Lemma 3.25 are valid for this transformation as well just by taking $A = B = 1$, where the second equation comes from the result (3.57). The first cumulant is obtained by computing the first derivative of the cumulant generating function in $t = 0$ (see [3]). The first derivative of the approximation of the cumulant generating function $(K^*)'(t)$ is given by

$$(K^*)'(t) = -\ln(\alpha) + \psi(\kappa + t) + \left(1 - \frac{1}{2}\kappa\right)\frac{\kappa - 1}{(\kappa + t - 1)^2}\alpha + H(t, \kappa) + O(\alpha^\kappa), \tag{5.18}$$

where $H(t, \kappa)$ represents the first derivative of the coefficient of α^2 with respect to t, where $A = B = 1$. Taking this derivative is tedious, yet not particularly technically interesting part of the proof and hence, the detailed form and derivation of the term $H(t, \kappa)$ is provided in the Maple Document LnPlusOneNBSecondDegreeExp.mw included in the digital appendix of this work. By evaluating in $t = 0$ we obtain

$$k_1 = -\ln(\alpha) + \psi(\kappa) + \frac{1 - \frac{1}{2}\kappa}{\kappa - 1}\alpha - \frac{1}{24}\frac{\kappa^5 - 13\kappa^4 + 53\kappa^3 - 95\kappa^2 + 78\kappa - 24}{(\kappa - 1)^3(\kappa - 2)^2}\alpha^2 + O(\alpha^\kappa) \tag{5.19}$$

The fact that the first cumulant of a random variable is equal to its first moment (see [3]) altogether with the first result of Lemma 3.26 concludes the proof. □

Theorem 5.10. *Let Y be a random variable obtained by transformation* (5.1)*. Let $\kappa > 2$ and let the cumulant function $K^*(t)$ be given by Lemma 3.25, then the variance of the random variable Y given by* (5.1) *can be approximated by*

$$\mathbf{var}Y = \psi'(\kappa) + \frac{\kappa-2}{(\kappa-1)^2}\alpha + \frac{1}{12}\frac{6\kappa^6 - 66\kappa^5 + 287\kappa^4 - 638\kappa^3 + 769\kappa^2 - 478\kappa + 120}{(\kappa-1)^4(\kappa-2)^3}\alpha^2 + O(\alpha^\kappa). \tag{5.20}$$

Proof. As mentioned already in the proof of Lemma 5.9 the moment and consequently the cumulant generating function approximations derived for random variable transformed via (3.57) stay valid, and the formulae for the transformation (5.1) are obtained by simply taking $A = B = 1$. In order to obtain the second cumulant we need to compute the second derivative of the cumulant generating function and evaluate it in $t = 0$ (see [3]). The second derivative of $K^*(t)$ is given by

$$(K^*)''(t) = \psi'(\kappa+t) - 2\left(1 - \frac{1}{2}\kappa\right)\frac{\kappa-1}{(\kappa+t-1)^3}\alpha + L(t,\kappa) + O(\alpha^\kappa), \tag{5.21}$$

where $L(t,\kappa)$ represents the second derivative of the coefficient of α^2 with respect to t, where $A = B = 1$. As before the computation of the derivative will not be presented in this proof, but can again be found in the Maple Document LnPlusOneNBSecondDegreeVar.mw included in the digital appendix of this work. By evaluating (5.21) in $t = 0$ we obtain

$$k_1 = \psi'(\kappa) + \frac{\kappa-2}{(\kappa-1)^2}\alpha + \frac{1}{12}\frac{6\kappa^6 - 66\kappa^5 + 287\kappa^4 - 638\kappa^3 + 769\kappa^2 - 478\kappa + 120}{(\kappa-1)^4(\kappa-2)^3}\alpha^2 + O(\alpha^\kappa). \tag{5.22}$$

The fact that the second cumulant of a random variable is equal to its second central moment (see [3]) altogether with the second result of Lemma 3.26 concludes the proof. □

As an immediate result of the variance approximation formula we have the result about inequality of variance approximations among p independent random samples $\mathbf{X}_1, ..., \mathbf{X}_p$ from distributions $NBi(\mu_1, \kappa)$, $i = 1, ..., p$ transformed via (5.1), when their expectations are not equal.

Corollary 5.11. *Let $\boldsymbol{X}_1, ..., \boldsymbol{X}_p$ be p independent samples from distributions $NBi(\mu_1, \kappa), ..., NBi(\mu_p, \kappa)$ in sequence. We assume that there exist $i, k \in \{1, ..., p\}$, $i \neq k$ such that $\mu_i \neq \mu_k$, i. e. the hypothesis H_1 of inequality of expectations holds (see* (4.2)*). Let $\boldsymbol{Y}_1, ..., \boldsymbol{Y}_p$ be the p random samples obtained from $\boldsymbol{X}_1, ..., \boldsymbol{X}_p$ by applying the transformation* (5.1)*. Let $\sigma^2_{11}, ..., \sigma^2_{1n}, ..., \sigma^2_{p1}, ..., \sigma^2_{pn}$ be the approximations of the variance of the random variables Y_{ij}, $i = 1, ..., p$ $j = 1, ..., n$ obtained via Theorem 5.5. Then $\sigma^2_{ij} \neq \sigma^2_{kj}$ for all $i \neq k$, $i, k \in \{1, ..., p\}$.*

Proof. We clearly have that for i fixed we have $\sigma^2_{ij} = \sigma^2_{il}$ for all $j, l \in \{1, ..., p\}$. From equation (5.12) we see, that $\sigma^2_{ij} = \sigma^2(\mu_{ij})$. Therefore given that there exist i, k such that if $i \neq k$ then $\mu_i \neq \mu_k$, the variance approximations $\sigma^2_{ij} = \sigma^2(\mu_i)$ and $\sigma^2_{kj} = \sigma^2(\mu_k)$ have to be different. □

Let us also present rather obvious, yet important result covering the case when the assumed random variables that undergo the transformation (5.1) have equal expectations.

Corollary 5.12. *Let $\boldsymbol{X}_1, ..., \boldsymbol{X}_p$ be p independent random samples from distributions $NBi(\mu_i, \kappa)$, $i = 1, ..., p$. We assume that the expectations μ_i is equal for all $i = 1, ..., p$ (i. e. the hypothesis H_0 of equality of expectations holds, see* (4.1)*). Let $\boldsymbol{Y}_1, ..., \boldsymbol{Y}_p$ be the p random samples obtained from $\boldsymbol{X}_1, ..., \boldsymbol{X}_p$ by applying the transformation* (5.1)*. Let $\sigma^2_{11}, ..., \sigma^2_{1n}, ..., \sigma^2_{p1}, ..., \sigma^2_{pn}$ be the approximations of the variance of the random variables Y_{ij}, $i = 1, ..., p$, $j = 1, ..., n$ obtained via Theorem 5.5. Then $\sigma^2_{11} = \sigma^2_{1n} = ... = \sigma^2_{p1} = \sigma^2_{pn} = \sigma^2$.*

Proof. We clearly have that for i fixed we have $\sigma^2_{ij} = \sigma^2_{il}$ for all $j, l \in \{1, ..., p\}$. We have already seen in proof of Corollary 5.6, that the variability in the variance approximations $\sigma^2_{ij} = \sigma^2(\mu_i)$ of the variance of random variables Y_{ij} is caused by the fact that there exist $i, k \in \{1, ..., p\}$ such that μ_i, μ_k different. If we assume the converse we obviously get $\sigma^2_{11} = \sigma^2_{1n} = ... = \sigma^2_{p1} = \sigma^2_{pn} = \sigma^2$. □

Based on Lemmata 3.9, 3.10 and the proofs of the Corollaries 5.6, 5.7 we will in a form of remark introduce the result on equality of variances when the transformation (3.4) is applied.

Remark 5.13. *Let $\boldsymbol{Y}_1, ..., \boldsymbol{Y}_p$ be p independent random samples of a size obtained by transforming independent random samples $\boldsymbol{X}_1, ..., \boldsymbol{X}_p$ of a size n from distributions $NBi(\mu_i, \kappa)$ via transformation* (3.4).

The idea of the method for finding the optimal values of the constants introduced in the transformation (3.4) *was to choose the constants in such way, that the coefficient of the term of the variance of Y of the smallest order that is dependant α (see Proposition 3.20) would be zero. The higher order terms dependant on α still survive, but for α small (i. e. μ large) their input will be not significant and hence, we may neglect them. Therefore we may say that the variance approximations of random variables Y_{ij} obtained via the transformation* (3.4) *from the random variables X_{ij} will be equal up to the term of the order we decide to neglect regardless of whether the hypothesis H_0 or H_1 holds.*

5.3 Simulation Study of Used Approximations, Effect of Transformations on Parameter Estimates of Poisson or Negative Binomial Distribution

In this section we will check the quality of the approximations of the numerical characteristics of the transformed variables introduced in the Sections 3.1, 3.2, 5.1, and 5.2 as well as the approximation of the ratio of two linear combinations of independent χ^2 distributed random variables introduced in the Subsection 4.2.3 via simulation. Additionally we will study via simulation the effect of the transformations on the expectation parameter of both Poisson and negative binomially distributed variables.

We will restrict ourselves to the comparison of the numerical characteristics of the random variable Y obtained via applying transformation (3.4) or (5.1) on random variable $X \sim Po(\lambda)$ for the Poisson case, and of the random variable Y obtained via transformation (3.56) and (5.1) applied on random variable $X \sim NBi(\mu, \kappa)$ in the negative binomial case.

We are aware that in the negative binomial case for small values of κ the normality assumption might be violated due to big absolute values of the skewness parameter of the transformed variable since we determined in Section 3.2, that the limiting value of skewness tends to zero for large values of κ (see Theorem 3.36). Hence, we will assume that $\kappa > 2$. Therefore we will also not study the transformation (3.57), since we consider it to be only an approximation of (3.56) (see Lemma 3.18), and the value of κ is sufficient for Theorem 3.33 and Corollary 3.30 to hold and so all the optimal values of the constants of (3.56) may be determined. When referring to the transformation (3.56) we will hereinafter assume that the constants c and d are chosen to be optimal.

Let us begin by checking goodness of the approximations of the numerical characteristics. Since the process is identical regardless of whether it is done for Poisson or negative binomial case we will describe it for the Poisson case, while the corresponding data used for the negative binomial case will be mentioned in brackets. We will do this by evaluating the formulae (3.20), (3.26), (5.8), (5.12), ((3.96), (3.108), (5.17), (5.20)) for values of λ (μ) in a given interval I and comparing them with the values of estimators of the corresponding numerical characteristic. The estimators are computed from generated sample of a size n of a distribution $Po(\lambda)$ ($NBi(\mu, \kappa)$) for $\lambda \in I_{Po}$ ($\mu \in I_{NBi}$), on which the transformation (3.4), or (5.1) ((3.56), or (5.1)) was applied. As an estimator of the expectation we use in both cases the arithmetic mean given by

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \tag{5.23}$$

and as an estimator of the variance we use in both cases the sample variance given by

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2. \tag{5.24}$$

We will also check the goodness of approximation given by Theorem 4.35, that will be done in following way. Assume we have p independent random samples $\mathbf{X}_1, ..., \mathbf{X}_p$ from $Po(\lambda_i)$ ($NBi(\mu_i, \kappa)$) , $i = 1, ..., p$. We choose the expectation λ_1 (μ_1) and a step h. Expectation parameters of all other distributions differ from λ_1 (μ_1) by multiples of h. We apply the transformation (5.1) on the independent random samples $\mathbf{X}_i$, from $Po(\lambda_i)$ ($NBi(\mu_i, \kappa)$) $i = 1, ..., p$ in order to obtain transformed random samples $\mathbf{Y}_i$, $i = 1, ..., p$ and determine their numerical characteristics via formulas (5.8), (5.12) ((5.17) and (5.20)). Using the numerical characteristics we compute the vector $\boldsymbol{\mu}$ and the matrix $\mathbf{S}^T\mathbf{S}$ (see Proposition 4.30). We find the eigenvalues λ_i and eigenvectors $\mathbf{p}_i$ for $i = 1, ..., p$ of $\mathbf{S}^T\mathbf{S}$. The nonzero eigenvalues of $\mathbf{S}^T\mathbf{S}$ are the coefficients of the linear combination K_1'. The degrees of freedom of the independent χ^2 distributed random variables of the linear combination K_1' are all equal to one. (see Corollary 4.33). Using the eigenvectors of $\mathbf{S}^T\mathbf{S}$ and the vector $\boldsymbol{\mu}$ the noncentrality parameters of the independent χ^2 distributed random variables of the linear combination K_1' are determined via formua $\delta_i = (\mathbf{p}_i^T\boldsymbol{\mu})^2$ for $i = 1, ..., p$ (see proof of Corollary 4.33). Due to Corollary 4.27 we know that the coefficients of the linear combination K_2' are actually variances σ_i^2 of the transformed random samples $\mathbf{Y}_i$ for $i = 1, ..., p$. In the computations we will use the approximations given by (5.12) ((5.20)). From the same Corollary we also obtain that the degrees of freedom of each independent central χ^2 distributed random variable in the linear combination K_2' are equal to $n-1$. With all the parameters of all the random variables of the quadratic forms K_1' and K_2' determined, we can generate them numerically and compute the ratio (4.40). We do so repeatedly in order to obtain a random sample of the ratio (4.40), from which we determine the empirical quantile function. Moreover, we use the parameters of the random variables of the quadratic forms K_1' and K_2' , to compute the values of the coefficients λ_1^*, λ_2^* (see (4.95)), the degrees of freedom ν_1^*, ν_2^* (see (4.96)), and the noncentrality parameter δ_1^* (see (4.97)) by applying Lemma 4.34 and Theorem 4.35. Hence, we determine the parameters of the approximation and we can generate its quantiles, which we compare graphically with the empirical quantiles. The empirical quantiles in the computation were computed in the R, using function "quantiles".

Finally we will study the effect of the transformations (3.4) and (5.1) ((3.56) and (5.1)) on the data via estimates of the parameter λ (μ) of the Poisson (negative binomial) distribution. This will be done as follows. Let $X_1, ..., X_n$ be a sample of Poisson distribution $Po(\lambda)$ (negative binomial distribution $NBi(\mu, \kappa)$). We transform this sample via transformation (3.4) or (5.1) ((3.56) or (5.1)) to obtain random sample $Y_1, ..., Y_n$. We estimate the expectation of the transformed random sample via arithmetic mean $\overline{Y}$ given by formula (5.23). We apply the respective transformation in reverse to the value of the arithmetic mean $\overline{Y}$ and obtain an estimate λ_Y of λ (μ_Y of μ). We will repeat the procedure described above k-times for the same setting of parameter λ (μ and κ) and using formulas (5.23) and (5.24) compute the arithmetical mean $\overline{\lambda}_Y$ ($\overline{\mu}_Y$) and sample variance s^2 of the sample of λ_Y (μ_Y) obtained via the repetition. When describing the results of this study we will frequently use the term bias defined in Definition 3.13.

5.3.1 Data Input and Results for Poisson Case

The interval I for the parameter λ was set to $I = [0, 500]$ the sample size used in the simulated sample was set on $n = 10000$. In all the figures the blue line represents the values of the approximations and red points represent the values obtained via simulations. We can conclude given the Figures 5.1a, 5.2a, 5.1b, and 5.2b that all the approximation fit rather well. We can observe that the transformation (3.4) provides a good variance stabilisation even for small values of λ (see Figure 5.2a). On the other hand, the variance for the transformation (5.1) continues to decrease for increasing values of λ, which might pose a significant problem namely for small values of λ (see Figure 5.2b).

We will continue by providing some numerical results on the goodness of approximation introduced in the Theorem 4.35. The empirical quantile function is computed from a random sample of the distribution of F statistic of a size $k = 1000$ according to the procedure described earlier in this Section. The number of assumed random subsamples is $p = 3$ and their size is $n = 100$. The graphical comparison is done for the following values of the parameters of the original Poisson distributed random variable $\lambda = 15, 50, 100$ and corresponding values of steps $h = 3, 10, 20$. The values of λ_2 and λ_3 are obtained as follows.

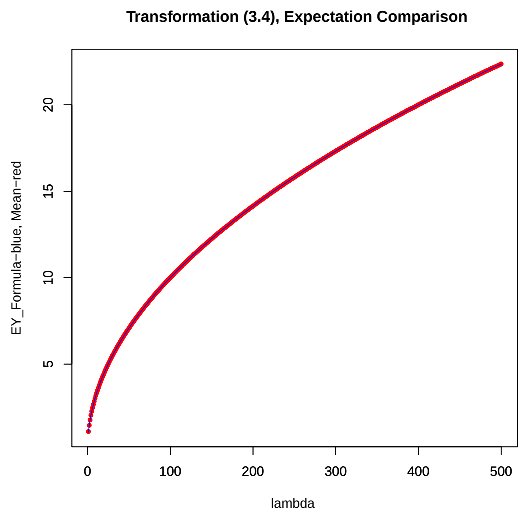
$$\lambda_j = \lambda_1 + (-1)^j (kh_0) \;\; k = 0, ..., 30, \tag{5.25}$$

so that the difference $\Delta\mu_j = |\mu_1 - \mu_j|$ for $j = 2, 3$ increases with the value of k. The blue line represents the quantile function of the random variable $\frac{\lambda_1^* \nu_1^*}{\lambda_2^* \nu_2^*} F^*$ (see Theorem 4.35), the red points are the values of the empirical quantile function. We conclude that for all used values of parameters the approximation has a good fit.

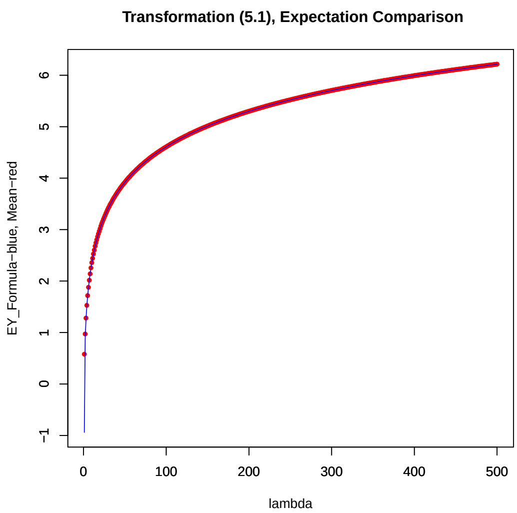
Finally, we will also provide the results of the study of the effect of the transformations (3.4) and (5.1) on the parameter λ. The study was done for the values $\lambda = 5, 10, 20, 50$, The sample size was set to $n = 100$, the number of repetitions was set to $k = 1000$. Histograms for each respective setting can be seen on Figures 5.4a, 5.4b, 5.5a, 5.5b, 5.6a, 5.6b, 5.7a, and 5.7b. The values of the arithmetical mean and sample variance of the samples from distribution identical to that of λ_Y for each respective setting are collected in the Table 5.1. From the histograms we can observe that for both transformations the estimate λ_Y is biased. The sample characteristics of the estimate provided in the table supports this statement. We can also observe that the bias is smaller for the transformation (3.4).

Transformation	Parameter	$\lambda = 5$	$\lambda = 10$	$\lambda = 20$	$\lambda = 50$
$Y = \sqrt{X + 3/8}$	$\overline{\lambda}_Y$	4.749	9.745	19.738	49.758
	s^2	0.053	0.098	0.196	0.520
$Y = \ln(X + 1)$	$\overline{\lambda}_Y$	4.553	9.519	19.503	49.514
	s^2	0.056	0.101	0.198	0.522

Table 5.1: Arithmetic mean and sample variance of samples of distribution identical to λ_Y.

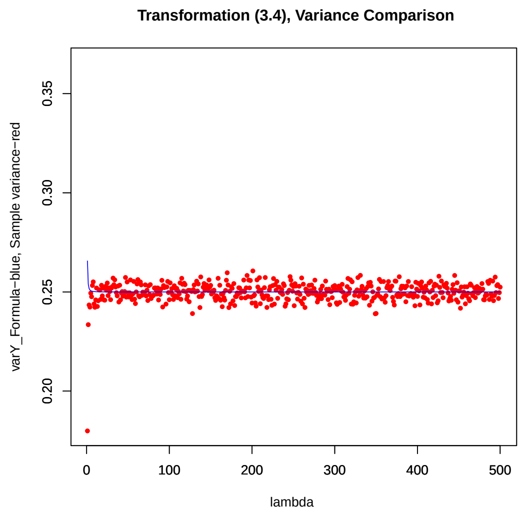


(a) Comparison of expectation approximation formula and arithmetic mean for transformation $Y = \sqrt{X + 3/8}$ (see (3.4)).

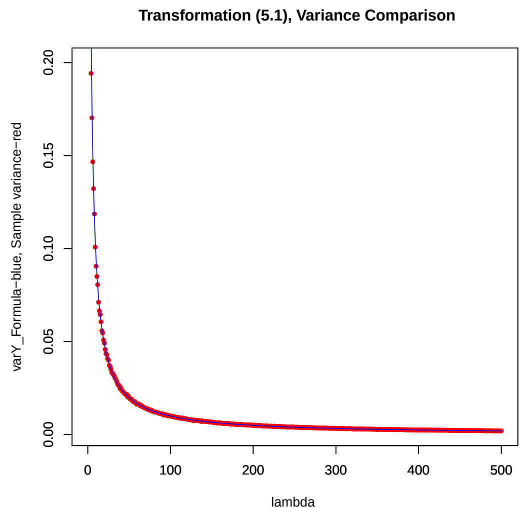


(b) Comparison of expectation approximation formula and arithmetic mean for transformation $Y = \ln(X + 1)$ (see (5.1)).

Figure 5.1: Comparison of expectation approximation formulae and expectation estimates via arithmetic mean for $Po(\lambda)$

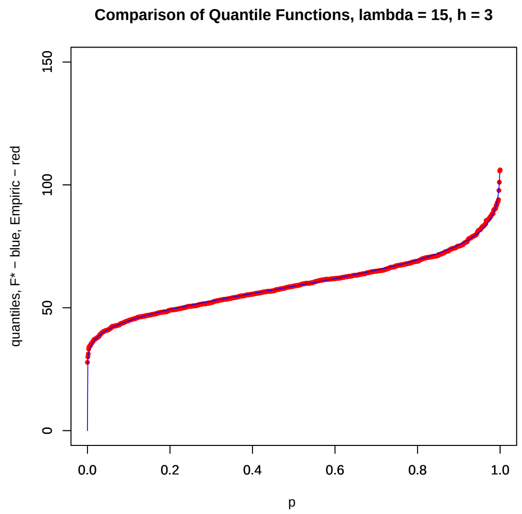


(a) Comparison of variance approximation formula and sample variance for transformation $Y = \sqrt{X + 3/8}$ (see (3.4)).

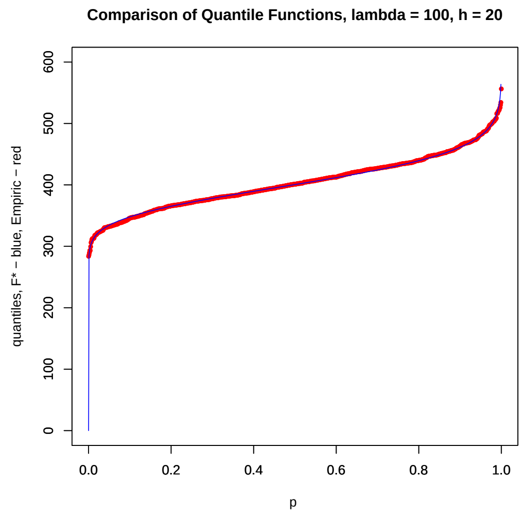


(b) Comparison of variance approximation formula and sample variance for transformation $Y = \ln(X + 1)$ (see (5.1)).

Figure 5.2: Comparison of variance approximation formulae and variance estimates via sample variance mean for $Po(\lambda)$

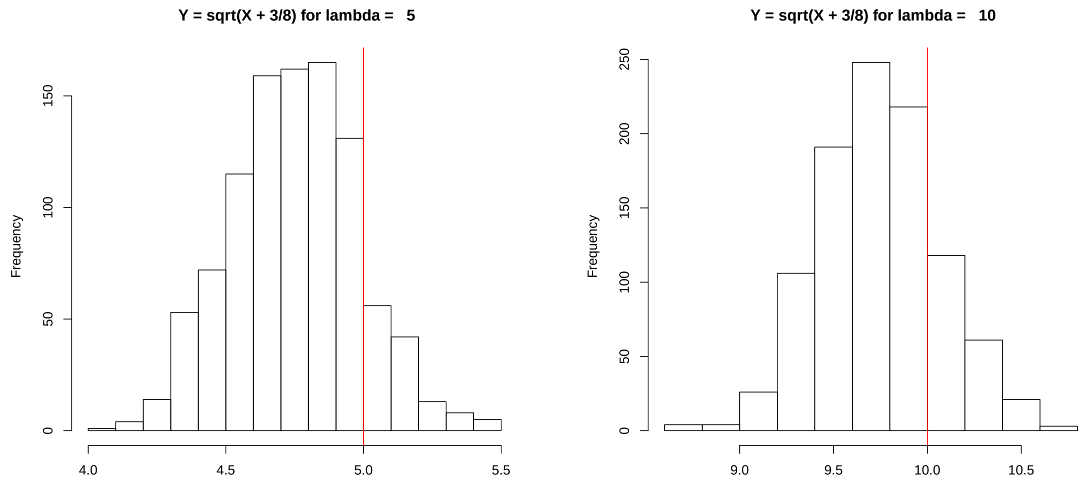


(a) Comparison of the quantile function of random variable F^* and empirical quantile function of the ratio (4.40) for $\lambda = 15$, $h = 3$.



(b) Comparison of the quantile function of random variable F^* and empirical quantile function of the ratio (4.40) for $\lambda = 100$, $h = 20$.

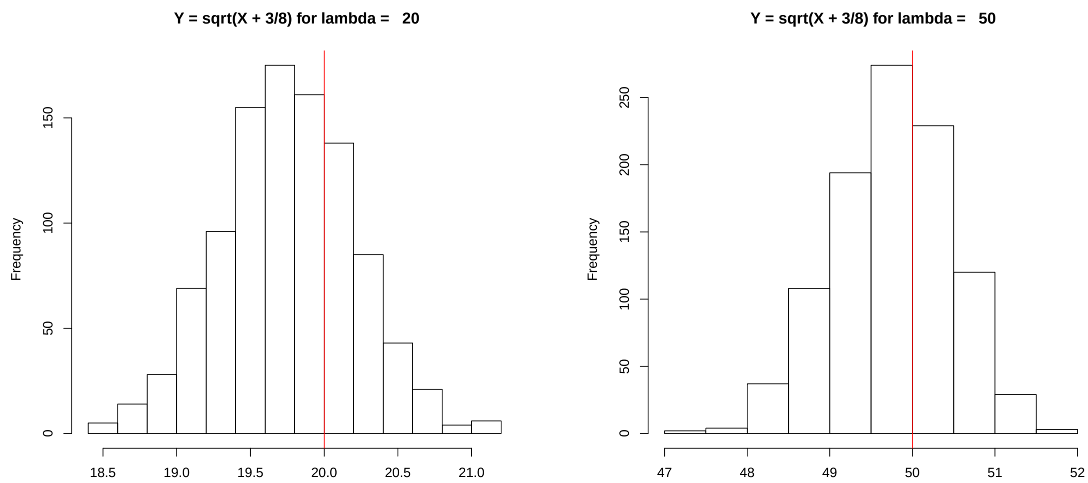
Figure 5.3: Comparison of the quantile function of random variable F^* and empirical quantile function of the ratio (4.40) for $Po(\lambda)$.



(a) Histogram of λ_Y for $\lambda = 5$

(b) Histogram of λ_Y for $\lambda = 10$

Figure 5.4: Histograms of λ_Y for transformation $Y = \sqrt{X + 3/8}$ for sample size $n = 100$ and number of repetitions $k = 1000$ for $Po(\lambda)$.



(a) Histogram of λ_Y for $\lambda = 20$

(b) Histogram of λ_Y for $\lambda = 50$

Figure 5.5: Histograms of λ_Y for transformation $Y = \sqrt{X + 3/8}$ for sample size $n = 100$ and number of repetitions $k = 1000$ for $Po(\lambda)$.

5.3.2 Data Input and Results for Negative Binomial Case

The interval I for the parameter μ was set to $I = [0, 250]$. The sample size used in the simulated sample was set on $n = 10000$. The comparison was done for few different values of the shape parameter κ. Namely $\kappa = 3, 5, 10$. The value of the shape parameter proved to affect both the quality of the approximation and the stability of the sample variance. On the other hand, the effect of varying parameter κ on the expectation approximation did not affect the quality of fit.

In all the figures the blue line represents the values of the approximations and red point represent the values obtained via simulations.

The Figures 5.8a, 5.8b display the comparison of the expectation approximations and their estimates for the transformation (3.56) and (5.1). We can see that the approximation formulae and the estimates have good fit on the whole interval I.

Let us now turn our attention towards the variances. From the Figures 5.9a, 5.9b we can observe that for small values of μ (up to approximately $\mu = 19$ for transformation (3.56) and $\mu = 38$ for transformation (5.1)) the variance approximation differs from the sample variance greatly. These values of μ also describe approximately the point where the sample variances become stable. An interesting result is that the value of μ at which the sample variance of the sample to which the transformation (3.56) was applied is significantly lesser than the value of μ for which the same happens when (5.1) is applied. This difference increases with the increasing values of κ as can be seen in Figures 5.10a, 5.10b and 5.11a, 5.11b, however the growth of the critical value of μ for the transformation (3.56) is significantly slower than for the transformation (5.1). For $\kappa = 5$ the critical value for (3.56) is still somewhere around $\mu = 19$, however for the transformation (5.10b) it is already $\mu = 75$ and for $\kappa = 5$ it is $\mu = 29$ for (3.56) versus $\mu = 90$ for (5.1).

Let us continue by providing some numerical results on the goodness of approximation introduced in the Theorem 4.35, when applied on the ratio $F = \frac{p(n-1)}{p-1}\frac{K_1'}{K_2'}$, where K_1' is given by (4.91) and K_2' is given by (4.67). Let us denote as before by F^* the random variable obtained by applying Theorem 4.35 onto (4.40). We will check the goodness of the approximation by comparing quantile function of the random variable F^* with a empirical quantile function of the random variable F. The empirical quantile function is computed from a random sample of F of a size $n = 1000$. The random sample of F is obtained by generating random samples from each χ^2 distributed random variable of the linear combination K_1' and K_2' and then computing their ratio F.

The graphical comparison is done for the following values of the parameters of the original negative binomially distributed random variable $\mu = 15, 50, 100$ and $\kappa = 3, 5$. The values of $\mu_j,\ \ j = 2, 3$ were computed as follows

$$\mu_j = \mu_1 + (-1)^j (kh) \ \ k = 0, ..., 30, \tag{5.26}$$

so that the difference $\Delta\mu_j = |\mu_1 - \mu_j|$ for $j = 2, 3$ increases with the value of k. The blue line represents the the quantile function of the random variable F^*, the red points are the values of the empirical quantile function.

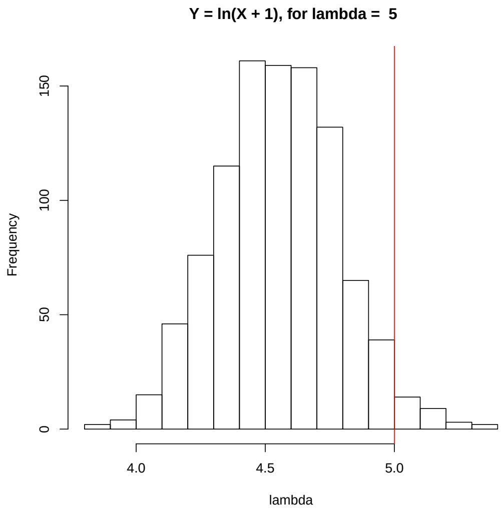
We see from the Figures 5.12a, 5.12b, 5.12c, 5.13a, 5.13b, and 5.13c that for all tested values we obtain a good fit.

Finally we will also provide the results of the study of the effect of the transformations (3.56) and (5.1) on the parameter μ. The study was done for the values of $\mu = 30, 50, 100$ with the respective setting of the shape parameter $\kappa = 3, 5, 10$. The sample size was set to $n = 100$. The number of repetitions was set to $k = 1000$. Histograms for each respective setting can be seen on Figures 5.14a, 5.14b, 5.14c, 5.15a, 5.15b, 5.15c, 5.16a,

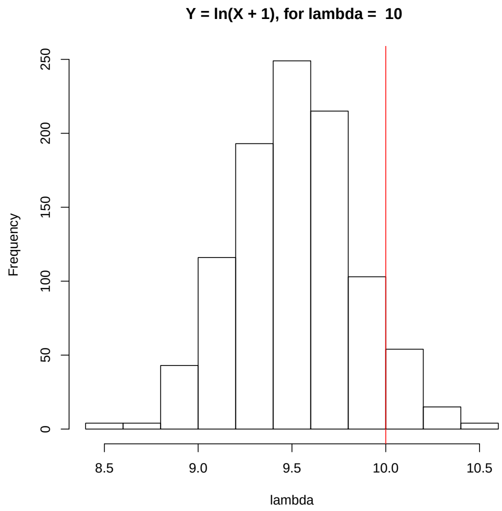
$Y = 2\sinh^{-1}\left(\sqrt{\frac{X+c}{\kappa+d}}\right)$	$\overline{\mu}_Y$	$\mu = 30$	$\mu = 50$	$\mu = 100$	s^2	$\mu = 30$	$\mu = 50$	$\mu = 100$
$\kappa = 3$		24.948	41.898	84.436		9.717	22.489	97.027
$\kappa = 5$		26.897	44.923	90.290		6.092	15.873	63.794
$\kappa = 10$		28.341	47.466	94.917		3.639	9.661	35.467
$Y = \ln(X+1)$	$\overline{\mu}_Y$	$\mu = 30$	$\mu = 50$	$\mu = 100$	s^2	$\mu = 30$	$\mu = 50$	$\mu = 100$
$\kappa = 3$		24.923	41.747	83.954		9.286	25.764	99.045
$\kappa = 5$		26.860	44.840	90.266		6.541	18.696	68.146
$\kappa = 10$		28.086	47.252	94.727		3.558	8.641	35.506

Table 5.2: Table of arithmetic mean and sample skewness of a sample from distribution identical to μ_Y.

5.16b, 5.16c, 5.17a, 5.17b, 5.17c, 5.18a, 5.18b, 5.18c, 5.19a, 5.19b, and 5.19c. The values of the arithmetical mean and sample variance of the samples from the distribution identical with the distribution of μ_Y are collected in Table 5.2. From the histograms we can observe that the estimate μ_Y obtained with the use of any of the two transformations (3.56) and (5.1) is biased. However we notice that the bias is bigger for the case when transformation (5.1) is used. Another interesting result is that for increasing values of κ the bias gets smaller. These observations are supported by the data collected in the table. Additionally we can observe from the table, that the sample variance of the estimates μ_Y decreases for both transformations as the parameter κ grows.

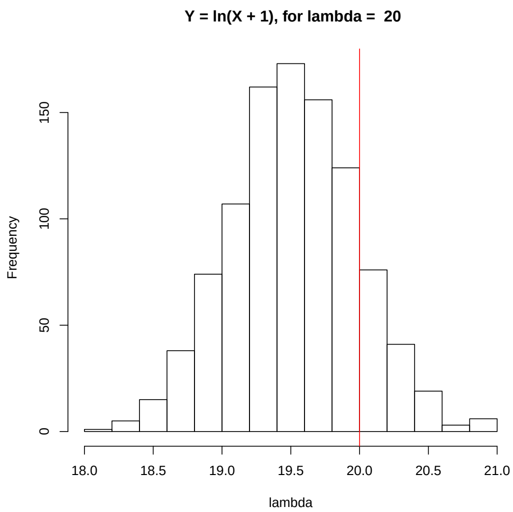


(a) Histogram of λ_Y for $\lambda = 5$

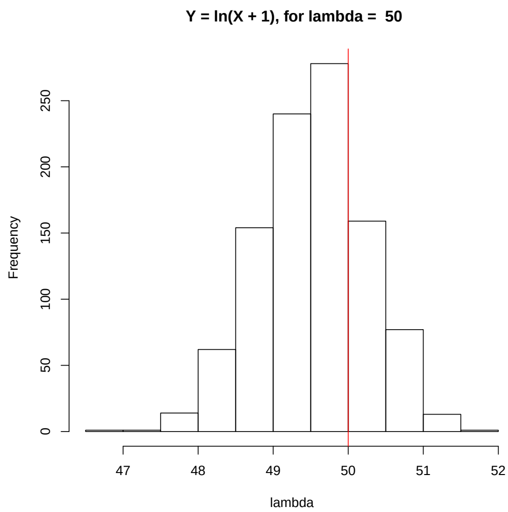


(b) Histogram of λ_Y for $\lambda = 10$

Figure 5.6: Histograms of λ_Y for transformation $Y = \ln(X + 1)$ for sample size $n = 100$ and number of repetitions $k = 1000$ for $Po(\lambda)$.

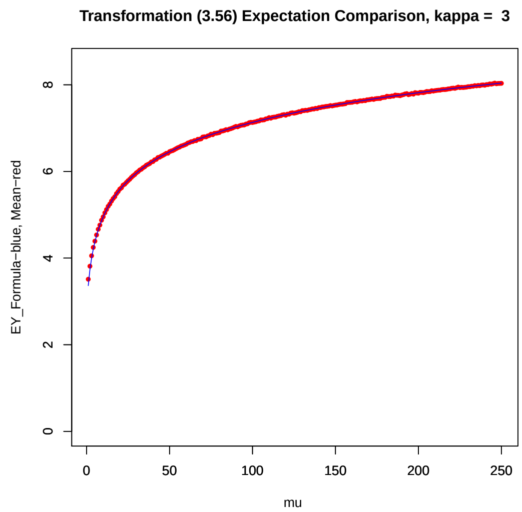


(a) Histogram of λ_Y for $\lambda = 20$

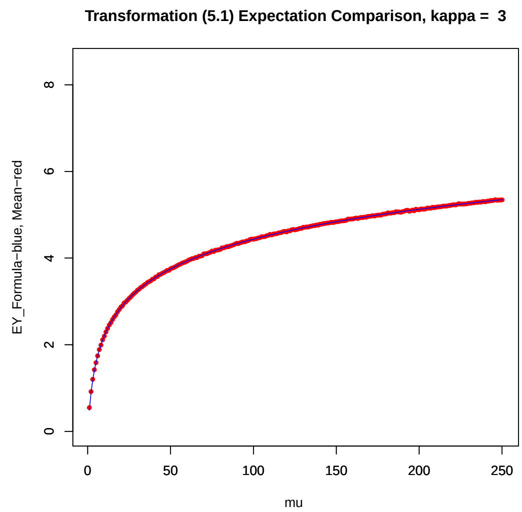


(b) Histogram of λ_Y for $\lambda = 50$

Figure 5.7: Histograms of λ_Y for transformation $Y = \ln X + 1$ for sample size $n = 100$ and number of repetitions $k = 1000$ for $Po(\lambda)$.

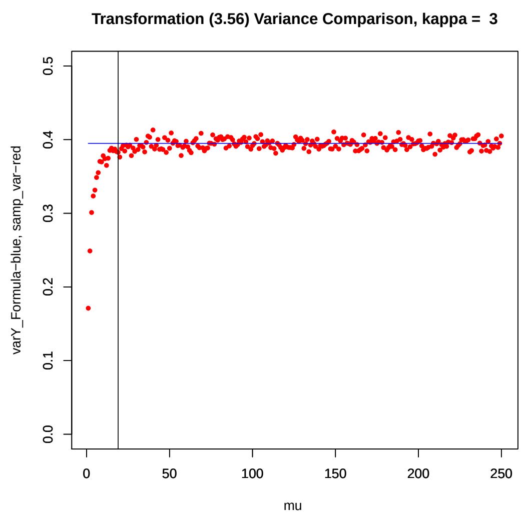


(a) Comparison of expectation approximation formula and arithmetic mean for $Y = 2\sinh^{-1}(\sqrt{(X+c)/(\kappa+d)})$ (see (3.56)) for $\kappa = 3$.

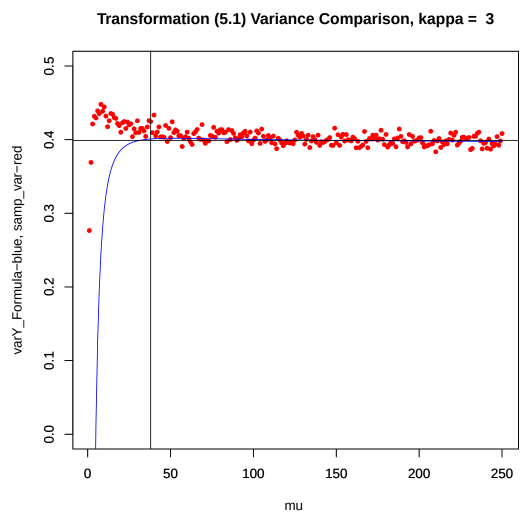


(b) Comparison of expectation approximation formula and arithmetic mean for transformation $Y = \ln(X+1)$ (see (5.1)) for $\kappa = 3$.

Figure 5.8: Comparison of expectation approximation formulae and arithmetic mean for $NBi(\mu, 3)$.

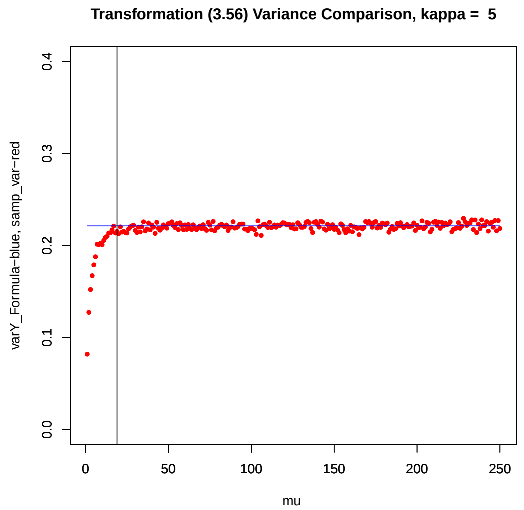


(a) Comparison of variance approximation formula and sample variance for $Y = 2\sinh^{-1}(\sqrt{(X+c)/(\kappa+d)})$ (see (3.56)) for $\kappa = 3$. The horizontal line is at $\mu = 19$.

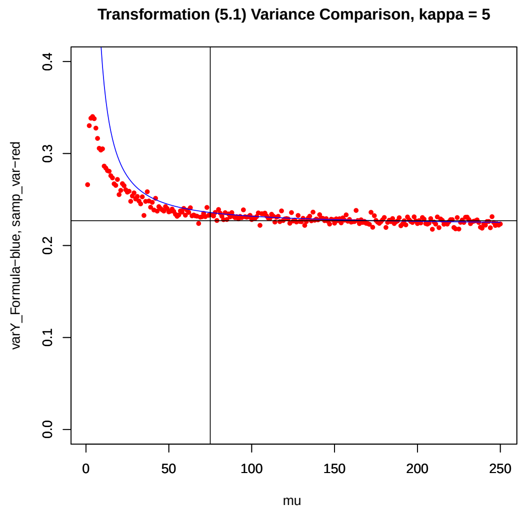


(b) Comparison of variance approximation formula and sample variance for $Y = \ln(X+1)$ (see (5.1)) for $\kappa = 3$. The horizontal line is at $\mu = 38$.

Figure 5.9: Comparison of variance approximation formulae and sample variance for $NBi(\mu, 3)$.

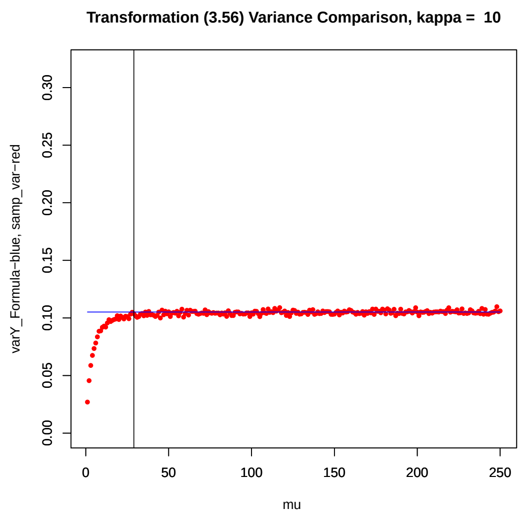


(a) Comparison of variance approximation formula and sample variance for $Y = 2\sinh^{-1}(\sqrt{(X+c)/(\kappa+d)})$ (see (3.56)) for $\kappa = 5$. The horizontal line is at $\mu = 19$.

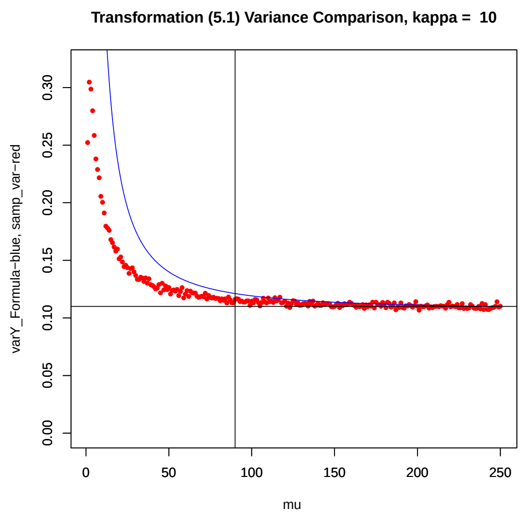


(b) Comparison of variance approximation formula and sample variance for $Y = \ln(X+1)$ (see (5.1)) for $\kappa = 5$. The horizontal line is at $\mu = 75$.

Figure 5.10: Comparison of variance approximation formulae and sample variance for $NBi(\mu, 5)$.

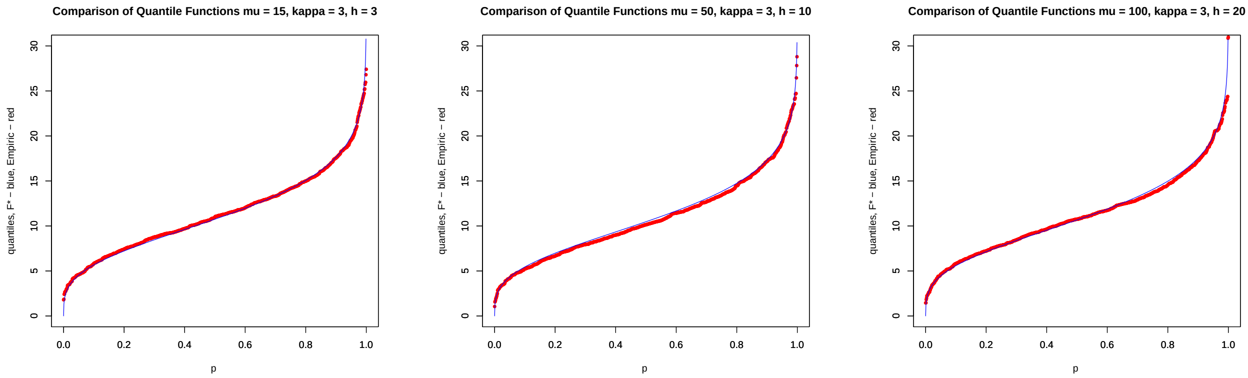


(a) Comparison of variance approximation formula and sample variance for $Y = 2\sinh^{-1}(\sqrt{(X+c)/(\kappa+d)})$ (see (3.56)) for $\kappa = 10$. The horizontal line is at $\mu = 29$.



(b) Comparison of variance approximation formula and sample variance for $Y = \ln(X+1)$ (see (5.1)) for $\kappa = 10$. The horizontal line is at $\mu = 90$.

Figure 5.11: Comparison of variance approximation formulae and sample variance for $NBi(\mu, 10)$.

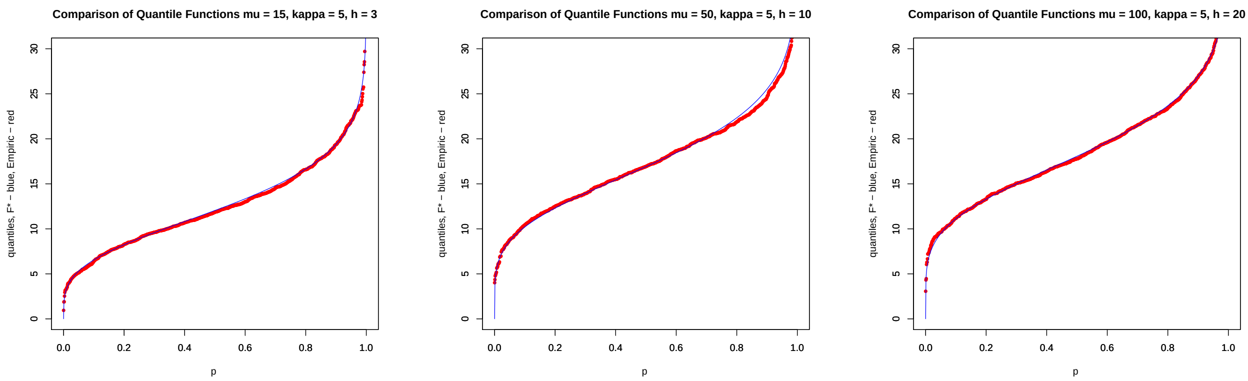


(a) Comparison of the quantile function of random variable F^* and empirical quantile function of the ratio(4.40) for $\mu = 15$, $\kappa = 3$, $h = 3$.

(b) Comparison of the quantile function of random variable F^* and empirical quantile function of the ratio (4.40) for $\mu = 50$, $\kappa = 3$, $h = 10$.

(c) Comparison of the quantile function of random variable F^* and empirical quantile function of the ratio (4.40) for $\mu = 100$, $\kappa = 3$, $h = 20$.

Figure 5.12: Comparison of the quantile function of random variable F^* and empirical quantile function of the ratio (4.40) for $NBi(\mu, 3)$.

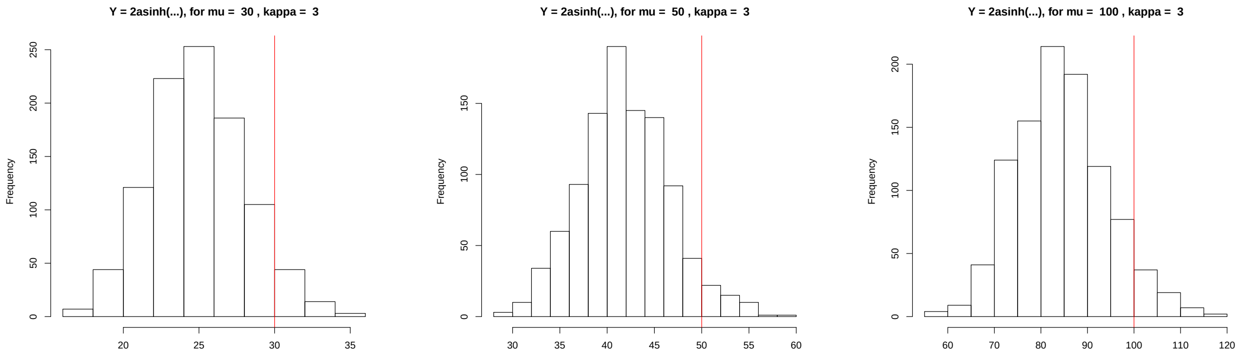


(a) Comparison of the quantile function of random variable F^* and empirical quantile function of the ratio(4.40) for $\mu = 15$, $\kappa = 5$, $h = 3$.

(b) Comparison of the quantile function of random variable F^* and empirical quantile function of the ratio (4.40) for $\mu = 50$, $\kappa = 5$, $h = 10$.

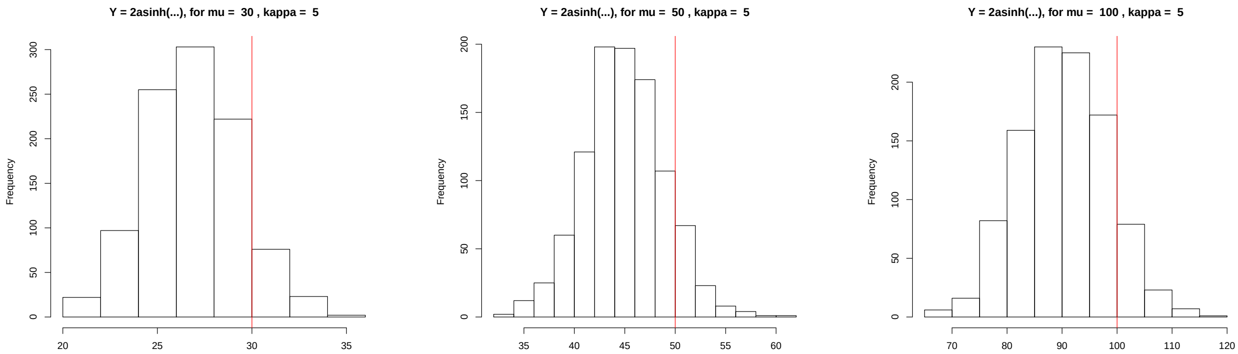
(c) Comparison of the quantile function of random variable F^* and empirical quantile function of the ratio (4.40) for $\mu = 100$, $\kappa = 5$, $h = 20$.

Figure 5.13: Comparison of the quantile function of random variable F^* and empirical quantile function of the ratio (4.40) for $NBi(\mu, 5)$.



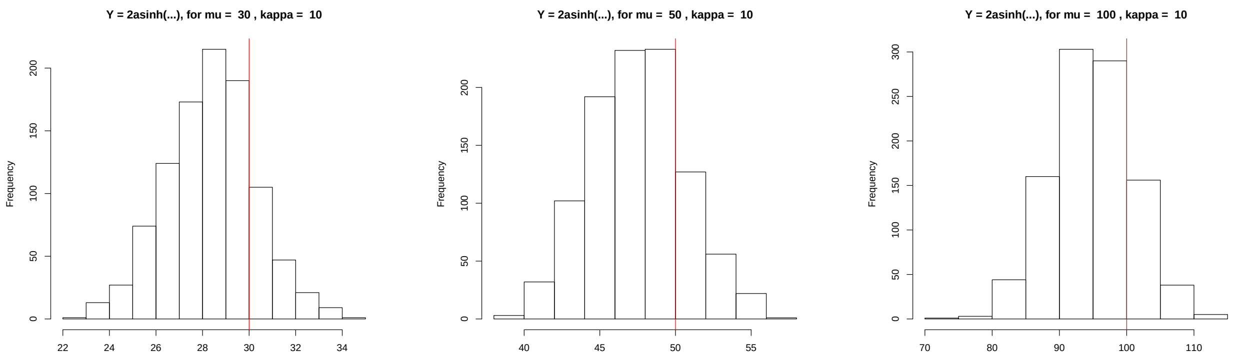
(a) Histogram of μ_Y for $\mu = 30$ (b) Histogram of μ_Y for $\mu = 50$ (c) Histogram of μ_Y for $\mu = 100$

Figure 5.14: Histograms of μ_Y for transformation $Y = 2\sinh^{-1}(\sqrt{(X+c)/(\kappa+d)})$ for $\kappa = 3$, sample size $n = 100$, and number of repetitions $k = 1000$



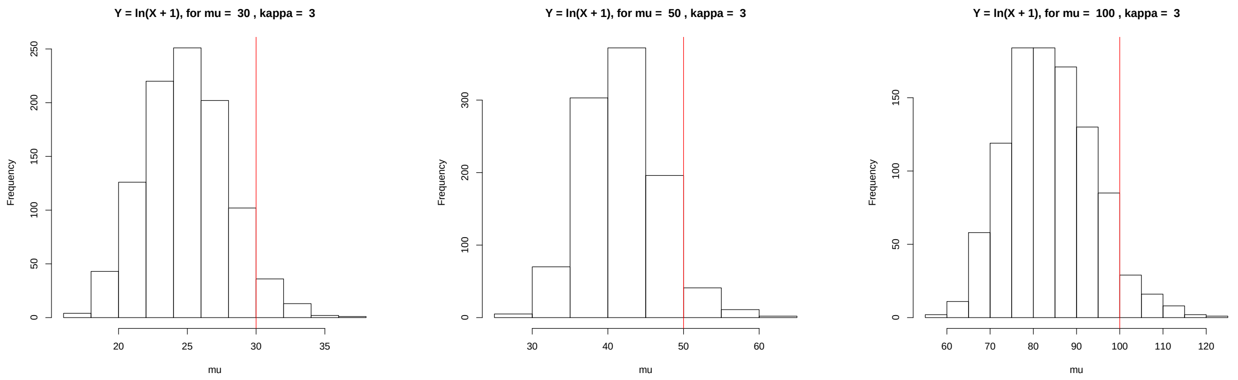
(a) Histogram of μ_Y for $\mu = 30$ (b) Histogram of μ_Y for $\mu = 50$ (c) Histogram of μ_Y for $\mu = 100$

Figure 5.15: Histograms of μ_Y for transformation $Y = 2\sinh^{-1}(\sqrt{(X+c)/(\kappa+d)})$ for $\kappa = 5$, sample size $n = 100$, and number of repetitions $k = 1000$



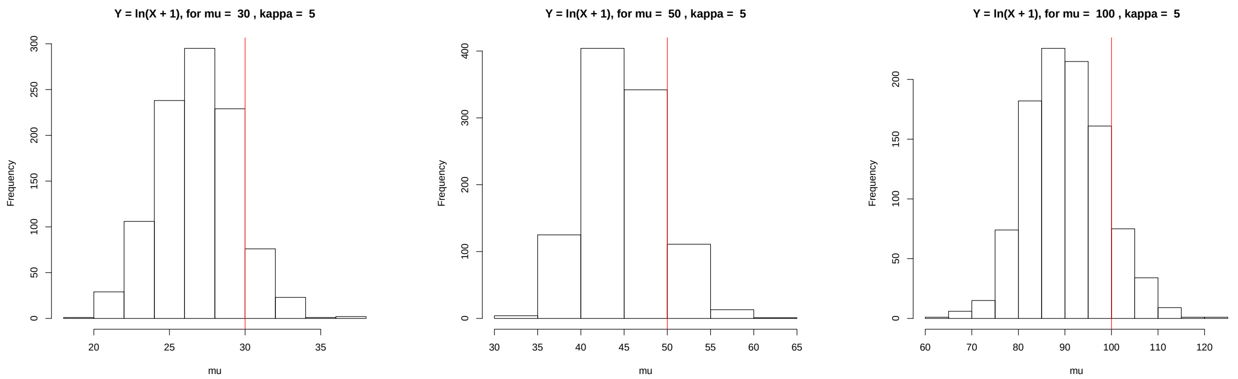
(a) Histogram of μ_Y for $\mu = 30$ (b) Histogram of μ_Y for $\mu = 50$ (c) Histogram of μ_Y for $\mu = 100$

Figure 5.16: Histograms of μ_Y for transformation $Y = 2\sinh^{-1}(\sqrt{(X+c)/(\kappa+d)})$ for $\kappa = 10$, sample size $n = 100$, and number of repetitions $k = 1000$



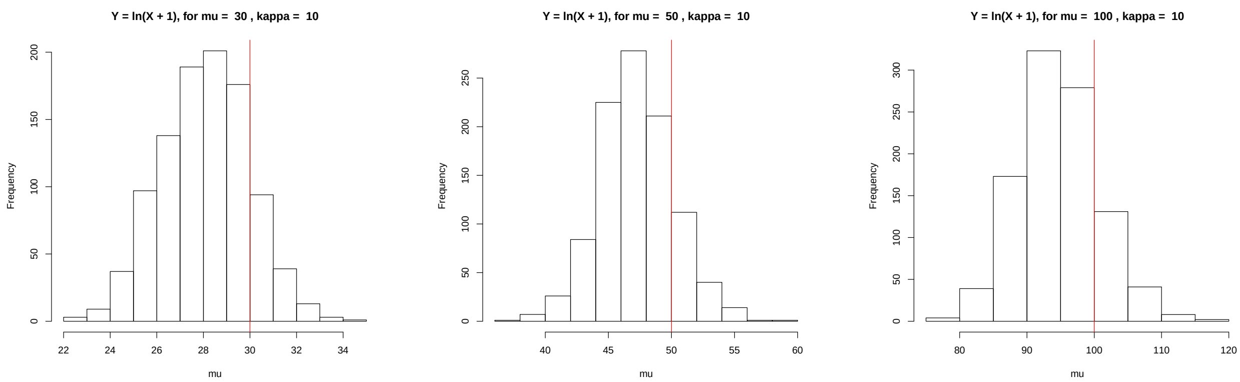
(a) Histogram of μ_Y for $\mu = 30$ (b) Histogram of μ_Y for $\mu = 50$ (c) Histogram of μ_Y for $\mu = 100$

Figure 5.17: Histograms of μ_Y for transformation $Y = \ln(X+1)$ for $\kappa = 3$, sample size $n = 100$, and number of repetitions $k = 1000$



(a) Histogram of μ_Y for $\mu = 30$ (b) Histogram of μ_Y for $\mu = 50$ (c) Histogram of μ_Y for $\mu = 100$

Figure 5.18: Histograms of μ_Y for transformation $Y = \ln(X+1)$ for $\kappa = 5$, sample size $n = 100$, and number of repetitions $k = 1000$



(a) Histogram of μ_Y for $\mu = 30$ (b) Histogram of μ_Y for $\mu = 50$ (c) Histogram of μ_Y for $\mu = 100$

Figure 5.19: Histograms of μ_Y for transformation $Y = \ln(X+1)$ for $\kappa = 10$, sample size $n = 100$, and number of repetitions $k = 1000$

5.4 Comparison of Power Functions by Simulation

In this section we will describe how the power functions of the F-test applied to test the hypothesis of equality of expectations (4.1) (see the beginning of the Chapter 4) of p random samples of the same size n are computed either by using formula (4.36), or via simulations. In the further text we will denote the power function computations based on the formula (4.36) as theoretical power functions. When the approach via simulations is used, we will refer to simulated power functions.

Let us start by describing how the simulated power functions are computed for either Poisson or negative binomial data. The simulated power functions are computed in the following way. We choose a step h and an initial value of expectation λ_1 for Poisson case or μ_1 for negative binomial case. Additionally for the negative binomial case we choose the value of the parameter κ. We generate p random samples $\mathbf{X}_i$ for $i = 1, ..., p$ of a size n from $Po(\lambda_i)$ for $i = 1, ..., p$ in the Poisson case, and from $NBi(\mu_i, \kappa)$ for $i = 1, ..., p$ in negative binomial case where, the values $\lambda_2, ..., \lambda_p$ or $\mu_2, ..., \mu_p$ for the negative binomial case are obtained from λ_1 or μ_1 by adding multiples of the step h. We transform the random samples using transformations (3.4), (5.1), and (2.9) for the Poisson case and transformations (3.56), (5.1), and (2.9) for the negative binomial case obtaining transformed random samples $\mathbf{Y}_i$ for $i = 1, ..., p$. The best value of the parameter of Yeo-Johnson transformation is determined via maximum likelihood method. Using formula (4.6) we compute the F statistic, where for input vector $\mathbf{Y}_{np} = (Y_{11}, ..., Y_{1n}, ..., Y_{p1}, ..., Y_{pn})^T$ we take the transformed samples $\mathbf{Y}_i = (Y_{i1}, ..., Y_{in})^T$ for $i = 1, ..., p$ stacked one above each other. We compare the value of the F statistic with quantile $Q_F(1 - \alpha, p - 1, p(n - 1))$ and decide about the result of the test. We repeat this process k times for the same setting of parameters and compute the relative frequency of rejecting hypothesis H_0. By increasing the value of step h and repeating the described procedure we obtain values of the simulated power function all across the interval $[0, 1]$.

5.4.1 Computation of Theoretical Power Function under Asumption of Equal Variances

In this subsection we will describe how the theoretical power function of the F test applied to test the hypothesis of equality of expectations (4.1), (see the beginning of the Chapter 4) of p random samples $\mathbf{X}_1, ..., \mathbf{X}_p$ of a size n of either Poisson or negative binomial distribution transformed via (3.4) or (3.56) is computed. The computation is based on the formulae and results provided in Chapter 3 and Section 4.1, and is done as follows. As in Section 5.3 the general procedure is identical for both Poisson and negative binomial cases, save the applied transformations. The description will be provided for Poisson distribution, while the corresponding data for negative binomial distribution will be provided in brackets following the data for Poisson distribution. Let $\mathbf{X}_1, ..., \mathbf{X}_p$ be p samples of a size n from a distribution $Po(\lambda_i)$ $(NBi(\mu_i, \kappa))$, $i = 1, ..., p$. We choose the value λ_1 (μ_1) and the step h. The values λ_i (μ_i) for $i = 2, ..., p$ differ from the value λ_1 (μ_1) by a multiples of the step.

We compute the numerical characteristics of random variables Y_{ij} $i = 1, ..., p$, $j = 1, ..., n$ of the random samples $\mathbf{Y}_1, ..., \mathbf{Y}_p$ obtained by applying the transformation (3.4) ((3.56)) on $\mathbf{X}_i$ for $i = 1, ..., p$ by formulae (3.20) and (3.26) ((3.108) and (3.96)). We have seen in the Subsection 4.1.3, Proposition 4.17 that, given that the assumption of equal variances holds, the power function $\beta_\alpha(\boldsymbol{\theta})$ on the significance level α is given by formula

$$\beta_\alpha(\boldsymbol{\theta}) = 1 - \mathcal{F}_{F_\delta}(Q_F(1-\alpha; s, t), s, t, \delta)), \tag{5.27}$$

where, $\boldsymbol{\theta}$ is the p-dimensional vector of expectations, $s = p-1$ and $t = p(n-1)$ are the degrees of freedom of the F distributed random variable under the hypothesis H_0 of equality of expectations and δ is the noncentrality parameter given by

$$\delta = \frac{n}{\sigma^2}\boldsymbol{\theta}^T \mathbf{H}_{p\times p}\boldsymbol{\theta}. \tag{5.28}$$

For the values of the entries of $\boldsymbol{\theta}$ we take the expectation approximations $\mathbf{E}Y_{ij}$ of the transformed random variables Y_{ij}, for $i = 1, ..., p$, and j fixed, obtained via (3.20) ((3.108)). Note that $\mathbf{E}Y_{ij} = \mathbf{E}Y_{il}$ for each $j, l \in \{1, ..., n\}$, so it does not matter which j we pick. For the value of σ^2 we take the variance approximation $\mathbf{var}Y_{ij}$ of the transformed random variables Y_{ij}, for i and j fixed, given by (3.26) ((3.96)). This is because $\mathbf{var}Y_{ij} = \mathbf{var}Y_{il}$ for each $j, l \in \{1, ..., n\}$, and by Remark 5.8 (Remark 5.13) also $\mathbf{var}Y_{ij} = \mathbf{var}Y_{kj}$ for each $i, k \in \{1, ..., p\}$ up to a term that we decide to neglect.

We start the computation by setting $h = 0$ and hence, $\theta_1 = ... = \theta_p$. We proceed to increase the difference between the expectations in order to obtain a range of values of $\beta_\alpha(\boldsymbol{\theta})$ from the whole interval of values $[\alpha, 1]$. The precise description of how this is done is provided in Subsection 5.4.3 (Subsection 5.4.4).

5.4.2 Computation of Theoretical Power Function Under Assumption of Unequal Variances

In this subsection we will describe how the theoretical power function of the F test applied to test the hypothesis of equality of expectations (4.1) (see the beginning of the Chapter 4) of p random samples $\mathbf{X}_1, ..., \mathbf{X}_p$ of a size n of either Poisson or negative binomial distribution transformed via for the transformation (5.1) is computed. the general procedure is identical for both Poisson and negative binomially distributed random variable samples, save the applied transformations. The description will be provided for Poisson distribution, while the corresponding data for negative binomial distribution will be provided in brackets following the data for Poisson distribution. The situation is a bit more complicated in this case because of the assumption of the equality of variances among the samples is violated, and is based mainly on the results developed in Section 4.2. Let $\mathbf{X}_1, ..., \mathbf{X}_p$ be p samples of a size n from distribution $Po(\lambda_i)$ $(NBi(\mu_i, \kappa))$, $i = 1, ..., p$. We again choose the value λ_1 (μ_1) and the step h. The values λ_i (μ_i) for $i = 2, ..., p$ differ from the value λ_1 (μ_1) by a multiples of the step. The precise description of how this is done in the computations is provided in Subsection 5.4.3 (Subsection 5.4.4).

Let $\mathbf{Y}_i$ $i = 1, ..., p$ be the random samples obtained by applying the transformation (5.1) on independent samples $\mathbf{X}_i$ for $i = 1, ..., p$ from distribution $Po(\lambda_i)$ $(NBi(\mu_i, \kappa))$. The numerical characteristics of the transformed random variables Y_{ij} are computed by approximation formulae (5.17), and (5.20). To compute the power function we need to determine the degrees of freedom of the F statistic under the hypothesis H_0. It can be easily seen, that the degrees of freedom of the F statistic under the hypothesis H_0, given the assumption that the variances are not equal, in the model we are assuming, where the variance approximation is a function of the expectation, is again $s = p - 1$, $t = p(n-1)$. This is summed up in the following Lemma.

Lemma 5.14. *Assume that we have p independent samples $\boldsymbol{X}_1, ..., \boldsymbol{X}_p$ of a size n from distributions $\mathcal{L}(\theta_1), ..., \mathcal{L}(\theta_p)$ (either Poisson or negative binomial) in sequence, where the parameters θ_i are chosen in such a way that for each $i = 1, ..., p$ and $j = 1, ..., n$, $\mathbf{E}X_{ij} = \theta_i$, Furthermore assume that the null hypothesis* (4.1) *of the equality of expectations holds, i. e. $\theta_1 = ... = \theta_p = \theta$. Let us denote $\boldsymbol{Y}_1 = (Y_{11}, ..., Y_{1n}), ..., \boldsymbol{Y}_p = (Y_{p1}, ..., Y_{pn})$ the p random samples obtained from random samples $\boldsymbol{X}_1, ..., \boldsymbol{X}_p$ by applying transformation* (5.1). *Let σ_{ij}^2 variances of random variables Y_{ij} for $i = 1, ..., p$, $j = 1, ..., n$. Then the distribution of the test statistics F given by* (4.40) *may be modelled by*

$$F \sim F_{p-1,p(n-1)}. \tag{5.29}$$

Proof. Obviously $\sigma_{ij}^2 = \sigma_{il}^2$ for each $j, l \in \{1, ..., n\}$. Let us denote $\boldsymbol{\sigma}^2 = (\sigma_1^2, ..., \sigma_p^2)$, where for any fixed $i \in \{1, ..., n\}$ we have $\sigma_i^2 = \sigma_{ij}^2$ for any $j = 1, ..., n$. Let us recall that through this chapter we consider the following model $\boldsymbol{\sigma}^2 = (\sigma_1^2, ..., \sigma_p^2) = (\sigma^2(\theta_1), ..., \sigma^2(\theta_p))$, where by abuse of notation we will for the sake of the proof identify the variances σ_i^2 with their respective approximations (5.12) or (5.20) depending on whether are we assuming the Poisson or the negative binomial case. I. e. all the variability in the variances among the p random samples is caused by the difference of expectations θ_i among the samples. Given that the hypothesis H_0 holds, this implies, that

$$\sigma_1^2 = ... = \sigma_p^2 = \sigma^2. \tag{5.30}$$

Let us first focus on K_1'. Applying this result to the matrix $\mathbf{S}^T\mathbf{S}$ (see Proposition 4.30) we obtain

$$\mathbf{S}^T\mathbf{S} = diag(\boldsymbol{\sigma})\mathbf{H}diag(\boldsymbol{\sigma}) = \sigma^2\mathbf{H}, \tag{5.31}$$

and since we have that for any matrix $\mathbf{A}$ with a nonzero eigenvector $\mathbf{x}$, a corresponding eigenvalue λ and a nonzero real α

$$(\alpha\mathbf{A})\mathbf{x} = \alpha(\mathbf{A}\mathbf{x}) = \alpha(\lambda\mathbf{x}) = (\alpha\lambda)\mathbf{x} \tag{5.32}$$

and so $\mathbf{x}$ is an eigenvector of $\alpha\mathbf{A}$ for the eigenvalue $\alpha\lambda$, and the matrix $\mathbf{H}$ is symmetric and idempotent of a rank $p-1$ (see Lemma 4.4) and therefore $p-1$ of its eigenvalues are equal to one and one to zero (see Proposition 1.19), the matrix $\mathbf{S}^T\mathbf{S}$ has $p-1$ eigenvalues equal to σ^2 and one eigenvalue equal to 0. Using this result with the one presented in Corollary 4.33 we obtain

$$K_1' = \sigma^2(V_1^2 + ... + V_{p-1}^2), \tag{5.33}$$

where the V_i are independent, identically distributed with variance equal to one. Moreover we have seen in the second step of the proof of the Theorem 4.36 that the expectation of each V_i is equal to zero, hence,

$$K_1' = \sigma^2 X_1, \tag{5.34}$$

where $X_1 \sim \chi_{p-1}^2$ (see Proposition 1.35).

A simpler situation is with K_2'. By Corollary 4.27, namely equation (4.68) and the result (5.30) we have that

$$K_2' = \sigma^2(V_1^2 + ... + V_{p(n-1)}^2), \tag{5.35}$$

where $V_i^2 \sim N(0, 1)$ are independent identically distributed random variables, and hence, due to Proposition 1.35 we can rewrite (5.35) in the form

$$K_2' = \sigma^2 X_2 \tag{5.36}$$

where $X_2 \sim \chi^2_{p(n-1)}$. Hence, combining the results (5.36) and (5.36) we get

$$F = \frac{p(n-1)}{p-1}\frac{X_1}{X_2} \tag{5.37}$$

It can be already seen from definition of the F distribution, that $F \sim F_{p-1,p(n-1)}$ (see [8]), moreover if we apply the approximation given by Theorem 4.35, we again obtain for this special case of ratio of two χ^2 distributed random variables that $\frac{\lambda_1^*\nu_1^*}{\lambda_2^*\nu_2^*}F^* \sim F_{p-1,p(n-1)}$ (see [8]). □

In the following statement we will introduce the power function of the F test applied to test the hypothesis of equality of expectations (4.1) among the p samples of a size n from Poisson or negative binomial distribution transformed via (5.1).

Theorem 5.15. *Let us denote $\boldsymbol{Y}_1 = (Y_{11}, ..., Y_{1n}), ..., \boldsymbol{Y}_p = (Y_{p1}, ..., Y_{pn})$ the p random samples obtained by applying transformation (5.1) to independent random samples $\boldsymbol{X}_1 = (X_{11}, ..., X_{1n}), ..., \boldsymbol{X}_p = (X_{p1}, ..., X_{pn})$ of a size n from given distributions $\mathcal{L}(\theta_i)$, $i = 1, ..., p$ (either Poisson or negative binomial) where the parameters θ_i are chosen in such a way that $\mathbf{E}X_i = \theta_i$ for $i = 1, ..., p$. Let $\sigma^2_{11}, ..., \sigma^2_{1n}, ..., \sigma^2_{p1}, ..., \sigma^2_{pn}$ be variances of random variables Y_{ij}, for $i = 1, ..., p$ and $j = 1, ..., n$. Let the F test statistic be given by (4.40) Then the power function of the F test on the level of significance α may be approximated by*

$$\beta_\alpha(\boldsymbol{\theta}) = 1 - \mathcal{F}_{F_\delta}(r^*; \nu_1^*, \nu_2^*, \delta_1^*), \tag{5.38}$$

where $\mathcal{F}_{F_\delta}$ is the distribution function of F^*, given by Theorem 4.35 and the constants $\nu_1^*, \nu_2^*, \delta_1^*$ and the parameter $r^* = \frac{\lambda_2^*\nu_2^*}{\lambda_1^*\nu_1^*}r$ are determined by Lemma 4.34 and Theorem 4.35. The value of r is given by

$$r = Q_F(1-\alpha; p-1, n(p-1)). \tag{5.39}$$

Proof. Obviously $\sigma^2_{ij} = \sigma^2_{il}$ for each $j, l \in \{1, ..., n\}$. Let us denote $\boldsymbol{\sigma}^2 = (\sigma_1^2, ..., \sigma_p^2)$, where for any fixed $i \in \{1, ..., n\}$ we have $\sigma_i^2 = \sigma_{ij}^2$ for any $j = 1, ..., n$. Let us recall that through this chapter we consider the following model $\boldsymbol{\sigma^2} = (\sigma_1^2, ..., \sigma_p^2) = (\sigma^2(\theta_1), ..., \sigma^2(\theta_p))$, where by abuse of notation we will for the sake of the proof identify the variances σ_i^2 with their respective approximations (5.12) or (5.20) depending on whether are we assuming the Poisson or the negative binomial case. I. e. all the variability in the variances among the p random samples is caused by the difference of expectations θ_i among the samples. Due to Proposition 4.17 the power function $\beta_\alpha(\boldsymbol{\theta})$ is in general case given by formula

$$\beta_\alpha(\boldsymbol{\theta}) = 1 - \mathcal{F}_{F_\delta}(Q_F(1-\alpha; s, t), \nu_1, \nu_2, \delta), \tag{5.40}$$

where $\mathcal{F}_{F_\delta}$ is the distribution function of the test statistic under the alternative (4.2), and the parameters s and t are the degrees of freedom of the test statistic under the null hypothesis (4.1). In Lemma 5.14 we have determined that $s = p-1$ and $t = p(n-1)$.

Due to Corollary 5.6 for Poisson case and Corollary 5.11 for negative binomial case if H_1 given by (4.2) holds there exist $i, k \in \{1, ..., p\}$, $i \neq k$ such that $\sigma_i^2 \neq \sigma_k^2$, therefore the assumption of equal variances is violated. Due to Theorem 4.36 we may approximate the F statistic by F^* according to Theorem 4.35, but in order to do so we need to compute the new value of quantile r^* corresponding to the quantile $r = Q_F(1-\alpha; p-1, n(p-1))$ of the original test statistic F which concludes the proof. □

The practical computation of the the power function given by (5.38) is done in the following way. The numerical characteristics of the transformed random variables Y_{ij} $i = 1, ..., p$, $j = 1, ...n$ introduced at the beginning of this subsection are used to compute the vector $\boldsymbol{\mu}$ and the matrix $\mathbf{S}^T\mathbf{S}$ (see Proposition 4.30). The eigenvalues λ_i and eigenvectors $\mathbf{p}_i$ for $i = 1, ..., p$ of $\mathbf{S}^T\mathbf{S}$ are found. The nonzero eigenvalues of $\mathbf{S}^T\mathbf{S}$ are the coefficients of the linear combination K_1'. The degrees of freedom of the independent χ^2 distributed random variables of the linear combination K_1' are all equal to one. (see Corollary 4.33). Using the eigenvectors of $\mathbf{S}^T\mathbf{S}$ and the vector $\boldsymbol{\mu}$ the noncentrality parameters of the independent χ^2 distributed random variables of the linear combination K_1' are determined via formula $\delta_i = (\mathbf{p}_i^T\boldsymbol{\mu})^2$ for $i = 1, ..., p$ (see proof of Corollary 4.33). Due to Corollary 4.27 we know that the coefficients of the linear combination K_2' are actually variances σ_i^2 of the transformed random variables Y_{ij} for $i = 1, ..., p$ $j = 1, ..., n$. In the computations we will use the approximations given by (5.12) for Poisson case and (5.20) for negative binomial case. From the same Corollary we also obtain that the degrees of freedom of each central χ^2 distributed random variable in the linear combination K_2' are equal to $n - 1$. From these data we, by applying Lemma 4.34 and Theorem 4.35 compute the values of the coefficients λ_1^*, λ_2^* (see (4.95)), the degrees of freedom ν_1^*, ν_2^* (see (4.96)), and the noncentrality parameter δ_1^* (see (4.97)) of the approximation F^*. Finally also the value of quantile r^* of the approximation, corresponding to the quantile r is determined by the formula introduced in the Theorem 4.35 where for r we take (5.39). As in the case of equal variances (see Subsection 5.4.1) obtaining values of the power function $\beta_\alpha(\boldsymbol{\mu})$ all across the interval $[\alpha, 1]$ is done by increasing the step h.

5.4.3 Power Functions for Case of Poisson Distribution

The computations in the work were done for $p = 3$, $n = 100$, the values of λ_1 were set to $5, 10, 20, 50$. The values of $\lambda_j,\ \ j = 2, 3$ were computed as follows

$$\lambda_j = \lambda_1 + (-1)^j(kh_0)\ \ k = 0, ..., 30, \tag{5.41}$$

so that the difference $\Delta\mu_j = |\mu_1 - \mu_j|$ for $j = 2, 3$ increases with the value of k. The value of the step h and the values of k were picked in such a way that we would obtain values of the power function from the whole interval of values $[0, 1]$ and also keep the computations time-feasible.

Additionally to the power functions of the F test when either of the two transformations (3.4) and (5.1) was applied, a power function of the F test when the Yeo-Johnson transformation (see Section 2.3) was applied was obtained via simulations. The parameter of the Yeo-Johnson transformation is traditionally denoted by λ, to avoid confusion with the parameter of the Poisson distribution but keep the tradition we will denote it λ_{YJ}. The best value of the parameter λ_{YJ} was estimated via by maximum likelihood method for each setting of λ_1 by applying the method on a sample from $Po(\lambda_1)$ of a size $n = 100$.

For all assumed values of λ both the theoretical and the simulated power function of the F test applied to a sample transformed via transformation (3.4) scored better than the theoretical and simulated power function of the F test applied to a sample transformed via transformation (5.1). The simulated power functions attain values close to their respective theoretical counterparts.

The simulated power function of the F test applied to a sample transformed via Yeo-Johnson transformation scores similarly to simulated and theoretical power function of the F test applied to a sample transformed via transformation (3.4) and slightly better than the theoretical and simulated power function of the F test applied to a sample transformed via transformation (5.1). The closest resemblance of power functions of Yeo-Johnson case and transformation (3.4) case is for $\lambda = 5$ (see Figure 5.20). We may explain this by the fact that the value of the Yeo-Johnson transformation parameter is $\lambda_{YJ} = 0.45$, which is close to 0.5-the power of the square root.

For increasing values of λ all the power functions tend to attain increasingly similar values.

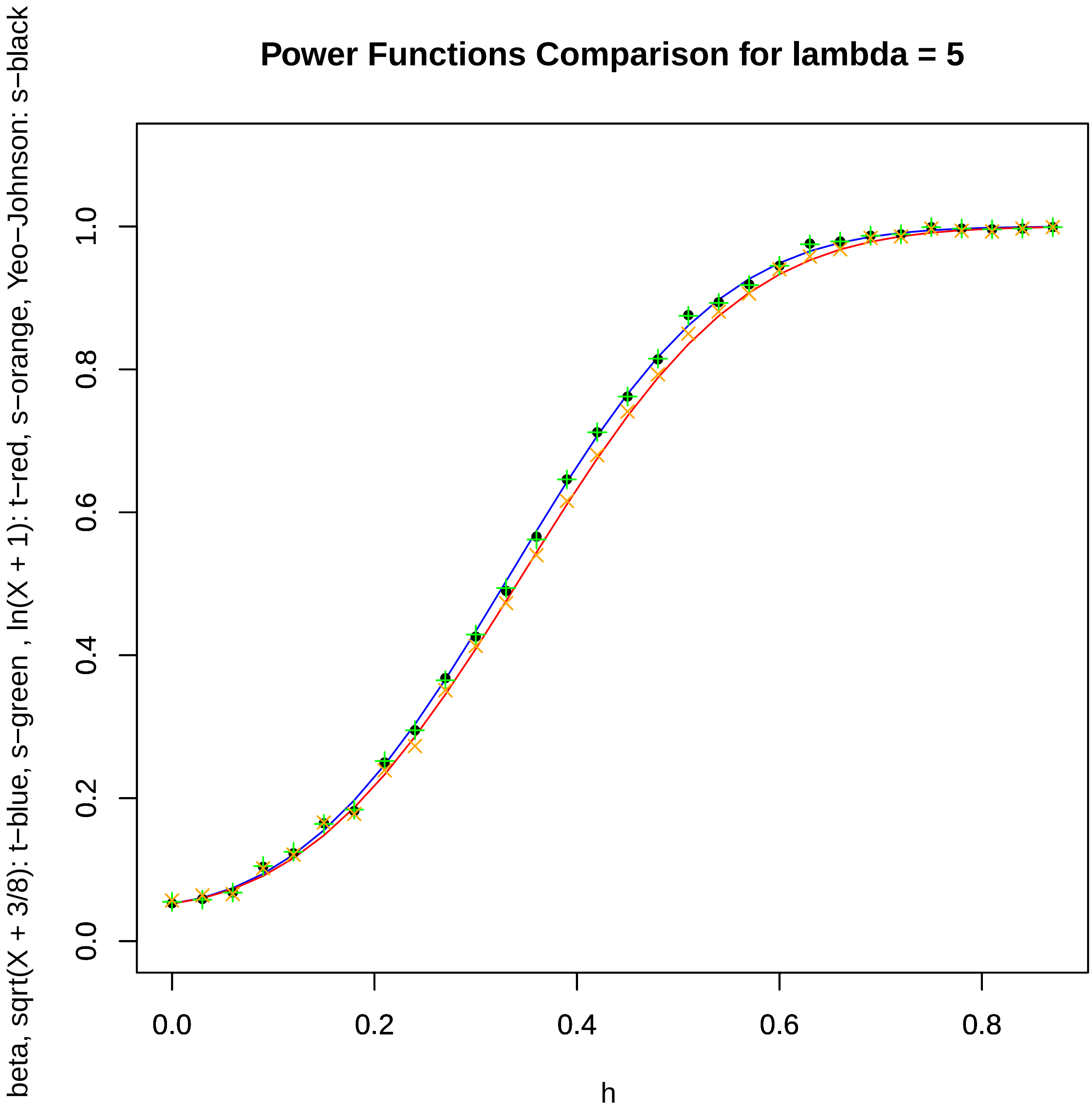


Figure 5.20: Comparison of the power functions. Transformation $Y = \sqrt{X + \frac{3}{8}}$ theoretical p. f.: blue line, simulated p.f.: green + symb., transformation $Y = \ln(X + 1)$ theoretical p. f.: red line, simulated p.f.: orange $\times$ symb., Yeo-Johnson transformation simulated p.f.: black dot. Computed for $\lambda_1 = 5$, and coefficient of Yeo Johnson transformation $\lambda_{YJ} = 0.45$.

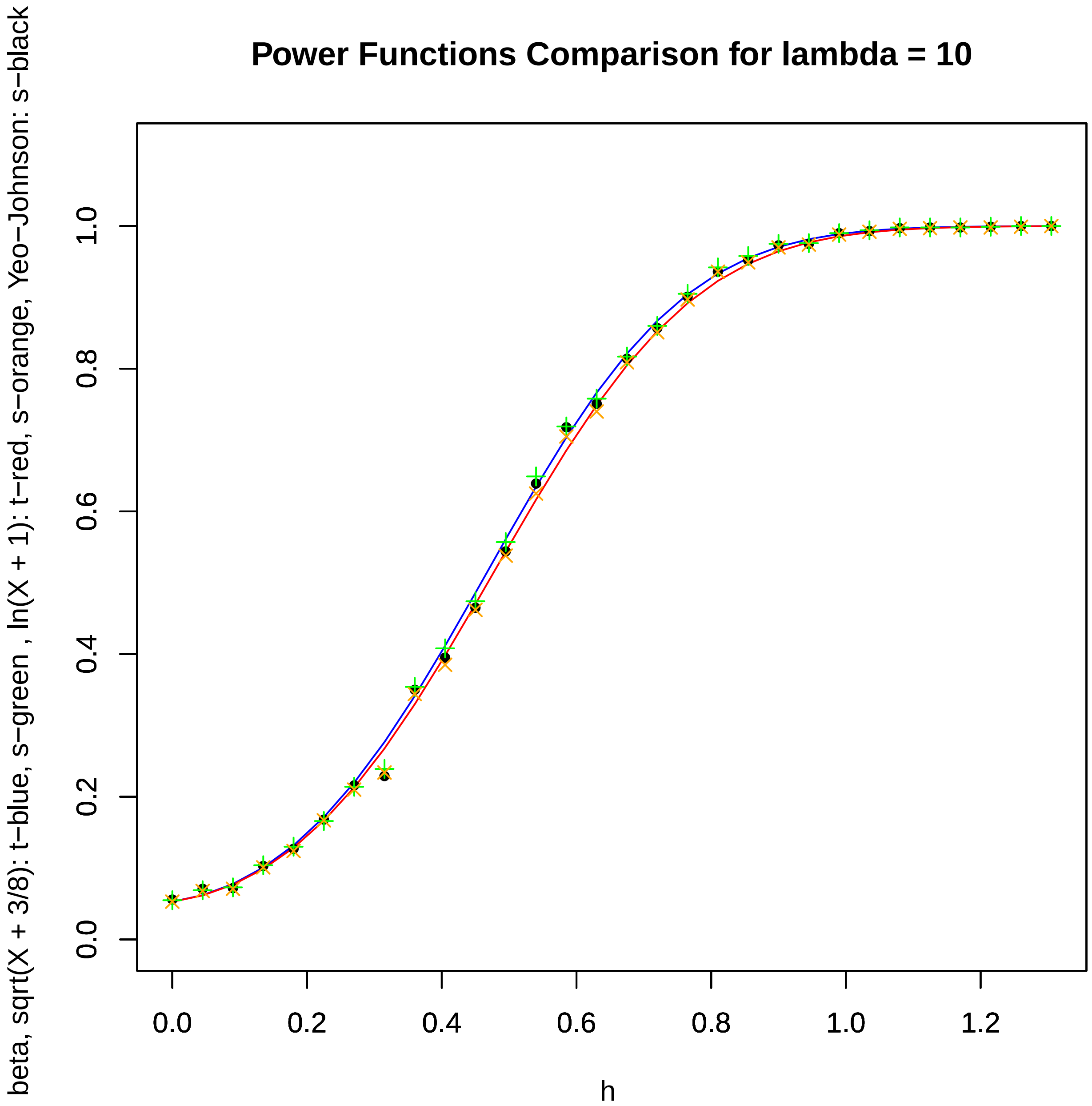


Figure 5.21: Comparison of the power functions. Transformation $Y = \sqrt{X + \frac{3}{8}}$ theoretical p. f.: blue line, simulated p.f.: green + symb., transformation $Y = \ln(X + 1)$ theoretical p. f.: red line, simulated p.f.: orange $\times$ symb., Yeo-Johnson transformation simulated p.f.: black dot. Computed for $\lambda_1 = 5$, and coefficient of Yeo Johnson transformation $\lambda_{YJ} = 0.2$.

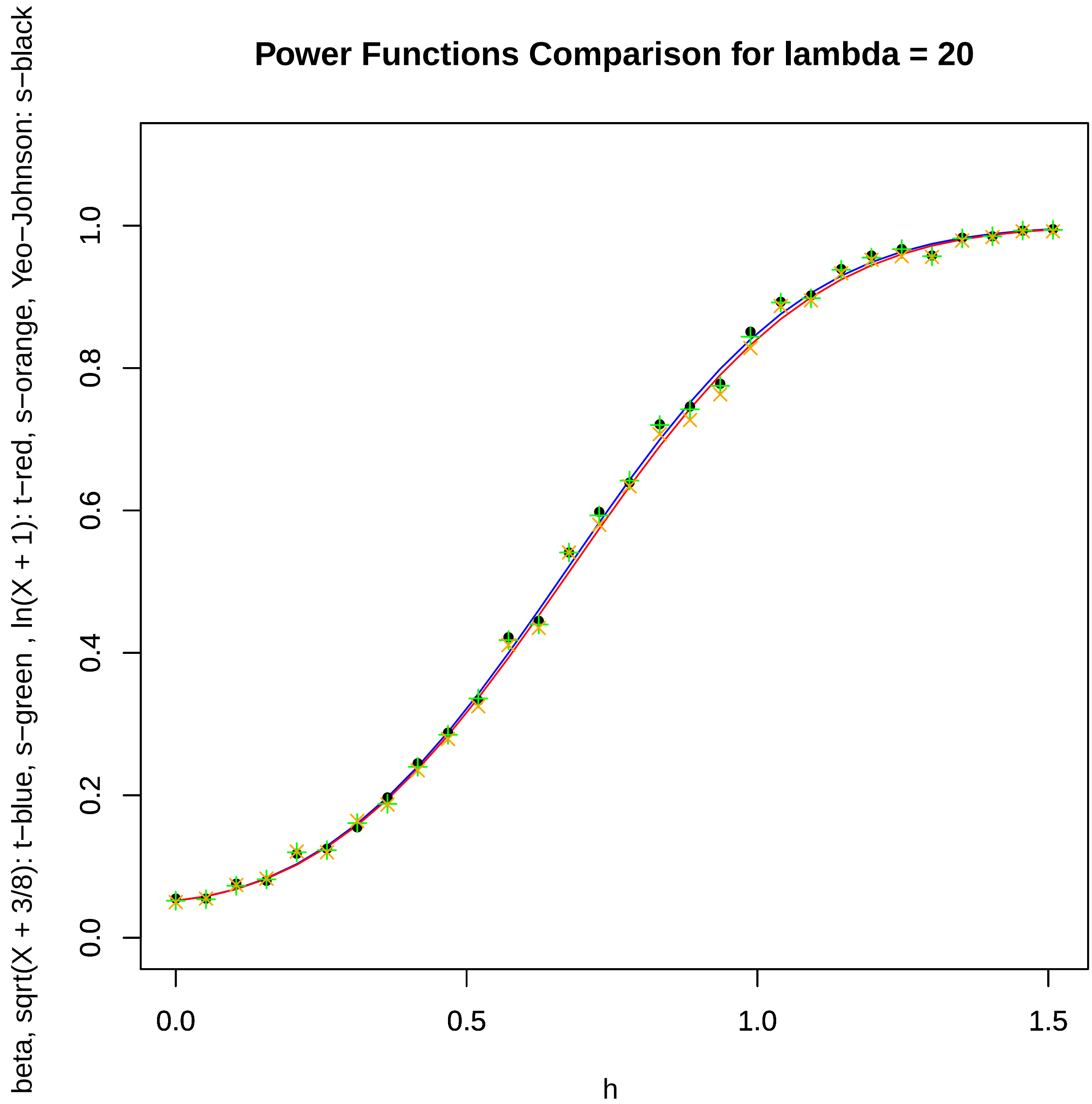


Figure 5.22: Comparison of the power functions. Transformation $Y = \sqrt{X + \frac{3}{8}}$ theoretical p. f.: blue line, simulated p.f.: green + symb., transformation $Y = \ln(X + 1)$ theoretical p. f.: red line, simulated p.f.: orange $\times$ symb., Yeo-Johnson transformation simulated p.f.: black dot. Computed for $\lambda_1 = 20$, and coefficient of Yeo Johnson transformation $\lambda_{YJ} = 0.95$.

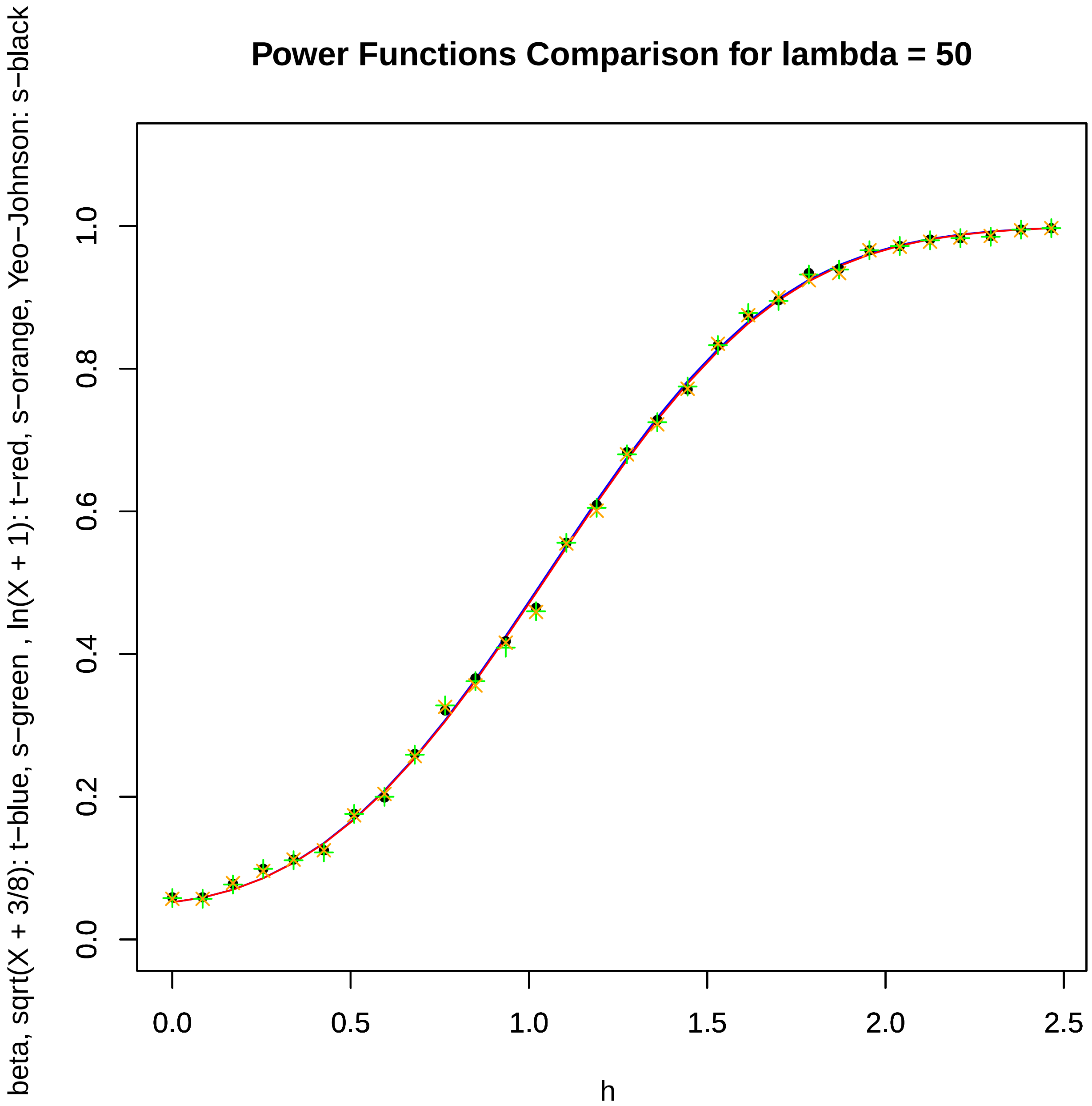


Figure 5.23: Comparison of the power functions. Transformation $Y = \sqrt{X + \frac{3}{8}}$ theoretical p. f.: blue line, simulated p.f.: green + symb., transformation $Y = \ln(X + 1)$ theoretical p. f.: red line, simulated p.f.: orange $\times$ symb., Yeo-Johnson transformation simulated p.f.: black dot. Computed for $\lambda_1 = 50$, and coefficient of Yeo Johnson transformation $\lambda_{YJ} = 0.9$.

5.4.4 Power Functions for Case of Negative Binomial Distribution

The computations in the work were done for $p = 3$, $n = 100$ the values of μ_1 were set to $30, 50$, and 100. The shape parameter κ was set to $\kappa = 3, 5$, and 10. The values of $\mu_j,\ \ j = 2, 3$ were computed as follows

$$\mu_j = \mu_1 + (-1)^j (kh)\ \ k = 0, ..., 30, \tag{5.42}$$

so that the difference $\Delta\mu_j = |\mu_1 - \mu_j|$ for $j = 2, 3$ increases with the value of k. The value of the step h and the values of k were picked in such a way that we would obtain values of the power function from the whole interval of values $[0, 1]$ and also keep the computations time-feasible

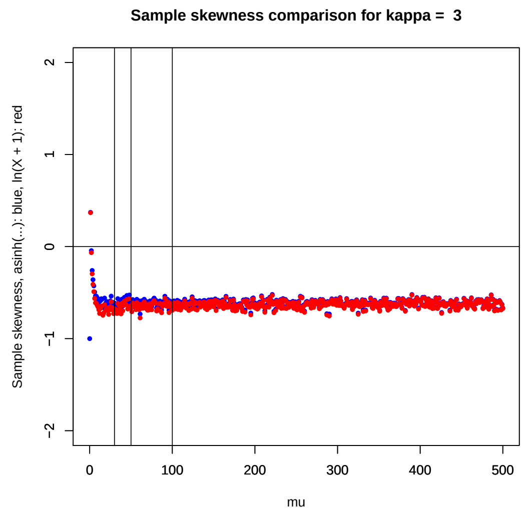
Additionally to power functions for the F test when either of the two transformations (3.56) (5.1) was applied, a power function of the F test when the Yeo-Johnson transformation (see Section 2.3) was applied, was computed via simulations. The value of the parameter λ_{YJ} of the Yeo-Johnson transformation was estimated via maximum likelihood method for each setting of μ_1 and κ, by applying the method on a sample from $NBi(\mu_1, \kappa)$ of a size $n = 100$. The values of λ_{YJ} for each setting may be found in the descriptions of each Figure.

Let us also additionally to the power functions comparison Figures add Figures of a sample skewness as a function of increasing parameter μ for values of $\kappa = 3, 5, 10$ to obtain a better insight on when a possible problem with normality may arise.

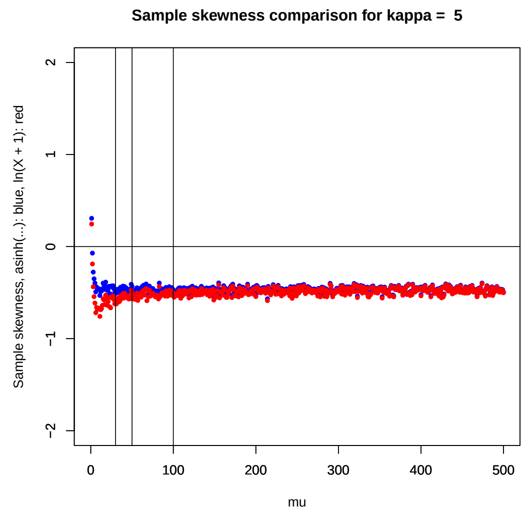
For the value of $\kappa = 3$ (see Figure 5.25) an interesting phenomena occurs for the value of $\mu_1 = 30$ where the theoretical power function of the F test applied to a random sample transformed via transformation (5.1) scores higher than the theoretical power function of the F test applied to a random sample transformed via transformation (3.56), and in most points even higher than both of the simulated power functions of F test applied to a random sample transformed via transformation (5.1) and (3.56), topped only by the power function of F test applied to a random sample transformed via Yeo-Johnson Transformation with $\lambda_{YJ} = 0.25$. Since this has not occurred for any other values of either κ or μ_1 we assume that the possible cause of this phenomena is rather the result of possible nonnormality of the transformed variables due to high absolute value of skewness (see Figure 5.24a) which is the highest of all for the values of κ considered and also because the variance approximation formula (5.20) neither does have a good fit nor "behaves well" for small values of μ_1 for $\kappa = 3$ (see Figure 5.9b).

Another pathological case occurs for smaller values of $\mu_1 = 30, 50$ for the largest assumed value of $\kappa = 10$ (see Figure 5.31 and 5.32) where the theoretical power function of the F test applied to a random sample transformed via transformation (5.1) scored significantly worse than both the corresponding simulated power function and all the other power functions. We explain this by the phenomena observed in the Subsection 5.3.2. We have seen that for increasing values of κ it takes larger values of μ for both sample variance and the variance obtained via the approximation to stabilise (see Figure 5.11b). The variance approximation (5.20) increases above all bounds rapidly as μ tends to 0 from the right hand side and does not become even close to stable for values of μ around $\mu = 50$.

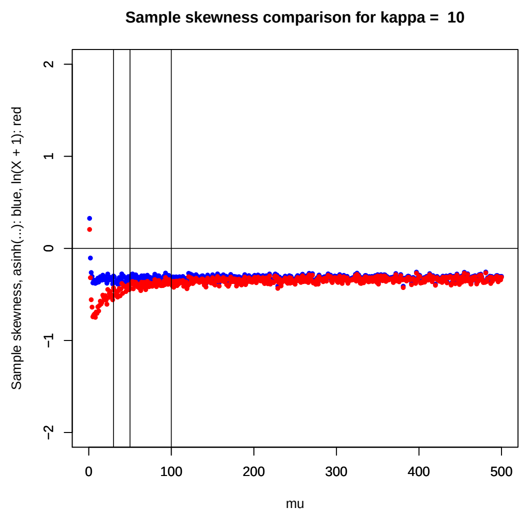
In general, save the pathological cases described above, the power function of the F test applied to a sample transformed via transformation (3.56) scores always a little better both for theoretical and simulated power functions. However the simulated power function of the F test applied to a sample transformed via Yeo-Johnson transformation scores the best in all cases but one. Indeed for $\mu_1 = 30$ and $\kappa = 5$ the simulated and theoretical power functions of the F test applied to a sample transformed via transformation (3.56) scores slightly better. That is caused by the fact that the parameter λ_{YJ} of the Yeo-Johnson transformation estimated via maximum likelihood estimation happened to be very close to 0 so the Yeo-Johnson transformation almost coincided with transformation (5.1).



(a) Comparison of sample of transformed samples for $\kappa = 3$.



(b) Comparison of skewness of transformed samples for $\kappa = 5$.



(c) Comparison of skewness of transformed samples for $\kappa = 10$.

Figure 5.24: Comparison of sample skewness of samples transformed via transformation $Y = 2\sinh^{-1}(\sqrt{(X+c)/(\kappa+d)})$ (blue) and transformation $Y = \ln(X+1)$ (red). The vertical lines are at $\mu = 30, 50, 100$.

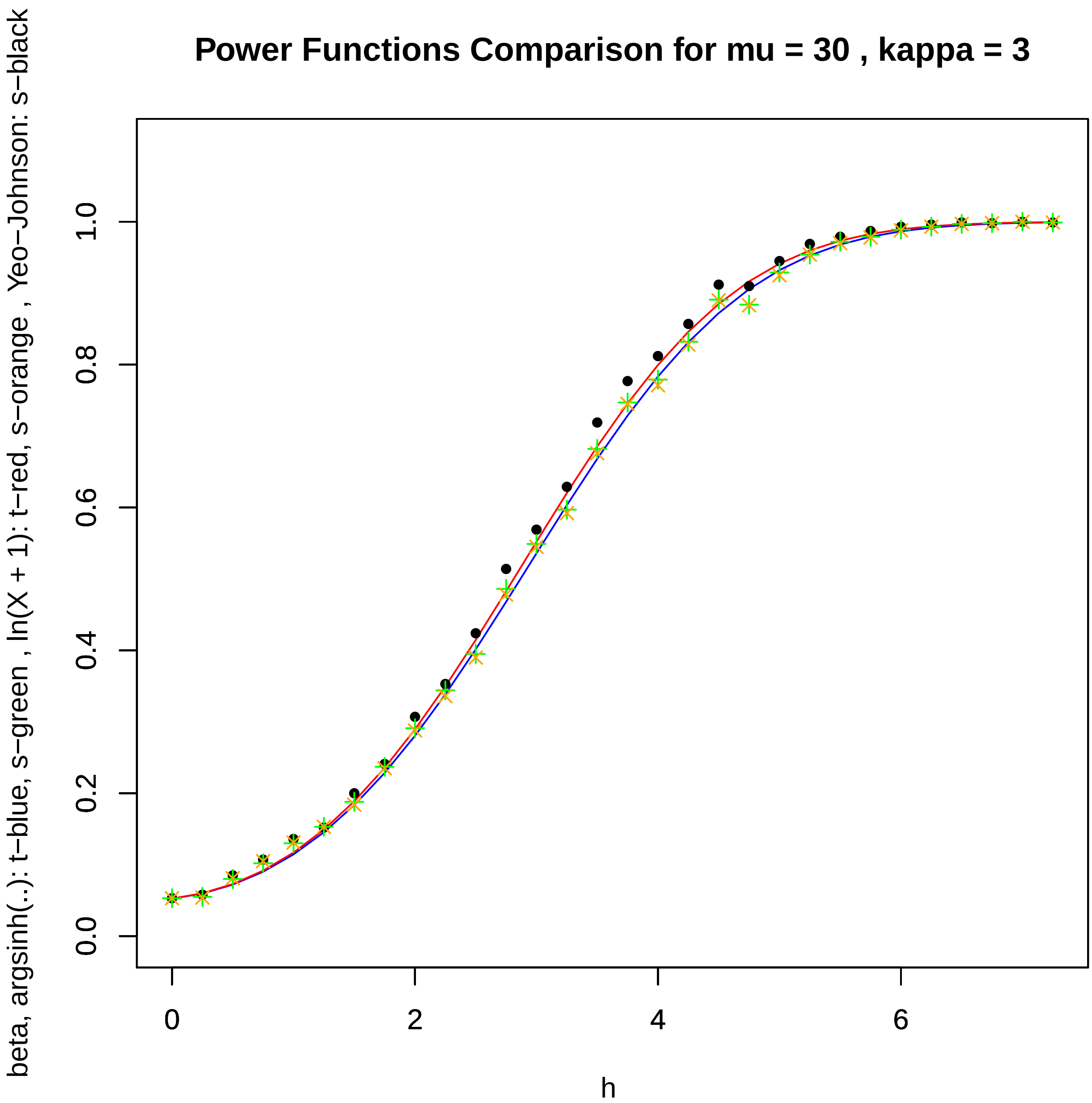


Figure 5.25: Comparison of the power functions. Transformation $Y = 2\sinh^{-1}(\sqrt{(X+c)/(\kappa+d)})$ theoretical p. f.: blue line, simulated p.f.: green + symb., transformation $Y = \ln(X+1)$ theoretical p. f.: red line, simulated p.f.: orange $\times$ symb., Yeo-Johnson transformation simulated p.f.: black dot. Computed for $\mu_1 = 30$, $\kappa = 3$, and coefficient of Yeo Johnson transformation $\lambda_{YJ} = 0.25$.

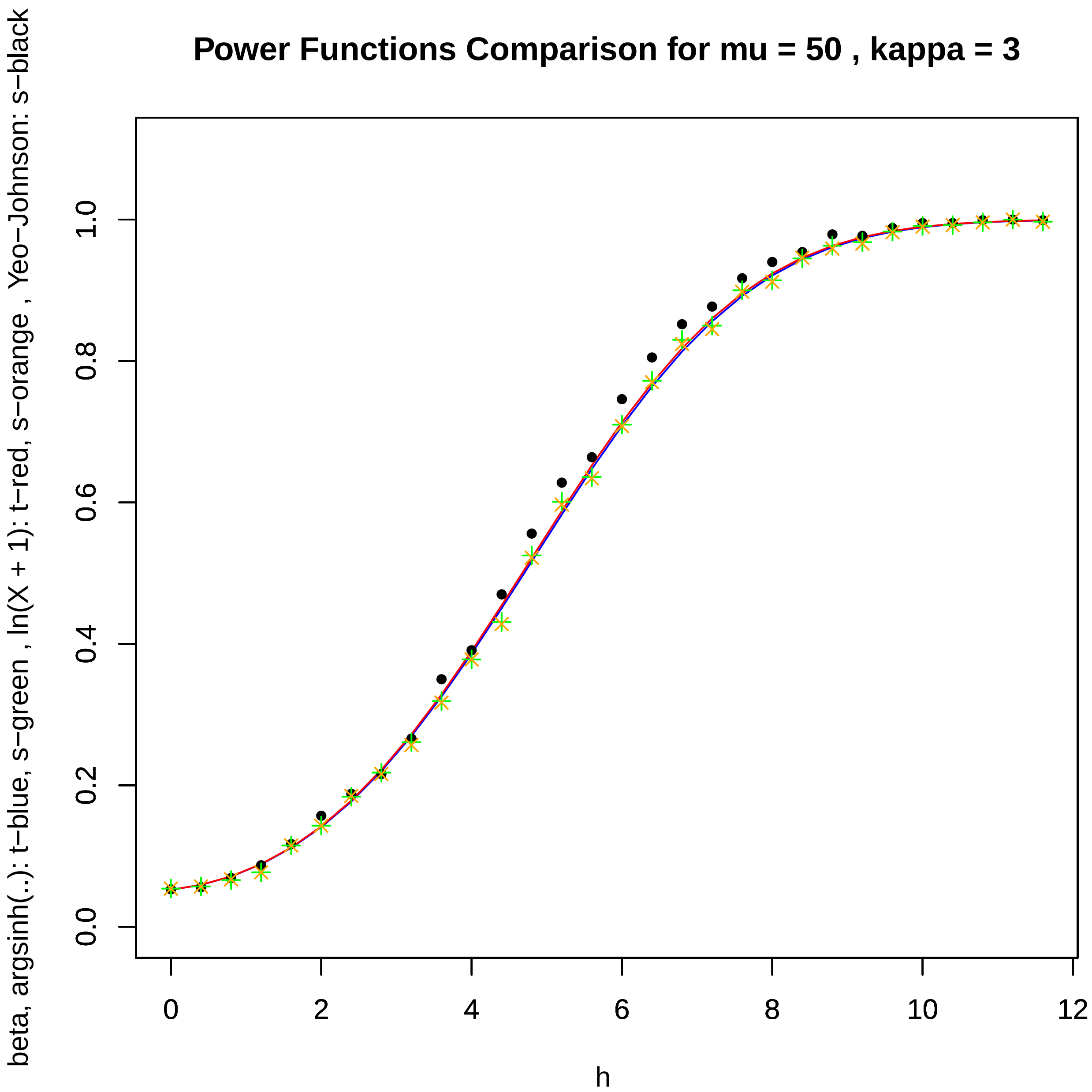


Figure 5.26: Comparison of the power functions. Transformation $Y = 2\sinh^{-1}(\sqrt{(X+c)/(\kappa+d)})$ theoretical p. f.: blue line, simulated p.f.: green + symb., transformation $Y = \ln(X+1)$ theoretical p. f.: red line, simulated p.f.: orange $\times$ symb., Yeo-Johnson transformation simulated p.f.: black dot. Computed for $\mu_1 = 50$, $\kappa = 3$, and coefficient of Yeo Johnson transformation $\lambda_{YJ} = 0.25$.

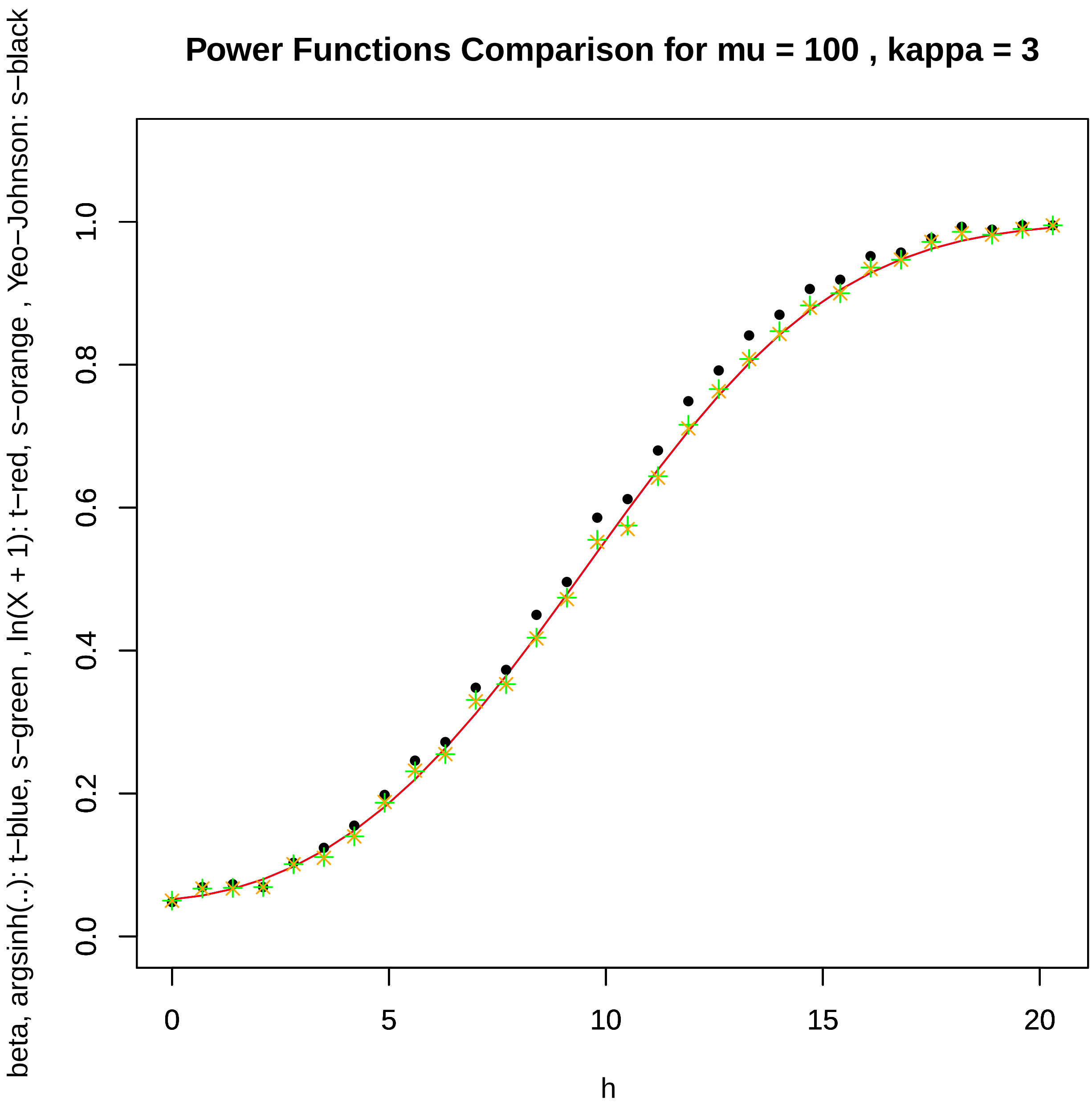


Figure 5.27: Comparison of the power functions. Transformation $Y = 2\sinh^{-1}(\sqrt{(X+c)/(\kappa+d)})$ theoretical p. f.: blue line, simulated p.f.: green + symb., transformation $Y = \ln(X+1)$ theoretical p. f.: red line, simulated p.f.: orange $\times$ symb., Yeo-Johnson transformation simulated p.f.: black dot. Computed for $\mu_1 = 100$, $\kappa = 3$, and coefficient of Yeo Johnson transformation $\lambda_{YJ} = 0.25$.

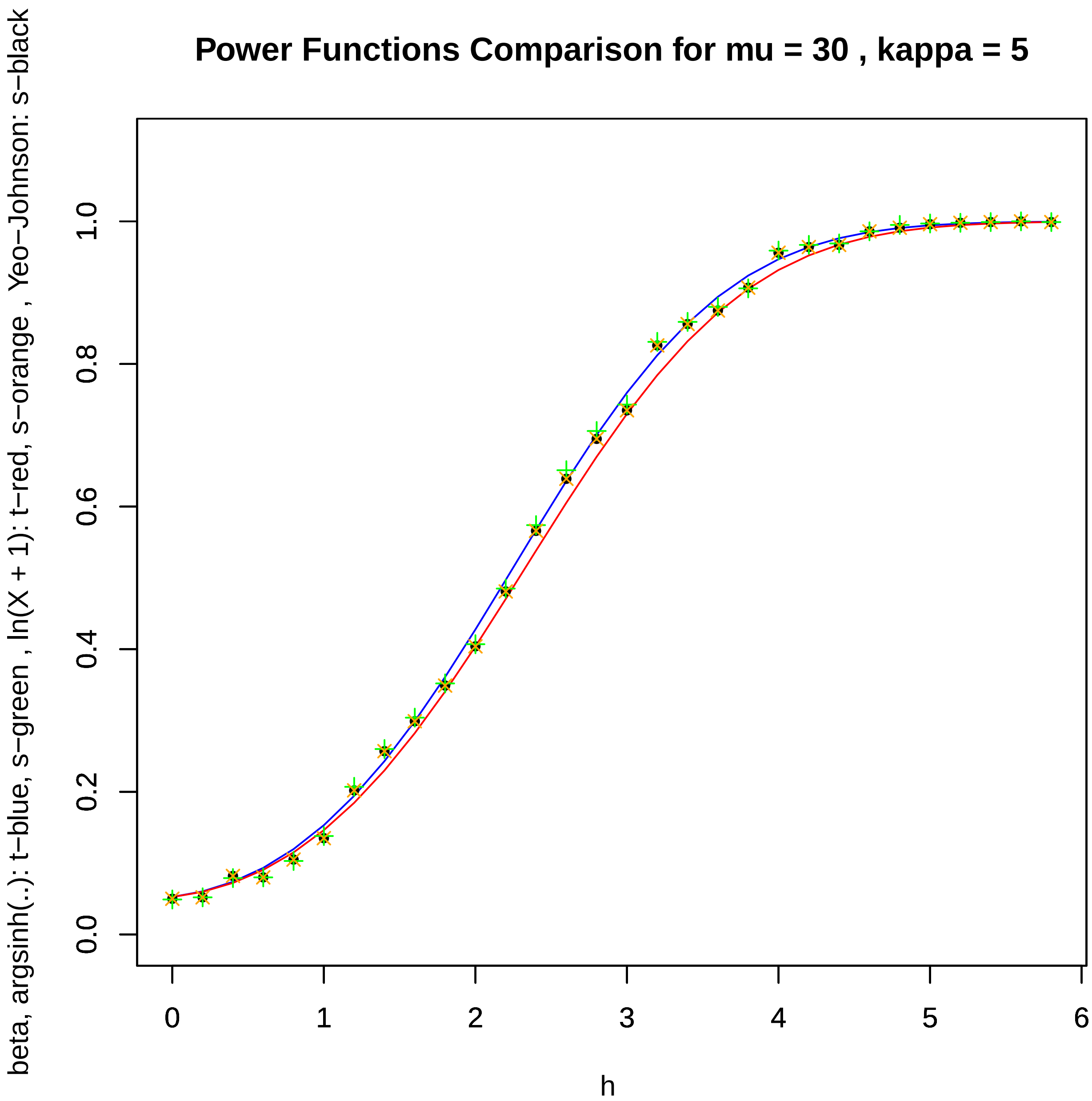


Figure 5.28: Comparison of the power functions. Transformation $Y = 2\sinh^{-1}(\sqrt{(X+c)/(\kappa+d)})$ theoretical p. f.: blue line, simulated p.f.: green + symb., transformation $Y = \ln(X+1)$ theoretical p. f.: red line, simulated p.f.: orange $\times$ symb., Yeo-Johnson transformation simulated p.f.: black dot. Computed for $\mu_1 = 30$, $\kappa = 5$, and coefficient of Yeo Johnson transformation $\lambda_{YJ} = 0.05$.

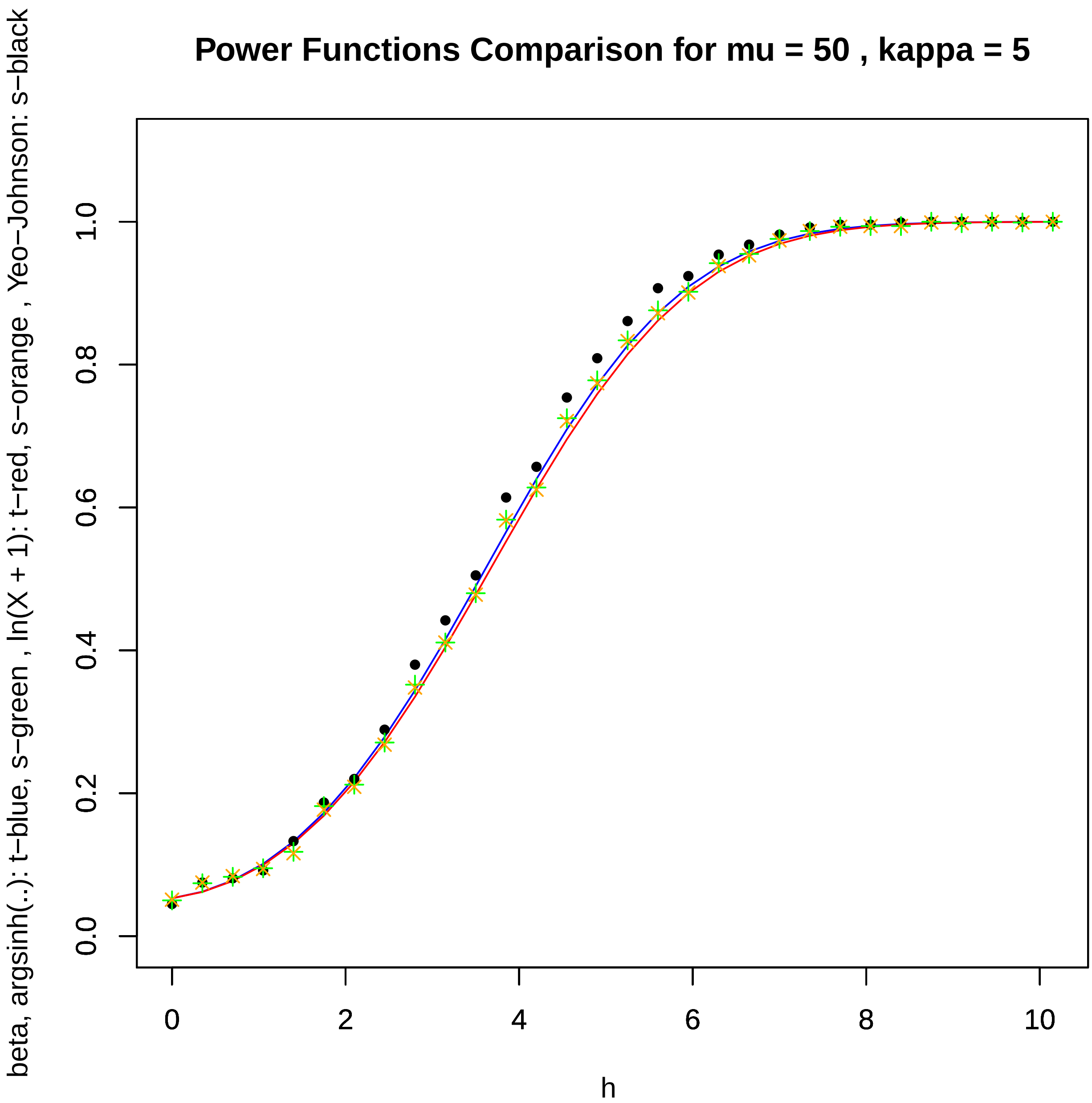


Figure 5.29: Comparison of the power functions. Transformation $Y = 2\sinh^{-1}(\sqrt{(X+c)/(\kappa+d)})$ theoretical p. f.: blue line, simulated p.f.: green + symb., transformation $Y = \ln(X+1)$ theoretical p. f.: red line, simulated p.f.: orange $\times$ symb., Yeo-Johnson transformation simulated p.f.: black dot. Computed for $\mu_1 = 50$, $\kappa = 5$, and coefficient of Yeo Johnson transformation $\lambda_{YJ} = 0.5$.

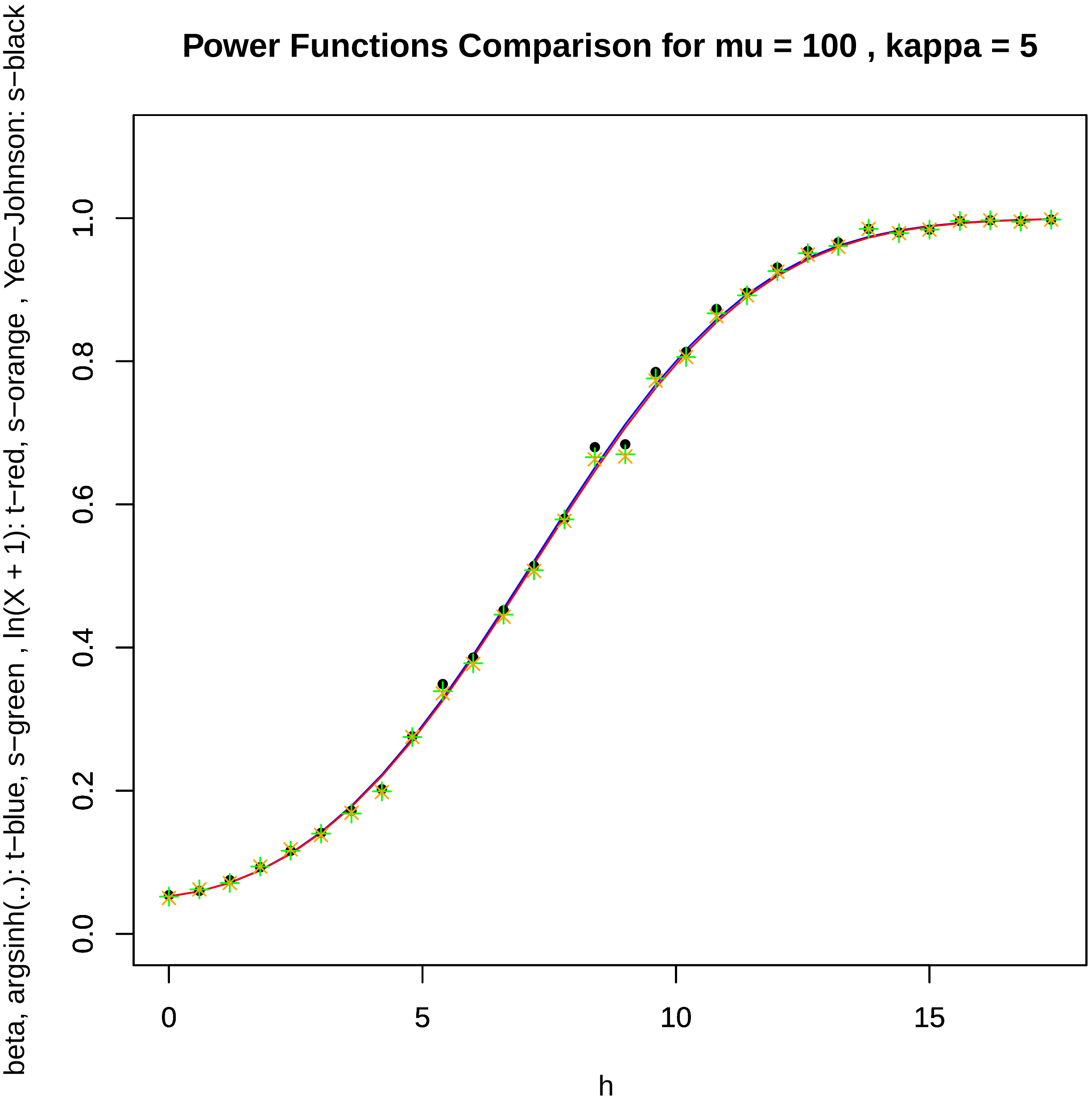


Figure 5.30: Comparison of the power functions. Transformation $Y = 2\sinh^{-1}(\sqrt{(X+c)/(\kappa+d)})$ theoretical p. f.: blue line, simulated p.f.: green + symb., transformation $Y = \ln(X+1)$ theoretical p. f.: red line, simulated p.f.: orange $\times$ symb., Yeo-Johnson transformation simulated p.f.: black dot. Computed for $\mu_1 = 100$, $\kappa = 5$, and coefficient of Yeo Johnson transformation $\lambda_{YJ} = 0.1$.

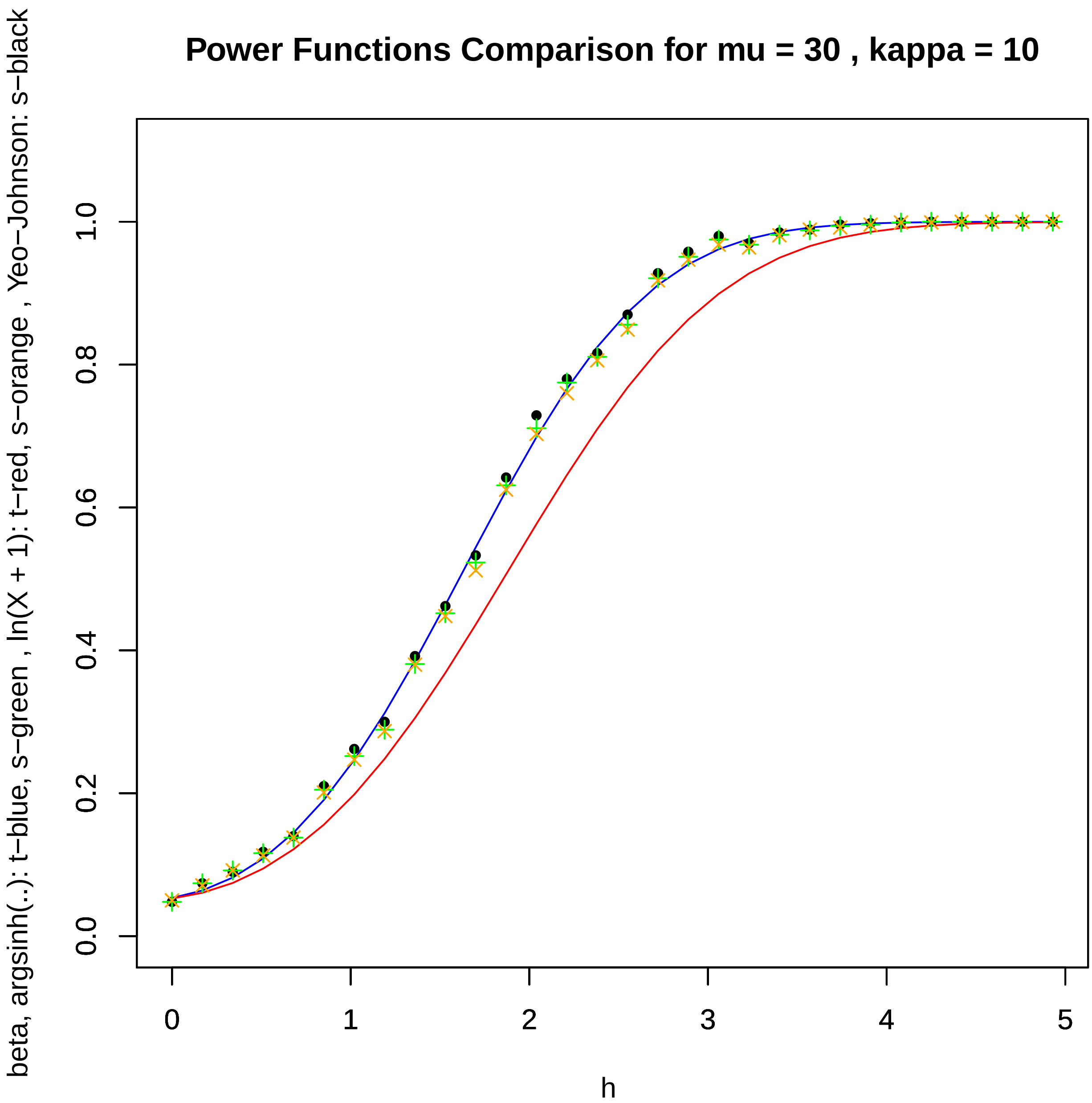


Figure 5.31: Comparison of the power functions. Transformation $Y = 2\sinh^{-1}(\sqrt{(X+c)/(\kappa+d)})$ theoretical p. f.: blue line, simulated p.f.: green + symb., transformation $Y = \ln(X+1)$ theoretical p. f.: red line, simulated p.f.: orange $\times$ symb., Yeo-Johnson transformation simulated p.f.: black dot. Computed for $\mu_1 = 30$, $\kappa = 10$, and coefficient of Yeo Johnson transformation $\lambda_{YJ} = 0.4$.

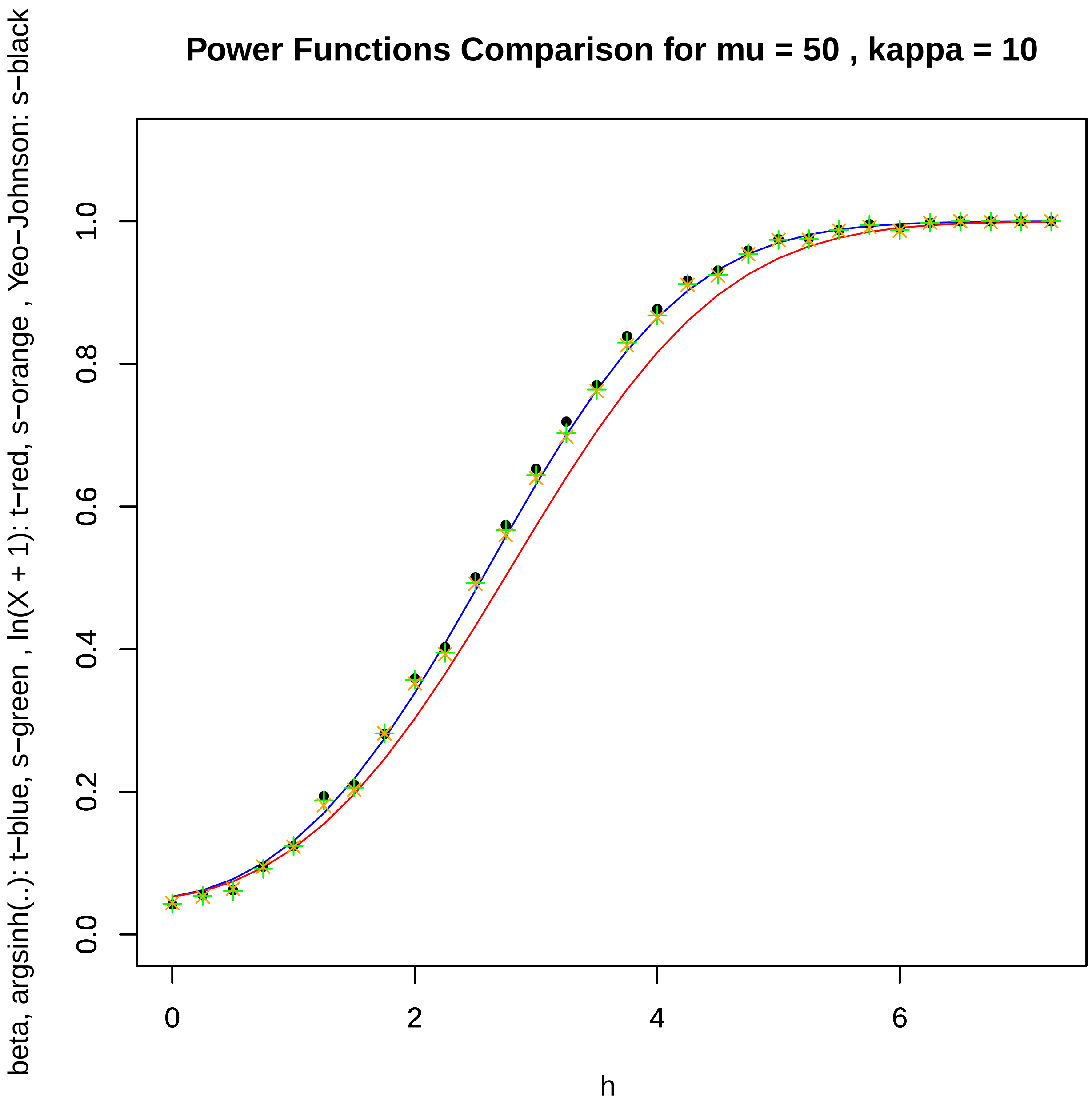


Figure 5.32: Comparison of the power functions. Transformation $Y = 2\sinh^{-1}(\sqrt{(X+c)/(\kappa+d)})$ theoretical p. f.: blue line, simulated p.f.: green + symb., transformation $Y = \ln(X+1)$ theoretical p. f.: red line, simulated p.f.: orange $\times$ symb., Yeo-Johnson transformation simulated p.f.: black dot. Computed for $\mu_1 = 50$, $\kappa = 10$, and coefficient of Yeo Johnson transformation $\lambda_{YJ} = 0.3$.

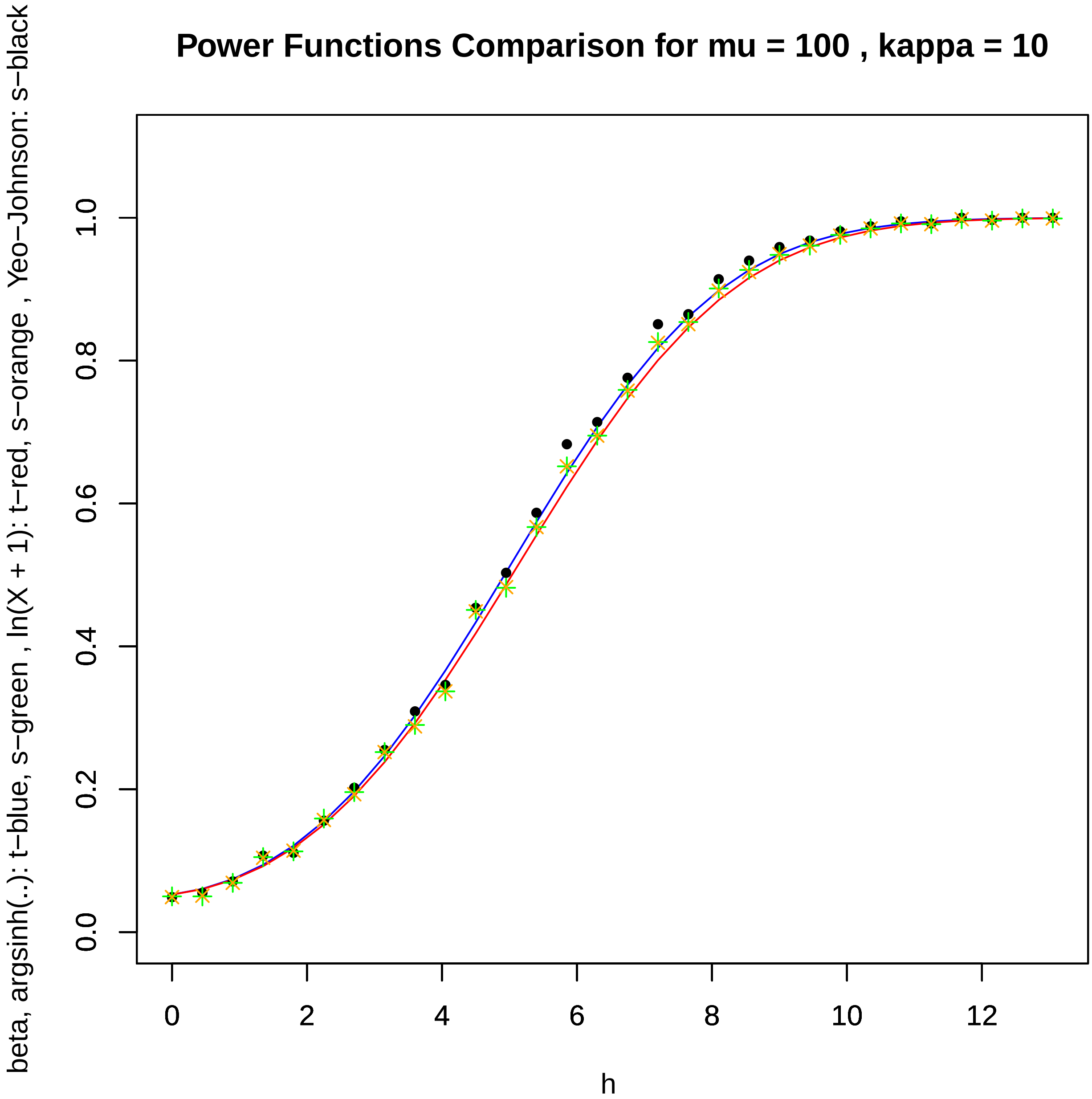


Figure 5.33: Comparison of the power functions. Transformation $Y = 2\sinh^{-1}(\sqrt{(X+c)/(\kappa+d)})$ theoretical p. f.: blue line, simulated p.f.: green + symb., transformation $Y = \ln(X+1)$ theoretical p. f.: red line, simulated p.f.: orange $\times$ symb., Yeo-Johnson transformation simulated p.f.: black dot. Computed for $\mu_1 = 100$, $\kappa = 10$, and coefficient of Yeo Johnson transformation $\lambda_{YJ} = 0.9$.

Appendix A

Computation of Numerical Characteristics of Selected Distributions

A.0.5 Poisson Probability Distrbution

Lemma A.1. *Let $X : \Omega \longrightarrow \mathbb{N}_0$ be a random variable with Poisson probability distribution, then the first moment of X is*

$$\mathbf{E}X = \lambda. \tag{A.1}$$

Proof. By the definition of expectation of discreet random variable (see [2]) we have

$$\mathbf{E}X = \sum_{x=0}^{\infty} x \cdot p(x) = \sum_{x=0}^{\infty} x \cdot e^{-\lambda} \frac{\lambda^x}{x!} \tag{A.2}$$

Since the term of the sum for $x = 0$ is equal to zero we have the following

$$e^{-\lambda} \sum_{x=1}^{\infty} x \cdot \frac{\lambda^x}{x!} = e^{-\lambda} \sum_{x=1}^{\infty} \lambda \cdot \frac{\lambda^{x-1}}{(x-1)!} = \lambda e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = \lambda e^{-\lambda} e^{\lambda} = \lambda \tag{A.3}$$

□

Lemma A.2. *Let $X : \Omega \longrightarrow \mathbb{N}_0$ be a random variable with Poisson probability distribution, then the second moment of X is*

$$\mathbf{E}X^2 = \lambda. \tag{A.4}$$

Proof. Second moment of the discreet random variable X is given by (see [2])

$$\mathbf{E}X^2 = \sum_{x=0}^{\infty} x^2 p(x) = \sum_{x=0}^{\infty} x^2 e^{-\lambda} \frac{\lambda^x}{x!} \tag{A.5}$$

Since the first term of the sum for $x = 0$ is equal to zero we have the following

$$\begin{aligned}
\mathbf{E}X^2 &= \sum_{x=1}^{\infty} x^2 \frac{\lambda^x}{x!} = e^{-\lambda} \sum_{x=1}^{\infty} x^2 \cdot \frac{\lambda \cdot \lambda^{x-1}}{x \cdot (x-1)!} = \lambda e^{-\lambda} \sum_{i=1}^{\infty} x \cdot \frac{\lambda^{x-1}}{(x-1)!} = \lambda e^{-\lambda} \sum_{i=1}^{\infty} x \frac{\lambda^{x-1}}{(n-1)!} \\
&= \lambda e^{-\lambda} \sum_{i=0}^{\infty} (x+1) \frac{\lambda^x}{x!} = \lambda e^{-\lambda} \sum_{i=0}^{\infty} \left[x \frac{\lambda^x}{x!} + \frac{\lambda^x}{x!} \right] = \lambda e^{-\lambda} \left[\sum_{i=0}^{\infty} x \frac{\lambda^x}{x!} + \sum_{i=0}^{\infty} \frac{\lambda^x}{x!} \right] \\
&= \lambda e^{-\lambda} \sum_{i=0}^{\infty} x \frac{\lambda^x}{x!} + \lambda e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^x}{x!}
\end{aligned} \tag{A.6}$$

In the proof of the Lemma A.1 we have seen that $e^{-\lambda}\sum_{i=0}^{\infty} x\frac{\lambda^x}{x!} = \lambda$. Using this result we finally obtain

$$\mathbf{E}X^2 = \lambda^2 + \lambda. \tag{A.7}$$

□

A.0.6 Negative Binomial Probability Disribution

Lemma A.3. *Let $X : \Omega \longrightarrow \mathbb{N}_0$ be a random variable with negative binomial probability distribution, then the first moment of X is*

$$\mathbf{E}X = \frac{\kappa(1-q)}{q}. \tag{A.8}$$

Proof. By the definition of first moment of random variable with discrete probability distribution (see [2]) we have

$$\mathbf{E}X = \sum_{x=0}^{\infty} x \cdot p(x). \tag{A.9}$$

For our case we obtain

$$\begin{aligned}
\mathbf{E}X &= \sum_{x=0}^{\infty} x \binom{\kappa + x - 1}{x} q^{\kappa}(1-q)^x = \sum_{x=0}^{\infty} x \frac{(x+\kappa-1)!}{x!(\kappa-1)!} q^{\kappa}(1-q)^x \\
&= \sum_{x=0}^{\infty} \frac{(x+\kappa-1)!}{(x-1)!(\kappa-1)!} \cdot \frac{\kappa}{\kappa} q^{\kappa}(1-q)^x = \sum_{x=0}^{\infty} \kappa \cdot \frac{(x+\kappa-1)!}{(x-1)!\kappa!} \cdot q^{\kappa}(1-q)^x \\
&= \sum_{x=0}^{\infty} \kappa \binom{\kappa + x - 1}{\kappa} q^{\kappa}(1-q)^x.
\end{aligned} \tag{A.10}$$

Since the term of the sum (A.10) for $x = 0$ is equal to zero we may write

$$\mathbf{E}X = \sum_{x=1}^{\infty} \kappa \binom{\kappa + x - 1}{\kappa} q^{\kappa}(1-q)^x. \tag{A.11}$$

Let us now introduce a following reparametrisation, let $y = x - 1$, and $\delta = \kappa + 1$,

$$\begin{aligned}
\mathbf{E}X &= \sum_{y=0}^{\infty} (\delta - 1) \binom{\delta + x - 1}{\delta - 1} q^{\delta-1}(1-q)(1-q)^y \\
&= (\delta-1)(1-q)q^{\delta-1} \sum_{y=0}^{\infty} \binom{\delta + x - 1}{\delta - 1} (1-q)^y \\
&= (\delta-1)(1-q)q^{\delta-1} \frac{1}{[1-(1-q)]^{\delta}}.
\end{aligned} \tag{A.12}$$

By returning back to the original parametrisation we obtain

$$\mathbf{E}X = \frac{\kappa(1-p)q^{\kappa}}{q^{-\kappa-1}} = \frac{\kappa(1-q)}{q}. \tag{A.13}$$

□

Lemma A.4. *Let $X : \Omega \longrightarrow \mathbb{N}_0$ be a random variable with negative binomial probability distribution, then the variance of X is*

$$\mathbf{var}X = \frac{\kappa(1-q)}{q^2}. \tag{A.14}$$

Proof. By the definition of k-th moment of random variable with discrete probability distribution (see [2]) we have that

$$\mathbf{E}X^k = \sum_{x=0}^{\infty} x^k \cdot p(x). \tag{A.15}$$

For our case we obtain

$$\begin{aligned}
\mathbf{E}X^2 &= \sum_{x=0}^{\infty} x^2 \binom{\kappa+x-1}{x} q^{\kappa}(1-q)^x = \sum_{x=0}^{\infty} x^2 \frac{(x+\kappa-1)!}{x!(\kappa-1)!} q^{\kappa}(1-q)^x \\
&= \sum_{x=0}^{\infty} x \cdot \frac{(x+\kappa-1)!}{(x-1)!(\kappa-1)!} \cdot \frac{\kappa}{\kappa} \cdot q^{\kappa}(1-q)^x = \sum_{x=0}^{\infty} x\kappa \cdot \frac{(x+\kappa-1)!}{(x-1)!\kappa!} \cdot q^{\kappa}(1-q)^x \\
&= \sum_{x=0}^{\infty} x\kappa \binom{\kappa+x-1}{\kappa} q^{\kappa}(1-q)^x.
\end{aligned} \tag{A.16}$$

Since the term of the sum (A.16) for $x = 0$ is equal to zero we may write

$$\mathbf{E}X^2 = \sum_{x=1}^{\infty} x\kappa \binom{\kappa+x-1}{\kappa} q^{\kappa}(1-q)^x. \tag{A.17}$$

We introduce following reparametrisation, let $y = x - 1$ and $\delta = \kappa + 1$, with the reparametrisation we have

$$\begin{aligned}
\mathbf{E}X^2 &= \sum_{y=0}^{\infty} (\delta-1)(y+1) \binom{\delta+y-1}{\delta-1} q^{\delta-1}(1-q)(1-q)^y \\
&= \sum_{y=0}^{\infty} (\delta-1) \left[y \binom{\delta+y-1}{\delta-1} q^{\delta-1}(1-q)(1-q)^y + \binom{\delta+y-1}{\delta-1} q^{\delta-1}(1-q)(1-q)^y \right] \\
&= \sum_{y=0}^{\infty} (\delta-1) y \binom{\delta+y-1}{\delta-1} q^{\delta-1}(1-q)(1-q)^y \\
&+ \sum_{y=0}^{\infty} (\delta-1) \binom{\delta+y-1}{\delta-1} q^{\delta-1}(1-q)(1-q)^y.
\end{aligned} \tag{A.18}$$

Let us denote

$$S_1 = \sum_{y=0}^{\infty} (\delta-1) y \binom{\delta+y-1}{\delta-1} q^{\delta-1}(1-q)(1-q)^y, \tag{A.19}$$

and

$$S_2 = \sum_{y=0}^{\infty} (\delta-1) \binom{\delta+y-1}{\delta-1} q^{\delta-1}(1-q)(1-q)^y \tag{A.20}$$

For S_1 we have

$$\begin{aligned}
S_1 &= \sum_{y=0}^{\infty} (\delta-1) y \binom{\delta+y-1}{\delta-1} q^{\delta-1}(1-q)(1-q)^y \\
&= \sum_{y=0}^{\infty} (\delta-1) y \frac{(\delta+y-1)!}{(\delta-1)!y!} q^{\delta-1}(1-q)(1-q)^y \\
&= \sum_{y=0}^{\infty} (\delta-1) \delta \frac{(\delta+y-1)!}{\delta!(y-1)!} q^{\delta-1}(1-q)(1-q)^y \\
&= \sum_{y=0}^{\infty} (\delta-1) \delta \binom{\delta+y-1}{\delta} q^{\delta-1}(1-q)(1-q)^y. \qquad \text{(A.21)}
\end{aligned}$$

Since the term of S_1 for $y = 0$ is equal to zero, we may write

$$S_1 = \sum_{y=1}^{\infty} (\delta-1) \delta \binom{\delta+y-1}{\delta} q^{\delta-1}(1-q)(1-q)^y. \qquad \text{(A.22)}$$

We introduce the following reparametrisation, let $z = y - 1$, and let $\tau = \delta + 1$. With the reparametrisation we have

$$\begin{aligned}
S_1 &= \sum_{z=0}^{\infty} (\tau-2)(\tau-1) \binom{\delta+y-1}{\tau-1} q^{\tau-2}(1-q)^2(1-q)^z \\
&= (\tau-2)(\tau-1) q^{\tau-2}(1-q)^2 \frac{1}{[1-(1-q)]^{\tau}}. \qquad \text{(A.23)}
\end{aligned}$$

If we return to the original parametrisation we obtain

$$S_1 = \frac{\kappa(\kappa+1) q^{\kappa}(1-q)^2}{q^{\kappa+2}} = \frac{\kappa(\kappa+1)(1-q)^2}{q^2}. \qquad \text{(A.24)}$$

By the proof of Proposition A.3 for S_2 we have

$$S_2 = \frac{\kappa(1-q)}{q}, \qquad \text{(A.25)}$$

so

$$\mathbf{E}X^2 = \frac{\kappa(\kappa+1)(1-q)^2}{q^2} + \frac{\kappa(1-q)}{q}. \qquad \text{(A.26)}$$

□

Conclusion

After introducing the necessary theoretical background for this work in Chapter 1 and 2, the variance stabilising transformations for random variables with Poisson and negative binomial distribution were studied. Based on the work [1] generalised versions of these transformations with additional parameters inside the arguments were presented, namely $Y = \sqrt{X+c}$, where $c > 0$ for Poisson case and $Y = 2\sinh^{-1}\left(\sqrt{(X+c)/(\kappa+d)}\right)$ and $Y = \ln(X+A)$, where $A, c, d > 0$ for the negative binomial case. A great part of the Chapter 3 tackles with the problematic of approximating numerical characteristics of the random variables transformed via the presented generalised transformations in order to determine the optimal value of the additional parameters.

For the Poisson distributed random variable the approximations of the numerical characteristics were found by taking Taylor expansion of the transformation, computing the numerical characteristic of the Taylor expansion term by term and then expanding each term asymptotically for large values of the parameter λ of Poisson distribution. In such way expansions dependant on the additional parameter were found. The optimal value of the parameter was chosen in such a way, so that the coefficient of the term dependant on λ of the highest degree in the variance approximation expansion would be zero. The optimal value of the parameter for the Poisson distributed random variable was determined to be $c = \frac{3}{8}$. Additionally it was seen in the end of the Section 3.1, that the optimal value of the parameter c minimises the bias b_Y, which was defined as the difference of the parameter λ and its estimate λ_Y derived by applying the transformation in reverse to arithmetic mean of transformed a random sample of a Poisson distribution with parameter λ.

For negative binomially distributed random variable $X \sim NBi(\mu, \kappa)$ case first the asymptotic expansion of the moment generating function and consequently the cumulant generating function of the transformed random variable were found. The approximations of the numerical characteristics were derived from the cumulant generating function approximation. In this case the approximations are dependant on the shape parameter κ of the negative binomial distribution and for higher values of κ we can obtain better approximations. The idea behind finding the optimal values of the parameters A, c, d is similar to the Poisson case. Again an approximation of variance of the transformed random variable in a form of an expansion was used. The optimal value of the constant was picked so that the coefficient of the term of the variance approximation dependant on the expectation parameter μ of the highest order would be zero. The optimal value of A was determined to be $A = \frac{1}{2}\kappa$. The optimal value of $d = -2c$ and finally $c = \frac{3}{8} + \frac{23}{192\kappa} + O\left(\frac{1}{\kappa^2}\right)$ for κ large. At the end of the Section 3.2 limiting values of skewness parameter for the transformed random variables were derived as a functions of κ that tend to zero for κ large.

The second goal of the thesis was to provide comparison of the transformations introduced in the Section 3 with some other commonly used transformations of the random variable both theoretically and via simulations. The comparison was done within One-

Way ANOVA framework by comparing power functions of the F test used to test the hypothesis of equality of expectations among p random samples of equal size n of Poisson or negative binomial distribution to which one of the transformations was applied. The Chapter 4 presents the theoretical background for the comparison. Additionally it was assumed that for some transformations the assumption of equal variances necessary for the One-Way ANOVA was violated. In order to proceed with the comparison in Section 4.2 the F statistic under the relaxed assumption of unequal variances is studied. With the results of Subsections 4.2.1 and 4.2.2 it is found out, that given the assumption of equal variances violated, the F statistic may be expressed as a ratio of linear combinations of χ^2 distributed random variables with positive coefficients. Using this result and the results of [12] the ratio is approximated by one F distributed random variable multiplied by a constant.

In the fifth and final chapter the numerical comparison itself is carried out by computing the power functions of the F test for transformations introduced in Chapter 3 and for a transformation $\ln(X+1)$, when applied to sample of either Poisson or negative binomial distribution. Two different approaches were used, a theoretical one, based on the definition of the power of a test, using the approximations of the numerical characteristics of the transformed random variables and a an approach via simulations. Both of the approaches were described in detail in Section 5.4. Additionally Yeo-Johnson was added to the comparison but only via the approach by simulations. To get better insight a goodness of approximations of the numerical characteristics of the transformed random variables was checked by comparing them with their respective estimations from generated samples.

A study of the properties of the parameter estimates of random variable of Poisson and negative binomial distribution via simulations was provided in this section as well. A sample of either Poisson or negative binomial distribution was generated and transformed via $Y = \sqrt{X + 3/8}$ and $Y = \ln(X + 1)$ for the Poisson case and via $Y = 2\sinh^{-1}(\sqrt{(X + c)/(\kappa + d)})$ and $Y = \ln(X + 1)$ for the negative binomial case. Expectation of the transformed sample was estimated using arithmetic mean. The respective transformation was then applied to the arithmetic mean in reverse to obtain an estimate λ_Y of parameter λ of the Poisson distribution or μ_Y of the parameter μ of negative binomial distribution in the respective case. This procedure was repeated k times for the same settings of parameters λ or μ to obtain a random samples of the estimate λ_Y or μ_Y. Two interesting results were obtained in this study. For both the Poisson and negative binomial case the respective estimates were biased. In the negative binomial case however the value of bias decreased as the value of the shape parameter κ grew.

Two interesting facts came as results of the check of the goodness of the approximations. First, for negative binomial case it was found out, that with increasing value of the shape parameter κ the value of μ for which the sample variance became approximately stable increased. This was more evident for the transformation $\ln(X+1)$, the increase of μ for $Y = 2\sinh^{-1}\left(\sqrt{(X + c)/(\kappa + d)}\right)$ was significantly slower. Second, for the Poisson case the variance of the random variable transformed via $\ln(X+1)$ did not become stable for values of λ up to approximately one hundred which might pose a problem when using this transformation for small values of the parameter λ. On the other hand, the variance of the random variable obtained via transformation $Y = \sqrt{X + 3/8}$ was stable even for relatively small values of λ.

Finally, in the last part of the Chapter 5 the power function comparison was carried out. For the Poisson case, random samples of Poisson distribution with values of parameter $\lambda = 5, 10, 20, 50$ transformed via transformations $Y = \sqrt{X + 3/8}$, $Y = \ln(X + 1)$

and the Yeo-Johnson transformation were considered. In general the simulated power functions attained values close to their theoretical counterparts for both $Y = \sqrt{X + 3/8}$ and $Y = \ln(X + 1)$. Both the simulated and the theoretical power function of the F test applied to a sample transformed via $Y = \sqrt{X + 3/8}$ scored a little better than the simulated and theoretical power function of the F test applied to a sample transformed via $Y = \ln(X+1)$ and similarly to the simulated the power function of the F test applied to a sample transformed via Yeo-Johnson transformation. The similarity of the Yeo-Johnson and the $Y = \sqrt{X + 3/8}$ case may be explained by the fact, that the optimal values of the parameter of Yeo-Johnson transformation obtained via maximum likelihood estimation were close to the power of the square root 0.5. Namely in the case $\lambda = 5$ the optimal value of the Yeo-Johnson transformation parameter was 0.45. For increasing values of the parameter λ the difference between all the power functions became smaller. For value of $\lambda = 50$ was the difference already practically negligible, however we conclude that for smaller values of λ one should tend to choose either Yeo-Johnson transformation or $Y = \sqrt{X + 3/8}$ over $Y = \ln(X + 1)$.

For the negative binomial case, random samples of negative binomial distribution with values of parameter $\mu = 30, 50, 100$, and values of shape parameter $\kappa = 3, 5, 10$, transformed via $Y = 2\sinh^{-1}\left(\sqrt{(X + c)/(\kappa + d)}\right)$, $Y = \ln(X + 1)$, and the Yeo-Johnson transformation were considered.

We observed, that for this case two possibly pathological phenomena occurs for certain settings of parameters μ and κ where the theoretical power function of the F test applied to a random sample transformed via $Y = \ln(X + 1)$ behaves oddly and differently than its simulated counterpart. For setting $\mu = 30$ and $\kappa = 3$ the above mentioned theoretical power function scores higher than its simulated counterpart and both theoretical and simulated power function of the F test applied to a random sample transformed via $Y = 2\sinh^{-1}\left(\sqrt{(X + c)/(\kappa + d)}\right)$, topped only by the power function of the F test applied to Yeo-Johnson transformed sample. We see two possible factors whose interplay led to the occurrence of this phenomena. First, as seen from the Figure 5.36, the absolute value of sample skewness of the sample transformed via $Y = \ln(X + 1)$ is the highest out of all values of parameter κ that were considered, so there might be a significant departure from normality of the transformed sample. The second factor is the quality of the variance approximation used to compute the theoretical power function, which is rather poor for small values of parameters μ and κ for the transformation $Y = \ln(X + 1)$. In fact the approximation drops rapidly below any bound as μ tends to zero.

A second pathological phenomena occurred for small values of μ and the largest assumed value of κ, where the theoretical power function of the F test applied to a sample transformed via $Y = \ln(X + 1)$ scored significantly worse than its simulated counterpart and all the other power functions. We explain this by what we observed in the Subsection 5.3.2, where we saw that for increasing values of κ it takes larger value of μ for both sample variance and variance approximation to become at least close to stable. This phenomena is more significant for the sample transformed via $Y = \ln(X + 1)$. Moreover the variance approximation for the case of $\kappa = 10$ grows above all bounds rapidly, hence for small values of μ the power function computation using this approximation might be heavily vitiated by error.

In general, save the pathological cases described above both the theoretical and the simulated power function of the F test applied to a $Y = 2\sinh^{-1}\left(\sqrt{(X + c)/(\kappa + d)}\right)$ transformed sample scores always slightly better than the power functions of the F test applied to a sample transformed via $Y = \ln(X + 1)$, but the Yeo-Johnson transformation outperforms both of the two other transformations in all cases but one, for the setting $\mu =$

30, $\kappa = 5$. This is however because by the parameter of the Yeo-Johnson transformation estimated via the maximum likelihood estimation was close to 0 for which the Yeo-Johnson transformation coincides with the transformation $Y = \ln(X + 1)$.

For the increasing values of the parameters κ and μ the difference between all the power functions became smaller.

We conclude for the negative binomial case that for the small values of κ and μ one shall tend to choose Yeo-Johnson transformation over the other two discussed transformations, save the case when the estimated best value of the parameter of the Yeo-Johnson transformation is very close to 0. In such case we suggest to use $Y = 2\sinh^{-1}\left(\sqrt{(X+c)/(\kappa+d)}\right)$. For small values of μ (say around $\mu = 30$) and large values of κ (say around $\kappa = 10$) one should choose preferably again the Yeo-Johnson transformation, possibly also $Y = 2\sinh^{-1}\left(\sqrt{(X+c)/(\kappa+d)}\right)$. For large values of the parameters the difference between the power functions becomes increasingly insignificant, we would favorize the more sophisticated transformations, however the transformation $Y = \ln(X + 1)$ should be sufficient as well.

Bibliography

[1] ANSCOMBE, Francis John. *The transformation of Poisson, binomial and negative-binomial data.* Biometrika. 1948, 35 (3–4), 246–254.

[2] ANDĚL, Jiří. *Základy matematické statistiky.* Praha: Matfyzpress, 2011. ISBN 978-80-7378-1620.

[3] PECCATI, Giovanni and Murad S. TAQQU. *Wiener chaos: moments, cumulants and diagrams : a survey with computer implementation.* New York: Springer, c2011. Bocconi and Springer series, 1. ISBN 8847016789.

[4] KLENKE, Achim. *Probability theory: a comprehensive course.* London: Springer, c2008. ISBN 9781848000483.

[5] ZWILLINGER, Daniel. *CRC standard mathematical tables and formulae.* 30th ed. Boca Raton, Fla.: CRC Press, c1996. ISBN 0-8493-2479-3.

[6] BARTLE, Robert Gardner and Donald R. SHERBERT. *Introduction to real analysis.* 3rd ed. New York: Wiley, c2000. ISBN 0471321486.

[7] FRANCŮ, Jan. *Funkcionální analýza I.* 2., dopl. vyd. Brno: Akademické nakladatelství CERM, 2014. ISBN 978-80-214-4920-6.

[8] SCHEFFE, Henry. *The analysis of variance.* Wiley classics library ed. New York: Wiley-Interscience Publication, 1999. ISBN 0471758345.

[9] MATHAI, A. M. and Serge B. PROVOST. *Quadratic forms in random variables: theory and applications.* New York: M. Dekker, c1992. ISBN 0824786912.

[10] LAUB, Alan J. *Matrix analysis for scientists and engineers.* Philadelphia: Society for Industrial and Applied Mathematics, c2005. ISBN 0898715768.

[11] MARDIA, K. V., J. T. KENT and J. M. BIBBY. *Multivariate analysis.* New York: Academic Press, 1979. ISBN 0124712525.

[12] KIM, Hae-Young, Matthew J GRIBBIN, Keith E MULLER a Douglas J TAYLOR. *Analytic, Computational, and Approximate Forms for Ratios of Noncentral and Central Gaussian Quadratic Forms.* Journal of Computational and Graphical Statistics [online]. 2006, 15(2), 443-459 [cit. 2017-03-20]. DOI: 10.1198/106186006X112954. ISSN 1061-8600. Available from: http://www.tandfonline.com/doi/abs/10.1198/106186006X112954

[13] RAO, C. Radhakrishna. *Linear statistical inference and its applications.* New York: John Wiley, 1965.

[14] KARÁSEK, Jiří and Ladislav SKULA. *Lineární algebra: teoretická část.* Brno: Akademické nakladatelství CERM, 2005. ISBN 80-214-3100-8.

[15] MOSS, D. and D. P. McPhee, *The impacts of recreational four-wheel driving on the abundance of the Ghost crab (Ocypode cordimanus) on sub- tropical sandy beaches in SE Queensland.* Coastal Management, 34:133–140, 2006

[16] DRAPER, R. Norman and Harry SMITH. *Applied Regression Analysis.* New York: Wiley, 1998. ISBN 0-471-17082-8.

[17] Box, G. E. P., and D. R. Cox. *An Analysis of Transformations.* Journal of the Royal Statistical Society. Series B (Methodological), vol. 26, no. 2, 1964, pp. 211–252. JSTOR, www.jstor.org/stable/2984418.

[18] Yeo, In-Kwon, and Richard A. Johnson. *A New Family of Power Transformations to Improve Normality or Symmetry.* Biometrika, vol. 87, no. 4, 2000, pp. 954–959. JSTOR, www.jstor.org/stable/2673623.

Notation Index

$\mathbb{R}$	The set of real numbers
$\mathbb{R}^n$	The n-dimensional real space
$\otimes$	Kronecker product
X, Y	Random variables
$\mathbf{E}X$	Expectation of a random variable X
$\mathbf{var}X$	Variance of a random variable X
$Po(\lambda)$	Poisson distribution of parameter λ
$NBi(r,p)$	Negative binomial distribution of parameters r and p
$N(\mu,\sigma^2)$	Normal distribution of expectation μ and variance σ^2
$N(0,1)$	Standard normal distribution
$\Phi(x)$	Distribution function of standard normal distribution
ϕ	Probability density function of standard normal distribution
χ^2_n	Pearson χ^2 square distribution with n degrees of freedom
$\chi^2_{n,\delta}$	Noncentral Pearson χ^2 square distribution with n degrees of freedom and noncentrality parameter δ
$F_{s,t}$	Fisher Snedecor distribution of parameters s and t
$F_{s,t,\delta}$	Noncentral Fisher Snedecor distribution of parameters s and t and noncentrality parameter δ
$\mathcal{F}_F$	Distribution function of F-distributed random variable
$\mathcal{F}_{F_\delta}$	Distribution function of noncentral Fisher Snedecor distributed random variable with n
$Q_F(r;s,t)$	r-th quantile of F distribution
$\mu'_{X,k}$	k-th general moment
$\mu_{X,k}$	k-th central moment
$\mathbf{x}_n$	n-dimensional real (deterministic) vector
$\mathbf{1}_n$	n-dimensional vector of ones
$\mathbf{X}_n$	n-dimensional random vector
$N_n(\boldsymbol{\mu},\mathbf{V})$	n-dimensional normal distribution of vector of expectations $\boldsymbol{\mu}$ and variance matrix $\mathbf{V}$
$\mathbf{EX}$	Expectation of a random vector $\mathbf{X}$
$\mathbf{varX}$	Variance of a random vector $\mathbf{X}$
$\mathbf{A}_{m\times n}$	A matrix of a type $n\times$
$diag(\mathbf{x}_p)$	Diagonal matrix with p dimensional vector $\mathbf{x}$ on the main diagonal
$\mathbf{A}^2_{p\times p}$	Matrix product $\mathbf{A}_{p\times p}\mathbf{A}_{p\times p}$
$\mathbf{I}_{n\times n}$	Identity matrix of a type $n\times n$
$Tr(\mathbf{A}_{n\times n})$	Trace of a matrix $\mathbf{A}_{n\times n}$
$rank(\mathbf{A}_{m\times n})$	Rank of a matrix $\mathbf{A}_{m\times n}$

Electronic Appendix Index

Po_Goodness_of_Approximations.R
NBi_Goodness_of_Approximations.R
Po_Power_Functions_Comparison.R
NBi_Power_Functions_Comparison.R
Po_Histograms.R
NBi_Histograms.R
CummulantGeneratingFunction.mw
ExpecationQuadraticTermApprox.mw
LnPlusOneNB2ndDegreeExp.mw
LnPlusOneNB2ndDegreeVar.mw
LnPlusOnePoExpecatation.mw
LnPlusOnePoVariance.mw