1-2020

Modified-half-normal distribution and different methods to estimate average treatment effect.

Jingchao Sun
University of Louisville

Recommended Citation

MODIFIED-HALF-NORMAL DISTRIBUTION AND DIFFERENT METHODS TO ESTIMATE AVERAGE TREATMENT EFFECT

By

Jingchao Sun
B.S., Southeast University, 2013
M.S., University of Illinois Urbana-Champaign, 2014

A Dissertation
Submitted to the Faculty of the
School of Public Health and Information Sciences
of the University of Louisville
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy
in Biostatistics

Department of Bioinformatics and Biostatistics
University of Louisville
Louisville, Kentucky

December, 2020

MODIFIED-HALF-NORMAL DISTRIBUTION AND DIFFERENT METHODS TO ESTIMATE AVERAGE TREATMENT EFFECT

By

Jingchao Sun
B.S., Southeast University, 2013
M.S., University of Illinois Urbana-Champaign, 2014

A Dissertation Approved on

August 28, 2020

by the following Dissertation Committee:

_____

Maiying Kong, Ph.D., Dissertation Director

_____

Subhadip Pal, Ph.D., Dissertation Co-director

_____

Scott D Duncan, M.D.

_____

Jeremy Gaskins, Ph.D

_____

Qi Zheng, Ph.D

DEDICATION

This dissertation is dedicated to my parents who have given me invaluable educational opportunities.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my advisors Dr. Maiying Kong and Dr. Subhadip Pal for their insightful guidance and constant inspiration over the last four years. I always deeply appreciate their support when I encounter difficulties, and their sincere advice when I get lost. Their supportive instruction and kind behavior help me explore the statistical world and show me a charming scenery comprised of methodology and shaped by information pieces.

I would like to thank M.D. Scott D Duncan, Dr. Jeremy Gaskins and Dr. Qi Zheng for their time to serve in my dissertation committee. I especially would like to thank M.D. Scott D Duncan for providing me the opportunity to work on medical area.

I would like to thank the Department of Bioinformatics and Biostatistics for providing me a great environment to do research, thank all the faculty members for enriching my knowledge in biostatistics, thank my fellow students in the biostatistics department and my wonderful colleagues in pediatric department who are always helpful and have made my life more enjoyable.

I am also grateful to my friends: Shuhong, You, Ruiqi, Xiaofang and Qian, who have given me great encouragement and stood by me. I always feel lucky and grateful to have them in my life.

Finally, I would like to express my thanks to my parents and my husband for their immeasurable supports in my life, so that I can grow to be a person who are stronger than yesterday.

ABSTRACT

MODIFIED-HALF-NORMAL DISTRIBUTION AND DIFFERENT
METHODS TO ESTIMATE AVERAGE TREATMENT EFFECT

Jingchao Sun

August 28, 2020

This dissertation consists of three projects related to Modified-Half-Normal distribution and causal inference.

In my first project, a new distribution called Modified-Half-Normal distribution was introduced. The Modified-Half-Normal distribution occurs in diverse areas of research, such as sandwich MCMC algorithm in Bayesian binary regression, inference with the Projected Normal distribution, data augmentation algorithm for the vonMises-Fisher distribution, and application in robust Bayesian graphical model. I explored a few of its distributional properties, the procedures for generating random samples based on Bayesian approaches, and the parameter estimation based on the method of moments.

The second project deals with the problem of selection bias of average treatment effect (ATE) if we use the observational data. I combined the propensity score based inverse probability of treatment weighting (IPTW) method and the directed acyclic graph (DAG) to solve this problem. In addition, I carried out simulations to examine the performance of the propensity score based IPTW method in estimating ATE when different sets of variables are included in the propensity score estimation. Moreover, a case study was conducted based on the 2016 HCUP KID database.

The third project solves the problem of bias ATE using observational data by combining the doubly robust methods with the super learner algorithm. Doubly robust method estimate is unbiased if either propensity score model or outcome model is correctly specified. I used super learner algorithm to replace the outcome model. Super learner is a prediction method which aims to find the best combination of a group of prediction algorithms, such as machine learning methods and generalized additive model, by minimizing the cross validation error. Simulations were carried out to examine the performance of different doubly robust methods and super learner assisted doubly robust methods. Furthermore, a case study was implemented based on the National Health and Nutrition Examination Survey Data I Epidemiologic Follow-up Study (NHEFS).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 The Modified-Half-Normal distribution: properties and an efficient sampling scheme

We introduced the Modified-Half-Normal (MHN) distribution, a novel family of continuous probability distributions, supported on the positive part of the real line. The probability density function of the distribution is proportional to the function $x \mapsto x^{\alpha-1}\exp(-\beta x^2 + \gamma x)\mathbb{I}(x > 0)$. The parameters α and β are positive real numbers while γ can be any real number. We studied a few of its distributional properties including the normalizing constant and moments of distribution in terms of the Fox-Wright function. We demonstrated its relevance by exhibiting its connection to a number of Bayesian statistical methods appearing from diverse areas of research. An efficient sampling algorithm is contributive to the Bayesian procedures, which require sampling from Modified-Half-Normal distribution. Therefore, a major focus of this article is the procedures of generating random samples from the Modified-Half-Normal distribution. To ensure efficiency, we have proved that the proposed Accept-Reject sampling algorithms are uniformly efficient irrespective of the choice of the parameter specifications.

1.2 Directed acyclic graph assisted methods for estimating average treatment effect

Observational data, such as electronic clinical records and claims data, can be very useful to examine average treatment effect (ATE) and help decision making, if used appropriately. Propensity score based inverse probability of treatment weighting (IPTW) method has been very powerful to estimate ATE if the assumptions of exchangeability, consistency, and positivity hold. Directed acyclic graph (DAG) provides a feasible approach to examine the exchangeability assumption, that is, the treatment assignment variable and the potential outcome are independent given a set of confounding variables which block all backdoor paths from treatment assignment variable to the potential outcome. Therefore, we do not need to adjust all confounding variables. It is common that there are several sets of such confounding variables. In the propensity score model, we consider to adjust a minimal set of confounding variables which block all backdoor paths from treatment assignment variable to the potential outcome. Recent literature also shows that including predictors (variables related to outcomes but not treatment) in the propensity score model also improves the accuracy in estimating ATE. We proposed to include a minimal set of confounding variables and predictors in the propensity score model. We carried out extensive simulations to examine the performance of the propensity score based IPTW method in estimating ATE when different sets of variables were included in the propensity score model. The simulation results indicated that the performance of ATE estimation based on the minimal set of confounding variables plus predictors was comparable with the maximal set of confounding variables plus predictors. We applied the proposed method to examine if tracheostomy (i.e., a medical procedure involving creating an opening in the neck in order to place a tube into a person's windpipe) was a cause of in-hospital death for infants based on the 2016 Healthcare

Cost and Utilization Project (HCUP) Kids' Inpatient Database (KID).

1.3 Different doubly robust methods for estimating ATE and their improvement using super learner

Average treatment effect (ATE) is an important parameter to measure the benefit of a treatment in medical research. Propensity score based methods, including stratification, matching, inverse probability of treatment weighting (IPTW), regression with propensity score as covariates, and doubly robust (DR) method are often used to estimate ATE. Among them, DR method has gained much attention due to the property that ATE estimate based on DR method is asymptotically consistent if either propensity score model or outcome model is correctly specified. We investigated different forms of doubly robust estimators and compared their performances via simulations. In addition, we investigated the super learner algorithm for outcome model, and applied the super learner algorithm in DR method. The resulting super learner assisted DR method provides less biased estimates for ATE even when both propensity score model and outcome model are falsely specified. Extensive simulations were carried out to examine the performance of super learner assisted doubly robust methods. Moreover, a case study based on the National Health and Nutrition Examination Survey Data I Epidemiologic Follow-up Study (NHEFS) was implemented to examine the weight gain due to quitting smoking.

CHAPTER 2

THE MODIFIED-HALF-NORMAL DISTRIBUTION: PROPERTIES AND AN EFFICIENT SAMPLING SCHEME

2.1 Introduction

The objective of this manuscript is to introduce the Modified-Half-Normal (MHN) distribution, a novel family of continuous probability distributions, supported on the positive part of the real line. The corresponding probability density function is specified as

$$f_{\mathrm{MHN}}(x \mid \alpha, \beta, \gamma) = \frac{2\beta^{\frac{\alpha}{2}} x^{\alpha-1} \exp(-\beta x^2 + \gamma x)}{{}_1\Psi_1 \left[\begin{matrix} (\frac{\alpha}{2}, \frac{1}{2}) \\ (1, 0) \end{matrix}; \frac{\gamma}{\sqrt{\beta}} \right]} \mathbb{I}(x > 0),$$

(2.1)

for $\alpha > 0, \beta > 0, \gamma \in \mathbb{R}$ where $\mathbb{I}(x > 0)$ denotes the indicator function and ${}_1\Psi_1 \left[\begin{matrix} (\frac{\alpha}{2}, \frac{1}{2}) \\ (1, 0) \end{matrix}; x \right]$ refers to a specific case of the Fox-Wright function (Wright, 1935; Fox, 1928) that we review in the Section 2.2.

The introduction of the Modified-Half-Normal distribution is primarily motivated by its connection to a number of MCMC based Bayesian procedures. Although in Bayesian analysis new distributions frequently crop up in the form of conditional posterior distributions, usage for many such probability distributions is too specific and they may not be relevant for a broader context. Additionally, most such dis-

tributions often lack a tractable form of their distributional attributes (such as the normalizing constant), resulting in difficulties to utilize them in other contexts. On the contrary, the Modified-Half-Normal distribution occurs in diverse areas of research signifying its relevance to contemporary statistical modeling and computation. In order to demonstrate its usefulness, in Section 2.1.1 we include a few examples that involve the $\mathrm{MHN}\,(\alpha, \beta, \gamma)$ distribution as one of their posterior conditionals to sample from. In particular, the distribution will play an important role in the development of data augmentation algorithms for the directional data (see Section 2.1.1), a budding technique to analyze directional data. Apart from the examples provided in Section 2.1.1, the $\mathrm{MHN}\,(\alpha, \beta, \gamma)$ is likely to transpire in many other statistical models due to the similarity of the Normal distribution. Moreover, the Modified-Half-Normal distribution has interesting properties including the availability of its normalizing constant.

From the perspective of its applicability towards MCMC algorithms, it is important to develop methods to generate random samples from the distribution. Therefore, we devote considerable efforts to develop algorithms for generating random samples from the Modified-Half-Normal distribution. Specifically, we employed the Accept-Reject sampling technique (Devroye, 2006; Casella, 1985; Ahrens and Dieter, 1982; Philippe, 1997) for the purpose. An Accept-Reject sampling algorithm may suffer from high rate of rejection; especially when the rejection rate is not controlled uniformly for all cases of the parameter values. Keeping this in mind, we devise "uniformly efficient" Accept-Reject algorithms that ensure high performance for all possible choices of its parameter values. Thus, we provide a reliable sampling algorithm that would guarantee a sample irrespective of the parameter specifications that may occur during a real data analysis. Establishing uniformly efficient Accept-Reject sampling algorithms has been a nontrivial challenge and is a major theoretical component of the article, we contrive a unique approach for proving the result.

On this occasion, we would like to point out that the algorithms developed by Devroye (2014), Hörmann and Leydold (2014) are examples of the "uniformly efficient" Accept-Reject algorithms. On a related note, the MHN (α, β, γ) is a log-concave distribution when $\alpha \geq 1$. Therefore, the generic sampling techniques applicable for a log-concave density (Devroye, 2006; Gilks and Wild, 1992; Martino and Míguez, 2011) are also relevant for the case. On the other hand, the sampling algorithms developed in this article, though conceptually generalizable for other distributions, are optimally designed to achieve high efficiency for sampling from MHN (α, β, γ). Additionally, unlike many generic sampling procedure for the log-concave densities, the algorithm we have designed does not require the computation of its normalizing constant. About a particular case, note that the class of Truncated Normal distributions whose support is restricted to the positive part of the real line is contained in the Modified-Half-Normal family of distributions. If we focus our attention on that specific class of distributions, the performance of our algorithms to sampling from Truncated Normal distributions is comparable to the ones that are found by Robert (1995), Damien and Walker (2001).

In addition to the sampling algorithm, we explored a number of properties of the Modified-Half-Normal distribution in Section 2.3. We recognize its normalizing constant to be a Fox-wright function (Fox, 1928; Wright, 1935; Mehrez and Sitnik, 2019). The moments and the moment based statistics such as variance, skewness for the distribution can be represented in terms of the Fox-Wright functions. Besides, we identify a recursive relation between three consecutive moments of the distribution which is not only helpful in establishing a few key characteristics of the distribution but also plays an important role in estimating the parameters of MHN (α, β, γ) that we discuss in Section 2.5. Note that the family of Modified-Half-Normal distributions can be viewed as a generalization of multiple families including Half Normal, Truncated Normal, Square root of a Gamma and Gamma distributions (see Section 2.3).

Therefore, the family of distributions possess decent flexibility that can capture attributes of a real-valued positive data. Although it is not the primary objective, for the sake of completeness we briefly discussed the inferential techniques for analyzing data using the Modified-Half-Normal distribution.

As an additional component of this manuscript, we develop a few properties of the Fox-wright function that are required for studying the Modified-Half-Normal distribution. Depending on different parameter choices, we designed appropriate computation strategies for the Fox-Wright function that pertains to the current context. Reliable computation of the function is not only important to evaluate the basic statistics such as mean, variance, skewness, etc. of the Modified-Half-Normal distribution but also useful for computing the efficiency measures of the sampling algorithms in Sections 2.4. However, the implementation of the sampling algorithms do not require for its computation. A few nontrivial inequalities for the Fox-Wright function are also designed as they are required to establish the uniform efficiency of the developed Accept-Reject sampling algorithms.

2.1.1 A few applications of the Modified-Half-Normal distribution

In this section we consider a few examples to demonstrate the usefulness of the Modified-Half-Normal distribution in various statistical models. For the following Bayesian procedures, the $\mathrm{MHN}\,(\alpha, \beta, \gamma)$ distribution appears as one of the posterior conditional distributions that constitute the associated Markov Chain Monte Carlo algorithms. The examples aggregated here are inspired by diverse statistical problems, thus showing the spontaneity of its occurrence in literature.

Sandwich MCMC algorithm in Bayesian binary regression

Let $y_{R,1}, y_{R,2}, \ldots, y_{R,n}$ be independent Bernoulli random variables such that $P(y_{R,i} = 1 \mid \boldsymbol{\beta}_R) = H_{\nu_R}(\boldsymbol{x}_{R,i}^T \, \boldsymbol{\beta}_R)$ where $\boldsymbol{x}_{R,i} \in \mathbb{R}^p$ denotes the corresponding covari-

ates for $i = 1, \ldots n$. $\boldsymbol{\beta}_R \in \mathbb{R}^p$ is a vector of unknown regression coefficients and $H_{\nu_R}(\cdot)$ is the cumulative distribution function of the t-random variable with ν_R degrees of freedom. To develop an MCMC based posterior inference, it is required to sample from the conditional posterior density of $\boldsymbol{\beta}_R$, which is propositional to

$$\prod_{i=1}^{n} \left[H_{\nu_R}(\boldsymbol{x}_{R,i}^T \, \boldsymbol{\beta}_R) \right]^{y_{R,i}} \left[1 - H_{\nu_R}(\boldsymbol{x}_{R,i}^T \, \boldsymbol{\beta}_R) \right]^{1-y_{R,i}}, \tag{2.2}$$

where an improper flat prior is used for $\boldsymbol{\beta}_R$. In Equation 2.2, we use the notation $\boldsymbol{y}_R = (y_{R,1}, y_{R,2}, \ldots, y_{R,n})$. A sufficient condition for the property of the posterior in Equation 2.2 can be found in (Pal et al., 2015).

Albert and Chib (1993) developed a Data Augmentation (DA) algorithm for exploring the above posterior (Equation 2.2). The technique is based on introducing a set of latent variables $\Lambda_R := \left(\lambda_{R,1}, \ldots, \lambda_{R,n} \right)^T \in \mathbb{R}_+^n$ and $Z_R := (z_{R,1}, \ldots z_{R,n})^T \in \mathbb{R}^n$ with appropriate probability distributions, so that the full conditional posteriors of $\boldsymbol{\beta}_R, Z_R$ and Λ_R reduced to be Normal, Truncated Normal and Gamma distributions respectively (Albert and Chib, 1993; Pal et al., 2015).

It is well known that the DA algorithm in (Albert and Chib, 1993) can be slow to converge and often leads to MCMC samples with high auto correlation (Pal et al., 2015; Roy et al., 2012; Van Dyk and Meng, 2001). To improve the performance of data augmentation algorithms, *sandwich data augmentation* algorithms can be useful (Meng and Van Dyk, 1999; Liu and Wu, 1999; Marchev and Hobert, 2004). The idea is to expedite the MCMC convergence by introducing an additional draw from a scalar random variable so that the stationary distribution of the Markov chain remains same. We here denote it by G. Pal et al. (2015) proposed a suitable version of the *sandwich algorithm* that is applicable for the case where the probability density function of G

$$\pi_G(g \mid \boldsymbol{\beta}_R, Z_R, \Lambda_R, \boldsymbol{y}_R) \propto g^{n-1} e^{-\sum_{i=1}^{n} \frac{1}{2}\lambda_{R,i}(g \, z_{R,i} - \boldsymbol{x}_{R,i}^T \, \boldsymbol{\beta}_R)^2} \mathbb{I}(g > 0). \tag{2.3}$$

Based on the Equation 2.1, it follows that the distribution corresponding to Equation 2.3 is

$$
\text{MHN}\left(\cdot; n, \left(\frac{1}{2}\sum_{i=1}^{n}\lambda_{R,i}\, z_{R,i}^2\right), \left(\sum_{i=1}^{n}\lambda_{R,i}\; z_{R,i}\; \boldsymbol{x}_{R,i}^T\, \boldsymbol{\beta}_R\right)\right).
$$

Hence, an efficient sampler for the Modified-Half-Normal is also contributive to the performance of the above sandwich algorithm.

Inference with the Projected Normal distribution

Projected Normal distribution is a popular choice to model the directional observations, $\boldsymbol{y}_{\textbf{PN},i} \in \mathbb{S}^{p-1}$ for $i = 1 \ldots n$ where $\mathbb{S}^{p-1} = \left\{\boldsymbol{y} \in \mathbb{R}^p : \boldsymbol{y}^T\boldsymbol{y} = 1\right\}$ denotes the unit sphere. It is assumed that there are variables $\boldsymbol{x}_{\textbf{PN},i} \in \mathbb{R}^p$ such that $\boldsymbol{x}_{\textbf{PN},i} = r_{\textbf{PN},i}\, \boldsymbol{y}_{\textbf{PN},i}$ where $r_{\textbf{PN},i} > 0$, $\|\boldsymbol{y}_{\textbf{PN},i}\| = 1$. The latent variables $r_{\textbf{PN},i}$ represents the norm of the vector $\boldsymbol{x}_{\textbf{PN},i}$. The random vectors $\boldsymbol{x}_{\textbf{PN},i}$ are assumed to be independent and Normally distributed with mean $\boldsymbol{\mu}_{\textbf{PN}} \in \mathbb{R}^p$ and variance $\Sigma_{\textbf{PN}}$, a $p \times p$ positive definite matrix. For the identifiability of the model parameters, it is customary to fix the last diagonal element of $\Sigma_{\textbf{PN}}$ at the unity. In the context of the $MCMC$ based posterior inference for the model, as discussed in Hernandez-Stumpfhauser et al. (2017); Chakraborty et al. (2017), the probability density for the full conditional posterior of $r_{\textbf{PN},i}$ is

$$
\pi_{\textbf{PN},i}(r) \propto r^{p-1}e^{-r^2\left(\frac{1}{2}\boldsymbol{y}_i^T\Sigma_{\textbf{PN}}^{-1}\boldsymbol{y}_{\textbf{PN},i}\right)+r\left(\boldsymbol{y}_{\textbf{PN},i}^T\Sigma_{\textbf{PN}}^{-1}\boldsymbol{\mu}_{\textbf{PN}}\right)}\mathbb{I}(r > 0). \tag{2.4}
$$

It can be seen that the distribution in (2.4) corresponds to the MHN $(p, \frac{1}{2}\boldsymbol{y}_{\textbf{PN},i}^T\Sigma_{\textbf{PN}}^{-1}\boldsymbol{y}_{\textbf{PN},i},$ $\boldsymbol{y}_{\textbf{PN},i}^T\Sigma_{\textbf{PN}}^{-1}\boldsymbol{\mu}_{\textbf{PN}})$ distribution. We refer to Hernandez-Stumpfhauser et al. (2017) for more details on the model, parameter specifications, and the posterior sampling distributions. Therefore, the study of the Modified-Half-Normal distribution including an efficient sampler is relevant for this area of research as well.

Data augmentation algorithm for the vonMises-Fisher distribution

Let $\boldsymbol{y}_{\mathbf{VF},1}, \ldots \boldsymbol{y}_{\mathbf{VF},n} \in \mathbb{S}^{p-1}$, be the observed directional data on the p dimensional Euclidean space. VonMises-Fisher distribution with the probability density function (with respect to the appropriate Haar measure on $\mathbb{S}^{p-1}$),

$$f_{\mathbf{VF}}(\boldsymbol{y}) = \left(\frac{\kappa_{\mathbf{VF}}^{\nu} \, \exp\left(\kappa_{\mathbf{VF}} \, \boldsymbol{\mu}_{\mathbf{VF}}^{T} \boldsymbol{y}_{\mathbf{VF}} \right)}{(2\pi)^{p/2} \, I_{\nu}(\kappa_{\mathbf{VF}})} \right) \mathbb{I}(\boldsymbol{y}_{\mathbf{VF}} \in \mathbb{S}^{p}),$$

is one of the most popular distributions to model such data. In the above equation $\nu = p/2 - 1$ and $I_{\nu}(\cdot)$ denotes the Modified Bessel function of the first kind. The parameter $\boldsymbol{\mu}_{\mathbf{VF}}$ represent the modal direction while $\kappa_{\mathbf{VF}}$ controls the concentration of probability around the mode of the distribution. For a full Bayesian analysis for the model, a data augmentation algorithm can be designed where the independent random variables $T_{\mathbf{VF},1}, \ldots T_{\mathbf{VF},n}$ distributed according to the probability density function,

$$f_{T,\nu}(t; \kappa_{\mathbf{VF}}) = \frac{I_{\nu}(\kappa_{\mathbf{VF}})}{(2\kappa_{\mathbf{VF}})^{\nu}} \sum_{k=1}^{\infty} \frac{2^{\frac{2+\nu}{2}} j_{\nu,k}^{\nu+1}}{J_{\nu+1}(j_{\nu,k})} e^{-\left(j_{\nu,k}^2 + \kappa_{\mathbf{VF}}^2 \right)t} \, \mathbb{I}(t > 0),$$

is augmented to deal with the intractable normalizing constant involving the Bessel function of the first kind. If a $\mathrm{Gamma}(\alpha_{\mathbf{VF}}, \beta_{\mathbf{VF}})$ prior is used for the parameter $\kappa_{\mathbf{VF}}$ then the the conditional posterior distribution for $\kappa_{\mathbf{VF}}$ results in the following density

$$\pi_{\kappa_{\mathbf{VF}}}(\kappa) \propto \kappa^{\alpha_{\mathbf{VF}}-1} e^{-\left(n\overline{T}_{\mathbf{VF}} \right)\kappa^2 + \kappa\left(n\boldsymbol{\mu}_{\mathbf{VF}}^{T}\overline{Y}_{\mathbf{VF}} - \beta_{\mathbf{VF}} \right)} \mathbb{I}(\kappa > 0), \qquad (2.5)$$

where $\overline{T}_{\mathbf{VF}} = \sum_{i=1}^{n} T_{\mathbf{VF},i}/n$ and $\overline{Y}_{\mathbf{VF}} = \sum_{i=1}^{n} \boldsymbol{y}_{\mathbf{VF},i}/n$. The full conditional distribution of $\kappa_{\mathbf{VF}}$ given the other variables follow a

$$\mathrm{MHN}\left(\alpha_{\mathbf{VF}}, \ n\overline{T}_{\mathbf{VF}}, \ n\boldsymbol{\mu}_{\mathbf{VF}}^{T}\overline{Y}_{\mathbf{VF}} - \beta_{\mathbf{VF}} \right)$$

(see Equation 2.5 and 2.1). Note that the Modified-Half-Normal distribution appears even if the conjugate prior is used for the pair $(\boldsymbol{\mu}_{\mathbf{VF}}, \kappa_{\mathbf{VF}})$. For the efficiency of the above data augmentation algorithm, it is imperative to have an efficient sampling scheme for the Modified-Half-Normal distribution. Note that the described data augmentation technique is a newly developed procedure and the sampling from the MHN (α, β, γ) distribution plays a crucial role in the related Bayesian procedures.

Application in robust bayesian graphical model

Consider the robust Bayesian graphical model as it is outlined in Finegold et al. (2014). The objective of this model is to learn the covariance structure Ψ_G of a set of multivariate observations assuming that the corresponding concentration matrix is likely to be a sparse positive definite matrix. To describe the model, let $\boldsymbol{y}_{\mathbf{G},1}, \boldsymbol{y}_{\mathbf{G},2}, \dots,$ $\boldsymbol{y}_{\mathbf{G},n} \in \mathbb{R}^p$ denotes the random sample from the alternative-t distribution (Finegold et al., 2014). To facilitate the parameter estimation in a tractable manner, Finegold et al. (2014) introduce a set of latent vectors that we denote by $\boldsymbol{\tau}_{\mathbf{G},1}, \dots, \boldsymbol{\tau}_{\mathbf{G},n} \in \mathbb{R}_+^p$ where $\boldsymbol{\tau}_{\mathbf{G},i} = \left(\tau_{\mathbf{G},i,1}, \dots, \tau_{\mathbf{G},i,p}\right)^T$ for $i = 1, \dots, n$. The model can be represented as

$$
\begin{aligned}
\boldsymbol{y}_{\mathbf{G},i} \mid \boldsymbol{\mu}_{\mathbf{G},i}, \tau_{\mathbf{G},i}, \Psi_{\mathbf{G}} &\sim \text{Normal}\left(\boldsymbol{\mu}_{\mathbf{G},i}, D_{\mathbf{G},\tau,i}^{-1} \Psi_{\mathbf{G}} D_{\mathbf{G},\tau,i}^{-1}\right) \text{ for } i = 1, \dots n, \\
\boldsymbol{\tau}_{\mathbf{G},i,j}^2 &\overset{iid}{\sim} \text{Gamma}(\frac{\nu_{\mathbf{G}}}{2}, \frac{\nu_{\mathbf{G}}}{2}),
\end{aligned}
$$

for $i = 1, \dots n$ and $j = 1, \dots p$ where $D_{\mathbf{G},\tau,i}$ denotes the diagonal matrix with the diagonal elements $\boldsymbol{\tau}_{\mathbf{G},i}$. The sparsity pattern governed by the concentration matrix $\Theta_{\mathbf{G}} := \Psi_{\mathbf{G}}^{-1}$ corresponds to a graph which represents the conditional independence structure between the different coordinates of the observations. For Bayesian formulation of the model, the prior distribution for the parameter Ψ_G is assumed to be Hyper Inverse Wishart distribution (Finegold et al., 2014; Roverato, 2002) where $\boldsymbol{\mu}_{\mathbf{G}}$ is considered to be Normally distributed. A Gibbs sampling algorithm is employed for

posterior inference where the sampling density of the parameters $\tau_{\mathbf{G},i,j}$ is proportional to

$$\tau^{\nu_{\mathbf{G}}-1} e^{-\beta_{\mathbf{G}}\tau^2 + \gamma_{\mathbf{G}}\tau}, \tag{2.6}$$

for appropriate values of $\beta_{\mathbf{G}}$ and $\gamma_{\mathbf{G}}$ (see Finegold et al. (2014) for details). Thus, the probability density in Equation 2.6 corresponds to a Modified-Half-Normal distribution.

So far in this section, we have discussed a few examples related to the topics of co-variance estimation, analysis of directional data and the Bayesian Binary regression. They shows the diversity of the applications where the Modified-Half-Normal distribution is utilized. The authors believe that the spontaneity for the occurrence of the MHN $(,,)$ is due to the similarity of its probability density with that of the Normal distribution and it will be useful for many other occasions apart from the examples described above.

2.2   The Fox-Wright function and a few of its properties

In this section, we present the general form of the Fox-Wright function (Fox, 1928; Wright, 1935) and identify a specific case that is required for studying the Modified-Half-Normal distribution. We include a brief discourse to familiarize the related notions and to investigate certain aspects of the function. If not interested, the reader may only consider the statements of the Lemmas and move to the next Section of the article. The general class of the Fox-Wright function, introduced by Fox (1928) and Wright (1935), is defined as

$$
{}_p\Psi_q \left[\begin{matrix} (a_1, A_1) & (a_2, A_2) & \dots & (a_p, A_p) \\ (b_1, B_1) & (b_2, B_2) & \dots & (b_q, B_q) \end{matrix} ; z \right] = \sum_{n=0}^{\infty} \frac{\Gamma(a_1 + A_1 n) \cdots \Gamma(a_p + A_p n)}{\Gamma(b_1 + B_1 n) \cdots \Gamma(b_q + B_q n)} \frac{z^n}{n!},
$$

where p, q are non-negative integers, $A_l \geq 0$ for $l = 1 \ldots p$ and $B_m \geq 0$ for $m = 1, \ldots q$. The constants a_l for $l = 1, \ldots, p$ and b_m for $m = 1, \ldots, q$ are allowed to be any nonzero complex numbers (Craven and Csordas, 2006; Mainardi and Pagnini, 2007), but for this manuscript we only need the case when they are positive real numbers. The above series converges absolutely for $z \in \mathbb{C}$ if $\sum_{l=1}^{q} B_l - \sum_{m=1}^{p} A_m + 1$. The usage of the function is seen in different branches of science including Mathematics (Mainardi and Pagnini, 2007), Physics, and Finance. Many of its properties are documented for years while the related scientific investigation is still an active area of research (Wang (2019); Wei (2019)). In the backdrop of the principal objectives of the current article, we streamline our focus on the

$$
{}_1\Psi_1 \left[\begin{array}{c} (\frac{\alpha}{2}, \frac{1}{2}) \\ (1, 0) \end{array} ; x \right] = \sum_{n=0}^{\infty} \frac{\Gamma(\frac{\alpha}{2} + \frac{1}{2}n)}{n!} x^n, \tag{2.7}
$$

a specific case of the Fox-Wright function which is required not only to represent the normalizing constant of the MHN (α, β, γ) distribution (see Section 2.3) but also to apprise the efficiency of the rejection sampling algorithms developed in Section 2.3. Additionally, the availability of computational algorithms for ${}_1\Psi_1 \left[\begin{array}{c} (\frac{\alpha}{2}, \frac{1}{2}) \\ (1, 0) \end{array} ; x \right]$ and/or the ratio of the appropriate Fox-wright functions are crucial for computing mean, variance, skewness and other moment based statistics (see Section 2.3 for details).

We would like to mention that the Fox-Wright Functions are closely connected with ${}_pF_q$, the hypergeometric functions (Srivastava, 2007; Mehrez and Sitnik, 2019). Specifically, if $A_1, ..., A_p = 1$ and $B_1, ..., B_p = 1$ then

$$
{}_p\Psi_q \left[\begin{array}{c} (a_1, 1) \ldots (a_p, 1) \\ (b_1, 1) \ldots (b_q, 1) \end{array} ; z \right] = \frac{\Gamma(a_1) \cdots \Gamma(a_p)}{\Gamma(b_1) \cdots \Gamma(b_q)} {}_pF_q \left[\begin{array}{c} a_1 \ldots a_p \\ b_1 \ldots b_q \end{array} ; z \right]. \tag{2.8}
$$

As a result, many of the attributes including computation of the Fox-Wright function

follows directly from that of the $_pF_q$ whenever the functional equality in Equation 2.8 is applicable. Unfortunately, it is not applicable to the specific case of the Fox-Wright function (as in Equation 2.7) that corresponds to the MHN (α, β, γ) distribution. Therefore, we invested a nontrivial effort to design strategies for its computation which is required for the subsequent developments. Additionally, we develop a few generic properties of Fox-Wright Functions that we utilize in the later sections.

Lemma 1. *Let* $_1\Psi_1 \begin{bmatrix} (\frac{\alpha+n}{2}, \frac{1}{2}) \\ ; x \\ (1,0) \end{bmatrix}$ *denotes the Fox-Wright Function. The following functional properties holds for all* $x \in \mathbb{R}$.

(a) *If* $\alpha + n > 2$ *then*

$$
_1\Psi_1 \begin{bmatrix} (\frac{\alpha+n}{2}, \frac{1}{2}) \\ ; x \\ (1,0) \end{bmatrix} = \frac{\alpha+n-2}{2} {}_1\Psi_1 \begin{bmatrix} (\frac{\alpha+n-2}{2}, \frac{1}{2}) \\ ; x \\ (1,0) \end{bmatrix} + \frac{x}{2} {}_1\Psi_1 \begin{bmatrix} (\frac{\alpha+n-1}{2}, \frac{1}{2}) \\ ; x \\ (1,0) \end{bmatrix}.
$$

(b) *If* $\alpha > 1$ *then*

$$
\frac{{}_1\Psi_1 \begin{bmatrix} (\frac{\alpha+1}{2}, \frac{1}{2}) \\ ; x \\ (1,0) \end{bmatrix}}{{}_1\Psi_1 \begin{bmatrix} (\frac{\alpha}{2}, \frac{1}{2}) \\ ; x \\ (1,0) \end{bmatrix}} = \frac{x}{2} + \frac{\alpha-1}{2} \cfrac{1}{\frac{x}{2} + \cdots \cfrac{1}{\frac{x}{2} + \frac{1+\delta_\alpha}{2} \cfrac{\frac{1}{{}_1\Psi_1 \begin{bmatrix} (\frac{1+\delta_\alpha}{2}, \frac{1}{2}) \\ ;x \\ (1,0) \end{bmatrix}}}{{}_1\Psi_1 \begin{bmatrix} (\frac{\delta_\alpha}{2}, \frac{1}{2}) \\ ;x \\ (1,0) \end{bmatrix}}}},
$$

$\delta_\alpha = 1$ *if* α *is an integer otherwise* $\delta_\alpha = \alpha - \lfloor \alpha \rfloor$ *where* $\lfloor \alpha \rfloor$ *denotes the largest integer less than equal to* α.

14

(c) If we denote the cumulative distribution function of the standard normal distribution as $\Phi(\cdot)$ then,

$$\dfrac{{}_1\Psi_1\left[\begin{array}{c}\left(\frac{2}{2},\frac{1}{2}\right)\\(1,0)\end{array};x\right]}{{}_1\Psi_1\left[\begin{array}{c}\left(\frac{1}{2},\frac{1}{2}\right)\\(1,0)\end{array};x\right]} = \frac{x}{2} + \frac{1}{2\sqrt{\pi}}\frac{\exp(-\frac{x^2}{4})}{1-\Phi(-\frac{x}{\sqrt{2}})}, \quad and \ {}_1\Psi_1\left[\begin{array}{c}\left(\frac{1}{2},\frac{1}{2}\right)\\(1,0)\end{array};x\right] =$$

$2\sqrt{\pi}e^{\frac{x^2}{4}}\left(1-\Phi(-\frac{x}{\sqrt{2}})\right).$

The proof of part(a) of the Lemma 1 relies on the recurrence relation of the Gamma function while part(b) is a consequence of the part(a). The proof of part (c) utilizes the properties of the truncated normal distribution.

Lemma 2. *Let* $\alpha > 0, x \in \mathbb{R}$ *then for arbitrary* $\epsilon > 0$, *let* $A(k) = \frac{\Gamma(\frac{\alpha}{2}+k)x^{2k}}{(2k)!}$ *and* $B(k) = \frac{\Gamma(\frac{\alpha+1}{2}+k)x^{2k+1}}{(2k+1)!}$. *Let* $\lceil x \rceil$ *denotes the smallest integer larger than* x. *For a given constant* $0 < q < 1$

(a) $\frac{A(k+1)}{A(k)} < q$ *if* $k > C_1$, *where* $s_1 = (6q - x^2)$

$$C_1 = \max\left\{\left\lceil \frac{-s_1 + \sqrt{s_1^2 - 8q(4q - \alpha x^2)}}{8q}\right\rceil, 1\right\},$$

(b) $\frac{B(k+1)}{B(k)} < q$ *if* $k > C_2$, *where* $s_2 = (10q - x^2)$,

$$C_2 = \max\left\{\left\lceil \frac{-s_2 + \sqrt{s_2^2 - 8q(12q - (\alpha+1)x^2)}}{8q}\right\rceil, 1\right\}.$$

(c) $A(k)$ *is strictly decreasing function when* $k \geq C_1$ *and* $B(k)$ *is strictly decreasing function when* $k \geq C_2$. *Moreover* $\lim_{k\to\infty} A(k) = 0$ *and* $\lim_{k\to\infty} B(k) = 0$.

(d) *For a given* $\epsilon > 0$, *if* $K = \max\{K_1, K_2\}$, $K_1 = \max\{\min\{k : A(K) \leq \frac{\epsilon}{4}$ *for all* $K \geq k\}, C_1\}$, $K_2 = \max\{\min\{k : |B(K)| \leq \frac{\epsilon}{4}$ *for all* $K \geq k\}, C_2\}$ *then*

15

$$\left| {}_1\Psi_1 \begin{bmatrix} (\frac{\alpha}{2}, \frac{1}{2}) \\ (1, 0) \end{bmatrix} ; x \right] - \left(\sum_{k=0}^{K} A(k) + \sum_{k=0}^{K} B(k) \right) \Bigg| \le \epsilon.$$

The Lemma 2 provides a way to compute ${}_1\Psi_1 \begin{bmatrix} (\frac{\alpha}{2}, \frac{1}{2}) \\ (1, 0) \end{bmatrix} ; x \right]$ by truncation of a infinite series in a systematic manner. Specifically the part(d) of the Lemma identifies a truncation point K so that the finite summation $\sum_{k=0}^{K} A(k) + \sum_{k=0}^{K} B(k)$ can be used to approximate the intended function with a maximum predetermined error bound $\epsilon > 0$.

In the case when $\gamma < 0$, the series approximation procedure appears to be inefficient. From empirical experiments, we observed that the accumulated errors contributed by the computation of the each terms of the $A(k)$ and $B(k)$ (computation error that involves 'lgamma' function that is implemented 'R') appears to be significant compared to the functional value of Fox-Wright function. Hence, we consider an alternative procedure premised on numerical integration. The strategy appears to perform efficiently for computing the function when $\gamma < 0$. In the following lemma, we truncate the infinite integral to that over a finite region in a systematic manner so that the error of the approximation can be controlled.

Lemma 3. *Let $\alpha > 0, \beta > 0$ and $\gamma < 0$. For any $m > 0$,*

$$\left| {}_1\Psi_1 \begin{bmatrix} (\frac{\alpha}{2}, \frac{1}{2}) \\ (1, 0) \end{bmatrix} ; \frac{\gamma}{\sqrt{\beta}} \right] - 2\beta^{\frac{\alpha}{2}} \int_{0}^{M_\epsilon} x^{\alpha-1} e^{-\beta x^2 - |\gamma| x} dx \right| < \epsilon.$$

when $M_\epsilon > \frac{1}{b} \gamma^{-1} \left(a, \frac{[\Gamma(a)]^2}{b^a} - \frac{\epsilon \Gamma(a)(2\beta m + |\gamma|)}{2\beta^{\frac{\alpha}{2}} m^\alpha (\beta m + |\gamma|)} \right)$. The constants $a = \frac{\alpha(\beta m + |\gamma|)}{2\beta m + |\gamma|}$, $b = \beta m^2 + |\gamma| m$, and the function $\gamma(a, bM_\epsilon)$ is a lower incomplete gamma function.

An implementation for the inverse of the lower incomplete gamma function is available in the R package 'zipfR' (Evert and Baroni, 2007). So far, we have developed

properties of the Fox-Wright function primarily aiming at its numerical computation for different cases. The following result provides functional approximation to the Fox-Wright function.

Lemma 4. *Let $\alpha > 1, \beta > 0$.*

(a) If $\gamma \leq 0$ then for any $m > 0$

$$
{}_1\Psi_1 \left[\begin{matrix} (\frac{\alpha}{2}, \frac{1}{2}), \\ (1, 0) \end{matrix}; \frac{-|\gamma|}{\sqrt{\beta}} \right] \geq \frac{\beta^{\frac{\alpha}{2}} \exp\left(-\frac{m|\gamma|}{2}\right) \Gamma(\frac{\alpha}{2})}{\left(\beta + \frac{|\gamma|}{2m}\right)^{\frac{\alpha}{2}}}.
$$

(b) If $\gamma > 0$ then for any positive $\alpha_0 > 0$ and $p, q > 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$,

$$
{}_1\Psi_1 \left[\begin{matrix} (\frac{\alpha}{2}, \frac{1}{2}), \\ (1, 0) \end{matrix}; \gamma \right] \geq \frac{2p^{\frac{\alpha+(2-\alpha_0)p-1}{2} - \frac{p}{q}} \left[\frac{\Gamma\left(\frac{\alpha+p\alpha_0-1}{2p}\right)}{2} \right]^p (q\gamma)^{p(\alpha_0-1)+\frac{p}{q}}}{\left[\Gamma\left(q\alpha_0 - q + 1\right) \right]^{\frac{p}{q}}}.
$$

In particular if $p = \frac{\alpha}{\alpha-1}, q = \alpha, \alpha_0 = \frac{\alpha+1}{p}$ then

$$
{}_1\Psi_1 \left[\begin{matrix} (\frac{\alpha}{2}, \frac{1}{2}), \\ (1, 0) \end{matrix}; \gamma \right] \geq \frac{[\Gamma(\alpha-1)]^{\frac{\alpha}{\alpha-1}} (\alpha\gamma)^\alpha}{[2\Gamma(\alpha(\alpha-1))]^{\frac{1}{\alpha-1}}}. \tag{2.9}
$$

The Lemma is utilized for establishing uniform efficiency of the Accept-Reject sampling algorithms that are designed later in this article.

2.3 Properties of the Modified-Half-Normal distribution

In this section, we survey a few properties of the Modified-Half-Normal distribution. We begin with the observation that the MHN (α, β, γ) is an exponential family of distributions (Casella and Berger, 2002). Therefore, the generic properties of the exponential family of distributions are automatically applicable to the

MHN (α, β, γ) distribution as well. Apart from those, we include a few specific characteristics of distribution in the next few Lemmas. By definition, the density function of the MHN (α, β, γ) distribution is proportional to the function $x \mapsto x^{\alpha-1} \exp(-\beta x^2 + \gamma x) \mathbb{I}(x > 0)$, while we identify the corresponding normalizing constant and a form of its cumulative distribution function in the result below.

Lemma 5. *Let $f_{MHN}(x \mid \alpha, \beta, \gamma)$ and $F_{MHN}(\cdot \mid \alpha, \beta, \gamma)$ denotes the probability density and the cumulative distribution function of the MHN(α, β, γ) distribution for $\alpha > 0, \beta > 0, \gamma \in \mathbb{R}$.*

(a) The corresponding normalizing constant is given as $\frac{1}{2\beta^{\frac{\alpha}{2}}} {}_1\Psi_1 \left[\begin{matrix} \left(\frac{\alpha}{2}, \frac{1}{2}\right) \\ (1, 0) \end{matrix}; \frac{\gamma}{\sqrt{\beta}} \right]$, and

$$f_{MHN}(x \mid \alpha, \beta, \gamma) = \frac{2\beta^{\frac{\alpha}{2}} x^{\alpha-1} \exp(-\beta x^2 + \gamma x)}{{}_1\Psi_1 \left[\begin{matrix} \left(\frac{\alpha}{2}, \frac{1}{2}\right) \\ (1, 0) \end{matrix}; \frac{\gamma}{\sqrt{\beta}} \right]} \mathbb{I}(x > 0).$$

(b) The cumulative distribution function

$$F_{MHN}(x \mid \alpha, \beta, \gamma)$$
$$= \frac{2\beta^{\frac{\alpha}{2}}}{{}_1\Psi_1 \left[\begin{matrix} \left(\frac{\alpha}{2}, \frac{1}{2}\right) \\ (1, 0) \end{matrix}; \frac{\gamma}{\sqrt{\beta}} \right]} \sum_{i=0}^{\infty} \frac{\gamma^i}{2i!} \beta^{-\frac{\alpha+i}{2}} \gamma\left(\frac{\alpha+i}{2}, \beta x^2\right),$$

where $\gamma(s, y) = \int_0^y t^{s-1} e^{-t} dt$, denotes the lower incomplete gamma function, and ${}_1\Psi_1 \left[\begin{matrix} \left(\frac{\alpha}{2}, \frac{1}{2}\right) \\ (1, 0) \end{matrix}; \frac{\gamma}{\sqrt{\beta}} \right]$ denotes the Fox-Wright function of the appropriate order (see Section 2.2).

The Figure 2.1 displays a few densities of the MHN (α, β, γ) distribution for different choices of $\alpha \geq 1$. The density become more similar to that of a Normal

probability density in appearance when the parameter γ is large. In Figure 2.1, we
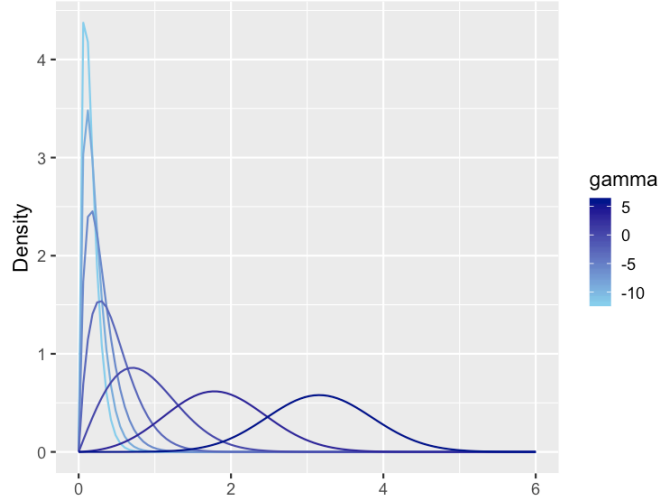
Figure 2.1: Density plot of the MHN (α, β, γ) distribution when $\alpha > 1$.

used a fixed positive β when similar plots can be seen if other values of $\beta > 0$ and $\alpha > 1$ are chosen. For the case when $\alpha < 1, \gamma > 0$, the distribution appears not to be unimodal. Before we study the modal characteristics of the distribution in Lemma 7, we include the following properties related to its moments.

Lemma 6. *Let* $X \sim MHN(\alpha, \beta, \gamma)$ *then for* $k \geq 0$, *then*

(a) *Assuming* $\alpha + k$ *to be a positive real number, the* k^{th} *moment of the distribution*

$$
E(X^k) = \frac{{}_1\Psi_1 \left[\begin{matrix} \left(\frac{\alpha+k}{2}, \frac{1}{2}\right) \\ (1, 0) \end{matrix} ; \frac{\gamma}{\sqrt{\beta}} \right]}{\beta^{\frac{k}{2}} \, {}_1\Psi_1 \left[\begin{matrix} \left(\frac{\alpha}{2}, \frac{1}{2}\right) \\ (1, 0) \end{matrix} ; \frac{\gamma}{\sqrt{\beta}} \right]}.
$$

(b) *If* $\alpha + k > 0$, *then* $E(X^{k+2}) = \frac{\alpha+k}{2\beta} E(X^k) + \frac{\gamma}{2\beta} E(X^{k+1})$.

(c) *The variance of the distribution*

$$
Var(X) = \frac{\alpha}{2\beta} + E(X) \left(\frac{\gamma}{2\beta} - E(X) \right).
$$

(d) The moment generating function of the distribution is given as

$$M_X(t) = \frac{{}_1\Psi_1\left[\begin{matrix}\left(\frac{\alpha}{2},\frac{1}{2}\right) \\ (1,0)\end{matrix} ; \frac{\gamma+t}{\sqrt{\beta}}\right]}{{}_1\Psi_1\left[\begin{matrix}\left(\frac{\alpha}{2},\frac{1}{2}\right) \\ (1,0)\end{matrix} ; \frac{\gamma}{\sqrt{\beta}}\right]}.$$

It is evident from Lemma 6 that all finite positive moments of the distribution are finite and can be represented using the ratio of the Fox-Wright functions. Regarding its computation, we refer to the procedures developed in Section 2.2. The moments have uncomplicated representation (see part(b) of Lemma 1) when the parameter α is a positive integer. The part(b) of the Lemma 6 provides a nontrivial recursive relation between the different moments of the distribution. The property played a key role in estimating the parameters of the Modified-Half-Normal distribution that we discuss in Section 2.5. The variance of the distribution can be represented in terms of the Fox-Wright functions using the part(a) whereas an implication of the part(c) is that the variance of the distribution is bounded above by $\alpha/(2\beta)$ if $\gamma < 0$.

Though it is an attribute inherited from the exponential family of distributions, we would like to point out a resemblance between the form of its moments and that of the Generalized Inverse Gaussian (GIG) distribution (Balakrishnan and Johnson, 1994). The moments of the GIG distribution can be represented as the ratio of Modified Bessel functions of the second kind (Abramowitz et al., 1988), that also appears in the normalizing constant of the GIG distribution. Likewise, for the MHN (α, β, γ) distribution, the moments are the ratios of Fox-Wright functions which corresponds to its normalizing constant. In the next lemma, we characterize the mode of the distribution.

Lemma 7. *Consider the MHN (α, β, γ) with $\alpha > 0$, $\beta > 0$ and $\gamma \in \mathbb{R}$.*

(a) *The probability density function of the distribution is log-concave if $\alpha \geq 1$.*

(b) *The mode of the distribution is $\frac{\gamma + \sqrt{\gamma^2 + 8\beta(\alpha-1)}}{4\beta}$ if $\alpha > 1$.*

(c) *If $\gamma > 0$ and $1 - \frac{\gamma^2}{8\beta} \leq \alpha < 1$ then the density has a local maxima at $\frac{\gamma + \sqrt{\gamma^2 + 8\beta(\alpha-1)}}{4\beta}$ and a local minima at $\frac{\gamma - \sqrt{\gamma^2 + 8\beta(\alpha-1)}}{4\beta}$.*

(d) *The density function is gradually decresing on $\mathbb{R}_+$ and mode of the distribution doesn't exist, if either $\gamma > 0$, $0 < \alpha < 1 - \frac{\gamma^2}{8\beta}$ or $\gamma < 0, \alpha \leq 1$.*

From the part(c) as wells as Figure 2.2, it can be seen that, the density function is not unimodal when $\alpha < 1$ and $\gamma > 0$. In terms of the shape of the density, it lacks the parity compared to the other cases of the parameters α, γ. Although, for the sake of completeness, we keep the distribution in the family, the specific case does not occur in the applications that we have included in Section 2.1.1.
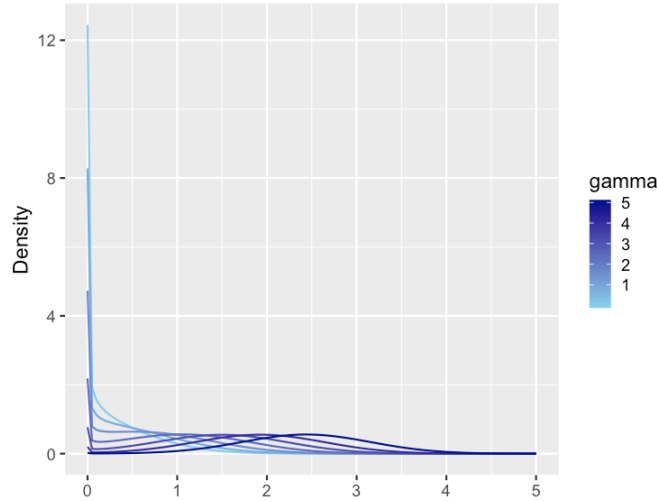
Figure 2.2: Density plot of the MHN (α, β, γ) distribution when $\alpha < 1$ and $\gamma > 0$.

Lemma 8. *Let $X \sim MHN(\alpha, \beta, \gamma)$ for $\alpha \geq 1$. $X_{mode} = \frac{\gamma + \sqrt{\gamma^2 + 8\beta(\alpha-1)}}{4\beta}$ denotes the mode of the distribution.*

(a) For all $\gamma \in \mathbb{R}$ if $\alpha > 1$ then,

$$X_{mode} \leq E(X) \leq \frac{\gamma + \sqrt{\gamma^2 + 8\alpha\beta}}{4\beta}.$$

(b) If $\gamma > 0$ and $\alpha \geq 4$,

$$\log(X_{mode}) \leq E(\log(X)) \leq \log\left(\frac{\gamma + \sqrt{\gamma^2 + 8\alpha\beta}}{4\beta}\right).$$

The difference between the upper and lower bound provided in part(a) approaches to zero as α gets larger. Therefore, part(a) of the lemma also provides high precision approximation of $E(X)$ when α is large. Regarding part(b), the lower bound does not follow from the lower bound in the part(a). Also, the condition $\alpha \geq 4$ is a sufficient condition for its validity. On the contrary, the upper bounds in part(a) as well as in part(b) are true for all $\alpha > 0$. An implication of the fact $E(X) \geq X_{\text{mode}}$ is that the distribution is positively skewed. Both the parts of the lemma plays a crucial role for apprising efficiency of the Accept-Reject sampling algorithms that we develop in the Section 2.4.

Lemma 9. *Let $X \sim MHN(\alpha, \beta, \gamma)$.*

(a) If $\gamma > 0$ then there exists a random variable V such that

$$V \mid X \sim Poission(\gamma X), \;\; and$$

$$X^2 \mid V \sim Gamma(\frac{\alpha + V}{2}, \beta).$$

(b) If $\gamma < 0$ then there exists a random variable U such that

$$U \mid X \sim GIG(\frac{1}{2}, 1, \gamma^2 X^2), \;\; and$$

$$X^2 \mid U \sim Gamma(\frac{\alpha}{2}, (\beta + \frac{\gamma^2}{U})).$$

The above result can be utilized to design hierarchical models by introducing additional variables U, V that can bypass the need for sampling from the MHN (α, β, γ) distribution. But the strategy of introducing additional latent variables may delay the mixing of the corresponding Markov chains. We conclude this section by identifying a few known families of distributions to be the special case of the Modified-Half-Normal family of distributions.

Lemma 10. *Let* $X \sim MHN(\alpha, \beta, \gamma)$.

(a) *If* $\gamma = 0$ *then* X *is distributed as a square root of a gamma distribution, i.e.* $X^2 \sim Gamma(\frac{\alpha}{2}, \beta)$.

(b) *If* $\alpha = 1$ *then* $X \sim Truncated\text{-}Normal(\frac{\gamma}{2\beta}, \frac{1}{\sqrt{2\beta}}, 0, \infty)$ *where the support of the distribution is restricted to* $(0, \infty)$.

(c) *If* $\alpha = 1$ *and* $\gamma = 0$ *then* $X \sim Half\text{-}Normal(\frac{1}{\sqrt{2\beta}})$.

Furthermore, if we allow the parameter β to take the value zero, then the case $\gamma < 0$ corresponds to the Gamma distribution with shape parameter α and rate parameter $-\gamma$.

2.4 Algorithms for generating random samples from the Modified-Half-Normal distribution

We devote this section for developing algorithms to sample from the Modified-Half-Normal distribution. The ability to sample efficiently from the distribution is crucial for the success of the Bayesian statistical models that bring about the MHN (α, β, γ) distribution as one of the posterior conditionals. In terms of the technique, we employ the rejection sampling strategy (Devroye (2006); Casella (1985);

Ahrens and Dieter (1982); Philippe (1997)) for the purpose. The technique is widely used while the associated notations are general. Therefore, we include a brief overview of the generic procedure along with the specifications for the particular terminologies that we use in the subsequent sections.

The rejection sampling algorithm is a technique to generate a random sample from a distribution that may not be sampled using a more direct approach. It is also commonly referred to as the "Accept-Reject Algorithm" (Casella et al., 2004). Let us assume that we need to sample from a distribution with the probability density function (or the probability mass function) $f(x)$. In order to do so, it is required to find a density $g(x)$ such that $f(x) \leq Mg(x)$ for all x in the support of the target density $f(x)$ and it is easy to sample from the distribution with probability density function $g(x)$ (Casella and Berger, 2002; Casella et al., 2004). The notation M is a suitable positive constant. Often, the above notion is implemented by finding a density kernel (or probability mass kernel) $g_{\text{kernel}}(x)$ so that $f(x) \leq g_{\text{kernel}}(x)$ for all x and $\int g_{\text{kernel}}(x)\mu(dx) < \infty$, where $\mu(\cdot)$ denotes the Lebesgue measure or appropriate counting measure depending on the context. Consequently we can get a proposal density $g(x)$ (or p.m.f) that is proportional to $g_{\text{kernel}}(x)$. Later in this article, we will refer to the function $g_{\text{kernel}}(x)$ by the proposal kernel and the associated density function as the proposal density.

Once an appropriate proposal density $g(x)$ is found, the steps involved to sample from the distribution corresponding to $f(x)$ is as follows; Sample $X \sim g(x)$ and $U \sim \text{Uniform}(0,1)$; If $\frac{f(X)}{g(X)} \leq U$ then accept X as a valid random sample from $f(x)$ otherwise reject it and repeat the procure from the beginning until a valid sample is obtained. The efficiency of the rejection sampling algorithm can be measured by the acceptance probability that transpires to be the reciprocal of the constant M (Casella and Berger, 2002). The acceptance probability, M^{-1}, can also be expressed as $\left(\int g_{\text{kernel}}(x)\mu(dx)\right)^{-1}$. A major component to design an efficient Accept-

Reject sampling algorithm is to find an appropriate $g_{\text{kernel}}(x)$ so that the corresponding acceptance probability remains large for all possible choices of the parameters of $f(x)$.

In the specific case of MHN (α, β, γ) distribution, we had to use different strategies for constructing the proposal kernels depending on the sign of the parameter γ. The description of the particularities associated with the algorithms are elaborated in the remaining parts of this section.

2.4.1 Sampling from the MHN (α, β, γ) when $\gamma > 0$.

From Lemma 7, we know that the MHN (α, β, γ) distribution is unimodal if $\alpha \geq 1$ and $\gamma > 0$ while its density appears to be idiosyncratic when $\alpha < 1$ and $\gamma > 0$. We first focus on the former case and develop the Accept-Reject sampling algorithm in Section 2.4.1 whereas we construct an alternative procedure in the Section 2.4.1 that is applicable for $\alpha > 0, \gamma > 0$ and works effectively for the case $\gamma > 0$ and $0 < \alpha < 1$ in particular.

Sampling from the MHN (α, β, γ) when $\gamma > 0, \alpha > 1$.

In this section, we discuss the Accept-Reject sampling algorithms to generate random samples from MHN (α, β, γ) which works best when $\alpha > 1, \beta > 0$ and $\gamma > 0$. We utilize either the square root of the Gamma or the Normal distribution as the candidates for the proposal distribution. The following Lemma provides the foundation for the algorithm.

Theorem 1. *Let $f_{MHN}(x \mid \alpha, \beta, \gamma)$ be the probability density function of MHN(α, β, γ) while $f_{Normal}, f_{\sqrt{Gam}}$ are probability density functions of the Normal and the square root of a Gamma distribution specified as*

$$f_{Normal}(x \mid \mu, \frac{1}{2\beta}) = \frac{\sqrt{\beta}}{\sqrt{\pi}} \exp\{-\beta(x - \mu)^2\},$$

$$f_{\sqrt{Gam}}(x \mid \alpha, \delta) = \frac{x^{\alpha-1}\delta^{\frac{\alpha}{2}}e^{-\delta x^2}}{\Gamma(\frac{\alpha}{2})}.$$

If $\gamma \geq 0$ then ,

(a) for any constants $\mu \in (\frac{\gamma}{2\beta}, \infty)$ and $\delta \in (0, \beta)$

$$f_{MHN}(x \mid \alpha, \beta, \gamma) \leq IK_1(\mu, \alpha, \beta, \gamma) \left\{ f_{Normal}(x \mid \mu, \frac{1}{2\beta}) \right\} + (1-I)K_2(\delta, \alpha, \beta, \gamma) \left\{ f_{\sqrt{Gam}}(x \mid \alpha, \delta) \right\}$$

where

$$K_1(\mu, \alpha, \beta, \gamma) = \frac{2\sqrt{\pi}\left(\frac{\sqrt{\beta}(\alpha-1)}{2\beta\mu-\gamma}\right)^{\alpha-1}e^{-(\alpha-1)+\beta\mu^2}}{{}_1\Psi_1\left[\begin{matrix}(\frac{\alpha}{2},\frac{1}{2})\\(1,0)\end{matrix};\frac{\gamma}{\sqrt{\beta}}\right]},$$

$$K_2(\delta, \alpha, \beta, \gamma) = \frac{\left(\sqrt{\beta}\right)^\alpha \Gamma(\frac{\alpha}{2})\,\exp\left(\frac{\gamma^2}{4(\beta-\delta)}\right)}{{}_1\Psi_1\left[\begin{matrix}(\frac{\alpha}{2},\frac{1}{2})\\(1,0)\end{matrix};\frac{\gamma}{\sqrt{\beta}}\right]\left(\sqrt{\delta}\right)^\alpha},$$

and $I = \mathbb{I}(K_1(\delta, \alpha, \beta, \gamma) \leq K_2(\mu, \alpha, \beta, \gamma))$.

(b) The optimum choices for the constants μ and δ are

$$\mu_{opt} = \frac{\gamma + \sqrt{\gamma^2 + 8(\alpha-1)\beta}}{4\beta}, \quad and$$

$$\delta_{opt} = \beta + \frac{\gamma^2 - \gamma\sqrt{\gamma^2 + 8\alpha\beta}}{4\alpha}.$$

(c) The constants $K_1(\mu_{opt}, \alpha, \beta, \gamma), K_2(\delta_{opt}, \alpha, \beta, \gamma)$ depends only on α and $\Delta :=$ $\frac{\gamma}{\sqrt{\beta}}$, Therefore we will denote $K_1(\alpha, \Delta) := K_1(\mu_{opt}, \alpha, \beta, \gamma)$ and $K_2(\alpha, \Delta) :=$ $K_1(\delta_{opt}, \alpha, \beta, \gamma)$.

Part(a) of the lemma provides the explicit form of the proposal kernels where the variable I identifies the more efficient strategy between the Normal or the square

Algorithm 1 Sampling strategy for $\alpha > 1$, $\beta > 0$ and $\gamma > 0$.
Input: $\alpha > 1, \beta > 0, \gamma > 0$.
Output: X, a sample from MHN (α, β, γ).

1: **if** $K_2 > K_1$ **then**
2:      $\mu = \frac{\gamma + \sqrt{\gamma^2 + 8\beta(\alpha - 1)}}{4\beta}$.
3:      **goto** *loop*
4:      Generate $X \sim \text{Normal}(\mu, \frac{1}{2\beta})$
5:      Generate $U \sim \text{Uniform}(0, 1)$
6:      **if** $X > 0$ and $U < \frac{X^{\alpha - 1}}{\mu} \exp[(2\beta\mu - \gamma)(\mu - X)]$ **then**
7:          accept X
8:      **end if**
9:      **end** *loop* until accepting a sample.
10: **else if** $K_2 \leq K_1$ **then**
11:      $\delta = \beta + \frac{\gamma^2 - \gamma\sqrt{\gamma^2 + 8\alpha\beta}}{4\alpha}$
12:      **goto** *loop*
13:      Generate $T \sim \text{Gamma}(\text{shape} = \frac{\alpha}{2}, \text{rate} = \delta)$
14:      $X = \sqrt{T}$
15:      Generate $U \sim \text{Uniform}(0, 1)$
16:      **if** $U < \exp[-(\beta - \delta)X^2 + \gamma X - \frac{\gamma^2}{4(\beta - \delta)}]$ **then**
17:          accept X
18:      **end if**
19:      **end** *loop* until accepting a sample.
20: **end if**

root of Gamma proposals. Although the constants K_1, K_2 involve the Fox-Wright function in its expression, the evaluation of the variable I does not require its computation. One of the implications of Part(c) is that, the scale change of the random variable does not alter its efficiency. Therefore, without the loss of generality, we may consider $\beta = 1$ when assessing the performance of the algorithm. The details of the sampling method is included in Algorithm 1 while the following theorem apprises its performance.

Theorem 2. *The Acceptance probability of the Algorithm 1 to generate a random sample from $MHN(\alpha, \beta, \gamma)$, $\alpha \geq 1, \beta > 0, \gamma > 0$ is given as*

(a) $\mathcal{A}_{pos}(\alpha, \Delta) = \max\{\frac{1}{K_1(\alpha, \Delta)}, \frac{1}{K_2(\alpha, \Delta)}\}$ *where* $\Delta = \frac{\gamma}{\sqrt{\beta}}$.

(b) For a fixed $\alpha > 1$ the function $\Delta \mapsto K_1(\alpha, \Delta)$ is an decreasing function while

$\Delta \mapsto K_2(\alpha, \Delta)$ *is an increasing function.*

(c) $K_1(\alpha, \sqrt{\alpha}) \geq K_2(\alpha, \sqrt{\alpha})$ *for* $\alpha \geq 1$

(d) $\alpha \mapsto K_1(\alpha, \sqrt{\alpha})$ *is a decreasing function for* $\alpha \geq 4$.

(e) $\mathcal{A}_{pos}(\alpha, \frac{\gamma}{\sqrt{\beta}}) \geq \frac{1}{K_1(\alpha, \sqrt{\alpha})}$ *for all* $\alpha \geq 4$, $\gamma > 0, \beta > 0$ *and* $\mathcal{A}_{pos}(\alpha, \frac{\gamma}{\sqrt{\beta}}) \geq \frac{1}{K_1(4,2)} \geq$
0.8 *for all* $\alpha \geq 4, \beta > 0$ *and* $\gamma > 0$.

Note that, the part(e) of the Theorem 2 implies the uniform efficiency of the Algorithm 1. The proof of Theorem 2 is quite involved and necessities Lemma 8 as a key component. A major difficulty appears due to the unavailability of the analytic expression of the point Δ_{crit} such that $K_1(\alpha, \Delta_{\text{crit}}) = K_2(\alpha, \Delta_{\text{crit}})$ for a given α. We contrive a unique approach where the following attribute played a crucial role. For a fixed $\alpha \geq 1$, there is a unique $\Delta_{\text{crit}} \geq \sqrt{\alpha}$ such that $K_1(\alpha, \Delta_{\text{crit}}) = K_2(\alpha, \Delta_{\text{crit}})$ and $\frac{\Delta_{\text{crit}}}{\sqrt{\alpha}} \to 1$ as $\alpha \to \infty$. In consonance with Theorem 2 the Figure 2.3 also displays that specific attribute of Δ_{crit}. Note that, in Figure 2.3, we plot $A(\alpha, \frac{\gamma}{\sqrt{\beta}})$, the acceptance probability of the Algorithm 1 verses $C = \frac{\gamma}{\sqrt{\beta\alpha}}$. The curves with different colors correspond to different values of $\alpha \geq 1$. As C increases, the acceptance rate of the Normal proposal kernel increases, while it decreases for the other case. The Algorithm uses the square root of the Gamma proposal when the value of C is closer to zero. For its larger magnitudes, the Normal proposal kernel is selected by the indicator I. Therefore, in Figure 2.3, we see that each blue curve, representing the function $A(\alpha, \frac{\gamma}{\sqrt{\beta}})$, decreases reach to an optima and increases when the Normal kernel outperforms the square root of the Gamma kernel. Part(d) of the Lemma 11 ensures that the acceptance probability of Algorithm 1 is always greater than 0.8, which we can notice from Figure 2.3 as well.
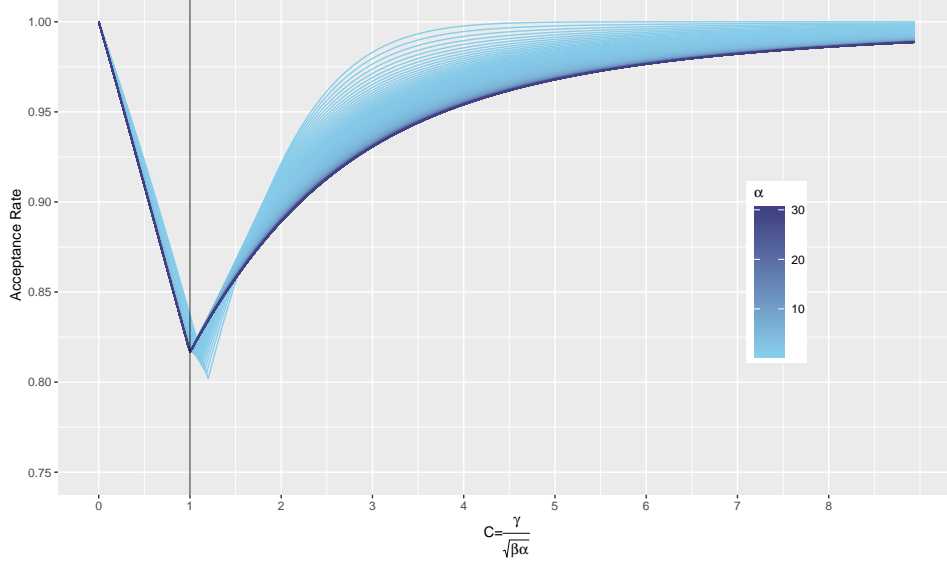
Figure 2.3: Acceptance rate of the Algorithm 1 for different values of α and $C = \frac{\gamma}{\sqrt{\beta\alpha}}$, where $\alpha \geq 1, \beta, \gamma > 0$.

A general sampling algorithm for MHN (α, β, γ) distribution when $\alpha > 0$ and $\gamma > 0$

We develop an alternative procedure for generating random samples from the MHN (α, β, γ) distribution when $\gamma > 0$ and $\alpha > 0$. This algorithm is applicable for the case when $0 \leq \alpha < 1$ and $\gamma > 0$. To discuss the different components of the algorithm, let us start with the following lemma.

Lemma 11. *Let $f_{MHN}(x \mid \alpha, \beta, \gamma)$ denotes the probability density function of the MHN(α, β, γ) distribution where $\alpha > 0, \beta > 0$. If $\gamma > 0$ then*

$$f_{MHN}(x \mid \alpha, \beta, \gamma) = \sum_{i=0}^{\infty} p_i f_i(x \mid \alpha, \beta)$$

where the positive constants $p_i = \dfrac{1}{2\beta^{\frac{\alpha}{2}}} \dfrac{\Gamma(\frac{\alpha}{2}+\frac{1}{2}i)(\frac{\gamma}{\sqrt{\beta}})^i}{i! \ _1\Psi_1\left[\begin{matrix}(\frac{\alpha}{2},\frac{1}{2})\\ (1,0)\end{matrix};\frac{\gamma}{\sqrt{\beta}}\right]}$ for $i \geq 0$ and

$$f_i(x \mid \alpha, \beta) = \frac{2\beta^{\frac{\alpha+i}{2}}}{\Gamma(\frac{\alpha+i}{2})}x^{\alpha+i-1}e^{-\beta x^2}\mathbb{1}(x > 0)$$

denotes the probability density function of the square root of Gamma distribution.

Let W be a discrete random variable, supported on the non-negative integers with probability mass function $P(W = i) = p_i$ for $i = 0, 1, \ldots, \infty$ where p_i's are given in Lemma 11. Then a straightforward approach to design a sampling scheme for MHN (α, β, γ) using Lemma 11 is to first sample W, thereafter sample from the square root of a Gamma distribution with shape $\frac{\alpha+W}{2}$ and rate β. Note that the numerical evaluations of p_i's require the computation of the Fox-Wright functions that we already discussed in Section 2.2. However, we design the following procedure that can circumvent the need for computing the Fox-Wright function as it can be time consuming.

Lemma 12. *Let* $q_i = \dfrac{\Gamma(\frac{\alpha+i}{2})\left(\frac{\gamma}{\sqrt{\beta}}\right)^i}{i!}$ *for* $i \geq 0$ *where* $\alpha > 0$, $\gamma > 0, \beta > 0$. *For an be arbitrary positive number* $\epsilon_1 > 0$, *suppose* $M^\dagger = \max\left\{[\alpha], \left[\frac{\gamma^2}{\epsilon_1^2\beta}\right]\right\}$ *where* $[\alpha]$ *and* $\left[\frac{\gamma^2}{\epsilon_1^2\beta}\right]$ *denotes the largest integer less than equal to* α *and* $\frac{\gamma^2}{\epsilon_1^2\beta}$ *respectively. Then*

(a) $\dfrac{q_{M+1}}{q_M} \leq \epsilon_1$ *when* $M \geq M^\dagger$,

(b) $\dfrac{q_{i+2}q_i}{q_{i+1}^2} < 1$ *for* $i \geq 0$,

(c) $\displaystyle\sum_{j=M+1}^{\infty} q_j \leq q_M \dfrac{\epsilon_1}{1-\epsilon_1}$ *when* $M \geq M^\dagger$ *and*

(d) $\displaystyle\lim_{i\to\infty} q_i = 0$.

We now construct a discrete probability distribution with support $\{0, 1, \ldots, \infty\}$ and the corresponding unnormalized probabilities $\{\bar{q}_i\}_{i\geq 0}$ are defined as

30

$$
\overline{q}_i =
\begin{cases}
q_i, & \text{if } i \leq M \\[2ex]
q_M \epsilon_1^{i-M}, & \text{if } i > M.
\end{cases}
$$

Note that $\sum_{i=1}^{\infty} \overline{q}_i = q_M \frac{\epsilon_1}{1-\epsilon_1} + \sum_{j=1}^{M} q_j$. The part(a) and the part(b) of the Lemma 12 imply that

$$
q_i \leq q_M \left\{ \prod_{j=M}^{i-1} \frac{q_{j+1}}{q_j} \right\} \leq q_M \epsilon_1^{i-M} = \overline{q}_i \text{ for } i > M. \tag{2.10}
$$

Therefore $q_i \leq \overline{q}_i$ for all $i \geq 0$. In order to generate sample from the a discrete distribution with probabilities $\{p_i\}_{i \geq 0}$, a rejection sampling algorithm can be developed by constructing the discrete proposal distribution supported on the nonnegative integers with the probabilities $\{\overline{p}_i\}_{i \geq 0}$ where

$$
\overline{p}_i =
\begin{cases}
\dfrac{q_i}{q_M \frac{\epsilon_1}{1-\epsilon_1} + \sum_{j=1}^{M} q_j}, & \text{if } i \leq M \\[3ex]
\dfrac{q_M \epsilon_1^{i-M}}{q_M \frac{\epsilon_1}{1-\epsilon_1} + \sum_{j=1}^{M} q_j}, & \text{if } i > M.
\end{cases}
$$

Accept a random sample from the proposal distribution as a valid sample with the probability $\frac{q_i}{\overline{q}_i}$. Note that the sampling algorithm can be made efficient with appropriate choice for M. For example, given any $\epsilon > 0$ we can choose M such that $q_M \leq \epsilon$ and $\frac{q_{M+1}}{q_M} < \epsilon_1 := \frac{1}{2}$. Part(a) and Part (d) of Lemma 12 ensures the existance of such M. In that case probability of rejection would be

$$
\sum_{j=M+1}^{\infty} q_M \epsilon_1^{j-M} = \frac{\epsilon_1 q_M}{1-\epsilon_1} \leq \epsilon \frac{\frac{1}{2}}{1 - \frac{1}{2}} = \epsilon.
$$

Based on the above discussion, we now assume that we can sample from the discrete probability distribution having the support $\{0, 1, 2, \ldots, \infty\}$ with corresponding probabilities $\{p_i\}_{i \geq 0}$ that are defined in Lemma 11. The following algorithm provides

the details of the steps required to sampling from the MHN (α, β, γ), $\alpha > 0, \gamma > 0$ distribution.

Algorithm 2 Sampling Strategy for $\gamma > 0$.
Input: $\alpha > 0, \beta > 0, \gamma > 0$.
Output: Y, a sample from MHN (α, β, γ).

1: Generate $W \sim \text{Discrete}(\{0, 1, 2, \ldots, \}, \{p_i\}_{i \geq 0})$
2: Generate $Y^\star \mid W \sim \text{Gamma}\left(\text{shape} = \frac{\alpha + W}{2}, \text{rate} = \beta\right)$
3: Set $Y = +\sqrt{Y^\star}$

If either of α or $\frac{\gamma^2}{\beta}$ is large then the computational time of the algorithm is significant. On the other hand, for small or moderate values of $\max\{\alpha, \frac{\gamma^2}{\beta}\}$, the algorithm is efficient. Unfortunately, the techniques of the above algorithm is inapplicable when $\gamma < 0$ because the odd order terms in the sequence $\{p_i\}_{i \geq 0}$ (defined in Lemma 11) would then be negative invalidating the discrete mixture representation provided in Lemma 11. An entirely different approach is considered to tackle the case $\gamma < 0$ that we discuss next.

2.4.2 Sampling from the MHN (α, β, γ) when $\gamma \leq 0$.

If $\gamma \leq 0$ then the MHN (α, β, γ) density is proportional to $x^{\alpha-1} \exp(-\beta x^2 - |\gamma| x) \mathbb{I}(x > 0)$. Using the generalized version of the AM-GM inequality (Steele, 2004) in the exponent part of the target density, we can construct a proposal kernel

$$x^{\alpha-1} \exp(-\beta x^2 - |\gamma| x) \leq x^{\alpha-1} \exp(-(\beta + |\gamma|) x^{\frac{2\beta + |\gamma|}{\beta + |\gamma|}}) \tag{2.11}$$

(see Theorem 3 with the choice $m = 1$), which can be normalized to make a proper proposal density. The functional value of the above proposal kernel matches with that of the target density (i.e. equality in the Equation 2.11) at the point $x = 1$ irrespective of the values of the parameters α, β, γ. It follows from the characteristics of the AM-GM inequality (Steele, 2004) that the proposal kernel approximates the target density better around the point $x = 1$ compared to that for the other points.
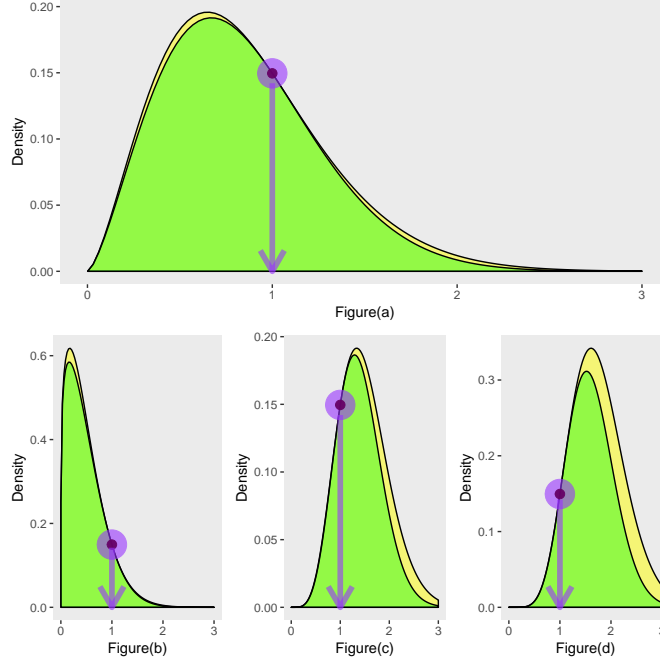
Figure 2.4: The target density and the proposal kernel (upto the proportionality constant, see Equation (2.11)) is plotted for four different cases of (α, β, γ). The marked point is the 'matching point', which transpires to be located at $x = 1$ irrespective of the values of the parameters $\alpha > 0, \beta > 0, \gamma < 0$.

In a quest for developing a more general class of proposal kernels, we introduce the notion of 'matching point', the point where the functional value of the proposal kernel becomes equal to that of the target density. For example, in Figure 2.4, the matching point for the target and the proposal kernel is located at the point $x = 1$ irrespective of the choices of the parameters α, β, γ. Furthermore, the Figure 2.4 depicts the functional inequality specified in Equation 2.11 while it indicates that the inequality may not be optimal when the matching point is always at $x = 1$.

Therefore we design a class of proposal kernels where the flexibility is introduced by allowing the matching point to vary while the crucial functional inequality between the target density and the proposal kernel remains unaffected. The following Theorem provides a way to construct such a class of proposal kernels.

Theorem 3. *Let $f_{MHN}(x \mid \alpha, \beta, \gamma)$ be the probability density function of the $MHN(\alpha, \beta, \gamma)$*

distribution where $\gamma \leq 0$. Let $m > 0$ be any positive number then,

$$f_{MHN}(x \mid \alpha, \beta, \gamma) \leq \quad K_0(m, \alpha, \beta, \gamma) x^{\alpha-1} \exp\left[-m(\beta m + |\gamma|)\left(\tfrac{x}{m}\right)^{\frac{2\beta m + |\gamma|}{\beta m + |\gamma|}}\right]$$

where $K_0(m, \alpha, \beta, \gamma) = \dfrac{2\beta^{\frac{\alpha}{2}}}{{}_1\Psi_1\left[\begin{array}{c}\left(\frac{\alpha}{2}, \frac{1}{2}\right) \\ (1, 0)\end{array}; \frac{-|\gamma|}{\sqrt{\beta}}\right]}.$

The proof of the Theorem 3 is reinforced on the generalized version of the AM-GM inequality (Steele, 2004). The parameter m which apparently appears as an abstract algebraic trick in the proof, actually have a tangible interpretation signifying the matching point that we have discussed earlier. The proposal kernel in the Theorem 3 can be normalized to a probability density that corresponds to the $(\frac{\beta m + |\gamma|}{2\beta m + |\gamma|})^{\text{th}}$ power of a Gamma distribution. In particular, if we consider

$$T \sim \text{Gamma}\left(\frac{\alpha(\beta m + |\gamma|)}{2\beta m + |\gamma|}, m(\beta m + |\gamma|)\right)$$

and set $Y = T^{\frac{\beta m + |\gamma|}{2\beta m + |\gamma|}}$ then the probability density function of the random variable Y

$$f_Y(y) \propto y^{\alpha-1} \exp\left(-m(\beta m + |\gamma|)y^{\frac{2\beta m + |\gamma|}{\beta m + |\gamma|}}\right),$$

which is identical (ignoring the constants) to the proposal kernel in Theorem 3. As a consequence, a random sample from the corresponding proposal density turns out to be an appropriately transformed random number generated from a suitable Gamma random variable. The specific details of the corresponding Accept-Reject sampling algorithm is provided in Algorithm 3. We would like to point out that the standard implementation of the 'rgamma' function in 'R' for generating Gamma random variable may not execute with a desired accuracy when the shape parameter is small (Liu et al., 2017). We use a procedure discussed in Liu et al. (2017) to tackle that specific

challenge.

Algorithm 3 Sampling Strategy for $\gamma < 0$.

Input: $\alpha > 0, \beta > 0, \gamma < 0, m > 0$.

Output: X, a sample from MHN (α, β, γ).

1: **goto** *loop*

2: Generate $T \sim \text{Gamma} \left(\frac{\beta m + |\gamma| \alpha}{2 \beta m + |\gamma|}, \text{rate} = m(\beta m + |\gamma|) \right)$

3: $X = m T^{\frac{\beta m + |\gamma|}{2 \beta m + |\gamma|}}$

4: Generate $U \sim \text{Uniform}(0, 1)$

5: **if** $U < \exp \left(m(\beta m + |\gamma|) \left(\frac{X}{m} \right)^{\frac{2\beta m + |\gamma|}{\beta m + |\gamma|}} - \beta X^2 - |\gamma| X \right)$ **then**

6:     accept X

7: **end if**

8: **end** *loop* until an acceptance.

The Algorithm 3 is valid for all positive values of the parameter m while an apposite choice for m can result in significant efficiency gain. In order to illustrate the role of the parameter m in Theorem 3, we refer to the Figure 2.5 where we considered MHN $(5, 1, -1)$ density as an example. We took four different choices of matching points $0.8, 1.4, 2.15$ and the mode of the distribution and plot the corresponding kernels that are obtained using Theorem 3.

It is seen that the difference between the areas under the MHN $(5, 1, -1)$ density and the proposal kernels alters as we change the matching point m. Specifically, the difference of the areas gradually decreases, reaches an optima and further increases when the magnitude of m gradually increases from zero. Thus, by selecting the optimal value for m, we can gain in efficiency. From an intuitive standpoint, matching the proposal kernel at the mode of the target density seem to be a prudent strategy. But in actuality, the optimal value for m transpires to be larger than mode, seemingly due to the fact that the target density, as well as the proposal kernel is rightly skewed. We reach to a congruous conclusion from the Figure 2.5 as well. Consolidating the heuristics that we have discussed so far in this paragraph, we contrive the following theorem where, along other results, we establish the uniqueness of the
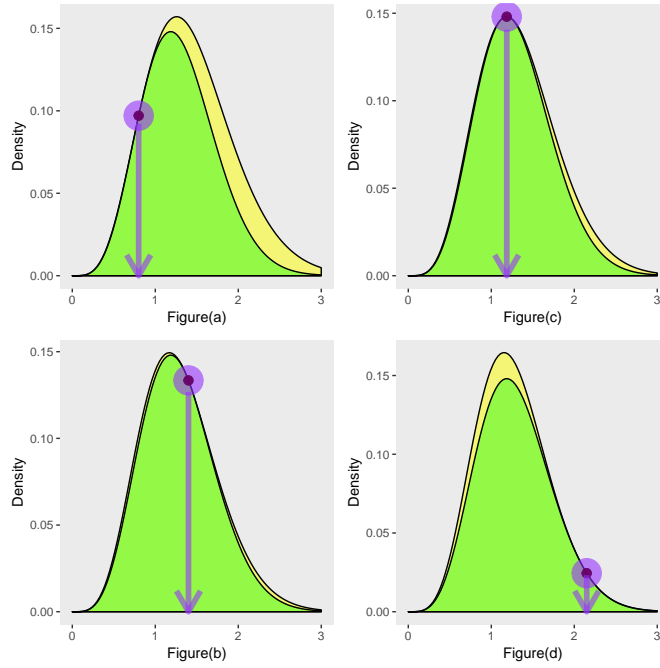
Figure 2.5: The MHN $(5, 1, -1)$ density and the proposal kernel in Theorem 3 is plotted for four different choices of the 'matching point' (m). The difference between the areas under the target density and the proposal kernel gradually decreases, reaches an optima and further increases when the value of m gradually increases from zero. The optimal value for m appears to be larger than the mode of the target density.

optimal matching point.

Theorem 4. *Let $\mathcal{A}_{neg}(m, \alpha, \beta, \gamma)$ denotes the acceptance probability of the Algorithm 3 when $\alpha > 1, \beta > 0$, $\gamma \leq 0$. If m is any positive constant*

(a)

$$\mathcal{A}_{neg}(m, \alpha, \beta, \gamma) = \frac{(2\beta m + |\gamma|)(\beta m + |\gamma|)^{\frac{\alpha(\beta m + |\gamma|)}{2\beta m + |\gamma|} - 1} {}_1\Psi_1 \left[\begin{array}{c} (\frac{\alpha}{2}, \frac{1}{2}) \\ (1, 0) \end{array}; \frac{-|\gamma|}{\sqrt{\beta}} \right]}{2\beta^{\frac{\alpha}{2}} \, m^{\frac{\alpha\beta m}{2\beta m + |\gamma|}} \, \Gamma(\frac{\alpha(\beta m + |\gamma|)}{2\beta m + |\gamma|})}.$$

(b) For any $\alpha > 1, \beta > 0$ and $\gamma < 0$, the function $m \mapsto \mathcal{A}_{neg}(m, \alpha, \beta, \gamma)$ has a unique maxima at a point m_{opt} where $m_{opt} > X_{mode}$, the mode of the distribution.

(c) For all $\alpha > 1, \beta > 0$ and $\gamma \leq 0$,

$$\mathcal{A}_{neg}(m_{opt}, \alpha, \beta, \gamma) \geq \mathcal{A}_{neg}(m_\star, \alpha, \beta, \gamma) \geq \frac{1}{\sqrt{2}}$$

where $m_\star = \frac{\gamma + \sqrt{\gamma^2 + 8\alpha\beta}}{4\beta}$.

The part(c) of the theorem implies uniform efficiency of the Algorithm 3. The proof of part(c) is highly nontrivial and requires several properties of Gamma and Digamma functions including the Ramanujan's double inequality for the Gamma function (Alzer, 1997, 2003; Batir, 2005, 2008). The acceptance probability $\mathcal{A}_{\text{neg}}(m, \alpha, \beta, \gamma)$ becomes unity when the parameter γ is set to zero irrespective of the choice for $m > 0$. The part(b) of the theorem ensures the existence of the optimal matching point whereas the optimization of $m \mapsto \mathcal{A}_{\text{neg}}(m, \alpha, \beta, \gamma)$ for finding m_{opt} is nontrivial. As it is seen in the part(a) of Theorem 4, the parameter m does not involve the Fox-Wright function in the expression of $\mathcal{A}_{\text{neg}}(m_{\text{opt}}, \alpha, \beta, \gamma)$. Therefore, the optimization does not involve Fox-Wright function. In the Appendix 5.0.17, we provide an

iterative procedure to find m_{opt}. We anticipate that the strategy to use an iterative procedure for finding m_{opt} and thereafter use it for sampling from the corresponding Modified-Half-Normal may be time consuming. Therefore, we utilize the properties of the Modified-Half-Normal density and the intuitive understanding about the location of m_{opt} to design m_{init} in a non iterative way to approximate m_{opt}. In the following paragraph we discuss the formulation to compute m_{init}.

Let X_{mode} and X_{inflex} denote the mode and the rightmost inflection point of the Modified-Half-Normal distribution. We refer to the Lemma 7 for the exact form of the X_{mode} while X_{inflex} can be obtained by finding the largest real root of a quartic polynomial (Jenkins and Traub, 1972; Venables et al., 2019). Thereafter, we define

$$
m_{\text{init}} = \begin{cases} \frac{\alpha^2}{1+\alpha} & \text{if } \alpha \leq 1.1, \\[2ex] \frac{3\lambda}{2} X_{\text{mode}} + (1 - \frac{3\lambda}{2}) X_{\text{inflex}} & \text{if } \alpha > 1.1, \end{cases}
$$

$$
\text{where } \lambda = \frac{f_{\text{MHN}}(2X_{\text{mode}} - X_{\text{inflex}})}{f_{\text{MHN}}(2X_{\text{mode}} - X_{\text{inflex}}) + f_{\text{MHN}}(X_{\text{inflex}})}.
$$

The candidate m_{init} is designed to be in between the mode and the rightmost inflection point of the distribution. Depending on the skewness of the distribution, the fraction λ carefully manages its proximity to the point X_{mode}. For the numerical evaluation of λ, there is no need to compute the Fox-Wright functions as its expression involves $f_{\text{MHN}}(\cdot)$ in the numerator and the denominator as well. Although the development of m_{init} is based on heuristics, it serves the need of the Algorithm 3.

In the Figure 2.6a, we plot the acceptance probabilities for different values of α, γ when β is set to 1. The values of γ are used in the X axis. The different colors represent different choices of α. It appears from the figure as wells as from Table 2.1, that the value of $\mathcal{A}_{\text{neg}}(m_{\text{init}}, \alpha, \beta, \gamma)$ is larger than 0.9. In Figure 2.6c we plot the acceptance probabilities when m_{opt} is used to implement the Accept Reject algorithm. Comparing the Figure 2.6a and 2.6c, it appears that there is an

increment in the acceptance probabilities whereas the improvement after employing the iterative procedure is not quite significant. From the 6th and the 8th columns of the Table 2.1, we see that improvement is less than 0.2% when comparing the acceptance probabilities corresponding to m_{init} and m_{opt}. In the fifth of column of the Table 2.1, we listed the number of steps required to get m_{opt} starting the iterative algorithm from m_{init}. We can see that the number of steps required for convergence is small if the iterative algorithm is started from m_{init}. Therefore, m_{init} appears to be in the vicinity of m_{opt}. We recommend to execute a single step of the iterative algorithm after starting it from the m_{init} and get m_{reco} to be used as the suggested value for m. Although, the current context is completely different, the strategy is motivated from of the one-step estimators (Jurečková and Sen, 1990). It is seen from the Table 2.1 (compare 7th and 8th columns) that the acceptance probability of the corresponding Accept-Reject sampling algorithms are not much different even if the computationally inexpensive recommended value, m_{reco} is used instead of the m_{opt}. Also, there is barely any visible differences between Figures 2.6b and 2.6c, which plots $\mathcal{A}_{\text{neg}}(m_{\text{init}}, \alpha, \beta, \gamma)$ and $\mathcal{A}_{\text{neg}}(m_{\text{reco}}, \alpha, \beta, \gamma)$.

The remaining of the section we discuss the efficiency of the Algorithm 3 when the optimal value of the parameter m is used. In the Figure 2.6 we plot the Acceptance Probability where in the X axis we plot the quantity $C = \frac{\gamma}{\sqrt{\alpha}}$. The different colored curves pertains to the different specification of the parameter $\alpha \geq 1$. The efficiency of the algorithm increases with the increment of the magnitude of α. It is clear from the plot that the worst possible acceptance probability is larger than 0.90.
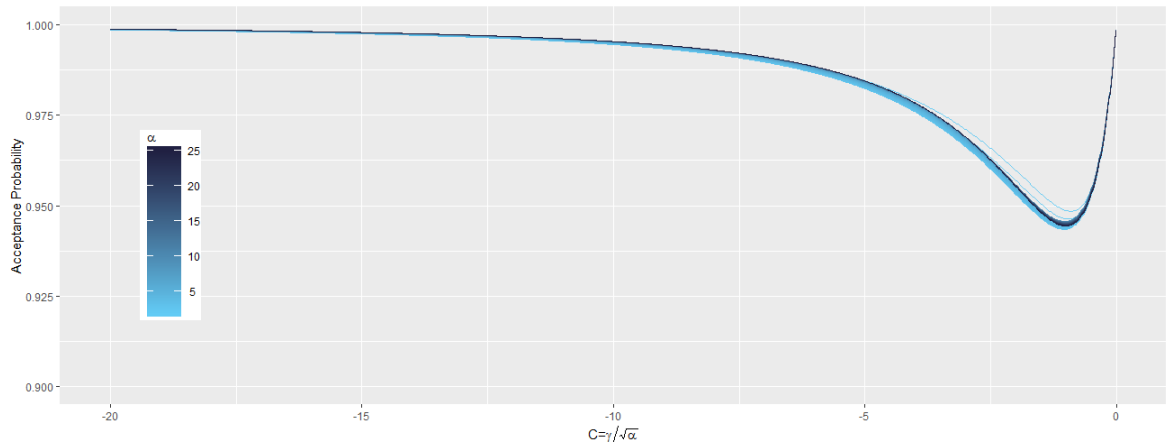
(a) Acceptance probability when using $m := m_{\text{init}}$.
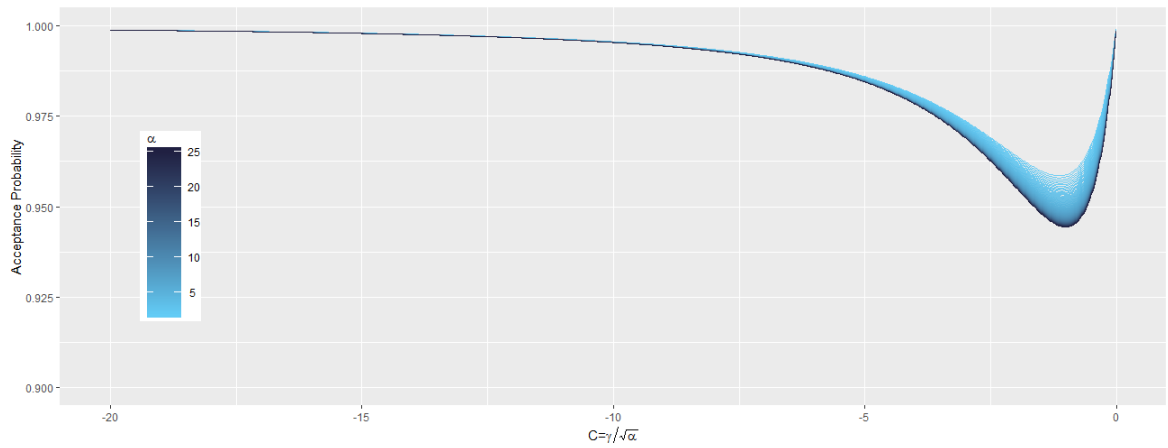
(b) Acceptance probability when using $m := m_{\text{reco}}$.
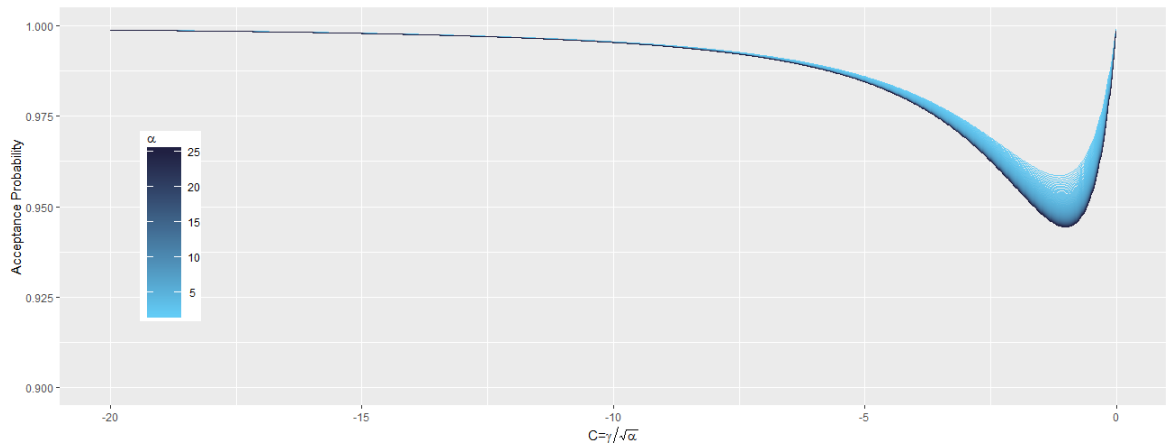
(c) Acceptance probability when using $m := m_{\text{opt}}$.

Figure 2.6: Acceptance rate of the Algorithm 3 for different value of α and $C = \frac{\gamma}{\sqrt{\beta}}$ when $\alpha > 0, \beta > 0, \gamma \le 0$. The figures (a), (b), (c) pertains to different choices for the parameter m.

α	β	γ	C	Number of Steps	$\mathcal{A}_{\text{neg}}$ at m_{init}	$\mathcal{A}_{\text{neg}}$ at m_{reco}	$\mathcal{A}_{\text{neg}}$ at m_{opt}
2	1	-20	-14.14	2	0.997	0.998	0.998
2	1	-16	-11.31	2	0.996	0.996	0.996
2	1	-12	-8.49	3	0.993	0.994	0.994
2	1	-8	-5.66	3	0.987	0.988	0.988
2	1	-4	-2.83	3	0.969	0.972	0.972
2	1	-1	0.71	5	0.95	0.96	0.962
2	1	-0.1	0.07	4	0.989	0.992	0.993
5	1	-20	-8.94	3	0.994	0.994	0.994
5	1	-16	-7.16	3	0.991	0.992	0.992
5	1	-12	-5.37	3	0.985	0.986	0.987
5	1	-8	-3.58	4	0.974	0.976	0.976
5	1	-4	-1.79	6	0.952	0.955	0.956
5	1	-1	0.45	8	0.957	0.96	0.961
5	1	-0.1	0.05	6	0.993	0.994	0.994
15	1	-20	-5.16	3	0.985	0.985	0.985
15	1	-16	-4.13	3	0.979	0.979	0.979
15	1	-12	-3.10	4	0.97	0.97	0.97
15	1	-8	-2.07	5	0.957	0.957	0.957
15	1	-4	-1.03	7	0.945	0.945	0.945
15	1	-1	0.26	7	0.969	0.969	0.969
15	1	-0.1	0.01	1	0.996	0.996	0.996

Table 2.1: Acceptance probabilities of the Algorithm 3 for different values of α, β, γ. The 6,7 and 8th column shows the accepence probability when the m_{init}, m_{reco} and m_{opt} is utilized as a specification of the matching point. The 5th column provides the number of iterative steps required to obtain m_{opt} when starting the algorithm from m_{init}.

2.5 Statistical inference for analyzing data using Modified-Half-Normal distribution as a probability model

Although it is not the primary objective of this article, we consider estimation for the parameters of MHN (α, β, γ) distribution. We see from Lemma 10 that the Truncated Normal, square root of Gamma and Half normal and Gamma distributions are specific cases of the MHN (α, β, γ) distribution. Therefore, the Modified-Half-Normal family of distributions is a reasonably flexible class of distributions which can

model continuous data on the positive part of the real line. As a part of the very basic inference procedure, we discuss the method of moments and the maximum likelihood estimation procedures, which has certain non-triviality in the current context.

2.5.1 Parameter estimation

Let $x_1, \ldots x_n$ be independent and identically distributed observations from MHN (α, β, γ) distribution. For denoting the kth sample moment of the data, we introduce the notation

$$\mathbb{m}_k = \frac{1}{n} \sum_{i=1}^{n} x_i^k, \text{ for } k > 0.$$

According to the standard practice, the method of moment estimators for the parameters are obtained by equating the sample moments with the corresponding population moments that are expressed as a function of the unknown parameters (Casella and Berger, 2002). In the current context, the implementation of the standard method of moment estimation would reduce to solving for α, β, γ using the equations

$$\mathbb{m}_k = \frac{{}_1\Psi_1 \left[\begin{matrix} (\frac{\alpha+k}{2}, \frac{1}{2}) \\ (1, 0) \end{matrix} ; \frac{\gamma}{\sqrt{\beta}} \right]}{\beta^{\frac{k}{2}} \, {}_1\Psi_1 \left[\begin{matrix} (\frac{\alpha}{2}, \frac{1}{2}) \\ (1, 0) \end{matrix} ; \frac{\gamma}{\sqrt{\beta}} \right]}$$

for $k = 1, 2, 3$ (see part(a) of the Lemma 6). Finding solutions to the above set of equations are complicated as it involves the Fox-Wright functions. In order to mitigate the difficulties, we utilize the recursion relation in the part(b) of the Lemma 6 instead of using the formula in part(a). In particular, we solve α, β, γ from the following set of equations

$$\mathbb{m}_{k+2} = \frac{\alpha+k}{2\beta} \mathbb{m}_k + \frac{\gamma}{2\beta} \mathbb{m}_{k+1} \text{ for } k = 0, 1, 2,$$

where $\mathtt{m}_0 = 1$. Note that the above set of equations refers to a set of linear equations that can be solved easily. The estimators obtained by the above procedure are statistically consistent and also asymptotically Normal.

Now we discuss the maximum likelihood estimation for the current context. The log likelihood function (see Equation 2.1) for α, β, γ based on the independent samples $x_1, \ldots x_n$ is given by

$$l(\alpha, \beta, \gamma) = \quad n \, \log(2) + \frac{n\alpha}{2} \log(\beta) - n \, \log\left({}_1\Psi_1 \left[\begin{matrix} (\frac{\alpha}{2}, \frac{1}{2}), \\ (1, 0) \end{matrix} ; \frac{\gamma}{\sqrt{\beta}} \right] \right)$$
$$+ (\alpha - 1) \sum_{j=1}^{n} \log(x_j) - \beta \sum_{j=1}^{n} x_j{}^2 + \gamma \sum_{j=1}^{n} x_j.$$

To utilize the Newton-Raphson procedure we require computation of the first and second-order derivatives of the Fox-Wright function. The computation is nontrivial especially to numerically evaluate the derivatives involving the parameter α. An implementation of the Newton-Raphson procedure, in this case, appears not to be very stable. Therefore, as a workable solution, we utilize the 'Nonsmooth Optimization by Mesh Adaptive Direct Search'(NOMAD) procedure (Audet and Dennis Jr, 2006; Abramson et al., 2011; Le Digabel, 2010; Audet and Hare, 2017) to minimize the negative of the log-likelihood function. Specifically, we use the 'R' interface via the function 'snomadr' implemented in the 'R' package 'crs' (Nie and Racine, 2012). The optimization procedure requires a set of initial values. We provide the estimates that we get from the method of moment procedure. The performance of the estimation procedures is assessed via simulations that we discuss next.

2.5.2 Simulation results

The objective of this section is to assess the performance of the proposed estimation procedures via simulation. We fix a set of true parameter $\alpha_{\text{true}}, \beta_{\text{true}},$

and γ_{true} to generate 1000 different data sets, each containing i.i.d samples from MHN ($\alpha_{\text{true}}, \beta_{\text{true}}, \gamma_{\text{true}}$). We perform the method of moment as well as the maximum likelihood procedure to get the corresponding estimated values. We calculate the average squared error difference between the estimated values and the true value of each of the parameters. We repeat the procedure thrice keeping everything same except for the sample sizes. In the three different sets up the sample sizes for each of the data set were chosen to be $n = 5000, 20000$ and 40000 correspondingly.

The average squared difference between the estimated and the true value of the parameters when using the method of moment estimation. The estimated values of the MSE are given as 0.471 0.1173 and 0.0525 respectively for the case $n = 5000, 20000$ and $40,000$ respectively. On the other hand for the case of the maximum likelihood estimation, the corresponding MSE estimates appear to be 0.168, 0.0403,0.0192 respectively. Additionally, the performance of the maximum likelihood procedure appears to be better than the method of moment procedures.

CHAPTER 3

DIRECTED ACYCLIC GRAPH ASSISTED METHODS FOR ESTIMATING AVERAGE TREATMENT EFFECT

3.1 Introduction

Randomized controlled trials (RCTs) are often considered as the gold standard to estimate treatment effect. However, RCTs may not be always feasible because (i) RCTs can be time-consuming and expensive, and (ii) RCTs can be unethical under some circumstances. On the other hand, observational studies abound. Observational studies can provide important information for treatment effect and personalized medication, provided that confounding variables are properly controlled.

In the literature, propensity score based methods have become very popular in estimating average treatment effect (ATE) since the seminal work was proposed by Rosenbaum and Rubin (1983). ATE is defined as the mean difference between the potential outcome under treatment and the potential outcome under control. That is, ATE is defined as $E(Y^{(1)}) - E(Y^{(0)})$. $Y^{(1)}$ denotes the potential outcome when a subject is in the treatment group (say $T = 1$), and $Y^{(0)}$ denotes the potential outcome when a subject is in the control group (say $T=0$). However, one can only observes one potential outcome, which depends on the treatment that a subject receives (say, T). In other words, the observed outcome $Y = Y^{(1)}$ if $T = 1$, and $Y = Y^{(0)}$ if $T = 0$, which is often referred as the consistency assumption (Abdia et al. (2017)). The additional key assumptions for estimating ATE include the exchangeability assumption (i.e., $T \perp\!\!\!\perp Y^{(t)}|X$, where X is a set of measured covariates) and positivity assumption. The

propensity score is defined as the probability of a subject receiving treatment given the covariates X. Under the positivity, consistency, and exchangeability assumptions, many propensity score based methods, such as matching, regression with propensity score as covariate, stratification, inverse probability of treatment weighting (IPTW), and doubly robust estimates have been used to estimate ATE (Rosenbaum and Rubin (1983), Lunceford and Davidian (2004), Yan et al. (2019)).

In observational studies, other than the treatment assignment variable (say, T) and outcome variable (say, Y), there are many other variables collected, which are illustrated in Figure 3.1a and referred as different types of covariates X: (i) confounding variables (say X_C), which are related to treatment assignment variable as well as outcome variable; (ii) instrument variables (say, X_I), which are related to treatment assignment variable but not outcome variable; (iii) predictor variables (say, X_P), which are related to outcome variable but not treatment assignment variable; (iv) mediator (say, X_M), which are in the downstream of treatment assignment variable but a cause of the outcome variable; (v) collider (say, X_N), which are in the downstream of treatment assignment variable and the outcome variables (Craycroft et al. (2020), Pearl (2000)); and (vi) spurious variables (say, X_S), which are relevant neither treatment assignment variable nor outcome variable. One question is what types of covariates X need to be included in the propensity model so that the ATE estimates are unbiased with high accuracy. Brookhart et al. (2006), Austin (2007), Franklin et al. (2015), and Craycroft et al. (2020) have pointed out that including variables related to confounding variables and predictor variables can improve the precision of ATE estimates. On the other hand, Pearl (2000) proposed using directed acyclic graph (DAG) to identify the variables, which block all backdoor paths from T to Y, to be included in the propensity score model. DAG can be used to represent the causal relationships between different variables in a complex system (Lauritzen (1996), Pearl (2000)). DAGs are often obtained from close collaboration between data analysts and

subject-matter experts. Once DAGs are properly constructed, Pearl (2000) proposed to adjust (in the propensity score model) a set of variables which block all backdoor paths from treatment assignment variable T to outcome variable Y. There could be multiple sets of such variables, and each set of variables block all backdoor paths from T to Y. To obtain a valid ATE estimate, one can use any set of such variables in the propensity score model. That is, one does not need to include all covariates in the propensity score model. Instead, one only needs to include a set of variables, which block all backdoor paths from T to Y, to estimate the propensity score. Recent literature has indicated that including both confounding variables and predictor variables in the propensity score model can improve the efficiency of the propensity score based ATE estimates (Craycroft et al. (2020)). In this project, we investigated the performance of the propensity score based IPTW method when a set of confounding variables (which block all backdoor paths from T to Y) and predictor variables are included in the propensity score model.

Marginal structural models (MSMs) are a class of causal models which can be used to estimate the causal effect (Robins et al. (2000)). In a MSM, only the marginal mean and causal parameters are presented. Applying the inverse probability of treatment weighting (IPTW) to MSM has been proposed to estimate the causal parameters (Robins et al. (2000)). In Section 3.2, we propose to include the confounding variables which block all backdoor paths from T to Y and the predictor variables in the propensity score model. In order to reduce the selection bias in the observational data, we introduce the IPTW method, which is based on the propensity score model. Then we apply the IPTW method to estimate the causal parameters in the MSM to get the unbiased ATE estimate. This process is referred as DAG-assisted method in this article. In Section 3.3, we apply this proposed method to study the effect of tracheostomy on children with bronchopulmonary dysplasia (BPD) based on the 2016 Healthcare Cost and Utilization Project (HCUP) Kids' Inpatient Database

47

(KID). In Section 3.4, we evaluate the performance of the proposed method via extensive simulations. The last section is devoted to conclusion and discussion.
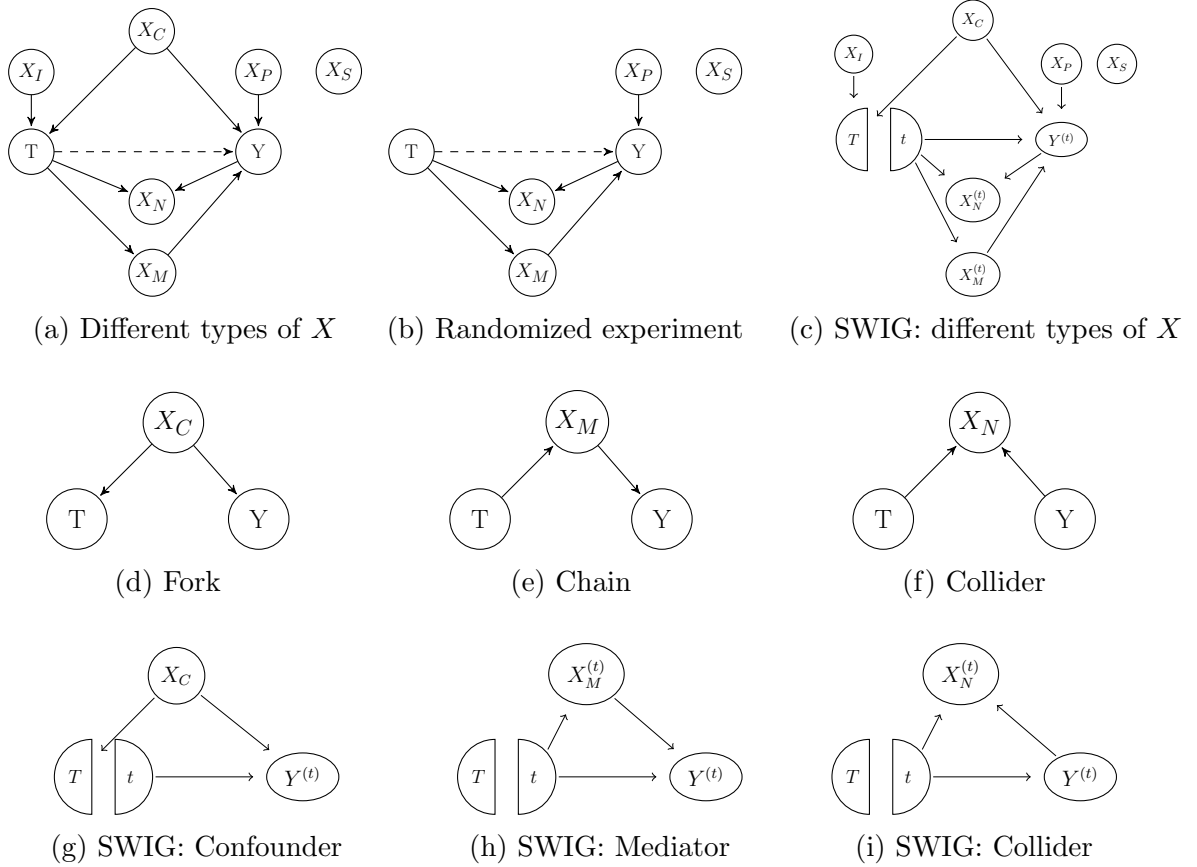
Figure 3.1: Basic relationships between different variables illustrated in DAGs and SWIGs

3.2 DAG-assisted method to estimate ATE

Let us assume that a triplet $\{X, T, Y\}$ is observed for each subject. T is the observed treatment assignment group. $T = 1$ if a subject is assigned to treatment group, and $T = 0$ if the subject is assigned to control group. X is the observed covariates which can be classified as confounding variables (say, X_C), mediators (say, X_M), colliders (X_N), instrumental variables (say, X_I), predictor variables (say X_P), and spurious variables (X_S) (Figure 3.1a). Here we assume the data is from observational studies and the treatment for a patient is not randomly assigned, instead the

treatment received by a patient may depend on his/her own characteristics. We limit our investigation on propensity score based approach. Including different types of covariates X in the propensity score model may result in different estimates of ATE. In this section, we first present the basic assumptions of obtaining unbiased ATE estimates, and then we present the link between the DAG and the basic assumption on exchangeability. Finally, we present our approach to estimate ATE.

3.2.1 Basic assumptions for causal inference

In this article, we focus on estimating ATE which measures the mean difference of the potential outcomes between all treated subjects and all untreated subject. That is,

$$ATE = E(Y^{(1)}) - E(Y^{(0)}),$$

where $Y^{(1)}$ is the potential outcome when a subject is treated, and $Y^{(0)}$ is the potential outcome when the subject is untreated. However, in observational studies, we only observe the outcome (say, Y) which corresponds to the treatment assigned. In general, we assume $Y = Y^{(0)}\mathbb{1}_{\{T=0\}} + Y^{(1)}\mathbb{1}_{\{T=1\}}$. That is, $Y = Y^{(1)}$ when a subject is treated, otherwise $Y = Y^{(0)}$. In RCTs, the subjects are randomly assigned to treatment or control group. So, there is no confounding variable in the RCT, that is, there is no arrow entering the treatment assignment variable (see Figure 3.1b). The subjects in the treatment group can be considered as a random sample from the study population, and the sample mean of the outcomes from treated subjects is an unbiased estimator of $E(Y^{(1)})$. Similarly, $E(Y^{(0)})$ can be estimated by the sample mean from the subjects assigned to control group. There is no selection bias because patients in treatment and control groups are similar in all aspects except the exposures (treatment or control) received. However, in observational studies, the treatment assignment variable is more likely related to patient's health condition, and patient's health condition may also related to the outcome variable. Therefore, there are con-

founding variables between treatment assignment variable and outcome variable. To obtain an unbiased ATE estimate, the confounding variables must be controlled. The commonly used propensity score based ATE estimates are valid under the following assumptions.

(1) Exchangeability (i.e. conditional independence):

$$T \perp\!\!\!\perp Y^{(t)}|X \ \text{ for } t \in \{0,1\}.$$

(2) Positivity:

$$Pr(T = t|X = x) > 0$$

for $t \in \{0,1\}$ and all values x with $Pr(X = x) \neq 0$. That is, given $X = x$, the subject has chance to receive treatment as well as control.

(3) Consistency:

$$Y = Y^{(0)}\mathbb{1}_{\{T=0\}} + Y^{(1)}\mathbb{1}_{\{T=1\}}.$$

The propensity score is defined as the probability of a subject receiving treatment given the covariates X, i.e., $Pr(T = 1|X)$. The inverse probability of treatment weighting (IPTW) is built on the propensity score model: the weight $w = \frac{1}{Pr(T=1|X)}$ if the subject with covariate X is in treatment group and $w = \frac{1}{1-Pr(T=1|X)}$ if the subject with covariate X is in control group. That is, $w = \mathbb{1}_{\{T=1\}}\frac{1}{Pr(T=1|X)} + \mathbb{1}_{\{T=0\}}\frac{1}{1-Pr(T=1|X)}$. Under the assumptions of positivity, consistency, and exchangeability, it has been proved that the IPTW weighted mean of observed outcome is equal to the counterfactual mean $E[Y^{(t)}]$ (Hernán and Robins (2019)):

$$E[Y^{(t)}] = E[\frac{\mathbb{1}_{\{T=t\}}}{Pr(T = t|X)} Y]. \tag{3.1}$$

This equality implies that the IPTW weighted mean of observed outcome can be used to estimate the mean of the potential outcomes, further to estimate ATE. This equality (3.1) can be obtained from the following derivation:

$$
\begin{aligned}
& E\left\{ \frac{\mathbb{1}_{\{T=t\}}}{Pr(T = t|X)} Y \right\} && \text{(by positivity assumption)} \\
& = E\left\{ \frac{\mathbb{1}_{\{T=t\}}}{Pr(T = t|X)} [Y^{(1)}T + Y^{(0)}(1 - T)] \right\} && \text{(by consistency assumption)} \\
& = E\left\{ \frac{\mathbb{1}_{\{T=t\}}}{Pr(T = t|X)} Y^{(t)} \right\} \\
& = E_X\left\{ E_{(Y^{(t)},T)|X}\left[\frac{\mathbb{1}_{\{T=t\}}}{Pr(T = t|X)} Y^{(t)} \bigg| X \right] \right\} \\
& = E_X\left\{ E_{T|X}\left[\frac{\mathbb{1}_{\{T=t\}}}{Pr(T = t|X)} \bigg| X \right] E_{Y^{(t)}|X}\left[Y^{(t)} \bigg| X \right] \right\} && \text{(by exchangeability assumption)} \\
& = E_X\left\{ 1 \cdot E_{Y^{(t)}|X}\left[Y^{(t)} \bigg| X \right] \right\} \\
& = E[Y^{(t)}].
\end{aligned}
$$

3.2.2 DAG and the exchangeability assumption

A DAG is both directed and acyclic, which displays the hypothetical relationships between different variables that are causally related (Pearl (2000)). DAGs have been developed to facilitate communication between data analysts and subject matter experts to make causal inference and estimate ATE (Pearl (2000)). In this subsection, we first introduce the d-separation and backdoor criteria under the scheme of DAG, which are closely related to the exchangeability assumption. This assumption is further illustrated by using the single world intervention graph (SWIG) (Richardson and

Robins (2013)) which unifies the DAG and potential outcomes (see Figure 3.1c).

In a DAG, d-separation (Pearl (2014)) plays an important role in assessing the independence and conditional independence between different variables or different sets of variables. Two variables being d-separated (blocked) by a set of nodes (say S) in a DAG indicates that these two variables are conditionally independent given the set of variables S (Pearl (2000)). The basic building blocks of a DAG are fork (Figure 3.1d), chain (Figure 3.1e) and collider (Figure 3.1f). A fork (Figure 3.1d) shows the relationship that two nodes (e.g., variables T and Y) have the same cause (or same parent) X_C. X_C is often referred as a confounding variable. T and Y are dependent even if T doesn't have direct causal effect on Y. However, if we condition on X_C (Figure 3.1d), T and Y are d-separated. That is, $T \perp\!\!\!\perp Y | X_C$ (i.e., T and Y are conditional independent). A chain (Figure 3.1e) is that a variable (say T) has a causal effect on the middle variable X_M which further has a causal effect on the third variable (say Y). X_M is termed as mediator. T and Y are d-separated given X_M (i.e., $T \perp\!\!\!\perp Y | X_M$). A collider (Figure 3.1f) is a descendant (say X_N) of two nodes (e.g., T and Y). That is, T and Y are the causes of X_N. In Figure 3.1f, T and Y are marginally independent, however, T and Y become dependent when we condition on X_N (Pearl et al. (2016)). A path from T to Y is said to be d-separated (or blocked) by a set of nodes in S if (i) the path contains a fork (Figure 3.1d) and the middle node of the fork (e.g., the confounding variable X_C) is in S; (ii) the path contains a chain (Figure 3.1e) and the middle node of the chain (e.g., the mediator X_M) is in S; (iii) the path contains a collider (Figure 3.1f) and the middle node of the collider (e.g., X_N) and its descendant are not in S (Pearl et al. (2016)).

Although d-separation can be used to assess dependence or conditional independence between two variables (or two sets of variables), if we would like to estimate the ATE (total causal effect), the set of the variables to be adjusted in the propensity score model should meet the backdoor criteria. A backdoor path from T to Y is a

path between T and Y with an arrow into T. If there is a backdoor path from T to Y, T and Y become dependent even if T does not have causal relationship with Y (Pearl (2000)). A set of nodes (say, S) satisfy the backdoor criteria for the ordered pair $\{T, Y\}$ if and only if (i) no node in S is a descendant of T; and (ii) nodes in S d-separate (block) every backdoor path between T and Y (Pearl et al. (2016)). If a set of nodes S satisfy the backdoor criteria from T to Y, the causal effect of T on Y can be obtained by propensity score based method, where the nodes in S are included in the propensity score model. Among all of the basic causal relationships in DAGs (Figures 3.1d- 3.1f), only Figure 3.1d shows a backdoor path from T to Y. There are no backdoor path from T to Y in Figures 3.1e and 3.1f. According to the backdoor criteria, we should only include the confounding variable X_C rather than mediator X_M and collider X_N in the set of nodes S.

The backdoor criteria provides a way to achieve the conditional independence that $T \perp\!\!\!\perp Y|S$. However, the notion of the conditional independence between T and observed outcome Y given S lack direct connection with the exchangeability assumption which uses potential outcome $Y^{(t)}(t = 0, 1)$. The SWIG is a unification of DAG and potential outcome via node-splitting (Richardson and Robins (2013)). In a SWIG, the node T is split into two nodes: the random variable T and the potential intervention t. Any arrow entering into T still remains entering T, and any arrow emanating from node T in the DAG becomes an arrow emanating from the node t in the SWIG. Furthermore, the child of the node t becomes potential outcome $Y^{(t)}$. For example, the SWIG of Figure 3.1a is shown in Figure 3.1c, and the SWIGs of Figures 3.1d, 3.1e, and 3.1f are shown in Figures 3.1g, 3.1h, and 3.1i, respectively. SWIG reveals the link between DAG and the exchangeability assumption explicitly.

Under the scheme of SWIG, we can use backdoor criteria to identify a set of nodes S which block the backdoor path from T to $Y^{(t)}$. For example, in Figures 3.1h and 3.1i, there is no backdoor path between T and $Y^{(t)}$. Therefore, a mediator or

collider should not be adjusted (or controlled) in the propensity score model when we examine the causal treatment effect of T on outcome variable. In Figures 3.1g, the confounding variable X_C is on the backdoor path between T and $Y^{(t)}$, and the treatment assignment variable T and potential outcome $Y^{(t)}$ are marginally dependent. However, T and $Y^{(t)}$ become independent (i.e., blocked or d-separated) once we condition on the confounding variable X_C, i.e.,

$$T \perp\!\!\!\perp Y^{(t)}|X_C.$$

In a SWIG, given a set of variables (say, S) which meet the backdoor criteria for the pair $(T, Y^{(t)})$, we have $T \perp\!\!\!\perp Y^{(t)}|S$. S is considered as a suitable adjustment set for propensity score model. It can be either a subset of confounding variables or all confounding variables, as long as variables in S block all backdoor paths from T to $Y^{(t)}$.

Besides including the variables in S in the propensity score model, including the predictor variables X_P (see Figures 3.1a-3.1c, variables related to outcome only) can improve the efficiency of the ATE estimates (Craycroft et al. (2020)). Based on the d-separation, we have $T \perp\!\!\!\perp Y^{(t)}|\{X_C, X_I\}$ and $T \perp\!\!\!\perp Y^{(t)}|\{X_C, X_P\}$ (see Figure 3.1c) (Richardson and Robins (2013), Craycroft et al. (2020)). It has been shown that ATE estimates based on IPTW using those three adjustment sets $\{X_C\}$, $\{X_C, X_I\}$, and $\{X_C, X_P\}$ are all unbiased. The variations of the ATE estimates with different adjustment sets have the relationship $Var(\widehat{ATE}_{\{X_C, X_P\}}) \leq Var(\widehat{ATE}_{\{X_C\}}) \leq Var(\widehat{ATE}_{\{X_C, X_I\}})$ (Craycroft et al. (2020)). We propose to estimate the propensity score using the confounding variable in S along with X_P. Then we apply the IPTW to marginal structure model to estimate ATE.

3.2.3 Proposed ATE estimator

Given a set of variables X such that $T \perp\!\!\!\perp Y^{(t)} | X$, the exchangeability assumption holds. In this article, we propose to include the confounding variables in S and the predictor variables X_P in the propensity score model. Once the propensity score models (say $Pr(T = 1 | S, X_P)$) are obtained, we can use the marginal structural model (MSM) and the IPTW technique to estimate the ATE (Robins et al. (2000)). The MSM can be written as

$$g(E(Y^{(t)})) = \gamma_0 + \gamma_1 t, \tag{3.2}$$

where g is a known link function. When Y is a continuous outcome, we may take the link function g as the identity link function, and model (3.2) has the form of $E(Y^{(t)}) = \gamma_0 + \gamma_1 t$. When Y is a binary outcome, we can take the link function as the logit link function, and model (3.2) has the form of $logit(Pr(Y^{(t)} = 1)) = \gamma_0 + \gamma_1 t$. The MSM model (3.2) implies that the mean of the potential outcome under control is captured by parameter γ_0, namely, $g(E(Y^{(0)})) = \gamma_0$. The mean of the potential outcome under treatment is captured by γ_0 and γ_1, that is, $g(E(Y^{(1)})) = \gamma_0 + \gamma_1$. In RCTs, treatment assignment is unconfounded with outcome and the association parameters in the association model (3.3) are often unbiased estimates of the causal parameters in MSM (3.2).

$$g(E(Y|T)) = \beta_0 + \beta_1 T. \tag{3.3}$$

However, in observational studies, the treatment assignment and outcome variables are often confounded by the confounding variables X_C. The estimated parameters $(\hat{\beta}_0, \hat{\beta}_1)$ in model (3.3) based on observational data are often biased estimates for (γ_0, γ_1) in the MSM (3.2). Robins et al. (2000) proposed to use a weighted sample to obtain the estimated parameters $(\hat{\beta}_0, \hat{\beta}_1)$ in model (3.3). The estimated association

parameters $(\hat{\beta}_0, \hat{\beta}_1)$ based on the weighted sample are unbiased estimators of the causal parameters (γ_0, γ_1) in model (3.2). The weighted sample is obtained by creating w_i copies of ith subject $(i = 1, \cdots, n)$, where w_i is the IPTW. Specifically, if the ith subject is in the treatment group, $w_i = \frac{1}{Pr(T_i=1|X_i)}$; and if the ith subject is in the control group, $w_i = \frac{1}{Pr(T_i=0|X_i)}$. We can prove that the sample size of the treatment group in the weighted sample (say N_1^*) is asymptotically same as the sample size of the control group in the weighted sample (say N_0^*), which is asymptotically same as the sample size in the original sample (say N). This can be shown from the following equations:

$$N_1^* = E\left(\sum_{i=1}^{N} w_i T_i\right) = E\left(\sum_{i=1}^{N} \frac{\mathbb{1}_{\{T_i=1\}}}{Pr(T_i = 1|X_i)}\right) = E\left(\sum_{i=1}^{N} \frac{E(\mathbb{1}_{\{T_i=1\}}|X_i)}{Pr(T_i = 1|X_i)}\right) = N$$

and similarly,

$$N_0^* = E\left(\sum_{i=1}^{N} w_i(1 - T_i)\right) = N.$$

We can also prove that the marginal distributions of X in the original sample is same as the marginal distribution of X in the treatment group of the weighted sample, as well as the marginal distribution of X in the control group of the weighted sample. That is,

$$
\begin{aligned}
f_w(x|T = 1) \quad &\propto \quad \frac{f(x|T = 1)}{Pr(T = 1|x)} \qquad \text{(by definition for the weighted sample)} \\
&\propto \quad \frac{f(x, T = 1)}{f(T = 1|x)f(T = 1)} \\
&\propto \quad \frac{f(T = 1|x)f(x)}{f(T = 1|x)f(T = 1)} \\
&\propto \quad f(x). \qquad\qquad\qquad\qquad\qquad\qquad\qquad (3.4)
\end{aligned}
$$

Similarly, $f_w(x|T = 0) \propto \frac{f(x|T=0)}{Pr(T=0|x)} \propto f(x)$. Thus, there is no selection bias in the weighted sample. The estimated association parameters $(\hat{\beta}_0, \hat{\beta}_1)$ based on the

weighted sample are unbiased estimates for the causal parameters (γ_0, γ_1). To carry out the estimation, we consider a weighted likelihood function with weight w_i for each subject i $(i = 1, \cdots, N)$. When Y is a binary variable, the likelihood function based on the weighted sample can be written as

$$L(\beta|\underline{y}) = \prod_{i=1}^{n} f(y_i)^{w_i} = \prod_{i=1}^{n} [p_i^{y_i}(1 - p_i)^{1-y_i}]^{w_i},$$

where $p_i = Pr(Y_i = 1|T_i)$ and $logit(p_i) = \beta_0 + \beta_1 T_i$ $(i = 1, \cdots, n)$. The corresponding weighted log-likelihood function can be written as:

$$l(\beta|\underline{y}) = \sum_{i=1}^{n} \{w_i y_i(\beta_0 + \beta_1 T_i) - w_i log[1 + exp(\beta_0 + \beta_1 T_i)]\}.$$

Using the Newton−Raphson method (Ypma (1995)), we can get the estimated parameters $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)^T$. The ATE estimate can be obtained as

$$\widehat{ATE} = \frac{exp(\hat{\beta}_0 + \hat{\beta}_1)}{1 + exp(\hat{\beta}_0 + \hat{\beta}_1)} - \frac{exp(\hat{\beta}_0)}{1 + exp(\hat{\beta}_0)}. \tag{3.5}$$

A robust (sandwich) variance estimator for $\hat{\beta}$ can be written as

$$\widehat{Var}(\hat{\beta}) = \hat{V}_{EH} \hat{B}_s \hat{V}_{EH}^{-1},$$

where $\hat{V}_{EH} = E_Y \left[-\frac{\partial l^2}{\partial \beta \partial \beta^T} \Big|_{\beta=\hat{\beta}} \right]$ and $\hat{B}_s = \begin{pmatrix} \sum_{i=1}^{N} \frac{\partial l_i}{\partial \beta_0} \frac{\partial l_i}{\partial \beta_0} & \sum_{i=1}^{N} \frac{\partial l_i}{\partial \beta_0} \frac{\partial l_i}{\partial \beta_1} \\ \sum_{i=1}^{N} \frac{\partial l_i}{\partial \beta_0} \frac{\partial l_i}{\partial \beta_1} & \sum_{i=1}^{N} \frac{\partial l_i}{\partial \beta_1} \frac{\partial l_i}{\partial \beta_1} \end{pmatrix} \Big|_{\beta=\hat{\beta}}$. The variance for $\widehat{ATE}$ can be estimated as

$$Var(\widehat{ATE}) = Var(\hat{\beta}_0) \left[\frac{\partial g}{\partial \beta_0} \Big|_{\beta=\hat{\beta}} \right]^2 + Var(\hat{\beta}_1) \left[\frac{\partial g}{\partial \beta_1} \Big|_{\beta=\hat{\beta}} \right]^2 + 2Cov(\hat{\beta}_0, \hat{\beta}_1) \left[\frac{\partial g}{\partial \beta_0} \frac{\partial g}{\partial \beta_1} \Big|_{\beta=\hat{\beta}} \right],$$

where

$$g(\beta_0, \beta_1) = \frac{exp(\beta_0 + \beta_1)}{1 + exp(\beta_0 + \beta_1)} - \frac{exp(\beta_0)}{1 + exp(\beta_0)},$$

$$\frac{\partial g}{\partial \beta_0} = \frac{exp(\beta_0 + \beta_1)}{(1 + exp(\beta_0 + \beta_1))^2} - \frac{exp(\beta_0)}{(1 + exp(\beta_0))^2},$$

and

$$\frac{\partial g}{\partial \beta_1} = \frac{exp(\beta_0 + \beta_1)}{(1 + exp(\beta_0 + \beta_1))^2}.$$

It is worth mentioning that the sandwich variance estimate may not take the variation of estimating the propensity scores into account. As an alternative, the variance estimate of $\widehat{ATE}$ can be obtained by bootstrap method. That is, B (say, 100) bootstrap samples are obtained from the original sample. For the bth bootstrap sample $(b = 1, \cdots, B)$, the $\widehat{ATE}^{*(b)}$ is estimated using the same procedure as outlined in equation (3.5). The variance of the $\widehat{ATE}$ is obtained by the variance of $\widehat{ATE}^{*(b)}$ $(b = 1, \cdots, B)$ based on B bootstrap samples.

Note that a simple ATE estimate of the following form has been used in the literature:

$$\widehat{ATE} = \frac{\sum_{i=1}^{n} \mathbb{1}_{\{T_i=1\}} Y_i w_i}{\sum_{i=1}^{n} \mathbb{1}_{\{T_i=1\}} w_i} - \frac{\sum_{i=1}^{n} \mathbb{1}_{\{T_i=0\}} Y_i w_i}{\sum_{i=1}^{n} \mathbb{1}_{\{T_i=0\}} w_i}.$$

The ATE estimate in equation (3.5) is equivalent to this simple ATE estimate when there are only treatment and control groups. However, the MSM approach can be extended to multiple groups, even to treatments in continuous scale such as different dose levels. Another advantage to use MSM is that we can easily obtain the robust sandwich variance estimate based on MSM, otherwise one has to use the bootstrap method to estimate the variance if the simple ATE estimate is used.

3.3 Case study

Bronchopulmonary dysplasia (BPD) is a form of chronic lung disease and a common sequelae in premature infants who require oxygen and assisted ventilation

for respiratory distress syndrome (Northway et al. (1967)). Oxygen delivery often requires assisted ventilation, which may cause damage to the developing respiratory system, in particular lung injury. The human respiratory system is composed of upper and lower airways. The upper airway includes the nose, nasal passages, sinuses, pharynx and a portion of the larynx. The lower airway includes the distal portion of the larynx, trachea, bronchi, and bronchioles leading to the alveoli. The trachea is a tubular structure that carries air from the larynx to the lungs. When a child requires a ventilator for a long time, a tracheostomy may be needed for airway management. Tracheostomy is a surgical procedure that creates an opening through the neck into the windpipe. For infants requiring prolonged assisted ventilation, placement of a tracheostomy tube is safer and more comfortable, and it also allows easier movement of the baby than using a standard endotracheal tube. However, some studies suggest that children with a tracheostomy may have higher mortality than those who do not (Watters (2017)).

In this section, we investigated the in-hospital mortality related to tracheostomy surgery and explored if tracheostomy may cause a higher mortality. We used the 2016 HCUP KID database and constructed a cohort of children with BPD based on the International Classification of Disease, Tenth Revision, Clinical Modification (ICD-10-CM) diagnoses codes. We excluded children who had tracheostomy prior to the hospital stays. We defined tracheostomy procedure based on the International Classification of Disease, Tenth Revision, Procedure Coding System (ICD-10-PCS) procedure codes and used the in-hospital death as outcome. The flow chart of generating the study cohort is shown in Figure 5.3 in Appendix and there were 12806 subjects in the study cohort.

Based on th expert knowledge, the causal paths of the variables are constructed as shown in Figure 3.2. We first examined the variables which may be causally related to tracheostomy surgery. Mechanical ventilation is a major treatment option for in-

fants with chronic respiratory failure (CRF) (Fedor (2017)). For patients with severe CRF, tracheostomy may be needed. Extreme low-birth-weight preterm infants requiring prolongs assisted ventilation may require a tracheostomy (Pereira et al. (2003)). Further, premature infants with BPD often have cardiovascular sequelae, such as pulmonary hypertension and CRF, which may result in the need for prolonged assisted ventilation (Kim (2010)). In adults, tracheostomy may reverse the associated cardiovascular comorbid conditions, such as pulmonary hypertension (Thatcher and Maisel (2003)). In addition, preterm delivery is intuitively associated with low birth weight. The long-term use of ventilation, low birth weight, preterm, CRF, and neonatal pulmonary hypertension (NPH) are important factors for doctors to consider in choosing tracheostomy as a treatment for patients with BPD.

Next, we came to examine the variables which are related to the outcome variable (i.e., in-hospital death). Tracheostomy might induce the tracheostomy complications. Preterm birth, CRF, NPH, tracheostomy complications, and sepsis may increase the mortality rate. The associations of different covariates with treatment (tracheostomy) and outcome (i.e. in-hospital death) are presented in Table 3.1. These associations clearly support our understanding of the relationships among these variables, which are depicted in a DAG (Figure 3.2). In the DAG, the five variables above the horizontal line from "tracheostomy" to "death" are all confounding variables. Specifically, "ventilation" is both a confounding variable (on the backdoor path $tracheostomy \leftarrow ventilation \rightarrow death$) and a collider (on the backdoor path $tracheostomy \leftarrow CRF \rightarrow ventilation \leftarrow NPH \rightarrow death$). To block the backdoor path $tracheostomy \leftarrow ventilation \rightarrow death$, we have to condition on "ventilation" in the propensity score model. To block the backdoor path $tracheostomy \leftarrow CRF \rightarrow ventilation \leftarrow NPH \rightarrow death$, we can condition on either CRF or NPH along with "ventilation". Thus, to block both backdoor paths we need to condition on either {CRF, ventilation} or {NPH,ventilation}. To block the backdoor path

tracheostomy ← ventilation ← NPH → death, either "ventilation" or NPH should be included in the propensity score model. To block the backdoor path *tracheostomy ← ventilation ← CRF → death*, either "ventilation" or CRF should be included in the propensity score model. To block the backdoor path *tracheostomy ← CRF → death*, the node CRF must be included in the propensity score model. To block the other backdoor paths, either the node extreme low birth weight or the node extreme preterm needs to be included in the propensity score model. Therefore, the minimal set of confounding variables which block all the backdoor paths from the node "tracheostomy" to the node "death" is either the set S_1={ventilation, CRF, NPH, extreme low birth weight} or the set S_2={ventilation, CRF, NPH, extreme preterm}. We only need to include the variables in set S_1 or S_2 in the propensity score model to obtain unbiased estimates for ATE. The variable "sepsis" is a predictor, and including the predictor "sepsis" in the propensity score model may improve the accuracy of ATE estimates. We used eight different sets of covariates in the propensity score models: (1) minimal set of confounding variables S_1={ventilation, CRF, NPH, extreme low birth weight}, (2) minimal set of confounding variables S_2={ventilation, CRF, NPH, extreme preterm}, (3) all confounding variables S_{AllX_C}={ventilation, CRF, NPH, extreme low birth weight, extreme preterm}, (4) $S_1 + X_P$, (5) $S_2 + X_P$, (6) $S_{AllX_C} + X_P$, (7) all covariates, and (8) no covariates. The results based on different sets of adjusted covariates in the propensity score models are presented in Table 3.2. Based on Table 3.2, we may conclude that (1) the ATE estimates were quite similar when the first six sets of variables were included in the propensity score model, and the ATE estimates due to tracheostomy were between 1.5% to 1.6%, with zero in the 95% confidence intervals, indicating that tracheostomy surgery may not cause higher in-hospital mortality; (2) the ATE estimate with all covariates $\{X_C, X_P, X_M\}$ in the propensity score model resulted in a larger ATE estimate at 2.6%; and (3) the ATE estimate without adjusting any covariates was 2.7%. Based on the theoretical consid-

eration and the simulation results in the next section, the ATE estimate in the latter two cases were more likely to be biased estimates for ATE.
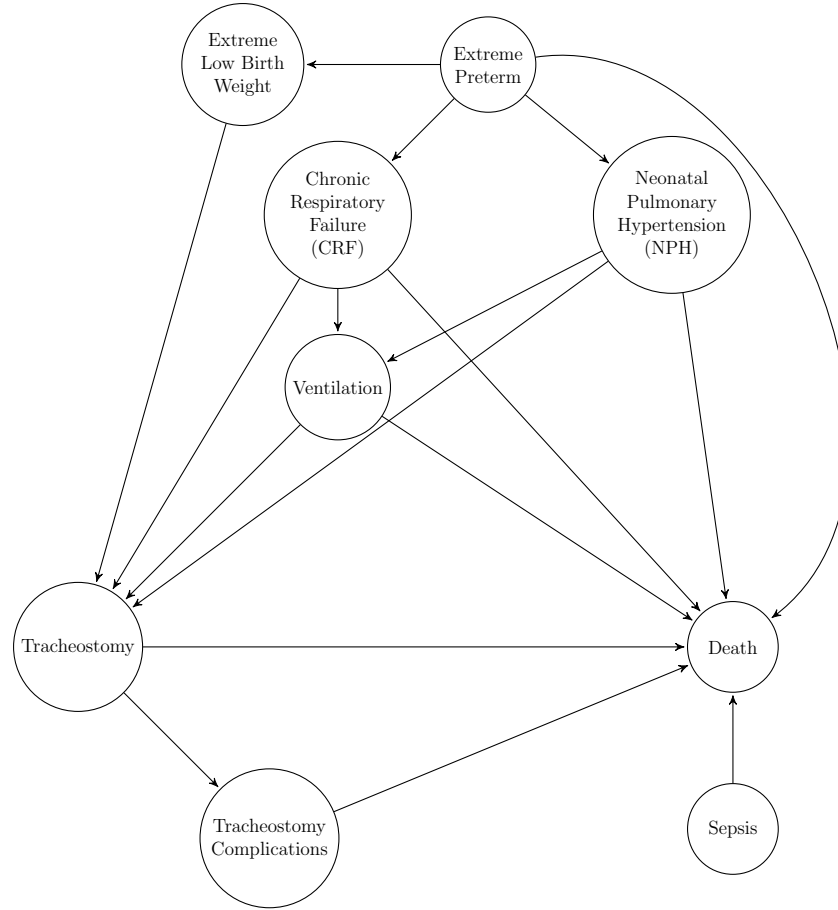
Figure 3.2: Directed acyclic graph (DAG) for children with BPD

Table 3.1: The associations of different covariates with the treatment assignment variable (i.e. tracheostomy) as well as the outcome variable (i.e. in-hospital death) based on 2016 HCUP KID database

Covariates		N (%)	Tracheostomy (%)	Mortality (%)
Tracheostomy	Yes	227(1.8%)	100.0%	5.7%
	No	12566(98.2%)	0.0%	3.0%
Tracheostomy Complications	Yes	74(0.6%)	100.0%	2.7%
	No	12719(99.4%)	1.2%	3.0%
Ventilation	Yes	487(3.8%)	21.8%	12.7%
	No	12306(96.2%)	1.0%	2.6%
Extreme Preterm ≤ 27 weeks	Yes	4841(37.8%)	1.5%	3.7%
	No	7952(62.2%)	1.9%	2.6%
Birth Weight < 1000g	Yes	4348(34%)	1.4%	4.2%
	No	8445(66%)	1.9%	2.4%
Chronic Respiratory Failure	Yes	363(2.8%)	16.8%	7.4%
(CRF)	No	12430(97.2%)	1.3%	2.9%
Neonatal Pulmonary Hypertention	Yes	1236(9.7%)	6.3%	6.5%
(NPH)	No	11557(90.3%)	1.3%	2.7%
Sepsis	Yes	1565(12.2%)	0.9%	6.7%
	No	11228(87.8%)	1.9%	2.5%

Table 3.2: The estimated average treatment effect (ATE) on in-hospital death due to tracheostomy based on 2016 HCUP KID database using different sets of variables in the propensity score models

Variables in propensity score	ATE Estimates	Bootstrap SE	Bootstrap 95% CI	Model based SE	Model based 95% CI
Minimal set S_1: Ventilation, CRF, NPH, Low Birth Weight	1.5%	2.0%	(-2.3%, 5.4%)	1.9%	(-2.2%, 5.3%)
Minimal set S_2: Ventilation, CRF, NPH, Extreme Preterm	1.6%	1.9%	(-2.2%, 5.4%)	1.9%	(-2.2%, 5.4%)
Set S_{AllX_C}: Ventilation, CRF, NPH, Low Birth Weight, Extreme Preterm	1.6%	2.1%	(-2.5%, 5.7%)	1.9%	(-2.2%, 5.4%)
Minimal set $S_1 + X_P$: Ventilation, CRF, NPH, Low Birth Weight, Sepsis	1.6%	1.9%	(-2.1%, 5.2%)	1.9%	(-2.1%, 5.2%)
Minimal set $S_2 + X_P$: Ventilation, CRF, NPH, Extreme Preterm, Sepsis	1.6%	1.8%	(-1.9%, 5.1%)	1.9%	(-2.1%, 5.3%)
Set $S_{AllX_C} + X_P$: Ventilation, CRF, NPH, Low Birth Weight, Extreme Preterm, Sepsis	1.6%	1.8%	(-1.9%, 5.2%)	1.9%	(-2.1%, 5.3%)
All covariates: Ventilation, CRF, NPH, Low Birth Weight, Extreme Preterm, Sepsis Tracheostomy Complications	2.6%	2.6%	(-2.5%, 7.8%)	2.6%	(-2.6%, 7.8%)
No covariates	2.7%	1.7%	(-0.5%, 6%)	1.5%	(-0.3%, 5.8%)

Note: Low Birth Weight indicates birth weight $<1000g$; Extreme Preterm indicates gestation age $\leq 27weeks$.

3.4 Simulation

In this section, we carried out simulation studies to examine the consistency and accuracy of ATE estimates when different types of covariates were included in the propensity score models. We used a simplified model (see Figure 3.1a) in Section 3.4.1 and a more complex model in Section 3.4.2 to illustrate the impact on ATE estimates due to different types of covariates in the propensity score models.

3.4.1 Simulation studies for different types of covariates in a simplified DAG

In this simulation study, we generated data based on the causal structure presented in Figure 3.1a and examined the performance in estimating ATE when different types of covariates were included in the propensity score models. In this simulation, we ignored the spurious variables X_S and generated one variable for each type of covariates. We performed the simulation studies for different sample sizes (say, n=200, 1000, and 10000). For each sample size, we carried out the simulation study using the following steps.

Step 1: Generated n observations of the confounding variable X_C from the uniform

distribution $U(0,5)$; generated n observations for the instrument variable X_I as well as the predictor X_P from the standard normal distribution.

Step 2: Generated n observations of the treatment T as a binary variable with probability obtained from

$$logit(Pr(T = 1|X_C, X_I)) = 0.01 + 0.7\ X_C - 2.5\ X_I.$$

Step 3: Generated n observations of the mediator X_M as a binary variable with probability obtained from

$$logit(Pr(X_M = 1|T)) = 0.01 + 0.8\ T.$$

Step 4: Generated n observations of the outcome Y which was causally related to confounding variable X_C, treatment T, mediator X_M, and predictor X_P. Here we assumed that Y followed the Bernoulli distribution with probability determined by the following model:

$$logit(Pr(Y = 1|X_C, X_M, X_P, T)) = 0.01 - 2.7\ X_C + 1.5\ X_M + 5\ X_P + \tau\ T.$$

Step 5: Generated n observations of the collider X_N as a binary variable with probability obtained from

$$logit(Pr(X_N = 1|T, Y)) = 0.01 + 0.7\ T + 0.8\ Y.$$

Step 6: Generated potential outcomes by using the outcome model in Step 4 but fixing the treatment as t (t=0 or 1) and taking X_M as $X_M^{(t)}$. $X_M^{(0)}$ and $X_M^{(1)}$ were obtained from Step 3. The potential outcome was a binary variable with prob-

ability obtained from

$$logit(Pr(Y^{(t)} = 1 | X_C, X_M^{(t)}, X_P, T = t)) = 0.01 - 2.7X_C + 1.5X_M^{(t)} + 5X_P + \tau t.$$

Step 7: Obtained the true sample ATE as the mean difference between potential outcomes $Y^{(1)}$ and $Y^{(0)}$ in Step 7, depending on the sample data generated from Steps 1-6.

Step 8: Estimated the propensity scores using each one of the seven different sets of covariates (see Table 3.3) based on the n observations generated from Steps 1-5. Depending on each set of covariates, we obtained their corresponding ATE estimates by using MSM. We also obtained the standard errors by using the sandwich estimate and the bootstrap method, respectively.

Step 9: Repeated Steps 1-8 for 1000 times. Here we set $\tau = 0.5$.

Step 10: Obtained the true ATE by the mean of the 1000 sample ATEs in Step 8.

Step 11: Presented the mean of the estimated ATEs, the bias, the empirical standard error, the standard error obtained from the bootstrap method, the standard error based on sandwich estimate, and the coverage rate (Table 3.3). The coverage rate was obtained as the percentage of true ATE falling in the 1000 95% confidence intervals of the estimated ATEs. The bias was obtained as the absolute difference between the true ATE and mean of the estimated ATEs. The empirical standard error was the standard error of the 1000 estimated ATEs. The standard error from bootstrap method and sandwich estimate is the mean of 1000 corresponding standard errors obtained from Step 8.

Table 3.3: Simulation results for including different types of covariates in the propensity score models (see the DAG in Figure 3.1a). Here τ=0.5 and true $ATE = 3.3\%$.

Covariates in propensity score model	Estimated ATE	Bias	Empirical Std.Error	Bootstrap Std.Error	Bootstrap Coverage Rate	Model based Std.Error	Model based Coverage Rate
sample size=200							
X_N	-7.7%	11.0%	6.8%	6.8%	64.5%	6.8%	65.6%
X_M	-7.0%	10.3%	6.9%	6.9%	71.7%	6.8%	71.5%
X_N, X_C	1.6%	1.7%	5.8%	5.7%	92.7%	6.4%	96.3%
X_M, X_C	2.1%	1.2%	5.8%	5.8%	94.0%	6.4%	96.3%
X_C	3.2%	0.1%	5.5%	5.5%	93.9%	6.1%	96.5%
X_C, X_P	3.3%	0.0%	4.3%	4.4%	95.9%	6.1%	99.4%
X_C, X_I	1.5%	1.8%	11.9%	9.6%	90.8%	9.5%	91.9%
sample size=1000							
X_N	-7.4%	10.7%	3.0%	3.0%	5.1%	3.1%	5.7%
X_M	-6.7%	10.0%	3.0%	3.0%	7.7%	3.1%	7.6%
X_N X_C	1.8%	1.5%	2.6%	2.5%	91.1%	2.8%	94.9%
X_M X_C	2.2%	1.1%	2.6%	2.5%	93.5%	2.8%	96.3%
X_C	3.4%	0.1%	2.5%	2.4%	94.9%	2.7%	97.4%
X_C X_P	3.3%	0.0%	2.0%	1.9%	93.7%	2.7%	99.4%
X_C X_I	2.9%	0.4%	7.4%	5.5%	92.0%	5.7%	92.8%
sample size=10000							
X_N	-7.6%	10.9%	1.0%	1.0%	0.0%	1.0%	0.0%
X_M	-6.9%	10.2%	1.0%	1.0%	0.0%	1.0%	0.0%
X_N X_C	1.7%	1.6%	0.8%	0.8%	46.9%	0.9%	55.9%
X_M X_C	2.1%	1.2%	0.8%	0.8%	67.0%	0.9%	75.3%
X_C	3.3%	0.0%	0.8%	0.8%	94.9%	0.9%	98.0%
X_C X_P	3.3%	0.0%	0.6%	0.6%	95.0%	0.9%	99.7%
X_C X_I	3.1%	0.2%	3.4%	2.5%	94.5%	2.6%	95.1%

X_C: confounding variable; X_I: instrumental variable; X_P: predictor variable; X_M: mediator; X_N: collider.

Based on the simulation results shown in Table 3.3, we concluded that (1) the true ATE under this setting was 3.3%; (2) the ATE estimates corresponding to the sets $\{X_C\}$, $\{X_C, X_P\}$, $\{X_C, X_I\}$ were close to the true ATE; their bias decreased and the accuracy improved as the sample size went large; (3) when only X_C was included in the propensity score model, ATE estimate was unbiased with coverage rate close to 95%; (4) when either mediator X_M or collider X_N was included in the propensity score model, the ATE estimate was biased, regardless the sample sizes; (5) when

both X_C and X_P were included in the propensity score model, the ATE estimate was unbiased and the accuracy was improved (i.e., smallest standard error); (6) when X_C and the instrument variable X_I were included in the propensity score model, the variance increased, though the ATE estimate was unbiased. In summary, including X_C and X_P in the propensity score model provided unbiased estimate for ATE and improved the accuracy of ATE estimate.

3.4.2 Simulation study for a complex model based on the DAG of BPD

In this simulation study, we generated data based on the DAG presented in the case study in Figure 3.2, where the treatment choice was determined by patients characteristics. For children with BPD, extreme low birth weight, CRF, NPH, and long-term ventilation often were causally related to a higher chance to receive the tracheostomy surgery, which were also causally related to the outcome (say, in-hospital death). In addition, the tracheostomy surgery was causally related to tracheostomy complications which might increase the mortality. Sepsis was only related to outcome and also might increase the mortality. We generated simulated datasets and examined the performance of the proposed method using the following simulation settings.

Step 1: Generated confounding variables: extreme preterm (EP) (i.e., gestational age $\leq$ 27 weeks), extreme low birth weight (ELBW) (i.e., birth weight $<$ 1000g), CRF, NPH, and ventilation based on Bernoulli distribution with probabilities obtained from

$$Pr(\text{EP=1}) = 0.3,$$

$$logit(Pr(\text{ELBW=1}|\text{EP})) = 0.02 + 0.4 \text{ EP},$$

$$logit(Pr(\text{CRF} = 1|\text{EP})) = 0.05 + 0.5 \text{ EP},$$

$$logit(Pr(\text{NPH} = 1|\text{EP})) = 0.05 + 0.45 \text{ EP},$$

and $logit(Pr(\text{ventilation} = 1|\text{CRF, NPH})) = 0.01 + 0.7 \text{ CRF} + 0.7 \text{ NPH}.$

Step 2: Generated the treatment assignment variable T (say, tracheostomy surgery) based on the Bernoulli distribution with probability obtained from

$$logit(Pr(T = 1|\text{ELBW, CRF, NPH, ventilation}))$$
$$= 0.3 \text{ ELBW} + 0.4 \text{ CRF} + 0.5 \text{ NPH} + 0.6 \text{ ventilation}.$$

Step 3: Generated the mediator variable postT (say, tracheostomy complications) from the Bernoulli distribution with probability obtained from

$$logit(Pr(\text{postT} = 1|T)) = 0.2 + 0.5 \text{ T}.$$

Step 4: Generated predictor variable sepsis from the Bernoulli distribution with probability obtained from

$$Pr(\text{sepsis} = 1) = 0.2.$$

Step 5: Generated outcome variable Y from Bernoulli distribution with probability obtained from

$$logit(Pr(Y = 1|\text{EP, CRF, NPH, ventilation, postT, sepsis, T}))$$
$$= -1 + 1.5 \text{ EP} + 2 \text{ CRF} + 2 \text{ NPH} + 2.5 \text{ ventilation} + 3 \text{ postT} + 4 \text{ sepsis} + \tau \text{ T}.$$

Step 6: Generated the potential outcomes $Y^{(t)}$ with the treatment fixed at t (0 or 1) by using the Bernoulli distribution with probability obtained from

$$logit(Pr(Y^{(t)} = 1|\text{EP, CRF, NPH, ventilation, postT}^{(t)}, \text{sepsis}, T = t))$$
$$= -1 + 1.5 \text{ EP} + 2 \text{ CRF} + 2 \text{ NPH} + 2.5 \text{ ventilation} + 3 \text{ postT}^{(t)} + 4 \text{ } sepsis + \tau \text{ } t.$$

The potential mediator postT$^{(t)}$ is generated based on the model in Step 3 with treatment fixed at t (0 or 1).

Step 7: Obtained the true sample ATE estimate based on potential outcomes in Step 6 for the sample data generated in Steps 1-5.

Step 8: Estimated ATE by using the MSM and the sample data generated in Steps 1-5. Calculated the standard error based on sandwich variance estimate as well as the bootstrap method. In this step, we included different sets of covariates shown in Table 3.4.

Step 9: Repeated Steps 1-8 1000 times. Here we took three different values of τ to examine the model performance on estimating ATEs.

Step 10: Calculated the true ATE as the mean of the 1000 true sample ATEs generated in Step 7.

Step 11: For each fixed $\tau \in \{0, \text{-}1, 1\}$, we summarized the simulation results by the mean of the estimated ATE, the bias, the empirical standard error, the standard error by using bootstrap method, the standard error by using sandwich estimates, and the coverage rate (See Table 3.4). The coverage rate was obtained as the percentage of true ATE falling in the 1000 95% confidence intervals of the estimated ATEs. The bias was obtained as the absolute difference between the true ATE and mean of the estimated ATEs. The empirical standard error was the standard error of the 1000 estimated ATEs. The standard error from bootstrap method and sandwich estimate is the mean of 1000 corresponding standard errors obtained from Step 8.

The results for this simulation study are presented in Table 3.4, where eight different sets of covariates were included in the propensity score models. They were minimal set of confounding variables S_1, minimal set of confounding variables S_2, all

69

confounding variables S_{AllX_C}, $S_1 + X_P$, $S_2 + X_P$, $S_{AllX_C} + X_P$, all covariates, and no covariates adjusted. From the results in Table 3.4, we concluded that (1) all ATE estimates based on the first six sets of adjusted covariates which included confounding variables blocking all backdoor paths from treatment assignment variable to outcome variable, with/without predictor variable, provided unbiased estimates for ATE; (2) the ATE estimates using all covariates (which included the mediator "tracheostomy complications") in the propensity score models provided biased estimates; (3) the ATE estimates using no covariates in the propensity score models provided biased ATE estimates; (4) when X_P along with the set of confounding variables (which block all backdoor paths from treatment assignment variable to outcome variable) were included in the propensity score model, the empirical standard error was smaller than its counterpart without X_p, indicating that including X_P in the propensity score model could improve the accuracy of ATE estimates; (5) the coverage rates for the first six sets of covariates were close to 95%, the nominal coverage rate; (6) the standard error based on bootstrap method was more close to the empirical standard error than sandwich variance estimate, thus bootstrap method is recommended to estimate standard error.

3.5   Conclusion and discussion

In this article, we have investigated the performance of the propensity score models using different sets of covariates. We have demonstrated that not all covariates should be included in the propensity score model, instead one should only include confounding variables (or a minimal set of confounding variables meeting backdoor criteria) and predictor variables. This approach is suitable for observational studies, when both the treatment assignment variable and outcome variable are impacted by the patient's characteristics. From the case study and simulation studies, when all covariates are included in the propensity score model, the ATE estimates are often

Table 3.4: Simulation results based on the DAG in Figure 3.2

Variables included in propensity score	Estimated ATE	Bias	Empirical Std.Error	Bootstrap Std.Error	Bootstrap Coverage Rate	Model based Std.Error	Model based Coverage Rate
					sample size=5000, $\tau = 0$, true ATE=3.8%		
Minimal set S_1: Ventilation, CRF, NPH, Low Birth Weight	3.8%	0.0%	1.3%	1.3%	93.9%	1.4%	96.0%
Minimal set S_2: Ventilation, CRF, NPH, Extreme Preterm	3.8%	0.0%	1.3%	1.3%	93.8%	1.4%	95.9%
Set S_{AllX_C}: Ventilation, CRF, NPH, Low Birth Weight, Extreme Preterm	3.8%	0.0%	1.3%	1.3%	94.1%	1.4%	95.8%
Minimal set $S_1 + X_P$: Ventilation, CRF, NPH, Low Birth Weight, Sepsis	3.8%	0.0%	1.3%	1.2%	94.3%	1.4%	96.5%
Minimal set $S_2 + X_P$: Ventilation, CRF, NPH, Extreme Preterm, Sepsis	3.8%	0.0%	1.2%	1.2%	93.9%	1.4%	96.6%
Set $S_{AllX_C} + X_P$: Ventilation, CRF, NPH, Low Birth Weight, Extreme Preterm, Sepsis	3.8%	0.0%	1.3%	1.2%	93.8%	1.4%	96.8%
All covariates: Ventilation, CRF, NPH, Low Birth Weight, Extreme Preterm, Sepsis Tracheostomy Complications	0.0%	3.8%	1.1%	1.1%	8.0%	1.3%	15.6%
No covariates	5.5%	1.8%	1.4%	1.4%	72.5%	1.4%	72.9%
					sample size=5000, $\tau = 1$, true ATE= 11.5%		
Minimal set S_1: Ventilation, CRF, NPH, Low Birth Weight	11.5%	0.0%	1.3%	1.3%	93.3%	1.3%	95.5%
Minimal set S_2: Ventilation, CRF, NPH, Extreme Preterm	11.5%	0.0%	1.3%	1.2%	93.5%	1.3%	96.0%
Set S_{AllX_C}: Ventilation, CRF, NPH, Low Birth Weight, Extreme Preterm	11.5%	0.0%	1.3%	1.2%	93.5%	1.3%	95.7%
Minimal set $S_1 + X_P$: Ventilation, CRF, NPH, Low Birth Weight, Sepsis	11.5%	0.0%	1.2%	1.2%	93.2%	1.3%	96.3%
Minimal set $S_2 + X_P$: Ventilation, CRF, NPH, Extreme Preterm, Sepsis	11.5%	0.0%	1.2%	1.2%	93.7%	1.3%	96.7%
Set $S_{AllX_C} + X_P$: Ventilation, CRF, NPH, Low Birth Weight, Extreme Preterm, Sepsis	11.5%	0.0%	1.2%	1.2%	93.3%	1.3%	96.6%
All covariates: Ventilation, CRF, NPH, Low Birth Weight, Extreme Preterm, Sepsis Tracheostomy Complications	8.0%	3.5%	1.1%	1.1%	10.0%	1.3%	17.6%
No covariates	13.1%	1.7%	1.4%	1.3%	74.6%	1.3%	74.7%
					sample size=5000, $\tau = -1$, true ATE=-6.0%		
Minimal set S_1: Ventilation, CRF, NPH, Low Birth Weight	-6.0%	0.0%	1.4%	1.3%	94.2%	1.4%	96.2%
Minimal set S_2: Ventilation, CRF, NPH, Extreme Preterm	-6.0%	0.0%	1.3%	1.3%	94.3%	1.4%	96.5%
Set S_{AllX_C}: Ventilation, CRF, NPH, Low Birth Weight, Extreme Preterm	-6.0%	0.0%	1.3%	1.3%	94.3%	1.4%	96.2%
Minimal set $S_1 + X_P$: Ventilation, CRF, NPH, Low Birth Weight, Sepsis	-6.0%	0.0%	1.3%	1.3%	94.2%	1.4%	96.8%
Minimal set $S_2 + X_P$: Ventilation, CRF, NPH, Extreme Preterm, Sepsis	-6.0%	0.0%	1.3%	1.3%	94.3%	1.4%	96.9%
Set $S_{AllX_C} + X_P$: Ventilation, CRF, NPH, Low Birth Weight, Extreme Preterm, Sepsis	-6.0%	0.0%	1.3%	1.3%	94.3%	1.4%	97.4%
All covariates: Ventilation, CRF, NPH, Low Birth Weight, Extreme Preterm, Sepsis Tracheostomy Complications	-9.9%	3.9%	1.1%	1.1%	7.4%	1.4%	16.2%
No covariates	-4.2%	1.8%	1.5%	1.4%	73.0%	1.4%	73.2%

Note: Low Birth Weight indicates birth weight <1000g; Extreme Preterm indicates gestation age $\leq 27 weeks$.

biased. When a set of confounding variables which block all backdoor paths from T to Y are included in the propensity score model, the ATE estimates are unbiased. Furthermore, including predictors along with confounding variables (meeting the backdoor criteria) in the propensity score model improves the accuracy of the ATE estimates.

In the case study and simulations, we only present the result of ATE estimates based on MSM, which is same as the result based on the simple ATE estimate. The MSM can be easily extended to multiple treatment groups and even continuous treatments, while the simple ATE can not be easily extended to these cases. In addition, the ATE in this paper is referred as the total causal effect which include the effect of treatment through mediator. If one were interested in the direct causal effect, different set of variables need to be adjusted (VanderWeele (2009)), which is beyond the investigation of this article.

There are various propensity score based methods to estimate ATE in observational studies, such as matching, stratification, IPTW, and using propensity score as covariates. Though each propensity score based method has its strengths and weaknesses, we did not compared them in this paper. Instead, we focus on the impact of including different types of covariates in the propensity score model. The investigation of these methods with reference to DAG would be our future research interests.

CHAPTER 4

DIFFERENT DOUBLY ROBUST METHODS FOR ESTIMATING

ATE AND THEIR IMPROVEMENT USING SUPER LEARNER

4.1   Introduction

In this project, we investigate different doubly robust methods for estimating average treatment effect (ATE) based on observational data. ATE measures the mean difference of the potential outcomes when all subjects are treated versus all subjects are untreated. Conceptually, we could assume that each patient has two potential outcomes: $Y^{(1)}$ and $Y^{(0)}$, where $Y^{(1)}$ is the potential outcome when the subject is treated and $Y^{(0)}$ is the potential outcome when the subject is untreated. ATE is defined as:

$$ATE = E(Y^{(1)}) - E(Y^{(0)}).$$

In practice, we can only observe one potential outcome for a subject, i.e., the potential outcome corresponding to the treatment the subject receives. The other potential outcome is unobserved. That is, the observed outcome Y equals $Y^{(0)}$ when the subject is untreated (say $T = 0$), and the observed outcome Y equals $Y^{(1)}$ when the subject is treated (say $T = 1$). In a randomized experiment, each subject is randomly assigned to either treatment group or control group, and the sample in the treatment group can be considered as a random sample from the population. Thus, the sample mean of the outcomes for the treated subjects is a consistent estimate of the population mean when all subjects are treated, say $E(Y^{(1)})$. Similarly, the sample mean of the outcomes

for the untreated subjects is a consistent estimate for the population mean when all subjects are untreated, say $E(Y^{(0)})$. In an observational study, the treatment (say T) is not randomly assigned, instead the treatment choice for a subject often depends on the patient's health conditions. Subsequently, the outcome depends on not only the treatment received but also the patient's health condition. That is, the treatment assignment and the outcome are confounded. Thus, the sample mean for treated subjects is not a consistent estimate for the population mean. Under the assumption that there are not unmeasured confounding variables, propensity score based methods, such as matching, stratification, inverse probability of treatment weighting (IPTW), regression model with propensity score as covariate, and DR methods, have been developed to estimate ATE (Rosenbaum and Rubin (1983), Lunceford and Davidian (2004), Kang et al. (2007), Abdia et al. (2017), and Yan et al. (2019),). Among them, the doubly robust methods use both propensity score model and outcome model, and their ATE estimates are asymptotically consistent if either propensity score model or outcome model is correctly specified. The doubly robust methods gained popularity due to this underlying property. There are different forms of doubly robust estimators, including bias-corrected methods, weighted least squares method, and an augmented doubly robust method (Lunceford and Davidian (2004), Kang et al. (2007), and Bang and Robins (2005)). The bias-corrected method has been widely used (Funk et al. (2011), Yan et al. (2019)), and its comparison with other forms of doubly robust estimators has not yet been made.

Doubly robust methods may provide biased estimates when both propensity score model and outcome model are mis-specified. Recently, the super learner algorithm (Van der Laan et al. (2007), Polley and Van Der Laan (2010)) has been developed to predict outcome with high accuracy. The super learner algorithm finds the best combination of a group of prediction algorithms by minimizing the cross validation error. These prediction algorithms could be different machine learning

methods and generalized additive model (Breiman et al. (1984), Quinlan (1986), Loh (2014), Breiman (1998), Breiman (2001), Breiman (1997), Friedman (2001), Mason et al. (2000), Hastie and Tibshirani (1990)). The super learner algorithm has been implemented in R-package "SuperLearner" (Polley and Van Der Laan (2010)). It is anticipated that doubly robust estimators, when outcome is obtained from super learner, are robust in estimating ATE even when both propensity score model and outcome model are mis-specified. In this project, we first review these doubly robust methods in their original forms in Section 4.2.1 and compare their performances via simulations in Section 4.3.1. It is known that doubly robust methods may fail when both the outcome model and propensity score model are mis-specified. We investigate whether the performance of the doubly robust methods improve when super learner algorithm is applied to obtain the outcome model in Section 4.2.2. Extensive simulation studies are carried out to examine their performances in Section 4.3.2. In Section 4.4, we examine whether quitting smoking could cause weight gain based on the NHEFS dataset. The last section is devoted to discussion and conclusions.

4.2 Doubly robust methods and super learner doubly robust methods

Let X, T, and Y denote, respectively, a vector of confounding variables, a treatment assignment variable, and an outcome variable. Let us assume that there are not unmeasured confounding variables. Propensity score based methods are often used to estimate ATE. The propensity score model is defined as the probability of receiving treatment given the confounding variables, that is,

$$\pi(X) = Pr(T = 1|X). \tag{4.6}$$

Doubly robust methods use both the propensity score (PS) model and outcome model. Let us assume the outcome variable Y is a continuous variable, the outcome model for an untreated subject with confounding variable X is $m_0(X|\beta) = E(Y^{(0)}|X)$, and the outcome model for a treated subject with confounding variable X is $m_1(X|\beta) = E(Y^{(1)}|X)$. For example, a model of the form $E(Y|T,X) = \beta_0 + \beta_T T + X^T \beta_X$ could be fit to the observational data, then $m_0(X|\hat{\beta}) = \hat{\beta}_0 + X^T \hat{\beta}_X$ and $m_1(X|\hat{\beta}) = \hat{\beta}_0 + \hat{\beta}_1 + X^T \hat{\beta}_X$. In the following section, we first present different forms of doubly robust estimators, and then we present the super learner doubly robust method to estimate ATE. We use the triplet (X_i, T_i, Y_i) $(i = 1, \cdots, N)$ to denote the N observations in an observational study.

4.2.1 Doubly robust methods

The Lunceford's doubly robust method (Lunceford and Davidian (2004)) based on N observations has the following form:

$$
\begin{aligned}
ATE_{(DR)}^{(Lunceford)} = {} & \frac{1}{N} \sum_{i=1}^{N} \left\{ \frac{T_i Y_i}{\pi(X_i)} - \frac{T_i - \pi(X_i)}{\pi(X_i)} m_1(X_i|\beta) \right\} \\
& - \frac{1}{N} \sum_{i=1}^{N} \left\{ \frac{(1-T_i)Y_i}{1-\pi(X_i)} + \frac{T_i - \pi(X_i)}{1-\pi(X_i)} m_0(X_i|\beta) \right\}.
\end{aligned}
\tag{4.7}
$$

One can prove that the expectation of the first term equals to $E(Y^{(1)})$ and the expectation of the second term equals $E(Y^{(0)})$ when either the propensity score model or outcome model is correctly specified. In practice, the ATE estimate is obtained by replacing the propensity scores and the outcome models by their estimated values:

$$
\begin{aligned}
\widehat{ATE}_{(DR)}^{(Lunceford)} = {} & \frac{1}{N} \sum_{i=1}^{N} \left\{ \frac{T_i Y_i}{\hat{\pi}(X_i)} - \frac{T_i - \hat{\pi}(X_i)}{\hat{\pi}(X_i)} m_1(X_i|\hat{\beta}) \right\} \\
& - \frac{1}{N} \sum_{i=1}^{N} \left\{ \frac{(1-T_i)Y_i}{1-\hat{\pi}(X_i)} + \frac{T_i - \hat{\pi}(X_i)}{1-\hat{\pi}(X_i)} m_0(X_i|\hat{\beta}) \right\}.
\end{aligned}
\tag{4.8}
$$

The doubly robust estimator in Equation (4.8) is often used in the literature. Kang et al. (2007) proposed two doubly robust methods to estimate ATE. One is based on the residuals (say, $\hat{\epsilon}_i = y_i - [T_i m_1(X_i|\hat{\beta}) + (1 - T_i)m_0(X_i|\hat{\beta})]$ $(i = 1, \cdots, N)$):

$$
\widehat{ATE}_{(DR)}^{(Kang.Res)} = \left\{ \frac{\sum_{i=1}^{N} m_1(X_i|\hat{\beta})}{N} + \frac{\sum_{i=1}^{N} T_i \hat{\pi}(X_i)^{-1} \hat{\epsilon}_i}{\sum_{i=1}^{N} T_i \hat{\pi}(X_i)^{-1}} \right\}
$$
$$
- \left\{ \frac{\sum_{i=1}^{N} m_0(X_i|\hat{\beta})}{N} + \frac{\sum_{i=1}^{N} (1 - T_i)(1 - \hat{\pi}(X_i))^{-1} \hat{\epsilon}_i}{\sum_{i=1}^{N} (1 - T_i)(1 - \hat{\pi}(X_i))^{-1}} \right\}. \tag{4.9}
$$

The Kang's doubly robust estimator (4.9) is asymptotic equivalent to Lunceford's doubly robust estimator (4.8).

The other doubly robust approach proposed by Kang (Kang et al. (2007)) is based on the weighted least squares (WLS) estimates of the outcome models:

$$
\widehat{ATE}_{(DR)}^{(Kang,WLS)} = \frac{1}{N} \sum_{i=1}^{N} \left\{ m_1(X_i|\hat{\beta}_{WLS}) - m_0(X_i|\hat{\beta}_{WLS}) \right\}. \tag{4.10}
$$

In the case that the outcome model is a linear regression model, the model coefficients are estimated as $\hat{\beta}_{WLS} = (H^T W H)^{-1} H^T W Y$, where $H = \{T, X\}$ and the weight matrix W is a diagonal matrix with IPTWs assigned as the diagonal elements.

Bang and Robins (2005) proposed the doubly robust estimator by considering the IPTW as a covariate in the outcome model. The doubly robust method can be written as

$$
\widehat{ATE}_{(DR)}^{(Bang)} = \frac{1}{N} \sum_{i=1}^{N} \left\{ e(T = 1, X_i|\hat{\beta}, \hat{\phi}) - e(T = 0, X_i|\hat{\beta}, \hat{\phi}) \right\}. \tag{4.11}
$$

Here the outcome model has the form $e(T, X_i|\hat{\beta}, \hat{\phi}) = g\left[s(T, X|\beta) + \phi(\frac{T}{\pi(X)} + \frac{(1-T)}{1-\pi(X)})\right]$, where g is a canonical link function, and $s(X|\beta)$ is a known regression function with unknown parameter β. $e(T = 1, X_i|\hat{\beta}, \hat{\phi})$ is the predicted potential outcomes when

the ith subjects is treated; while $e(T = 0, X_i|\hat{\beta}, \hat{\phi})$ is the predicted potential outcome when the ith subjects is untreated. $\hat{\beta}$ and $\hat{\phi}$ are the solutions of the estimating equation $0 = \sum_{i=1}^{N} \frac{\partial e(T_i, X_i|\beta, \phi)}{\partial(\beta^T, \phi)}\{Y_i - e(T_i, X_i|\beta, \phi)\}$.

4.2.2 Super learner doubly robust methods

Super learner algorithm ensembles various algorithms, such as machine learning methods and generalized additive model, to predict outcomes more accurately (Van der Laan et al. (2007), Polley and Van Der Laan (2010)). In the super learner algorithm, we consider regression tree (Breiman et al. (1984), Quinlan (1986), Loh (2014)), bagging tree (Breiman (1998)), random forest (Breiman (2001)), gradient boosting model (GBM) (Breiman (1997), Friedman (2001), Mason et al. (2000)), and generalized additive model (GAM) (Hastie and Tibshirani (1990)). The basic idea is that we first obtain the predictive values based on each single algorithm, and then we obtain the super learner predicted values as a linear combination of the predicted values from different single algorithms. The predicted outcome values from super learner are plugged into different doubly robust methods to obtain the ATE estimates, which are referred as the super learner doubly robust estimates. In this subsection, we provide details for the proposed super learner doubly robust estimates.

We first introduce the single algorithms in the super learner algorithm. Regression tree constructs a large tree on the full data using the rpart function in the rpart package in R, then the large tree is trimmed down to a smaller tree by minimizing the cost-complexity score which is equals to sum of squared residual plus αT (Terry et al. (2015)). In the rpart function, we can get the complexity parameter table. Under each tree size (T), the table shows the best complexity parameter (α) , the corresponding cross validated error (xerror) and the standard deviation of the error (xstd). As the complexity parameter increases, the tree size decreases. We prefer to choose a smaller tree but keep the the xerror small, that is, we choose the maximum

complexity parameter such that xerror is smaller than the sum of the minimized xerror and the corresponding xstd. We prune the tree based on this optimal complexity parameter value.

The bagging tree and random forest can be implemented using the randomForest function in the randomForest package in R (Leo et al. (2018)). We consider the mtry (i.e., the optimal number of variables randomly sampled as the candidates at each split) as a tuning parameter. Mtry is chosen such that the mean square errors (MSE) is minimized. When we set mtry as the number of all covariates, we implement the bagging algorithm to obtain predicted values.

The gradient boosting model (GBM) is constructed using the gbm package in R (Brandon et al. (2013)). First, we set several candidate values for those parameters: (1) shrinkage (i.e., learning rate), (2) interaction.depth (i.e., maximum number of splits in each tree), (3) n.minobsinnode (i.e., minimum number of observations in the terminal nodes of trees), and (4) bag.fraction (i.e., the training set observations randomly selected to generate next tree in the expansion). We set n.trees=6000, that is, we grow 6000 trees (maximum number of iterations) to fit the model. Under each parameter setting, we get a minimum RMSE with corresponding optimal n.trees. Finally, we choose the set of parameters such that the RMSE is minimized.

GAM uses smoothing splines and additive model to obtain the outcome model by minimizing the (weighted) penalized sum of square (Hastie and Tibshirani (1990)):

$$\sum_{i=1}^{N} w_i(y_i - T_i m_1(x_i) - (1 - T_i)m_0(x_i))^2 + \lambda \int \left((m_0''(x))^2 + (m_1''(x))^2\right) dx.$$

GAM can be implemented using the gam function in the mgcv R-package (Wood (2015)).

We propose a revised super learner algorithm once we get all the predicted outcomes from different algorithms mentioned above. Let us denote the predicted

potential outcomes as $\{\hat{m}_{0,k}, \hat{m}_{1,k}\}$ $(k = 1, \ldots, K)$ from K different algorithms, e.g., $K=5$ for regression tree, bagging tree, random forest, GBM, and GAM. Here $\hat{m}_{1,k}$ and $\hat{m}_{0,k}$ denote the predicted potential outcomes for the subject being treated and being untreated, respectively, based on the kth single algorithm. For each algorithm, we can obtain the predicted outcome $\hat{m}_k = T_i \hat{m}_{1,k} + (1 - T_i)\hat{m}_{0,k}$ for $k = 1, \ldots, K$. The super learner estimated outcome is obtained as the linear combination of the predicted outcomes from the K algorithms, say, $\hat{m}_{SL} = \sum_{k=1}^{K} \alpha_k \hat{m}_k$ subject to the constrains that all weights $\alpha_k \geq 0$ $(k = 1, \ldots, K)$, and $\sum_{k=1}^{K} \alpha_k = 1$. The optimal weights $\hat{\alpha}_k$ $(k = 1, \ldots, K)$ are estimated by minimizing the loss function:

$$\min \sum_{i=1}^{N}(Y_i - \hat{m}_{SL,i})^2.$$

The super learner doubly robust methods for Lunceford's and Kang's residual version are simply obtained from replacing the predicted potential outcome m_0 and m_1 by the super learner predicted outcomes. As to the Kang's WLS doubly robust method, the outcome model for each single algorithm is obtained by using the weighted machine learning methods and weighted GAM. The potential outcomes of the super learner is also obtained by the optimization strategy based on the single algorithm's predicted values. Here the weights are the IPTW of each subject in the sample. As to Bang's version of super learner doubly robust method, each single algorithm (machine learning methods and GAM) takes the covariate X and T along with the IPTW as input to get the predicted potential outcomes. Then the two predicted potential outcomes e_0 and e_1 obtained by super learner algorithm are plugged into equation (4.11) to get the ATE estimates based on Bang's method. In the following section, we carried out simulation studies to examine the performances of these proposed methods.

4.3 Simulation studies

In this section, simulation studies were carried out to compare the performance of different doubly robust methods. We generated data (X, T, Y) from a complex propensity score model and an outcome regression model using the following simulation settings.

Step 1: Generated the covariates X_{c1}, X_{c2}, and X_p from the normal distribution N(0,1), N(0,1), and N(3,3), respectively.

Step 2: Generated the treatment T from the Bernoulli distribution with probability obtained from

$$logit[P(T = 1|X)] = 0.5 + 0.5X_{c1} - 0.3X_{c2} + 0.5X_{c1}^2 - 0.3X_{c2}^2 + 0.2X_{c1}X_{c2} \quad (4.12)$$

Step 3: Generated outcome Y from the normal distribution:

$$
\begin{aligned}
Y &= 0.06 + T + 0.1X_{c1} + 0.2X_{c2} + 0.15X_p - 0.3X_{c1}^2 + 0.3X_{c2}^2 \\
&\quad - 0.2X_p^2 + 0.08X_{c1}X_{c2} + 0.24X_{c1}X_p - 0.1X_{c2}X_p + N(0,1). \quad (4.13)
\end{aligned}
$$

4.3.1 Simulation studies for doubly robust methods

To examine the performance of the four doubly robust methods presented in Section 4.2.1, we compared the correctly specified model with the mis-specified model. Under the data generating mechanism, the correctly specified propensity score model was defined by equation (4.12):

$$logit[P(T = 1|X)] = \gamma_0 + \gamma_1 X_{c1} + \gamma_2 X_{c2} + \gamma_3 X_{c1}^2 + \gamma_4 X_{c2}^2 + \gamma_5 X_{c1}X_{c2}.$$

The mis-specified propensity score model was defined as

$$logit[P(T = 1|X)] = \gamma_0^* + \gamma_1^* X_{c1} + \gamma_2^* X_{c2}. \tag{4.14}$$

Similarly, the correctly specified outcome regression model was defined by equation (4.13):

$$Y = \mathbf{X}\beta + N(0, 1), \tag{4.15}$$

where $\mathbf{X} = \{\mathbf{1}, T, X_{c1}, X_{c2}, X_p, X_{c1}^2, X_{c2}^2, X_p^2, X_{c1}X_{c2}, X_{c1}X_p, X_{c2}X_p\}$ and β is the vector of parameters. The mis-specified outcome regression model was defined as

$$E(Y|T, X) = \beta_0^* + \beta_1^* T + \beta_2^* X_{c1} + \beta_3^* X_{c2} + \beta_4^* X_p. \tag{4.16}$$

For each doubly robust method, we generated data from the underlying propensity score model (4.12) and outcome regression model (4.13), then estimated the propensity score with the true propensity score model (4.12) as well as the mis-specified propensity score model (4.14). We also used the underlying outcome model (4.13) and mis-specified outcome model (4.16), respectively. Thus, we obtained four doubly robust estimators based on four combinations of the the propensity score model and outcome model used: (i) both propensity score and outcome models were correctly specified (True OR, True PS); (ii) outcome model was correctly specified, but propensity score model was mis-specified (True OR, False PS); (iii) outcome model was not correctly specified while propensity score model was correctly specified (False OR, True PS); (iv) neither outcome model nor propensity score model were correctly specified (False OR, False PS).

1000 simulated data sets were generated from Steps 1-3. In each simulation, we obtained an estimated ATE and a bootstrap standard error which was the stan-

dard deviation of the estimated ATEs from 100 re-sampling bootstrap samples. The simulation results are presented in Table 4.1. The estimated ATE and empirical standard error (Std.error) were the mean and standard deviation of the 1000 estimated ATEs. The Bootstrap Std.Error was the mean of the 1000 Bootstrap standard errors. The coverage rate was the percentage of true ATE falling in the 1000 95% confidence intervals of the estimated ATEs. The bias was the absolute difference between the true ATE and the mean of the estimated ATEs.

Table 4.1: The simulation results of the four different doubly robust methods under four different model specifications, where the sample size=1000 and the true ATE is zero.

Model Specifications	Estimated ATE	Bias	Empirial Std.Error	Bootstrap Std.Error	Coverage Rate
Lunceford's method					
True OR, True PS	0.000	0.000	0.000	0.000	99.2%
True OR, False PS	0.000	0.000	0.000	0.000	98.8%
False OR, True PS	-0.005	0.005	0.277	0.225	95.1%
False OR, False PS	-0.351	0.351	0.180	0.185	50.8%
Kang's residual method					
True OR, True PS	0.000	0.000	0.000	0.000	98.8%
True OR, False PS	0.000	0.000	0.000	0.000	99.1%
False OR, True PS	-0.008	0.008	0.250	0.215	94.9%
False OR, False PS	-0.351	0.351	0.180	0.184	50.8%
Kang's WLS method					
True OR, True PS	0.000	0.000	0.000	0.000	97.4%
True OR, False PS	0.000	0.000	0.000	0.000	97.0%
False OR, True PS	-0.021	0.021	0.217	0.202	94.5%
False OR, False PS	-0.352	0.352	0.179	0.183	50.2%
Bang's method					
True OR, True PS	0.000	0.000	0.000	0.000	99.0%
True OR, False PS	0.000	0.000	0.000	0.000	97.8%
False OR, True PS	1.653	1.653	35.699	39.610	97.1%
False OR, False PS	-0.323	0.323	0.178	0.183	56.7%

Note: OR indicates outcome model, and PS indicates propensity score model.

Based on Table 4.1, all doubly robust methods except Bang's method performed well when either outcome model or propensity score model was correctly specified, that is, the estimated ATE were unbiased with coverage rate close to 95%.

However, if both the outcome model and propensity score model were mis-specified, all doubly robust methods resulted in biased ATE estimates with large variances and low coverage rates. The Bang's DR method worked well when the outcome model was correctly specified. However, when outcome model was mis-specified, the estimated ATEs were biased regardless of the propensity score model used. This is probably because Bang's method assumes an structural outcome model. In the following subsection, we examined whether the super learner doubly robust methods can alleviate the situations when neither propensity score model nor outcome model are specified correctly.

4.3.2 Super learner doubly robust methods

In this subsection, we obtained the predicted potential outcomes by using machine learning methods, GAM, and super learner algorithm. We ran 1000 simulations and the simulation results are presented in Table 4.2 under two scenarios: (i) propensity score models was correctly specified; and (ii) propensity score model was mis-specified.

Based on the simulation result in Table 4.2, the best ATE estimates for Lunceford method and Kang's residual method were from the GBM, with the coverage rate close to 95%, no matter the propensity score model was correctly specified or not. The second best ATE estimates were from the super learner algorithm, which usually performed near the best. For Kang's WLS method, the top 2 best single outcome prediction models were the regression tree model and GBM with the coverage rate larger than 95%, no matter the propensity score model was correctly specified or not. The performances of the super learner doubly robust method performed between the regression tree model and GBM. For Bang's DR method, only the regression tree assisted Bang's DR method worked well (Table 4.2). The super learner algorithm, GAM, and most of the machine learning did not contribute to the accuracy of es-

Table 4.2: The simulation results when super learner doubly robust methods were used. The true ATE was zero.

Outcome Models	DR Method	PS Models	Estimated ATE	Bias	Empirical Std.Error	Boostrap Std.Error	Coverage Rate
Regression Tree	Lunceford method	True PS	0.026	0.026	0.233	0.195	95.6%
		False PS	0.220	0.220	0.098	0.101	39.6%
	Kang's residual method	True PS	0.037	0.037	0.189	0.158	95.1%
		False PS	0.220	0.220	0.098	0.101	39.6%
	Kang's WLS method	True PS	0.000	0.000	0.000	0.001	100.0%
		False PS	0.000	0.000	0.000	0.001	100.0%
	Bang's method	True PS	0.000	0.000	0.004	0.006	99.9%
		False PS	0.000	0.000	0.009	0.019	99.3%
Bagging Tree	Lunceford method	True PS	0.013	0.013	0.051	0.034	82.7%
		False PS	0.053	0.053	0.021	0.016	12.1%
	Kang's residual method	True PS	0.015	0.015	0.045	0.027	82.5%
		False PS	0.053	0.053	0.021	0.016	12.1%
	Kang's WLS method	True PS	0.013	0.013	0.006	0.009	82.9%
		False PS	0.013	0.013	0.006	0.009	82.9%
	Bang's method	True PS	0.116	0.116	0.031	0.023	3.4%
		False PS	0.030	0.030	0.041	0.038	84.1%
Random Forest	Lunceford method	True PS	0.017	0.017	0.052	0.033	79.8%
		False PS	0.055	0.055	0.023	0.017	12.3%
	Kang's residual method	True PS	0.019	0.019	0.046	0.026	79.4%
		False PS	0.055	0.055	0.023	0.017	12.3%
	Kang's WLS method	True PS	0.020	0.020	0.008	0.010	48.3%
		False PS	0.020	0.020	0.008	0.010	50.5%
	Bang's method	True PS	0.132	0.132	0.038	0.030	5.4%
		False PS	0.045	0.045	0.045	0.041	76.1%
GBM	Lunceford method	True PS	0.001	0.001	0.018	0.018	97.5%
		False PS	0.001	0.001	0.018	0.018	97.4%
	Kang's residual method	True PS	0.001	0.001	0.018	0.018	97.5%
		False PS	0.001	0.001	0.018	0.018	97.4%
	Kang's WLS method	True PS	0.004	0.004	0.017	0.016	96.1%
		False PS	0.000	0.000	0.017	0.016	95.8%
	Bang's method	True PS	0.057	0.057	0.036	0.032	54.1%
		False PS	0.003	0.003	0.030	0.029	97.3%
GAM	Lunceford method	True PS	0.000	0.000	0.115	0.095	95.7%
		False PS	0.001	0.001	0.056	0.055	95.2%
	Kang's residual method	True PS	0.001	0.001	0.101	0.080	95.1%
		False PS	0.001	0.001	0.056	0.055	95.2%
	Kang's WLS method	True PS	0.000	0.000	0.071	0.061	91.9%
		False PS	0.002	0.002	0.057	0.056	93.5%
	Bang's method	True PS	9.4E+09	9.4E+09	2.9E+11	4.3E+19	100.0%
		False PS	0.002	0.002	0.076	0.104	98.3%
SuperLearner	Lunceford method	True PS	0.001	0.001	0.024	0.019	96.6%
		False PS	0.003	0.003	0.019	0.018	95.3%
	Kang's residual method	True PS	0.001	0.001	0.021	0.018	96.5%
		False PS	0.003	0.003	0.019	0.018	95.3%
	Kang's WLS method	True PS	0.005	0.005	0.017	0.016	96.3%
		False PS	0.001	0.001	0.017	0.016	96.2%
	Bang's method	True PS	9.9E+06	9.9E+06	3.2E+08	1.4E+16	89.2%
		False PS	0.004	0.004	0.031	0.029	96.6%

Note: PS represents propensity score model.

timated ATEs. Including the IPTW as covariate in a model, we cannot take full advantages of the propensity score's feature. That is, we cannot balance the distribution of the covariates in treatment group and control group to reduce the selection bias in observational data.

4.4 Case study

NHEFS is a dataset containing data from the National Health and Nutrition Examination Survey Data I Epidemiologic Follow-up Study (NHEFS). The NHEFS was jointly initiated by the National Center for Health Statistics and the National Institute on Aging in collaboration with other agencies of the United States Public Health Service. A detailed description of the NHEFS, together with publicly available data sets and documentation, can be found at wwwn.cdc.gov/nchs/nhanes/nhefs/. Our case study is a subset of NHEFS data, which is composed of 1629 cigarette smokers aged from 25 to 74 years old who had a baseline visit in 1971 and a follow-up visit in 1982.

The outcome is the weight change in kilograms from 1971 to 1982, where 1566 individuals had their weight measured at the follow-up visit. The treatment is a binary variable which indicates if a subject quitted smoking between 1971 and 1982. The baseline covariates includes age, sex, education, physical activity, total family income, amount of alcohol drink, and serum cholesterol (mg/100ml). We included all subjects who had reading values for all these covariates, quitting smoking or not, and the weight changes, which resulted in 1110 subjects in our study cohort.

We applied the super learner doubly robust methods to estimate the weight gain due to quitting smoking, and the results are presented in Table 4.3. From Table 4.3, quitting smoking was more likely to cause weight gain. The super learner doubly robust methods showed that the weight gains due to quitting smoking ranged from 2.80kg to 2.95kg from 1971 to 1982, which were significantly different from zero.

Table 4.3: The estimated average weight gained due to quit smoking based on super learner doubly robust methods and the NHEFS dataset.

Outcome Models	Estimated ATE	95% CI
Lunceford method		
Regression Tree	2.568	(1.524, 3.612)
Bagging Tree	3.137	(2.080, 4.194)
Random Forest	2.704	(1.763, 3.644)
GBM	2.985	(1.899, 4.072)
GAM	3.007	(1.891, 4.123)
SuperLearner	2.850	(1.804, 3.896)
Kang's residual method		
Regression Tree	2.918	(1.783, 4.053)
Bagging Tree	3.128	(2.074, 4.183)
Random Forest	2.866	(1.882, 3.850)
GBM	2.970	(1.871, 4.069)
GAM	2.970	(1.844, 4.096)
SuperLearner	2.954	(1.909, 3.998)
Kang's WLS method		
Regression Tree	0.000	(-2.206, 2.206)
Bagging Tree	3.126	(1.943, 4.31)
Random Forest	2.687	(1.732, 3.642)
GBM	2.868	(1.799, 3.938)
GAM	2.979	(1.895, 4.064)
SuperLearner	2.803	(1.609, 3.996)

4.5   Discussion

In the project, we compared different doubly robust methods, and proposed super learner doubly robust methods. The simulation studies showed that super learner doubly robust methods could provide unbiased ATE estimates regardless the propensity score models are correctly specified or not. The super learner algorithm, which combines various machine learning methods and GAM, performs nearly as well as the individual best model. In addition, the weights in super learner algorithm can provide information on the best outcome prediction model, which is often the single algorithm with the largest weight.

REFERENCES

Younathan Abdia, KB Kulasekera, Somnath Datta, Maxwell Boakye, and Maiying Kong. Propensity scores based methods for estimating average treatment effect and average treatment effect among treated: A comparative study. *Biometrical Journal*, 59(5):967–985, 2017.

Milton Abramowitz, Irene A Stegun, and Robert H Romer. Handbook of mathematical functions with formulas, graphs, and mathematical tables, 1988.

Mark A Abramson, Charles Audet, G Couture, John E Dennis Jr, Sébastien Le Digabel, and C Tribes. The nomad project, 2011.

Joachim H. Ahrens and Ulrich Dieter. Generating gamma variates by a modified rejection technique. *Communications of the ACM*, 25(1):47–54, 1982.

James H Albert and Siddhartha Chib. Bayesian analysis of binary and polychotomous response data. *Journal of the American statistical Association*, 88(422):669–679, 1993.

Horst Alzer. On some inequalities for the gamma and psi functions. *Mathematics of Computation*, 66(217):373–389, 1997.

Horst Alzer. On ramanujan's double inequality for the gamma function. *Bulletin of the London Mathematical Society*, 35(5):601–607, 2003.

GD Anderson, RW Barnard, KC Richards, MK Vamanamurthy, and M Vuorinen.

Inequalities for zero-balanced hypergeometric functions. *Transactions of the American Mathematical Society*, pages 1713–1723, 1995.

Charles Audet and John E Dennis Jr. Mesh adaptive direct search algorithms for constrained optimization. *SIAM Journal on optimization*, 17(1):188–217, 2006.

Charles Audet and Warren Hare. *Derivative-Free and Blackbox Optimization*. Springer Series in Operations Research and Financial Engineering. Springer International Publishing, 2017. doi: 10.1007/978-3-319-68913-5.

Peter C Austin. The performance of different propensity score methods for estimating marginal odds ratios. *Statistics in Medicine*, 26(16):3078–3094, 2007.

N Balakrishnan and N Johnson. Samuel kotz. continuous univariate distributions. vol. 1, 1994.

Heejung Bang and James M Robins. Doubly robust estimation in missing data and causal inference models. *Biometrics*, 61(4):962–973, 2005.

Necdet Batir. Some new inequalities for gamma and polygamma functions. *Journal of Inequalities in Pure and Applied Mathematics*, 6(4):1–9, 2005.

Necdet Batir. Inequalities for the gamma function. *Archiv der Mathematik*, 91(6): 554–563, 2008.

Greenwell Brandon, Boehmke Bradley, and Cunningham Jay. Package âgbmâ. *Vitattu*, 10(2013):40, 2013.

Leo Breiman. Arcing the edge. *Technical Report, University of California, Berkeley, CA*, 1997.

Leo Breiman. Arcing classifier (with discussion and a rejoinder by the author). *The Annals of Statistics*, 26(3):801–849, 1998.

Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

Leo Breiman, Jerome Friedman, Charles J Stone, and Richard A Olshen. *Classification and Regression Trees*. CRC Press, 1984.

Alan M Brookhart, Sebastian Schneeweiss, Kenneth J Rothman, Robert J Glynn, Jerry Avorn, and Til Stürmer. Variable selection for propensity score models. *American journal of epidemiology*, 163(12):1149–1156, 2006.

George Casella. An introduction to empirical bayes data analysis. *The American Statistician*, 39(2):83–87, 1985.

George Casella and Roger L Berger. *Statistical Inference*, volume 2. Duxbury Pacific Grove, CA, 2002.

George Casella, Christian P Robert, Martin T Wells, et al. Generalized accept-reject sampling schemes. In *A Festschrift for Herman Rubin*, pages 342–347. Institute of Mathematical Statistics, 2004.

Saptarshi Chakraborty, Kshitij Khare, et al. Convergence properties of gibbs samplers for bayesian probit regression with proper priors. *Electronic Journal of Statistics*, 11(1):177–210, 2017.

Thomas Craven and George Csordas. The fox–wright functions and laguerre multiplier sequences. *Journal of Mathematical Analysis and Applications*, 314(1):109–125, 2006.

John A Craycroft, Jiapeng Huang, and Maiying Kong. Propensity score specification for optimal estimation of average treatment effect with binary response. *Statistical Methods in Medical Research*, 2020.

Paul Damien and Stephen G Walker. Sampling truncated normal, beta, and gamma densities. *Journal of Computational and Graphical Statistics*, 10(2):206–215, 2001.

Luc Devroye. Nonuniform random variate generation. *Handbooks in Operations Research and Management Science*, 13:83–121, 2006.

Luc Devroye. Random variate generation for the generalized inverse gaussian distribution. *Statistics and Computing*, 24(2):239–246, 2014.

Stefan Evert and Marco Baroni. *zipfR*: Word frequency distributions in R. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, Posters and Demonstrations Sessions*, pages 29–32, Prague, Czech Republic, 2007. (R package version 0.6-66 of 2019-09-30).

Katherine L Fedor. Noninvasive respiratory support in infants and children. *Respiratory care*, 62(6):699–717, 2017.

Michael Finegold, Mathias Drton, et al. Robust bayesian graphical modeling using dirichlet *t*-distributions. *Bayesian Analysis*, 9(3):521–550, 2014.

C. Fox. The asymptotic expansion of generalized hypergeometric functions. *Proceedings of the London Mathematical Society*, s2-27(1):389–400, 1928. doi: 10.1112/plms/s2-27.1.389. URL https://londmathsoc.onlinelibrary.wiley.com/doi/abs/10.1112/plms/s2-27.1.389.

Jessica M Franklin, Wesley Eddings, Robert J Glynn, and Sebastian Schneeweiss. Regularized regression versus the high-dimensional propensity score for confounding adjustment in secondary database analyses. *American Journal of Epidemiology*, 182(7):651–659, 2015.

Jerome H Friedman. Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 1189–1232, 2001.

Michele Jonsson Funk, Daniel Westreich, Chris Wiesen, Til Stürmer, M Alan

Brookhart, and Marie Davidian. Doubly robust estimation of causal effects. *American Journal of Epidemiology*, 173(7):761–767, 2011.

Walter R Gilks and Pascal Wild. Adaptive rejection sampling for gibbs sampling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 41(2):337–348, 1992.

Trevor J Hastie and Robert J Tibshirani. *Generalized Additive Models*, volume 43. CRC Press, 1990.

Miguel A Hernán and James M Robins. *Causal Inference: What If.* 2019.

Daniel Hernandez-Stumpfhauser, F Jay Breidt, Mark J van der Woerd, et al. The general projected normal distribution of arbitrary dimension: modeling and bayesian inference. *Bayesian Analysis*, 12(1):113–133, 2017.

Wolfgang Hörmann and Josef Leydold. Generating generalized inverse gaussian random variates. *Statistics and Computing*, 24(4):547–557, 2014.

Michael A. Jenkins and Joseph F Traub. Algorithm 419: zeros of a complex polynomial [c2]. *Communications of the ACM*, 15(2):97–99, 1972.

Jana Jurečková and Pranab Kumar Sen. Effect of the initial estimator on the asymptotic behavior of one-step m-estimator. *Annals of the Institute of Statistical Mathematics*, 42(2):345–357, Jun 1990. ISSN 1572-9052. doi: 10.1007/BF00050841. URL https://doi.org/10.1007/BF00050841.

Joseph DY Kang, Joseph L Schafer, et al. Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science*, 22(4):523–539, 2007.

Donat K Kazarinoff. On wallis' formula. *Edinburgh Mathematical Notes*, 40:19–21, 1956.

Gi Beom Kim. Pulmonary hypertension in infants with bronchopulmonary dysplasia. *Korean Journal of Pediatrics*, 53(6):688, 2010.

Steffen L Lauritzen. *Graphical Models*, volume 17. Clarendon Press, 1996.

Sébastien Le Digabel. *NOMAD: Nonlinear optimization with the MADS algorithm*. Groupe d'études et de recherche en analyse des décisions, 2010.

Breiman Leo, Cutler Adele, Liaw Andy, and Wiener Matthew. Package ârandomforestâ. *University of California, Berkeley: Berkeley, CA, USA*, 2018.

Chuanhai Liu, Ryan Martin, and Nick Syring. Efficient simulation from a gamma distribution with small shape parameter. *Computational Statistics*, 32(4):1767–1775, 2017.

Jun S Liu and Ying Nian Wu. Parameter expansion for data augmentation. *Journal of the American Statistical Association*, 94(448):1264–1274, 1999.

Wei-Yin Loh. Fifty years of classification and regression trees. *International Statistical Review*, 82(3):329–348, 2014.

Jared K Lunceford and Marie Davidian. Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. *Statistics in Medicine*, 23(19):2937–2960, 2004.

Francesco Mainardi and Gianni Pagnini. The role of the fox–wright functions in fractional sub-diffusion of distributed order. *Journal of Computational and Applied Mathematics*, 207(2):245–257, 2007.

Dobrin Marchev and James P Hobert. Geometric ergodicity of van dyk and meng's algorithm for the multivariate student's t model. *Journal of the American Statistical Association*, 99(465):228–238, 2004.

Luca Martino and Joaquín Míguez. A generalization of the adaptive rejection sampling algorithm. *Statistics and Computing*, 21(4):633–647, 2011.

Llew Mason, Jonathan Baxter, Peter L Bartlett, and Marcus R Frean. Boosting algorithms as gradient descent. In *Advances in Neural Information Processing Systems*, pages 512–518, 2000.

Khaled Mehrez and Sergei M. Sitnik. Functional inequalities for the fox–wright functions. *The Ramanujan Journal*, 50(2):263–287, Nov 2019. ISSN 1572-9303. doi: 10.1007/s11139-018-0071-2. URL https://doi.org/10.1007/s11139-018-0071-2.

Xiao-Li Meng and David A Van Dyk. Seeking efficient data augmentation schemes via conditional and marginal augmentation. *Biometrika*, 86(2):301–320, 1999.

Zhenghua Nie and Jeffrey S Racine. The crs package: Nonparametric regression splines for continuous and categorical predictors. *R Journal*, 4(2), 2012.

William H Northway, Robert C Rosan, and David Y Porter. Pulmonary disease following respirator therapy of hyaline-membrane disease: bronchopulmonary dysplasia. *New England Journal of Medicine*, 276(7):357–368, 1967.

Subhadip Pal, Kshitij Khare, and James P Hobert. Improving the data augmentation algorithm in the two-block setup. *Journal of Computational and Graphical Statistics*, 24(4):1114–1133, 2015.

Judea Pearl. *Models, Reasoning and Inference*. Cambridge, UK: Cambridge University Press, 2000.

Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Elsevier, 2014.

Judea Pearl, Madelyn Glymour, and Nicholas P Jewell. *Causal Inference in Statistics: A primer*. John Wiley & Sons, 2016.

Kevin D Pereira, Allison R MacGregor, Chad M McDuffie, and Ron B Mitchell. Tracheostomy in preterm infants: current trends. *Archives of Otolaryngology–Head & Neck Surgery*, 129(12):1268–1271, 2003.

Anne Philippe. Simulation of right and left truncated gamma distributions by mixtures. *Statistics and Computing*, 7(3):173–181, 1997.

Eric C Polley and Mark J Van Der Laan. Super learner in prediction. *U.C. Berkeley Division of Biostatistics Working Paper Series*, 2010.

Feng Qi and Cristinel Mortici. Some inequalities for the trigamma function in terms of the digamma function. *Applied Mathematics and Computation*, 271:502–511, 2015.

Feng Qi, Qiu-Ming Luo, et al. Bounds for the ratio of two gamma functions—from wendel's and related inequalities to logarithmically completely monotonic functions. *Banach Journal of Mathematical Analysis*, 6(2):132–158, 2012.

J. Ross Quinlan. Induction of decision trees. *Machine Learning*, 1(1):81–106, 1986.

Thomas S Richardson and James M Robins. Single world intervention graphs (swigs): A unification of the counterfactual and graphical approaches to causality. *Center for the Statistics and the Social Sciences, University of Washington Series. Working Paper*, 128(30), 2013.

Christian P Robert. Simulation of truncated normal variables. *Statistics and Computing*, 5(2):121–125, 1995.

James M Robins, Miguel Angel Hernan, and Babette Brumback. Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11(5):550–560, 2000.

Paul R Rosenbaum and Donald B Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

Alberto Roverato. Hyper inverse wishart distribution for non-decomposable graphs and its application to bayesian inference for gaussian graphical models. *Scandinavian Journal of Statistics*, 29(3):391–411, 2002.

Vivekananda Roy et al. Convergence rates for mcmc algorithms for a robust bayesian binary regression model. *Electronic Journal of Statistics*, 6:2463–2485, 2012.

HM Srivastava. Some fox-wright generalized hypergeometric functions and associated families of convolution operators. *Applicable Analysis and Discrete Mathematics*, pages 56–71, 2007.

J. Michael Steele. *The Cauchy-Schwarz Master Class: An Introduction to the Art of Mathematical Inequalities.* Cambridge University Press, New York, NY, USA, 2004. ISBN 052154677X.

Therneau Terry, Atkinson Beth, and Brian Ripley. Package ârpartâ. *Available online: cran. ma. ic. ac. uk/web/packages/rpart/rpart. pdf (accessed on 20 April 2016)*, 2015.

Gentry W Thatcher and Robert H Maisel. The long-term evaluation of tracheostomy in the management of severe obstructive sleep apnea. *The Laryngoscope*, 113(2): 201–204, 2003.

Mark J Van der Laan, Eric C Polley, and Alan E Hubbard. Super learner. *Statistical Applications in Genetics and Molecular Biology*, 6(1), 2007.

David A Van Dyk and Xiao-Li Meng. The art of data augmentation. *Journal of Computational and Graphical Statistics*, 10(1):1–50, 2001.

Tyler J VanderWeele. Marginal structural models for the estimation of direct and indirect effects. *Epidemiology*, 20(1):18–26, 2009.

Bill Venables, Kurt Hornik, and Martin Maechler. *polynom: A Collection of Functions to Implement a Class for Univariate Polynomial Manipulations*, 2019. URL https://CRAN.R-project.org/package=polynom. R package version 1.4-0. S original by Bill Venables, packages for R by Kurt Hornik and Martin Maechler.

Xiaoxia Wang. Contiguous relations for the fox–wright function. *Journal of Mathematical Analysis and Applications*, 475(1):203–214, 2019.

G. N. Watson. A note on gamma functions. *Edinburgh Mathematical Notes*, 42: 7, nov 1959. doi: 10.1017/s0950184300003207. URL https://doi.org/10.1017%2Fs0950184300003207.

Karen F Watters. Tracheostomy in infants and children. *Respiratory Care*, 62(6): 799–825, 2017.

Chuanan Wei. Strange evaluations of fox–wright function. *Integral Transforms and Special Functions*, 30(1):6–27, 2019.

Simon Wood. Package âmgcvâ. *R package version*, 1:29, 2015.

E. Maitland Wright. The asymptotic expansion of the generalized hypergeometric function. *Journal of the London Mathematical Society*, s1-10(4):286–293, 1935. doi: 10.1112/jlms/s1-10.40.286. URL https://londmathsoc.onlinelibrary.wiley.com/doi/abs/10.1112/jlms/s1-10.40.286.

Xiaofang Yan, Younathan Abdia, Somnath Datta, KB Kulasekera, Beatrice Ugiliweneza, Maxwell Boakye, and Maiying Kong. Estimation of average treatment effects among multiple treatment groups by using an ensemble approach. *Statistics in Medicine*, 38(15):2828–2846, 2019.

Zhen-Hang Yang, Yu-Ming Chu, and Xiao-Jing Tao. A double inequality for the

trigamma function and its applications. In *Abstract and Applied Analysis*, volume 2014. Hindawi, 2014.

Tjalling J Ypma. Historical development of the newton–raphson method. *SIAM Review*, 37(4):531–551, 1995.

APPENDIX

APPENDIX A: SUPPLEMENTS OF THE MODIFIED-HALF-NORMAL DISTRIBUTION

5.0.1 Proof of Lemma 1

Proof of Lemma 1 (a)

In the case when $\alpha > 1$, from the standard recurrence formula of the Gamma function we obtain that $\Gamma(\frac{\alpha+1+i}{2}) = \left(\frac{\alpha-1+i}{2}\right)\Gamma(\frac{\alpha-1+i}{2})$ for all $i \geq 0$. As a result

$$
\begin{aligned}
{}_1\Psi_1\left[\begin{array}{c}(\frac{\alpha+1}{2},\frac{1}{2}),\\ (1,0)\end{array};x\right] &= \sum_{i=0}^{\infty}\frac{x^i\Gamma(\frac{\alpha+1+i}{2})}{i!}\\
&= \sum_{i=0}^{\infty}\frac{x^i\left(\frac{\alpha-1+i}{2}\right)\Gamma(\frac{\alpha-1+i}{2})}{i!}\\
&= \left(\frac{\alpha-1}{2}\right)\sum_{i=0}^{\infty}\frac{x^i\Gamma(\frac{\alpha-1+i}{2})}{i!} + \frac{x}{2}\sum_{i=1}^{\infty}\frac{x^{i-1}\Gamma(\frac{\alpha+i-1}{2})}{(i-1)!}\\
&= \left(\frac{\alpha-1}{2}\right)\sum_{i=0}^{\infty}\frac{x^i\Gamma(\frac{\alpha-1+i}{2})}{i!} + \frac{x}{2}\sum_{i=0}^{\infty}\frac{x^i\Gamma(\frac{\alpha+i}{2})}{i!}\\
&= \left(\frac{\alpha-1}{2}\right){}_1\Psi_1\left[\begin{array}{c}(\frac{\alpha-1}{2},\frac{1}{2}),\\ (1,0)\end{array};x\right] + \frac{x}{2}{}_1\Psi_1\left[\begin{array}{c}(\frac{\alpha}{2},\frac{1}{2}),\\ (1,0)\end{array};x\right].
\end{aligned}
$$

Proof of Lemma 1 (b)

The continued fraction representation can be obtained via repeated use of the recursive relation established in part(a) the Lemma.

Proof of Lemma 1 (c)

From Lemma 5, we get that $_1\Psi_1 \begin{bmatrix} (\frac{\alpha}{2}, \frac{1}{2}), \\ ; \frac{\gamma}{\sqrt{\beta}} \\ (1,0) \end{bmatrix} = 2\beta^{\frac{\alpha}{2}} \int_0^\infty y^{\alpha-1} \exp\left(-\beta y^2 + \gamma y\right) dy.$

Therefore,

$$_1\Psi_1 \begin{bmatrix} (\frac{1}{2}, \frac{1}{2}), \\ ; x \\ (1,0) \end{bmatrix} = 2e^{\frac{x^2}{4}} \int_0^\infty e^{-(y-\frac{x}{2})^2} dy = 2\sqrt{\pi}e^{\frac{x^2}{4}} \int_{-\frac{x}{\sqrt{2}}}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = 2\sqrt{\pi}e^{\frac{x^2}{4}}[1 - \Phi(-\frac{x}{\sqrt{2}})]$$

and

$$\begin{aligned} _1\Psi_1 \begin{bmatrix} (\frac{2}{2}, \frac{1}{2}), \\ ; x \\ (1,0) \end{bmatrix} = 2e^{\frac{x^2}{4}} \int_0^\infty y e^{-(y-\frac{x}{2})^2} dy &= 2\sqrt{\pi}e^{\frac{x^2}{4}} \int_{-\frac{x}{\sqrt{2}}}^\infty (\frac{t}{2} + \frac{x}{2\sqrt{2}}) e^{-\frac{t^2}{2}} dt \\ &= 2\sqrt{\pi}e^{\frac{x^2}{4}} \left[-\frac{1}{2}e^{-\frac{t^2}{2}} \Big|_{-\frac{x}{\sqrt{2}}}^\infty + \frac{x}{2}\sqrt{\pi}(1 - \Phi(-\frac{x}{\sqrt{2}})) \right] \\ &= 1 + \sqrt{\pi}x e^{\frac{x^2}{4}}[1 - \Phi(-\frac{x}{\sqrt{2}})]. \end{aligned}$$

5.0.2 Proof of Lemma 2

Let $\alpha > 0, \in \mathbb{R}$ then for arbitrary $\epsilon > 0$, let $A(k) = \frac{\Gamma(\frac{\alpha}{2}+k)x^{2k}}{(2k)!}$ and $B(k) = \frac{\Gamma(\frac{\alpha+1}{2}+k)x^{2k+1}}{(2k+1)!}$. Let $\lceil x \rceil$ denotes the smallest integer larger than x.

Proof of part(a) of Lemma 2

If $0 < q < 1$ then

$$\frac{A(k+1)}{A(k)} \leq q \implies \frac{(\frac{\alpha}{2}+k)x^2}{(2k+2)(2k+1)} \leq q \implies 4qk^2 + (6q - x^2)k + (2q - \frac{\alpha}{2}x^2) \geq 0 \implies k \geq C_1,$$

where $k \geq C_1$, where $C_1 = \max\left\{ \left\lceil \frac{-(6q-x^2)+\sqrt{(6q-x^2)^2 - 8q(4q-\alpha x^2)}}{8q} \right\rceil, 1 \right\}.$

Proof of part(b) of Lemma 2

If $0 < q < 1$ then

$$\frac{B(k+1)}{B(k)} \leq q \implies \frac{(\frac{\alpha+1}{2}+k)x^2}{(2k+3)(2k+2)} \leq q \implies 4qk^2 + (10q - x^2)k + (6q - \frac{\alpha+1}{2}x^2) \geq 0$$

$$\implies k \geq C_2,$$

where $C_2 = \max\left\{ \left\lceil \frac{-(10q-x^2)+\sqrt{(10q-x^2)^2-8q(12q-(\alpha+1)x^2)}}{8q} \right\rceil, 1 \right\}$.

Proof of part(c) of Lemma 2

By Lemma 2 (a) and Lemma 2 (b), we get that the sequences $\{A(k)\}_{k \geq C_1}$ and $\{B(k)\}_{k \geq C_2}$ are strictly decreasing. Also, if $k > C_1$ then $A(k) \leq q^{k-C_1}A(C_1)$. Hence

$$\lim_{k \to \infty} A(k) \leq A(C_1) \lim_{k \to \infty} q^{k-C_1} = 0$$

as $0 < q < 1$ and $|A(C_1)| < \infty$. In a similar fashion we can show that $\lim_{k \to \infty}|B(k)| = 0$.

Proof of part(d) of Lemma 2

Given any $\epsilon > 0$, let $K_1 = \min\{k : A(k) \leq (1-q)\frac{\epsilon}{2} \text{ for all } k \geq C_1\}$ where $0 < q < 1$ is a fraction of our choice and C_1 is the constant as it is define in the previosu part. A possible value for q is $\frac{1}{2}$. The integer K_1 is well defined because $\lim_{k \to \infty} A(k) = 0$.

As $K_1 \geq C_1$, we get from the previous part of the Lemma that $A(n) \leq A(K_1)q^{n-K_1} \leq (1-q)\frac{\epsilon}{2}q^{n-K_1}$ for all $n \geq K_1$. As a result,

$$\left| \sum_{k=0}^{\infty} A(k) - \sum_{k=0}^{K_1} A(k) \right| = \left| \sum_{k=K_1}^{\infty} A(k) \right| \leq \sum_{k=K_1}^{\infty} \left((1-q)\frac{\epsilon}{2} \right) q^{k-K_1} = \frac{\epsilon}{2}.$$

102

In a similar fashion, it can be shown that $\left|\sum_{k=0}^{\infty} B(k) - \sum_{k=0}^{K_2} B(k)\right| \leq \frac{\epsilon}{2}$, when $K_2 = \min\{k : |B(k)| \leq (1-q)\frac{\epsilon}{2} \text{ for all } k \geq C_2\}$. Define $K = \max\{K_1, K_2\}$.

$$\left|{}_1\Psi_1\left[\begin{array}{c}(\frac{\alpha}{2}, \frac{1}{2}), \\ (1,0)\end{array}; x\right] - \left(\sum_{k=0}^{K} A(k) + \sum_{k=0}^{K} B(k)\right)\right| = \left|\sum_{k=0}^{\infty} A(k) - \sum_{k=0}^{K} A(k) + \left[\sum_{k=0}^{\infty} B(k) - \sum_{k=0}^{K} B(k)\right]\right|$$

$$\leq \left|\sum_{k=0}^{\infty} A(k) - \sum_{k=0}^{K} A(k)\right| + \left|\sum_{k=0}^{\infty} B(k) - \sum_{k=0}^{K} B(k)\right|$$

$$\leq \epsilon$$

$$(5.17)$$

for a given value for the error of the approximation ϵ and $K = \max\{K_1, K_2\}$ is a truncated point for the infinite series for even as well as odd order terms.

5.0.3   Proof of Lemma 3

From Lemma 5, we get that

$${}_1\Psi_1\left[\begin{array}{c}(\frac{\alpha}{2}, \frac{1}{2}), \\ (1,0)\end{array}; \frac{\gamma}{\sqrt{\beta}}\right] = 2\beta^{\frac{\alpha}{2}} \int_0^{\infty} x^{\alpha-1} \exp\left(-\beta x^2 + \gamma x\right) dx$$

If we would like to find a positive constant T, depending on α, β, γ such that

$$\left|{}_1\Psi_1\left[\begin{array}{c}(\frac{\alpha}{2}, \frac{1}{2}), \\ (1,0)\end{array}; \frac{\gamma}{\sqrt{\beta}}\right] - 2\beta^{\frac{\alpha}{2}} \int_0^{T} x^{\alpha-1} \exp\left(-\beta x^2 + \gamma x\right) dx\right| = 2\beta^{\frac{\alpha}{2}} \int_T^{\infty} x^{\alpha-1} \exp\left(-\beta x^2 + \gamma x\right) dx \leq \epsilon.$$

Using the inequality in part(a) of Theorem 3, we get

$$\int_T^\infty x^{\alpha-1} \exp\left(-\beta x^2 + \gamma x\right) dx \le \int_T^\infty x^{\alpha-1} \exp\left(-(\beta m^2 + |\gamma|m)(\frac{x}{m})^{\frac{2\beta m+|\gamma|}{\beta m+|\gamma|}}\right) dx,$$

It follows from the change of variable $t = \left(\frac{x}{m}\right)^{\frac{2\beta m+|\gamma|}{\beta m+|\gamma|}}$ that

$$
\begin{aligned}
\int_T^\infty x^{\alpha-1} \exp\left(-\beta x^2 + \gamma x\right) dx &\le& m^\alpha \frac{\beta m + |\gamma|}{2\beta m + |\gamma|} \int_T^\infty t^{\alpha-1} e^{-bt} dt \\
&=& m^\alpha \frac{\beta m + |\gamma|}{2\beta m + |\gamma|} \left[\int_0^\infty t^{\alpha-1} e^{-bt} - \int_0^T t^{\alpha-1} e^{-bt}\right] \\
&=& m^\alpha \frac{\beta m + |\gamma|}{2\beta m + |\gamma|} \left[\frac{\Gamma(a)}{b^a} - \frac{\gamma(a, bT)}{\Gamma(a)}\right]
\end{aligned}
$$

Let $2\beta^{\frac{\alpha}{2}} m^\alpha \frac{\beta m+|\gamma|}{2\beta m+|\gamma|} \left[\frac{\Gamma(a)}{b^a} - \frac{\gamma(a,bN)}{\Gamma(a)}\right] = \epsilon$, the truncated point of the numerical integral is

$$T = \frac{1}{b}\gamma^{-1}\left(a, \frac{[\Gamma(a)]^2}{b^a} - \frac{\epsilon\Gamma(a)}{2\beta^{\frac{\alpha}{2}} m^\alpha \frac{\beta m+|\gamma|}{2\beta m+|\gamma|}}\right),$$

where $a = \frac{\alpha(\beta m+|\gamma|)}{2\beta m+|\gamma|}$, $b = \beta m^2 + |\gamma|m)$, and $\gamma(a, bT)$ is a lower incomplete gamma function. The inverse of the lower incomplete gamma function can be calculated by the Igamma.inv() in the "zipfR" package in R.

5.0.4   Proof of Lemma 4

Proof of part(a) of Lemma 4

Consider the fact that

$$x = \frac{x}{\sqrt{m}}\sqrt{m} \le \frac{1}{2}[(\frac{x}{\sqrt{m}})^2 + (\sqrt{m})^2] = \frac{x^2}{2m} + \frac{m}{2}.$$

Therefore,

$$
{}_1\Psi_1 \left[\begin{array}{c} (\frac{\alpha}{2}, \frac{1}{2}), \\ (1,0) \end{array}; \frac{-|\gamma|}{\sqrt{\beta}} \right] = 2\beta^{\frac{\alpha}{2}} \int_0^\infty x^{\alpha-1} \exp\left(-\beta x^2 - |\gamma| x\right) dx
$$

$$
\geq 2\beta^{\frac{\alpha}{2}} \int_0^\infty x^{\alpha-1} \exp\left(-\beta x^2 - \frac{|\gamma| x^2}{2m} - \frac{|\gamma| m}{2}\right) dx
$$

$$
= \beta^{\frac{\alpha}{2}} \exp\left(-\frac{|\gamma| m}{2}\right) \int_0^\infty y^{\frac{\alpha}{2}-1} \exp\left(-(\beta + \frac{|\gamma|}{2m})y\right) dy
$$

Let $\alpha > 1, \beta > 0$ and $\gamma < 0$, then for any $m > 0$,

$$
{}_1\Psi_1 \left[\begin{array}{c} (\frac{\alpha}{2}, \frac{1}{2}), \\ (1,0) \end{array}; \frac{-|\gamma|}{\sqrt{\beta}} \right] \geq \frac{\beta^{\frac{\alpha}{2}} \exp\left(-\frac{m|\gamma|}{2}\right) \Gamma(\frac{\alpha}{2})}{\left(\beta + \frac{|\gamma|}{2m}\right)^{\frac{\alpha}{2}}}. \tag{5.18}
$$

The developed lower bound closely approximates the Fox-Wright function, especially for large values of α when the parameter m in Equation 5.18 chosen appropriately. In Figure 5.1, we plot the the log of the functional value and log of the lower bound choosing $m = m_\star$ where $m_\star$ satisfy the equation $2\beta m_\star^2 + |\gamma| m_\star = \alpha$. The red colored curve represent the log FoxWright function while the blue colored curve represents the log of lower bound in the Equation 5.18. In the plot we vary α in x axis while keeping the keeping the parameters γ and β to be fixed at the point -1, 1 respectively.
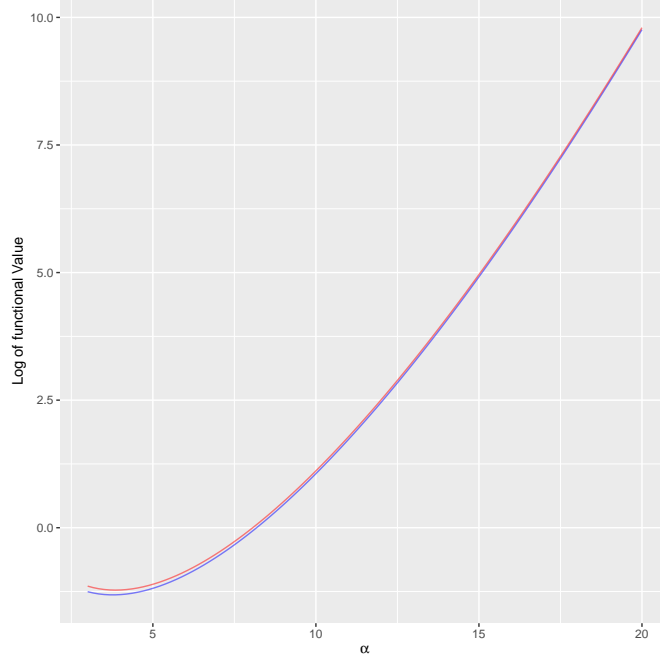
Figure 5.1: The red colored curve represent the log transformed Fox-Wright function while the blue colored curve depict the log of the lower bound that is shown in Equation 5.18 where $\gamma = -1$ and $\beta = 1$.

Proof of part(b) of Lemma 4

Let $p, q > 1$ be such that $\frac{1}{p} + \frac{1}{q} = 1$ then, using Holder's inequality we get that

$$
\int_0^\infty x^{\frac{\alpha + p\alpha_0 - 1}{p} - 1} e^{-\frac{x^2}{p}} dx
$$
$$
= \int_0^\infty \left[x^{\frac{\alpha - 1}{p}} e^{-\frac{x^2}{p} + \frac{\gamma x}{p}} \right] \left[x^{\alpha_0 - 1} e^{-\frac{\gamma x}{p}} \right] dx
$$
$$
\leq \left[\int_0^\infty x^{\alpha - 1} e^{-x^2 + \gamma x} dx \right]^{\frac{1}{p}} \left[\int_0^\infty x^{q(\alpha_0 - 1)} e^{-\frac{q\gamma x}{p}} dx \right]^{\frac{1}{q}}.
$$

106

Therefore, it follows that

$$
{}_1\Psi_1\left[\begin{array}{c}\left(\frac{\alpha}{2},\frac{1}{2}\right)\\ (1,0)\end{array};\gamma\right] = 2\int_0^\infty x^{\alpha-1}e^{-x^2+\gamma x}dx
$$

$$
\geq \frac{2\left[\int_0^\infty x^{\frac{\alpha+p\alpha_0-1}{p}-1}e^{-\frac{x^2}{p}}dx\right]^p}{\left[\int_0^\infty x^{(q\alpha_0-q+1)-1}e^{-\frac{q\gamma x}{p}}dx\right]^{\frac{p}{q}}}
$$

$$
= \frac{2\left[\frac{\Gamma\left(\frac{\alpha+p\alpha_0-1}{2p}\right)}{2}p^{\frac{\alpha+p\alpha_0-1}{2p}}\right]^p}{\left[\Gamma\left(q\alpha_0-q+1\right)\left(\frac{p}{q\gamma}\right)^{q\alpha_0-q+1}\right]^{\frac{p}{q}}}
$$

$$
= \frac{2p^{\frac{\alpha+p\alpha_0-1}{2}-\frac{p(q\alpha_0-q+1)}{q}}\left[\frac{\Gamma\left(\frac{\alpha+p\alpha_0-1}{2p}\right)}{2}\right]^p}{\left[\Gamma\left(q\alpha_0-q+1\right)\right]^{\frac{p}{q}}(q\gamma)^{-\frac{p(q\alpha_0-q+1)}{q}}}
$$

$$
= \frac{2p^{\frac{\alpha+(2-\alpha_0)p-1}{2}-\frac{p}{q}}\left[\frac{\Gamma\left(\frac{\alpha+p\alpha_0-1}{2p}\right)}{2}\right]^p(q\gamma)^{p(\alpha_0-1)+\frac{p}{q}}}{\left[\Gamma\left(q\alpha_0-q+1\right)\right]^{\frac{p}{q}}}. \tag{5.19}
$$

Our recommended values are $p = \frac{\alpha}{\alpha-1}, q = \alpha, \alpha_0 = \frac{\alpha+1}{p}$. Then the bound takes the following form

$$
{}_1\Psi_1\left[\begin{array}{c}\left(\frac{\alpha}{2},\frac{1}{2}\right)\\ (1,0)\end{array};\gamma\right] \geq \frac{[\Gamma(\alpha-1)]^{\frac{\alpha}{\alpha-1}}(\alpha\gamma)^\alpha}{[2\Gamma(\alpha(\alpha-1))]^{\frac{1}{\alpha-1}}}.
$$

In particular when $\gamma = \sqrt{\alpha}$ then

$$
{}_1\Psi_1\left[\begin{array}{c}\left(\frac{\alpha}{2},\frac{1}{2}\right)\\ (1,0)\end{array};\sqrt{\alpha}\right] \geq \frac{[\Gamma(\alpha-1)]^{\frac{\alpha}{\alpha-1}}\alpha^{\frac{3\alpha}{2}}}{[2\Gamma(\alpha(\alpha-1))]^{\frac{1}{\alpha-1}}}. \tag{5.20}
$$

The developed lower bound closely approximates the Fox-Wright function, especially for large values of α. We can get a sense of the approximation from the Plot 5.2, where the red colored curve represent the FoxWright function while the blue colored curve represents the lower bound in the Equation 5.20.
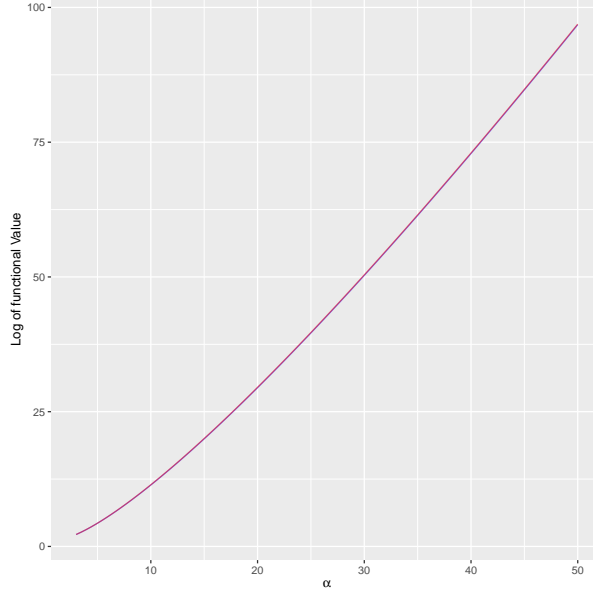
Figure 5.2: The red colored curve represent the log transformed FoxWright function while the blue colored curve depict the log of lower bound that is shown in Equation 5.20.

5.0.5 Proof of Lemma 5

Proof of Lemma 5 (a)

The density function of the MHN (α, β, γ), $f_{\text{MHN}}(x \mid \alpha, \beta, \gamma) \propto \cdot x^{\alpha-1} \exp(-\beta x^2 + \gamma x)\mathbb{I}(x > 0)$. The corresponding normalizing constant is

$$\int_0^\infty x^{\alpha-1} \exp(-\beta x^2 + \gamma x)dx = \int_0^\infty x^{\alpha-1} \exp(-\beta x^2) \sum_{i=0}^\infty \frac{\gamma^i x^i}{i!}dx.$$

Using a change of variable $t = x^2$, the above integral turns out to be

$$\sum_{i=0}^\infty \frac{\gamma^i \Gamma(\frac{\alpha+i}{2})}{2i!\beta^{\frac{\alpha+i}{2}}} \int_0^\infty \frac{\beta^{\frac{\alpha+i}{2}}}{\Gamma(\frac{\alpha+i}{2})} t^{\frac{\alpha+i}{2}-1} \exp(-\beta t)dt = \sum_{i=0}^\infty \frac{\gamma^i \Gamma(\frac{\alpha+i}{2})}{2i!\beta^{\frac{\alpha+i}{2}}} = \frac{2\beta^{\frac{\alpha}{2}}}{{}_1\Psi_1 \begin{bmatrix} (\frac{\alpha}{2}, \frac{1}{2}) \\ (1, 0) \end{bmatrix}; \frac{\gamma}{\sqrt{\beta}}},$$

where $_1\Psi_1 \left[\begin{array}{c}(\frac{\alpha}{2}, \frac{1}{2}) \\ (1,0)\end{array}; \frac{\gamma}{\sqrt{\beta}}\right] = \sum_{i=0}^{\infty} \frac{\Gamma(\frac{\alpha}{2}+\frac{i}{2})(\frac{\gamma}{\sqrt{\beta}})^i}{\Gamma(1)i!}$, which denotes the Fox-Wright function

of the appropriate order. Consequently the corresponding density function is given

as

$$f_{\text{MHN}}(x \mid \alpha, \beta, \gamma) = \frac{2\beta^{\frac{\alpha}{2}} x^{\alpha-1} \exp(-\beta x^2 + \gamma x)}{_1\Psi_1 \left[\begin{array}{c}(\frac{\alpha}{2}, \frac{1}{2}) \\ (1,0)\end{array}; \frac{\gamma}{\sqrt{\beta}}\right]} \mathbb{I}(x > 0).$$

Proof of Lemma 5 (b)

The cumulative distribution function $F(t)$ is given as

$$F_{\text{MHN}}(t \mid \alpha, \beta, \gamma) = \int_0^t \frac{2\beta^{\frac{\alpha}{2}}}{_1\Psi_1 \left[\begin{array}{c}(\frac{\alpha}{2}, \frac{1}{2}) \\ (1,0)\end{array}; \frac{\gamma}{\sqrt{\beta}}\right]} x^{\alpha-1} \exp(-\beta x^2 + \gamma x) dx$$

$$= \frac{2\beta^{\frac{\alpha}{2}}}{_1\Psi_1 \left[\begin{array}{c}(\frac{\alpha}{2}, \frac{1}{2}) \\ (1,0)\end{array}; \frac{\gamma}{\sqrt{\beta}}\right]} \int_0^t x^{\alpha-1} \exp(-\beta x^2) \sum_{i=0}^{\infty} \frac{\gamma^i x^i}{i!} dx$$

by considering th change of variable $y = \beta x^2$

$$= \frac{2\beta^{\frac{\alpha}{2}}}{_1\Psi_1 \left[\begin{array}{c}(\frac{\alpha}{2}, \frac{1}{2}) \\ (1,0)\end{array}; \frac{\gamma}{\sqrt{\beta}}\right]} \sum_{i=0}^{\infty} \frac{\gamma^i}{2i!} \beta^{-\frac{\alpha+i}{2}} \int_0^{\beta t^2} y^{\frac{\alpha+i}{2}-1} e^{-y} dy$$

$$= \frac{2\beta^{\frac{\alpha}{2}}}{_1\Psi_1 \left[\begin{array}{c}(\frac{\alpha}{2}, \frac{1}{2}) \\ (1,0)\end{array}; \frac{\gamma}{\sqrt{\beta}}\right]} \sum_{i=0}^{\infty} \frac{\gamma^i}{2i!} \beta^{-\frac{\alpha+i}{2}} \gamma(\frac{\alpha+i}{2}, \beta t^2),$$

where $\gamma(s, t) = \int_0^x t^{s-1} e^{-t} dt$ denotes the lower incomplete gamma function.

5.0.6   Proof of Lemma 6

Proof of Lemma 6 (a)

A moment is a specific quantitative measure of the shape of a function. The kth moment of the random variable following a MHN (α, β, γ) distribution can be expressed as:

$$
\begin{aligned}
E(X^k) &= \int_0^\infty x^k f(x)dx \\
&= \int_0^\infty x^k \frac{2\beta^{\frac{\alpha}{2}}}{{}_1\Psi_1 \left[\begin{matrix} (\frac{\alpha}{2}, \frac{1}{2}) \\ (1, 0) \end{matrix} ; \frac{\gamma}{\sqrt{\beta}} \right]} x^{\alpha-1} \exp(-\beta x^2 + \gamma x)dx \\
&= \frac{{}_1\Psi_1 \left[\begin{matrix} (\frac{\alpha+k}{2}, \frac{1}{2}) \\ (1, 0) \end{matrix} ; \frac{\gamma}{\sqrt{\beta}} \right]}{\beta^{\frac{k}{2}} {}_1\Psi_1 \left[\begin{matrix} (\frac{\alpha}{2}, \frac{1}{2}) \\ (1, 0) \end{matrix} ; \frac{\gamma}{\sqrt{\beta}} \right]}.
\end{aligned}
$$

Proof of Lemma 6 (b)

Based on the property of Fox-Wright function

$$
{}_1\Psi_1 \left[\begin{matrix} (\frac{\alpha+k}{2}, \frac{1}{2}) \\ (1, 0) \end{matrix} ; x \right] = \frac{\alpha + k - 2}{2} {}_1\Psi_1 \left[\begin{matrix} (\frac{\alpha+k-2}{2}, \frac{1}{2}) \\ (1, 0) \end{matrix} ; x \right] + \frac{x}{2} {}_1\Psi_1 \left[\begin{matrix} (\frac{\alpha+k-1}{2}, \frac{1}{2}) \\ (1, 0) \end{matrix} ; x \right],
$$

From part(a) we get that

$$
E(X^k) = \frac{\alpha + k - 2}{2\beta} E(X^{k-2}) + \frac{\gamma}{2\beta} E(X^{k-1}).
$$

Proof of Lemma 6 (c)

The variance of a random variable X, which follows a $\text{ModHalfNormal}(\alpha, \beta, \gamma)$ distribution is

$$\text{Var}(X) = E(X^2) - [E(X)]^2$$

$$= \frac{\sum_{i=0}^{\infty} \frac{\gamma^i \Gamma(\frac{\alpha+i+2}{2})}{i! \beta^{\frac{i}{2}}}}{\beta \sum_{i=0}^{\infty} \frac{\gamma^i \Gamma(\frac{\alpha+i}{2})}{i! \beta^{\frac{i}{2}}}} - \left[\frac{\sum_{i=0}^{\infty} \frac{\gamma^i \Gamma(\frac{\alpha+i+1}{2})}{i! \beta^{\frac{i}{2}}}}{\beta^{\frac{1}{2}} \sum_{i=0}^{\infty} \frac{\gamma^i \Gamma(\frac{\alpha+i}{2})}{i! \beta^{\frac{i}{2}}}} \right]^2$$

$$= \frac{{}_1\Psi_1\left[\begin{matrix} (\frac{\alpha+2}{2}, \frac{1}{2}) \\ (1,0) \end{matrix}; \frac{\gamma}{\sqrt{\beta}} \right]}{\beta\, {}_1\Psi_1\left[\begin{matrix} (\frac{\alpha}{2}, \frac{1}{2}) \\ (1,0) \end{matrix}; \frac{\gamma}{\sqrt{\beta}} \right]} - \left[\frac{{}_1\Psi_1\left[\begin{matrix} (\frac{\alpha+1}{2}, \frac{1}{2}) \\ (1,0) \end{matrix}; \frac{\gamma}{\sqrt{\beta}} \right]}{\sqrt{\beta}\, {}_1\Psi_1\left[\begin{matrix} (\frac{\alpha}{2}, \frac{1}{2}) \\ (1,0) \end{matrix}; \frac{\gamma}{\sqrt{\beta}} \right]} \right]^2. \tag{5.21}$$

Using the property of Fox-Wright function

$${}_1\Psi_1\left[\begin{matrix} (\frac{\alpha+k}{2}, \frac{1}{2}) \\ (1,0) \end{matrix}; x \right] = \frac{\alpha+k-2}{2}\, {}_1\Psi_1\left[\begin{matrix} (\frac{\alpha+k-2}{2}, \frac{1}{2}) \\ (1,0) \end{matrix}; x \right] + \frac{x}{2}\, {}_1\Psi_1\left[\begin{matrix} (\frac{\alpha+k-1}{2}, \frac{1}{2}) \\ (1,0) \end{matrix}; x \right],$$

$$\text{Var}(X) = \frac{\alpha}{2\beta} + E(X)\left(\frac{\gamma}{2\beta} - E(X) \right). \tag{5.22}$$

Proof of Lemma 6 (d)

The moment-generating function of the random variable X from $\text{MHN}(\alpha, \beta, \gamma)$ distribution is

$$M_x(t) = E[e^{tX}]$$

$$= \int_0^\infty e^{tx} \frac{2\beta^{\frac{\alpha}{2}}}{{}_1\Psi_1 \left[\begin{matrix} \left(\frac{\alpha}{2}, \frac{1}{2}\right) \\ (1,0) \end{matrix} ; \frac{\gamma}{\sqrt{\beta}} \right]} x^{\alpha-1} \exp(-\beta x^2 + \gamma x) dx$$

$$= \frac{2\beta^{\frac{\alpha}{2}}}{{}_1\Psi_1 \left[\begin{matrix} \left(\frac{\alpha}{2}, \frac{1}{2}\right) \\ (1,0) \end{matrix} ; \frac{\gamma}{\sqrt{\beta}} \right]} \frac{{}_1\Psi_1 \left[\begin{matrix} \left(\frac{\alpha}{2}, \frac{1}{2}\right) \\ (1,0) \end{matrix} ; \frac{\gamma+t}{\sqrt{\beta}} \right]}{2\beta^{\frac{\alpha}{2}}} \int_0^\infty \frac{2\beta^{\frac{\alpha}{2}}}{{}_1\Psi_1 \left[\begin{matrix} \left(\frac{\alpha}{2}, \frac{1}{2}\right) \\ (1,0) \end{matrix} ; \frac{\gamma+t}{\sqrt{\beta}} \right]} x^{\alpha-1} \exp(-\beta x^2 + (\gamma+t)x) dx$$

$$= \frac{{}_1\Psi_1 \left[\begin{matrix} \left(\frac{\alpha}{2}, \frac{1}{2}\right) \\ (1,0) \end{matrix} ; \frac{\gamma+t}{\sqrt{\beta}} \right]}{{}_1\Psi_1 \left[\begin{matrix} \left(\frac{\alpha}{2}, \frac{1}{2}\right) \\ (1,0) \end{matrix} ; \frac{\gamma}{\sqrt{\beta}} \right]}.$$

5.0.7   Proof of Lemma 7

(a) The density function for the MHN (α, β, γ) distribution is

$$f_{\text{MHN}}(x \mid \alpha, \beta, \gamma) = \frac{2\beta^{\frac{\alpha}{2}}}{{}_1\Psi_1 \left[\begin{matrix} \left(\frac{\alpha}{2}, \frac{1}{2}\right) \\ (1,0) \end{matrix} ; \frac{\gamma}{\sqrt{\beta}} \right]} x^{\alpha-1} \exp(-\beta x^2 + \gamma x).$$

Therefore $\frac{d}{dx} \log\left(f_{\text{MHN}}(x \mid \alpha, \beta, \gamma)\right) = \frac{\alpha-1}{x} - 2\beta x + \gamma$. If X_{mode} denotes the mode of the distribution then

$$\frac{\alpha-1}{X_{\text{mode}}} - 2\beta X_{\text{mode}} + \gamma = 0 \implies X_{\text{mode}} = \frac{\gamma + \sqrt{\gamma^2 + 8(\alpha-1)\beta}}{4\beta},$$

while we ignore the other solution $\frac{\gamma - \sqrt{\gamma^2 + 8(\alpha-1)\beta}}{4\beta}$ as it does not belong to the support of the distribution. Additionally,

$$\frac{d^2}{dx^2} \log\left(f_{\text{MHN}}(x \mid \alpha, \beta, \gamma)\right)\bigg|_{x=X_{\text{mode}}} = -\frac{\alpha-1}{X_{\text{mode}}^2} - 2\beta < 0 \text{ for } \alpha \geq 1.$$

(b) If $1 - \frac{\gamma^2}{8\beta} \leq \alpha < 1$ and $\gamma > 0$ the equation $\frac{\alpha-1}{x} - 2\beta x + \gamma = 0$ has two real positive solutions for x, corresponding to the local minima and the point of local maxima. As a consequence, the density has a local maxima at $\frac{\gamma + \sqrt{\gamma^2 + 8\beta(\alpha-1)}}{4\beta}$ and a local minima at $\frac{\gamma - \sqrt{\gamma^2 + 8\beta(\alpha-1)}}{4\beta}$ because

$$\frac{d^2}{dx^2} \log\left(f_{\text{MHN}}(x \mid \alpha, \beta, \gamma)\right)\bigg|_{x=\frac{\gamma + \sqrt{\gamma^2 + 8\beta(\alpha-1)}}{4\beta}} < 0, \text{ and}$$

$$\frac{d^2}{dx^2} \log\left(f_{\text{MHN}}(x \mid \alpha, \beta, \gamma)\right)\bigg|_{x=\frac{\gamma - \sqrt{\gamma^2 + 8\beta(\alpha-1)}}{4\beta}} > 0.$$

(c) If $0 < \alpha < 1 - \frac{\gamma^2}{8\beta}$ and $\gamma > 0$ then $\frac{d}{dx} \log\left(f_{\text{MHN}}(x \mid \alpha, \beta, \gamma)\right) = -\frac{1}{x}\left[(\sqrt{2\beta x} - \frac{\gamma}{\sqrt{2\beta x}})^2 + \left(1 - \frac{\gamma^2}{8\beta} - \alpha\right)\right] < 0$ for all $x > 0$. Therefore, in this case, the density function is gradually decreasing on $\mathbb{R}_+$ and mode of the distribution doesn't exist. The same is true for the case $\gamma \leq 0$ and $\alpha \leq 1$.

5.0.8 Proof of Lemma 8

Proof of part(a) of Lemma 8

If $\alpha > 1$, then from Lemma 6, we have that $E(X) = \frac{\alpha-1}{2\beta}E(\frac{1}{X}) + \frac{\gamma}{2\beta}$. As the function $x \mapsto \frac{1}{x}$ is convex on $\mathbb{R}_+$, using Jensen's inequality $E(\frac{1}{X}) > \frac{1}{E(X)}$ when $X \sim \text{MHN}(\alpha, \beta, \gamma)$. Hence, $E(X) \geq \frac{\alpha-1}{2\beta}\frac{1}{E(X)} + \frac{\gamma}{2\beta}$. Consequently,

$$E(X) \geq \frac{\gamma + \sqrt{\gamma^2 + 8\beta(\alpha-1)}}{4\beta}. \tag{5.23}$$

113

On the other hand, using lemma 6 we get the recursive relation $E(X^2) = \frac{\alpha}{2\beta}E(X^0) +$

$\frac{\gamma}{2\beta}E(X) = \frac{\alpha}{2\beta} + \frac{\gamma}{2\beta}E(X)$. As $E(X^2) \geq E(X)^2$, $E(X)^2 \leq \frac{\alpha}{2\beta} + \frac{\gamma}{2\beta}E(X)$. Therefore,

$$E(X) \leq \frac{\gamma + \sqrt{\gamma^2 + 8\beta\alpha}}{4\beta}. \tag{5.24}$$

From Lemma 7, we know that the mode of the distribution is $X_{\text{mode}} = \frac{\gamma + \sqrt{\gamma^2 + 8\beta(\alpha-1)}}{4\beta}$

when $\alpha > 1$. An implication of the inequality in 5.23 is that $E(X) > X_{\text{mode}}$. Thus

MHN (α, β, γ) is a positively skewed distribution when $\alpha > 1$.

Proof of part(b) of Lemma 8

Let $X \sim$ MHN (α, β, γ) for $\alpha \geq 4$, and $\gamma > 0$, then $E(\log(X)) \geq \log(X_{\text{mode}})$

where $X_{\text{mode}} = \frac{\gamma + \sqrt{\gamma^2 + 8\beta(\alpha-1)}}{4\beta}$. Without loss of generality we assume β to be 1 for

this proof. Define the function

$$h(\gamma) = E(\log(X)) - \log(X_{\text{mode}}) = \int_{\mathbb{R}_+} \log(x) f_{\text{MHN}}(x \mid \alpha, \beta, \gamma)dx - \log(X_{\text{mode}}).$$

Consider that

$$\frac{\partial X_{\text{mode}}}{\partial \gamma} = \frac{1}{4} + \frac{\gamma}{4\sqrt{\gamma^2 + 8(\alpha-1)}} = \frac{X_{\text{mode}}}{\sqrt{\gamma^2 + 8(\alpha-1)}}.$$

Therefor, we get

$$\begin{aligned}
\frac{\partial h}{\partial \gamma} &= \int_{\mathbb{R}_+} x \log(x) f_{\text{MHN}}(x \mid \alpha, \beta, \gamma)dx - \frac{1}{X_{\text{mode}}}\frac{\partial X_{\text{mode}}}{\partial \gamma} \\
&= E(X \log(X)) - \frac{1}{X_{\text{mode}}}\frac{\partial X_{\text{mode}}}{\partial \gamma} \\
&\geq E(X)\log(E(X)) - \frac{1}{\sqrt{\gamma^2 + 8(\alpha-1)}}, \tag{5.25}
\end{aligned}$$

where the step us due to the Jensen's inequality because $x \mapsto x \log(x)$ is a convex

function ($\frac{\partial^2}{\partial x^2}(x \log(x)) = \frac{1}{x} > 0$ for $x > 0$). Also, it follows from Lemma 6 that

$E(X) \geq X_{\text{mode}} = \frac{\gamma + \sqrt{\gamma^2 + 8(\alpha-1)}}{4}$. Therefore, from Equation 5.25 we get that

$$
\begin{aligned}
\frac{\partial h}{\partial \gamma} &\geq \frac{\gamma + \sqrt{\gamma^2 + 8(\alpha-1)}}{4} \log\left(\frac{\gamma + \sqrt{\gamma^2 + 8(\alpha-1)}}{4}\right) - \frac{1}{\sqrt{\gamma^2 + 8(\alpha-1)}} \\
&\overset{(\dagger\dagger)}{\geq} \log\left(\frac{\sqrt{8(\alpha-1)}}{4}\right)\frac{\sqrt{8(\alpha-1)}}{4} - \frac{1}{\sqrt{8(\alpha-1)}} \\
&\geq 0. \tag{5.26}
\end{aligned}
$$

for $\alpha \geq 4$. The inequality in ($\dagger\dagger$) is an implication of the fact that $\frac{\gamma + \sqrt{\gamma^2 + 8(\alpha-1)}}{4} \log$ $\left(\frac{\gamma + \sqrt{\gamma^2 + 8(\alpha-1)}}{4}\right) - \frac{1}{\sqrt{\gamma^2 + 8(\alpha-1)}}$ is an increasing function in $\gamma > 0$. Altogether, form Equation 5.26, it follows that the function $\gamma \mapsto h(\gamma)$ is an increasing function in γ. As a result, for $\alpha \geq 4$ and $\gamma \geq 0$,

$$
\begin{aligned}
h(\gamma) &\geq h(0) \\
&= \int_0^\infty \log(x)\frac{2x^{\alpha-1}e^{-x^2}}{\Gamma(\frac{\alpha}{2})}dx - \log\left(\frac{\sqrt{8(\alpha-1)}}{4}\right) \\
&= \frac{\frac{\partial}{\partial \alpha}\int_0^\infty 2x^{\alpha-1}e^{-x^2}dx}{\Gamma(\frac{\alpha}{2})} - \log\left(\frac{\sqrt{8(\alpha-1)}}{4}\right) \\
&= \frac{1}{2}\frac{\frac{\partial \Gamma(\frac{\alpha}{2})}{\partial \alpha}}{\Gamma(\frac{\alpha}{2})} - \log\left(\sqrt{\frac{(\alpha-1)}{2}}\right) \\
&\overset{(\dagger\star)}{>} \frac{1}{2}\log\left(\frac{\alpha}{\alpha-1}\right) - \frac{1}{2\alpha} - \frac{1}{6\alpha^2} \\
&\geq 0. \tag{5.27}
\end{aligned}
$$

The inequality in ($\dagger\star$) is due to the fact that $\frac{\frac{d}{d\alpha}(\Gamma(\frac{\alpha}{2}))}{\Gamma(\frac{\alpha}{2})} \geq \log(\frac{\alpha}{2}) - \frac{1}{\alpha} - \frac{1}{3\alpha^2}$ (Batir, 2005). On the other hand,

$$
\frac{d}{d\alpha}\left[\frac{1}{2}\log\left(\frac{\alpha}{\alpha-1}\right) - \frac{1}{2\alpha} - \frac{1}{6\alpha^2}\right] = -\frac{1}{2\alpha(\alpha-1)} + \frac{1}{2\alpha^2} + \frac{1}{3\alpha^3} = -\frac{\alpha+2}{6\alpha^3(\alpha-1)} < 0 \text{ for } \alpha > 1.
$$

Therefore the function $\alpha \mapsto \frac{1}{2}\log(\frac{\alpha}{\alpha-1}) - \frac{1}{2\alpha} - \frac{1}{6\alpha^2}$ is decreasing in $\alpha > 1$ and

$$\frac{1}{2}\log(\frac{\alpha}{\alpha-1}) - \frac{1}{2\alpha} - \frac{1}{6\alpha^2} \geq \lim_{\alpha\to\infty} \frac{1}{2}\log(\frac{\alpha}{\alpha-1}) - \frac{1}{2\alpha} - \frac{1}{6\alpha^2} = 0.$$

Finally, if follows from the the definition of the function $\gamma \mapsto h(\gamma)$ and the Equation 5.27 that

$$E(\log(X)) \geq \log(X_{\mathrm{mode}}),$$

for all $\gamma \geq 0$ when $X \sim \mathrm{MHN}(\alpha, 1, \gamma)$, $\alpha \geq 4$. On the other hand, using Jensen's inequality the previous part of the lemma we get that

$$E(\log(X)) \leq \log(E(X)) \leq \log\left(\frac{\gamma + \sqrt{\gamma^2 + 8\alpha\beta}}{4\beta}\right). \tag{5.28}$$

5.0.9 Proof of Lemma 9

Proof of Lemma 9 case(a):

Let $\gamma > 0$ when $X \sim \mathrm{MHN}(\alpha, \beta, \gamma)$. Consider a random variable V such that the conditional probability distribution of V given X is a Poisson distribution with parameter γX which has the probability mass function

$$f_{Poi}(V = v \mid X) = \frac{e^{-\gamma X}(\gamma X)^v}{v!}. \tag{5.29}$$

Consequently the conditional probability density of the random variable X given V

$$
\begin{aligned}
f_{X|V}(x \mid v) &= \frac{f_{X,V}(x,v)}{f_V(v)} \\[2em]
&= \frac{\dfrac{2\beta^{\frac{\alpha}{2}} x^{\alpha-1} e^{-\beta x^2 + \gamma x}}{{}_1\Psi_1\!\left[\begin{matrix}(\frac{\alpha}{2},\frac{1}{2})\\(1,0)\end{matrix};\frac{\gamma}{\sqrt{\beta}}\right]} \left\{\dfrac{e^{-\gamma x}(\gamma x)^v}{v!}\right\} \mathbb{I}(x>0)}{\dfrac{\gamma^v \Gamma(\frac{\alpha+v}{2})}{{}_1\Psi_1\!\left[\begin{matrix}(\frac{\alpha}{2},\frac{1}{2})\\(1,0)\end{matrix};\frac{\gamma}{\sqrt{\beta}}\right] v!\, \beta^{\frac{v}{2}}}} \\[2em]
&= \frac{2\beta^{\frac{\alpha+v}{2}}}{\Gamma(\frac{\alpha+v}{2})} x^{\alpha+v-1} e^{-\beta x^2} \mathbb{I}(x>0). \quad\quad (5.30)
\end{aligned}
$$

As a result, the conditional distribution of the random variable X given V is the square root of a Gamma random variable with shape parameter $\frac{\alpha+v}{2}$ and rate parameter β. Obviously the conditional distribution of V given X is a Poisson random variable with parameter γX.

Proof of Lemma 9 case(b):

Let $\gamma < 0$ when $X \sim \text{MHN}(\alpha, \beta, \gamma)$. Consider a random variable U such that the conditional probability distribution of U given X is a Generalized Inverse Gaussian distribution, i.e.

$$
U \mid X \sim GIG(\frac{1}{2}, 1, \gamma^2 X^2)
$$

with the probability density function

$$
f_{U|X}(u \mid x) = \frac{1}{\sqrt{2\pi}} e^{|\gamma|x} u^{\frac{1}{2}-1} e^{-\frac{1}{2}\left(u + \frac{\gamma^2 x^2}{u}\right)}. \quad\quad (5.31)
$$

Hence the conditional probability density of the random variable X given U

$$
\begin{aligned}
f_{X|U}(x \mid u) &= \frac{f_{X,U}(x,u)}{f_U(u)} \\
&= \frac{\frac{2\beta^{\frac{\alpha}{2}} x^{\alpha-1} e^{-\beta x^2 - |\gamma| x}}{{}_1\Psi_1\left[\begin{matrix} \left(\frac{\alpha}{2}, \frac{1}{2}\right) \\ (1,0) \end{matrix}; \frac{\gamma}{\sqrt{\beta}}\right]} \left\{ \frac{1}{\sqrt{2\pi}} e^{|\gamma| x} u^{\frac{1}{2}-1} e^{-\frac{1}{2}\left(u + \frac{\gamma^2 x^2}{u}\right)} \right\} \mathbb{I}(x > 0)}{\frac{\Gamma(\frac{\alpha}{2})}{2(\beta + \frac{\gamma^2}{u})^{\frac{\alpha}{2}}} \frac{2\beta^{\frac{\alpha}{2}}}{\sqrt{2\pi}\,{}_1\Psi_1\left[\begin{matrix} \left(\frac{\alpha}{2}, \frac{1}{2}\right) \\ (1,0) \end{matrix}; \frac{\gamma}{\sqrt{\beta}}\right]} e^{-\frac{1}{2}u} u^{\frac{1}{2}-1}} \\
&= \frac{2(\beta + \frac{\gamma^2}{2})^{\frac{\alpha}{2}}}{\Gamma(\frac{\alpha}{u})} x^{\alpha-1} e^{-\left(\beta + \frac{\gamma^2}{u}\right) x^2} \mathbb{I}(x > 0). \tag{5.32}
\end{aligned}
$$

As a result, the conditional distribution of the random variable X given U is actually the square root of a Gamma random variable with shape parameter $\frac{\alpha}{2}$ and rate parameter $\beta + \frac{\gamma^2}{u}$. The above result can be utilized to design hierarchical models by introducing additional variables U, V that can bypass the sampling step that involves sampling from the MHN (α, β, γ) distribution directly. But, this strategy of introducing additional variables is expected to lead to slower mixing Markov chains.

5.0.10 Proof of Lemma 10

The proof of the lemma follows from the definition of the square root of Gamma, truncated Normal and the Half normal distribution.

5.0.11   Proof of Theorem 1

Proof of Theorem 1 (a)

Let $\gamma > 0$.

$$
\begin{aligned}
f_{\text{MHN}}(x \mid \alpha, \beta, \gamma) &= \frac{2\beta^{\frac{\alpha}{2}}}{{}_1\Psi_1\left[\begin{matrix}\left(\frac{\alpha}{2}, \frac{1}{2}\right) \\ (1,0)\end{matrix}; \frac{\gamma}{\sqrt{\beta}}\right]} x^{\alpha-1} \exp\{-\beta x^2 + \gamma x\} \\
&= \frac{2\beta^{\frac{\alpha}{2}}}{{}_1\Psi_1\left[\begin{matrix}\left(\frac{\alpha}{2}, \frac{1}{2}\right) \\ (1,0)\end{matrix}; \frac{\gamma}{\sqrt{\beta}}\right]} x^{\alpha-1} \exp\{-\beta x^2 + \gamma x\} \exp\{\beta(x-\mu)^2\} \exp\{-\beta(x-\mu)^2\} \\
&= \frac{2\beta^{\frac{\alpha}{2}}}{{}_1\Psi_1\left[\begin{matrix}\left(\frac{\alpha}{2}, \frac{1}{2}\right) \\ (1,0)\end{matrix}; \frac{\gamma}{\sqrt{\beta}}\right]} x^{\alpha-1} \exp\{-(2\beta\mu - \gamma)x + \beta\mu^2\} \exp\{-\beta(x-\mu)^2\}
\end{aligned}
$$

$$(5.33)$$

Assuming $\mu > \frac{\gamma}{2\beta}$, it can be shown that $x^{\alpha-1} \exp\{-(2\beta\mu-\gamma)x\} \leq \left(\frac{\alpha-1}{2\beta\mu-\gamma}\right)^{\alpha-1} \exp\{-(\alpha-1)\}$ for all $x > 0$. Therefore,

$$
\begin{aligned}
f_{\text{MHN}}(x \mid \alpha, \beta, \gamma) &\leq \frac{2\beta^{\frac{\alpha}{2}}}{{}_1\Psi_1\left[\begin{matrix}\left(\frac{\alpha}{2}, \frac{1}{2}\right) \\ (1,0)\end{matrix}; \frac{\gamma}{\sqrt{\beta}}\right]} \left(\frac{\alpha-1}{2\beta\mu-\gamma}\right)^{\alpha-1} \exp\{-(\alpha-1) + \beta\mu^2\} \exp\{-\beta(x-\mu)^2\} \\
&= K_1 f_{\text{Normal}}\left(x \mid \mu, \frac{1}{2\beta}\right),
\end{aligned}
$$

$$(5.34)$$

where $K_1 = \frac{2\sqrt{\pi}\left(\frac{\sqrt{\beta}(\alpha-1)}{2\beta\mu-\gamma}\right)^{\alpha-1}}{{}_1\Psi_1\left[\begin{matrix}\left(\frac{\alpha}{2}, \frac{1}{2}\right) \\ (1,0)\end{matrix}; \frac{\gamma}{\sqrt{\beta}}\right]} e^{\{-(\alpha-1)+\beta\mu^2\}}$. On the other hand, if $0 < \delta < \beta$ be any constant then

$$f_{\mathrm{MHN}}(x \mid \alpha, \beta, \gamma) \;=\; \frac{2\beta^{\frac{\alpha}{2}}}{{}_1\Psi_1\left[\begin{matrix}\left(\frac{\alpha}{2},\frac{1}{2}\right) \\ (1,0)\end{matrix}; \frac{\gamma}{\sqrt{\beta}}\right]}\, x^{\alpha-1}\exp\left(-\beta x^2 + \gamma x\right)$$

$$=\; \frac{2\beta^{\frac{\alpha}{2}}}{{}_1\Psi_1\left[\begin{matrix}\left(\frac{\alpha}{2},\frac{1}{2}\right) \\ (1,0)\end{matrix}; \frac{\gamma}{\sqrt{\beta}}\right]}\,\exp\left(-(\beta-\delta)x^2 + \gamma x\right) x^{\alpha-1}e^{-\delta x^2}$$

$$(5.35)$$

As $\gamma x - (\beta-\delta)x^2 = \frac{\gamma^2}{4(\beta-\delta)} - \left\{\frac{\gamma}{2\sqrt{(\beta-\delta)}} - \sqrt{(\beta-\delta)}x\right\}^2 \leq \frac{\gamma^2}{4(\beta-\delta)}$, from Equation 5.35
we get that

$$f_{\mathrm{MHN}}(x \mid \alpha, \beta, \gamma) \;\leq\; \frac{2\beta^{\frac{\alpha}{2}}}{{}_1\Psi_1\left[\begin{matrix}\left(\frac{\alpha}{2},\frac{1}{2}\right) \\ (1,0)\end{matrix}; \frac{\gamma}{\sqrt{\beta}}\right]}\, e^{\frac{\gamma^2}{4(\beta-\delta)}}\, x^{\alpha-1}e^{-\delta x^2}$$

$$=\; \frac{2\beta^{\frac{\alpha}{2}}}{{}_1\Psi_1\left[\begin{matrix}\left(\frac{\alpha}{2},\frac{1}{2}\right) \\ (1,0)\end{matrix}; \frac{\gamma}{\sqrt{\beta}}\right]}\, e^{\frac{\gamma^2}{4(\beta-\delta)}}\, \frac{\Gamma(\alpha/2)}{2\delta^{\frac{\alpha}{2}}} f_{\sqrt{\mathrm{Gam}}}(x|\alpha,\delta)$$

$$=\; K_2 f_{\sqrt{\mathrm{Gam}}}(x|\alpha,\delta), \qquad\qquad (5.36)$$

where $K_2 = \dfrac{\Gamma(\frac{\alpha}{2})\beta^{\frac{\alpha}{2}}\exp\left(\frac{\gamma^2}{4(\beta-\delta)}\right)}{{}_1\Psi_1\left[\begin{matrix}\left(\frac{\alpha}{2},\frac{1}{2}\right) \\ (1,0)\end{matrix}; \frac{\gamma}{\sqrt{\beta}}\right]\delta^{\frac{\alpha}{2}}}.$

Proof of part(b), Theorem 1

The inequalities in part(a) is efficient when the constant $K_1(\mu,\alpha,\beta,\gamma)$ and $K_2(\delta,\alpha,\beta,\gamma)$ are small. Therefore, we find the optimum values for μ and δ by minimizing $K_1(\mu,\alpha,\beta,\gamma)$ and $K_2(\delta,\alpha,\beta,\gamma)$ correspondingly. The function $\mu \mapsto \frac{\exp(\beta\mu^2)}{(2\beta\mu-\gamma)^{\alpha-1}}$

from $(\frac{\gamma}{2\beta}, \infty)$ to $\mathbb{R}_+$ is minimized when $\mu = \frac{\gamma + \sqrt{\gamma^2 + 8(\alpha-1)\beta}}{4\beta}$. Thus, the best possible choice for μ is $\mu_{\mathrm{opt}} = \frac{\gamma + \sqrt{\gamma^2 + 8(\alpha-1)\beta}}{4\beta}$ and $K_1(\mu_{\mathrm{opt}}) = \frac{2\sqrt{\pi}\beta^{\frac{\alpha-1}{2}}\mu^{\alpha-1}\exp\{-(\alpha-1)+\beta\mu^2\}}{{}_1\Psi_1\left[\begin{matrix}(\frac{\alpha}{2},\frac{1}{2})\\(1,0)\end{matrix};\frac{\gamma}{\sqrt{\beta}}\right]}$.

On the other, the function $\delta \mapsto \frac{\exp\left(\frac{\gamma^2}{4(\beta-\delta)}\right)}{\delta^{\frac{\alpha}{2}}}$ from $(0,\beta)$ to $\mathbb{R}_+$ is minimized when $\delta = \beta + \frac{\gamma^2 - \gamma\sqrt{\gamma^2 + 8\alpha\beta}}{4\alpha}$. Therefore, the optimum choice for δ is given as $\delta_{\mathrm{opt}} = \beta + \frac{\gamma^2 - \gamma\sqrt{\gamma^2 + 8\alpha\beta}}{4\alpha}$.

Proof of part(c), Theorem 1

Let $\Delta = \frac{\gamma}{\sqrt{\beta}}$. As $\sqrt{\beta}\mu_{\mathrm{opt}} = \frac{\Delta + \sqrt{\Delta^2 + 8(\alpha-1)}}{4}$,

$$
\begin{aligned}
K_1(\mu_{\mathrm{opt}}, \alpha, \beta, \gamma) &= \frac{2\sqrt{\pi}\left(\frac{\sqrt{\beta}(\alpha-1)}{2\beta\mu_{\mathrm{opt}}-\gamma}\right)^{\alpha-1}\exp\left(-(\alpha-1)+\beta\mu_{\mathrm{opt}}^2\right)}{{}_1\Psi_1\left[\begin{matrix}(\frac{\alpha}{2},\frac{1}{2})\\(1,0)\end{matrix};\frac{\gamma}{\sqrt{\beta}}\right]} \\
&= \frac{2\sqrt{\pi}\left(\sqrt{\beta}\mu_{\mathrm{opt}}\right)^{\alpha-1}\exp\left(-(\alpha-1)+\beta\mu_{\mathrm{opt}}^2\right)}{{}_1\Psi_1\left[\begin{matrix}(\frac{\alpha}{2},\frac{1}{2})\\(1,0)\end{matrix};\frac{\gamma}{\sqrt{\beta}}\right]} \\
&= \frac{2\sqrt{\pi}\left(\frac{\Delta + \sqrt{\Delta^2 + 8(\alpha-1)}}{4}\right)^{\alpha-1}\exp\left(-\frac{(\alpha-1)}{2}+\frac{\Delta^2 + \Delta\sqrt{\Delta^2 + 8(\alpha-1)}}{8}\right)}{{}_1\Psi_1\left[\begin{matrix}(\frac{\alpha}{2},\frac{1}{2})\\(1,0)\end{matrix};\Delta\right]} \quad (5.37)
\end{aligned}
$$

$$K_1(\delta_{\text{opt}}, \alpha, \beta, \gamma) = \frac{\left(\sqrt{\beta}\right)^\alpha \Gamma(\tfrac{\alpha}{2})\ \exp\left(\frac{\gamma^2}{4(\beta-\delta_{\text{opt}})}\right)}{{}_1\Psi_1\left[\begin{matrix}(\tfrac{\alpha}{2},\tfrac{1}{2})\\(1,0)\end{matrix}; \tfrac{\gamma}{\sqrt{\beta}}\right]\left(\sqrt{\delta_{\text{opt}}}\right)^\alpha}$$

$$= \frac{\left(\tfrac{\beta}{\delta_{\text{opt}}}\right)^{\tfrac{\alpha}{2}} \Gamma(\tfrac{\alpha}{2})\ \exp\left(\frac{4\alpha\gamma^2}{4(\gamma\sqrt{\gamma^2+8\alpha\beta}-\gamma^2)}\right)}{{}_1\Psi_1\left[\begin{matrix}(\tfrac{\alpha}{2},\tfrac{1}{2})\\(1,0)\end{matrix}; \tfrac{\gamma}{\sqrt{\beta}}\right]}$$

$$= \frac{\Gamma(\tfrac{\alpha}{2})\ \exp\left(\frac{\alpha}{(\sqrt{1+8\frac{\alpha}{\Delta^2}}-1)}\right)}{\left(1+\frac{\Delta^2-\Delta\sqrt{\Delta^2+8\alpha}}{4\alpha}\right)^{\tfrac{\alpha}{2}} {}_1\Psi_1\left[\begin{matrix}(\tfrac{\alpha}{2},\tfrac{1}{2})\\(1,0)\end{matrix}; \Delta\right]}. \tag{5.38}$$

Consequently, we will denote $K_1(\alpha, \Delta) := K_1(\mu_{\text{opt}}, \alpha, \beta, \gamma)$ and $K_2(\alpha, \Delta) := K_1(\delta_{\text{opt}}, \alpha, \beta, \gamma)$ where $\Delta = \frac{\gamma}{\sqrt{\beta}}$.

Result 1. *Let $\alpha > 0$ and $\Delta \in \mathbb{R}$.*

(a) Then $\frac{\partial}{\partial\Delta}\left\{{}_1\Psi_1\left[\begin{matrix}(\tfrac{\alpha}{2},\tfrac{1}{2})\\(1,0)\end{matrix}; \Delta\right]\right\} = {}_1\Psi_1\left[\begin{matrix}(\tfrac{\alpha+1}{2},\tfrac{1}{2})\\(1,0)\end{matrix}; \Delta\right]$

(b) $\frac{\Delta+\sqrt{\Delta^2+8(\alpha-1)}}{4} \leq \frac{{}_1\Psi_1\left[\begin{matrix}(\tfrac{\alpha+1}{2},\tfrac{1}{2})\\(1,0)\end{matrix}; \Delta\right]}{{}_1\Psi_1\left[\begin{matrix}(\tfrac{\alpha}{2},\tfrac{1}{2})\\(1,0)\end{matrix}; \Delta\right]} \leq \frac{\Delta+\sqrt{\Delta^2+8(\alpha)}}{4}$

Proof of the part(a) of Result 1

$$\frac{\partial}{\partial \Delta} \left\{ {}_1\Psi_1 \begin{bmatrix} \left(\frac{\alpha}{2}, \frac{1}{2}\right) \\ (1,0) \end{bmatrix}; \Delta \right\} = \frac{\partial}{\partial \Delta} \left\{ 2 \int_0^\infty y^{\alpha-1} \exp\left(-y^2 + \Delta y\right) dy \right\}$$

$$= 2 \int_0^\infty y^{\alpha+1-1} \exp\left(-y^2 + \Delta y\right) dy$$

$$= {}_1\Psi_1 \begin{bmatrix} \left(\frac{\alpha+1}{2}, \frac{1}{2}\right) \\ (1,0) \end{bmatrix}; \Delta . \tag{5.39}$$

Proof of the part(b) of Result 1

Let $X \sim \text{MHN}\,(\alpha, 1, \Delta)$ then from Lemma 6, we get that $E(X) = \dfrac{{}_1\Psi_1 \begin{bmatrix} \left(\frac{\alpha+1}{2}, \frac{1}{2}\right) \\ (1,0) \end{bmatrix}; \Delta}{{}_1\Psi_1 \begin{bmatrix} \left(\frac{\alpha}{2}, \frac{1}{2}\right) \\ (1,0) \end{bmatrix}; \Delta}.$

From Lemma 8 we get that

$$\frac{\Delta + \sqrt{\Delta^2 + 8(\alpha-1)}}{4} \leq \frac{{}_1\Psi_1 \begin{bmatrix} \left(\frac{\alpha+1}{2}, \frac{1}{2}\right) \\ (1,0) \end{bmatrix}; \Delta}{{}_1\Psi_1 \begin{bmatrix} \left(\frac{\alpha}{2}, \frac{1}{2}\right) \\ (1,0) \end{bmatrix}; \Delta} \leq \frac{\Delta + \sqrt{\Delta^2 + 8(\alpha)}}{4}.$$

5.0.12 Proof of Theorem 2

Proof of part(a), Theorem 2

From Theorem 1, we get that

$$f_{\text{MHN}}(x \mid \alpha, \beta, \gamma) \leq IK_1(\alpha, \Delta) \left\{ f_{\text{Normal}}(x \mid \mu, \frac{1}{2\beta}) \right\} + (1-I)K_2(\alpha, \Delta) \left\{ f_{\sqrt{\text{Gam}}}(x \mid \alpha, \delta) \right\},$$

where $\Delta = \frac{\gamma}{\sqrt{\beta}}$ and $I = \mathbb{I}(K_1(\alpha, \Delta) \leq K_2(\alpha, \Delta))$. Therefore

$$
\begin{aligned}
\mathcal{A}_{\text{pos}}(\alpha, \Delta) &= \frac{1}{\int_0^\infty \left\{ IK_1(\alpha, \Delta) f_{\text{Normal}}(x \mid \mu, \frac{1}{2\beta}) + (1-I)K_2(\alpha, \Delta) f_{\sqrt{\text{Gam}}}(x \mid \alpha, \delta) \right\} dx} \\
&= \frac{1}{IK_1(\alpha, \Delta) + (1-I)K_2(\alpha, \Delta)} \\
&= \frac{I}{K_1(\alpha, \Delta)} + \frac{1-I}{K_2(\alpha, \Delta)} \\
&= \max \left\{ \frac{1}{K_1(\alpha, \Delta)}, \frac{1}{K_2(\alpha, \Delta)} \right\}.
\end{aligned}
\tag{5.40}
$$

Proof of part(b), Theorem 2

To prove that $\Delta \mapsto K_1(\alpha, \Delta)$ is decreasing in Δ.

$$K_1(\alpha, \Delta) = \frac{2\sqrt{\pi} \, (B_\Delta)^{\alpha-1} \, \exp\left(-(\alpha-1) + B_\Delta^2\right)}{{}_1\Psi_1 \left[\begin{matrix} (\frac{\alpha}{2}, \frac{1}{2}) \\ (1,0) \end{matrix}; \Delta \right]}$$

$$
\begin{aligned}
\implies \log(K_1(\alpha, \Delta)) &= (\alpha-1)\log(B_\Delta) + B_\Delta^2 - \log\left({}_1\Psi_1 \left[\begin{matrix} (\frac{\alpha}{2}, \frac{1}{2}) \\ (1,0) \end{matrix}; \Delta \right] \right) \\
&+ \log(2\sqrt{\pi} \exp(-(\alpha-1))).
\end{aligned}
\tag{5.41}
$$

where $B_\Delta = \frac{\Delta + \sqrt{\Delta^2 + 8(\alpha - 1)}}{4}$. Note that

$$
\begin{aligned}
\frac{\partial B_\Delta}{\partial \Delta} &= \frac{\partial}{\partial \Delta}\left(\frac{\Delta + \sqrt{\Delta^2 + 8(\alpha - 1)}}{4}\right) \\
&= \frac{1}{4}\left(1 + \frac{\Delta}{\sqrt{\Delta^2 + 8(\alpha - 1)}}\right) \\
&= \left(\frac{\Delta + \sqrt{\Delta^2 + 8(\alpha - 1)}}{4}\right)\left(\frac{1}{\sqrt{\Delta^2 + 8(\alpha - 1)}}\right) \\
&= \left(\frac{B_\Delta}{\sqrt{\Delta^2 + 8(\alpha - 1)}}\right).
\end{aligned}
\tag{5.42}
$$

Therefore, Result 1 and Equation 5.41, it follows that

$$
\begin{aligned}
\frac{\partial \log(K_1(\alpha, \Delta))}{\partial \Delta} &= \left\{\frac{(\alpha - 1)}{B_\Delta} + 2B_\Delta\right\}\frac{\partial B_\Delta}{\partial \Delta} - \frac{\frac{\partial}{\partial \Delta}\left\{{}_1\Psi_1\left[\begin{matrix}(\frac{\alpha}{2}, \frac{1}{2}) \\ (1, 0)\end{matrix}; \Delta\right]\right\}}{{}_1\Psi_1\left[\begin{matrix}(\frac{\alpha}{2}, \frac{1}{2}) \\ (1, 0)\end{matrix}; \Delta\right]} \\
&= \left\{\frac{(\alpha - 1)}{B_\Delta} + 2B_\Delta\right\}\frac{\partial B_\Delta}{\partial \Delta} - \frac{{}_1\Psi_1\left[\begin{matrix}(\frac{\alpha+1}{2}, \frac{1}{2}) \\ (1, 0)\end{matrix}; \Delta\right]}{{}_1\Psi_1\left[\begin{matrix}(\frac{\alpha}{2}, \frac{1}{2}) \\ (1, 0)\end{matrix}; \Delta\right]} \\
&< \frac{(\alpha - 1) + 2B_\Delta^2}{\sqrt{\Delta^2 + 8(\alpha - 1)}} - \frac{\Delta + \sqrt{\Delta^2 + 8(\alpha - 1)}}{4} \\
&< \frac{(\alpha - 1) + 2B_\Delta^2}{\sqrt{\Delta^2 + 8(\alpha - 1)}} - B_\Delta \\
&= \frac{2(\alpha - 1) + \Delta B_\Delta}{\sqrt{\Delta^2 + 8(\alpha - 1)}} - B_\Delta,
\end{aligned}
\tag{5.43}
$$

because B_Δ satisfy the condition $2B_\Delta^2 - \Delta B_\Delta - (\alpha - 1) = 0$. Consider the fact that

$$
\begin{aligned}
(2(\alpha - 1) + \Delta B_\Delta)^2 &= 4(\alpha - 1)^2 + 4(\alpha - 1)\Delta B_\Delta + \Delta^2 B_\Delta^2 \\
&= 4(\alpha - 1)\left[(\alpha - 1) + \Delta B_\Delta\right] + \Delta^2 B_\Delta^2 \\
&= 4(\alpha - 1)\left[2B_\Delta^2\right] + \Delta^2 B_\Delta^2 \\
&= B_\Delta^2\left(\Delta^2 + 8(\alpha - 1)\right).
\end{aligned}
\tag{5.44}
$$

Therefore it follows from Equation 5.44 and Equation 5.43 it follows that,

$$
\begin{aligned}
\frac{\partial \log(K_1(\alpha, \Delta))}{\partial \Delta} &< \frac{2(\alpha - 1) + \Delta B_\Delta}{\sqrt{\Delta^2 + 8(\alpha - 1)}} - B_\Delta \\
&= \frac{\sqrt{B_\Delta^2\left(\Delta^2 + 8(\alpha - 1)\right)}}{\sqrt{\Delta^2 + 8(\alpha - 1)}} - B_\Delta \\
&= 0.
\end{aligned}
$$

To Prove that $\Delta \mapsto K_2(\alpha, \Delta)$ is increasing in Δ. Consider that

$$
K_2(\alpha, \Delta) = \frac{\left(\frac{\sqrt{\Delta^2 + 8\alpha} + \Delta}{2}\right)^\alpha \Gamma(\frac{\alpha}{2}) \exp\left(\frac{\Delta(\sqrt{\Delta^2 + 8\alpha} + \Delta)}{8}\right)}{(2\alpha)^{\frac{\alpha}{2}} {}_1\Psi_1\left[\begin{matrix}(\frac{\alpha}{2}, \frac{1}{2}), \\ (1, 0)\end{matrix}; \frac{\gamma}{\sqrt{\beta}}\right]} = \frac{(2D_\Delta)^\alpha \Gamma(\frac{\alpha}{2}) \exp\left(\frac{\Delta D_\Delta}{2}\right)}{(2\alpha)^{\frac{\alpha}{2}} {}_1\Psi_1\left[\begin{matrix}(\frac{\alpha}{2}, \frac{1}{2}), \\ (1, 0)\end{matrix}; \frac{\gamma}{\sqrt{\beta}}\right]},
$$

where $D_\Delta = \frac{\Delta + \sqrt{\Delta^2 + 8(\alpha)}}{4}$. Hence

$$
\begin{aligned}
&\log(K_2(\alpha, \Delta)) \\
&= \alpha \log(D_\Delta) + \frac{\Delta D_\Delta}{2} - \log({}_1\Psi_1\left[\begin{matrix}(\frac{\alpha}{2}, \frac{1}{2}), \\ (1, 0)\end{matrix}; \frac{\gamma}{\sqrt{\beta}}\right]) + \log\left(\frac{(2)^\alpha \Gamma(\frac{\alpha}{2})}{(2\alpha)^{\frac{\alpha}{2}}}\right) \\
&= \alpha \log(D_\Delta) + D_\Delta^2 - \frac{\alpha}{2} - \log({}_1\Psi_1\left[\begin{matrix}(\frac{\alpha}{2}, \frac{1}{2}), \\ (1, 0)\end{matrix}; \frac{\gamma}{\sqrt{\beta}}\right]) + \log\left(\frac{(2)^\alpha \Gamma(\frac{\alpha}{2})}{(2\alpha)^{\frac{\alpha}{2}}}\right),
\end{aligned}
\tag{5.45}
$$

because D_Δ satisfy the condition $2D_\Delta^2 - \Delta D_\Delta - \alpha = 0$. Note that

$$
\begin{aligned}
\frac{\partial D_\Delta}{\partial \Delta} &= \frac{\partial}{\partial \Delta}\left(\frac{\Delta + \sqrt{\Delta^2 + 8\alpha}}{4}\right) \\
&= \frac{1}{4}\left(1 + \frac{\Delta}{\sqrt{\Delta^2 + 8\alpha}}\right) \\
&= \left(\frac{\Delta + \sqrt{\Delta^2 + 8\alpha}}{4}\right)\left(\frac{1}{\sqrt{\Delta^2 + 8\alpha}}\right) \\
&= \left(\frac{D_\Delta}{\sqrt{\Delta^2 + 8\alpha}}\right).
\end{aligned}
\tag{5.46}
$$

Therefore, from Result 1 and Equation 5.45, we infer that

$$
\begin{aligned}
\frac{\partial \log(K_2(\alpha, \Delta))}{\partial \Delta} &= \left\{\frac{\alpha}{D_\Delta} + 2D_\Delta\right\}\frac{\partial D_\Delta}{\partial \Delta} - \frac{\frac{\partial}{\partial \Delta}\left\{{}_1\Psi_1\left[\begin{matrix}(\frac{\alpha}{2}, \frac{1}{2}), \\ (1, 0)\end{matrix}; \Delta\right]\right\}}{{}_1\Psi_1\left[\begin{matrix}(\frac{\alpha}{2}, \frac{1}{2}), \\ (1, 0)\end{matrix}; \Delta\right]} \\
&= \left\{\frac{\alpha}{D_\Delta} + 2D_\Delta\right\}\frac{\partial D_\Delta}{\partial \Delta} - \frac{{}_1\Psi_1\left[\begin{matrix}(\frac{\alpha+1}{2}, \frac{1}{2}), \\ (1, 0)\end{matrix}; \Delta\right]}{{}_1\Psi_1\left[\begin{matrix}(\frac{\alpha}{2}, \frac{1}{2}), \\ (1, 0)\end{matrix}; \Delta\right]} \\
&> \frac{\alpha + 2D_\Delta^2}{\sqrt{\Delta^2 + 8\alpha}} - \frac{\Delta + \sqrt{\Delta^2 + 8\alpha}}{4} \\
&> \frac{\alpha + 2D_\Delta^2}{\sqrt{\Delta^2 + 8\alpha}} - D_\Delta \\
&= \frac{2\alpha + \Delta D_\Delta}{\sqrt{\Delta^2 + 8\alpha}} - D_\Delta,
\end{aligned}
\tag{5.47}
$$

where the last equality follows as the constant D_Δ satisfy the condition $2D_\Delta^2 - \Delta D_\Delta = \alpha$. Consider the fact that

$$
\begin{aligned}
(2\alpha + \Delta D_\Delta)^2 &= 4\alpha^2 + 4\alpha\Delta D_\Delta + \Delta^2 D_\Delta^2 \\
&= 4\alpha\left[\alpha + \Delta D_\Delta\right] + \Delta^2 D_\Delta^2 \\
&= 4\alpha\left[2D_\Delta^2\right] + \Delta^2 D_\Delta^2 \\
&= D_\Delta^2\left(\Delta^2 + 8\alpha\right). \quad (5.48)
\end{aligned}
$$

Consequently, it follows from Equation 5.47 and Equation 5.48 that

$$
\frac{\partial \log(K_2(\alpha, \Delta))}{\partial \Delta} \quad > \quad \frac{2\alpha + \Delta D_\Delta}{\sqrt{\Delta^2 + 8\alpha}} - D_\Delta = \frac{\sqrt{D_\Delta^2\left(\Delta^2 + 8\alpha\right)}}{\sqrt{\Delta^2 + 8\alpha}} - D_\Delta = 0. \quad (5.49)
$$

Proof of part(c), Theorem 2

Let $B_\alpha = \frac{\sqrt{\alpha} + \sqrt{\alpha + 8(\alpha - 1)}}{4} = \frac{\sqrt{\alpha} + \sqrt{9\alpha - 8}}{4}$ for $\alpha \geq 1$.

$$
\begin{aligned}
\lim_{\alpha \to \infty} \frac{K_1(\alpha, \sqrt{\alpha})}{K_2(\alpha, \sqrt{\alpha})} &= \lim_{\alpha \to \infty} \frac{2\sqrt{\pi}\,(B_\alpha)^{\alpha-1}\,\exp\left(-(\alpha - 1) + B_\alpha^2\right)}{2^{\frac{\alpha}{2}}\Gamma(\frac{\alpha}{2})\exp(\frac{\alpha}{2})} \\
&= \lim_{\alpha \to \infty} \frac{2\sqrt{\pi}\,(B_\alpha)^{\alpha-1}\,\exp\left(-(\alpha - 1) + B_\alpha^2\right)}{2^{\frac{\alpha}{2}}\left[2\sqrt{\pi}\alpha^{\frac{\alpha-1}{2}}\exp(-\frac{\alpha}{2})(2)^{-\frac{\alpha}{2}}\right]\exp(\frac{\alpha}{2})} \times \left\{\lim_{\alpha \to \infty} \frac{2\sqrt{\frac{\pi}{\alpha}}\left(\frac{\alpha}{2e}\right)^{\frac{\alpha}{2}}}{\Gamma(\frac{\alpha}{2})}\right\} \\
&= \lim_{\alpha \to \infty} \frac{2\sqrt{\pi}\,(B_\alpha)^{\alpha-1}\,\exp\left(-(\alpha - 1) + B_\alpha^2\right)}{2^{\frac{\alpha}{2}}\left[2\sqrt{\pi}\alpha^{\frac{\alpha-1}{2}}\exp(-\frac{\alpha}{2})(2)^{-\frac{\alpha}{2}}\right]\exp(\frac{\alpha}{2})},
\end{aligned}
$$

$$
(5.50)
$$

because $\lim_{\alpha\to\infty}\frac{2\sqrt{\frac{\pi}{\alpha}}\left(\frac{\alpha}{2e}\right)^{\frac{\alpha}{2}}}{\Gamma\left(\frac{\alpha}{2}\right)}=1$ using the Starling's approximation for the Gamma function. Consequently,

$$
\begin{aligned}
\lim_{\alpha\to\infty}\frac{K_1(\alpha,\sqrt{\alpha})}{K_2(\alpha,\sqrt{\alpha})} &= \lim_{\alpha\to\infty}\frac{(B_\alpha)^{\alpha-1}\exp\left(-(\alpha-1)+B_\alpha^2\right)}{\alpha^{\frac{\alpha-1}{2}}}\times\{1\}\\
&= \lim_{\alpha\to\infty}\left(\frac{B_\alpha}{\sqrt{\alpha}}\right)^{\alpha-1}\exp\left(-(\alpha-1)+B_\alpha^2\right)\\
&= \lim_{\alpha\to\infty}\left(\frac{B_\alpha}{\sqrt{\alpha}}\right)^{\alpha-1}\exp\left(-\frac{(\alpha-1)}{2}+\frac{\sqrt{\alpha}B_\alpha}{2}\right), \quad (5.51)
\end{aligned}
$$

because B_α satisfies $2B_\alpha^2-\sqrt{\alpha}B_\alpha=(\alpha-1)$. Now consider that

$$
\begin{aligned}
\left(\frac{B_\alpha}{\sqrt{\alpha}}\right)^{\alpha-1} = \left(\frac{\sqrt{\alpha}+\sqrt{9\alpha-8}}{4\sqrt{\alpha}}\right)^{\alpha-1} &= \left(1-\frac{\sqrt{9\alpha}-\sqrt{9\alpha-8}}{4\sqrt{\alpha}}\right)^{\alpha-1}\\
&= \left(1-\frac{2}{\sqrt{\alpha}(\sqrt{9\alpha}+\sqrt{9\alpha-8})}\right)^{\alpha-1}\\
&= \left(1-\frac{2}{(3\alpha+\sqrt{\alpha(9\alpha-8)})}\right)^{\alpha-1} .\quad(5.52)
\end{aligned}
$$

Therefore

$$
\lim_{\alpha\to\infty}\left(\frac{B_\alpha}{\sqrt{\alpha}}\right)^{\alpha-1}=\exp\left(-\lim_{\alpha\to\infty}\frac{2(\alpha-1)}{(3\alpha+\sqrt{\alpha(9\alpha-8)})}\right)=\exp(-\frac{1}{3}).\quad(5.53)
$$

On the other hand

$$
\begin{aligned}
\lim_{\alpha\to\infty}-\frac{(\alpha-1)}{2}+\frac{\sqrt{\alpha}B_\alpha}{2} &= \lim_{\alpha\to\infty}\frac{1}{2}-\frac{\alpha}{2}+\frac{\alpha+\sqrt{\alpha(9\alpha-8)}}{8}\\
&= \lim_{\alpha\to\infty}\frac{1}{2}-\sqrt{\alpha}\left(\frac{3\sqrt{\alpha}-\sqrt{\alpha(9\alpha-8)}}{8}\right)\\
&= \lim_{\alpha\to\infty}\frac{1}{2}-\left(\frac{\sqrt{\alpha}}{(3\sqrt{\alpha}+\sqrt{\alpha(9\alpha-8)})}\right)\\
&= \frac{1}{2}-\frac{1}{6}\\
&= \frac{1}{3}.\quad(5.54)
\end{aligned}
$$

Hence, it follows from Equations 5.51, 5.53, 5.54 that

$$\lim_{\alpha \to \infty} \frac{K_1(\alpha, \sqrt{\alpha})}{K_2(\alpha, \sqrt{\alpha})} = 1. \tag{5.55}$$

From Equation 5.41, we get that

$$\log(K_1(\alpha, \sqrt{\alpha})) = (\alpha-1)\log(B_\alpha)+B_\alpha^2-(\alpha-1)-\log\left({}_1\Psi_1\left[\begin{matrix}(\frac{\alpha}{2}, \frac{1}{2}) \\ (1, 0)\end{matrix}; \sqrt{\alpha}\right]\right)+\log(2\sqrt{\pi}).$$

On the other hand,

$$\log(K_2(\alpha, \sqrt{\alpha})) = \log\left(\frac{2^{\frac{\alpha}{2}}\Gamma(\frac{\alpha}{2})\exp(\frac{\alpha}{2})}{{}_1\Psi_1\left[\begin{matrix}(\frac{\alpha}{2}, \frac{1}{2}) \\ (1, 0)\end{matrix}; \sqrt{\alpha}\right]}\right).$$

Therefore,

$$\begin{aligned}
&\frac{d}{d\alpha}\log(\frac{K_1(\alpha, \sqrt{\alpha})}{K_2(\alpha, \sqrt{\alpha})}) \\
&= \frac{d}{d\alpha}\log(K_1(\alpha, \sqrt{\alpha})) - \frac{d}{d\alpha}\log(K_2(\alpha, \sqrt{\alpha})) \\
&= \log(B_\alpha) + \frac{B_\alpha}{2\sqrt{\alpha}} - \frac{d}{d\alpha}\left(\frac{\alpha}{2}\log(2) + \log(\Gamma(\frac{\alpha}{2})) + \frac{\alpha}{2}\right) \\
&= \log(B_\alpha) + \frac{B_\alpha}{2\sqrt{\alpha}} - \left(\frac{1}{2}\log(2) + \frac{1}{2}\frac{\frac{d}{d\alpha}(\Gamma(\frac{\alpha}{2}))}{\Gamma(\frac{\alpha}{2})} + \frac{1}{2}\right) \\
&\leq \log(B_\alpha) + \frac{B_\alpha}{2\sqrt{\alpha}} - \left(\frac{1}{2}\log(2) + \frac{1}{2}\left(\log(\frac{\alpha}{2}) - \frac{1}{\alpha} - \frac{1}{3\alpha^2}\right) + \frac{1}{2}\right) \tag{5.56}
\end{aligned}$$

because $\log(\frac{\alpha}{2}) - \frac{1}{\alpha} - \frac{1}{3\alpha^2} \leq \frac{\frac{d}{d\alpha}(\Gamma(\frac{\alpha}{2}))}{\Gamma(\frac{\alpha}{2})} \leq \log(\frac{\alpha}{2}) - \frac{1}{\alpha}$ that follows from an inequality in (Batir, 2005). Therefore,

$$
\begin{aligned}
\frac{d}{d\alpha} \log \left(\frac{K_1(\alpha, \sqrt{\alpha})}{K_2(\alpha, \sqrt{\alpha})} \right) &\leq \log \left(\frac{B_\alpha}{\sqrt{\alpha}} \right) + \frac{B_\alpha}{2\sqrt{\alpha}} + \frac{1}{2\alpha} + \frac{1}{6\alpha^2} - \frac{1}{2} \\
&\leq 0. \qquad\qquad (5.57)
\end{aligned}
$$

In order to establish the last inequality, we define $p(\alpha) = \log \left(\frac{B_\alpha}{\sqrt{\alpha}} \right) + \frac{B_\alpha}{2\sqrt{\alpha}} + \frac{1}{2\alpha} + \frac{1}{6\alpha^2} - \frac{1}{2}$. Note that $p(\alpha)$ is an increasing function as its derivative $\frac{dp(\alpha)}{\alpha} = \frac{1}{B_\alpha \alpha \sqrt{\alpha(9\alpha-8)}} + \frac{1}{2\alpha\sqrt{\alpha(9\alpha-8)}} - \frac{1}{2\alpha^2} - \frac{1}{3\alpha^3} > 0$ and $\lim_{\alpha \to \infty} p(\alpha) = 0$. An implication of Equation 5.57 is that the function $\alpha \mapsto \frac{K_1(\alpha, \sqrt{\alpha})}{K_2(\alpha, \sqrt{\alpha})}$ is decreasing in $\alpha > 1$. As a result, it follows from Equation 5.57 and Equation 5.55

$$
\frac{K_1(\alpha, \sqrt{\alpha})}{K_2(\alpha, \sqrt{\alpha})} \geq \lim_{\alpha \to \infty} \frac{K_1(\alpha, \sqrt{\alpha})}{K_2(\alpha, \sqrt{\alpha})} = 1 \text{ for all } \alpha > 1.
$$

Consequently, $K_1(\alpha, \sqrt{\alpha}) \geq K_2(\alpha, \sqrt{\alpha})$ for all $\alpha > 1$.

Proof of part(d), Theorem 2

To prove $K_1(\alpha, \sqrt{\alpha})$ is decreasing. From the definition of $K_1(\alpha, \Delta)$, we get that

$$
K_1(\alpha, \sqrt{\alpha}) = \frac{2\sqrt{\pi} \, (B_\alpha)^{\alpha-1} \, \exp\left(-(\alpha-1) + B_\alpha^2\right)}{{}_1\Psi_1 \left[\begin{matrix} (\frac{\alpha}{2}, \frac{1}{2}) \\ (1, 0) \end{matrix} ; \sqrt{\alpha} \right]}
$$

$$(5.58)$$

where $B_\alpha = \frac{\sqrt{\alpha}+\sqrt{9\alpha-8}}{4}$. As

$$
\begin{aligned}
(\alpha - 1) &+ 2B_\alpha^2 \\
&= (\alpha - 1) + \left(\frac{5\alpha - 4 + \sqrt{\alpha(9\alpha - 8)}}{4} \right) \\
&= \left(\frac{9\alpha - 8 + \sqrt{\alpha(9\alpha - 8)}}{4} \right) \\
&= (\sqrt{9\alpha - 8}) B_\alpha,
\end{aligned}
$$

it turns out that

$$
\frac{\partial B_\alpha}{\partial \Delta} = \frac{1}{4}\left[\frac{1}{2\sqrt{\alpha}} + \frac{9}{2\sqrt{9\alpha - 8}} \right] = \frac{1}{8}\left[\frac{\sqrt{9\alpha - 8} + 9\sqrt{\alpha}}{\sqrt{\alpha(9\alpha - 8)}} \right] = \frac{B_\alpha}{2\sqrt{\alpha(9\alpha - 8)}} + \frac{1}{\sqrt{(9\alpha - 8)}}.
$$

Observe that

$$
\log(K_1(\alpha, \Delta)) = (\alpha - 1)\log(B_\alpha) + B_\alpha^2 - (\alpha - 1) - \log\left({}_1\Psi_1 \left[\begin{matrix} \left(\frac{\alpha}{2}, \frac{1}{2}\right) \\ (1, 0) \end{matrix}; \sqrt{\alpha} \right] \right) + \log(2\sqrt{\pi}).
$$

Therefore,

132

$$\frac{\partial \log(K_1(\alpha, \Delta))}{\partial \alpha}$$

$$= \log(B_\alpha) - 1 + \left\{\frac{(\alpha-1)}{B_\alpha} + 2B_\alpha\right\}\frac{\partial B_\alpha}{\partial \alpha} - \frac{\frac{\partial}{\partial \alpha}\left\{{}_1\Psi_1\begin{bmatrix}(\frac{\alpha}{2}, \frac{1}{2}) \\ (1,0)\end{bmatrix}; \sqrt{\alpha}\end{bmatrix}\right\}}{{}_1\Psi_1\begin{bmatrix}(\frac{\alpha}{2}, \frac{1}{2}) \\ (1,0)\end{bmatrix}; \sqrt{\alpha}\end{bmatrix}}$$

$$= \log(B_\alpha) - 1 + \left\{\frac{(\alpha-1)}{B_\alpha} + 2B_\alpha\right\}\frac{\partial B_\alpha}{\partial \alpha} - \frac{\frac{\partial}{\partial \alpha}\left\{{}_1\Psi_1\begin{bmatrix}(\frac{\alpha}{2}, \frac{1}{2}) \\ (1,0)\end{bmatrix}; \sqrt{\alpha}\end{bmatrix}\right\}}{{}_1\Psi_1\begin{bmatrix}(\frac{\alpha}{2}, \frac{1}{2}) \\ (1,0)\end{bmatrix}; \sqrt{\alpha}\end{bmatrix}}$$

$$= \log(B_\alpha) - 1 + \frac{(\alpha-1) + 2B_\alpha^2}{B_\alpha}\left(\frac{B_\alpha}{2\sqrt{\alpha(9\alpha-8)}} + \frac{1}{\sqrt{(9\alpha-8)}}\right) - \frac{\frac{\partial}{\partial \alpha}\left\{{}_1\Psi_1\begin{bmatrix}(\frac{\alpha}{2}, \frac{1}{2}) \\ (1,0)\end{bmatrix}; \sqrt{\alpha}\end{bmatrix}\right\}}{{}_1\Psi_1\begin{bmatrix}(\frac{\alpha}{2}, \frac{1}{2}) \\ (1,0)\end{bmatrix}; \sqrt{\alpha}\end{bmatrix}}$$

$$\overset{(\dagger)}{=} \log(B_\alpha) - 1 + \frac{(\sqrt{9\alpha-8})B_\alpha}{B_\alpha}\left(\frac{B_\alpha}{2\sqrt{\alpha(9\alpha-8)}} + \frac{1}{\sqrt{(9\alpha-8)}}\right) - \frac{\frac{\partial}{\partial \alpha}\left\{{}_1\Psi_1\begin{bmatrix}(\frac{\alpha}{2}, \frac{1}{2}) \\ (1,0)\end{bmatrix}; \sqrt{\alpha}\end{bmatrix}\right\}}{{}_1\Psi_1\begin{bmatrix}(\frac{\alpha}{2}, \frac{1}{2}) \\ (1,0)\end{bmatrix}; \sqrt{\alpha}\end{bmatrix}}$$

$$= \log(B_\alpha) - 1 + \left(\frac{B_\alpha}{2\sqrt{\alpha}} + 1\right) - \frac{\frac{\partial}{\partial \alpha}\left\{{}_1\Psi_1\begin{bmatrix}(\frac{\alpha}{2}, \frac{1}{2}) \\ (1,0)\end{bmatrix}; \sqrt{\alpha}\end{bmatrix}\right\}}{{}_1\Psi_1\begin{bmatrix}(\frac{\alpha}{2}, \frac{1}{2}) \\ (1,0)\end{bmatrix}; \sqrt{\alpha}\end{bmatrix}}$$

$$= \log(B_\alpha) + \frac{B_\alpha}{2\sqrt{\alpha}} - \frac{\frac{\partial}{\partial \alpha}\left\{{}_1\Psi_1\begin{bmatrix}(\frac{\alpha}{2}, \frac{1}{2}) \\ (1,0)\end{bmatrix}; \sqrt{\alpha}\end{bmatrix}\right\}}{{}_1\Psi_1\begin{bmatrix}(\frac{\alpha}{2}, \frac{1}{2}) \\ (1,0)\end{bmatrix}; \frac{133}{\sqrt{\alpha}}\end{bmatrix}} \qquad (5.59)$$

where the equality in (†) is due to Equation 5.59. On the other hand,

$$
\frac{\frac{\partial}{\partial \alpha}\left\{ {}_1\Psi_1\left[\begin{matrix} \left(\frac{\alpha}{2},\frac{1}{2}\right) \\ (1,0) \end{matrix}; \sqrt{\alpha} \right] \right\}}{{}_1\Psi_1\left[\begin{matrix} \left(\frac{\alpha}{2},\frac{1}{2}\right) \\ (1,0) \end{matrix}; \sqrt{\alpha} \right]} = E(\log(X)) + \frac{E(X)}{2\sqrt{\alpha}} \tag{5.60}
$$

where $X \sim \mathrm{MHN}\left(\alpha, 1, \sqrt{\alpha}\right)$. Therefore, it follows from Equation 5.59 that

$$
\begin{aligned}
\frac{\partial \log(K_1(\alpha, \Delta))}{\partial \alpha} &= \log(B_\alpha) + \frac{B_\alpha}{2\sqrt{\alpha}} - E(\log(X)) - \frac{E(X)}{2\sqrt{\alpha}} \\
&= -[E(\log(X)) - \log(B_\alpha)] - \frac{[E(X) - B_\alpha]}{2\sqrt{\alpha}} \\
&= -E\left(\log\left(\frac{X}{B_\alpha}\right)\right) - \frac{[E(X) - B_\alpha]}{2\sqrt{\alpha}} \\
&< 0, \tag{5.61}
\end{aligned}
$$

when $\alpha \geq 4$. The last inequality in Equation 5.61 follows from part(a) and part(b) of the Lemma 8 as B_α is also the mode of the MHN $(\alpha, 1, \sqrt{\alpha})$ distribution.

Proof of part(e), Theorem 2

Result 2. *For arbitrary $\alpha > 1$, $\inf\limits_{\Delta > 0} \max\left\{ \frac{1}{K_1(\alpha, \Delta)}, \frac{1}{K_2(\alpha, \Delta)} \right\} \geq \sup\limits_{\Delta > 0} \min\left\{ \frac{1}{K_1(\alpha, \Delta)}, \frac{1}{K_2(\alpha, \Delta)} \right\}$.*

Proof of Result 2: According to the part(b) of the Theorem 2, the functions $\Delta \mapsto K_1(\alpha, \Delta)$ is non-increasing and $\Delta \mapsto K_2(\alpha, \Delta)$ is non-decreasing for all $\alpha > 1$. Therefore,

$$
\left(\frac{1}{K_1(\alpha, \Delta_1)} - \frac{1}{K_1(\alpha, \Delta_2)} \right)\left(\frac{1}{K_2(\alpha, \Delta_1)} - \frac{1}{K_2(\alpha, \Delta_2)} \right) \leq 0
$$

for arbitrary $\Delta_1, \Delta_2 > 0$ and $\alpha > 1$. If the statement of the result were not true then

$$\inf_{\Delta>0} \max \left\{ \frac{1}{K_1(\alpha, \Delta)}, \frac{1}{K_2(\alpha, \Delta)} \right\} < \sup_{\Delta>0} \min \left\{ \frac{1}{K_1(\alpha, \Delta)}, \frac{1}{K_2(\alpha, \Delta)} \right\}.$$

In that case, there would exist $\Delta_1, \Delta_2 > 0, \alpha > 1$ such that

$$\max \left\{ \frac{1}{K_1(\alpha, \Delta_1)}, \frac{1}{K_2(\alpha, \Delta_1)} \right\} < \min \left\{ \frac{1}{K_1(\alpha, \Delta_2)}, \frac{1}{K_2(\alpha, \Delta_2)} \right\}$$

$$\implies \begin{cases} \frac{1}{K_1(\alpha,\Delta_1)} < \min \left\{ \frac{1}{K_1(\alpha,\Delta_2)}, \frac{1}{K_2(\alpha,\Delta_2)} \right\} \quad \text{and ,} \\[2mm] \frac{1}{K_2(\alpha,\Delta_1)} < \min \left\{ \frac{1}{K_1(\alpha,\Delta_2)}, \frac{1}{K_2(\alpha,\Delta_2)} \right\} \end{cases}$$

$$\implies \begin{cases} \frac{1}{K_1(\alpha,\Delta_1)} < \frac{1}{K_1(\alpha,\Delta_2)} \quad \text{and ,} \\[2mm] \frac{1}{K_2(\alpha,\Delta_1)} < \frac{1}{K_2(\alpha,\Delta_2)} \end{cases}$$

$$\implies \left(\frac{1}{K_1(\alpha, \Delta_1)} - \frac{1}{K_1(\alpha, \Delta_2)} \right) \left(\frac{1}{K_2(\alpha, \Delta_1)} - \frac{1}{K_2(\alpha, \Delta_2)} \right) > 0, \quad (5.62)$$

which would contradict the fact in Equation 5.62. Therefore the statement of the result holds true.

Proof of part(e), Theorem 2:

From part(a) of the Theorem 2, we have $\mathcal{A}_{\text{pos}}(\alpha, \Delta) = \max \left\{ \frac{1}{K_1(\alpha,\Delta)}, \frac{1}{K_2(\alpha,\Delta)} \right\}$. An implication of the Part(b) of the Theorem 2 is that, for any $\alpha > 1$ the functions $\Delta \mapsto \frac{1}{K_1(\alpha,\Delta)}$ and $\Delta \mapsto \frac{1}{K_2(\alpha,\Delta)}$ are increasing and decreasing in $\Delta > 0$ correspondingly. On the other hand, Result 2 implicates that

$$\max \left\{ \frac{1}{K_1(\alpha, \Delta)}, \frac{1}{K_2(\alpha, \Delta)} \right\} \geq \min \left\{ \frac{1}{K_1(\alpha, \Delta_\star)}, \frac{1}{K_2(\alpha, \Delta_\star)} \right\} \quad (5.63)$$

for arbitrary $\Delta > 0$ and $\Delta_\star > 0$.

In particular, if we choose $\Delta_\star$ to be $\sqrt{\alpha}$ then for any $\Delta > 0$,

$$\max\left\{\frac{1}{K_1(\alpha, \Delta)}, \frac{1}{K_2(\alpha, \Delta)}\right\} \geq \min\left\{\frac{1}{K_1(\alpha, \sqrt{\alpha})}, \frac{1}{K_2(\alpha, \sqrt{\alpha})}\right\} = \frac{1}{K_1(\alpha, \sqrt{\alpha})}, \tag{5.64}$$

where the last equality is due to the Part(c) of the Theorem 2 which conveys that $\frac{1}{K_1(\alpha, \sqrt{\alpha})} \leq \frac{1}{K_2(\alpha, \sqrt{\alpha})}$ for all $\alpha \geq 4$. Additionally, from the part(d) of Theorem 2 we infer that $\frac{1}{K_1(4, \sqrt{4})} \leq \frac{1}{K_1(\alpha, \sqrt{\alpha})}$ for all $\alpha \geq 4$. Altogether, we conclude that

$$\mathcal{A}_{\text{pos}}(\alpha, \Delta) = \max\left\{\frac{1}{K_1(\alpha, \Delta)}, \frac{1}{K_2(\alpha, \Delta)}\right\} \geq \frac{1}{K_1(\alpha, \sqrt{\alpha})} \geq \frac{1}{K_1(4, 2)} \geq 0.8$$

for all $\Delta > 0$ and $\alpha \geq 4$. As the point $\Delta > 0$ is arbitrary, we can write

$$\mathcal{A}_{\text{pos}}(\alpha, \frac{\gamma}{\sqrt{\beta}}) \geq \frac{1}{K_1(\alpha, \sqrt{\alpha})} \geq \frac{1}{K_1(4, 2)} \geq 0.8 \text{ for all } \alpha \geq 4, \beta > 0, \gamma > 0.$$

5.0.13   Proof of Lemma 11

$$
\begin{aligned}
f_{\text{MHN}}(x \mid \alpha, \beta, \gamma) &= \frac{2\beta^{\frac{\alpha}{2}}}{{}_1\Psi_1\left[\begin{matrix}(\frac{\alpha}{2}, \frac{1}{2}), \\ (1, 0)\end{matrix}; \frac{\gamma}{\sqrt{\beta}}\right]} x^{\alpha-1} e^{\gamma x} e^{-\beta x^2} \\[2ex]
&= \frac{2\beta^{\frac{\alpha}{2}}}{{}_1\Psi_1\left[\begin{matrix}(\frac{\alpha}{2}, \frac{1}{2}), \\ (1, 0)\end{matrix}; \frac{\gamma}{\sqrt{\beta}}\right]} x^{\alpha-1} \sum_{i=0}^{\infty} \frac{(\gamma x)^i}{i!} e^{-\beta x^2} \\[2ex]
&= \sum_{i=0}^{\infty} \frac{2\beta^{\frac{\alpha}{2}}}{{}_1\Psi_1\left[\begin{matrix}(\frac{\alpha}{2}, \frac{1}{2}), \\ (1, 0)\end{matrix}; \frac{\gamma}{\sqrt{\beta}}\right]} \frac{\gamma^i}{i!} \frac{\Gamma(\frac{\alpha+i}{2})}{2\beta^{\frac{\alpha+i}{2}}} \frac{2\beta^{\frac{\alpha+i}{2}}}{\Gamma(\frac{\alpha+i}{2})} x^{\alpha+i-1} e^{-\beta x^2} \\[2ex]
&= \sum_{i=0}^{\infty} p_i f_i(x|\alpha, \beta), \tag{5.65}
\end{aligned}
$$

where $p_i = \dfrac{\Gamma(\frac{\alpha+i}{2})(\frac{\gamma}{\sqrt{\beta}})^i}{{}_1\Psi_1\left[\begin{array}{c}(\frac{\alpha}{2},\frac{1}{2})\\[2pt] ;\frac{\gamma}{\sqrt{\beta}}\\[2pt](1,0)\end{array}\right]i!}$ and $f_i(x|\alpha,\beta) = \dfrac{2\beta^{\frac{\alpha+i}{2}}}{\Gamma(\frac{\alpha+i}{2})}x^{\alpha+i-1}e^{-\beta x^2}$. Here $f_i(x|\alpha,\beta)$ de-

notes the density of the square root of the Gamma distribution.

5.0.14 Proof of Lemma 12

Proof of Lemma 12 (a)

$$q_i = \frac{\Gamma(\frac{\alpha+i}{2})\left(\frac{\gamma}{\sqrt{\beta}}\right)^i}{i!} \text{ for } i \geq 0 \text{ where } \alpha \geq 1, \gamma > 0, \beta > 0. \text{ For an arbitrary positive}$$

number $\epsilon_1 > 0$, suppose $M^\dagger = \max\left\{[\alpha], \left[\frac{\gamma^2}{\epsilon_1^2\beta}\right]\right\}$ where $[\alpha]$ and $\left[\frac{\gamma^2}{\epsilon_1^2\beta}\right]$ denotes the

largest integer less than equal to α and $\frac{\gamma^2}{\epsilon_1^2\beta}$ respectively.

$$\frac{q_{M+1}}{q_M} = \frac{\frac{\Gamma(\frac{\alpha+M+1}{2})\left(\frac{\gamma}{\sqrt{\beta}}\right)^{M+1}}{(M+1)!}}{\frac{\Gamma(\frac{\alpha+M}{2})\left(\frac{\gamma}{\sqrt{\beta}}\right)^{M}}{M!}} = \frac{\gamma\Gamma(\frac{\alpha+M+1}{2})}{\sqrt{\beta}(M+1)\Gamma(\frac{\alpha+M}{2})}, \tag{5.66}$$

Using the fact that $\frac{\Gamma(x+1)}{\Gamma(x+\frac{1}{2})} \leq \sqrt{x + \frac{1}{4} + \left[\frac{\Gamma(\frac{3}{4})}{\Gamma(\frac{1}{4})}\right]^2}$ for $x \geq -\frac{1}{2}$ (see (Kazarinoff, 1956; Watson, 1959; Qi et al., 2012)) we get that,

$$
\begin{aligned}
\frac{q_{M+1}}{q_M} &\leq \frac{\gamma\sqrt{\frac{\alpha+M-1}{2} + \frac{1}{4} + \left[\frac{\Gamma(\frac{3}{4})}{\Gamma(\frac{1}{4})}\right]^2}}{\sqrt{\beta}(M+1)} = \frac{\gamma\sqrt{\frac{M}{2} + \left\{\frac{\alpha-1}{2} + \frac{1}{4} + \left[\frac{\Gamma(\frac{3}{4})}{\Gamma(\frac{1}{4})}\right]^2\right\}}}{\sqrt{\beta}(M+1)} \\[10pt]
&\leq \frac{\gamma\sqrt{\frac{M}{2} + \left\{\frac{\alpha-1}{2} + \frac{1}{2}\right\}}}{\sqrt{\beta}(M+1)} \\[10pt]
&= \frac{\gamma\sqrt{\frac{M}{2} + \frac{\alpha}{2}}}{\sqrt{\beta}(M+1)} \\[10pt]
&\overset{(ix)}{\leq} \frac{\gamma\sqrt{M+1}}{\sqrt{\beta}(M+1)} \\[10pt]
&= \frac{\gamma}{\sqrt{\beta}(M+1)}, \tag{5.67}
\end{aligned}
$$

where the inequality in (ix) follows from the assumption that $M \geq [\alpha]$. From equation 5.67 and the assumption $M \geq \left[\frac{\gamma^2}{\epsilon_1^2 \beta}\right]$, it follows that

$$M + 1 \geq \frac{\gamma^2}{\epsilon_1^2 \beta} \implies \frac{q_{M+1}}{q_M} \leq \epsilon_1.$$

Proof of Lemma 12 (b)

$$
\begin{aligned}
\frac{q_{i+2}q_i}{q_{i+1}^2} &= \frac{\frac{\Gamma(\frac{\alpha+i+2}{2})\left(\frac{\gamma}{\sqrt{\beta}}\right)^{i+2}}{(i+2)!} \frac{\Gamma(\frac{\alpha+i}{2})\left(\frac{\gamma}{\sqrt{\beta}}\right)^{i}}{i!}}{\left[\frac{\Gamma(\frac{\alpha+i+1}{2})\left(\frac{\gamma}{\sqrt{\beta}}\right)^{i+1}}{(i+1)!}\right]^2} \\
&= \frac{\frac{\Gamma(\frac{\alpha+i+2}{2})}{(i+2)!} \frac{\Gamma(\frac{\alpha+i}{2})}{i!}}{\left[\frac{\Gamma(\frac{\alpha+i+1}{2})}{(i+1)!}\right]^2} \\
&= \frac{[(i+1)!]^2}{(i+2)!(i)!} \frac{\Gamma(\frac{\alpha+i+2}{2})\Gamma(\frac{\alpha+i}{2})}{\left[\Gamma(\frac{\alpha+i+1}{2})\right]^2} \\
&= \frac{(i+1)}{(i+2)} \frac{\Gamma(\frac{\alpha+i+2}{2})\Gamma(\frac{\alpha+i}{2})}{\left[\Gamma(\frac{\alpha+i+1}{2})\right]^2} \\
&= \frac{(i+1)}{(i+2)} \frac{\Gamma(\frac{\alpha+i}{2} + 1)}{\Gamma(\frac{\alpha+i}{2} + \frac{1}{2})} \frac{\Gamma(\frac{\alpha+i-1}{2} + \frac{1}{2})}{\Gamma(\frac{\alpha+i-1}{2} + 1)}.
\end{aligned}
\tag{5.68}
$$

From Watson's monotonicity properties (Kazarinoff, 1956; Watson, 1959; Qi et al., 2012) of the ratio of Gamma functions, we have the double inequality

$$\sqrt{x + \frac{1}{4}} \leq \frac{\Gamma(x+1)}{\Gamma(x+\frac{1}{2})} \leq \sqrt{x + \frac{1}{4} + \left[\frac{\Gamma(\frac{3}{4})}{\Gamma(\frac{1}{4})}\right]^2} \quad \text{for } x \geq -\frac{1}{2}. \tag{5.69}$$

Using equation 5.69 in Equation 5.68 we get that

$$\frac{q_{i+2}q_i}{q_{i+1}^2} \le \frac{i+1}{i+2} \frac{\sqrt{\frac{\alpha+i}{2} + \frac{1}{4} + \left[\frac{\Gamma(\frac{3}{4})}{\Gamma(\frac{1}{4})}\right]^2}}{\sqrt{\frac{\alpha+i-1}{2} + \frac{1}{4}}}$$

$$\le \frac{i+1}{i+2} \frac{\sqrt{\frac{1+i}{2} + \frac{1}{4} + \left[\frac{\Gamma(\frac{3}{4})}{\Gamma(\frac{1}{4})}\right]^2}}{\sqrt{\frac{i}{2} + \frac{1}{4}}}$$

$$\le \frac{i+1}{i+2} \frac{\sqrt{1 + i + \frac{1}{2} + 2\left[\frac{\Gamma(\frac{3}{4})}{\Gamma(\frac{1}{4})}\right]^2}}{\sqrt{i + \frac{1}{2}}}$$

$$\le \frac{i+1}{i+2} \frac{\sqrt{i + \frac{3}{2} + 2\left[\frac{\Gamma(\frac{3}{4})}{\Gamma(\frac{1}{4})}\right]^2}}{\sqrt{i + \frac{1}{2}}}. \tag{5.70}$$

As $\left[\frac{\Gamma(\frac{3}{4})}{\Gamma(\frac{1}{4})}\right]^2 < \frac{1}{4}$ we get that

$$\frac{q_{i+2}q_i}{q_{i+1}^2} < \frac{i+1}{i+2} \frac{\sqrt{i+2}}{\sqrt{i + \frac{1}{2}}} = \frac{i+1}{\sqrt{(i+2)(i+\frac{1}{2})}} = \frac{i+1}{\sqrt{(i+1)^2 + \frac{i}{2}}} \le 1$$

for all $i \ge 0$.

Proof of Lemma 12 (c)

Let $M \ge M^\dagger$, then using part (a) we get that

$$\frac{q_{M+1}}{q_M} \le \epsilon_1, \tag{5.71}$$

while from part (b) it follows that

$$\frac{q_{M+1}}{q_M} > \frac{q_{M+2}}{q_{M+1}} > \ldots$$

Consequently, $q_{M+1} < q_M \epsilon_1, \ldots, q_{j+1} \leq q_M \epsilon_1^{j-M}$ for $j > M$. As a result

$$\sum_{j=M}^{\infty} q_j = \sum_{j=M}^{\infty} \epsilon_1^{j-M} q_M = q_M \sum_{j=0}^{\infty} \epsilon_1^j = \frac{q_M}{1 - \epsilon_1}.$$

part(d): From the definition of the the Fox-Wright function we have

$$\sum_{i=1}^{\infty} q_i = \sum_{i=0}^{\infty} \frac{\Gamma(\frac{\alpha+i}{2}) \left(\frac{\gamma}{\sqrt{\beta}}\right)^i}{i!} {}_1\Psi_1 \left[\begin{array}{c} (\frac{\alpha}{2}, \frac{1}{2}), \\ (1, 0) \end{array}; \frac{\gamma}{\sqrt{\beta}} \right] < \infty,$$

which is a (absolutely) convergent series. Hence $\lim_{i \to \infty} q_i = 0$.

5.0.15 Proof of Theorem 3

We start by stating the following theorem that we need in the subsequent proof.

Theorem: (AM-GM inequality with non-uniform weights (Steele, 2004))

Let $p_1, p_2, \cdots, p_n$ be positive real numbers such that $\sum_{i=1}^{n} p_i = 1$. If $a_1, a_2 \cdots, a_n$ be arbitrary positive real numbers then

$$\sum_{i=1}^{n} p_i a_i \geq \prod_{i=1}^{n} a_i^{p_i}.$$

Now consider that

$$
\begin{aligned}
f_{\mathrm{MHN}}(x|\alpha,\beta,\gamma) &= \frac{2\beta^{\frac{\alpha}{2}}}{{}_1\Psi_1\!\left[\begin{array}{c}\left(\frac{\alpha}{2},\frac{1}{2}\right)\\ (1,0)\end{array};\frac{\gamma}{\sqrt{\beta}}\right]}\, x^{\alpha-1}e^{\gamma x}e^{-\beta x^2}\\[2mm]
&= \frac{2\beta^{\frac{\alpha}{2}}}{{}_1\Psi_1\!\left[\begin{array}{c}\left(\frac{\alpha}{2},\frac{1}{2}\right)\\ (1,0)\end{array};\frac{\gamma}{\sqrt{\beta}}\right]}\, x^{\alpha-1}\exp\!\left(-\beta m^2(x/m)^2-|\gamma|m(x/m)\right)\\[2mm]
&= \frac{2\beta^{\frac{\alpha}{2}}}{{}_1\Psi_1\!\left[\begin{array}{c}\left(\frac{\alpha}{2},\frac{1}{2}\right)\\ (1,0)\end{array};\frac{\gamma}{\sqrt{\beta}}\right]}\, x^{\alpha-1}\exp\!\left(-(\beta m^2+|\gamma|m)(\tfrac{\beta m^2}{\beta m^2+|\gamma|m}(x/m)^2+\tfrac{|\gamma|m}{\beta m^2+|\gamma|m}(x/m)\right) \quad (5.72)
\end{aligned}
$$

Applying the AM-GM Inequality with non uniform Weights(that we stated above) we get that

$$
f_{\mathrm{MHN}}(x\mid\alpha,\beta,\gamma)\le g_{\mathrm{kernel}}(x;\alpha,\beta,\gamma) \tag{5.73}
$$

where $g_{\mathrm{kernel}}(x|\alpha,\beta,\gamma)=K_0(m,\alpha,\beta,\gamma)x^{\alpha-1}\ \exp\left(-(\beta m^2+|\gamma|m)(\tfrac{x}{m})^{\frac{2\beta m+|\gamma|}{\beta m+|\gamma|}}\right),$

$$
K_0(m,\alpha,\beta,\gamma)=\frac{2\beta^{\frac{\alpha}{2}}}{{}_1\Psi_1\!\left[\begin{array}{c}\left(\frac{\alpha}{2},\frac{1}{2}\right)\\ (1,0)\end{array};\frac{\gamma}{\sqrt{\beta}}\right]}.
$$

5.0.16 Proof of Theorem 4

Proof of part(a), Theorem 4:

The proposal kernel for the Accept-Reject algorithm is (see Theorem 3)

$$g_{\text{kernel}}(x \mid \alpha, \beta, \gamma) = \frac{2\beta^{\frac{\alpha}{2}}}{{}_1\Psi_1\left[\begin{array}{c}\left(\frac{\alpha}{2}, \frac{1}{2}\right) \\ (1,0)\end{array}; \frac{\gamma}{\sqrt{\beta}}\right]} x^{\alpha-1} \exp\left(-(\beta m^2 + |\gamma|m)(\frac{x}{m})^{\frac{2\beta m + |\gamma|}{\beta m + |\gamma|}}\right).$$

Therefore,

$$\int_0^\infty g_{\text{kernel}}(x|\alpha, \beta, \gamma)dx = \frac{2\beta^{\frac{\alpha}{2}}}{{}_1\Psi_1\left[\begin{array}{c}\left(\frac{\alpha}{2}, \frac{1}{2}\right) \\ (1,0)\end{array}; \frac{\gamma}{\sqrt{\beta}}\right]} \int_0^\infty x^{\alpha-1} \exp\left(-(\beta m^2 + |\gamma|m)(\frac{x}{m})^{\frac{2\beta m + |\gamma|}{\beta m + |\gamma|}}\right) dx.$$

Applying the change of variable $y = \left(\frac{x}{m}\right)^{\frac{2\beta m + |\gamma|}{\beta m + |\gamma|}}$, we get that

$$\int_0^\infty g(x|\alpha, \beta, \gamma)dx = \frac{2\beta^{\frac{\alpha}{2}}}{{}_1\Psi_1\left[\begin{array}{c}\left(\frac{\alpha}{2}, \frac{1}{2}\right) \\ (1,0)\end{array}; \frac{\gamma}{\sqrt{\beta}}\right]} m^\alpha \frac{\beta m + |\gamma|}{2\beta m + |\gamma|} \int_0^\infty y^{\frac{\alpha(\beta m + |\gamma|)}{2\beta m + |\gamma|} - 1} \exp[-(\beta m^2 + |\gamma|m)y]dy$$

$$= \frac{2\beta^{\frac{\alpha}{2}}}{{}_1\Psi_1\left[\begin{array}{c}\left(\frac{\alpha}{2}, \frac{1}{2}\right) \\ (1,0)\end{array}; \frac{\gamma}{\sqrt{\beta}}\right]} m^\alpha \frac{\beta m + |\gamma|}{2\beta m + |\gamma|} \frac{\Gamma(\frac{\alpha(\beta m + |\gamma|)}{2\beta m + |\gamma|})}{[m(\beta m + |\gamma|)]^{\frac{\alpha(\beta m + |\gamma|)}{2\beta m + |\gamma|}}}.$$

Then the acceptance rate

$$\mathcal{A}_{\mathrm{neg}}(m, \alpha, \beta, \gamma) = \frac{1}{\int_0^\infty g_{\mathrm{kernel}}(x|\alpha, \beta, \gamma) dx}$$

$$= \frac{\frac{1}{2(\beta m^2)^{\frac{\alpha}{2}}} {}_1\Psi_1 \left[\begin{array}{c} (\frac{\alpha}{2}, \frac{1}{2}) \\ (1, 0) \end{array}; \frac{-|\gamma|}{\sqrt{\beta}} \right]}{\frac{\beta m + |\gamma|}{2\beta m + |\gamma|} \frac{\Gamma(\frac{\alpha(\beta m + |\gamma|)}{2\beta m + |\gamma|})}{[m(\beta m + |\gamma|)]^{\frac{\alpha(\beta m + |\gamma|)}{2\beta m + |\gamma|}}}}.$$

Proof of part(b), Theorem 4:

$$\frac{\partial}{\partial m} \log(\mathcal{A}_{\mathrm{neg}}(m, \alpha, \beta, \gamma))$$

$$= \frac{\alpha\beta|\gamma|}{(2\beta m + |\gamma|)^2} \left[\psi\left(\frac{\alpha(\beta m + |\gamma|)}{2\beta m + |\gamma|} \right) - log(\beta m^2 + m|\gamma|) \right] + \frac{2\beta}{2\beta m + |\gamma|} - \frac{\beta}{\beta m + |\gamma|}$$

$$= \frac{\alpha\beta|\gamma|}{(2\beta m + |\gamma|)^2} \left[\psi\left(\frac{\alpha(\beta m + |\gamma|)}{2\beta m + |\gamma|} \right) - log(\beta m^2 + m|\gamma|) \right] + \frac{\alpha\beta|\gamma|}{(\beta m + |\gamma|)(2\beta m + |\gamma|)}$$

$$= \frac{\alpha\beta|\gamma|}{(2\beta m + |\gamma|)^2} \left[\psi\left(\frac{\alpha(\beta m + |\gamma|)}{2\beta m + |\gamma|} \right) + \frac{2\beta m + |\gamma|}{\alpha(\beta m + |\gamma|)} - log(\beta m^2 + m|\gamma|) \right], \quad (5.74)$$

where $\psi(\cdot)$ denotes the digamma function. The function $x \mapsto \psi(x) + \frac{1}{x}$ is a strictly increasing function because

$$\frac{d}{dx}\left(\psi(x) + \frac{1}{x} \right) = \left(\frac{d\psi(x)}{dx} - \frac{1}{x^2} \right) \overset{(\star\star)}{>} \frac{e^{\frac{1}{x+1}} - e^{-\frac{1}{x}}}{2} > 0,$$

where the inequality in $(\star\star)$ is due to Equation 1.12 in Qi and Mortici (2015). ((Yang et al., 2014))

The function $m \mapsto \frac{\alpha(\beta m + |\gamma|)}{2\beta m + |\gamma|}$ is strictly decreasing. Therefore their composition, $m \mapsto \psi(\frac{\alpha(\beta m + |\gamma|)}{2\beta m + |\gamma|}) + \frac{2\beta m + |\gamma|}{\alpha(\beta m + |\gamma|)}$ is a strictly decreasing function. As the functions $m \mapsto \frac{\alpha\beta|\gamma|}{(2\beta m + |\gamma|)^2}$ and $m \mapsto -\log(\beta m^2 + |\gamma|m)$ are strictly decreasing as well, it follows from Equation 5.74 that $\frac{\partial}{\partial m} \log(\mathcal{A}_{\mathrm{neg}}(m, \alpha, \beta, \gamma))$ is strictly decreasing in m for $m > 0$.

Moreover,

$$\lim_{m\to 0} \frac{\partial}{\partial m} \log(\mathcal{A}_{\mathrm{neg}}(m,\alpha,\beta,\gamma)) = \infty, \ \ \lim_{m\to\infty} \frac{\partial}{\partial m} \log(\mathcal{A}_{\mathrm{neg}}(m,\alpha,\beta,\gamma)) = -\infty,$$

and $m \mapsto \frac{\partial}{\partial m} \log(\mathcal{A}_{\mathrm{neg}}(m,\alpha,\beta,\gamma))$ is a continuous function. Consequently, we conclude that that a unique $m_{\mathrm{opt}} > 0$ such that $\frac{\partial}{\partial m} \log(\mathcal{A}_{\mathrm{neg}}(m,\alpha,\beta,\gamma))\big|_{m=m_{\mathrm{opt}}} = 0$.

Now we will prove that $m_{\mathrm{opt}} > X_{\mathrm{mode}}$ the mode of the MHN (α,β,γ) distribution. X_{mode} satisfies the equation

$$\frac{\alpha-1}{X_{\mathrm{mode}}} - 2\beta X_{\mathrm{mode}} - |\gamma| = 0 \ \implies \ 2\beta X_{\mathrm{mode}}^2 + |\gamma| X_{\mathrm{mode}} = \alpha - 1. \qquad (5.75)$$

As a result, it follows that

$$\left[\psi\left(\frac{\alpha(\beta X_{\mathrm{mode}} + |\gamma|)}{2\beta X_{\mathrm{mode}} + |\gamma|} \right) - \log(\beta X_{\mathrm{mode}}^2 + X_{\mathrm{mode}}|\gamma|) \right]$$
$$= \left[\psi\left(\frac{\alpha(\beta X_{\mathrm{mode}}^2 + |\gamma| X_{\mathrm{mode}})}{\alpha - 1} \right) - \log(\beta X_{\mathrm{mode}}^2 + X_{\mathrm{mode}}|\gamma|) \right]. \qquad (5.76)$$

Consider the inequality $\log(x) - \psi(x) < \frac{1}{x}$ for all $x > 0$ (Alzer, 1997; Anderson et al., 1995). Thus it follows from Equation 5.76 that

$$\frac{\alpha\beta|\gamma|}{(2\beta X_{\mathrm{mode}} + |\gamma|)^2} \left[\psi\left(\frac{\alpha(\beta X_{\mathrm{mode}} + |\gamma|)}{2\beta X_{\mathrm{mode}} + |\gamma|} \right) - \log(\beta X_{\mathrm{mode}}^2 + X_{\mathrm{mode}}|\gamma|) \right]$$

$$> \frac{\alpha\beta|\gamma|}{(2\beta X_{\mathrm{mode}} + |\gamma|)^2} \left\{ \log\left(\frac{\alpha(\beta X_{\mathrm{mode}}^2 + |\gamma| X_{\mathrm{mode}})}{\alpha - 1} \right) - \frac{\alpha-1}{\alpha(\beta X_{\mathrm{mode}}^2 + |\gamma| X_{\mathrm{mode}})} - \log(\beta X_{\mathrm{mode}}^2 + X_{\mathrm{mode}}|\gamma|) \right\}$$

$$= \frac{\alpha\beta|\gamma|}{(2\beta X_{\mathrm{mode}} + |\gamma|)^2} \log\left(\frac{\alpha}{\alpha-1} \right) - \frac{\alpha\beta|\gamma|}{(2\beta X_{\mathrm{mode}} + |\gamma|)^2} \frac{\alpha-1}{\alpha(\beta X_{\mathrm{mode}}^2 + |\gamma| X_{\mathrm{mode}})}$$

$$= \frac{\alpha\beta|\gamma|}{(2\beta X_{\mathrm{mode}} + |\gamma|)^2} \log\left(\frac{\alpha}{\alpha-1} \right) - \frac{\beta|\gamma|}{(2\beta X_{\mathrm{mode}} + |\gamma|)(2\beta X_{\mathrm{mode}}^2 + |\gamma| X_{\mathrm{mode}})} \frac{\alpha-1}{(\beta X_{\mathrm{mode}} + |\gamma|)}$$

$$\overset{(*)}{=} \frac{\alpha\beta|\gamma|}{(2\beta X_{\mathrm{mode}} + |\gamma|)^2} \log\left(\frac{\alpha}{\alpha-1} \right) - \frac{\beta|\gamma|}{(2\beta X_{\mathrm{mode}} + |\gamma|)(\alpha-1)} \frac{\alpha-1}{(\beta X_{\mathrm{mode}} + |\gamma|)}$$

$$= \frac{\alpha\beta|\gamma|}{(2\beta X_{\mathrm{mode}} + |\gamma|)^2} \log\left(\frac{\alpha}{\alpha-1} \right) - \frac{\beta|\gamma|}{(2\beta X_{\mathrm{mode}} + |\gamma|)(\beta X_{\mathrm{mode}} + |\gamma|)},$$

144

where the equality in $(*)$ is a consequence of the Equation 5.75. Therefore,

$$
\begin{aligned}
& \tfrac{\partial}{\partial m} \log(\mathcal{A}_{\text{neg}}(m,\alpha,\beta,\gamma))\big|_{m=X_{\text{mode}}} \\
& = \frac{\alpha\beta|\gamma|}{(2\beta X_{\text{mode}}+|\gamma|)^2}\left[\psi\left(\frac{\alpha(\beta X_{\text{mode}}+|\gamma|)}{2\beta X_{\text{mode}}+|\gamma|}\right)-log(\beta X^2_{\text{mode}}+X_{\text{mode}}|\gamma|)\right]+\frac{2\beta}{2\beta X_{\text{mode}}+|\gamma|}-\frac{\beta}{\beta X_{\text{mode}}+|\gamma|} \\
& > \frac{\alpha\beta|\gamma|}{(2\beta X_{\text{mode}}+|\gamma|)^2}\log\left(\frac{\alpha}{\alpha-1}\right)-\frac{\beta|\gamma|}{(2\beta X_{\text{mode}}+|\gamma|)(\beta X_{\text{mode}}+|\gamma|)}+\frac{\beta|\gamma|}{(2\beta X_{\text{mode}}+|\gamma|)(\beta X_{\text{mode}}+|\gamma|)} \\
& = \frac{\alpha\beta|\gamma|}{(2\beta X_{\text{mode}}+|\gamma|)^2}\log\left(\frac{\alpha}{\alpha-1}\right) \\
& > 0.
\end{aligned}
\tag{5.77}
$$

The function $\log(\mathcal{A}_{\text{neg}}(m,\alpha,\beta,\gamma)))$ is increasing in the region $(0, m_{\text{opt}})$ whereas it decreases on (m_{opt}, ∞). As the slope of the function at X_{mode} is positive (see Equation 5.77) we conclude that $X_{\text{mode}} \in (0, m_{\text{opt}})$ and $m_{\text{opt}} > X_{\text{mode}}$.

Proof of part(c), Theorem 4:

The arguments used to prove this part of the Themorem utilizes the following Theorem on the Gamma function (Ramanujan's Double Inequality(Alzer, 2003)). For $z > 0$,

$$
\sqrt{\pi}\left(\frac{z}{e}\right)^z\left(8z^3 + 4z^2 + z + \frac{1}{100}\right)^{\frac{1}{6}} < \Gamma(1+z) < \sqrt{\pi}\left(\frac{z}{e}\right)^z\left(8z^3 + 4z^2 + z + \frac{1}{30}\right)^{\frac{1}{6}}.
$$

From Theorem 3 and the Lemma 4 we get that,

$$
\begin{aligned}
\mathcal{A}_{\text{neg}}(m,\alpha,\beta,\gamma) &= \frac{(2\beta m + |\gamma|)(\beta m + |\gamma|)^{\frac{\alpha(\beta m+|\gamma|)}{2\beta m+|\gamma|}-1}\,{}_1\Psi_1\left[\begin{array}{c}(\frac{\alpha}{2},\frac{1}{2})\\(1,0)\end{array};\frac{-|\gamma|}{\sqrt{\beta}}\right]}{2\beta^{\frac{\alpha}{2}}\,m^{\frac{\alpha\beta m}{2\beta m+|\gamma|}}\,\Gamma(\frac{\alpha(\beta m+|\gamma|)}{2\beta m+|\gamma|})}\\[2mm]
&\geq \frac{(2\beta m + |\gamma|)(\beta m + |\gamma|)^{\frac{\alpha(\beta m+|\gamma|)}{2\beta m+|\gamma|}-1}\left[\frac{\beta^{\frac{\alpha}{2}}\exp\left(-\frac{m|\gamma|}{2}\right)\Gamma(\frac{\alpha}{2})}{\left(\beta+\frac{|\gamma|}{2m}\right)^{\frac{\alpha}{2}}}\right]}{2\beta^{\frac{\alpha}{2}}\,m^{\frac{\alpha\beta m}{2\beta m+|\gamma|}}\,\Gamma(\frac{\alpha(\beta m+|\gamma|)}{2\beta m+|\gamma|})}\\[2mm]
&\geq \frac{(2\beta m + |\gamma|)(\beta m + |\gamma|)^{\frac{\alpha(\beta m+|\gamma|)}{2\beta m+|\gamma|}-1}\left[\exp\left(-\frac{m|\gamma|}{2}\right)\Gamma(\frac{\alpha}{2})\right]}{2\,m^{\frac{\alpha\beta m}{2\beta m+|\gamma|}}\,\Gamma(\frac{\alpha(\beta m+|\gamma|)}{2\beta m+|\gamma|})\left[\left(\beta+\frac{|\gamma|}{2m}\right)^{\frac{\alpha}{2}}\right]}\\[2mm]
&\geq \frac{(2\beta m + |\gamma|)(\beta m + |\gamma|)^{\frac{\alpha(\beta m+|\gamma|)}{2\beta m+|\gamma|}-1}}{2\,m^{\frac{\alpha\beta m}{2\beta m+|\gamma|}}\left[\left(\beta+\frac{|\gamma|}{2m}\right)^{\frac{\alpha}{2}}\right]}\times\frac{\left[\exp\left(-\frac{m|\gamma|}{2}\right)\Gamma(\frac{\alpha}{2})\right]}{\Gamma(\frac{\alpha(\beta m+|\gamma|)}{2\beta m+|\gamma|})}\\[2mm]
&\geq \frac{(2\beta m + |\gamma|)(\beta m + |\gamma|)^{\frac{\alpha(\beta m+|\gamma|)}{2\beta m+|\gamma|}-1}}{2^{-\frac{\alpha}{2}+1}\,m^{\frac{\alpha\beta m}{2\beta m+|\gamma|}-\frac{\alpha}{2}}\left[(2\beta m+|\gamma|)^{\frac{\alpha}{2}}\right]}\times\frac{\left[\exp\left(-\frac{m|\gamma|}{2}\right)\Gamma(\frac{\alpha}{2})\right]}{\Gamma(\frac{\alpha(\beta m+|\gamma|)}{2\beta m+|\gamma|})}\\[2mm]
&\geq \frac{2^{\frac{\alpha}{2}-1}(\beta m + |\gamma|)^{\frac{\alpha(\beta m+|\gamma|)}{2\beta m+|\gamma|}-1}}{m^{\frac{\alpha\beta m}{2\beta m+|\gamma|}-\frac{\alpha}{2}}\,(2\beta m+|\gamma|)^{\frac{\alpha}{2}-1}}\times\frac{\left[\exp\left(-\frac{m|\gamma|}{2}\right)\Gamma(\frac{\alpha}{2})\right]}{\Gamma(\frac{\alpha(\beta m+|\gamma|)}{2\beta m+|\gamma|})}. \quad (5.78)
\end{aligned}
$$

Consider a point $m_\star > 0$ such that $2\beta m_\star^2 + |\gamma|m_\star = \alpha$. Note that, X_{mode}, the mode of the MHN (α,β,γ), $\alpha > 1, \gamma < 0$ satisfies the equation $2\beta X_{\text{mode}}^2 + |\gamma|X_{\text{mode}} = \alpha - 1$. Therefore $m_\star$ is greater than X_{mode}. It follows from Equation 5.78 that

$$
\begin{aligned}
\mathcal{A}_{\text{neg}}(m_\star,\alpha,\beta,\gamma) &\geq \frac{2^{\frac{\alpha}{2}-1}(\beta m_\star + |\gamma|)^{\frac{\alpha(\beta m_\star+|\gamma|)}{2\beta m_\star+|\gamma|}-1}}{m_\star^{\frac{\alpha\beta m_\star}{2\beta m_\star+|\gamma|}-\frac{\alpha}{2}}\,(2\beta m_\star+|\gamma|)^{\frac{\alpha}{2}-1}}\times\frac{\left[\exp\left(-\frac{m_\star|\gamma|}{2}\right)\Gamma(\frac{\alpha}{2})\right]}{\Gamma(\frac{\alpha(\beta m_\star+|\gamma|)}{2\beta m_\star+|\gamma|})}\\[2mm]
&= \frac{2^{\frac{\alpha}{2}-1}(\beta m_\star^2 + |\gamma|m_\star)^{\frac{\alpha(\beta m_\star+|\gamma|)}{2\beta m_\star+|\gamma|}-1}}{(2\beta m_\star^2+|\gamma|m_\star)^{\frac{\alpha}{2}-1}}\times\frac{\left[\exp\left(-\frac{m_\star|\gamma|}{2}\right)\Gamma(\frac{\alpha}{2})\right]}{\Gamma(\frac{\alpha(\beta m_\star+|\gamma|)}{2\beta m_\star+|\gamma|})}.
\end{aligned}
$$

Applying Ramanujan's Double Inequality for the Gamma function (stated above), we get that

$$\frac{\Gamma(\frac{\alpha}{2})}{\Gamma(\frac{\alpha(\beta m_\star + |\gamma|)}{2\beta m_\star + |\gamma|})} > \left(\frac{\alpha-2}{2e}\right)^{\frac{\alpha}{2}-1} \left(\frac{1}{e}\left\{\frac{\alpha(\beta m_\star + |\gamma|)}{(2\beta m_\star + |\gamma|)} - 1\right\}\right)^{-\frac{\alpha(\beta m_\star + |\gamma|)}{2\beta m_\star + |\gamma|}+1} \Upsilon(\alpha, \beta, \gamma, m_\star),$$

$$(5.79)$$

where

$$\Upsilon(\alpha, \beta, \gamma, m_\star) = \frac{\left((\alpha-2)^3 + (\alpha-2)^2 + \frac{\alpha-2}{2} + \frac{1}{100}\right)^{\frac{1}{6}}}{\left[(\alpha-2)^3(1+c_{\beta,\gamma,m_\star})^3 + (\alpha-2)^2(1+c_{\beta,\gamma,m_\star})^2 + \frac{\alpha-2}{2}(1+c_{\beta,\gamma,m_\star}) + \frac{1}{30}\right]^{\frac{1}{6}}},$$

and

$$c_{\beta,\gamma,m_\star} = \frac{\alpha\gamma}{(\alpha-2)(2\beta m_\star + \gamma)}.$$

In the case when $\alpha \geq 3$

$$
\begin{aligned}
c_{\beta,\gamma,m_\star} = \frac{\alpha\gamma}{(\alpha-2)(2\beta m_\star + \gamma)} &= \frac{\alpha\gamma m_\star}{(\alpha-2)(2\beta m_\star^2 + \gamma m_\star)} \\
&= \frac{\alpha\gamma m_\star}{(\alpha-2)(\alpha)} \\
&= \frac{2}{(\alpha-2)\left(\sqrt{(1+\frac{8\alpha\beta}{\gamma^2})}+1\right)} \\
&\leq 1.
\end{aligned}
$$

$$(5.80)$$

Moreover, $\Upsilon(\alpha, \beta, \gamma, m_\star) \geq \frac{\left((\alpha-2)^3 + (\alpha-2)^2 + \frac{\alpha-2}{2} + \frac{1}{100}\right)^{\frac{1}{6}}}{\left[(\alpha-2)^3(2)^3 + (\alpha-2)^2(2)^2 + \frac{\alpha-2}{2}(2) + \frac{1}{30}\right]^{\frac{1}{6}}} \geq \frac{1}{\sqrt{2}}$ because $\alpha \mapsto \frac{\left(\alpha^3 + \alpha^2 + \frac{\alpha}{2} + \frac{1}{100}\right)}{\left(8\alpha^3 + 4\alpha^2 + \alpha + \frac{1}{30}\right)}$ is strictly decreasing in α and $\lim_{\alpha\to\infty} \frac{\left(\alpha^3 + \alpha^2 + \frac{\alpha}{2} + \frac{1}{100}\right)}{\left(8\alpha^3 + 4\alpha^2 + \alpha + \frac{1}{30}\right)} = \frac{1}{8}$. Therefore,

$$\frac{\Gamma(\frac{\alpha}{2})}{\Gamma(\frac{\alpha(\beta m_\star + |\gamma|)}{2\beta m_\star + |\gamma|})} > \left(\frac{\alpha-2}{2e}\right)^{\frac{\alpha}{2}-1} \left(\frac{1}{e}\left\{\frac{\alpha(\beta m_\star + |\gamma|)}{(2\beta m_\star + |\gamma|)} - 1\right\}\right)^{-\frac{\alpha(\beta m_\star + |\gamma|)}{2\beta m_\star + |\gamma|}+1} \frac{1}{\sqrt{2}}.$$

Consider the fact that, for any $c > 0, x > 0$,

$$(1 + \frac{1}{x+c})^{x+c} > (1 + \frac{1}{x})^x$$

$$\implies \quad x^x(x+c+1)^{x+c} > (x+1)^x(x+c)^{x+c}$$

$$\implies \quad x^x(x+c)^{-(x+c)} > (x+1)^x(x+c+1)^{-(x+c)}.$$

Utilizing the inequality with $x = \frac{\alpha}{2} - 1$ and $x + c = \frac{\alpha(\beta m_\star + |\gamma|)}{(2\beta m_\star + |\gamma|)} - 1$, it follows from Equation 5.81 that

$$\frac{\Gamma(\frac{\alpha}{2})}{\Gamma(\frac{\alpha(\beta m_\star + |\gamma|)}{2\beta m_\star + |\gamma|})} > \left(\frac{\alpha}{2e}\right)^{\frac{\alpha}{2}-1} \left(\frac{1}{e}\left\{\frac{\alpha(\beta m_\star + |\gamma|)}{(2\beta m_\star + |\gamma|)}\right\}\right)^{-\frac{\alpha(\beta m_\star + |\gamma|)}{2\beta m_\star + |\gamma|}+1} \frac{1}{\sqrt{2}}$$

$$= \left(\frac{e}{\alpha}\right)^{\frac{\alpha(\beta m_\star + |\gamma|)}{2\beta m_\star + |\gamma|}-\frac{\alpha}{2}} \left(\frac{1}{2}\right)^{\frac{\alpha}{2}-1} \left(\frac{(\beta m_\star + |\gamma|)}{(2\beta m_\star + |\gamma|)}\right)^{-\frac{\alpha(\beta m_\star + |\gamma|)}{2\beta m_\star + |\gamma|}+1} \frac{1}{\sqrt{2}}$$

$$= 2^{-\frac{\alpha-1}{2}} \left(\frac{e}{\alpha}\right)^{\frac{\alpha|\gamma|}{2(2\beta m_\star + |\gamma|)}} \left(\frac{(\beta m_\star + |\gamma|)}{(2\beta m_\star + |\gamma|)}\right)^{-\frac{\alpha(\beta m_\star + |\gamma|)}{2\beta m_\star + |\gamma|}+1}. \quad (5.81)$$

It follows from Equations 5.79 and 5.81 that

$$\mathcal{A}_{\text{neg}}(m_\star, \alpha, \beta, \gamma)$$

$$\geq \frac{2^{\frac{\alpha}{2}-1}(\beta m_\star^2 + |\gamma|m_\star)^{\frac{\alpha(\beta m_\star + |\gamma|)}{2\beta m_\star + |\gamma|}-1}}{(2\beta m_\star^2 + |\gamma|m_\star)^{\frac{\alpha}{2}-1}} \times \frac{\left[\exp\left(-\frac{m_\star|\gamma|}{2}\right)\Gamma(\frac{\alpha}{2})\right]}{\Gamma(\frac{\alpha(\beta m_\star + |\gamma|)}{2\beta m_\star + |\gamma|})}.$$

$$\geq \frac{2^{\frac{\alpha}{2}-1}(\beta m_\star^2 + |\gamma|m_\star)^{\frac{\alpha(\beta m_\star + |\gamma|)}{2\beta m_\star + |\gamma|}-1} \exp\left(-\frac{m_\star|\gamma|}{2}\right)}{(2\beta m_\star^2 + |\gamma|m_\star)^{\frac{\alpha}{2}-1}}$$

$$\times \left[2^{-\frac{\alpha-1}{2}}\left(\frac{e}{\alpha}\right)^{\frac{\alpha|\gamma|}{2(2\beta m_\star + |\gamma|)}}\left(\frac{(\beta m_\star + |\gamma|)}{(2\beta m_\star + |\gamma|)}\right)^{-\frac{\alpha(\beta m_\star + |\gamma|)}{2\beta m_\star + |\gamma|}+1}\right]$$

$$\geq \frac{2^{\frac{\alpha}{2}-1}\exp\left(-\frac{m_\star|\gamma|}{2}\right)}{(2\beta m_\star^2 + |\gamma|m_\star)^{-\frac{\alpha|\gamma|}{2(2\beta m_\star + |\gamma|)}}} \times \left[2^{-\frac{\alpha-1}{2}}\left(\frac{e}{\alpha}\right)^{\frac{\alpha|\gamma|}{2(2\beta m_\star + |\gamma|)}}\right]$$

$$= \frac{2^{-\frac{1}{2}}\exp\left(-\frac{m_\star|\gamma|}{2}\right)}{(2\beta m_\star^2 + |\gamma|m_\star)^{-\frac{\alpha|\gamma|}{2(2\beta m_\star + |\gamma|)}}} \times \left[\left(\frac{e}{\alpha}\right)^{\frac{\alpha|\gamma|}{2(2\beta m_\star + |\gamma|)}}\right]. \quad (5.82)$$

As the point $m_\star$ is such that, $2\beta m_\star^2 + |\gamma|m_\star = \alpha$. It appears that

$$
\begin{aligned}
\mathcal{A}_{\text{neg}}(m_\star, \alpha, \beta, \gamma) &\geq \frac{2^{-\frac{1}{2}} \exp\left(-\frac{m_\star|\gamma|}{2}\right)}{(2\beta m_\star^2 + |\gamma|m_\star)^{-\frac{\alpha|\gamma|}{2(2\beta m_\star + |\gamma|)}}} \times \left[\left(\frac{e}{\alpha}\right)^{\frac{\alpha|\gamma|}{2(2\beta m_\star + |\gamma|)}}\right]. \\
&= \frac{2^{-\frac{1}{2}} \exp\left(-\frac{m_\star|\gamma|}{2}\right)}{(\alpha)^{-\frac{\alpha|\gamma|}{2(2\beta m_\star + |\gamma|)}}} \times \left[\left(\frac{e}{\alpha}\right)^{\frac{\alpha|\gamma|}{2(2\beta m_\star + |\gamma|)}}\right]. \\
&= 2^{-\frac{1}{2}} \exp\left(-\frac{m_\star|\gamma|}{2}\right) \times \left[(e)^{\frac{\alpha|\gamma|m_\star}{2(\alpha)}}\right]. \\
&= \frac{1}{\sqrt{2}}.
\end{aligned}
\tag{5.83}
$$

The point m_{opt} maximizes the function $m \mapsto \mathcal{A}_{\text{neg}}(m, \alpha, \beta, \gamma)$. Therefore, we conclude that

$$
\mathcal{A}_{\text{neg}}(m_{\text{opt}}, \alpha, \beta, \gamma) \geq \mathcal{A}_{\text{neg}}(m_\star, \alpha, \beta, \gamma) \geq \frac{1}{\sqrt{2}}.
$$

5.0.17 Iterative algorithm to find the optimal matching point.

In order to find m_{opt}, we utilize Newton-Raphson Method to maximize $\log(\mathcal{A}_{\text{neg}}($ $m_{\text{opt}}, \alpha, \beta, \gamma))$ with respect to m. We start the algorithm at m_{init} Note that, we discussed the construction of m_{init} in the Section 2.4.2. Apart from the initial value, we require the following quantities for describing the specific steps of the algorithm.

$l(m, \alpha, \beta, \gamma)$

$= \log(\mathcal{A}_{\text{neg}}(m_{\text{opt}}, \alpha, \beta, \gamma))$

$= \log({}_1\Psi_1 \left[\begin{matrix} (\frac{\alpha}{2}, \frac{1}{2}), \\ (1, 0) \end{matrix} ; \frac{-|\gamma|}{\sqrt{\beta}} \right]) + \frac{\alpha(\beta m + |\gamma|)}{2\beta m + |\gamma|} \log(\beta m^2 + m|\gamma|) + \log(2\beta m + |\gamma|) - \log 2$

$- \frac{\alpha}{2} \log(\beta m^2) - \log \Gamma(\frac{\alpha(\beta m + |\gamma|)}{2\beta m + |\gamma|}) - \log(\beta m + |\gamma|).$

If we denote $l'(m, \alpha, \beta, \gamma)$ and $l''(m, \alpha, \beta, \gamma)$ to be the first and the second derivative of the function $l(m, \alpha, \beta, \gamma)$ then

$$l'(m, \alpha, \beta, \gamma) = \frac{\alpha\beta|\gamma|}{(2\beta m + |\gamma|)^2} \left[\psi\left(\frac{\alpha(\beta m + |\gamma|)}{2\beta m + |\gamma|} \right) - \log(\beta m^2 + m|\gamma|) \right] + \frac{2\beta}{2\beta m + |\gamma|} - \frac{\beta}{\beta m + |\gamma|},$$

$$l''(m, \alpha, \beta, \gamma) =$$
$$\frac{4\alpha\beta^2|\gamma|}{(2\beta m + |\gamma|)^3} \left[\psi\left(\frac{\alpha(\beta m + |\gamma|)}{2\beta m + |\gamma|} \right) - \log(\beta m^2 + m|\gamma|) \right]$$
$$+ \frac{\alpha\beta|\gamma|}{(2\beta m + |\gamma|)^2} \left[\frac{-\alpha\beta|\gamma|}{(2\beta m + |\gamma|)^2} \psi'\left(\frac{\alpha(\beta m + |\gamma|)}{2\beta m + |\gamma|} \right) - \frac{2\beta m + |\gamma|}{\beta m^2 + m|\gamma|} \right] - \frac{1}{(m + \frac{|\gamma|}{2\beta})^2} + \frac{1}{(m + \frac{|\gamma|}{\beta})^2}.$$

The notation ψ and ψ' used in the above expressions to refer to the digamma and trigamma functions respectively. The steps of the iterative procedure to find m_{opt} is given as follows

- If we denote the starting value of the parameter m to be $m_{(0)}$ then set $m_{(0)} = m_{\text{init}}$.

- At the $(t+1)^{\text{th}}$ step of the iteration, compute $m_{t+1} = m_t - \frac{l'(m_t, \alpha, \beta, \gamma)}{l''(m_t, \alpha, \beta, \gamma)}$ for $t \geq 0$.

- Repeat the previous step until $|\mathcal{A}_{\text{neg}}(m_{t+1}, \alpha, \beta, \gamma) - \mathcal{A}_{\text{neg}}(m_t, \alpha, \beta, \gamma)| < \epsilon$, where $\epsilon > 0$ is a predetermined small positive number to identify the convergence of the algorithm. In our implementation we used ϵ to be 0.001.
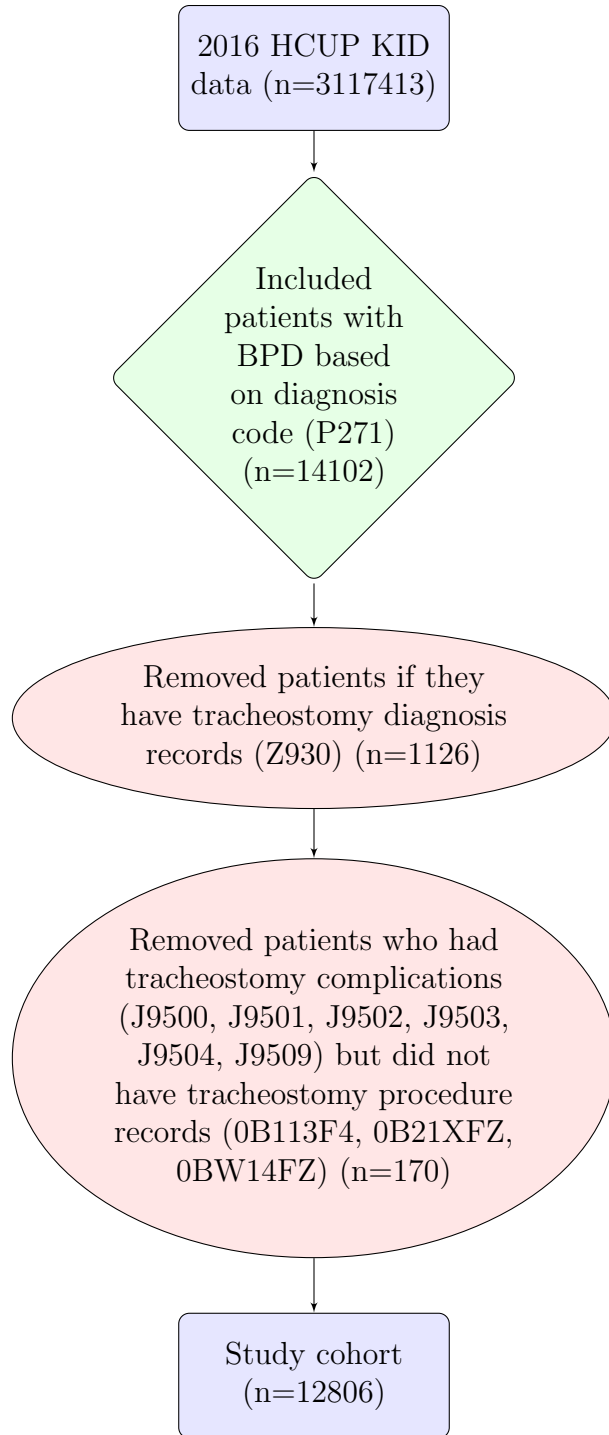
Figure 5.3: Flow chart of generating case study cohort

Figure 5.3 shows how we generated the cohort based on the 2016 HCUP KID data for the case study.

CURRICULUM VITA

NAME: Jingchao Sun

ADDRESS: Department of Biostatistics and Bioinformatics

                  University of Louisville

                  Louisville, KY 40202

EDUCATION: Bachelor of Science in Finance,

                  Southeast University, 2013

                  Masters of Science in Economics,

                  University of Illinois at Urbana-Champaign, 2014

PUBLICATIONS: Jingchao Sun, Subhadip Pal,

                  The Modified-Half-Normal distribution: properties and an efficient

                  sampling scheme, (Submitted).

                  Jingchao Sun, Subhidip Pal, Scott Duncan, Maiying Kong,

                  Estimation of treatment effect by directed acyclic graph and

                  propensity score, (To be submitted).

                  Jingchao Sun, Maiying Kong,

                  Different doubly robust methods for estimating ATE and

their improvement using super learner (In process).

Mike Smith, Jingchao Sun, Yana Feygin,
Maiying Kong, Bethany Wattles,
Outpatient antibiotic prescribing in a cohort of children
with 5 years continuous enrollment Kentucky medicaid,
(PAS meeting accepted).

Julia Springate, Duncan,Scott Davis, Feygin,Yana,
Kong,Maiying, Jingchao Sun, Jawad,Kahir,
An epidemiologic survey of neonatal intensive care unit
capacity and utilization in Kentucky, 1999-2018,
(PAS meeting accepted)

Deborah Winders Davis, Yana Feygin, Liza Creel,
Maiying Kong, Kahir Jawad, Jingchao Sun, Nathan J. Blum,
W. David Lohr, P. Gail Williams, Jennifer Le,
V. Faye Jones, and Natalie Pasquenza (2020).
Epidemiology of treatment for preschoolers on Kentucky
medicaid diagnosed with attention-deficit/hyperactivity disorder.
Journal of Child and Adolescent Psychopharmacology.

PRESENTATIONS: The Modified-Half-Normal distribution:
properties and efficient sampling scheme
Seminar Series, Department of Bioinformatics and Biostatistics,
University of Louisville, Louisville, KY, USA, March 03, 2019.

The Modified-Half-Normal distribution:

properties and an efficient sampling scheme,

ASA-KY Chapter Meeting, Louisville, KY, April 05, 2019.

Directed acyclic graph assisted methods for

estimating average treatment effect

Eastern North American Region (ENAR), March 24, 2020.

HONORS AND

AWARDS Graduate Research Assistantship,

Department of Bioinformatics and Biostatistics,

University of Louisville,

August 2016-July 2020

Best Student Presentation Award,

ASA-KY Chapter Meeting.

March 2019

WORK

EXPERIENCE Biostatistician Internship

Amgen, Thousand Oaks, CA

June 2019-August 2019