Measure of dependence for length-biased survival data

Rachid Bentoumi

Thesis submitted to the Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Mathematics [1]

Department of Mathematics and Statistics
Faculty of Science
University of Ottawa

[1] The Ph.D. program is a joint program with Carleton University, administered by the Ottawa-Carleton Institute of Mathematics and Statistics

Abstract

In epidemiological studies, subjects with disease (prevalent cases) differ from newly diseased (incident cases). They tend to survive longer due to sampling bias, and related covariates will also be biased. Methods for regression analyses have recently been proposed to measure the potential effects of covariates on survival. The goal is to extend the dependence measure of Kent [33], based on the information gain, in the context of length-biased sampling. In this regard, to estimate information gain and dependence measure for length-biased data, we propose two different methods namely kernel density estimation with a regression procedure and parametric copulas. We will assess the consistency for all proposed estimators. Algorithms detailing how to generate length-biased data, using kernel density estimation with a regression procedure and parametric copulas approaches, are given. Finally, the performances of the estimated information gain and dependence measure, under length-biased sampling, are demonstrated through simulation studies.

Acknowledgements

First and foremost, I would like to express my sincere gratitude and my very great appreciation to my supervisors Dr. Mayer Alvo and Dr. Mhamed Mesfioui for all their contributions of times, ideas, assistance, motivation, patience, immense knowledge and enthusiasm throughout my Ph.D study and research. I could not have imagined finishing my Ph.D without the continuous support of Dr. Mayer Alvo and Dr. Mhamed Mesfioui . Thank you deeply. My sincere thanks also goes to my thesis examining committee.

This work was supported by grants from Fonds québécois de la recherche sur la nature et les technologies. They find here the expression of my gratitude. Also, University of Ottawa Admission Scholarship and Faculty of Graduate and Postdoctoral Studies are acknowledged and greatly appreciated.

Last but not the least, I would like to offer my special thanks to my wife, Ghita, for her personal support and endless patience at all times. A heartfelt thank you to my children Aicha, Yassmine and Youness who have been encouraging me with their smiles and understanding of how busy I was. They have been a great source of inspiration and motivation. My parents, brothers and sisters are also to be thanked for their support, prayers and understanding. I am especially grateful my wonderful and generous friends at the Department of Mathematics and Statistics, University of Ottawa for stimulating a rich and a welcoming social and academic environment throughout.

Dedication

This work is dedicated to my dear parents Abdelaziz and Najia , to my “little ones” Aicha, Yassmine and Youness, to my lovely wife Ghita and to the loving memory of my grand-parents.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Survival analysis is a branch of statistics, generally defined as a set of statistical techniques for analyzing a positive-valued random variable. Typically, the random variable describes the time until the occurrence of a specific event such as death, relapse, failure, response or the development of a given disease. Survival data, which are often referred to as time-to-event-data or lifetime data, occur in many areas such as medicine, epidemiology, biology, economics and manufacturing. The principal goal in survival analysis is the study of the occurrence of a specific event. In epidemiology, this analysis is based on the study of incident and prevalent cases. The following diagram exhibits some possible incident cases.

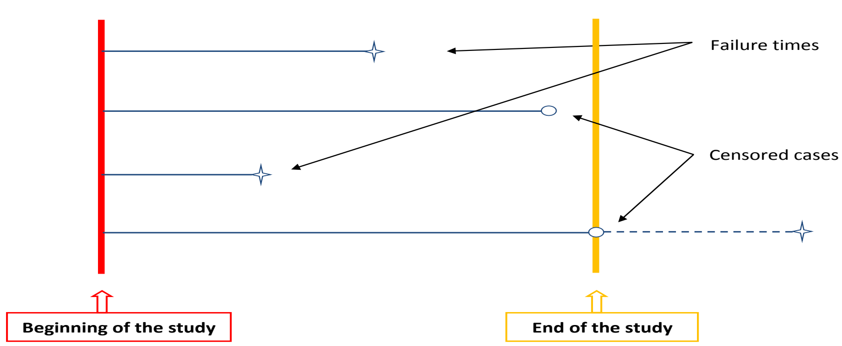

Figure 1.1: Study of incident cases.

In the study of incident cases, subjects are observed from the time of initiation of a specific event, such as onset of a disease and followed until occurrence of the event or censoring. In such studies, incidence is the rate of new cases of the disease in a given population, generally reported as the number of new cases occurring within a period of time. In addition, the censoring process is noninformative since it does not depend on the survival time. When a disease is so rare, or simply due to certain time and cost constraints, an alternative approach is suggested which is the study of a prevalent cohort, collected through cross-sectional surveys. The prevalence is the actual number of cases still alive with disease, in some population, at a particular date in time (point prevalence). The following diagram presents some possible observed prevalent cases.

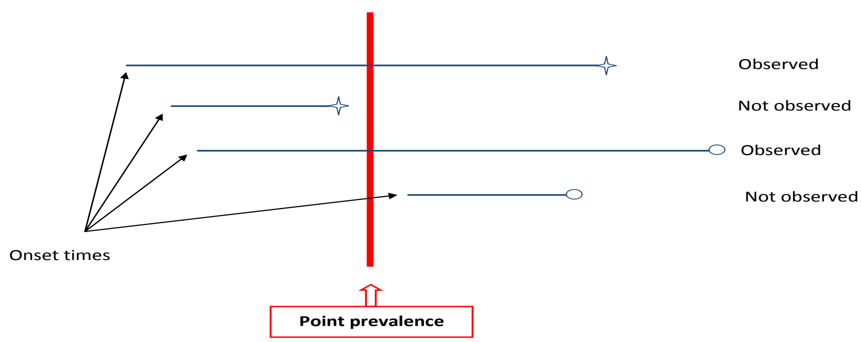

Figure 1.2: Study of prevalent cases.

A cross-sectional study, as shown in Figure 1.2, allows for the identification of prevalent cases with disease. The observed subjects must already have the disease in question before entering the study (this is called left truncation) and for some fixed period of time, they are followed until failure or censoring. The collected observed data from prevalent cases form a biased sample, due to the bias that stems from the lifetimes being left-truncated (the event has already occurred). In addition, when we assume that the onset times stem from a stationary Poisson process (if there has been no epidemic of the disease during the onset times of the subjects [6]) then the observed failure lifetimes are length-biased. Under the assumption of stationarity, the truncation time is uniformly distributed and the term "length-biased" is used instead of left truncated [50]. For the stationarity, an informal test was investigated by Asgharian et al. [6] and the first formal test for the stationarity of the incidence rate using data from a prevalent cohort study with follow-up was developed by Addona and Wolfson [1].

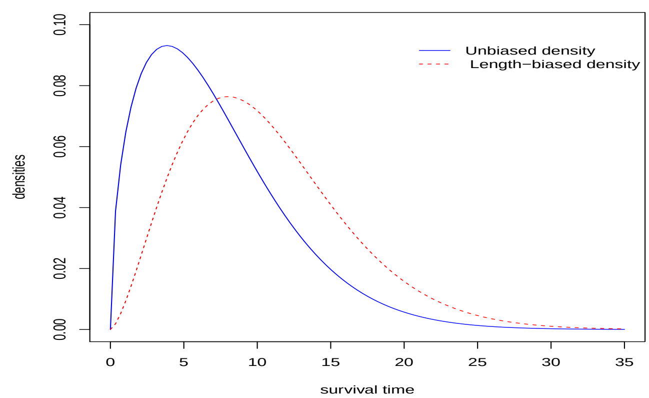

Figure 1.3: Unbiased density versus length-biased density.

Under a cross-sectional study, the probability of recruiting a longer-lived individual is higher than that of recruiting a shorter-lived individual and consequently, the prevalent population is not representative of the incident population because the survival times associated with the prevalent cases can be considered as a biased sample. In this direction, as shown in [9] covariates that accompany length-biased survival times follow a biased density and cannot be representative of covariates in the general population. Corresponding with Figure 1.3, which illustrates Weibull density along with its associated length-biased density, Figure 1.4 shows that using the prevalent cases instead of incident cases one can overestimate the survival function of the true population.

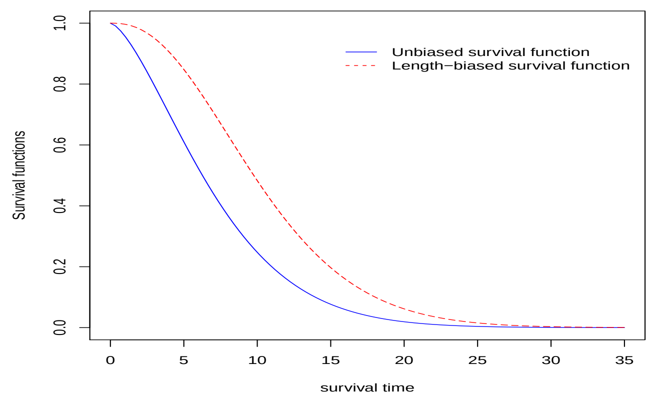

Figure 1.4: Unbiased survival function versus length-biased survival function.

Most of the literature on length-biased sampled data concentrates on statistical methods for survival function [13] and [48], estimating density function [7] and [30], kernel smoothing [49], proportional hazards models [51], covariate bias induced by length-biased sampling of failure times [9]. However, under length-biased sampling, measures of the degree of dependence between the survival time and the covariates appear to have been neglected in the literature. In this direction, the principal objective in many studies, when regression models are used in survival analysis, is to extract the relationship between the survival time and the covariates. For example, it is of interest to know if there exists any correlation between survival times with dementia and associated covariates such that age at onset, sex, years of education. In fact, using the multiple linear regression model when the errors follow the normal distribution, the multiple correlation coefficient is the most familiar measure of dependence between the dependent and the independent variable. However, this measure cannot be employed in the presence of the censoring and truncation or under non normality of the errors. For more general models used in survival analysis such as Weibull regression model or Cox's proportional hazards model, a measure of dependence between a censored time and covariates can be defined using the concept of information gain (see Kent [33], Kent and O'Quigley [34]). This concept generalizes more common measures such as the multiple correlation coefficient. Kent and O'Quigley [34] used Fraser information [18] to extend the work of Linfoot [37] and provided a dependence measure, based on information gain, for right-censored survival data. We propose two different methods to extend Kent's [33] dependence measure. This thesis is organized as follows: in Chapter 2, we first review the basic notions of survival analysis and then we expose the concept of information gain. In addition, we examine the dependence measure, for right-censored survival data, proposed by Kent and O'Quigley [34]. We end this second chapter by presenting under length-biased sampling, the length-biased distribution of the survival time and the biased distribution of the covariates. In Chapter 3, we extend Kent's [33] dependence measure in the context of length-biased sampling, without censoring for the case of one continuous covariate. We establish a link between the conditional and joint information gain. We further develop the first method : kernel density estimation with a regression procedure to estimate the dependence measure for length-biased data. An extension of the first method and results of the last chapter will be detailed in Chapter 4: we derive the dependence measure for length-biased data without censoring for the case of several continuous covariates. We focus our attention on the general case: partial dependence measure. To estimate this measure, we generalize the first method for the univariate case (one covariate) in Chapter 3. The last section is devoted to examining the consistency of the proposed estimators. In Chapter 5, we review some general notions of copulas. Based on the concept of information gain, we develop the second method: parametric copula to obtain the dependence measure between a survival time and one continuous covariate, without censoring. We adapt this method under length-biased sampling. For the purpose of implementation, we propose in Chapter 6 some new simulation algorithms for the two proposed methods: kernel density estimation with a regression procedure and parametric copula. Chapter 7 is devoted to applications. We investigate the performance of the two proposed methods. We conclude with a summary of the contributions of the thesis and discuss new avenues of research.

Chapter 2

Preliminaries

In this chapter, we recall the basic notions of survival analysis, in particular survival functions and regression models frequently used for lifetime data analysis. The concept of information gain for general statistical models and length-biased sampling will be exposed to derive a dependence measure for length-biased survival data.

2.1 Some notions of survival analysis

In the current section, we review some quantities and relations used in survival analysis. Next, we consider regression models for survival data.

2.1.1 Survival time functions

Suppose that the random variable (r.v.) T, which denotes survival time, is absolutely continuous. From [36], the distribution of T is usually characterized by equivalently the survival function, the probability density function and the hazard function.

Definition 2.1.1 *The survival function, denoted by $S(t)$, is defined as the probability that an individual survives up to time t:*

$$S(t) = \mathbb{P}\left(T > t\right) = 1 - F(t), \tag{2.1.1}$$

where $F(t)$, the distribution function of T, is the probability that an individual fails before t.

Here, the survival function $S(t)$ is a nonincreasing function of time t with the properties $S(t) = 1$ if $t = 0$ and $S(t) = 0$ if $t = \infty$.

Definition 2.1.2 *The probability density function of T is defined as the probability of failure in a small interval per unit time. It can be expressed as*

$$f(t) = \lim_{\Delta t \to 0} \frac{\mathbb{P}\left[\textit{an individual dying in the interval } (t, t+\Delta t)\right]}{\Delta t}. \tag{2.1.2}$$

If the distribution function of T has a derivative at t then

$$f(t) = \lim_{\Delta t \to 0} \frac{\mathbb{P}\left(t < T < t + \Delta t\right)}{\Delta t} = F'\left(t\right) = -S'\left(t\right). \tag{2.1.3}$$

Here, $f(t)$ is the probability density function (PDF) of T. As shown in Klein and Moeschberger [31], a very useful property of the mean of f is

$$\mu = \mathrm{E}\left[T\right] = \int_0^\infty S\left(t\right) dt. \tag{2.1.4}$$

Definition 2.1.3 *The hazard function of survival time T is defined as the limit of the probability that an individual fails in a very short time interval given that the individual has survived to time t:*

$$h(t) = \lim_{\Delta t \to 0} \frac{\mathbb{P}\left(t < T < t + \Delta t \,| T > t\right)}{\Delta t}. \tag{2.1.5}$$

The hazard function $h(t)$ can be expressed in terms of the survival function $S(t)$ and the probability density function $f(t)$:

$$h(t) = \frac{f(t)}{S(t)} = -\frac{S'(t)}{S(t)} = -\frac{d}{dt}\log\left\{S(t)\right\}. \tag{2.1.6}$$

The hazard function is also known as the instantaneous failure rate, force of mortality, conditional mortality rate and age-specific failure rate.

2.1.2 Right-censored and left-truncated data

One of the characteristics of survival data is the existence of incomplete observations. In fact, data are often collected partially, especially in the presence of various type of censoring and truncation. In survival analysis, the most frequent are left truncation and right censoring having, respectively, the following definitions [31].

Definition 2.1.4 *Left truncation occurs when subjects enter a study at a specific time (not necessarily the origin for the event of interest) and are followed from this delayed-entry time until the event occurs or the subject is censored.*

Definition 2.1.5 *Right censoring occurs when a subject leaves the study before an event occurs or the study ends before the event has occurred.*

When the experiment involves a right-censoring process, the corresponding observations can be represented by random vector (T, δ), where δ indicates whether the survival time X is observed $(\delta = 1)$ or not $(\delta = 0)$ and T equal to X if the survival time is observed and to C_r if it is the right-censored time, i.e., $T = \min(X, C_r)$. In this case the sample for n individuals takes the form of the pairs (T_i, δ_i), $i = 1, \ldots, n$.

2.1.3 Regression models for survival data

The use of regression models is an important way to understand and exploit a relationship between a survival time and covariates. As before, let T denote the time to some specific event. The data based on a sample of size n, consist of the triple $(T_i, \delta_i, \mathbf{Z}_i)$, $i = 1, \ldots, n$, where T_i is the time on study for the ith individual, δ_i is the corresponding event indicator ($\delta_i = 1$ if the event has occurred and $\delta_i = 0$ if the survival time is right-censored) and $\mathbf{Z}_i(t) = (Z_{i1}(t), \ldots, Z_{id}(t))'$ is the vector of covariates or risk factors at time t. Here, the covariates may be serial blood pressure measurements, current disease status, etc., or they may not depend on the time such as treatment, race, disease status, age, weight, and temperature, etc. For all that follows, we consider only the fixed-covariate vector $\mathbf{Z}_i = (Z_{i1}, \ldots, Z_{id})'$ independent of time for the modeling of covariates effects on survival time.

Two popular approaches in the statistical literature ([31], [36]) are Accelerated Failure Time (AFT) and Proportional Hazards models (PH). The Accelerated Failure Time model can be considered as the classical linear regression approach, where the log survival time is modelled. A linear regression model for $Y = \log\{T\}$ is

$$Y = \mu + \boldsymbol{\beta}'\mathbf{Z} + \sigma\varepsilon, \tag{2.1.7}$$

where μ is an intercept, $\boldsymbol{\beta}'$ is the transpose of a vector of regression coefficients $\boldsymbol{\beta}$, σ is a scale parameter and ε is the error variate, independent of $\mathbf{Z}$. Under the AFT model (2.1.7), the distribution of the error ε can be identified once the distribution of the survival time T is known. The following table describes some useful distributions of the lifetime T when the AFT model is used.

T	$\log\{T\}$
Exponential	Extreme value
Weibull	Extreme value
Log-logistic	Logistic
Lognormal	Normal

Table 2.1: Useful densities under the AFT model.

To see why this model is called the AFT, let $S_0(t)$ be the survival function of $T = e^Y$ when $\mathbf{Z}$ equals zero. So, $S_0(t)$ is the survival function of $T = e^{\mu+\sigma\varepsilon}$. Now, as shown by Lawless [36], the survival time of T conditional on $\mathbf{Z} = \mathbf{z}$ can be deduced from the model (2.1.7) as follows

$$S\left(t|\mathbf{z}\right) = S_0\left(te^{-\boldsymbol{\beta}'\mathbf{z}}\right). \tag{2.1.8}$$

It is easy to see by the last equation that, the effect of the covariates in the original time scale is to change the time scale by a factor $e^{-\boldsymbol{\beta}'\mathbf{z}}$ and that the time is either accelerated or decelerated, depending on the sign of $\boldsymbol{\beta}'\mathbf{z}$. Note that, based on (2.1.6), the hazard function of T given $\mathbf{Z} = \mathbf{z}$ can be expressed by

$$h(t|\mathbf{z}) = h_0\left(te^{-\boldsymbol{\beta}'\mathbf{z}}\right)e^{-\boldsymbol{\beta}'\mathbf{z}}, \tag{2.1.9}$$

where $h_0(t)$ is the baseline hazard function of T when $\mathbf{Z}$ equals zero.

The Proportional hazards model, is a class of models with the interesting property that different individuals have hazard functions proportional to one another. The ratio of hazard functions $h(t|\mathbf{z}_1)/h(t|\mathbf{z}_2)$ for two individuals with covariate vectors $\mathbf{z}_1,\ \mathbf{z}_2$ does not vary with time t which implies that $h(t|\mathbf{z})$ can be written as

$$h(t|\mathbf{z}) = h_0(t)\,\varphi(\mathbf{z}), \tag{2.1.10}$$

where $\varphi(\mathbf{z})$ is any positive function and $h_0(t)$ is the baseline hazard function of T when $\varphi(\mathbf{z}) = 1$. From (2.1.6) and (2.1.10), the survival function of T given $\mathbf{Z} = \mathbf{z}$ can be expressed in terms of a baseline survival time as

$$S(t|\mathbf{z}) = (S_0(t))^{\varphi(\mathbf{z})}. \tag{2.1.11}$$

A very important special case is the Cox model [14] which assumes that $\varphi(\mathbf{z}) = e^{\boldsymbol{\beta}'\mathbf{z}}$. In this case, (2.1.10) takes the form

$$h(t|\mathbf{z}\ ) = h_0(t)\,e^{\boldsymbol{\beta}'\mathbf{z}}. \tag{2.1.12}$$

The partial likelihood [14] can be constructed from the data sets as

$$L_{\text{Cox}}(\boldsymbol{\beta}) = \prod_{i=1}^{D} \frac{e^{\boldsymbol{\beta}'\mathbf{z}_i}}{\sum\limits_{j\in R(T_i)} e^{\boldsymbol{\beta}'\mathbf{z}_j}}, \tag{2.1.13}$$

where D is the number of deaths observed among the n subjects in the study, $T_1 < \cdots < T_D$ are distinct failure times and $R(T_i)$ is the set of all individuals still at risk just before T_i. The Partial likelihood, given in (2.1.13), takes into account the right censoring process and does not depend on the baseline hazard function $h_0(t)$. We can estimate $\boldsymbol{\beta}$ without knowing $h_0(t)$ by maximizing (2.1.13). A positive regression coefficient for an explanatory variable means that the hazard is higher, and thus the prognosis worse while, a negative regression coefficient implies a better prognosis for patients with higher values of that variable.

A difference between PH and AFT models is that, AFT models compare survival functions while PH models compare hazard functions. In addition, the effect of covariates in AFT models is proportional with respect to time while in PH models it is multiplicative with respect to hazard function. Note that, both the exponential and Weibull distributions satisfy the assumption of both the AFT and PH models since these distributions can be written in the form of (2.1.8) and (2.1.11).

2.2 Dependence measure based on the concept of information gain

If the dependence between two random variables is modelled parametrically then the concept of information gain can be used to define a measure of dependence which is, essentially, based on the definition of likelihood. We first explain this concept and then discuss the dependence measure, for right-censored survival data, proposed by Kent and O'Quigley [34].

2.2.1 Concept of information gain

Let X be a r.v. with true fixed density $g(x)$ and consider two families of parametric models $\{f(x;\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta_i\}$ $(i = 0, 1)$ with $\Theta_0 \subset \Theta_1$. The Fraser information [18] of $\boldsymbol{\theta}$ under $g(x)$ is defined by the expected log-likelihood

$$\Phi(\boldsymbol{\theta}) = \int \log\{f(x;\boldsymbol{\theta})\}\, g(x)dx. \tag{2.2.1}$$

For the comparison between the best fitting models under Θ_0 and Θ_1, Kent [33] defines the information gain to be

$$\Gamma(\boldsymbol{\theta}_1, \boldsymbol{\theta}_0) = 2\{\Phi(\boldsymbol{\theta}_1) - \Phi(\boldsymbol{\theta}_0)\}, \tag{2.2.2}$$

where $\boldsymbol{\theta}_i$ maximizes $\Phi(\boldsymbol{\theta})$ over Θ_i. Here, $\Gamma(\boldsymbol{\theta}_1, \boldsymbol{\theta}_0)$ is always nonnegative since $\Theta_0 \subset \Theta_1$ and if $g(x) = f(x; \boldsymbol{\theta}^*)$ for some $\boldsymbol{\theta}^* \in \Theta_1$, then (2.2.2) reduces to twice the Kullback-Leibler [35] information gain.

Example 2.2.1 Let X be a r.v. with true fixed density $g(x)$ and consider two families of parametric models $\{f(x;\theta), \theta \in \Theta_i\}$ $(i = 0, 1)$ with $\Theta_0 \subset \Theta_1$. Suppose that

$$f(x;\theta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \tag{2.2.3}$$

where $\theta = \mu$. By using (2.2.1), the information gain given in (2.2.2) under $g(x)$ is

$$\begin{aligned}
\Gamma(\theta_1, \theta_0) &= 2\{\Phi(\theta_1) - \Phi(\theta_0)\} \\
&= 2\left\{\int_{-\infty}^{\infty} \log\{f(x;\theta_1)\} g(x)dx - \int_{-\infty}^{\infty} \log\{f(x;\theta_0)\} g(x)dx\right\},
\end{aligned}$$

where $\theta_1 = \mu_1$, $\theta_0 = \mu_0$ and $\mu_0 \leq \mu_1$. The last equation becomes

$$\begin{aligned}
\Gamma(\theta_1, \theta_0) &= 2\left\{\int_{-\infty}^{\infty} \log\left\{\frac{f(x;\mu_1)}{f(x;\mu_0)}\right\} g(x)dx\right\} \\
&= 2\mathrm{E}_g\left[\log\left\{\frac{f(X;\mu_1)}{f(X;\mu_0)}\right\}\right]. \qquad (2.2.4)
\end{aligned}$$

Now,

$$\begin{aligned}
\log\left\{\frac{f(x;\mu_1)}{f(x;\mu_0)}\right\} &= \log\left\{\frac{\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma}\right)^2}}{\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}\right\} \\
&= -\frac{1}{2}\left(\frac{x-\mu_1}{\sigma}\right)^2 + \frac{1}{2}\left(\frac{x-\mu_0}{\sigma}\right)^2 \\
&= \frac{1}{2\sigma^2}\left(2x\mu_1 - \mu_1^2 - 2x\mu_0 + \mu_0^2\right) \\
&= \frac{1}{2\sigma^2}\left((2\mu_1 - 2\mu_0)x - \mu_1^2 + \mu_0^2\right). \qquad (2.2.5)
\end{aligned}$$

Substituting (2.2.5) into (2.2.4), we get

$$\begin{aligned}\Gamma\left(\theta_1,\theta_0\right) &= 2\mathrm{E}_g\left[\frac{1}{2\sigma^2}\left((2\mu_1-2\mu_0)X-\mu_1^2+\mu_0^2\right)\right]\\ &= \frac{2}{\sigma^2}\left(\mu_1-\mu_0\right)\left[\mathrm{E}_g\left[X\right]-\frac{\mu_0+\mu_1}{2}\right].\end{aligned}$$

If the true density $g(x)=f(x,\theta_1)$ then the information gain is

$$\Gamma\left(\theta_1,\theta_0\right)=\frac{1}{\sigma^2}\left(\mu_1-\mu_0\right)^2.$$

So, the information gain under two Gaussian distributions with different means and the same variance is proportional to the squared distance between the two means.

As information gain increases, the model under Θ_1 gets closer to the true density $g(x)$ compared with the model under Θ_0 but, how does this relate to dependence?
Let (Y,Z) be a random vector which plays the role of X. Suppose that Y and Z have true joint density $g(y,z)$, modelled by a parametric family $\{f(y,z;\boldsymbol{\theta}),\boldsymbol{\theta}\in\Theta_1\}$ such that $\boldsymbol{\theta}=(\boldsymbol{\alpha},\boldsymbol{\lambda})$, where $\boldsymbol{\alpha}$ and $\boldsymbol{\lambda}$ are p-dimensional and q-dimensional vectors, respectively. Suppose that Y and Z are modelled as independent random variables under $\Theta_0=\boldsymbol{\theta}_0:\boldsymbol{\alpha}=\mathbf{0}\}$. Thus, $\boldsymbol{\alpha}$ measures the parametric dependence between Y and Z. The joint information gain is

$$\Gamma\left(\boldsymbol{\theta}_1,\boldsymbol{\theta}_0\right)=2\left\{\Phi(\boldsymbol{\theta}_1)-\Phi(\boldsymbol{\theta}_0)\right\}, \tag{2.2.6}$$

where

$$\Phi(\boldsymbol{\theta})=\iint\log\left\{f(y,z;\boldsymbol{\theta})\right\}g(y,z)dydz. \tag{2.2.7}$$

Kent [33] proposes

$$\rho_J^2\left(Y,Z\right)=1-\exp\left\{-\Gamma\left(\boldsymbol{\theta}_1,\boldsymbol{\theta}_0\right)\right\}, \tag{2.2.8}$$

as a measure of dependence between Y and Z. If Y is modelled conditionally on Z by a parametric family $\{f(y|z;\boldsymbol{\theta}),\boldsymbol{\theta}\in\Theta_1\}$, Kent [33] uses conditional Fraser information on the expected conditional log-likelihood

$$\Phi_C(\boldsymbol{\theta})=\iint\log\left\{f(y|z;\boldsymbol{\theta})\right\}g(y,z)dydz, \tag{2.2.9}$$

to adapt the joint information gain (2.2.6) to a conditional information gain. The conditional measure of dependence of Kent [33] is

$$\rho_C^2\left(Y|Z\right) = 1 - \exp\left\{-\Gamma_C\left(\boldsymbol{\theta}_1, \boldsymbol{\theta}_0\right)\right\}, \tag{2.2.10}$$

where $\Gamma_C\left(\boldsymbol{\theta}_1, \boldsymbol{\theta}_0\right) = 2\left\{\Phi_C(\boldsymbol{\theta}_1) - \Phi_C(\boldsymbol{\theta}_0)\right\}$. The measures ρ_J^2 and ρ_C^2 have the following properties:

- if Y and Z are two independent random variables denoted $(Y \perp Z)$, then $\rho_J^2 = 0$ ($\rho_C^2 = 0$ in conditional models).
- $0 \leq \rho_J^2 < 1$. This is also true for ρ_C^2.
- under normal models, ρ_J^2 reduces to the product-moment correlation and ρ_C^2 is the squared multiple correlation coefficient.

The next important step is to provide an estimator of information gain. Suppose that $Y_1, \ldots, Y_n$ is a sequence of independent observations from $g(y)$ and we wish to estimate $\Gamma\left(\boldsymbol{\theta}_1, \boldsymbol{\theta}_0\right)$ in (2.2.2). For n large, Kent [33] suggests

$$\hat{\Gamma}\left(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_0\right) = \frac{2}{n}\left\{\sum_{i=1}^{n}\log\left\{f\left(Y_i; \hat{\boldsymbol{\theta}}_1\right)\right\} - \sum_{i=1}^{n}\log\left\{f\left(Y_i; \hat{\boldsymbol{\theta}}_0\right)\right\}\right\}, \tag{2.2.11}$$

as an estimator of $\Gamma\left(\boldsymbol{\theta}_1, \boldsymbol{\theta}_0\right)$, where $\hat{\boldsymbol{\theta}}_1$ is the maximum likelihood estimator of $\boldsymbol{\theta}_1$ under Θ_1 and $\hat{\boldsymbol{\theta}}_0$ is the maximum likelihood estimator of $\boldsymbol{\theta}_0$ under Θ_0 and $\hat{\Gamma}\left(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_0\right)$ converges in probability to $\Gamma\left(\boldsymbol{\theta}_1, \boldsymbol{\theta}_0\right)$, see Kent [33]. Note that, $n\hat{\Gamma}\left(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_0\right)$ is the usual likelihood ratio test statistic for testing $\boldsymbol{\theta}_0 \in \Theta_0$ against $\boldsymbol{\theta}_1 \in \Theta_1$. In the case where the sample size n is small, Kent [33] uses some rather strong assumptions to provide a different estimator for the information gain.

2.2.2 Dependence measure for right-censored data

The Cox model is most popular for analyzing censored survival data in medical research. Cox's model specifies the conditional hazard function of a continuous survival time T given the $(p+q)$-dimensional explanatory variable $\mathbf{Z} = \left(\mathbf{Z}^{(1)}, \mathbf{Z}^{(2)}\right)$ as

$$h\left(t\left|\mathbf{z}\right.\right) = h_0\left(t\right) \exp\left\{\boldsymbol{\beta}^{(1)\prime}\mathbf{z}^{(1)} + \boldsymbol{\beta}^{(2)\prime}\mathbf{z}^{(2)}\right\}, \tag{2.2.12}$$

where $h_0\left(t\right)$ is an unspecified baseline hazard function and $\boldsymbol{\beta} = \left(\boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)}\right)$ is the vector of regression coefficients. Kent and O'Quigley [34] mentioned that the model given in (2.2.12) can be reduced through a monotone increasing transformation $T^* = \phi(T)$ to a model with the same regression vector $\boldsymbol{\beta}$. In particular, if $T^* \sim \text{Weibull}\left(\alpha, \mu\right)$ then the conditional distribution of T^* given $\mathbf{Z} = \mathbf{z}$ is $\text{Weibull}\left(\alpha, \mu \exp\{-\boldsymbol{\beta}'\mathbf{z}/\alpha\}\right)$ and $Y^* = \log\left(T^*\right)$ follows an AFT model given in (2.1.7) which is a Weibull linear regression model

$$Y^* = -\sigma\mu - \sigma\boldsymbol{\beta}^{(1)\prime}\mathbf{z}^{(1)} - \sigma\boldsymbol{\beta}^{(2)\prime}\mathbf{z}^{(2)} + \sigma\varepsilon, \tag{2.2.13}$$

where $\alpha{=}\sigma^{-1}$ and the error variate ε follows a Gumbel distribution with variance $\psi'(1) = 1.645$. Here, $\psi'(\cdot)$ is the derivative of the digamma function, see section 3 in Kent and O'Quigley [34]. We seek a measure of partial dependence between Y^* and $\mathbf{Z}^{(1)}$, allowing for the regression on $\mathbf{Z}^{(2)}$ denoted by $\rho^2\left(Y^*, \mathbf{Z}^{(1)}\left|\mathbf{Z}^{(2)}\right.\right)$. Some possible measures of dependence are the squared multiple partial product-moment correlation coefficient ρ^2_{PM} and a useful approximation for the Weibull regression model $\rho^2_{W.A}$ [34] given, respectively, by

$$\rho^2_{PM} = \frac{A}{A + 1.645} \quad \text{and} \quad \rho^2_{W.A} = \frac{A}{A+1}, \tag{2.2.14}$$

where $A = \boldsymbol{\beta}^{(1)\prime}\Omega_{11.2}\boldsymbol{\beta}^{(1)}$ and $\Omega_{11.2} = \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}$. Here, Ω is the covariance matrix of $\mathbf{Z}$ partitioned in the usual way. We note that, ρ^2_{PM} and $\rho^2_{W.A}$ can be estimated, respectively, by

$$\hat{\rho}^2_{PM} = \frac{\hat{A}}{\hat{A} + 1.645} \quad \text{and} \quad \rho^2_{W.A} = \frac{\hat{A}}{\hat{A}+1}, \tag{2.2.15}$$

where $\hat{A} = \hat{\boldsymbol{\beta}}^{(1)\prime}S_{11.2}\hat{\boldsymbol{\beta}}^{(1)}$, $S_{11.2} = S_{11} - S_{12}S_{22}^{-1}S_{21}$, $S = \text{cov}(\mathbf{Z})$ and $\hat{\boldsymbol{\beta}}^{(1)}$ is the estimator of $\boldsymbol{\beta}^{(1)}$ that maximizes $L_{\text{Cox}}\left(\boldsymbol{\beta}^{(1)}\right)$ given in (2.1.13).

A stronger notion of dependence can be defined using the concept of information gain. For that, let $f_T(t)$ and $G\left(dz\right)$ denote the density function and marginal distribution, respectively, of the right-censored time T and vector of the covariates $\mathbf{Z}$ partitioned above. Suppose that, the conditional distribution of $Y = \log(T)$ given $\mathbf{Z}$ follows the AFT model (2.2.13), where ε has some specified density function $f_{\varepsilon}\left(\varepsilon\right)$ and ε is independent of $\mathbf{Z}$. Let $\boldsymbol{\theta} = (\boldsymbol{\beta}, \mu, \sigma^2)$ denote the parameter of the model ($\sigma > 0$ and $\boldsymbol{\beta}$ is partitioned with respect to $\mathbf{Z}$) and $\boldsymbol{\theta}_1 = (\boldsymbol{\beta}_1, \mu_1, \sigma_1^2)$ denotes the true value of the parameter. Generally $\boldsymbol{\beta}_1^{(1)} \neq 0$. The objective here, is to measure $\rho_C^2\left(Y, \mathbf{Z}^{(1)} \middle| \mathbf{Z}^{(2)}\right)$. So, we have to test

$$H_0:\ \boldsymbol{\beta}_1^{(1)} = 0 \quad \text{vs} \quad H_1:\ \boldsymbol{\beta}_1^{(1)} \neq 0.$$

The measure of the distance between H_0 and H_1 is given by twice the Kullback-Leibler [35] information gain as

$$\Gamma_C = 2\left\{\Phi\left(\boldsymbol{\theta}_1, \boldsymbol{\theta}_1\right) - \Phi\left(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1\right)\right\}, \tag{2.2.16}$$

where $\Phi\left(\boldsymbol{\theta}, \boldsymbol{\theta}_1\right)$ is the expected log-likelihood given by

$$\Phi\left(\boldsymbol{\theta}, \boldsymbol{\theta}_1\right) = \iint \log\left\{f\left(y\left|\mathbf{z}; \boldsymbol{\theta}\right.\right)\right\} f\left(y\left|\mathbf{z}; \boldsymbol{\theta}_1\right.\right) dy G\left(d\mathbf{z}\right), \tag{2.2.17}$$

and $\boldsymbol{\theta}_0$ is the value of $\boldsymbol{\theta}$ maximizing $\Phi\left(\boldsymbol{\theta}, \boldsymbol{\theta}_1\right)$ under H_0. Based on the conditional Fraser information, Kent and O'Quigley [34] proposed a measure of dependence between Y and $\mathbf{Z}^{(1)}$ after allowing for the regression on $\mathbf{Z}^{(2)}$ as

$$\rho_C^2\left(Y, \mathbf{Z}^{(1)} \middle| \mathbf{Z}^{(2)}\right) = 1 - \mathrm{e}^{-\Gamma_C}. \tag{2.2.18}$$

To estimate the conditional information gain given in (2.2.16), Kent and O'Quigley [34] suggested the following two approaches.

Approach 1 (Without censoring lifetimes using log-likelihood)

Let $(y_i, \mathbf{z}_i)$, $i = 1, \ldots, n$ be a sample from the model (2.2.13). The conditional information gain based on the observed distribution of $(Y, \mathbf{Z})$ can be estimated by

$$\hat{\Gamma}_C = \frac{2}{n}\left(\sum_{i=1}^{n} \log\left\{f\left(y_i \,|\mathbf{z}_i;\ \hat{\boldsymbol{\theta}}_1\right)\right\} - \sum_{i=1}^{n} \log\left\{f\left(y_i \,|\mathbf{z}_i;\ \hat{\boldsymbol{\theta}}_0\right)\right\}\right), \tag{2.2.19}$$

where $\hat{\boldsymbol{\theta}}_1$ and $\hat{\boldsymbol{\theta}}_0$ maximize the observed log-likelihood, $\log\left\{\prod_{i=1}^{n} f\left(y_i \,|\mathbf{z}_i; \boldsymbol{\theta}\right)\right\}$ over $\boldsymbol{\theta}$ satisfying H_1 and H_0, respectively. In this case, we have $\hat{\Gamma}_C = \Lambda/n$, where Λ is the usual log-likelihood ratio statistic for testing H_0 against H_1.

Approach 2 (With censoring lifetimes and/or unknown monotone transformation)

This approach is based on the fitted density for Y given $\mathbf{Z}$, with any estimate $\tilde{\boldsymbol{\theta}}_1$ of $\boldsymbol{\theta}_1$. So, given $\tilde{\boldsymbol{\theta}}_1$ and under hypothesis H_0 let $\tilde{\boldsymbol{\theta}}_0$ maximize

$$\frac{1}{n}\sum_{i=1}^{n}\int \log\left\{f\left(y \,|\mathbf{z}_i; \boldsymbol{\theta}\right)\right\} f\left(y \,|\mathbf{z}_i;\ \tilde{\boldsymbol{\theta}}_1\right) dy.$$

Then, the conditional information gain can be estimated by

$$\tilde{\Gamma}_C = \frac{2}{n}\sum_{i=1}^{n}\int \log\left\{f\left(y \,|\mathbf{z}_i;\ \tilde{\boldsymbol{\theta}}_1\right) / f\left(y \,|\mathbf{z}_i;\ \tilde{\boldsymbol{\theta}}_0\right)\right\} f\left(y \,|\mathbf{z}_i;\ \tilde{\boldsymbol{\theta}}_1\right) dy. \tag{2.2.20}$$

According to Kent and O'Quigley [34], ρ_C^2 and Γ_C have the following properties:

- $0 \leq \rho_C^2 < 1$, $\rho_C^2 \to 1$ as $\|\boldsymbol{\beta}\| \to \infty$ and $\rho_C^2 = 0$ under H_0.
- ρ_C^2 is invariant under linear transformations of $Y, Z^{(1)}$ and $Z^{(2)}$.
- ρ_C^2 depends only on the scaled regression coefficient $\boldsymbol{\beta}$ and the marginal distribution $G\left(d\mathbf{z}\right)$ of $\mathbf{Z}$, but not on μ or σ.
- Under H_0, the limiting distribution of $n\hat{\rho}_C^2$ is χ_p^2.
- Under H_1, $\sqrt{n}\left(\tilde{\Gamma}_C - \Gamma_C\right) \sim N\left(0,\ v\right)$ for some $v > 0$.

2.3 Weighted and length-biased distributions

Consider a natural mechanism generating a r.v. X with PDF $f_{uw}(x)$. For drawing a random sample of observation on X, a specific method of selection is used which gives the same chance of including in the sample any observation produced by the original mechanism. In practice it may happen that the relative chances of inclusion of two observations x_1 and x_2 are $w(x_1)/w(x_2)$, where $w(x)$ is non-negative weight function. Then, the recorded X to be denoted by X^w has the PDF

$$g_w(x) = \frac{w(x) f_{uw}(x)}{\mu_w}, \quad w(x) > 0, \tag{2.3.1}$$

where $\mu_w = \int w(x) f_{uw}(x) dx < \infty$ is a normalizing factor obtained to make the total probability equal to one, $g_w(x)$ is called weighted PDF and $f_{uw}(x)$ is the original or unweighted PDF. Rao [41] introduced distributions of the type (2.3.1) with an arbitrary non-negative weight function $w(x)$ and gave some practical examples, where $w(x) = x$ or $w(x) = x^\alpha$ are appropriate. When $w(x) = x$, the weighted density $g_w(x)$ is also called length-biased PDF, defined by

$$f_{LB}(x) = \frac{x f_U(x)}{\mu}, \quad x > 0, \tag{2.3.2}$$

where $\mu = \int x f_U(x) dx < \infty$, and the corresponding unbiased density is denoted by f_U.

2.3.1 Length-biased sampling

From Asgharian et al. [5], the observed data for the prevalent cases under cross-sectional study denoted by (X_i, δ_i), $i = 1, \ldots, n$, are described in the following diagram.

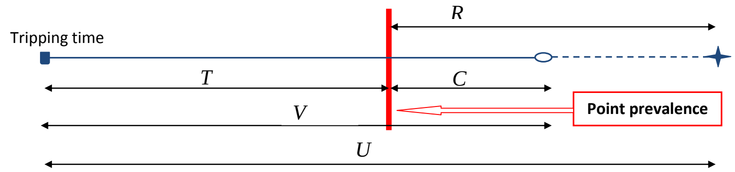

Figure 2.1: Observation of prevalent case.

where for the ith subject

$$\tilde{X}_i = \begin{cases} U_i = T_i + R_i & if\ \delta_i = 1, \\ V_i = T_i + C_i & if\ \delta_i = 0, \end{cases}$$

- U_i - total failure lifetime (complete observation).
- T_i - truncation variable (recurrent time), measures the time between onset and a fixed recruitment time.
- R_i - residual lifetime, measures the time between recruitment and failure.
- V_i - total censoring lifetime (incomplete observation).
- C_i - residual censored lifetime, measures the time between recruitment and censoring.
- $\delta_i = 1_{\{R_i \le C_i\}}$.

We note that, under a cross-sectional study, the observations (X_i, δ_i), $i = 1, \ldots, n$ are independent but, U_i and V_i are not since they have a common left truncation time T_i. In these cases, the censoring process is informative. In addition, C_i and (T_i, R_i) are independent. To see why the U_i's have a length-biased density, let U be a r.v. which denotes the true failure time with density function $f_U(u)$ and let T be the left truncation time with density function $g(t)$. Under a cross sectional study, the subjects are observed only if $U \ge T$. Suppose that U and T are independent. Then, the joint density of (U, T) given $U \ge T$ can be expressed as

$$f_{U,T}\left(u, t | U \ge T\right) = \frac{f_{U,T}\left(u, t\right)}{\mathbb{P}\left(U \ge T\right)} = \frac{f_U\left(u\right) g\left(t\right)}{\mathbb{P}\left(U \ge T\right)}, \tag{2.3.3}$$

if $U \geq T$ and 0 otherwise. Now,

$$\begin{aligned}\mathbb{P}(U \geq T) &= \int_0^\infty \mathbb{P}(U \geq t|T = t)\, g(t)dt \\ &= \int_0^\infty \mathbb{P}(U \geq t) g(t)dt \\ &= \int_0^\infty S_U(t) g(t)dt.\end{aligned}$$

If the onset times follow a stationary Poisson process, the truncation times are uniformly distributed over the interval $(0, c)$ and $\mathbb{P}(U \geq c) = 0$, see Wang [50]. From (2.1.4), it follows that

$$\mathbb{P}(U \geq T) = \frac{\mu}{c}, \tag{2.3.4}$$

where μ is the mean failure time. Therefore, Equation (2.3.3) becomes

$$f_{U,T}(u, t|U \geq T) = \frac{f_U(u)}{\mu}\cdot \tag{2.3.5}$$

The density function of U conditional on $U \geq T$ is then

$$f(u|U \geq T) = \int_0^u f_{U,T}(u, t|U \geq T)dt = \int_0^u \frac{f_U(u)}{\mu} dt,$$

and hence,

$$f(u|U \geq T) = \frac{u f_U(u)}{\mu} = f_{LB}(u). \tag{2.3.6}$$

2.3.2 Likelihood approaches under length-biased sampling

Let f be a joint density function of the observed vector of data $w_i = (t_i, r_i \wedge c_i, \delta_i)$, $i = 1, \ldots, n$. Vardi [48] derived the following likelihood as

$$\begin{aligned}L(\boldsymbol{\theta}) &= \prod_{i=1}^n f(t_i, r_i \wedge c_i, \delta_i | U \geq T; \boldsymbol{\theta}) \\ &= \prod_{i=1}^n \left(\frac{f_U(t_i + r_i; \boldsymbol{\theta})}{\mu(\boldsymbol{\theta})}\right)^{\delta_i} \left(\int_{s \geq t_i + c_i} \frac{f_U(s; \boldsymbol{\theta})}{\mu(\boldsymbol{\theta})} ds\right)^{1-\delta_i},\end{aligned} \tag{2.3.7}$$

where f_U is the unbiased density function and $\mu(\boldsymbol{\theta})$ is the mean of f_U. The asymptotic properties of the maximum likelihood estimators (MLE's) obtained from (2.3.7) under cross-sectional sampling are derived by Asgharian et al. [5]. When covariates are introduced in the model, the conditional likelihood for $(w_i, \mathbf{z}_i)$, $i = 1, \ldots, n$ simply extends the above likelihood as follows

$$\begin{aligned} L_C(\boldsymbol{\theta}) &= \prod_{i=1}^{n} f(t_i, r_i \wedge c_i, \delta_i | \mathbf{z}_i, U \geq T; \boldsymbol{\theta}) \\ &= \prod_{i=1}^{n} \left(\frac{f_U(t_i + r_i | \mathbf{z}_i; \boldsymbol{\theta})}{\mu(\mathbf{z}_i; \boldsymbol{\theta})} \right)^{\delta_i} \left(\int_{s \geq t_i + c_i} \frac{f_U(s | \mathbf{z}_i; \boldsymbol{\theta})}{\mu(\mathbf{z}_i; \boldsymbol{\theta})} ds \right)^{1-\delta_i}, \end{aligned} \tag{2.3.8}$$

where $\mu(\mathbf{z}_i; \boldsymbol{\theta}) = \mathrm{E}[U | \mathbf{z}_i]$. Here, the likelihood ignores the sampling distribution of the covariates. In order to incorporate the covariates in a likelihood function, we work with the joint likelihood [9]

$$L_J(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(w_i, \mathbf{z}_i | U \geq T; \boldsymbol{\theta}) = \prod_{i=1}^{n} f(t_i, r_i \wedge c_i, \mathbf{z}_i, \delta_i | U \geq T; \boldsymbol{\theta}). \tag{2.3.9}$$

By using the relation between the joint and conditional density functions we can write the likelihood, given in (2.3.9), for the observation $(w_i, \mathbf{z}_i)$ as

$$\begin{aligned} L_{J,i}(\boldsymbol{\theta}) &= f(t_i, r_i \wedge c_i, \mathbf{z}_i, \delta_i | U \geq T; \boldsymbol{\theta}) \\ &= f(t_i, r_i \wedge c_i, \delta_i | \mathbf{z}_i, U \geq T; \boldsymbol{\theta}) f(\mathbf{z}_i | U \geq T; \boldsymbol{\theta}) \\ &= L_{C,i}(\boldsymbol{\theta}) f(\mathbf{z}_i | U \geq T; \boldsymbol{\theta}). \end{aligned}$$

Hence,

$$L_J(\boldsymbol{\theta}) = L_C(\boldsymbol{\theta}) \prod_{i=1}^{n} f(\mathbf{z}_i | U \geq T; \boldsymbol{\theta}). \tag{2.3.10}$$

Definition 2.3.1 *Under length-biased sampling, the density of the covariate $\boldsymbol{Z}$ conditional on $U \geq T$, denoted by $f_B(\boldsymbol{z}; \boldsymbol{\theta})$, is the biased density of the covariate.*

The biased density $f_B(\mathbf{z}; \boldsymbol{\theta})$ [9] can be expressed as

$$f_B(\mathbf{z}; \boldsymbol{\theta}) = f(\mathbf{z} | U \geq T; \boldsymbol{\theta}) = \frac{\mathbb{P}(U \geq T | \mathbf{z}; \boldsymbol{\theta}) f_{\mathbf{Z}}(\mathbf{z})}{P(U \geq T; \boldsymbol{\theta})}, \tag{2.3.11}$$

where $f_{\mathbf{Z}}(\mathbf{z})$ is the unbiased density of the covariate $\mathbf{Z}$. By using the fact that the r.v. U is independent of the truncation time T which follows a uniform distribution $g(t)$

over the interval $(0, c)$ and does not depend on the covariate, then

$$\begin{aligned}\mathbb{P}\left(U \geq T\left|\mathbf{z};\boldsymbol{\theta}\right.\right) &= \int_0^\infty \int_0^u f\left(u, t\left|\mathbf{z};\boldsymbol{\theta}\right.\right) dt du \\ &= \int_0^\infty \int_0^u f_U\left(u\left|\mathbf{z};\boldsymbol{\theta}\right.\right) g\left(t\right) dt du.\end{aligned}$$

It follows that,

$$\mathbb{P}\left(U \geq T\left|\mathbf{z};\boldsymbol{\theta}\right.\right) = \int_0^\infty \frac{u}{c} f_U\left(u\left|\mathbf{z};\boldsymbol{\theta}\right.\right) du = \frac{\mu\left(\mathbf{z};\boldsymbol{\theta}\right)}{c}. \tag{2.3.12}$$

Now, from (2.3.12) one has

$$\begin{aligned}\mathbb{P}\left(U \geq T;\boldsymbol{\theta}\right) &= \int_{\mathbf{z}} P\left(U \geq T, \mathbf{z};\boldsymbol{\theta}\right) d\mathbf{z} \\ &= \int_{\mathbf{z}} P\left(U \geq T\left|\mathbf{z}\right.;\boldsymbol{\theta}\right) f_{\mathbf{Z}}\left(\mathbf{z}\right) d\mathbf{z} \\ &= \frac{1}{c}\int_{\mathbf{z}} \mu\left(\mathbf{z};\boldsymbol{\theta}\right) f_{\mathbf{Z}}\left(\mathbf{z}\right) d\mathbf{z}.\end{aligned}$$

Therefore,

$$\mathbb{P}\left(U \geq T;\boldsymbol{\theta}\right) = \frac{\mathrm{E}\left[\mu\left(\mathbf{Z};\boldsymbol{\theta}\right)\right]}{c} = \frac{\mu\left(\boldsymbol{\theta}\right)}{c}. \tag{2.3.13}$$

Substituting (2.3.12) and (2.3.13) into (2.3.11), we obtain

$$f_B\left(\mathbf{z};\boldsymbol{\theta}\right) = \frac{\mu\left(\mathbf{z};\boldsymbol{\theta}\right) f_{\mathbf{Z}}\left(\mathbf{z}\right)}{\int_{\mathbf{z}} \mu\left(\mathbf{z};\boldsymbol{\theta}\right) f_{\mathbf{Z}}\left(\mathbf{z}\right) d\mathbf{z}} = \frac{\mu\left(\mathbf{z};\boldsymbol{\theta}\right) f_{\mathbf{Z}}\left(\mathbf{z}\right)}{\mu\left(\boldsymbol{\theta}\right)}. \tag{2.3.14}$$

Since $f_{\mathbf{Z}}\left(\mathbf{z}\right)$ is independent of $\boldsymbol{\theta}$, the joint likelihood (2.3.10) becomes

$$\begin{aligned}L_J\left(\boldsymbol{\theta}\right) &\propto L_C\left(\boldsymbol{\theta}\right) \times \prod_{i=1}^n \frac{\mu\left(\mathbf{z}_i;\boldsymbol{\theta}\right)}{\mu\left(\boldsymbol{\theta}\right)} \\ &= \prod_{i=1}^n \left(\frac{f_U\left(t_i + r_i\left|\mathbf{z}_i;\boldsymbol{\theta}\right.\right)}{\mu\left(\boldsymbol{\theta}\right)}\right)^{\delta_i} \left(\int_{w \geq t_i + c_i} \frac{f_U\left(w\left|\mathbf{z}_i;\boldsymbol{\theta}\right.\right)}{\mu\left(\boldsymbol{\theta}\right)} dw\right)^{1-\delta_i}.\end{aligned}$$

We note that any likelihood inference based on $L_I\left(\boldsymbol{\theta}\right)$ or $L_J\left(\boldsymbol{\theta}\right)$ is conditional on $\mathbf{Z} = \mathbf{z}$. In addition, the corresponding MLE's $\hat{\boldsymbol{\theta}}_{J,n}$ and $\hat{\boldsymbol{\theta}}_{C,n}$ are asymptotically similar. However, the asymptotic efficiencies of those MLE's can be quite different since, $L_J\left(\boldsymbol{\theta}\right)$ incorporates the information ignored by $L_I\left(\boldsymbol{\theta}\right)$ [9]. It can be shown by an analytic example in [9] that $\hat{\boldsymbol{\theta}}_{J,n}$ can be 50% more efficient than $\hat{\boldsymbol{\theta}}_{C,n}$.

Chapter 3

Measure of dependence for length-biased data: one continuous covariate

Our goal in this chapter is to extend the measure of dependence proposed by Kent [33] in the context of length-biased sampling without censoring for the case of one continuous covariate. In this direction, we begin by establishing a link between the conditional information gain and joint information gain. To estimate the measure of dependence between survival time U and a single covariate Z, we propose to use the method based on the concept of kernel density estimator with a regression procedure. In particular, the estimation of the length-biased density of U conditional on Z, estimation of the unbiased density of the covariate Z and estimation of the corresponding biased density will be considered in this chapter.

3.1 Conditional and joint dependence measures under length-biased sampling

In this section, we investigate the form of the joint length-biased density under both the dependence model (survival time and covariate are dependent) and under the independence model (survival time and covariate are independent). In the context of length-biased sampling, we provide the relationship between conditional information gain and joint information gain. Also, we adapt the conditional and joint measures of dependence proposed by Kent [33] in this context.

3.1.1 Joint length-biased density under the dependence and independence models

Theorem 3.1.1 *Let U be a survival time with length-biased density $f_{LB}(u)$ given in (2.3.2) and let Z be a covariate with biased continuous density $f_B(z)$ given in (2.3.14).*

(a) If U and Z are dependent random variables then the joint length-biased density takes the following form

$$f_{LB}(u,z) = f_{LB}(u|z)f_B(z) = \frac{uf_U(u,z)}{\mu}, \tag{3.1.1}$$

where $f_{LB}(u|z)$ is the length-biased density of U conditional on $Z=z$, $f_U(u,z)$ is the joint unbiased density of the random vector (U,Z) and the overall mean lifetime of the unbiased population is $\mu = \iint uf_U(u,z)dudz = \int uf_U(u)du < \infty$.

(b) If the random variables U and Z are independent then the joint length-biased density can be written as

$$f_{LB}(u,z) = f_{LB}(u)f_Z(z) = \frac{uf_U(u)}{\mu}f_Z(z), \tag{3.1.2}$$

where $f_Z(z)$ is the unbiased density of the covariate.

Proof: (a) Based on Equations (2.3.2) and (2.3.14), the joint length-biased density of (U, Z) under the dependence model can be written as

$$f_{LB}(u, z) = f_{LB}(u|z) f_B(z) = \frac{u f_U(u|z)}{\mu(z)} \frac{\mu(z) f_Z(z)}{\mu}, \tag{3.1.3}$$

where $\mu(z) = \mathrm{E}[U|Z = z] = \int u f_U(u|z) du < \infty$ and $\mu = \mathrm{E}[\mathrm{E}[U|Z = z]] = \mathrm{E}[U] = \int u f_U(u) du$. Therefore,

$$f_{LB}(u, z) = f_{LB}(u|z) f_B(z) = \frac{u f_U(u, z)}{\mu}. \tag{3.1.4}$$

(b) From the independence of U and Z, we have

$$f_{LB}(u, z) = f_{LB}(u) f_B(z) = f_{LB}(u) f_Z(z), \tag{3.1.5}$$

where in Equation (2.3.14), we used the fact $\mu(z) = \mathrm{E}[U|Z = z] = \mu$. From (2.3.2), this leads to

$$f_{LB}(u, z) = f_{LB}(u) f_Z(z) = \frac{u f_U(u)}{\mu} f_Z(z). \tag{3.1.6}$$

∎

3.1.2 Conditional information gain versus joint information gain under length-biased sampling

Let (U, Z) be a pair of random variables possibly dependent with true joint density $f_{LB}(u, z)$. Based on the concept of information gain [33] and Theorem 3.1.1, the following two propositions establish a link between the conditional information gain and joint information gain in the context of length-biased sampling.

Proposition 3.1.2 *The conditional information gain under length-biased sampling can be expressed as*

$$\Gamma_C = 2\left\{\iint \log\left\{f_{LB}\left(u|z\right)\right\} f_{LB}\left(u,z\right) dudz - \int \log\left\{f_{LB}\left(u\right)\right\} f_{LB}\left(u\right) du\right\}, \quad (3.1.7)$$

and the adapted conditional measure of dependence of Kent [33] is

$$\rho_C^2\left(U|Z\right) = 1 - \exp\left\{-\Gamma_C\right\}. \quad (3.1.8)$$

Proof: To obtain a conditional measure of dependence $\rho_C^2\left(U|Z\right)$, we consider the following models

$$\text{Independence}: \; f_{LB}\left(u|z\right) = f_{LB}\left(u\right), \text{ for all } u,$$

$$\text{Dependence}: \; f_{LB}\left(u|z\right) \neq f_{LB}\left(u\right), \text{ for some } u.$$

The conditional information under the dependence model can be expressed as

$$\Phi_{C,1} = \iint \log\left\{f_{LB}\left(u|z\right)\right\} f_{LB}\left(u,z\right) dudz,$$

and the conditional information under the independence model is

$$\Phi_{C,0} = \iint \log\left\{f_{LB}\left(u\right)\right\} f_{LB}\left(u,z\right) dudz = \int \log\left\{f_{LB}\left(u\right)\right\} f_{LB}\left(u\right) du.$$

To measure the conditional information gain we use twice the Kullback-Leibler [35] information gain as

$$\begin{aligned}\Gamma_C &= 2\left\{\Phi_{C,1} - \Phi_{C,0}\right\} \\ &= 2\left\{\iint \log\left\{f_{LB}\left(u|z\right)\right\} f_{LB}\left(u,z\right) dudz - \int \log\left\{f_{LB}\left(u\right)\right\} f_{LB}\left(u\right) du\right\}.\end{aligned}$$

Now, we can adapt the conditional measure of dependence of Kent [33] as

$$\rho_C^2\left(U|Z\right) = 1 - \exp\left\{-\Gamma_C\right\}.$$

∎

Proposition 3.1.3 *The joint information gain under length-biased sampling is*

$$\Gamma = \Gamma_C + \Gamma_B, \tag{3.1.9}$$

where Γ_C is given by (3.1.7) and Γ_B is the information gain obtained through knowledge of the bias of covariate

$$\Gamma_B = 2\left\{\int \log\{f_B(z)\} f_B(z)\, dz - \int \log\{f_Z(z)\} f_B(z) dz\right\}, \tag{3.1.10}$$

and the adapted joint measure of dependence of Kent [33] is

$$\rho_J^2(U, Z) = 1 - \exp\{-(\Gamma_C + \Gamma_B)\}. \tag{3.1.11}$$

Proof: To obtain a joint measure of dependence $\rho_J^2(U, Z)$, we consider the following models

$$\text{Independence}: \ f_{LB}(u, z) = f_{LB}(u) f_Z(z), \ \text{for all } u, z,$$

$$\text{Dependence}: \ f_{LB}(u, z) \neq f_{LB}(u) f_Z(z), \ \text{for some } u, z.$$

Under length-biased sampling, the joint information under the dependence and independence models are given, respectively, by

$$\Phi_1 = \iint \log\{f_{LB}(u, z)\} f_{LB}(u, z) du dz, \tag{3.1.12}$$

$$\Phi_0 = \iint \log\{f_{LB}(u) f_Z(z)\} f_{LB}(u, z) du dz. \tag{3.1.13}$$

Equation (3.1.12) can be expressed as

$$\begin{aligned}
\Phi_1 &= \iint \log\{f_{LB}(u|z) f_B(z)\} f_{LB}(u, z) du dz \\
&= \iint \log\{f_{LB}(u|z)\} f_{LB}(u, z) du dz + \iint \log\{f_B(z)\} f_{LB}(u, z)\, du dz \\
&= \iint \log\{f_{LB}(u|z)\} f_{LB}(u, z) du dz + \int \log\{f_B(z)\} f_B(z)\, dz, \qquad (3.1.14)
\end{aligned}$$

and Equation (3.1.13) can be written as

$$\begin{aligned}
\Phi_0 &= \iint \log\left\{f_{LB}\left(u\right)f_Z\left(z\right)\right\} f_{LB}\left(u,z\right)dudz \\
&= \iint \log\left\{f_{LB}\left(u\right)\right\} f_{LB}\left(u,z\right)dudz + \iint \log\left\{f_Z\left(z\right)\right\} f_{LB}\left(u,z\right)dudz \\
&= \int \log\left\{f_{LB}\left(u\right)\right\} f_{LB}\left(u\right)du + \int \log\left\{f_Z\left(z\right)\right\} f_B\left(z\right)dz. \qquad (3.1.15)
\end{aligned}$$

To measure the joint information gain we use twice the Kullback-Leibler [35] information gain as

$$\begin{aligned}
\Gamma &= 2\left\{\Phi_1 - \Phi_0\right\} \\
&= 2\left\{\iint \log\left\{f_{LB}\left(u|z\right)\right\} f_{LB}\left(u,z\right)dudz + \int \log\left\{f_B\left(z\right)\right\} f_B\left(z\right)dz \right.\\
&\left. - \int \log\left\{f_{LB}\left(u\right)\right\} f_{LB}\left(u\right)du - \int \log\left\{f_Z\left(z\right)\right\} f_B\left(z\right)dz\right\} \\
&= 2\left\{\iint \log\left\{f_{LB}\left(u|z\right)\right\} f_{LB}\left(u,z\right)dudz - \int \log\left\{f_{LB}\left(u\right)\right\} f_{LB}\left(u\right)du \right.\\
&\left. + \int \log\left\{f_B\left(z\right)\right\} f_B\left(z\right)dz - \int \log\left\{f_Z\left(z\right)\right\} f_B\left(z\right)dz\right\}. \qquad (3.1.16)
\end{aligned}$$

It follows that the information gain under length-biased sampling is

$$\Gamma = \Gamma_C + \Gamma_B, \qquad (3.1.17)$$

where Γ_C is the conditional information gain given by (3.1.7) and Γ_B is the information gain obtained through knowledge of the bias of the covariate

$$\Gamma_B = 2\left\{\int \log\left\{f_B\left(z\right)\right\} f_B\left(z\right)dz - \int \log\left\{f_Z\left(z\right)\right\} f_B\left(z\right)dz\right\}.$$

Here, $f_Z(z)$ denotes the unbiased density of the covariate under independence and $f_B(z)$ denotes the biased density of the covariate under dependence. Hence, the adapted joint measure of the dependence of Kent [33] is

$$\rho_J^2\left(U,Z\right) = 1 - \exp\left\{-\left(\Gamma_C + \Gamma_B\right)\right\}.$$

∎

Estimation of the conditional and joint measures of dependence given, respectively, by (3.1.8) and (3.1.11) are carried out by estimating the corresponding conditional information gain and information gain obtained through knowledge of the bias of the covariate. To estimate Γ_C in (3.1.7), we require estimators of $f_{LB}(u|z)$ and $f_{LB}(u)$. In addition, to estimate Γ_B we need to estimate $f_B(z)$ and $f_Z(z)$.

Given length-biased data, we propose to use the kernel density estimator, to find non-parametric estimators of $f_{LB}(u)$, $f_Z(z)$ and semiparametric estimators of $f_{LB}(u|z)$, $f_B(z)$. First, recall the concept of kernel density estimator and its properties. Since, $f_{LB}(u|z)$ and $f_B(z)$ are of the form of a weighted density (2.3.1), we make use of the method for unweighted and weighted densities given weighted data.

3.2 Kernel density estimator and its properties

Here, we fisrt describe the univariate density estimation based on kernel methods and then we examine some useful properties of the kernel density estimator (KDE) discussed in [49].

3.2.1 Kernel density estimator

Kernel density estimation is a non-parametric method to estimate the PDF of a random variable. Rosenblatt [43] and Parzen [40] provided the main ideas which are described in [3]. To this end, let $X_1, \ldots, X_n$ be independent and identically distributed ($i.i.d$) observations from a random variable with a cumulative distribution function $F(x)$ (CDF) and probability density function (PDF) $f(x) = dF(x)/dx$. The goal is to estimate $f(x)$ without imposing any functional form (parametric) assumptions on the PDF. First, we note that a natural estimator of the CDF $F(x)$ is the empirical cumulative distribution function (ECDF) given as

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}}. \tag{3.2.1}$$

In addition, by the strong law of large numbers, the ECDF $F_n(x)$ converges almost surely to $F(x)\,,\ \forall x \in \mathbb{R}$ as $n \to \infty$. Therefore, $F_n(x)$ is a consistent estimator of $F(x)\,,\ \forall x \in \mathbb{R}$. The question here is how can we estimate the PDF, $f(x)$? To estimate $f(x)$, we note that intuitively

$$f(x) \approx \frac{F(x+h) - F(x-h)}{2h}, \text{ for small } h > 0.$$

We replace $F(x)$ by the estimate $F_n(x)$ and we define

$$f_n^R(x) = \frac{F_n(x+h) - F_n(x-h)}{2h},$$

where the function $f_n^R(x)$ is an estimate of $f(x)$ called the Rosenblatt-Parzen [4] kernel estimator which takes this form

$$f_n^R(x) = \frac{1}{2nh} \sum_{i=1}^{n} \mathbb{1}_{\{x-h \le X_i \le x+h\}} = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right). \tag{3.2.2}$$

Here, $K(s) = \frac{1}{2}\mathbb{1}_{\{|s| \le 1\}}$ is simply the uniform density function and h is the smoothing parameter or the bandwidth of the estimator. We note that the estimator $f_n^R(x)$ is the percentage of observations around x and the bandwidth h controls the degree of smoothing applied to the data. A simple generalization of (3.2.2) is given by

$$f_n(x) = \frac{1}{n} \sum_{i=1}^{n} K_h(x - X_i), \tag{3.2.3}$$

where the function $K_h(s) = h^{-1}K(h^{-1}s)$ and $K(\cdot)$ is called a kernel function. The function $f_n(x)$ is called the standard kernel density estimator which is the average of the kernel centred over data points $X_i,\ i = 1, \ldots, n$.

3.2.2 Kernel functions

For the following sections a kernel function $K : \mathbb{R} \to \mathbb{R}$ is defined to be any smooth function satisfying the following assumptions.

Assumptions 3.2.1

(a) $K(s)$ is a probability density function.

(b) $K(-s) = K(s)$.

(c) $\int sK(s)ds = 0$.

(d) $\|K\|_2^2 = \int K^2(s)ds < \infty$.

(e) $\mu_2(K) = \int s^2 K(s) < \infty$.

The following kernel functions satisfy Assumptions 3.2.1:

- Epanechnikov : $K(s) = \frac{3}{4}(1 - s^2)\mathbb{1}_{\{|s|\leq 1\}}$.
- Gaussian: $K(s) = (2\pi)^{-1/2} e^{-s^2/2}$, $-\infty < s < \infty$.
- $K(s) = \frac{1}{2}\mathbb{1}_{\{|s|\leq 1\}}$.

3.2.3 Some properties of the kernel density estimator

A basic measure of the accuracy of estimator f_n, at an arbitrary fixed point x, called the mean squared error (MSE) of kernel estimator, is defined by

$$\mathrm{MSE}(f_n(x)) = \mathrm{E}\left[(f_n(x) - f(x))^2\right] = \mathrm{Bias}^2(f_n(x)) + \mathrm{Var}(f_n(x)), \tag{3.2.4}$$

where $\mathrm{Bias}(f_n(x))$ and $\mathrm{Var}(f_n(x))$ are the bias and the variance of f_n at a point x, respectively, defined as

$$\mathrm{Bias}(f_n(x)) = \mathrm{E}[f_n(x)] - f(x), \tag{3.2.5}$$

$$\mathrm{Var}\left(f_n\left(x\right)\right) = \mathrm{E}\left[\left(f_n\left(x\right) - \mathrm{E}\left[f_n\left(x\right)\right]\right)^2\right]. \tag{3.2.6}$$

Based on (3.2.3), the last two equations become respectively,

$$\mathrm{Bias}\left(f_n\left(x\right)\right) = \left(K_h * f\right)\left(x\right) - f(x), \tag{3.2.7}$$

$$\mathrm{Var}\left(f_n\left(x\right)\right) = n^{-1}\left\{\left(K_h^2 * f\right)\left(x\right) - \left(K_h * f\right)^2\left(x\right)\right\}, \tag{3.2.8}$$

with the convolution notation

$$(K_h * f)(x) = \int K_h(x-y)f(y)dy.$$

These may be combined to give

$$\mathrm{MSE}\left(f_n\left(x\right)\right) = n^{-1}\left\{\left(K_h^2 * f\right)\left(x\right) - \left(K_h * f\right)^2\left(x\right)\right\} + \left\{\left(K_h * f\right)\left(x\right) - f(x)\right\}^2. \tag{3.2.9}$$

A means of judging the overall error of the kernel density estimator is to use the global criterion of mean integrated squared error (MISE) which is

$$\mathrm{MISE}\left(f_n\right) = \int \mathrm{MSE}\left(f_n\left(x\right)\right)dx. \tag{3.2.10}$$

Using (3.2.9) into (3.2.10) leads to

$$\mathrm{MISE}\left(f_n\right) = n^{-1}\int\left\{\left(K_h^2 * f\right)\left(x\right) - \left(K_h * f\right)^2\left(x\right)\right\}dx + \int\left\{\left(K_h * f\right)\left(x\right) - f(x)\right\}^2dx. \tag{3.2.11}$$

One problem with the MSE and MISE is that both depend on the bandwidth h in a complicated way making it difficult to interpret the influence of the smoothing parameter on the kernel density estimator $f_n(x)$. To solve this problem, we can derive a large sample approximation for leading variance and bias terms. In this direction, we show that these approximations play an important role to obtain the MISE-optimal bandwidth and can be used to prove the consistency of the kernel density estimator. First, we make the following assumptions for the density f and for the smoothing parameter h.

Assumptions 3.2.2

(a) The density f is such that its second derivative f'' is continuous, bounded and square integrable.

(b) The smoothing parameter h is a function of n such that $\lim_{n\to\infty} h = 0$ and $\lim_{n\to\infty} nh = \infty$, which is equivalent to saying that h approaches zero, but at a slower rate than n^{-1}.

We first consider the estimation of f at $x \in \mathbb{R}$. Expanding $f(x+ht)$ in a Taylor series around x

$$f(x+ht) = f(x) + htf'(x) + \frac{1}{2}h^2t^2f''(x) + o(h^2). \tag{3.2.12}$$

Based on (3.2.7), the bias of the function $f_n(x)$ can be written as

$$\text{Bias}\,(f_n(x)) = \int K(t)f(x+ht)dt - f(x), \tag{3.2.13}$$

by letting $h^{-1}(x-y) = -t$ and using Assumption 3.2.1 (c). Hence, the bias expression becomes

$$\text{Bias}\,(f_n(x)) = \frac{1}{2}h^2\mu_2(K)f''(x) + o(h^2), \tag{3.2.14}$$

where we used (3.2.12) and Assumptions 3.2.1. We note that the bias is of order h^2 which implies that the kernel density estimator is asymptotically unbiased.
For the variance, we have from (3.2.7) and (3.2.8),

$$\begin{aligned}
\text{Var}\,(f_n(x)) &= (nh)^{-1}\int K^2(t)f(x+ht)dt - n^{-1}\left\{\text{E}\left[f_n(x)\right]\right\}^2 \\
&= (nh)^{-1}\int K^2(t)\left\{f(x)+o(1)\right\}dt - n^{-1}\left\{f(x)+o(1)\right\}^2 \\
&= (nh)^{-1}\|K\|_2^2 f(x) + o\left((nh)^{-1}\right),
\end{aligned} \tag{3.2.15}$$

where $\|K\|_2^2 = \int K^2(s)ds$. Since the variance of $f_n(x)$ is of order $(nh)^{-1}$, Assumption 3.2.2 (b) ensures that $\text{Var}\,(f_n(x))$ converges to zero as $n \to \infty$. Consequently,

$$\text{MSE}\,(f_n(x)) = \frac{1}{nh}\|K\|_2^2 f(x) + \frac{h^4}{4}\mu_2^2(K)\left(f''(x)\right)^2 + o\left((nh)^{-1}\right) + o(h^4). \tag{3.2.16}$$

Integrating this expression and using Assumption 3.2.2 (a), lead to

$$\mathrm{MISE}\,(f_n) = \mathrm{AMISE}\,(f_n) + o\left((nh)^{-1}\right) + o(h^4), \tag{3.2.17}$$

where the asymptotic MISE (AMISE) is

$$\mathrm{AMISE}\,(f_n) = \frac{1}{nh}\|K\|_2^2 + \frac{h^4}{4}\mu_2^2\,(K)\,\|f''\|_2^2. \tag{3.2.18}$$

The latter provides a useful large sample approximation to the MISE. We note that, taking h very small in the last equation, the integrated variance increases whereas the integrated squared bias decreases. This is known as the variance-bias trade-off. An optimal bandwidth for the kernel density estimator obtained by minimizing (3.2.18) over h is

$$h_{\mathrm{AMISE}} = \left(\frac{\|K\|_2^2}{\mu_2^2\,(K)\,\|f''\|_2^2 n}\right)^{1/5}. \tag{3.2.19}$$

A practical estimator of the optimal bandwidth h, based on the normal rule, was proposed by Silverman [46]

$$\hat{h}_{opt} = \frac{0.9\hat{\sigma}}{n^5}, \tag{3.2.20}$$

where $\hat{\sigma} = min\,(s, R/1.34)$. Here, s and R are the standard deviation and interquartile range of the data, respectively.

Theorem 3.2.3 *Under Assumptions 3.2.2, f_n is a consistent estimator of f.*

Proof: By the Markov's inequality, we have

$$\begin{aligned}\mathbb{P}\,(|f_n\,(x) - f\,(x)| > \varepsilon) &= \mathbb{P}\left(|f_n\,(x) - f\,(x)|^2 > \varepsilon^2\right)\\ &\leq \frac{\mathrm{E}[f_n\,(x) - f\,(x)]^2}{\varepsilon^2}\\ &= \frac{\mathrm{MSE}\,(f_n\,(x))}{\varepsilon^2}.\end{aligned}$$

As $n \to \infty$, $h \to 0$ and $nh \to 0$. It follows by (3.2.16) that $\mathrm{MSE}\,(f_n(x)) \to 0$. Consequently, $f_n \xrightarrow{P} f$ and hence f_n is a consistent estimator of f. ∎

3.3 Unbiased density estimator given length-biased data

In this section, we provide three useful methods for estimating the unbiased density given data from the length-biased density. Let $Y_1, \ldots, Y_n$ be positive $i.i.d$ observations from a length-biased density

$$f_{LB}(y) = \frac{y f_U(y)}{\mu}, \quad y > 0, \tag{3.3.1}$$

where $f_U(y)$ is the unbiased density and $\mu = \int y f_U(y)\, dy < \infty$. Let

$$\hat{f}_{LB}(y) = \frac{1}{n}\sum_{i=1}^{n} K_h(y - Y_i) \tag{3.3.2}$$

be a kernel density estimator of $f_{LB}(y)$. A first intuitive estimator of the unbiased density

$$f_U(y) = \mu \frac{f_{LB}(y)}{y}, \tag{3.3.3}$$

proposed by Bhattacharyya et al. [7] is

$$\tilde{f}_U(y) = \hat{\mu}\frac{\hat{f}_{LB}(y)}{y} = (ny)^{-1}\hat{\mu}\sum_{i=1}^{n} K_h(y - Y_i), \tag{3.3.4}$$

where $\hat{\mu}$ is an estimator of μ. Since from (3.3.3)

$$\mu \int_0^\infty \frac{f_{LB}(y)}{y} dy = \int_0^\infty f_U(y)\, dy = 1,$$

it follows that

$$\mu = \left(\mathrm{E}_{f_{LB}}\left[\frac{1}{Y}\right]\right)^{-1}.$$

Hence, an estimator of μ proposed by Cox [13] is

$$\hat{\mu} = \left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{Y_i}\right)^{-1}. \tag{3.3.5}$$

Jones [30] provided a new kernel density estimation procedure for length-biased data as follows:

- μ can be estimated by $\hat{\mu}$ given in (3.3.5),
- $f_{LB}(y)/y$ can be estimated by

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{Y_i}K_h(y-Y_i).$$

Based on (3.3.3) and the two results above, Jones [30] proposed

$$\hat{f}_U(y) = n^{-1}\hat{\mu}\sum_{i=1}^{n}Y_i^{-1}K_h(y-Y_i), \tag{3.3.6}$$

as a second estimator of $f_U(y)$. The new kernel density estimator of Jones [30] has various advantages over an alternative suggested by Bhattacharyya et al. [7] since $\hat{f}_U(y)$ is always a density itself while $\tilde{f}_U(y)$ may well not have a finite integral. In addition, $\hat{f}_U(y)$ has better asymptotic mean integrated squared error properties [30]. A third approach to estimate $f_U(y)$ can be constructed as follows. First, consider a length-biased sample $\mathcal{Y} = (Y_1, \ldots, Y_n)$ from $f_{LB}(y)$. Then, use the bootstrap techniques with replacement for the original sample $\mathcal{Y}$ to obtain a new sample $\mathcal{Y}^* = (Y_1^*, \ldots, Y_n^*)$. The idea is that, Y_i is chosen to be included in the new sample $\mathcal{Y}^*$ with probability p_i. For $j = 1, \ldots, n$, the probability p_i, $i = 1, \ldots, n$ can be found using (3.3.3) as

$$\begin{aligned} p_i &= \mathbb{P}\left(Y_j^* = Y_i | Y_1, \ldots, Y_n\right) \\ &= \hat{\mu}\frac{\mathbb{P}\left(Y_j^* = Y_i\right)}{Y_i} \\ &= \hat{\mu}\frac{1/n}{Y_i} \\ &= \left(\frac{1}{n}\sum_{i=1}^{n}Y_i^{-1}\right)^{-1}\frac{n^{-1}}{Y_i}. \end{aligned}$$

Consequently,

$$p_i = \frac{Y_i^{-1}}{\sum_{i=1}^{n} Y_i^{-1}}. \tag{3.3.7}$$

Hence, the sample $Y_1^*, \ldots, Y_n^*$ obtained previously can be used to estimate $f_U(y)$ by the standard kernel density estimator

$$\check{f}_U(y) = \frac{1}{n} \sum_{i=1}^{n} K_h(y - Y_i^*), \tag{3.3.8}$$

which has the same properties discussed in Section 3.2.3. However, some properties of $\hat{\mu}$ and $\hat{f}_U(y)$ will be given, in detail, in the next section when our interest is to estimate the unweighted density given weighted data. Length-biased distribution is a particular case of weighted distribution.

3.4 Unweighted density estimator given weighted data and some properties of the estimators

We provide, in this section, two methods in common use to estimate unweighted density given data from weighted density. These approaches can be viewed as a generalization of those exposed in the previous section. Also, we give some useful properties of the proposed estimators.

3.4.1 Unweighted density estimation given weighted data

Let $Y_1, \ldots, Y_n$ be a random sample from the weighted density given by (2.3.1)

$$g_w(y) = \frac{w(y) f_{uw}(y)}{\mu_w}, \quad w(y) > 0, \tag{3.4.1}$$

where $\mu_w = \int w(y) f_{uw}(y)\, dy < \infty$. From (3.4.1), the unweighted density can be expressed as

$$f_{uw}(y) = \mu_w \frac{g_w(y)}{w(y)}. \tag{3.4.2}$$

Given a sample described above, Jones [30] suggested a similar approach as for (3.3.6) to find an estimator for the unweighted density $f_{uw}(y)$:

- μ_w can be estimated by

$$\hat{\mu}_w = n\left(\sum_{i=1}^{n} w(Y_i)^{-1}\right)^{-1}, \tag{3.4.3}$$

  since by (3.4.2) we have $\mu_w \int_{\mathbb{R}} w\,(y)^{-1} g_w\,(y)\,dy = 1$ which implies that

$$\mu_w = \left(\mathrm{E}_{g_w}\left[w\,(Y)^{-1}\right]\right)^{-1}. \tag{3.4.4}$$

- $g_w\,(y)/w\,(y)$ can be estimated by

$$\frac{1}{n}\sum_{i=1}^{n} w(Y_i)^{-1} K_h\,(y - Y_i)\,.$$

Based on (3.4.2), an estimator of $f_{uw}\,(y)$ is

$$\hat{f}_{uw}\,(y) = n^{-1}\hat{\mu}_w \sum_{i=1}^{n} w(Y_i)^{-1} K_h\,(y - Y_i)\,. \tag{3.4.5}$$

Another estimator for $f_{uw}\,(y)$ is to use the standard kernel density estimator

$$\check{f}_{uw}\,(y) = \frac{1}{n}\sum_{i=1}^{n} K_h\,(y - Y_i^*)\,, \tag{3.4.6}$$

where $\mathcal{Y}^* = (Y_1^*, \ldots, Y_n^*)$ is a new sample obtained by using the bootstrap techniques with replacement, from the original sample $\mathcal{Y} = (Y_1, \ldots, Y_n)$ and Y_i is chosen to be included in the new sample $\mathcal{Y}^*$ with probability p_i. For $j = 1, \ldots, n$ the form of p_i, $i = 1, \ldots, n$ can be found by using (3.4.2) as follows

$$\begin{aligned} p_i &= \mathbb{P}\left(Y_j^* = Y_i | Y_1, \ldots, Y_n\right) \\ &= \hat{\mu}_w \frac{\mathbb{P}\left(Y_j^* = Y_i\right)}{w(Y_i)} \\ &= \hat{\mu}_w \frac{(1/n)}{w(Y_i)}. \end{aligned}$$

So that,

$$
\begin{aligned}
p_i &= \left(\frac{1}{n}\sum_{i=1}^{n} w(Y_i)^{-1}\right)^{-1} \frac{n^{-1}}{w(Y_i)} \\
&= \frac{w(Y_i)^{-1}}{\sum\limits_{i=1}^{n} w(Y_i)^{-1}}. \qquad (3.4.7)
\end{aligned}
$$

3.4.2 Some properties of the estimators

There are many interesting results for the estimators $\hat{\mu}_w^{-1}$ and $\hat{f}_{uw}$, in the literature especially in [30]. In this section, we give some properties of those estimators and their corresponding proofs.

Property 3.4.1 *Let $Y_1, \ldots, Y_n$ be a random sample from the weighted density $g_w(y)$. Suppose that $E_{g_w}\left[w(Y_1)^{-1}\right] < \infty$. Then,*

(a) $Bias\,(\hat{\mu}_w^{-1}) = 0$.

(b) $Var\,(\hat{\mu}_w^{-1}) = n^{-1}\mu_w^{-2}\left(E_{f_{uw}}\left[w(Y_1)^{-1}\right] E_{f_{uw}}\left[w(Y_1)\right] - 1\right)$.

(c) $E_{f_{uw}}\left[w(Y_1)^{-1}\right] E_{f_{uw}}\left[w(Y_1)\right] \geq 1$.

Proof: (a) Since, $Y_1, \ldots, Y_n$ are $i.i.d$ then by (3.4.3) and (3.4.4), we have

$$
\mathrm{E}_{g_w}\left[\hat{\mu}_w^{-1}\right] = \mathrm{E}_{g_w}\left[n^{-1}\sum_{i=1}^{n} w(Y_i)^{-1}\right] = \mathrm{E}_{g_w}\left[w(Y_1)^{-1}\right] = \mu_w^{-1}.
$$

Therefore,

$$
\mathrm{Bias}\left(\hat{\mu}_w^{-1}\right) = \mathrm{E}_{g_w}\left[\hat{\mu}_w^{-1}\right] - \mu_w^{-1} = 0.
$$

(b) Using the fact that $Y_1, \ldots, Y_n$ are $i.i.d.$ then

$$
\mathrm{Var}\left(\hat{\mu}_w^{-1}\right) = \mathrm{Var}\left(n^{-1}\sum_{i=1}^{n} \frac{1}{w(Y_i)}\right) = \frac{1}{n}\mathrm{Var}\left(\frac{1}{w(Y_1)}\right).
$$

This leads to,

$$
\begin{aligned}
\operatorname{Var}\left(\hat{\mu}_w^{-1}\right) &= \frac{1}{n}\left(\mathrm{E}_{g_w}\left[\left(\frac{1}{w\left(Y_1\right)}\right)^2\right] - \left(\mathrm{E}_{g_w}\left[\frac{1}{w\left(Y_1\right)}\right]\right)^2\right) \\
&= \frac{1}{n}\left(\int\left(\frac{1}{w\left(y_1\right)}\right)^2 g_w\left(y_1\right) dy_1 - \mu_w^{-2}\right) \\
&= \frac{1}{n}\left(\int\left(\frac{1}{w\left(y_1\right)}\right)^2 \frac{w\left(y_1\right) f_{uw}\left(y_1\right)}{\mu_w} dy_1 - \mu_w^{-2}\right) \\
&= \frac{1}{n}\left(\int \frac{1}{w\left(y_1\right)} \frac{f_{uw}\left(y_1\right)}{\mu_w} dy_1 - \mu_w^{-2}\right) \\
&= \frac{1}{n}\mu_w^{-2}\left(\mu_w \int \frac{1}{w\left(y_1\right)} f_{uw}\left(y_1\right) dy_1 - 1\right) \\
&= \frac{1}{n}\mu_w^{-2}\left(\mathrm{E}_{f_{uw}}\left[\frac{1}{w\left(Y_1\right)}\right]\mu_w - 1\right).
\end{aligned}
$$

Hence,

$$
\operatorname{Var}\left(\hat{\mu}_w^{-1}\right) = \frac{1}{n}\mu_w^{-2}\left(\mathrm{E}_{f_{uw}}\left[\frac{1}{w\left(Y_1\right)}\right]\mathrm{E}_{f_{uw}}\left[w\left(Y_1\right)\right] - 1\right).
$$

(c) For a positive r.v. X and a convex function $\varphi(x) = 1/x,\ x \in]0, \infty[$, we have by Jensen's inequality

$$
\varphi\left(\mathrm{E}[X]\right) \leq \mathrm{E}\left[\varphi(X)\right],
$$

so that

$$
\frac{1}{\mathrm{E}[X]} \leq \mathrm{E}\left[\frac{1}{X}\right]. \tag{3.4.8}
$$

Consequently, one obtains

$$
\frac{1}{\mathrm{E}_{f_{uw}}\left[w\left(Y_1\right)\right]} \leq \mathrm{E}_{f_{uw}}\left[\frac{1}{w\left(Y_1\right)}\right].
$$

∎

We note that, $\hat{\mu}_w^{-1}$ is an unbiased estimator of μ_w^{-1} . However, as we will see in the next property, $\hat{\mu}_w$ is a biased estimator of μ_w.

Property 3.4.2 *Let $Y_1, \ldots, Y_n$ be a random sample from the weighted density $g_w(y)$. Suppose that $E_{g_w}\left[w(Y_1)^{-1}\right] < \infty$. Then,*

(a) $\mu_w \leq E_{g_w}[\hat{\mu}_w]$.

(b) $Bias(\hat{\mu}_w) \simeq n^{-1}\mu_w \left(E_{f_{uw}}\left[w(Y_1)^{-1}\right] E_{f_{uw}}[w(Y_1)] - 1\right)$.

(c) $\hat{\mu}_w$ is a consistent estimator of μ_w.

Proof: (a) From (3.4.8) we have

$$\frac{1}{\mathrm{E}_{g_w}[\hat{\mu}_w]} \leq \mathrm{E}_{g_w}\left[\frac{1}{\hat{\mu}_w}\right],$$

and since $\mathrm{E}_{g_w}\left[w(Y_1)^{-1}\right] = \mu_w^{-1}$, we get

$$\frac{1}{\mathrm{E}_{g_w}[\hat{\mu}_w]} \leq \frac{1}{\mu_w}.$$

(b) Let $\varphi(x) = 1/x$ be a twice differentiable function. The Taylor expansion of $\varphi(\hat{\mu}_w^{-1})$ around μ_w^{-1} is

$$\varphi\left(\hat{\mu}_w^{-1}\right) \simeq \varphi\left(\mu_w^{-1}\right) + \varphi'\left(\mu_w^{-1}\right)\left(\hat{\mu}_w^{-1} - \mu_w^{-1}\right) + \frac{\varphi''\left(\mu_w^{-1}\right)}{2}\left(\hat{\mu}_w^{-1} - \mu_w^{-1}\right)^2,$$

so that

$$\hat{\mu}_w \simeq \mu_w - \mu_w^2\left(\hat{\mu}_w^{-1} - \mu_w^{-1}\right) + \frac{2\mu_w^3}{2}\left(\hat{\mu}_w^{-1} - \mu_w^{-1}\right)^2.$$

Taking expectation with respect to the weighted density, we get

$$\begin{aligned}
\mathrm{E}_{g_w}[\hat{\mu}_w] &\simeq \mathrm{E}_{g_w}\left[\mu_w - \mu_w^2\left(\hat{\mu}_w^{-1} - \mu_w^{-1}\right) + \mu_w^3\left(\hat{\mu}_w^{-1} - \mu_w^{-1}\right)^2\right] \\
&= \mu_w + \mu_w^3 \mathrm{Var}\left(\hat{\mu}_w^{-1}\right).
\end{aligned}$$

Using Property 3.4.1 (b), we then obtain

$$\mathrm{E}_{g_w}[\hat{\mu}_w] \simeq \mu_w + \frac{1}{n}\mu_w\left(\mathrm{E}_{f_{uw}}\left[\frac{1}{w(Y_1)}\right]\mathrm{E}_{f_{uw}}[w(Y_1)] - 1\right).$$

Hence,

$$\mathrm{Bias}(\hat{\mu}_w) \simeq \frac{1}{n}\mu_w\left(\mathrm{E}_{f_{uw}}\left[\frac{1}{w(Y_1)}\right]\mathrm{E}_{f_{uw}}[w(Y_1)] - 1\right).$$

(c) Since Y_i, $i = 1, \ldots, n$ are *i.i.d* and $\mu_w^{-1} = E_{g_w}\left[w\left(Y_i\right)^{-1}\right] < \infty$, by using the weak law of large numbers, one has

$$\hat{\mu}_w^{-1} = \frac{\sum_{i=1}^{n} w\left(Y_i\right)^{-1}}{n} \xrightarrow{P} \mu_w^{-1} \text{ as } n \to \infty, \tag{3.4.9}$$

which implies that $\hat{\mu}_w^{-1}$ is a consistent estimator of μ_w^{-1}. It follows that, as $n \to \infty$ $\hat{\mu}_w \xrightarrow{P} \mu_w$. Therefore, $\hat{\mu}_w$ is a consistent estimator of μ_w. ∎

Property 3.4.3 *Let $Y_1, \ldots, Y_n$ be i.i.d observations from the weighted density $g_w(y)$. Suppose that $E_{g_w}\left[w\left(Y_1\right)^{-1}\right] < \infty$. Then,*

(a) the kernel density estimator $\hat{f}_{uw}(y)$ of the unweighed density is a PDF.

(b) $Bias\left(\hat{f}_{uw}(y)\right) = \left(K_h * f_{uw}\right)(y) - f_{uw}(y)$.

(c) $Var\left(\hat{f}_{uw}(y)\right) = n^{-1}\left\{\left(K_h^2 * \gamma_w\right)(y) - \left(K_h * f_{uw}\right)^2(y)\right\}$ *where* $\gamma_w(y) = \mu_w w(y)^{-1} f_{uw}(y)$.

Property 3.4.4 *Under Assumptions (3.2.1)*

(a) $Bias\left(\hat{f}_{uw}(y)\right) = \frac{h^2}{2}\mu_2(K) f''(y) + o(h^2)$.

(b) $Var\left(\hat{f}_{uw}(y)\right) = \frac{1}{nh}\gamma_w(y)\, \|K\|_2^2 f(y) + o\left(\frac{1}{nh}\right)$.

(c) $MSE\left(\hat{f}_{uw}(y)\right) = \frac{1}{nh}\gamma_w(y)\, \|K\|_2^2 f(y) + \frac{h^4}{4}\mu_2^2(K)\left(f''(y)\right)^2 + o\left(\frac{1}{nh}\right) + o(h^4)$.

(d) $MISE\left(\hat{f}_{uw}\right) = \frac{1}{nh}\gamma_w(y)\, \|K\|_2^2 + \frac{h^4}{4}\mu_2^2(K)\|f''\|_2^2 + o\left(\frac{1}{nh}\right) + o(h^4)$.

Theorem 3.4.5 *Suppose that $h \to 0$ and $nh \to 0$ as $n \to \infty$. Then, $\hat{f}_{uw}$ is a consistent estimator of f_{uw}.*

The proofs of Properties 3.4.3, 3.4.4 and Theorem 3.4.5 can be obtained in the same way as for the standard kernel density estimator in Section 3.2.3.

3.5 Kernel density estimation procedure under the independence and dependence models

Here, we develop the kernel density estimation with a regression procedure to find, under the independence model, nonparametric estimators of $f_{LB}(u)$, $f_Z(z)$ and, under the dependence model, semiparametric estimators of $f_{LB}(u\,|z)$, $f_B(z)$.

3.5.1 Estimation procedure for the length-biased density conditional on a fixed covariate

Let $U_1, \ldots, U_n$ be $i.i.d$ positive observations, of a survival time, from a length-biased density $f_{LB}(u)$ and let $Z_1, \ldots, Z_n$ denote a random sample from a biased density $f_B(z)$. A kernel density estimator of $f_{LB}(u)$ can be obtained from (3.2.3) as follows

$$\hat{f}_{LB}(u) = \frac{1}{n}\sum_{i=1}^{n} K_h(u - U_i). \tag{3.5.1}$$

The length-biased density of U conditional on $Z = z$ is

$$f_{LB}(u|z) = \frac{u f_U(u|z)}{\mu(z)}, \tag{3.5.2}$$

where

- $f_U(u|z)$ is the unbiased density of U conditional on $Z = z$.
- $\mu(z) = \int\ u f_U(u|z)\, du < \infty$.

The question here is how can we adapt the estimator of Jones [30] to estimate the density of U conditional on a fixed covariate $Z = z$ denoted by $f_U(u\,|z)$.
We can do this by way of a linear regression model

$$\phi(U) = Y = \alpha + \beta Z + \varepsilon, \tag{3.5.3}$$

where ϕ is a monotone increasing transformation, α is an intercept, β is a coefficient of regression and ε is a random variable (error variate) independent of Z. The next step is to obtain, by the following algorithm, the pseudo-observations from $f_{LB}(u|z)$.

Algorithm 3.5.1

1. Define the linear model
$$Y_i = \alpha + \beta Z_i + \varepsilon_i,\ i = 1, \ldots, n.$$
2. Estimate α and β by the Least squares method, say $\hat{\alpha}$ and $\hat{\beta}$.
3. Estimate the errors $\varepsilon_i,\ i = 1, \ldots, n$ by
$$\hat{\varepsilon}_i = Y_i - \hat{\alpha} - \hat{\beta} Z_i,\ i = 1, \ldots, n.$$
4. Based on the sample $\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n$, use a goodness-of-fit to identify a parametric model for f_ε.
5. Generate a random sample $\tilde{\varepsilon}_i,\ i = 1, \ldots, n$ from f_ε.
6. For a fixed value $Z = z$, compute
$$\tilde{Y}_i = \hat{\alpha} + \hat{\beta} z + \tilde{\varepsilon}_i,\ i = 1, \ldots, n.$$
7. The pseudo-observations from $f_{LB}(u|z)$ can be obtained as follows
$$\tilde{U}_i = \phi^{-1}\left(\tilde{Y}_i\right) = \phi^{-1}\left(\hat{\alpha} + \hat{\beta} z + \tilde{\varepsilon}_i\right),\ i = 1, \ldots, n. \tag{3.5.4}$$

So, the adapted estimator $\hat{f}_U(u\,|z)$ of Jones [30], given in (3.3.6), would be
$$\hat{f}_U(u\,|z) = n^{-1}\hat{\mu}(z)\sum_{i=1}^{n} \tilde{U}_i^{-1} K_h\left(u - \tilde{U}_i\right), \tag{3.5.5}$$

where

$$\hat{\mu}(z) = n\left(\sum_{i=1}^{n} \tilde{U}_i^{-1}\right)^{-1} = n\left(\sum_{i=1}^{n} \frac{1}{\phi^{-1}\left(\hat{\alpha} + \hat{\beta}z + \tilde{\varepsilon}_i\right)}\right)^{-1}. \qquad (3.5.6)$$

Hence,

$$\hat{f}_U(u\,|z) = n^{-1}\hat{\mu}(z)\sum_{i=1}^{n} \frac{1}{\phi^{-1}\left(\hat{\alpha} + \hat{\beta}z + \tilde{\varepsilon}_i\right)} K_h\left(u - \phi^{-1}\left(\hat{\alpha} + \hat{\beta}z + \tilde{\varepsilon}_i\right)\right). \qquad (3.5.7)$$

Lemma 3.5.2 *If the kernel function K satisfies Assumptions 3.2.1 (a) and (c) then*

$$\int u\hat{f}_U(u\,|z)du = \hat{\mu}(z), \qquad (3.5.8)$$

where $\hat{\mu}(z)$, given by (3.5.6), is the estimator of $\mu(z)$.

Proof: From (3.5.5), one has

$$\begin{aligned}
\int u\hat{f}_U(u\,|z)du &= \int u\frac{1}{n}\hat{\mu}(z)\sum_{i=1}^{n}\tilde{U}_i^{-1}K_h\left(u-\tilde{U}_i\right)du \\
&= \frac{1}{n}\hat{\mu}(z)\sum_{i=1}^{n}\int \frac{u}{\tilde{U}_i}K_h\left(u-\tilde{U}_i\right)du \\
&= \frac{1}{n}\hat{\mu}(z)\sum_{i=1}^{n}\int \frac{w+\tilde{U}_i}{\tilde{U}_i}K_h(w)dw \\
&= \frac{1}{n}\hat{\mu}(z)\sum_{i=1}^{n}\left(\frac{1}{\tilde{U}_i}\int wK_h(w)dw + \int K_h(w)dw\right).
\end{aligned}$$

Therefore,

$$\int u\hat{f}_U(u\,|z)du = \hat{\mu}(z),$$

where we used $u - \tilde{U}_i = w$ and Assumptions 3.2.1 (a) and (c). ∎

Now based on (3.5.2), we propose to use

$$\hat{f}_{LB}(u\,|z) = \frac{u\hat{f}_U(u\,|z)}{\int u\hat{f}_U(u\,|z)du}, \qquad (3.5.9)$$

as a density estimator of $f_{LB}(u\,|z)$, where $\hat{f}_U(u\,|z)$ is given by Equation (3.5.7). Using lemma 3.5.2, this leads to

$$\hat{f}_{LB}(u\,|z) = \frac{u\hat{f}_U(u\,|z)}{\hat{\mu}(z)}. \tag{3.5.10}$$

Substituting (3.5.7) into (3.5.10), one gets

$$\hat{f}_{LB}(u\,|z) = n^{-1}\sum_{i=1}^{n}\frac{u}{\phi^{-1}\left(\hat{\alpha}+\hat{\beta}z+\tilde{\varepsilon}_i\right)}K_h\left(u-\phi^{-1}\left(\hat{\alpha}+\hat{\beta}z+\tilde{\varepsilon}_i\right)\right). \tag{3.5.11}$$

In the case where $\phi(U) = \log\{U\}$, the linear regression model (3.5.3) is just an Accelerated Failure Time model. It follows that the theoretical density of the error, f_ε, can be identified once the distribution of $\log\{U\}$ is known. Hence, in Algorithm 3.5.1 we can replace steps 3, 4, 5 by the following step:

- Generate a random sample $\tilde{\varepsilon}_i,\ i = 1,\ldots,n$ directly from f_ε.

3.5.2 Density estimation of the covariate under the independence and dependence models

Given a length-biased random sample $(U_1, Z_1),\ldots,(U_n, Z_n)$ from $f_{LB}(u, z)$, our goal is to provide a density estimator of the covariate Z, under the independence model (U and Z are independent) and under the dependence model (U and Z are dependent). Recall that the biased density of the covariate under the dependence model is

$$f_B(z) = \frac{\mu(z)\,f_Z(z)}{\mu}, \tag{3.5.12}$$

where $\mu = \int \mu(z)\,f_Z(z)dz < \infty$ and $f_Z(z)$ is the unbiased density of the covariate Z. Under the independence model, we have by Theorem 3.1.1 $f_{LB}(u, z) = f_{LB}(u)f_Z(z)$ and Equation (3.5.12) becomes

$$f_B(z) = f_Z(z)\,, \tag{3.5.13}$$

since $\mu(z) = \mathrm{E}[U|Z=z] = \mathrm{E}[U] = \mu$. It follows that, the estimator of the unbiased density must take into account the fact that U and Z are independent random variables. However, the estimator of the biased density should contain some estimator of $\mu(z)$ because the weight function $\mu(z)$ involved in (3.5.12) contains some dependence between U and Z. In this context, we propose to use a linear regression model, described in the above section

$$\phi(U) = Y = \alpha + \beta Z + \varepsilon.$$

Let $S_0(u)$ denote the survival function of $U = \phi^{-1}(Y)$ when Z is zero. It follows that $S_0(u)$ is the survival function of $U = \phi^{-1}(\alpha + \varepsilon)$ and by (2.1.4), the expectation of U when Z equals zero can be expressed as

$$\mu(0) = \mathrm{E}[U|Z=0] = \int_0^\infty S_0(u)du. \tag{3.5.14}$$

The survival function of U given $Z = z$ is

$$\begin{aligned} S(u\,|z) &= \mathbb{P}(U \geq u\,|z) \\ &= \mathbb{P}(\phi(U) \geq \phi(u)\,|z) \\ &= \mathbb{P}(\alpha + \beta z + \varepsilon \geq \phi(u)) \\ &= \mathbb{P}(\alpha + \varepsilon \geq \phi(u) - \beta z) \\ &= \mathbb{P}\left(\phi^{-1}(\alpha + \varepsilon) \geq \phi^{-1}(\phi(u) - \beta z)\right) \\ &= \mathbb{P}\left(U \geq \phi^{-1}(\phi(u) - \beta z)\right). \end{aligned}$$

Hence,

$$S(u\,|z) = S_0\left(\phi^{-1}(\phi(u) - \beta z)\right). \tag{3.5.15}$$

Based on (2.1.4), the expectation of U conditional on $Z = z$ is

$$\mu(z) = \mathrm{E}[U\,|Z=z] = \int_0^\infty S(u\,|z)du = \int_0^\infty S_0\left(\phi^{-1}(\phi(u) - \beta z)\right)du.$$

We can obtain a closed form of $\mu(z)$ by using an AFT model thus, when $\phi(\cdot) = \log\{\cdot\}$. In this case

$$\mu(z) = \exp\{\beta z\} \int_0^\infty S_0(v)\, dv, \tag{3.5.16}$$

by letting $v = u \exp\{-\beta z\}$. Using (3.5.14), this leads to

$$\mu(z) = \exp\{\beta z\}\, \mu(0). \tag{3.5.17}$$

Now from (3.5.17), the biased density of covariate given in (3.5.12) becomes

$$f_B(z) = \frac{\mu(z)\, f_Z(z)}{\int_{\mathbb{R}} \mu(z)\, f_Z(z)\, dz} = \frac{\exp\{\beta z\}\mu(0)\, f_Z(z)}{\int_{\mathbb{R}} \exp\{\beta z\}\mu(0)\, f_Z(z)\, dz}. \tag{3.5.18}$$

It follows that,

$$f_B(z) = \frac{\exp\{\beta z\} f_Z(z)}{\nu_\beta}, \tag{3.5.19}$$

where $\nu_\beta = \int_{\mathbb{R}} \exp\{\beta z\} f_Z(z)\, dz < \infty$.

We note that, even if Equation (3.5.19) is of the form of (3.4.1), we cannot use

$$\hat{f}_{uw}(z) = n^{-1} \hat{\mu}_w \sum_{i=1}^{n} w(Z_i)^{-1} K_h(z - Z_i), \tag{3.5.20}$$

as an estimator of $f_Z(z)$ because the weight function $w(z) = \exp\{\beta z\}$ contains a regression coefficient β which is considered as a parameter of dependence between U and Z. In this case, based on Equation (3.2.3), we can estimate $f_Z(z)$ as follows

$$\hat{f}_Z(z) = \frac{1}{n} \sum_{i=1}^{n} K_h(z - Z_i^*), \tag{3.5.21}$$

where $\mathcal{Z}^* = (Z_1^*, \ldots, Z_n^*)$ is a new sample obtained by using the bootstrap techniques with replacement, from the original sample $\mathcal{Z} = (Z_1, \ldots, Z_n)$ choosing Z_i with probability p_i. In such a case, we have from (3.4.7)

$$p_i = \frac{w(Z_i)^{-1}}{\sum_{i=1}^{n} w(Z_i)^{-1}} = \frac{\exp\{-\beta Z_i\}}{\sum_{i=1}^{n} \exp\{-\beta Z_i\}}. \tag{3.5.22}$$

Based on Equation (3.5.19), an estimator of $f_B(z)$ is

$$\hat{f}_B(z) = \frac{\exp\left\{\hat{\beta}z\right\}\hat{f}_Z(z)}{\hat{\nu}_{\hat{\beta}}}, \tag{3.5.23}$$

where $\hat{\beta}$ is the estimator of β obtained in Algorithm 3.5.1, $\hat{f}_Z(z)$ is the estimator of the unbiased density $f_Z(z)$ and $\hat{\nu}_{\hat{\beta}} = \int_{\mathbb{R}} \exp\left\{\hat{\beta}z\right\}\hat{f}_Z(z)\,dz < \infty$. An estimator of the probability p_i, $i = 1, \dots, n$ given by (3.5.22) can be obtained as follows

$$\hat{p}_i = \frac{\exp\left\{-\hat{\beta}Z_i\right\}}{\sum_{i=1}^{n}\exp\left\{-\hat{\beta}Z_i\right\}}. \tag{3.5.24}$$

A closed form of $\hat{\nu}_{\hat{\beta}}$ can be obtained from the moment generating function (MGF) defined below.

Definition 3.5.3 *Let X be a random variable with density $f(x)$, $x \in \mathbb{R}$. The moment generating function of X is defined as*

$$\mathcal{M}_X(t) = E[\exp\{tX\}] = \int_{\mathbb{R}} \exp\{tx\} f(x)dx, \tag{3.5.25}$$

for all t for which the expectation exists. In particular, if $X \sim \mathcal{N}(\mu, \sigma^2)$ then

$$\mathcal{M}_X(t) = \exp\left\{\mu t + \frac{1}{2}\sigma^2 t^2\right\}. \tag{3.5.26}$$

Now, $\hat{\nu}_{\hat{\beta}}$ can be expressed as

$$\begin{aligned}
\hat{\nu}_{\hat{\beta}} &= \int_{\mathbb{R}} \exp\left\{\hat{\beta}z\right\}\hat{f}_Z(z)\,dz \\
&= \int_{\mathbb{R}} \exp\left\{\hat{\beta}z\right\}\frac{1}{n}\sum_{i=1}^{n} K_h(z - Z_i^*)dz \\
&= \frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{R}} \exp\left\{\hat{\beta}z\right\} K_h(z - Z_i^*)dz \\
&= \frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{R}} \exp\left\{\hat{\beta}z\right\} \frac{1}{h}K\left(\frac{z - Z_i^*}{h}\right)dz.
\end{aligned}$$

Letting $z = hs + Z_i^*$, we get

$$\begin{aligned}\hat{\nu}_{\hat{\beta}} &= \frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{R}} \exp\left\{\hat{\beta}\left(hs + Z_i^*\right)\right\} K\left(s\right)ds \\ &= \frac{1}{n}\sum_{i=1}^{n}\exp\left\{\hat{\beta}Z_i^*\right\}\left(\int_{\mathbb{R}}\exp\left\{\left(\hat{\beta}h\right)s\right\}K\left(s\right)ds\right).\end{aligned}$$

Following Definition 3.5.3, this leads to

$$\hat{\nu}_{\hat{\beta}} = \left(\frac{1}{n}\sum_{i=1}^{n}\exp\left\{\hat{\beta}Z_i^*\right\}\right)\mathcal{M}_S\left(\hat{\beta}h\right), \tag{3.5.27}$$

where S is a r.v. with kernel function $K(s)$. Hence, using (3.5.21) and (3.5.27) into (3.5.23), an estimator of $f_B\left(z\right)$ becomes

$$\hat{f}_B\left(z\right) = \frac{\exp\left\{\hat{\beta}z\right\}\sum_{i=1}^{n}K_h\left(z - Z_i^*\right)}{\mathcal{M}_S\left(\hat{\beta}h\right)\sum_{i=1}^{n}\exp\left\{\hat{\beta}Z_i^*\right\}}. \tag{3.5.28}$$

If the kernel function K is a standard normal density then by Definition 3.5.3, we have

$$\mathcal{M}_S\left(\hat{\beta}h\right) = \exp\left\{\frac{1}{2}\hat{\beta}^2h^2\right\}. \tag{3.5.29}$$

3.6 Estimation of the conditional and joint dependence measures for length-biased data

Our objective in this section is to estimate the conditional and joint measure of dependence given length-biased data $(U_1, Z_1), \ldots, (U_n, Z_n)$ from the joint length-biased density $f_{LB}\left(u, z\right)$. First, we use the fact that Γ_C given in (3.1.7) and Γ_B given by (3.1.10) can be written, respectively, as

$$\Gamma_C = 2\left\{\mathrm{E}\left[\log\left\{f_{LB}\left(U|Z\right)\right\}\right] - \mathrm{E}\left[\log\left\{f_{LB}\left(U\right)\right\}\right]\right\}, \tag{3.6.1}$$

$$\Gamma_B = 2\left\{\mathrm{E}\left[\log\left\{f_B\left(Z\right)\right\}\right] - \mathrm{E}\left[\log\left\{f_Z\left(Z\right)\right\}\right]\right\}. \tag{3.6.2}$$

From Equation (3.6.1), Γ_C can be estimated by

$$\hat{\Gamma}_C = 2\left\{\frac{1}{n}\sum_{j=1}^{n}\log\left\{\hat{f}_{LB}\left(U_j|Z_j\right)\right\} - \frac{1}{n}\sum_{j=1}^{n}\log\left\{\hat{f}_{LB}\left(U_j\right)\right\}\right\}, \tag{3.6.3}$$

where for $j = 1, \ldots, n$, $\hat{f}_{LB}\left(U_j|Z_j\right)$ and $\hat{f}_{LB}\left(U_j\right)$ can be computed, respectively, by using (3.5.11) and (3.5.1). Similarly, Γ_B given in (3.6.2) can be estimated as follows

$$\hat{\Gamma}_B = 2\left\{\frac{1}{n}\sum_{j=1}^{n}\log\left\{\hat{f}_B\left(Z_j\right)\right\} - \frac{1}{n}\sum_{j=1}^{n}\log\left\{\hat{f}_Z\left(Z_j\right)\right\}\right\}, \tag{3.6.4}$$

where for $j = 1, \ldots, n$, $\hat{f}_B\left(Z_j\right)$ and $\hat{f}_Z\left(Z_j\right)$ can be computed, respectively, from (3.5.23) and (3.5.21).

Based on (3.1.8) and (3.6.3) an estimator of the conditional dependence measure is

$$\hat{\rho}_C^2\left(U|Z\right) = 1 - \exp\left\{-\hat{\Gamma}_C\right\}. \tag{3.6.5}$$

Also, based on (3.1.11), (3.6.3) and (3.6.4) an estimator of the joint dependence measure is

$$\hat{\rho}_J^2\left(U, Z\right) = 1 - \exp\left\{-\hat{\Gamma}\right\}, \tag{3.6.6}$$

where $\hat{\Gamma}$ denotes estimator of the joint information gain given by the following equation

$$\hat{\Gamma} = \hat{\Gamma}_C + \hat{\Gamma}_B. \tag{3.6.7}$$

Chapter 4

Measure of dependence for length-biased data: several continuous covariates

In the previous chapter, we provided under length-biased sampling a relationship between the conditional information gain and joint information gain. In this sense, we developed the kernel density estimation with a regression procedure to estimate the conditional and joint dependence measures between survival time and one continuous covariate, without censoring. However, often in some practical situation, especially in survival analysis, we are interested in the measure of the dependence between a survival time and p-covariates conditional on q-covariates, named partial measure of dependence. Our goal in this chapter is to obtain this measure given length-biased data without censoring for the case of several continuous covariates. First, we establish link between the partial information gain, conditional information gain and joint information gain. To estimate the partial measure of dependence, we generalize the first method discussed in Chapter 3. In particular, the consistency of all estimators that we propose, in this chapter, will be considered.

4.1 Multivariate kernel density estimator and its properties

The multivariate kernel density estimator that we study in this section is a direct extension of the univariate estimator discussed in Chapter 3. However, this extension requires the specification of many more bandwidth parameters than in the univariate setting and some simplifying structure of the multivariate function.

4.1.1 Multivariate kernel density estimator

Consider a d-dimensional random vector $\mathbf{Y} = (Y_1, \ldots, Y_d)'$, where $Y_1, \ldots, Y_d$ are one dimensional random variables. Suppose that we collect the ith observation of each of the d-dimensional random variables in the vector $\mathbf{Y}_i$, such that

$$\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{id})',\ i = 1, \ldots, n,$$

where Y_{ij} is the ith observation of the random variable Y_j.

Let $f(\mathbf{y}) = f(y_1, \ldots, y_d)$ be a PDF of the random vector $\mathbf{Y}$. The adapted kernel density estimator of $f(\mathbf{y})$ [49] is

$$\hat{f}_h(\mathbf{y}) = n^{-1}h^{-d}\sum_{i=1}^{n}\mathcal{K}\left(\frac{\mathbf{y}-\mathbf{Y}_i}{h}\right) = n^{-1}h^{-d}\sum_{i=1}^{n}\mathcal{K}\left(\frac{y_1-Y_{i1}}{h}, \ldots, \frac{y_d-Y_{id}}{h}\right), \quad (4.1.1)$$

where $\mathcal{K}$ is a multivariate kernel function and h is the same bandwidth for each component. If we assume that there exists a vector of bandwidth $\mathbf{h} = (h_1, \ldots, h_d)'$ then the multivariate kernel estimator becomes:

$$\hat{f}_{\mathbf{h}}(\mathbf{y}) = n^{-1}\left(\prod_{\ell=1}^{d}h_\ell\right)^{-1}\sum_{i=1}^{n}\mathcal{K}\left(\frac{y_1-Y_{i1}}{h_1}, \ldots, \frac{y_d-Y_{id}}{h_d}\right). \quad (4.1.2)$$

One way to get a form of the multidimensional function $\mathcal{K}(\mathbf{s}) = \mathcal{K}(s_1, \ldots, s_d)$ is to use a multiplicative kernel [49]

$$\mathcal{K}(\mathbf{s}) = \mathcal{K}(s_1, \ldots, s_d) = K(s_1)\cdots K(s_d), \quad (4.1.3)$$

where K denotes a univariate kernel function satisfying Assumptions 3.2.1. In this case, (4.1.2) becomes

$$\hat{f}_{\mathbf{h}}(\mathbf{y}) = \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{d}\frac{1}{h_j}K\left(\frac{y_j - Y_{ij}}{h_j}\right). \tag{4.1.4}$$

The general form for the multivariate kernel density estimator with bandwidth non-singular matrix $\mathbf{H}$ [15] is

$$\hat{f}_{\mathbf{H}}(\mathbf{y}) = \frac{1}{n}\sum_{i=1}^{n}\mathcal{K}_{\mathbf{H}}(\mathbf{y} - \mathbf{Y}_i), \tag{4.1.5}$$

where

- $\mathcal{K}_{\mathbf{H}}(\mathbf{u}) = \frac{1}{|\mathbf{H}|}\mathcal{K}\left(\mathbf{H}^{-1}\mathbf{u}\right)$ and $\mathbf{H}^{-1}$ is the inverse of the matrix $\mathbf{H}$.
- $|\mathbf{H}| = \det(\mathbf{H})$ (determinant of the matrix $\mathbf{H}$).

The bandwidth matrix $\mathbf{H}$ includes all simpler cases:

- **Equal bandwidth** h: $\mathbf{H} = h\mathbf{I}_d$, where $\mathbf{I}_d$ is the $d \times d$ identity matrix.
- **Different bandwidths** $h_1, \ldots, h_d$: $\mathbf{H} = \operatorname{diag}(h_1, \ldots, h_d)$, diagonal matrix with elements $h_1, \ldots, h_d$.

4.1.2 Multivariate kernel functions

For the following sections, a multivariate kernel function $\mathcal{K} : \mathbb{R}^d \to \mathbb{R}$ is defined to be any smooth multivariate function satisfying the following assumptions.

Assumptions 4.1.1

(a) $\mathcal{K}$ is a multivariate density function.

(b) $\mathcal{K}$ is symmetric.

(c) $\int_{\mathbb{R}^d}\mathbf{s}\mathcal{K}(\mathbf{s})\,d\mathbf{s} = \mathbf{0}_d$.

(d) $\mathcal{K}$ has a second moment (matrix): $\int_{\mathbb{R}^d} \mathbf{s}\mathbf{s}'\mathcal{K}(\mathbf{s})\, d\mathbf{s} = \mu_2(\mathcal{K})\,\mathbf{I}_d$.

(e) $\mathcal{K}$ has a kernel norm: $||\mathcal{K}||_2^2 = \int_{\mathbb{R}^d} \mathcal{K}^2(\mathbf{s})\, d\mathbf{s}$.

The following multivariate kernel functions satisfy Assumptions 4.1.1:

- Epanechnikov : $\mathcal{K}(\mathbf{s}\ ) = \left(\frac{3}{4}\right)^d \prod_{i=1}^{d} \left(1 - s_i^2\right) \mathbb{1}_{\{|s_i|<1\}}$.
- Standard d-variate normal density: $\mathcal{K}(\mathbf{s}) = (2\pi)^{-d/2} \exp\left\{-\frac{1}{2}\mathbf{s}'\mathbf{s}\right\}$.

4.1.3 Some properties of the multivariate kernel density estimator

As for the standard univariate kernel density estimator, the standard multivariate kernel density estimator has very useful properties [49]:

1. $\hat{f}_{\mathbf{H}}(\mathbf{y})$ is a multivariate density function.

2. **Asymptotic Bias** (ABias):

$$A\text{Bias}\left(\hat{f}_{\mathbf{H}}(\mathbf{y})\right) = \frac{1}{2}\mu_2(\mathcal{K})\,\text{tr}\left\{\mathbf{H}'\mathcal{H}_f(\mathbf{y})\mathbf{H}\right\},$$

   where $\mathcal{H}_f(\mathbf{y}) = \left(\partial^2 f/\partial x_i x_j\right)_{i,j=1,\ldots,d}$ is the Hessian matrix and $\text{tr}\{B\}$ is the trace of the square matrix B.

3. **Asymptotic Variance** (AVar):

$$A\text{Var}\left(\hat{f}_{\mathbf{H}}(\mathbf{y})\right) = \lim_{n\to\infty} \text{Var}\left(\hat{f}_{uw}(\mathbf{y})\right) = \frac{1}{n|\mathbf{H}|}\|\mathcal{K}\|_2^2 f(\mathbf{y}).$$

4. **Asymptotic Mean Integrated Squared Error** (AMISE):

$$\text{AMISE}(\mathbf{H}) = \frac{1}{4}\mu_2^2(\mathcal{K}) \int_{\mathbb{R}^d} \text{tr}^2\left\{\mathbf{H}'\mathcal{H}_f(\mathbf{y})\mathbf{H}\right\} d\mathbf{y} + \frac{1}{n|\mathbf{H}|}\|\mathcal{K}\|_2^2.$$

5. At any point $\mathbf{y}$, $\hat{f}_{\mathbf{H}}(\mathbf{y})$ is a consistent estimator of $f(\mathbf{y})$:

$$\hat{f}_{\mathbf{H}}(\mathbf{y}) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{K}_{\mathbf{H}}(\mathbf{y} - \mathbf{Y}_i) \xrightarrow{P} f(\mathbf{y}) \text{ as } n \to \infty.$$

4.2 Multivariate unweighted density estimator given multivariate weighted data and its properties

In Chapter 3, we have shown that given a sample from a weighted density the unweighted density can be estimated using two different methods in common use. Those approaches play an important role in the context of length-biased sampling. Our main objective in the current section is to make an extension of those methods, especially when our data come from a multivariate weighted density. Moreover, we give a generalization of some properties, proposed by Jones [30], with proofs for the multivariate unweighted density estimator.

4.2.1 Estimation of the multivariate unweighted density given multivariate weighted data

Let $\mathbf{Y} = (Y_1, \ldots, Y_d)'$ be a random vector with multivariate weighted density $g_w(\mathbf{y})$ which is defined as

$$g_w(\mathbf{y}) = \frac{w(\mathbf{y}) f_{uw}(\mathbf{y})}{\mu_w}, \tag{4.2.1}$$

where

- $w(\mathbf{y})$ is a multivariate weight function such that $w(\mathbf{y}) > 0$.
- $f_{uw}(\mathbf{y})$ is a multivariate unweighted density.
- $\mu_w = \int_{\mathbb{R}^d} w(\mathbf{y}) f_{uw}(\mathbf{y}) \, d\mathbf{y} < \infty$.

The goal is to estimate the multivariate unweighted density $f_{uw}(\mathbf{y})$ given a d-dimensional data set $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{id})'$, $i = 1, \ldots, n$ from a multivariate weighted density $g_w(\mathbf{y})$. By using (4.2.1), we can write $f_{uw}(\mathbf{y})$ as

$$f_{uw}(\mathbf{y}) = \mu_w \frac{g_w(\mathbf{y})}{w(\mathbf{y})}. \tag{4.2.2}$$

As in the univariate case,

- μ_w can be estimated by

$$\hat{\mu}_w = n\left(\sum_{i=1}^{n} w(\mathbf{Y}_i)^{-1}\right)^{-1}, \tag{4.2.3}$$

since by (4.2.2) we have $\mu_w \int_{\mathbb{R}^d} w\left(\mathbf{y}\right)^{-1} g_w\left(\mathbf{y}\right) d\mathbf{y} = 1$ which implies that

$$\mu_w = \left(\mathrm{E}_{g_w}\left[\frac{1}{w\left(\mathbf{Y}\right)}\right]\right)^{-1}. \tag{4.2.4}$$

- $g_w\left(\mathbf{y}\right)/w\left(\mathbf{y}\right)$ can be estimated by

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{w(\mathbf{Y}_i)}\mathcal{K}_{\mathbf{H}}\left(\mathbf{y}-\mathbf{Y}_i\right). \tag{4.2.5}$$

Based on (4.2.2), an estimator of $f_{uw}\left(\mathbf{y}\right)$ becomes

$$\hat{f}_{uw}\left(\mathbf{y}\right) = n^{-1}\hat{\mu}_w \sum_{i=1}^{n} w(\mathbf{Y}_i)^{-1}\mathcal{K}_{\mathbf{H}}\left(\mathbf{y}-\mathbf{Y}_i\right). \tag{4.2.6}$$

which is proposed by Ibrahim [28] and can be viewed as the multivariate extension of the work of Jones [30].

Another way to estimate $f_{uw}\left(\mathbf{y}\right)$ is to use the standard multivariate kernel density estimator given by (4.1.5)

$$\check{f}_{uw}\left(\mathbf{y}\right) = \frac{1}{n}\sum_{i=1}^{n}\mathcal{K}_{\mathbf{H}}\left(\mathbf{y}-\mathbf{Y}_i^*\right), \tag{4.2.7}$$

where $\mathcal{Y}^* = (\mathbf{Y}_1^*, \ldots, \mathbf{Y}_n^*)$ is a new sample obtained, using the bootstrap techniques with replacement, from the original sample $\mathcal{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_n)$ and $\mathbf{Y}_i$ is chosen to be included in the new sample $\mathcal{Y}^*$ with probability p_i. For $j = 1, \ldots, n$ the probability p_i, $i = 1, \ldots, n$ can be found, by using (4.2.2), as follows

$$\begin{aligned}
p_i &= \mathbb{P}\left(\mathbf{Y}_j^* = \mathbf{Y}_i | \mathbf{Y}_1, \ldots, \mathbf{Y}_n\right) \\
&= \hat{\mu}_w \frac{\mathbb{P}\left(\mathbf{Y}_j^* = \mathbf{Y}_i\right)}{w(\mathbf{Y}_i)} \\
&= \hat{\mu}_w \frac{1/n}{w(\mathbf{Y}_i)} \\
&= \left(\frac{1}{n}\sum_{i=1}^{n} w(\mathbf{Y}_i)^{-1}\right)^{-1} \frac{n^{-1}}{w(\mathbf{Y}_i)} \\
&= \frac{w(\mathbf{Y}_i)^{-1}}{\sum_{i=1}^{n} w(\mathbf{Y}_i)^{-1}}. \qquad (4.2.8)
\end{aligned}$$

4.2.2 Some properties of the multivariate unweighted density estimator

Before starting to examine the properties of unweighted density estimator, we note that all properties discussed in Chapter 3 for $\hat{\mu}_w = n\left(\sum\limits_{i=1}^{n} w(Y_i)^{-1}\right)^{-1}$ still hold for $\hat{\mu}_w = n\left(\sum\limits_{i=1}^{n} w(\mathbf{Y}_i)^{-1}\right)^{-1}$ by using the fact that, for $i = 1, \ldots, n$, Y_i can be simply replaced by $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{id})'$.

Property 4.2.1 *Let $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$ be a random sample from the multivariate weighted density $g_w(\boldsymbol{y})$ and suppose that μ_w is known.*

(a) The bias of the estimator $\hat{f}_{uw}(\boldsymbol{y})$ is

$$Bias\left(\hat{f}_{uw}(\boldsymbol{y})\right) = (\mathcal{K}_{\boldsymbol{H}} * f_{uw})(\boldsymbol{y}) - f_{uw}(\boldsymbol{y}).$$

(b) The variance of $\hat{f}_{uw}(\boldsymbol{y})$ can be expressed as

$$Var\left(\hat{f}_{uw}(\boldsymbol{y})\right) = n^{-1}\left\{\left(\mathcal{K}_{\boldsymbol{H}}^2 * \gamma_w\right)(\boldsymbol{y}) - \left(\mathcal{K}_{\boldsymbol{H}} * f_{uw}\right)^2(\boldsymbol{y})\right\},$$

where

$$\gamma_w(\boldsymbol{y}) = \frac{\mu_w f_{uw}(\boldsymbol{y})}{w(\boldsymbol{y})}.$$

Proof: (a) From (4.2.6), one has

$$
\begin{aligned}
\mathrm{E}\left[\hat{f}_{uw}(\mathbf{y})\right] &= \mathrm{E}\left[n^{-1}\mu_w \sum_{i=1}^{n} w(\mathbf{Y}_i)^{-1}\mathcal{K}_H(\mathbf{y}-\mathbf{Y}_i)\right] \\
&= n^{-1}\sum_{i=1}^{n}\mathrm{E}\left[\mu_w w(\mathbf{Y}_i)^{-1}\mathcal{K}_H(\mathbf{y}-\mathbf{Y}_i)\right] \\
&= \mathrm{E}_{g_w}\left[\mu_w w(\mathbf{Y}_1)^{-1}\mathcal{K}_{\mathbf{H}}(\mathbf{y}-\mathbf{Y}_1)\right] \\
&= \int_{\mathbb{R}^d}\mu_w w(\mathbf{y}_1)^{-1}\mathcal{K}_{\mathbf{H}}(\mathbf{y}-\mathbf{y}_1)g_w(\mathbf{y}_1)\,d\mathbf{y}_1 \\
&= \int_{\mathbb{R}^d}\mathcal{K}_{\mathbf{H}}(\mathbf{y}-\mathbf{y}_1)\frac{\mu_w}{w(\mathbf{y}_1)}g_w(\mathbf{y}_1)\,d\mathbf{y}_1 \\
&= \int_{\mathbb{R}^d}\mathcal{K}_{\mathbf{H}}(\mathbf{y}-\mathbf{y}_1)f_{uw}(\mathbf{y}_1)\,d\mathbf{y}_1 \\
&= \mathcal{K}_{\mathbf{H}} * f_{uw}(\mathbf{y}),
\end{aligned}
$$

and consequently,

$$
\mathrm{Bias}\left(\hat{f}_{uw}(\mathbf{y})\right) = (\mathcal{K}_{\mathbf{H}} * f_{uw})(\mathbf{y}) - f_{uw}(\mathbf{y}).
$$

(b) The use of Equation (4.2.6) leads to

$$
\begin{aligned}
\mathrm{Var}\left(\hat{f}_{uw}(\mathbf{y})\right) &= \mathrm{Var}\left(n^{-1}\mu_w\sum_{i=1}^{n} w(\mathbf{Y}_i)^{-1}\mathcal{K}_{\mathbf{H}}(\mathbf{y}-\mathbf{Y}_i)\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}\left(\mu_w w(\mathbf{Y}_i)^{-1}\mathcal{K}_{\mathbf{H}}(\mathbf{y}-\mathbf{Y}_i)\right) \\
&= \frac{1}{n}\mathrm{Var}\left(\mu_w w(\mathbf{Y}_1)^{-1}\mathcal{K}_{\mathbf{H}}(\mathbf{y}-\mathbf{Y}_1)\right) \\
&= \frac{1}{n}\mathrm{E}_{g_w}\left[\mu_w w(\mathbf{Y}_1)^{-1}\mathcal{K}_{\mathbf{H}}(\mathbf{y}-\mathbf{Y}_1)\right]^2 \\
&\quad -\frac{1}{n}\left(\mathrm{E}_{g_w}\left[\mu_w w(\mathbf{Y}_1)^{-1}\mathcal{K}_{\mathbf{H}}(\mathbf{y}-\mathbf{Y}_1)\right]\right)^2 \\
&= \frac{1}{n}\int_{\mathbb{R}^d}\frac{\mu_w^2}{w(\mathbf{y}_1)^2}\mathcal{K}_{\mathbf{H}}^2(\mathbf{y}-\mathbf{y}_1)\,g_w(\mathbf{y}_1)\,d\mathbf{y}_1 \\
&\quad -\frac{1}{n}\left(\int_{\mathbb{R}^d}\frac{\mu_w}{w(\mathbf{y}_1)}\mathcal{K}_{\mathbf{H}}(\mathbf{y}-\mathbf{y}_1)\,g_w(\mathbf{y}_1)\,d\mathbf{y}_1\right)^2.
\end{aligned}
$$

So that,

$$\begin{aligned}\operatorname{Var}\left(\hat{f}_{uw}\left(\mathbf{y}\right)\right) &= \frac{1}{n}\int_{\mathbb{R}^d}\mathcal{K}^2_{\mathbf{H}}\left(\mathbf{y}-\mathbf{y}_1\right)\frac{\mu_w}{w\left(\mathbf{y}_1\right)}f_{uw}\left(\mathbf{y}_1\right)d\mathbf{y}_1\\ &\quad -\frac{1}{n}\left(\int_{\mathbb{R}^d}\mathcal{K}_{\mathbf{H}}\left(\mathbf{y}-\mathbf{y}_1\right)f_{uw}\left(\mathbf{y}_1\right)d\mathbf{y}_1\right)^2.\end{aligned}$$

Now, by letting

$$\gamma_w\left(\mathbf{y}_1\right)=\frac{\mu_w}{w\left(\mathbf{y}_1\right)}f_{uw}\left(\mathbf{y}_1\right),$$

one concludes that

$$\operatorname{Var}\left(\hat{f}_{uw}\left(\mathbf{y}\right)\right)=n^{-1}\left\{\left(\mathcal{K}^2_{\mathbf{H}}*\gamma_w\right)\left(\mathbf{y}\right)-\left(\mathcal{K}_{\mathbf{H}}*f_{uw}\right)^2\left(\mathbf{y}\right)\right\}.$$

∎

The following properties of $\hat{f}_{uw}\left(\mathbf{y}\right)$, proposed by Ibrahim [28], hold when μ_w is known and can be viewed as a generalization of those developed by Jones [30].

Property 4.2.2 *Let $\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_n$ be a random sample from the multivariate weighted density $g_w\left(\boldsymbol{y}\right)$.*

(a) The estimator $\hat{f}_{uw}\left(\boldsymbol{y}\right)$ of the multivariate density $f_{uw}\left(\boldsymbol{y}\right)$ is a PDF.

(b) Let $\nabla_{f_{uw}}\left(\boldsymbol{y}\right)$ be the vector of first-order partial derivatives of f_{uw}. If all entries of $\mathcal{H}_{f_{uw}}\left(\boldsymbol{y}\right)$ exist, then the asymptotic bias of $\hat{f}_{uw}\left(\boldsymbol{y}\right)$ can be expressed by:

$$ABias\left(\hat{f}_{uw}\left(\boldsymbol{y}\right)\right)=\frac{1}{2}\mu_2\left(\mathcal{K}\right)tr\left\{\boldsymbol{H}'\mathcal{H}_f(\boldsymbol{y})\boldsymbol{H}\right\}\cdot$$

(c) If all entries of $\boldsymbol{H}$ approach $\boldsymbol{0}$ as $n\to\infty$, then the asymptotic variance of the estimator $\hat{f}_{uw}\left(\boldsymbol{y}\right)$ is:

$$AVar\left(\hat{f}_{uw}\left(\boldsymbol{y}\right)\right)=\lim_{n\to\infty}Var\left(\hat{f}_{uw}\left(\boldsymbol{y}\right)\right)=\frac{1}{n|\boldsymbol{H}|}\|\mathcal{K}\|_2^2\gamma_w\left(\boldsymbol{y}\right)\cdot$$

(d) *Suppose that* $\mathcal{H}_{f_{uw}}(\boldsymbol{y})$ *exist. Then,*

$$MSE\left(\hat{f}_{uw}(\boldsymbol{y})\right) \simeq \frac{1}{4}\mu_2^2(\mathcal{K})\, tr^2\left\{\boldsymbol{H}'\mathcal{H}_f(\boldsymbol{y})\boldsymbol{H}\right\} + \frac{1}{n|\boldsymbol{H}|}\|\mathcal{K}\|_2^2\gamma_w(\boldsymbol{y}).$$

Moreover, if all entries of $\boldsymbol{H}$ *approach* $\boldsymbol{0}$ *such that* $n|\boldsymbol{H}| \to \infty$ *as* $n \to \infty$, *then*

$$\lim_{n\to\infty} MSE\left(\hat{f}_{uw}(\boldsymbol{y})\right) = 0.$$

(e) *If all entries of* $\mathcal{H}_{f_{uw}}(\boldsymbol{y})$ *exist and all entries of* $\boldsymbol{H}$ *approach* $\boldsymbol{0}$ *and* $n|\boldsymbol{H}| \to \infty$ *as* $n \to \infty$, *then* $\hat{f}_{uw}(\boldsymbol{y})$ *is a consistent estimator of* $f_{uw}(\boldsymbol{y})$.

4.3 Partial, conditional and joint measures of dependence for length-biased data

In this section, we start by providing a form of the multivariate length-biased density under both the dependence model (survival time and vector of covariates are dependent) and under the independence model (survival time and vector of covariates are independent) and then we examine some links between the partial, conditional and joint information gain.

4.3.1 Multivariate length-biased density under the dependence and independence models

For all that follows in this chapter, let $(U, \mathbf{Z})$ be a random vector with multivariate length-biased density $f_{LB}(u, \mathbf{z})$, where U denotes a survival time and $\mathbf{Z}$ is a vector of d-covariates. Suppose that $\mathbf{Z}$ is partitioned as follows $\mathbf{Z} = \left(\mathbf{Z}^{(1)}, \mathbf{Z}^{(2)}\right)$, where $\mathbf{Z}^{(1)} = (Z_1, \ldots, Z_p)'$ and $\mathbf{Z}^{(2)} = (Z_{p+1}, \ldots, Z_{p+q})'$ with $d = p+q$. Suppose that the multivariate data $(U_k, \mathbf{Z}_k)$, $k = 1, \ldots, n$ come from $f_{LB}(u, \mathbf{z})$, where $\mathbf{Z}_k = (Z_{k1}, \ldots, Z_{kd})$ which can be partitioned with respect to the partition of $\mathbf{Z}$ as $\mathbf{Z}_k = \left(\mathbf{Z}_k^{(1)}, \mathbf{Z}_k^{(2)}\right)$ with

$\mathbf{Z}_k^{(1)} = (Z_{k1}, \ldots, Z_{kp})'$ and $\mathbf{Z}_k^{(2)} = \left(Z_{k(p+1)}, \ldots, Z_{k(p+q)}\right)'$.

First, we generalize Theorem 3.1.1, given in Chapter 3, as follows.

Theorem 4.3.1

(a) *If U and $\boldsymbol{Z}$ are dependent then the multivariate length-biased density takes the following form*

$$f_{LB}(u, \boldsymbol{z}) = f_{LB}(u|\boldsymbol{z}) f_B(\boldsymbol{z}) = \frac{u f_U(u, \boldsymbol{z})}{\mu}, \tag{4.3.1}$$

where $f_{LB}(u|\boldsymbol{z})$ is the length-biased density of U conditional on $\boldsymbol{Z} = \boldsymbol{z}$, $f_B(\boldsymbol{z})$ is the multivariate biased density of $\boldsymbol{Z}$, $f_U(u, \boldsymbol{z})$ is the multivariate unbiased density of the random vector $(U, \boldsymbol{Z})$ and the overall mean lifetime of the unbiased population is $\mu = \int_{\mathbb{R}^{d+1}} u f_U(u, \boldsymbol{z}) du d\boldsymbol{z} = \int u f_U(u) du < \infty$.

(b) *If U and $\boldsymbol{Z}$ are independent then the multivariate length-biased density can be written as*

$$f_{LB}(u, \boldsymbol{z}) = f_{LB}(u) f_{\boldsymbol{Z}}(\boldsymbol{z}) = \frac{u f_U(u)}{\mu} f_{\boldsymbol{Z}}(\boldsymbol{z}), \tag{4.3.2}$$

where $f_{\boldsymbol{Z}}(\boldsymbol{z})$ is the unbiased density of the covariate.

The proof of the theorem above is similar to the proof of Theorem 3.1.1.

4.3.2 Partial information gain under several covariates

Proposition 4.3.2 *The measure of dependence between a survival time U and p-covariates $\boldsymbol{Z}^{(1)}$ conditional on q-covariates $\boldsymbol{Z}^{(2)}$ is*

$$\rho^2_{PA}\left(U, \boldsymbol{Z}^{(1)} | \boldsymbol{Z}^{(2)}\right) = 1 - \exp\left\{-\Gamma_{PA}\right\}, \tag{4.3.3}$$

where Γ_{PA} is the partial information gain given by

$$\begin{aligned}\Gamma_{PA} &= 2\left\{\int_{\mathbb{R}^{d+1}} \log\left\{f_{LB}\left(u, \boldsymbol{z}^{(1)} \left| \boldsymbol{z}^{(2)}\right.\right)\right\} f_{LB}(u, \boldsymbol{z})\, du d\boldsymbol{z}\right. \\ &\quad \left. - \int_{\mathbb{R}^{d+1}} \log\left\{f_{LB}\left(u, \boldsymbol{z}^{(1)}\right)\right\} f_{LB}(u, \boldsymbol{z})\, du d\boldsymbol{z}\right\}. \end{aligned} \tag{4.3.4}$$

Proof: We consider two models

$$\text{Independence}: \; f_{LB}\left(u, \mathbf{z}^{(1)} \left| \mathbf{z}^{(2)} \right.\right) = f_{LB}\left(u, \mathbf{z}^{(1)}\right), \text{ for all } u, \mathbf{z}^{(1)},$$

$$\text{Dependence}: \; f_{LB}\left(u, \mathbf{z}^{(1)} \left| \mathbf{z}^{(2)} \right.\right) \neq f_{LB}\left(u, \mathbf{z}^{(1)}\right), \text{ for some } u, \mathbf{z}^{(1)}.$$

The conditional information gain is

$$\Gamma_{\mathtt{PA}} = 2\left\{\phi_1 - \phi_0\right\}, \tag{4.3.5}$$

where under independence

$$\phi_0 = \int_{\mathbb{R}^{d+1}} \log\left\{f_{LB}\left(u, \mathbf{z}^{(1)}\right)\right\} f_{LB}\left(u, \mathbf{z}\right) du d\mathbf{z}, \tag{4.3.6}$$

and under dependence

$$\phi_1 = \int_{\mathbb{R}^{d+1}} \log\left\{f_{LB}\left(u, \mathbf{z}^{(1)} \left| \mathbf{z}^{(2)} \right.\right)\right\} f_{LB}\left(u, \mathbf{z}\right) du d\mathbf{z}. \tag{4.3.7}$$

The partial information gain can be expressed as

$$\begin{aligned} \Gamma_{\mathtt{PA}} &= 2\left\{\int_{\mathbb{R}^{d+1}} \log\left\{f_{LB}\left(u, \mathbf{z}^{(1)} \left| \mathbf{z}^{(2)} \right.\right)\right\} f_{LB}\left(u, \mathbf{z}\right) du d\mathbf{z} \right. \\ &\quad \left. - \int_{\mathbb{R}^{d+1}} \log\left\{f_{LB}\left(u, \mathbf{z}^{(1)}\right)\right\} f_{LB}\left(u, \mathbf{z}\right) du d\mathbf{z}\right\}, \end{aligned}$$

and the partial measure of dependence is

$$\rho^2_{\mathtt{PA}}\left(U, \mathbf{Z}^{(1)} | \mathbf{Z}^{(2)}\right) = 1 - \exp\left\{-\Gamma_{\mathtt{PA}}\right\}.$$

∎

The following two corollaries can be viewed both as special cases of Proposition 4.3.2. Hence, for the next sections, we restrict our attention on the general case: partial information gain and partial dependence measure.

4.3.3 Conditional and joint information gain under several covariates

Corollary 4.3.3 *If the vector of p-covariates* $\boldsymbol{Z}^{(1)}$ *is absent then the partial measure of dependence and partial information gain given, respectively, by (4.3.3) and (4.3.4) become*

$$\rho_C^2\left(U|\boldsymbol{Z}\right) = 1 - \exp\left\{-\Gamma_C\right\}, \tag{4.3.8}$$

$$\Gamma_C = 2\left\{\int_{\mathbb{R}^{q+1}} \log\left\{f_{LB}\left(u\left|\boldsymbol{z}\right.\right)\right\} f_{LB}\left(u,\boldsymbol{z}\right) du d\boldsymbol{z} - \int_{\mathbb{R}} \log\left\{f_{LB}\left(u\right)\right\} f_{LB}\left(u\right) du\right\}. \tag{4.3.9}$$

Proof: If $\mathbf{Z}^{(1)}$ is absent then $\mathbf{Z} = \mathbf{Z}^{(2)}$. This means, by Proposition 4.3.2, that it returns to seek a conditional measure of dependence which is based on the conditional information gain. Under this case, Equations (4.3.3) and (4.3.4) in Proposition 4.3.2, respectively, become

$$\rho_C^2\left(U|\mathbf{Z}\right) = 1 - \exp\left\{-\Gamma_C\right\},$$

and

$$\begin{aligned}
\Gamma_C &= 2\left\{\int_{\mathbb{R}^{q+1}} \log\left\{f_{LB}\left(u\left|\mathbf{z}^{(2)}\right.\right)\right\} f_{LB}\left(u,\mathbf{z}\right) du d\mathbf{z}\right. \\
&\quad \left. - \int_{\mathbb{R}^{q+1}} \log\left\{f_{LB}\left(u\right)\right\} f_{LB}\left(u,\mathbf{z}\right) du d\mathbf{z}\right\} \\
&= 2\left\{\int_{\mathbb{R}^{q+1}} \log\left\{f_{LB}\left(u\left|\mathbf{z}\right.\right)\right\} f_{LB}\left(u,\mathbf{z}\right) du d\mathbf{z}\right. \\
&\quad \left. - \int_{\mathbb{R}} \log\left\{f_{LB}\left(u\right)\right\} f_{LB}\left(u\right) du\right\},
\end{aligned}$$

where we used $f_{LB}\left(u\right) = \int_{\mathbb{R}^q} f_{LB}\left(u,\mathbf{z}\right) d\mathbf{z}$. ∎

Corollary 4.3.4 *If the vector of q-covariates* $\boldsymbol{Z}^{(2)}$ *is absent then the partial information gain given in (4.3.4) converts to joint information gain*

$$\Gamma = \Gamma_C + \Gamma_B, \tag{4.3.10}$$

where

$$\Gamma_C = 2\left\{\int_{\mathbb{R}^{p+1}} \log\{f_{LB}(u|\boldsymbol{z})\} f_{LB}(u,\boldsymbol{z})\, du d\boldsymbol{z} - \int_{\mathbb{R}} \log\{f_{LB}(u)\} f_{LB}(u)\, du\right\}, \tag{4.3.11}$$

and

$$\Gamma_B = 2\left\{\int_{\mathbb{R}^{p}} \log\{f_{B}(\boldsymbol{z})\} f_{B}(\boldsymbol{z})\, d\boldsymbol{z} - \int_{\mathbb{R}^p} \log\{f_{\boldsymbol{Z}}(\boldsymbol{z})\} f_{B}(\boldsymbol{z})\, d\boldsymbol{z}\right\}. \tag{4.3.12}$$

Moreover, the partial measure of dependence given by (4.3.3) converts to the joint measure of dependence as

$$\rho_J^2(U,\boldsymbol{Z}) = 1 - \exp\{-(\Gamma_C + \Gamma_B)\}. \tag{4.3.13}$$

Proof: If $\mathbf{Z}^{(2)}$ is absent then $\mathbf{Z} = \mathbf{Z}^{(1)}$. So that, by Proposition 4.3.2, we are in the case where we seek a joint measure of dependence. Equation (4.3.4) in Proposition 4.3.2 becomes

$$\begin{aligned}
\Gamma &= 2\left\{\int_{\mathbb{R}^{p+1}} \log\{f_{LB}(u,\mathbf{z})\} f_{LB}(u,\mathbf{z})\, du d\mathbf{z}\right. \\
&\quad \left. - \int_{\mathbb{R}^{p+1}} \log\{f_{LB}(u)\, f_{\mathbf{Z}}(\mathbf{z})\} f_{LB}(u,\mathbf{z})\, du d\mathbf{z}\right\} \\
&= 2\left\{\int_{\mathbb{R}^{p+1}} \log\{f_{LB}(u|\mathbf{z})\, f_{B}(\mathbf{z})\} f_{LB}(u,\mathbf{z})\, du d\mathbf{z}\right. \\
&\quad \left. - \int_{\mathbb{R}^{p+1}} \log\{f_{LB}(u)\} f_{LB}(u,\mathbf{z})\, du d\mathbf{z} - \int_{\mathbb{R}^{p+1}} \log\{f_{\mathbf{Z}}(\mathbf{z})\} f_{LB}(u,\mathbf{z})\, du d\mathbf{z}\right\} \\
&= 2\left\{\int_{\mathbb{R}^{p+1}} \log\{f_{LB}(u|\mathbf{z})\} f_{LB}(u,\mathbf{z})\, du d\mathbf{z} + \int_{\mathbb{R}^{p+1}} \log\{f_{B}(\mathbf{z})\} f_{LB}(u,\mathbf{z})\, du d\mathbf{z}\right. \\
&\quad \left. - \int_{\mathbb{R}^{p+1}} \log\{f_{LB}(u)\} f_{LB}(u,\mathbf{z})\, du d\mathbf{z} - \int_{\mathbb{R}^{p+1}} \log\{f_{\mathbf{Z}}(\mathbf{z})\} f_{LB}(u,\mathbf{z})\, du d\mathbf{z}\right\} \\
&= 2\left\{\int_{\mathbb{R}^{p+1}} \log\{f_{LB}(u|\mathbf{z})\} f_{LB}(u,\mathbf{z})\, du d\mathbf{z} + \int_{\mathbb{R}^{p}} \log\{f_{B}(\mathbf{z})\} f_{B}(\mathbf{z})\, d\mathbf{z}\right. \\
&\quad \left. - \int_{\mathbb{R}} \log\{f_{LB}(u)\} f_{LB}(u)\, du - \int_{\mathbb{R}^{p}} \log\{f_{\mathbf{Z}}(\mathbf{z})\} f_{B}(\mathbf{z})\, d\mathbf{z}\right\},
\end{aligned}$$

where we used $f_{LB}(u) = \int_{\mathbb{R}^p} f_{LB}(u,\mathbf{z}) d\mathbf{z}$ and $f_B(\mathbf{z}) = \int_{\mathbb{R}} f_{LB}(u,\mathbf{z}) du$. It follows that,

$$\Gamma = \Gamma_C + \Gamma_B,$$

where

$$\Gamma_C = 2\left\{\int_{\mathbb{R}^{p+1}} \log\left\{f_{LB}\left(u|\mathbf{z}\right)\right\} f_{LB}\left(u,\mathbf{z}\right) du d\mathbf{z} - \int_{\mathbb{R}} \log\left\{f_{LB}\left(u\right)\right\} f_{LB}\left(u\right) du\right\},$$

and

$$\Gamma_B = 2\left\{\int_{\mathbb{R}^p} \log\left\{f_B\left(\mathbf{z}\right)\right\} f_B\left(\mathbf{z}\right) d\mathbf{z} - \int_{\mathbb{R}^p} \log\left\{f_{\mathbf{Z}}\left(\mathbf{z}\right)\right\} f_B\left(\mathbf{z}\right) d\mathbf{z}\right\}.$$

Therefore, the joint measure of dependence is

$$\rho_J^2\left(U,\mathbf{Z}\right) = 1 - \exp\left\{-\left(\Gamma_C + \Gamma_B\right)\right\}.$$

∎

4.4 Estimation procedure for the partial information gain and partial measure of dependence

In this section, we start by writing the partial information gain in terms of expectations and then we generalize the kernel density estimation with a regression procedure, used in Chapter 3, to obtain an estimator of $\Gamma_{\mathtt{PA}}$ and $\rho^2_{\mathtt{PA}}\left(U,\mathbf{Z}^{(1)}|\mathbf{Z}^{(2)}\right)$.

Proposition 4.4.1 *The partial information gain can be expressed as*

$$\Gamma_{PA} = \Gamma_{PA,1} + \Gamma_{PA,2}, \tag{4.4.1}$$

where

$$\Gamma_{PA,1} = 2\left\{E\left[\log\left\{f_{LB}\left(U|\boldsymbol{Z}\right)\right\}\right] - E\left[\log\left\{f_{LB}\left(U|\boldsymbol{Z}^{(1)}\right)\right\}\right]\right\}, \tag{4.4.2}$$

and

$$\Gamma_{PA,2} = 2\left\{E\left[\log\left\{f_B\left(\boldsymbol{Z}^{(1)}|\boldsymbol{Z}^{(2)}\right)\right\}\right] - E\left[\log\left\{f_B\left(\boldsymbol{Z}^{(1)}\right)\right\}\right]\right\}. \tag{4.4.3}$$

Proof: Based on (4.3.4), we write $\Gamma_{\texttt{PA}}$ as

$$\begin{aligned}
\Gamma_{\texttt{PA}} &= 2\left\{\mathrm{E}\left[\log\left\{f_{LB}\left(U,\mathbf{Z}^{(1)}|\mathbf{Z}^{(2)}\right)\right\}\right] - \mathrm{E}\left[\log\left\{f_{LB}\left(U,\mathbf{Z}^{(1)}\right)\right\}\right]\right\} \\
&= 2\left\{\mathrm{E}\left[\log\left\{f_{LB}\left(U|\mathbf{Z}^{(1)},\mathbf{Z}^{(2)}\right) f_{B}\left(\mathbf{Z}^{(1)}|\mathbf{Z}^{(2)}\right)\right\}\right]\right. \\
&\quad \left. -\mathrm{E}\left[\log\left\{f_{LB}\left(U|\mathbf{Z}^{(1)}\right) f_{B}\left(\mathbf{Z}^{(1)}\right)\right\}\right]\right\} \\
&= 2\left\{\mathrm{E}\left[\log\left\{f_{LB}\left(U|\mathbf{Z}\right)\right\}\right] + \mathrm{E}\left[\log\left\{f_{B}\left(\mathbf{Z}^{(1)}|\mathbf{Z}^{(2)}\right)\right\}\right]\right. \\
&\quad \left. -\mathrm{E}\left[\log\left\{f_{LB}\left(U|\mathbf{Z}^{(1)}\right)\right\}\right] -\mathrm{E}\left[\log\left\{f_{B}\left(\mathbf{Z}^{(1)}\right)\right\}\right]\right\} \\
&= 2\left\{\mathrm{E}\left[\log\left\{f_{LB}\left(U|\mathbf{Z}\right)\right\}\right] - \mathrm{E}\left[\log\left\{f_{LB}\left(U|\mathbf{Z}^{(1)}\right)\right\}\right]\right. \\
&\quad \left. + \mathrm{E}\left[\log\left\{f_{B}\left(\mathbf{Z}^{(1)}|\mathbf{Z}^{(2)}\right)\right\}\right] - \mathrm{E}\left[\log\left\{f_{B}\left(\mathbf{Z}^{(1)}\right)\right\}\right]\right\}.
\end{aligned}$$

Letting

$$\Gamma_{\texttt{PA},1} = 2\left\{\mathrm{E}\left[\log\left\{f_{LB}\left(U|\mathbf{Z}\right)\right\}\right] - \mathrm{E}\left[\log\left\{f_{LB}\left(U|\mathbf{Z}^{(1)}\right)\right\}\right]\right\},$$

and

$$\Gamma_{\texttt{PA},2} = 2\left\{\mathrm{E}\left[\log\left\{f_{B}\left(\mathbf{Z}^{(1)}|\mathbf{Z}^{(2)}\right)\right\}\right] - \mathrm{E}\left[\log\left\{f_{B}\left(\mathbf{Z}^{(1)}\right)\right\}\right]\right\},$$

we get

$$\Gamma_{\texttt{PA}} = \Gamma_{\texttt{PA},1} + \Gamma_{\texttt{PA},2}.$$

∎

From Proposition 4.4.1, we can see that the decomposition of the partial information gain is related to the densities $f_{LB}\left(u|\mathbf{z}\right)$, $f_{LB}\left(u|\mathbf{z}^{(1)}\right)$, $f_{B}\left(\mathbf{z}^{(1)}|\mathbf{z}^{(2)}\right)$ and $f_{B}\left(\mathbf{z}^{(1)}\right)$. Therefore, to estimate $\Gamma_{\texttt{PA}}$, we propose to use the multivariate kernel density estimation with a multiple regression procedure. This approach can be viewed as a generalization of the first method developed in the previous chapter.

4.4.1 Estimation procedure for the length-biased density of lifetime conditional on a fixed vector of covariates

Recall that $f_{LB}(u,\mathbf{z}) = f_{LB}(u|\mathbf{z})\, f_B(\mathbf{z})$, where $f_{LB}(u|\mathbf{z})$ is the length-biased density of U given a vector of d-covariates $\mathbf{Z} = \mathbf{z}$ given by

$$f_{LB}(u|\mathbf{z}) = \frac{u f_U(u|\mathbf{z})}{\mu(\mathbf{z})}, \tag{4.4.4}$$

where $\mu(\mathbf{z}) = \int u f_U(u|\mathbf{z}) du$ is finite and $f_U(u|\mathbf{z})$ is the unbiased density of U conditional on $\mathbf{Z} = \mathbf{z}$. We can adapt the estimator of Jones [30] to find an estimator of $f_U(u\,|\mathbf{z})$, by using the multiple linear regression model

$$\phi(U) = Y = \alpha + \boldsymbol{\beta}'\mathbf{Z} + \varepsilon, \tag{4.4.5}$$

where ϕ is a monotone increasing transformation, α is an intercept, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_d)'$ is a vector of regression coefficients and ε is a r.v. (error variate) independent of $\mathbf{Z}$. To obtain the pseudo-observations from $f_{LB}(u\,|\mathbf{z})$, we generalize Algorithm 3.5.1 given in Section 3.5.1, simply by replacing one covariate Z by the vector of d-covariates $\mathbf{Z} = (Z_1, \ldots, Z_d)'$, β by $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_d)'$ and $Z = z$ by $\mathbf{Z} = \mathbf{z} = (z_1, \ldots, z_d)'$. In fact, the pseudo-observations from $f_{LB}(u\,|\mathbf{z})$ can be obtained as follows

$$\tilde{U}_i = \phi^{-1}\left(\tilde{Y}_i\right) = \phi^{-1}\left(\hat{\alpha} + \hat{\boldsymbol{\beta}}'\mathbf{z} + \tilde{\varepsilon}_i\right),\ i = 1, \ldots, n. \tag{4.4.6}$$

The adapted estimator, $\hat{f}_U(u\,|\mathbf{z})$, of Jones [30] becomes

$$\hat{f}_U(u\,|\mathbf{z}) = n^{-1}\hat{\mu}(\mathbf{z}) \sum_{i=1}^{n} \tilde{U}_i^{-1} K_h\left(u - \tilde{U}_i\right), \tag{4.4.7}$$

where

$$\hat{\mu}(\mathbf{z}) = n\left(\sum_{i=1}^{n} \tilde{U}_i^{-1}\right)^{-1} = n\left(\sum_{i=1}^{n} \frac{1}{\phi^{-1}\left(\hat{\alpha} + \hat{\boldsymbol{\beta}}'\mathbf{z} + \tilde{\varepsilon}_i\right)}\right)^{-1}. \tag{4.4.8}$$

Substituting (4.4.6) into (4.4.7), leads to

$$\hat{f}_U(u\,|\mathbf{z}) = n^{-1}\hat{\mu}(\mathbf{z}) \sum_{i=1}^{n} \frac{1}{\phi^{-1}\left(\hat{\alpha} + \hat{\boldsymbol{\beta}}'\mathbf{z} + \tilde{\varepsilon}_i\right)} K_h\left(u - \phi^{-1}\left(\hat{\alpha} + \hat{\boldsymbol{\beta}}'\mathbf{z} + \tilde{\varepsilon}_i\right)\right). \tag{4.4.9}$$

From (4.4.4), a density estimator of $f_{LB}(u\,|\mathbf{z})$ is

$$\hat{f}_{LB}(u\,|\mathbf{z}) = \frac{u\hat{f}_U(u\,|\mathbf{z})}{\hat{\mu}(\mathbf{z})},$$

where $\hat{\mu}(\mathbf{z})$ and $\hat{f}_U(u\,|\mathbf{z})$ are, respectively, given by (4.4.8) and (4.4.9). Consequently,

$$\hat{f}_{LB}(u\,|\mathbf{z}) = n^{-1}\sum_{i=1}^{n}\frac{u}{\phi^{-1}\left(\hat{\alpha}+\hat{\boldsymbol{\beta}}'\mathbf{z}+\tilde{\varepsilon}_i\right)}K_h\left(u-\phi^{-1}\left(\hat{\alpha}+\hat{\boldsymbol{\beta}}'\mathbf{z}+\tilde{\varepsilon}_i\right)\right). \quad (4.4.10)$$

We note that, the estimator given above is a generalization of (3.5.11). Now, following the partition of $\mathbf{Z}=\left(\mathbf{Z}^{(1)},\mathbf{Z}^{(2)}\right)$, let $\boldsymbol{\beta}=\left(\boldsymbol{\beta}^{(1)},\boldsymbol{\beta}^{(2)}\right)$ and $\mathbf{z}=\left(\mathbf{z}^{(1)},\mathbf{z}^{(2)}\right)$. An estimator of $f_{LB}\left(u\,\middle|\mathbf{z}^{(1)}\right)$ is

$$\hat{f}_{LB}\left(u\,\middle|\mathbf{z}^{(1)}\right) = \frac{u\hat{f}_U\left(u\,\middle|\mathbf{z}^{(1)}\right)}{\hat{\mu}\left(\mathbf{z}^{(1)}\right)},$$

where $\hat{f}_U\left(u\,\middle|\mathbf{z}^{(1)}\right)$ and $\hat{\mu}\left(\mathbf{z}^{(1)}\right)$ can be deduced from the model (4.4.5) and the Algorithm 3.5.1 through a random sample $\left(U_i,\mathbf{Z}_i^{(1)}\right)$, $i=1,\ldots,n$. Hence,

4.4.2 Estimation procedure for the multivariate density of several covariates under the independence and dependence models

Given a length-biased data $(U_1,\mathbf{Z}_1),\ldots,(U_n,\mathbf{Z}_n)$ from $f_{LB}(u,\mathbf{z})$, our goal is to provide an estimator for the multivariate density of d-covariates $\mathbf{Z}$ under the independence model (U and $\mathbf{Z}$ are independent) and under the dependence model (U and $\mathbf{Z}$ are dependent). Recall that the multivariate biased density of $\mathbf{Z}$ under the dependence model is

$$f_B(\mathbf{z}) = \frac{\mu(\mathbf{z})\,f_{\mathbf{Z}}(\mathbf{z})}{\mu}, \quad (4.4.11)$$

where $f_{\mathbf{Z}}(\mathbf{z})$ denotes the multivariate unbiased density of $\mathbf{Z}$ and $\mu=\int_{\mathbb{R}^d}\mu(\mathbf{z})\,f_{\mathbf{Z}}(\mathbf{z})d\mathbf{z}$ is finite.

Under the independence model, Equation (4.4.11) becomes

$$f_B(\mathbf{z}) = f_{\mathbf{Z}}(\mathbf{z}), \tag{4.4.12}$$

since $\mu(\mathbf{z}) = \mathrm{E}[U|\mathbf{Z} = \mathbf{z}] = \mathrm{E}[U] = \mu$. It follows that, the estimator of $f_{\mathbf{Z}}(\mathbf{z})$ must take into account the fact that U and $\mathbf{Z}$ are independent. However, an estimator of $f_B(\mathbf{z})$ should contain an appropriate estimator of $\mu(\mathbf{z})$ because the weight function $\mu(\mathbf{z})$ involved in (4.4.11) reflects some dependence between U and $\mathbf{Z}$. In this way, we propose to use the multiple linear regression model given in (4.4.5)

$$\phi(U) = Y = \alpha + \boldsymbol{\beta}'\mathbf{Z} + \varepsilon.$$

A simple way to get a closed form of $\mu(\mathbf{z})$ in (4.4.11) is to use the AFT model. Therefore, a generalization of (3.5.19) gives a new form of (4.4.11) as follows

$$f_B(\mathbf{z}) = \frac{\exp\{\boldsymbol{\beta}'\mathbf{z}\} f_{\mathbf{Z}}(\mathbf{z})}{\nu_{\boldsymbol{\beta}}}, \tag{4.4.13}$$

where $\nu_{\boldsymbol{\beta}} = \int_{\mathbb{R}^d} \exp\{\boldsymbol{\beta}'\mathbf{z}\} f_{\mathbf{Z}}(\mathbf{z})d\mathbf{z} < \infty$. We can see that, Equation (4.4.13) is of the form of (4.2.1). In others words, $f_B(\mathbf{z})$ is a multivariate weighted density, $f_{\mathbf{Z}}(\mathbf{z})$ is a multivariate unweighted density and the weight function is $w(\mathbf{z}) = \exp\{\boldsymbol{\beta}'\mathbf{z}\}$. As in the univariate case, we cannot use the unweighed density estimator given in (4.2.6)

$$\hat{f}_{uw}(\mathbf{z}) = n^{-1}\hat{\mu}_w \sum_{i=1}^{n} w(\mathbf{Z}_i)^{-1}\mathcal{K}_{\mathbf{H}}(\mathbf{z} - \mathbf{Z}_i),$$

to estimate $f_{\mathbf{Z}}(\mathbf{z})$ because the weight function $w(\mathbf{z}) = \exp\{\boldsymbol{\beta}'\mathbf{z}\}$ in (4.4.13) depends on the vector of regression coefficients $\boldsymbol{\beta}$, considered as a vector of dependence parameters between U and $\mathbf{Z}$. In this case, an appropriate estimator of $f_{\mathbf{Z}}(\mathbf{z})$ can be based on (4.2.7) as

$$\hat{f}_{\mathbf{Z}}(\mathbf{z}) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{K}_{\mathbf{H}}(\mathbf{z} - \mathbf{Z}_i^*), \tag{4.4.14}$$

where $\mathcal{Z}^* = (\mathbf{Z}_1^*, \ldots, \mathbf{Z}_n^*)$ is obtained, using the bootstrap techniques with replacement, from the original sample $\mathcal{Z} = (\mathbf{Z}_1, \ldots, \mathbf{Z}_n)$ choosing $\mathbf{Z}_i$ with probability p_i given by (4.2.8)

$$p_i = \frac{w(\mathbf{Z}_i)^{-1}}{\sum\limits_{i=1}^{n} w(\mathbf{Z}_i)^{-1}} = \frac{\exp\{-\boldsymbol{\beta}'\mathbf{Z}_i\}}{\sum\limits_{i=1}^{n} \exp\{-\boldsymbol{\beta}'\mathbf{Z}_i\}}. \tag{4.4.15}$$

So that, once $\boldsymbol{\beta}$ is estimated by $\hat{\boldsymbol{\beta}}$ using the same way discussed in Algorithm 3.5.1, we can estimate $f_B(\mathbf{z})$ by

$$\hat{f}_B(\mathbf{z}) = \frac{\exp\left\{\hat{\boldsymbol{\beta}}'\mathbf{z}\right\} \hat{f}_{\mathbf{Z}}(\mathbf{z})}{\hat{\nu}_{\hat{\boldsymbol{\beta}}}}, \tag{4.4.16}$$

where $\hat{f}_{\mathbf{Z}}(\mathbf{z})$ is given by (4.4.14) and $\hat{\nu}_{\hat{\boldsymbol{\beta}}} = \int_{\mathbb{R}^d} \exp\left\{\hat{\boldsymbol{\beta}}'\mathbf{z}\right\} \hat{f}_{\mathbf{Z}}(\mathbf{z})d\mathbf{z} < \infty$ is the estimator of $\nu_{\boldsymbol{\beta}}$. In this case, an estimator of the probability p_i given in (4.4.15) is

$$\hat{p}_i = \frac{\exp\left\{-\hat{\boldsymbol{\beta}}'\mathbf{Z}_i\right\}}{\sum\limits_{i=1}^{n} \exp\left\{-\hat{\boldsymbol{\beta}}'\mathbf{Z}_i\right\}}. \tag{4.4.17}$$

For a closed form of $\hat{\nu}_{\hat{\boldsymbol{\beta}}}$ we use the MGF.

Definition 4.4.2 *Let $\boldsymbol{X} = (X_1, \ldots, X_d)$ be a random vector with multivariate density $f(\boldsymbol{x})$ and $\boldsymbol{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$. The moment generating function of $\boldsymbol{X}$ is defined by*

$$\mathcal{M}_{\boldsymbol{X}}(\boldsymbol{t}) = E[\exp\{\boldsymbol{t}'\boldsymbol{X}\}] = \int_{\mathbb{R}^d} \exp\{\boldsymbol{t}'\boldsymbol{x}\} f(\boldsymbol{x})d\boldsymbol{x}, \tag{4.4.18}$$

for all $\boldsymbol{t}$ which the expectation exists. In particular, if $\boldsymbol{X} \sim \mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ then

$$\mathcal{M}_{\boldsymbol{X}}(\boldsymbol{t}) = \exp\left\{\boldsymbol{t}'\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t}'\boldsymbol{\Sigma}\boldsymbol{t}\right\}. \tag{4.4.19}$$

Now, $\hat{\nu}_{\hat{\boldsymbol{\beta}}}$ can be expressed as

$$\begin{aligned}
\hat{\nu}_{\hat{\boldsymbol{\beta}}} &= \int_{\mathbb{R}^d} \exp\left\{\hat{\boldsymbol{\beta}}'\mathbf{z}\right\} \hat{f}_{\mathbf{Z}}(\mathbf{z})d\mathbf{z} \\
&= \int_{\mathbb{R}^d} \exp\left\{\hat{\boldsymbol{\beta}}'\mathbf{z}\right\} \frac{1}{n}\sum_{i=1}^{n} \mathcal{K}_{\mathbf{H}}(\mathbf{z} - \mathbf{Z}_i^*)d\mathbf{z}
\end{aligned}$$

$$\begin{aligned} &= \frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{R}^d}\exp\left\{\hat{\boldsymbol{\beta}}'\mathbf{z}\right\}\mathcal{K}_{\mathbf{H}}\left(\mathbf{z}-\mathbf{Z}_i^*\right)d\mathbf{z} \\ &= \frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{R}^d}\exp\left\{\hat{\boldsymbol{\beta}}'\mathbf{z}\right\}|\mathbf{H}|^{-1}\mathcal{K}\left(\mathbf{H}^{-1}\left(\mathbf{z}-\mathbf{Z}_i^*\right)\right)d\mathbf{z}\cdot \end{aligned}$$

Letting $\mathbf{z} = \mathbf{H}\mathbf{s} + \mathbf{Z}_i^*$, one has

$$\begin{aligned} \hat{\nu}_{\hat{\boldsymbol{\beta}}} &= \frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{R}^d}\exp\left\{\hat{\boldsymbol{\beta}}'\left(\mathbf{H}\mathbf{s}+\mathbf{Z}_i^*\right)\right\}\mathcal{K}(\mathbf{s})\,d\mathbf{s} \\ &= \frac{1}{n}\sum_{i=1}^{n}\exp\left\{\hat{\boldsymbol{\beta}}'\mathbf{Z}_i^*\right\}\int_{\mathbb{R}^d}\exp\left\{\left(\mathbf{H}'\hat{\boldsymbol{\beta}}\right)'\mathbf{s}\right\}\mathcal{K}(\mathbf{s})\,d\mathbf{s}. \end{aligned}$$

Following Definition 4.4.2, this leads to

$$\hat{\nu}_{\hat{\boldsymbol{\beta}}} = \left(\frac{1}{n}\sum_{i=1}^{n}\exp\left\{\hat{\boldsymbol{\beta}}'\mathbf{Z}_i^*\right\}\right)\mathcal{M}_{\mathbf{S}}\left(\mathbf{H}'\hat{\boldsymbol{\beta}}\right), \tag{4.4.20}$$

where $\mathbf{S}$ is a random vector with multivariate kernel density function $\mathcal{K}(\mathbf{s})$. Hence, using (4.4.14) and (4.4.20) into (4.4.16) an estimator of $f_B(\mathbf{z})$ becomes

$$\hat{f}_B(\mathbf{z}) = \frac{\exp\left\{\hat{\boldsymbol{\beta}}'\mathbf{z}\right\}\sum_{i=1}^{n}\mathcal{K}_{\mathbf{H}}\left(\mathbf{z}-\mathbf{Z}_i^*\right)}{\mathcal{M}_{\mathbf{S}}\left(\mathbf{H}'\hat{\boldsymbol{\beta}}\right)\sum_{i=1}^{n}\exp\left\{\hat{\boldsymbol{\beta}}'\mathbf{Z}_i^*\right\}}\cdot \tag{4.4.21}$$

In particular, if $\mathcal{K}(\mathbf{s})$ is a standard multivariate normal density then by Definition 4.4.2, we have

$$\mathcal{M}_{\mathbf{S}}\left(\mathbf{H}'\hat{\boldsymbol{\beta}}\right) = \exp\left\{\frac{1}{2}\left(\mathbf{H}'\hat{\boldsymbol{\beta}}\right)'\mathbf{H}'\hat{\boldsymbol{\beta}}\right\}\cdot \tag{4.4.22}$$

Following the partition of $\mathbf{Z} = \left(\mathbf{Z}^{(1)}, \mathbf{Z}^{(2)}\right)$, let $\mathbf{H}^{(1)}$ be the $p\times p$ bandwidth matrix with respect to $\mathbf{Z}^{(1)} = (Z_1, \ldots, Z_p)'$ and let $\mathbf{H}^{(2)}$ be the $q\times q$ bandwidth matrix with respect to $\mathbf{Z}^{(2)} = (Z_{p+1}, \ldots, Z_{p+q})'$. Also, suppose that the random vector $\mathbf{S}$ which follows $\mathcal{K}(\mathbf{s})$ is partitioned with respect to $\mathbf{Z}$ as $\mathbf{S} = \left(\mathbf{S}^{(1)}, \mathbf{S}^{(2)}\right)$.

Based on (4.4.21) an estimator of $f_B\left(\mathbf{z}^{(1)}\right)$ is

$$\hat{f}_B\left(\mathbf{z}^{(1)}\right) = \frac{\exp\left\{\hat{\boldsymbol{\beta}}^{(1)\prime}\mathbf{z}^{(1)}\right\}\sum_{i=1}^{n}\mathcal{K}_{\mathbf{H}^{(1)}}\left(\mathbf{z}^{(1)}-\mathbf{Z}_i^{*(1)}\right)}{\mathcal{M}_{\mathbf{S}^{(1)}}\left(\mathbf{H}^{(1)\prime}\hat{\boldsymbol{\beta}}^{(1)}\right)\sum_{i=1}^{n}\exp\left\{\hat{\boldsymbol{\beta}}^{(1)\prime}\mathbf{Z}_i^{*(1)}\right\}}\cdot \tag{4.4.23}$$

If $\mathcal{K}\left(\mathbf{s}^{(1)}\right)$ is a standard multivariate normal density then

$$\mathcal{M}_{\mathbf{S}^{(1)}}\left(\mathbf{H}^{(1)\prime}\hat{\boldsymbol{\beta}}^{(1)}\right) = \exp\left\{\frac{1}{2}\left(\mathbf{H}^{(1)\prime}\hat{\boldsymbol{\beta}}^{(1)}\right)^{\prime}\mathbf{H}^{(1)\prime}\hat{\boldsymbol{\beta}}^{(1)}\right\}. \qquad (4.4.24)$$

To obtain an estimator of $f_B\left(\mathbf{z}^{(1)}|\mathbf{z}^{(2)}\right)$, we use the fact that

$$f_B\left(\mathbf{z}^{(1)}|\mathbf{z}^{(2)}\right) = \frac{f_B\left(\mathbf{z}^{(1)},\mathbf{z}^{(2)}\right)}{f_B\left(\mathbf{z}^{(2)}\right)} = \frac{f_B\left(\mathbf{z}\right)}{f_B\left(\mathbf{z}^{(2)}\right)}, \qquad (4.4.25)$$

and an estimator of $f_B\left(\mathbf{z}^{(1)}|\mathbf{z}^{(2)}\right)$ is

$$\hat{f}_B\left(\mathbf{z}^{(1)}|\mathbf{z}^{(2)}\right) = \frac{\hat{f}_B\left(\mathbf{z}^{(1)},\mathbf{z}^{(2)}\right)}{\hat{f}_B\left(\mathbf{z}^{(2)}\right)} = \frac{\hat{f}_B\left(\mathbf{z}\right)}{\hat{f}_B\left(\mathbf{z}^{(2)}\right)}, \qquad (4.4.26)$$

where $\hat{f}_B\left(\mathbf{z}\right)$ is given by (4.4.21) and $\hat{f}_B\left(\mathbf{z}^{(2)}\right)$ can be obtained by the same way as for (4.4.23). Hence,

$$\hat{f}_B\left(\mathbf{z}^{(1)}|\mathbf{z}^{(2)}\right) = C\exp\left\{\hat{\boldsymbol{\beta}}^{(1)\prime}\mathbf{z}^{(1)}\right\}\frac{\sum_{i=1}^{n}\mathcal{K}_{\mathbf{H}}\left(\mathbf{z}-\mathbf{Z}_i^{*}\right)}{\sum_{i=1}^{n}\mathcal{K}_{\mathbf{H}^{(2)}}\left(\mathbf{z}^{(2)}-\mathbf{Z}_i^{*(2)}\right)}, \qquad (4.4.27)$$

where

$$C = \frac{\mathcal{M}_{\mathbf{S}^{(2)}}\left(\mathbf{H}^{(2)\prime}\hat{\boldsymbol{\beta}}^{(2)}\right)\sum_{i=1}^{n}\exp\left\{\hat{\boldsymbol{\beta}}^{(2)\prime}\mathbf{Z}_i^{*(2)}\right\}}{\mathcal{M}_{\mathbf{S}}\left(\mathbf{H}^{\prime}\hat{\boldsymbol{\beta}}\right)\sum_{i=1}^{n}\exp\left\{\hat{\boldsymbol{\beta}}^{\prime}\mathbf{Z}_i^{*}\right\}}. \qquad (4.4.28)$$

4.4.3 Estimation of the partial information gain and partial dependence measure

A natural estimator of $\Gamma_{\texttt{PA}}$ is based on Proposition 4.4.1 as

$$\hat{\Gamma}_{\texttt{PA}} = \hat{\Gamma}_{\texttt{PA},1} + \hat{\Gamma}_{\texttt{PA},2}, \qquad (4.4.29)$$

where

$$\hat{\Gamma}_{\texttt{PA},1} = 2\left\{\frac{1}{n}\sum_{j=1}^{n}\log\left\{\hat{f}_{LB}\left(U_j|\mathbf{Z}_j\right)\right\} - \frac{1}{n}\sum_{j=1}^{n}\log\left\{\hat{f}_{LB}\left(U_j|\mathbf{Z}_j^{(1)}\right)\right\}\right\}, \qquad (4.4.30)$$

denotes the estimator of the partial information gain $\Gamma_{\mathtt{PA},1}$ given by (4.4.2) and

$$\hat{\Gamma}_{\mathtt{PA},2} = 2\left\{\frac{1}{n}\sum_{j=1}^{n}\log\left\{\hat{f}_B\left(\mathbf{Z}_j^{(1)}|\mathbf{Z}_j^{(2)}\right)\right\} - \frac{1}{n}\sum_{j=1}^{n}\log\left\{\hat{f}_B\left(\mathbf{Z}_j^{(1)}\right)\right\}\right\}, \tag{4.4.31}$$

is the estimator of the partial information gain $\Gamma_{\mathtt{PA},2}$ given by (4.4.3).

An estimator of the partial information gain becomes

$$\begin{aligned}\hat{\Gamma}_{\mathtt{PA}} &= 2\left\{\frac{1}{n}\sum_{j=1}^{n}\log\left\{\hat{f}_{LB}\left(U_j|\mathbf{Z}_j\right)\right\} - \frac{1}{n}\sum_{j=1}^{n}\log\left\{\hat{f}_{LB}\left(U_j|\mathbf{Z}_j^{(1)}\right)\right\}\right. \\ &\quad \left. + \frac{1}{n}\sum_{j=1}^{n}\log\left\{\hat{f}_B\left(\mathbf{Z}_j^{(1)}|\mathbf{Z}_j^{(2)}\right)\right\} - \frac{1}{n}\sum_{j=1}^{n}\log\left\{\hat{f}_B\left(\mathbf{Z}_j^{(1)}\right)\right\}\right\}, \end{aligned} \tag{4.4.32}$$

where $\hat{f}_{LB}\left(U_j|\mathbf{Z}_j\right)$, $\hat{f}_{LB}\left(U_j|\mathbf{Z}_j^{(1)}\right)$, $\hat{f}_B\left(\mathbf{Z}_j^{(1)}|\mathbf{Z}_j^{(2)}\right)$ and $\hat{f}_B\left(\mathbf{Z}_j^{(1)}\right)$, $j = 1,\ldots,n$, can be computed using, respectively, Equations (4.4.10), (**??**), (4.4.26) and (4.4.23).

An estimator of the partial measure of dependence given in (4.3.3) is then

$$\hat{\rho}^2_{\mathtt{PA}}\left(U,\mathbf{Z}^{(1)}|\mathbf{Z}^{(2)}\right) = 1 - \exp\left\{-\hat{\Gamma}_{\mathtt{PA}}\right\}, \tag{4.4.33}$$

where $\hat{\Gamma}_{\mathtt{PA}}$ is given by (4.4.32).

4.5 Consistency of the estimators

In this section, we study the consistency of the estimated partial information gain and partial dependence measure given, respectively, by (4.4.32) and (4.4.33). We start by proving the consistency of the estimators related to the partial information gain in particular, $\hat{f}_{LB}(u|\mathbf{z})$, $\hat{f}_{LB}\left(u|\mathbf{z}^{(1)}\right)$, $\hat{f}_B(\mathbf{z})$ and $\hat{f}_B\left(\mathbf{z}^{(1)}|\mathbf{z}^{(2)}\right)$.

Theorem 4.5.1 *The length-biased density of the survival time U conditional on a fixed vector of d-covariates $\boldsymbol{Z} = \boldsymbol{z}$, $\hat{f}_{LB}(u|\boldsymbol{z})$, is a consistent estimator of $f_{LB}(u|\boldsymbol{z})$.*

Proof: By construction in Algorithm 3.5.1, the random sample $\tilde{\varepsilon}_1,\ldots,\tilde{\varepsilon}_n$ is from f_ε. So that, $\tilde{U}_1,\ldots,\tilde{U}_n$ are also length-biased $i.i.d$ observations from $f_{LB}(u|\mathbf{z})$. Therefore, from [2] we conclude that $\hat{f}_U(u|\mathbf{z}) \xrightarrow{P} f_U(u|\mathbf{z})$ as $n \to \infty$. In addition, we have

$$\hat{f}_{LB}(u|\mathbf{z}) = \frac{u\hat{f}_U(u|\mathbf{z})}{\hat{\mu}(\mathbf{z})},$$

and $\hat{\mu}(\mathbf{z})$ is a particular case of $\hat{\mu}_w$, given by (4.2.3). Moreover, $\hat{\mu}_w$ is a consistent estimator of μ_w, given by (4.2.4). Consequently, $\hat{\mu}(\mathbf{z}) \xrightarrow{P} \mu(\mathbf{z})$ and $\hat{\mu}^{-1}(\mathbf{z}) \xrightarrow{P} \mu^{-1}(\mathbf{z})$ as $n \to \infty$. Hence, $\hat{f}_{LB}(u|\mathbf{z}) \xrightarrow{P} \hat{f}_{LB}(u|\mathbf{z})$ as $n \to \infty$, where we used Slutsky's theorem. This means that $\hat{f}_{LB}(u|\mathbf{z})$ is a consistent estimator of $f_{LB}(u|\mathbf{z})$. Similarly, we can show that $\hat{f}_{LB}\left(u|\mathbf{z}^{(1)}\right)$ is a consistent estimator of $f_{LB}\left(u|\mathbf{z}^{(1)}\right)$. ∎

Assumptions 4.5.2

(a) $\hat{\beta}_j \xrightarrow{a.s.} \beta_j$ for $j = 1, \dots, d$, as $n \to \infty$.

(b) For all $\mathbf{t} \in \mathbb{R}^d$, the MGF $\mathcal{M}(\mathbf{t})$, is finite and continuous.

(c) $\mathbf{Z}_1^*, \dots, \mathbf{Z}_n^*$ is the new sample, following $f_Z(\mathbf{z})$, obtained using the bootstrap techniques, from the original biased sample.

Theorem 4.5.3 *If Assumptions 4.5.2 hold, then as $n \to \infty$*

$$\hat{\nu}_{\hat{\boldsymbol{\beta}}} = \int_{\mathbb{R}^d} \exp\left\{\hat{\boldsymbol{\beta}}'\boldsymbol{z}\right\}\hat{f}_{\boldsymbol{Z}}(\boldsymbol{z})\, d\boldsymbol{z} \xrightarrow{a.s.} \nu_{\boldsymbol{\beta}} = \int_{\mathbb{R}^d} \exp\left\{\boldsymbol{\beta}'\boldsymbol{z}\right\} f_{\boldsymbol{Z}}(\boldsymbol{z})\, d\boldsymbol{z}\cdot$$

Proof: Under Assumption 4.5.2 (c), we have by (4.4.20)

$$\hat{\nu}_{\hat{\boldsymbol{\beta}}} = \left(\frac{1}{n}\sum_{i=1}^{n} \exp\left\{\hat{\boldsymbol{\beta}}'\mathbf{Z}_i^*\right\}\right)\mathcal{M}_{\mathbf{S}}\left(\mathbf{H}'\hat{\boldsymbol{\beta}}\right), \tag{4.5.1}$$

where $\mathbf{S}$ is a random vector with multivariate kernel density function $\mathcal{K}(\mathbf{s})$. From Chen et al. [11], we have that $\hat{\boldsymbol{\beta}} \xrightarrow{a.s.} \boldsymbol{\beta}$ as $n \to \infty$. In addition, $\mathbf{H} \to \mathbf{0}$ as $n \to \infty$. Using Slutsky's theorem and the continuity of the MGF $\mathcal{M}_{\mathbf{S}}$, this leads to

$$\mathcal{M}_{\mathbf{S}}\left(\mathbf{H}'\hat{\boldsymbol{\beta}}\right) \xrightarrow{a.s.} \mathcal{M}_{\mathbf{S}}(0) = 1 \text{ as } n \to \infty\cdot \tag{4.5.2}$$

Now, we show by two different methods that

$$\frac{1}{n}\sum_{i=1}^{n}\exp\left\{\hat{\boldsymbol{\beta}}'\mathbf{Z}_i^*\right\}\xrightarrow{a.s.}\mathrm{E}_{f_{\mathbf{Z}}}\left[\exp\left\{\boldsymbol{\beta}'\mathbf{Z}\right\}\right]\ \text{as } n\to\infty.$$

Approach 1

Let $\varepsilon>0$. Since $\hat{\boldsymbol{\beta}}\xrightarrow{a.s.}\boldsymbol{\beta}$ as $n\to\infty$, there exists δ_1 such that for all $n\geq\delta_1$

$$\boldsymbol{\beta}-\varepsilon\leq\hat{\boldsymbol{\beta}}\leq\boldsymbol{\beta}+\varepsilon,\ \text{a.s.}$$

and we can write $\hat{\boldsymbol{\beta}}=\boldsymbol{\beta}+o\left(\varepsilon\right).$ Now, using the fact that $\nu_{\boldsymbol{\beta}}<\infty$ and $\mathbf{Z}_1^*,\ldots,\mathbf{Z}_n^*$ are $i.i.d$ with PDF $f_{\mathbf{Z}}\left(\mathbf{z}\right)$, by the strong law of large numbers we have

$$\mathcal{M}_n\left(\boldsymbol{\beta}\right)=\frac{1}{n}\sum_{i=1}^{n}\exp\left\{\boldsymbol{\beta}'\mathbf{Z}_i^*\right\}\xrightarrow{a.s.}\mathcal{M}\left(\boldsymbol{\beta}\right)=\mathrm{E}_{f_{\mathbf{Z}}}\left[\exp\left\{\boldsymbol{\beta}'\mathbf{Z}\right\}\right]\ \text{as } n\to\infty. \tag{4.5.3}$$

It follows that there exists δ_2 such that for all $n\geq\delta_2$

$$\mathcal{M}_n\left(\boldsymbol{\beta}\right)=\mathcal{M}\left(\boldsymbol{\beta}\right)+o\left(\varepsilon\right),\ \text{a.s.}$$

Now, $\forall n\geq\delta_2$

$$\mathcal{M}_n\left(\hat{\boldsymbol{\beta}}\right)=\mathcal{M}_n\left(\boldsymbol{\beta}+o\left(\varepsilon\right)\right),\ \text{a.s.}$$

It follows that, for $n\geq\max\left(\delta_1,\delta_2\right)$, we have:

$$\mathcal{M}_n\left(\hat{\boldsymbol{\beta}}\right)=\mathcal{M}\left(\boldsymbol{\beta}\right)+o\left(\varepsilon\right),\ \text{a.s.}$$

which implies that

$$\frac{1}{n}\sum_{i=1}^{n}\exp\left\{\hat{\boldsymbol{\beta}}'\mathbf{Z}_i^*\right\}\xrightarrow{a.s.}\mathrm{E}_{f_{\mathbf{Z}}}\left[\exp\left\{\boldsymbol{\beta}'\mathbf{Z}\right\}\right]=\int_{\mathbb{R}^d}\exp\left\{\boldsymbol{\beta}'\mathbf{z}\right\}f_{\mathbf{Z}}\left(\mathbf{z}\right)d\mathbf{z}\ \text{as } n\to\infty. \tag{4.5.4}$$

Approach 2

For a simple notation, we work in this approach with the random sample $\mathbf{Z}_1,\ldots,\mathbf{Z}_n$ instead of $\mathbf{Z}_1^*,\ldots,\mathbf{Z}_n^*$. To this end, let $\varepsilon>0$. From Assumption 4.5.2 (a), there exists $\boldsymbol{\delta}=\left(\delta_1,\ldots,\delta_d\right)'$ such that for all $n\geq\delta_j$, we have:

$$\beta_j-\varepsilon\leq\hat{\beta}_j\leq\beta_j+\varepsilon,\quad j=1,\ldots,d\quad\text{a.s.}$$

Let $\pi^+_{ij}(\gamma) = \exp\{\gamma Z_{ij}\}\mathbb{1}_{(Z_{ij}>0)}$ and $\pi^-_{ij}(\gamma) = \exp\{\gamma Z_{ij}\}\mathbb{1}_{(Z_{ij}\leq 0)}$. Therefore,

$$\pi^+_{ij}(\beta_j - \varepsilon) \leq \exp\left\{\hat{\beta}_j Z_{ij}\right\}\mathbb{1}_{(Z_{ij}>0)} \leq \pi^+_{ij}(\beta_j + \varepsilon) \quad \text{a.s.}$$

and

$$\pi^-_{ij}(\beta_j + \varepsilon) \leq \exp\left\{\hat{\beta}_j Z_{ij}\right\}\mathbb{1}_{(Z_{ij}\leq 0)} \leq \pi^-_{ij}(\beta_j + \varepsilon) \quad \text{a.s.}$$

So that,

$$\pi^+_{ij}(\beta_j - \varepsilon) + \pi^-_{ij}(\beta_j + \varepsilon) \leq \exp\left\{\hat{\beta}_j Z_{ij}\right\} \leq \pi^+_{ij}(\beta_j + \varepsilon) + \pi^-_{ij}(\beta_j - \varepsilon) \quad \text{a.s.}$$

It follows that,

$$\prod_{j=1}^{d}\left[\pi^+_{ij}(\beta_j - \varepsilon) + \pi^-_{ij}(\beta_j + \varepsilon)\right] \leq \exp\left\{\hat{\boldsymbol{\beta}}'\mathbf{Z}_i\right\} \leq \prod_{j=1}^{d}\left[\pi^+_{ij}(\beta_j + \varepsilon) + \pi^-_{ij}(\beta_j - \varepsilon)\right].$$

From the above inequality, one has

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{d}\left[\pi^+_{ij}(\beta_j - \varepsilon) + \pi^-_{ij}(\beta_j + \varepsilon)\right]\\ &\leq \frac{1}{n}\sum_{i=1}^{n}\exp\left\{\hat{\boldsymbol{\beta}}'\mathbf{Z}_i\right\} \leq \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{d}\left[\pi^+_{ij}(\beta_j + \varepsilon) + \pi^-_{ij}(\beta_j - \varepsilon)\right].\end{aligned} \tag{4.5.5}$$

Since, $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ are $i.i.d$, by the strong law of large numbers, we have as $n \to \infty$

$$\frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{d}\left[\pi^+_{ij}(\beta_j - \varepsilon) + \pi^-_{ij}(\beta_j + \varepsilon)\right]$$

converges almost surely to

$$\mathrm{E}\left[\prod_{j=1}^{d}\left\{\left[\exp\{(\beta_j - \varepsilon)Z_j\}\mathbb{1}_{(Z_j>0)}\right] + \exp\{(\beta_j + \varepsilon)Z_j\}\mathbb{1}_{(Z_j\leq 0)}\right\}\right] = g_1(\boldsymbol{\beta}, \varepsilon),$$

and

$$\frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{d}\left[\pi^+_{ij}(\beta_j + \varepsilon) + \pi^-_{ij}(\beta_j - \varepsilon)\right]$$

converges almost surely to

$$\mathrm{E}\left[\prod_{j=1}^{d}\left\{\left[\exp\{(\beta_j + \varepsilon)Z_j\}\mathbb{1}_{(Z_j>0)}\right] + \exp\{(\beta_j - \varepsilon)Z_j\}\mathbb{1}_{(Z_j\leq 0)}\right\}\right] = g_2(\boldsymbol{\beta}, \varepsilon)\cdot$$

Letting $n \to \infty$ in (4.5.5), we then get for any $\varepsilon > 0$:

$$g_1(\boldsymbol{\beta}, \varepsilon) - \varepsilon \leq \ell^- \leq \ell^+ \leq g_2(\boldsymbol{\beta}, \varepsilon) + \varepsilon \quad \text{a.s.} \tag{4.5.6}$$

where

$$\ell^+ = \limsup_{n\to\infty} \left(\frac{1}{n} \sum_{i=1}^{n} \exp\left\{ \hat{\boldsymbol{\beta}}' \mathbf{Z}_i \right\} \right),$$

and

$$\ell^- = \liminf_{n\to\infty} \left(\frac{1}{n} \sum_{i=1}^{n} \exp\left\{ \hat{\boldsymbol{\beta}}' \mathbf{Z}_i \right\} \right).$$

Now, we apply the dominated convergence theorem to prove that

$$\lim_{\varepsilon\to 0} g_k(\boldsymbol{\beta}, \varepsilon) = \mathrm{E}\left[\exp\left\{\boldsymbol{\beta}'\mathbf{Z}\right\}\right], \quad k = 1, 2.$$

Indeed, one observes

$$\prod_{j=1}^{d} \left\{ \left[\exp\left\{(\beta_j - \varepsilon)Z_j\right\}\mathbb{1}_{(Z_j>0)}\right] + \exp\left\{(\beta_j + \varepsilon)Z_j\right\}\mathbb{1}_{(Z_j\leq 0)} \right\} \leq \exp\left\{\boldsymbol{\beta}'\mathbf{Z}\right\},$$

with $\mathrm{E}[\exp\{\boldsymbol{\beta}'\mathbf{Z}\}] = \mathcal{M}(\boldsymbol{\beta}) < \infty$. So that, the dominated convergence theorem ensures that

$$\lim_{\varepsilon\to 0} \mathrm{E}\left[\prod_{j=1}^{d} \left\{ \left[\exp\left\{(\beta_j - \varepsilon)Z_j\right\}\mathbb{1}_{(Z_j>0)}\right] + \exp\left\{(\beta_j + \varepsilon)Z_j\right\}\mathbb{1}_{(Z_j\leq 0)} \right\} \right] = \mathrm{E}\left[\exp\left\{\boldsymbol{\beta}'\mathbf{Z}\right\}\right],$$

that is, $\lim_{\varepsilon\to 0} g_1(\boldsymbol{\beta}, \varepsilon) = \mathrm{E}\left[\exp\left\{\boldsymbol{\beta}'\mathbf{Z}\right\}\right]$. Now, without loss of generality, one can take, $\varepsilon \leq 1$. Therefore,

$$\begin{aligned}
&\prod_{j=1}^{d} \left\{ \left[\exp\left\{(\beta_j + \varepsilon)Z_j\right\}\mathbb{1}_{(Z_j>0)}\right] + \exp\left\{(\beta_j - \varepsilon)Z_j\right\}\mathbb{1}_{(Z_j\leq 0)} \right\} \\
\leq\ & \exp\left\{\boldsymbol{\beta}'\mathbf{Z}\right\} \prod_{j=1}^{d} \left\{ \exp\left\{Z_j\right\} + \exp\left\{-Z_j\right\} \right\} \\
=\ & \exp\left\{\boldsymbol{\beta}'\mathbf{Z}\right\} \sum_{\boldsymbol{e}\in\mathcal{A}} \exp\left\{\boldsymbol{e}'\mathbf{Z}\right\} \\
=\ & \sum_{\boldsymbol{e}\in\mathcal{A}} \exp\left\{(\boldsymbol{\beta} + \boldsymbol{e})'\mathbf{Z}\right\},
\end{aligned} \tag{4.5.7}$$

where $\mathcal{A}$ denotes the set $\{(\tau_1, \ldots, \tau_d)', \tau_i = -1, 1, i = 1, \ldots, d\}$. Note that the expectation of the random variable given in (4.5.7) is finite, because

$$\mathrm{E}\left\{\sum_{e\in\mathcal{A}} \exp\left\{(\boldsymbol{\beta}+\boldsymbol{e})'\mathbf{Z}\right\}\right\} = \sum_{e\in\mathcal{A}} \mathcal{M}(\boldsymbol{\beta}+\boldsymbol{e}) < \infty.$$

We conclude from the dominated convergence theorem that $\lim_{\varepsilon\to 0} g_2(\boldsymbol{\beta}, \varepsilon) = \mathrm{E}\left[\exp\left\{\boldsymbol{\beta}'\mathbf{Z}\right\}\right]$. Finally, letting $\varepsilon \to 0$ in (4.5.6) then

$$\ell^+ = \ell^- = \ell = \mathrm{E}[\exp\{\boldsymbol{\beta}'\mathbf{Z}\}] \quad \text{a.s.}$$

where

$$\ell = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \exp\left\{\hat{\boldsymbol{\beta}}'\mathbf{Z}_i\right\}.$$

Hence,

$$\frac{1}{n}\sum_{i=1}^{n} \exp\left\{\hat{\boldsymbol{\beta}}'\mathbf{Z}_i\right\} \xrightarrow{a.s.} \mathrm{E}\left[\exp\left\{\boldsymbol{\beta}'\mathbf{Z}\right\}\right] \text{ as } n\to\infty. \tag{4.5.8}$$

From (4.5.1), (4.5.2) and (4.5.8), we conclude by Slutsky's theorem that

$$\int_{\mathbb{R}^d} \exp\left\{\hat{\boldsymbol{\beta}}'\mathbf{z}\right\} \hat{f}_{\mathbf{Z}}(\mathbf{z})\, d\mathbf{z} \xrightarrow{a.s.} \mathrm{E}\left[\exp\left\{\boldsymbol{\beta}'\mathbf{Z}\right\}\right] = \int_{\mathbb{R}^d} \exp\left\{\boldsymbol{\beta}'\mathbf{z}\right\} f_{\mathbf{Z}}(\mathbf{z})\, d\mathbf{z} \text{ as } n\to\infty.$$

Therefore, $\hat{\nu}_{\hat{\boldsymbol{\beta}}}$ is a consistent estimator of $\nu_{\boldsymbol{\beta}}$. ∎

Theorem 4.5.4 *Under the dependence model, $\hat{f}_B(\boldsymbol{z})$ is a consistent estimator of $f_B(\boldsymbol{z})$.*

Proof: From (4.4.16) we have

$$\hat{f}_B(\mathbf{z}) = \frac{\exp\left\{\hat{\boldsymbol{\beta}}'\mathbf{z}\right\} \hat{f}_{\mathbf{Z}}(\mathbf{z})}{\hat{\nu}_{\hat{\boldsymbol{\beta}}}},$$

where

- $\hat{f}_{\mathbf{Z}}(\mathbf{z})$ is the standard multivariate density estimator which it is a consistent estimator of $f_{\mathbf{Z}}(\mathbf{z})$.

- $\hat{\nu}_{\hat{\boldsymbol{\beta}}} = \int_{\mathbb{R}^d} \exp\left(\hat{\boldsymbol{\beta}}'\mathbf{z}\right) \hat{f}_{\mathbf{Z}}\left(\mathbf{z}\right) d\mathbf{z}$ which is a consistent estimator of $\nu_{\boldsymbol{\beta}}$ by Theorem 4.5.3.

Since from Chen et al. [11] we have $\hat{\boldsymbol{\beta}} \xrightarrow{a.s.} \boldsymbol{\beta}$ as $n \to \infty$, $\exp\left\{\hat{\boldsymbol{\beta}}'\mathbf{z}\right\} \xrightarrow{a.s.} \exp\left\{\boldsymbol{\beta}'\mathbf{z}\right\}$ as $n \to \infty$. Applying Slutsky's theorem, one gets

$$\frac{\exp\left\{\hat{\boldsymbol{\beta}}'\mathbf{z}\right\} \hat{f}_{\mathbf{Z}}\left(\mathbf{z}\right)}{\hat{\nu}_{\hat{\boldsymbol{\beta}}}} \quad \xrightarrow{P} \quad \frac{\exp\left\{\boldsymbol{\beta}'\mathbf{z}\right\} f_{\mathbf{Z}}\left(\mathbf{z}\right)}{\nu_{\boldsymbol{\beta}}} \text{ as } n \to \infty.$$

Therefore $\hat{f}_B\left(\mathbf{z}\right) \xrightarrow{P} f_B\left(\mathbf{z}\right)$ as $n \to \infty$. Hence, $\hat{f}_B\left(\mathbf{z}\right)$ is a consistent estimator of $f_B\left(\mathbf{z}\right)$. In particular, $\hat{f}_B\left(\mathbf{z}^{(1)}\right)$ and $\hat{f}_B\left(\mathbf{z}^{(2)}\right)$ are the both consistent estimators of $f_B\left(\mathbf{z}^{(1)}\right)$ and $f_B\left(\mathbf{z}^{(2)}\right)$, respectively. ∎

Corollary 4.5.5 $\hat{f}_B\left(\boldsymbol{z}^{(1)}|\boldsymbol{z}^{(2)}\right)$ *is a consistent estimator of* $f_B\left(\boldsymbol{z}^{(1)}|\boldsymbol{z}^{(2)}\right)$.

Proof: Recall that by (4.4.26), we used

$$\hat{f}_B\left(\mathbf{z}^{(1)}|\mathbf{z}^{(2)}\right) = \frac{\hat{f}_B\left(\mathbf{z}\right)}{\hat{f}_B\left(\mathbf{z}^{(2)}\right)}.$$

as an estimator of $f_B\left(\mathbf{z}^{(1)}|\mathbf{z}^{(2)}\right)$. Moreover, we have showed by Theorem 4.5.4 that

$$\hat{f}_B\left(\mathbf{z}\right) \xrightarrow{P} f_B\left(\mathbf{z}\right) \text{ as } n \to \infty,$$

and in particular,

$$\hat{f}_B\left(\mathbf{z}^{(2)}\right) \xrightarrow{P} f_B\left(\mathbf{z}^{(2)}\right) \text{ as } n \to \infty.$$

Now, provided $\hat{f}_B\left(\mathbf{z}^{(2)}\right)$ and $f_B\left(\mathbf{z}^{(2)}\right)$ are different from zero, one gets

$$\frac{\hat{f}_B\left(\mathbf{z}\right)}{\hat{f}_B\left(\mathbf{z}^{(2)}\right)} \xrightarrow{P} \frac{f_B\left(\mathbf{z}\right)}{f_B\left(\mathbf{z}^{(2)}\right)} \text{ as } n \to \infty,$$

where Slutsky's theorem is used. This leads to,

$$\hat{f}_B\left(\mathbf{z}^{(1)}|\mathbf{z}^{(2)}\right) \xrightarrow{P} f_B\left(\mathbf{z}^{(1)}|\mathbf{z}^{(2)}\right) \text{ as } n \to \infty,$$

and hence $\hat{f}_B\left(\mathbf{z}^{(1)}|\mathbf{z}^{(2)}\right)$ is a consistent estimator of $f_B\left(\mathbf{z}^{(1)}|\mathbf{z}^{(2)}\right)$. ∎

Theorem 4.5.6 *The estimated partial information gain $\hat{\Gamma}_{\mathtt{PA}}$ is a consistent estimator of $\Gamma_{\mathtt{PA}}$.*

Proof: From Equation (4.4.32), we have

$$\hat{\Gamma}_{\mathtt{PA}} = \hat{\Gamma}_{\mathtt{PA},1} + \hat{\Gamma}_{\mathtt{PA},2}, \tag{4.5.9}$$

where $\hat{\Gamma}_{\mathtt{PA},1}$ and $\hat{\Gamma}_{\mathtt{PA},2}$ are given by Equations (4.4.2) and (4.4.3), respectively. We can rearrange $\hat{\Gamma}_{\mathtt{PA},1}$ as follows

$$\begin{aligned}\hat{\Gamma}_{\mathtt{PA},1} = 2 \Bigg\{ & \frac{1}{n}\sum_{j=1}^{n} \log\left\{\hat{f}_{LB}\left(U_j\,|\mathbf{Z}_j\,\right)\right\} - \frac{1}{n}\sum_{j=1}^{n} \log\left\{\hat{f}_{LB}\left(U_j|\mathbf{Z}_j^{(1)}\right)\right\} \\ & + \left(\frac{1}{n}\sum_{j=1}^{n} \log\left\{f_{LB}\left(U_j\,|\mathbf{Z}_j\,\right)\right\} - \frac{1}{n}\sum_{j=1}^{n} \log\left\{f_{LB}\left(U_j\,|\mathbf{Z}_j\,\right)\right\}\right) \\ & - \left(\frac{1}{n}\sum_{j=1}^{n} \log\left\{f_{LB}\left(U_j|\mathbf{Z}_j^{(1)}\right)\right\} - \frac{1}{n}\sum_{j=1}^{n} \log\left\{f_{LB}\left(U_j|\mathbf{Z}_j^{(1)}\right)\right\}\right)\Bigg\}.\end{aligned}$$

It follows that,

$$\begin{aligned}\hat{\Gamma}_{\mathtt{PA},1} = 2 \Bigg\{ & \frac{1}{n}\sum_{j=1}^{n} \log\left\{f_{LB}\left(U_j\,|\mathbf{Z}_j\,\right)\right\} - \frac{1}{n}\sum_{j=1}^{n} \log\left\{f_{LB}\left(U_j|\mathbf{Z}_j^{(1)}\right)\right\} \\ & + \left(\frac{1}{n}\sum_{j=1}^{n} \log\left\{\hat{f}_{LB}\left(U_j\,|\mathbf{Z}_j\,\right)\right\} - \frac{1}{n}\sum_{j=1}^{n} \log\left\{f_{LB}\left(U_j\,|\mathbf{Z}_j\,\right)\right\}\right) \\ & - \left(\frac{1}{n}\sum_{j=1}^{n} \log\left\{\hat{f}_{LB}\left(U_j|\mathbf{Z}_j^{(1)}\right)\right\} - \frac{1}{n}\sum_{j=1}^{n} \log\left\{f_{LB}\left(U_j|\mathbf{Z}_j^{(1)}\right)\right\}\right)\Bigg\},\end{aligned}$$

and

$$\hat{\Gamma}_{\mathtt{PA},1} = 2\left\{V_1 + W_{1,1} - W_{1,0}\right\},$$

where

- $V_1 = \frac{1}{n}\sum\limits_{j=1}^{n} \log\left\{f_{LB}\left(U_j\left|\mathbf{Z}_j\right.\right)\right\} - \frac{1}{n}\sum\limits_{j=1}^{n} \log\left\{f_{LB}\left(U_j|\mathbf{Z}_j^{(1)}\right)\right\}$.
- $W_{1,1} = \frac{1}{n}\sum\limits_{j=1}^{n} \log\left\{\hat{f}_{LB}\left(U_j\left|\mathbf{Z}_j\right.\right)\right\} - \frac{1}{n}\sum\limits_{j=1}^{n} \log\left\{f_{LB}\left(U_j\left|\mathbf{Z}_j\right.\right)\right\}$.
- $W_{1,0} = \frac{1}{n}\sum\limits_{j=1}^{n} \log\left\{\hat{f}_{LB}\left(U_j|\mathbf{Z}_j^{(1)}\right)\right\} - \frac{1}{n}\sum\limits_{j=1}^{n} \log\left\{f_{LB}\left(U_j|\mathbf{Z}_j^{(1)}\right)\right\}$.

By using the weak law of large numbers and Slutsky's theorem, we conclude that

$$V_1 \xrightarrow{P} \mathrm{E}\left[\log\left\{f_{LB}\left(U\left|\mathbf{Z}\right.\right)\right\}\right] - \mathrm{E}\left[\log\left\{f_{LB}\left(U|\mathbf{Z}^{(1)}\right)\right\}\right] \text{ as } n \to \infty.$$

Moreover, by Theorem 4.5.1 we have $W_{1,1} \xrightarrow{P} 0$ and $W_{1,0} \xrightarrow{P} 0$ as $n \to \infty$. So that,

$$\hat{\Gamma}_{\texttt{PA},1} \xrightarrow{P} 2\left\{\mathrm{E}\left[\log\left\{f_{LB}\left(U\left|\mathbf{Z}\right.\right)\right\}\right] - \mathrm{E}\left[\log\left\{f_{LB}\left(U|\mathbf{Z}^{(1)}\right)\right\}\right]\right\} = \Gamma_{\texttt{PA},1} \text{ as } n \to \infty.$$

Rearrange $\hat{\Gamma}_{\texttt{PA},2}$ as follows

$$\begin{aligned}
\hat{\Gamma}_{\texttt{PA},2} = 2\Bigg\{ & \frac{1}{n}\sum_{j=1}^{n} \log\left\{\hat{f}_B\left(\mathbf{Z}_j^{(1)}\left|\mathbf{Z}_j^{(2)}\right.\right)\right\} - \frac{1}{n}\sum_{j=1}^{n} \log\left\{\hat{f}_B\left(\mathbf{Z}_j^{(1)}\right)\right\} \\
& + \left(\frac{1}{n}\sum_{j=1}^{n} \log\left\{f_B\left(\mathbf{Z}_j^{(1)}\left|\mathbf{Z}_j^{(2)}\right.\right)\right\} - \frac{1}{n}\sum_{j=1}^{n} \log\left\{f_B\left(\mathbf{Z}_j^{(1)}\left|\mathbf{Z}_j^{(2)}\right.\right)\right\}\right) \\
& - \left(\frac{1}{n}\sum_{j=1}^{n} \log\left\{f_B\left(\mathbf{Z}_j^{(1)}\right)\right\} - \frac{1}{n}\sum_{j=1}^{n} \log\left\{f_{LB}\left(\mathbf{Z}_j^{(1)}\right)\right\}\right)\Bigg\}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\hat{\Gamma}_{\texttt{PA},2} = 2\Bigg\{ & \frac{1}{n}\sum_{j=1}^{n} \log\left\{f_B\left(\mathbf{Z}_j^{(1)}\left|\mathbf{Z}_j^{(2)}\right.\right)\right\} - \frac{1}{n}\sum_{j=1}^{n} \log\left\{f_B\left(\mathbf{Z}_j^{(1)}\right)\right\} \\
& + \left(\frac{1}{n}\sum_{j=1}^{n} \log\left\{\hat{f}_B\left(\mathbf{Z}_j^{(1)}\left|\mathbf{Z}_j^{(2)}\right.\right)\right\} - \frac{1}{n}\sum_{j=1}^{n} \log\left\{f_B\left(\mathbf{Z}_j^{(1)}\left|\mathbf{Z}_j^{(2)}\right.\right)\right\}\right) \\
& - \left(\frac{1}{n}\sum_{j=1}^{n} \log\left\{\hat{f}_B\left(\mathbf{Z}_j^{(1)}\right)\right\} - \frac{1}{n}\sum_{j=1}^{n} \log\left\{f_B\left(\mathbf{Z}_j^{(1)}\right)\right\}\right)\Bigg\},
\end{aligned}$$

and

$$\hat{\Gamma}_{\texttt{PA},2} = 2\left\{V_2 + W_{2,1} - W_{2,0}\right\},$$

where

- $V_2 = \frac{1}{n}\sum_{j=1}^{n} \log\left\{f_B\left(\mathbf{Z}_j^{(1)}\middle|\mathbf{Z}_j^{(2)}\right)\right\} - \frac{1}{n}\sum_{j=1}^{n}\log\left\{f_B\left(\mathbf{Z}_j^{(1)}\right)\right\}$.
- $W_{2,1} = \frac{1}{n}\sum_{j=1}^{n} \log\left\{\hat{f}_B\left(\mathbf{Z}_j^{(1)}\middle|\mathbf{Z}_j^{(2)}\right)\right\} - \frac{1}{n}\sum_{j=1}^{n}\log\left\{f_B\left(\mathbf{Z}_j^{(1)}\middle|\mathbf{Z}_j^{(2)}\right)\right\}$.
- $W_{2,0} = \frac{1}{n}\sum_{j=1}^{n} \log\left\{\hat{f}_B\left(\mathbf{Z}_j^{(1)}\right)\right\} - \frac{1}{n}\sum_{j=1}^{n}\log\left\{f_B\left(\mathbf{Z}_j^{(1)}\right)\right\}$.

By using the weak law of large numbers and Slutsky's theorem, we conclude that

$$V_2 \xrightarrow{P} \mathrm{E}\left[\log\left\{f_B\left(\mathbf{Z}^{(1)}\middle|\mathbf{Z}^{(2)}\right)\right\}\right] - \mathrm{E}\left[\log\left\{f_B\left(\mathbf{Z}^{(1)}\right)\right\}\right] \text{ as } n\to\infty.$$

In addition, from Theorem 4.5.4 we have $W_{2,1} \xrightarrow{P} 0$ and $W_{2,0} \xrightarrow{P} 0$ as $n\to\infty$. So that,

$$\hat{\Gamma}_{\mathtt{PA},2} \xrightarrow{P} 2\mathrm{E}\left[\log\left\{f_B\left(\mathbf{Z}^{(1)}\middle|\mathbf{Z}^{(2)}\right)\right\}\right] - \mathrm{E}\left[\log\left\{f_B\left(\mathbf{Z}^{(1)}\right)\right\}\right] = \Gamma_{\mathtt{PA},2} \text{ as } n\to\infty.$$

Now, since $\hat{\Gamma}_{\mathtt{PA},1} + \hat{\Gamma}_{\mathtt{PA},2} \xrightarrow{P} \Gamma_{\mathtt{PA},1} + \Gamma_{\mathtt{PA},2}$ as $\to\infty$, this leads to $\hat{\Gamma}_{\mathtt{PA}} \xrightarrow{P} \Gamma_{\mathtt{PA}}$ as $n\to\infty$. Therefore, $\hat{\Gamma}_{\mathtt{PA}}$ is a consistent estimator of $\Gamma_{\mathtt{PA}}$. ∎

Corollary 4.5.7 *The estimated partial dependence measure $\hat{\rho}^2_{\mathtt{PA}}\left(U, \boldsymbol{Z}^{(1)}|\boldsymbol{Z}^{(2)}\right)$ is a consistent estimator of $\rho^2_{\mathtt{PA}}\left(U, \boldsymbol{Z}^{(1)}|\boldsymbol{Z}^{(2)}\right)$.*

Proof: By definition,

$$\hat{\rho}^2_{\mathtt{PA}}\left(U, \mathbf{Z}^{(1)}|\mathbf{Z}^{(2)}\right) = 1 - \exp\left\{-\hat{\Gamma}_{\mathtt{PA}}\right\}.$$

We have proved by Theorem 4.5.6 that, $\hat{\Gamma}_{\mathtt{PA}} \xrightarrow{P} \Gamma_{\mathtt{PA}}$ as $n\to\infty$. Hence

$$1 - \exp\left\{-\hat{\Gamma}_{\mathtt{PA}}\right\} \xrightarrow{P} 1 - \exp\left\{-\Gamma_{\mathtt{PA}}\right\} \text{ as } n\to\infty.$$

This leads to

$$\hat{\rho}^2_{\mathtt{PA}}\left(U, \mathbf{Z}^{(1)}|\mathbf{Z}^{(2)}\right) \xrightarrow{P} \rho^2_{\mathtt{PA}}\left(U, \mathbf{Z}^{(1)}|\mathbf{Z}^{(2)}\right) \text{ as } n\to\infty.$$

Hence, $\hat{\rho}^2_{\mathtt{PA}}\left(U, \mathbf{Z}^{(1)}|\mathbf{Z}^{(2)}\right)$ is a consistent estimator of $\rho^2_{\mathtt{PA}}\left(U, \mathbf{Z}^{(1)}|\mathbf{Z}^{(2)}\right)$. ∎

Chapter 5

Dependence measure for length-biased data using copulas

In terms of dependence measure indicators, the linear correlation coefficient of Bravais-Pearson is considered a powerful indicator when the dependency relationship is linear and the error variate is normally distributed. Unfortunately in finance and in survival analysis the dependency relationship may not be linear. To remedy this, we use other indicators based on the concordance and discordance observed in a sample. We use non-linear correlation coefficients as Kendall's tau or Spearman rho. Our goal in this chapter, is to provide an alternative indicator of dependence measure, based on the concept of information gain, using the parametric copulas. Obtaining conditional and joint measure of dependence between survival time and one continuous covariate in particular, for length-biased data will be considered in this chapter.

5.1 Some general notions of copulas

In this section, we first study the basic definitions and properties of copulas by providing some examples of parametric copulas and then we analyze practical implementation methods through simulations.

5.1.1 Introduction

In several research areas such as finance, medicine and biology, researchers are constantly striving to understand the dependence structure between two or more random variables, relationship described by the joint distribution function. However, determining this joint CDF can be a very tedious task. The concept of copulas is an innovative tool for modeling this dependence structure. Indeed, the knowledge of this concept is essential to understanding many areas of application in particular, survival analysis. Thus, whenever it is necessary to model the dependence structure, we can use the copulas.

5.1.2 Sklar's Theorem

Let X and Y be two random variables with CDF's $F(x)$ and $G(y)$, respectively, and joint CDF $H(x, y)$. We note that this bivariate distribution contains simultaneously the information about marginal distributions and dependence structure. The marginal distributions can be modelled by CDF's F and G. However, the dependence structure between r.v.'s X and Y can be modelled by the so-called copula function, denoted by C. The following famous theorem of Sklar gives a link between the joint distribution H, the CDF's F and G, and copula C.

Theorem 5.1.1 (*Sklar's Theorem [47]*). *Let H be a joint distribution function with marginal distribution F and G. Then, there exists a copula C such that*

$$H(x, y) = C(F(x), G(y)), \quad \forall (x, y) \in \mathbb{R}^2. \tag{5.1.1}$$

If F and G are continuous then C is unique; otherwise, C is uniquely determined on $RanF \times RanG$, where $RanF = F([-\infty, \infty])$ is the range of F. Also, the converse is true, that is if C is a copula and F and G are univariate CDF's then the joint CDF with margins F and G is defined by (5.1.1).

Theorem 5.1.2 *The copula C is simply the distribution corresponding to the random vector (U, V) with uniform margins defined by*

$$U = F(X) \sim \mathcal{U}_{[0,1]} \quad \textit{and} \quad V = G(Y) \sim \mathcal{U}_{[0,1]}.$$

Proof:

$$\begin{aligned}
\mathbb{P}(F(X) \leq u, G(Y) \leq v) &= \mathbb{P}\left(X \leq F^{-1}(u), Y \leq G^{-1}(v)\right) \\
&= H\left(F^{-1}(u), G^{-1}(v)\right) \\
&= C\left(F \circ F^{-1}(u), G \circ G^{-1}(v)\right) \\
&= C(u, v).
\end{aligned}$$

∎

This leads to the next practical definition of copula.

Definition 5.1.3 *A function $C : [0,1]^2 \mapsto [0,1]$ is a copula if and only if there exists random variables $U \sim \mathcal{U}_{[0,1]}$ and $V \sim \mathcal{U}_{[0,1]}$ such that*

$$C(u, v) = \mathbb{P}(U \leq u, V \leq v), \quad \forall u, v \in [0,1],$$

satisfying the following conditions:

(i) $C(u, 0) = C(0, v) = 0.$

(ii) $C(u, 1) = C(1, u) = u.$

(iii) $C(u_2, v_2) - C(u_1, v_2) - C(u_2, v_1) + C(u_1, v_1) \geq 0$ *for all* $u_1 \leq u_2,\ v_1 \leq v_2.$

Note that, Condition (iii) in definition above ensures that if copula C is twice differentiable then C admit a copula density defined by

$$c(u,v) = \frac{\partial^2 C(u,v)}{\partial u \partial v}\cdot \tag{5.1.2}$$

From Theorem 5.1.1, we can see that the copula C is independent of the marginal distributions. In addition, C is considered as the dependence function associated to the random vector (X,Y). In practice, Sklar's Theorem is very interesting because it models F, G and the dependence structure separately. The following examples illustrate some applications of this theorem.

5.1.3 Application examples of Sklar's Theorem

Firstly, we show how to construct a bivariate distribution H from appropriate margins F, G and copula C. On the other hand, we give a method allowing to extract copula function C from a given bivariate distribution H and univariate marginal distributions F and G. In fact, from the representation

$$H(x,y) = C(F(x), G(y)),$$

one can construct bivariate distribution H in two steps:

Step 1: choose appropriate marginal distributions of X and Y.
Step 2: select an appropriate copula function C.

As shown below, we can construct two bivariate distributions with the same copula but with different marginal distributions. For that, consider the following copula which is given in [39]

$$C(u,v) = \frac{uv}{u+v-uv}\cdot \tag{5.1.3}$$

Example 5.1.4 *: (Construction of bivariate distribution). If we take* $F(x) = G(x) = 1 - e^{-x}$*,* $x \geq 0$ *then from (5.1.1), we get the next joint distribution*

$$H(x, y) = C\left(1 - e^{-x}, 1 - e^{-y}\right) = \left(\frac{1}{1 - e^{-x}} + \frac{1}{1 - e^{-y}} - 1\right)^{-1}.$$

Now, if we set $F(x) = G(x) = x^2$*,* $x \in [0, 1]$ *then by using copula defined by (5.1.3), we obtain another distribution expressed by*

$$H(x, y) = C\left(x^2, y^2\right) = \frac{x^2y^2}{x^2 + y^2 - x^2y^2}.$$

These two distributions has the same copula but with different marginal distributions.

When, our interest is to obtain a copula from a given joint distribution, we suggest to use the following approach:

- Letting $u = F(x)$ and $v = G(y)$ or equivalently, $x = F^{-1}(u)$ and $y = G^{-1}(v)$.

- Using the formula $H(x, y) = C(F(x), G(y))$, then for all $u, v \in [0, 1]$, we get

$$C(u, v) = H(F^{-1}(u), G^{-1}(v)).$$

Example 5.1.5 *(Extraction of copula from a given joint distribution). Let* $H(x, y)$ *be a bivariate distribution defined by*

$$H(x, y) = \sqrt{\frac{xy}{x + y - xy}} \quad \textit{for} \quad x, y \in [0, 1]\cdot$$

In fact, the marginal distributions are

$$F(x) = H(x, 1) = \sqrt{x} \quad \textit{and} \quad G(y) = H(1, y) = \sqrt{y}.$$

Consequently, the inverse of these distributions are

$$F^{-1}(u) = u^2 \quad \textit{and} \quad G^{-1}(v) = v^2.$$

It follows that, the corresponding copula of $H(x, y)$ *is*

$$C(u, v) \;=\; H(F^{-1}(u), G^{-1}(v))$$

$$
\begin{aligned}
&= H\left(u^2, v^2\right) \\
&= \sqrt{\frac{u^2v^2}{u^2+v^2-u^2v^2}} \\
&= \left(u^{-2}+v^{-2}-1\right)^{-\frac{1}{2}}.
\end{aligned}
$$

Example 5.1.6 *(Extraction of copula from a given joint distribution). Let $H_\theta(x,y)$ be the joint distribution function of Gumbel's bivariate exponential distribution [26] given by*

$$
H_\theta(x,y) = \begin{cases} 1-e^{-x}-e^{-y}-e^{-(x+y+\theta xy)}, & x,y>0, \\ \\ 0, & otherwise. \end{cases}
$$

where θ is a parameter in $[0,1]$. Clearly, the marginals are exponentially distributed:

$$
F(x) = H(x,\infty) = 1-e^{-x} \quad and \quad G(y) = H(\infty, y) = 1-e^{-y},
$$

with inverses

$$
F^{-1}(u) = -\ln\{1-u\} \quad and \quad G^{-1}(v) = -\ln\{1-v\}, \quad u,v \in [0,1].
$$

Hence the corresponding copula is

$$
C_\theta(u,v) = H_\theta(F^{-1}(u), G^{-1}(v)) = H_\theta\left(-\ln\{1-u\}, -\ln\{1-v\}\right).
$$

That is,

$$
C_\theta(u,v) = u+v-1+(1-u)(1-v)e^{-\theta\ln\{1-u\}\ln\{1-v\}}.
$$

Noting that, the parameter $\theta \in [0,1]$ of the copula C_θ can be viewed as a dependence parameter.

5.1.4 Some fundamentals properties of copulas

Here, we list some interesting properties of copulas discussed in Nelson [39].

Property 5.1.7 *Let X and Y be two random variables with continuous marginal distribution F and G, respectively, and joint distribution H. If X and Y are independent then the corresponding copula is*

$$\Pi(u, v) = uv. \tag{5.1.4}$$

Theorem 5.1.8 *For any copula C, one has for all $u, v \in [0, 1]$:*

$$W(u, v) \leq C(u, v) \leq M(u, v), \tag{5.1.5}$$

where

- $W(u, v) = \max(u + v - 1, 0)$ *is called Fréchet's lower bound. It describes the perfect negative dependence.*
- $M(u, v) = \min(u, v)$ *is Fréchet's upper bound. It corresponds to the perfect positive dependence.*

An important property of copulas comes from the fact that for strictly monotone transformations of the random variables, copulas are invariant. The following property is very useful in the study of nonparametric statistics.

Theorem 5.1.9 *Let X and Y be continuous random variables with copula C_{XY}. If f and g are strictly increasing transformations on Ran$X$ and Ran$Y$, respectively, then the random vectors (X, Y) and $(f(X), g(Y))$ have the same copula*

$$C_{XY}(u, v) = C_{f(X)g(Y)}(u, v). \tag{5.1.6}$$

5.1.5 Survival copulas

In many applications in medicine and reliability, the random variables of interest represent the lifetimes of individuals or devices in some population. For that, let X be the lifetime of an individual. The probability of an individual (with lifetime X) living or surviving beyond time x is given by the survival function $S(x) = \mathbb{P}(X > x)$. In this direction, the joint survival function corresponding to a pair of lifetime (X, Y) is given by

$$S(x, y) = \mathbb{P}(X > x, Y > y). \tag{5.1.7}$$

We are now in position to define the survival copula. To this end, let H be the distribution function of the random pair (X, Y). Using the well known relation between S and H:

$$\begin{aligned} S(x, y) &= \mathbb{P}(X > x, Y > y) \\ &= 1 - \mathbb{P}(X \leq x \,\text{or}\, Y \leq y) \\ &= 1 - \mathbb{P}(X \leq x) - \mathbb{P}(Y \leq y) + \mathbb{P}(X \leq x, Y \leq y) \\ &= S(x) + S(y) - 1 + C(F(x), G(y)) \\ &= S(x) + S(y) - 1 + C(1 - S(x), 1 - S(y)). \end{aligned}$$

Define a function $\hat{C}$ from $[0, 1] \times [0, 1]$ into $[0, 1]$ by

$$\hat{C}(u, v) = u + v - 1 + C(1 - u, 1 - v). \tag{5.1.8}$$

Then, one has

$$S(x, y) = \hat{C}(S(x), S(y)). \tag{5.1.9}$$

The copula $\hat{C}$ obtained in this way is called the survival copula of X and Y [39].

5.1.6 Usual copulas families

Gaussian copula

Recall that, the density and distribution function of the standard normal distribution $\mathcal{N}(0,1)$ are, respectively,

$$\phi_1(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \quad \text{and} \quad \Phi_1(x) = \int_{-\infty}^{x} \phi(t)dt.$$

The density and distribution function of the bivariate standard normal distribution $\mathcal{N}_2(0, I_\rho)$, with correlation parameter $\rho \in (-1,1)$, are

$$\phi_2(x,y,\rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right),$$

$$\Phi_2(x,y,\rho) = \int_{-\infty}^{x} \int_{-\infty}^{y} \phi_2(t,s,\rho)dtds.$$

The bivariate normal copula with parameter ρ is then defined by application of Sklar's Theorem

$$C_\rho(u,v) = \Phi_2\left(\Phi_1^{-1}(u), \Phi_1^{-1}(v), \rho\right),$$

or equivalently

$$C_\rho(u,v) = \int_{-\infty}^{\Phi_1^{-1}(u)} \int_{-\infty}^{\Phi_1^{-1}(v)} \phi_2(t,s,\rho)dtds.$$

We note that, we can construct a model where the dependence is normal, but the margins are not necessary normally distributed. This can be done by using normal copula. For example, let F and G be any continuous distribution functions then

$$H(x,y) = C_\rho(F(x), G(y)),$$

is a bivariate distribution with marginal distributions F and G and normal dependence described by the normal copula $C_\rho(F(x), G(y), \rho)$.

Student copula

Student copula is extracted in the same way as the Gaussian copula using bivariate student distribution such that

$$C_{\nu,\rho}(u,v) = \frac{1}{2\pi\nu\sqrt{1-\rho^2}} \int_{-\infty}^{T_{1,\nu}^{-1}(u)} \int_{-\infty}^{T_{1,\nu}^{-1}(v)} T_{2,\nu}(t,s,\rho)dtds,$$

where $T_{1,\nu}$ and $T_{2,\nu}$ are the univariate and the bivariate student density, respectively, ν is the number of degrees of freedom and $\rho \in (-1,1)$ is the correlation coefficient.

Archimedean copulas

Archimedean copulas defined by Genest and Mackay [19] is an important class of copulas with several applications in practice. The main reason is the ease with which they can be constructed. In addition, there is a great variety of families of copulas which belong to this class. Finally, this class of copulas possesses many nice properties.

Definition 5.1.10 *The Archimedean copulas are expressed as*

$$C_\phi(u,v) = \phi^{-1}\{\phi(u)+\phi(v)\}, \quad u,v \in [0,1], \tag{5.1.10}$$

where ϕ denotes a continuous, strictly decreasing convex function defined from $[0,1]$ to $[0,\infty[$ such that $\phi(1)=0$. The function ϕ^{-1} represents the inverse of ϕ. The mapping ϕ is so-called the generator of the copula C_ϕ.

Example 5.1.11 *: (Independence copula). The independence copula $\Pi(u,v)=uv$ is an Archimedean copula with generator $\phi(t) = -\ln\{t\}$. In fact, $\phi^{-1}(t) = \exp\{-t\}$ and from Definition 5.1.10, we have for all $u,v \in [0,1]$*

$$\begin{aligned} C_\phi(u,v) &= \phi^{-1}\{\phi(u)+\phi(v)\} \\ &= \exp\{-(-\ln\{u\})+(-\ln\{v\})\} \\ &= uv. \end{aligned}$$

Example 5.1.12 : *(Clayton copula). The family of Clayton copulas are expressed by*

$$C_\theta(u,v) = \left(u^{-\theta} + v^{-\theta} - 1\right)^{-\frac{1}{\theta}} \quad \textit{for} \quad \theta \in [-1, \infty[\setminus \{0\}\,, \tag{5.1.11}$$

where the generator of this family is given by

$$\phi_\theta(t) = \frac{t^{-\theta} - 1}{\theta}\cdot \tag{5.1.12}$$

Some interesting properties of Clayton copula are:

- *The independence copula* Π *is reached when* θ *goes to* 0*, that is:*

$$\lim_{\theta \to 0} C_\theta(u,v) = uv.$$

- *The lower bound copula* W*, perfect negative dependence copula, is reached when* $\theta = -1$*, that is:*

$$C_{-1}(u,v) = W(u,v).$$

- *The upper bound copula* M*, perfect positive dependence copula, is achieved when* θ *goes to infinity, that is:*

$$\lim_{\theta \to \infty} C_\theta(u,v) = M(u,v).$$

Example 5.1.13 : *(Frank's copula). The analytic expression of Frank's copula is*

$$C_\theta(u,v) = -\frac{1}{\theta} \ln\left\{1 - \frac{(1-e^{-\theta u})(1-e^{-\theta v})}{1-e^{-\theta}}\right\} \quad \textit{for} \quad \theta \in \mathbb{R}\setminus \{0\}\,, \tag{5.1.13}$$

where the generator of this family is

$$\phi_\theta(t) = -\ln\left\{\frac{1-e^{-\theta}}{1-e^{-\theta t}}\right\}. \tag{5.1.14}$$

Frank's copula has the following properties:

- *The dependence copula is reached when* θ *converges to* 0*:* $\lim\limits_{\theta \to 0} C_\theta(u,v) = uv.$

- *The lower and upper bounds W and M, respectively, are achieved. Because,* $\lim_{\theta \to -\infty} C_\theta(u,v) = W(u,v)$ *and* $\lim_{\theta \to \infty} C_\theta(u,v) = M(u,v)$.

Example 5.1.14 : (*Gumbel's copula*). *Gumbel's copula is formulated as*

$$C_\theta(u,v) = \exp\left\{ -\left((-\log\{u\})^\theta + (-\log\{v\})^\theta \right)^{\frac{1}{\theta}} \right\} \quad \text{for} \quad \theta \in [1, \infty[, \tag{5.1.15}$$

with generator

$$\phi_\theta(t) = (-\log\{t\})^\theta . \tag{5.1.16}$$

Gumbel's copula has very intersecting properties:

- *The independence copula Π is achieved when $\theta = 1$.*
- *The perfect positive dependence copula M is obtained when θ goes to infinity.*
- *Gumbel's copula provides only the positive dependence since*

$$\Pi(u,v) \leq C_\theta(u,v) \leq M(u,v).$$

5.1.7 Simulation of copulas

Simulation of copulas is very useful in practice. In particular, simulation plays an essential role to establish goodness-of-fit (GOF) for copula. In this section, we show how to simulate data from a given copula. From [39], the theorem below provides an algorithm to generate data from a given copula.

Theorem 5.1.15 *Let (U,V) be uniform random vector with copula C. Define for every $u, v \in [0,1]$:*

$$L_v(u) = \mathbb{P}\left[U \leq u | V = v\right] = \frac{\partial}{\partial v} C(u,v).$$

Let T_1 and T_2 be two independent r.v.'s such that $T_1 \sim \mathcal{U}_{[0,1]}$ and $T_2 \sim \mathcal{U}_{[0,1]}$. Put $X = L_{T_2}^{-1}(T_1)$ and $Y = T_2$. Then, the copula of the random pair (X,Y) is C.

Proof: We have,

$$\begin{aligned}
\mathbb{P}(X \leq u, Y \leq v) &= \mathbb{P}(L_{T_2}^{-1}(T_1) \leq u, T_2 \leq v) \\
&= \int_0^1 \mathbb{P}(L_{T_2}^{-1}(T_1) \leq u, T_2 \leq v | T_2 = s) ds \\
&= \int_0^v \mathbb{P}(L_s^{-1}(T_1) \leq u) ds \\
&= \int_0^v L_s(u) ds \\
&= \int_0^v \frac{\partial}{\partial s} C(u, s) ds \\
&= C(u, v).
\end{aligned}$$

Hence, the copula of the random pair (X, Y), where $X = L_{T_2}^{-1}(T_1)$ and $Y = T_2$, is exactly C. ∎

Based on Thorem 5.1.15 a simulation algorithm is

- Generate two independent variates u, t from $\mathcal{U}_{[0,1]}$.
- Set $v = L_u^{-1}(t)$.
- The desired pair is then (u, v).

Let us now apply this algorithm to simulate an Archimedean copula. To this end, let C be an Archimedean copula with generator ϕ,

$$C(u, v) = \phi^{-1}\{\phi(u) + \phi(v)\}.$$

Firstly, we need to compute $L_v(u)$ in terms of the generator ϕ. After standard computation, we get

$$L_v(u) = \frac{\partial}{\partial v} C(u, v) = \frac{\phi'(v)}{\phi'[\phi^{-1}\{\phi(u) + \phi(v)\}]}.$$

Taking the inverse of $L_v(u)$, we obtain

$$L_v^{-1}(t) = \phi^{-1}\left[\phi\left\{(\phi')^{-1}\left(\frac{\phi'(v)}{t}\right)\right\} - \phi(v)\right].$$

It follows that, an Archimedean copulas simulation algorithm is

- Generate two independent variables u, t from $\mathcal{U}_{[0,1]}$.
- Set
$$v = L_u^{-1}(t) = \phi^{-1}\left[\phi\left\{(\phi')^{-1}\left(\frac{\phi'(u)}{t}\right)\right\} - \phi(u)\right].$$
- The desired pair is then (u, v).

Example 5.1.16 *: Algorithm and simulation of Clayton's copula*
Recall that the generator of Clayton copula is

$$\phi(t) = \frac{t^{-\theta} - 1}{\theta} \quad \textit{with inverse} \quad \phi^{-1}(s) = (\theta s + 1)^{-1/\theta}.$$

Hence,
$$\phi'(t) = -t^{-\theta-1} \quad \textit{and} \quad (\phi')^{-1}(s) = (-s)^{-1/(\theta+1)}.$$

Consequently,
$$L_u^{-1}(t) = \left(u^{-\theta} t^{-\theta/(\theta+1)} - u^{-\theta} + 1\right)^{-1/\theta}.$$

This leads to the next algorithm allowing to generate data from Clayton copula

- *Generate independent variables u, t from $\mathcal{U}_{[0,1]}$.*
- *Set*
$$v = \left(u^{-\theta} t^{-\theta/(\theta+1)} - u^{-\theta} + 1\right)^{-1/\theta}.$$
- *The desired pair is then (u, v).*

The next figure shows how the parameter θ, of Clayton copula, illustrates the dependence between the random variables U and V.

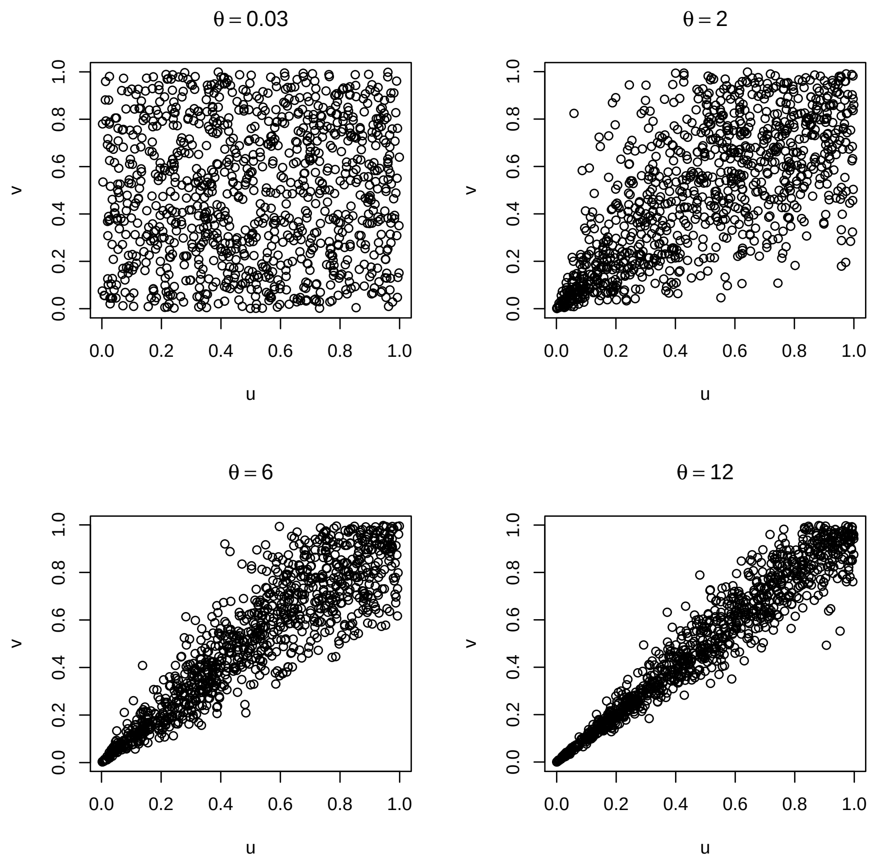

Figure 5.1: Simulation of (U_i, V_i), $i = 1, \dots, 1000$ from Clayton copula with different values of θ.

5.1.8 Goodness-of-fit procedures for copula

Consider a continuous random vector (X, Y) with margins F, G and bivariate CDF H. Assume that the copula C of (X, Y) belongs to a class of parametric copula $C_0 = \{C_\theta, \theta \in \Theta\}$, where Θ is the parameter space. Let (X_i, Y_i), $i = 1, \dots, n$ denote independent copies of (X, Y). Suppose one wants to choose between the null and alternative hypotheses of belonging or not to a given parametric family, namely

$$H_0 : C \in C_0 \quad \text{versus} \quad H_1 : C \notin C_0. \tag{5.1.17}$$

Many goodness-of-fit procedures to confront H_0 and H_1 have been developed recently, e.g. Genest and Rivest [20], Shih [45], Breymann et al. [10], Fermanian [16], Genest et al. [22], Scaillet [44], Mesfioui et al. [38] and Genest et al. [24]. The study of some GOF tests for copula and their implementation using the copula package leads to describe two methods that are very useful in survival analysis. The first approach is based on the empirical copula which gave the best results overall, as mentioned by Genest et al. [24] and later, Berg [8] confirmed this remark resulting from examination and comparison of several GOF tests. The second approach is based on the Rosenblatt transformation. The two formal GOF tests that we examine in this section are rank-based. In other words, instead of using the observations (X_i, Y_i), $i = 1, \dots, n$ one uses the pseudo-observations $\mathbf{U}_i = (U_{i1}, U_{i2}) = (R_i/(n+1), S_i/(n+1))$, $i = 1, \dots, n$, where $R_i = nF_n(X_i)$ is the rank of X_i among $X_1, \dots, X_n$ and $S_i = nG_n(Y_i)$ is the rank of Y_i among $Y_1, \dots, Y_n$. Here, F_n and G_n denote ECDF of X and Y, respectively. Note that, the pseudo-observations can be expressed as

$$\mathbf{U}_i = \left(\frac{n}{n+1} F_n(X_i), \frac{n}{n+1} G_n(Y_i) \right) \quad \text{for} \quad i = 1, \dots, n, \tag{5.1.18}$$

and considered as a sample from the copula C. In addition, they are not mutually independent and that their components are only approximately uniform on $(0, 1)$. We note that, the factor $n/(n+1)$ in (5.1.18) is introduced to avoid problems with C_θ blowing up at the boundary $[0, 1]^2$. The idea behind using the pseudo-observations is that the copula C of a random vector is invariant by continuous, strictly increasing transformations of its components.

Testing (5.1.17) involves the estimation of the dependence parameter θ by some consistent estimator $\hat{\theta}$. When the copula C admits a density c_θ, an estimator of θ can be found by maximizing the log-likelihood

$$\ell(\theta) = \sum_{i=1}^{n} \log \left[c_\theta \left\{ F_{\hat{\boldsymbol{\lambda}}}(X_i), G_{\hat{\boldsymbol{\psi}}}(Y_i) \right\} \right], \tag{5.1.19}$$

where $F_{\hat{\boldsymbol{\lambda}}}$ and $G_{\hat{\boldsymbol{\psi}}}$ are first obtained and represent the parametric estimates of the margins F and G respectively. This approach is due to Joe [29]. Another way for estimating θ is to use the semi-parametric method, which considers the ECDF's F_n and G_n. Then, an estimator of θ is the value of $\hat{\theta}$ that maximizes the log pseudo-likelihood

$$\ell\left(\theta\right)=\sum_{i=1}^{n}\log\left[c_{\theta}\left\{F_n(X_i),G_n(Y_i)\right\}\right]. \tag{5.1.20}$$

The asymptotic normality of $\hat{\theta}$ was established by Genest et al. [21].

Approach $\mathcal{A}_1$: (Test based on the empirical copula)

For testing $H_0 : C \in C_0$, Genest et al. [24] used the pseudo-observations $\mathbf{U}_1,\ldots,\mathbf{U}_n$ and proposed to work with a consistent estimation of an unknown copula C. Particulary, the empirical copula

$$C_n\left(\mathbf{u}\right)=\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\left(U_{i1}\le u_1,U_{i2}\le u_2\right),\quad \mathbf{u}=(u_1,u_2)\in[0,1]^2. \tag{5.1.21}$$

Fermanian et al. [17] showed under various conditions that C_n is a consistent estimator of the true underlying copula C. The idea in this approach is to compare C_n with an estimator of C under $H_0 : C \in C_0$. In a goodness-of-fit setting, Genest et al. [24] suggested to use the empirical process

$$\mathbb{C}_n=\sqrt{n}\left(C_n-C_{\hat{\theta}_n}\right), \tag{5.1.22}$$

where $\hat{\theta}_n=\Upsilon_n\left(\mathbf{U}_1,...,\mathbf{U}_n\right)$ is an estimator of θ. From [24] a Cramér-von Mises statistic for approach $\mathcal{A}_1$ is

$$S_n^{(E)}=\int_{[0,1]^2}\mathbb{C}_n\left(\mathbf{u}\right)^2 d\mathbb{C}_n\left(\mathbf{u}\right)=\sum_{i=1}^{n}\left\{C_n\left(\mathbf{U}_i\right)-C_{\hat{\theta}_n}\left(\mathbf{U}_i\right)\right\}^2. \tag{5.1.23}$$

An approximate P-value can be deduced from the limiting distribution of (5.1.23) which depends on the asymptotic behavior of the empirical process given in (5.1.22). Under appropriate regularity conditions on the assumed parametric family C_0 and the sequence $\hat{\theta}_n$ of estimators, Genest and Rémillard [23] established the convergence of (5.1.22), and showed that the test based on $S_n^{(E)}$ is consistent. In practice, the limiting distribution of the statistic given in (5.1.23) depends on C_0 and on the unknown parameter θ. A specific parametric bootstrap procedure, developed in [24], can be used to approximate the P-value for this statistic. The validity of this method is established by Genest and Rémillard [23].

Approach $\mathcal{A}_2$: (Test based on the Rosenblatt's transform)

Genest et al. [24] proposed to apply approach $\mathcal{A}_1$ to the new vector $\mathbf{V} = \mathcal{R}(\mathbf{U})$, where $\mathcal{R}$ is the Rosenblatt transformation. This mapping is frequently used for simulation and represents a simple way to transform a set of dependent variables with a given distribution into a new set of independent U(0,1) variables. The concept was first introduced by Rosenblatt [42]. Its standard definition [24] is recalled below.

Definition 5.1.17 *Rosenblatt's probability transform of a copula C is the mapping $\mathcal{R} : (0,1)^d \to (0,1)^d$ which to every $\boldsymbol{u} = (u_1, \ldots, u_d) \in (0,1)^d$ assigns another vector $\mathcal{R}(\boldsymbol{u}) = (v_1, \ldots, v_d)$ with $v_1 = u_1$ and*

$$v_i = \frac{\partial^{i-1} C\left(u_1, \ldots, u_i, 1, \ldots, 1\right)}{\partial u_1 \cdots \partial u_{i-1}} \Big/ \frac{\partial^{i-1} C\left(u_1, \ldots, u_{i-1}, 1, \ldots, 1\right)}{\partial u_1 \cdots \partial u_{i-1}} \quad \text{for} \quad i = 1, \ldots, d.$$

It follows that, the hypotheses $H_0 : \mathbf{U} \sim C \in C_0$ and $H_0^* : \mathcal{R}_\theta(\mathbf{U}) \sim C_\perp$ are equivalent for some $\theta \in \Theta$, where $C_\perp$ denotes the independent copula. To test this hypothesis, Genest et al. [24] suggested to use the pseudo-observations

$$\mathbf{V}_i = \mathcal{R}_{\hat{\theta}_n}\left(\mathbf{U}_i\right), \quad i = 1, \ldots, n, \tag{5.1.24}$$

which can be interpreted as a sample from $C_\perp$. So that under the null hypothesis, the empirical distribution associated with $\mathbf{V}_1, \ldots, \mathbf{V}_n$ is given by (5.1.21) as

$$D_n\left(\mathbf{u}\right) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\left(\mathbf{V}_i \leq \mathbf{u}\right), \quad \mathbf{u} \in [0,1]^2. \tag{5.1.25}$$

The idea here, is to compare D_n with the independence copula $C_\perp$. Genest et al. [24] proposed to use the following empirical process

$$\mathbb{D}_n = \sqrt{n}\left(D_n - C_\perp\right). \tag{5.1.26}$$

From [24] a Cramér-von Mises statistic for approach $\mathcal{A}_2$ is

$$S_n^{(R)} = \int_{[0,1]^2} \mathbb{D}_n\left(\mathbf{u}\right)^2 d\mathbb{D}_n\left(\mathbf{u}\right) = \sum_{i=1}^{n} \left\{D_n\left(\mathbf{V}_i\right) - C_\perp\left(\mathbf{V}_i\right)\right\}^2. \tag{5.1.27}$$

Genest et al. [24] mentioned that the asymptotic null behavior of (5.1.26) can be easily determined using the tools given in [25] and this leads to the convergence of $S_n^{(R)}$. As in the previous approach, the asymptotic distribution of $S_n^{(R)}$ depends both on the unknown copula C_θ and θ. The approximate P-value for this statistic can be found via the parametric bootstrap procedure given in [24].

5.2 Information gain and dependence measure using parametric copulas method

In this section, we exploit the general notions of copulas, given in the previous section, to develop an alternative dependence measure based on the concept of information gain using the second method: parametric copulas.

5.2.1 Introduction

Let X be r.v. with PDF $f_X(x;\boldsymbol{\lambda})$ and CDF $F_X(x;\boldsymbol{\lambda})$. Let Y be another r.v. with PDF $f_Y(y;\boldsymbol{\psi})$ and CDF $F_Y(y;\boldsymbol{\psi})$. Suppose that the random vector (X, Y) is associated with some parametric copula C_α, where α denotes dependence parameter. Using Sklar's Theorem, the joint CDF of (X, Y) can be written in terms of the copula C_α as follows

$$F\left(x, y;\boldsymbol{\theta}\right) = C_\alpha\left(F_X(x;\boldsymbol{\lambda}), F_Y(y;\boldsymbol{\psi})\right), \tag{5.2.1}$$

where $\boldsymbol{\theta} = (\alpha, \boldsymbol{\lambda}, \boldsymbol{\psi})$ denotes the parameter of the model. It follows that, the joint density function of (X, Y) takes the following form

$$f(x, y; \boldsymbol{\theta}) = c_\alpha \left(F_X(x; \boldsymbol{\lambda}), F_Y(y; \boldsymbol{\psi})\right) f_X(x; \boldsymbol{\lambda}) f_Y(y; \boldsymbol{\psi}), \tag{5.2.2}$$

where c_α is the copula density given by (5.1.2). Now, since

$$f(x, y; \boldsymbol{\theta}) = f(x|y; \boldsymbol{\theta}) f_Y(y; \boldsymbol{\psi}) = c_\alpha \left(F_X(x; \boldsymbol{\lambda}), F_Y(y; \boldsymbol{\psi})\right) f_X(x; \boldsymbol{\lambda}) f_Y(y; \boldsymbol{\psi}), \tag{5.2.3}$$

the density of X conditional on $Y = y$, in terms of the parametric copula density, is

$$f(x|y; \boldsymbol{\theta}) = c_\alpha \left(F_X(x; \boldsymbol{\lambda}), F_Y(y; \boldsymbol{\psi})\right) f(x; \boldsymbol{\lambda}) \cdot \tag{5.2.4}$$

For the following sections $\boldsymbol{\theta}_0 = (\alpha_0, \boldsymbol{\lambda}_0, \boldsymbol{\psi}_0)$ and $\boldsymbol{\theta}_1 = (\alpha_1, \boldsymbol{\lambda}_1, \boldsymbol{\psi}_1)$ denote parameters of the model under independence and dependence models, respectively. It should be noted that, for some independence parameter $\alpha = \alpha_0$ of copula C_α we have

- $C_{\alpha_0} \left(F_X(x; \boldsymbol{\lambda}_0), F_Y(y; \boldsymbol{\psi}_0)\right) = F_X(x; \boldsymbol{\lambda}_0) F_Y(y; \boldsymbol{\psi}_0)$.
- $c_{\alpha_0} \left(F_X(x; \boldsymbol{\lambda}_0), F_Y(y; \boldsymbol{\psi}_0)\right) = 1$.
- $f(x, y; \boldsymbol{\theta}_0) = f_X(x; \boldsymbol{\lambda}_0) f_Y(y; \boldsymbol{\psi}_0)$.
- $f(x|y; \boldsymbol{\theta}_0) = f_X(x; \boldsymbol{\lambda}_0)$.

5.2.2 Conditional information gain

Proposition 5.2.1 *Let (X, Y) be a pair of random variables possibly dependent with true density $f(x, y; \boldsymbol{\theta}_1)$ given in (5.2.2). The conditional information gain, based on the parametric copula density, is*

$$\begin{aligned} \Gamma_C &= 2\left\{\iint \log\left\{c_{\alpha_1}\left(F_X(x; \boldsymbol{\lambda}_1), F_Y(y; \boldsymbol{\psi}_1)\right) f_X(x; \boldsymbol{\lambda}_1)\right\} f(x, y; \boldsymbol{\theta}_1) dx dy \right. \\ &\quad \left. - \int \log\left\{f_X(x; \boldsymbol{\lambda}_0)\right\} f_X(x; \boldsymbol{\lambda}_1) dx \right\}. \end{aligned} \tag{5.2.5}$$

Proof: By testing the two hypotheses $H_0 : \alpha = \alpha_0$ versus $H_1 : \alpha \neq \alpha_0$, the twice Kullback-Leibler [35] information gain is

$$\begin{aligned}
\Gamma_C &= 2\left\{\iint \log\left\{f\left(x|y;\boldsymbol{\theta}_1\right)\right\} f\left(x,y;\boldsymbol{\theta}_1\right)dxdy \right.\\
&\quad \left. -\iint \log\left\{f\left(x|y;\boldsymbol{\theta}_0\right)\right\} f\left(x,y;\boldsymbol{\theta}_1\right)dxdy\right\}\\
&= 2\left\{\iint \log\left\{c_{\alpha_1}\left(F_X(x;\boldsymbol{\lambda}_1),F_Y(y;\boldsymbol{\psi}_1)\right)f_X(x;\boldsymbol{\lambda}_1)\right\} f\left(x,y;\boldsymbol{\theta}_1\right)dxdy \right.\\
&\quad \left. -\int \log\left\{f_X\left(x;\boldsymbol{\lambda}_0\right)\right\} f_X\left(x;\boldsymbol{\lambda}_1\right)dx\right\},
\end{aligned}$$

where $f\left(x|y;\boldsymbol{\theta}_1\right)$ is given by (5.2.4) and we used the fact that under the independence model: $f\left(x|y;\boldsymbol{\theta}_0\right) = f\left(x;\boldsymbol{\theta}_0\right) = f_X\left(x;\boldsymbol{\lambda}_0\right)$. ∎

Consequently, from Proposition 5.2.1, the conditional dependence measure with respect to the work of Kent [33] is

$$\rho_C^2\left(X|Y\right) = 1 - \exp\left\{-\Gamma_C\right\}\cdot \tag{5.2.6}$$

5.2.3 Estimation of the conditional information gain and conditional measure of dependence

Let (X_i, Y_i), $i = 1, \ldots, n$ be a random sample from the true joint density $f\left(x,y;\boldsymbol{\theta}_1\right)$ given in (5.2.2). Based on Proposition 5.2.1, the conditional information gain can be written, in terms of expectation, as

$$\Gamma_C = 2\left\{\mathrm{E}\left[\log\left\{c_{\alpha_1}\left(F_X(X;\boldsymbol{\lambda}_1),F_Y(Y;\boldsymbol{\psi}_1)\right)f_X(X;\boldsymbol{\lambda}_1)\right\}\right] - \mathrm{E}\left[\log\left(f_X\left(X;\boldsymbol{\lambda}_0\right)\right)\right]\right\}\cdot \tag{5.2.7}$$

It follows that, an estimator of Γ_C is

$$\begin{aligned}
\hat{\Gamma}_C &= \frac{2}{n}\left\{\sum_{i=1}^{n}\log\left\{c_{\hat{\alpha}_1}\left(F_X\left(X_i;\hat{\boldsymbol{\lambda}}_1\right),F_Y\left(Y_i;\hat{\boldsymbol{\psi}}_1\right)\right)f_X\left(X_i;\hat{\boldsymbol{\lambda}}_1\right)\right\}\right.\\
&\quad \left. -\sum_{i=1}^{n}\log\left\{f_X\left(X_i;\hat{\boldsymbol{\lambda}}_0\right)\right\}\right\},
\end{aligned} \tag{5.2.8}$$

where $\hat{\boldsymbol{\theta}}_1 = \left(\hat{\alpha}_1, \hat{\boldsymbol{\lambda}}_1, \hat{\boldsymbol{\psi}}_1\right)$ and $\hat{\boldsymbol{\theta}}_0 = \left(\alpha_0, \hat{\boldsymbol{\lambda}}_0, \boldsymbol{\psi}_0\right)$ are the parameter values that maximize the observed log-likelihood, respectively,

$$\log\left\{\prod_{i=1}^{n} f\left(X_i|Y_i; \boldsymbol{\theta}_1\right)\right\} = \sum_{i=1}^{n} \log\left\{c_{\alpha_1}\left(F_X(X_i; \boldsymbol{\lambda}_1), F_Y(Y_i; \boldsymbol{\psi}_1)\right) f_X(X_i; \boldsymbol{\lambda}_1)\right\},$$

and

$$\log\left\{\prod_{i=1}^{n} f\left(X_i|Y_i; \boldsymbol{\theta}_0\right)\right\} = \sum_{i=1}^{n} \log\left\{f_X\left(X_i; \boldsymbol{\lambda}_0\right)\right\}.$$

Therefore, an estimator of the conditional measure of dependence (5.2.6) is

$$\hat{\rho}^2_C\left(X|Y\right) = 1 - \exp\left\{-\hat{\Gamma}_C\right\}, \tag{5.2.9}$$

where $\hat{\Gamma}_C$ is given by (5.2.8).

5.2.4 Joint information gain

Proposition 5.2.2 *Let (X, Y) be a pair of random variables possibly dependent with true density $f\left(x, y; \boldsymbol{\theta}_1\right)$ given in (5.2.2). The joint information gain, based on the parametric copula density, is*

$$\begin{aligned}\Gamma &= 2\left\{\iint \log\left\{c_{\alpha_1}\left(F_X(x; \boldsymbol{\lambda}_1), F_Y(y; \boldsymbol{\psi}_1)\right) f_X(x; \boldsymbol{\lambda}_1) f_Y(y; \boldsymbol{\psi}_1)\right\} f\left(x, y; \boldsymbol{\theta}_1\right) dxdy\right.\\ &\quad \left.- \iint \log\left\{f_X\left(x; \boldsymbol{\lambda}_0\right) f_Y\left(y; \boldsymbol{\psi}_0\right)\right\} f\left(x, y; \boldsymbol{\theta}_1\right) dxdy\right\}.\end{aligned} \tag{5.2.10}$$

Proof: The twice Kullback-Leibler [35] information gain, by testing $H_0 : \alpha = \alpha_0$ versus $H_1 : \alpha \neq \alpha_0$, would be

$$\begin{aligned}\Gamma &= 2\left\{\iint \log\left\{f\left(x, y; \boldsymbol{\theta}_1\right)\right\} f\left(x, y; \boldsymbol{\theta}_1\right) dxdy\right.\\ &\quad \left.- \iint \log\left\{f\left(x, y; \boldsymbol{\theta}_0\right)\right\} f\left(x, y; \boldsymbol{\theta}_1\right) dxdy\right\}\\ &= 2\left\{\iint \log\left\{c_{\alpha_1}\left(F_X(x; \boldsymbol{\lambda}_1), F_Y(y; \boldsymbol{\psi}_1)\right) f_X(x; \boldsymbol{\lambda}_1) f_Y(y; \boldsymbol{\psi}_1)\right\} f\left(x, y; \boldsymbol{\theta}_1\right) dxdy\right.\\ &\quad \left.- \iint \log\left\{f_X\left(x; \boldsymbol{\lambda}_0\right) f_Y(y; \boldsymbol{\psi}_0)\right\} f\left(x, y; \boldsymbol{\theta}_1\right) dxdy\right\},\end{aligned}$$

where $f(x,y;\boldsymbol{\theta}_1)$ is given by (5.2.2) and we used the fact that under the independence model: $f(x,y;\boldsymbol{\theta}_0) = f_X(x;\boldsymbol{\lambda}_0)\, f_Y(y;\boldsymbol{\psi}_0)$. ∎

From Proposition 5.2.2, the joint dependence measure with respect to the work of Kent [33], is

$$\rho_J^2(X,Y) = 1 - \exp\{-\Gamma\}\cdot \tag{5.2.11}$$

5.2.5 Estimation of the joint information gain and joint measure of dependence

Let (X_i, Y_i), $i = 1, \ldots, n$ be a random sample from the joint density $f(x,y;\boldsymbol{\theta}_1)$ given in (5.2.2). From Proposition 5.2.2, the joint information gain can be expressed as

$$\begin{aligned}\Gamma &= 2\left\{\mathrm{E}\left[\log\left\{c_{\alpha_1}\left(F_X(X;\boldsymbol{\lambda}_1), F_Y(Y;\boldsymbol{\psi}_1)\right) f_X(X;\boldsymbol{\lambda}_1) f_Y(Y;\boldsymbol{\psi}_1)\right\}\right]\right.\\ &\quad \left.-\mathrm{E}\left[\log\left(f_X(X;\boldsymbol{\lambda}_0)\, f_Y(Y;\boldsymbol{\psi}_0)\right)\right]\right\}\cdot\end{aligned} \tag{5.2.12}$$

Consequently, an estimator of Γ is

$$\begin{aligned}\hat{\Gamma} &= \frac{2}{n}\left\{\sum_{i=1}^{n}\log\left\{c_{\hat{\alpha}_1}\left(F_X\left(X_i;\hat{\boldsymbol{\lambda}}_1\right), F_Y\left(Y_i;\hat{\boldsymbol{\psi}}_1\right)\right) f_X\left(X_i;\hat{\boldsymbol{\lambda}}_1\right) f_Y\left(Y_i;\hat{\boldsymbol{\psi}}_1\right)\right\}\right.\\ &\quad \left.-\sum_{i=1}^{n}\log\left\{f_X\left(X_i;\hat{\boldsymbol{\lambda}}_0\right) f_Y\left(Y_i;\hat{\boldsymbol{\psi}}_0\right)\right\}\right\},\end{aligned} \tag{5.2.13}$$

where $\hat{\boldsymbol{\theta}}_1 = \left(\hat{\alpha}_1, \hat{\boldsymbol{\lambda}}_1, \hat{\boldsymbol{\psi}}_1\right)$ and $\hat{\boldsymbol{\theta}}_0 = \left(\alpha_0, \hat{\boldsymbol{\lambda}}_0, \hat{\boldsymbol{\psi}}_0\right)$ are the parameter values that maximize, respectively, the observed log-likelihood

$$\log\left\{\prod_{i=1}^{n} f(X_i, Y_i;\boldsymbol{\theta}_1)\right\} = \sum_{i=1}^{n}\log\left\{c_{\alpha_1}\left(F_X(X_i;\boldsymbol{\lambda}_1), F_Y(Y_i;\boldsymbol{\psi}_1)\right) f_X(X_i;\boldsymbol{\lambda}_1) f_Y(Y_i;\boldsymbol{\psi}_1)\right\},$$

and

$$\log\left\{\prod_{i=1}^{n} f(X_i, Y_i;\boldsymbol{\theta}_0)\right\} = \sum_{i=1}^{n}\log\left\{f_X(X_i;\boldsymbol{\lambda}_0)\, f_Y(Y_i;\boldsymbol{\psi}_0)\right\}.$$

Hence, an estimator of the joint dependence measure (5.2.11) is

$$\hat{\rho}_J^2(X,Y) = 1 - \exp\left\{-\hat{\Gamma}\right\}, \tag{5.2.14}$$

where $\hat{\Gamma}$ is given by (5.2.13).

5.3 Information gain and dependence measure under length-biased sampling using parametric copulas method

Here, we exploit the concept of information gain to derive a dependence measure for length-biased data, using the second method: parametric copulas.

5.3.1 Introduction

Recall that, under length-biased sampling, U denotes length-biased survival time with CDF $F_{LB}(u, \boldsymbol{\lambda})$ and PDF $f_{LB}(u, \boldsymbol{\lambda})$ while Z denotes biased covariate with CDF $F_B(z, \boldsymbol{\psi})$ and PDF $f_B(z, \boldsymbol{\psi})$. Suppose that the random vector (U, Z) having a parametric copula C_α. Using Sklar's Theorem, a joint length-biased CDF of (U, Z) is

$$F_{LB}\left(u, z; \boldsymbol{\theta}\right) = C_\alpha\left(F_{LB}(u; \boldsymbol{\lambda}), F_B(z; \boldsymbol{\psi})\right), \tag{5.3.1}$$

and the corresponding joint length-biased density of (U, Z) can be found as follows

$$f_{LB}\left(u, z; \boldsymbol{\theta}\right) = c_\alpha\left(F_{LB}(u; \boldsymbol{\lambda}), F_B(z; \boldsymbol{\psi})\right) f_{LB}(u; \boldsymbol{\lambda}) f_B(z; \boldsymbol{\psi}), \tag{5.3.2}$$

where c_α is the parametric copula density given in (5.1.2). Consequently, the conditional density of U conditional on $Z = z$, in terms of the parametric copula density, is

$$f_{LB}\left(u|z; \boldsymbol{\theta}\right) = c_\alpha\left(F_{LB}(u; \boldsymbol{\lambda}), F_B(z; \boldsymbol{\psi})\right) f_{LB}(u; \boldsymbol{\lambda}) \cdot \tag{5.3.3}$$

Note that, for some independence parameter α_0 of copula C_α, the r.v.'s U and Z are independent. This implies that $f_B(z; \boldsymbol{\psi}_0) = f_Z(z; \boldsymbol{\psi}_0)$ and $F_B(z; \boldsymbol{\psi}_0) = F_Z(z; \boldsymbol{\psi}_0)$, where $F_Z(z; \boldsymbol{\psi}_0)$ and $f_Z(z; \boldsymbol{\psi}_0)$ are, respectively, CDF and PDF of the unbiased covariate under the independence model. Therefore, if the covariate sample from the incident cases is available, one can estimate $\boldsymbol{\psi}_0$ by the MLE $\hat{\boldsymbol{\psi}}_0$. In this case, the parameter of the independence model becomes $\boldsymbol{\theta}_0 = (\alpha_0, \boldsymbol{\lambda}_0)$ and this leads to

- $C_{\alpha_0}\left(F_{LB}(u;\boldsymbol{\lambda}_0), F_Z(z;\hat{\boldsymbol{\psi}}_0)\right) = F_{LB}(u;\boldsymbol{\lambda}_0)F_Z(z;\hat{\boldsymbol{\psi}}_0)$.
- $c_{\alpha_0}\left(F_{LB}(u;\boldsymbol{\lambda}_0), F_Z(y;\hat{\boldsymbol{\psi}}_0)\right) = 1$.
- $f_{LB}\left(u, z;\boldsymbol{\theta}_0\right) = f_{LB}(u;\boldsymbol{\lambda}_0)f_Z(z;\hat{\boldsymbol{\psi}}_0)$.
- $f_{LB}\left(u|z;\boldsymbol{\theta}_0\right) = f_{LB}\left(u;\boldsymbol{\lambda}_0\right)$.

5.3.2 Conditional information gain under length-biased sampling

The conditional information gain under length-biased sampling gain, based on the parametric copula density, can be derived similarly as in Proposition 5.2.1, simply by using the fact that under the dependence model ($\alpha \neq \alpha_0$): $f_{LB}\left(u|z;\boldsymbol{\theta}_1\right) = c_{\alpha_1}\left(F_{LB}(u;\boldsymbol{\lambda}_1), F_B(z;\boldsymbol{\psi}_1)\right) f_{LB}(u;\boldsymbol{\lambda}_1)$ and under the independence model ($\alpha = \alpha_0$): $f_{LB}\left(u|z;\boldsymbol{\theta}_0\right) = f_{LB}(u;\boldsymbol{\lambda}_0)$.

Proposition 5.3.1 *Let (U, Z) be a pair of random variables possibly dependent with true density $f_{LB}\left(u, z;\boldsymbol{\theta}_1\right)$ given in (5.3.2). Under length-biased sampling, the conditional information, based on the parametric copula density, can be expressed as*

$$\begin{aligned}\Gamma_C &= 2\left\{\iint \log\left\{c_{\alpha_1}\left(F_{LB}(u;\boldsymbol{\lambda}_1), F_B(z;\boldsymbol{\psi}_1)\right) f_{LB}(u;\boldsymbol{\lambda}_1)\right\} f_{LB}\left(u, z;\boldsymbol{\theta}_1\right) du dz\right. \\ &\quad \left. - \int \log\left\{f_{LB}\left(u;\boldsymbol{\lambda}_0\right)\right\} f_{LB}\left(u;\boldsymbol{\lambda}_1\right) du\right\}. \end{aligned} \tag{5.3.4}$$

5.3.3 Estimation of the conditional information gain and conditional measure of dependence for length-biased data

Let (U_i, Z_i), $i = 1, \ldots, n$ be a random sample from $f_{LB}(u, z;\theta_1)$ given in (5.3.2). Based on Proposition 5.3.1, the conditional information gain can be formulated as

$$\Gamma_C = 2\left\{\mathrm{E}\left[\log\left\{c_{\alpha_1}\left(F_{LB}(U;\boldsymbol{\lambda}_1), F_B(Z;\boldsymbol{\psi}_1)\right) f_{LB}(U;\boldsymbol{\lambda}_1)\right\}\right] - \mathrm{E}\left[\log\left\{f_{LB}\left(U;\boldsymbol{\lambda}_0\right)\right\}\right]\right\}. \tag{5.3.5}$$

An estimator of Γ_C is

$$\begin{aligned}\hat{\Gamma}_C &= \frac{2}{n}\left\{\sum_{i=1}^{n}\log\left\{c_{\hat{\alpha}_1}\left(F_{LB}\left(U_i;\hat{\boldsymbol{\lambda}}_1\right), F_B\left(Z_i;\hat{\boldsymbol{\psi}}_1\right)\right) f_{LB}\left(U_i;\hat{\boldsymbol{\lambda}}_1\right)\right\}\right. \\ &\quad \left. -\sum_{i=1}^{n}\log\left\{f_{LB}\left(U_i;\hat{\boldsymbol{\lambda}}_0\right)\right\}\right\}, \qquad (5.3.6)\end{aligned}$$

where $\hat{\boldsymbol{\theta}}_1 = \left(\hat{\alpha}_1, \hat{\boldsymbol{\lambda}}_1, \hat{\boldsymbol{\psi}}_1\right)$ and $\hat{\boldsymbol{\theta}}_0 = \left(\alpha_0, \hat{\boldsymbol{\lambda}}_0\right)$ are the parameter values that maximize, respectively, the observed log-likelihood

$$\sum_{i=1}^{n}\log\left\{c_{\alpha_1}\left(F_{LB}(U_i;\boldsymbol{\lambda}_1), F_B(Z_i;\boldsymbol{\psi}_1)\right) f_{LB}(U_i;\boldsymbol{\lambda}_1)\right\},$$

and

$$\sum_{i=1}^{n}\log\left\{f_{LB}\left(U_i;\boldsymbol{\lambda}_0\right)\right\}.$$

Therefore, an estimator of the conditional measure of dependence is then

$$\hat{\rho}_C^2\left(U|Z\right) = 1 - \exp\left(-\hat{\Gamma}_C\right), \qquad (5.3.7)$$

where $\hat{\Gamma}_C$ is given by (5.3.6).

5.3.4 Joint information gain under length-biased sampling

The joint information gain under length-biased sampling, based on the parametric copula density, can be obtained in two ways. The first approach is a direct consequence of Proposition 5.2.2 since under the independence model ($\alpha = \alpha_0$): $f_{LB}\left(u, z; \boldsymbol{\theta}_0\right) = f_{LB}(u;\boldsymbol{\lambda}_0) f_Z(z;\hat{\boldsymbol{\psi}}_0)$ and under the dependence model ($\alpha \neq \alpha_0$): $f_{LB}\left(u, z; \boldsymbol{\theta}_1\right)$ is given by (5.3.2).

Proposition 5.3.2 *Let (U, Z) be a pair of random variables possibly dependent with true density $f_{LB}\left(u, z; \boldsymbol{\theta}_1\right)$ given in (5.3.2). Under length-biased sampling, the joint information gain, based on the parametric copula density, is*

$$\begin{aligned}\Gamma &= 2\left\{\iint \log\left\{c_{\alpha_1}\left(F_{LB}(u;\boldsymbol{\lambda}_1), F_B(z;\boldsymbol{\psi}_1)\right) f_{LB}(u;\boldsymbol{\lambda}_1) f_B(z;\boldsymbol{\psi}_1)\right\} f_{LB}\left(u,z;\boldsymbol{\theta}_1\right) du dz\right.\\ &\left. -\iint \log\left\{f_{LB}\left(u;\boldsymbol{\lambda}_0\right) f_Z(z;\hat{\boldsymbol{\psi}}_0)\right\} f_{LB}\left(u,z;\boldsymbol{\theta}_1\right) du dz\right\}. \qquad (5.3.8)\end{aligned}$$

The second approach recalls from (3.1.9), we have

$$\Gamma = \Gamma_C + \Gamma_B, \qquad (5.3.9)$$

where Γ_C is given by (5.3.4) and

$$\Gamma_B = 2\left\{\int \log\left\{f_B(z;\boldsymbol{\psi}_1)\right\} f_B(z;\boldsymbol{\psi}_1) dz - \int \log\left\{f_Z(z;\hat{\boldsymbol{\psi}}_0)\right\} f_B(z;\boldsymbol{\psi}_1) dz\right\} \qquad (5.3.10)$$

is the information gain obtained through knowledge of the bias of covariate. Note that, the second approach can be obtained directly from Proposition 5.3.2.

5.3.5 Estimation of the joint information gain and joint measure of dependence for length-biased data

Let (U_i, Z_i), $i = 1, \ldots, n$ be a random sample from $f_{LB}(u, z; \theta_1)$ given in (5.3.2). There exist two ways for estimating the joint information. The first method is based on (5.3.8). So that, an estimator of Γ is

$$\begin{aligned}\hat{\Gamma} &= 2\left\{\frac{1}{n}\sum_{i=1}^{n} \log\left\{c_{\hat{\alpha}_1}\left(F_{LB}\left(U_i;\hat{\boldsymbol{\lambda}}_1\right), F_B\left(Z_i;\hat{\boldsymbol{\psi}}_1\right)\right) f_{LB}\left(U_i;\hat{\boldsymbol{\lambda}}_1\right) f_B\left(Z_i;\hat{\boldsymbol{\psi}}_1\right)\right\}\right.\\ &\left. -\frac{1}{n}\sum_{i=1}^{n} \log\left\{f_{LB}\left(U_i;\hat{\boldsymbol{\lambda}}_0\right) f_Z\left(Z_i;\hat{\boldsymbol{\psi}}_0\right)\right\}\right\}, \qquad (5.3.11)\end{aligned}$$

where $\hat{\boldsymbol{\theta}}_1 = \left(\hat{\boldsymbol{\alpha}}_1, \hat{\boldsymbol{\lambda}}_1, \hat{\boldsymbol{\psi}}_1\right)$ and $\hat{\boldsymbol{\theta}}_0 = \left(\alpha_0, \hat{\boldsymbol{\lambda}}_0\right)$ are the parameter values that maximize the observed log-likelihood, respectively,

$$\sum_{i=1}^{n} \log\left\{c_{\alpha_1}\left(F_{LB}(U_i;\boldsymbol{\lambda}_1), F_B(Z_i;\boldsymbol{\psi}_1)\right) f_{LB}(U_i;\boldsymbol{\lambda}_1) f_B(Z_i;\boldsymbol{\psi}_1)\right\},$$

and

$$\sum_{i=1}^{n} \log \left\{ f_{LB}\left(U_i; \boldsymbol{\lambda}_0\right) f_Z\left(Z_i; \hat{\boldsymbol{\psi}}_0\right) \right\}.$$

The second method is based on (5.3.9). In this direction, an estimator of the joint information gain is

$$\hat{\Gamma} = \hat{\Gamma}_C + \hat{\Gamma}_B, \tag{5.3.12}$$

where $\hat{\Gamma}_C$ is given by (5.3.6) and the estimator of Γ_B is

$$\hat{\Gamma}_B = \frac{2}{n} \left\{ \sum_{i=1}^{n} \log \left\{ f_B\left(Z_i; \hat{\boldsymbol{\psi}}_1\right) \right\} - \sum_{i=1}^{n} \log \left\{ f_Z\left(Z_i; \hat{\boldsymbol{\psi}}_0\right) \right\} \right\}. \tag{5.3.13}$$

Hence, an estimator of the joint measure of dependence is

$$\hat{\rho}_J^2\left(U, Z\right) = 1 - \exp\left\{ -\left(\hat{\Gamma}_C + \hat{\Gamma}_B\right) \right\}. \tag{5.3.14}$$

We note that, in the case where the covariate sample from the incident cases is not available, a natural estimator of the unbiased density of the covariate, f_Z, is given by (3.5.21) as

$$\hat{f}_Z\left(z\right) = \frac{1}{n} \sum_{i=1}^{n} K_h\left(z - Z_i^*\right).$$

Chapter 6

Algorithms

In this chapter, we show how to simulate length-biased survival times with covariate using kernel density estimation with a regression procedure and parametric copulas methods. In this way, we will develop in detail some useful algorithms which can be used, in the next chapter, to study the behaviour of the information gain and dependence measure especially under length-biased sampling.

6.1 Algorithms for the kernel density estimation with a regression procedure

In this section, we develop some useful algorithms for the first method. Firstly, we study the form of the length-biased distribution of the generalized gamma distributions. Secondly, we use a simple way to generate length-biased data using the relationship between the unbiased generalized gamma density and the corresponding length-biased generalized gamma density. We provide also a simulation algorithm based on the Cox model to simulate survival time with covariate from the joint length-biased density.

6.1.1 Simulating length-biased survival times

Let X be a positive random variable which follows a generalized gamma distribution $GG(r, p, k)$ defined by

$$f(x) = \frac{r}{p\Gamma(k)} \left(\frac{x}{p}\right)^{rk-1} \exp\left\{-\left(\frac{x}{p}\right)^{r}\right\}, \quad x, r, p, k > 0, \tag{6.1.1}$$

where Γ is the gamma function, r and p are shape parameters and k is the scale parameter. The generalized gamma family is very flexible and includes several well-known models as sub-models [32]. This includes $Gamma(k, p)$, $Weibull(r, p)$ and $Exp(p)$ by letting in (6.1.1), $r = 1$, $k = 1$, and $r = k = 1$, respectively. As shown by Hashimoto et al. [27], if the random variable $X \sim GG(r, p, k)$ then $Y = \log\{X\}$ follows the generalized log-gamma distribution denoted by $GLG(r^*, p^*, k^*)$, where $r^* = \left(\sqrt{k}\right)^{-1}$ is the shape parameter, $p^* = \left(r\sqrt{k}\right)^{-1}$ is the scale parameter and the location parameter is $k^* = \log\{p\} + r^{-1} \log\left\{(p^*)^{-2}\right\}$. Now, using the relationship between length-biased density and unbiased density

$$f_{LB}(x) = \frac{x f_U(x)}{\mu}, \tag{6.1.2}$$

Andres and Wolfson [12] showed that if the unbiased density is $GG(r, p, k)$ then the corresponding length-biased density can be expressed as follows

$$f_{LB}(x) = \frac{r}{p\Gamma(k + r^{-1})} \left(\frac{x}{p}\right)^{r\left(k+r^{-1}\right)-1} \exp\left\{-\left(\frac{x}{p}\right)^{r}\right\}, \tag{6.1.3}$$

which is $GG(r, p, k + r^{-1})$. In particular,

- If $k = 1$ then $GG(r, p, 1)$ is $Weibull(r, p)$ and the corresponding length-biased distribution is $GG(r, p, 1 + r^{-1})$.

- If $r = k = 1$ then $GG(1, p, 1)$ is $Exp(p)$ and the corresponding length-biased distribution is $Gamma(2, p)$.

- If $r = 1$ then $GG(1, p, k)$ is $Gamma(k, p)$ and the corresponding length-biased distribution is $Gamma(k + 1, p)$.

We note that, if the r.v. $X \sim f_{LB}(x)$, given in (6.1.3), which is $GG(r, p, 1 + r^{-1})$ then $Y = \log\{X\} \sim GLG(r^*, p^*, k^*)$, where r^* , p^* and k^* can be obtained by the same way as described above.

The following figure illustrates unbiased density, $GG(r, p, k)$, and corresponding length-biased density, $GG(r, p, k + r^{-1})$.

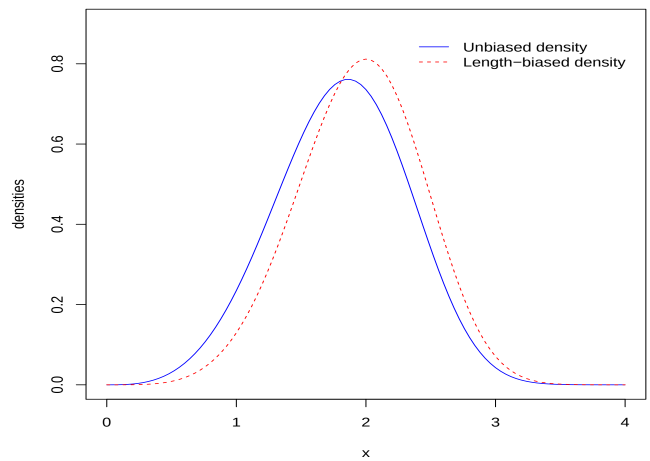

Figure 6.1: Unbiased density, $GG(r, p, k)$, versus length-biased density, $GG(r, p, k + r^{-1})$, for $r = 4$, $p = 2$ and $k = 1$.

Let $F_{LB}(x)$ denote the length-biased distribution function corresponding to (6.1.3). Using the following algorithm, we can easily generate length-biased survival times directly from $GG(r, p, k + r^{-1})$ given that the corresponding unbiased density is $GG(r, p, k)$.

Algorithm 6.1.1

For $i = 1, \ldots, n$

1. $W_i \sim U(0, 1)$.

2. $X_i = F_{LB}^{-1}(W_i)$.

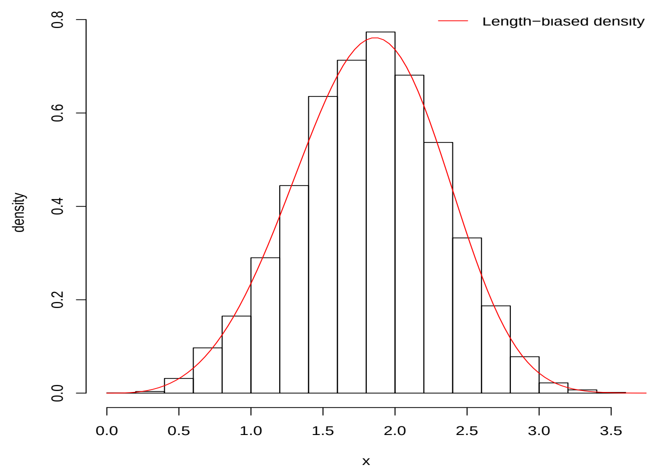

Figure 6.2: Histogram of the simulated sample $X_1, \ldots, X_n$ and corresponding length-biased density, $GG\left(r, p, k + r^{-1}\right)$, for $n = 1000$, $r = 4$, $p = 2$ and $k = 1$.

Figure 6.2 shows the histogram of the simulated length-biased survival times $X_1, \ldots, X_n$ obtained from $GG\left(r, p, k + r^{-1}\right)$.

Often, it is difficult to simulate length-biased data directly from length-biased distribution because in general the CDF $F_{LB}(x)$ and its inverse $F_{LB}^{-1}(x)$ may not have a closed form. In this case, we can use the following algorithm which is based on the bootstrap techniques.

Algorithm 6.1.2

1. Simulate a large sample $X_1^*, \ldots, X_N^*$ from a given unbiased density $f_U(x)$.

2. For $i = 1, \ldots, n$, use the bootstrap techniques from the original sample $X_1^*, \ldots, X_N^*$ with probability $p_i = X_i^* \left(\sum X_i^*\right)^{-1}$ to obtain a new sample $X_1, \ldots, X_n$ from the length-biased density $f_{LB}(x)$.

We note that the probability p_i, given in the latter algorithm, can be obtained in the same way as for (3.3.7).

6.1.2 Simulating length-biased survival times with covariate

Our main objective here is to simulate a random length-biased sample (U_i, Z_i), $i = 1, \ldots, n$ from $f_{LB}(u, z)$. For that, we develop a simulation algorithm based on the Cox model and the relationship between unbiased generalized gamma distribution and its corresponding length-biased generalized gamma distribution. To this end, we suppose that:

- The unbiased density, $f_U(u)$, of the survival time is $Weibull(r, p)$.
- The unbiased density, $f_Z(z)$, of the covariate is $U(0, 1)$.

From (3.5.19) the corresponding biased density of the covariate becomes

$$f_B(z) = \frac{e^{\beta z}}{\int_0^1 e^{\beta z} dz} = \frac{\beta e^{\beta z}}{e^{\beta} - 1}. \tag{6.1.4}$$

The biased distribution function associated with (6.1.4) is

$$F_B(z) = \int_0^z f_B(t)\, dt = \frac{e^{\beta z} - 1}{e^{\beta} - 1}, \tag{6.1.5}$$

with the following inverse

$$F_B^{-1}(z) = \frac{1}{\beta} \log\left\{1 - \left(1 - e^{\beta}\right) z\right\}. \tag{6.1.6}$$

Since $U \sim Weibull(r, p)$, which is $GG(r, p, 1)$, from (6.1.1) the density of U is

$$f(u) = \frac{r}{p}\left(\frac{u}{p}\right)^{r-1} \exp\left\{-\left(\frac{u}{p}\right)^{r}\right\}. \tag{6.1.7}$$

Based on (2.1.6), the hazard function of U can be expressed as

$$h(u) = \frac{f(u)}{S(u)} = r\frac{1}{p^r} u^{r-1}, \tag{6.1.8}$$

where $S(u)$ denotes the survival function of U given by

$$S(u) = \exp\left\{-\left(\frac{u}{p}\right)^{r}\right\}. \tag{6.1.9}$$

Since the Weibull distributions satisfy the assumption of both the AFT and PH models, the hazard function of U given a covariate Z can be written as

$$h\left(u\,|z\right) = h_0\left(u\right)\exp\left\{\beta_c z\right\}, \tag{6.1.10}$$

where $h_0\left(u\right)$ is the baseline hazard function (when $Z = 0$) and β_c is the regression coefficient under the PH model. In this case, the regression coefficient under the AFT model given in (6.1.4) is related to β_c by the following equation [36]

$$\beta = -\frac{\beta_c}{r}\cdot \tag{6.1.11}$$

This leads to

$$h\left(u\,|z\right) = r\frac{1}{p^r}u^{r-1}\exp\left\{\beta_c z\right\} = r\frac{1}{\left(p\exp\left\{-\frac{\beta_c}{r}z\right\}\right)^r}u^{r-1}\cdot \tag{6.1.12}$$

From (6.1.8) and (6.1.12), we conclude that the unbiased density of U conditional on the covariate Z denoted by $f_U\left(u\,|z\right)$ is $Weibull\left(r, p\exp\left\{-\frac{\beta_c}{r}z\right\}\right)$. This means that $U|Z = z$ follows $GG\left(r, p\exp\left\{-\frac{\beta_c}{r}z\right\}, 1\right)$. Therefore, the corresponding length-biased density denoted by $f_{LB}\left(u\,|z\right)$ is $GG\left(r, p\exp\left\{-\frac{\beta_c}{r}z\right\}, 1 + r^{-1}\right)$. The following two algorithms represent a good way to simulate data from the joint unbiased density $f_U\left(u, z\right)$ and from the joint length-biased density $f_{LB}\left(u, z\right)$, respectively.

Algorithm 6.1.3

For $i = 1, \ldots, N$

1. $Z_i^* \sim U(0, 1)$.
2. $U_i^* \sim GG\left(r, p\exp\left\{-\frac{\beta_c}{r}Z_i^*\right\}, 1\right)$.
3. $\left(U_i^*, Z_i^*\right)$ is the desired observation from $f_U\left(u, z\right)$.

Algorithm 6.1.4

For $i = 1, \ldots, n$

1. $W_i \sim U(0,1)$.
2. $Z_i = F_B^{-1}(W_i)$.
3. $U_i \sim GG\left(r, p\exp\left\{-\frac{\beta_c}{r}Z_i\right\}, 1 + r^{-1}\right)$.
4. The desired observation from $f_{LB}(u,z)$ is (U_i, Z_i).

To better understand the behavior of information gain and dependence measure estimators, under length-biased sampling, we propose to use the following algorithm.

Algorithm 6.1.5

For $k = 1, \ldots, m$ and for $i = 1, \ldots, n$

1. Simulate length-biased data (U_i, Z_i) from Algorithm 6.1.4.
2. Use $\log\{U_i\} = Y_i = \alpha + \beta Z_i + \varepsilon_i$, $i = 1, \ldots, n$ to compute $\hat{\alpha}$ and $\hat{\beta}$.
3. Given $\hat{\varepsilon}_i = Y_i - \hat{\alpha} - \hat{\beta}Z_i$, find the MLE's $\hat{r}^*$, $\hat{p}^*$ and $\hat{k}^*$.
4. Generate $\tilde{\varepsilon}_i$ from $GLG\left(\hat{r}^*, \hat{p}^*, \hat{k}^*\right)$.
5. Compute $\tilde{Y}_i = \hat{\alpha} + \hat{\beta}Z_i + \tilde{\varepsilon}_i$ and $\tilde{U}_i = \exp\left\{\tilde{Y}_i\right\}$.
6. Based on (3.5.11) and (3.3.2) compute $f_{LB}(U_i|Z_i)$ and $f_{LB}(U_i)$, respectively.
7. Use formula (3.5.21) to compute $\hat{f}_Z(Z_i)$, where $Z_1^*, \ldots, Z_n^*$ is a generated sample from $U(0,1)$.
8. Calculate $\hat{f}_B(Z_i)$ by (3.5.23).
9. Compute estimators of information gain $\hat{\Gamma}_{C,k}$, $\hat{\Gamma}_{B,k}$ and $\hat{\Gamma}_k = \hat{\Gamma}_{C,k} + \hat{\Gamma}_{B,k}$, respectively, by (3.6.3), (3.6.4) and (3.6.7).

10. Use (3.6.5) and (3.6.6) to obtain, respectively, conditional and joint dependence measure estimators $\hat{\rho}^2_{C,k}\left(U|Z\right)$ and $\hat{\rho}^2_{J,k}\left(U,Z\right)$.

6.2 Algorithms for the parametric copulas

Here, we develop some useful algorithms for the second method. Using some known parametric copula, we show how to simulate data from the joint unbiased density and then we develop a practical approach, based on the bootstrap techniques, to simulated length-biased data from the joint length-biased density.

6.2.1 Data simulation using copulas

Let $f_X(x;\boldsymbol{\lambda})$ and $F_X\left(x;\boldsymbol{\lambda}\right)$ denote PDF and CDF of the continuous r.v. X. Also, let $g(y;\boldsymbol{\psi})$ and $G_Y\left(y;\boldsymbol{\psi}\right)$ be PDF and CDF of the continuous r.v. Y. From Theorem 5.1.1, the joint CDF of the random vector (X,Y) can be written as a function of a parametric copula as follows

$$F\left(x,y,\boldsymbol{\theta}\right)=C_\alpha\left(F\left(x;\boldsymbol{\lambda}\right),G\left(y;\boldsymbol{\psi}\right)\right),\quad \forall(x,y)\in\mathbb{R}^2, \tag{6.2.1}$$

where $\boldsymbol{\theta}=\left(\alpha,\boldsymbol{\lambda},\boldsymbol{\psi}\right)$. The joint density of the random vector (X,Y), denoted by $f\left(x,y,\boldsymbol{\theta}\right)$, can be derived from (5.1.1) provided $\partial^2C_\alpha\left(u,v\right)/\partial u\partial v$ exists. Algorithm 6.2.1 can be used to simulate a random sample $\left(X_i,Y_i\right)$, $i=1,\ldots,N$ from the joint unbiased density $f\left(x,y;\boldsymbol{\theta}\right)$.

Algorithm 6.2.1

For $i = 1, \ldots, N$

1. $(U_i^*, V_i^*) \sim C_\alpha (u, v)$.
2. $X_i = F_X^{-1} (U_i^*, \boldsymbol{\lambda})$.
3. $Y_i = G_Y^{-1} (V_i^*, \boldsymbol{\psi})$.
4. The desired observation from $f (x, y; \boldsymbol{\theta})$ is (X_i, Y_i).

Now, consider a random sample $(X_1, Y_1), \ldots, (X_N, Y_N)$ obtained from the joint unbiased density $f (x, y, \boldsymbol{\theta})$ using Algorithm 6.2.1. To study the behavior of the estimated information gain and dependence measure under different values of α, dependence parameters of some known parametric copula C_α, we suggest to use the algorithm described below.

Algorithm 6.2.2

For $k = 1, \ldots, m$ and for $i = 1, \ldots, N$

1. For the conditional model, find $\hat{\boldsymbol{\theta}}_1 = \left(\hat{\alpha}_1, \hat{\boldsymbol{\lambda}}_1, \hat{\boldsymbol{\psi}}_1\right)$ and $\hat{\boldsymbol{\theta}}_0 = \hat{\boldsymbol{\lambda}}_0$ that maximize, respectively, the observed log-likelihood
$$\sum_{i=1}^{N} \log \{c_{\alpha_1} (F_X(X_i; \boldsymbol{\lambda}_1), F_Y(Y_i; \boldsymbol{\psi}_1))\, f_X(X_i; \boldsymbol{\lambda}_1)\}$$
and
$$\sum_{i=1}^{N} \log \{f_X (X_i; \boldsymbol{\lambda}_0)\}.$$
2. For the conditional model, calculate $\hat{\Gamma}_{C,k}$ and $\hat{\rho}^2_{C,k} (X|Y)$, respectively, by (5.2.8) and (5.2.9).
3. For the joint model, find $\hat{\boldsymbol{\theta}}_1 = \left(\hat{\boldsymbol{\alpha}}_1, \hat{\boldsymbol{\lambda}}_1, \hat{\boldsymbol{\psi}}_1\right)$ and $\hat{\boldsymbol{\theta}}_0 = \left(\hat{\boldsymbol{\lambda}}_0, \hat{\boldsymbol{\psi}}_0\right)$ that maximize the observed log-likelihood, respectively,
$$\sum_{i=1}^{N} \log \{c_{\alpha_1} (F_X(X_i; \boldsymbol{\lambda}_1), F_Y(Y_i; \boldsymbol{\psi}_1))\, f_X(X_i; \boldsymbol{\lambda}_1) f_Y(Y_i; \boldsymbol{\psi}_1)\},$$

and

$$\sum_{i=1}^{N} \log \{f_X(X_i; \boldsymbol{\lambda}_0) f_Y(Y_i; \boldsymbol{\psi}_0)\}.$$

4. For the joint model, calculate $\hat{\Gamma}_k$ and $\hat{\rho}^2_{J,k}(X,Y)$, respectively, by (5.2.13) and (5.2.14).

6.2.2 Length-biased data simulation using copulas

Using the copula provided in Section 6.2.1, a simulation algorithm was used to obtain a random sample from the joint unbiased density. However, as we will show in the current section, we cannot simulate data directly from the joint length-biased using copula. A bootstrap techniques will be proposed as a simple solution for this simulation problem. First, recall that

$$f_B(z) = \frac{\mu(z) f_Z(z)}{\mu}, \tag{6.2.2}$$

where

- $\mu(z) = \mathrm{E}[U|Z=z] = \int_0^\infty u f_U(u|z) du.$
- $\mu = \int_0^1 \mu(z) f_Z(z) dz = \int_0^1 \mu(z) dz.$

Using the fact that

$$f_U(u|z) = c_\alpha(F_U(u), F_Z(z)) f_U(u), \tag{6.2.3}$$

then

$$\mu(z) = \int_0^\infty u c_\alpha(F_U(u), F_Z(z)) f_U(u) du, \tag{6.2.4}$$

and hence, (6.2.2) becomes

$$f_B(z) = \frac{\int_0^\infty u c_\alpha(F_U(u), F_Z(z)) f_U(u) du}{\mu} f_Z(z)\cdot \tag{6.2.5}$$

The length-biased density of U conditional on $Z = z$ becomes

$$f_{LB}(u|z) = \frac{u f_U(u|z)}{\mu(z)} = \frac{u c_\alpha(F_U(u), F_Z(z)) f_U(u)}{\int_0^\infty u c_\alpha(F_U(u), F_Z(z)) f_U(u) du}\cdot \tag{6.2.6}$$

Even for a given parametric copula associated with some known unbiased CDF $F_U(u,z)$, Equations (6.2.5) and (6.2.6) cannot be used to simulate, directly, a random sample $(U_1, Z_1), \ldots, (U_n, Z_n)$ from $f_{LB}(u,z)$. The general problem is that we cannot find a closed forms of $f_B(z)$, $F_B(z)$, $F_B^{-1}(z)$, $f_{LB}(u|z)$, $F_{LB}(u|z)$ and $F_{LB}^{-1}(u|z)$. In this situation, we suggest to use the following algorithm, based on the bootstrap techniques, which is considered as a simple way to simulate length-biased data.

Algorithm 6.2.3

For $i = 1, \ldots, N$

1. Use Algorithm 6.2.1 to simulate (U_i^*, Z_i^*) from the joint unbiased density $f_U(u,z)$.

2. Use Algorithm 6.1.2 to obtain length-biased survival times $U_1, \ldots, U_n$.

3. From (U_i^*, Z_i^*), $i = 1 \ldots, N$ find $Z_1, \ldots, Z_n$ associated with $U_1, \ldots, U_n$.

4. The desired random sample from $f_{LB}(u,z)$ is (U_j, Z_j), $j = 1 \ldots, n$.

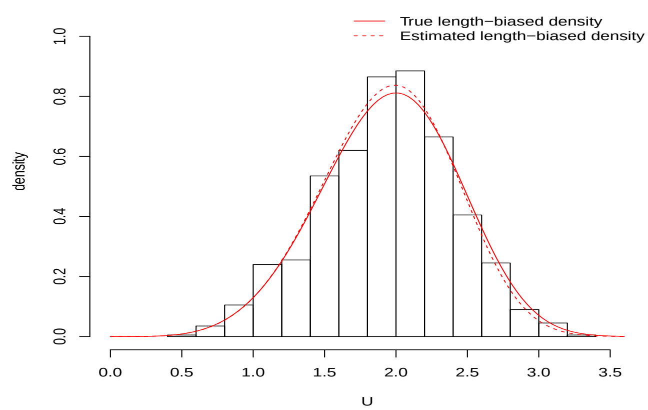

Figure 6.3: Observed frequencies of the length-biased survival times, true length-biased density $GG(r, p, 1 + r^{-1})$ and $GG(\hat{r}, \hat{p}, \hat{k})$ with $N = 5000$, $n = 1000$, $r = 4$, $p = 2$ and $\alpha = 8$.

Figure 6.3 describes the histogram of the length-biased survival times, $U_1, \ldots, U_n$, obtained from Algorithm 6.2.3, true length-biased density $GG(r, p, 1 + r^{-1})$ and corresponding $GG(\hat{r}, \hat{p}, \hat{k})$, where

- The unbiased density of the survival time , $f_U\left(u\right)$, is $Weibull(r, p)$.
- The unbiased density of the covariate , $f_Z\left(z\right)$, is $U(0, 1)$.
- The parametric copula associated with the joint unbiased CDF $F_U\left(u, z\right)$ is Clayton copula.

Chapter 7

Simulation studies

This chapter exhibits the results of several simulations assessing the behaviour of the estimated information gain and dependence measure given length-biased data. In this regard, we implement the proposed methods discussed in the previous chapters particularly, kernel density estimation with a regression procedure and parametric copulas. Also, we will show by simulation that the information gain and dependence measure properties discussed in Chapter 2, Section 2.2, continue to hold especially under length-biased sampling.

7.1 Simulation studies for the kernel density estimation with a regression procedure

In the current section, we investigate the performance of the first method. We first study the quality of the estimators proposed in Chapter 3. In particular, we compare the true densities $f_U(u|z)$, $f_{LB}(u|z)$ and $f_B(z)$ with their estimators $\hat{f}_U(u|z)$, $\hat{f}_{LB}(u|z)$ and $\hat{f}_B(z)$, respectively. Secondly, we give for different values of β, estimators of information gain and dependence measure. To this end, let (U_i, Z_i), $i = 1, \ldots, n$ be a simulated random length-biased sample, obtained from $f_{LB}(u, z)$, using Algorithm 6.1.4, where

- $n = 1000$ is the number of simulated observations.
- $r = 4$ and $p = 2$ are parameters of the Weibull distribution.
- $\beta_c = 20$ denotes the regression coefficient under the PH model.
- $\beta = -5$ is the regression coefficient under the AFT model using (6.1.11).

We will use the standard normal density as the kernel function. Recall that,

- the true unbiased density, $f_Z(z)$, of the covariate is $U(0,1)$.
- the true unbiased density of the survival time is $Weibull(r,p)$ and the true conditional unbiased density $f_U(u|z)$ is $GG\left(r, p\exp\left\{-\frac{\beta_c}{r}z\right\}, 1\right)$.
- $f_B(z)$ denotes the true biased density, given by (6.1.4).
- $GG\left(r, p\exp\left\{-\frac{\beta_c}{r}z\right\}, 1+r^{-1}\right)$ is the true conditional length-biased density $f_{LB}(u|z)$.

Now following Algorithm 3.5.1 proposed in Chapter 3, we first give the steps and results for the comparison between $f_U(u\,|z)$, $f_{LB}(u\,|z)$ and their estimators $\hat{f}_U(u\,|z)$, $\hat{f}_{LB}(u\,|z)$, respectively.
For $i = 1, \ldots, n$

1. using the linear regression $Y_i = \log\{U_i\} = \alpha + \beta Z_i + \varepsilon_i$ we find the estimators of α and β, $\hat{\alpha} = 0.639$ and $\hat{\beta} = -4.979$, respectively;
2. estimate the error ε_i by $\hat{\varepsilon}_i = Y_i - \hat{\alpha} - \hat{\beta}Z_i$;
3. since the parametric density of the errors, f_ε, is generalized log-gamma density $GLG(r^*, p^*, k^*)$, we use the sample obtained from the previous step to find the maximum likelihood estimators $\hat{r}^* = 0.130$, $\hat{p}^* = 0.214$ and $\hat{k}^* = 1.041$;
4. generate $\tilde{\varepsilon}_i$ from $GLG\left(\hat{r}^*, \hat{p}^*, \hat{k}^*\right)$;

5. generate $V \sim U(0,1)$;

6. for a fixed covariate $z = F_B^{-1}(V)$, we compute $\tilde{Y}_i = \hat{\alpha} + \hat{\beta} z + \tilde{\varepsilon}_i$;

7. the pseud-observations from $f_{LB}(u\,|z)$ are $\tilde{U}_i = \exp\left\{\tilde{Y}_i\right\}$;

8. based on (3.5.6), we estimate $\mu(z)$ by $\hat{\mu}(z) = n\left(\sum_{i=1}^{n} \tilde{U}_i^{-1}\right)^{-1}$;

9. an estimator of $f_U(u\,|z)$ is given by (3.5.7)

$$\hat{f}_U(u\,|z) = n^{-1}\hat{\mu}(z)\sum_{i=1}^{n}\frac{1}{\exp\left\{\hat{\alpha}+\hat{\beta}z+\tilde{\varepsilon}_i\right\}}K_h\left(u-\exp\left\{\hat{\alpha}+\hat{\beta}z+\tilde{\varepsilon}_i\right\}\right) ;$$

10. an estimator of $f_{LB}(u\,|z)$ can be obtained by (3.5.11)

$$\hat{f}_{LB}(u\,|z) = n^{-1}\sum_{i=1}^{n}\frac{u}{\exp\left\{\hat{\alpha}+\hat{\beta}z+\tilde{\varepsilon}_i\right\}}K_h\left(u-\exp\left\{\hat{\alpha}+\hat{\beta}z+\tilde{\varepsilon}_i\right\}\right).$$

Figure 7.1: Observed frequencies of the estimated error and its corresponding density $GLG\left(\hat{r}^*, \hat{p}^*, \hat{k}^*\right)$.

Figure 7.2 and Figure 7.3 provide a good comparison between $f_U(u\,|z)$, $f_{LB}(u\,|z)$ and $\hat{f}_U(u\,|z)$, $\hat{f}_{LB}(u\,|z)$, respectively. In both figures, we can see clearly that the true density is very close to its density estimator.

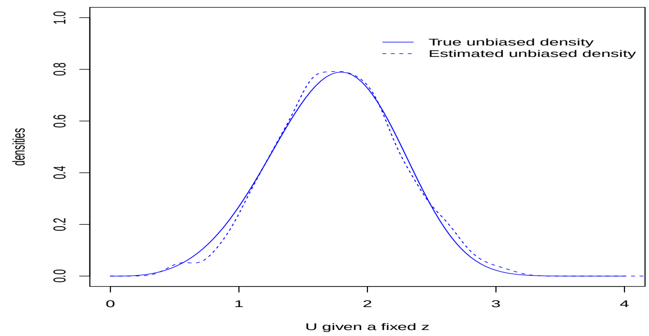

Figure 7.2: True unbiased density $f_U\left(u\,|z\right)$ and its estimator $\hat{f}_U\left(u\,|z\right)$.

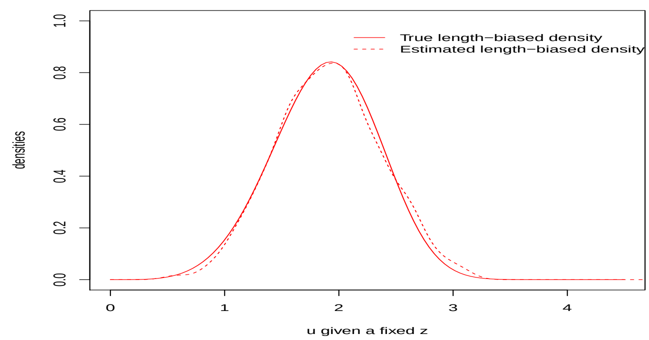

Figure 7.3: True length-biased density $f_{LB}\left(u\,|z\right)$ and its estimator $\hat{f}_{LB}\left(u\,|z\right)$.

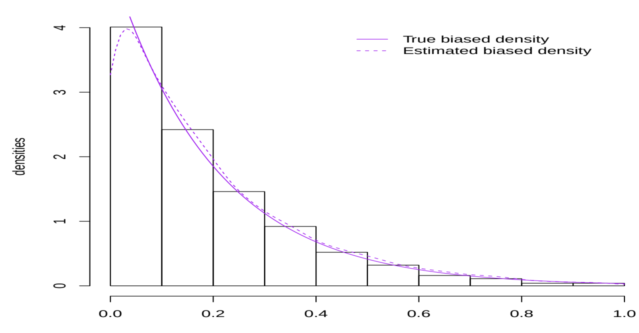

Figure 7.4: Observed frequencies of the biased covariate, true biased density $f_B\left(z\right)$ and its estimator $\hat{f}_B\left(z\right)$.

Figure 7.4 compares the true biased density $f_B(z)$ in (6.1.4) to the estimated biased density $\hat{f}_B(z)$ given by (3.5.23).

Algorithm 6.1.5, where $m = 1000$ denotes the number of the simulated samples, is used to study the behavior of the estimated information gain and dependence measure under length-biased sampling. Here, $\hat{\Gamma} = \hat{\Gamma}_C + \hat{\Gamma}_B$, $\hat{\rho}^2_C(U|Z) = 1 - e^{-\hat{\Gamma}_C}$ and $\hat{\rho}^2_J(U,Z) = 1 - e^{-\hat{\Gamma}}$.

True β	Av. $\hat{\beta}$	Av. $\hat{\Gamma}_C$	Av. $\hat{\Gamma}_B$	Av. $\hat{\Gamma}$	Av. $\hat{\rho}^2_C(U\|Z)$	Av. $\hat{\rho}^2_J(U,Z)$
0.25	0.264	0.072	0.004	0.078	0.069	0.074
0.5	0.451	0.291	0.088	0.379	0.252	0.315
1	0.989	0.802	0.153	0.955	0.551	0.615
8	8.089	2.369	2.392	4.762	0.906	0.991

Table 7.1: The average information gain and dependence measure estimates given length-biased data, using kernel density estimation with a regression procedure, for $n = m = 1000$, $r = 4$ and $p = 2$.

In fact Table 7.1 indicates, for different values of the regression coefficient β, the average (Av.) information gain and dependence measure estimates given simulated length-biased data. Here, we compare conditional information with joint information gain and conditional dependence measure with joint dependence measure. Indeed, for $\beta = 0.25$ the estimate of the information gain is small and results in an estimated dependence measure that is almost identical to the estimated information gain. This comes from the fact that for Γ close to zero, $1 - e^{-\Gamma} \approx \Gamma$ by using a Taylor expansion. Also, we can see that once β increases, $\hat{\Gamma}_C$, $\hat{\Gamma}_B$ and $\hat{\Gamma}$ increase also and consequently, the dependence measures estimators $\hat{\rho}^2_C(U|Z)$ and $\hat{\rho}^2_J(U,Z)$ increase. Moreover, for $\beta = 8$ the joint measure of dependence estimator becomes close to one.

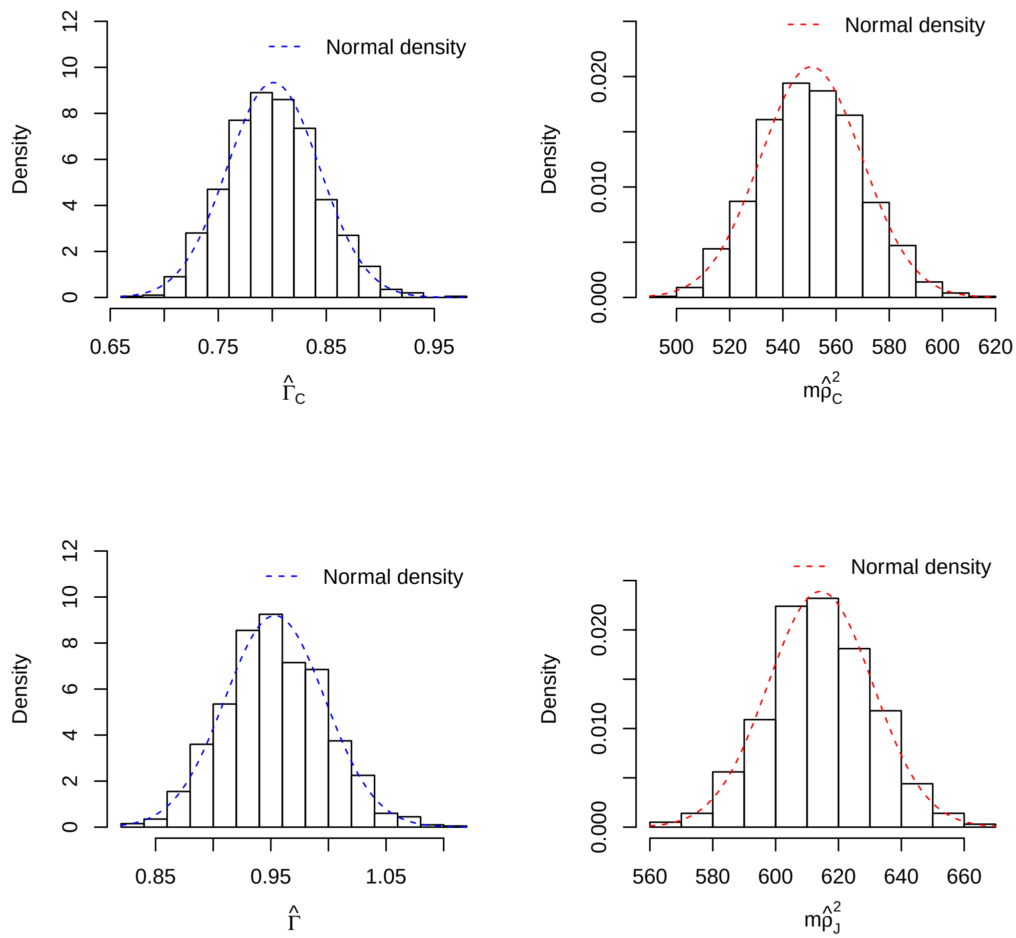

Figure 7.5: Histograms of $\hat{\Gamma}_C$, $m\hat{\rho}_C^2\,(U|Z)$, $\hat{\Gamma}$ and $m\hat{\rho}_J^2\,(U,Z)$, using kernel density estimation with a regression procedure, compared with the normal density for $n = m = 1000$, $r = 4$, $p = 2$ and $\beta = 1$.

When a regression coefficient β does not satisfy the null hypothesis ($\beta \neq 0$), $\hat{\Gamma}_C$, $\hat{\Gamma}$, $m\hat{\rho}_C^2\,(U|Z)$ and $m\hat{\rho}_J^2\,(U,Z)$ are normally distributed as shown by Figure 7.5 for $\beta = 1$. However, we cannot confirm, by simulation, that under the null hypothesis (β close to zero), $m\hat{\Gamma}_C$, $m\hat{\Gamma}$, $m\hat{\rho}_C^2\,(U|Z)$ and $m\hat{\rho}_J^2\,(U,Z)$ follow Chi-squared distribution with $k = 1$ degrees of freedom. The main problem lies in the estimation of the regression coefficient when the true value of β is close to zero.

7.2 Simulation studies for the parametric copulas

In this section, we examine the performance of the parametric copulas method when the data come from an unbiased density. Especially, under length-biased sampling the performance of this approach will be considered. As for the kernel density estimation with a regression procedure, we suppose that

- the true unbiased density of the survival time is $Weibull(r, p)$, where $r = 4$ and $p = 2$;
- the true unbiased density of the covariate is $U(0, 1)$;

and consider

- C_α: Clayton copula defined in Chapter 5;
- $\boldsymbol{\theta} = (\alpha, r, p)$: parameter of the model;
- $N = 1000$: number of the simulated observations;
- $m = 1000$: number of the simulated samples.

We first start, to simulate a random sample (U_i^*, Z_i^*) $i = 1, \ldots, N$ from the joint unbiased density $f_U(u, z)$ using Algorithm 6.1.3 and then we apply Algorithm 6.1.5 to investigate the behavior of the estimated information gain and dependence measure, for different values of the dependence parameter α of the Clayton copula. We note that, since the unbiased density of the covariate is uniformly distributed, the conditional and joint dependence measure are equal for any $\boldsymbol{\theta} = (\alpha, r, p)$. Later, we will see that under length-biased sampling they are not equal and the difference may be very significant.

True $\boldsymbol{\theta}$			Av. $\hat{\boldsymbol{\theta}}_1$			Av. $\hat{\boldsymbol{\theta}}_0$	
α	r	p	$\hat{\alpha}_1$	$\hat{r}_1$	$\hat{p}_1$	$\hat{r}_0$	$\hat{p}_0$
0.005	4	2	0.0067	4.0042	1.9987	4.0040	1.9986
0.5	4	2	0.5024	4.0040	1.9996	4.0044	1.9996
2	4	2	2.0172	4.0030	2.0004	4.0065	2.0004
10	4	2	10.0253	3.9997	2.0001	4.0098	2.0009

Table 7.2: Av. MLE's for $\boldsymbol{\theta}$ under hypotheses H_1 and H_0, for $N = m = 1000$.

True α	Av. $\hat{\Gamma}_C$	Av. $\hat{\rho}^2_C\,(X\|Y)$
0.005	0.0010	0.0010
0.5	0.1461	0.1357
2	0.8664	0.5789
10	2.9757	0.9488

Table 7.3: Av. information gain and dependence measure estimators, using parametric copula method, for $N = m = 1000$, $r = 4$ and $p = 2$.

First from Table 7.3, for $\alpha = 0.005$ the conditional dependence measure is close to 0. Second, this dependence measure increases when α increases and approaches 1 under higher dependence. Moreover, for any dependence parameter α which does not satisfy the null hypothesis, in particular for $\alpha = 10$, $\hat{\Gamma}_C$ and $m\hat{\rho}^2_C\,(U|Z)$ are normally distributed, as shown by Figure 7.6. On other hand for α close to zero (under the null hypothesis) we can see clearly from the Figure 7.7 that $m\hat{\Gamma}_C$ and $m\hat{\rho}^2_C\,(U|Z)$ follow Chi-squared distributions with $k = 1$ degrees of freedom.

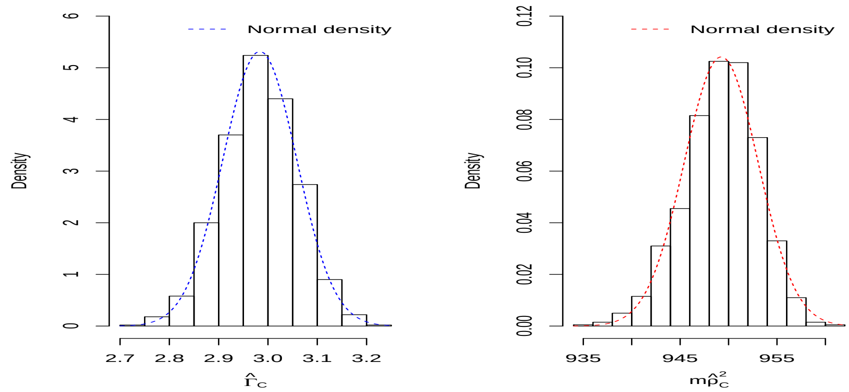

Figure 7.6: Histograms of $\hat{\Gamma}_C$ and $m\hat{\rho}_C^2\,(U|Z)$, using parametric copula method, compared with the normal density for $\alpha = 10$.

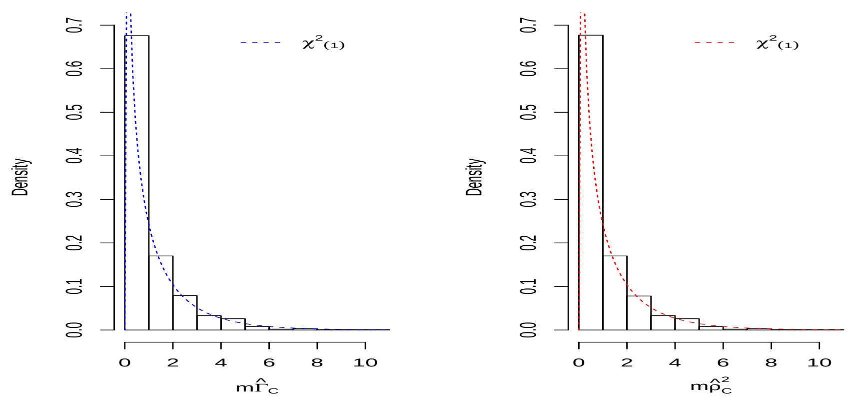

Figure 7.7: Histograms of $m\hat{\Gamma}_C$ and $m\hat{\rho}_C^2\,(U|Z)$, using parametric copula method, compared with the Chi-squared density for $\alpha = 0.005$

Now, given length-biased data (U_i, Z_i), $i = 1\ldots, n$ obtained by using Algorithm 6.2.3, the question is which copula family is associated with the joint CDF $F_{LB}\,(u, z)$? A practical answer to this question is to use the goodness-of-fit procedures for copula to find the parametric copula is associated with that length-biased data. In such a case, we suggest to use the goodness-of-fit statistic computed from the empirical copula processes, $S_n^{(E)}$, discussed in Section 5.1.8.

True copula		Families under H_0				
		Clayton	Frank	Gumbel	Gaussian	Student
Clayton	α	$S_n^{(E)}$	$S_n^{(E)}$	$S_n^{(E)}$	$S_n^{(E)}$	$S_n^{(E)}$
	0.005	4,3	4.4	-	5.4	49.7
	0.5	4.7	99.5	100	88.6	97.5
	2	4.2	100	100	100	100
	10	3.8	100	100	100	100

Table 7.4: Percentage of rejection at 5%, based on 1000 replicates, of the null hypothesis of belonging to a given family of copulas with $N = 5000$, $n = m = 1000$, $r = 4$ and $p = 2$.

Table 7.4 leads to the conclusion that the test based on $S_n^{(E)}$ confirms that the Clayton family copula associated with the unbiased CDF $F_U(u, z)$ is the same as for the length-biased CDF $F_{LB}(u, z)$, but with different estimated values of dependence parameter, denoted by $\hat{\alpha}_{LB}$, as shown by the Table 7.5.

True α	Av. $\hat{\alpha}$	Av. $\hat{\alpha}_{LB}$
0.005	0.0052	0.0048
0.5	0.4996	0.3853
2	1.9980	1.5219
10	10.021	7.5899

Table 7.5: Av. estimated dependence parameters $\hat{\alpha}$ and $\hat{\alpha}_{LB}$, based on 1000 replicates, for Clayton copula associated with the CDF's $F_U(u, z)$ and $F_{LB}(u, z)$, respectively, for $N = 5000$, $n = m = 1000$, $r = 4$ and $p = 2$.

Now our principal objective is to examine, for different values of α given in Table 7.5, the behavior of information gain and dependence measure estimators. Recall that, the copula family under length-biased sampling is Clayton copula with dependence parameter, denoted by α_{LB}, and the length-biased density of the survival time $f_{LB}(u)$ is $GG(r, p, k)$, where $k = 1 + r^{-1}$. Let $\boldsymbol{\theta}_{LB} = (\alpha_{LB}, r, p, k)$ denote the parameter of the model under length-biased sampling. A simple choice used to estimate the unbiased density $f_Z(z)$ and the biased density $f_B(z)$ is the kernel density estimator as follows

$$\hat{f}_Z(z) = \frac{1}{N}\sum_{i=1}^{N} K_h(z - Z_i^*), \tag{7.2.1}$$

$$\hat{f}_B(z) = \frac{1}{n}\sum_{i=1}^{n} K_h(z - Z_i). \tag{7.2.2}$$

To estimate the joint dependence measure, Algorithm 6.2.2 can be adapted for length-biased data provided

$$\hat{\Gamma}_B = \frac{2}{n}\left\{\sum_{i=1}^{n}\log\left\{\hat{f}_B(Z_i)\right\} - \sum_{i=1}^{n}\log\left\{\hat{f}_Z(Z_i)\right\}\right\}. \tag{7.2.3}$$

Av. $\hat{\boldsymbol{\theta}}_{LB,1}$				Av. $\hat{\boldsymbol{\theta}}_{LB,0}$		
$\hat{\alpha}_{LB,1}$	$\hat{r}_1$	$\hat{p}_1$	$\hat{k}_1$	$\hat{r}_0$	$\hat{p}_0$	$\hat{k}_0$
0.0021	3.9751	1.9749	1.2788	3.9766	1.9755	1.2776
0.3553	3.9396	1.9597	1.3266	3.9932	1.9799	1.2692
1.4696	3.9430	1.9628	1.3235	3.9811	1.9753	1.2783
7.4200	3.8997	1.931	1.3600	3.9770	1.9742	1.2802

Table 7.6: A.v MLE's for $\boldsymbol{\theta}_{LB}$, using parametric copula method, under hypotheses H_1 and H_0 for $N = 5000$, $n = m = 1000$, $r = 4$ and $p = 2$.

The following table gives, for different estimated values of dependence parameter $\hat{\alpha}_{LB}$ the average of information gain and dependence measure estimators under length-biased sampling.

Av. $\hat{\alpha}_{LB}$	Av. $\hat{\Gamma}_C$	Av. $\hat{\Gamma}_B$	Av. $\hat{\Gamma}$	Av. $\hat{\rho}_C^2\,(U\|Z)$	Av. $\hat{\rho}_J^2\,(U,Z)$
0.0021	0.0010	0.0028	0.0299	0.0010	0.0295
0.3553	0.0951	0.0145	0.1097	0.0905	0.1036
1.4696	0.6472	0.0215	0.6688	0.4758	0.4871
7.4200	2.5431	0.0426	2.5858	0.9211	0.9245

Table 7.7: Av. estimated information gain and dependence measure given simulated length-biased data, using parametric copula method, for $N = 5000$, $n = m = 1000$, $r = 4$ and $p = 2$.

The most important remark from Table 7.7 is that the estimated conditional and joint dependence measures are slightly different due to the small values of $\hat{\Gamma}_B$ for all estimated values of $\hat{\alpha}_{LB}$. This can be explained, simply, by the initial choice of the model parameters. In particular, if the shape parameter $r = 0.6$, the parametric copula associated with the CDF $F_{LB}(u, z)$ is always Clayton copula.

True copula		Families under H_0				
		Clayton	Frank	Gumbel	Gaussian	Student
Clayton	α	$S_n^{(E)}$	$S_n^{(E)}$	$S_n^{(E)}$	$S_n^{(E)}$	$S_n^{(E)}$
	0.005	5	5	-	5.2	49.6
	0.5	5.4	22.7	48.3	17.6	62.5
	2	6.9	99.4	100	97.7	99.7
	10	4	100	100	100	100

Table 7.8: Percentage of rejection at 5%, based on 1000 replicates, of the null hypothesis of belonging to a given family of copulas for $N = 5000$, $n = m = 1000$, $r = 0.6$ and $p = 2$.

However we can see clearly, by Table 7.8, how this new value of the shape $r = 0.6$ may influence considerably $\hat{\Gamma}_B$ and $\hat{\Gamma}_C$.

Av. $\hat{\alpha}_{LB}$	Av. $\hat{\Gamma}_C$	Av. $\hat{\Gamma}_B$	Av. $\hat{\Gamma}$	Av. $\hat{\rho}^2_C\,(U\vert Z)$	Av. $\hat{\rho}^2_J\,(U,Z)$
0.0009	0.0009	0.0288	0.0298	0.0009	0.0293
0.1300	0.0145	0.0664	0.0809	0.0143	0.0776
0.5017	0.1410	0.4576	0.5987	0.1312	0.4501
2.5136	1.1282	1.0504	2.1786	0.6757	0.8866

Table 7.9: Av. estimated information gain and dependence measure given simulated length-biased data, using parametric copula method, for $N = 5000$, $n = m = 1000$, $r = 0.6$ and $p = 2$.

Consequently, the difference between estimated conditional and joint dependence measure is very significant and hence we can conclude that given length-biased data we cannot ignore the potential effect of the covariate on the survival time.

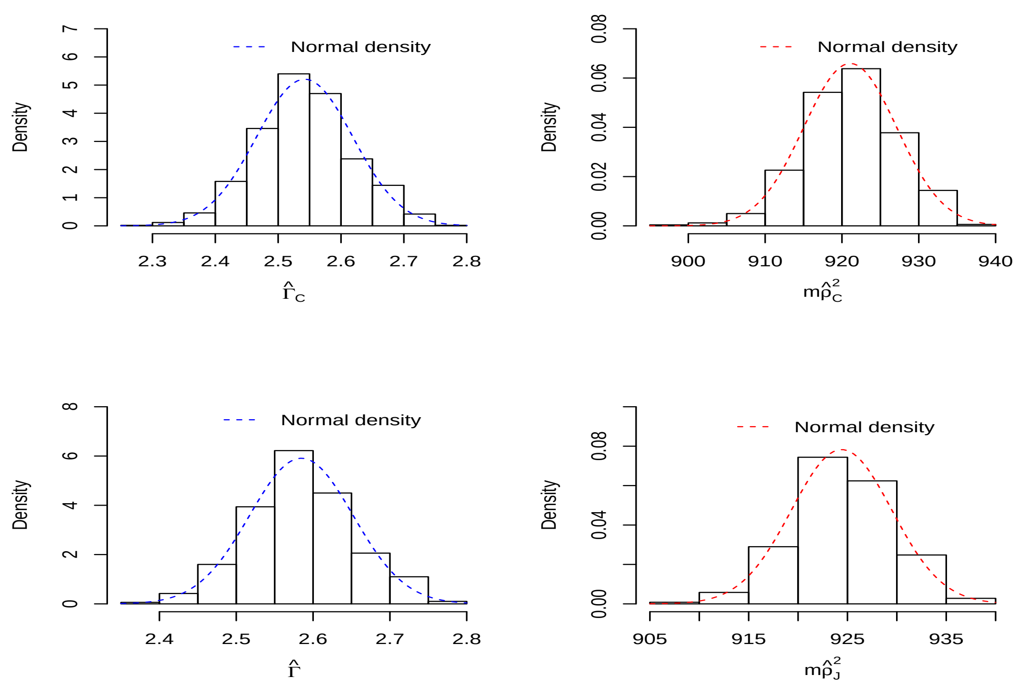

Figure 7.8: Histograms of $\hat{\Gamma}_C$, $m\hat{\rho}^2_C\,(U|Z)$, $\hat{\Gamma}$, $m\hat{\rho}^2_J\,(U,Z)$ given length-biased data, using parametric copula method, compared with the normal density for $N = 5000$, $n = m = 1000$, $r = 4$, $p = 2$ and $\alpha = 10$.

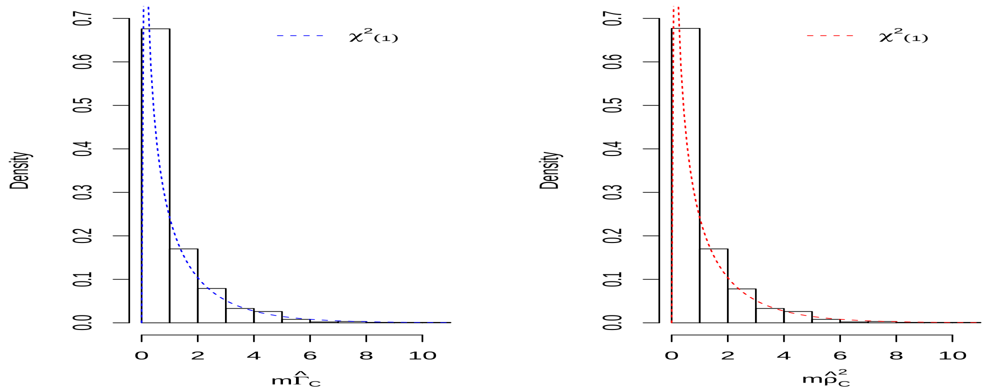

Figure 7.9: Histograms of $m\hat{\Gamma}_C$ and $m\hat{\rho}^2_C(U|Z)$ given length-biased data, using parametric copula method, compared with the Chi-squared density for $N = 5000$, $n = m = 1000$, $r = 4$, $p = 2$ and $\alpha = 0.005$.

Figure 7.8 leads to conclude that for $\alpha = 10$, which does not satisfy the null hypothesis, $\hat{\Gamma}_C$, $m\hat{\rho}^2_C(U|Z)$, $\hat{\Gamma}$ and $m\hat{\rho}^2_J(U,Z)$ are normally distributed, while under the null hypothesis such that $\alpha = 0.005$, Figure 7.9 shows that $m\hat{\Gamma}_C$ and $m\hat{\rho}^2_C(U|Z)$ are distributed according to $\chi^2_{(1)}$. However, since $\hat{\Gamma} = \hat{\Gamma}_C + \hat{\Gamma}_B$ and a kernel density estimators are used to compute $\hat{\Gamma}_B$ by (7.2.3) we could not obtain a Chi-squared distribution for $m\hat{\Gamma}$ and $m\hat{\rho}^2_J(U,Z)$. In this direction, if the parametric models for the biased density $f_B(z)$ and for the unbiased density $f_Z(z)$ can be used to estimate Γ_B we can then obtain under the null hypothesis $\hat{\Gamma}_B \approx 0$, since under the independent model we have $f_B(z) = f_Z(z)$. Consequently, we can get $\hat{\Gamma} \approx \hat{\Gamma}_C$ and in such a case $m\hat{\Gamma}$ and $m\hat{\rho}^2_J(U,Z)$ can be also distributed according to $\chi^2_{(1)}$.

Conclusion and future works

The present work of research lies in the survival analysis theory. The main contribution of this work lies in the development of two different methods to obtain a dependence measure for length-biased survival data. In the first proposed approach, kernel density estimation with a regression procedure, we used the concept of information gain, the kernel methods and linear regression procedure while in the second proposed approach, parametric copulas, we used the concept of information and parametric families of copulas frequently used in survival analysis.

In Chapter 2, we stated some basic notions of survival analysis and dependence measure for survival data, based on the concept of information gain. Under cross sectional study, we exposed length-biased distribution of the survival time and biased distribution of the covariates. In Chapter 3, we adapted Kent's [33] dependence measure in the context of length-biased sampling, without censoring for the case of one continuous covariate. In this sense, we provided a link between the conditional and joint information gain and then we developed kernel density estimation with a regression procedure to estimate the dependence measure for length-biased survival data. In Chapter 4 we generalized the results obtained in Chapter 3. Particulary, we focused our attention on the general case: partial dependence measure, without censoring for the case of several continuous covariates. In the last section of this chapter, we presented a study of the consistency of the proposed estimators which are related to the dependence measure. In Chapter 5, we reviewed some notions of copulas and then,

based on the concept of information gain, we used the parametric copulas to obtain the dependence measure between a survival time and one continuous covariate, without censoring. Furthermore, this approach was adapted under cross sectional study. New simulation algorithms were developed, in Chapter 6, for the implementation of the two proposed methods. In Chapter 7, simulation studies demonstrate the performance of the estimators related to the dependence measure, using both methods kernel density estimation with a regression procedure and parametric copulas.
This thesis addressed the problem of the measure of dependence for length-biased survival data. The goal concentrated on an extension of Kent's [33] dependence measure under length-biased sampling. More specifically, we looked at a measure of dependence between survival time and continuous covariates without censoring. For the future work it would interesting to :

- extend the parametric copulas for the case of several continuous covariates;
- adapt both proposed methods, under censoring;
- consider others type of covariates in the model.

Bibliography

[1] Addona, V. and Wolfson, D. (2006). A formal test for the stationarity of the incidence rate using data from a prevalent cohort study with follow-up. *Lifetime Data Analysis.* **12(3)**, 267-284.

[2] Ajami. M., Fakoor. V. and Jomhoori. S. (2013). Some asymptotic results of kernel density estimator in length-biased sampling. *Journal of Sciences.* **24(1)**, 55-62.

[3] Akaike, H. (1954). An approximation to the density function *Ann. Inst. Statist. Math.* **6(2)**, 127132.

[4] Tsybakov, A.B. (2009). Introduction to nonparametric estimation. *Springer, New York.*

[5] Asgharian, M., M'Lan, C., and Wolfson, D.B. (2002). Length-biased sampling with right censoring: An unconditional approach. *Journal of the American Statistical Association.* **97(457)**, 201-209.

[6] Asgharian, M., Wolfson, D.B. and Zhang, X. (2006). Checking stationarity of the incidence rate using prevalent cohort survival data. *Statistics in medicine.* **25(10)**, 1751-1767.

[7] Bhattacharyya, B.B., Franklin, L.A. and Richardson, G. D. (1988). A comparison of nonparametric unweighted and length-biased density estimation of fibres. *Comm. Statist.* **17(11)**, 3629-3644.

[8] Berg, D. (2009). Goodness-of-fit testing: An overview and power comparison. *The European Journal of Finance.* **15(7-8)**, 675-701.

[9] Bergeron, P.-J, Ashgarian, M. and Wolfson, D.B. (2008). Covariate bias induced by length-biased sampling of failure times. *Journal of the American Statistical Association.* **103(482)**, 737-742.

[10] Breymann, W., Dias, A. and Embrechts P. (2003). Dependence structures for multivariate high-frequency data in finance. *Quantitative Finance.* **3(1)**, 114.

[11] Chen, G. J., Lai, T. L. and Wei, C. Z. (1981). Convergence system and strong consistency of least squares estimates in regression models. *Journal of Multivariate Analysis.* **11(3)**, 319-333.

[12] Andres, C.J. and Wolfson, D.B. (1999). Length-bias: some characterizations and applications. *Journal of statistical computation and simulation.* **64(3)**, 209-219.

[13] Cox, D. R. (1969). Some sampling problems in technology. *In new developments in survey sampling.* Wiley.

[14] Cox, D. R. (1975). Partial likelihood. *Biometrika.* **62(2)**, 269-276.

[15] Deheuvels, P. (1977). Estimation non parametrique de la densité par histogrammes generalisés. *Publ. l'Inst. Statist. l'Univ.* **(22)**, 1-23.

[16] Fermanian, J. (2005). Goodness-of-fit tests for copulas. *Journal of Multivariate Analysis.* **95(1)**, 119-152.

[17] Fermanian, J., Radulovic, D. and Wegkamp, M. (2004). Weak convergence of empirical copula processes. *Bernoulli.* **10(5)**, 84860.

[18] Fraser, D.A.S. (1965). On information in statistics. *The Annals of Mathematical Statistics.* **36(3)**, 890-896.

[19] Genest, C. and Mackay R. J. (1986). The joy of copulas: Bivariate distributions with uniform marginals. *The American Statistician.* **(40)**, 280-283.

[20] Genest, C. and Rivest L.-P. (1993). Statistical inference procedures for bivariate archimedean copulas. *Journal of the American Statistical Association.* **88(423)**, 1034-1043.

[21] Genest, C., Ghoudi K. and Rivest, L. (1995). A semi-parametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika.* **82(3)**, 543552.

[22] Genest, C., Quessy, J.-F. and Rémillard B. (2006). Goodness-of-fit procedures for copula models based on the probability integral transform. *Scandinavian Journal of Statistics.* **33(2)**, 337-366.

[23] Genest, C., and Rémillard B. (2008). Validity of the parametric bootstrap for goodness-of- fit testing in semiparametric models. *Annals de l'institut Henri Poincaré-Probabilités et statitistiques.* **44(2)**, 199-213.

[24] Genest, C., Rémillard B. and Beaudoin D. (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics.* **(44)**, 199-213.

[25] Ghoudi, K., and Rémillard B. and (2004). Empirical processes based on pseudo-observations. II. The multivariate case. In: Asymptotic Methods in Stochastics. *Amer. Math. Soc.* **(44)**, 381406.

[26] Gumbel E. J. (1960). Bivariate exponential distributions. *Journal of the American Statistical Association.* **55(292)**, 698-707.

[27] Hashimoto, E.M, Ortega, E.M. , Cancho, V.G. and Cordeiro, G.M. (2013). On estimation and diagnostics analysis in log-generalized gamma regression model for interval-censored data. *Statistics* **47(2)**, 379-398.

[28] Ibrahim A. A. (1995). On multivariate kernel estimation for samples from weighted distributions. *Statistics & Probability Letters.* **22(2)**, 121-129.

[29] Joe, H. (1997). Multivariate models and dependence concepts. *Chapman & Hall, London.*

[30] Jones, M.C. (1991). Kernel density estimation for length-biased data. *Biometrika.* **78(3)**, 511-519.

[31] Klein, J.P. and Moeschberger, M.(2003). Survival analysis techniques for censored and truncated data, second edition. *Springer, New York.*

[32] Johnson, N.L., Kotz, S. and Balakrishnan, N.(1994). Continuous univariate distributions. *John Wiley and Sons, New York.*

[33] Kent, J.T. (1983). Information gain and a general measure of correlation. *Biometrika.* **70(1)**, 163-173.

[34] Kent, J.T. and O'Quigley, J. (1988). Measure of dependence for censored survival data. *Biometrika.* **75(3)**, 525-534.

[35] Kullback, S. and Leibler, R.A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics.* **22(1)**, 79-86.

[36] Lawless, J. F. (2003). Statistical models and methods for lifetime data. *Wiley, New York.*

[37] Linfoot, E.H (1957). On informational measure of correlation. *Information and Control.* **1(1)**, 85-89.

[38] Mesfioui M., Quessy, J.-F. and Toupin M.-H. (2009). On a new goodness-of-fit process for families of copulas . *The Canadian Journal of Statistics.* **37(1)**, 80-101.

[39] Nelson, R. B. (2006). An introduction to copulas. *Springer, New York.*

[40] Parzen, E. (1962). On estimation of a probability density function and the mode. *Ann. Math. Statist.* **33(3)**, 10651076.

[41] Rao, C.R. (1965). In a celebration of statistics. *Atkinson A.C.* & *Fienberg S.E., Springer, New York.*

[42] Rosenblatt, M. (1952). Remarks on a multivariate transformation. *Ann. Math. Statist.* **23(2)**, 470472.

[43] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27(3)**, 832835.

[44] Scaillet, O. (2007). Kernel based goodness-of-fit tests for copulas with fixed smoothing parameters. *Journal of Multivariate Analysis.* **98(3)**, 533-543.

[45] Shih, J. H. (1998). A goodness-of-fit test for association in a bivariate survival model. *Biometrika.* **85(1)**, 189-200.

[46] Silverman, B. W. (1986). Density estimation for statistics and data analysis. *Chapman & Hall, London.*

[47] Sklar, A. (1959). Fonction de répartion á n dimensions et leurs marges . *Publ. Inst. Statist. Univ. Paris.* **(8)**, 229-231

[48] Vardi, Y. (1989). Multiplicative censoring, renewal processes, convolution and decreasing density: nonparametric estimation. *Biometrika.* **76(4)**, 751-61.

[49] Wand, M.P. and Jones, M.C. (1995). Kernel smoothing. *Chapman* & *Hall, London.*

[50] Wang, M.-C. (1991). Nonparametric estimation from cross-sectional survival data. *Journal of the American Statistical Association.* **86(413)**, 130-143.

[51] Wang, M.-C. (1996). Hazards regression analysis for length-biased data. *Biometrika.* **83(2)**, 343-354.