Quantification and Estimation of Regression to The Mean for Bivariate Distributions

Manzoor Khan

School of Mathematics and Statistics
The University of New South Wales

August, 2019

Submitted in total fulfillment of the requirements
of the degree of Doctor of Philosophy

Thesis/Dissertation Sheet

Surname/Family Name	:	**Khan**
Given Name/s	:	**Manzoor**
Abbreviation for degree as give in the University calendar	:	**PhD**
Faculty	:	**Science**
School	:	**Mathematics and Statistics**
Thesis Title	:	**Quantification and Estimation of Regression to The Mean for Bivariate Distributions**

Abstract 350 words maximum: (PLEASE TYPE)

Regression to the mean (RTM) occurs when relatively high or low observations upon remeasurement are found closer to the population mean. When an intervention is applied to subjects selected in the tail of a distribution, an observed mean difference of the pre-post variables is called the total effect. The total effect is the sum of RTM and intervention/treatment effects, and estimation of RTM helps to accurately estimate the intervention/treatment effect.

The first study considers the bivariate Poisson distribution. Formulae for the total effect are derived and decomposed into RTM and intervention effects. The behaviour of RTM is demonstrated for homogeneous and inhomogeneous Poisson processes. Maximum likelihood estimators (MLE) for the total, RTM, and intervention effects are derived and their asymptotic properties are theoretically studied and verified through simulations. Using NSW data on road fatalities, the total, RTM, and intervention effects are estimated.

The second study considers the bivariate binomial distribution. Due to the dependence structure of the true and error components, subtracting RTM from the total effect does not give an unbiased estimator for the intervention effect. The correlation coefficient can take values in its full range, and RTM inflates comparatively more for negative correlation coefficient values. The Poisson and normal approximations to the binomial distribution underestimate the RTM effect. The MLE of the total, RTM and intervention effects are derived and their asymptotic properties are studied theoretically and verified through simulations. Data on obese individuals and cardboard cans are used to estimate the total, RTM and intervention effects.

Finally, we derive general formulae for the total, RTM and intervention effects under any bivariate distribution, while relaxing potentially restrictive assumptions commonly used in past research. An expression for the total effect is derived in general and decomposed into RTM and intervention effects. Derivation for a p parameter exponential family is separately considered. Examples of some selected bivariate distributions are given for illustrative purposes. Statistical properties of the MLE of the total, RTM, and intervention effects are established theoretically where possible. The proposed and existing methods are compared using data on cholesterol levels by estimating the total, RTM and intervention effects.

FOR OFFICE USE ONLY Date of completion of requirements for Award:

Originality statement

I hereby declare that this submission is -my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.

Signed ....

Date ....

Copyright statement

Authenticity Statement

INCLUSION OF PUBLICATIONS STATEMENT

UNSW is supportive of candidates publishing their research results during their candidature as detailed in the UNSW Thesis Examination Procedure.

Publications can be used in their thesis in lieu of a Chapter if:

- The student contributed greater than 50% of the content in the publication and is the "primary author", ie. the student was responsible primarily for the planning, execution and preparation of the work for publication
- The student has approval to include the publication in their thesis in lieu of a Chapter from their supervisor and Postgraduate Coordinator.
- The publication is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in the thesis

Please indicate whether this thesis contains published material or not.

☐ This thesis contains no publications, either published or submitted for publication (if this box is checked, you may delete all the material on page 2)

Some of the work described in this thesis has been published and it has been documented in the relevant Chapters with acknowledgement (if this box is checked, you may delete all the material on page 2)

☐ *This thesis has publications (either published or submitted for publication) incorporated into it in lieu of a chapter and the details are presented below*

CANDIDATE'S DECLARATION

I declare that:

- I have complied with the Thesis Examination Procedure
- where I have used a publication in lieu of a Chapter, the listed publication(s) below meet(s) the requirements to be included in the thesis.

Name	Signature	Date (dd/mm/yy)
Manzoor Khan		03/03/2020

Abstract

Regression to the mean (RTM) occurs when relatively high or low observations upon remeasurement are found closer to the population mean. When an intervention is applied to subjects selected in the tail of a distribution, an observed mean difference of the pre-post variables is called the total effect. The total effect is the sum of RTM and intervention/treatment effects, and estimation of RTM helps to accurately estimate the intervention/treatment effect.

The first study considers the bivariate Poisson distribution. Formulae for the total effect are derived and decomposed into RTM and intervention effects. The behaviour of RTM is demonstrated for homogeneous and inhomogeneous Poisson processes. Maximum likelihood estimators (MLE) for the total, RTM, and intervention effects are derived and their asymptotic properties are theoretically studied and verified through simulations. Using NSW data on road fatalities, the total, RTM, and intervention effects are estimated.

The second study considers the bivariate binomial distribution. Due to the dependence structure of the true and error components, subtracting RTM from the total effect does not give an unbiased estimator for the intervention effect. The correlation coefficient can take values in its full range, and RTM inflates comparatively more for negative correlation coefficient values. The Poisson and normal approximations to the binomial distribution underestimate the RTM effect. The MLE of the total, RTM and intervention effects are derived and their asymptotic properties are studied theoretically and verified through simulations. Data on obese individuals and cardboard cans are used to estimate the total, RTM and intervention effects.

Finally, we derive general formulae for the total, RTM and intervention effects under any bivariate distribution, while relaxing potentially restrictive assumptions commonly used in past research. An expression for the total effect is derived in general and decomposed into RTM and intervention effects. Derivation for a p parameter exponential family is separately considered. Examples of some selected bivariate distributions are given for illustrative purposes. Statistical properties of the MLE of the total, RTM and intervention effects are established theoretically where possible. The proposed and existing methods are compared using data on cholesterol levels by estimating the total, RTM and intervention effects.

Acknowledgements

I am truly grateful to:

- Allah Almighty for giving me the opportunity, strength, and ability to complete this research project.
- my family for their encouragement, care, moral support, and their never ending love.
- my supervisor Prof Jake Olivier for his valuable and constructive suggestions with patience during the planning and development of this research work, and moral support.
- staff members of the school for their support.
- Dr Damien Wee for his help with writing in Latex at the beginning.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

When repeated measurements are made on the same individual or subject at two different times, relatively high or low observations are likely to be followed by less extreme observations nearer to the true mean. This statistical phenomenon is called *regression to the mean* (RTM), or regression towards the mean, and was first discovered by Sir Francis Galton (1886). Galton noted that parents who were taller than the population mean height had children who were shorter than them, but were closer to the population mean. Similarly, parents who were shorter on average than the population mean height had children whose heights were closer to the population mean.

Figure 1.1 illustrates the RTM phenomenon for standardized height $z = (x - \mu)/\sigma$, where μ and σ are respectively the population mean and standard deviation. The height of parents are in the left or right tails while the height of children are nearer the true mean. The length of the arrow is the RTM effect.

Random error or within subject variability gives rise to RTM. Data without random error is uncommon making RTM a ubiquitous problem in data analysis. Further, the magnitude of RTM is proportional to the measure of dispersion of the random error component (Barnett et al., 2005). RTM can be a group phenomenon where subjects are selected for a study when their baseline measurements are in the extreme of a distribution (Johnson and George, 1991).

RTM has influenced studies in many diverse research areas where repeated mea-

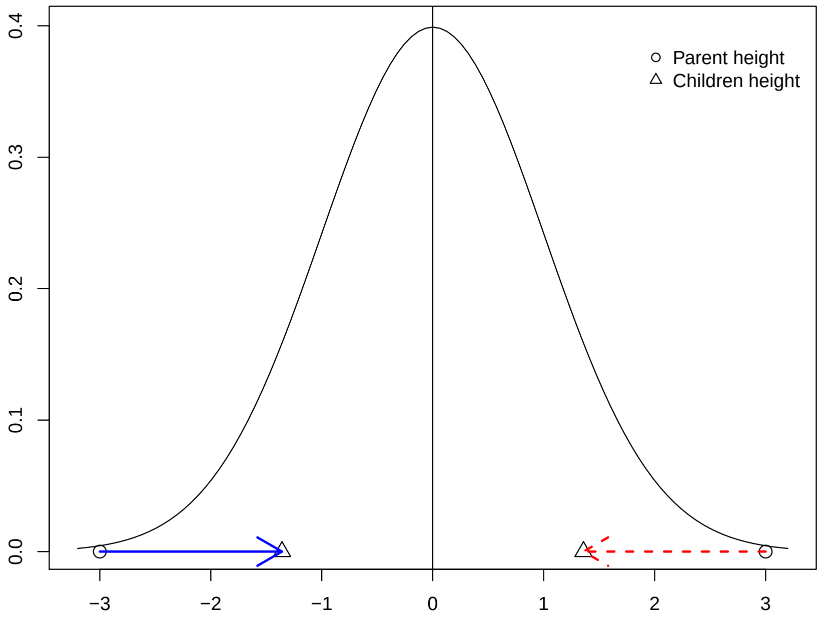

Figure 1.1. Graph of RTM effect for standardized height on parents and their children.

surements or observations are collected. In social psychology, for example, Yu and Chen (2014) provided evidence in favour of the efficacy of social conformity and unrealistic optimism effects, but the effects were no longer evident after controlling for RTM.

In public health, the prevalence of childhood obesity is of serious concern and effective strategies are introduced to prevent or reduce its rate. Burke et al. (2014) evaluated the effectiveness of one such program called HealthMPowers and concluded that it was effective in reducing childhood obesity. Skinner et al. (2015) were critical of the effectiveness of HealthMPowers and demonstrated the apparent change was due to RTM. In another study, Moores et al. (2018) concluded their intervention program, called the Parenting, Eating, and Activity for Child Health (PEACH), was effective on the basis of a statistically significant decrease in standardized BMI and waist scores. Hannon et al. (2018) argued that the conclusion was mistaken as the observed decrease was likely due to the RTM effect.

In economics, unusually rapid economic growth rates are rarely persistent and are often punctuated by a discontinuous drop-off. Thus, forecasting economic growth without accounting for RTM can be misleading (Pritchett and Summers, 2014). In their working paper, Pritchett and Summers warn about the likelihood of a decrease in the current growth rates of Asian giants China and India after accounting for RTM effects.

Other areas where RTM effects have been reported include alcohol consumption (McCambridge et al., 2014), birth weights (Wilcox et al., 1996), blood pressure (Kario et al., 2000), cholesterol (Schectman and Hoffmann., 1988), road crashes (Retting et al., 2003), economic evaluation (Schilling et al., 2017), and sports management decisions (Lee and Smith, 2002). Bland and Altman (1994) and Morton and Torgerson (2005) discuss several examples where RTM could influence statistical inference.

Natural events can induce RTM such as when chronically ill patients only seek treatment when their conditions are at their worst. Likewise, strong enforcement of speed laws and the vaccination of children may only be administered when road crashes and the incidence of tractable disease are at their peaks (Morton and Torgerson, 2003). The effect of RTM can be mitigated through randomization; however, ethical or logistical constraints often limit the ability to randomize participants to control and treatment groups.

1.1 A brief history of RTM

RTM as a concept was the culmination of many years of work by Sir Francis Galton (Stigler, 1997). In his book on heredity genius, Galton (1869) approached the concept by studying the way talent ran in families. For this purpose, Galton selected some notable people including great scientists (e.g., the Bernoullis), musicians (e.g., the Bachs), and their relatives. Galton observed that there was a noticeable propensity for a steady decrease in eminence the further down or up the family tree one moved from the notable person. This phenomenon also appeared to be true with dogs.

Galton made many attempts at explaining his observed ancestral peculiarities. After revisiting the problem over many years, Galton produced a formulation of RTM in terms of the bivariate normal distribution with help from Cambridge mathematician JH Dickson (Galton, 1889). Complete details can be found in Stigler (1986).

RTM has been expressed verbally, mathematically, and geometrically since the early days of its development (Stigler, 1997). Brief descriptions of each of these are given below.

Verbal description

Verbally, RTM can be considered a stochastic time-varying phenomenon, where successive observations are made on the same subject at different times. For instance, consider the number of points scored by basketball team A in two successive matches against the same opposition with an exceptionally high score on the first occasion. Team A would perform less well on a second match relative to the first occasion due to the RTM effect.

A random variable can be often written as the sum of two components: (i) a permanent or true component and (ii) a transient or random/measurement error component. In the basketball context, skill is the permanent component and luck is the error component, and scores are the combination of skill and luck. Thus, for an exceptionally high score, the contributions of skill and luck are both high, whereas, on the second occasion, the skill part persists but the luck part is reduced. The luck component on the second occasion does not become bad luck and can rarely improve, but on average there would be no luck at all. Thus, a transition can occur from skill plus luck on the first occasion to skill alone on the second occasion resulting in a net decrease.

On the other hand, if the first score is exceptionally low, i.e., in the left tail of a distribution, the initial score can perhaps be thought of as the sum of below average skill and bad luck. On the second occasion there would be only below average skill and no luck (bad or good), which would then result in a net increase in the score. The net decrease or increase in scores are due to RTM and not due to a change in the basketball team's skill level.

Mathematical description

Mathematically, RTM can be derived in several equivalent ways. Let X and Y be the respective pre and post variables, and let their joint distribution be the standard bivariate normal with correlation ρ. The probability density function is then

$$f(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}(x^2+y^2-2\rho xy)\right), \quad -\infty < x, y < \infty.$$

After some algebraic manipulation, the conditional distribution of Y given X can be expressed as

$$f(y|x) = \frac{f(x,y)}{f(x)} = \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2}\left(\frac{y-\rho x}{\sqrt{1-\rho^2}}\right)\right), \quad -\infty < y < \infty.$$

Note that $f(y|x)$ is equivalent to the probability density function of a normal distribution, $N(\rho x, 1-\rho^2)$. As the conditional expectation is $E(Y|X=x) = \rho x$ and $\rho < 1$, this implies regression towards 0, the mean of X.

In another mathematical representation, pre and post variables X and Y can be expressed mathematically as

$$X = T + E_1, \qquad Y = T + E_2,$$

where T is the true or permanent component and E_1 and E_2 are the transient or random error components that are mutually independent of each other. For the sake of simplicity, let us suppose T and E_i are identically distributed with $E(T) = E(E_i) = 0$ for $i = 1, 2$. The conditional expectation of Y given $X = x$ can be simplified to

$$\begin{aligned} E(Y|X=x) &= E(T+E_2|T+E_1=x) \\ &= E(T|T+E_1=x). \end{aligned}$$

However, the value x is equivalent to

$$\begin{aligned} x &= E(T+E_1|T+E_1=x) \\ &= 2E(T|T+E_1=x), \end{aligned}$$

and so $E(Y|X = x) = x/2$. Note that no specific distribution was assumed for the pre-post variables and this derivation does not require the existence of second moments. If the correlation exists, we would have $\rho = 1/2$, which is in agreement with the previous formula for the conditional expectation.

Geometrical description

To clarify the concept of RTM geometrically, consider the diagram of the bivariate standard normal distribution with correlation ρ given in Figure 1.2. A cross-sectional slice, perpendicular to the xy plane and parallel to the y axis, is taken at $X = x > 0$. This may represent an unusually high first score if x is in the tail of the distribution of X. The surface is then decapitated parallel to the xy plane, such that the level curve of intersection, which is an ellipse, is exactly tangent to the curve of intersection of the first slice. The resulting diagram is shown in Figure 1.3.

The major and minor axes of the ellipse are formed by the respective red line $Y = X$ and green line $Y = -X$. The black line passing through the origin and the point of tangency of the two curves is the conditional expectation of Y given X, i.e., $E(Y|X = x) = \rho x$. The curve of intersection of the first slice with the surface is the conditional distribution of $Y|X = x$ as depicted in Figure 1.4.

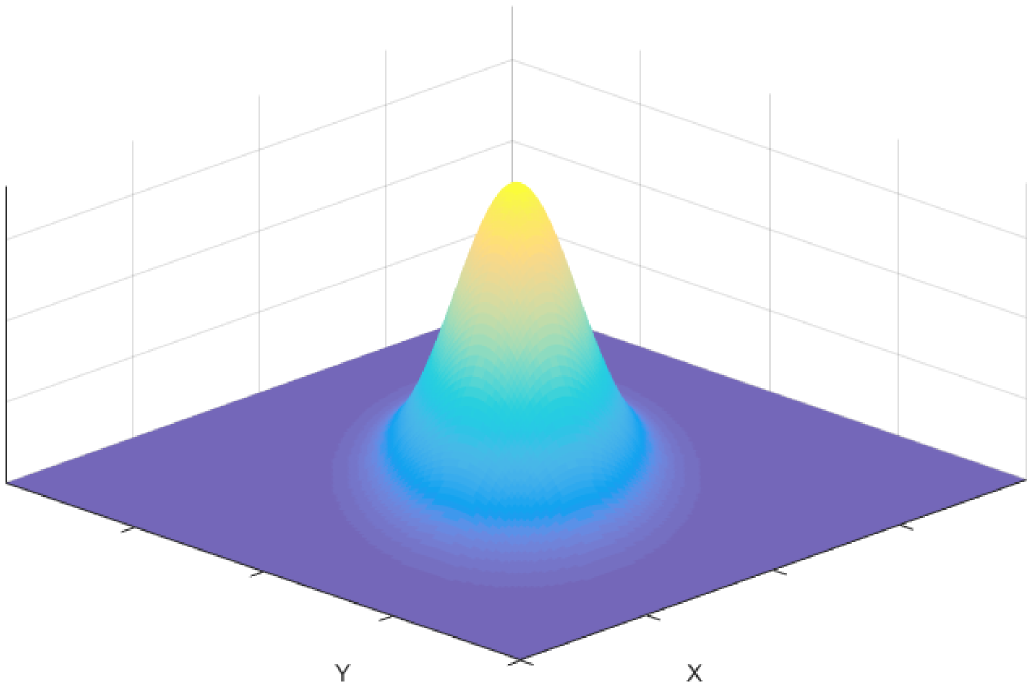

Figure 1.2. Surface of the bivariate normal distribution.

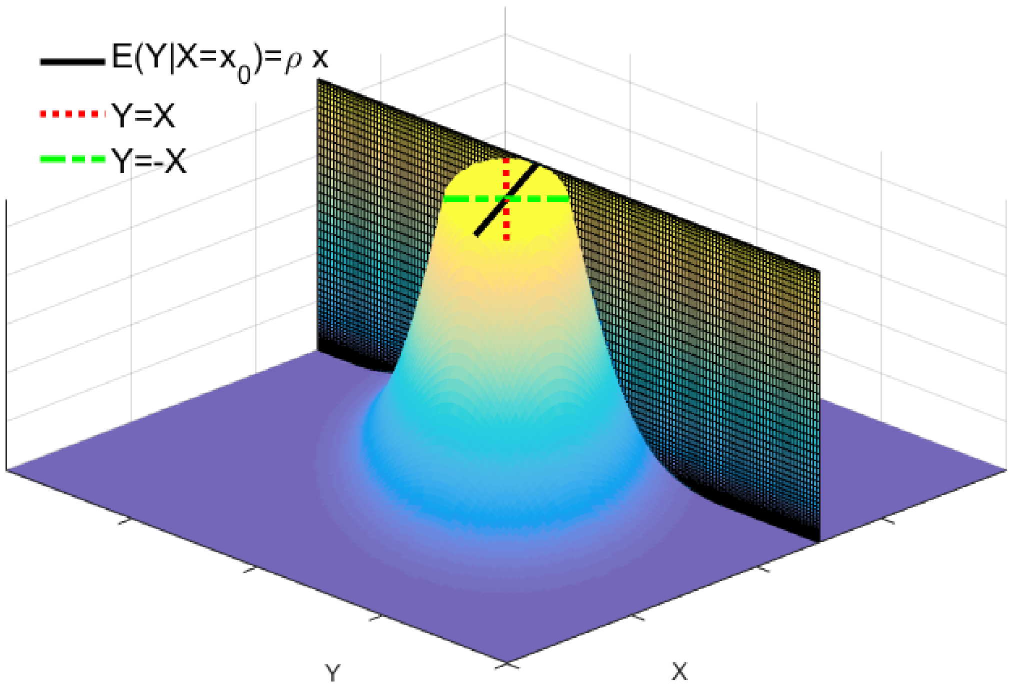

Figure 1.3. Geometric interpretation of RTM for the bivariate normal distribution.

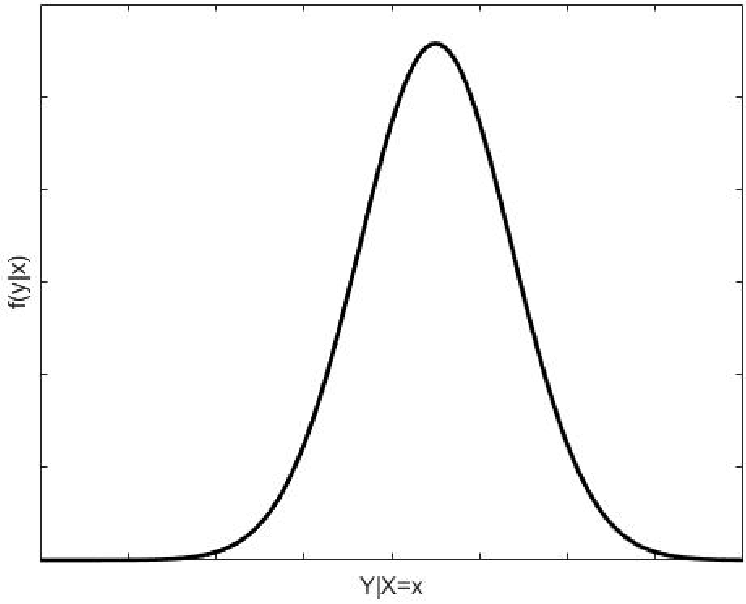

Figure 1.4. Conditional normal distribution of $Y|X = x$

In terms of this diagram, the RTM phenomenon consists of the observation that the line of conditional expectation must be closer to the x-axis than the major axis of the ellipse for RTM to happen. It is clear from the diagram that it is unlikely for the first slice to touch the ellipse at the point the major axis crosses it. This could only happen if X and Y were perfectly positively correlated (i.e., lines of the major axis of ellipse and the conditional expectation coincide), thereby collapsing the ellipse into a line segment. When $\rho = 1$, RTM is zero, and RTM is not zero for all other values of ρ.

1.2 An overview of existing methods

In uncontrolled clinical trials and intervention studies, quantification and estimation of RTM is necessary to accurately estimate treatment or intervention effects. James (1973) and Gardner and Heady (1973) derived RTM formulae for bivariate normally distributed random variables with stationary mean and variance, and strictly positive correlation $\rho > 0$. Davis (1976) extended the derivation of RTM formulae when multiple measurements were taken before applying a treatment to subjects and discussed how this approach was useful in reducing the RTM effect.

James (1973) derived a method of moments estimator for regression to the mean by assuming that the percent of the population in the truncated portion is known. Senn and Brown (1985) improved the derivation by James (1973) and also generalized the maximum likelihood estimator of parameters to various sampling schemes associated with the bivariate normal distribution. Chen and Cox (1992) derived a maximum likelihood estimator of the intervention effect assuming that the pre and post variables were identically distributed and the treatment was designed to change the post measurements in the direction of the population mean.

A key assumption in the above derivations was bivariate normality of the pre and post random variables. Clearly, not all data are normally distributed and this assumption could lead to inaccurate RTM estimation. Das and Mulder (1983) derived a general formula of regression to the mode for arbitrary continuous distributions in a stationary population of subjects. Senn (1990) was critical to using the terminology 'regression to the mode' and argued with the help of examples to use 'regression to the mean' instead. Importantly, the Das and Mulder (1983) method is not directly applicable to empirical distributions as the problem of unidentifiablity of distributions arise (Müller et al., 2003).

Beath and Dobson (1991) derived estimates for regression to the mean for non-normal data based on Edgeworth series and saddlepoint approximations. Edgeworth series approximations may become negative or multimodal for certain values of skewness and kurtosis (Barton and Dennis, 1952), and the saddlepoint approximation is more complicated from a calculation point of view. Both Das and Mulder (1983) and Beath and Dobson (1991) assumed normality for the random error component with zero mean and constant variance. Under fewer assumptions regarding the underlying distribution, Müller et al. (2003) proposed a non-parametric method for estimating the RTM effect. John and Jawad (2010) improved the Das and Mulder (1983) method by making it adaptive to empirical distributions via kernel estimation approaches, while still retaining their error component assumption.

Not all variables are continuous in nature, e.g., count and binary variables which follow discrete probability distributions. The distributions of discrete random variables can be approximated by continuous probability distributions, but under suitable conditions these approximations may be inaccurate. Examples of count variables which are modelled as the Poisson distribution can be found in Anderson (2013), Jones and Smith (2010), and Tse (2014). Similarly, the total number of correct answers in a standardized test, obese individuals in a cohort of fixed size, and absentees in a month may follow the binomial distribution. Along with bivariate Poisson and binomial random variables, formulae for RTM are missing in the literature for other non-normal bivariate random variables.

Existing methods for RTM make assumptions about the direction of the treatment effect and the parameters of the pre-post variables. The effect of treatment is assumed to move the post measurents in the direction of the mean, and pre-post observations are assumed to be identically distributed. Interventions may have effects in the direction opposite of what was intended. Ter Weel (2006), in a study of the Dutch soccer league, found no improvements in team performance after manager turnover. In another study, a negative relationship between employee turnover and performance was observed by Ton and Huckman (2008). Changzheng and Kai (2010) discussed different effects of employee turnover on firm performance including positive, negative and no effects. Thus, an intervention or treatment could change the composition of a population in any direction including away from the population mean.

The derivations of RTM formulae when the pre and post variables follow the bivariate Poisson or bivariate binomial distributions constitute the major part of this research project. When an intervention or treatment is applied to subjects screened on the basis of a cut-off point, the expected difference between pre and post variables is the total effect which is shown to be decomposable into intervention and RTM effects. Notably, in the presence of RTM, the difference in sample means is a biased estimator of the intervention effect, but the bias can then be used to derive an unbiased intervention or treatment effect. To achieve this objective, the other aims of this project are to derive expressions for the total effect and its decomposition into intervention and RTM effects.

Additionally, existing RTM formulae are based on certain, potentially restrictive assumptions including (i) identical distribution of the pre and post variables, (ii) strictly positive correlation, (iii) the direction of the post measurement to change towards the population mean, and (iv) the normal distribution with zero mean and constant variance of the error component. The penultimate goal is to derive formulae for RTM relaxing these assumptions. The last objective is to derive maximum likelihood estimators of the total, intervention and RTM effects and establish their statistical properties of unbiasedness, consistency and asymptotic normality.

1.3 Outline of thesis

This thesis aims to achieve the above mentioned objectives as follows. In Chapter 2, existing methods for estimating and/or mitigating the RTM effect at the design stage of a study and its derivation and estimation in data analyses are discussed. In Chapter 3, we derive RTM formulae for the bivariate Poisson distribution, homogeneous and inhomogeneous bivariate Poisson processes, and their estimators. Chapter 4 is devoted to deriving RTM formulae for the bivariate binomial distribution. Through a comparative study, it is shown that the normal or Poisson approximations to the binomial distribution are not suitable alternatives for quantifying RTM even when the usual conditions of approximation are satisfied. In Chapter 5, general formulae for any bivariate distribution are derived for the total, intervention, and RTM effects. Formulae for the exponential family of distributions are also derived when the vector of sufficient statistics includes the identity function. Chapter 6 concludes the thesis with a discussion, conclusions and directions for future work.

Chapter 2

Existing methods for regression to the mean

RTM is a common phenomenon in repeated data, which may bias the estimated effectiveness of an intervention or treatment (Barnett et al., 2005). Past research has derived formulae to account for RTM, while other research has devised strategies to avoid and/or mitigate the RTM effect at the design or data analysis stages of a pre-post study. A detailed account of existing RTM methods is presented in the following sections.

2.1 RTM reduction method by Ederer (1972)

Ederer (1972) developed a method which can be used to reduce the RTM effect. The author considered the successive observations Y_1, Y_2 and Y_3 on the same subject at time points t_i, such that $Y_i \sim N(\mu_i, \sigma_i^2)$ for $i = 1, 2, 3$. The author planned to select or classify participants on the basis of a cut-off point y_0 at t_1 using the first observation Y_1, i.e., $Y_1 > y_0$ and measure changes from Y_2. In other words, the aim was to explore changes in Y_i from t_2 to t_3 after classifying Y_1 by the cut point y_0.

Let the distribution of successive observations Y_1 and Y_2, and Y_1 and Y_3 be bivariate normal $(Y_1, Y_j) \sim N(\mu_1, \mu_j, \sigma_1^2, \sigma_j^2, \rho_{j1})$ for $j = 2, 3$. Then, Y_j can be expressed in terms of Y_1 as

$$Y_j = \alpha_j + \beta_{j1} Y_1 + E_j, \quad \text{for } j = 2, 3$$

where Y_1 and E_j are independent, $E_j \sim N\left(0, \left(1-\beta_{j1}^2\right)\sigma^2\right)$, α_j and β_{j1} are the intercepts and regression coefficients, for $j = 2, 3$, respectively.

Classifying subjects on $Y_1 > y_0$, the respective expected values of Y_2 and Y_3 for each value of Y_1 are

$$E(Y_2) = \alpha_2 + \beta_{21} y_1 \quad \text{and} \quad E(Y_3) = \alpha_3 + \beta_{31} y_1$$

where $y_1 \in (y_0, \infty)$. Solving the first equation for y_1 and substituting in the latter equation, we get

$$E(Y_3) = \alpha_3 + \frac{\beta_{31}}{\beta_{21}} \left(E(Y_2) - \alpha_2\right).$$

This representation gives the expected change obtained from Y_2 to Y_3 after having classified on $Y_1 > y_0$. The corresponding regression coefficient for Y_3 on Y_2, having classified on Y_1, is given by

$$\beta_{32(1)} = \frac{\beta_{31}}{\beta_{21}} = \frac{\rho_{31}\sigma_3/\sigma_1}{\rho_{21}\sigma_2/\sigma_1} = \frac{\rho_{31}\sigma_3}{\rho_{21}\sigma_2}.$$

Note that $\beta_{32(1)}$ is different from $\beta_{32} = \rho_{32} \cdot \sigma_3/\sigma_2$ which is the regression coefficient of Y_3 on Y_2 without classifying on $Y_1 > y_0$. RTM is zero when $\beta_{32(1)} = 1$, and its magnitude increases as $\beta_{32(1)}$ decreases, and reaches maximum when $\beta_{32(1)} = 0$.

A reduction in the RTM effect in the change from t_2 to t_3 obtained by changing the classification point from t_2 to t_1 implies that $\beta_{32} < \beta_{32(1)}$. Hence, a necessary condition for the reduction in RTM effect under bivariate normality is $\rho_{32} < \rho_{31}/\rho_{21}$.

The complete elimination of RTM implies that $\beta_{32} < \beta_{32(1)} = 1$ or, equivalently, $\rho_{31}\sigma_3 = \rho_{21}\sigma_2$ or $\text{cov}(Y_1, Y_3) = \text{cov}(Y_1, Y_2)$.

2.2 Derivation of RTM under bivariate normality

The existing RTM literature is primarily focused on the bivariate normal distribution, positive correlation, and stationary distributions of the pre-post variables. These assumptions are potentially limiting, although later research has relaxed some but not all assumptions made in earlier studies.

James (1973) and Gardner and Heady (1973) assumed the observed variable was the sum of true and random error components. Let X_i be an observed variable which is composed of the variables X_0 and E_i which are respectively the true and random error components on the i^{th} replication for the same individual, for $i = 1, ..., n$. That is, the model that connects X_i, X_0, and E_i is

$$X_i = X_0 + E_i, \tag{2.1}$$

where X_0 and E_i are independent of each other, X_0 is normally distributed as $N(\mu, \sigma_0^2)$, the error terms E_i are independent and identically distributed as $N(0, \sigma_e^2)$, for $i = 1, ..., n$. As a result, X_i are also identically distributed as $N(\mu, \sigma^2)$, where $\sigma^2 = \sigma_0^2 + \sigma_e^2$.

Many existing methods assume the bivariate normal distribution for derivation of RTM formulae (James, 1973; Gardner and Heady, 1973; Davis, 1976; Johnson and George, 1991). These authors have utilized model (2.1) or some extensions to derive RTM formulae. Details of each author's work are given below.

2.2.1 James (1973)

James (1973) considered successive random variables X_1 and X_2, representing some characteristics on the same subject before and after an intervention. Here, $X_i \sim N(\mu, \sigma^2)$ for $i = 1, 2$, and $\text{cov}(X_1, X_2) = \sigma_0^2$. The joint distribution of X_1 and X_2 is bivariate normal, where $\rho = \sigma_0^2/\sigma^2$ is the correlation of X_1 and X_2.

In clinical or intervention studies, participants with measurements above or below a cut-off or truncation point, say x_0, are selected for treatment or an intervention. Considering only the right cut-off point for demonstrative purposes, the ensuing truncated bivariate normal distribution is

$$f(X_1, X_2 | X_1 > x_0) = \frac{\exp\left(-\frac{1}{1-\rho^2}\left(\left(\frac{x_1-\mu}{\sigma}\right)^2 + \left(\frac{x_2-\mu}{\sigma}\right)^2 - 2\rho\left(\frac{x_1-\mu}{\sigma}\right)\left(\frac{x_2-\mu}{\sigma}\right)\right)\right)}{(1 - \Phi(z_0))\sigma^2\sqrt{1-\rho^2}},$$

where $x_0 < X_1 < \infty, -\infty < X_2 < \infty$, $\Phi(\cdot)$ is the standard normal cumulative distribution function (CDF) and $z_0 = (x_0 - \mu)/\sigma$.

The difference between the conditional means of the identically distributed variables X_1 and X_2 is defined to be RTM, $R(x_0)$, as

$$\begin{aligned} R(x_0) &= E(X_1 - X_2 | X_1 > x_0) \\ &= \int_{x_0}^{\infty} \int_{-\infty}^{\infty} (X_1 - X_2) f(X_1, X_2 | X_1 > x_0) dx_2 dx_1. \end{aligned} \tag{2.2}$$

Assuming a null treatment effect, a model describing the relationship between the pre-post variables can be written as

$$X_2 - \mu = \rho(X_1 - \mu) + E_1,$$

where $E_1|x_1 \sim N(0, (1-\rho^2)\sigma^2)$. James (1973) derived an expression for RTM under bivariate normality as

$$\begin{aligned} R(x_0) &= \frac{\sigma(1-\rho)\phi(z_0)}{1-\Phi(z_0)} \\ &= \frac{\sigma_e^2}{\sqrt{\sigma_0^2+\sigma_e^2}} \cdot \frac{\phi(z_0)}{1-\Phi(z_0)}, \end{aligned} \tag{2.3}$$

where $\phi(\cdot)$ is the standard normal density. James speculated that if a treatment was effective, then it would alter the post measurements in the direction of the pre measurement mean. A model with an effective treatment for the pre-post variables takes the form

$$X_2 - \mu = \gamma\rho(X_1 - \mu) + E_2,$$

where γ is a treatment parameter and $E_2|x_1 \sim N(0, (1-\rho^2)\sigma^2)$. The treatment is considered effective if $\gamma < 1$. Using this model, James (1973) derived a formula for the observed change, assuming the standard bivariate normal distribution as

$$E(Z_1 - Z_2|Z_1 > z_0) = \frac{(1-\gamma\rho)\phi(z_0)}{1-\Phi(z_0)}, \tag{2.4}$$

where $Z_i = (X_i - \mu)/\sigma$ for $i = 1, 2$. James (1973) defined the total proportional reduction TPR in the mean at level x_0 of X_1, due to both RTM and treatment effects as

$$TPR = \frac{x_0 - \gamma\rho x_0}{x_0} = 1 - \gamma\rho.$$

The proportional reduction observed in the mean, due to RTM alone would then be

$$\text{proportional reduction due to RTM} = 1 - \rho.$$

Thus, the proportional reduction due to RTM relative to the total reduction can be obtained as

$$\frac{\text{proportional reduction due to RTM}}{\text{TPR}} = \frac{1-\rho}{1-\rho\gamma}. \tag{2.5}$$

Equation (2.5) helps in decomposing the total change into treatment and RTM effects. However, if $\rho = 0$, that is when the pre-post variables are independent, then equation (2.5) equals one indicating that the percent reduction is solely due to RTM.

2.2.2 Gardner and Heady (1973)

Gardner and Heady (1973) considered the statistical model (2.1) for their derivations, but allowed n replicate measurements on the same subject instead of only two measurements. The distributional assumptions of X_i, X_0, and E_i were retained along with the dependence structure. Unlike the successive observed variables X_i for $i = 1, 2$, Gardner and Heady (1973) considered the joint distribution of the observed variable X_i and the true variable X_0. The joint distribution of X_i and X_0 is assumed bivariate normal, denoted by $f(X_i, X_0)$, with correlation $\mathrm{cor}(X_i, X_0) = \sigma_0/\sqrt{\sigma_0^2 + \sigma_e^2}$.

Gardner and Heady (1973) considered a group of individuals who were in the right tail of the distribution, i.e., $X_i > x_0$. Consequently, the distribution of the observed variable X_i is a univariate truncated normal distribution, $f(X_i|X_i > x_0)$. The expected value of $X_i|X_i > x_0$ is then

$$E(X_i|X_i > x_0) = \mu + \sigma \frac{\phi(z_0)}{1 - \Phi(z_0)}. \tag{2.6}$$

Similarly, the truncated distribution of X_0 given $X_i > x_0$ is

$$f(X_0|X_i > x_0) = \frac{\int_{x_0}^{\infty} f(x_i, x_0)dx_i}{\int_{x_0}^{\infty} f(x_i)dx_i},$$

and the conditional expected value of X_0 is

$$E(X_0|X_i > x_0) = \mu + \frac{\sigma_0^2}{\sigma} \cdot \frac{\phi(z_0)}{1 - \Phi(z_0)}. \tag{2.7}$$

Comparing equations (2.6) and (2.7), it can be verified that $\sigma > \sigma_0^2/\sigma$ when $\sigma_e^2 > 0$. Hence, the mean of the observed values will always be greater than the mean of their true values due to RTM unless $\sigma_e^2 = 0$, which corresponds to the case of perfect correlation between X_i and X_0, i.e., no within subject variability.

Gardner and Heady (1973) also considered taking multiple measurements on the same subject to reduce the RTM effect. Let the number of replicated measurements on the same subject be n, and the selection of subjects for treatment be based on the average $\overline{X} = \sum_{i=1}^{n} X_i/n$. Then, it can be shown that

$$E(\overline{X}|\overline{X} > x_0) = \mu + \sqrt{\sigma_0^2 + \sigma_e^2/n} \cdot \frac{\phi(z_{0n})}{1 - \Phi(z_{0n})}, \tag{2.8}$$

and

$$E(X_0|\overline{X} > x_0) = \mu + \frac{\sigma_0^2}{\sqrt{\sigma_0^2 + \sigma_e^2/n}} \cdot \frac{\phi(z_{0n})}{1 - \Phi(z_{0n})}, \tag{2.9}$$

where the selection criterion is based on the cut-off point x_0, i.e., $\overline{X} > x_0$, and $z_{0n} = (x_0 - \mu)/\sqrt{\sigma_0^2 + \sigma_e^2/n}$. From equation (2.8), it is clear that as n increases the conditional expected value of the sample mean of observed values decreases. Whereas, the expected value of the true variable increases as n increases in equation (2.9). Thus, for large n, there would be little difference between the mean of the observed and true values conditioned on the event $\overline{X} > x_0$. Mathematically, as $n \to \infty$, $\sigma_e^2/n \to 0$, or equivalently,

$$\lim_{n\to\infty} E(\overline{X}|\overline{X} > x_0) = \lim_{n\to\infty} E(X_0|\overline{X} > x_0) = \mu + \sigma_0 \frac{\phi(z_{0n})}{1 - \Phi(z_{0n})}.$$

To demonstrate this fact graphically, let the parameters of the bivariate normal distribution be $\mu = 5$, $\sigma_0^2 = 0.3$, and $\sigma_e^2 = 0.7$ and the right cut-off point be $x_0 = 7$. Equations (2.8) and (2.9) are plotted in Figure 2.1 for different values of n. As the value of n increases, the gap between the curves of the observed and true expected values decreases. For multiple measurements on each individual, the RTM formula was obtained by subtracting equation (2.8) from equation (2.9) as

$$\begin{aligned} R(x_0) &= \left(\sqrt{\sigma_0^2 + \sigma_e^2/n} - \frac{\sigma_0^2}{\sqrt{\sigma_0^2 + \sigma_e^2/n}} \right) \cdot \frac{\phi(z_{0n})}{1 - \Phi(z_{0n})} \\ &= \frac{\sigma_e^2/n}{\sqrt{\sigma_0^2 + \sigma_e^2/n}} \cdot \frac{\phi(z_{0n})}{1 - \Phi(z_{0n})}. \end{aligned} \tag{2.10}$$

For $n = 1$, equation (2.10) reduces to equation (2.3) from James (1973).

2.2.3 Davis (1976)

When planning an intervention study, it may be helpful to reduce the RTM effect through the study design. To formulate this, Davis (1976) revisited the work done by Gardner and Heady (1973) and Ederer (1972) which were based on

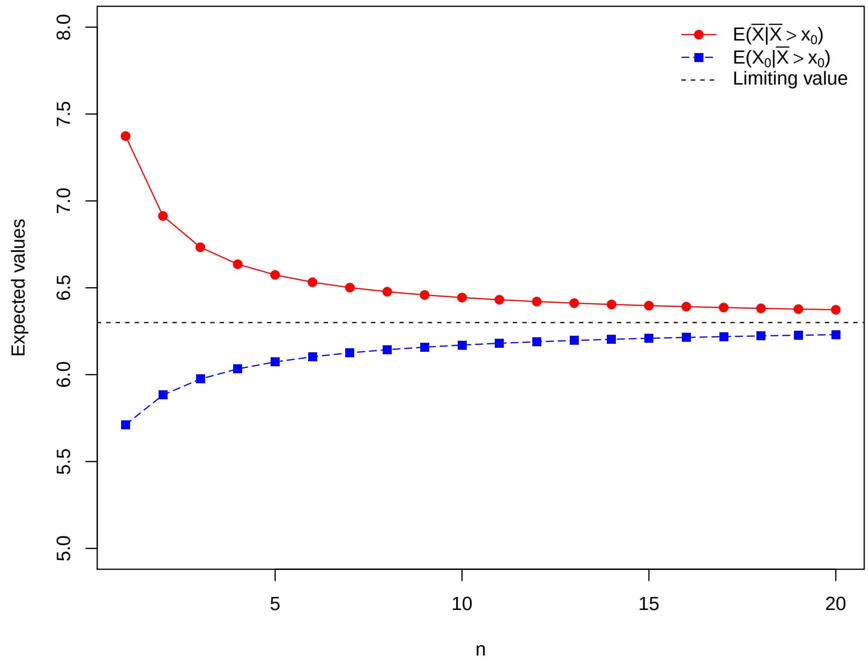

Figure 2.1. Expected values of the observed and true variables as function of n

taking two and two or more measurements on the same subject, respectively. The author considered the same statistical model given in equation (2.1). As before, let $\overline{X}$ be the sample mean of multiple observations taken on the same subject such that $\overline{X} \sim N(\mu, \sigma_0^2 + \sigma_e^2/n)$. Let X^* be the subsequent observation such that $X^* \sim N(\mu, \sigma_0^2 + \sigma_e^2)$. Then, the correlation between $\overline{X}$ and X^* is

$$\rho^* = \frac{\sigma_0^2}{\sqrt{(\sigma_0^2 + \sigma_e^2/n)(\sigma_0^2 + \sigma_e^2)}}.$$

The expectations of $\overline{X}$ and X^* conditioned on a cut-off point x_0 are given by

$$E(\overline{X}|\overline{X} > x_0) = \mu + \sqrt{\sigma_0^2 + \sigma_e^2/n} \cdot \frac{\phi(z_{0n})}{1 - \Phi(z_{0n})}, \tag{2.11}$$

and

$$E(X^*|\overline{X} > x_0) = \mu + \frac{\sigma_0^2}{\sqrt{\sigma_0^2 + \sigma_e^2/n}} \cdot \frac{\phi(z_{0n})}{1 - \Phi(z_{0n})}. \tag{2.12}$$

A formula for RTM can be obtained by subtracting equation (2.12) from equation (2.11) as

$$\begin{aligned} R(x_0, n) &= E(\overline{X} - X^*|\overline{X} > x_0) \\ &= \frac{\sigma_e^2/n}{\sqrt{\sigma_0^2 + \sigma_e^2/n}} \cdot \frac{\phi(z_{0n})}{1 - \Phi(z_{0n})}, \end{aligned} \tag{2.13}$$

which is the same result derived by Gardner and Heady (1973).

The magnitude of RTM can be reduced by taking multiple measurements on the same subject before applying an intervention as $R(x_0, n)$ is a decreasing function of n. RTM as a function of n is shown in Figure 2.2 for values $\mu = 5$, $\sigma_0^2 = 0.3$, $\sigma_e^2 = 0.7$ and $x_0 = 7$. The asymptotic value of $R(x_0, n)$ is depicted by the dashed horizontal line at zero. RTM decreases steeply for the first four to five measurements, but afterwards the decrease in not substantial.

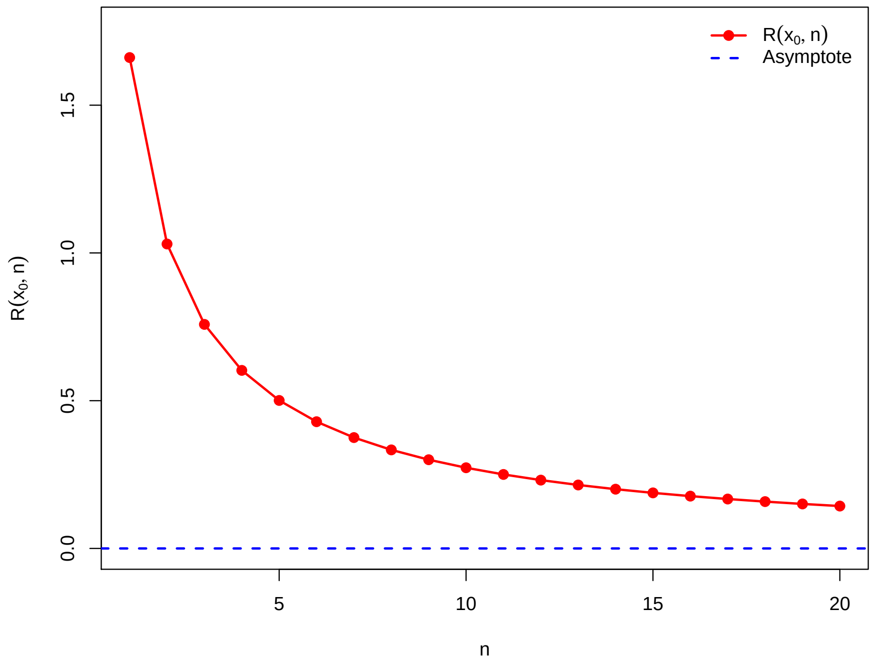

Figure 2.2. RTM as a function of n for $\sigma_0^2 = 0.3$, $\sigma_e^2 = 0.7$ and $x_0 = 7$.

Another way to reduce RTM can be accomplished by using the first observation X_1 as a baseline measurement for classification purposes, i.e., selecting a subject on the basis of the event $X_1 > x_0$, and the second observation X_2 on the same subject as the baseline from which the treatment effect can be evaluated. For example, cholesterol can be classified according to one baseline observation and measuring the change from another baseline point several weeks later (Ederer, 1972). Let X_3 be the post intervention measurements. Davis (1976) translated this mathematically in terms of conditional expectations as

$$E(X_2|X_1 > x_0) = \mu + \rho_{12} \cdot \sigma \frac{\phi(z_0)}{1 - \Phi(z_0)}, \tag{2.14}$$

and

$$E(X_3|X_1 > x_0) = \mu + \rho_{13} \cdot \sigma \frac{\phi(z_0)}{1 - \Phi(z_0)}, \tag{2.15}$$

where ρ_{12} and ρ_{13} are correlation coefficients, defined as $\mathrm{cor}(X_1, X_2) = \rho_{12}$ and $\mathrm{cor}(X_1, X_3) = \rho_{13}$. The resulting RTM formula is

$$\begin{aligned} R(x_0, \rho_{12}, \rho_{13}) &= E(X_2 - X_3|X_1 > x_0) \\ &= (\rho_{12} - \rho_{13}) \cdot \sigma \frac{\phi(z_0)}{1 - \Phi(z_0)}. \end{aligned} \tag{2.16}$$

The RTM effect will be zero when $\rho_{12} = \rho_{13}$, and if the difference $(\rho_{12} - \rho_{13})$ is small, multiple measurements may not be required for reducing the RTM effect.

2.2.4 Johnson and George (1991)

Gardner and Heady (1973) assumed that independent measurement errors are the only source of variability in repeated observations. In practice, there may be many factors which can influence within subject variability, such as the subject's emotional state at the time of measuring blood pressure and biological variation (Musini and Wright, 2009). Johnson and George (1991) extended model (2.1) to include a subject effect, S_i, as

$$Y_i = X_0 + S_i + E_i \quad \text{for } i = 1, \ldots, m, \tag{2.17}$$

where S_i and S_j may be correlated $\mathrm{cor}(S_i, S_j) = \rho_s > 0$ for $i \neq j$, but independent of X_0 and E_i, and $S_i \sim N(0, \sigma_s^2)$. The assumptions of model (2.1) regarding X_0 and E_i are retained in the derivation.

Johnson and George (1991) argued that model (2.1) would be appropriate if measurement error was the only attributable source of variability when repeated measurements were taken under identical conditions. However, measurements taken at different times under different conditions would lead to within subject fluctuations, and model (2.17) would be appropriate. For example, if Y_1 and Y_2 denote a successive characteristic measured on parents and offspring, respectively, then S_1 and S_2 as defined in model (2.17) may be correlated genetic effects with correlation coefficient $\rho_s < 1$. Even in the absence of measurement errors, S_i components will lead to within subject variability, for $i = 1, 2$, and would ultimately induce RTM.

Under model (2.17), the respective formulae for correlation between the successive variables and RTM are

$$\mathrm{cor}(Y_1, Y_2) = \frac{\sigma_0^2 + \rho_s \sigma_s^2}{\sigma_0^2 + \sigma_s^2 + \sigma_e^2},$$

and

$$R(y_0) = \frac{\sigma_0^2 + (1 - \rho_s)\sigma_s^2}{\sqrt{\sigma_0^2 + \sigma_s^2 + \sigma_e^2}} \cdot \frac{\phi(z_1)}{1 - \Phi(z_1)}, \tag{2.18}$$

where $z_1 = (y_0 - \mu)/\sqrt{\sigma_0^2 + \sigma_s^2 + \sigma_e^2}$.

Suppose for each subject in a study, repeated measurements are taken at m different times, and at each time, n replicate measurements are taken. Then, model (2.17) can be written as

$$Y_{ij} = X_0 + S_i + E_{ij}, \qquad \text{for} \quad i = 1, \ldots, m, \quad j = 1, \ldots, n, \tag{2.19}$$

where, as before, $X_0 \sim N(\mu, \sigma_0^2)$ and also $S = (S_1, \ldots, S_m) \sim N(\mathbf{0}, \Sigma)$, $\mathbf{0} = (0, \ldots, 0)^T$, Σ is compound symmetric as

$$\Sigma = \sigma_u^2 \begin{pmatrix} 1 & \rho_s & \cdots & \rho_s \\ & 1 & \cdots & \rho_s \\ & & \ddots & \vdots \\ & & & 1 \end{pmatrix},$$

$E_{ij} \sim N(0, \sigma_e^2)$ and are independent of X_0 and S. Let the sample mean $\overline{Y} = \sum_{i=1}^{m} \sum_{j=1}^{n} y_{ij}/(nm)$ be used to classify subjects for inclusion in the study to administer treatment if $\overline{Y} > y_0$. Then

$$\begin{aligned} \mathrm{var}(\overline{Y}) &= \sigma_{\bar{y}}^2 \\ &= \sigma_0^2 + \frac{\sigma_s^2}{m}(1 + (m-1)\rho_s) + \frac{\sigma_e^2}{mn}. \end{aligned} \tag{2.20}$$

Let Y^* be the observation after administering treatment, then the respective formulae for the correlation of Y^* and $\overline{Y}$, and RTM are

$$\operatorname{cor}(\overline{Y}, Y^*) = \frac{\sigma_0^2 + \rho_s \sigma_s^2}{\sqrt{\left(\sigma_0^2 + \sigma_s^2/m\,(1 + (m-1)\rho_s) + \sigma_e^2/mn\right)\left(\sigma_0^2 + \sigma_s^2 + \sigma_e^2\right)}},$$

and

$$R_T(y_0) = \frac{\left((1-\rho_s)\sigma_s^2 + \sigma_e^2/n\right)/m}{\sigma_{\overline{y}}} \cdot \frac{\phi(z_2)}{1 - \Phi(z_2)}, \tag{2.21}$$

where the subscript T stands for total in $R_T(y_0)$ and $z_2 = (y_0 - \mu)/\sigma_{\overline{y}}$. The measurement error and subject effect both contribute to RTM, and their individual contributions can be obtained by decomposing the total RTM, $R_T(y_0)$, as

$$\begin{aligned} R_T(y_0) &= R_S(y_0) + R_E(y_0) \\ &= \frac{(1-\rho_s)\sigma_s^2/m}{\sigma_{\overline{y}}} \cdot \frac{\phi(z_2)}{1-\Phi(z_2)} + \frac{\sigma_e^2/(mn)}{\sigma_{\overline{y}}} \cdot \frac{\phi(z_2)}{1-\Phi(z_2)}. \end{aligned} \tag{2.22}$$

If $R_T(y_0)$ and $R_E(y_0)$ are known, then $R_S(y_0)$ can be estimated without knowing the correlation structure among subjects. The measurement error component $R_E(y_0)$ of RTM can be reduced by either increasing the number of repeated measurements m, and/or by increasing the number of replications n of each measurement, whereas increasing the number of repeated measurements m at different times is the only option for reducing $R_S(y_0)$. However, in practice, it may not be possible in terms of cost and time to take a large number of measurements to classify subjects for administering treatment to them. So, a reasonable alternative is to replicate measurements at a given time to reduce RTM attributable to measurement error before applying a treatment to subjects on the basis of the condition $\overline{Y} > y_0$. Under these conditions, equations (2.20) and (2.21) for the variance of $\overline{Y}$ and RTM, respectively, simplify to

$$\operatorname{var}(\overline{Y}) = \sigma_0^2 + \sigma_s^2 + \sigma_e^2/n,$$

and

$$R_T(y_0) = \frac{(1-\rho_s)\sigma_s^2 + \sigma_e^2/n}{\sqrt{\sigma_0^2 + \sigma_s^2 + \sigma_e^2/n}} \cdot \frac{\phi(z_2)}{1-\Phi(z_2)}. \tag{2.23}$$

Taking the limit of equation (2.2), $R_T(y_0)$ reduces to $R_S(y_0)$ as

$$\lim_{n\to\infty} R_T(y_0) = \frac{(1-\rho_s)\sigma_s^2}{\sqrt{\sigma_0^2 + \sigma_s^2 + \sigma_e^2/n}} \cdot \frac{\phi(z_2)}{1-\Phi(z_2)} = R_S(y_0). \tag{2.24}$$

Thus, it is impossible to completely eliminate the RTM effect by increasing the number of replications unless $\rho_s = 1$.

2.3 Estimation of RTM under bivariate normality

Let $(x_{11}, x_{21}), (x_{12}, x_{22}), \ldots, (x_{1n}, x_{2n})$ be a random sample of pairs of observations of size n from a truncated bivariate normal distribution. James (1973) used the method of moments to estimate μ, σ^2, ρ, and γ. The percent of the population in the truncated portion c_0 was assumed to be known such that $c_0 = \phi(z_0)/(1 - \Phi(z_0)$. The obtained estimates are

$$\begin{aligned}
\hat{\mu} &= \overline{x}_1 - c_0\hat{\sigma}, \\
\hat{\sigma}^2 &= \frac{s^2_{x_1}}{c_0(x_0 - c_0) + 1} \\
\hat{\rho} &= \left[b^2\left(c_0(x_0 - c_0) + 1\right) - \frac{s^2_{x_2}}{\hat{\sigma}^2} + 1\right]^{1/2} \\
\hat{\gamma} &= \frac{\hat{\beta}}{\hat{\rho}},
\end{aligned}$$

where $\overline{x}_1 = \sum_{i=1}^n x_{1i}/n$, $s_j^2 = \sum_{i=1}^n (x_{ji} - \overline{x}_j)^2/n$, for $j = 1, 2$, and the estimated slope of X_2 on X_1 is $\hat{\beta} = \sum_{i=1}^n (x_{1i} - \overline{x}_1)(x_{2i} - \overline{x}_2)/\sum_{i=1}^n (x_{1i} - \overline{x}_1)^2$. The respective variances of estimates $\hat{\sigma}$, $\hat{\mu}$, $\hat{\rho}$ and $\hat{\gamma}$ are

$$\begin{aligned}
\text{var}(\hat{\sigma}) &= \frac{\sigma^4_{x_1}}{2\sigma^2\left(c_0(x_0 - c_0) + 1\right)^2(n-1)}, \\
\text{var}(\hat{\mu}) &= \frac{c_0^2\sigma^2_{x_1}}{n} + \frac{\sigma^4_{x_1}}{2\sigma^2\left(c_0(x_0 - c_0) + 1\right)^2(n-1)} \\
\text{var}(\hat{\rho}) &= \frac{(c_0(x_0 - c_0) + 1)^2\gamma^2\sigma^2}{\sum_{i=1}^n (x_{1i} - \overline{x}_2)^2} + \frac{\sigma^4_{x_2}}{2\rho^2\sigma^4(n-1)}\left[1 + \frac{\sigma^4_{x_1}}{\sigma^4\left(c_0(x_0 - c_0) + 1\right)^2}\right] \\
\text{var}(\hat{\gamma}) &= \frac{\sigma^2\left(1 + \left(c_0(x_0 - c_0) + 1\right)\gamma^4\right)}{\rho^2\sum_{i=1}^n (x_{1i} - \overline{x}_1)^2} + \frac{\gamma^2\sigma^4_{x_2}}{2\rho^4\sigma^4(n-1)}\left[1 + \frac{\sigma^4_{x_1}}{\sigma^4\left(c_0(x_0 - c_0) + 1\right)^2}\right].
\end{aligned}$$

Cohen (1955) derived maximum likelihood estimators (MLE) for the bivariate normal distribution, but Senn and Brown (1985) argued that they cannot be used for two reasons. First, the marginal distributions of X and Y are identical and second, Cohen did not allow for a treatment effect.

Senn and Brown (1985) relaxed the assumption of known percent of the population in the truncated portion, to derive the MLE by writing the likelihood function in the form

$$L(\boldsymbol{\theta}, \boldsymbol{x}) = \exp\left(-\frac{1}{2}\frac{\sum(x_i-\mu)^2}{\sigma^2}\right) \Big/ \left((2\pi\sigma^2)^{n/2} \times (1-\Phi(z_0))^n\right)$$
$$\times \exp\left(-\frac{1}{2}\frac{\sum[(y_i-\mu)-\gamma\rho(x_i-\mu)]^2}{\sigma^2(1-\rho^2)}\right) \Big/ \left(2\pi\sigma^2(1-\rho^2)\right)^{n/2},$$

where $\boldsymbol{\theta} = (\mu, \sigma^2, \rho, \gamma)$ and

$$\boldsymbol{x} = \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \cdots & \cdots \\ x_n & y_n \end{pmatrix}.$$

The MLE of $\boldsymbol{\theta}$ were then obtained by numerical methods. Additionally, Senn and Brown (1985) corrected the expression of variances derived by James (1973) using the method of moments estimation. However, the authors did not study the statistical properties of the derived estimators for RTM.

2.4 RTM for non-normal populations

2.4.1 Das and Mulder (1983)

There are many variables of interest in health, educational and social sciences which do not follow the normal distribution (Bono et al., 2017). Das and Mulder (1983) considered a statistical model for pre-post non-normal variables as

$$Y_i = W + e_i \qquad \text{for } i = 1, 2, \tag{2.25}$$

where Y_i are the observed, identically distributed variables with stationary means μ and variances σ^2, and are jointly distributed with a positive correlation coefficient, $0 < \rho_d < 1$, W is the true component which is arbitrarily distributed with density function $f(w)$, mean μ, and variance $\rho_d\sigma^2$, while the e_i are normally distributed random errors with mean zero and variance $(1-\rho_d)\sigma^2$, for $i = 1, 2$. Further, W and e_i are mutually independent of each other.

Let the respective common density functions of e_i and Y_i be $h(e)$ and $g(y)$, for $i = 1, 2$. Expressing $g(y)$ as the convolution of $f(w)$ and $h(e)$, we have

$$g(y) = \int_{-\infty}^{\infty} f(w)h(y-w)dw, \qquad -\infty < w < \infty. \tag{2.26}$$

Differentiating equation (2.26) with respect to y, we have

$$\frac{dg(y)}{dy} = -\frac{1}{(1-\rho_d)\sigma^2}\int_{-\infty}^{\infty}(y-w)f(w)h(y-w)dw. \tag{2.27}$$

Using definition (2.2) and equation (2.25), it can be shown that an equivalent form for the derivation of RTM $R(y)$ is

$$R(y) = E(Y_1 - Y_2|Y_1 = y) = E(e_1|Y_1 = y). \tag{2.28}$$

The conditional density of $e_1|Y_1 = y$ is

$$h(e_1|Y_1 = y) = \frac{h(e_1)f(y-e_1)}{g(y)}, \qquad -\infty < e_1 < \infty, \tag{2.29}$$

where $h(e)f(y-e)$ is the joint density of e_1 and Y_1. By evaluating the conditional expectation in equation (2.28), Das and Mulder (1983) derived a formula for RTM by utilizing equations (2.27) and (2.29) as

$$R_d(y) = -\sigma^2(1-\rho_d)\frac{d\log\left(g(y)\right)}{dy}. \tag{2.30}$$

When subjects are selected on the basis of the event $Y_1 > y_0$, the formula for RTM can be found by evaluating the conditional expectation $E(e_1|Y_1 > y_0)$. Using equation (2.30), Das and Mulder (1983) simplified the conditional expectation to

$$R_d(y_0) = \sigma^2(1-\rho_d)\cdot\frac{g(y_0)}{1-G(y_0)}, \tag{2.31}$$

where $G(y)$ is the distribution function and y_0 is the cut-off point. Note that this method is not directly applicable to empirical distributions (Beath and Dobson, 1991; John and Jawad, 2010).

2.4.2 Beath and Dobson (1991)

Beath and Dobson (1991) were motivated to estimate $g(x)$ and $G(x)$ from Das and Mulder (1983), and hence RTM, for empirical non-normal distributions using Edgeworth and saddlepoint approximations. Beath and Dobson (1991) considered the same model in equation (2.25) and retained the associated assumptions from Das and Mulder (1983), to obtain the RTM formula as

$$R_d(y_0) = (1-\rho_d)\sigma^2\cdot\frac{g(y_0)}{1-G(y_0)}.$$

Note that the notations used in Das and Mulder (1983) have been retained here with the same interpretation. It is well known that a distribution can be expressed in terms of the normal probability function and its derivative as

$$h(s) = \sum_{j=0}^{\infty} c_j H_j(s)\phi(s), \tag{2.32}$$

where H_j is the j^{th} hermite polynomial, c_j are constants determined by $h(s)$ (Kendall et al., 1987) and $\phi(\cdot)$ is the standard normal probability distribution function. Expressing equation (2.26) in terms of $\phi(\cdot)$, we get

$$g(y) = \frac{1}{\Delta}\int_{-\infty}^{\infty} f(w)\phi\left(\frac{y-w}{\Delta}\right)dw, \qquad -\infty < w < \infty, \tag{2.33}$$

where $\Delta = \sigma\sqrt{(1-\rho_d)}$ is used for brevity. Let $\theta = \sigma\sqrt{\rho_d}$, then substituting $s = (w-\mu)/\theta$ and $t = (y-\mu)/\theta$ into equation (2.33), we get

$$g(\theta t+\mu) = \frac{\theta}{\Delta}\int_{-\infty}^{\infty} h(s)\phi\left(\frac{\theta}{\Delta}(t-s)\right)ds, \tag{2.34}$$

where $h(s)$ is the standard density function of s. Substituting equation (2.32) in (2.34), using the result of Erdélyi (1954) and simplifying, we get

$$g(y_0) = \frac{1}{\sigma}\phi\left(\frac{y_0-\mu}{\sigma}\right)\sum_{i=0}^{\infty} c_i\left(\frac{\theta}{\sigma}\right)^i H_i\left(\frac{y_0-\mu}{\sigma}\right). \tag{2.35}$$

Using the result, $\int H_i(v)\phi(v)dv = -H_{i-1}(v)\phi(v)$ for $i \geq 1$, an expression for $G(y)$ can be obtained as

$$G(y_0) = \Phi\left(\frac{y_0\mu}{\sigma}\right) - \phi\left(\frac{y_0-\mu}{\sigma}\right)\sum_{i=1}^{\infty} c_i\left(\frac{\theta}{\sigma}\right)^i H_{i-1}\left(\frac{y_0-\mu}{\sigma}\right). \tag{2.36}$$

As $h(s)$ is in standard measure with zero mean and unit variance, the corresponding values of constants c_i, for $i = 0,1,2,3,4$, are $1,0,0,\gamma_1/6$ and $\gamma_2/24$, respectively (Kendall et al., 1987), where γ_1 and γ_2 are the respective coefficient of skewness and kurtosis of the distribution. Truncating the series in equation (2.35) and (2.36), approximations for $g(y)$ and $G(y)$ based on the Edgeworth series becomes

$$g(y_0) = \frac{1}{\sigma}\phi(z_3)\left(1+\frac{\gamma_1}{6}H_3(z_3)+\frac{\gamma_2}{24}H_4(z_3)+\frac{\gamma_1^2}{72}H_6(z_3)\right),$$

and

$$G(y_0) = \Phi(z_3) - \phi(z_3)\left(\frac{\gamma_1}{6}\left(\frac{\theta}{\Delta}\right)^3 H_2(z_3)+\frac{\gamma_2}{24}\left(\frac{\theta}{\Delta}\right)^4 H_3(z_3)+\frac{\gamma_1^2}{72}\left(\frac{\theta}{\Delta}\right)^6 H_5(z_3)\right),$$

where $z_3 = (y_0 - \mu)/\sigma$. These approximations can then be used for calculating $R_d(y_0)$.

The Edgeworth series may result in negative approximations or multi-modality for certain values of skewness and kurtosis (Barton and Dennis, 1952). To overcome these limitations, the method of saddlepoint approximation by Daniels (1954) can be used to estimate $R_d(y_0)$. In this approach, the probability distribution function $g(y)$ is approximated as

$$g(y) = \frac{\exp\left(K(t_0) - t_0 y\right)}{\sqrt{2\pi K''(t_0)}},$$

where $K(t)$ is the cumulant generating function and $K'(t_0) = y$. Using Easton and Ronchetti (1986), $K(t)$ can be approximated by

$$\widetilde{K}(t) = \mu t + \frac{\sigma^2 t^2}{2} + \frac{k_3 t^3}{6} + \frac{k_4 t^4}{24},$$

where $k_3 = \theta^3 \gamma_1$ and $k_4 = \theta^4 \gamma_2$ are the respective third and fourth cumulants and γ_1 and γ_2 are as defined earlier. This approximation requires normalization of $g(y)$ to be a probability density function. $G(y)$ is then numerically integrated to complete the estimation of RTM. This method is more complicated than others from calculation a point of view.

2.4.3 John and Jawad (2010)

As mentioned earlier, the Das and Mulder (1983) method cannot be applied directly to empirical distributions for estimation of RTM. John and Jawad (2010) aimed at making Das and Mulder's method data adaptive via kernel density estimation and kernel estimation approaches for the hazard rate function. Consider the formula derived by Das and Mulder (1983) for RTM as

$$R(y_0) = (1 - \rho_d)\sigma^2 \cdot \frac{g(y_0)}{1 - G(y_0)}.$$

Let $\widehat{g}_h(y)$ be a kernel density estimator of $g(y)$ and let $\widehat{G}(y)$ be the empirical distribution function. Then, the kernel estimator density function for the initial values X_{1i}, for $i = 1, \ldots, n$, is defined as

$$\widehat{g}_h(x) = \frac{1}{n} \sum_{i=1}^{n} K_h(x - x_{1i}), \tag{2.37}$$

where $K_h(\cdot) = K(\cdot|h)/h$ is the kernel function and h is the bandwidth or smoothing parameter. The mean integrated square error (MISE) measures the estimation error of $\widehat{g}_h(x)$ as

$$MISE(h) = E\int_{-\infty}^{\infty}(\widehat{g}_h(x) - g(x))^2 dx.$$

Asymptotic analysis of the MISE provides simple insight into how the bandwidth works as a smoothing parameter. MISE is asymptotically approximated by the asymptotic mean integrated square error (AMISE) as

$$AMISE(h) = R(K)/nh + h^4 R(g'')\left(\int_{-\infty}^{\infty} x^2 K(x)dx/2\right)^2,$$

where

$$R(\psi) = \int_{-\infty}^{\infty} \psi(x)dx.$$

An optimum value, h_{AMISE}, that minimizes $AMISE(h)$ provides a good approximation to h_{MISE}, which also minimizes $MISE(h)$ and can be calculated as

$$h_{AMISE} = \left(\frac{R(K)}{nR(g'')(\int_{-\infty}^{\infty} x^2 K(x)dx/2)^2}\right)^{1/5}.$$

Jones et al. (1996) reviewed the first and the second generation methods for optimal bandwidth selection methods and suggested that the Sheather and Jones (1991) method is stable and consistent among the existing methods, including the rule of thumb (Läuter, 1988), least squares cross-validation (Bowan, 1984; Rudemo, 1982; Hall and Marron, 1991), biased cross-validation (Scott and Terrell, 1987), solve the equation plug in approach (Sheather and Jones, 1991), and bootstrap (Faraway and Jhun, 1990; Taylor, 1989). The Sheather and Jones (1991) method chooses the bandwidth that is a solution of the fixed point equation

$$h_{AMISE} = \left(\frac{R(K)}{nR(g''_{f(h)})(\int_{-\infty}^{\infty} x^2 K(x)dx/2)^2}\right)^{1/5},$$

where $f(h)$ is the pilot bandwidth. John and Jawad (2010) used this bandwidth for the kernel density estimation of $g(x)$, which is the default method in many statistical software packages.

The problem of estimating RTM can be simplified to kernel based estimation methods of the hazard function. The *hazard function* $u(x)$ is the ratio of the probability density function $g(x)$ to the survival function $S(x) = 1-G(x)$. The hazard function can be written as $u(x) = g(x)/(1-G(x))$ when there is no censoring for the variable X_{1i} for $i = 1, ..., n$. Ramlau-Hansen (1983) generalized kernel estimators for the hazard function and studied asymptotic properties of the generalized form of estimators. A kernel estimator for u based on the generalization of Ramlau-Hansen (1983) takes the form

$$\widehat{u}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x - X_{1(i)}}{h}\right)\frac{1}{n-i+1}, \tag{2.38}$$

where $X_{1(i)}$ are the ordered observations of X_{1i}.

Researchers have suggested various methods for optimal bandwidth selection. These methods include the maximum likelihood cross validation method (Tanner and Wong, 1984), least square cross validation method (Cao et al., 1994; Patil, 1993; Sarda and Vieu, 1991), and bootstrap method (Gonzàlez-Manteiga et al., 1996) which are based on fixed-bandwidth fixed-kernel methods for estimating the hazard function. But, for unevenly distributed data over the range of interest, the degree of smoothness achieved via a fixed-bandwidth method will not be uniform. This non-adaptive behaviour of fixed bandwidth estimators can be fixed by using the varying bandwidth estimator as suggested by Muller and Wang (1994).

An alternative approach incorporates the idea of the nearest neighbour into the definition of bandwidth (Olaf and Holger, 1992). The boundary effects near the endpoints in the domain of the hazard function are not taken into account by the fixed kernel estimators, which can be fixed by changing the kernels at the boundary (Hougaard, 1988; Hougaard et al., 1989; Hall and Wehrly, 1991). Hess et al. (1999) conducted a simulation study and found that Muller and Wang (1994) and Olaf and Holger (1992) are advantageous over other existing methods for estimating the hazard function via kernel estimators. These two methods are also available in the `R` package called `muhaz`.

2.4.4 Müller et al. (2003)

In the preceding models for non-normal populations, the random error component was assumed to be normally distributed with zero mean and fixed variance. With a goal of predicting the true value X from the observed value Y, Müller et al.

(2003) allowed the random error component δ to be arbitrarily distributed within the model

$$Y_i = X_i + \delta_i, \quad \text{for} \quad i = 1, \ldots, n, \tag{2.39}$$

where X_i and δ_i are independent, X_i are distributed with a common density function $f_X(x)$, and δ_i also have a common density function $f_\delta(x)$ with zero mean and fixed variance. Let $f_Y(y)$ be the probability density function of the observed variable, then by convolution of $f_X(x)$ and $f_\delta(x)$, we get

$$f_Y(y) = \int_{-\infty}^{\infty} f_X(x) f_\delta(y-x)dx, \tag{2.40}$$

and the joint density $f_{Y,X}(y,x)$ of Y and X is

$$f_{Y,X}(y,x) = f_\delta(y-x)f_X(x). \tag{2.41}$$

Müller et al. (2003) aimed at predicting X from Y. The Bayes estimator, $E(X|Y)$, can be used to achieve this goal, which is also the best linear unbiased predictor. This leads us to the following RTM function

$$E(X|Y=y_0) = \frac{\int_{-\infty}^{\infty} x f_\delta(y_0-x)f_X(x)dx}{\int_{-\infty}^{\infty} f_\delta(y_0-x)f_X(x)dx}. \tag{2.42}$$

Difficulty arises in solving the right hand side of equation (2.42) when neither $f_X(x)$ nor $f_\delta(x)$ are contained in a parametric family of distributions. This problem can be addressed by using a non-parametric method.

The following assumptions were made by Müller et al. (2003) about $f_X(x)$ and $f_\delta(x)$ to complete their derivation. Both the functions $f_X(x)$ and $f_\delta(x)$ are twice continuously differentiable, and $f_\delta(x)$ is given by

$$f_\delta(x) = \frac{1}{\sigma_e}\psi\left(\frac{x}{\sigma_e}\right),$$

where σ_e is the standard deviation of δ, and the density function $\psi(\cdot)$ satisfies the moment conditions,

$$\int_{-\infty}^{\infty} x\psi(x)dx = 0,$$
$$\int_{-\infty}^{\infty} x^2\psi(x)dx = \mu_2 = 1$$
$$\int_{-\infty}^{\infty} x^3\psi(x)dx = \mu_3 < \infty.$$

Substituting $f_\delta(x)$ in equation (2.42), we get

$$\begin{aligned}E(X|Y=y_0) &= \int_{-\infty}^{\infty} x\frac{1}{\sigma_e}\psi\left(\frac{y_0-x}{\sigma_e}\right) f_X(x)dx \Bigg/ \int_{-\infty}^{\infty} \frac{1}{\sigma_e}\psi\left(\frac{y_0-x}{\sigma_e}\right) f_X(x)dx \\ &= \int_{-\infty}^{\infty} (y_0-\sigma_e z)\psi(z) f_X(y_0-\sigma_e z)dz \Bigg/ \int_{-\infty}^{\infty} \psi(z) f_X(y_0-\sigma_e z)dz.\end{aligned} \tag{2.43}$$

Using Taylor expansion along with the results obtained by differentiating the moments as $\int \psi'(x)dx = 0$, $\int x\psi'(x)dx = -\int \psi(x)dx = -1$, $\int x^2\psi'(x)dx = 0$, and $\int x^3\psi'(x)dx = -3\mu_2$, where $\psi' = d\psi/dx$, then equation (2.43) simplifies to

$$E(X|Y=y_0) = y_0 + \sigma_e^2 \frac{f_Y'(y_0)}{f_Y(y_0)} + \frac{1}{2}\sigma_e^3\mu_3\frac{f_Y''(y_0)}{f_Y(y_0)} + o(\sigma_e^3), \tag{2.44}$$

where $f' = df/dx$, and $f'' = d^2f/d^2x$. If $\mu_3 = 0$, then the leading remainder term is $\sigma_e^4(3\mu_3 - \mu_4)f_Y^3(y_0)/((y_0))$. The multivariate version of equation (2.44) is

$$E(X|Y=\boldsymbol{y_0}) = \boldsymbol{y_0} + V\frac{\nabla f_Y(\boldsymbol{y_0})}{f_Y(\boldsymbol{y_0})} + o(V^{3/2}), \tag{2.45}$$

where V is the $p \times p$ covariance matrix of δ.

Choi and Hall (1999) introduced a data sharpening method for density estimation. The relationship of local moments and local sample moments to the data sharpening method was formulated by Müller and Yan (2001). The mean update mode finding algorithm (Fukunaga and Hostetler, 1975) implicitly uses a special case of the local sample mean which can be useful to derive non-parametric RTM.

For some $p \geq 1$, let a random vector $\boldsymbol{Z}$ be the starting point for the local moment with density function f_Z which is twice continuously differentiable. Let $\boldsymbol{y_0} = (y_{01}, \ldots, y_{0p})$ be an arbitrary point such that $\boldsymbol{y_0} \in \mathbb{R}^p$, and let $\gamma = \gamma_n > 0$ be a sequence of window widths such that $\gamma \to 0$ as $n \to \infty$. Then, a sequence of local neighbourhoods S is defined as

$$S = S_n = \prod_{j=1}^{p}[y_{0j} - \gamma, y_{0j} + \gamma].$$

The local mean $\boldsymbol{\mu}_Z = (\mu_{z_1}, \ldots, \mu_{z_p})$ at $\boldsymbol{y_0}$ is defined as

$$\mu_{z_j} = \lim_{\gamma\to 0}\frac{1}{\gamma^2}E\left((\boldsymbol{Z}-\boldsymbol{y_0})^{e_j}|\boldsymbol{Z}\in S\right), \tag{2.46}$$

where $\boldsymbol{e_j} = (0, \ldots, 1, \ldots, 0)$ and 1 occurs in the j^{th} position. According to Müller and Yan (2001), μ_{z_j} can be written as

$$\mu_{z_j} = \frac{1}{3} D^{e_j} f_Z(\boldsymbol{y_0}) / f_Z(\boldsymbol{y_0}) \tag{2.47}$$

where

$$D^{\alpha} = \frac{\partial^{\alpha_1 + \cdots + \alpha_p}}{\partial^{\alpha_1} y_{01} \cdots \partial^{\alpha_p} y_{0p}} \quad \text{and} \quad \alpha = \alpha_1 + \cdots + \alpha_p.$$

Let $\boldsymbol{Z}_i = (z_{i1}, \ldots, z_{ip})$ be an $\mathbb{R}^p$ valued random sample of size n from the distribution $f_{\boldsymbol{Z}}$, then the local sample mean is $\hat{\boldsymbol{\mu}}_z = (\hat{\mu}_{z_1}, \ldots, \hat{\mu}_{z_p})$, with $\hat{\mu}_{z_j}$ defined as

$$\hat{\mu}_{z_j} = \frac{1}{\gamma^2} \sum_{i=1}^{n} (Z_{ij} - y_{0j}) \Bigg/ \sum_{i=1}^{n} I(\boldsymbol{Z}_i), \qquad j = 1, \ldots, p, \tag{2.48}$$

where $I(\boldsymbol{Z}_i) = 1$ if $\boldsymbol{Z}_i \in S$, $I(\boldsymbol{Z}_i) = 0$ if $\boldsymbol{Z}_i \notin S$. Using results from Müller and Yan (2001), we have

$$\hat{\boldsymbol{\mu}}_z = \frac{1}{3} \frac{\nabla f_Z(\boldsymbol{y_0})}{f_Z(\boldsymbol{y_0})} + o_p\left((n\gamma^{2+p})^{-1/2}\right). \tag{2.49}$$

The covariance matrix can be estimated by the sample covariance matrix from the observation with repeated measurements $(Y_{ik1}, \ldots, Y_{ikp})$, for $1 \le i \le n$ and $\le k \le m_i$ as

$$\widehat{V} = \left(\tfrac{1}{n} \textstyle\sum_{k=1}^{m_i} (Y_{ikr} - \overline{Y}_{i.r})(Y_{iks} - \overline{Y}_{i.s})\right)_{rs}, \qquad 1 \le r,\ s \le p, \tag{2.50}$$

where $\overline{Y}_{i.s} = \sum_{k=1}^{m_i} Y_{ikr}/m_i$, $2 \le m_i$, and $1 \le r \le p$. Moreover, using equation (2.49) and observations with repeated measurements, the mean vector μ_Y can be estimated as

$$\hat{\boldsymbol{\mu}}_Y = \frac{1}{3} \frac{\nabla f_Y(\boldsymbol{y_0})}{f_Y(\boldsymbol{y_0})} + o_p\left((n\gamma^{2+p})^{-1/2}\right). \tag{2.51}$$

Substituting equations (2.50) and (2.51) into equation (2.45), we get a non-parametric estimate of RTM as

$$\hat{E}(X|Y = \boldsymbol{y_0}) = \boldsymbol{y_0} + 3\widehat{V}\hat{\boldsymbol{\mu}}_Y. \tag{2.52}$$

2.5 Study designs to mitigate RTM

The design of a study can help mitigate the RTM effect in intervention studies (Yudkin and Stratton, 1996; Linden, 2013). Some well known study designs and their potential effect on RTM are described in the following subsections.

2.5.1 Randomized control trials

Random allocation of subjects to treatment groups (e.g., placebo and treatment) can minimize selection bias, and help to balance the influence of RTM across groups. The mean change in the placebo group gives an estimate of the RTM effect. The treatment effect can then account for RTM by finding the difference of the mean change in the treatment group and the mean change in the placebo group. However, randomization is not always possible due to ethical and/or logistical constraints.

2.5.2 Regression discontinuity designs

When randomization is not possible, an alternative approach is the regression discontinuity design (Lee and Lemieux, 2010; Linden and Adams, 2012). In this approach, subjects are assigned to a treatment group on the basis of a pre intervention continuous cut-off point. Subjects to the right and left of the cut-off point are assumed to be exchangeable and can be classified into control and treatment groups depending on the study. As subjects do not have precise control over their assignment score, and are unaware of the value of the cut-off point, they cannot self-select into the treatment groups. Thus, RTM would equally effect both the groups in the neighbourhood of the cut-off point (Linden, 2013).

2.5.3 Two measurements approach of Ederer (1972)

The Ederer (1972) method for mitigating RTM consists of taking two measurements on each subject before applying an intervention. The first measurement is used for selecting subjects, and the second measurement is used as a baseline from which the treatment effect is assessed. Assuming that RTM has happened between the first and second measurement, the mean change measured from baseline is the intervention effect. Denke and Frantz (1993) used this approach for mitigating the RTM effect in their study to assess the relationship between the starting level of cholesterol and response to treatment. In spite of adjusting for RTM using the Ederer (1972) approach, Denke and Frantz (1993) found that subjects with hypercholesterolemia were more diet-responsive than subjects with lower cholesterol levels.

2.5.4 Selection based on multiple measurements

The RTM formula in equation (2.3) indicates that RTM is proportional to measurement variability. Selecting two or more baseline measurements reduces the measurement variability. The study selection criterion can then be based on the average of multiple measurements, assuming the RTM effect has taken place between the first and later measurements (Gardner and Heady, 1973; Davis, 1976). Reducing variability can then be used to get a better estimate of the true component of each subject before applying an intervention. The choice of taking multiple measurements depends on the cost of obtaining them, and is not always an executable option when resources are limited.

2.6 Accounting for RTM through data analysis

As stated earlier, ethical and/or logistical constraints may not allow us to conduct a randomized control trial. Also, multiple measurements may not be feasible due to the time and/or cost of obtaining those observations. Consequently, observational data may be collected where the pre intervention observations were baseline measurements. In this situation, several techniques may be considered to account for RTM as described below.

2.6.1 Matching techniques

When only retrospective observational data are available, matching techniques (Stuart, 2010) can be used to create a control group for comparison purposes. Based on the observed pre intervention characteristics (especially one which may lead to RTM) of the treatment group, an analyst tries to replicate the randomization process to create a control group. The effect of the treatment can then be evaluated by comparing the treatment and control groups. Matching techniques allow the analyst to assess how well the pre-intervention variable overlaps in distribution between groups using graphical or numerical diagnostics (Stuart, 2010). A higher degree of overlap in the distribution increases our confidence that RTM is effectively controlled for, as we would expect in a randomized control trial. However, the farther the cut-off point is in the tail of the distribution, the more difficult it would be to create a control group on the basis of pre-intervention characteristics, because of the smaller fraction of available subjects.

2.6.2 Analysis of covariance

Analysis of covariance (ANCOVA) is the most common analytic approach used to account for RTM. In this procedure, each subject's follow-up measurement is adjusted according to their baseline measurements by including the pre-intervention measurements as a covariate in the model (Twisk, 2003). The following regression equation summarizes the approach

$$y_i = \alpha + \beta_1(x_i - \bar{x}) + \beta_2 G_i + \varepsilon_i,$$

where $G_i = 1$ for the treatment group and zero otherwise and the corresponding regression coefficient β_2 is the treatment effect adjusted for RTM, x_i is the baseline measurement with $\bar{x} = \sum_{i=1}^{n} x_i/n$ its sample mean, y_i is the follow-up measurement and $\epsilon_i \sim N(0, \sigma^2)$ for $i = 1, 2, ..., n$. Additionally, each subject's pre measurement score x can be adjusted using an RTM correction factor (Irwig et al., 1991; Trochim, 2001), and the adjusted baseline score x_{adj} can then be used in an ANCOVA model. A subject's adjusted score is

$$x_{adj} = \overline{x} + \rho(x - \overline{x}),$$

where ρ is the correlation of pre-post variables in the treatment group. However, the ANCOVA assumptions, for instance, linearity between outcome and covariates and normality, may not be valid. Moreover, there is no assurance that the treatment groups are comparable on all baseline covariates (Linden, 2013).

2.6.3 Subtracting the estimated RTM effect

Finally, the simplest approach to account for RTM is to subtract the estimated RTM effect from the total effect in the treatment group (Barnett et al., 2005). Appropriate statistical methods developed for this approach for the bivariate Poisson and binomial distributions will be discussed in Chapters 3 and 4, and generalised to any bivariate distribution in Chapter 5.

Chapter 3

Regression to the mean for the bivariate Poisson distribution

In this chapter, regression to the mean (RTM) formulae are derived assuming the bivariate Poisson distribution and for both homogeneous and inhomogeneous Poisson processes. The asymptotic distributions of RTM estimators have been derived and statistical properties of derivations have been evaluated through a simulation study. The total effect for the number of people killed in road crashes in different regions of New South Wales (Australia) is estimated and decomposed into the RTM and intervention effects using maximum likelihood. The contents of this chapter are reproduced from a published paper (Khan and Olivier, 2018), and the contents and notation have been slightly modified.

3.1 Introduction

RTM can influence inference about the effectiveness of an intervention/treatment applied to subjects in the tail of a distribution. Accounting for RTM can improve estimation of treatment or intervention effects, thereby assisting the researcher in drawing appropriate conclusions. The formulae for calculating the expected RTM effect based on an assumption of bivariate normality are well known in the literature (James, 1973; Gardner and Heady, 1973; Davis, 1976).

There are situations in which the underlying distribution may not be continuous and can be modelled as a Poisson distribution such as counting processes or rates

(Anderson, 2013; Jones and Smith, 2010; Tse, 2014). Further, Poisson processes can be categorized according to the homogeneity of their means. If the average arrival rate of a Poisson process is time/location invariant, then it is referred to as homogeneous. On the other hand, a Poisson process with a time/location varying average arrival rate is called an inhomogeneous Poisson process.

In pre-post study designs, interventions are implemented to change the rate of occurrence (Chaspari et al., 2014; Ruggeri and Sivaganesan, 2005). The change in the rate of occurrence due to RTM may be erroneously attributed to the intervention. However, the quantification of the RTM effect is missing in the literature when the underlying distribution in pre/post studies is bivariate Poisson. Therefore, in this chapter we derive expressions to quantify the RTM effect for the bivariate Poisson distribution and extend the results to both the homogeneous and inhomogeneous Poisson processes.

The remainder of this chapter is comprised of ten sections. In Section 2, the total effect and its relation with RTM and intervention effect is discussed and exemplified with the help of the bivariate normal distribution in Section 3. Formulae for the total effect assuming the bivariate Poisson distribution are derived in Section 4. Section 5 discusses the decomposition of the total effect into RTM and intervention effects and the results are extended to homogeneous and inhomogeneous Poisson processes in Section 6. The maximum likelihood estimation of the total, RTM and treatment/intervention effects and their asymptotic distributions are discussed in Section 7. A simulation study is carried out to investigate the statistical properties of the sample RTM effect and its probability distribution in Section 8. The RTM effect for the number of fatalities in NSW road crashes is calculated using maximum likelihood in Section 9. Estimation of RTM for log-transformed Poisson distributed data is investigated in Section 10. The chapter concludes with a discussion in Section 11.

3.2 The total, RTM and intervention effects

In clinical or intervention studies, patients/subjects with measurements above or below a cut-off or truncation point, say x_0, are selected for treatment or an intervention. Without loss of generality, only right cut-off points are presented here. Let X_1 and X_2 be some characteristic on the same subject before and after an intervention. The joint distribution of pre/post measurements for a truncated bivariate distribution is given by

$$f(X_1, X_2|X_1 > x_0) = \frac{f(X_1, X_2)}{f(X_1 > x_0)} \qquad \text{where} \qquad x_0 < X_1 < \infty, -\infty < X_2 < \infty.$$

The total effect $T(x_0, \boldsymbol{\theta})$ can be obtained by evaluating the conditional expectation of the difference of pre and post variables

$$\begin{aligned} T(x_0, \boldsymbol{\theta}) &= E(X_1 - X_2|X_1 > x_0) \\ &= \int_{x_0}^{\infty} \int_{-\infty}^{\infty} (X_1 - X_2) f(X_1, X_2|X_1 > x_0) dx_2 dx_1, \end{aligned} \tag{3.1}$$

where $\boldsymbol{\theta}$ is the parameter vector. Similarly, an expression of $T(x_0, \boldsymbol{\theta})$ for a bivariate discrete distribution can be obtained using equation (3.1) by replacing integrals with summations.

The total effect, $T(x_0, \boldsymbol{\theta})$, could be partially or totally due to RTM, depending on the effectiveness or non-effectiveness of an intervention effect. Thus, when X_1 and X_2 are identically distributed, or equivalently the intervention effect is zero $E(X_1) = E(X_2)$, then the difference of the conditional means of X_1 and X_2 is defined to be the RTM effect, denoted by $R(x_0, \boldsymbol{\theta})$,

$$R(x_0, \boldsymbol{\theta}) = E\left(X_1 - X_2|X_1 > x_0, E(X_1) = E(X_2)\right). \tag{3.2}$$

Let $\delta(\boldsymbol{\mu}) = E(X_1 - X_2)$ be the intervention effect, then $T(x_0, \boldsymbol{\theta})$ can be expressed as

$$T(x_0, \boldsymbol{\theta}) = R(x_0, \boldsymbol{\theta}) + \delta(\boldsymbol{\mu}),$$

where $\boldsymbol{\mu} = (\mu_1, \mu_2)^T \subseteq \boldsymbol{\theta}$ and $E(X_i) = \mu_i$ for $i = 1, 2$.

3.3 An example: the bivariate normal distribution

To exemplify the total, RTM and intervention effects and express $T(x_0, \boldsymbol{\theta})$ as the sum of $R(x_0, \boldsymbol{\theta})$ and $\delta(\boldsymbol{\mu})$, let the random variables $X_1 = X_0 + e_1$ and $X_2 = X_0 + e_2$ represent successive measurements of some characteristics on the same subject before and after an intervention. It is assumed that X_0 represents true measurements and is distributed normally as $N(\mu, \sigma_0^2)$, whereas e_1 and e_2 are random errors/fluctuations, identically distributed $N(0, \sigma_e^2)$ and independent of X_0. Thus, the resulting distributions of both X_1 and X_2 are $N(\mu, \sigma^2)$, where $\sigma^2 = \sigma_0^2 + \sigma_e^2$ and $\mathrm{cov}(X_1, X_2) = \sigma_0^2$. The joint distribution of X_1 and X_2 is bivariate normal, where $\rho = \sigma_0^2/\sigma^2$ is the correlation of X_1 and X_2. However, when subjects for an intervention are selected on the basis of a cut-off point, then the joint distribution of X_1 and X_2 is the truncated bivariate normal.

Using the joint truncated bivariate normal distribution, James (1973) derived an expression for the RTM effect by evaluating the conditional difference between the means of pre and post random variables (with identical marginal distribution) as

$$R(x_0, \boldsymbol{\theta}) = \frac{\sigma(1-\rho)\phi(z_0)}{1-\Phi(z_0)},$$

where ϕ and Φ are the standard normal density and distribution functions respectively, and $z_0 = (x_0 - \mu)/\sigma$.

This derivation can be extended to allow for unequal group means where $X_i \sim N(\mu_i, \sigma^2)$ for $i = 1, 2$. Under this set up, an expression for $T(x_0, \boldsymbol{\theta})$ can be shown to be

$$T(x_0, \boldsymbol{\theta}) = \frac{\sigma(1-\rho)\phi(z_0)}{1-\Phi(z_0)} + (\mu_1 - \mu_2),$$

where $\delta(\boldsymbol{\mu}) = \mu_1 - \mu_2$, is the intervention effect. The James (1973) formula is a special case where $\mu_1 = \mu_2 = \mu$ and hence, $\delta(\boldsymbol{\mu}) = 0$.

The influence of the covariance of X_1 and X_2 and the choice of cut-off point on the RTM effect is illustrated in Figures 3.1 for the bivariate normal distribution. As the correlation ρ increases, the RTM effect decreases (top-left panel) while the opposite effect is observed when the random error component σ_e^2 increases (top-right panel). As the cut-off point z_0 moves away from mean of the distribution on either side, the resulting RTM increases symmetrically (bottom-left panel).

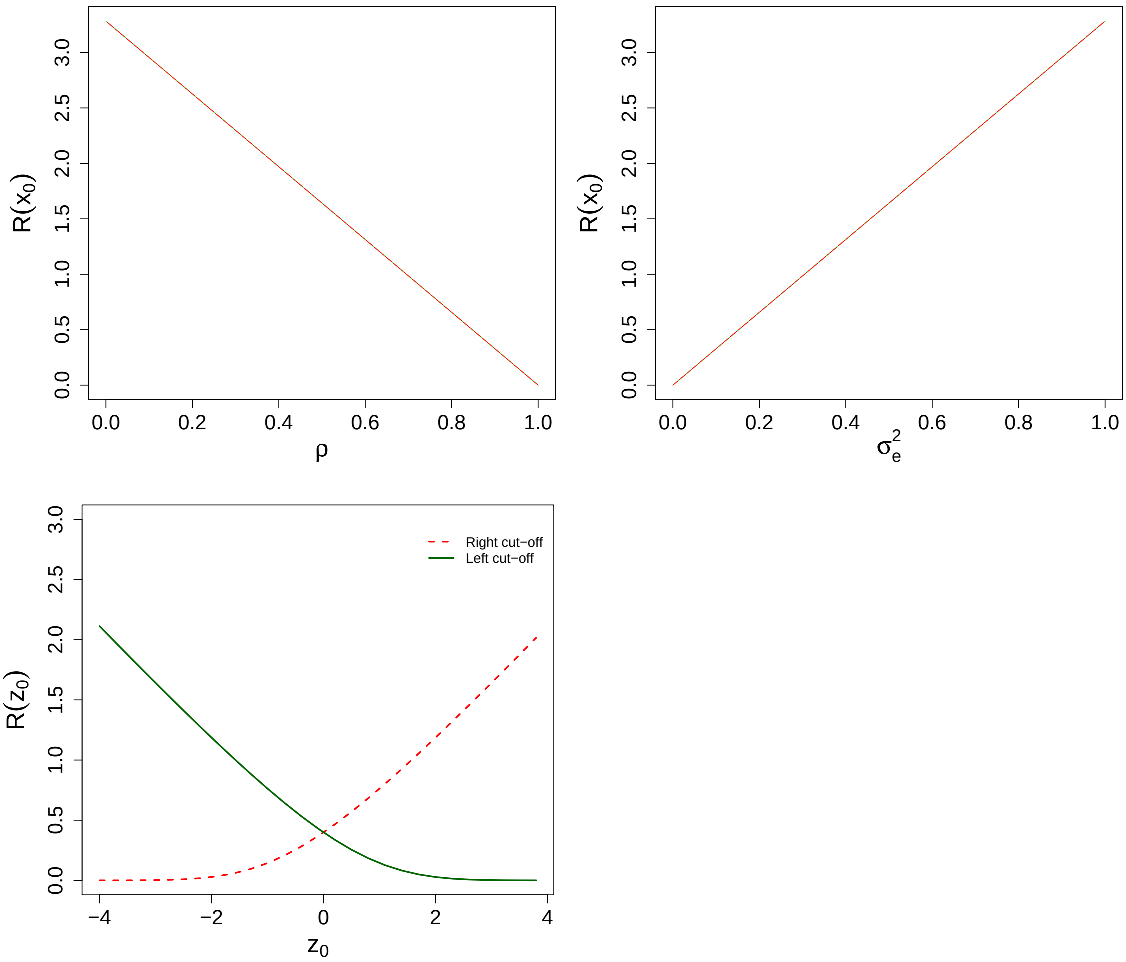

Figure 3.1. Top-left panel: Graph of the RTM effects as function of covariance/correlation, Top-right panel: Graph of the RTM as function of random error component, Bottom-left panel: Graph of RTM as a function of cut-off points when the underlying distribution is standard normal.

3.4 Total effect and the Bivariate Poisson Distribution

The normal distribution is one of the most important continuous probability distribution since it provides the basis for statistical inference in a large number of studies. Likewise, the Poisson distribution has numerous applications when the variable of interest is discrete. Count variables which are functions of time and/or space are commonly modelled as Poisson processes. Additionally, other discrete probability distributions are well approximated by the Poisson distribution. Formula to quantify the RTM effect can be derived under a bivariate Poisson distribution assumption using an approach similar to the bivariate normal distribution.

Let Y_1 and Y_2 be two random variables representing the successive number of occurrences of the same phenomenon. Define $Y_1 = X_0 + X_1$ and $Y_2 = X_0 + X_2$, where the random variable X_0 is the true number of occurrences and X_1 and X_2 represent random fluctuations or counting errors. Here, X_0, X_1 and X_2 are independent Poisson random variables each with parameter (rate of occurrence) θ_i for $i = 0, 1, 2$. Y_i is then Poisson distributed with parameter $\theta_0 + \theta_i$ for $i = 1, 2$. The bivariate Poisson distribution of Y_1 and Y_2, first discussed by Campbell (1934), is

$$P(Y_1 = y_1, Y_2 = y_2) = e^{-(\theta_0+\theta_1+\theta_2)} \frac{\theta_1^{y_1}}{y_1!} \frac{\theta_2^{y_2}}{y_2!} \sum_{x_0=0}^{\min(y_1,y_2)} x_0! \left(\frac{\theta_0}{\theta_1\theta_2}\right)^{x_0} \binom{y_1}{x_0}\binom{y_2}{x_0}.$$

The covariance of Y_1 and Y_2 is θ_0, so the correlation is

$$\mathrm{cor}(Y_1, Y_2) = \frac{\theta_0}{\sqrt{(\theta_0+\theta_1)(\theta_0+\theta_2)}}.$$

An intervention or treatment may be applied to extreme situations based on some cut-off point say, y_0. Depending on the problem under study, the cut-off point can be in either the left or right tail of the distribution. Due to the asymmetric shape of the Poisson distribution, right and left cut-off points are considered separately.

3.4.1 Case 1: Right cut-off point

Suppose an intervention is decided on the basis that the initial count Y_1 was greater than some cut-off value y_0, then the truncated joint probability distribution of Y_1 and Y_2 is given by

$$P(Y_1 = y_1, Y_2 = y_2 | Y_1 > y_0) = \frac{e^{-(\theta_0+\theta_1+\theta_2)}}{1 - P(Y_1 \leqslant y_0)} \frac{\theta_1^{y_1}}{y_1!} \frac{\theta_2^{y_2}}{y_2!} \sum_{x_0=0}^{\min(y_1,y_2)} x_0! \left(\frac{\theta_0}{\theta_1\theta_2}\right)^{x_0} \binom{y_1}{x_0}\binom{y_2}{x_0}.$$

Let $T_r(y_0;\boldsymbol{\theta})$ denote the total effect as defined in equation (3.1) for $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2)$. Consider the conditional expectation of $Y_1|Y_1 > y_0$ for $y_1 \leq y_2$,

$$E(Y_1|Y_1 > y_0) = \frac{e^{-(\theta_0+\theta_1+\theta_2)}}{1 - P(Y_1 \leqslant y_0)} \sum_{y_1=y_0+1}^{\infty} \sum_{y_2=x_0}^{\infty} y_1 \frac{\theta_1^{y_1}}{y_1!} \frac{\theta_2^{y_2}}{y_2!} \sum_{x_0=0}^{y_1} x_0! \left(\frac{\theta_0}{\theta_1\theta_2}\right)^{x_0} \binom{y_1}{x_0}\binom{y_2}{x_0}.$$

It can be shown this can be simplified to

$$E(Y_1|Y_1 > y_0) = \frac{e^{-(\theta_0+\theta_1)}}{1 - P(Y_1 \leqslant y_0)} \sum_{y_1=y_0+1}^{\infty} y_1 \frac{\theta_1^{y_1}}{y_1!} \left(1 + \frac{\theta_0}{\theta_1}\right)^{y_1},$$

by using the identity

$$\sum_{y_2=x_0}^{\infty}\frac{\theta_2^{y_2}}{y_2!}\sum_{x_0=0}^{y_1}x_0!\left(\frac{\theta_0}{\theta_1\theta_2}\right)^{x_0}\binom{y_1}{x_0}\binom{y_2}{x_0}=\left(1+\frac{\theta_0}{\theta_1}\right)^{y_1}e^{\theta_2}.$$

This expression can be further simplified to

$$E(Y_1|Y_1>y_0)=\left(\theta_0+\theta_1\right)\frac{1-F(y_0-1|\theta_0+\theta_1)}{1-F(y_0|\theta_0+\theta_1)},\tag{3.3}$$

where $F(y_0|\lambda)=\sum\limits_{r=0}^{y_0}\frac{\lambda^r e^{-\lambda}}{r!}$ is the cumulative distribution function (CDF) of the Poisson distribution with parameter λ. Similarly, when $y_2\leq y_1$ we get

$$E(Y_1|Y_1>y_0)=\frac{e^{-(\theta_0+\theta_1+\theta_2)}}{1-P(Y_1\leqslant y_0)}\sum_{y_1=y_0+1}^{\infty}\sum_{y_2=0}^{\infty}y_1\frac{\theta_1^{y_1}}{y_1!}\frac{\theta_2^{y_2}}{y_2!}\sum_{x_0=0}^{y_2}x_0!\left(\frac{\theta_0}{\theta_1\theta_2}\right)^{x_0}\binom{y_1}{x_0}\binom{y_2}{x_0}.\tag{3.4}$$

Expanding the last part on the right hand side of equation (3.4), we have

$$\begin{aligned}\sum_{y_2=0}^{\infty}\frac{\theta_2^{y_2}}{y_2!}\sum_{x_0=0}^{y_2}x_0!\left(\frac{\theta_0}{\theta_1\theta_2}\right)^{x_0}\binom{y_1}{x_0}\binom{y_2}{x_0}&=1+\theta_2\sum_{x_0=0}^{1}x_0!\left(\frac{\theta_0}{\theta_1\theta_2}\right)^{x_0}\binom{y_1}{x_0}\binom{1}{x_0}+\\&\quad\frac{\theta_2^2}{2!}\sum_{x_0=0}^{2}x_0!\left(\frac{\theta_0}{\theta_1\theta_2}\right)^{x_0}\binom{y_1}{x_0}\binom{2}{x_0}+\cdots\end{aligned}$$

After simplification, the above expression reduces to

$$\sum_{y_2=0}^{\infty}\frac{\theta_2^{y_2}}{y_2!}\sum_{x_0=0}^{y_2}x_0!\left(\frac{\theta_0}{\theta_1\theta_2}\right)^{x_0}\binom{y_1}{x_0}\binom{y_2}{x_0}=\left(1+\frac{\theta_0}{\theta_1}\right)^{y_1}e^{\theta_2}.\tag{3.5}$$

Substituting equation (3.5) into equation (3.4), we get

$$E(Y_1|Y_1>y_0)=\frac{e^{-(\theta_0+\theta_1)}}{1-P(Y_1\leqslant y_0)}\sum_{y_1=y_0+1}^{\infty}y_1\frac{\theta_1^{y_1}}{y_1!}\left(1+\frac{\theta_0}{\theta_1}\right)^{y_1}\tag{3.6}$$

which after simplification will also result in equation (3.3).

Similarly solving $E(Y_2|Y_1>y_0)$ using the same procedure, we get

$$E(Y_2|Y_1>y_0)=\theta_2+\theta_0\frac{1-F(y_0-1|\theta_0+\theta_1)}{1-F(y_0|\theta_0+\theta_1)}.\tag{3.7}$$

Substituting equations (3.3) and (3.7) into (3.1), we get the total effect for the bivariate Poisson assuming a right cut-off point.

$$T_r(y_0;\boldsymbol{\theta})=\theta_1\frac{1-F(y_0-1|\theta_0+\theta_1)}{1-F(y_0|\theta_0+\theta_1)}-\theta_2.\tag{3.8}$$

3.4.2 Case 2: Left cut-off point

Another possibility is when an intervention is carried out on the basis of some cut-off value less than or equal to say y_0. In this situation, the truncated probability distribution of Y_1 and Y_2 is

$$P(Y_1 = y_1, Y_2 = y_2 | Y_1 \leqslant y_0) = \frac{e^{-(\theta_0+\theta_1+\theta_2)}}{P(Y_1 \leqslant y_0)} \frac{\theta_1^{y_1}}{y_1!} \frac{\theta_2^{y_2}}{y_2!} \sum_{x_0=0}^{\min(y_1,y_2)} x_0! \left(\frac{\theta_0}{\theta_1\theta_2}\right)^{x_0} \binom{y_1}{x_0}\binom{y_2}{x_0}.$$

$T_\ell(y_0; \boldsymbol{\theta})$ for this case can be quantified using the formula

$$T_\ell(y_0; \boldsymbol{\theta}) = E(Y_2 - Y_1 | Y_1 \leqslant y_0) = E(Y_2 | Y_1 \leqslant y_0) - E(Y_1 | Y_1 \leqslant y_0), \tag{3.9}$$

where the subscript ℓ is for the left cut-off point. Following the same steps of evaluation as in right truncation, the resulting expression for equation (3.9) is

$$T_\ell(y_0; \boldsymbol{\theta}) = \theta_2 - \theta_1 \frac{F(y_0 - 1 | \theta_0 + \theta_1)}{F(y_0 | \theta_0 + \theta_1)}. \tag{3.10}$$

3.4.3 Variance formulae for $T_k(y_0; \boldsymbol{\theta})$

Expressions for the variance of $T_r(y_0; \boldsymbol{\theta})$ and $T_\ell(y_0; \boldsymbol{\theta})$ can be obtained by combining $\mathrm{var}(Y_1 | Y_1 > y_0)$, $\mathrm{var}(Y_2 | Y_1 > y_0)$ and $\mathrm{cov}(Y_1, Y_2 | Y_1 > y_0)$ as

$$\mathrm{var}(Y_1 - Y_2 | Y_1 > y_0) = \mathrm{var}(Y_1 | Y_1 > y_0) + \mathrm{var}(Y_2 | Y_1 > y_0) - 2 \times \mathrm{cov}(Y_1, Y_2 | Y_1 > y_0). \tag{3.11}$$

Some essential results are first derived to evaluate equation (3.11). The expression

$$E(Y_1(Y_1 - 1) | Y_1 > y_0) = \sum_{y_1 = y_0+1}^{\infty} \sum_{y_2 = x_0}^{\infty} Y_1(Y_1 - 1) P(Y_1 = y_1, Y_2 = y_2 | Y_1 > y_0),$$

can be simplified to

$$E(Y_1(Y_1 - 1) | Y_1 > y_0) = (\theta_0 + \theta_1)^2 \frac{1 - F(y_0 - 2 | \theta_0 + \theta_1)}{1 - F(y_0 | \theta_0 + \theta_1)}.$$

Likewise,

$$E(Y_2(Y_2 - 1) | Y_1 > y_0) = \theta_2^2 - \theta_0 \frac{\theta_0 F(y_0 - 2 | \theta_0 + \theta_1) + 2\theta_2 F(y_0 - 1 | \theta_0 + \theta_1) - \theta_0 - 2\theta_2}{1 - F(y_0 | \theta_0 + \theta_1)},$$

and

$$\begin{aligned} E(Y_1 Y_2 | Y_1 > y_0) = & \frac{\theta_0^2 + (\theta_1 + \theta_2 + 1)\theta_0 + \theta_1\theta_2 - ((\theta_2 + 1)\theta_0 + \theta_1\theta_2) F(y_0 - 1 | \theta_0 + \theta_1)}{1 - F(y_0 | \theta_0 + \theta_1)} \\ & - \frac{\theta_0(\theta_0 + \theta_1) F(y_0 - 2 | \theta_0 + \theta_1)}{1 - F(y_0 | \theta_0 + \theta_1)}. \end{aligned}$$

We know that $\text{var}(Y_i|Y_1 > y_0) = E(Y_i(Y_i-1)|Y_1 > y_0)+E(Y_i|Y_1 > y_0)-(E(Y_i|Y_1 > y_0))^2$ for $i = 1, 2$ and $\text{cov}(Y_1, Y_2|Y_1 > y_0) = E(Y_1Y_2|Y_1 > y_0) - E(Y_1|Y_1 > y_0)E(Y_2|Y_1 > y_0)$. Using the derived results and equations (3.3) and (3.7), the expression in equation (3.11) reduces to

$$\begin{aligned}\text{var}(Y_1 - Y_2|Y_1 > y_0) &= \theta_2 + \theta_1 \frac{1 - F(y_0 - 1|\theta_0 + \theta_1)}{1 - F(y_0|\theta_0 + \theta_1)} \\ &+ \theta_1^2 \sum_{i=0}^{1} \left((-1)^{i+1} \frac{P(Y_1 = y_0 - i)(1 - F(y_0 - 1 + i|\theta_0 + \theta_1))}{(1 - F(y_0|\theta_0 + \theta_1))^2} \right).\end{aligned} \tag{3.12}$$

Similarly, the expression of variance for the conditional difference of Y_1 and Y_2 for the left cut-off point is given by

$$\begin{aligned}\text{var}(Y_2 - Y_1|Y_1 \le y_0) &= \theta_2 + \theta_1 \frac{F(y_0 - 1|\theta_0 + \theta_1)}{F(y_0|\theta_0 + \theta_1)} \\ &+ \theta_1^2 \sum_{i=0}^{1} \left((-1)^{i} \frac{P(Y_1 = y_0 - i)(F(y_0 - 1 + i|\theta_0 + \theta_1))}{(F(y_0|\theta_0 + \theta_1))^2} \right).\end{aligned} \tag{3.13}$$

3.5 RTM and Intervention/Treatment Effects

In a pre/post study design, the average intervention/treatment effect $\delta_r(\boldsymbol{\theta})$, for a right cut-off point, is the expected difference of events before and after the intervention. Mathematically, this is

$$\delta_r(\boldsymbol{\theta}) = E(Y_1 - Y_2).$$

Assuming the pre/post observations follow the bivariate Poisson distribution and using the fact $E(Y_i) = \theta_0 + \theta_i$ for $i = 1, 2$, the expression for the intervention effect is

$$\delta_r(\boldsymbol{\theta}) = \theta_1 - \theta_2. \tag{3.14}$$

Suppose an intervention is applied to subjects above a certain threshold y_0. For a null intervention effect, the pre and post observations are identically distributed and the expected conditional difference of Y_1 and Y_2 conditioned on y_0 is the RTM effect as defined in equation (3.2). Formula for $R_r(y_0; \boldsymbol{\theta})$ can be obtained by letting $\theta_2 = \theta_1$ in equation (3.8) as

$$\begin{aligned}R_r(y_0; \boldsymbol{\theta}) &= \theta_1 \frac{1 - F(y_0 - 1|\theta_0 + \theta_1)}{1 - F(y_0|\theta_0 + \theta_1)} - \theta_1, \\ &= \theta_1 \cdot \frac{P(y_0|\theta_0 + \theta_1)}{1 - F(y_0|\theta_0 + \theta_1)},\end{aligned} \tag{3.15}$$

where $P(y|\lambda) = e^{-\lambda}\lambda^y/y!$ is the Poisson probability mass function for parameter λ.

For a non-zero intervention effect, the distribution of pre and post observations are not identical, i.e, $\theta_1 \neq \theta_2$ and the expected conditional difference, $E(Y_1 - Y_2|Y_1 > y_0)$, can be decomposed into RTM and intervention effects. To prove the argument mathematically, adding equations (3.14) and (3.15), we get

$$\begin{aligned} R_r(y_0;\boldsymbol{\theta}) + \delta_r(\boldsymbol{\theta}) &= \theta_1 \cdot \frac{P(y_0|\theta_0+\theta_1)}{1-F(y_0|\theta_0+\theta_1)} + (\theta_1 - \theta_2) \\ &= T_r(y_0;\boldsymbol{\theta}). \end{aligned} \tag{3.16}$$

Following the same steps, similar equations for the left cut-off point can be proved as

$$\delta_\ell(\boldsymbol{\theta}) = \theta_2 - \theta_1, \tag{3.17}$$

$$R_\ell(y_0;\boldsymbol{\theta}) = \theta_1 \cdot \frac{P(y_0|\theta_0+\theta_1)}{F(y_0|\theta_0+\theta_1)} \tag{3.18}$$

$$T_\ell(y_0;\boldsymbol{\theta}) = R_\ell(y_0;\boldsymbol{\theta}) + \delta_\ell(\boldsymbol{\theta}). \tag{3.19}$$

3.5.1 Variances of RTM and intervention/treatment effects

Using the properties var$(Y_i) = \theta_0 + \theta_i$ for $i = 1, 2$ and cov$(Y_1, Y_2) = \theta_0$ for the bivariate Poisson distribution, the variance of $Y_1 - Y_2$ is

$$\text{var}(Y_1 - Y_2) = \theta_1 + \theta_2. \tag{3.20}$$

Variances of $R_k(y_0;\boldsymbol{\theta})$ for $k = r, \ell$, can be obtained by substituting $\theta_2 = \theta_1$ in equations (3.12) and (3.13) as

$$\begin{aligned} \text{var}(Y_1 - Y_2|Y_1 > y_0, \theta_2 = \theta_1) &= \theta_1 + \theta_1 \frac{1-F(y_0-1|\theta_0+\theta_1)}{1-F(y_0|\theta_0+\theta_1)} \\ &+ \theta_1^2 \sum_{i=0}^{1}\left((-1)^{i+1}\frac{P(Y_1 = y_0 - i)(1-F(y_0-1+i|\theta_0+\theta_1))}{(1-F(y_0|\theta_0+\theta_1))^2}\right), \end{aligned} \tag{3.21}$$

and

$$\begin{aligned} \text{var}(Y_2 - Y_1|Y_1 \leq y_0, \theta_2 = \theta_1) &= \theta_1 + \theta_1 \frac{F(y_0-1|\theta_0+\theta_1)}{F(y_0|\theta_0+\theta_1)} \\ &+ \theta_1^2 \sum_{i=0}^{1}\left((-1)^{i}\frac{P(Y_1 = y_0 - i)(F(y_0-1+i|\theta_0+\theta_1))}{(F(y_0|\theta_0+\theta_1))^2}\right). \end{aligned} \tag{3.22}$$

3.5.2 RTM as a function of cut-off point y_0

Using equations (3.15) and (3.18), the graph for different cut-off values is given in Figure 3.2. For illustrative purposes, specific values of $(\theta_0, \theta_1, \theta_2) = (6, 3, 3)$ are considered. It is evident from the graph that the RTM effect is at its peak for cut-off values at the extremes on either side. For a right cut-off point, as the value of y_0 increases, the probability $P(Y_1 > y_0)$ decreases and the associated RTM increases. For left cut-off points, the probability $P(Y_1 \leq y_0)$ increases as y_0 increases and the reverse relationship is observed.

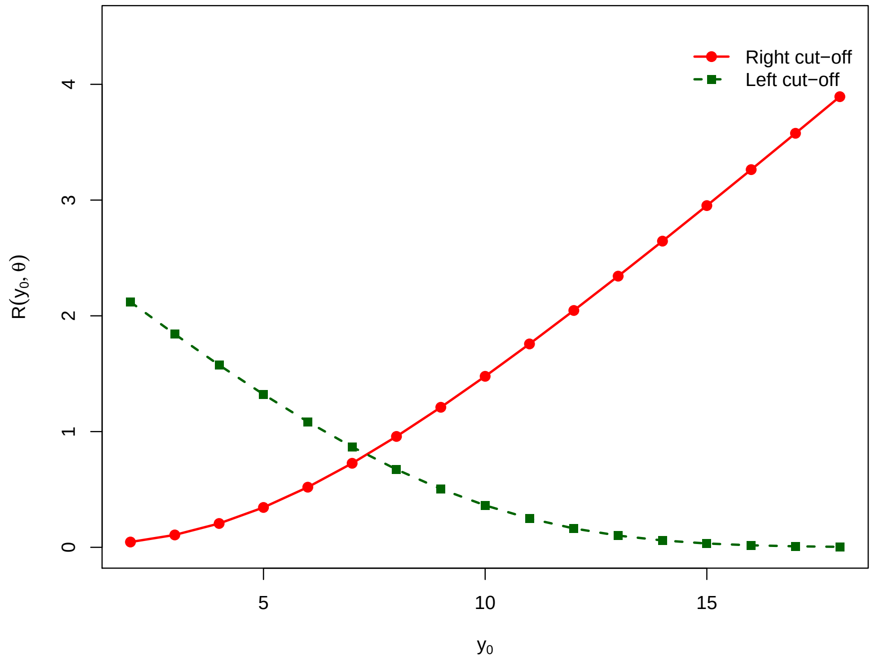

Figure 3.2. Graph of the RTM effect constructed on the basis of derived formula for points greater than or less than a cut-off value y_0 when the underlying distribution is bivariate Poisson with parameters $\theta_0 = 6$, and $\theta_1 = \theta_2 = 3$.

3.5.3 RTM as a function of covariance

The RTM effect as a function of the covariance θ_0 (the true parameter) is given in Figure 3.3. A fixed right cut-off point of $y_0 = 8$ and specific values of $(\theta_1, \theta_2) = (3, 3)$ are considered for demonstration purposes, though the general pattern is similar for other values. When θ_0 increases, the correlation between Y_1 and Y_2 also increases. In the case of the normal distribution, as the covariance/correlation between Y_1 and Y_2 increases, the RTM effect decreases. But, for the Poisson distribution, the situation is quite different due to the equality of mean and variance of the true variable X_0 and covariance of pre and post variables, i.e, $\text{cov}(Y_1, Y_2) = \theta_0 = \text{var}(X_0) = E(X_0)$. For the normal distribution, the mean/variance identity does not hold $\text{cov}(X_1, X_2) = \sigma_0^2 = \text{var}(X_0) \neq E(X_0)$.

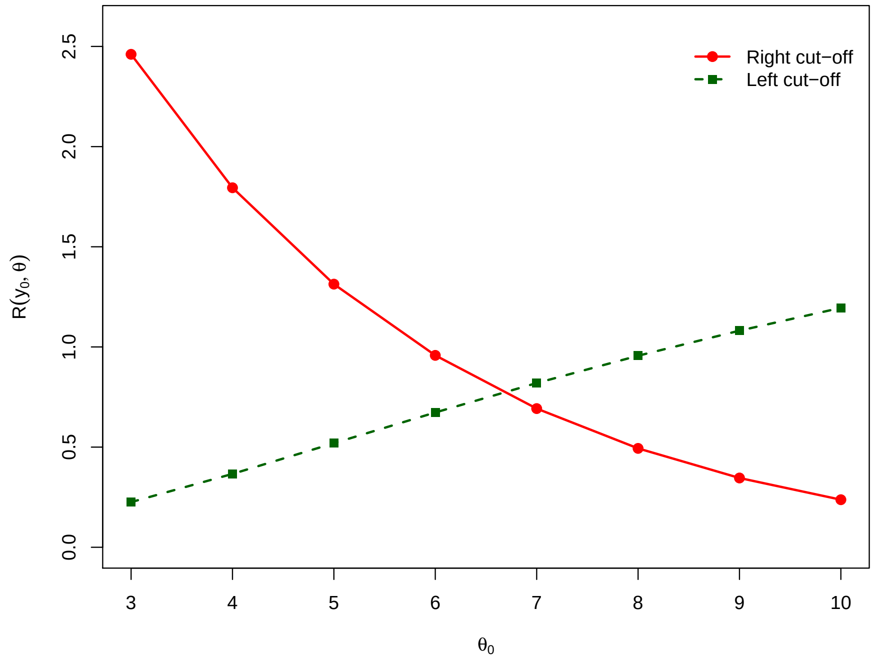

Figure 3.3. Graph of the RTM effect for different values of θ_0 and at fixed cut-off point y_0 and $\theta_1 = \theta_2 = 3$.

It is evident from the graph that as θ_0 increases, the RTM effect decreases due to the increase in the probability $P(Y_1 > y_0)$. On the other hand, when the fixed cut-off point is on the left side, the RTM effect increases as the value of θ_0 increases because we are moving farther away from the cut-off point to the right. Stated differently, the probability $P(Y_1 \leq y_0)$ decreases with increasing values of θ_0 which ultimately causes the RTM effect to increase.

3.6 RTM and Poisson Processes

A Poisson process is a collection of random variables $\{N(t) : t \geq 0\}$ where $N(t)$ is the number of events that have occurred up to time t. Many real world situations are modelled as a Poisson process. For example, the number of failures in repairable systems, sighting of invasive species, page view requests to a website during a time interval of length t and the number of plants of a particular species in a given location of area a.

Let $N_1(t)$ and $N_2(t)$ be two Poisson processes representing the successive number of occurrences of the same event over a specified interval of time before and after an intervention. Further, let $N_1(t) = M_0(t)+M_1(t)$ and $N_2(t) = M_0(t)+M_2(t)$ where $M_i(t)$ are independent Poisson processes with parameters $\theta_i(t)$ for $i = 0, 1, 2$. $M_0(t)$ represents the true occurrences and $M_1(t)$ and $M_2(t)$ are random errors. The joint truncated probability distribution function of $N_1(t)$ and $N_2(t)$ is given by

$$P(N_1(t) = k_1, N_2(t) = k_2 | N_1(t) > k_0)$$

$$= \frac{e^{-(\theta_0(t)+\theta_1(t)+\theta_2(t))}}{1 - P(N_1(t) \leqslant k_0)} \frac{\theta_1(t)^{k_1}}{k_1!} \frac{\theta_2(t)^{k_2}}{k_2!} \sum_{k_0=0}^{\min(k_1,k_2)} k_0! \left(\frac{\theta_0(t)}{\theta_1(t)\theta_2(t)} \right)^{k_0} \binom{k_1}{k_0} \binom{k_2}{k_0}.$$

On the basis of an arrival rate, Poisson processes can be divided into two categories known as homogeneous and inhomogeneous Poisson processes. In the following sections, the total, RTM and intervention effects are discussed separately for the two types of Poisson processes.

3.6.1 $T_k^{(H)}(y_0; \boldsymbol{\theta})$, $R_k^{(H)}(y_0; \boldsymbol{\theta})$, and $\delta_k(\boldsymbol{\theta})$ for homogeneous Poisson process

For a homogeneous Poisson process, the mean rate of occurrence $\theta(t)$ is constant and independent of the location of the interval. That is, it does not vary with time or space and depends only on the length of the interval. The number of events/occurrences in any interval of length t is Poisson distributed with mean $E(N(t)) = \theta(t) = t\theta$.

The statistical properties of the Poisson distribution and homogeneous Poisson process differ only by a multiple of the time interval t. Therefore, the total, intervention/treatment and RTM effects can be quantified simply by replacing $\theta_i(t) = t\theta_i$ for $i = 0, 1, 2$ in equations (3.8), (3.14), (3.15) for the right cut-off point and equations (3.10), (3.17), and (3.18) for the left cut-off point, respectively. The resulting formulae for both cases when truncation is on the right side, i.e., $N(t) > y_0$, and when truncation is on the left side, i.e., $N(t) \leqslant y_0$, are respectively given by

$$\begin{aligned} T_r^{(H)}(y_0;\boldsymbol{\theta}) &= t\theta_1 \frac{1 - F(y_0 - 1|t(\theta_0+\theta_1))}{1 - F(y_0|t(\theta_0+\theta_1))} - t\theta_2, \\ \delta_r^{(H)}(\boldsymbol{\theta}) &= t\theta_1 - t\theta_2, \\ R_r^{(H)}(y_0;\boldsymbol{\theta}) &= t\theta_1 \cdot \frac{P(y_0|t(\theta_0+\theta_1))}{1 - F(y_0|t(\theta_0+\theta_1))}, \end{aligned} \tag{3.23}$$

and

$$\begin{aligned} T_\ell^{(H)}(y_0;\boldsymbol{\theta}) &= t\theta_2 - t\theta_1 \frac{F(y_0 - 1|t(\theta_0+\theta_1))}{F(y_0|t(\theta_0+\theta_1))}, \\ \delta_\ell^{(H)}(\boldsymbol{\theta}) &= t\theta_2 - t\theta_1, \\ R_\ell^{(H)}(y_0;\boldsymbol{\theta}) &= t\theta_1 \cdot \frac{P(y_0|t(\theta_0+\theta_1))}{F(y_0|t(\theta_0+\theta_1))}, \end{aligned} \tag{3.24}$$

where $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2)$.

3.6.2 $T_k^{(I)}(y_0;\boldsymbol{\theta})$, $R_k^{(I)}(y_0;\boldsymbol{\theta})$, and $\delta_k(\boldsymbol{\theta})$ for inhomogeneous Poisson process

The arrival rate may depend on the location or time of an interval. For example, the arrival rate of calls to a telephone answering service varies with time as there are more calls during the day than the night. In this situation, the number of occurrences/events in an interval of length t is said to follow an inhomogeneous Poisson process with mean $\Theta(t) = \int_0^t \theta(t)dt$ where $\theta(t)$ is some function of time t. The joint probability distribution function of $N_1(t)$ and $N_2(t)$ will be a bivariate Poisson process with parameters $\Theta_i(t) = \int_0^t \theta_i(t)dt$ for $i = 0, 1, 2$.

The statistical properties of both the Poisson distribution and the inhomogeneous Poisson process are the same apart from the differences in parameter structure. Therefore, to quantify the total, intervention and RTM effects for inhomogeneous Poisson processes, we simply need to replace $\Theta_i(t) = \int_0^t \theta_i(t)dt$ for corresponding θ_i, for $i = 0, 1, 2$, in the respective equations (3.8), (3.14), (3.15) for a right cut-off point and equations (3.10), (3.17), and (3.18) for a left cut-off point. The resulting formulae for a right cut-off are given by

$$\begin{aligned}
T_r^{(I)}(y_0;\boldsymbol{\Theta}) &= \Theta_1(t)\frac{1-F(y_0-1|\Theta_0(t)+\Theta_1(t))}{1-F(y_0|\Theta_0(t)+\Theta_1(t))} - \Theta_2(t),\\
\delta_r^{(I)}(\boldsymbol{\Theta}) &= \Theta_1(t) - \Theta_2(t),\\
R_r^{(I)}(y_0;\boldsymbol{\Theta}) &= \Theta_1(t)\cdot\frac{P(y_0|\Theta_0(t)+\Theta_1(t))}{1-F(y_0|\Theta_0(t)+\Theta_1(t))},
\end{aligned} \tag{3.25}$$

where $\boldsymbol{\Theta} = (\Theta_0, \Theta_1, \Theta_2)$. Similarly,

$$\begin{aligned}
T_\ell^{(I)}(y_0;\boldsymbol{\Theta}) &= \Theta_2(t) - \Theta_1(t)\frac{F(y_0-1|\Theta_0(t)+\Theta_1(t))}{F(y_0|\Theta_0(t)+\Theta_1(t))},\\
\delta_\ell^{(I)}(\boldsymbol{\Theta}) &= \Theta_2(t) - \Theta_1(t)\\
R_\ell^{(I)}(y_0;\boldsymbol{\Theta}) &= \Theta_1(t)\cdot\frac{P(y_0|\Theta_0(t)+\Theta_1(t))}{F(y_0|\Theta_0(t)+\Theta_1(t))},
\end{aligned} \tag{3.26}$$

when truncation is on the left side of the distribution.

3.6.3 Numerical example of homogeneous and inhomogeneous Poisson processes

Let us consider the mean arrival rate of an inhomogeneous process to be $\theta_i(t) = \theta_i + b \times \cos(wt)$ for $i = 0, 1, 2$ where θ_i, b and w are constants. The resulting mean value of the inhomogeneous Poisson process is $\Theta_i(t) = \theta_i t + b/w \times \sin(wt)$ for $i = 0, 1, 2$. Further, for demonstrative purposes assume that $\theta_i = 6$, $b = 3$ and $w = 1$, then the graph of constant ($\theta_i = 6$) and varying arrival ($\Theta_i(t)$) rates are given in Figure 3.4. The peak and trough of the inhomogeneous Poisson process are 9 and 3 respectively, while the homogeneous Poisson process is constant at 6.

Without loss of generality, let us assume that the width of the time interval is unity, then the expression for the corresponding mean value of an inhomogeneous Poisson processes is given by

$$\Theta_i(t) = \int_{t-1}^{t}(\theta_i + b\times\cos(wt))dt = \theta_i + b/w\times\sin(wt) - b/w\times\sin(w(t-1)).$$

For illustration purposes, the graphs of constant and varying arrival rates are given in Figures 3.5-3.7 for $\theta_0 = 6$, $\theta_i = 3$ for $(i = 1, 2)$, $b = 30$ and $w = 1$.

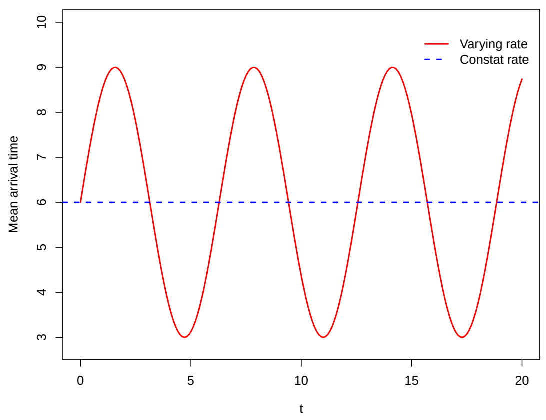

Figure 3.4. Graph of arrival rates for homogeneous and inhomogeneous Poisson processes.

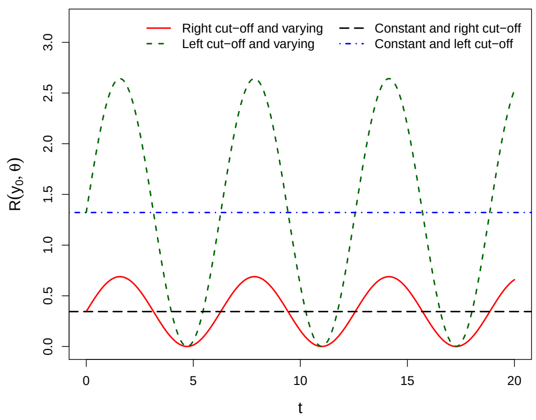

Figure 3.5. The RTM effects for constant and varying arrival rates for fixed cut-off point $y_0 = 5$

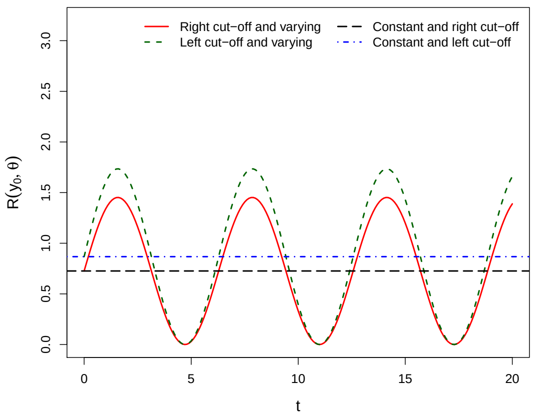

Figure 3.6. The RTM effects for constant and varying arrival rates for fixed cut-off point $y_0 = 7$

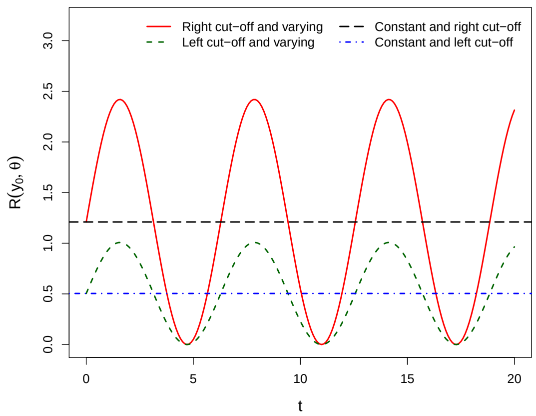

Figure 3.7. The RTM effects for constant and varying arrival rates for fixed cut-off point $y_0 = 9$

As the cut-off point increases from 5 to 9, the RTM effect for right truncation increases, while on the other hand, it decreases for left truncation both for constant and varying arrival rates. The varying arrival rate shows periodicity for left cut-off points, starts climbing up from a value of RTM for constant arrival rate (the blue horizontal line) reaching a maximum and then starts descending to zero and the cycles are repeated for left cut-off points. Likewise, for right cut-off points, the RTM effect for the varying arrival rate starts from a constant value of RTM (the black horizontal line) reaching a maximum and then declining to a low level.

3.7 Maximum Likelihood Estimation (MLE) of RTM Effect

Let $(y_{11}, y_{21}), (y_{12}, y_{22}), ..., (y_{1n}, y_{2n})$ be independent pairs of observations of size n from the truncated bivariate Poisson distribution. Let us denote $P(Y_1 = y_1, Y_2 = y_2|Y_1 > y_0)$ by $P_T(y_1, y_2)$ for brevity. The likelihood and log likelihood functions are respectively given by

$$L(\boldsymbol{\theta}, \boldsymbol{y}) = \prod_{i=1}^{n} P_T(y_{1i}, y_{2i}),$$

and

$$\ell(\boldsymbol{\theta}, \boldsymbol{y}) = \sum_{i=1}^{n} \log(P_T(y_{1i}, y_{2i})),$$

where

$$\boldsymbol{y} = \begin{pmatrix} y_{11} & y_{21} \\ y_{12} & y_{22} \\ \cdots & \cdots \\ y_{1n} & y_{2n} \end{pmatrix}.$$

Differentiating the log likelihood function with respect to θ_i and then equating to zero for $i = 0, 1, 2$, we get the equations

$$\frac{1}{n}\sum_{i=1}^{n} \frac{P_T(y_{1i}-1, y_{2i}-1)}{P_T(y_{1i}, y_{2i})} = \frac{1-F(y_0-1|\theta_0+\theta_1)}{1-F(y_0|\theta_0+\theta_1)}, \tag{3.27}$$

$$\frac{1}{n}\sum_{i=1}^{n} \frac{P_T(y_{1i}-1, y_{2i})}{P_T(y_{1i}, y_{2i})} = \frac{1-F(y_0-1|\theta_0+\theta_1)}{1-F(y_0|\theta_0+\theta_1)}, \tag{3.28}$$

and

$$\frac{1}{n}\sum_{i=1}^{n} \frac{P_T(y_1, y_2-1)}{P_T(y_1, y_2)} = 1. \tag{3.29}$$

A solution to this system of equations can be found using the Teicher (1954) recursive relationships given by

$$y_{1i}P_T(y_{1i}, y_{2i}) = \theta_1 P_T(y_{1i}-1, y_{12}) + \theta_0 P_T(y_{1i}-1, y_{2i}-1), \tag{3.30}$$

$$y_{2i}P_T(y_{1i}, y_{2i}) = \theta_2 P_T(y_{1i}, y_{2i}-1) + \theta_0 P_T(y_{1i}-1, y_{2i}-1). \tag{3.31}$$

Dividing equation (3.30) by $n \cdot P_T(y_1, y_2)$, summing over the sample, and using equations (3.27) and (3.28) we get

$$\overline{y}_{1|y_1>y_0} = (\theta_0 + \theta_1) \cdot \frac{1 - F(y_0 - 1||\theta_0 + \theta_1)}{1 - F(y_0|\theta_0 + \theta_1)}, \tag{3.32}$$

where $\overline{y}_{1|y_1>y_0} = \sum_{y_{1i}>y_0} y_{1i}/n$ is the conditional sample mean of pre observations. Similarly,

$$\overline{y}_{2|y_1>y_0} = \theta_2 + \theta_0 \cdot \frac{1 - F(y_0 - 1||\theta_0 + \theta_1)}{1 - F(y_0|\theta_0 + \theta_1)}, \tag{3.33}$$

where $\overline{y}_{2|y_1>y_0} = \sum_{y_{2i}>y_0} y_{2i}/n$ is the conditional sample mean of post observations.

Subtracting (3.33) from (3.32) and rearranging terms, we get the MLE of $T_r(y_0; \boldsymbol{\theta})$, given by

$$\widehat{T}_r(y_0, \boldsymbol{y}) = \hat{\theta}_1 \cdot \frac{1 - F(y_0 - 1|\hat{\theta}_0 + \hat{\theta}_1)}{1 - F(y_0|\hat{\theta}_0 + \hat{\theta}_1)} - \hat{\theta}_2 = \overline{y}_{1|y_1>y_0} - \overline{y}_{2|y_1>y_0}. \tag{3.34}$$

Likewise, the MLE of $T_\ell(y_0; \boldsymbol{\theta})$ is

$$\widehat{T}_\ell(y_0, \boldsymbol{y}) = \hat{\theta}_2 - \hat{\theta}_1 \cdot \frac{F(y_0 - 1|\hat{\theta}_0 + \hat{\theta}_1)}{F(y_0|\hat{\theta}_0 + \hat{\theta}_1)} = \overline{y}_{2|y_1\leq y_0} - \overline{y}_{1|y_1\leq y_0}. \tag{3.35}$$

The parameter vector $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2)$ of the truncated bivariate Poisson distribution can be estimated by modifying the direct method of maximum likelihood (Kawamura, 1984), which in turn can be used to estimate the intervention and RTM effects. Let $\hat{\boldsymbol{\theta}} = (\hat{\theta}_0, \hat{\theta}_1, \hat{\theta}_2)$ be the estimate of the parameter vector, then the estimators of RTM and intervention effects, for $k = r, \ell$, are

$$\widehat{R}_k(y_0, \boldsymbol{y}) = R_k(y_0, \hat{\boldsymbol{\theta}}), \quad \text{and} \quad \delta_k(\boldsymbol{y}) = \delta_k(\hat{\boldsymbol{\theta}}).$$

3.7.1 Variances of $\widehat{T}_k(y_0, \boldsymbol{y})$ and $\widehat{R}_k(y_0, \boldsymbol{y})$

To obtain the variance of $\widehat{T}_r(y_0, \boldsymbol{y})$, subtracting $T_r(y_0; \boldsymbol{\theta})$ on both sides of equation (3.34), squaring and taking expectations, we get

$$\text{var}(\widehat{T}_r(y_0, \boldsymbol{y})) = \text{var}(\overline{y}_{1|y_1>y_0} - \overline{y}_{2|y_1>y_0}) = \text{var}(Y_1 - Y_2 | Y_1 > y_0)/n. \tag{3.36}$$

In a similar manner, for the left cut-off point, we get

$$\text{var}(\widehat{T}_\ell(y_0, \boldsymbol{y})) = \text{var}(Y_2 - Y_1 | Y_1 \leq y_0)/n. \tag{3.37}$$

To obtain the variance of $\widehat{R}_k(y_0, \boldsymbol{y})$ for $k = r, \ell$, replace θ_2 with θ_1 in equations (3.36) and (3.37) respectively. Equivalently,

$$\mathrm{var}(\widehat{R}_r(y_0, \boldsymbol{y})) = \mathrm{var}\left(Y_1 - Y_2 | Y_1 > y_0, \theta_2 = \theta_1\right)/n,$$

and

$$\mathrm{var}(\widehat{R}_\ell(y_0, \boldsymbol{y})) = \mathrm{var}\left(Y_2 - Y_1 | Y_1 \leq y_0, \theta_2 = \theta_1\right)/n.$$

3.7.2 Unbiasedness of $\widehat{T}_k(y_0, \boldsymbol{y})$ and $\widehat{R}_k(y_0, \boldsymbol{y})$

The unbiasedness of $\widehat{T}_k(y_0, \boldsymbol{y})$ for $k = r, \ell$ can be established by using equations (3.3-3.10) as

$$E(\widehat{T}_k(y_0, \boldsymbol{y})) = T_k(y_0; \boldsymbol{\theta}). \tag{3.38}$$

For a null effect, i.e., $\delta(\boldsymbol{\theta}) = 0$, $\widehat{T}_k(y_0, \boldsymbol{y})$ and $\widehat{R}_k(y_0, \boldsymbol{y})$ are equivalent. Thus, $\widehat{R}_k(y_0, \boldsymbol{y})$ can be written as

$$E(\widehat{T}_k(y_0, \boldsymbol{y}) | \theta_1 = \theta_2) = E(\widehat{R}_k(y_0, \boldsymbol{y})) = R_k(y_0; \boldsymbol{\theta}). \tag{3.39}$$

3.7.3 Asymptotic distribution of $\widehat{T}_k(y_0, \boldsymbol{y})$ and $\widehat{R}_k(y_0, \boldsymbol{y})$

It is clear from equations (3.32) and (3.33) that $\widehat{T}_k(y_0, \boldsymbol{y})$ for $k = r, \ell$ are the differences of the conditional sample means of Y_1 and Y_2. Using the Central Limit Theorem and considering the right cut-off point, $\widehat{T}_r(y_0, \boldsymbol{y})$ is asymptotically normally distributed as

$$\sqrt{n}\left(\widehat{T}_r(y_0, \boldsymbol{y})) - T_r(y_0; \boldsymbol{\theta})\right) \overset{d}{\sim} N(0, \mathrm{var}(Y_1 - Y_2 | Y_1 > y_0)).$$

It is well known that the additive components of a normal random variable are necessarily normally distributed (Cramér, 1936). So, the components $\hat{R}_r(y_0, \boldsymbol{y})$ and $\hat{\delta}_r(\boldsymbol{y})$ of $\widehat{T}_r(y_0, \boldsymbol{y})$ are also asymptotically normally distributed, and hence

$$\sqrt{n}\left(\widehat{R}_r(y_0, \boldsymbol{y}) - R_r(y_0, \boldsymbol{\theta})\right) \overset{d}{\sim} N(0, \mathrm{var}(Y_2 - Y_1 | Y_1 > y_0, \theta_2 = \theta_1)).$$

The results also hold for the left cut-off point.

3.8 Simulation Study for Quantifying the RTM Effect

A simulation study was carried out for estimating the RTM effect and comparing it with the true RTM effect for specified parameters of the truncated bivariate Poisson distribution. The following steps were taken to generate two sets of observations, representing measurements before and after an intervention.

1. The probabilities $P(Y > y_0)$ or $P(y_0 \leq Y)$ are small if y_0 is farther in the tail of a probability distribution. So, the number of observations beyond/below a cut-off point y_0, i.e., $n \cdot P(Y > y_0)$ or $n \cdot P(Y \leq y_0)$ in a sample generated from the distribution would be small. Therefore, to get random samples of size $n = 10, 20, ..., 200$ beyond/below y_0, sufficiently large random samples were generated from a Poisson distribution with mean $\theta_0 = 6$. These realizations are denoted by x_{0j} for $j = 1, 2, ..., n$.

2. Sets of random samples of corresponding sizes were generated from the Poisson distribution with means $\theta_i = 3$ for $i = 1, 2$. These realizations of random errors are denoted by x_{ij} for $i = 1, 2$ and $j = 1, 2, ..., n$.

3. Pre and post observations were obtained by $y_{1j} = x_{0j} + x_{1j}$ and $y_{2j} = x_{0j} + x_{2j}$.

4. The first n observations of y_{1j} beyond/below y_0 and the corresponding y_{2i} were considered as random samples from a truncated bivariate Poisson distribution.

5. The sampling procedure was repeated $m = 1000$ times and the RTM effect for each sample, was estimated using maximum likelihood.

3.8.1 Empirical distribution of $\hat{R}_k(y_0, \boldsymbol{y})$

The normal quantile-quantile plots given in Figure 3.8 indicate that the sampling distributions of $\hat{R}_k(y_0, \boldsymbol{y})$ for a right cut-off point $y_0 = 15$ and different sample sizes are approximately normal. The normal quantile plots for the cases $\theta_1 > \theta_2$ and $\theta_1 < \theta_2$ are not given for brevity, but they also support approximate normality of the distribution of RTM estimators.

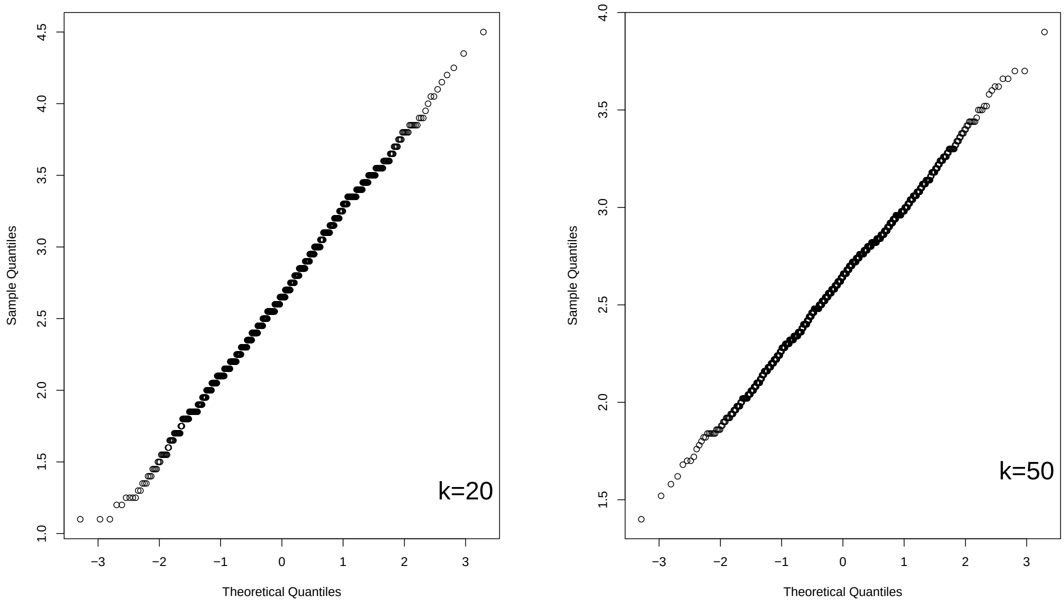

Figure 3.8. Left panel: Normal quantile plot of the sampling distribution of RTM effect for $n = 20$, $y_0 = 15$, $\theta_1 = \theta_2 = 3$ and $\theta_0 = 6$ Right panel: Normal quantile plot of the sampling distribution of RTM effect for $n = 50$.

3.8.2 Empirical unbiasedness and consistency of $\hat{R}_k(y_0, \boldsymbol{y})$

Estimates of RTM are given in Figure 3.9. The means of the sampling distributions of RTM are very close to the true value for different sample sizes and choices of the parameters θ_1 and θ_2, suggesting unbiasedness. As the sample size increases the spread around the centre decreases, suggesting consistency of the estimator.

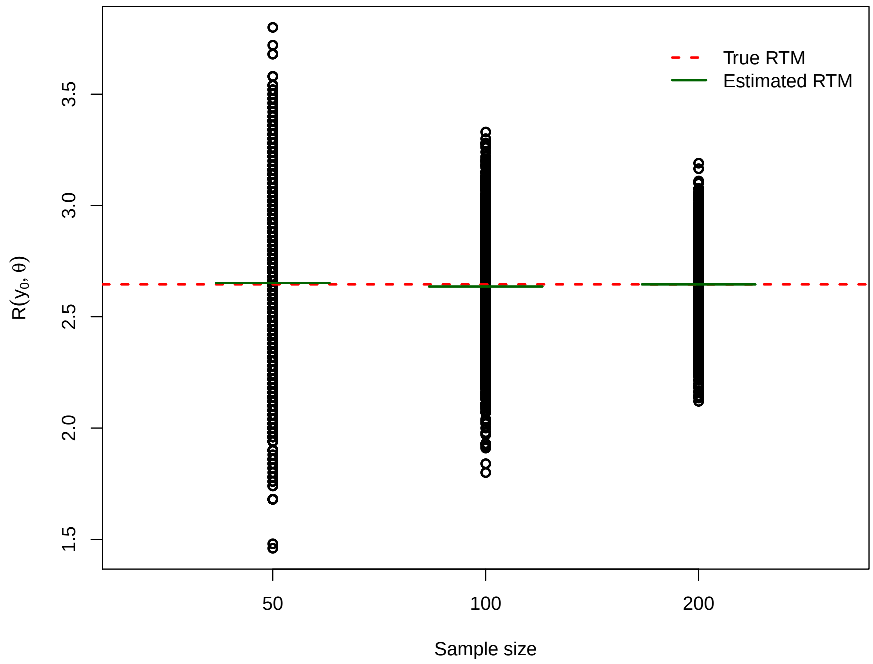

Figure 3.9. Estimates of RTM and its sampling distribution for different sample sizes.

3.8.3 Confidence intervals and coverage probabilities

Coverage probability is a useful tool for evaluating the performance of an estimator for parameters of a discrete distribution. Coverage probability of a $(1-\alpha)100\%$ confidence interval is the probability it contains the true parameter. Let $\widehat{L}_i$ and $\widehat{U}_i$ be the respective lower and upper limits of a confidence interval for $R_k(y_0;\boldsymbol{\theta})$ for $k = r, \ell$ estimated from a sample of size n, and let $I(\cdot)$ be the indicator function. The true coverage probability is given by

$$C(\boldsymbol{\theta}, n) = \sum_x I(\widehat{L}_i < R_i(y_0;\boldsymbol{\theta}) < \widehat{U}_i)P(x;\boldsymbol{\theta}).$$

$C(\boldsymbol{\theta}, n)$ is a function of $(\theta_0, \theta_1, \theta_2, n)$ and it cannot be displayed on a graph in a two dimensional plane without holding some parameters constant.

The coverage probability can be used to investigate how well asymptotic confidence intervals work for $R_k(y_0, \boldsymbol{\theta})$, and also to explore its behaviour for finite sample sizes. The simulated coverage probability is defined as the proportion of times confidence intervals contain the true parameter from a series of simulated datasets, given by

$$\widehat{C}(\boldsymbol{\theta}, n) = \frac{\sum_{i=1}^{m} I_i(\hat{L} < R_i(y_0; \boldsymbol{\theta}) < \hat{U})}{m},$$

where m is the number of simulated datasets.

Assuming normality of the estimates, 95% confidence intervals were constructed and their coverage probabilities were computed for sample sizes $n = 10, 20, ..., 200$, using maximum likelihood estimates as

$$\hat{R}_k(y_0, \boldsymbol{y}) \pm z_{1-\alpha/2} \times \sqrt{\widehat{\text{var}}(\hat{R}_k(y_0, \boldsymbol{y}))} \qquad \text{for} \qquad k = r, \ell. \tag{3.40}$$

For sample sizes of at least 20, the nominal coverage probability level (95%) is well approximated on average as depicted in Figure 3.10. For sample size $n = 10$ the simulated coverage probability remains consistently around 92%.

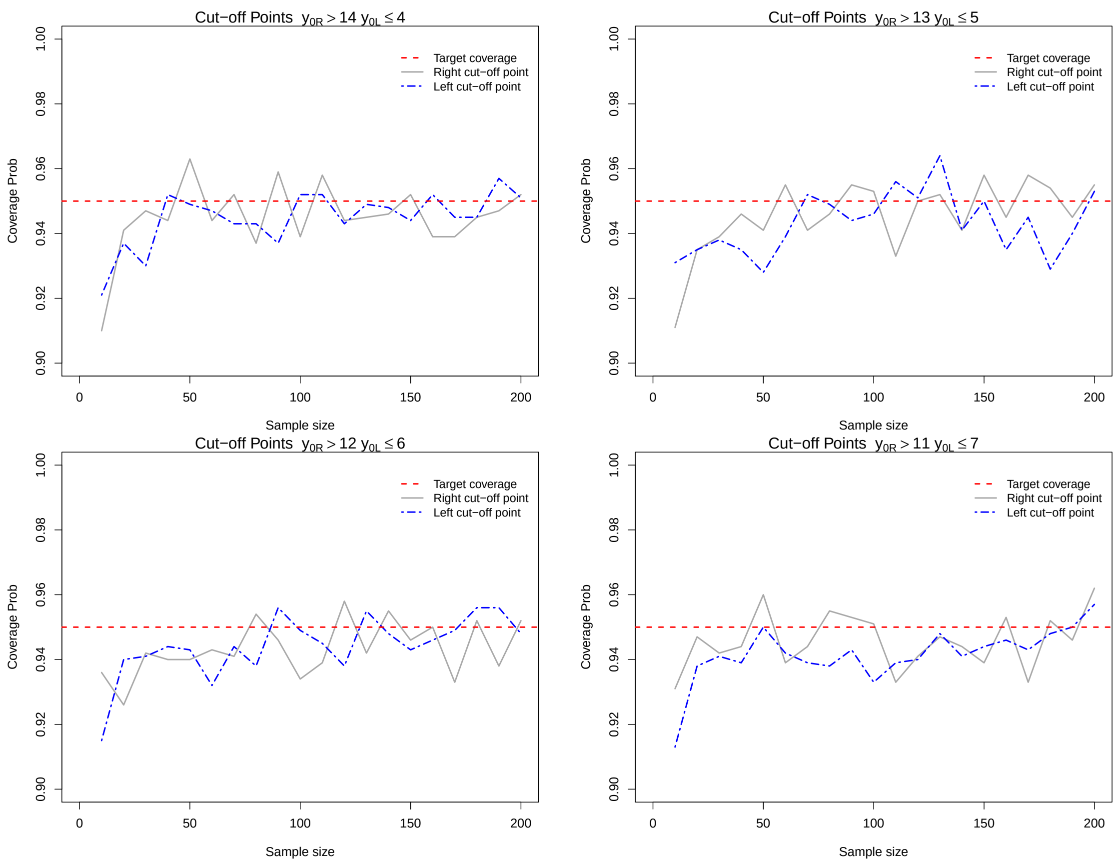

Figure 3.10. Simulated coverage probabilities for different sample sizes and cut-off points.

3.9 Data Example

Yearly aggregated data on the number of road crash fatalities for 130 different regions of New South Wales (NSW) for years 2011-2016 were provided by the NSW Centre for Road Safety (Transport for NSW, 2018). Note there are plans for enhancing NSW road fatality data through the linkage of multiple data sets. This may impact previously published data and would explain discrepancies between the data used in here and future data releases.

RTM is unlikely to occur when average yearly fatalities are small as the Poisson variance would also be small and relatively high number of annual fatalities would be inconsistent with a constant parameter vector $\boldsymbol{\theta}$ as assumed in the RTM derivations. For the purposes of this analytic demonstration, regions with less than two fatalities per year and regions with relatively large annual fatalities have been excluded. These exclusions have resulted in data from 67 regions for the analysis.

For the NSW road fatality data, successive observations within a region are assumed to be correlated, whereas different regions and observations more than one year apart are assumed to be independent. This appears to be a reasonable assumption as the estimated autocorrelation for successive observations was 0.226 for this data set, and 0.102, 0.114, 0.129 and 0.049 for lags of 2, 3, 4 and 5 years respectively.

For a cut-off point $y_0 = 2$, the estimates of the parameters of the truncated bivariate Poisson distribution are $\hat{\theta}_0 = 0.76$, $\hat{\theta}_1 = 3.938$ and $\hat{\theta}_2 = 3.585$. Based on these estimates, the total, intervention and RTM effects against different cut-off values are given in Figure 3.11.

Areas where the initial number of casualties were eight or more (left panel) experienced a decrease of almost five casualties on average the next year which are mostly due to the RTM effect with a contribution of more than four on average. The opposite effect was observed in places where casualties happened infrequently (right panel). For example, areas where the number of casualties were two or fewer, experienced on average an increase of more than two casualties the next year which is mainly due to RTM. Generally, as the cut-off point goes farther in

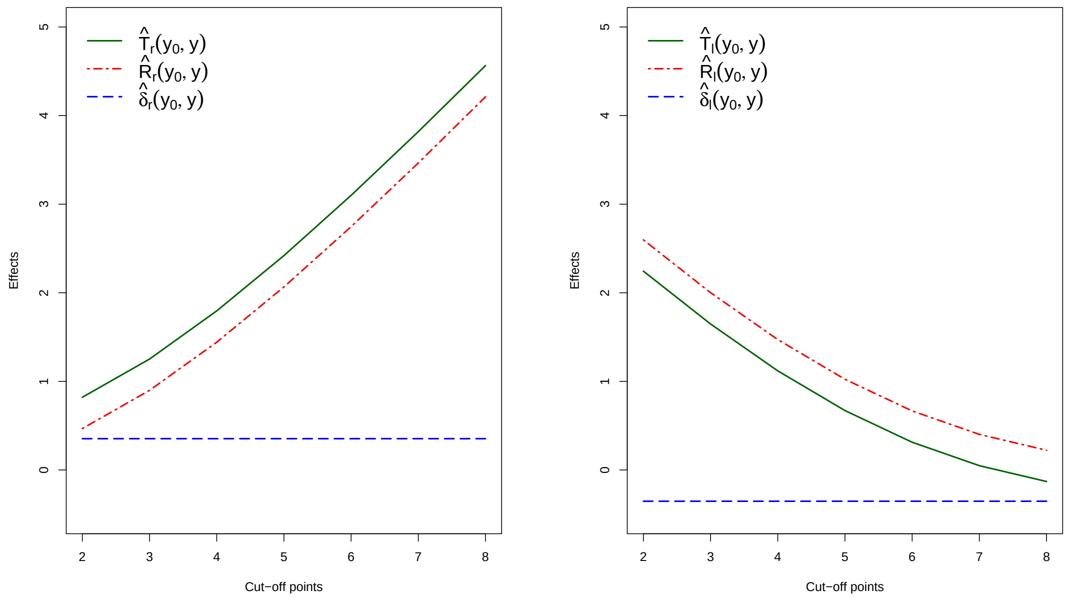

Figure 3.11. Left panel: Graph for the RTM effects for points greater than y_0, Right panel: Graph for the RTM effects for points less than or equal to y_0.

the tail, the RTM effect increases causing the total effect to increase, whereas the treatment effect remains constant. Consequently, an observed average increase or average decrease which is the additive effect of the RTM and treatment, may be mistaken for a real change.

3.10 Log-transformation of Poisson distributed data

Log-transformation is a widely used tool for dealing with positively skewed data in different research areas, e.g, image processing and biomedical research. The resulting observations may be well approximated by a normal distribution, thereby allowing for methods and formulae based on a normal assumption.

Log-transformation can be problematic when zeros have been observed, which makes it difficult to estimate RTM when selection is based on subjects below a certain threshold, i.e. $y \leq y_0$, for log-transformed data. However, if selection is based on subjects above a certain threshold (and zero has not been observed in the data set), then RTM can be evaluated for log-transformed data. Moreover, Feng et al. (2014) highlighted the limitations of log-transformation when dealing with skewed data.

For our data example, we considered the cut-off points of 7 and 8 because road safety interventions are not enacted when there are too few fatalities. RTM was calculated against these cut-off points for log-transformed data using maximum likelihood. The results obtained are given in Table 3.1. The estimated RTM effect was exponentiated to get it back in the original units. These estimates were then compared to those computed from the untransformed data. The percentage relative change was calculated using the formula by Tornqvist et al. (1985)

$$PRC = (RTM - \exp(\log RTM))\,/RTM \times 100\%.$$

The respective percentage relative change ranged from 46% to 53% for cut-off points 7 and 8. This example suggests that RTM formulae assuming the bivariate normal distribution on log-transformed data can severely underestimate RTM when the data are generated from a bivariate Poisson distribution.

Table 3.1. Comparison of RTM for the NSW road fatality log-transformed data

		Estimates	
	Formula	$y_0 = 7$	$y_0 = 8$
RTM for log-transformed data	$\log R(y_0)$	0.715	0.839
Exponentiated logRTM	$\exp(\log R(y_0))$	2.044	2.315
RTM for the Original data	$R(y_0)$	3.818	5.000
Difference	$R(y_0) - \exp(\log R(y_0))$	1.774	2.685
Percentage relative change	PRC	46.454$\%$	53.696$\%$

To check the amount of bias for the log transformed data, a simulation study was carried out for different sample sizes and parameters $\theta_0 = 6$, $\theta_1 = 3$, $\theta_2 = 3$ and $y_0 = 12$ (a right cut off point). The log of the true RTM, i.e., $\log(R(y_0))$ is 0.564 for these parameters represented by the green line in Figure 3.12. The RTM effect for the log transformed data was estimated using the log cut off point $\log(y_0) = 2.48$ and the sampling distribution of estimates are presented in Figure 3.12. The mean of the sampling distribution of $\log \hat{R}(y_0)$ for the log transformed data was 0.15 (the red line segments in Figure 3.12) and the corresponding percentage relative difference observed was 73%.

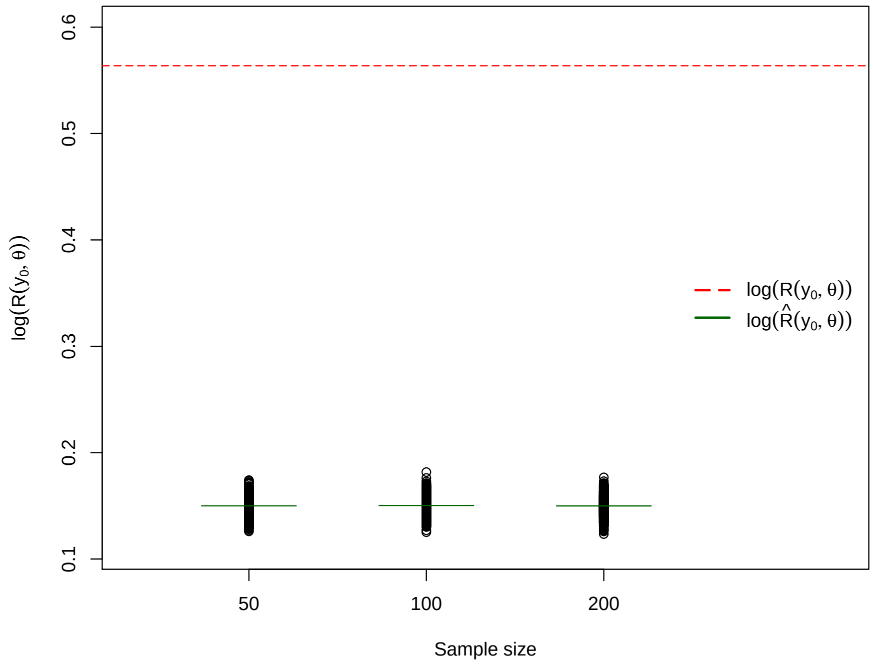

Figure 3.12. The estimated RTM effects for simulated log-transformed data for sample sizes $n = 50, 100, 200$.

3.11 Discussion

Regression to the mean is an important issue in data analysis that can lead to erroneous conclusions and therefore warrants consideration. For the normal distribution, expressions for RTM are available in the literature. However, there are many situations where the underlying distribution is Poisson. The evaluation of the impact of any intervention or policy-change aimed at changing the rate of occurrence could be improved by accounting for potential RTM effects. Therefore, quantification of the RTM effect for the Poisson distribution/process is an important research problem.

In a pre/post study design when an intervention or treatment is applied to subjects selected based on certain thresholds, RTM is likely to occur. The severity of RTM increases as the cut-off point is farther into the tail of the baseline distribution. The intervention or treatment effect can be estimated by decomposing the total effect into RTM and the intervention/treatment effects.

Our derivations assuming the bivariate Poisson differ from the bivariate normal in terms of the influence of covariance. For the normal distribution, the RTM effect decreases linearly as the covariance/correlation between variables before and after an intervention increases. On the contrary for the Poisson distribution, as the covariance/correlation increases, RTM for left and right cut-off points behave differently. As the covariance increases, RTM decreases non-linearly for a right cut-off point, whereas RTM increases for a left cut-off point. A possible reason for this difference is the equality of mean and variance of the Poisson distribution.

A log-transformation is often useful for positively skewed data as it may result in observations that are well approximated by a normal distribution thereby allowing for methods and formulae based on a normal assumption. For the NSW road fatality data and the simulated bivariate Poisson data, RTM estimates were computed assuming a bivariate normal distribution after a log transformation. In both instances, RTM was severely underestimated using this approach and therefore the log-transformation is not recommended when estimating RTM.

Our simulation study suggests that the maximum likelihood estimators of RTM are not only consistent and unbiased, but also approximately normally distributed confirming the asymptotic results.

Further, the behaviour of the RTM effect is markedly dissimilar for homogeneous and inhomogeneous Poisson processes which can be easily corroborated. It is therefore recommended to take into account the varying nature of arrival rates of Poisson processes for calculating the RTM effects in a data analysis.

Chapter 4

Regression to the mean for the bivariate binomial distribution

The binomial distribution is often used to describe the number of successes in a fixed number of trials. In an intervention study, the pre-post variables for number of successes may follow the bivariate binomial distribution. This chapter derives expression for RTM when the underlying distribution is the bivariate binomial. It highlights the differences resulting from the dependence structure of the true and random error components, and its impact on the intervention/treatment effect and the correlation.

This chapter also demonstrates that RTM is underestimated when normal and Poisson approximations to the bivariate binomial distribution are used. The maximum likelihood estimates of the total, RTM, and intervention effects are derived and the statistical properties of unbiasedness, consistency, and asymptotic normality are established. A simulation study is conducted to empirically assess the statistical properties of the RTM estimator and its asymptotic distribution.

Data on the number of obese individuals and the number of nonconforming cardboard cans are used to decompose the total effect into the RTM and intervention effects. The contents of this chapter are reproduced from a published paper (Khan and Olivier, 2019) with some minor modifications.

4.1 Introduction

The conclusions of pre-post intervention studies may be influenced by RTM when subjects are selected in the tail of a distribution. One approach to mitigate this issue is to quantify the RTM effect and subtract it from the total effect. When the pre-post variables follow the bivariate normal distribution, James (1973), Gardner and Heady (1973), and Davis (1976) derived formulae for calculating the expected RTM effect. Similarly, Khan and Olivier (2018) derived expressions for the RTM effect assuming the bivariate Poisson distribution for pre and post counts.

Many real life situations exist where the response variables are binary and could be decomposed into two components that generate the event of success. This decomposition includes (i) the true source (ii) and random fluctuations or luck/chance. For instance, in a standardized test, the number of correct answers scored by a candidate can be decomposed into the questions the student knows the answer and questions the student does not know the answer and guesses. In another example, the prevalence of obesity could be decomposed into two sources: (1) individual factors such as genetics and personal choices, and (2) collective behaviour such as social pressure and global economic drivers (Gallos et al., 2012). Similarly, skill and luck/chance play important roles in sports (Frans, 1985; Filip, 2014). The total number of matches a team wins in a fixed number of games could be the sum of the matches won by skill and those matches determined by luck/chance.

Other examples exist where the characteristic of interest is binary which could be decomposed into two different sources. For example, the Government of Khyber Pakhtunkhwa, Pakistan has established independent monitoring units to regularly evaluate the performance of public sector schools and hospitals (IMU, 2018a,b). One of the many objectives of this organization is to ensure the presence of the working staff for each day. The total absentee days by an employee in a month is the sum of official leaves and unauthorized absences, and the government makes decisions based on this data. Similarly, in statistical process control, the p-chart is used for monitoring the fraction of defective items in the manufacturing process (Montgomery, 2013). The total number of defective items produced could be due to some assignable cause or to chance variation.

In the above examples, the outcome of interest is the number of successes in a fixed number of trials, e.g., obese individuals, correct answers, wins, absentees, or defective items, which may follow the binomial distribution. The random component part in each case could induce the RTM effect. The conclusions of intervention studies such as the effectiveness of a program at reducing the prevalence of obesity (Hannon et al., 2018; Skinner et al., 2015), improving student's performance on a standardized test (Rothman and Henderson, 2011; Good et al., 2003), key decisions about changing strategies in sports for improving performance, and interrupting manufacturing processes for decreasing the production of defective items could be impacted by RTM and, therefore, estimates of the effect of an intervention may be inaccurate.

Further, normal or Poisson approximations to the binomial distribution are appropriate under certain conditions which may not always hold true. That is, the estimation of RTM under normal or Poisson approximations to the binomial distribution could be invalid. In pre/post studies involving binomial experiments, the quantification of RTM is missing in the literature. Therefore, the purpose of this chapter is to derive expressions to quantify the RTM effect when the underlying distribution of pre/post observations is a bivariate binomial distribution.

The remainder of the chapter is organized into seven sections. Formulae quantifying RTM effects are derived under the assumption of the bivariate binomial distribution in Section 4.2. The effect of the correlation between pre and post observations on RTM is studied in Section 4.3 and comparisons of RTM under the bivariate binomial distribution and normal or Poisson approximations to the binomial are carried out in Section 4.4. Section 4.5 is devoted to estimation of the RTM effect, and a simulation study is conducted in Section 4.6 to investigate the statistical properties and sampling distribution of the RTM estimator. Data examples for the number of obese individuals and the number of nonconforming cardboard cans is demonstrated in Section 4.7. A discussion in Section 4.8 concludes the chapter.

4.2 The Bivariate Binomial Distribution and Regression to the Mean

The normal and Poisson distributions are, respectively, important continuous and discrete probability distributions in statistics as they are relevant in a wide range of applications. There are also situations which can be modelled as binomial, e.g., the number of correct answers scored by a student, the number of matches won by a team, and the number of defective items produced in a manufacturing process in a fixed number of trials.

Similarly, in medicine and public health, the number of obese, the number of patients with allergies reporting symptomatic relief with a specific medication, and the number of coronary stents successfully transplanted in a fixed number of patients selected for treatment may follow the binomial distribution. Apart from this, normal or Poisson approximations to the binomial are not always appropriate, thus making the quantification of RTM under the bivariate binomial distribution a relevant problem to study.

In a set up similar to the bivariate Poisson, let $Y_1 = X_0^{(1)} + X_1$ and $Y_2 = X_0^{(2)} + X_2$ be the total number of successes in a pre/post study design with a fixed number of trials, say n. Here, $X_0^{(i)}$ represents the true number of successes, and X_i are random numbers of successes due to luck/chance, for $i = 1, 2$. For example, in a standardized test with multiple choice questions, $X_0^{(i)}$ would be the number of correct answers that the student knows and X_i for $i = 1, 2$, would be the number of correct answers from guessing. Here, $X_0^{(i)} \sim Bin(n, \pi_0)$ and the conditional distribution of X_i given that $X_0^{(i)} = x_0$ is also binomial, i.e., $X_i|x_0 \sim Bin(n - x_0, \pi_i)$ where $\pi_0 = P(X_0^{(i)} = 1)$ and $\pi_i = P(X_i = 1|X_0^{(i)} = x_0)$ for $i = 1, 2$.

In the normal and Poisson set up, the true and random component of measurements/counts are independent of each other, whereas they are not in the binomial case presented here. Y_i is the sum of two dependent binomial random variables $X_0^{(i)}$ and X_i, so its distribution is not straightforward. To derive the distribution of

Y_i, consider the joint distribution of $X_0^{(i)}$ and X_i given by

$$\begin{aligned} f_{X_0^{(i)},X_i}(x_0, x_i) &= f_{X_0^{(i)}}(x_0) f_{X_i|x_0}(x_i|x_0) \\ &= \binom{n}{x_0} \pi_0^{x_0} (1-\pi_0)^{n-x_0} \binom{n-x_0}{x_i} \pi_i^{x_i} (1-\pi_i)^{n-x_0-x_i}, \end{aligned}$$

where $x_0 = 0, 1, ..., n$, $x_1 = 0, 1, ..., n - x_0$ and $0 \leq X_0^{(i)} + X_i \leq n$.

The probability generating function (PGF) of Y_i is

$$\begin{aligned} P_{Y_i}(s) &= E(s^{Y_i}) = E(s^{X_0^{(i)}+X_i}) \\ &= \sum_{x_0=0}^{n} \sum_{x_i=0}^{n-x_0} s^{X_0^{(i)}+X_i} \binom{n}{x_0} \pi_0^{x_0} (1-\pi_0)^{n-x_0} \binom{n-x_0}{x_i} \pi_i^{x_i} (1-\pi_i)^{n-x_0-x_i} \\ &= \sum_{x_0=0}^{n} \binom{n}{x_0} (s\pi_0)^{x_0} (1-\pi_0)^{n-x_0} \sum_{x_i=0}^{n-x_0} \binom{n-x_0}{x_i} (s\pi_i)^{x_i} (1-\pi_i)^{n-x_0-x_i}. \end{aligned}$$

Summing the series first with respect to x_i and then x_0, we get

$$P_{Y_i}(s) = (s(\pi_0 + (1-\pi_0)\pi_i) + (1-\pi_0)(1-\pi_i))^n,$$

which is the PGF of a binomial distribution, i.e., $Y_i \sim Bin(n, \pi_0 + (1-\pi_0)\pi_i)$.

The component $(1-\pi_0)\pi_i$ can be interpreted as the probability of success due to chance. In this case, the total probability of success on an individual item in a pre/post trial cannot be explicitly decomposed into two parts like the parameters of the bivariate Poisson distribution. However, the notations can be eased by re-parametrizing them according to the outcomes of the bivariate Bernoulli distribution (Marshall and Olkin, 1985). Let Z_i be Bernoulli distributed random variables for $i = 1, 2$, then (Z_1, Z_2) has the four possible outcomes $(1,1), (1,0), (0,1)$ and $(0,0)$ in a bivariate set up. The probabilities of these outcomes are the sum of mutually exclusive events

$$\begin{aligned} P\{(Z_1, Z_2) = (1,1)\} &= \phi_0 = p_{TT} + p_{TR} + p_{RT} + p_{RR}, \\ P\{(Z_1, Z_2) = (1,0)\} &= \phi_1 = p_{T0} + p_{R0} \\ P\{(Z_1, Z_2) = (0,1)\} &= \phi_2 = p_{0T} + p_{0R} \\ P\{(Z_1, Z_2) = (0,0)\} &= \phi_3 = p_{00}, \end{aligned}$$

where the subscripts T and R denote successes generated from the true and random sources respectively, whereas 0 represents a failure and $\phi_3 = 1 - \phi_0 - \phi_1 - \phi_2$.

In a pre/post set up, let Z_1 and Z_2 be the pre and post Bernoulli variables. The probability of success in pre/post marginal Bernoulli trials are

$$P(Z_1 = 1) = \phi_0 + \phi_1,$$

and

$$P(Z_2 = 1) = \phi_0 + \phi_2.$$

Note that since there are only two possible outcomes of a Bernoulli trial, the selection of subjects for a pre-post study design can be based on either the presence (i.e., $Z_1 = 1$) or absence (i.e., $Z_1 = 0$) of a characteristic of interest.

The binomial random variables Y_1 and Y_2 can be represented as the sums of Bernoulli random variables as

$$Y_1 = \sum_{i=1}^{n} Z_{1i}, \quad Y_2 = \sum_{i=1}^{n} Z_{2i}.$$

In this new pre/post set up, $Y_i \sim Bin(n, \phi_0 + \phi_i)$ is an equivalent form of the distribution of the total number of successes Y_i, for $i = 1, 2$.

The classification of the pre/post number of successes in n bivariate Bernoulli trials are presented in Table 4.1. Here, α denotes the number of successes on both the pre and post occasions, i.e., $\alpha = |(Z_1, Z_2) = (1, 1)|$.

Table 4.1. 2×2 table for successes and failures in a distribution

	Pre successes	Pre failures	Totals
Post successes	α	$y_2 - \alpha$	y_2
Post failures	$y_1 - \alpha$	$n + \alpha - y_1 - y_2$	$n - y_2$
Totals	y_1	$n - y_1$	n

The joint distribution of Y_1 and Y_2, first discussed by Aitken and Gonin (1936), is

$$f_{Y_1,Y_2}(y_1, y_2, n) = \sum_{\alpha=0}^{\min(y_1,y_2)} f(\alpha, y_1 - \alpha, y_2 - \alpha, \phi_0, \phi_1, \phi_1, n),$$

where

$$f(\alpha, y_1 - \alpha, y_2 - \alpha, n) = \binom{n}{\alpha, y_1 - \alpha, y_2 - \alpha, n + \alpha - y_1 - y_2} \phi_0^{\alpha} \phi_1^{y_1 - \alpha} \phi_2^{y_2 - \alpha} (1 - \phi_0 - \phi_1 - \phi_2)^{n + \alpha - y_1 - y_2}$$

is a multinomial-type probability mass function. The covariance of Y_1 and Y_2 in this set up is $\text{cov}(Y_1, Y_2) = n(\phi_0 - (\phi_0 + \phi_1)(\phi_0 + \phi_2))$.

4.2.1 The Total Effect Under the Bivariate Binomial Distribution

In pre/post intervention studies, RTM may arise when subjects beyond/below a baseline point, say y_0, are selected. For example, schools/hospitals in Khyber Pukhtunkhwa, Pakistan, where absentees of staff members are greater than y_0 in a month triggers interventions like salary deduction and on-site inspections. Two situations arise as a result. Firstly, the intervention could be ineffective and the effect could be due to RTM. In other words, Y_1 and Y_2 are identically distributed with $\phi_1 = \phi_2$. Secondly, the intervention could be effective and the observed change could be a combination of RTM and intervention effects. Under the latter scenario, Y_1 and Y_2 are not necessarily identically distributed.

An intervention/treatment is potentially applied to extreme situations based on a cut-off value y_0, which could be either in the left or right tail of the distribution. The shape of the binomial distribution is asymmetric for most parametric values, so right and left cut-off points are considered separately. Based on a right cut-off point, let $T_r(y_0, \boldsymbol{\phi})$ be the total effect which is the difference of the conditional means of pre and post variables. Mathematically, this is

$$T_r(y_0, \boldsymbol{\phi}) = E(Y_1 - Y_2 | Y_1 > y_0), \tag{4.1}$$

where $\boldsymbol{\phi} = (\phi_0, \phi_1, \phi_2)$.

The truncated bivariate binomial distribution is

$$f_t(y_1, y_2, n) = \begin{cases} \frac{f_{Y_1,Y_2}(y_1,y_2,n)}{P(Y_1>y_0)}, & \text{if } Y_1 = y_0+1, y_0+2, ..., n \text{ and } Y_2 = 0, 1, ..., n \\ 0 & \text{otherwise.} \end{cases}$$

Considering the conditional expectation of Y_1,

$$E(Y_1 | Y_1 > y_0) = \sum_{y_1=y_0+1}^{n} y_1 \sum_{y_2=\alpha}^{n-y_1+\alpha} \sum_{\alpha=0}^{y_1} f(\alpha, y_1 - \alpha, y_2 - \alpha, n) \Big/ P(Y_1 > y_0)$$

and using the identity

$$\sum_{y_2=\alpha}^{n-y_1+\alpha} \sum_{\alpha=0}^{y_1} f(\alpha, y_1 - \alpha, y_2 - \alpha, n) = \binom{n}{y_1} (\phi_0 + \phi_1)^{y_1} (1 - \phi_0 - \phi_1)^{n-y_1},$$

it can be shown that this expression simplifies to

$$\begin{aligned} E(Y_1|Y_1 > y_0) &= \sum_{y_1=y_0+1}^{n} y_1 \binom{n}{y_1} (\phi_0+\phi_1)^{y_1}(1-\phi_0-\phi_1)^{n-y_1} \Big/ P(Y_1 > y_0) \\ &= n(\phi_0+\phi_1) \cdot \frac{1-F_{n-1}(y_0-1|\phi_0+\phi_1)}{1-F_n(y_0|\phi_0+\phi_1)}, \end{aligned} \tag{4.2}$$

where $F_n(y|p) = \sum_{t=0}^{y} \binom{n}{t} p^t (1-p)^{n-t}$ is the cumulative distribution function (CDF) of the binomial distribution. Now, considering the conditional expectation of Y_2, we have

$$E(Y_2|Y_1 > y_0) = \sum_{y_1=y_0+1}^{n} \sum_{y_2=\alpha}^{n-y_1+\alpha} \sum_{\alpha=0}^{y_1} y_2 f(\alpha, y_1-\alpha, y_2-\alpha, n) \Big/ P(Y_1 > y_0). \tag{4.3}$$

Expanding the inner summations,

$$\sum_{y_2=0}^{n-y_1} y_2 f(0, y_1, y_2, n) + \sum_{y_2=1}^{n-y_1+1} y_2 f(1, y_1-1, y_2-1, n) + \cdots + \sum_{y_2=y_1}^{n} y_2 f(y_1, y_1-y_1, y_2-y_1, n)$$

then substituting $y_2' = y_2 - i$ for $i = 1, 2, ..., y_1$ recursively, rearranging the expression and using the binomial theorem $\sum_{i=1}^{n} \binom{n}{i} a^x b^{n-x} = (a+b)^n$, the inner summations reduce to

$$\binom{n}{y_1}(n-y_1)\cdot\phi_2(1-\phi_0-\phi_1)^{n-y_1-1}(\phi_0+\phi_1)^{y_1} + \binom{n}{y_1} y_1 \cdot \phi_0 (1-\phi_0-\phi_1)^{n-y_1}(\phi_0+\phi_1)^{y_1-1}.$$

Substituting this result into equation (4.3), we get

$$\begin{aligned} E(Y_2|Y_1 > y_0) = \Bigg[&\sum_{y_1=y_0+1}^{n} \binom{n}{y_1}(n-y_1)\frac{\phi_2}{1-\phi_0-\phi_1}(1-\phi_0-\phi_1)^{n-y_1}(\phi_0+\phi_1)^{y_1} + \\ &\sum_{y_1=y_0+1}^{n} \binom{n}{y_1} y_1 \frac{\phi_0}{\phi_0+\phi_1}(1-\phi_0-\phi_1)^{n-y_1}(\phi_0+\phi_1)^{y_1} \Bigg] \Big/ P(Y_1 > y_0). \end{aligned}$$

Simplifying and using equation (4.2), we get

$$E(Y_2|Y_1 > y_0) = n(\phi_0+\phi_1) \cdot \frac{1-F_{n-1}(y_0-1|\phi_0+\phi_1)}{1-F_n(y_0|\phi_0+\phi_1)} \cdot \frac{\phi_0-(\phi_0+\phi_1)(\phi_0+\phi_2)}{(\phi_0+\phi_1)(1-\phi_0-\phi_1)} + n\frac{\phi_2}{1-\phi_0-\phi_1}. \tag{4.4}$$

Combining equations (4.2) and (4.4) in equation (4.1), and using the recursive relation

$$F_n(y_0|\phi_0+\phi_1) = F_{n-1}(y_0-1|\phi_0+\phi_1) + (1-\phi_0-\phi_1)P_{n-1}(Y_1 = y_0),$$

the expression for $T_r(y_0, \boldsymbol{\phi})$ is

$$T_r(y_0, \boldsymbol{\phi}) = n \cdot \frac{\phi_1(1-F_{n-1}(y_0-1|\phi_0+\phi_1)) - \phi_2\Big[1-F_n(y_0|\phi_0+\phi_1)-(\phi_0+\phi_1)P_{n-1}(Y_1 = y_0)\Big]}{1-F_n(y_0|\phi_0+\phi_1)}, \tag{4.5}$$

where $P_{n-1}(Y_1 = y_0) = \binom{n-1}{y_0}(\phi_0 + \phi_1)^{y_0}(1 - \phi_0 - \phi_1)^{n-1-y_0}$.

If selection criterion is based on all subjects equal to or less than a cut point, then $T_\ell(y_0, \boldsymbol{\phi})$ can be quantified by evaluating the difference of conditional means as

$$T_\ell(y_0, \boldsymbol{\phi}) = E(Y_2 - Y_1 | Y_1 \le y_0).$$

Following similar steps, the expression for $T_\ell(y_0, \boldsymbol{\phi})$ is

$$T_\ell(y_0, \boldsymbol{\phi}) = n \cdot \frac{\phi_2\Big[F_n(y_0|\phi_0 + \phi_1) + (\phi_0 + \phi_1)P_{n-1}(Y_1 = y_0)\Big] - \phi_1 F_{n-1}(y_0 - 1|\phi_0 + \phi_1)}{F_n(y_0|\phi_0 + \phi_1)}. \tag{4.6}$$

4.2.2 RTM and intervention effects under the bivariate binomial distribution

Let $R_i(y_0, \boldsymbol{\phi})$ and $\delta_i(\boldsymbol{\phi})$ be, respectively, the RTM and intervention effects for $i = r, \ell$. Using the fact $E(Y_i) = n(\phi_0 + \phi_i)$, for $i = 1, 2$, $\delta_r(\boldsymbol{\phi})$ can be expressed as

$$\delta_r(\boldsymbol{\phi}) = E(Y_1 - Y_2) = n(\phi_1 - \phi_2).$$

For a null intervention effect, $\delta_i(\boldsymbol{\phi}) = 0$ or $\phi_1 = \phi_2$, and the total effect is due to RTM. So, expressions of RTM for the right and left cut-off points can be derived by substituting $\phi_2 = \phi_1$ in equations (4.5) and (4.6) as

$$R_r(y_0, \boldsymbol{\phi}) = E(Y_1 - Y_2 | Y_1 > y_0, \phi_1 = \phi_2) = n\phi_1 \cdot \frac{P_{n-1}(Y_1 = y_0)}{1 - F_n(y_0|\phi_0 + \phi_1)}, \tag{4.7}$$

and

$$R_\ell(y_0, \boldsymbol{\phi}) = E(Y_1 - Y_2 | Y_1 \le y_0, \phi_1 = \phi_2) = n\phi_1 \cdot \frac{P_{n-1}(Y_1 = y_0)}{F_n(y_0|\phi_0 + \phi_1)}. \tag{4.8}$$

For the bivariate normal and Poisson distributions, the total effect $T_r^{P,N}(y_0, \boldsymbol{\theta})$ can be expressed as the sum of the RTM and intervention effects

$$T_r^{P,N}(y_0, \boldsymbol{\theta}) = R_r^{P,N}(y_0, \boldsymbol{\theta}) + \delta_r^{P,N}(\boldsymbol{\theta}), \tag{4.9}$$

where $R_r^{P,N}(y_0, \boldsymbol{\theta})$ and $\delta_r^{P,N}(\boldsymbol{\theta})$ are RTM and intervention effects, respectively. For a non-null case, $\delta_r^{P,N}(\boldsymbol{\theta})$ can be obtained simply by subtracting RTM from the total effect.

For the bivariate binomial distribution and a right cut-off point, the difference of the total and RTM effects can be written as

$$T_r(y_0, \boldsymbol{\phi}) - R_r(y_0, \boldsymbol{\phi}) = n(\phi_1 - \phi_2) \cdot \left[1 - (\phi_0 + \phi_1)\frac{P_{n-1}(Y_1 = y_0)}{1 - F_n(y_0|\phi_0 + \phi_1)}\right]$$
$$= \delta_r(\boldsymbol{\phi}) \cdot \Big[1 - B_r(y_0, \phi_0 + \phi_1)\Big], \tag{4.10}$$

where

$$B_r(y_0, \phi_0 + \phi_1) = (\phi_0 + \phi_1)\frac{P_{n-1}(Y_1 = y_0)}{1 - F_n(y_0|\phi_0 + \phi_1)}$$

is a factor by which the intervention effect is underestimated because $B_r(y_0, \phi_0 + \phi_1) \in (0, 1)$. Simple subtraction does not work for the bivariate binomial distribution to get the unbiased intervention effect from the total and RTM effects, and instead can be obtained as

$$\delta_r(\boldsymbol{\phi}) = \frac{T_r(y_0, \boldsymbol{\phi}) - R_r(y_0, \boldsymbol{\phi})}{1 - B_r(y_0, \phi_0 + \phi_1)} = n(\phi_1 - \phi_2).$$

The non-equivalence to the difference of total and RTM effects makes the bivariate binomial distribution distinct from the bivariate normal and Poisson distributions which could be attributed to the dependency of $X_0^{(i)}$ and X_i for $i = 1, 2$. Further, it can be shown that

$$\lim_{n\to\infty} F_{n-1}(y_0 - 1|\phi_0 + \phi_1) \longrightarrow F_n(y_0|\phi_0 + \phi_1)$$

and

$$\lim_{n\to\infty} B_r(y_0, \phi_0 + \phi_1) \longrightarrow 0$$

and hence for large n,

$$T_r(y_0, \boldsymbol{\phi}) - R_r(y_0, \boldsymbol{\phi}) \approx n(\phi_1 - \phi_2).$$

Therefore, $X_0^{(i)}$ and X_i are asymptotically independent for $i = 1, 2$.

Similarly, for a left cut-off point, the intervention effect $\delta_\ell(\boldsymbol{\phi})$, is

$$\delta_\ell(\boldsymbol{\phi}) = \frac{T_\ell(y_0, \boldsymbol{\phi}) - R_\ell(y_0, \boldsymbol{\phi})}{1 + B_\ell(y_0, \phi_0 + \phi_1)},$$

where

$$B_\ell(y_0, \phi_0 + \phi_1) = (\phi_0 + \phi_1) \cdot \frac{P_{n-1}(Y_1 = y_0)}{F_n(y_0|\phi_0 + \phi_1)} > 0$$

and the intervention effect would be overestimated if RTM is only subtracted from the total effect.

4.2.3 Variance of RTM

Expressions for the variance of $(Y_1 - Y_2|Y_1 > y_0)$ can be obtained by evaluating the conditional variance of the difference of pre and post variables given by

$$\text{var}(Y_1-Y_2|Y_1 > y_0) = \text{var}(Y_1|Y_1 > y_0)+\text{var}(Y_2|Y_1 > y_0)-2\times\text{cov}(Y_1, Y_2|Y_1 > y_0). \quad (4.11)$$

To complete the derivation, some helpful results are derived as follows. First,

$$E(Y_1(Y_1-1)|Y_1 > y_0) = \sum_{y_1=y_0+1}^{n} y_1(y_1-1) \sum_{y_2=\alpha}^{n-y_1+\alpha} f_T(y_1, y_2, n)$$

using the identity $\sum_{y_2=\alpha}^{n-y_1+\alpha} f_{Y_1,Y_2}(y_1, y_2, n) = \binom{n}{y_1}(\phi_0+\phi_1)^{y_1}(1-\phi_0-\phi_1)^{n-y_1}$, and then simplifying, we get

$$E(Y_1(Y_1-1)|Y_1 > y_0) = n(n-1)(\phi_0+\phi_1)^2 \cdot \frac{1-F_{n-2}(y_0-2|\phi_0+\phi_1)}{1-F_n(y_0|\phi_0+\phi_1)}. \quad (4.12)$$

Similarly, for a left cut-off point, it can be shown that

$$E(Y_1|Y_1 \le y_0) = n(\phi_0+\phi_1) \cdot \frac{F_{n-1}(y_0-1|\phi_0+\phi_1)}{F_n(y_0|\phi_0+\phi_1)}, \quad (4.13)$$

and

$$E(Y_1(Y_1-1)|Y_1 \le y_0) = n(n-1)(\phi_0+\phi_1)^2 \cdot \frac{F_{n-2}(y_0-2|\phi_0+\phi_1)}{F_n(y_0|\phi_0+\phi_1)}. \quad (4.14)$$

Now consider

$$E(Y_2(Y_2-1)|Y_1 > y_0) = \sum_{y_1=y_0+1}^{n} \sum_{y_2=\alpha}^{n-y_1+\alpha} y_2(y_2-1) f_T(y_1, y_2). \quad (4.15)$$

Expanding the inner summation, rearranging and simplifying, we get

$$\begin{aligned} &E(Y_2(Y_2-1)|Y_1 > y_0) \\ &= E(Y_1(Y_1-1)|Y_1 > y_0) \cdot \frac{(\phi_1\phi_2-\phi_0(1-\phi_0-\phi_1-\phi_2))^2}{(\phi_0+\phi_1)^2(1-\phi_0-\phi_1)^2} + n(n-1)\frac{\phi_2^2}{(1-\phi_0-\phi_1)^2} \\ &- 2\phi_2 \cdot \frac{n(\phi_1\phi_2-\phi_0(1-\phi_0-\phi_1-\phi_2))-\phi_1\phi_2}{(\phi_0+\phi_1)(1-\phi_0-\phi_1)^2} \cdot E(Y_1|Y_1 > y_0). \end{aligned} \quad (4.16)$$

The crossproduct expectation can be found by

$$\begin{aligned} E(Y_1Y_2|Y_1 > y_0) &= \sum_{y_1=y_0+1}^{n} y_1 \sum_{y_2=\alpha}^{n-y_1+\alpha} y_2 f_T(y_1, y_2, n), \\ &= \frac{\phi_0-(\phi_0+\phi_1)(\phi_0+\phi_2)}{(\phi_0+\phi_1)(1-\phi_0-\phi_1)} \cdot E(Y_1(Y_1-1)|Y_1 > y_0) \\ &+ \left(\frac{\phi_0}{\phi_0+\phi_1} + \frac{(n-1)\phi_2}{1-\phi_0-\phi_1}\right) \cdot E(Y_1|Y_1 > y_0), \end{aligned} \quad (4.17)$$

by using the identity

$$\sum_{y_2=\alpha}^{n-y_1+\alpha} y_2 f_T(y_1, y_2, n) = \binom{n}{y_1}(n-y_1)\frac{\phi_2}{1-\phi_0-\phi_1}(1-\phi_0-\phi_1)^{n-y_1}(\phi_0+\phi_1)^{y_1} + \binom{n}{y_1}y_1\frac{\phi_0}{\phi_0+\phi_1}(1-\phi_0-\phi_1)^{n-y_1}(\phi_0+\phi_1)^{y_1}.$$

Using the formulae of variances and covariance of Y_1 and Y_2, i.e., $\text{var}(Y_i|Y_1 > y_0) = E(Y_i(Y_i-1)|Y_1 > y_0) + E(Y_i|Y_1 > y_0) - (E(Y_i|Y_1 > y_0))^2$ for $i = 1, 2$ and $\text{cov}(Y_1, Y_2|Y_1 > y_0) = E(Y_1Y_2|Y_1 > y_0) - E(Y_1|Y_1 > y_0)E(Y_2|Y_1 > y_0)$, the expression of variance in equation (4.11) simplifies to

$$\begin{aligned}
&\text{var}(Y_1 - Y_2|Y_1 > y_0)\\
&= n\left(\frac{\phi_1 + (\phi_0+\phi_1)(\phi_2-\phi_1)}{(1-\phi_0-\phi_1)}\right)^2 \cdot \left[(n-1)\cdot\frac{1-F_{n-2}(y_0-2|\phi_0+\phi_1)}{1-F_n(y_0|\phi_0+\phi_1)}\right.\\
&\left. - n\cdot\left(\frac{1-F_{n-1}(y_0-1|\phi_0+\phi_1)}{1-F_n(y_0|\phi_0+\phi_1)}\right)^2\right] + \frac{n\phi_2}{(1-\phi_0-\phi_1)}\cdot\left(1-\frac{\phi_2}{(1-\phi_0-\phi_1)}\right)\\
&+ \frac{(\phi_0 - (\phi_0+\phi_1)(\phi_0+\phi_1))(1-\phi_0-\phi_1) + 2\phi_1\phi_2^2)}{(1-\phi_0-\phi_1)^2}\cdot\left(n\frac{1-F_{n-1}(y_0-1|\phi_0+\phi_1)}{1-F_n(y_0|\phi_0+\phi_1)}\right).
\end{aligned} \tag{4.18}$$

Likewise, the expression of the variance for the left cut-off point can be derived as

$$\begin{aligned}
&\text{var}(Y_2 - Y_2|Y_1 \le y_0)\\
&= n\left(\frac{\phi_1 + (\phi_0+\phi_1)(\phi_2-\phi_1)}{(1-\phi_0-\phi_1)}\right)^2 \cdot \left[(n-1)\cdot\frac{F_{n-2}(y_0-2|\phi_0+\phi_1)}{F_n(y_0|\phi_0+\phi_1)}\right.\\
&\left. - n\cdot\left(\frac{F_{n-1}(y_0-1|\phi_0+\phi_1)}{F_n(y_0|\phi_0+\phi_1)}\right)^2\right] + \frac{n\phi_2}{(1-\phi_0-\phi_1)}\cdot\left(1-\frac{\phi_2}{(1-\phi_0-\phi_1)}\right)\\
&+ \frac{(\phi_0 - (\phi_0+\phi_1)(\phi_0+\phi_2))(1-\phi_0-\phi_1) + 2\phi_1\phi_2^2)}{(1-\phi_0-\phi_1)^2}\cdot\left(n\frac{F_{n-1}(y_0-1|\phi_0+\phi_1)}{F_n(y_0|\phi_0+\phi_1)}\right).
\end{aligned} \tag{4.19}$$

Expression of the variances for RTM can be found by substituting $\phi_2 = \phi_1$ in equations (4.18) and (4.19).

4.3 RTM as a Function of Correlation Coefficient ρ

The magnitude of RTM is related to the correlation coefficient ρ of pre and post observations. The pattern of this relationship depends on the probability distribution under consideration. For the normal distribution, it varies linearly with ρ, whereas for the Poisson distribution the pattern is non-linear and is also affected by the direction of y_0 relative to the mean.

As ρ is a function of different parameters (ϕ_0, ϕ_1, ϕ_2), for illustrative purposes it is depicted in the left panel of Figure 4.1 for $\phi_1 = \phi_2$ and $\phi_0 = 0.1$. It is evident from the graph that ρ decreases non-linearly as ϕ_1 increases. To demonstrate RTM as a function of ρ, some specific values of y_0 both in the right and left tail are considered and the result is depicted in the right panel of Figure 4.1. For a left cut-off point, RTM decreases steeply as ρ takes values from -0.6 to 0.6, while for a right cut-off point it increases reaching maximum when ρ is around zero and then starts decreasing.

For normal and Poisson distributions, the RTM effect never exceeds the difference of y_0 and $E(Y_1)$, but for the bivariate binomial distribution when $\rho < 0$, RTM could have a relatively greater range thereby exaggerating the intervention effect comparatively more. Consequently, the conclusion would be more seriously in error if RTM is not accounted for when the pre and post observations follow the bivariate binomial distribution.

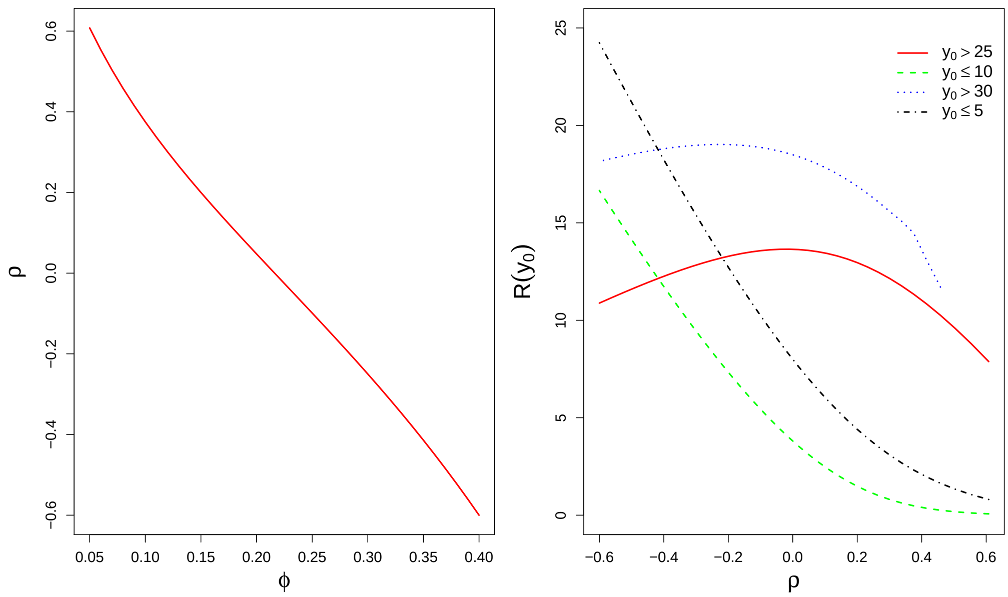

Figure 4.1. Left panel: Correlation as a function of $\phi = \phi_1 = \phi_2$ for fixed $n = 40$ and $\phi_0 = 0.1$, Right panel: RTM as a function of correlation and different cut-off points.

4.4 RTM for Normal and Poisson Approximations

As a rule of thumb, the normal approximation to the binomial distribution is deemed appropriate if both $n\pi$ and $n(1-\pi)$ are greater than ten (Moore et al., 2017). For demonstrative purposes, specific values of the parameters are considered here to satisfy this condition of approximation. Let $\phi_0 = 0.10$, $\phi_1 = \phi_2 = 0.05$ and $n = 80$. Assuming the normal is a good approximation to the binomial, the mean and variance of the normal distribution are $\mu = n(\phi_0 + \phi_1) = 12$ and $\sigma^2 = n(\phi_0 + \phi_1)(1 - \phi_0 - \phi_1) = 10.2$ respectively and the correlation coefficient is $\rho = n(\phi_0 - (\phi_0 + \phi_1)(\phi_0 + \phi_2))/\sigma^2 = 0.608$.

Similarly, the Poisson approximation to the binomial distribution is considered appropriate when the sample size n is large and the probability of success p is very small. The parameters of the bivariate Poisson distribution for the same values are $\theta_0 = n\phi_0 = 8$, $\theta_1 = n\phi_1 = 4$ and $\theta_2 = n\phi_2 = 4$.

The percentage relative change (PRC) can be used for quantitative comparison of two quantities/estimators by taking into account the size of things being compared and removing the units of measurement. One of the quantities being compared is the standard/reference/true value. With $R_i(y_0, \boldsymbol{\phi})$ under the bivariate binomial distribution as the reference value and $R_i(x_0)$ and $R_i(y_0, \boldsymbol{\theta})$ as RTM under the normal and Poisson approximations to the bivariate binomial distribution respectively, the PRC was calculated using the formula $PRC_i = (R_i(y_0, \boldsymbol{\phi}) - R_i(x_0))/R_i(y_0, \boldsymbol{\phi})$ for $i = r, \ell$ (Tornqvist et al., 1985). When PRC is around zero, the approximation works well in estimating RTM, whereas positive/negative values of PRC can be interpreted as underestimation/overestimation of RTM.

It is worth mentioning here that the formula of RTM for a left cut-off point remains the same for the normal approximation due to symmetry, i.e., $\Phi(-z_0) = 1 - \Phi(z_0)$.

The resulting graph of PRC_i for $i = r, \ell$, is given in Figure 4.2. Importantly, the PRC is greater than zero for all cut-off points, so the normal and Poisson approximations to the bivariate binomial distribution consistently underestimate RTM. The severity of underestimation stabilizes at 69 and 14 as the right cut-off point increases for the normal and Poisson approximations, respectively. On the other hand, as the left cut-off point decreases, PRC_ℓ first decreases and then increases touching 100 at y_0 = 36 for the normal approximation, and it stays at 15 for the Poisson approximation. Though the Poisson approximation works better than the normal approximation under suitable conditions, neither of them would be considered an appropriate alternative to the bivariate binomial distribution for quantifying RTM.

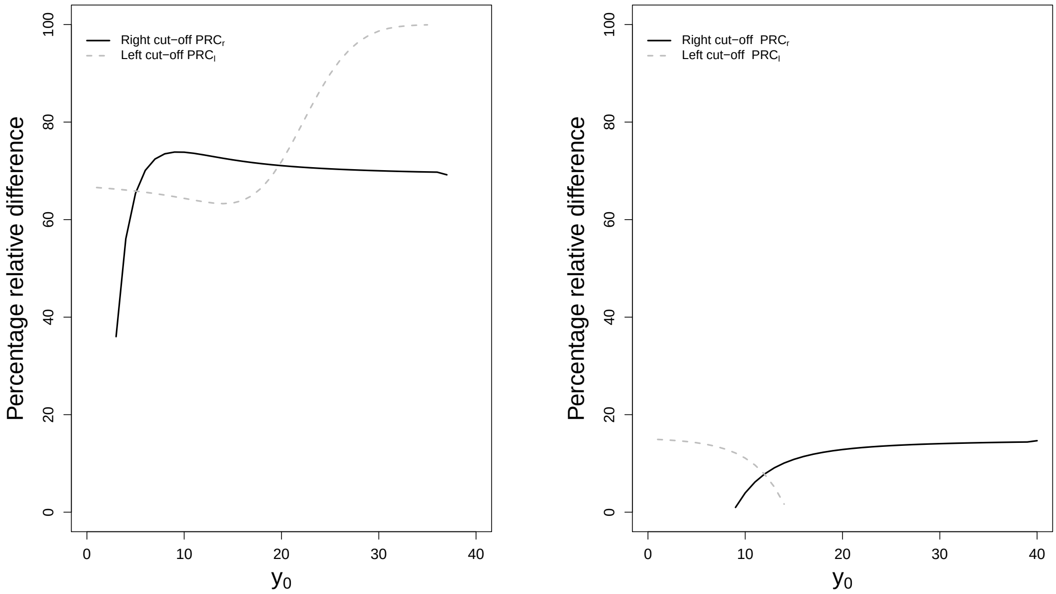

Figure 4.2. Left panel: Graph of percentage relative difference of RTM under normal approximation to bivariate binomial distribution and bivariate binomial distribution for different cut-off points, Right panel: Graph of percentage relative difference of RTM under Poisson approximation to bivariate binomial distribution and bivariate binomial distribution.

Correlation for the bivariate normal or Poisson of pre and post observations is always positive, so a comparative study cannot be performed when the bivariate binomial distribution has negative correlation.

4.5 Maximum Likelihood Estimation of the Total, RTM and Intervention Effects

A sample from the bivariate binomial distribution can contain information in two different ways. Firstly, the sample may contain thorough information on the number of successes on both the pre and post occasions on same subjects (α) and the total number of successes separately on the pre-post occasions, i.e., (Y_1, Y_2). Secondly, the sample may contain information only on (Y_1, Y_2). Estimation for both cases are considered separately.

4.5.1 MLE of ϕ when α is known

When information on the number of successes on both pre and post occasions, α_i, is available for each sample taken from the bivariate binomial distribution without a cut-off point, then ϕ can be estimated from the relevant sample proportions (Hamdan and Nasro, 1986). Let $(\alpha_1, y_{11}, y_{21}), (\alpha_2, y_{12}, y_{22}), ..., (\alpha_k, y_{1k}, y_{2k})$ be a random sample of size k from the bivariate binomial distribution. The MLE of $\boldsymbol{\phi} = (\phi_0, \phi_1, \phi_2, \phi_3)$ is given by

$$\begin{aligned}
\hat{\phi}_0 &= \frac{\sum_{i=1}^{k} \alpha_i}{k}, \\
\hat{\phi}_1 &= \frac{\sum_{i=1}^{k}(y_{1i} - \alpha_i)}{k} \\
\hat{\phi}_2 &= \frac{\sum_{i=1}^{k}(y_{2i} - \alpha_i)}{k} \\
\hat{\phi}_3 &= 1 - \hat{\phi}_0 - \hat{\phi}_1 - \hat{\phi}_2.
\end{aligned}$$

The total, RTM and intervention effects can be estimated by substituting estimates of (ϕ_0, ϕ_1, ϕ_2) into their relevant formulae.

4.5.2 MLE of ϕ when α is unknown

Let $(y_{11}, y_{21}), (y_{12}, y_{22}), ..., (y_{1k}, y_{2k})$ be a random sample of size k from a truncated bivariate binomial distribution. The likelihood function is given by

$$L(\boldsymbol{\phi_0}, \boldsymbol{y}) = \prod_{i=1}^{k} f_T(y_{1i}, y_{2i}, n),$$

where

$$\boldsymbol{y} = \begin{pmatrix} y_{11} & y_{21} \\ y_{12} & y_{22} \\ \dots & \dots \\ y_{1n} & y_{2k} \end{pmatrix}.$$

and the log of the likelihood is

$$\ell(\boldsymbol{\phi_0}, \boldsymbol{y}) = \sum_{i=1}^{k} log(f_T(y_{1i}, y_{2i}, n)).$$

Differentiating $\ell(\boldsymbol{\phi_0}, \boldsymbol{y})$ with respect to ϕ_0, ϕ_1 and ϕ_2, and then setting these equations to zero, we get

$$\frac{n}{k}\sum_{i=1}^{k}\frac{f_T(y_{1i}-1, y_{2i}-1, n-1)}{f_T(y_{1i}, y_{2i}, n)} = \frac{n}{k}\sum_{i=1}^{k}\frac{f_T(y_{1i}, y_{2i}, n-1)}{f_T(y_{1i}, y_{2i}, n)} + \frac{d\log(P(Y_1 > y_0))}{d\phi_0}, \quad (4.20)$$

$$\frac{n\phi_0}{k\phi_1}\sum_{i=1}^{k}\frac{f_T(y_{1i}-1, y_{2i}-1, n-1)}{f_T(y_{1i}, y_{2i}, n)} + \frac{n}{k}\sum_{i=1}^{k}\frac{f_T(y_{1i}, y_{2i}, n-1)}{f_T(y_{1i}, y_{2i}, n)} = \frac{\overline{y}_1}{\phi_1} - \frac{d\log(P(Y_1 > y_0))}{d\phi_1}, \quad (4.21)$$

and

$$\frac{n\phi_0}{k\phi_2}\sum_{i=1}^{k}\frac{f_T(y_{1i}-1, y_{2i}-1, n-1)}{f_T(y_{1i}, y_{2i}, n)} + \frac{n}{k}\sum_{i=1}^{k}\frac{f_T(y_{1i}, y_{2i}, n-1)}{f_T(y_{1i}, y_{2i}, n)} = \frac{\overline{y}_2}{\phi_2}, \quad (4.22)$$

where $\overline{y}_j = \sum_i^k y_{ji}/k$ are the sample means for $j = 1, 2$.

The derivative $d\log(P(Y_1 > y_0))/d\phi_i$ for $i = 0, 1$, can be expressed in terms of the CDF as

$$\frac{d\log(P(Y_1 > y_0))}{d\phi_i} = n \cdot \frac{P_{n-1}(y_0)}{1 - F_n(y_0|\phi_0 + \phi_1)}.$$

A recursive relation for the bivariate binomial distribution can be derived as

$$nf_T(y_{1i}, y_{2i}, n-1) = \frac{n - y_{1i} - y_{2i}}{1 - \phi_0 - \phi_1 - \phi_2}\cdot f_T(y_{1i}, y_{2i}, n) + \frac{n\phi_0}{1 - \phi_0 - \phi_1 - \phi_2}\cdot f_T(y_{1i}-1, y_{2i}-1, n-1). \quad (4.23)$$

Using the identity in equation (4.23), equation (4.22) simplifies to

$$\frac{n}{k}\sum_{i=1}^{k}\frac{f_T(y_{1i}-1, y_{2i}-1, n-1)}{f_T(y_{1i}, y_{2i}, n)} = \frac{\overline{y}_1\phi_2 + \overline{y}_2(1 - \phi_0 - \phi_1) - n\phi_2}{\phi_0(1 - \phi_0 - \phi_1)}. \quad (4.24)$$

Dividing equation (4.23) by $f_T(y_{1i}, y_{2i}, n)$, summing over the sample and using equation (4.24), we get the identity

$$\frac{n}{k}\sum_{i=1}^{k}\frac{f_T(y_{1i}, y_{2i}, n-1)}{f_T(y_{1i}, y_{2i}, n)} = \frac{n - \overline{y}_1}{1 - \phi_0 - \phi_1}. \quad (4.25)$$

Substituting the right hand side of equation (4.20) into equation (4.21), using equation (4.25) and solving for $\overline{y}_1$, we get

$$\overline{y}_1 = n(\phi_0 + \phi_1) \cdot \frac{1 - F_{n-1}(y_0 - 1|\phi_0 + \phi_1)}{1 - F_n(y_0|\phi_0 + \phi_1)}. \quad (4.26)$$

Combining equations (4.20) and (4.22), and using the identities (4.24-4.26), we get

$$\overline{y}_2 = n \cdot \frac{\phi_0 - (\phi_0 + \phi_1)(\phi_0 + \phi_2)}{1 - \phi_0 - \phi_1} \cdot \frac{1 - F_{n-1}(y_0 - 1|\phi_0 + \phi_1)}{1 - F_n(y_0|\phi_0 + \phi_1)} + \frac{n\phi_2}{1 - \phi_0 - \phi_1}. \tag{4.27}$$

Subtracting equation (4.27) from equation (4.26) and using the identity $F_n(y_0|\phi_0 + \phi_1) = F_{n-1}(y_0 - 1|\phi_0 + \phi_1) + (1 - \phi_0 - \phi_1)P_{n-1}(Y_1 = y_0)$, we get

$$n\phi_1 \cdot \frac{1 - F_{n-1}(y_0 - 1|\phi_0 + \phi_1)}{1 - F_n(y_0|\phi_0 + \phi_1)} - n\phi_2 + n\phi_2 \cdot \frac{(\phi_0 + \phi_1)P_{n-1}(Y_1 = y_0)}{1 - F_n(y_0|\phi_0 + \phi_1)} = \overline{y}_1 - \overline{y}_2. \tag{4.28}$$

By equation (4.5), the left hand side of equation (4.28) is $T_r(y_0, \boldsymbol{y})$, and hence the difference of the sample means is its estimate, given by

$$\begin{aligned}\widehat{T}_r(y_0, \boldsymbol{y}) &= n \cdot \frac{\hat{\phi}_1(1 - F_{n-1}(y_0 - 1|\hat{\phi}_0 + \hat{\phi}_1)) - \hat{\phi}_2\Big[1 - F_n(y_0|\hat{\phi}_0 + \hat{\phi}_1) - (\hat{\phi}_0 + \hat{\phi}_1)\widehat{P}_{n-1}(Y_1 = y_0)\Big]}{1 - F_n(y_0|\hat{\phi}_0 + \hat{\phi}_1)} \\ &= \overline{y}_1 - \overline{y}_2. \end{aligned} \tag{4.29}$$

Similarly, for a left cut point, the MLE of $T_\ell(y_0, \boldsymbol{\phi})$ can be obtained as

$$\begin{aligned}\widehat{T}_\ell(y_0, \boldsymbol{y}) &= n \cdot \frac{\hat{\phi}_2\Big[F_n(y_0|\hat{\phi}_0 + \hat{\phi}_1) + (\hat{\phi}_0 + \hat{\phi}_1)\widehat{P}_{n-1}(Y_1 = y_0)\Big] - \hat{\phi}_1 F_{n-1}(y_0 - 1|\hat{\phi}_0 + \hat{\phi}_1)}{F_n(y_0|\hat{\phi}_0 + \hat{\phi}_1)} \\ &= \overline{y}_2 - \overline{y}_1. \end{aligned} \tag{4.30}$$

For the estimation of the RTM and intervention effects, point estimates of ϕ_0, ϕ_1 and ϕ_2 are required, which consists of a two step procedure. In the first step, the MLE of $(\phi_0 + \phi_1)$ is obtained by solving equation (4.26) iteratively. The MLE of $(\phi_0 + \phi_1)$ can also be obtained from the marginal distribution of Y_1, the truncated univariate binomial distribution. In the second step, estimates of ϕ_i are obtained, for $i = 0, 1, 2$.

Once $(\phi_0 + \phi_1)$ is estimated, equations (4.20-4.22) can be re-organized after some algebraic manipulations to estimate ϕ_i, for $i = 0, 1, 2$. An equivalent form of equation (4.24) can be obtained by substituting equations (4.25) and (4.26) in equation (4.20) and simplifying

$$\frac{n}{k}\sum_{i=1}^{k} \frac{f_T(y_{1i} - 1, y_{2i} - 1, n - 1)}{f_T(y_{1i}, y_{2i}, n)} = n \cdot \frac{1 - F_{n-1}(y_0 - 1|\phi_0 + \phi_1)}{1 - F_n(y_0|\phi_0 + \phi_1)}. \tag{4.31}$$

For brevity, let the expression $d\log(P(Y_1 > y_0))/d\phi_i$ be $A_0 = n \cdot \frac{P_{n-1}(y_0)}{1-F_n(y_0|\phi_0+\phi_1)}$ for $i = 0, 1$. Similarly, let equations (4.25) and (4.31) be denoted by $A_1 = \frac{n}{k}\sum_{i=1}^{k} \frac{f_T(y_{1i}, y_{2i}, n-1)}{f_T(y_{1i}, y_{2i}, n)}$ and $A_2 = \frac{n}{k}\sum_{i=1}^{k} \frac{f_T(y_{1i}-1, y_{2i}-1, n-1)}{f_T(y_{1i}, y_{2i}, n)}$, respectively.

The sample estimates of A_0, A_1, and A_2 can be obtained by substituting $(\hat{\phi}_0 + \hat{\phi}_1)$ into their respective equations as

$$\begin{aligned}
\hat{A}_0 &= n \cdot \frac{\widehat{P}_{n-1}(y_0)}{1 - F_n(y_0|\hat{\phi}_0 + \hat{\phi}_1)}, \\
\hat{A}_1 &= \frac{n - \overline{y}_1}{1 - (\hat{\phi}_0 + \hat{\phi}_1)} \\
\hat{A}_2 &= n \cdot \frac{1 - F_{n-1}(y_0 - 1|\hat{\phi}_0 + \hat{\phi}_1)}{1 - F_n(y_0|\hat{\phi}_0 + \hat{\phi}_1)}.
\end{aligned}$$

Using the identity $\hat{A}_2 = \hat{A}_0 + \hat{A}_1$ and $\hat{A}_i$ for $i = 0, 1, 2$ in equation (4.21) and solving for ϕ_1, we get

$$\hat{\phi}_1 = \frac{\overline{y}_1}{\hat{A}_0 + \hat{A}_1} - \hat{\phi}_0. \tag{4.32}$$

Similarly, a solution of equation (4.22) for $\hat{\phi}_2$,

$$\hat{\phi}_2 = \frac{\overline{y}_2 - \hat{\phi}_0 \hat{A}_2}{\hat{A}_1}. \tag{4.33}$$

Rewriting equation (4.31) in terms of $\hat{A}_2$, we have

$$\frac{n}{k}\sum_{i=1}^{k} \frac{f_T(y_{1i} - 1, y_{2i} - 1, n - 1)}{f_T(y_{1i}, y_{2i}, n)} = \hat{A}_2. \tag{4.34}$$

Equation (4.34) is a polynomial in ϕ_0 and has to be solved numerically. Estimates of ϕ_1 and ϕ_2 can be obtained by substituting $\hat{\phi}_0$ in equations (4.32) and (4.33) respectively.

Substituting $(\hat{\phi}_0, \hat{\phi}_1, \hat{\phi}_2)$ in equations (4.7) and (4.9), we get the MLE of the RTM and intervention effects respectively as

$$\hat{R}_r(y_0, \boldsymbol{y}) = n\hat{\phi}_1 \cdot \frac{\widehat{P}_{n-1}(Y_1 = y_0)}{1 - F_n(y_0|\hat{\phi}_0 + \hat{\phi}_1)}, \tag{4.35}$$

and

$$\hat{\delta}_r(\boldsymbol{y}) = n(\hat{\phi}_1 - \hat{\phi}_2). \tag{4.36}$$

The variances of $\hat{R}_r(y_0, \boldsymbol{y})$ can be obtained by substituting $\phi_2 = \phi_1$ in equation (4.18) and dividing by the sample size k

$$\text{var}(\hat{R}_r(y_0)) = \text{var}(Y_1 - Y_2|Y_1 > y_0, \phi_1 = \phi_2)/k. \tag{4.37}$$

Similarly, for the left cut-off point we have

$$\text{var}(\hat{R}_\ell(y_0)) = \text{var}(Y_2 - Y_1 | Y_1 \le y_0, \phi_1 = \phi_2)/k. \tag{4.38}$$

4.5.3 Unbiasedness of $\widehat{T}_i(y_0, \boldsymbol{y})$ and $\widehat{R}_i(y_0, \boldsymbol{y})$

Applying expectation on both sides of equation (4.29) and using equations (4.2) and (4.4), we get

$$E(\widehat{T}_r(y_0, \boldsymbol{y})) = T_r(y_0, \boldsymbol{\phi}). \tag{4.39}$$

For the null effect case, i.e., $\phi_1 = \phi_2$, $\widehat{T}_r(y_0, \boldsymbol{y})$ and $\widehat{R}_r(y_0, \boldsymbol{\phi})$ are equivalent and we can write

$$E(\widehat{T}_r(y_0, \boldsymbol{y} | \phi_1 = \phi_2)) = E(\widehat{R}_r(y_0, \boldsymbol{y})) = R_r(y_0, \boldsymbol{\phi}). \tag{4.40}$$

For the non-null case, i.e., $\phi_1 \ne \phi_2$, the component parts of $T_r(y_0, \boldsymbol{\phi})$ are $R_r(y_0, \boldsymbol{\phi})$ and $\delta_r(\boldsymbol{\phi})\Big[1 - B_r(y_0, \phi_0 + \phi_1)\Big]$. Writing $\widehat{T}_r(y_0, \boldsymbol{y})$ into its constituent parts in equation (4.39), we get

$$E\Big[\widehat{R}_r(y_0, \boldsymbol{y}) + \hat{\delta}_r(\boldsymbol{y})\Big[1 - B_r(y_0, \hat{\phi}_0 + \hat{\phi}_1)\Big]\Big] = R_r(y_0, \boldsymbol{\phi}) + \delta_r(\boldsymbol{y})\Big[1 - B_r(y_0, \phi_0 + \phi_1)\Big]. \tag{4.41}$$

Thus, from equations (4.40) and (4.41), the unbiasedness property of $\widehat{R}_r(y_0, \boldsymbol{y})$ is established.

The unbiasedness of $\widehat{T}_\ell(y_0, \boldsymbol{y})$ and $\widehat{R}_\ell(y_0, \boldsymbol{y})$, for a left cut-off point can be established in a similar way.

4.5.4 Asymptotic distribution of $\widehat{T}_i(y_0, \boldsymbol{y})$ and $\widehat{R}_i(y_0, \boldsymbol{y})$

$\widehat{T}_i(y_0, \boldsymbol{y})$ for $i = r, \ell$ are the differences of the sample means of Y_1 and Y_2 generated from the truncated bivariate binomial distribution. So for a right cut-off point, by the Central Limit Theorem, $\widehat{T}_r(y_0, \boldsymbol{y})$ are asymptotically normally distributed as

$$\sqrt{k}\left(\widehat{T}_r(y_0, \boldsymbol{y}) - T_r(y_0, \boldsymbol{\phi})\right) \overset{d}{\sim} N\left(0, \text{var}(Y_1 - Y_2 | Y_1 > y_0)\right).$$

As additive components of a normal random variable are necessarily normally distributed (Cramér, 1936), so the components $\hat{R}_r(y_0, \boldsymbol{y})$ and $\hat{\delta}_r(\boldsymbol{y})\Big[1 \pm B_r(y_0, \hat{\phi}_0 + \hat{\phi}_1)\Big]$ of $\widehat{T}_r(y_0, \boldsymbol{y})$ are also asymptotically normally distributed. In particular, the RTM estimator is

$$\sqrt{k}\left(\widehat{R}_r(y_0, \boldsymbol{y}) - R_r(y_0, \boldsymbol{\phi})\right) \overset{d}{\sim} N\left(0, \text{var}(Y_1 - Y_2 | Y_1 > y_0, \phi_2 = \phi_1)\right).$$

The results also hold for the left cut-off point.

4.6 Simulation Study of RTM

To estimate the RTM effect and study its properties empirically, a simulation study was performed. The procedure of Hamdan and Nasro (1986) to generate bivariate sample observations from the bivariate binomial distribution was adopted. If a cut-off point is selected far in the tail on either side, the associated probability $P(Y_1 > y_0)$ is very small, thus the number of observations beyond/below a cut-off is very small. A sample of size $n = 1000$ from the binomial would suffice to get an expected sample of size 50 beyond a cut-off point greater than 12 for $n = 20$ and $\phi_0 + \phi_1 = 0.45$. Initially, samples of sufficiently large sizes were generated to get adequate number of realizations beyond/below a cut-off point y_0. First, $k = 10, 20, 50, 100$ realizations above/below y_0 were considered as random samples from the truncated bivariate binomial distribution with parameters $n = 20$, $y_0 = 12$, $\phi_0 = 0.20$ and $\phi_1 = \phi_2 = 0.25$ for demonstrative purposes. This procedure was repeated 1000 times.

4.6.1 Empirical distribution of $\hat{R}_r(y_0, \boldsymbol{y})$

The resulting normal quantile-quantile plots of the sampling distribution of $\hat{R}_r(y_0, \boldsymbol{y})$ are given in Figure 4.3 which suggests approximate normality of the distribution of $\hat{R}_r(y_0, \boldsymbol{y})$ for $k = 10, 20, 50, 100$.

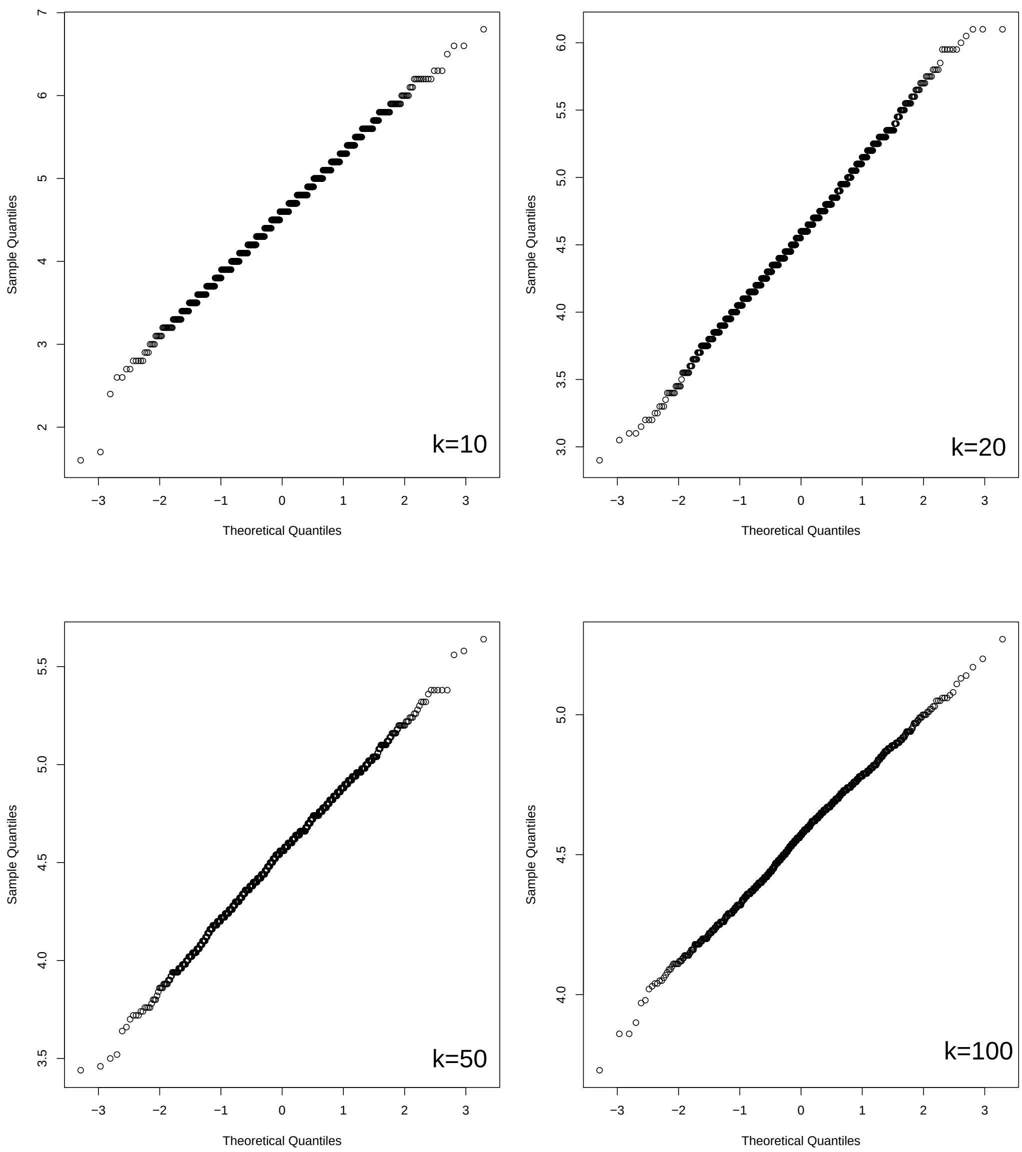

Figure 4.3. Normal qq-plot of the sampling distribution of RTM for $k = (10, 20, 50, 100)$ and $y_0 = 12$, $n = 20$, $\phi_0 = 0.20$, $\phi_1 = 0.25$, $\phi_2 = 0.25$ and $m = 1000$ simulations.

4.6.2 Empirical unbiasedness and consistency of $\hat{R}_r(y_0, \boldsymbol{y})$

The graph of RTM estimates for different sample sizes is given in Figure 4.4. The means of estimates shown by green line segments, coincide with the true RTM effect indicated by the dashed red line, confirming the theoretically derived result of unbiasedness of the RTM estimator. As the sample size k increases, the spread around the centre decreases, verifying consistency of the estimator shown theoretically in equations (4.38) and (4.39).

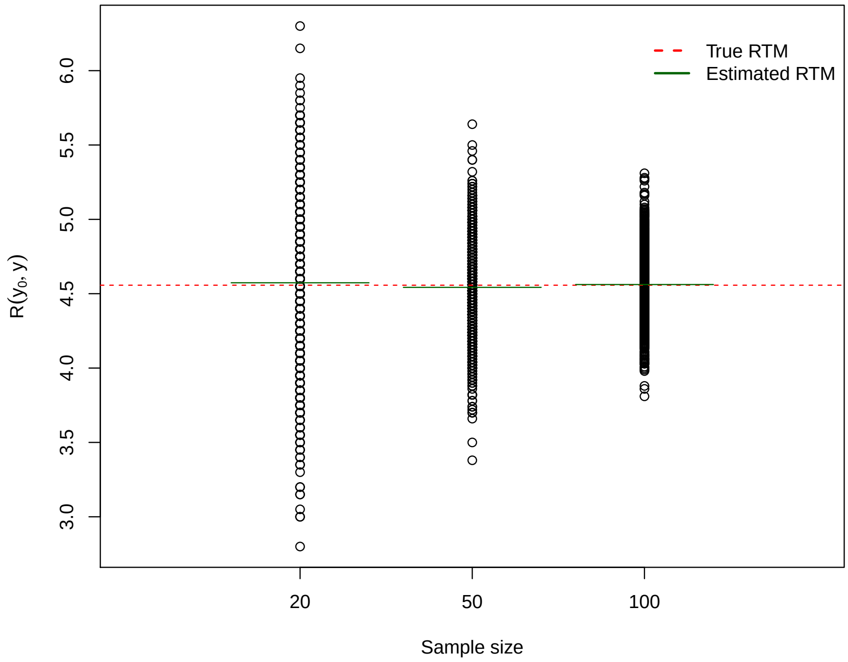

Figure 4.4. Estimates of RTM and its sampling distribution for different sample choices and parameters $y_0 = 12$, $n = 20$, $\phi_0 = 0.20$ and $\phi_1 = \phi_2 = 0.25$.

4.6.3 Confidence intervals and coverage probabilities

Coverage probabilities are often used for evaluating the performance of an estimator for parameters of a discrete distribution. The concepts of confidence and coverage probability are interrelated, and coverage probability of a $(1 - \alpha)100\%$ confidence interval is the probability it contains the true parameter. Let the respective upper and lower limits of a confidence interval for $R_i(y_0, \boldsymbol{\phi})$, be given by $\widehat{U}_i$ and $\widehat{L}_i$, estimated from a sample of size k, for $i = r, \ell$. The true coverage probability $C(\boldsymbol{\phi}, n, k)$ is

$$C(\boldsymbol{\phi}, n, k) = \sum_x I(\widehat{L}_i < R_k(y_0, \boldsymbol{\phi}) < \widehat{U}_i) P(x; \boldsymbol{\phi}) \tag{4.42}$$

where $I(\cdot)$ is the indicator function.

$C(\boldsymbol{\phi}, n, k)$ is a function of five parameters $(\phi_0, \phi_1, \phi_2, n, k)$ and hence cannot be displayed on a two dimensional plane. An alternative option is simulated coverage probability (SCP) which is the proportion of times an estimated confidence interval contains the true parameter from a series of m simulated datasets,

$$\widehat{C}(\boldsymbol{\phi}, n, k) = \frac{\sum_{i=1}^{m} I(\widehat{L}_i < R_i(y_0, \boldsymbol{\phi}) < \widehat{U}_i)}{m}.$$

SCP can be used for studying large sample properties such as asymptotic normality and consistency of the RTM estimator and also its behaviour for finite sample sizes. Assuming normality for the distribution of $\hat{R}_r(y_0, \boldsymbol{y})$, 95% confidence intervals were constructed for different sample sizes, using maximum likelihood estimates as

$$\hat{R}_k(y_0, \boldsymbol{y}) \pm 1.96\sqrt{\widehat{\text{var}}(\hat{R}_k(y_0, \boldsymbol{y}))} \quad \text{for} \quad k = r, \ell. \tag{4.43}$$

The resulting SCP for different sample sizes and cut-off points in both the right and left tails, is given in Figure 4.5. The pattern of SCP remains similar for different cut-off points and approaches the target value of 95% as the sample size increases. Whereas, for a sample size of 10, SCP misses the target value, although the SCP is greater than 90% for each scenario.

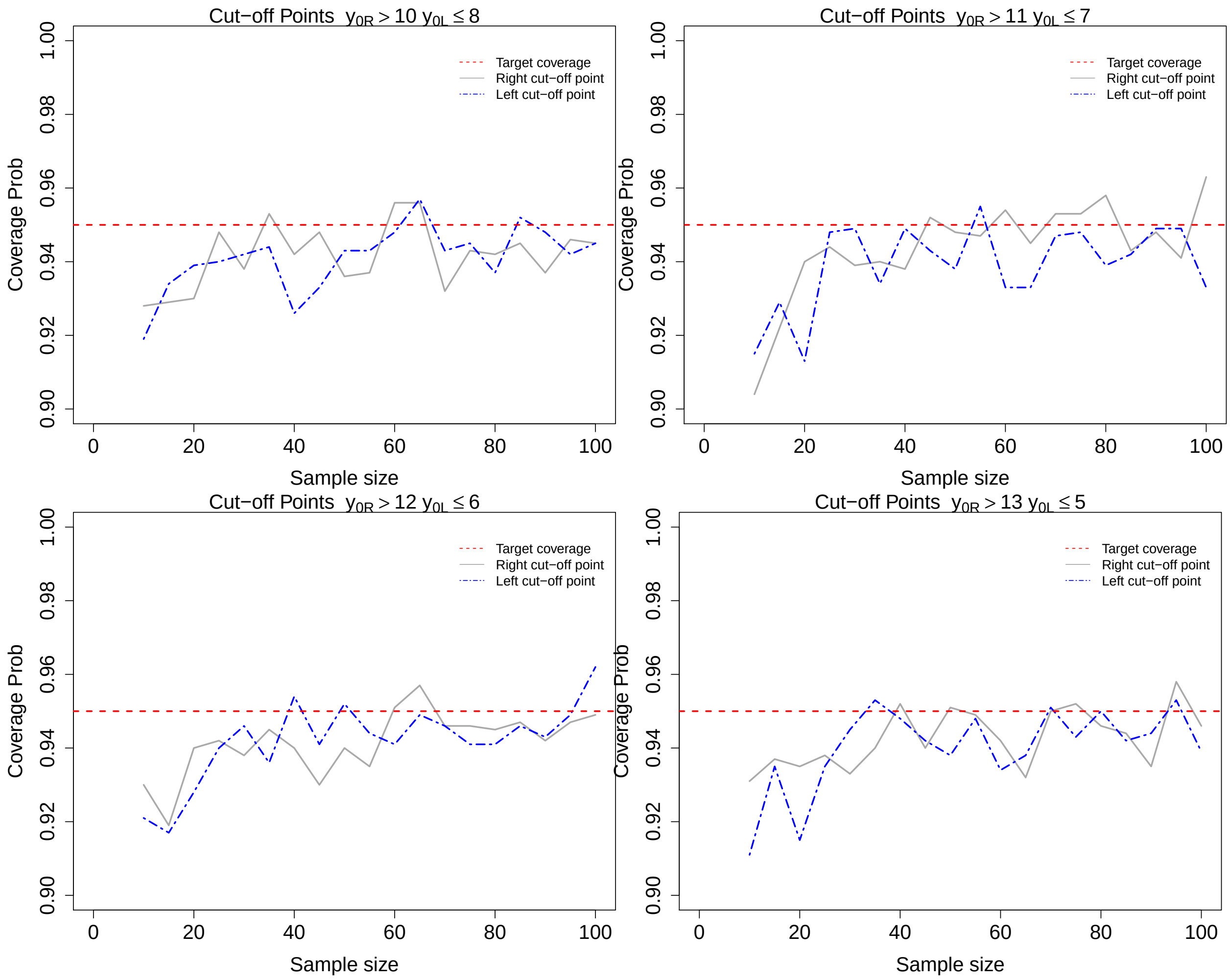

Figure 4.5. Simulated coverage probabilities for different sample sizes and cut-off points

4.7 Data Examples

4.7.1 Data Accessibility

The data* used in this manuscript were accessed from a published obesity study (Woolson and Clarke, 1984) and a text on statistical quality control (Montgomery, 2013). Note that the raw data have been re-organized for the current study as detailed below.

*https://content.sph.harvard.edu/fitzmaur/ala/obesity.txt

4.7.2 Data on Obesity in Different Age Cohorts

Data on weight and height measurements of five cohorts of children, initially aged 5-7, 7-9, 9-11, 11-13, and 13-15 years, were obtained biennially from 1977 to 1981 in Muscatine, Iowa (Lauer et al., 1997; Woolson and Clarke, 1984). One of the study aims was to assess whether risk of obesity increased with age. There was evidence that the log-odds of obesity increased from 6 to 12 years, levelled off from 12-14 years, and declined from 14 to 18 years. The structure of the longitudinal data collection is given in Figure 4.6.

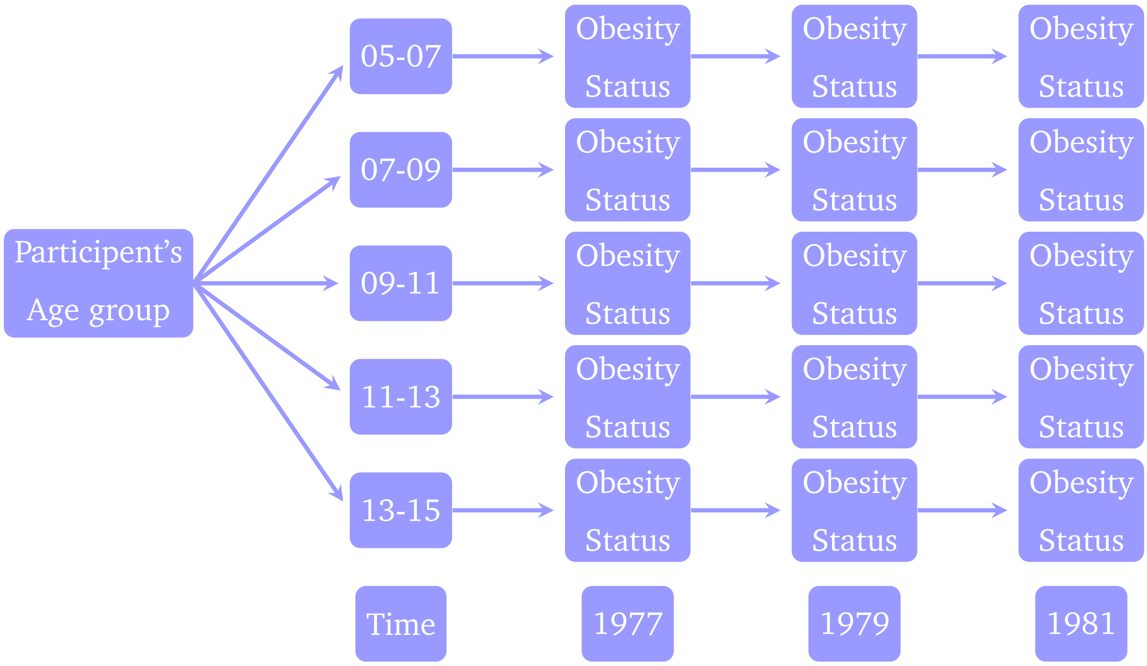

Figure 4.6. Longitudinal study flowchart

To demonstrate how RTM may affect the Muscatine study's conclusion, data were re-organized. Successive observations were considered as outcomes of a bivariate Bernoulli trial (Z_1, Z_2) for each individual and the resulting total number of individuals n and data pairs are given in Table 4.2 for each age group. Each individual can at most contribute two data pairs of information across all three occasions. Incomplete data for children were excluded. The sum of the data pairs is assumed to follow the bivariate binomial distribution. In this example, information on the number of obese individuals on both occasions, α, are available and subject selection is not based on cut-off points. So, the MLE of the parameters of the bivariate binomial distribution were obtained using the Hamdan and Nasro (1986) method for each age cohort. The estimates of the parameters of the bivariate binomial distribution for different age groups are given in Table 4.2.

Table 4.2. Estimates of the parameters of the bivariate binomial distribution

Age Cohort	n	Data pairs	$\hat{\phi}_0$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\rho}$
5-7	513	819	0.105	0.031	0.104	0.639
7-9	662	1137	0.145	0.057	0.095	0.515
9-11	673	1143	0.163	0.071	0.071	0.463
11-13	524	904	0.132	0.079	0.076	0.416
13-15	498	833	0.146	0.061	0.072	0.498

To study the contribution of RTM to the total difference of obesity risk on pre-post occasions, cut-off points for the number of obese individuals above the means of the respective cohorts were considered for demonstrative purposes. The estimated total, RTM, observed intervention and intervention effects (which is the age effect) for different cut-off points in the right tail, and different age cohorts are given in Figure 4.7.

For age cohorts (5-7, 7-9 and 13-15) where $\hat{\phi}_1 < \hat{\phi}_2$, as depicted in Figure 4.7 for a right cut-off point, the age effect is negative, indicating that obesity has increased with age. Importantly, the total effect decreases as the cut-off point increases, and the adverse age effect on obesity would be underestimated without accounting for RTM, in the age cohorts 5-7 and 7-9 years of age. Additionally, the difference between $\hat{\delta}(\boldsymbol{y})(1 - \hat{B}(y_0, \boldsymbol{y}))$ and $\hat{\delta}(\boldsymbol{y})$ is proportional to $(\hat{\phi}_1 - \hat{\phi}_2)$ and the cut-off points. It is at a maximum for groups 5-7 and 7-9 years of age and almost zero for the remaining age groups.

The curves of the total and RTM effects coincide in Figure 4.7 for age cohorts 9-11 and 11-13, indicating a possible null age effect, i.e., $\hat{\phi}_1 = \hat{\phi}_2$. The observed decrease up to 25 units on average in obesity is due to RTM and could erroneously be attributed to the age effect for reducing obesity in age cohorts 9-11 and 11-13 years.

Overall, RTM increases as the cut-off points on either side are selected farther in the tail of the distribution thereby depleting or inflating the total effect. Consequently an observed mean change, which is the sum of RTM and the age effect, may be mistaken for a real mean change in the prevalence of obesity due to ageing.

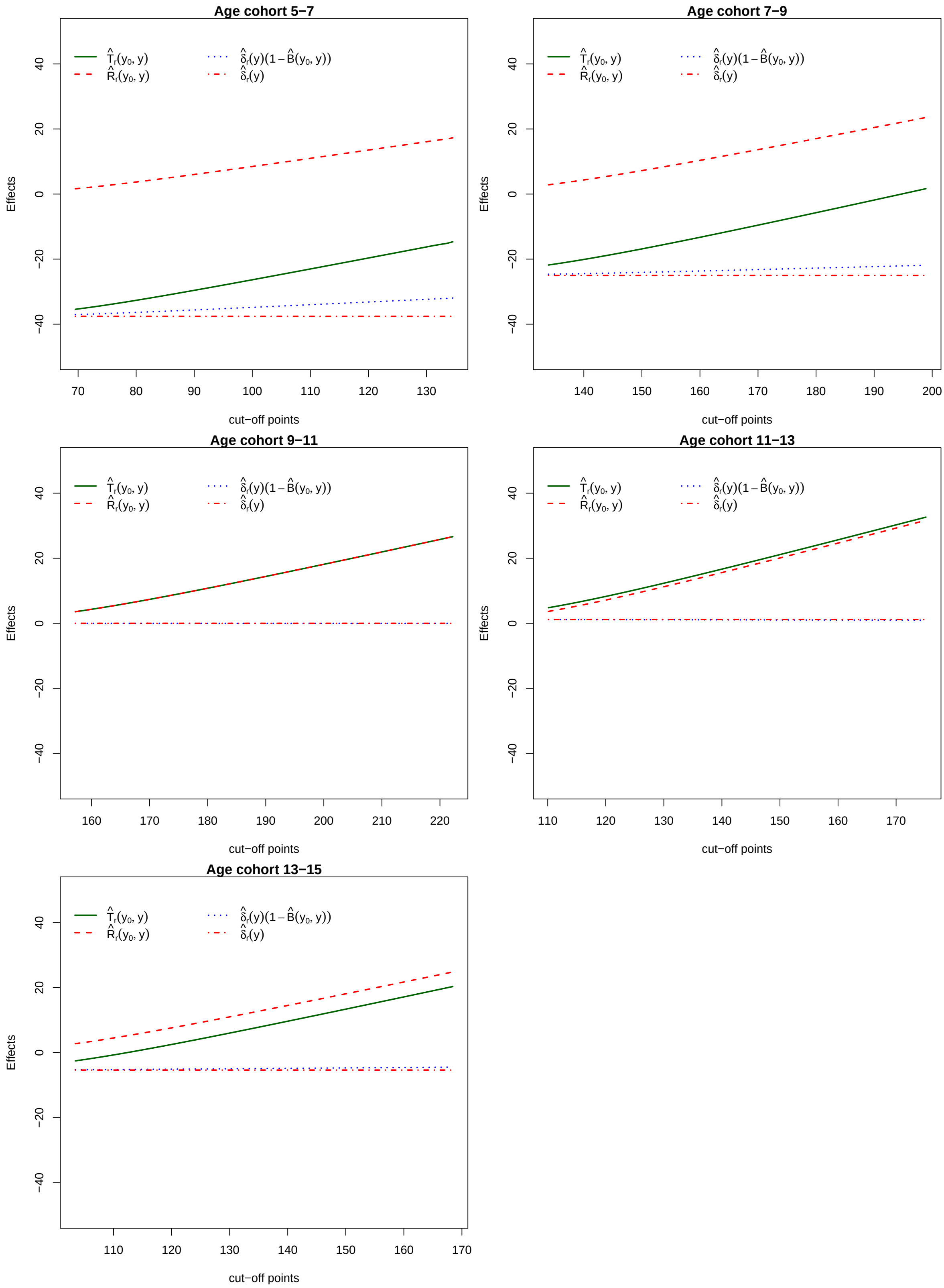

Figure 4.7. Graph of the RTM effects for obesity example for different age cohorts and right cut-off points.

4.7.3 Number of Nonconforming Cardboard Cans

Data on the number of nonconforming cardboard cans is considered here, which can be obtained from Tables 7.2 and 7.3 of Montgomery (2013). Sixty four samples each of size $n = 50$ were collected after half an hour over a three-shift period in the control state of the manufacturing process. The successive observations were treated as a bivariate sample of size $k = 63$ from the bivariate binomial distribution. The maximum likelihood method was used for estimating the parameters for a right cut-off point $y_0 = 3$ and estimates of the truncated bivariate binomial distribution are $\hat{\phi}_0 = 0.080$, $\hat{\phi}_1 = 0.029$ and $\hat{\phi}_2 = 0.021$.

In statistical process control, when the number of nonconforming units exceed a pre-specified limit (e.g., three sigma), the control chart signals an out of control situation. This is followed by an intervention (e.g., adjustment of machine, material checking, or adjustment of controllable variables) to bring the process back into a control state. For example, let the number of nonconforming units in a sample be equal to or greater than 7. This is an out of control situation as per three sigma limits and would trigger an intervention. The cumulative probability of having between 7 and 12 nonconforming units in a sample of size $n = 50$ are 0.173, 0.089, 0.041, 0.017, 0.006, and 0.002 respectively. Based on the estimated parameters, $\widehat{T}(y_0, \boldsymbol{y})$, $\widehat{R}(y_0, \boldsymbol{y})$, $\hat{\delta}_r(\boldsymbol{y})$ and $\hat{\delta}_r(\boldsymbol{y})(1 - \hat{B}(y_0, \boldsymbol{y}))$ for different right cut-off points (2-12) are given in Figure 4.8.

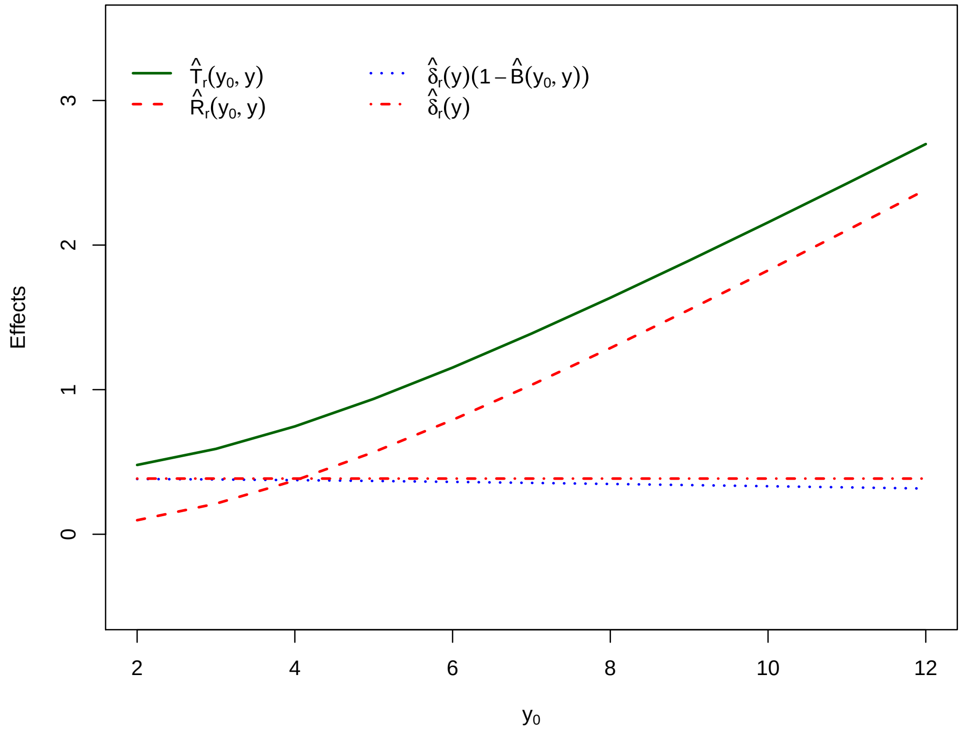

Figure 4.8. Graph of the RTM effects for cardboard can example for different right cut-off points y_0.

As the cut-off point increases, $\widehat{R}_r(y_0, \boldsymbol{y})$ increases and as a result $\widehat{T}(y_0, \boldsymbol{y})$ also increases, whereas $\hat{\delta}_r(\boldsymbol{y})(1 - \hat{B}(y_0, \boldsymbol{y}))$ stays virtually constant at $\hat{\delta}_r(\boldsymbol{y})$. Samples where the number of nonconforming cans y_0 were greater than 12, decreased on average by more than 3 in the next sample. This decrease in nonconforming units could be due to RTM and might have resulted in unnecessary machine adjustment, material checking or adjustment of controllable variables for reducing the nonconforming cardboard cans. This could have a potentially negative effect on production.

4.8 Discussion

In data analysis, RTM can potentially affect the conclusions of a study by exaggerating the intervention effect. The strategies of random allocation or multiple baseline measurements used for guarding against RTM are not always possible due to ethical/logistical constraints or associated costs. Therefore, quantifying and accounting for RTM in an analysis, is an important statistical issue. Expressions for

RTM are available in the literature when the underlying distributions are the bivariate normal and Poisson. However, expressions for RTM are missing when the pre and post variables follow the bivariate binomial distribution.

The correlation of pre and post variables under bivariate normality or bivariate Poisson are always positive, whereas correlation can be negative or positive when pre and post variables follow the bivariate binomial distribution. The RTM effect is more severe when ρ is negative. RTM is underestimated when the normal/Poisson approximations to the binomial are used, and therefore these approximations are not recommended. Apart from this, the intervention effect would be biased if it is obtained by subtracting RTM from the total effect.

Expressions for the MLE of RTM and intervention effects were derived assuming a bivariate binomial distribution. The properties of unbiasedness, consistency and asymptotic normality of the estimators were demonstrated theoretically. The asymptotic properties were verified through simulation by studying its empirical distribution, mean, the spread around its mean and simulated coverage probabilities.

The RTM effect for different cut-off points was evaluated for the number of obese individuals in different age groups, using the maximum likelihood method. Without accounting for RTM, observed differences may be mistaken for real differences in the prevalence of obesity. Likewise, the change in the number of nonconforming cardboard cans could be due to RTM and which may be incorrectly attributed to an intervention aimed at reducing the nonconforming units. Besides this, the intervention effect obtained by subtracting RTM from the total effect was biased in some cases.

Chapter 5

Regression to the mean for bivariate families of distributions

The earlier derivations for regression to the mean discussed in Chapter 2 were based on assumptions of normality, positive correlation, and a null intervention or treatment effect. In Chapters 3 and 4, formulae for RTM were derived to obtain an unbiased intervention effect by decomposing the total effect into RTM and intervention/treatment effects, while relaxing restrictive assumptions. This chapter derives expressions for the total effect for any bivariate distribution, while also providing a solution for decomposing the total effect into RTM and intervention or treatment effects. Maximum likelihood estimates are derived and the unbiasedness, consistency and normality of these estimators are established for exponential families, where possible. Data on the cholesterol levels in men aged 35 to 39 are used for decomposing the total change in cholesterol level on pre-post occasions into regression to the mean and intervention or treatment effects. The contents of this chapter are reproduced from a drafted paper with minor modification.

5.1 Introduction

As discussed in Chapter 2, early research derived formulae to account for RTM assuming the bivariate normal distribution for the pre-post variables in an intervention study (James, 1973; Gardner and Heady, 1973; Davis, 1976). Apart from the distributional assumption, the pre-post variables were restricted to be identically distributed and positively correlated, and the treatment effect was assumed to move the post measurements in the direction of the mean.

RTM formulae for non-normal populations were derived by Das and Mulder (1983), but this derivation was of limited use as it was not directly applicable to empirical distributions. Beath and Dobson (1991) derived estimates for RTM for non-normal data based on Edgeworth series and saddlepoint approximations. However, in these derivations the error term was allowed to be normally distributed with zero mean and constant variance and the pre-post variables were assumed to be stationary. John and Jawad (2010) aimed at making the Das and Mulder (1983) derivation data adaptive for estimation of RTM using kernel density estimation for the density and hazard rate functions.

James (1973) derived the method of moments estimator for RTM assuming that the percent of the population in the truncated portion is known. Senn and Brown (1985) improved the derivation by James (1973) and also generalized the maximum likelihood estimation of parameters to various sampling schemes associated with the bivariate normal distribution. Chen and Cox (1992) derived a maximum likelihood estimator of the intervention effect assuming that the pre-post parameters were identical and the treatment was designed to change the post measurements in the direction of the population mean. The authors did not study the statistical properties of their derived estimators.

Interventions may have effects in the direction opposite than intended. Ter Weel (2006), in a study of the Dutch soccer league, found no improvements in team performance after manager turnover. In a similar study, a negative relationship between employee turnover and performance was observed by Ton and Huckman (2008). Changzheng and Kai (2010) discussed different effects of employee turnover on firm performance including positive, negative and no effects. Thus, an intervention could change the composition of a population including its mean, variance and correlation parameters.

Khan and Olivier (2018, 2019) relaxed the identical distributional assumption for the pre-post variables, thereby allowing the treatment to change the post measurements towards the mean of the population or in the opposite direction. Moreover, an intervention or treatment applied to subjects screened on the bases of a cut-off point produces a compound effect called the total effect (Khan and Olivier, 2018). This could be the combined effect of the intervention and RTM. To accurately estimate an intervention effect, RTM should be accounted for. The authors achieved this objective by decomposing the total effect into the RTM and intervention effects, and obtaining the maximum likelihood estimates for the constituent parts.

The outcomes of a pre-post study design are rarely independent and follow a bivariate distribution which could be continuous or discrete depending on the problem under study. Balakrishnan and Lai (2009) discussed continuous bivariate distributions with applications in various research areas. Similarly, Norman et al. (1997) shed light on bivariate and multivariate discrete distributions along with applications. However, a general approach for accurately estimating the intervention effect by accounting for RTM is missing in literature.

Therefore, the aims of this chapter are to derive expressions and maximum likelihood estimators for the total, RTM and intervention effects while relaxing assumptions about the bivariate distribution, the direction of the treatment effect and the pre-post parameters. Additionally, exploring the statistical properties of the maximum likelihood estimators from a theoretical point of view, where possible, is another objective of this work.

For the remainder of this chapter, Section 5.2 generalizes the dependence structure to allow the pre-post variables to be negatively correlated where possible, and have non-stationary distributions. Section 5.3 derives an expression for the total effect which is decomposed into its constituent parts. Derivation of the RTM formulae for the exponential family of distributions is discussed in Section 5.4 and demonstrated with the help of some well known examples of bivariate distributions in Section 5.5. Maximum likelihood estimators of the total, RTM and intervention effects are derived in Section 5.6, while these effects are estimated using data on cholesterol levels in Section 5.7. Section 5.8 concludes the chapter with a discussion.

5.2 Successive random variables

Let X_1 and X_2 be characteristics of interest on the same subject before and after an intervention or treatment. Stigler (1997), in a review of the work by Sir Francis Galton, decomposed X_1 and X_2 into persistent traits X_0 and transient traits E_i as $X_1 = X_0 + E_1$ and $X_2 = X_0 + E_2$, for $i = 1, 2$. James (1973) used the terminology of true and random error components for the persistent and transient traits, respectively, and derived a formula for RTM. The variables X_0, E_1 and E_2 were assumed mutually independent and identically normally distributed. This formulation also restricts X_1 and X_2 to be identically distributed $N(\mu, \sigma^2)$ and positively correlated $\rho = \text{var}(X_0)/\text{var}(X_1) > 0$. After applying treatment to individuals with $X_1 > x_0$, James (1973) assumed a model for post measurements as

$$X_2 - \mu = \gamma\rho(X_1 - \mu) + E,$$

where $E \sim N(0, (1 - \rho^2)\sigma^2)$ is independent of X_1 and γ is the treatment parameter designed to move the post measurements towards the mean of the untruncated population. The treatment is deemed effective when $\gamma < 1$.

To add more flexibility, let the successive variables X_1 and X_2 be decomposed as $X_1 = X_0 + E_1$ and $X_2 = a + bX_0 + E_2$, where X_0 is the true component part as before, a and b are constants, and E_1 and E_2 are within subject or random errors. The variables E_1 and E_2 are mutually independent but may not be independent of X_0 or identically distributed (Khan and Olivier, 2019). For continuous probability distributions, the constants a and b allow the pre-post means to differ as a result of the intervention effect and be correlated either negatively or positively. Whereas, for discrete distributions, $a = 0$, $b = 1$ and the error components E_1 and E_2 allow the successive variables to have different means as a result of a possible intervention effect.

It is not always possible for a random variable to be decomposed as the sum of other random variables, e.g., the Pareto and Weibull distributed random variables (Nadarajah, 2008; Zaliapin et al., 2005). Consequently, the successive variables cannot be expressed as a linear sum of the true and random error components which is helpful in specifying their unique joint distribution. This potentially gives rise to non-uniqueness problems as there exist many bivariate distributions (Paduthol et al., 2014). However, formulae for RTM can be determined for these variables assuming any relevant bivariate distribution.

5.3 Derivations and decomposition the total effect

Let the respective distributions of X_1 and X_2 be $f(x_1;\boldsymbol{\theta}_1)$ and $f(x_2;\boldsymbol{\theta}_2)$ where $(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) \subseteq \boldsymbol{\theta}$ is the parameter vector, for $i = 1,2$. Let the means, variances and correlation of successive variables be $E(X_1) = \mu_{x_1}(\boldsymbol{\theta}_1)$, $E(X_2) = \mu_{x_2}(\boldsymbol{\theta}_2)$, $\text{var}(X_1) = \sigma^2_{x_1}(\boldsymbol{\theta}_1)$, $\text{var}(X_2) = \sigma^2_{x_2}(\boldsymbol{\theta}_2)$ and $\text{cor}(X_1, X_2) = \rho$. Let the joint distribution of X_1 and X_2 be $f(x_1, x_2;\boldsymbol{\theta})$ where $-\infty < X_1 < \infty, -\infty < X_2 < \infty$.

Assume that a treatment or intervention is applied to subjects selected from a population with parameter vector $\boldsymbol{\theta}_1$. The treatment effect $\delta(\boldsymbol{\theta})$ can be evaluated by finding the expected difference of successive variables X_1 and X_2 and variance as

$$\delta(\boldsymbol{\theta}) = E(X_1 - X_2) = \mu_{x_1}(\boldsymbol{\theta}_1) - \mu_{x_2}(\boldsymbol{\theta}_2) \tag{5.1}$$

$$\text{var}(X_1 - X_2) = \sigma^2_{x_1}(\boldsymbol{\theta}_1) + \sigma^2_{x_2}(\boldsymbol{\theta}_2) - 2\rho\sigma_{x_1}(\boldsymbol{\theta}_1)\sigma_{x_2}(\boldsymbol{\theta}_2). \tag{5.2}$$

Suppose subjects with measurements above or below a cut-off or truncation point, say x_0, are selected for an intervention or treatment. For demonstrative purposes, let the selection of subjects be based on a right cut-off point, then the joint distribution of successive measurements X_1 and X_2 is the truncated distribution

$$f_t(x_1, x_2;\theta) = \frac{f(x_1, x_2;\boldsymbol{\theta})}{1 - F(x_0;\boldsymbol{\theta}_1)} = f_t(x_1;\boldsymbol{\theta}_1) f(x_2 \mid x_1, \boldsymbol{\theta}) \qquad x_0 < X_1 < \infty, -\infty < X_2 < \infty, \tag{5.3}$$

where $f_t(x_1;\boldsymbol{\theta}_1) = f(x_1;\boldsymbol{\theta}_1)/\{1 - F(x_0;\boldsymbol{\theta}_1)\}$ and the subscript t stands for truncated.

For a truncated distribution, the expected difference of X_1 and X_2 is not equivalent to the treatment effect alone, and is instead the total effect, $T(x_0;\boldsymbol{\theta})$ as

$$T(x_0;\boldsymbol{\theta}) = E(X_1 - X_2 \mid X_1 > x_0) = \int_{x_0}^{\infty} \int_{-\infty}^{\infty} (x_1 - x_2) f_t(x_1, x_2;\boldsymbol{\theta}) dx_2 dx_1. \tag{5.4}$$

Evaluating the inner integral in equation (5.4), and using the fact $\int_{-\infty}^{\infty} f(x_2|x, \boldsymbol{\theta}) dx_2 =$

1, we get

$$T(x_0;\boldsymbol{\theta}) = \int_{x_0}^{\infty} x_1 f_t(x_1;\boldsymbol{\theta}_1)dx_1 - \int_{x_0}^{\infty} f_t(x_1;\boldsymbol{\theta}_1)\left(\int_{-\infty}^{\infty} x_2 f(x_2 \mid x_1,\boldsymbol{\theta})dx_2\right)dx_1$$
$$= E_{x_1}\left(X_1 - E(X_2 \mid X_1)\right), \qquad (5.5)$$

where $E(X_2|X_1) = \int_{-\infty}^{\infty} x_2 f(x_2 \mid x,\boldsymbol{\theta})dx_2$ is the conditional expectation of X_2 given X_1 and E_{x_1} denotes expectation with respect to X_1. The conditional expectation $E(X_2 \mid X_1)$ can be written as

$$E(X_2 \mid X_1) = E(X_2) + \beta_{x_2,x_1}\left(X_1 - E(X_1)\right), \qquad (5.6)$$

where $\beta_{x_2,x_1} = \text{cov}(X_1,X_2)/\text{var}(X_1)$ is the regression coefficient.

The total effect $T(x_0;\boldsymbol{\theta})$ can be decomposed into RTM and intervention effects. Substituting equation (5.6) into equation (5.5) and rearranging terms, we get

$$T(x_0;\boldsymbol{\theta}) = \{E(X_1 \mid X_1 > x_0) - E(X_1)\}(1-\beta_{x_2,x_1}) + E(X_1 - X_2)$$
$$= R(x_0;\boldsymbol{\theta}) + \delta(\boldsymbol{\theta}), \qquad (5.7)$$

where $R(x_0;\boldsymbol{\theta})$ is regression to the mean and $\delta(\boldsymbol{\theta})$ is the average intervention effect. A null treatment effect, $\delta(\boldsymbol{\theta}) = 0$, implies that X_1 and X_2 are identically distributed, that is $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$ or $E(X_1) = E(X_2)$. Thus, for a null intervention effect, the total effect is identical to RTM, $T(x_0;\boldsymbol{\theta}) = R(x_0;\boldsymbol{\theta})$.

By definition, the variance of $(X_1 - X_2 \mid X_1 > x_0)$ is

$$\text{var}(X_1 - X_2 \mid X_1 > x_0) = E\left((X_1 - X_2 \mid X_1 > x_0) - E(X_1 - X_2 \mid X_1 > x_0)\right)^2$$
$$= E\{X_1^2 + X_2^2 - 2X_1X_2 + \left(E(X_1 - X_2 \mid X_1 > x_0)\right)^2$$
$$- 2(X_1 - X_2)E(X_1 - X_2 \mid X_1 > x_0)\}. \qquad (5.8)$$

To complete the evaluation of equation (5.8), we derive some helpful identities. First, the second conditional moment of X_2 is

$$E(X_2^2 \mid X_1 > x_0) = \int_{x_0}^{\infty} f_t(x_1;\boldsymbol{\theta}_1)\left(\int_{-\infty}^{\infty} x_2^2 f(x_2 \mid x,\boldsymbol{\theta})dx_2\right)dx_1. \qquad (5.9)$$

As $E(X_2^2|X_1) = \int_{-\infty}^{\infty} x_2^2 f(x_2 \mid x_1,\boldsymbol{\theta})dx_2$, and $E(X_2^2|X_1) = \text{var}(X_2|X_1) + (E(X_2|X_1))^2$, equation (5.9) simplifies to

$$E(X_2^2 \mid X_1 > x_0) = \int_{x_0}^{\infty} (\text{var}(X_2|X_1) + (E(X_2|X_1))^2) f_t(x_1;\boldsymbol{\theta}_1)dx_1$$
$$= \text{var}(X_2 \mid X_1) + E_{x_1}\left(E(X_2|X_1)^2)\right). \qquad (5.10)$$

Following the same steps, it can be shown that

$$E(X_1X_2|X_1 > x_0) = E_{x_1}\{X_1 \times E(X_2 \mid X_1)\}. \tag{5.11}$$

Substituting equations (5.7), (5.10) and (5.11) in equation (5.8) and simplifying, we get

$$\begin{aligned}\text{var}(X_1 - X_2 \mid X_1 > x_0) &= \text{var}(X_2 \mid X_1) + Var\{X_1 - E(X_2 \mid X_1)\}\\ &= \text{var}(X_2 \mid X_1) + (1-\beta_{x_2,x_1})^2\text{var}(X_1 \mid X_1 > x_0).\end{aligned} \tag{5.12}$$

5.4 Expression of $T(x_0;\theta)$ for exponential family of distributions

The exponential family unifies many distributions into one framework. This helps in generalizing the derivation and estimation of $T(x_0;\boldsymbol{\theta})$ and its constituent parts $R(x_0;\boldsymbol{\theta})$ and $\delta(\boldsymbol{\theta})$. A probability distribution $f(x;\boldsymbol{\theta})$ is said to be a p parameter exponential family if it can be represented in the form

$$f(x;\boldsymbol{\theta}) = h(x)e^{\eta(\boldsymbol{\theta})^T t(x) - A(\eta)}, \qquad \boldsymbol{\theta} \in \boldsymbol{\Theta}, \tag{5.13}$$

where h is a known function, $t(X) = \{t_1(X),\dots,t_p(X)\}^T$ is a vector of sufficient statistics, $\eta(\boldsymbol{\theta}) = \{\eta_1(\boldsymbol{\theta}),\dots,\eta_p(\boldsymbol{\theta})\}^T$ is the natural parameter vector which is a twice continuously differentiable function of $\boldsymbol{\theta}$, and A is the $\log$ of a normalization factor. Here, $\boldsymbol{\Theta}$ is open and connected. An exponential family is said to be in canonical form if $\eta(\boldsymbol{\theta}) = \boldsymbol{\theta}$.

The respective equations for finding the mean vector and variance-covariance matrix of the sufficient statistics of an exponential family are

$$\begin{aligned}E\{t(X)\} &= \frac{dA(\eta)}{d\eta}, \quad \text{and}\\ \text{var}(t(X)) &= \frac{d^2A(\eta)}{d\eta^T d\eta}.\end{aligned}$$

Let $t(X_1,X_2)$ be the joint vector of known real-valued functions sufficient for the parameters of the truncated bivariate density function of the pre-post variables. Let the first and second elements of $t(X_1,X_2)$ be $t_1(X_1,X_2) = X_1$ and $t_2(X_1,X_2) = X_2$ with respective natural parameters $\eta_1(\boldsymbol{\theta})$ and $\eta_2(\boldsymbol{\theta})$. This holds for many members of the exponential family, but is not true in general, e.g, the beta distribution. Then, the means of the pre-post variables are

$$\begin{aligned} E(X_1 \mid X_1 > x_0) &= E\{t_1(X_1, X_2)\} = \frac{dA(\eta)}{d\eta_1}, \\ E_{x_1}\{E(X_2 \mid X_1)\} &= E\{t_2(X_1, X_2)\} = \frac{dA(\eta)}{d\eta_2}. \end{aligned}$$

Hence, $T(x_0;\boldsymbol{\theta})$ for an exponential family can be expressed as

$$T(x_0;\boldsymbol{\theta}) = \frac{dA(\eta)}{d\eta_1} - \frac{dA(\eta)}{d\eta_2}. \tag{5.14}$$

An expression for $R(x_0;\boldsymbol{\theta})$ can be deduced from equation (5.14) by assuming a null effect, i.e., $\delta(\boldsymbol{\theta}) = 0$ or $E(X_1) = E(X_2)$, and $\delta(\boldsymbol{\theta})$ can be obtained through subtraction when the true and random error components are independent. Similarly, var$(X_1 - X_2 \mid X_1 > x_0)$ for the exponential family can be expressed as

$$\text{var}(X_1 - X_2 \mid X_1 > x_0) = \frac{d^2A(\eta)}{d\eta_1^2} + \frac{d^2A(\eta)}{d\eta_2^2} - 2 \cdot \frac{d^2A(\eta)}{d\eta_1 d\eta_2}. \tag{5.15}$$

The bivariate Poisson (Campbell, 1934) and binomial (Aitken and Gonin, 1936) distributions cannot be expressed similarly to equation (5.13), and instead can be expressed as

$$f(x;\boldsymbol{\theta}) = \sum_{\alpha=0}^{\min(x)} h(x, \boldsymbol{\theta}, \alpha) e^{\eta(\boldsymbol{\theta})^T t(x) - A(\eta)}, \qquad \boldsymbol{\theta} \in \boldsymbol{\Theta}. \tag{5.16}$$

However, equations (5.14) and (5.15) can still be used for derivation of the total, RTM and intervention effects and their variances.

5.5 Expression of $R(x_0;\boldsymbol{\theta})$ for selected bivariate distributions

Based on the nature and dependence structure of pre-post variables, $X_1 = X_0 + E_1$ and $X_2 = a + bX_0 + E_2$, different cases arise. For each case, RTM formulae for some well known distributions, including members of the exponential family are discussed in the following subsections. Note that only a right cut-off point is considered here for demonstrative purposes, and formulae for RTM would be different for a left cut-off point when the distribution is not symmetric.

5.5.1 The bivariate Poisson distribution

For the bivariate Poisson distribution, the error terms E_i are independent of X_0 but not identically distributed, for $i = 1, 2$. Here, $a = 0$ and $b = 1$ as X_1 and X_2 are count variables. The probability mass function of the bivariate Poisson is

$$f(y_1, y_2) = e^{-(\theta_0+\theta_1+\theta_2)} \frac{\theta_1^{y_1}}{y_1!} \frac{\theta_2^{y_2}}{y_2!} \sum_{x=0}^{\min(y_1,y_2)} x! \left(\frac{\theta_0}{\theta_1\theta_2}\right)^x \binom{y_1}{x}\binom{y_2}{x} \quad y_1,\ y_2 = 0, 1, 2...$$

As stated earlier, the bivariate Poisson distribution is not explicitly expressible like equation (5.13); however, the truncated bivariate Poisson distribution can be represented as

$$f(y_1, y_2) = h(x, \theta) \exp\Big[y_1 \log\theta_1 + y_2 \log\theta_2 - (\theta_0 + \theta_1 + \theta_2) - \log\{1 - F(x_0 \mid \theta_0 + \theta_1)\}\Big].$$

In this representation,

$$h(x, \boldsymbol{\theta}) = \sum_{x=0}^{\min(y_1,y_2)} x! \left(\frac{\theta_0}{\theta_1\theta_2}\right)^x \binom{y_1}{x}\binom{y_2}{x} / y_1! y_2!,$$

$\theta = \{\theta_0, \theta_1, \theta_2\}$, $p = 3$, $t(X, Y) = \{y_1, y_2\}^T$, $\eta = \{\eta_1 = \log\theta_1, \eta_2 = \log\theta_2\}^T$ and the log normalizing factor is

$$A(\eta) = \theta_0 + e^{\eta_1} + e^{\eta_2} + \log\{1 - F(x_0 \mid \theta_0 + e^{\eta_1})\}.$$

The respective formulae for the total, RTM and intervention effects are then

$$\begin{aligned} T_P(x_0; \boldsymbol{\theta}) &= \theta_1 \cdot \frac{1 - F(x_0 - 1 \mid \theta_0 + \theta_1)}{1 - F(x_0 \mid \theta_0 + \theta_1)} - \theta_2, \\ R_P(x_0; \boldsymbol{\theta}) &= \theta_1 \cdot \frac{f(x_0 \mid \theta_0 + \theta_1)}{1 - F(x_0 \mid \theta_0 + \theta_1)} \\ \delta_P(\boldsymbol{\theta}) &= \theta_1 - \theta_2, \end{aligned}$$

where $f(x \mid \lambda)$ and $F(x \mid \lambda)$ are the respective Poisson probability mass and distribution functions.

5.5.2 The bivariate binomial distribution

For the bivariate binomial distribution, X_0 and e_i are not independent, although e_i for $i = 1, 2$, are mutually independent but not identically distributed. As pre-post variables are discrete, this restricts the range of the real constants to $a = 0$ and $b = 1$. The joint distribution of pre-post variables is

$$f(y_1, y_2; \boldsymbol{\phi}, n) = \sum_{\alpha=0}^{\min(y_1,y_2)} \binom{n}{\alpha, y_1-\alpha, y_2-\alpha, y_3} \phi_0^{\alpha}\phi_1^{y_1-\alpha}\phi_2^{y_2-\alpha}\phi_3^{y_3}, \quad y_1,\ y_2 = 0,1,2...n$$

where $\boldsymbol{\phi} = (\phi_0, \phi_1, \phi_2)$, $\phi_3 = 1 - \phi_0 - \phi_1 - \phi_2$ and $y_3 = n + \alpha - y_1 - y_2$. Likewise, the truncated bivariate binomial distribution can be expressed as equation (5.16), and formulae for the total, RTM and intervention effects are then

$$T_B(y_0, \boldsymbol{\phi}) = n \cdot \frac{\phi_1(1 - F_{n-1}(y_0 - 1|\phi_0 + \phi_1)) - \phi_2\left\{1 - F_n(y_0|\phi_0+\phi_1) - (\phi_0+\phi_1)P_{n-1}(Y_1 = y_0)\right\}}{1 - F_n(y_0|\phi_0+\phi_1)},$$

$$R_B(y_0, \boldsymbol{\phi}) = n\phi_1 \cdot \frac{P_{n-1}(Y_1 = y_0)}{1 - F_n(y_0|\phi_0+\phi_1)}$$

$$\delta_B(\boldsymbol{\phi}) = \frac{T_B(y_0,\phi) - R_B(y_0,\phi)}{1 - B(y_0, \phi_0+\phi_1)} = n(\phi_1 - \phi_2),$$

where $P_n(X = x)$ and $F_n(x|p)$ are the binomial probability mass and distribution functions for n trials, and $B(y_0, \phi_0{+}\phi_1) = (\phi_0{+}\phi_1)P_{n-1}(Y_1 = y_0)/\left\{1 - F_n(y_0|\phi_0+\phi_1)\right\}$. The intervention effect cannot simply be obtained by subtracting $R_B(y_0, \boldsymbol{\phi})$ from $T_B(y_0, \boldsymbol{\phi})$ due to the dependence of X_0 and e_i. However, $\lim_{n\to\infty} B(y_0, \phi_0+\phi_1) = 0$ and simple subtraction is approximately correct for large n.

5.5.3 The bivariate normal distribution

The bivariate normal distribution is a member of the exponential family of distributions and $t_i(X_1, X_2) \in t(X_1, X_2)$ for $i = 1, 2$. Here, the error terms E_i for $i = 1, 2$ are independently and identically distributed and also independent of the true component X_0. In the formulation of James (1973), X_1 and X_2 are strictly positively correlated $\rho > 0$, but the real constants a and b allows X_1 and X_2 to be negatively correlated and have different population means. The probability density function of the bivariate normal is

$$f(x_1, x_2) = \frac{\exp\left[-\frac{1}{2(1-\rho^2)}\left\{(\frac{x_1-\mu_1}{\sigma_1})^2 + (\frac{x_2-\mu_2}{\sigma_2})^2 - 2\rho \times (\frac{x_1-\mu_1}{\sigma_1})(\frac{x_2-\mu_2}{\sigma_2})\right\}\right]}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}},$$

where $-\infty < X_1, X_2 < \infty$. Expressions for the total, RTM and intervention effects are

$$\begin{aligned} T_N(x_0;\boldsymbol{\theta}) &= (\sigma_1 - \rho\sigma_2)\cdot\frac{\phi(z_0)}{1-\Phi(z_0)} + \mu_1 - \mu_2, \\ R_N(x_0;\boldsymbol{\theta}) &= (\sigma_1 - \rho\sigma_2)\cdot\frac{\phi(z_0)}{1-\Phi(z_0)} \\ \delta_N(\boldsymbol{\theta}) &= \mu_1 - \mu_2, \end{aligned}$$

where $z_0 = (x_0 - \mu_1)/\sigma_1$. The James (1973) formula for RTM is a special case of $R_N(x_0;\boldsymbol{\theta})$ when $b = 1$, which implies that $\sigma_1 = \sigma_2$.

5.5.4 The bivariate log-normal distribution

The lognormal distribution is a member of the exponential family, but $t_i(X,Y) \notin t(X,Y)$ for $i = 1,2$ and equation (5.7) should be used for derivation of RTM instead of equation (5.14). The dependence structure and the real constants remain the same as in the case of the normal distribution to add flexibility. The joint distribution of the pre-post observations is

$$f(x_1,x_2) = \frac{\exp\left[-\frac{1}{(1-\rho^2)}\left\{\left(\frac{\log x_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{\log x_2-\mu_2}{\sigma_2}\right)^2 - 2\rho\times\left(\frac{\log x_1-\mu_1}{\sigma_1}\right)\left(\frac{\log x_2-\mu_2}{\sigma_2}\right)\right\}\right]}{2\pi\sigma_1\sigma_2 x_1 x_2\sqrt{1-\rho^2}},$$

where $0 < x_1, x_2 < \infty$. Using equation (5.7), expressions for $T_{LN}(y_0,\theta)$, $R_{LN}(y_0,\theta)$ and $\delta_{LN}(\theta)$ are then

$$\begin{aligned} T_{LN}(y_0,\boldsymbol{\theta}) &= E(X_1)\left\{\frac{1-\Phi(z_0'-\sigma_1)}{1-\Phi(z_0')} - 1\right\}\left\{1 - \frac{E(X_2)}{E(X_1)}\cdot\frac{\exp(\rho\sigma_1\sigma_2)-1}{\exp(\sigma_1^2)-1}\right\} + \delta_{LN}(\boldsymbol{\theta}), \\ R_{LN}(y_0,\boldsymbol{\theta}) &= E(X_1)\left\{\frac{1-\Phi(z_0'-\sigma_1)}{1-\Phi(z_0')} - 1\right\}\left\{1 - \frac{E(X_2)}{E(X_1)}\cdot\frac{\exp(\rho\sigma_1\sigma_2)-1}{\exp(\sigma_1^2)-1}\right\} \\ \delta_{LN}(\boldsymbol{\theta}) &= E(X_1) - E(X_2), \end{aligned}$$

where $E(X_i) = \exp(\mu_i + \sigma_i^2/2)$ for $i = 1,2$ and $z_0' = (\log x_0 - \mu_1)/\sigma_1$.

5.5.5 The bivariate Pareto-I distribution

The Pareto distribution belongs neither to the exponential family nor can it be expressed as the linear sum of random variables. However, expressions for RTM can be derived under a suitable bivariate distribution. For illustrative purposes, consider the bivariate Pareto-I distribution (Mardia, 1962)

$$f(x_1,x_2) = \frac{p(p+1)(a_1a_2)^{p+1}}{(a_2x_1 + a_1x_2 - a_1a_2)^{p+2}}, \quad 0 < a_1 < x_1,\ 0 < a_2 < x_2,\ p > 0.$$

Using equation (5.7), the respective total, RTM and intervention effects are

$$T_{Pr}(x_0;\boldsymbol{\theta}) = \frac{(x_0 - a_1)(pa_1 - a_2) + a_1 p(a_1 - a_2)}{a_1(p-1)},$$
$$R_{Pr}(x_0;\boldsymbol{\theta}) = \frac{(x_0 - a_1)(pa_1 - a_2)}{a_1(p-1)}$$
$$\delta_{Pr}(\boldsymbol{\theta}) = \frac{p(a_1 - a_2)}{p-1}.$$

5.6 Maximum likelihood estimation

Let $(x_{11}, x_{21}), (x_{12}, x_{22}), \ldots, (x_{1n}, x_{2n})$ be a random sample of pairs of observations of size n from a truncated bivariate distribution. The respective likelihood and log likelihood functions are

$$L(\boldsymbol{\theta}, \boldsymbol{x}) = \prod_{i=1}^{n} f_t(x_{1i}, x_{2i}; \boldsymbol{\theta}), \quad \text{and} \tag{5.17}$$
$$\ell(\boldsymbol{\theta}, \boldsymbol{x}) = \sum_{i=1}^{n} \log f_t(x_{1i}, x_{2i}; \boldsymbol{\theta}), \tag{5.18}$$

where

$$\boldsymbol{x} = \begin{pmatrix} x_{11} & x_{21} \\ x_{12} & x_{22} \\ \vdots & \vdots \\ x_{1n} & x_{2n} \end{pmatrix}.$$

To find the maximum likelihood estimate of $T(x_0;\boldsymbol{\theta})$, $R(x_0;\boldsymbol{\theta})$ and $\delta(\boldsymbol{\theta})$, a point estimate of $\boldsymbol{\theta}$ is required. This can be obtained by differentiating $\ell(\boldsymbol{\theta}, \boldsymbol{x})$ with respect to θ_j and setting the partial derivatives equal to zero to get a set of estimating equations

$$\frac{d\ell(\boldsymbol{\theta}, \boldsymbol{x})}{d\theta_j} = 0 \qquad \text{for} \qquad j = 1, ..., p. \tag{5.19}$$

A solution to this system of equations does not have a closed form and has to be solved numerically. Once $\boldsymbol{\theta}$ is estimated, the maximum likelihood estimates of $T(x_0;\boldsymbol{\theta})$, $R(x_0;\boldsymbol{\theta})$ and $\delta(\boldsymbol{\theta})$ can be obtained by substituting $\hat{\boldsymbol{\theta}}$ into their respective equations using the invariance property of maximum likelihood estimation as

$$\widehat{T}(x_0;\boldsymbol{x}) = T(x_0;\hat{\boldsymbol{\theta}}),$$
$$\widehat{R}(x_0;\boldsymbol{x}) = R(x_0;\hat{\boldsymbol{\theta}})$$
$$\hat{\delta}(\boldsymbol{x}) = \delta(\hat{\boldsymbol{\theta}}). \tag{5.20}$$

When the joint truncated distribution $f_t(x_1, x_2; \boldsymbol{\theta})$ is expressible as in equations (5.13) or (5.16), $t_1(X_1, X_2) = X_1$ and $t_2(X_1, X_2) = X_2$, then the respective likelihood and log likelihood functions are

$$L(\boldsymbol{\theta}, \boldsymbol{x}) = \left\{ \prod_{i=1}^{n} h(x) \right\} e^{\sum_{i=1}^{n} \eta(\boldsymbol{\theta})^T t(x_1, x_2) - n \cdot A(\eta)},$$

and

$$\ell(\boldsymbol{\theta}, \boldsymbol{x}) = \sum_{i=1}^{n} \log h(x) + \sum_{i=1}^{n} \eta(\boldsymbol{\theta})^T t(x_1, x_2) - n \cdot A(\eta). \tag{5.21}$$

To find the maximum likelihood estimate of $T(x_0; \boldsymbol{\theta})$, differentiating $\ell(\boldsymbol{\theta}, \boldsymbol{x})$ with respect to η_1 and η_2 respectively, and equating the partial derivatives to zero we get

$$\overline{x}_1 = \frac{dA(\eta)}{d\eta_1} = E(X_1 \mid X_1 > x_0) \tag{5.22}$$

$$\overline{x}_2 = \frac{dA(\eta)}{d\eta_2} = E_{x_1} \left\{ E(X_2 \mid X_1) \right\}, \tag{5.23}$$

where $\overline{x}_1 = \sum_{i=1}^{n} x_{1i}/n$ and $\overline{x}_2 = \sum_{i=1}^{n} x_{2i}/n$. Subtracting equation (5.23) from equation (5.22), we get

$$\widehat{T}(x_0; \boldsymbol{x}) = \overline{x}_1 - \overline{x}_2. \tag{5.24}$$

If $\widehat{R}(x_0; \boldsymbol{x}) \neq 0$, then $\overline{x}_1 - \overline{x}_2$ is not the maximum likelihood estimate of the intervention effect. The estimates of $R(x_0; \boldsymbol{\theta})$ and $\delta(\boldsymbol{\theta})$ can be obtained by finding the maximum likelihood estimate of $\boldsymbol{\theta}$ and then substituting $\hat{\boldsymbol{\theta}}$ in their respective equations as in equation (5.19).

5.6.1 Variances of the estimators

Dividing equation (5.15) by n, the variance of $\widehat{T}(x_0; \boldsymbol{x})$ is

$$\text{var}\{\widehat{T}(x_0; \boldsymbol{x})\} = \text{var}(X_1 - X_2 \mid X_1 > x_0)/n. \tag{5.25}$$

For the variance of $\widehat{R}(x_0; \boldsymbol{x})$, we only need to replace θ_2 with θ_1 in equation (5.25) as

$$\text{var}(\widehat{R}(x_0; \boldsymbol{x})) = \text{var}\left(X_1 - X_2 \mid X_1 > x_0, E(X_1) = E(X_2)\right)/n. \tag{5.26}$$

Lastly, dividing equation (5.2) by n, the variance of $\hat{\delta}(\boldsymbol{x})$ is

$$\text{var}\{\hat{\delta}(\boldsymbol{x})\} = \text{var}(X_1 - X_2)/n. \tag{5.27}$$

5.6.2 Unbiasedness of the estimators

In general, with a notable exception of an exponential family, where $t_i(X_1, X_2) \in t(X_1, X_2)$ for $i = 1, 2$, the unbiasedness property of $\widehat{T}(x_0; \boldsymbol{x})$, $\widehat{R}(x_0; \boldsymbol{x})$ and $\hat{\delta}(\boldsymbol{x})$ cannot be proved theoretically. Assuming the true and error component are independent, $t_1(X_1, X_2) = X_1$ and $t_2(X_1, X_2) = X_2$ hold true, taking expectation of equation (5.24) and using equation (5.14), we get

$$E\{\widehat{T}(x_0; \boldsymbol{x})\} = T(x_0; \boldsymbol{\theta}). \tag{5.28}$$

The RTM and total effects are equal for a null intervention effect, i.e., $E(X_1) = E(X_2)$, and we can write

$$E\{\widehat{T}(x_0; \boldsymbol{x} \mid E(X_1) = E(X_2))\} = E\{\widehat{R}(x_0; \boldsymbol{x})\} = R(x_0; \boldsymbol{\theta}). \tag{5.29}$$

Writing equation (5.28) in terms of its constituent parts, we get

$$E\{\widehat{R}(x_0; \boldsymbol{x}) + \hat{\delta}(\boldsymbol{x})\} = R(x_0; \boldsymbol{\theta}) + \delta(\boldsymbol{\theta}). \tag{5.30}$$

Subtracting equation (5.29) from equation (5.30), we have

$$E\{\hat{\delta}(\boldsymbol{x})\} = \delta(\boldsymbol{\theta}), \tag{5.31}$$

which completes the proof of unbiasedness of estimators.

5.6.3 Asymptotic distributions

By the Central Limit Theorem, the distribution of the sample mean is asymptotically normally distributed irrespective of its parent distribution, if $E(X^2) < \infty$. As $\widehat{T}(x_0; \boldsymbol{x})$ is the difference of the sample means of X and Y, it is asymptotically normally distributed, $\sqrt{n}\{\widehat{T}(x_0; \boldsymbol{x}) - T(x_0; \boldsymbol{\theta})\} \overset{d}{\sim} N(0, \text{var}(X_1 - X_2 \mid X > x_0))$. The additive components of a normally distributed random variable are also normally distributed (Cramér, 1936). Hence, $\sqrt{n}\{\widehat{R}(x_0; \boldsymbol{x}) - R(x_0; \boldsymbol{\theta})\} \overset{d}{\sim} N(0, \text{var}(X_1 - X_2 \mid X > x_0, E(X_1) = E(X_2))$ and $\sqrt{n}\{\hat{\delta}(\boldsymbol{x}) - \delta(\boldsymbol{\theta})\} \overset{d}{\sim} N(0, \text{var}(X_1 - X_2))$.

5.7 Data Example: Cholesterol Levels

Data on cholesterol levels from a study undertaken by the Lipid Research Clinics (Senn and Brown, 1985) are used for highlighting the differences between the proposed and existing methods (James, 1973; Senn and Brown, 1985; Beath and Dobson, 1991) for estimating the RTM and intervention effects. Men aged 35 to 39 years were screened for cholesterol levels and those in excess of 265 mg% proceeded further in the trial. Cholesterol levels were remeasured for the screened participants before undergoing any treatment. The respective sample means, variances and regression coefficient of the pre-post log cholesterol levels were as follows:

$$\overline{x}_1 = 5.676,\ \overline{x}_2 = 5.634,\ s_1^2 = 0.00728,\ s_2^2 = 0.01348, \text{and } \widehat{\beta} = 0.792.$$

The cut-off point, x_0, for log cholesterol was 5.58. Assuming bivariate normality for the pre-post cholesterol levels, the estimated parameters obtained by Davis (1976) using the James (1973) method of estimation, and the method of maximum likelihood developed by Senn and Brown (1985) are presented in Table 1, where μ, σ, γ and ρ are the mean, standard deviation, treatment parameter and

Table 5.1. Parameter estimates for James's model

Methods	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\gamma}$	$\hat{\rho}$
James	5.207	0.228	0.870	0.910
Senn & Brown	5.390	0.186	0.985	0.861

correlation coefficients, respectively. Note that this model allows the post measurements to move only in the direction of the mean. To estimate the total effect, RTM and intervention effects by the proposed method for comparison with the estimates obtained by existing methods, the parameters of the bivariate normal distributions were estimated as

$$\hat{\mu}_1 = 5.205,\ \hat{\sigma}_1 = 0.229,\ \hat{\mu}_2 = 5.261,\ \hat{\sigma}_2 = 0.204,\ \text{and } \hat{\rho} = 0.887.$$

The maximum likelihood estimates of the total, RTM and intervention effects for both methods are given in Table 2. For the Beath and Dobson (1991) method using Edgeworth approximation, $\gamma_1 = 0$ and $\gamma_2 = 3$ were used as the respective coefficient of skewness and kurtosis for the normal distributions along with the Senn and Brown (1985) estimated parameters. The Beath and Dobson method lacks a mechanism for decomposing the total effect into its constituent parts, though it was done here through subtraction for demonstrative purposes. For comparison purposes, the total effect was decomposed into the RTM and interventions effects by the proportional reduction formula of James (1973)

$$\frac{\text{proportional reduction due to regression}}{\text{total proportional reduction}} = \frac{1-\rho}{1-\gamma\rho}.$$

The methods produced differing results in terms of the estimated intervention

Table 5.2. Comparison of Suggested and existing methods

Methods	$\widehat{T}(x_0;\boldsymbol{x})$	$\widehat{R}(x_0;\boldsymbol{x})$	$\hat{\delta}(\boldsymbol{x})$
James	0.042	0.018	0.024
Senn & Brown	0.042	0.038	0.004
Beath & Dobson (Edgeworth)	0.042	0.045	-0.003
Beath & Dobson (Saddlepoint)	0.042	0.040	0.002
Proposed	0.042	0.098	-0.056

effects. Due to the constraint imposed on the direction of the post measurement mean, a non-negative intervention effect will always be observed, $\hat{\delta}(\boldsymbol{x}) \geq 0$, using the James (1973) approach. Whereas, the proposed method allows an intervention effect to be in either direction. A negative intervention effect could lessen the observed change or hide an adverse effect. In particular, the James (1973) method estimates a positive intervention effect, Senn and Brown (1985) and Beath and Dobson (1991) estimate a nearly null effect, and the proposed method estimates a negative effect which is more than twice the magnitude estimated by the James method. That is, the method for accounting for RTM could influence the conclusion in any direction.

5.8 Discussion

Regression to the mean can occur whenever an intervention or treatment is applied to subjects selected in the extreme of a distribution. Ignoring regression to the mean in data analysis can potentially affect statistical inferences by exaggerating results. Further, the use of restrictive methods can potentially conceal an adverse intervention effect.

In clinical trials, treatments are often designed to change the post measurement mean in the direction of the population mean, although this change can potentially occur in any direction. Existing methods for RTM are limited due to restrictive assumptions about the distribution of pre-post variables and associated models for the intervention effect.

The maximum likelihood estimators were derived for the total, RTM and intervention effects. The statistical properties of unbiasedness, consistency and asymptotic normality were established where possible. In the presence of a regression to the mean effect, the difference of the sample means is not the maximum likelihood estimate of an intervention effect, as is usually the case.

The total, RTM and intervention effects were evaluated for cholesterol levels using the maximum likelihood method. A comparison of the proposed method with existing methods for RTM and intervention effects gave substantially different results. The proposed method allows more flexibility for an intervention study in terms of the direction of the intervention effect and allows for negative correlation between pre-post variables. Accounting for RTM increases accuracy in estimating an intervention effect.

Intervention or treatment studies where subjects are selected based on a cut-off point should account for RTM to avoid erroneous conclusions. The expressions derived in this study allow the intervention effect to be either favourable or adverse and could be used to estimate an unbiased intervention effect by accounting for RTM under any bivariate distribution.

Chapter 6

Discussion and future work

Pre-post study designs are often used to measure the within participant change in a variable of interest after the introduction of a treatment or intervention. When subjects are selected for study based on a cut-off point in the tail of a distribution, the inference drawn could be susceptible to the regression to the mean effect. This fact has been reported in diverse research areas including but not limited to public health, social psychology, economics, and sports management decisions.

The RTM effect could be mitigated by randomly assigning subjects to comparison groups (placebo and treatment), but ethical and/or logistical constraints limit its applicability. Estimating and accounting for RTM is another option to accurately estimate the intervention effect. However, existing methods for RTM are based on some restrictive assumptions, including bivariate normality, which may not hold true. Current methods developed for non-normal populations have limitations such as inapplicability to the empirical distribution, non-negativeness of the probability density function and multi-modality distributional problems, and can be computationally expensive.

In a pre-post study design, when an intervention or treatment is applied to subjects selected on the basis of a certain threshold, RTM could exaggerate the observed change called the total effect. The total effect is the sum of RTM and a function of intervention intervention effects, and RTM should be accounted for to accurately estimate the intervention effect.

Count or rate of occurrence data is often modelled by the Poisson distribution, and an objective of this thesis was to derive RTM formulae for the bivariate Poisson distribution. For the Poisson cases, it was also demonstrated that the conditional mean difference was the sum of the RTM and treatment effects.

Generally, when the correlation ρ between pre-post variables increases, the RTM effect decreases. However, RTM as a function of ρ behaves differently for different distributions. It is linearly related to ρ for the bivariate normal distribution irrespective of the direction of the cut-off point. Whereas, for the bivariate Poisson distribution, it decreases as ρ increases but the behaviour is non-linear and depends on the direction of cut-off point. Moreover, RTM increases as the cut-off point is selected farther in the tail of a distribution, and the behavior of RTM is markedly dissimilar for homogeneous and inhomogeneous Poisson processes.

A log-transformation is often useful for transforming positively skewed data to an approximate normal distribution. Methods and formulae based on a normal assumption can then be applied to log transformed data. For NSW road fatality data and the simulated bivariate Poisson data, RTM estimates were computed assuming a bivariate normal distribution for the log transformed data. In both instances, RTM was severely underestimated using the log-transformation approach and therefore this approach is not recommended.

The correlation ρ is strictly positive for the bivariate normal and Poisson distributions, although this assumption does not hold generally. In particular, for the bivariate binomial distribution, ρ can take both positive or negative values. For a left cut-off point, RTM decreases steeply as ρ takes on values in its range from the lower to upper limits. For a right cut-off, RTM increases reaching a maximum when ρ is around zero and then starts decreasing as ρ increases. When $\rho < 0$, RTM could have a relatively greater range and, consequently, the total effect is comparatively more exaggerated.

Distributional approximations make calculation easier when the relevant assumptions are satisfied. Using real and simulated data, however, RTM was underestimated under normal and Poisson approximations to the binomial distribution when conditions for approximations were suitable. The magnitude of underestimation is greater for the normal approximation compared to the Poisson.

Furthermore, simple subtraction of RTM from the total effect would give a biased intervention effect when the pre-post variables follow the bivariate binomial distribution. The bias term is proportional to the difference of the pre-post parameters and approaches zero as the sample size increases. Existing methods assume the treatment changes the post measurement mean in the direction of the population mean. Potentially, an adverse treatment effect could change the post measurement mean in the opposite direction than intended. For the Muscatine data on the number of obese individuals, the age effect is negative for the first two age cohorts, indicating that obesity has increased with age. The total effect could conceal an adverse effect if analytic methods do not account for RTM.

In the existing literature, RTM formulae have been derived assuming (i) identical distribution of the pre-post variables, (ii) strictly positive correlation, (iii) the direction of the post measurements to change towards the population mean, and (iv) the error components being normally distributed with zero mean and constant variance. Relaxing those assumptions, formulae for the total effect are derived and decomposed into RTM and treatment/intervention effects. The generalized derivations allow the pre-post variables to be distributed with different parameters, thereby allowing the treatment effect to be either positive or negative. This fact was observed with the Lipid Research Clinics data on cholesterol levels where the treatment effect was negative in contrast with the existing methods. In the proposed set up, the correlation can take any value in its range, where possible. For negative correlation coefficients, the range of RTM increases and could further exaggerate the observed change away from the true treatment effect.

RTM formulae and maximum likelihood estimators can be simplified for bivariate distributions that belong to the p parameter exponential family and can be written in canonical form. The unbiasedness, consistency and asymptotic normality of the maximum likelihood estimators of RTM have been established theoretically in general for the exponential family. The asymptotic properties have been verified through simulations for the bivariate Poisson and binomial distributions.

Using data on cholesterol levels, the total, RTM and intervention effects were estimated by the maximum likelihood method assuming the bivariate normal distribution. A comparison of the proposed method with existing methods for RTM and intervention effects gave substantially different results. In particular, existing methods estimate either a positive or nearly null intervention effect, and the proposed method estimates a negative intervention effect which is more than twice the magnitude of the highest estimated intervention effect by one of the existing methods.

In sum, the proposed methods derived in this thesis allow for more flexibility in estimating the regression to the mean effects which, in turn, allows for more accurate estimation of the intervention or treatment effects.

6.1 Future work

The methods developed in this thesis can be further extended, but the time and/or resources do not allow me to explore them in full detail. In the future, I plan to work on the following research topics.

6.1.1 Interrupted time series

Interrupted time series analysis is frequently used in quasi-experimental designs for assessing the impact of interventions when a randomized controlled design cannot be conducted. Pairs of time series observations, with lag h are often modelled by a bivariate distribution. Potentially, the ideas and techniques developed in this thesis could be applied to the joint distribution of the pairs of time series observations to decompose the effect into its constituent parts to accurately estimate the intervention effect.

6.1.2 Statistical process control

In statistical process control, control charts are important tools used for improving the quality of products and/or services by reducing assignable process variability. Whenever a control chart detects an out of control situation based on the extreme values of a charting statistic, an interruption is made to bring the process back in control state by some necessary adjustments, thereby producing products at the nominal value. The nominal value is usually the mean of the charting statistic. Accounting for RTM could potentially be helpful in making accurate adjustments and hence increasing productivity.

6.1.3 Revisiting the bivariate normal distribution

In this thesis, RTM formulae have been derived in general for bivariate distributions by relaxing some restrictive assumptions. However, general derivation of maximum likelihood estimators of RTM, and the effect of the RTM and its variance on statistical inference when the correlation is negative are worthy of further exploration. In particular, although several approaches exist for RTM assuming the bivariate normal distribution, very little attention has been given to the negatively correlated case.

6.1.4 Writing an `R` package

Estimation of the RTM and intervention effects requires estimation of the parameters of truncated bivariate distributions. The methods developed in this thesis are not straight forward, and a companion `R` package would increase the likelihood these methods would be used by other researchers.

The variable of interest in intervention or treatment studies could be binary, count, or continuous, and not necessarily normally distributed. The expressions derived in this study allow the intervention effect to be either favourable or adverse and could be used to estimate an unbiased intervention effect by accounting for RTM under any bivariate distribution, particularly, the bivariate Poisson, binomial, normal, log-normal and Pareto-I distributions.

Bibliography

Aitken, A. C. and Gonin, H. T. (1936). On fourfold sampling with and without replacement. *Proceedings of the Royal Society of Edinburgh*, 55:114–125.

Anderson, C. (2013). Poisson regression or regression of counts and rates. http://courses.education.illinois.edu/EdPsy589/lectures/4glm3-ha-online.pdf.

Balakrishnan, N. and Lai, C.-D. (2009). *Continuous Bivariate Distributions*. Springer, New York, second edition.

Barnett, A. G., van der Pols, J. C., and Dobson, A. J. (2005). Regression to the mean: what it is and how to deal with it. *International Journal of Epidemiology*, 34(1):215–220.

Barton, D. E. and Dennis, K. E. (1952). The conditions under which gram-charlier and edgeworth curves are positive definite and unimodal. *Biometrika*, 39(3/4):425–427.

Beath, K. J. and Dobson, A. J. (1991). Regression to the mean for nonnormal populations. *Biometrika*, 78(2):431–435.

Bland, J. M. and Altman, D. G. (1994). Statistics notes: Some examples of regression towards the mean. *BMJ*, 309(6957):780.

Bono, R., Blanca, M. J., Arnau, J., and Gómez-Benito, J. (2017). Non-normal distributions commonly used in health, education, and social sciences: A systematic review. *Frontiers in psychology*, 8:1602.

Bowan, A. W. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, 71(2):353–360.

Burke, R., Meyer, A., Kay, C., Allensworth, D., and Gazmararian, J. (2014). A holistic school-based intervention for improving health-related knowledge, body composition, and fitness in elementary school students: an evaluation of the healthmpowers program. *International Journal of Behavioral Nutrition and Physical Activity*, 11(1):78.

Campbell, J. T. (1934). The Poisson correlation function. *Proceedings of the Edinburgh Mathematical Society*, 4(1):18–26.

Cao, R., Cuevas, A., and Manteiga, W. G. (1994). A comparative study of several smoothing methods in density estimation. *Computational Statistics and Data Analysis*, 17(2):153 – 176.

Changzheng, Z. and Kai, G. (2010). The relationship between employees' turnover rate and industrial firm performance: A literature review. In *2010 IEEE International Conference on Advanced Management Science (ICAMS 2010)*, volume 2, pages 383–387.

Chaspari, T., Goodwin, M., Wilder-Smith, O., Gulsrud, A., Mucchetti, C., Kasari, C., and Narayanan, S. (2014). A non-homogeneous poisson process model of skin conductance responses integrated with observed regulatory behaviors for autism intervention. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1611–1615.

Chen, S. and Cox, C. (1992). Use of baseline data for estimation of treatment effects in the presence of regression to the mean. *Biometrics*, 48(2):593–598.

Choi, E. and Hall, P. (1999). Data sharpening as a prelude to density estimation. *Biometrika*, 86(4):941–947.

Cohen, A. C. (1955). Restriction and selection in samples from bivariate normal distributions. *Journal of the American Statistical Association*, 50(271):884–893.

Cramér, H. (1936). Über eine eigenschaft der normalen verteilungsfunktion. *Mathematische Zeitschrift*, 41(1):405–414.

Daniels, H. E. (1954). Saddlepoint approximations in statistics. *The Annals of Mathematical Statistics*, 25(4):631–650.

Das, P. and Mulder, P. (1983). Regression to the mode. *Statistica Neerlandica*, 37(1):15–20.

Davis, C. E. (1976). The effect of regression to the meam in epidemiologic and clinic studies. *American Journal of Epidemiology*, 104(5):493–498.

Denke, M. A. and Frantz, I. D. (1993). Response to a cholesterol-lowering diet: Efficacy is greater in hypercholesterolemic subjects even after adjustment for regression to the mean. *The American Journal of Medicine*, 94(6):626 – 631.

Easton, G. S. and Ronchetti, E. (1986). General saddlepoint approximations with applications to l statistics. *Journal of the American Statistical Association*, 81(394):420–430.

Ederer, F. (1972). Serum cholesterol changes: effects of diet and regression toward the mean. *Journal of Chronic Diseases*, 25:277–89.

Erdélyi, A. e. a., editor (1954). *Tables of integral transforms*. McGraw, Hill, New York.

Faraway, J. J. and Jhun, M. (1990). Bootstrap choice of bandwidth for density estimation. *Journal of the American Statistical Association*, 85(412):1119–1122.

Feng, C., Hongyue, W., Lu, N., Chen, T., He, H., Lu, Y., and Tu, X. (2014). Log-transformation and its implications for data analysis. *Shanghai Archives of Psychiatry*, 26:105–9.

Filip, K. (2014). Kinds of chance in games and sports. *Sport, Ethics and Philosophy*, 8(1):65–76.

Frans, D. W. (1985). In praise of chance: A philosophical analysis of the element of chance in sports. *Journal of the Philosophy of Sport*, 12(1):52–61.

Fukunaga, K. and Hostetler, L. (1975). The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Transactions on Information Theory*, 21(1):32–40.

Gallos, L., Barttfeld, P., Shlomo, H., Sigman, M., and Hernán, A. (2012). Collective behavior in the spatial spreading of obesity. *Scientific Reports*, 2.

Galton, F. (1869). *Heredity genius: An inquiry into its laws and consequences*. London:Macmillan.

Galton, F. (1886). Regression towards mediocrity in hereditary stature. *Journal of the Anthropological Institute*, 15:246–263.

Galton, F. (1889). *Natural inheretence*. London:Macmillan.

Gardner, M. J. and Heady, J. A. (1973). Some effects of within-person variability in epidemiological studies. *Journal of Chronic Diseases*, 26(12):785–791.

Gonzàlez-Manteiga, W., Cao, R., and Marron, J. S. (1996). Bootstrap selection of the smoothing parameter in nonparametric hazard rate estimation. *Journal of the American Statistical Association*, 91(435):1130–1140.

Good, C., Aronson, J., and Inzlicht, M. (2003). Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat. *Journal of Applied Developmental Psychology*, 24(6):645–662.

Hall, P. and Marron, J. S. (1991). Local minima in cross-validation functions. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(1):245–252.

Hall, P. and Wehrly, T. E. (1991). A geometrical method for removing edge effects from kernel-type nonparametric regression estimators. *Journal of the American Statistical Association*, 86(415):665–672.

Hamdan, M. and Nasro, M. (1986). Maximum likelihood estimation of the parameters of the bivariate binomial distribution. *Communications in Statistics - Theory and Methods*, 15(3):747–754.

Hannon, B., Thomas, D., Siu, C., and Allison, D. (2018). The claim that effectiveness has been demonstrated in the parenting, eating and activity for child health (peach) childhood obesity intervention is unsubstantiated by the data. *British Journal of Nutrition*, page 1–2.

Hess, K. R., Serachitopol, D. M., and Brown, B. W. (1999). Hazard function estimators: a simulation study. *Statistics in Medicine*, 18(22):3075–3088.

Hougaard, P. (1988). A boundary modification of kernel function smoothing, with application to insulin absorption kinetics. In *Compstat*, pages 31–36, Heidelberg. Physica-Verlag HD.

Hougaard, P., Plum, A., and Ribel, U. (1989). Kernel function smoothing of insulin absorption kinetics. *Biometrics*, 45(4):1041–1052.

IMU (2018a). Independent monitoring unit, elementary and secondary education department government of khyber pakhtunkhwa, pakistan. http://175.107.63.45/NewIMUSite/index.aspx.

IMU (2018b). Independent monitoring unit health department khyber pakhtunkhwa, pakistan. http://www.healthkp.gov.pk/index.php/independent-monitoring-unit-health-department-kp/.

Irwig, L., Glasziou, P., Wilson, A., and Macaskill, P. (1991). Estimating an individual's true cholesterol level and response to intervention. *Chest*, 266(12):1678 – 1685.

James, K. E. (1973). Regression toward the mean in uncontrolled clinical studies. *Biometrics*, 29(1):121–130.

John, M. and Jawad, A. (2010). Assessing the regression to the mean for non-normal populations via kernel estimators. *North American Journal of Medical Sciences*, 2(7):288 – 292.

Johnson, W. D. and George, V. T. (1991). Effect of regression to the mean in the presence of within-subject variability. *Statistics in Medicine*, 10(8):1295–1302.

Jones, M. C., Marron, J. S., and Sheather, S. J. (1996). A brief survey of bandwidth selection for density estimation. *Journal of the American Statistical Association*, 91(433):401–407.

Jones, P. and Smith, P. (2010). *Stochastic Processes: An Introduction, Second Edition*.

Kario, K., Schwartz, J. E., and Pickering, T. G. (2000). Changes of nocturnal blood pressure dipping status in hypertensives by nighttime dosing of alpha-adrenergic blocker, doxazosin. *Hypertension*, 35(3):787–794.

Kawamura, K. (1984). Direct calculation of maximum likelihood estimator for the bivariate poisson distribution. *Kodai Mathematical Journal*, 7(2):211–221.

Kendall, M. G., Stuart, A., and Ord, J. K., editors (1987). *Kendall's Advanced Theory of Statistics*. Oxford University Press, Inc., New York, NY, USA.

Khan, M. and Olivier, J. (2018). Quantifying the regression to the mean effect in Poisson processes. *Statistics in Medicine*, 37(26):3832–3848.

Khan, M. and Olivier, J. (2019). Regression to the mean for the bivariate binomial distribution. *Statistics in Medicine*, 38(13):2391–2412.

Lauer, R. M., Clarke, W., and Burns, T. L. (1997). Obesity in childhood: the muscatine study. *Zhonghua Minguo xiao er ke yi xue hui za zhi [Journal]. Zhonghua Minguo xiao er ke yi xue hui*, 38(6):432–7.

Lee, D. S. and Lemieux, T. (2010). Regression Discontinuity Designs in Economics. *Journal of Economic Literature*, 48(2):281–355.

Lee, M. and Smith, G. (2002). Regression to the mean and football wagers. *Journal of Behavioral Decision Making*, 15(4):329–342.

Linden, A. (2013). Assessing regression to the mean effects in health care initiatives. *BMC Medical Research Methodology*, 13(1):119.

Linden, A. and Adams, J. L. (2012). Combining the regression discontinuity design and propensity score-based weighting to improve causal inference in program evaluation. *Journal of Evaluation in Clinical Practice*, 18(2):317–325.

Läuter, H. (1988). Silverman, b. w.: Density estimation for statistics and data analysis. chapman & hall, london – new york 1986, 175 pp., £12.—. *Biometrical Journal*, 30(7):876–877.

Mardia, K. V. (1962). Multivariate pareto distributions. *The Annals of Mathematical Statistics*, 33(3):1008–1015.

Marshall, A. W. and Olkin, I. (1985). A family of bivariate distributions generated by the bivariate bernoulli distribution. *Journal of the American Statistical Association*, 80(390):332–338.

McCambridge, J., Kypri, K., and McElduff, P. (2014). Regression to the mean and alcohol consumption: A cohort study exploring implications for the interpretation of change in control groups in brief intervention trials. *Drug and Alcohol Dependence*, 135:156 – 159.

Müller, H., Abramson, I., and Azari, R. (2003). Nonparametric regression to the mean. *Proceedings of the Nationa Academy of Sciences of the United States of America*, 100(17):9715 – 9720.

Müller, H.-G. and Yan, X. (2001). On local moments. *Journal of Multivariate Analysis*, 76(1):90 – 109.

Montgomery, D. C. (2013). *Introduction to Statistical Quality Control*. Hoboken, NJ:Wiley.

Moore, D. S., MacCabe, G. P., and Craig, B. A. (2017). *Introduction to the practice of statistics*, volume 8. New York : W.H. Freeman, Macmillan Learning.

Moores, C. J., Miller, J., Daniels, L. A., Vidgen, H. A., and Magarey, A. M. (2018). Pre–post evaluation of a weight management service for families with overweight and obese children, translated from the efficacious lifestyle intervention parenting, eating and activity for child health (peach). *British Journal of Nutrition*, 119(12):1434–1445.

Morton, V. and Torgerson, D. J. (2003). Effect of regression to the mean on decision making in health care. *BMJ*, 326(7398):1083–1084.

Morton, V. and Torgerson, D. J. (2005). Regression to the mean: treatment effect without the intervention. *Journal of Evaluation in Clinical Practice*, 11(1):59–65.

Muller, H.-G. and Wang, J.-L. (1994). Hazard rate estimation under random censoring with varying kernels and bandwidths. *Biometrics*, 50(1):61–76.

Musini, V. M. and Wright, J. M. (2009). Factors affecting blood pressure variability: Lessons learned from two systematic reviews of randomized controlled trials. *PLOS ONE*, 4(5):1–6.

Nadarajah, S. (2008). A review of results on sums of random variables. *Acta Applicandae Mathematicae*, 103(2):131–140.

Norman, L., Samuel, K., and Balakrishnan, N. (1997). *Discrete Multivariate Distributions*. John Wiley and Sons.

Olaf, G. and Holger, D. (1992). Nearest neighbour kernel estimation of the hazard function from censored data. *Journal of Statistical Computation and Simulation*, 43(1-2):93–101.

Paduthol, S., Nair, N., and Preethi, J. (2014). A family of bivariate Pareto distributions. *Statistica*, 74(2):199–215.

Patil, P. (1993). Bandwidth choice for nonparametric hazard rate estimation. *Journal of Statistical Planning and Inference*, 35(1):15 – 30.

Pritchett, L. and Summers, L. H. (2014). Asiaphoria meets regression to the mean. Working Paper 20573, National Bureau of Economic Research.

Ramlau-Hansen, H. (1983). Smoothing counting process intensities by means of kernel functions. *The Annals of Statistics*, 11(2):453–466.

Retting, R. A., Ferguson, S. A., and Hakkert, A. S. (2003). Effects of red light cameras on violations and crashes: A review of the international literature. *Traffic Injury Prevention*, 4(1):17–23. PMID: 14522657.

Rothman, T. and Henderson, M. (2011). Do school-based tutoring programs significantly improve student performance on standardized tests? *RMLE Online*, 34(6):1–10.

Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics*, 9(2):65–78.

Ruggeri, F. and Sivaganesan, S. (2005). On modeling change points in non-homogeneous poisson processes. *Statistical Inference for Stochastic Processes*, 8(3):311–329.

Sarda, P. and Vieu, P. (1991). Smoothing parameter selection in hazard estimation. *Statistics and Probability Letters*, 11(5):429 – 434.

Schectman, G. and Hoffmann., R. G. (1988). A history of hypercholesterolemia influences cholesterol measurements. *Archives of Internal Medicine*, 148(5):1169–1171.

Schilling, C., Petrie, D., Dowsey, M. M., Choong, P. F., and Clarke, P. (2017). The impact of regression to the mean on economic evaluation in quasi-experimental pre–post studies: The example of total knee replacement using data from the osteoarthritis initiative. *Health Economics*, 26(12):e35–e51.

Scott, D. W. and Terrell, G. R. (1987). Biased and unbiased cross-validation in density estimation. *Journal of the American Statistical Association*, 82(400):1131–1146.

Senn, S. (1990). Regression: A new mode for an old meaning? *The American Statistician*, 44(2):181–183.

Senn, S. J. and Brown, R. A. (1985). Estimating treatment effects in clinical trials subject to regression to the mean. *Biometrics*, 41(2):555–560.

Sheather, S. J. and Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(3):683–690.

Skinner, A. C., Heymsfield, S. B., Pietrobelli, A., Faith, M. S., and Allison, D. B. (2015). Ignoring regression to the mean leads to unsupported conclusion about obesity. *International Journal of Behavioral Nutrition and Physical Activity*, 12(1):56.

Stigler, S. M. (1986). *The History of Statistics: The Measurement of Uncertainty Before 1900*. Harvard University Press: Cambridge.

Stigler, S. M. (1997). Regression towards the mean, historically considered. *Statistical Methods in Medical Research*, 6(2):103–114. PMID: 9261910.

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical science : a review journal of the Institute of Mathematical Statistics*, 25:1–21.

Tanner, M. A. and Wong, W. H. (1984). Data-based nonparametric estimation of the hazard function with applications to model diagnostics and exploratory analysis. *Journal of the American Statistical Association*, 79(385):174–182.

Taylor, C. C. (1989). Bootstrap choice of the smoothing parameter in kernel density estimation. *Biometrika*, 76(4):705–712.

Teicher, H. (1954). On the multivariate poisson distribution. *Scandinavian Actuarial Journal*, 1954(1):1–9.

Ter Weel, B. (2006). Does manager turnover improve firm performance? new evidence using information from dutch soccer, 1986-2004. *IZA Discussion Papers 2483*.

Ton, Z. and Huckman, R. S. (2008). Managing the impact of employee turnover on performance: The role of process conformance. *Organization Science*, 19(1):56–68.

Tornqvist, L., Vartia, P., and Vartia, Y. (1985). How should relative changes be measured? *The American Statistician*, 39(1):43–46.

Transport for NSW (March 2018). NSW Centre for Road Safety. Crash and casualty statistics. http://roadsafety.transport.nsw.gov.au/statistics/interactivecrashstats/nsw.html?tabnsw=5.

Trochim, W. M. K. (2001). *Research methods knowledge base*. Cincinnati, OH : Atomic Dog Publishing, 2nd ed edition. Includes index.

Tse, K. (2014). Some applications of the Poisson process. *Applied Mathematics*, 5:3011–3017.

Twisk, J. (2003). *Applied longitudinal data analysis for epidemiology: A practical guide*. Cambridge University Press.

Wilcox, M. A., Chang, A. M. Z., and Johnson, I. R. (1996). The effects of parity on birthweight using successive pregnancies. *Acta Obstetricia et Gynecologica Scandinavica*, 75(5):459–463.

Woolson, R. and Clarke, W. (1984). Analysis of categorical incomplete longitudinal data. *Journal of the Royal Statistical Society. Series A (General)*, 147(1):87–99.

Yu, R. and Chen, L. (2014). The need to control for regression to the mean in social psychology studies. *Frontiers in Psychology*, 5(1574).

Yudkin, P. and Stratton, I. (1996). How to deal with regression to the mean in intervention studies. *The Lancet*, 347(8996):241 – 243.

Zaliapin, I., Kagan, Y. Y., and Schoenberg, F. P. (2005). Approximating the distribution of pareto sums. *Pure and Applied Geophysics*, 162(6):1187–1228.