

MODELING A CAUSAL RELATIONSHIP BETWEEN POISSON PROCESSES

by

Zeliha Bicak, B.S.

A thesis submitted to the Graduate Council of
Texas State University in partial fulfillment
of the requirements for the degree of
Master of Science
with a Major in Mathematics
August 2020

Committee Members:

Alexander White, Chair

Qiang Zhao

Kathleen Melhuish

COPYRIGHT

by

Zeliha Bicak

2020

FAIR USE AND AUTHOR'S PERMISSION STATEMENT

Fair Use

This work is protected by the Copyright Laws of the United States (Public Law 94-553, section 107). Consistent with fair use as defined in the Copyright Laws, brief quotations from this material are allowed with proper acknowledgment. Use of this material for financial gain without the author's express written permission is not allowed.

Duplication Permission

As the copyright holder of this work I, Zeliha Bicak, authorize duplication of this work, in whole or in part, for educational or scholarly purposes only.

ACKNOWLEDGMENTS

First, I would like to thank my husband, who supported me all the time and encouraged me to write this thesis and my family, who came all the way from Turkey to visit me during a tough time.

I would also like to thank Dr. Alexander White for all the help and guidance. Dr. White never lost his patience with me and was always willing to help when I needed it. This thesis would not be completed without his help.

Moreover, I would like to thank my committee members, Dr. Kathleen Melhuish and Dr. Qiang Zhao for being in my thesis committee and reviewing my work.

Last but not least, I would like to thank my friends, who are graduate students at Texas State University. They were always very supportive.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS.....	iv
LIST OF TABLES.....	vii
LIST OF FIGURES.....	viii
CHAPTER	
I. INTRODUCTION.....	1
Capturing the Nature of the Interaction	1
Homogeneous Poisson Process	3
Non-Homogeneous Poisson Process	9
II. SINGLE CLASSROOM MODEL	11
Teacher Process.....	11
Student Process.....	12
Asymptotic Distribution	15
III. SIMULATION STUDY.....	24
Simulation Description.....	24
Analysis	26
IV. META-ANALYSIS OF MULTIPLE CLASSROOMS SIMULATION	37
Motivation	37
Method.....	37
Parameter Estimation	38

	Parameter Estimation	38
	Asymptotic Distribution	39
	Meta-Analysis Study of Multiple Classrooms Simulation Description.....	40
	Analysis for Meta-Analysis of Multiple Classrooms Simulation	41
V.	APPLICATION	53
	Data Description.....	53
	Parameter Estimation	53
VI.	FUTURE RESEARCH	60
	APPENDIX SECTION.....	61
	REFERENCES	78

LIST OF TABLES

Table	Page
1 p Summation Formulas for small values of p	17
2 First 5 Rows of Results	25
3 Summary Statistics for Small Sample Distribution of $\hat{\gamma}$	27
4 Summary Statistics for Small Sample Distribution for $\hat{\alpha}_0$	28
5 Summary Statistics for the Distribution of $\hat{\gamma}$	29
6 Summary Statistics for Small Sample Distribution for $\hat{\alpha}_0$	30
7 First 5 Rows of Results	41
8 Summary Statistics for Small Distribution for $\hat{\gamma}$	43
9 Summary Statistics for Small Sample Distribution for $\hat{\alpha}_0$	44
10 Summary Statistics for Small Sample Distribution for $\hat{\gamma}$	46
11 Summary Statistics for Small Sample Distribution for $\hat{\alpha}_0$	47
12 The First Model (All Teacher Moves) Results For One Classroom . .	55
13 The Second Model (Low Moves) Results for One Classroom	55
14 The Third Model (High Moves) Results for One Classroom	56
15 The First Model (All Teacher Moves) Results For 24 Classrooms . . .	56
16 The Second Model (Low Teacher Moves) Results for 24 Classrooms .	57
17 The Third Model (High Moves) Results For 24 Classrooms	57

LIST OF FIGURES

Figure	Page
1 Poisson Process	4
2 Poisson Process Interarrival Times	6
3 Pseudo Code for Simulation	25
4 Number of Valid Simulations	26
5 Density of $\hat{\gamma}$ with $\alpha_0 = -1.4$ and $\gamma = .01$	27
6 Density of $\hat{\alpha}_0$ with $\alpha_0 = -1.4$ and $\gamma = .01$	28
7 The Effect of α_0 on the Distribution of $\hat{\gamma}$	29
8 Effect of γ on Distribution of $\hat{\alpha}_0$	31
9 Effect of α_0 on $Bias(\hat{\gamma})$	32
10 Effect of α_0 on $MSE(\hat{\gamma})$	32
11 Effect of γ on $Bias(\hat{\alpha}_0)$	33
12 Effect of γ on $MSE(\hat{\alpha}_0)$	34
13 $Var(\hat{\alpha}_0)$ vs. $Var(\hat{\gamma})$	35
14 $MSE(\hat{\alpha}_0)$ vs. $MSE(\hat{\gamma})$	35
15 $Bias(\hat{\alpha}_0)$ vs. $Bias(\hat{\gamma})$	36
16 Number of Valid Simulations	42
17 Density of $\hat{\gamma}$ with $\alpha_0 = -1.4$ and $\gamma = .01$	42
18 Density of $\hat{\alpha}_0$ with $\alpha_0 = -1.4$ and $\gamma = .01$	44
19 The Effect of α_0 on the Distribution of $\hat{\gamma}$	45
20 Effect of γ on Distribution of $\hat{\alpha}_0$	47
21 Effect of α_0 on $Bias(\hat{\gamma})$	48
22 Effect of α_0 on $MSE(\hat{\gamma})$	49

23	Effect of γ on $Bias(\hat{\alpha}_0)$	50
24	Effect of γ on $MSE(\hat{\alpha}_0)$	50
25	$MSE(\hat{\alpha}_0)$ vs. $MSE(\hat{\gamma})$	51
26	$Var(\hat{\alpha}_0)$ vs. $Var(\hat{\gamma})$	52
27	$Bias(\hat{\alpha}_0)$ vs. $Bias(\hat{\gamma})$	52
28	First Model (All teacher moves)	54
29	Second Model (Low teacher moves)	54
30	Third Model (High teacher moves)	55
31	All six models together	57
32	Expected Value of the Student Rate Function versus Time	59

I. INTRODUCTION

In this thesis, we will develop a statistical model that captures in-the-moment interaction between teacher and student in mathematics classrooms. We will model this interaction by considering the teacher moves as a homogeneous Poisson process and the student actions as a non-homogeneous Poisson process. Our model will assess the causal relationship between these two Poisson processes.

Capturing the Nature of the Interaction

Pedagogical, productive, and meaningful moves on the part of teachers enable students to elaborate their own ideas and engage with their peer's ideas. Moreover, meaningful teacher moves can help students draw conclusions or successfully solve problems in classroom activities. Focusing specifically on mathematics classrooms, productive teacher moves provide opportunities for students to engage in deep mathematical reasoning and discourse, which will lead to higher levels of engagement and achievement outcomes (Boaler and Staples, 2008). To understand how teachers' moves shape student engagement and achievement in mathematics classrooms, it is important to investigate their initial and subsequent moves and their effects on the nature of student and teacher interactions (Webb et al., 2014).

Capturing interactions between teacher and students in math classrooms is a crucial area in mathematics education. Researchers have recently developed a variety of instruments that capture particular aspects of these interactions (Jacobs and Spangler, 2017). For example, one instrument may measure teacher moves and students' responses separately, while another instrument captures the teacher role in discourse and the teacher reactions (Jacobs and Spangler, 2017). However, few instruments explore how teacher and student activity unfolds over time.

To address this need, Melhuish et al. (2020) have developed a well-constructed instrument that captures the interactions in the moment in mathematics classrooms, The Math Habits Tool. This instrument is used to collect data and define the events occurring in the moment about the nature of the students and teacher relationships in mathematics classrooms (Sorto et al., 2019). However, to make use of this type of data, there is a need to develop statistical models to explain the interactions in the moment in mathematics classrooms.

Why do we need a statistical model to capture the interactions between teacher moves and student actions? Previous studies have employed qualitative methods to investigate the processes of teacher-student interactions in mathematics education (Jacobs and Spangler, 2017). To identify these interactions, researchers typically follow certain steps. First, they observe classrooms over a period of time and identify teacher and student moves as data. To analyze these data, they examine interactions between teacher and students with a conceptual or instrumental framework, using codes on discrete instances or holistic rubrics. While qualitative analysis can serve to develop theories of how teacher and student activity unfolds, and score-based instruments can link overall teaching to student outcomes, there is a need to measure instruction at a fine-grained level to quantitatively test theorized relationships between teacher and student activity. Existing statistical models do not utilize timing precisely, such as whether particular teacher moves precede students' moves or how long after teacher makes a productive and meaningful move, students respond to it. Thus, current models do not adequately capture the interactive nature of teachers' and students' moves as a process in time. However, to test theories of teacher-student interactions, it is necessary to pay attention to the sequence and timing of the interactions (Jacobs and Spangler, 2017). The theories being proposed are about interactions that are happening over time, prior to the Math Habits Tool, but the current tools

available were not precise enough to measure the actual processes these theories purport to explain. Therefore, there is a need for a well-constructed model that captures in-the-moment interactions between teachers and students in mathematics classrooms.

The Math Habits Tool captures in-the-moment interactions between the teacher and students. As these two processes (teacher's moves and students' actions) occur concurrently over time, a model should capture the interactions while taking into account the time dimension. In our model, while the teacher's moves are considered homogeneous processes, the students' responses are considered non-homogeneous Poisson processes. This statistical model can therefore be used to examine interactions over the time of their duration.

Homogeneous Poisson Process

The Poisson process is used broadly in many research areas, such as health, engineering, and education. It is a counting process used to model the occurrences of an event within a specified amount of time. For example, the receipt of daily email from the administration or the times a user checks their social media account. The plot of the Poisson process (see Figure 1) is a step function with jumps of size 1 at each time when event is observed. Note that the jump size is constant but the times of the jumps are random. In the example shown in Figure 1, seven events are observed before time $t = 40$. In this section, we define and explain the Poisson process with its properties, which is crucial for future chapters which focus on simulating the homogeneous and non homogeneous Poisson processes and developing a model that assesses the relationship between them. After testing the model with simulated data, we will apply the model to real-world data. For further background on Poisson processes see *The Statistical Analysis of Series of Events* (Cox and Lewis, 1966), *Introduction to Stochastic Processes* (Cinlar, 1975), and

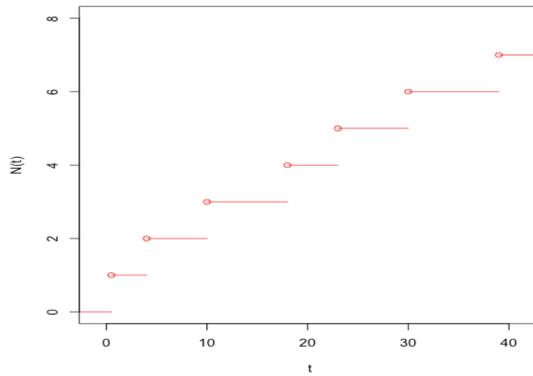


Figure 1: Poisson Process

Introduction to Stochastic Process with R (Dobrow, 2016).

In this counting process, let $N(t)$ denote the total number of arrivals over $[0, t]$, and λ be a constant to represent the rate of occurrences. $N(t, t + h)$ denotes total number of occurrences or events that have occurred in $[t, t + h]$, where $h > 0$.

Definition 1 (Poisson Process Dobrow (2016)). For $\lambda > 0$ and $t > 0$, a counting process is called a Poisson process if the following criteria are met

1. N is non-decreasing with $N(0) = 0$.
2. The rate of occurrences of an event λ is constant over $[0, t]$.
3. Two events cannot occur simultaneously. $P(N(t + h) - N(t) > 1) \rightarrow 0$ as $h \rightarrow 0$ for all $t > 0$.
4. Events in non-overlapping intervals are independent: $N(t, t+h)$ is independent of $N(t)$.
5. The process has stationary increments, that is, for every $t \geq 0$ and $h \geq 0$, the $N(t + h) - N(t)$ depends only on h , the length of the interval $[t, t + h]$.

To present several important properties of Poisson processes, let T_k be the time of the k^{th} event and $X_k = T_k - T_{k-1}$ be the *Interarrival* time between

the $(k - 1)^{th}$ and k^{th} event. See Figure 2 for a labeled sketch. In Theorem 1, we derive the distributions of $N(t)$, T_k and X_k . Of special importance, the interarrival times of the Poisson process follow an exponential distribution, and thus exhibit the memoryless property. The memoryless property is key to understanding the uniqueness and limitations of the Poisson process. To understand the implications of the memoryless properties, let us consider the following examples. Let N represent the number of boys in a family and suppose there are four boys in this family, and the parents are expecting their next child in three months. We know that the probability is 50% for the next child to be either a boy or girl. That their previous children are all boys does not affect the probability of the gender of next child. The probability of having a girl is independent (forgets) the past. For another example, consider placing an order for masks from an online store given the mask shortage in the world because of Covid-19. Let us also say there is an online store, and that it takes pre-orders for masks. Let us assume that we ordered some masks and waited for a month and did not receive anything and called the factory and asked them what the probability of receiving the masks in the next ten-day is. Someone from customer service informed us that it is 20%, then we have waited another ten days and did not receive anything again. We called the company and asked the same question. The customer service gave us the same answer. The probability of receiving the masks in the next ten days is 20% because it is exponential and it has memoryless property. This story is a good representation of the memoryless property.

Theorem 1 (Properties of Poisson Process). *Let $N(t)$ be a Poisson process with rate λ for $t > 0$, T_1, T_2, \dots be the event times then*

1. $N(t) \sim \text{Poisson}(\lambda t)$.
2. The interarrival times $X_k \sim \text{Exponential}(\frac{1}{\lambda})$.

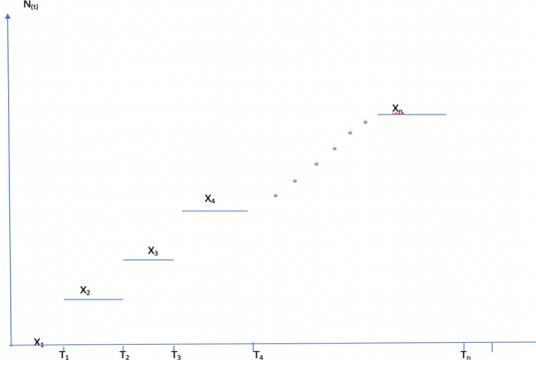


Figure 2: Poisson Process Interarrival Times

3. $T_k \sim \text{Erlang}(k, \frac{1}{\lambda})$.

4. $P(X_k > h + s | X_k > s) = P(X_k > h)$ for $h > 0$ and $s > 0$.

Proof. The main ideas of the proof of the properties are shown below, see Dobrow (2016) and Cox and Lewis (1966) for more details.

1. From the criteria of Definition 1 we can form a partition of $[0, t]$

$$0 < t_1 = \frac{t}{m} < \dots < t_k = \frac{kt}{m} < \dots < t_m = t$$

with a sufficiently fine mesh so that $N(t_k) - N(t_{k-1})$ are independent and approximately Bernoulli with probability of success $p = \frac{\lambda t}{m}$ and $N(t) = \sum_{k=1}^m [N(t_k) - N(t_{k-1})] \approx \text{Bin}(m, p)$. Since $mp = \lambda t$ is constant as $m \rightarrow \infty$, the result follows from the Poisson approximation to the Binomial Distribution: $N(t) = \sum_{k=1}^m [N(t_k) - N(t_{k-1})] \Rightarrow \text{Poisson}(\lambda t)$ as $m \rightarrow \infty$. Since Poisson variables are approximately normally distributed when the mean is large:

$$\frac{N(t) - \lambda t}{\sqrt{\lambda t}} \Rightarrow N(0, 1) \text{ as } \lambda t \rightarrow \infty$$

2. We begin with proof of the result for the first the interarrival time X_1 . Since $N(t) \sim \text{Poisson}(\lambda t)$ and $X_1 > x$ if and only if $N(x) = 0$, $P(X_1 >$

$x) = P(N(x) = 0) = \frac{e^{-\lambda x}(\lambda x)^0}{0!} = e^{-\lambda x}$ which recognize as the survival function of the exponential distribution with mean $\frac{1}{\lambda}$. By condition 4 and the linearity property of independent Poisson variables, $N(t) - N(t_k)|T_k = t_k \sim \text{Poisson}(\lambda(t - t_k))$. Hence, $P(X_k > x) = P(N(t_k + x) - N(t_k) = 0|T_k = t_k) = e^{-\lambda x}$.

3. By condition 5, X_k are independent. $T_k = \sum_{j=1}^k X_j$ is the sum of independent exponential random variables and hence, follows an Erlang distribution.
4. Since the $X_t \sim \text{Exponential}(\frac{1}{\lambda})$,

$$P(X_k > h + s | X_k > s) = \frac{e^{-\lambda(h+s)} - e^{-\lambda(h+s)}}{e^{-\lambda s}} = e^{-\lambda h} = P(X_k > h).$$

□

Another key property, presented in Theorem 3, describes the relationship between the uniform order statistics and inter arrival times. This property provides an algorithm to simulate Poisson processes and compute the likelihood given an observed process. To provide intuition for the property consider the simple case with only 1 event observed in the interval $[0, t]$.

Example 2. Suppose $N(t) = 1$ and investigate the distribution of the arrival time T_1 . For every $0 \leq s \leq t$,

$$\begin{aligned} P[T_1 \leq s | N(t) = 1] &= \frac{P[T_1 \leq s | N(t) = 1]}{P[N(t) = 1]} \\ &= \frac{P[N(s) = 1, N(t) = 1]}{P[N(t) = 1]} \\ &= \frac{P[N(s) = 1, N(t) - N(s) = 0]}{P[N(t) = 1]} \\ &= \frac{P[N(s) = 1] P[N(t - s) = 0]}{P[N(t) = 1]} \\ &= \frac{\exp(-\lambda s) \lambda s \exp(-\lambda (t - s))}{\exp(-\lambda t) \lambda t} = \frac{s}{t} \end{aligned} \tag{1}$$

which is the cumulative distribution of the uniform distribution over $[0, t]$. Hence, $T_1 \sim \text{Uniform}(0, t)$.

To generalize to the case with more observed events we need to consider uniform order statistics.

Definition 2 (Order Statistics). Let Y_1, \dots, Y_n be a sample of random variables. The *order statistics* are the collection of the ordered variables

$$Y_{(1)} \leq Y_{(2)} \leq \dots \leq Y_{(n)}$$

Note that $Y_{(1)}$ is the minimum and $Y_{(n)}$ is the maximum.

An important special case of the order statistics occurs when the variables are i.i.d. $\text{Uniform}(0, t)$, which we denote

$$U_{(1)} < U_{(2)} < \dots < U_{(n)}.$$

Note that since the uniform distribution is continuous, the probability of ties is 0. Further, it can be shown that the joint density function of the uniform order statistics is

$$f_{U_{(1)}, \dots, U_{(n)}}(u_1, \dots, u_n) = \frac{n!}{t^n}$$

Dobrow (2016).

Using an extension of the argument in Example 2, Dobrow (2016) proves the next theorem.

Theorem 3. For any Poisson process, conditional on the event $N(t) = n$, the joint distribution of the n arrival times t_1, \dots, t_n is the same as the joint distribution of $U_{(1)}, U_{(2)}, \dots, U_{(n)}$, the order statistics of n i.i.d. Uniform(0, t) r.v.s. is given by

$$f_{(s_1, \dots, s_n)} = \frac{n!}{t^n}, 0 < s_1 < s_2 < s_3 < \dots < s_n < t$$

Non-Homogeneous Poisson Process

A counting process $N(t)$ is a non-homogenous Poisson process with the rate function $\lambda(t)$, $t > 0$ if:

- $N(0) = 0$, and if $h < t$, then $N(h) \leq N(t)$ for every h and $t \geq 0$.
- This process has independent increments.
- The random variable of $N_{t+h} - N_t$ is independent of any random variable that occurred before time t where t and $h \geq 0$. That means future events are independent of the past history.
- The rate is not constant and it is a function of time for this process. Let the number of events in the time of interval $[0, t]$, these events are Poisson distributed with the expected value of

$$E(N_t) = \Lambda(t)$$

where $\Lambda(t)$ is the cumulative rate function and it is a non-decreasing right-continuous function. It represents the expected number of events until time t . Hence,

$$\Lambda(t) = \int_0^t \lambda(t) dt$$

Cinlar (1975).

Pasupathy (2011) includes Theorem 4, the non homogeneous analog of Theorem 3. In Chapter 2, Theorems 3 and 4 are used to model the Student and Teacher Processes, while in Chapter 3 they are used to simulate the process and study the estimates. In addition to given theorems,

Theorem 4 (Pasupathy (2011)). *Let $\Lambda(t), t \geq 0$ be a positive, non decreasing continuous function, and the random variables $T_1, T_2, T_3, \dots, T_n$ are event times that are non homogeneous Poisson process with expectation function $\Lambda(t)$, and let N_t represent the number of events that are occurring before time t so $t \in [0, t_0]$. Then, conditional on the number of events $N_{t_0} = n$, the event times $T_1, T_2, T_3, \dots, T_n$ are distributed as order statistics of a sample from distribution function $F(t) = \frac{\Lambda(t)}{\Lambda(t_0)}$.*

II. SINGLE CLASSROOM MODEL

In this chapter, we describe the simple model used for two stochastic processes, the *Teacher Process* and the *Student Process*, and capture the relationship between the processes in the parameter γ . Time is continuous and outcomes are discrete in the Teacher Process, and it is a counting process. To make the problem tractable, the Teacher Process is modeled with a simple homogeneous Poisson process. Since the rate of the Student Process depends on the teacher moves, the Student Process is modeled with a non-homogeneous Poisson process. As a result, the Teacher Process is a homogeneous Poisson process with rate μ , while the Student Process is a non-homogeneous Poisson Process whose rate is stochastic and a function of the Teacher Process. The maximum likelihood estimate (m.l.e.) and its asymptotic distribution are derived as described in Hogg et al. (2005).

Teacher Process

The Teacher Process is the driver of the two processes and follows a homogenous Poisson process with rate μ over the interval $[0, T]$:

$$V_t \sim \text{PoissonProcess}(\mu, [0, T]).$$

Using Theorem 3, we can represent the process as,

$$V_t = \{M, U_{(1)}, \dots, U_{(M)}\}$$

where $M \sim \text{Poisson}(\mu T)$, $U_1, \dots, U_m | M = m \sim \text{Uniform}[0, T]$, and $U_{(1)}, \dots, U_{(m)}$ are the corresponding order statistics. The likelihood is

$$L(\mu | V_t) = \frac{e^{-\mu T} (\mu T)^M}{M!} \cdot M! T^{-M} = e^{-\mu T} (\mu)^M.$$

To find the m.l.e, take the log, take the derivative, set to zero, and solve.

$$l = \log(L) = -\mu T + M \log(\mu)$$

and

$$l' = -T + \frac{M}{\mu} = 0.$$

Solving for μ yields the m.l.e., $\hat{\mu} = \frac{M}{T}$.

Student Process

The Student Process Y_t is a non homogeneous Poisson process whose rate λ_t is a function of the Teacher Process:

$$\log(\lambda_t) = \alpha_0 + \gamma V_t,$$

where $-\infty < \alpha_0 < \infty$, $\exp(\alpha_0)$ is the base rate of occurrences and $-\infty < \gamma < \infty$ is the catalytic constant. Hence, when $\gamma > 0$ the rate of future occurrences in the Student Process increases after each observed event in the Teacher Process. Use Theorem 4 to represent the Student Process. Let

$$\Lambda_t = \int_0^t \lambda_s ds = \int_0^t \exp(\alpha_0 + \gamma V_s) ds = \exp(\alpha_0) \int_0^t \exp(\gamma V_s) ds.$$

Note that λ_t is a step function, hence, Λ_t is a piecewise continuous linear function with change points at each occurrence, u_j , of the Teacher Process. For convenience of notation, let $u_{(0)} = 0$ and $u_{(M+1)} = T$ represent the endpoints of the interval $[0, T]$. The value of Λ_t at each of the change points is given by

$$w_j = \sum_{k=1}^j \exp(\gamma(k-1)) (u_{(k)} - u_{(k-1)}), j = 1, \dots, M+1$$

with $w_0 = 0$. To find the formula of Λ_t throughout the interval $[0, T]$, define $k_t^* = \max\{k : u_{(k)} \leq t\}$ so that $u_{(k_t^*)}$ is the time of the last jump prior to time t . Then, noting that Λ_t is linear between each change point, we see that

$$\Lambda_t = \exp(\alpha_0) (w_k + \exp(\gamma k)(t - u_{(k)})) \text{ when } u_{(k)} \leq t \leq u_{(k+1)}.$$

To use results from Chapter 1, define a c.d.f. by normalizing the increasing function Λ_t ,

$$F(t) = \frac{\Lambda_t}{\Lambda_T} = w_{M+1}^{-1} (w_k + \exp(\gamma k)(t - u_{(k)})) \text{ when } u_{(k)} \leq t \leq u_{(k+1)}.$$

The quantile function (inverse of c.d.f.) is also piecewise linear and is given by

$$Q(y) = \frac{w_{M+1}y - w_k}{\exp(\gamma k)} + u_{(k)} \text{ when } \frac{w_k}{w_{M+1}} \leq y \leq \frac{w_{k+1}}{w_{M+1}}.$$

Conditional on the Teacher Process, the Student Process follows a nonhomogeneous Poisson Process, $Y_t|V_t \sim \text{PoissonProcess}(\lambda_t, [0, T])$ which we represent as Refer to Chapter 1 result, $Y_t = \{N, X_{(1)}, \dots, X_{(N)}\}$ where $N \sim \text{Poisson}(\Lambda_T = e^{\alpha_0} w_{M+1})$ and

$$0 < X_{(1)} < \dots < X_{(N)} < T$$

are the order statistics from a sample of size N with c.d.f. F . We can generate the sample by first taking a sample of uniform random variables from the unit interval and applying the quantile function Q .

The likelihood function for the μ, α_0 and γ is given by

$$\begin{aligned}
L(\mu, \alpha_0, \gamma | V_t, Y_t) &= L(\mu | V_t) L(\alpha_0, \gamma | V_t, Y_t) \\
&= e^{-\mu T} (\mu)^M \frac{e^{-w_{M+1} \exp(\alpha_0)} (\exp(\alpha_0) w_{M+1})^N}{N!} \\
&\quad N! \frac{e^{\gamma k_{X(1)}^*} \dots e^{\gamma k_{X(N)}^*}}{w_{M+1}^N} \\
&= e^{-\mu T} (\mu)^M e^{-w_{M+1} \exp(\alpha_0)} e^{N\alpha_0} e^{\gamma \sum_{j=1}^N k_{X(j)}^*}
\end{aligned}$$

So the log likelihood is given by

$$l = -\mu T + M \log(\mu) - w_{M+1} \exp(\alpha_0) + N\alpha_0 + \gamma \sum_{j=1}^N k_{X(j)}^*. \quad (2)$$

Taking the derivative with respect to μ ,

$$\frac{\partial l}{\partial \mu} = -T + \frac{M}{\mu}, \quad (3)$$

we find that the m.l.e., $\hat{\mu} = \frac{M}{T}$, is the same as when considering the Teacher Process alone. To find the estimates of α_0 and γ , recall that $w_{M+1} = \sum_{k=1}^{M+1} e^{\gamma(k-1)} (u_{(k)} - u_{(k-1)})$. From the derivatives,

$$\frac{\partial l}{\partial \gamma} = -e^{\alpha_0} \sum_{k=1}^{M+1} (k-1) e^{\gamma(k-1)} (u_{(k)} - u_{(k-1)}) + \sum_{j=1}^N k_{X(j)}^* \quad (4)$$

and

$$\frac{\partial l}{\partial \alpha_0} = -e^{\alpha_0} \sum_{k=1}^{M+1} e^{\gamma(k-1)} (u_{(k)} - u_{(k-1)}) + N, \quad (5)$$

we see that the sufficient statistics are $M, N, u_{(1)}, \dots, u_{(M)}, \sum_{j=1}^N k_{X(j)}^*$. Due to the highly nonlinear form of (4) and (5), the likelihood equations have no closed form solutions. In Chapter 3, the `multroot` from the R package `rootsolve` will provide numerical estimates of α_0 and γ .

Asymptotic Distribution

Under general regularity conditions, m.l.e.'s are consistent and asymptotically normally distributed (Hogg et al., 2005). Furthermore, the asymptotic variance can be computed from the Fisher Information which is the negative of the expected value of the Hessian of the log likelihood (2).

Theorem 5. *The Fisher Information contained in (V_t, Y_t) is equal to*

$$I = I(\mu, \lambda, \gamma) = - \begin{bmatrix} E \left[\frac{\partial^2 l}{\partial \mu^2} \right] & 0 & 0 \\ 0 & E \left[\frac{\partial^2 l}{\partial \alpha_0^2} \right] & E \left[\frac{\partial^2 l}{\partial \gamma \partial \alpha_0} \right] \\ 0 & E \left[\frac{\partial^2 l}{\partial \gamma \partial \alpha_0} \right] & E \left[\frac{\partial^2 l}{\partial \gamma^2} \right] \end{bmatrix}$$

where

$$E \left[\frac{\partial^2 l}{\partial \mu^2} \right] = -\frac{T}{\mu} \quad (6)$$

$$E \left[\frac{\partial^2 l}{\partial \alpha_0^2} \right] = -\frac{T e^{\alpha_0}}{\lambda(1-e^\gamma)} (1 - e^{\lambda(e^\gamma-1)}) \quad (7)$$

$$E \left[\frac{\partial^2 l}{\partial \gamma \partial \alpha_0} \right] = -\frac{T e^{\alpha_0+\gamma}}{(1-e^\gamma)^2} \left(\frac{1-e^{-\lambda}}{\lambda} - e^{\lambda(e^\gamma-1)} + \frac{e^{-\lambda}}{\lambda} (e^{e^\gamma \lambda} (e^\gamma \lambda - 1) + 1) \right) \quad (8)$$

$$\begin{aligned} E \left[\frac{\partial^2 l}{\partial \gamma^2} \right] &= -\frac{T e^{\alpha_0+\gamma}}{(1-e^\gamma)^3} \left((1+e^\gamma) \frac{(1-e^{-\lambda})}{\lambda} - \lambda e^{\gamma+\lambda(e^\gamma-1)} - e^{\lambda(e^\gamma-1)} \right. \\ &\quad \left. + (2-e^\gamma) \frac{e^{-\lambda}}{\lambda} (e^{e^\gamma \lambda} (1 - e^\gamma \lambda + e^{2\gamma} \lambda^2) - 1) \right. \\ &\quad \left. + 2 \frac{e^{-\lambda}}{\lambda} (e^{e^\gamma \lambda} (e^\gamma \lambda - 1) + 1) - \frac{e^{-\lambda}}{\lambda} (e^{e^\gamma \lambda} - 1) \right) \end{aligned} \quad (9)$$

and $\lambda = \mu T$.

The proof of Theorem 5 requires lengthy calculations involving exponential sums (Boyadzhiev, 2009), Touchard Polynomials (Chrysaphinou, 1985) and

low-order polylogarithms (Chrysaphinou, 1985). To ease exposition, the proof is broken into several lemmas.

Lemma 6 (*p* Summation Formulas). *For p in N, the p summation formula is given by*

$$\sum_{m=0}^{\infty} m^p \frac{z^{m+1}}{(m+1)!} = \sum_{k=0}^p (-1)^{p-k} \binom{p}{k} e^z T_k(z)$$

where $T_k(z) = \sum_{j=0}^k \binom{k}{j} z^j$ is the Touchard Polynomial of order k .

The values of the summation formula for $p = 0, 1, 2$ are shown in Table 1.

Proof. The results follows from a change of variable $n = m + 1$, the binomial theorem and interchanging the order of summation.

$$\begin{aligned} \sum_{m=0}^{\infty} m^p \frac{z^{m+1}}{(m+1)!} &= \sum_{n=1}^{\infty} (n-1)^p \frac{z^n}{n!} = \sum_{n=1}^{\infty} \sum_{k=0}^p n^k (-1)^{p-k} \binom{p}{k} \frac{z^n}{n!} \\ &= \sum_{k=0}^p (-1)^{p-k} \binom{p}{k} \sum_{n=1}^{\infty} n^k \frac{z^n}{n!} \\ &= \sum_{k=0}^p (-1)^{p-k} \binom{p}{k} \left(\sum_{n=0}^{\infty} n^k \frac{z^n}{n!} \right) \\ &= \sum_{k=0}^p (-1)^{p-k} \binom{p}{k} e^z T_k(z) \end{aligned}$$

where the final equality follows from the relationship between the exponential function and Touchard Polynomials (Boyadzhiev, 2009):

$$\sum_{n=0}^{\infty} n^k \frac{z^n}{n!} = e^z T_k(z). \quad (10)$$

□

The following lemma includes needed results for expected values of functions of Poisson random variables.

Table 1: p Summation Formulas for small values of p

p	$\sum_{m=0}^{\infty} m^p \frac{z^{m+1}}{(m+1)!}$
0	$e^z - 1$
1	$e^z(z - 1) + 1$
2	$e^z(1 - z + z^2) - 1$

Lemma 7. *If $M \sim \text{Poisson}(\lambda)$, then*

$$E\left[\frac{1}{M+1}\right] = \frac{1 - e^{-\lambda}}{\lambda} \quad (11)$$

$$E\left[\frac{z^{M+1}}{M+1}\right] = \frac{e^{-\lambda}}{\lambda}(e^{e^{\gamma}\lambda} - 1) \quad (12)$$

$$E[z^M] = e^{-\lambda}e^{e^{\gamma}\lambda} \quad (13)$$

$$E\left[M\frac{z^{(M+1)}}{M+1}\right] = \frac{e^{-\lambda}}{\lambda}((e^{e^{\gamma}\lambda}(e^{\gamma}\lambda - 1) + 1)) \quad (14)$$

$$E[Mz^M] = \lambda e^{\gamma-\lambda+e^{\gamma}\lambda} \quad (15)$$

$$E\left[M^2\frac{z^{(M+1)}}{M+1}\right] = \frac{e^{-\lambda}}{\lambda}(e^{e^{\gamma}\lambda}(1 - e^{\gamma}\lambda + (e^{\gamma}\lambda)^2) - 1) \quad (16)$$

Proof. The probability mass function for M is $f(m) = \frac{e^{-\lambda}(\lambda)^m}{m!}$ for $m = 0, 1, 2, \dots$

The expected values are computed by manipulating the series using formulas from Table 1 and (10).

Equation 11: Using a change of variable and the series representation of e^{λ} we see that

$$E\left[\frac{1}{M+1}\right] = \sum_{m=0}^{\infty} \frac{1}{m+1} \frac{e^{-\lambda}(\lambda)^m}{m!} = \frac{e^{-\lambda}}{\lambda} \sum_{m=0}^{\infty} \frac{(\lambda)^{m+1}}{(m+1)!}.$$

Change the index variable, $m + 1 = n$, and then

$$\mathbb{E} \left[\frac{1}{M+1} \right] = \frac{e^{-\lambda}}{\lambda} \sum_{n=1}^{\infty} \frac{(\lambda)^n}{n!} = \frac{e^{-\lambda}}{\lambda} \left(\sum_{n=0}^{\infty} \frac{(\lambda)^n}{n!} - 1 \right).$$

Recognizing the series representation of e^λ provides the result.

Equation 12: Rewriting the series to match the p summation formula yields

$$\mathbb{E} \left[\frac{z^{(M+1)}}{M+1} \right] = \sum_{m=0}^{\infty} \frac{z^{m+1}}{m+1} \frac{e^{-\lambda}(\lambda)^m}{m!} = \frac{e^{-\lambda}}{\lambda} \sum_{m=0}^{\infty} \frac{(z\lambda)^{m+1}}{(m+1)!}.$$

The result follows from the p- cases formula in Table 1 with $p = 0$ and $z = z\lambda$.

Equation 13: Using the series representation of the exponential function $e^{z\lambda}$ we see that

$$\mathbb{E} [z^M] = \sum_{m=0}^{\infty} z^m \frac{e^{-\lambda}(\lambda)^m}{m!} = e^{-\lambda} \sum_{m=0}^{\infty} \frac{(z\lambda)^m}{m!} = e^{-\lambda} e^{z\lambda}.$$

Equation 14: Again we can rewrite the series to match a p summation formula.

$$\mathbb{E} \left[M \frac{z^{(M+1)}}{M+1} \right] = \sum_{m=0}^{\infty} \frac{mz^{m+1}}{m+1} \frac{e^{-\lambda}(\lambda)^m}{m!} = \frac{e^{-\lambda}}{\lambda} \sum_{m=0}^{\infty} \frac{m(z\lambda)^{m+1}}{(m+1)!}.$$

The result follows from the p- cases formula in Table 1 with $p = 1$ and $z = z\lambda$.

Equation 15:

$$E(Mz^M) = \sum_{m=0}^{\infty} mz^m \frac{e^{-\lambda}(\lambda)^m}{m!} = e^{-\lambda} \sum_{m=0}^{\infty} \frac{m(z\lambda)^m}{m!}.$$

The result follows from (10) with $k = 1$.

Equation 16:

$$\mathbb{E} \left[M^2 \frac{z^{(M+1)}}{M+1} \right] = \sum_{m=0}^{\infty} \frac{m^2 z^{m+1}}{m+1} \frac{e^{-\lambda} (\lambda)^m}{m!} = \frac{e^{-\lambda}}{\lambda} \sum_{m=0}^{\infty} \frac{m^2 (z\lambda)^{m+1}}{(m+1)!}.$$

The result follows from the p- cases formula in Table 1 with $p = 2$ and $z = z\lambda$.

□

We are now ready to prove Theorem 5.

Proof. Starting with the first derivatives (3), (4) and (5), the second derivatives of log likelihood function are given by

$$\frac{\partial^2 l}{\partial \mu^2} = -\frac{M}{\mu^2} \quad (17)$$

$$\frac{\partial^2 l}{\partial \mu \gamma} = \frac{\partial^2 l}{\partial \mu \alpha_0} = 0 \quad (18)$$

$$\frac{\partial^2 l}{\partial \alpha_0^2} = -e^{\alpha_0} \sum_{k=1}^{M+1} e^{\gamma(k-1)} (u_{(k)} - u_{(k-1)}) \quad (19)$$

$$\frac{\partial^2 l}{\partial \gamma^2} = -e^{\alpha_0} \sum_{k=1}^{M+1} (k-1)(k-1) e^{\gamma(k-1)} (u_{(k)} - u_{(k-1)}) \quad (20)$$

$$\frac{\partial^2 l}{\partial \gamma \partial \alpha_0} = -e^{\alpha_0} \sum_{k=1}^{M+1} (k-1) e^{\gamma(k-1)} (u_{(k)} - u_{(k-1)}) \quad (21)$$

Begin by noting that the zero elements of the information matrix follow from (18),

i.e. $\mathbf{I}_{1,2} = \mathbf{I}_{1,3} = \mathbf{I}_{2,1} = \mathbf{I}_{3,1} = 0$.

We now derive the other components of the information matrix.

Equation 6: Since $M \sim \text{Poisson}(\mu T)$, $\mathbb{E}[M] = \mu T$ and $\mathbb{E} \left[\frac{\partial^2 l}{\partial \mu^2} \right] = -\frac{T}{\mu}$.

To find the expected values of (19), (20), and (21), we first condition on M , the number of occurrences in the Teacher Process, noting that

$$\mathbb{E}[(u_{(k)} - u_{(k-1)}) | M] = \frac{T}{M+1}.$$

Then we manipulate resulting series using the results from Lemmas 7 and Table 1.

Equation 7: Using a change of variable, $n = k - 1$, and recognizing the geometric series, we see that

$$\begin{aligned} \mathbb{E} \left[\frac{\partial^2 l}{\partial \alpha_0^2} \right] &= \mathbb{E} \left[-\frac{e^{\alpha_0 T}}{M+1} \sum_{k=1}^{M+1} z^{(k-1)} \right] \\ &= \mathbb{E} \left[-\frac{e^{\alpha_0 T}}{M+1} \sum_{n=0}^M z^n \right] \\ &= \mathbb{E} \left[\frac{-e^{\alpha_0 T}}{M+1} \left(\frac{1 - z^{M+1}}{1 - z} \right) \right] \\ &= \frac{-e^{\alpha_0 T}}{1 - z} \mathbb{E} \left[\frac{1 - z^{M+1}}{M+1} \right] \\ &= \frac{-e^{\alpha_0 T}}{1 - z} \left(\mathbb{E} \left[\frac{1}{M+1} \right] - \mathbb{E} \left[\frac{z^{M+1}}{M+1} \right] \right). \end{aligned}$$

Substituting (11) and (12) for expectations above gives

$$\begin{aligned} \mathbb{E} \left[\frac{\partial^2 l}{\partial \alpha_0^2} \right] &= \frac{-e^{\alpha_0 T}}{1 - z} \left(\frac{1 - e^{-\lambda}}{\lambda} - \frac{e^{-\lambda}}{\lambda} (e^{z\lambda} - 1) \right) \\ &= \frac{-e^{\alpha_0 T}}{1 - z} \left(\frac{1 - e^{-\lambda}}{\lambda} - \frac{e^{-\lambda}}{\lambda} (e^{e^\gamma \lambda} - 1) \right) \\ &= \frac{-e^{\alpha_0 T}}{(1 - z) \lambda} (1 - e^{-\lambda} - e^{-\lambda} e^{e^\gamma \lambda} + e^{-\lambda}) \\ &= \frac{-e^{\alpha_0 T}}{(1 - z) \lambda} (1 - e^{-\lambda} e^{e^\gamma \lambda}) \\ &= \frac{-e^{\alpha_0 T}}{(1 - z) \lambda} (1 - e^{\lambda(e^\gamma - 1)}). \end{aligned}$$

Equation 7 follows by substituting $e^\gamma = z$.

Equation 8: Conditioning on M and applying a change of variable, $n = k - 1$ yields

$$\begin{aligned}
\mathbb{E} \left[\frac{\partial^2 l}{\partial \alpha_0 \partial \gamma} \right] &= \mathbb{E} \left[-e^{\alpha_0} \sum_{k=1}^{M+1} (k-1) z^{k-1} \frac{T}{M+1} \right] \\
&= -e^{\alpha_0} T \mathbb{E} \left[\frac{1}{M+1} \sum_{k=1}^{M+1} (k-1) z^{k-1} \right] \\
&= -e^{\alpha_0} T \mathbb{E} \left[\frac{1}{M+1} \sum_{n=1}^M n z^n \right] \\
&= -e^{\alpha_0} T \mathbb{E} \left[\frac{1}{M+1} z \frac{1 - (M+1)z^M + Mz^{M+1}}{(1-z)^2} \right]
\end{aligned}$$

where the last equality follows from the low-order poly logarithm (Weisstein, 2002)

$$\sum_{k=1}^n k z^k = z \frac{1 - (n+1)z^n + n z^{n+1}}{(1-z)^2}.$$

Algebraic manipulation gives

$$\begin{aligned}
\mathbb{E} \left[\frac{\partial^2 l}{\partial \alpha_0 \partial \gamma} \right] &= -\frac{e^{\alpha_0} T z}{(1-z)^2} \mathbb{E} \left[\frac{1}{M+1} - \frac{z^M (M+1)}{M+1} + \frac{M z^{M+1}}{M+1} \right] \\
&= -\frac{e^{\alpha_0} T z}{(1-z)^2} \left(\mathbb{E} \left[\frac{1}{M+1} \right] - \mathbb{E} [z^M] + \mathbb{E} \left[\frac{M z^{M+1}}{M+1} \right] \right).
\end{aligned}$$

Substituting (11), (13) and (14) for expectations above gives

$$\mathbb{E} \left[\frac{\partial^2 l}{\partial \alpha_0 \partial \gamma} \right] = -\frac{e^{\alpha_0} T z}{(1-z)^2} \left(\frac{1 - e^{-\lambda}}{\lambda} - e^{-\lambda} e^{z\lambda} + \frac{e^{-\lambda}}{\lambda} (e^{z\lambda} (z\lambda - 1) + 1) \right).$$

Equation 8 follows by substituting $e^\gamma = z$.

Equation 9:

$$\begin{aligned} \mathbb{E} \left[\frac{\partial^2 l}{\partial \gamma^2} \right] &= \mathbb{E} \left[-\frac{e^{\alpha_0 T}}{M+1} \sum_{k=1}^{M+1} (k-1)^2 e^{\gamma(k-1)} \right] \\ &= \mathbb{E} \left[\frac{-e^{\alpha_0 T}}{M+1} \sum_{n=0}^M n^2 z^n \right]. \end{aligned}$$

Using the low-order poly logarithm result (Sykora, 2006; Weisstein, 2002)

$$\sum_{k=1}^n k^2 z^k = z \frac{1+z-(n+1)^2 z^n + (2n^2+2n-1)z^{n+1} - n^2 z^{n+2}}{(1-z)^3}$$

we can rewrite the expected value as

$$\begin{aligned} \mathbb{E} \left[\frac{\partial^2 l}{\partial \gamma^2} \right] &= -\frac{e^{\alpha_0 T} z}{(1-z)^3} \mathbb{E} \left[\frac{1}{M+1} + \frac{z}{M+1} - \frac{(M+1)^2 z^M}{M+1} \right. \\ &\quad \left. + \frac{(2M^2+2M-1)z^{M+1}}{M+1} - \frac{M^2 z^{M+2}}{M+1} \right] \\ &= -\frac{e^{\alpha_0 T} z}{(1-z)^3} \left((1+z) \mathbb{E} \left[\frac{1}{M+1} \right] - \mathbb{E} [(M+1)z^M] \right. \\ &\quad \left. + \mathbb{E} \left[\frac{(2M^2+2M-1)z^{M+1}}{M+1} \right] - \mathbb{E} \left[\frac{M^2 z^{M+2}}{M+1} \right] \right) \\ &= -\frac{e^{\alpha_0 T} z}{(1-z)^3} \left(\mathbb{E} \left[(1+z) \frac{1}{M+1} \right] - \mathbb{E} [Mz^M] - \mathbb{E} [z^M] \right. \\ &\quad \left. + 2\mathbb{E} \left[\frac{M^2 z^{M+1}}{M+1} \right] + \mathbb{E} \left[\frac{Mz^{M+1}}{M+1} \right] - \mathbb{E} \left[\frac{z^{M+1}}{M+1} \right] \right. \\ &\quad \left. - z\mathbb{E} \left[\frac{M^2 z^{M+1}}{M+1} \right] \right). \end{aligned}$$

Substituting the results from Lemma 7 yields

$$\begin{aligned} \mathbb{E} \left[\frac{\partial^2 l}{\partial \gamma^2} \right] &= \frac{-e^{\alpha_0} T z}{(1-z)^3} \left((1+z) \frac{(1-e^{-\lambda})}{\lambda} - (e^{-\lambda}(z\lambda)e^{z\lambda}) - (e^{-\lambda}e^{z\lambda}) \right. \\ &\quad \left. + (2-z) \frac{e^{-\lambda}}{\lambda} (e^{z\lambda}(1-e^{\gamma\lambda} + (e^{\gamma\lambda})^2) - 1) \right. \\ &\quad \left. + 2 \frac{e^{-\lambda}}{\lambda} (e^{z\lambda}(z\lambda - 1) + 1) - \frac{e^{-\lambda}}{\lambda} (e^{z\lambda} - 1) \right). \end{aligned}$$

Equation 9 follows by substituting $e^\gamma = z$.

□

Now that we have derived the Fisher Information, maximum likelihood theory provides the asymptotic (as $\mu T \rightarrow \infty$) distribution

$$\hat{\boldsymbol{\theta}} = (\hat{\mu}, \hat{\alpha}_0, \hat{\gamma})' \sim N(\boldsymbol{\theta}, \mathbf{I}(\boldsymbol{\theta})^{-1})$$

where $\boldsymbol{\theta} = (\mu, \alpha_0, \gamma)'$. In applications, we substitute $\hat{\boldsymbol{\theta}}$ in $\mathbf{I}(\boldsymbol{\theta})$ to estimate the variance. However, it is not clear how large T must be in order for the normal approximation to be useful. This issue is investigated via a simulation study in Chapter 3.

III. SIMULATION STUDY

In Chapter 2, we derived the m.l.e. and its asymptotic distribution for the parameters of the paired Teacher and Student Processes. We now describe a simulation study programmed in R to investigate the small sample (i.e. small T) distribution. In the Analysis section, the results are analyzed and compared with the theoretical asymptotic distribution.

Simulation Description

Implementing the results from Chapter 2 in R, the order statistic representation of Poisson processes was used to simulate the paired Teacher and Student Processes for given choices of T , μ , α_0 and γ . The function `multiroot` from the R package `rootsolve` was used to compute numerical zeroes of score functions (4) and (5) and find estimates of $\hat{\alpha}_0$ and $\hat{\gamma}$. For each choice of the parameters, 50 processes were simulated and the following summary statistics were computed:

$$MSE(\hat{\theta}) = 50^{-1} \sum_{i=1}^{50} (\hat{\theta}_i - \theta_i)^2 \quad (22)$$

$$Bias(\hat{\theta}) = 50^{-1} \sum_{i=1}^{50} (\hat{\theta}_i - \theta_i) \quad (23)$$

$$Var(\hat{\theta}) = MSE(\hat{\theta}) - Bias(\hat{\theta})^2 \quad (24)$$

where θ is either α_0 or γ . When we simulated the data, we used $T=50$ and $\mu = 0.25$ because these values were suitable for the real classroom environment and they were based on the data that's used in the application section. The values from -3 to 1 increasing by 0.01 used for α_0 and the values from 0 to 0.4 increasing by 0.1 used for γ in the simulation. The complete R code is included in Appendix A and a summary pseudocode in Figure 3.

```

for (i in 1:50) do-,
   $M \leftarrow \text{rpois}(\mu)$ 
   $V \leftarrow \text{sort}(\text{runif}(M, \text{min} = 0, \text{max} = T))$  ▷ Teacher Process
  Compute  $w_j$  defined in Chapter 2
  Compute  $F_t$  defined in Chapter 2
   $N \leftarrow \text{rpois}(1, w_{M+1})$ 
   $Y \leftarrow \text{sort}(F_t^{-1}(\text{runif}(N)))$  ▷ Student Process
  Compute sufficient statistics  $u_{(1)}, \dots, u_{(M)}, \sum_{j=1}^M k_{X(j)}^*$ 
  Compute score function from Equations (4) and (5)
   $(\hat{\alpha}_0, \hat{\gamma})_i \leftarrow \text{multroot}(\text{score})$  ▷ m.l.e. estimates

   $\text{Bias}_{\hat{\gamma}} \leftarrow \text{mean}(\hat{\gamma} - \gamma)$ 
   $\text{Bias}_{\hat{\alpha}_0} \leftarrow \text{mean}(\hat{\alpha}_0 - \alpha_0)$ 
   $\text{MSE}_{\hat{\gamma}} \leftarrow \text{mean}((\hat{\gamma} - \gamma)^2)$ 
   $\text{MSE}_{\hat{\alpha}_0} \leftarrow \text{mean}((\hat{\alpha}_0 - \alpha_0)^2)$ 
  Compute Information,  $I$  from Theorem 5
   $\text{var} \leftarrow \text{solve}(I) = 0$ 

```

Figure 3: Pseudo Code for Simulation

Table 2 shows the first 5 rows of the results of the simulation.

Table 2: First 5 Rows of Results

gamma	alpha0	cnt	g.bias	a.bias	g.mse	a.mse	95%CI($\hat{\alpha}_0$)	95% CI($\hat{\gamma}$)
0	-3.0	46	-0.36	-0.04	4.97	2.42	(-6.10,0.00)	(-4.67,3.95)
0	-2.9	45	-0.18	-1.34	2.52	15.11	(-11.39,2.90)	(-3.27,2.90)
0	-2.8	49	-0.02	-0.03	0.02	1.43	(-5.17,-0.48)	(-0.32,0.26)
0	-2.7	49	-0.01	-0.30	0.02	3.24	(-6.48,0.47)	(-0.30,0.27)
0	-2.6	49	-0.01	0.04	0.00	1.22	(-4.72,-0.39)	(-0.17,0.13)

If the simulation results in no observed events for either process, no estimate of the parameters is possible. Figure 4 illustrates the values of (α_0, γ) that resulted in fewer than 90% valid estimates. As α_0 and γ decreases, the chances are of observing events decreases considerably.

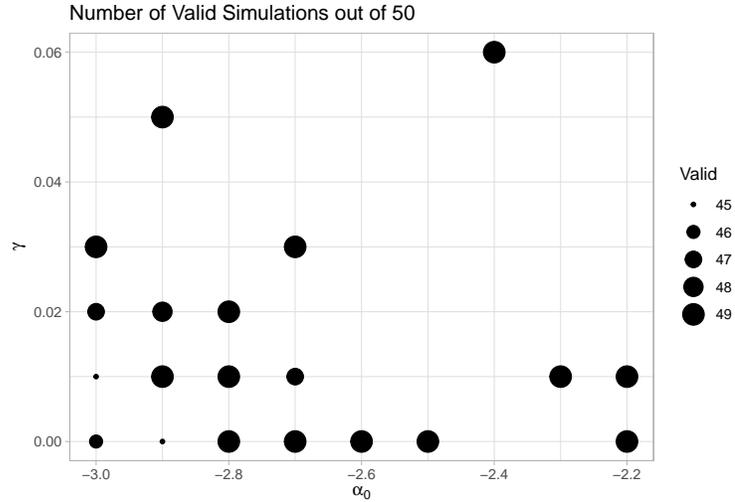


Figure 4: Number of Valid Simulations

Analysis

We begin the investigation of the small sample distribution of $\hat{\alpha}_0$ and $\hat{\gamma}$, by examining the distribution when $\alpha_0 = -1.4$ and $\gamma = .01$. We picked these values because these values are based on the data we used in the application section. From the density estimate of $\hat{\gamma}$ shown in Figure 5, we see that it looks approximately normal. We have performed the D'Agostino-Pearson test and Anscombe-Glynn test of kurtosis test to determine if our $\hat{\gamma}$ values are exactly normally distributed. The p-values (> 0.05) from both tests indicate that we have failed to reject the null hypothesis that the data is approximately normal.

Table 3 presents the summary statistics from distribution of $\hat{\gamma}$. As shown in the graph of the estimated density, the bias equals to 0.003, and the distribution is slightly positively skewed. However, the skewness is less than 0.5, and kurtosis is almost 3, the kurtosis of the normal distribution. We can conclude that the distribution of $\hat{\gamma}$ values are approximately normal. There is no conflict between the test results, Figure 5, and Table 3 in terms of being approximately normal for $\hat{\gamma}$ values.

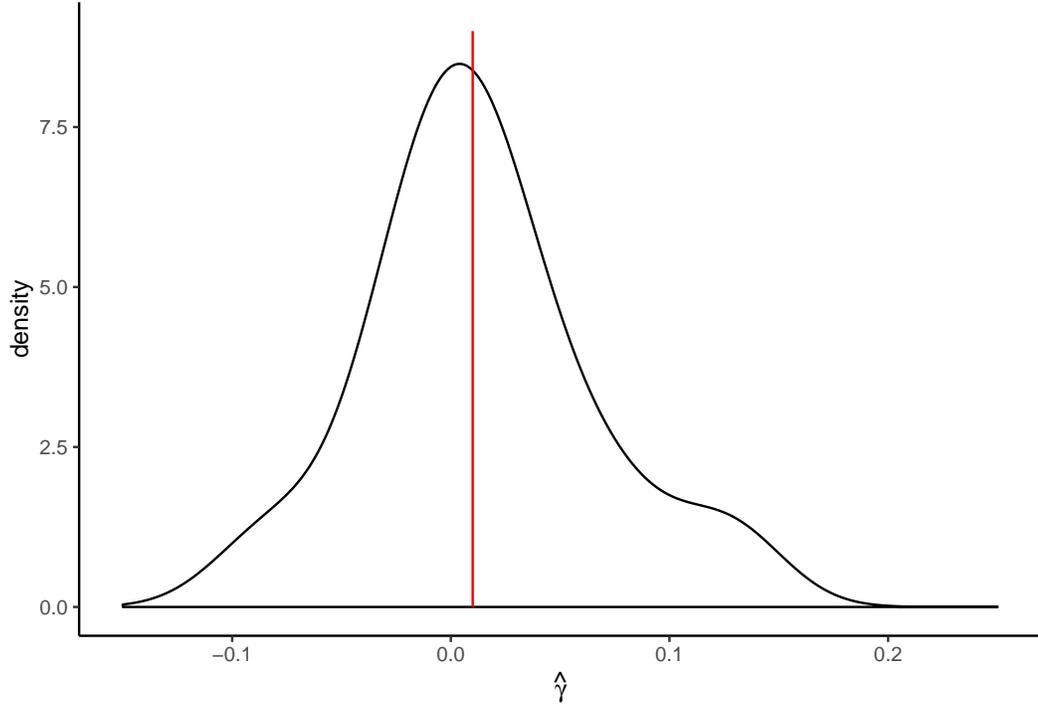


Figure 5: Density of $\hat{\gamma}$ with $\alpha_0 = -1.4$ and $\gamma = .01$

Table 3: Summary Statistics for Small Sample Distribution of $\hat{\gamma}$

mean	var	skew	kurt	mse
0.013	0.003	0.435	3.318	0.003

Now we consider the density plot for $\hat{\alpha}_0$ for the case where $\gamma = .01$ and $\alpha_0 = -1.4$ shown in Figure 6. In this density graph, the distribution is left skewed. Also, p-values (< 0.05) from the D'Agostino-Pearson test and Anscombe-Glynn test of kurtosis test indicates that we reject null hypothesis and conclude that the distribution of $\hat{\alpha}_0$ values are not approximately normal. The statistics in Table 4 shows the summary statistics from distribution of $\hat{\alpha}_o$. As shown in the Figure 6, the bias equals to -0.091 , the distribution is slightly left skewed. Since the skewness is -1.557 and kurtosis is very different from 3, the distribution of $\hat{\alpha}_o$ is not approximately normal. We do not know the cause of this non normality, and we need to investigate whether it is normal when μT goes to infinity.

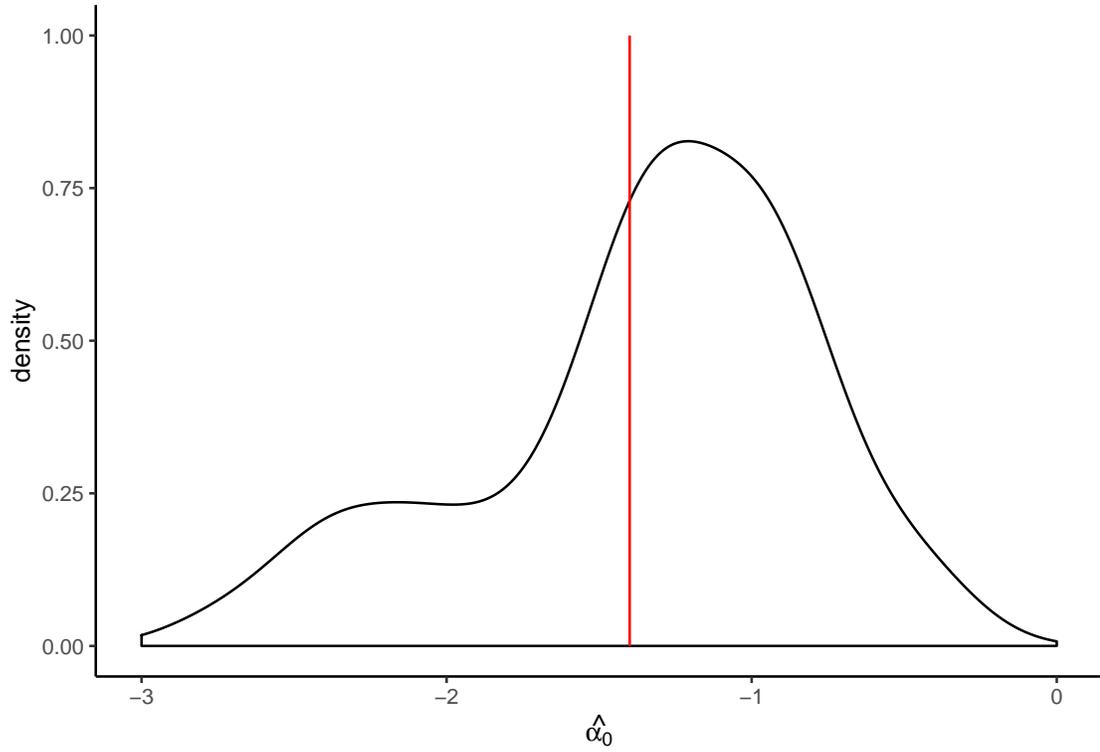


Figure 6: Density of $\hat{\alpha}_0$ with $\alpha_0 = -1.4$ and $\gamma = .01$

Table 4: Summary Statistics for Small Sample Distribution for $\hat{\alpha}_0$

mean	var	skew	kurt	mse
-1.491	0.629	-1.557	5.537	0.638

Figure 7 shows the effects of several α_0 values on the distribution of $\hat{\gamma}$ values where $\gamma = 0.01$.

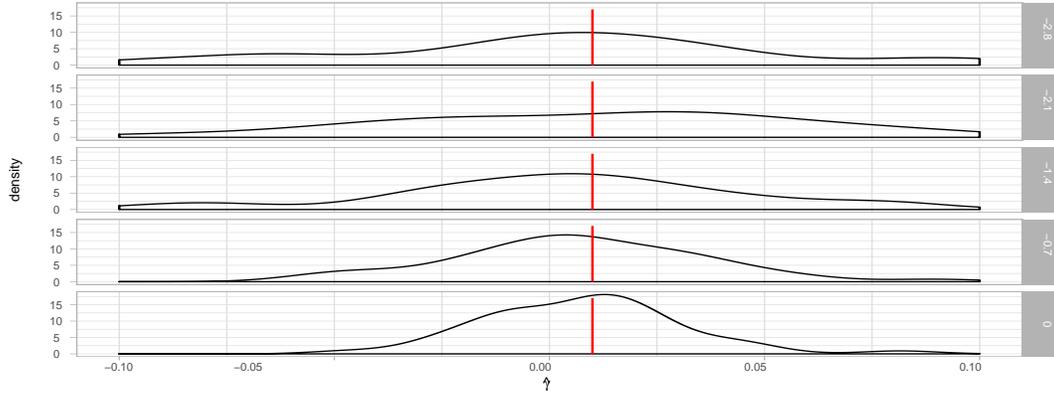


Figure 7: The Effect of α_0 on the Distribution of $\hat{\gamma}$

Table 5: Summary Statistics for the Distribution of $\hat{\gamma}$

alpha0	mean	var	skew	kurt	mse
-3.0	0.383	4.059	6.431	42.576	4.198
-2.8	0.026	0.012	-0.370	6.789	0.012
-2.6	-0.020	0.038	-1.720	14.220	0.038
-2.4	0.001	0.008	-0.146	4.849	0.009
-2.2	0.017	0.007	0.334	4.175	0.007
-2.0	0.018	0.004	0.696	4.537	0.004
-1.8	0.019	0.003	0.196	2.327	0.003
-1.6	-0.003	0.005	-0.779	4.905	0.005
-1.4	0.013	0.003	0.435	3.318	0.003
-1.2	0.021	0.002	1.875	9.454	0.002
-1.0	0.013	0.001	0.748	4.368	0.001
-0.8	0.013	0.002	1.208	7.656	0.002
-0.6	0.007	0.000	0.001	2.835	0.000
-0.4	0.013	0.001	0.091	2.877	0.001
-0.2	0.011	0.000	-0.194	2.492	0.000
0.0	0.008	0.001	0.438	4.065	0.001

Figure 7 demonstrates that the distribution of $\hat{\gamma}$ differs as α_o values vary in their range. It will enable us to investigate the distribution of $\hat{\gamma}$ values. We can conclude that the distribution of $\hat{\gamma}$ is not approaching a normal distribution for all the values of α_0 in the simulation by looking at the heavy tails from Figure 7. Moreover, we see from the Table 5 that $\hat{\gamma}$ values with the different α_o values are not always approaching a normal distribution. For instance, in the first row of the table, we see that skewness and the kurtosis are very large compare to the cut of values for a normal distribution. Another example is when $\alpha_0 = -2.6$, skewness and kurtosis are very large in magnitude, and the distribution of the $\hat{\gamma}$ is not behaving like a normal distribution. Figure 7 and the Table 5 are crucial tools for investigating the distribution of $\hat{\gamma}$.

Table 6: Summary Statistics for Small Sample Distribution for $\hat{\alpha}_0$

gamma	mean	var	skew	kurt	mse
0.00	-1.406	0.353	-0.116	2.646	0.353
0.05	-1.532	0.233	-0.272	2.441	0.250
0.10	-1.436	0.173	-0.532	3.447	0.174
0.15	-1.401	0.098	-0.600	2.946	0.098
0.20	-1.394	0.066	-1.058	5.125	0.066
0.25	-1.463	0.034	0.340	4.005	0.038
0.30	-1.410	0.013	-0.533	3.463	0.013
0.35	-1.395	0.009	0.871	5.996	0.010
0.40	-1.412	0.005	-0.481	4.705	0.005

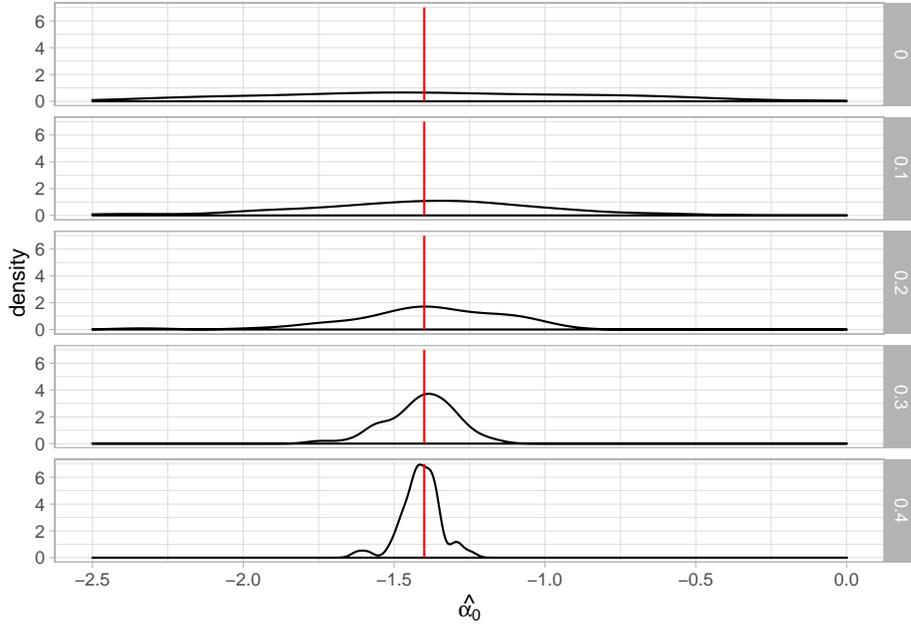


Figure 8: Effect of γ on Distribution of $\hat{\alpha}_0$

Figure 8 shows the effects of a couple of γ values on the distribution of $\hat{\alpha}_0$ values where $\alpha_0 = -1.4$. It would be appropriate to say that the most of the density graphs are not representing the normal distribution. They have very heavy tails and especially top density graph almost demonstrates a straight line. In addition to Figure 8, we can also assess the distribution of $\hat{\alpha}_0$ values by looking at the Table 6. We see from Table 6 that the kurtosis values are not equal to 3 and some of them are bigger than 3, and these values indicate that distributions have heavy to moderate tails.

Figure 9 demonstrates the effects of $\alpha_0 (> -2.5)$ values on the $\text{Bias}(\hat{\gamma})$ values. When α_0 values increase, the absolute $\text{Bias}(\hat{\gamma})$ decreases. Also, overall, $\text{Bias}(\hat{\gamma})$ values are small in magnitude ($< .03$).

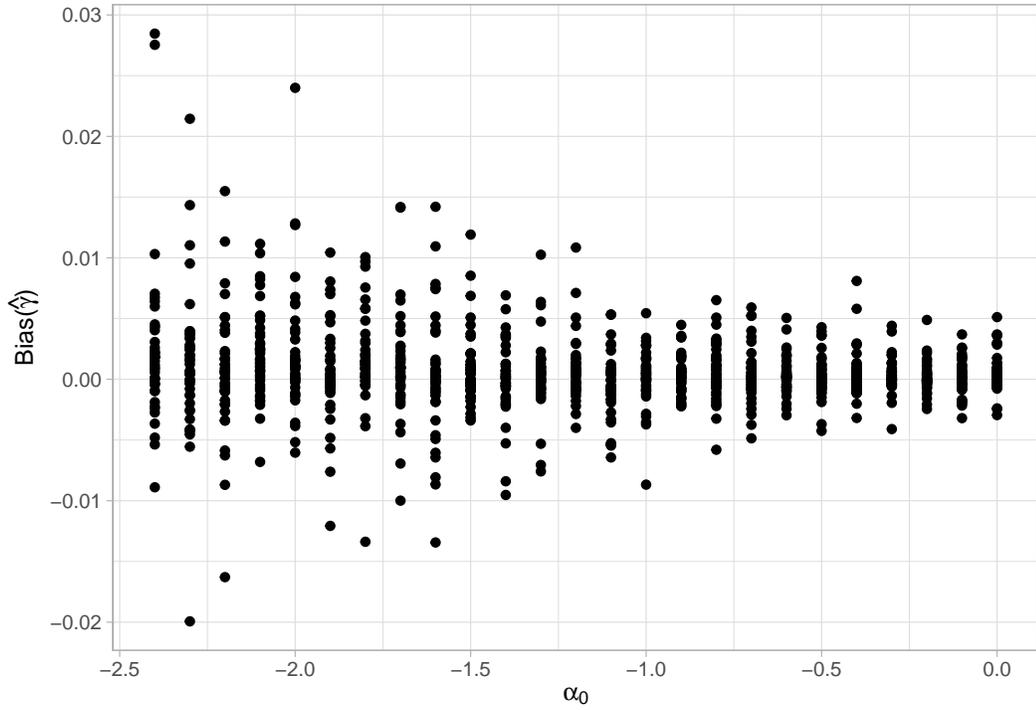


Figure 9: Effect of α_0 on $Bias(\hat{\gamma})$

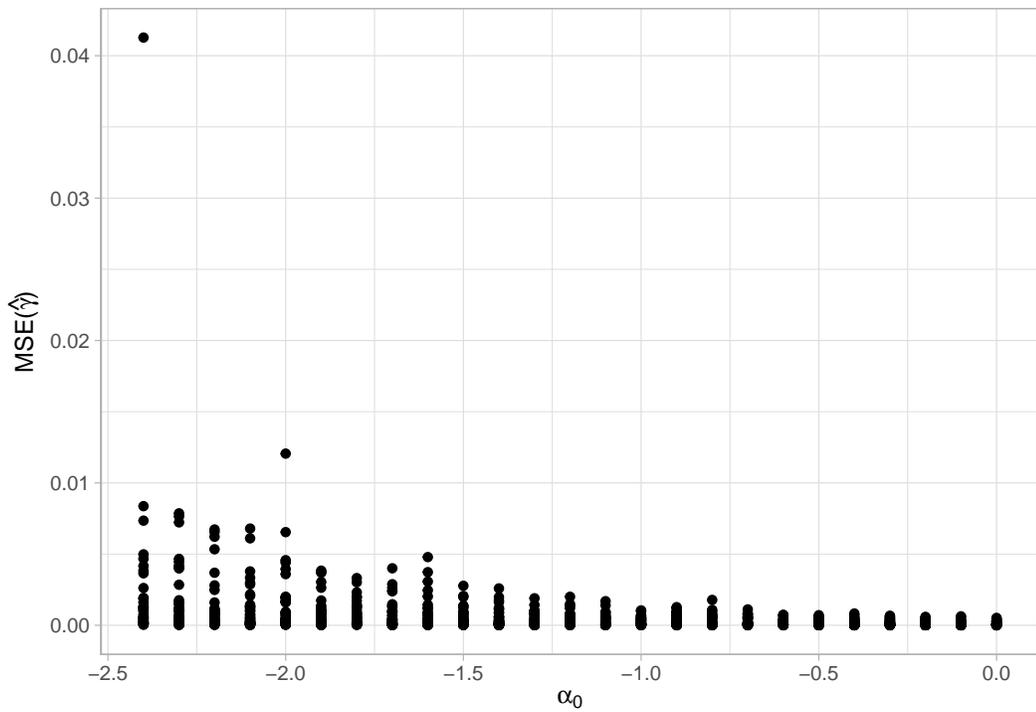


Figure 10: Effect of α_0 on $MSE(\hat{\gamma})$

Figure 10 presents the effects of $\alpha_0 (> -2.5)$ values on the $MSE(\hat{\gamma})$ values. $MSE(\hat{\gamma})$ values tend to decrease when α_0 values increase.

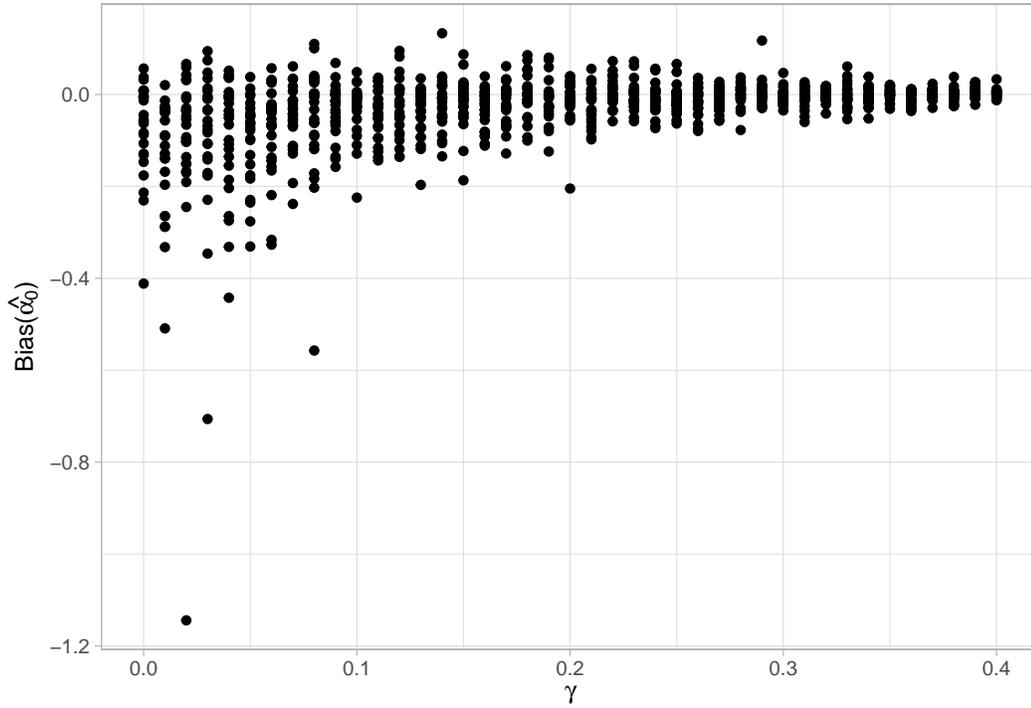


Figure 11: Effect of γ on $Bias(\hat{\alpha}_0)$

Figure 11 demonstrates the effects of γ values on the $Bias(\hat{\alpha}_0)$ values where $\alpha_0 > -2.5$. We see that when γ values increase, $Bias(\hat{\alpha}_0)$ values decrease. The big difference between $Bias(\hat{\alpha}_0)$ values and $Bias(\hat{\gamma})$ values is that $Bias(\hat{\alpha}_0)$ have more negative values and slightly larger magnitude than $Bias(\hat{\gamma})$. However, when γ values increase, $Bias(\hat{\alpha}_0)$ values are approaching 0 more than $Bias(\hat{\gamma})$. The picture seems more stable.

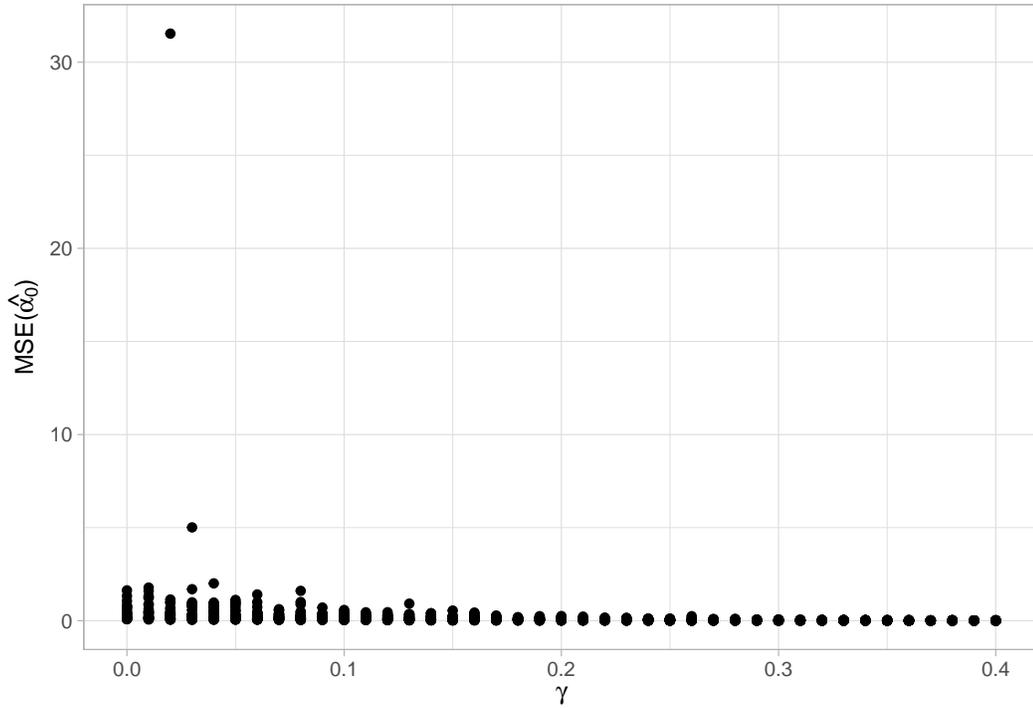


Figure 12: Effect of γ on $MSE(\hat{\alpha}_0)$

Figure 12 exhibits the effects of γ values on the $MSE(\hat{\alpha}_0)$ values where $\alpha_0 > -2.5$. When γ values increase, $MSE(\hat{\alpha}_0)$ values decrease. Overall, $MSE(\hat{\alpha}_0)$ are larger than $MSE(\hat{\gamma})$ values.

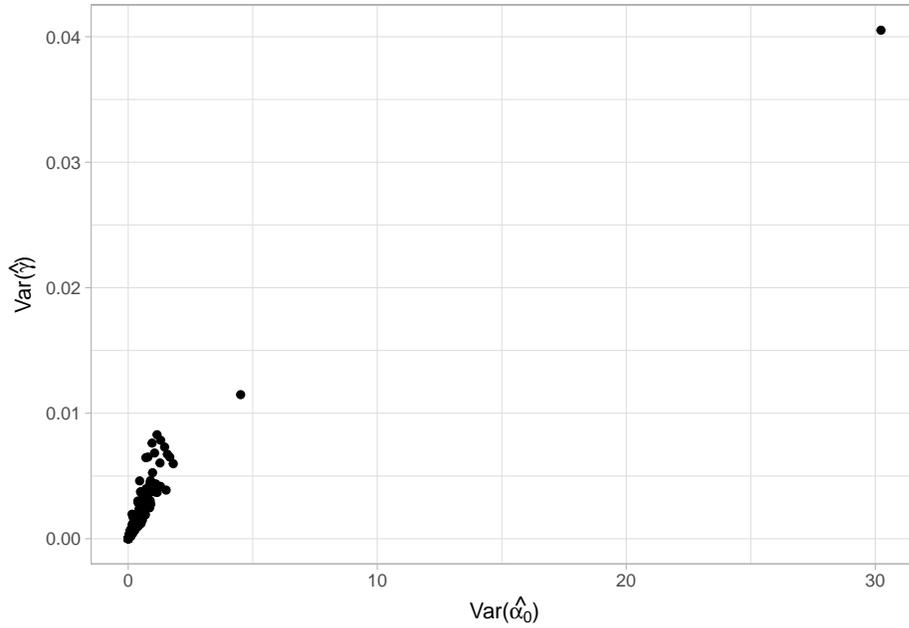


Figure 13: $Var(\hat{\alpha}_0)$ vs. $Var(\hat{\gamma})$

Figure 14 illustrates the correlation between $MSE(\hat{\alpha}_0)$ and $MSE(\hat{\gamma})$ where $\alpha_0 > -2.5$. As $MSE(\hat{\alpha}_0)$ increases, $MSE(\hat{\gamma})$ also tends to increase. We can call this correlation as a fairly strong positive relationship. The interpretation of the

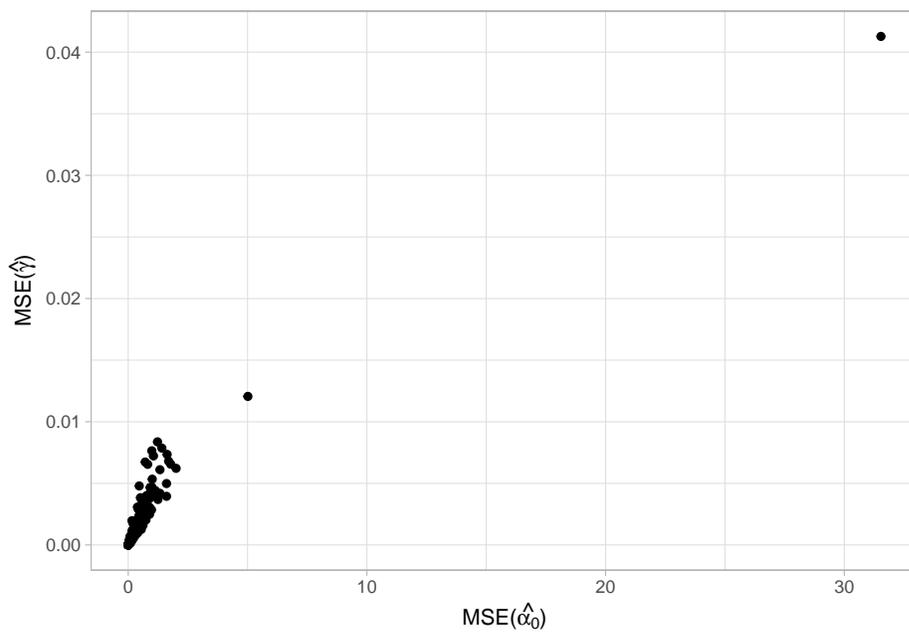


Figure 14: $MSE(\hat{\alpha}_0)$ vs. $MSE(\hat{\gamma})$

scatter-plot of $\text{Var}(\hat{\alpha}_0)$ versus $\text{Var}(\hat{\gamma})$ is same as the interpretation of MSE graphs. However, Figure 15 graph indicates a different result. As $\text{Bias}(\hat{\alpha}_0)$ increases, $\text{Bias}(\hat{\gamma})$ tends to decrease. We can call it as a fairly strong negative relationship.

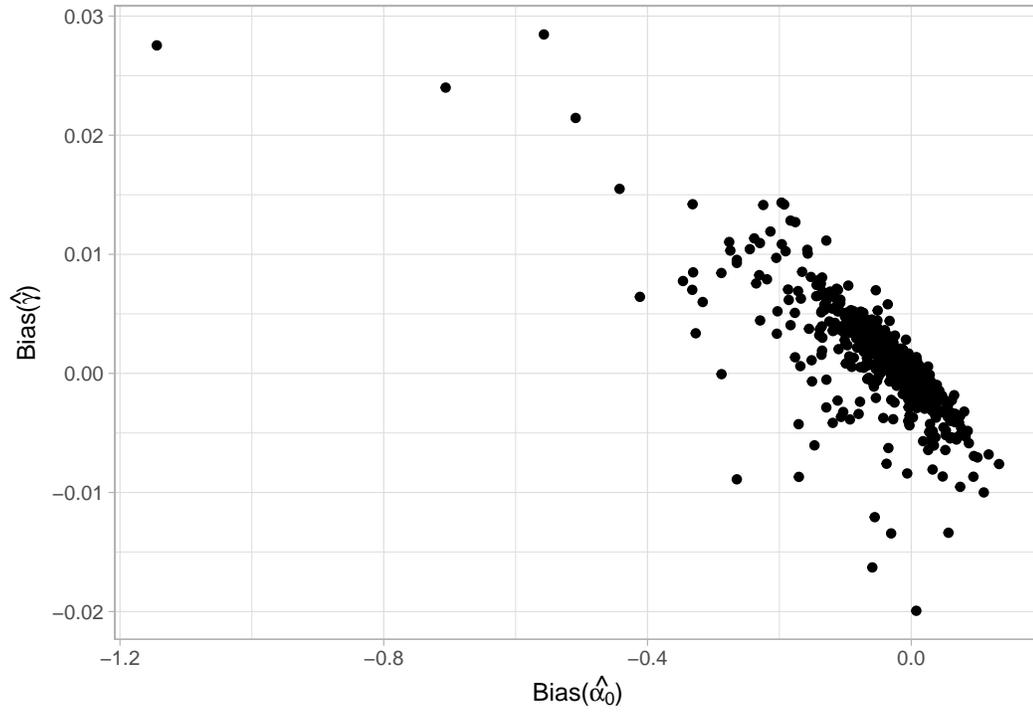


Figure 15: $\text{Bias}(\hat{\alpha}_0)$ vs. $\text{Bias}(\hat{\gamma})$

IV. META-ANALYSIS OF MULTIPLE CLASSROOMS SIMULATION

Motivation

As described in the preceding chapters, the goal of this project is to model the relationship between two stochastic processes and estimate the catalytic constant γ . In Chapter 2, we derived the maximum likelihood estimates and the asymptotic distribution for the parameter estimates. However, the simulation study of the small sample distribution described in Chapter 3 showed that data from one classroom was insufficient to allow for the use of the asymptotic results. The distribution was biased and the asymptotic variance estimates were too small compared to the computed variance from the simulated data. In this chapter, we extend the results shown above to a "meta-analysis" of data collected from observations of independent classrooms. A conventional meta-analysis is a statistical method that combines the results of multiple studies because results typically vary from one study to the next. In this research, we did not use different studies, but we considered each classroom as an independent observation and combined them to perform meta-analysis. Therefore, we will combine analyses from separate classrooms to derive a pooled estimate of γ .

Method

The assumption of independence between the classrooms implies that the log-likelihood for the combined data set will be the sum of the log-likelihoods from each classroom. This provides a straightforward road map to extend the results from Chapter 2 to the meta-analysis problem.

1. Extend the log-likelihood from (2.1) to include independent data from P classrooms.

2. Use the linearity of the derivatives to compute the gradient of the log-likelihood and determine the likelihood equations for the meta-analysis.
3. Find the Hessian and compute the Fisher Information for meta-analysis problem.
4. Adapt the R code from Chapter 3 to allow for data from multiple classrooms.
5. Repeat the simulation study described in Chapter 3 for the simulations of data from P classrooms. Compare the results between two simulation studies.

Parameter Estimation

We start showing the extension of the log-likelihood function of one classroom case to the log-likelihood function of P classrooms case by using the meta-analysis method. We will repeat our previous experiment $p = 1, \dots, P$ times with fixed α_0, γ and μ . The log-likelihood of meta-analysis experiment is given by

$$l_p = \sum_{p=1}^P \left(-\mu T_p + M_p \log(\mu T_p) - w_{M+1_p} \exp(\alpha_0) + N_p \alpha_0 + \gamma \sum_{j=1}^{N_p} k_{X(j)_p}^* \right). \quad (25)$$

Taking the derivative with respect to μ ,

$$\frac{\partial l_p}{\partial \mu} = \sum_{p=1}^P \left(-T_p + \frac{M_p}{\mu} \right), \quad (26)$$

we find that the m.l.e., $\hat{\mu} = \sum_{p=1}^P \frac{M_p}{T_p}$, is very similar to one classroom case. It is appropriate to say that finding $\hat{\mu}$ was straightforward. To find the estimates of α_0 and γ , we take the derivative of the log-likelihood function for P classrooms with respect to α_0 and γ . From the derivatives,

$$\frac{\partial l_p}{\partial \alpha_0} = -e^{\alpha_0} \sum_{k=1}^{M_p+1} e^{\gamma(k-1)} (u_{(k)_p} - u_{(k-1)_p}) + N_p, \quad (27)$$

and

$$\frac{\partial l_p}{\partial \gamma} = -e^{\alpha_0} \sum_{k=1}^{M_p+1} (((k-1)e^{\gamma(k-1)} (u_{(k)_p} - u_{(k-1)_p})) + \sum_{j=1}^{N_p} k_{X(j)_p}^*) \quad (28)$$

we see that the sufficient statistics are $\sum_{p=1}^P M_p, \sum_{p=1}^P N_p, \{u_{(1)_p}, \dots, u_{(M)_p}\}, \sum_{p=1}^P \sum_{j=1}^{N_p} k_{X(j)_p}^*$. Due to the highly nonlinear form of (27) and (28), the likelihood equations have no closed form solutions. As in Chapter 2 and 3, the `multiroot` from the R package `rootsolve` will provide numerical estimates of α_0 and γ . We did not give all the details for this section since all the functions are extensions of those in Chapter 2 and 3.

Asymptotic Distribution

These P classrooms are independent and Fisher information for meta-analysis can be obtained adding each classrooms Fisher information,

$$I = I(\mu, \lambda, \gamma) = \sum_{p=1}^P I_p(\mu, \lambda, \gamma)$$

$$I_p(\mu, \lambda, \gamma) = - \begin{bmatrix} E \left[\frac{\partial^2 l_p}{\partial \mu^2} \right] & 0 & 0 \\ 0 & E \left[\frac{\partial^2 l_p}{\partial \alpha_0^2} \right] & E \left[\frac{\partial^2 l_p}{\partial \gamma \partial \alpha_0} \right] \\ 0 & E \left[\frac{\partial^2 l_p}{\partial \gamma \partial \alpha_0} \right] & E \left[\frac{\partial^2 l_p}{\partial \gamma^2} \right] \end{bmatrix}$$

where

$$E \left[\frac{\partial^2 l_p}{\partial \mu^2} \right] = \sum_{p=1}^P -\frac{T_p}{\mu} \quad (29)$$

$$E \left[\frac{\partial^2 l_p}{\partial \alpha_0^2} \right] = \sum_{p=1}^P -\frac{T_p e^{\alpha_0}}{\lambda(1-e^\gamma)} (1 - e^{\lambda(e^\gamma-1)}) \quad (30)$$

$$E \left[\frac{\partial^2 l_p}{\partial \gamma \partial \alpha_0} \right] = \sum_{p=1}^P -\frac{T_p e^{\alpha_0+\gamma}}{(1-e^\gamma)^2} \left(\frac{1-e^{-\lambda}}{\lambda} - e^{\lambda(e^\gamma-1)} + \frac{e^{-\lambda}}{\lambda} (e^{e^\gamma \lambda} (e^\gamma \lambda - 1) + 1) \right) \quad (31)$$

$$E \left[\frac{\partial^2 l_p}{\partial \gamma^2} \right] = \sum_{p=1}^P -\frac{T_p e^{\alpha_0+\gamma}}{(1-e^\gamma)^3} \left((1+e^\gamma) \frac{(1-e^{-\lambda})}{\lambda} - \lambda e^{\gamma+\lambda(e^\gamma-1)} - e^{\lambda(e^\gamma-1)} \right. \\ \left. + (2-e^\gamma) \frac{e^{-\lambda}}{\lambda} (e^{e^\gamma \lambda} (1-e^\gamma \lambda + e^{2\gamma} \lambda^2) - 1) \right. \\ \left. + 2 \frac{e^{-\lambda}}{\lambda} (e^{e^\gamma \lambda} (e^\gamma \lambda - 1) + 1) - \frac{e^{-\lambda}}{\lambda} (e^{e^\gamma \lambda} - 1) \right) \quad (32)$$

and $\lambda = \mu T_p$ by using the extension of chapter 2 results, the Fisher information, maximum likelihood theory provides the asymptotic (as $T \rightarrow \infty$) distribution

$$\hat{\boldsymbol{\theta}} = (\hat{\mu}, \hat{\alpha}_0, \hat{\gamma})' \sim N(\boldsymbol{\theta}, \mathbf{I}(\boldsymbol{\theta})^{-1})$$

where $\boldsymbol{\theta} = (\mu, \alpha_0, \gamma)'$. In applications, we substitute $\hat{\boldsymbol{\theta}}$ in $\mathbf{I}(\boldsymbol{\theta})$ to estimate the variance and $\hat{Var}(\hat{\boldsymbol{\theta}}) = \mathbf{I}(\hat{\boldsymbol{\theta}})^{-1} = \left\{ \sum_{p=1}^P I_p \right\}^{-1}$.

Meta-Analysis Study of Multiple Classrooms Simulation Description

Our description for simulation study is same as the previous one classroom simulation except some extension. We have kept the first part of Figure 3 same and we made a loop from 1 to P to find $\sum_{p=1}^P M_p, \sum_{p=1}^P N_p, \{u_{(1)_p}, \dots, u_{(M)_p}\}, \sum_{p=1}^P \sum_{j=1}^{N_p} k_{X(j)_p}^*$. Also, we made another loop for (25) from 1 to P to estimate our parameters. We have used (29),(30),(31),(32) and the Fisher information matrix for meta analysis to compute to variance estimation. The complete R code is included

in Appendix B.

Analysis for Meta-Analysis of Multiple Classrooms Simulation

Table 7: First 5 Rows of Results

gamma	alpha0	cnt	g.bias	a.bias	g.mse	a.mse	95%CI($\hat{\alpha}_0$)	95% CI($\hat{\gamma}$)
0	-3.0	12	-0.012	0.18	0.00	0.08	(-3.26,-2.35)	(-0.04,0.01)
0	-2.9	11	-0.001	0.09	0.00	0.03	(-3.13,-2.48)	(-0.01,0.01)
0	-2.8	14	0.007	-0.03	0.00	0.02	(-3.15,-2.50)	(-0.01,0.02)
0	-2.7	31	-0.001	-0.00	0.00	0.03	(-3.08,-2.31)	(-0.03,0.02)
0	-2.6	30	0.003	-0.02	0.00	0.07	(-3.15,-2.08)	(-0.02,0.03)

Table 7 shows the first 5 rows of the results of the meta analysis of multiple classroom simulation. From Table 7, $MSE(\hat{\gamma})$ values look 0. They are actually not zero, some of them have 7 decimal places, therefore, they are too minuscule. Our software program did not show them. We have calculated some of them individually and observed that they are very small.

Figure 16 illustrates the number of valid simulations out of 50. As we see, when α_0 and γ gets smaller, we have less valid simulations. We also have many number of 50 valid simulation when $\alpha_0 > -2$ with γ values. As we see from Figure 16, we have more valid number of simulations compared to the Figure 4 due to the meta analysis experiment.

We now begin the investigation of the small sample distribution of $\hat{\alpha}_0$ and $\hat{\gamma}$, by examining the distribution when $\alpha_0 = -1.4$ and $\gamma = .01$. From the density of $\hat{\gamma}$ shown in Figure 17, we see that it looks approximately normal. We have performed the D'Agostino-Pearson test and Anscombe-Glynn test of kurtosis test to determine if our $\hat{\gamma}$ values are normally distributed. The p-values (> 0.05) from both outputs

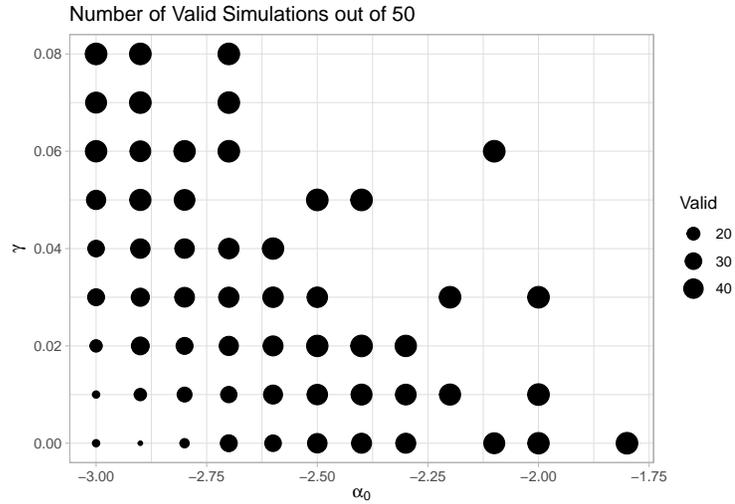


Figure 16: Number of Valid Simulations

indicate that we have failed to reject the null hypothesis, and we can conclude that the data is normally distributed. It would be appropriate to state that we have much higher p-values compared to the one classroom case tests.

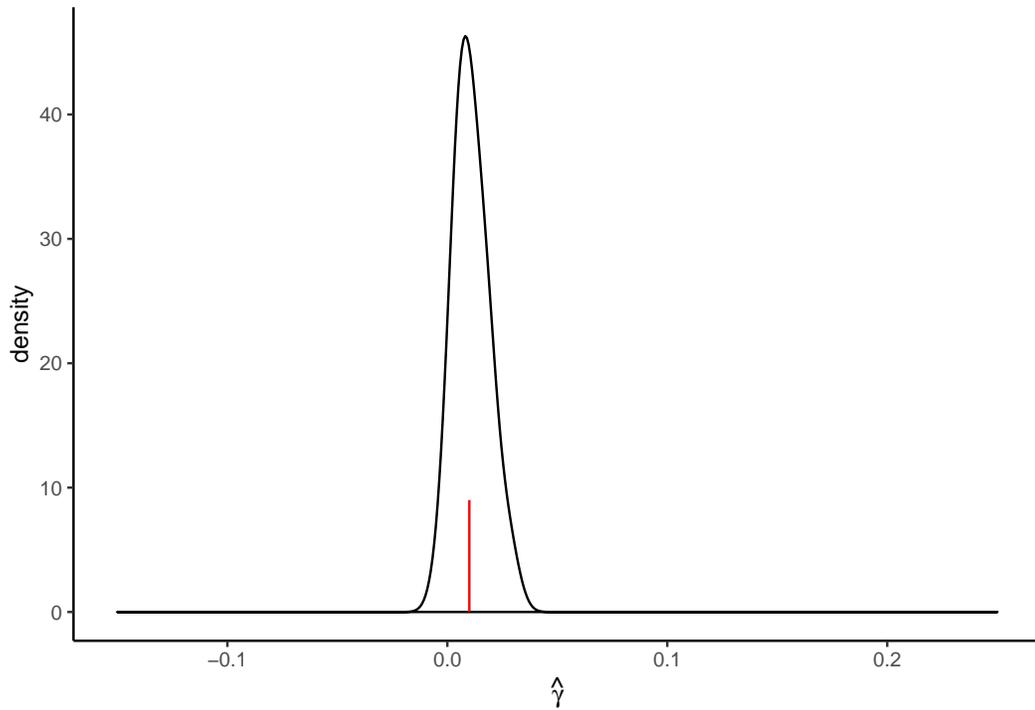


Figure 17: Density of $\hat{\gamma}$ with $\alpha_0 = -1.4$ and $\gamma = .01$

Table 8: Summary Statistics for Small Distribution for $\hat{\gamma}$

mean	skew	kurt	mse*1000	var*1000
0.011	0.529	2.798	0.0057	0.056

We can also see it from Table 8 that kurtosis is very close to 3 and skewness is almost 0.5. This summary statistics align with our test results. As we mentioned before variance and mse values are very small and our software program produced them as 0. For the report purposes, we have multiplied variance and mse values by 1000.

Now we consider the density plot for $\hat{\alpha}_0$ for the case where $\gamma = .01$ and $\alpha_0 = -1.4$ shown in Figure 18. It seems that $\hat{\alpha}_0$ values are slightly left skewed. Also, p-values (> 0.05) from the D'Agostino-Pearson test and Anscombe-Glynn test of kurtosis test indicates that we have failed to reject null hypothesis and conclude that the distribution of $\hat{\alpha}_0$ is approximately normal. The statistics in Table 9 shows the summary statistics from distribution of $\hat{\alpha}_0$. As shown in the Figure 18, the bias equals to -0.002 , the distribution is normaly distributed. Since the skewness is -0.603 and kurtosis is very close to 3, the distribution of $\hat{\alpha}_0$ is approximately normal.

If we compare Figure 18 and Figure 6, we see that Figure 18 looks closer to normal. Also our test results and Table 9 supports our decision. We see that meta-analysis method gave us statistically better results. Both of $\hat{\alpha}_0$ and $\hat{\gamma}$ values are normally distributed where $\alpha_0 = -1.4$ and $\gamma = 0.01$.

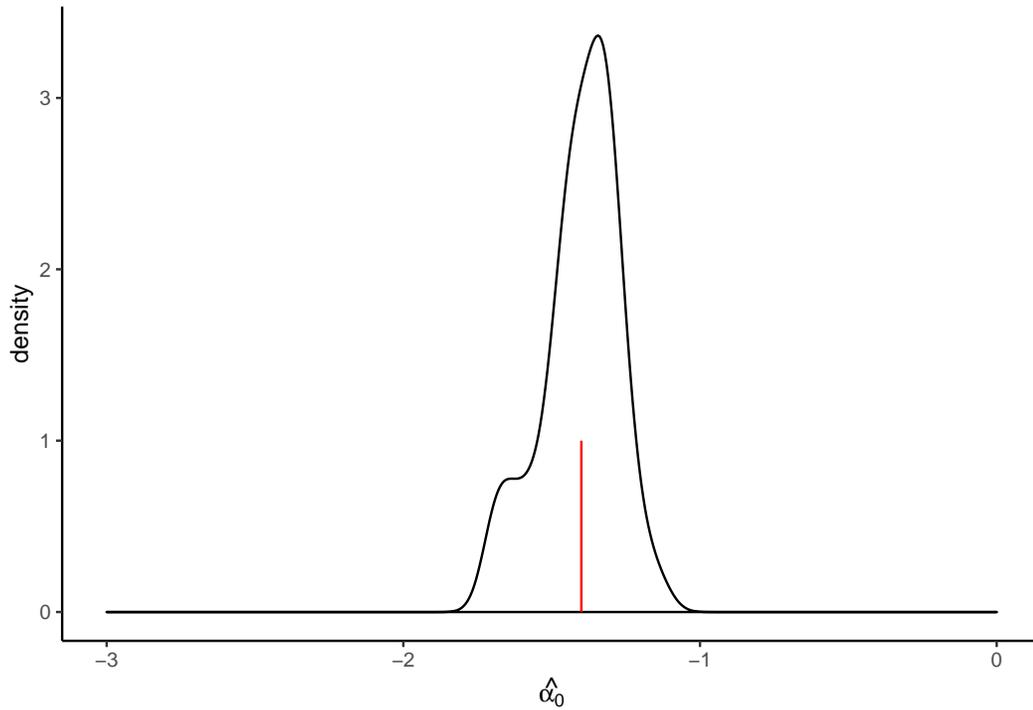


Figure 18: Density of $\hat{\alpha}_0$ with $\alpha_0 = -1.4$ and $\gamma = .01$

Table 9: Summary Statistics for Small Sample Distribution for $\hat{\alpha}_0$

mean	var	skew	kurt	mse
-1.402	0.016	-0.603	3.003	0.016

Figure 19 shows the effects of several α_0 values on the distribution of $\hat{\gamma}$ values where $\gamma = 0.01$. We see that Figure 19 looks more normal and has a smaller spread compared to the Figure 7. Also, from Table 10, most of the kurtosis values are very close to 3 and skewness values are very small. When we compare Table 10 and Table 5, we see that kurtosis and skewness values indicate more normality for meta analysis experiment. As we mentioned before, we multiplied original variance and mse values by 1000 due the fact having too minuscule variance and mse values.

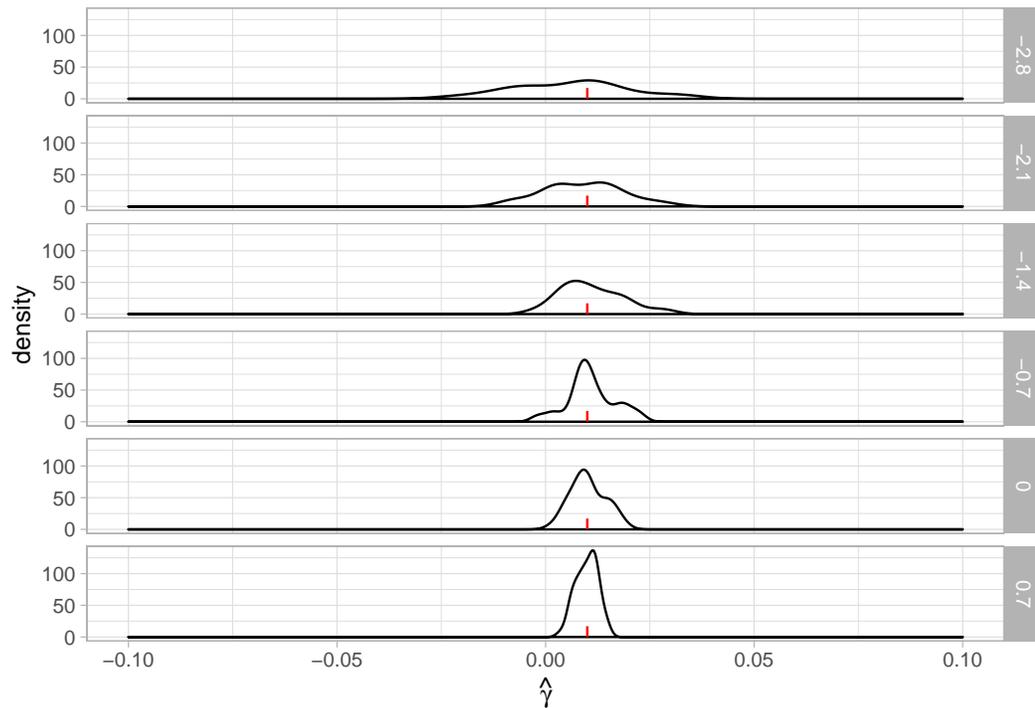


Figure 19: The Effect of α_0 on the Distribution of $\hat{\gamma}$

Table 10: Summary Statistics for Small Sample Distribution for $\hat{\gamma}$

alpha0	mean	var*1000	skew	kurt	mse*1000
-3.0	0.012	0.145	-0.737	3.004	0.149
-2.8	0.006	0.193	0.030	2.551	0.210
-2.6	0.009	0.171	0.228	2.435	0.172
-2.4	0.010	0.109	-0.639	3.636	0.109
-2.2	0.011	0.183	-0.275	2.824	0.185
-2.0	0.010	0.125	-0.064	3.228	0.125
-1.8	0.009	0.083	-0.410	2.974	0.084
-1.6	0.008	0.068	0.316	2.248	0.073
-1.4	0.011	0.056	0.529	2.798	0.057
-1.2	0.009	0.054	-0.529	3.584	0.054
-1.0	0.009	0.032	0.134	2.334	0.032
-0.8	0.011	0.038	0.223	2.417	0.038
-0.6	0.010	0.019	-0.152	2.761	0.019
-0.4	0.009	0.019	0.234	2.752	0.021
-0.2	0.010	0.013	0.167	3.099	0.014
0.0	0.010	0.018	0.208	2.386	0.018
0.2	0.010	0.010	0.054	3.349	0.010
0.4	0.009	0.010	0.128	2.577	0.011
0.6	0.010	0.005	-0.186	2.671	0.005
0.8	0.009	0.008	-0.199	2.581	0.009
1.0	0.009	0.006	-0.100	2.779	0.006

Table 11: Summary Statistics for Small Sample Distribution for $\hat{\alpha}_0$

gamma	mean	var	skew	kurt	mse
0.00	-1.382	0.016	0.027	2.809	0.016
0.05	-1.421	0.009	0.079	2.138	0.010
0.10	-1.406	0.007	0.013	2.837	0.007
0.15	-1.395	0.003	0.148	2.036	0.003
0.20	-1.402	0.002	-0.183	2.484	0.002
0.25	-1.398	0.001	-0.219	3.570	0.001
0.30	-1.401	0.000	0.210	2.592	0.000
0.35	-1.398	0.000	0.567	4.053	0.000
0.40	-1.400	0.000	0.210	2.893	0.000

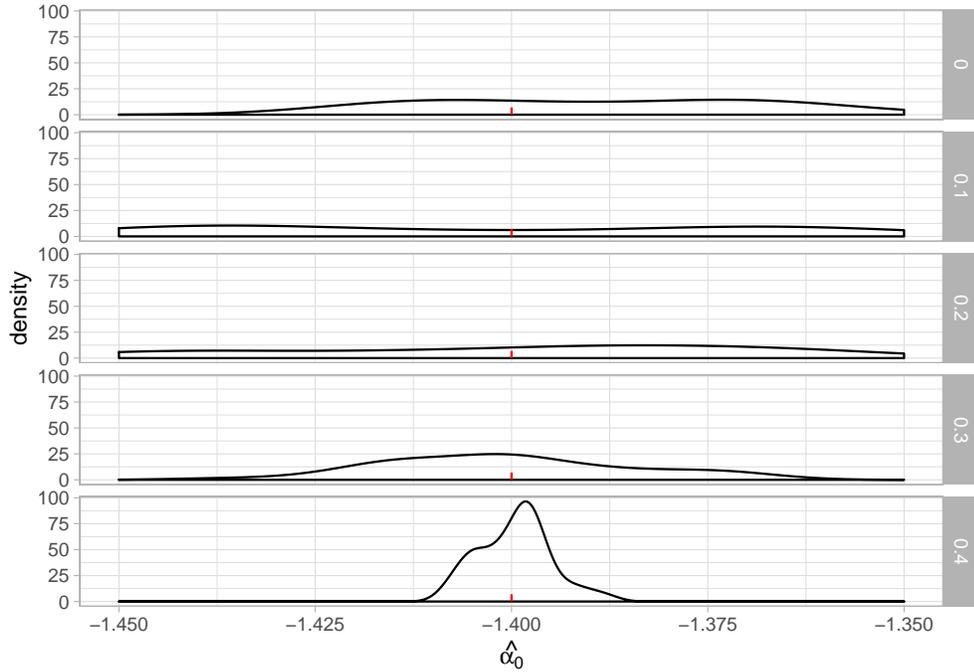


Figure 20: Effect of γ on Distribution of $\hat{\alpha}_0$

Figure 20 illustrates the effects of a couple of γ values on the distribution of $\hat{\alpha}_0$ values where $\alpha_0 = -1.4$. It is not very clear to see from Figure 20 that if

$\hat{\alpha}_0$ values behave normal because of the scaling, but performed tests and Table 11 show that the distribution of $\hat{\alpha}_0$ values from meta-analysis experiment behave more normal. When we compare Table 11 and Table 6, we have less skewness and most of the kurtosis values are around 3, and we have much smaller MSE values. For instance, if we compare the fifth row of Table 11 and Table 6, we see a significant difference. Table 11 shows smaller skewness values, better kurtosis values, much more smaller variance and slightly less bias. However we could not see it from the Figure 20 because of the scaling.

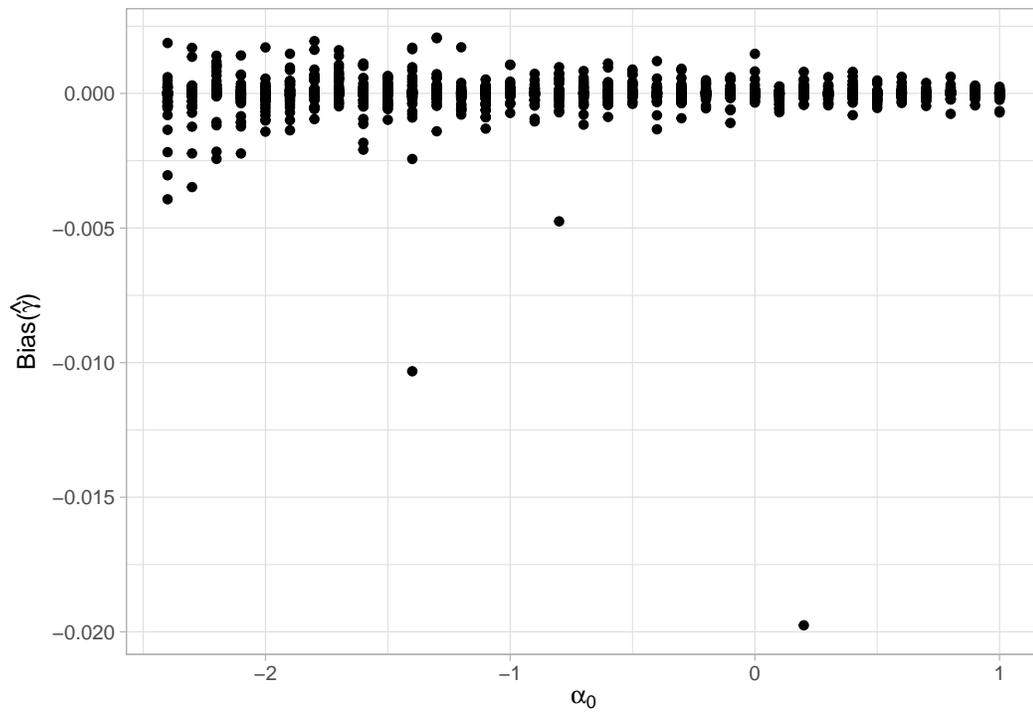


Figure 21: Effect of α_0 on $Bias(\hat{\gamma})$

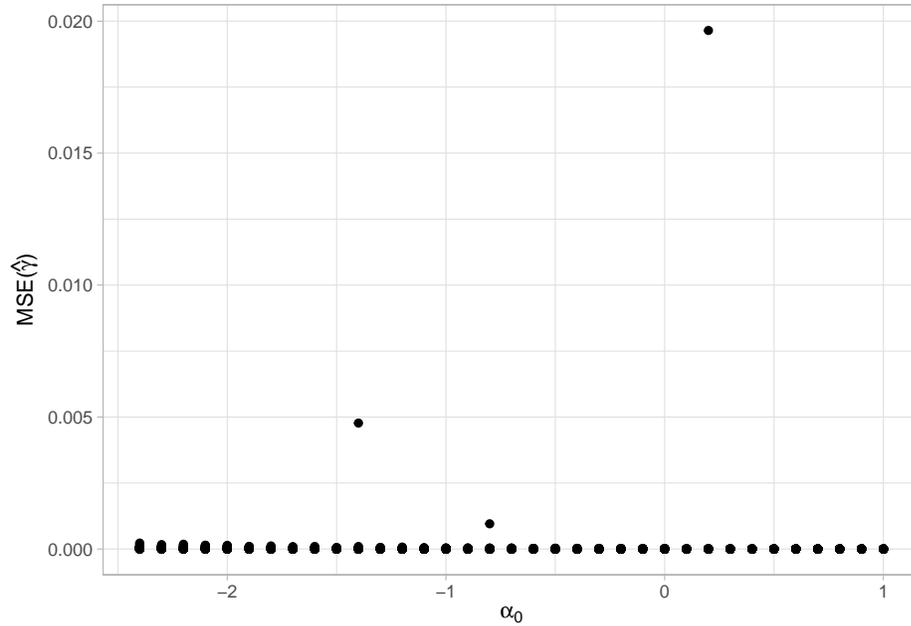


Figure 22: Effect of α_0 on $MSE(\hat{\gamma})$

Figure 21 presents the effects of $\alpha_0 > -2.5$ values on the $Bias(\hat{\gamma})$ values. $Bias(\hat{\gamma})$ values tend to decrease when α_0 values increase. When we compare Figure 21 and Figure 9, we see notable differences that $Bias(\hat{\gamma})$ values are mostly around 0 and they are smaller in magnitude on Figure 21. There is also much less spread in $Bias(\hat{\gamma})$ values in the meta analysis experiment.

Figure 22 illustrates the effects of $\alpha_0 > -2.5$ values on the $MSE(\hat{\gamma})$ values. It seems that $MSE(\hat{\gamma})$ values decrease when α_0 values increase. We also have a couple of values that do not behave like the most of them. However, if we compare Figure 22 and Figure 10, we have much smaller $MSE(\hat{\gamma})$ values which indicates that meta-analysis experiment results are statistically more powerful than one classroom case study.

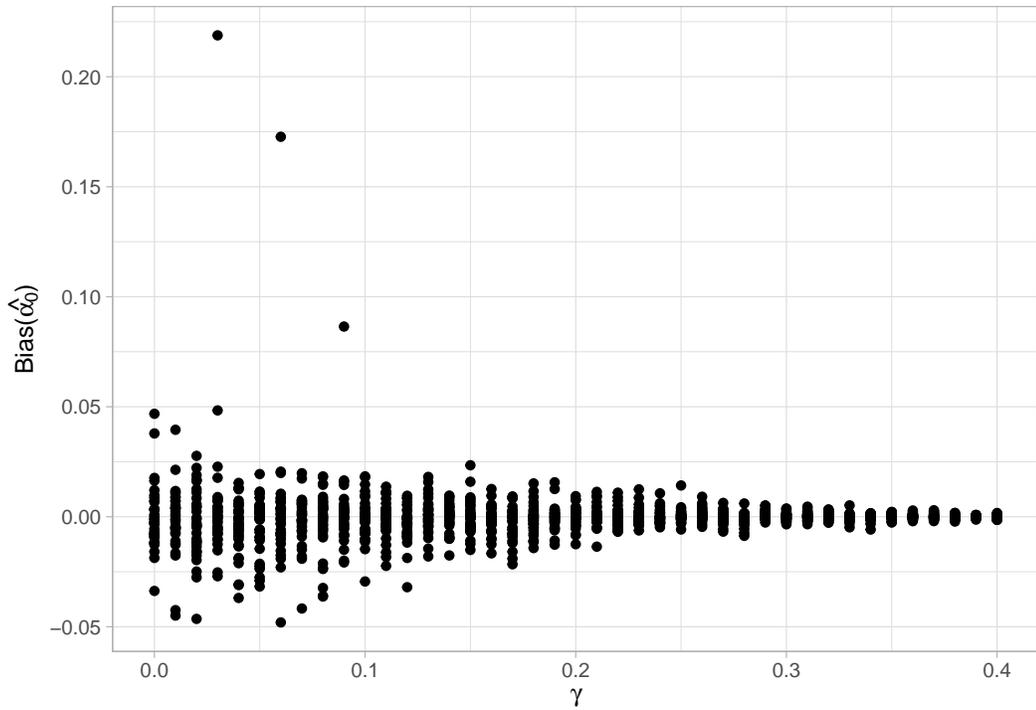


Figure 23: Effect of γ on $Bias(\hat{\alpha}_0)$

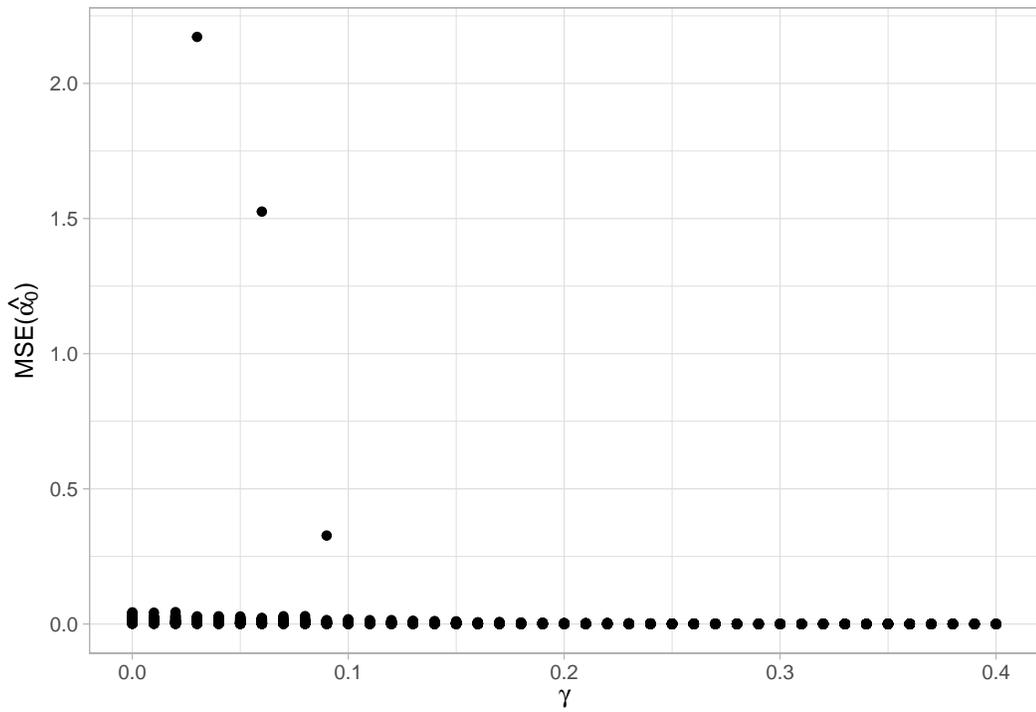


Figure 24: Effect of γ on $MSE(\hat{\alpha}_0)$

Figure 23 exhibits the effects of γ values on the $\text{Bias}(\hat{\alpha}_0)$ values. It seems that $\text{Bias}(\hat{\alpha}_0)$ values decrease when γ values increase. We also have a couple of values that do not behave like the most of them. However, if we compare Figure 23 and Figure11, we have smaller $\text{Bias}(\hat{\alpha}_0)$ values, and they are mostly around 0 which is a good sign for our meta-analysis experiment. Also, $\text{Bias}(\hat{\alpha}_0)$ values are smaller in magnitude on Figure 23.

Figure 24 shows the effects of γ values on the $\text{MSE}(\hat{\alpha}_0)$ values. It seems that $\text{MSE}(\hat{\alpha}_0)$ values decrease when γ values increase. If we compare Figure 24 and Figure12, we have smaller $\text{MSE}(\hat{\alpha}_0)$ values and they are mostly around 0 which is an important indication for our meta-analysis experiment.

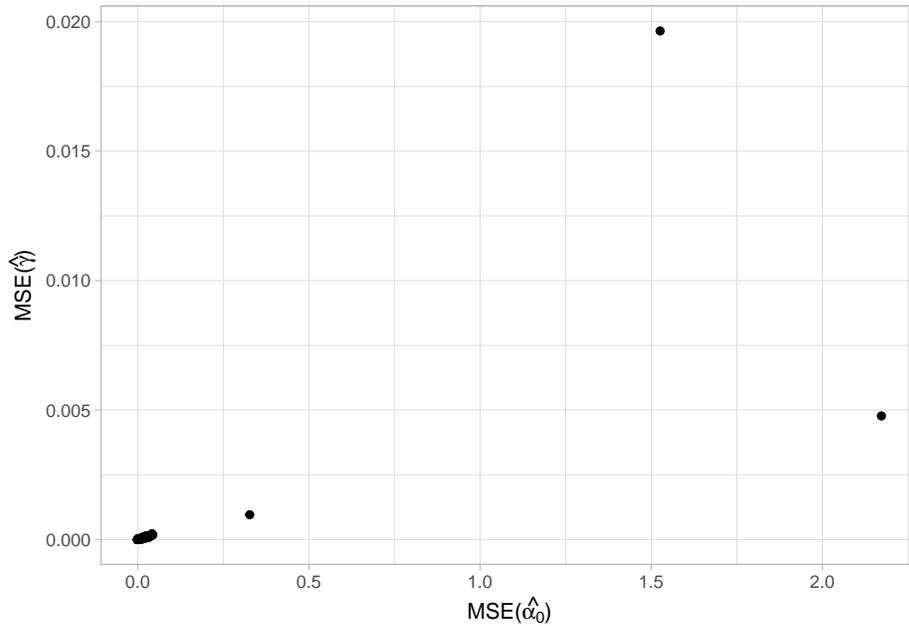


Figure 25: $\text{MSE}(\hat{\alpha}_0)$ vs. $\text{MSE}(\hat{\gamma})$

Figure 25 illustrates the correlation between $\text{MSE}(\hat{\gamma})$ and $\text{MSE}(\hat{\alpha}_0)$ where $\alpha_0 > 2.5$. As $\text{MSE}(\hat{\gamma})$ increases, $\text{MSE}(\hat{\alpha}_0)$ also tends to increase. We can call this correlation as a fairly strong positive relationship by looking at the graph and the correlation coefficient value.

Figure 26 has same interpretation as Figure 25. However, Figure 27 indicates

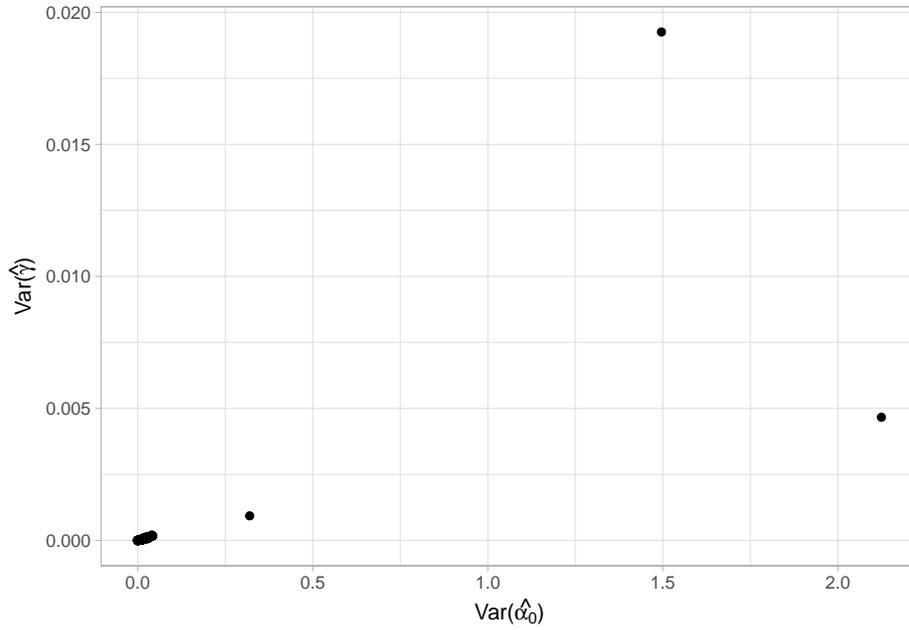


Figure 26: $Var(\hat{\alpha}_0)$ vs. $Var(\hat{\gamma})$

a different result. As $Bias(\hat{\gamma})$ decreases, $Bias(\hat{\alpha}_0)$ tends to increase. We can call this correlation as a fairly strong negative relationship.

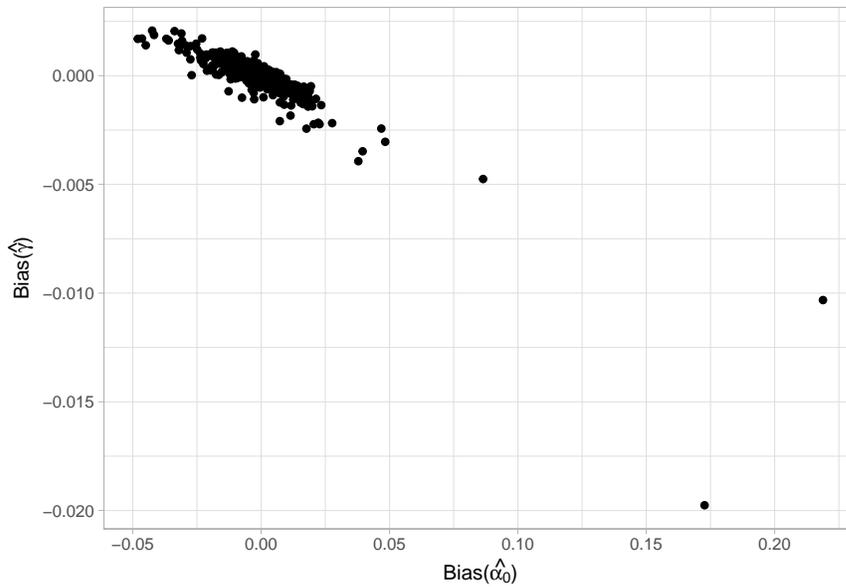


Figure 27: $Bias(\hat{\alpha}_0)$ vs. $Bias(\hat{\gamma})$

Overall, our results from meta-analysis experiment are statistically better than one classroom case.

V. APPLICATION

Data Description

The National Science Foundation funded project (DRL-1814114), *Using Technology to Capture Classroom Interactions: The Design, Validation, and Dissemination of a Formative Assessment of Instruction Tool for Diverse K-8 Mathematics Classrooms*, led by Dr. Kathleen Melhuish is developing a formative assessment tool called the Math Habits Tool (MHT). The MHT is intended to capture patterns of in-the-moment teacher-student and student-student classroom interactions. As part of this work, the team has coded video observations for teaching moves that can serve to catalyze student productive engagement with mathematics (catalytic teaching habits) and productive ways students engage in mathematics. Teachers actions consist of teacher routines and catalytic teaching habits which are classified as either high or low. Students responses are coded as students habit of interaction and student habits of mind. There are 24 independent mathematics classrooms with different teachers. We have coded all teachers moves separately, but we coded students moves without differentiating them.

Parameter Estimation

In this chapter, we will use the previously described method to analyze the data from 24 independent mathematics classrooms. We will estimate the parameters of the log-linear rate function of Student Processes as well as the constant rate of Teacher Process. For this purpose, we will start by analyzing one classroom data and then apply the meta-analysis methods to estimate $\hat{\mu}$, $\hat{\alpha}_0$ and the $\hat{\gamma}$ values and compute asymptotic variance. Our primary interest is to estimate the catalytic constant γ to assess the effect of the catalytic teaching habits on the observed examples of productive student engagement.

In this thesis, we have developed three models. The first model is coded by including all the teacher moves as a teacher process. It means that we did not differentiate teacher moves. We coded teacher routine, catalytic teaching habit high, and catalytic teaching habit low as teacher moves and student actions by combining student's habit of interaction and student's habit of mind.

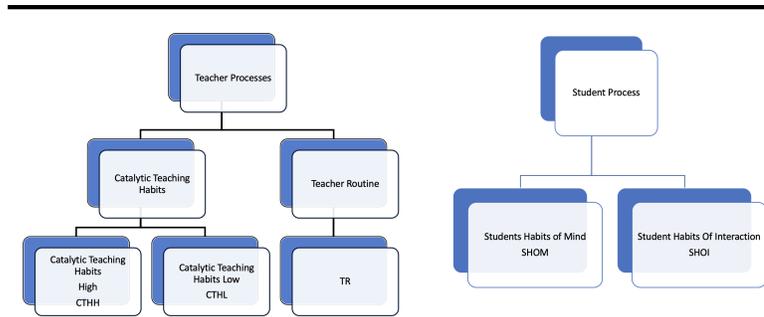


Figure 28: First Model (All teacher moves)

In our second model, we included only catalytic teaching habit low moves as Teacher Process and kept the Student Process the same as the first model.



Figure 29: Second Model (Low teacher moves)

In our third model, we included only catalytic teaching habit high moves as Teacher Process and kept the Student Process the same as the first model and the second model.

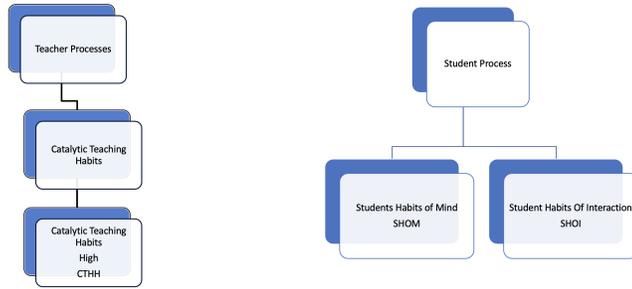


Figure 30: Third Model (High teacher moves)

First of all, we have prepared the data frame that we can use in our models. In our data frame, we have enumerated 24 independent classrooms, such as the first classroom, second classroom, and third classroom. We randomly have chosen the fifth classroom and estimate μ , α_0 and γ values for one classroom analysis. Also, we have computed the variance of estimations. To obtain these results, we have used R programming software. It provided us many numerical methods to maximize our log-likelihood function and find the maximum likelihood estimators of the parameters. The complete R code is included in Appendix C

The following table is the summary of the fifth classroom by using the first model. As we see from Table 12, $\hat{\alpha}_0$ and $\hat{\gamma}$ are negative. It is acceptable for $\hat{\alpha}_0$ to be less than 0, but $\hat{\gamma}$ being less than zero implies that teacher moves are not helping the students.

Table 12: The First Model (All Teacher Moves) Results For One Classroom

$\hat{\mu}$	$\hat{\alpha}_0$	$\hat{\gamma}$	$sd(\hat{\alpha}_0)$	$sd(\hat{\gamma})$
0.5	-1.429146	-0.05644022	0.6225259	0.05033712

The following table is the summary of the fifth classroom by applying the second model.

Table 13: The Second Model (Low Moves) Results for One Classroom

$\hat{\mu}$	$\hat{\alpha}_0$	$\hat{\gamma}$	$sd(\hat{\alpha}_0)$	$sd(\hat{\gamma})$
0.1481481	-1.699543	-0.1130635	0.590564	0.1449812

The next table is the summary of the fifth classroom by applying the third model.

Table 14: The Third Model (High Moves) Results for One Classroom

$\hat{\mu}$	$\hat{\alpha}_0$	$\hat{\gamma}$	sd($\hat{\alpha}_0$)	sd($\hat{\gamma}$)
0.1296296	-1.384556	-0.2081086	0.5235135	0.1679655

As we can see from the Table 12, the Table 13, and the Table 14, all $\hat{\gamma}$ values are negative in three models for fifth classroom that we randomly selected. For sensitivity analysis, we also ran these three models for a couple of classrooms individually. We have found that sometimes $\hat{\gamma}$ values are greater than 0, but sometimes they were slightly less than 0. We know from Chapter 3 that single classroom data did not allow us to use asymptotic results. Hence, we have applied the meta-analysis method that has explained in Chapter 4. We combined 24 classrooms to perform meta-analysis experiment. We have used the same codes that are used in Chapter 4 for the log-likelihood function and the Fisher information matrix. We have derived $\hat{\mu}$, $\hat{\alpha}_0$, $\hat{\gamma}$, and their standard deviation from the fisher information theory. We have used same three models that are used with individual classrooms. The followings are results of the first model:

Table 15: The First Model (All Teacher Moves) Results For 24 Classrooms

$\hat{\mu}$	$\hat{\alpha}_0$	$\hat{\gamma}$	sd($\hat{\alpha}_0$)	sd($\hat{\gamma}$)
0.609834	-1.1817	0.01400656	0.066822	0.002262891

The Table 15 illustrates that we have a higher $\hat{\gamma}$ value compared to the Table 12. Also we have much smaller standard deviations compared to the Table 12.

The following table presents the results of the second model for 24 classrooms. When we compare the Table 13 and the Table 16, we see that we have a higher $\hat{\gamma}$ value and smaller standard deviations for $\hat{\gamma}$ and $\hat{\alpha}_0$ in the meta-analysis experiment compared to one classroom analysis.

Table 16: The Second Model (Low Teacher Moves) Results for 24 Classrooms

$\hat{\mu}$	$\hat{\alpha}_0$	$\hat{\gamma}$	sd($\hat{\alpha}_0$)	sd($\hat{\gamma}$)
0.2745849	-1.137404	0.02692045	0.06355882	0.004541115

The new table illustrates the results of the third model for 24 classrooms. The Table 17 depicts that we have a positive $\hat{\gamma}$ value and higher $\hat{\alpha}_0$ value compared to the Table 14. They have both show the statistics summary of the third model. Also we have much smaller standard deviations in the Table 17.

Table 17: The Third Model (High Moves) Results For 24 Classrooms

$\hat{\mu}$	$\hat{\alpha}_0$	$\hat{\gamma}$	sd($\hat{\alpha}_0$)	sd($\hat{\gamma}$)
0.2418582	-1.12592	0.02948545	0.06255076	0.005125693

As we mentioned before, our main interest is to estimate catalytic constant gamma. Let us present the new figure that explains all these six models.

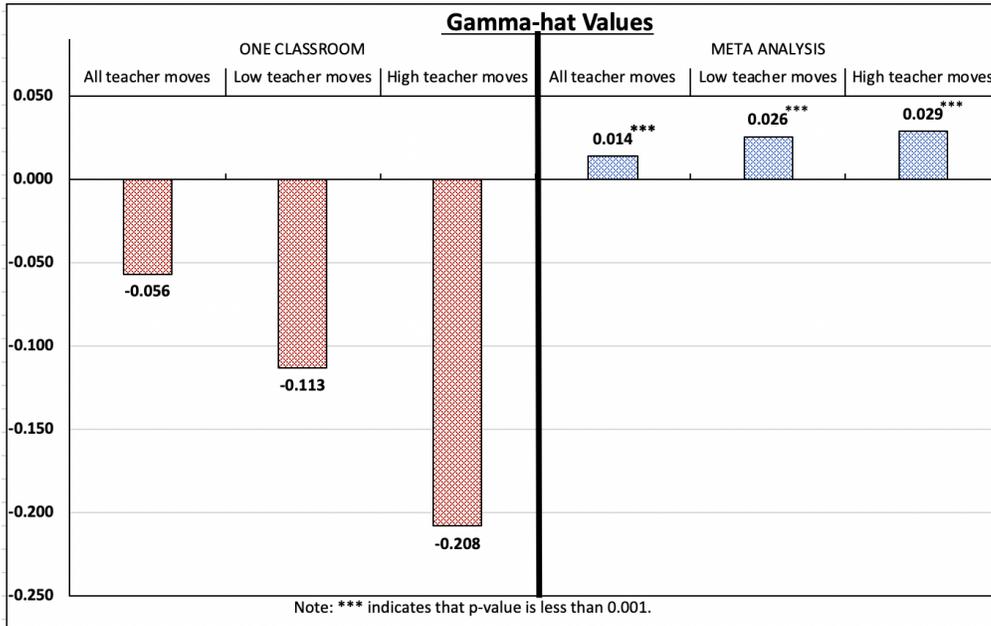


Figure 31: All six models together

We see from Figure 31 that the results of the models that used the meta-analysis experiment demonstrate higher $\hat{\gamma}$ values. We also tested each $\hat{\gamma}$ values from the models that used meta-analysis method to see if these $\hat{\gamma}$ are

significantly greater than 0. As we see from the Figure 31, p-values(< 0.001) for $\hat{\gamma}$ values in three meta-analysis models indicate that $\hat{\gamma}$ values are significantly greater than 0.

If we compare the results among three meta-analysis models, we see that first model has the highest $\hat{\mu}$ value. We anticipated this because first model has the highest teacher moves. We have reduced the number of teacher moves in the second and the third models. Since our primary interest is to estimate gamma, we would like to compare these three models that used meta-analysis experiment results in terms of $\hat{\gamma}$ values. We see that third model(high teacher moves) has the highest $\hat{\gamma}$ value. We also wanted to investigate and the visualize the student rate versus classroom time. Since second model and third model have the higher $\hat{\gamma}$ values compare to the first model for 24 classrooms models, we wanted to compare the effect of high and low moves on the student rate in the mathematics classrooms. We have calculated the expected value of the student rate

$$\lambda_t = e^{\alpha_0 + \gamma V_t}$$

by using (13), we have computed the following expected value.

$$E(\lambda_t) = e^{\alpha_0} e^{uT(e^\gamma - 1)}$$

The Figure 32 presents the expected value of the student rate versus time for second and third models that employed meta-analysis method.

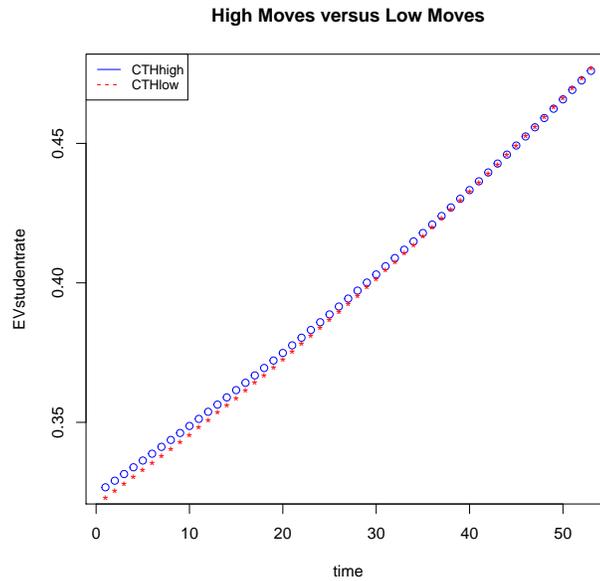


Figure 32: Expected Value of the Student Rate Function versus Time

As the Figure 32 illustrates, we have obtained the higher student rate with CTHH moves compared to CTHL until some time. However, after some times, the lower teacher moves slightly lead to a higher student rate. We are not able to explain the reason now, but we will look at the different characteristic of the teacher moves in the future research. Overall, our newly developed basic model works to explain in-the-moment interaction between teacher and student by taking the time into account. We hypothesized to see whether catalytic gamma value is greater than 0, and productive, meaningful pedagogical teacher actions lead to a higher student rate. We found that $\hat{\gamma}$ values from three meta-analysis models are significantly greater than 0. However, we could not explain why higher moves do not lead to a higher student rate in mathematics classrooms all the time, but we hope to explain it in the future research.

VI. FUTURE RESEARCH

- Use the second model with more classrooms and observe the value of the parameters.
- Add a set of predictor variables x_1, \dots, x_k which take on possibly different values in each classroom x_{k_j} with $j = 1, \dots, P$ classrooms. Adjust definition of intensity rate to take predictor variables into account:

$$\log(\lambda_t) = \alpha_0 + \beta_1 x_{1_j} + \dots + \beta_k x_{k_j} + \gamma V_t$$

- Adjust likelihood from previous section and use it to write the maximum likelihood for α_0 , β and γ . Modify the R code to find maximum likelihood and test it out.
- Power analysis should perform to determine how large P needs to be to be able to estimate parameters in simplest case which applies to the real world data.
- Create a different log linear rate function such as exponential polynomial rate function and investigate the range of ϕ

$$\lambda_t = e^{\alpha_0 + \gamma V_t^\phi}$$

or

$$\log(\lambda_t) = \alpha_0 + \gamma \log(V_t + 1).$$

APPENDIX SECTION

APPENDIX A

```
set.seed(4000)

library(tidyverse)

sim_aw <- function(time,mu) {

  path <- matrix(0, nrow = 1, ncol = 2)

  jumps_number <- rpois(1, lambda = mu * time)
  jumps_time <- runif(n = jumps_number, min = 0, max = time) %>% sort()

  jumps_time

}

two.pois.sim2 <- function(time, mu, alpha0, gamma){

  tset <- sim_aw(time, mu)
  M <- length(tset)
  w<-rep(0,M+2)
  if(M>0){
    u.diff = c(tset, time)- c(0, tset)
    w[2:(M+2)] = exp(alpha0)*cumsum(exp(gamma*0:M)*u.diff)
    Ftinv<- function(u){
      k = findInterval(u, w/w[M+2])+1
      (w[M+2]*u- w[k-1])*exp(-alpha0-gamma*(k-2))+c(0, tset)[k-1]
    }
  } else{
```

```

tset = NA
Ftinv = function(u){time *u}
w[2] = time*exp(alpha0)
  }

N=rpois(1, w[M+2])
if(N>0){
X0=rep(NA,N)
for(i in 1:N){
X0[i]=Ftinv(runif(1))
  }
}else{
  X0= NA
}

list(t.times = tset, s.times = if(is.numeric(X0)){sort(X0)}else{NA})

}

library(rootSolve)
mle.sim.gamma<- function(gamma, n.sim =50,time= 50, mu=0.5, alpha0=0){
sim.res = matrix(0, nrow = n.sim, ncol = 9)
colnames(sim.res)<- c("time", "mu", "alpha0",
"gamma", "M", "N", "r", "alpha0hat", "gammahat")
  param <- c(time, mu, alpha0, gamma)
  for(i in 1:n.sim){

```

```

two.pois.sim2(time, mu, alpha0, gamma)-> res
if(is.numeric(res$t.times)&is.numeric(res$s.times) ){
r= sum(findInterval(res$s.times, c(0, res$t.times, 50))-1)
N= length(res$s.times)
M = length(res$t.times)
u.diff = c(res$t.times, time)- c(0, res$t.times)
model=function(par){
l.prime.ad=(-exp(par[1])*sum(exp(par[2]*0:M)*u.diff) +N)
l.prime.gd=(-exp(par[1])*sum((0:M)*exp(par[2]*0:M)*u.diff) +r)
c(l.prime.ad=l.prime.ad,l.prime.gd=l.prime.gd) }
p3<- multiroot(model,c(-1, 1),maxiter=1000000)$root[1]
p5<-multiroot(model,c(-1, 1),maxiter=1000000)$root[2]

sim.res[i,] = c(param, c(M, N, r,p3,p5))}else{
sim.res[i,] = c(param,
length(res$t.times),
length(res$s.times),
rep(NA, 3))
}
}

return(sim.res)

}

mle.sim.gamma(0.3,50,50,0.5,0)
sim.multi <- mle.sim.gamma(0,50,50,0.5,0)
for( i in seq(from=0, to=0.4, by=0.01)){ for( j in seq(from=-3, to=0,
by=0.1)){

```

```

sim.multi= rbind(sim.multi,mle.sim.gamma(i,50,50,0.5,j) )

}

}

modvar<- function ( p3,p5,time,M)
{
EValphaa<- (exp(p3)*((1-exp(p5))^-1)*time*(M)^-1*
(1-exp(M)^(exp(p5)-1)))
EValphaandgam<- (exp(p3)*time*exp(p5)*(1-exp(p5))^-2
*(((1-exp(-M))*(M)^-1)
-(exp(-M)*exp(M*exp(p5))))+exp(-M)*(M)^-1*
(exp(exp(p5)*M)*(exp(p5)*M-1)+1)))
EVgamma<- exp(p3)*time*exp(p5)*(1-exp(p5))^-3*(
(1+exp(p5))*(1-exp(-M))*(M)^-1-
exp(-M)*exp(exp(p5)*M)*(exp(p5)*M+1)+(2-exp(p5))*exp(-M)*M^-1*(
exp(exp(p5)*M)*(1-exp(p5)*M+(exp(p5)*M)^2)-1
)+ 2*exp(-M)*M^-1*(
exp(exp(p5)*M)*(exp(p5)*M-1)+1
)-exp(-M)*(M)^-1*(exp(exp(p5)*M)-1))
return(matrix(c(EValphaa, EValphaandgam,
EValphaandgam, EVgamma),nrow=2,ncol=2))
}

modvar(p3,p5,time,M)

my.mean = function(x){mean(x, na.rm = TRUE)}

zeljun11<-as.data.frame(sim.multi)%>%mutate(g.diff = gammahat - gamma,
a.diff = alpha0hat-alpha0)%>%
group_by(gamma,alpha0)%>%

```

```

summarise(cnt = sum(!is.na(gammahat)), g.bias = my.mean(g.diff),
a.bias = my.mean(a.diff), g.mse = my.mean(g.diff^2) ,
a.mse= my.mean(a.diff^2) )

```

APPENDIX B

```

library(tidyverse)
sim_aw <- function(time,mu) {

path <- matrix(0, nrow = 1, ncol = 2)

jumps_number <- rpois(1, lambd = mu * time)
jumps_time <- runif(n = jumps_number, min = 0, max = time) %>% sort()

jumps_time

}

two.pois.sim2 <- function(time, mu, alpha0, gamma){

tset <- sim_aw(time, mu)
M <- length(tset)
w<-rep(0,M+2)
if(M>0){
u.diff = c(tset, time)- c(0, tset)
w[2:(M+2)] = exp(alpha0)*cumsum(exp(gamma*0:M)*u.diff)
Ftinv<- function(u){
k = findInterval(u, w/w[M+2])+1
(w[M+2]*u- w[k-1])*exp(-alpha0-gamma*(k-2))+c(0, tset)[k-1]
}
}

```

```

} else{
tset = NA
Ftinv = function(u){time *u}
w[2] = time*exp(alpha0)
}

N=rpois(1, w[M+2])
if(N>0){
X0=rep(NA,N)
for(i in 1:N){
X0[i]=Ftinv(runif(1))
}
}else{
X0= NA
}

list(t.times = tset, s.times = if(is.numeric(X0)){sort(X0)}else{NA})

}

library(rootSolve)
mle.sim.gamma<- function(gamma, n.sim =50,time= 50, mu=0.5, alpha0=0){
sim.res = matrix(0, nrow = n.sim, ncol = 6)
colnames(sim.res)<- c(
"time", "mu", "alpha0"

```

```

"gamma","alpha0hat","gammahat")
param <- c(time, mu, alpha0, gamma)
u.diff = list(0)
r = rep(0, 20)
N= rep(0, 20)
M = rep(0, 20)
for(i in 1:n.sim){for( p in seq(from=1, to=20, by=1)){
two.pois.sim2(time, mu, alpha0, gamma)-> res
if(is.numeric(res$t.times)&is.numeric(res$s.times) ){
r[p]= sum(findInterval(res$s.times, c(0, res$t.times, 50))-1)
N[p]= length(res$s.times)
M[p] = length(res$t.times)
u.diff[[p]] = c(res$t.times, time)- c(0, res$t.times)
}else{
r[p] = NA
}

}

if(sum(is.na(r))<1){
model=function(par){
l.prime.adp = 0
l.prime.gdp = 0
for (p in 1:20){l.prime.adp=(-
exp(par[1])*sum(exp(par[2]*0:M[p])*u.diff[[p]])+N[p]) +
l.prime.adp l.prime.gdp=(-
exp(par[1])*sum((0:M[p])*exp(par[2]*0:M[p]) *u.diff[[p]])
+r[p]) + l.prime.gdp }
}
}

```

```

c(l.prime.adp=l.prime.adp,l.prime.gdp=l.prime.gdp) }
p3<- multiroot(model,c(-2,2))$root[1]
p5<-multiroot(model,c(-2,2))$root[2]

sim.res[i,] = c(param, c(p3,p5)) } else {
sim.res[i,] = c(param, rep(NA,2))
}

}

return(sim.res)
}

sim.multi <- mle.sim.gamma(0,50,50,0.5,0)
for( i in seq(from=0, to=0.4, by=0.01))
{ for( j in seq(from=-3, to=1, by=0.1)){
sim.multi= rbind(sim.multi,mle.sim.gamma(i,50,50,0.5,j) )

}

}

summultimeta<-as.data.frame(sim.multi)
my.mean = function(x){mean(x, na.rm = TRUE)}
zelmeta<-as.data.frame(sim.multi)%>%mutate(g.diff = gammahat - gamma,
a.diff = alpha0hat-alpha0)%>%
group_by(gamma,alpha0)%>%
summarise(cnt = sum(!is.na(gamma.hat)), g.bias = my.mean(g.diff),
a.bias = my.mean(a.diff), g.mse = my.mean(g.diff^2) ,
a.mse= my.mean(a.diff^2) )

```

View(zelmeta)

APPENDIX C

All Teacher Moves

```
library(tidyverse)
library(rootSolve)
for (i in 1:24){
  assign(paste0("s.times", i), (dataframe%>%
mutate(time=(Minutes+Seconds/60))
%>%filter(LessonCode == i, category ==1 ))$time)
  assign(paste0("t.times", i),
(dataframe%>%mutate(time =
(Minutes+Seconds/60))
%>%filter(LessonCode == i, category >1 ))$time)
}

Time<-list(58.8333333,63.6,53.25,59.28333333,54,44.8,64.33333333,
44.51666667,40.4666667,71.88333333,74,57.03333333,47.56666667,
60,56.06666667,
43.15,66.83333333,46.85,41.06666667,
44.23333333,43.0333333,
41.15,42,34.85)
myteacher<-list(t.times1,t.times2,t.times3,t.times4,t.times5,
t.times6,
t.times7,t.times8,t.times9,t.times10,t.times11,t.times12,t.times13,
t.times14,t.times15,t.times16,t.times17,t.times18,
t.times19,t.times20,t.times21,
```

```

t.times22,
t.times23,t.times24)
mystudent<-list(s.times1,s.times2,s.times3,s.times4,s.times5,s.times6,
s.times7,s.times8,s.times9,s.times10,s.times11,s.times12,
s.times13,s.times14,s.times15,s.times16,s.times17,s.times18,
s.times19,
s.times20,s.times21,s.times22,s.times23,s.times24)

N=NULL

u.diff=NULL

M=NULL

r=NULL

for(i in 1:24){
N[i]=length(mystudent[[i]])
M[i]=length(myteacher[[i]])
u.diff[[i]] = c(myteacher[[i]], Time[[i]])- c(0, myteacher[[i]])
r[i]= sum(findInterval(mystudent[[i]],
c(0, myteacher[[i]], Time[[i]]))-1)
}

model=function(par){
l.prime.adp = 0
l.prime.gdp = 0
for (p in 1:24){
l.prime.adp=(-exp(par[1])*sum(exp(par[2]*0:M[p])*
u.diff[[p]])
+N[p]) + l.prime.adp
l.prime.gdp=(-exp(par[1])*sum((0:M[p])*exp(par[2]*
0:M[p])*u.diff[[p]]))

```

```

+r[p]) + 1.prime.gdp }
c(1.prime.adp=1.prime.adp,1.prime.gdp=1.prime.gdp) }
p3<- multiroot(model,c(-1,0))$root[1]
p5<-multiroot(model,c(-1,0))$root[2]
Tim=c(58.83333333,63.6,53.25,59.28333333,54,44.8,64.33333333,
44.51666667,
40.46666667,71.88333333,74,57.03333333,47.56666667,60,
56.06666667,43.15,
66.83333333,46.85,41.06666667,44.23333333,43.03333333,41.15
,42,34.85)
modvar <- function ( p3,p5,Tim,M)
{
EValphaa<- sum((exp(p3)*((1-exp(p5))^-1)*Tim*(M)^-1*
(1-exp(M)^(exp(p5)-1))))
EValphaandgam<- sum((exp(p3)*Tim*exp(p5)*(1-exp(p5))^-2*(((1-exp(-M))*
(M)^-1)-
(exp(-M)*exp(M*exp(p5))))+exp(-M)*(M)^-1*(exp(exp(p5)*M)*
(exp(p5)*M-1)+1))))
EVgamma<- sum( exp(p3)*Tim*exp(p5)*(1-exp(p5))^-3*(
(1+exp(p5))*(1-exp(-M))*(M)^-1-
exp(-M)*exp(exp(p5)*M)*(exp(p5)*M+1)+(2-exp(p5))*exp(-M)*M^-1*(
exp(exp(p5)*M)*(1-exp(p5)*M+(exp(p5)*M)^2)-1
)+ 2*exp(-M)*M^-1*(
exp(exp(p5)*M)*(exp(p5)*M-1)+1
)-exp(-M)*(M)^-1*(exp(exp(p5)*M)-1)))
return(matrix(c(EValphaa, EValphaandgam, EValphaandgam,
EVgamma),nrow=2,ncol=2))

```

```

}

sum(M)/sum(Tim)

mod1=solve(modvar(p3,p5,Tim,M))
mod1
High Moves
library(tidyverse)
library(rootSolve)
for (i in 1:24){
  assign(paste0("s.times", i), (dataframe
  %>%mutate(time = (Minutes+Seconds/60))
  %>%filter(LessonCode == i, category ==1 ))$time)
  assign(paste0("t.times", i), (dataframe
  %>%mutate(time = (Minutes+Seconds/60))
  %>%filter(LessonCode == i, category ==4 ))$time)

}

Time<-list(58.83333333,63.6,53.25,59.28333333,54,44.8,64.33333333,
44.51666667,40.46666667,71.88333333,74,57.03333333,47.56666667,
60,56.06666667,43.15,66.83333333,46.85
,41.06666667,44.23333333,
43.03333333,41.15,42,34.85)
myteacher<-list(t.times1,t.times2,t.times3,t.times4,t.times5,t.times6,
t.times7,t.times8,t.times9,t.times10,t.times11,t.times12,t.times13,
t.times14,t.times15,t.times16,t.times17,
t.times18,t.times19,t.times20,

```

```

t.times21,t.times22,t.times23,t.times24)
mystudent<-list(s.times1,s.times2,s.times3,s.times4,s.times5,s.times6,
s.times7,s.times8,s.times9,s.times10,s.times11,s.times12,s.times13,
s.times14,s.times15,s.times16,s.times17,s.times18,s.times19,
s.times20,
s.times21,s.times22,s.times23,s.times24)
N=NULL
u.diff=NULL
M=NULL
r=NULL
for(i in 1:24){
N[i]=length(mystudent[[i]])
M[i]=length(myteacher[[i]])
u.diff[[i]] = c(myteacher[[i]],
Time[[i]])- c(0, myteacher[[i]])
r[i]= sum(findInterval(mystudent[[i]],
c(0, myteacher[[i]], Time[[i]]))-1)
}
model=function(par){
l.prime.adp = 0
l.prime.gdp = 0
for (p in 1:24){
l.prime.adp=(-exp(par[1])*sum(exp(par[2]*0:M[p])
*u.diff[[p]])
+N[p]) + l.prime.adp
l.prime.gdp=(-exp(par[1])*sum((0:M[p])*exp(par[2]*0:M[p])
*u.diff[[p]]))

```

```

+r[p]) + 1.prime.gdp }
c(1.prime.adp=1.prime.adp,1.prime.gdp=1.prime.gdp) }
p3<- multiroot(model,c(-1,0))$root[1]
p5<-multiroot(model,c(-1,0))$root[2]
Tim=c(58.83333333,63.6,53.25,59.28333333,54,
44.8,64.33333333,44.51666667,
40.46666667,71.88333333,74,57.03333333,
47.56666667,60,56.06666667,43.15,
66.83333333,46.85,41.06666667,44.23333333,
43.03333333,41.15,42,
34.85)
modvar <- function ( p3,p5,Tim,M)
{
EValphaa<- sum((exp(p3)*((1-exp(p5))^-1)*Tim*(M)^-1*
(1-exp(M)^(exp(p5)-1))))
EValphaandgam<- sum((exp(p3)*Tim*exp(p5)*(1-exp(p5))^-2*
(((1-exp(-M))*M^-1)-(exp(-M)*exp(M*exp(p5))))+exp(-M)*
(M)^-1*(exp(exp(p5)*M)*(exp(p5)*M-1)+1))))
EVgamma<- sum( exp(p3)*Tim*exp(p5)*(1-exp(p5))^-3*(
(1+exp(p5))*(1-exp(-M))*(M)^-1-
exp(-M)*exp(exp(p5)*M)*(exp(p5)*M+1)+(2-exp(p5))*exp(-M)*M^-1*(
exp(exp(p5)*M)*(1-exp(p5)*M+(exp(p5)*M)^2)-1
)+ 2*exp(-M)*M^-1*(
exp(exp(p5)*M)*(exp(p5)*M-1)+1
)-exp(-M)*(M)^-1*(exp(exp(p5)*M)-1)))
return(matrix(c(EValphaa, EValphaandgam, EValphaandgam,
EVgamma),nrow=2,ncol=2))

```

```

}
sum(M)/sum(Tim)*
Low Moves
mod1=solve(modvar(p3,p5,Tim,M))
mod1
library(tidyverse)
library(rootSolve)
for (i in 1:24){
assign(paste0("s.times", i), (dataframe%>%mutate(time =
(Minutes+Seconds/60))%>%filter(LessonCode == i, category ==1 ))$time)
assign(paste0("t.times", i), (dataframe%>%
mutate(time = (Minutes+Seconds/60))%>%
filter(LessonCode == i, category==3 ))$time)
}

Time<-list(58.83333333,63.6,53.25,59.28333333,54,44.8,64.33333333,
44.51666667,40.46666667,71.88333333,74,57.03333333,
47.56666667,60,56.06666667,43.15,66.83333333,
46.85,41.06666667,44.23333333,43.03333333,
41.15,42,34.85)
myteacher<-list(t.times1,t.times2,t.times3,t.times4,t.times5,t.times6,
t.times7,t.times8,t.times9,t.times10,
t.times11,t.times12,t.times13,t.times14,
t.times15,t.times16,t.times17,t.times18,
t.times19,t.times20,t.times21,t.times22,

```

```

t.times23,t.times24)
mystudent<-list(s.times1,s.times2,s.times3,s.times4,s.times5,s.times6,
s.times7,s.times8,s.times9,s.times10,s.times11,s.times12,
s.times13,s.times14,s.times15,s.times16,s.times17,s.times18,
s.times19,s.times20,s.times21,s.times22,
s.times23,s.times24)

N=NULL

u.diff=NULL

M=NULL

r=NULL

for(i in 1:24){
N[i]=length(mystudent[[i]])
M[i]=length(myteacher[[i]])
u.diff[[i]] = c(myteacher[[i]], Time[[i]])- c(0, myteacher[[i]])
r[i]= sum(findInterval(mystudent[[i]],
c(0, myteacher[[i]], Time[[i]]))-1)
}

model=function(par){
l.prime.adp = 0
l.prime.gdp = 0
for (p in 1:24){ l.prime.adp=(-exp(par[1])*sum(exp(par[2]*0:M[p])*
u.diff[[p]]) +N[p]) + l.prime.adp
l.prime.gdp=(-exp(par[1])*sum((0:M[p])*exp(par[2]*
0:M[p])*u.diff[[p]]) +r[p]) + l.prime.gdp }
c(l.prime.adp=l.prime.adp,
l.prime.gdp=l.prime.gdp) }
p3<- multiroot(model,c(-1,0))$root[1]

```

```

p5<-multiroot(model,c(-1,0))$root[2]
Tim=c(58.8333333,63.6,53.25,59.28333333,54,44.8,64.33333333,
44.51666667,
40.4666667,71.88333333,74,57.03333333,47.56666667,60,56.06666667,
43.15,
66.83333333,46.85,41.06666667,44.23333333,
43.0333333,41.15,42,34.85)
modvar <- function ( p3,p5,Tim,M)
{
EValphaa<- sum((exp(p3)*((1-exp(p5))^-1)*Tim*(M)^-1*
(1-exp(M)^(exp(p5)-1))))
EValphaandgam<- sum((exp(p3)*Tim*exp(p5)*(1-exp(p5))^-2*(((1-exp(-M))*
(M)^-1)-(exp(-M)*exp(M*exp(p5))))+exp(-M)*
(M)^-1*(exp(exp(p5)*M)*(exp(p5)*M-1)+1))))
EVgamma<- sum( exp(p3)*Tim*exp(p5)*(1-exp(p5))^-3*(
(1+exp(p5))*(1-exp(-M))*(M)^-1-
exp(-M)*exp(exp(p5)*M)*(exp(p5)*M+1)+(2-exp(p5))*exp(-M)*M^-1*(
exp(exp(p5)*M)*(1-exp(p5)*M+(exp(p5)*M)^2)-1
)+ 2*exp(-M)*M^-1*(
exp(exp(p5)*M)*(exp(p5)*M-1)+1
)-exp(-M)*(M)^-1*(exp(exp(p5)*M)-1)))
return(matrix(c(EValphaa, EValphaandgam, EValphaandgam, EVgamma),
nrow=2,ncol=2))
}
sum(M)/sum(Tim)
mod1=solve(modvar(p3,p5,Tim,M))
mod1

```

REFERENCES

- Boaler, J. and Staples, M. (2008). Creating Mathematical Futures through an Equitable Teaching Approach: The Case of Railside School. *Teachers College Record*, 110(3).
- Boyadzhiev, K. N. (2009). Exponential Polynomials, Stirling Numbers, and Evaluation of Some Gamma Integrals. *Abstract and Applied Analysis*.
- Chrysaphinou, O. (1985). On Touchard Polynomials. *Discrete mathematics*, 54(2).
- Cinlar, E. (1975). *Introduction to Stochastic Processes*. Englewood Cliffs Prentice-Hall.
- Cox, D. R. and Lewis, P. A. W. (1966). *The Statistical Analysis of Series of Events*. Chapman and Hall.
- Dobrow, R. P. (2016). *Introduction to Stochastic Processes with R*. Wiley.
- Hogg, R. V., McKean, J., and Craig, A. T. (2005). *Introduction to Mathematical Statistics*. Pearson Education.
- Jacobs, V. and Spangler, D. (2017). Research on core practices in k-12 mathematics teaching. In Cai, J., editor, *Compendium for Research in Mathematics Education*, chapter 28, pages 766–792. National Council of Teachers of Mathematics, Reston, VA.
- Melhuish, K., Thanheiser, E., Heaton, R., Sorto, M., Strickland, S., and Sugimoto, A. (2020). The math habits tool - research version [measurement instrument]. <http://mathhabits.wp.txstate.edu>.
- Pasupathy, R. (2011). Generating Nonhomogeneous Poisson Processes. *Wiley encyclopedia of operations research and management science*.

Sorto, M. A., Melhuish, K., Thanheiser, E., Zied, K., Koehne, C., Sugimoto, A., Pham, A., Han, S. B., and Strickland, S. (2019). Components of High-quality Mathematics Classrooms: Attending to Learning Opportunities for English Language Learners. In Otten, S., Candela, A. G., de Araujo, Z., Haines, C., and Munter, C., editors, *Proceedings of the forty-first annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*, pages 1594 – 1603, St Louis, MO: University of Missouri.

Sykora, S. (2006). Finite and infinite sums of the power series $(k^p)(x^k)$. <http://www.ebyte.it/library/docs/math06b/PowerSeriesKpXk.html>.

Webb, N. M., Franke, M. L., Ing, M., Wong, J., Fernandez, C. H., Shin, N., and Turrou, A. C. (2014). Engaging with Others' Mathematical Ideas: Interrelationships Among Student Participation, Teachers' Instructional Practices, and Learning. *International Journal of Educational Research*.

Weisstein, E. (2002). Power sum. <https://mathworld.wolfram.com/PowerSum.html>.