APPLICATIONS OF POINT PROCESS MODELS TO IMAGING AND BIOLOGY

by

Burcin Simsek

BS, Ankara University, 2008

MS, Akdeniz University, 2010

MA, University of Pittsburgh, 2012

Submitted to the Graduate Faculty of

the Department of Statistics, Dietrich Graduate School of Arts and

Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2016

UNIVERSITY OF PITTSBURGH

DIETRICH GRADUATE SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Burcin Simsek

It was defended on

November 16, 2015

and approved by

Satish Iyengar, Department of Statistics

Henry Block, Department of Statistics

Kehui Chen, Department of Statistics

John Chadam, Department of Mathematics

Dissertation Director: Satish Iyengar, Department of Statistics

APPLICATIONS OF POINT PROCESS MODELS TO IMAGING AND BIOLOGY

Burcin Simsek, PhD

University of Pittsburgh, 2016

This dissertation deals with point process models and their applications to imaging and messenger RNA (mRNA) transcription. We address three problems. The first problem arises in two-photon laser scanning microscopy. We model the process by which photons are counted by a detector which suffers from a dead period upon registration of a photon. In this model, we assume that there are a Poisson (α) number of excited molecules, with exponentially distributed waiting times for the emissions of photons. We derive the exact distribution of all observed counts, rather than grouped counts which were used earlier. We use it to get improved estimates of the Poisson intensity, which leads to images with higher signal-to-noise ratio. This improvement is because grouping of count data results in loss of information. We illustrate this improvement on imaging data of paper fibers. Next, we study two variants of this model: the first uses a finite time horizon and the second considers gamma waiting times for the emissions.

The second problem concerns the Conway-Maxwell-Poisson distribution for count data. This family has been proposed as a generalization of the Poisson for handling overdispersion and underdisperson. Because the normalizing constant of this family is hard to compute, good approximations for it are needed. We provide a statistical approach to derive an existing approximation more simply. However, this approximation does not perform well across all the parameter ranges. Therefore, we introduce correction terms to improve its performance. For other parts of the parameter space, we use the geometric and Bernoulli distributions, with correction terms based on Taylor expansions. Using numerical examples, we show that

our approximations are much better than earlier proposed methods.

In the last problem, we present a new application for Conway-Maxwell-Poisson family. We use the generalized linear model setting of this family to study mRNA counts. We then compare its performance with the existing methods used for modeling mRNAs, such as the negative binomial. This empirical model can be a good modeling tool for dispersed mRNA count data when a biophysically based model is not available.

TABLE OF CONTENTS

1.0 INTRODUCTION 1

1.1 Motivation and Objectives on Photon Counting Problem 2

1.2 Motivation and Objectives of Conway-Maxwell-Poisson problem 3

1.3 Outline 4

2.0 PHOTON COUNTING WITH DEAD TIME 5

2.1 BACKGROUND ON COUNTERS 5

2.1.1 Definitions 5

2.1.2 Brief literature review on counters 5

2.2 TWO-PHOTON LASER SCANNING MICROSCOPY 6

2.2.1 Introduction 6

2.2.2 How TPLSM works. 8

2.2.3 Effect of dead period on photon counting process 8

2.3 Earlier efforts 10

2.4 Chapter Summary 11

3.0 THE DISTRIBUTION OF THE NUMBER OF PHOTONS DETECTED 12

3.1 Notation and assumptions 12

3.2 Deriving the distribution of D 13

3.2.1 First approach 14

3.2.2 Second approach 19

3.3 First two moments of D 24

3.4 INFERENCE ABOUT THE UNDERLYING INTENSITY PARAMETER . 24

3.4.1 MLE 25

3.4.1.1 Simulated data . . . 26
3.4.1.2 For the paper fiber data . . . 26
3.4.2 Fisher information . . . 28
3.4.3 Loss of information due to grouping . . . 29
3.4.3.1 Simulated data . . . 30
3.4.3.2 Paper fiber data . . . 31
3.5 Chapter Summary . . . 33
4.0 MODEL EXTENSIONS . . . 35
4.1 Finite time assumption . . . 35
4.1.1 $P(D = d)$ for finite horizon based on approach in Section 3.2.1 . . . 35
4.1.2 Finite time assumption for approach in Section 3.2.2 . . . 39
4.1.3 How large T should be? . . . 43
4.2 Gamma waiting times assumption . . . 46
4.3 Chapter Summary . . . 50
5.0 APPROXIMATING THE CONWAY-MAXWELL-POISSON NORMALIZING CONSTANT . . . 51
5.1 CONWAY-MAXWELL-POISSON DISTRIBUTION . . . 51
5.1.1 Special cases of Conway-Maxwell-Poisson . . . 52
5.1.2 Moments . . . 54
5.1.3 MLEs . . . 56
5.2 APPROXIMATION OF THE NORMALIZING CONSTANT . . . 56
5.3 Earlier Approach: Laplace approximation for integer ν . . . 57
5.4 Our approach . . . 59
5.4.1 A statistical approach for all $\nu > 0$ . . . 59
5.4.1.1 Express normalizing constant as an expectation . . . 59
5.4.1.2 Use Normal approximation to $2\sqrt{X}$ and expansions . . . 60
5.4.1.3 Take expectations . . . 62
5.4.2 GENERALIZED HYPERGEOMETRIC FUNCTION APPROACH . . . 63
5.4.3 CORRECTION TERMS . . . 65
5.4.3.1 Use the expansions for $\ln(1+t)$ and $\ln\Gamma(x+1)$ with more terms 65

5.4.3.2 Take expectations . . . 66
5.4.4 NUMERICAL EXAMPLES . . . 67
5.4.5 Approximation of $C(\lambda, \nu)$ when λ and ν small . . . 69
5.5 Chapter Summary . . . 70
6.0 APPLICATION OF CONWAY-MAXWELL-POISSON GLM FOR MRNA COUNTS . . . 72
6.1 Literature Review . . . 73
6.1.1 On Conway-Maxwell-Poisson distribution . . . 73
6.1.2 On mRNA counts . . . 74
6.2 Conway-Maxwell-Poisson GLM for mRNA counts . . . 76
6.2.1 Why use Conway-Maxwell-Poisson GLM for mRNA counts? . . . 77
6.2.2 Conway-Maxwell-Poisson GLM and its zero-inflated version . . . 77
6.2.3 The other models . . . 78
6.3 mRNA count data sets . . . 79
6.3.1 Comparing E. coli expression levels for different promoters . . . 79
6.3.2 Comparing different doxycycline levels in mammalian cells . . . 82
6.4 Chapter Summary . . . 83
7.0 CONCLUSION . . . 85
7.1 PHOTON COUNTING WITH DEAD TIMES . . . 85
7.2 APPROXIMATION OF THE NORMALIZING CONSTANT OF CONWAY-MAXWELL-POISSON . . . 87
7.3 CONWAY-MAXWELL-POISSON GLM FOR mRNA COUNTS . . . 88
8.0 FUTURE WORK . . . 89
8.1 For photon counting problem . . . 89
8.2 For approximation of the normalizing constant of the Conway-Maxwell-Poisson 90
8.3 For application of the Conway-Maxwell-Poisson to mRNA production . . . 91
9.0 APPENDIX: FOR PHOTON COUNTING PROBLEM . . . 92
9.1 Proofs . . . 92
9.2 Simulation Steps . . . 98
9.2.1 For infinite observation period . . . 98

9.2.2 For finite observation period 99
BIBLIOGRAPHY 100

LIST OF TABLES

1 Newton Raphson iterations 26

2 Comparison of equation 4.1 with equation 4.2 38

3 Value $P(D = 0)$ under different model assumption 43

4 When $\delta = 1$, the value p_1 under different model assumption 46

5 First four probabilities based on gamma assumption 48

6 Probabilities for Poisson-gamma model 49

7 Percentage errors when approximating $C(\lambda, \nu)$ by using (5.12) 62

8 Relative errors in (5.13) for leading term and for first-order correction; $\nu = 2$. 64

9 Relative error using the leading term in (5.17) 68

10 Relative errors using first correction term in (5.18) 69

11 Comparison of relative errors for (5.4) and (5.19) 70

12 BIC of each model fit for E. coli data 80

13 BIC for each zero-inflated model fit 81

14 BIC of each model fit for mammalian cells 83

15 For even case $c = d/2$. 93

16 For odd case $c = \frac{d-1}{2}$. 94

LIST OF FIGURES

1 Schematic of two-photon microscopy. 7
2 First step of TPLSM mechanism 8
3 Second step of TPLSM mechanism 9
4 No dead period. 9
5 Dead periods produced while recording photons. 10
6 A depiction of $[D = 1]$. 14
7 A depiction of $[D = 2]$. 15
8 Graph of the (a) mean, (b) variance, (c) Fano factor. 24
9 Log-likelihood plot. 27
10 Fisher Information for different groupings. 30
11 Contour plot of $\hat{\alpha}_i$ estimates based on (m_0, m_1, m_{2+}) 32
12 Contour plot of improvement in $\hat{\alpha}_i$ estimates 33
13 First way to record a single photon when the observation period is finite. . . . 36
14 Second way to record a single photon when the observation period is finite. . 37
15 Comparison of values of $P(D = 0)$ under finite observation period assumption (colored line) and infinite observation period (black line). 44
16 Comparison of values of $P(D = 1)$ under finite observation period (colored line) and infinite observation period (black line). 45
17 Gene activation and inactivation. 75
18 Bursty behavior of mRNA. 76

ACKNOWLEDGEMENTS

I would like to express my sincere thanks and appreciation to all the people who have guided and inspired me through my PhD research at University of Pittsburgh.

First and for most, I thank my advisor Dr. Satish Iyengar for his constant support. If it was not for his help and dedication, I could not have conducted this research. I appreciate all his effort in investing his time and energy to answer my questions and sharing his wisdom, not only in my research but also in all of his fields of expertise. His patience in elaborating the details is truly exceptional. I would also like to thank the rest of my dissertation committee members for their time and guidance.

I am pleased to express my special thanks to my parents Saniye and Yilmaz, and my sister Buket for their constant support throughout my life. Thanks to all my instructors who helped me build up my knowledge during PhD. Last but not least, I thank my friends and colleagues for their friendly support during all these years. They are the ones whom I spent the most part of my days with while working on our research. I value their friendship along with their technical advice and supports.

1.0 INTRODUCTION

This dissertation concerns point process models and their applications to imaging and biology. It consists of three parts: a study of models for photon counting, a study of the Conway-Maxwell-Poisson distribution for count data, and the application of this distribution to mRNA transcription. The common thread is that counting problems are difficult when there are censoring mechanisms like dead periods in the photon counting problem and off states gene activity in mRNA transcription.

In recent years, one of the most significant biomedical research tools is optical microscopy which has applications to molecular biology. Many variants of microscopy have been developed to obtain high-resolution and high-sensitivity three-dimensional images, especially for deep-tissue imaging. One of these is two-photon laser scanning microscopy (TPLSM, [13, 25]), which is an essential development in biological imaging [59]. Moreover, it allows imaging of living tissue up to a very high depth (about one millimeter) with high-resolution and high-sensitivity fluorescence three-dimensional images [37, 58, 63]. In the first part of this thesis, we deal with stochastic models for TPLSM and statistical inference for them. The motivation and objectives of this problem given in Section 1.1.

Next, direct detection of gene activity is often not possible because new proteins from individual activation events are masked by proteins remaining from previous events [47]. Thus, researchers observe mRNA because it gives information about gene activity. Modeling such phenomena is complicated because there are many factors that affect outcomes and are hard to control. In the absence of detailed biophysical knowledge, we propose the use of the Conway-Maxwell-Poisson family as a potentially good candidate for fitting dispersed mRNA count data. Before we pursue this, we address an interesting technical problem of obtaining better approximations to the normalizing constant of the this family: see Section 1.2.

1.1 MOTIVATION AND OBJECTIVES ON PHOTON COUNTING PROBLEM

Two-photon microscopy enables imaging of living tissue by minimizing the confounds of scattered and absorbed light, up to a depth of about one millimeter below the surface of the brain and with sub-micrometer resolution [14]. Thus, it can be used in many areas such as physiology, neurobiology, and tissue engineering, for which imaging of highly scattering tissue is required [59]. Neuroscientists have used TPLSM to image physiological functioning in microscopic and subcellular neural compartments [40].

As pointed out in Oheim et al. [43], TPLSM is a fluorescence imaging technique. The duration of time between absorption of an excited photon and emitted photon is called the fluorescence lifetime. If another photon is emitted during the fluorescence lifetime, the device cannot detect it. This period is therefore considered a dead period of the device. Since dead periods lead to undercounts of emitted photons, it is of interest to do statistical inference for the photon counting problem with dead time [55].

Starting in early 1950s, there have been many studies of probabilistic modeling and statistical inference for this problem. Almost all of the work that has been done is concerned with asymptotic results, and none of them give exact results. Knowing the exact results on this photon counting with dead time problem can not only improve the quality of images, but can also lead to images with higher resolution and greater sensitivity in three dimensions. The literature review we have done indicates that there are no relevant probability calculations and resulting statistical models that address the dead period. Therefore, it is clear that a mathematical model that relates the number of photons recorded and the actual emitted photon counts is needed.

In this study, we first address the problem of photon counting with dead time which appears in TPLSM. We use two separate but equivalent approaches. The first approach is based on an inhomogeneous Poisson process to capture the underlying biophysics. The second approach uses an equivalent representation by obtaining the marginal distribution of D, the number of photons recorded, from its conditional distribution given N, the total number of emitted photons [56]. We then use maximum likelihood to estimate the underlying

intensity parameter and dead time of the mechanism. We illustrate our method on imaging data. We also study several variants of the basic model.

1.2 MOTIVATION AND OBJECTIVES OF CONWAY-MAXWELL-POISSON PROBLEM

Messenger RNA (mRNA) transcription is a product of gene activity, so it can provide valuable information about gene activation and inactivation patterns [59]. mRNA transcription occurs at a high rate when the gene is in its on state, and no mRNA transcription when its in its off state. The off state of the gene is analogous to the dead period of TPLSM: see [44]. Because of this common feature, we are interested in modeling mRNA counts as our second research direction. We propose the use of the Conway-Maxwell-Poisson, which is an empirical distribution for modeling mRNA counts.

There are several reasons why we propose the use of the Conway-Maxwell-Poisson family for mRNA counts. First, the mRNA count data is overdispersed because the mRNA produced at high rate when the gene is on-state. We believe that the Conway-Maxwell-Poisson family could potentially be a good candidate for modeling mRNA counts because it is well known as good for a variety of dispersed data. Second, the distributions used to model the mRNA counts are special cases of Conway-Maxwell-Poisson family. Therefore, we want to see how a broad family would perform in modeling mRNA counts. Third, the Conway-Maxwell-Poisson model was first proposed in the queueing theory context [9]. Queueing models have also been considered for gene expression and mRNA transcription [17].

Before we address modeling mRNA counts, we deal with another interesting problem on the Conway-Maxwell-Poisson distribution. The normalizing constant of this distribution plays an important role in the computation of probabilities, moments, maximum likelihood estimates and their standard errors. Since it is not always easy to compute the normalizing constant, good approximations are needed. An earlier approximation for the normalizing constant is good only in some certain parameter ranges. Below, we improve the quality of this approximation by introducing some correction terms, and suggesting alternative approx-

imations for other extreme parameter ranges.

As a result, in the second part of this thesis, we improve the quality of the approximation of the normalizing constant of Conway-Maxwell-Poisson distribution. We then use this family to study mRNA counts, thereby introducing it to a new area.

1.3 OUTLINE

The first five chapters of this thesis deals with the problem that is addressed in Section 1.1. In Chapter 2, we give a brief review of the literature on counters and TPLSM. We then describe the problem that motivated us in detail. In Chapter 3, we derive the exact distribution of D, the number of photons detected: we first give the details of two approaches that we used; we then turn to the problem of estimating the model parameters; third, we relate this problem with the more general concept of loss of information due to grouping; finally, we present our results on both simulated data and paper fiber data. In Chapter 4, we derive the distribution of D under different model assumptions to study possible extensions of the work presented in Chapter 3. We are specifically interested in comparing finite versus infinite observation periods, and exponential vs. gamma waiting times. The proofs and simulation algorithms of Chapter 3 are given in Chapter 9.

The remaining chapters are for the problems described in Section 1.2: approximating the normalizing Conway-Maxwell-Poisson distribution and use of this on mRNA counts. In Chapter 5, after introducing Conway-Maxwell-Poisson family, we investigate the approximation of the normalizing constant of this family after we present briefly the earlier efforts. To illustrate the performance of our approximation to the normalizing constant, we present numerical examples. In Chapter 6, we introduce literature on both the Conway-Maxwell-Poisson GLM and mRNA counts. We then study the use of the Conway-Maxwell-Poisson on mRNA counts. In Chapter 7, we summarize our conclusions on both of these two problems presented in Section 1.1 and Section 1.2. In Chapter 7, we present our conclusions. Finally, in Chapter 8, we suggest several open problems that are worth pursuing in these areas.

2.0 PHOTON COUNTING WITH DEAD TIME

2.1 BACKGROUND ON COUNTERS

2.1.1 Definitions

A counter is a machine that can detect and record pulses [30]. Some of the arrivals may not be recorded by the counter because of its inertia. This is called the dead period of the counting machine [46]. Depending on the type of dead period, the counters can be classifed as either Type I counters or Type II counters.

If the dead period only occurs after registration of a pulse, we have a Type I counter or non-paralyzable counter. If the dead period occurs after each pulse (whether it is recorded or not), we have a Type II counter or paralyzable counter. Common examples of Type I and Type II counters are Geiger-Miuller counters [19, 46, 64]. For TPLSM, we are only interested in Type I counters.

2.1.2 Brief literature review on counters

Type I and Type II counters have been used in various fields, so the statistics literature for them goes back to early 1950s. Almost all these previous studies about the counters deal with asymptotic results. Also, most of them used the homogeneous Poisson assumption for the pulse process, and renewal theory methods are the main analytic tool.

Pyke [46] and Takàcs [64] considered several renewal processes that are connected to Type I and Type II counters, and derived the limiting distributions of observed counts. Kao and Smith [30] studied the counter models given by Pyke, and approached the problem by assuming a phase type interarrival distribution. Albert and Nelson [2] gave the confidence

intervals for the mean rate of such counters.

Bentley [4] assumed that the arrivals followed a homogeneous Poisson process, and studied the limiting distributions of the dead and live periods of the detector. The expected length of the process being in the dead period was also calculated. Yu and Fessler [67] obtained the first two moments of the observed counts for single photon counting problem for Type I and Type II counters. They also defined a Type III counter, in which two photons arrive within dead period of each other, then neither photon will be recorded; they obtained the exact mean and variance for Type III counters.

Driscoll et. all [14] dealt with the photon counting problem by considering "zero counts" and "others". They compared the results with that of "zero counts", "one counts", and "others" on their companion paper [15]. The images and their signal-to-noise ratios (SNR) were considerably better when using the one counts also. However, they did not compute the distribution of the observed counts; we turn to that problem in the next section.

2.2 TWO-PHOTON LASER SCANNING MICROSCOPY

2.2.1 Introduction

One of the most significant biomedical research tools in cellular and systems biology is optical microscopy. Many variants of microscopy have been developed to obtain high-resolution and high-sensitivity three-dimensional images [18]. One of these, of particular relevance for deep imaging of brain structure and function, involves pulsed laser light and is dominated by two-photon laser scanning microscopy (TPLSM, [13, 25]). Two-photon microscopy enables imaging of living tissue by minimizing the confounds of scattered and absorbed light, up to a depth of about one millimeter below the surface of the brain and with sub-micrometer resolution. Beyond this, pulsed laser light is a powerful tool to image tissue based on harmonic generation and vibration spectra through coherent anti-Stokes Raman spectroscopy [55].

TPLSM is a valuable tool for biological imaging for living tissues, involves the periodic excitation of molecules by hundred-femtosecond laser pulses, followed by the emission of

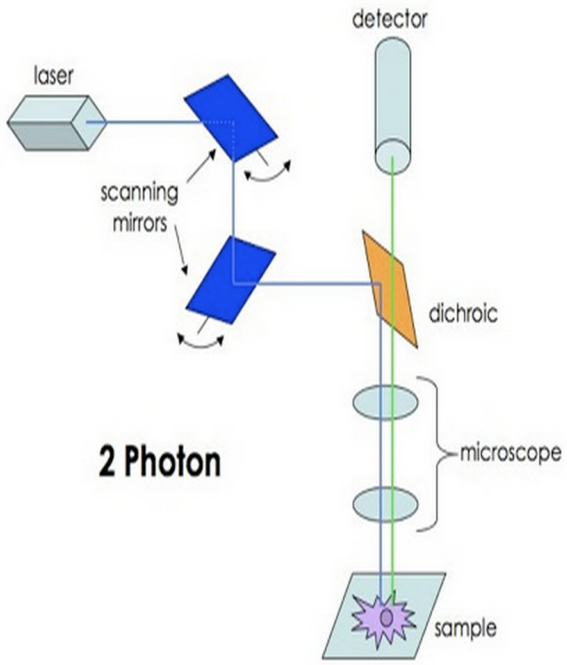

Figure 1: Schematic of two-photon microscopy.

fluorescent light before the arrival of the next pulse [14, 65]. In TPLSM, the use of a pulsed laser is necessary to achieve the extreme electric fields that lead to two-photon absorption. Absorption can only occur within the lifetime of the pulse, which is significantly less than one picosecond. In contrast, the time between absorption of an excitation photon and the emission of a photon, called the fluorescence lifetime, is on the order of a few nanoseconds. Thus, the laser pulse provides a well defined start signal for the emission process. Detection of a single photon is a mature process [59]. However, every detector and associated threshold electronics has a dead period, during which the composite system is insensitive to the arrival of an additional photon. The dead period is typically on the order of one nanosecond and depends on details of the detector and electronics.

2.2.2 How TPLSM works.

Figure 1, from [61], demonstrates that TPLSM is a complicated engineering feat. Here, we describe the main features that are relevant for us from a modeling perspective. First, the laser excites N molecules on the surface that is to be imaged. Each of those molecules emits photons. Next, these photons are recorded by the detector, and the resulting photon counts are used to construct the image. In another words, when the laser beam reaches to the atom of the tissue: see Figure 2, it relese photons each which has different frequencies: see Figure 3. Thus, the photons arrives at the detector at different times.

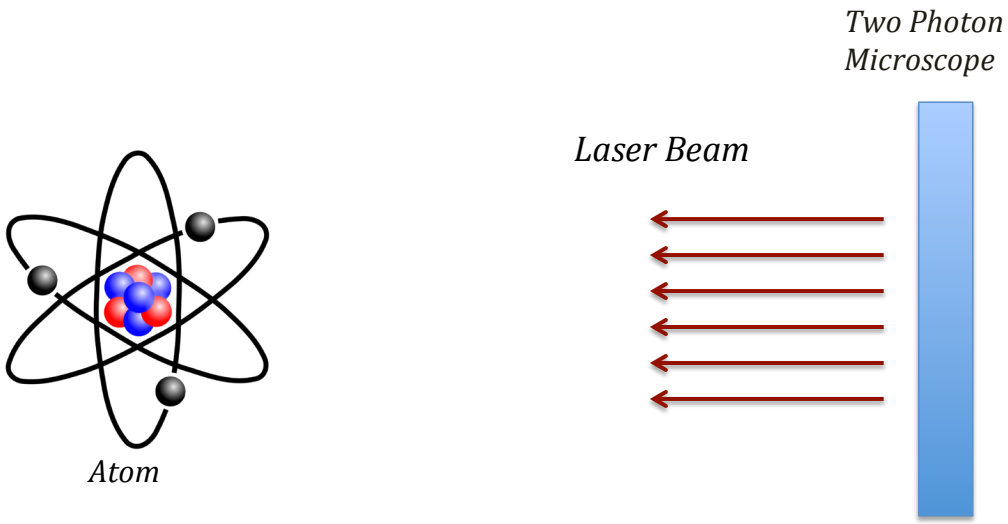

Figure 2: First step of TPLSM mechanism

2.2.3 Effect of dead period on photon counting process

When recording photons, the detector has a dead period. Thus, if a second photon is emitted during the fluorescence lifetime, the device will not detect it if it arrives too soon after the first photon. This process continues so that detection of a second photon leads to a second dead period, etc. Since dead periods lead to an undercount of the number of emitted photons, it is of interest to develop statistical methods for the photon counting problem with dead

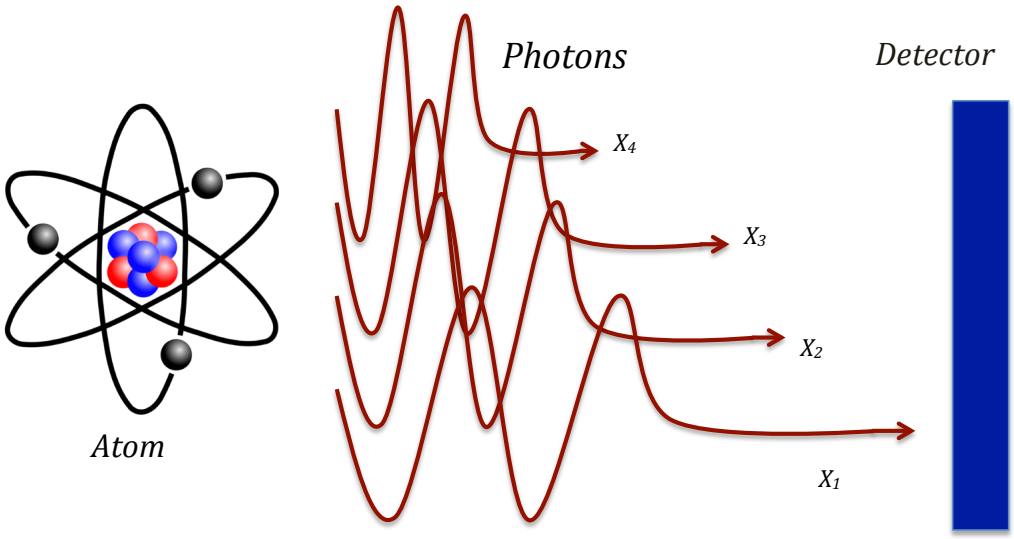

Figure 3: Second step of TPLSM mechanism

time.

Figure 4 below is a depiction that illustrates how the dead period of TPLSM affects the photon counting process. Suppose that 6 photons emitted at times t_1 to t_6. The detector would have recorded all these emitted photons if it did not have a dead period after the registration of each photon. Thus, this would have been a simple point process to analyze.

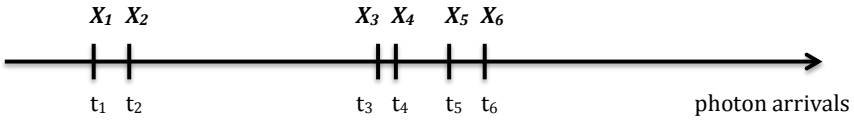

Figure 4: No dead period.

However, after each photon arrival, the detector of the machine has a dead period. In Figure 5, it can be seen that X_2, X_4, and X_5 arrived at the detector during its dead periods. We end up recording $D = 3$ out of $N = 6$ emitted photons; this undercount in turn results in degraded images. As we shall see below, the existence of dead periods for TPLSM makes the photon counting process challenging.

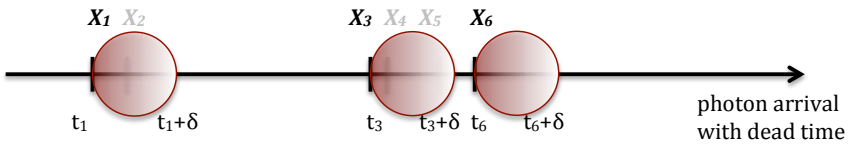

Figure 5: Dead periods produced while recording photons.

2.3 EARLIER EFFORTS

To avoid the problem of undercounts, investigators engineered the laser power to obtain small photon emission rates so that each emitted photon is recorded (with very high probability) before the next photon arrives. In earlier work, Driscoll et al. [14] collected the data by assuming that the rate of emitted photons is small and that the dead periods are a known constant. They used low laser intensities to ensure that the concentration of excited molecules, and thus the rate of emitted photons was small. At that time, the distribution of D, the number of detected photons, was only partially known, so they worked with counts grouped in the categories of zero, one, and greater than one. They compared the images and signal-to-noise ratios (SNR) by considering zero counts only, zero and one counts only, and analog methods; for analog methods the total photocurrent is measured as opposed to the arrival times of individual photons. They showed that the use of zero and one counts gave considerably better images than the use of zero counts only; we expect further improvement with the use of higher counts. Hence, we must use higher rates of photon emission, and

detect or infer as many photons as possible to get better images.

2.4 CHAPTER SUMMARY

In this chapter, we give brief overview of the problem that motivated us to pursue the distribution of photon counts with dead time. We first address the problem of photon counting with dead time in TPLSM. By using a simple example, we demonstrate the effect of the dead period on the counting process. From this example, we can see that the dead period makes the modeling of the photon counts more complicated. We then summarized the literature on the earlier efforts in modeling photon counts when there are dead periods. This literature review indicates that there are no appropriate probability calculations and resulting statistical models that address the dead period, resulting in bad image quality. Therefore, it is clear that a mathematical model that relates the number of photons recorded and the actual emitted photon counts is needed. In the next chapter, we pursue this model.

3.0 THE DISTRIBUTION OF THE NUMBER OF PHOTONS DETECTED

Our main aim is to obtain the distribution of D, the number of observed photons. We use two approaches to compute it. In the first approach, the process is represented by an inhomogeneous Poisson process, and the calculations are done accordingly. In the second approach, the problem is solved by obtaining the marginal distribution of D from the conditional distribution of D given N, the total number of emitted photons. We give the details of the second approach because it leads to several interesting polynomial identities using the marginal distribution of D derived from the first approach.

This chapter is outlined as follows. We start with our assumptions and notation. Second, we describe the details of our model, then derive the distribution of D, and study its properties. Third, we give details of maximum likelihood estimates of the key parameters, namely the emission rate and the dead period, along with their estimated standard errors; we also compare the estimates' standard errors with different degrees of grouping. We use simulated data and paper fiber data to illustrate the performance of our model. We end this chapter with a short summary of our work.

3.1 NOTATION AND ASSUMPTIONS

In this work, we make distributional assumptions as in past studies, namely, the Poisson assumption for the pulse process and exponentially distributed fluorescence lifetimes as in Oheim et al. [43].

Let N represent the number of excited molecules. We assume that N is a Poisson random variable with mean α, and let $t_1, \ldots, t_n$ be the photon arrival times. When $N = n \geq 1$, let

$W_1, \ldots, W_n$ be the waiting times for the photon emissions, which are independent exponential random variables with time constant τ nanoseconds (ns). The order statistics$W_{(n:1)} < \ldots < W_{(n:n)}$ represent the waiting times for the emission of the photons.

Suppose that the photon detector has a dead period of Δ ns; the photons that are emitted during a dead period do not affect that dead period, so this model is a Type I counter. For our computations below, it is convenient to use a clock with units in terms of the time constant, so we use the rescaled dead period $\delta = \Delta/\tau$, and assume that $\tau = 1$ without loss of generality. Although the observation period T is finite in our application, it is often large enough that we can assume that it is infinite to get good approximations: see [14]. In Chapter 4 we will study the finite time horizon case, and assess how large T should be for the infinite horizon approximation to be good. Let D be the number of photons detected.

In order to express the probabilities $P(D = d)$ in a compact form, we define the following quantities: for $k = 1, 2, \cdots$, let

$$A_k = A_k(\delta) = (1 - e^{-k\delta}), \quad B_k = B_k(x) = 1 + x + \cdots + x^k$$

$$\{k\} = \binom{k}{2} \quad \text{and} \quad \zeta_k = \zeta_k(\alpha, \delta) = e^{-\alpha e^{-k\delta}}.$$

We also use the convention that the product $\prod_{k=a}^{b} A_k = 1$ whenever $b < a$.

3.2 DERIVING THE DISTRIBUTION OF D

We have used two different approaches to derive the distribution of D. When $N = n \geq 1$, let $W_{(n:1)} < \ldots < W_{(n:n)}$ be the ordered waiting times. The model that we have just described is equivalent to an inhomogeneous Poisson process $N(t)$ with intensity $\lambda(t) = \alpha f(t) = \alpha e^{-t}$ and cumulative intensity $\Lambda(u) = \alpha(1 - e^{-u})$: see Daley and Vere-Jones [12]. Thus, the order statistics of the arrivals $W_{(n:1)}, \ldots, W_{(n:n)}$ have the same distribution as the ordered points from the inhomogeneous process when $N(\infty) = n$. We first use the inhomogeneous Poisson process formulation to compute the distribution of D. After that, we give a lengthy analysis

of the order statistics formulation to derive this distribution again; this alternative approach yields several interesting polynomial identities.

3.2.1 First approach

This approach represents the counter process as the inhomogeneous Poisson process $N(t)$ with intensity $\lambda(t) = \alpha f(t) = \alpha e^{-t}$ and cumulative intensity $\Lambda(u) = \alpha(1-e^{-u})$. We compute the first few probabilities $P(D = d)$ to illustrate our general approach. The notation defined in the previous section is used to express these probabilities in a compact form.

Since $[D = 0] = [N = 0]$, we have

$$P[D = 0] = P[N = 0] = e^{-\alpha} = \zeta_0.$$

The event $[D = 1]$ occurs when there are no emissions in some interval $[0, t)$, one emission in the interval $[t, t + dt)$, and no emissions beyond the dead period $[t + \delta, \infty]$: see Figure 6.

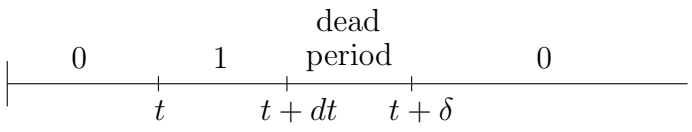

Figure 6: A depiction of $[D = 1]$.

The probability of no emissions in $[0, t)$ is

$$\exp\left(-\int_0^t \alpha e^{-x} dx\right) = e^{-\alpha(1-e^{-t})};$$

the probability of an emission in $[t, t + dt)$ is $\alpha e^{-t} dt$; and the probability of no emissions after the dead period is

$$\exp\left(-\int_{t+\delta}^{\infty} \alpha e^{-x} dx\right) = e^{-\alpha e^{-(t+\delta)}}.$$

Putting these pieces together and using the substitution $x = \alpha e^{-t}$, we have

$$\begin{aligned} P[D=1] &= \int_0^\infty e^{-\alpha(1-e^{-t})} e^{-\alpha e^{-(t+\delta)}} \alpha e^{-t} dt = e^{-\alpha} \int_0^\alpha e^{x\left(1-e^{-\delta}\right)} dx \\ &= \frac{e^{-\alpha e^{-\delta}} - e^{-\alpha}}{1-e^{-\delta}} = \frac{\zeta_1}{A_1} - \frac{\zeta_0}{A_1}. \end{aligned}$$

Similarly, the event [D=2] is depicted in Figure 7.

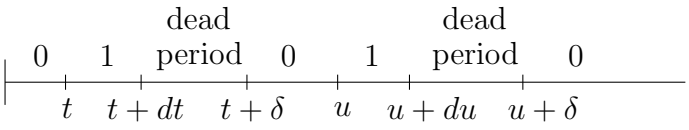

Figure 7: A depiction of $[D = 2]$.

For the first case, the intervals with an observation $[t, t+dt)$ and $[u, u+du)$, has probability $\alpha e^{-t} dt$, and $\alpha e^{-u} du$, respectively. For the second case, the intervals with no emissions are

$$[0, t) \cup [t+\delta, u) \cup [t+\delta, \infty),$$

each of these intervals has, respectively, the following probabilities

$$\exp\left(-\int_0^t \alpha e^{-x} dx\right) = e^{-\alpha(1-e^{-t_1})},$$

$$\exp\left(-\int_{t+\delta}^u \alpha e^{-x} dx\right) = e^{-\alpha[e^{-(t+\delta)} - e^{-u}]},$$

and

$$\exp\left(-\int_{u+\delta}^\infty \alpha e^{-x} dx\right) = e^{-\alpha e^{-(u+\delta)}}.$$

Putting these together, we have

$$P[D=2] = \int_{t=0}^\infty \int_{u=t+\delta}^\infty e^{-\alpha(1-e^{-t})} \alpha\, e^{-t} e^{-\alpha(e^{-(t+\delta)} - e^{-u})} \alpha\, e^{-u} e^{-\alpha e^{-(u+\delta)}} du\, dt.$$

We first integrate with respect to u by substituting $u = s + t + \delta$ and then $x = \alpha e^{-s}$ to get

$$\begin{aligned}
\int_{t+\delta}^{\infty} e^{-\alpha(e^{-(t+\delta)}-e^{-u})} e^{-\alpha e^{-(u+\delta)}} \alpha\, e^{-u} du &= e^{-\alpha e^{-(t+\delta)}} e^{-(t+\delta)} \int_0^{\infty} e^{\alpha\left[e^{-(s+t+\delta)}-e^{-(s+t+2\delta)}\right]} \alpha e^{-s} ds \\
&= e^{-\alpha e^{-(t+\delta)}} e^{-(t+\delta)} \int_0^{\alpha} e^{x\left[e^{-(t+\delta)}-e^{-(t+2\delta)}\right]} dx \\
&= \frac{e^{-\alpha\, e^{-(t+2\delta)}} - e^{-\alpha\, e^{-(t+\delta)}}}{A_1} = \frac{U_2(t) - U_1(t)}{A_1}
\end{aligned}$$

where $U_m = e^{-\alpha\, e^{-(t+m\delta)}}$. Next, we integrate with respect to t by substituting $x = \alpha e^{-t}$ to get

$$\begin{aligned}
P[D=2] &= \int_0^{\infty} e^{-\alpha(1-e^{-t})} \alpha\, e^{-t} \left(\frac{U_2(t) - U_1(t)}{A_1}\right) dt = \frac{e^{-\alpha}}{A_1} \int_0^{\alpha} \left(e^{xA_2} - e^{xA_1}\right) dx \\
&= \frac{\zeta_2 - \zeta_0}{A_1 A_2} - \frac{\zeta_1 - \zeta_0}{A_1^2} = \frac{\zeta_2}{A_1 A_2} - \frac{\zeta_1}{A_1^2} + \frac{e^{-\delta}\zeta_0}{A_1 A_2}.
\end{aligned}$$

Notice that, we combined the coefficients of ζ_0 above. And for $[D = 3]$ below, we combine the coefficients of ζ_0 and ζ_1 thus:

$$\begin{aligned}
P[D=3] &= \frac{\zeta_3 - \zeta_0}{A_1 A_2 A_3} - \frac{\zeta_1 - \zeta_0}{A_1^2 A_2} + \frac{\zeta_2 - \zeta_0}{A_1^2 A_2} - \frac{\zeta_1 - \zeta_0}{A_1^3} \\
&= \frac{\zeta_3}{A_1 A_2 A_3} - \frac{\zeta_2}{A_1^2 A_2} + \frac{e^{-\{2\}\delta}\zeta_1}{A_1^2 A_2} - \frac{e^{-\{3\}\delta}\zeta_0}{A_1 A_2 A_3}.
\end{aligned}$$

In general, for $[D = d]$, combining the coefficients of $\zeta_0, \ldots, \zeta_{d-2}$ in similar expressions is facilitated by the following key result, the proof of which is given in the Appendix.

Lemma 1. *For $d \geq 1$,*

$$\sum_{k=0}^{d-1} (-1)^{k+1} \frac{e^{-\{k\}\delta}}{\prod_{i=1}^{k} A_i \prod_{i=1}^{d-k} A_i} = (-1)^d \frac{e^{-\{d\}\delta}}{\prod_{i=1}^{d} A_i}. \tag{3.1}$$

More generally, the event $[D = d]$ is described by infinitesimal intervals with one observed emission in each and intervals with no observations after the dead periods. For the first case, an observation in $[t_i, t_i + dt_i)$ has probability $\alpha e^{-t_i} dt_i$, for $i = 1, \ldots, d$. For the second case, the intervals with no emissions are

$$[0, t_1) \cup \left[\bigcup_{i=1}^{d-1} [t_i + \delta, t_{i+1}) \right] \cup [t_d + \delta, \infty).$$

The cumulative intensities for these intervals are

$$\Lambda[0, t_1) = \int_0^{t_1} \alpha e^{-x} dx = \alpha(1 - e^{-t_1}),$$

$$\Lambda[t_i + \delta, t_{i+1}) = \alpha[e^{-(t_i+\delta)} - e^{-t_{i+1}}] \quad \text{for} \quad i = 1, \ldots, d-1,$$

and

$$\Lambda[t_d, \infty) = \alpha e^{-(t_d+\delta)}.$$

Putting these together, we get

$$P(D = d) \;=\; \int_{t_1=0}^{\infty} \cdots \int_{t_d=t_{d-1}+\delta}^{\infty} H(t_1, \ldots, t_d | \alpha, \delta)\, dt_1 \cdots dt_d, \tag{3.2}$$

where

$$H(t_1, \ldots, t_d | \alpha, \delta) \;=\; e^{-\alpha(1-e^{-t_1})} \prod_{i=1}^{d-1} e^{-\alpha[e^{-(t_i+\delta)} - e^{-t_{i+1}}]} e^{-\alpha e^{-(t_d+\delta)}} \prod_{i=1}^{d} \alpha e^{-t_i}.$$

By applying Lemma 1 for the iterated integrals in (3.2), we get

Theorem 1. *For $d = 0, 1, \ldots$, the distribution of D is*

$$p_d = P(D = d) = \sum_{k=0}^{d} (-1)^k \frac{e^{-\{k\}\delta} \zeta_{d-k}}{\prod_{i=1}^{k} A_i \prod_{j=1}^{d-k} A_j}. \tag{3.3}$$

Now that entire distribution of D is known, for TPLSM we can use all available photon counts without grouping to construct the likelihood and make inferences about the parameters (α, δ); in addition, we are no longer restricted to the use of low intensities. The expressions for these probabilities are rather involved, but they satisfy a simple set of equations: the first three are

$$\zeta_0 = p_0, \quad \zeta_1 = A_1 p_1 + p_0, \quad \zeta_2 = A_1 A_2 p_2 + A_2 p_1 + p_0.$$

The first equation gives us p_0, which we can use with the second to get p_1, which we can then use to solve for p_2. In general, we have

Corollary 1. *The probabilities in (3.3) satisfy the recursions*

$$\zeta_d = \sum_{i=0}^{d} \left(\prod_{j=i+1}^{d} A_j \right) p_{d-i}.$$

This result can be used for easy computation of the probabilities. Notice that the equations given in Theorem 1 and Corollary 1 can be represented in matrix notation: define the vectors $\boldsymbol{p}_d = (p_0, \ldots, p_d)'$ and $\boldsymbol{\zeta}_d = (\zeta_0, \ldots, \zeta_d)'$, so that $\boldsymbol{p} = Q_d \boldsymbol{\zeta}$ and $\boldsymbol{\zeta} = R_d \boldsymbol{p}$, where

$$Q_d = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 & 0 \\ -\frac{1}{A_1} & \frac{1}{A_1} & 0 & \cdots & 0 & 0 \\ \frac{1}{A_1 A_2} & -\frac{1}{A_1^2} & \frac{1}{A_1 A_2} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{1}{A_1 A_2 \ldots A_{d-1}} & -\frac{1}{A_1^2 A_2 \ldots A_{d-2}} & \frac{1}{A_1^2 A_2 \ldots A_{d-3}} & \cdots & \frac{(-1)^{d-1}}{A_1 A_2 \ldots A_{d-1}} & \frac{(-1)^d}{A_1 A_2 \ldots A_{d-1} A_d} \end{pmatrix}$$

$$R_d = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & A_1 & 0 & \cdots & 0 & 0 \\ 1 & A_2 & A_1 A_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & A_{d-1} & A_1 A_{d-1} & \cdots & A_1 \ldots A_{d-1} & 0 \\ 1 & A_d & A_1 A_d & \cdots & A_1 \ldots A_{d-2} A_d & A_1 \ldots A_{d-1} A_d \end{pmatrix}.$$

The proof of Corollary 1 is in the Appendix. We show there that $Q_d R_d = I$, once again using Lemma 1.

Corollary 1 also easily establishes the fact that distinct values of α and δ lead to distinct values of the probabilities. Because of the one-to-one correspondence between $\{p_k\}$ and $\{\zeta_k\}$, it is enough to show that distinct values of α and δ lead to distinct values of $\{\zeta_k\}$. That is trivial because of the relations $\zeta_0 = e^{-\alpha}$ and $\zeta_1 = e^{-\alpha e^{-\delta}}$. Thus, these two parameters are identifiable; we address the simultaneous estimation when we are making inference on the underlying intensity paramter and the dead period.

3.2.2 Second approach

The second approach uses the equivalent representation which leads to a lengthy analysis of the order statistics formulation. Here, using the same model and notation, we first condition on N to derive the conditional distribution of D given $[N = n]$ and then get the marginal distribution of D.

Let $W_{(n:1)} < \ldots < W_{(n:n)}$ be the ordered waiting times when $N = n \geq 1$. These order statistics can be represented thus [11]:

$$W_{(n:k)} \stackrel{d}{=} \frac{X_1}{n} + \cdots + \frac{X_k}{n-k+1}, \tag{3.4}$$

where the X_k are independent exponential random variables with time constant 1. The first few probabilities are easy. As before, because the observation period is infinite,

$$P(D = 0) = P(N = 0) = e^{-\alpha}.$$

Next, for $N \geq 1$, write

$$P(D = d) = \sum_{n=d}^{\infty} P(D = d|N = n)P(N = n) = \sum_{n=d}^{\infty} P_n(D = d)P(N = n).$$

Using the lack of memory property of the exponential distribution, we have

$$P_n(D = 1) = P(W_{(n:n)} \leq W_{(1:n)} + \delta) = (1 - e^{-\delta})^{n-1};$$

and since the generating function of N is

$$E(u^N) = e^{-\alpha(1-u)},$$

we have

$$\begin{aligned} P(D=1) &= \sum_{1}^{\infty}(1-e^{-\delta})^{n-1}P(N=n) \\ &= \sum_{0}^{\infty}(1-e^{-\delta})^{n-1}P(N=n) - \frac{P(N=0)}{1-e^{-\delta}} \\ &= \frac{E(1-e^{-\delta})^N - e^{-\alpha}}{1-e^{-\delta}} = \frac{e^{-\alpha e^{-\delta}} - e^{-\alpha}}{1-e^{-\delta}} = \frac{\zeta_1}{A_1} - \frac{\zeta_0}{A_1}. \end{aligned}$$

For larger values of d, we decompose the event $(D=d)$. For the d-tuple κ where $\kappa = \kappa(d) = (k_1, \ldots, k_d)$ with $1 = k_1 < \cdots < k_d \le n$, we observe $W_{(k_1)} < \cdots < W_{(k_d)}$ if and only if

$$Q_i^\kappa = \begin{cases} (W_{(k_{i+1}-1)} \le W_{(k_i)} + \delta) \cap (W_{(k_{i+1})} > W_{(k_i)} + \delta) & \text{if } i = 1, \ldots d-1 \\ (W_{(k_{i+1}-1)} \le (W_{(k_i)} + \delta) & \text{if } i = d \end{cases}$$

all occur.

By representation (3.4), for $i = 1, \ldots, d-1$, Q_i^κ depends only on $(X_{k_i+1}, \cdots, X_{k_{i+1}})$, and Q_d^κ depends on $X_{k_d+1}, \ldots, X_n$; hence, the events $\{Q_i^\kappa : 1 \le i \le d\}$ are independent. By the lack of memory property,

$$P(Q_d^\kappa) = P(W_{(n)} \le W_{(k_d)} + \delta) = (1-e^{-\delta})^{n-k_d};$$

and for $d \ge 2$ and $i = 1, \ldots, d-1$,

$$P(Q_i^\kappa) = \binom{n-k_i}{k_{i+1}-k_i-1} e^{-(n-k_{i+1}+1)\delta}(1-e^{-\delta})^{k_{i+1}-k_i-1}.$$

Summing over all the d-tuples, we have

$$P_n(D=d) = \sum_\kappa (1-e^{-\delta})^{n-k_d} \prod_{i=1}^{d-1} P(Q_i^\kappa) \tag{3.5}$$

$$= (1-e^{-\delta})^{n-d} e^{-(n+1)(d-1)\delta} \sum_{i=2}^{d} \cdots \sum_{k_i=k_{i-1}+1}^{n-d+i} \binom{n-k_i}{k_{i+1}-k_i-1} e^{k_i\delta}.$$

With this expression, we use

$$P(D=d) = \sum_d^\infty P_n(D=d)P(N=n).$$

to derive the marginal distribution of D. Before deriving the first few probabilities, we recall the following notation in order to express the probabilities in a compact form:

$$B_k(x) = 1 + x + \cdots + x^k \quad \text{for } k = 1, 2, \cdots$$

and let $B_k = B_k(e^{-\delta})$.

First, sum over k_d in equation (3.5) to get $P(D = 2)$.

$$\begin{aligned} \sum_{k_d=k_{d-1}+1}^{n} \binom{n-k_d}{k_d-k_{d-1}-1} e^{k_d\delta} &= \sum_{m=0}^{n-k_{d-1}-1} \binom{n-k_{d-1}}{m} e^{(m+k_{d-1}+1)\delta} \\ &= e^{(k_{d-1}+1)\delta}(1+e^{\delta})^{n-k_{d-1}} - e^{(n+1)\delta} \qquad (3.6) \\ &= e^{2\delta}B_1^{n-1} - e^{(n+1)\delta} \quad \text{for } d = 2. \end{aligned}$$

To get (3.6), we set $m = k_d - k_{d-1} - 1$ and then used the binomial theorem. Next, let $d = 2$, and substitute $k_1 = 1$. Then the conditional probability for $d = 2$ is

$$P_n(D=2) = (1-e^{-\delta})^{n-2}e^{-(n+1)\delta}[e^{2\delta}B_1^{n-1} - e^{(n+1)\delta}] = \frac{A_2^n}{A_1A_2} - \frac{A_1^n}{A_1^2}.$$

Notice that this expression is zero for $n = 1$ (see Lemma 3 below). Thus, we have

$$P(D=2) = \sum_{n=2}^{\infty}\left(\frac{A_2^n}{A_1A_2} - \frac{A_1^n}{A_1^2}\right)\frac{e^{-\alpha}\alpha^n}{n!} = \sum_{n=2}^{\infty}\frac{(A_2\alpha)^n e^{-\alpha}}{n!A_1A_2} - \sum_{n=2}^{\infty}\frac{(A_1\alpha)^n e^{-\alpha}}{n!A_1^2}$$

$$\begin{aligned} &= \frac{e^{-\alpha}}{A_1A_2}\left[e^{(A_2\alpha)} - \frac{(A_2\alpha)^1}{1!} - \frac{(A_2\alpha)^0}{0!}\right] - \frac{e^{-\alpha}}{A_1^2}\left[e^{(A_1\alpha)} - \frac{(A_1\alpha)^1}{1!} - \frac{(A_1\alpha)^0}{0!}\right] \\ &= \frac{\zeta_2}{A_1A_2} - \frac{\zeta_1}{A_1^2} + \frac{\zeta_0}{A_1}\left(\frac{1}{A_1} - \frac{1}{A_2}\right) = \frac{\zeta_2}{A_1A_2} - \frac{\zeta_1}{A_1^2} + \frac{e^{-\delta}\zeta_0}{A_1A_2}. \end{aligned}$$

Note that Lemma 1 appears in the calculation of the coefficient of ζ_0.

Next, sum over k_d, k_{d-1} in equation (3.5). Now we define $m = k_{d-1} - k_{d-2} - 1$ and use the binomial theorem again to get

$$\sum_{k_{d-1}=k_{d-2}+1}^{n-1}\ \sum_{k_d=k_{d-1}+1}^{n}\binom{n-k_{d-1}}{k_{d-1}-k_{d-2}-1}\binom{n-k_d}{k_d-k_{d-1}-1}e^{k_d\delta}e^{k_{d-1}\delta}$$

$$
\begin{aligned}
&= \sum_{k_{d-1}=k_{d-2}+1}^{n-1} \binom{n-k_{d-1}}{k_{d-1}-k_{d-2}-1} \left(e^{(k_{d-1}+1)\delta} B_1^{n-k_{d-1}} - e^{(n+1)\delta}\right) e^{k_{d-1}\delta} \\
&= \sum_{m=0}^{n-k_{d-2}-2} \binom{n-k_{d-2}}{m} \left[\frac{e^{(2k_{d-1}+3)\delta} e^{2\delta m} B_1^{n-k_{d-2}-m}}{B_1} - e^{\delta m} e^{(n+k_{d-2}+2)\delta}\right] \\
&= S_1 - S_2
\end{aligned}
$$

where

$$
S_1 = \frac{e^{(2k_{d-2}+3)\delta}}{B_1}\left[B_2^{n-k_{d-2}} - \binom{n-k_{d-2}}{1} e^{2\delta(n-k_{d-2}-1)} B_1 - \binom{n-k_{d-2}}{0} e^{2\delta(n-k_{d-2})}\right]
$$

and

$$
S_2 = e^{(n+k_{d-2}+2)\delta}\left[B_1^{n-k_{d-2}} - \binom{n-k_{d-2}}{1} e^{\delta(n-k_{d-2}-1)} - \binom{n-k_{d-2}}{0} e^{\delta(n-k_{d-2})}\right].
$$

Then

$$
S_1 - S_2 = \frac{e^{(2k_{d-2}+3)\delta} B_2^{n-k_{d-2}}}{B_1} - e^{(n+k_{d-2}+2)\delta} B_1^{n-k_{d-2}} - \binom{n-k_{d-2}}{1} c_1 - \binom{n-k_{d-2}}{0} c_0.
$$

In this case, the coefficients c_i of $\binom{n-k_{d-2}}{i}$ for $i = 0, 1$ can be generalized using the following result, the proof of which is given in the Appendix.

Lemma 2. *In the calculation of $P_n(D = i)$ for $i > 2$, the coefficient c_{d-j+1} of $\binom{n-k_{d-i-1}}{d-j+1}$ for $j = 1, 2, \cdots, d-2$ is*

$$
e^{[(d-1)n+(d-2)]\delta}\left[\sum_{k=1}^{d-1}(-1)^k \frac{e^{\{d-k-1\}\delta}\ e^{-(j-1)(d-k)\delta} B_{d-k-1}^j}{\prod_{i=1}^{d-k-1} B_i \prod_{i=1}^{k-2} B_i}\right] = 0 \tag{3.7}
$$

and the coefficient c_0 of $\binom{n-k_{d-i-1}}{0}$ is

$$
e^{(d-1)(n+1)\delta}\left[\sum_{k=1}^{d-1}(-1)^k \frac{e^{\{d-k\}\delta}}{\prod_{i=1}^{d-k-1} B_i \prod_{i=1}^{k-2} B_i}\right] = (-1)^{d-1}\frac{e^{(d-1)(n+1)\delta}}{\prod_{i=1}^{d-2} B_i}. \tag{3.8}
$$

Thus, $c_1 = 0$ and $c_0 = \frac{e^{2\delta(n+1)}}{B_1}$, by Lemma 2. This yields

$$S_1 - S_2 = \frac{e^{(2k_{d-2}+3)\delta} B_2^{n-k_{d-2}}}{B_1} - e^{(n+k_{d-2}+2)\delta} B_1^{n-k_{d-2}} + \frac{e^{2\delta(n+1)}}{B_1}. \tag{3.9}$$

For $d = 3$, $k_d - 2 = k_1 = 1$ so substitute this in (3.9) to get $P_n(D = 3)$ as follows

$$\begin{aligned} P_n(D=3) &= (1-e^{-\delta})^{n-3} e^{-2(n+1)\delta} \left[\frac{e^{5\delta} B_2^{n-1}}{B_1} - e^{(n+3)\delta} B_1^{n-1} + \frac{e^{2\delta(n+1)}}{B_1} \right] \\ &= \frac{A_3^n}{A_1 A_2 A_3} - \frac{A_2^n}{A_1^2 A_2} + \frac{A_1^n e^{-\delta}}{A_1^2 A_2}. \end{aligned}$$

Note this time that this expression is zero for $n = 1, 2$ (see Lemma 3 below). By multiplying the conditional probability of $D = 3$ given $N = n$ with the marginal probability function of $N = n$, we get $P(D = 3)$ as follows

$$\begin{aligned} P(D=3) &= \sum_{n=3}^{\infty} \frac{(A_3\alpha)^n e^{-\alpha}}{n!\, A_1 A_2 A_3} - \sum_{n=3}^{\infty} \frac{(A_2\alpha)^n e^{-\alpha}}{n!\, A_1^2 A_2} + \sum_{n=3}^{\infty} \frac{(A_1\alpha)^n e^{-\alpha} e^{-\delta}}{n!\, A_1^2 A_2} \\ &= \frac{\zeta_3}{A_1 A_2 A_3} - \frac{\zeta_2}{A_1^2 A_2} + \frac{e^{-\delta}\zeta_1}{A_1^2 A_2} + \frac{\alpha^2 e^{-\alpha}}{2!} h_2 + \frac{\alpha e^{-\alpha}}{1!} h_1 + e^{-\alpha} h_0 \\ &= \frac{\zeta_3}{A_1 A_2 A_3} - \frac{\zeta_2}{A_1^2 A_2} + \frac{e^{-\{2\}\delta}\zeta_1}{A_1^2 A_2} - \frac{e^{-\{3\}\delta}\zeta_0}{A_1 A_2 A_3}, \end{aligned}$$

where

$$\begin{aligned} h_0 &= -\frac{1}{A_1 A_2 A_3} + \frac{1}{A_1^2 A_2} - \frac{e^{-\delta}}{A_1^2 A_2}, \\ h_1 &= -\frac{A_3}{A_1 A_2 A_3} + \frac{A_2}{A_1^2 A_2} - \frac{A_1 e^{-\delta}}{A_1^2 A_2} \\ h_2 &= -\frac{A_3^2}{A_1 A_2 A_3} + \frac{A_2^2}{A_1^2 A_2} - \frac{A_1^2 e^{-\delta}}{A_1^2 A_2} \end{aligned}$$

Notice that Lemma 1 appears when we calculate h_0 which is the coefficient of the $\zeta_0 = e^{-\alpha}$ term. The coefficients of $\frac{\alpha^j e^{-\alpha}}{j!}$ for $j = 1, 2, \cdots, d-1$ are given in the following result, the proof of which uses Theorem 1; we do not yet have a direct proof of it.

Lemma 3. *For $d > 1$, the coefficient of $\frac{\alpha^n e^{-\alpha}}{n!}$ is*

$$h_n = \sum_{k=0}^{d-1} (-1)^k \frac{e^{-\{k\}\delta} A_{d-k}^n}{\prod_{i=1}^{k} A_i \prod_{i=1}^{d-k} A_i} = \sum_{k=0}^{d-1} h_{nk} = 0 \quad \textit{for } n = 1, 2, \ldots, d-1 \tag{3.10}$$

To get $P(D = d)$, sum over all $k_d, k_{d-1}, \cdots k_1$ and do the calculation in the same manner by using (3.1), (3.7), (3.8), and (3.10) at each step to get the result stated in Theorem 1.

3.3 FIRST TWO MOMENTS OF D

We now compare the mean μ and variance σ^2 of D with those of N, the actual Poisson(α) number of photons emitted. The expressions for these features are not tractable but are easy to compute using Corollary 1. We also compare the Fano factor $FF = \sigma^2/\mu$. They all are depicted in FIGURE 8. It can be seen there that the mean is smaller than α; the

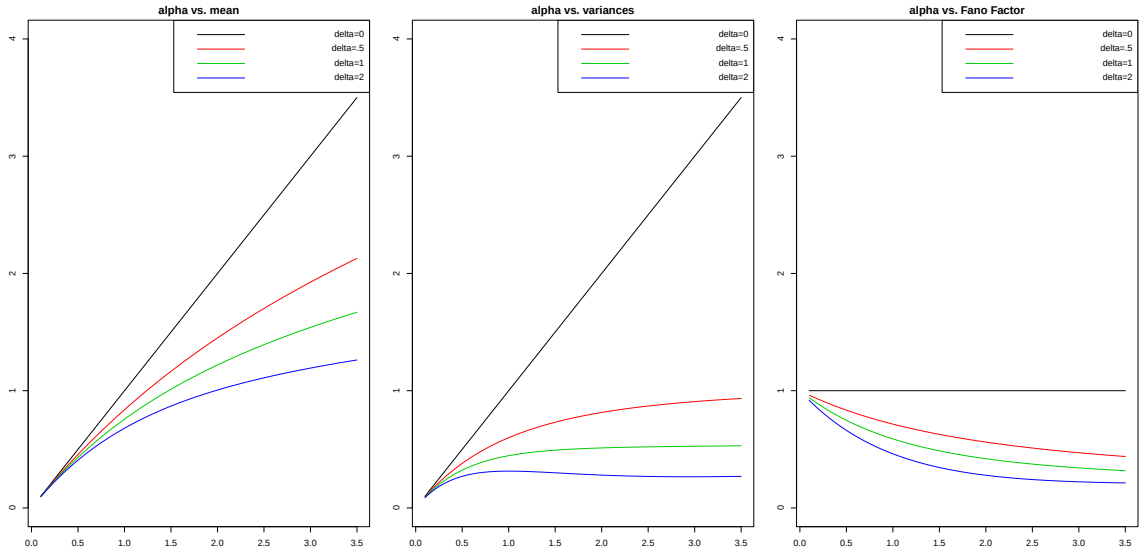

Figure 8: Graph of the (a) mean, (b) variance, (c) Fano factor.

variance is even smaller, so that the Fano factor is less than 1, indicating that the D point process is more regular than a Poisson process. In addition, as the dead period gets larger the moments of D gets further away from the moments of Poisson process. These results are expected because of the dead period.

3.4 INFERENCE ABOUT THE UNDERLYING INTENSITY PARAMETER

We first estimate both (α, δ) simultaneously. We then study inference about α by assuming that the dead period δ is known.

3.4.1 MLE

In applications, when we are unsure about the value of δ, we must estimate both (α, δ) simultaneously given the observed counts $\mathbf{D} = (D_1, \ldots, D_m)$. Write $p_k = p_k(\alpha, \delta)$, and assume that the observations are independent with common distribution (3). The log-likelihood for (α, δ) is

$$\ln L(\alpha, \delta|\mathbf{D}) = \ln \binom{m}{m_0, \cdots, m_C} + \sum_{k=0}^{K} m_k \ln p_k(\alpha, \delta), \tag{3.11}$$

where $K = \max_{1 \le i \le m} D_i$, and $m_k = \sum_{i=1}^{m} I(D_i = k)$. We use standard Newton-Raphson methods to get the maximum likelihood estimates (MLEs) $(\hat{\alpha}, \hat{\delta})$ and approximate standard errors from the Fisher information matrix $\hat{I}(\hat{\alpha}, \hat{\delta})$ using second differences [11]. For numerical purposes, it is more convenient to use the parametrization $(\eta, \theta) = (e^{-\alpha}, e^{-\delta})$, which restricts the search to the unit square $[0, 1]^2$; of course, the MLEs transform easily, as does the Fisher information matrix. In our discussion below we use the original parametrization.

Next, assuming δ is known, and that the observations are independent with common distribution (3), we turn to the problem of estimating the underlying Poisson intensity parameter α given the observed counts $\mathbf{D} = (D_1, \ldots, D_m)$ using standard methods to get the maximum likelihood estimate and its approximate standard error. Then log-likelihood in (3.11) becomes

$$l(\alpha) = \ln L(\alpha|\mathbf{D}) = \ln \binom{m}{m_0, \cdots, m_K} + \sum_{k=0}^{K} m_k \ln p_k(\alpha), \tag{3.12}$$

where $K = \max_{1 \le i \le m} D_i$, and $m_k = \sum_{i=1}^{m} I(D_i = k)$. Unless $K = 1$, the log-likelihood is not tractable, so we use Newton-Raphson iterations to compute the MLE $\hat{\alpha}$ and its estimated standard deviaton $\hat{s}(\hat{\alpha})$ using the observed Fisher information (see Efron and Hinkley (1978)).

The Newton-Raphson steps are

$$\alpha_{n+1} = \alpha_n - \frac{l'(\alpha)}{l''(\alpha)}, \quad \text{where} \quad l'(\alpha) = \frac{dl(\alpha)}{d\alpha} \text{ and } l''(\alpha) = \frac{d^2 l(\alpha)}{d\alpha^2}.$$

Thus, the standard error of $\hat{\alpha}$ is

$$\hat{s}(\hat{\alpha}) = \frac{1}{\sqrt{m\hat{I}(\hat{\alpha})}}.$$

In this section, our main aim is to estimate the underlying intensity parameter α since its estimate is used to get an estimate of the $SNR = \sqrt{\alpha}$ under our model assumptions. Therefore, summaries for the δ known case is summarized in the rest of this section. When there is doubt about the magnitude of the dead period, we estimate both (α, δ) simultaneously: see Section 3.4.1.2.

3.4.1.1 Simulated data

The details of the R code used for our simulations are in the Appendix. One run using the true values $\alpha = 2$, $\delta = 1$, $m = 100$, yielded the following counts: $m_0 = 5$, $m_1 = 50$, $m_2 = 34$, $m_3 = 11$, $m_j = 0$ for $j \geq 4$. The estimated standard deviation of $\hat{\alpha}$ is based on the observed Fisher information: $\hat{I}(\hat{\alpha}) = 0.16$. Thus, the standard error of $\hat{\alpha}$ is

$$\hat{s}(\hat{\alpha}) = \frac{1}{\sqrt{m\hat{I}(\hat{\alpha})}} = .25.$$

Table 8 below is an illustration of Newton Raphson steps when maximizing (3.12).

Table 1: Newton Raphson iterations

Iteration	0	1	2	3	4	5	6
α_n	0.50000	0.86593	1.33038	1.73799	1.96664	2.05825	2.08968
Iteration	7	8	9	10	11	12	13
α_n	2.09989	2.10314	2.10418	2.10451	2.10461	2.10464	2.10465

Figure 9 depicts the log-likelihood which shows that a quadratic approximation is good near the MLE.

3.4.1.2 For the paper fiber data

Our colleagues David Kleinfeld and Philbert Tsai of the Physics Department at UC San Diego provided count data by imaging filter paper under several power settings: 20%, 30%, 40%, 50%, 75%, 100%, which represent the percent (of maximum) laser power used in each

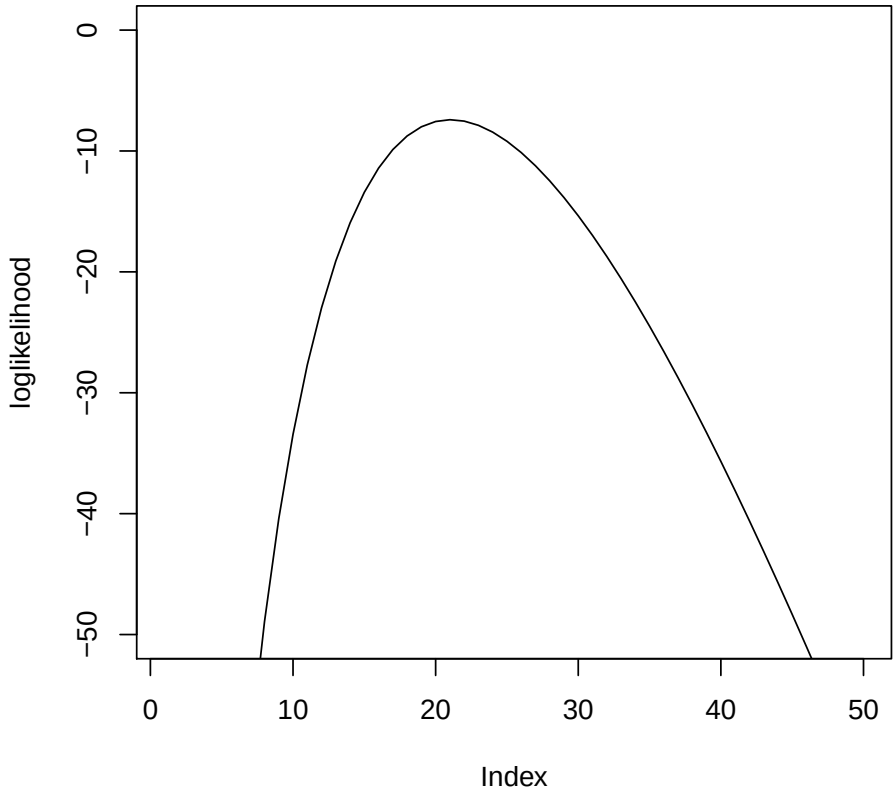

Figure 9: Log-likelihood plot.

run. When processing the filter paper data under each power setting, the pixel dwell time for each data point is 1.6 microseconds and the laser repetition rate is approximately 76 megahertz. This yields a maximum count of 122 laser pulse triggers per pixel. The resulting data sets consist of fluorescence taken from 400-by-400 pixel frame which yields 160,000 unique points along a line through the sample, with the line repeated approximately 50 times. Therefore, we have 160,000 columns for each of which represents 50 repeated trials of data taken from one particular position in the sample.

For illustrative purposes, we use the data from the 20% power setting. In this case, the majority of pixels had 0 counts, indicating dark regions; in the brighter regions, the counter

recorded 1 photon most of the time. The number of 2 counts were considerable even in this low power setting, and there were relatively few 3-or-more counts. Under this low power setting, for each pixel we use the model (3.11) to estimate (α_i, δ_i) simultaneously for pixel i. The majority of the estimates of δ_i were between 1.0 to 1.5 ns, with a median value of approximately 1.5 ns. Thus, we assume here that the dead period is constant across all pixels, so we set $\delta_i \equiv 1.5$ to estimate α_i.

In the dark regions, there were no photon arrivals at all. For such cases, we set $\hat{\alpha} = 0$. For pixels which had at least one photon (among the approximately 122 trials) the range of $\hat{\alpha}$ varies from (.001, 3.159). The smaller α estimates correspond to the dark part of the image and the higher α values correspond to the brighter regions.

3.4.2 Fisher information

We now turn to the problem of obtaining the standard error of the parameter estimates. The earlier papers [14, 15] used the information only about 0 counts and 0 and 1 counts and compared the Fisher information for them. Since the the distribution of D was not known in that study, they used grouping as follows to get the Fisher Information:

- 0 counts and ≥ 1 counts ($I_0 =$ its Fisher information)
- 0 counts, 1 counts, and ≥ 2 counts ($I_1 =$ its Fisher information)

They reported that grouping the data into three groups led to a large improvement in image quality because I_1 is much larger than I_0. We can see that grouping in effect creates incomplete data which results loss of information. Thus, the Fisher information grows as we add more actual (as opposed to grouped) counts in the original data.

When estimating (α, δ) simultaneously, the expected Fisher information is

$$I(\alpha, \delta) = -E\left[H(\alpha, \delta)\right] = -E\begin{bmatrix} \frac{\partial^2 l(\alpha,\delta)}{\partial \alpha^2} & \frac{\partial^2 l(\alpha,\delta)}{\partial \alpha \partial \delta} \\ \frac{\partial^2 l(\alpha,\delta)}{\partial \alpha \partial \delta} & \frac{\partial^2 l(\alpha,\delta)}{\partial \delta^2} \end{bmatrix},$$

where H is the Hessian matrix of the log-likelihood. The corresponding observed Fisher information is

$$\hat{I}(\hat{\alpha}, \hat{\delta}) = -H(\hat{\alpha}, \hat{\delta}) = -\begin{bmatrix} \frac{\partial^2 l(\hat{\alpha},\hat{\delta})}{\partial \alpha^2} & \frac{\partial^2 l(\hat{\alpha},\hat{\delta})}{\partial \alpha \partial \delta} \\ \frac{\partial^2 l(\hat{\alpha},\hat{\delta})}{\partial \alpha \partial \delta} & \frac{\partial^2 l(\hat{\alpha},\hat{\delta})}{\partial \delta^2} \end{bmatrix},$$

where the complete log-likelihood is given in (3.11).

When estimating α assuming that δ is known, the expected and observed Fisher information are, respectively,

$$I(\alpha) = E\left[-l''(\alpha)\right] \quad \text{and} \quad \hat{I}(\hat{\alpha}) = -l''(\hat{\alpha}),$$

where the complete log-likelihood is given in (3.12).

3.4.3 Loss of information due to grouping

Now that we know the general formula for the distribution of the D, we can use this to derive the Fisher Information of the complete data, and it is given in (3.12). As stated in McLachlan and Krishnan [38], grouping the data leads to loss of information; in this section we use their notation. Let $\mathcal{X}$ represent the complete data, and $\mathcal{Y}$ represent the grouped (incomplete) data. Let K represent the conditional distribution of $\mathcal{Y}$ given $\mathcal{X}$. Because the complete data log-likelihood can be written as

$$\ln L_{inc}(\alpha) = \ln L_c(\alpha) - \ln K(d|\alpha),$$

the expected Fisher information is

$$I_{inc}(\alpha) = I_c(\alpha) - I_m(\alpha; d),$$

where $I_{inc}(\alpha)$ is the incomplete information, $I_c(\alpha)$ is the complete information, and $I_m(\alpha; d)$ is the missing information.

Now suppose that we categorize the data into $k+1$ groups: $0, 1, \cdots, k$ counts and all greater counts collapsed into the $\geq k$ group. The monotonicity of the information follows from the fact that $I_k \leq I_{k+1}$, where I_k is the information by based on k groups. Here, I_k is a collapsed version of I_{k+1}. Since the missing information for $k+1$ group is smaller than that of k groups, the monotonicity follows.

3.4.3.1 Simulated data

In this section, we examine how the observed Fisher information changes by estimating I_0, I_1, and I_2. For this purpose, data are simulated by using $\delta = 1$, $\alpha = 2, 2.2, \ldots, 3.0$, and total sample size $m = 100$ for each case. Let m_0 be the total number of 0 counts, m_1 be the total number of 1 counts, m_2 be the total number of 2 counts, and $m_{3+} = 1 - m_0 - m_1 - m_2$ be the total number of ≥ 3 counts. Thus, the incomplete log-likelihood is

$$l_{inc}(\alpha) = \ln L(\alpha|\mathbf{D}) = c + \sum_{k=0}^{2} m_k \ln p_k(\alpha) + m_{3+} \ln p_{3+}(\alpha)$$

where

$$c = \ln \binom{m}{m_0, m_1, m_2, m_{3+}} \quad \text{and} \quad p_{3+}(\alpha) = 1 - p_0 - p_1 - p_2.$$

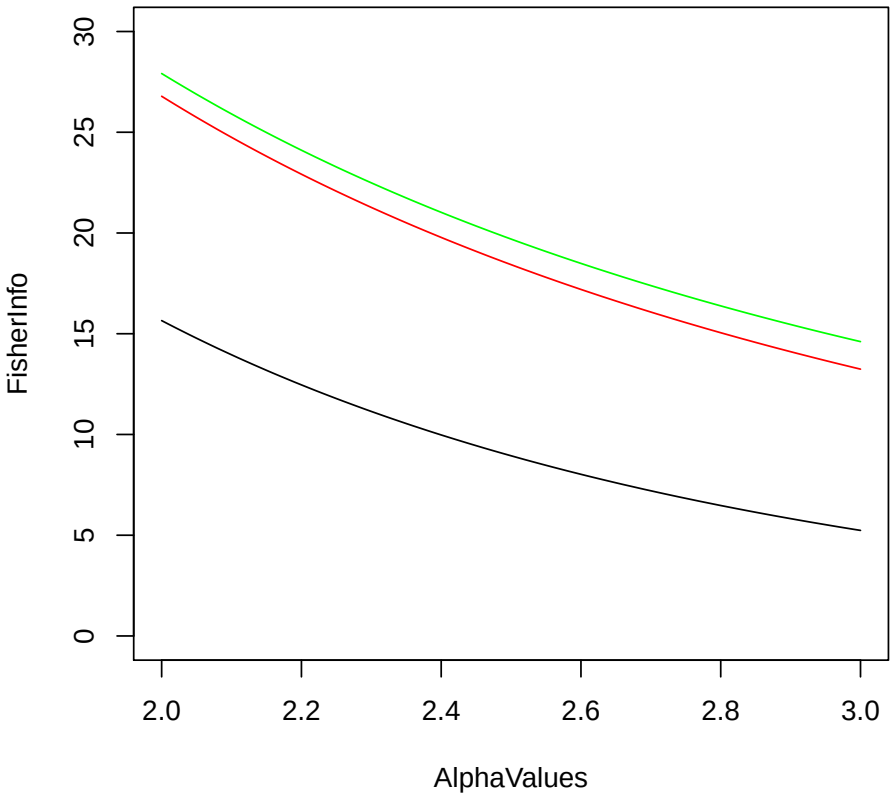

Figure 10: Fisher Information for different groupings.

In Figure 10, we plot the Fisher information for three different groupings: the black curve is for I_0, the red curve is I_1, and the green curve is I_2. Notice that there is a large increase in going from I_0 to I_1, and a much smaller increase to I_2. We will see later that even this smaller improvement in Fisher information leads to improvement in the SNR: see Figure 12.

Next, if we were to study a the finer grouping with counts 0, 1, 2, 3, and 4+ and redo these calculations, the Fisher information I_3 would be quite close to I_2. The reason is that when small α values are of interest, the probabilities $p_3, p_4, \cdots$ are very small, and so we do not gain much information even if we consider additional counts. Thus, we only consider grouping on 0, 1, 2, and ≥ 3 in this study.

For the simulated data in Section 3.4.1.1, using the grouping 0, 1, 2, and 3+, the expected Fisher information is $I_{inc}(\hat{\alpha}) = 25.85$ and the observed Fisher information is obtained as $\hat{I}(\hat{\alpha}) = 37.89$. Following the recommendation of Efron and Hinkley [16], we use the observed Fisher information result to get the standard error of $\hat{\alpha}$.

3.4.3.2 Paper fiber data

The paper fiber data sets consist of fluorescence taken from 400-by-400 pixel frame which yields 160,000 unique points along a line through the sample, with the line repeated approximately 50 times. Therefore, we have 160,000 columns for each of which represents 50 repeated trials of data taken from one particular position in the sample. For illustrative purposes, we use the data from the 20% power setting. In this case, the majority of pixels had 0 counts, indicating dark regions; in the brighter regions, the counter recorded 1 photon most of the time. The number of 2 counts were considerable even in this low power setting, and there were relatively few 3-or-more counts.

In particular, we focus on comparing estimates of α_i from various groupings of counts because their apparatus had difficulty in distinguishing among counts higher than 3. Let $m_{i+} = \sum_{j \geq i} m_j$ be the number of trials with at least i photon counts. Previous work [14] compared the grouping (m_0, m_{1+}) with (m_0, m_1, m_{2+}). Figure 11 depicts the estimates of α_i based only on the zero, one, and two-or-more count summaries. Figure 12 summarizes the differences in estimates of α_i between the grouping (m_0, m_1, m_{2+}) against (m_0, m_1, m_2, m_{3+}). We note that even though there are relatively few three-or-more counts, all of the estimates

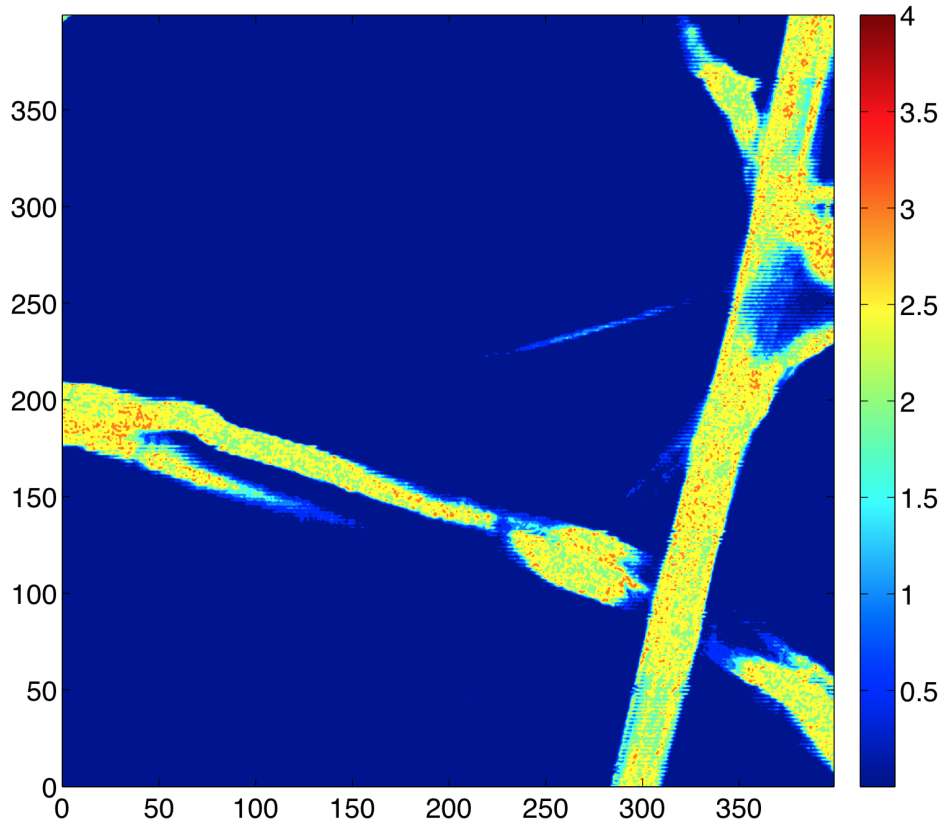

Figure 11: Contour plot of $\hat{\alpha}_i$ estimates based on (m_0, m_1, m_{2+})

for the latter grouping are at least as large as for the former groupings. In the lighter regions the increase in the estimates of α_i is considerably larger.

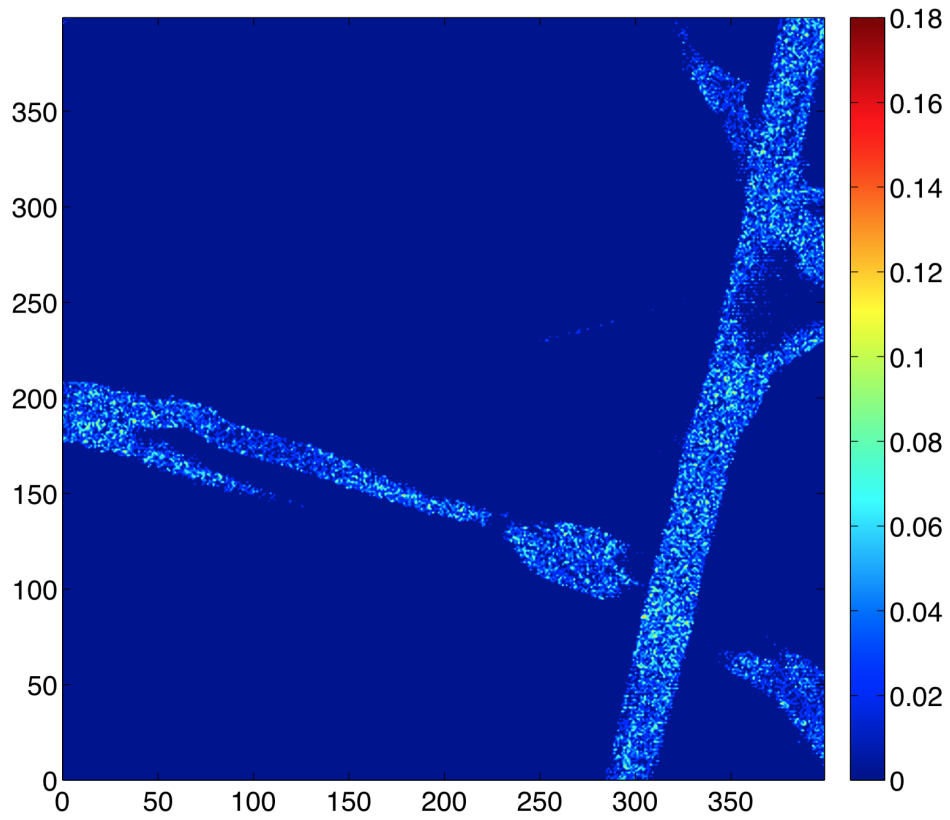

Figure 12: Contour plot of improvement in $\hat{\alpha}_i$ estimates

3.5 CHAPTER SUMMARY

In this chapter, our main aim is to derive a mathematical model that relates the number of photons recorded and the actual emitted photon counts is needed. In particular, we derive the exact distribution of D, the number of photons deteced by using two equivalent approaches. We assumed an infinite observation horizon and a fixed dead period for both of them. Using this model, we are primarily interested in estimating the underlying intensity parameter because its estimate determines the intensity of a single pixel in an image, and it also gives a measure of the SNR. Therefore, we turn to inference on the underlying intensity Poisson parameter (α) and standardized dead time (δ) of the mechanism. Then, we relate our

results to the more general principle of loss of information due to grouping. Next, we used our model on simulated data. Finally, our collaborators extended earlier work by modifying the digital counting electronics used in Driscoll, et al. [14] so that counts could now be grouped in the categories of zero, one, two, three and greater than three. Their data were obtained using higher laser intensities which yield higher emission intensities, or higher Poisson rates α. They initially collected the data on paper fiber for us to get the rescontructed images based on our method. Currently, our collaborators are working on getting us the data on other applications. All proofs of the results presented in this chapter are in the Appendix.

4.0 MODEL EXTENSIONS

In Chapter 3, we developed our model assuming exponential waiting times assumption and an infinite observation period. Our aim in this chapter is to investigate the effect of the variants of these assumptions that are of practical interest. First, the experiments have a finite time horizon T, so it is of interest to know how large T must be in order for the infinite horizon results are good approximations. Next, although exponential waiting times are well supported by theoretical considerations, it is useful to know the effect upon our results of plausible other waiting time distributions, specifically the gamma family. In this chapter, we study these two issues.

4.1 FINITE TIME ASSUMPTION

Earlier, we assumed that the observation period is infinite to derive the distribution of D. In this section, we assume that the observation period is a finite horizon, $[0, T]$, because actual experiments are of course finite length. The calculations for finite time are considerably more complicated because of the possibility of an emission close to T which leads to a truncated dead period. Thus, we only provide details for $d \leq 2$ and use numerical calculations to compare with the infinite horizon results.

4.1.1 $P(D = d)$ for finite horizon based on approach in Section 3.2.1

We use the same assumptions and notation as in Section 3.2.1. If there is no photon arrival to the detector before our observation period T, then we would have $[D = 0]$. Thus, we can

calculate the probability of this event as

$$P(D=0) = \exp\left(-\int_0^T \alpha e^{-x} dx\right) = e^{-\alpha\left(1-e^{-T}\right)}.$$

There are two possible ways to record a single photon. One way is that the first dead period ends before the observation period, T: see Figure 13. In this case, the event $[D=1]$ occurs before time T and there is no emission between the end of the dead period and T.

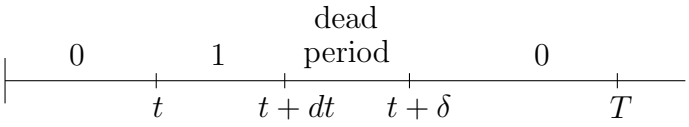

Figure 13: First way to record a single photon when the observation period is finite.

Therefore, the probability of each these intervals given in Figure 13:

- no emissions in $[0, t)$ is

$$\exp\left(-\int_0^t \alpha e^{-x} dx\right) = e^{-\alpha(1-e^{-t})};$$

- an emission in $[t, t+dt)$ is $\alpha e^{-t} dt$
- no emissions after the dead period is

$$\exp\left(-\int_{t+\delta}^T \alpha e^{-x} dx\right) = e^{-\alpha e^{-(t+\delta)}} e^{\alpha e^{-T}}.$$

Note that the first photon must arrive at the detector before time $T-\delta$ in order for $t+\delta < T$ in Figure 13. Thus, $0 < t < T-\delta$.

Putting these pieces together and using the substitution $x = \alpha e^{-t}$, we have the probability of the first way, fw, as

$$\begin{aligned} \text{fw} = e^{\alpha e^{-T}} \int_0^{(T-\delta)} e^{-\alpha(1-e^{-t})} e^{-\alpha e^{-(t+\delta)}} \alpha e^{-t} dt &= e^{\alpha e^{-T}} e^{-\alpha} \int_{\alpha e^{-(T-\delta)}}^{\alpha} e^{x\left(1-e^{-\delta}\right)} dx \\ &= \frac{e^{-\alpha(e^{-\delta}-e^{-T})} - e^{-\alpha(1-e^{-(T-\delta)})}}{1-e^{-\delta}}. \end{aligned} \quad (4.1)$$

The second way of recording a single photon is emitted close to T so that its dead period exceeds T: see Figure 14.

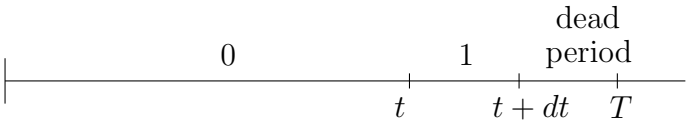

Figure 14: Second way to record a single photon when the observation period is finite.

The probability of each of these intervals given in Figure 14:

- no emissions in $[0,t)$ is
$$\exp\left(-\int_0^t \alpha e^{-x}dx\right) = e^{-\alpha(1-e^{-t})};$$
- an emission in $[t,t+dt)$ is $\alpha e^{-t}dt$.

This happens if a photon arrives after time $T-\delta$, so that $T-\delta < t < T$. Thus, the probability of second way, sw, is

$$\text{sw}= = \int_{(T-\delta)}^{T} e^{-\alpha(1-e^{-t})}\alpha e^{-t}dt = e^{-\alpha}\int_{\alpha e^{-T}}^{\alpha e^{-(T-\delta)}} e^{x}dx = e^{-\alpha(1-e^{-(T-\delta)})} - e^{-\alpha(1-e^{-T})}. \quad (4.2)$$

By summing (4.1) and (4.2), we get the probability of recording one photon:

$$\begin{aligned} P[D=1] &= \text{fw} + \text{sw} \\ &= \frac{e^{-\alpha(e^{-\delta}-e^{-T})} - e^{-\alpha(1-e^{-(T-\delta)})}}{1-e^{-\delta}} + e^{-\alpha(1-e^{-(T-\delta)})} - e^{-\alpha(1-e^{-T})}, \end{aligned} \quad (4.3)$$

which agrees with the formula we derive below using our second approach. Notice that as $T\to\infty$, the expression in (4.3) tends to

$$\frac{\zeta_1}{A_1} - \frac{\zeta_0}{A_1}.$$

We expect that the probability in (4.1) is much larger than that in (4.2). We now want to explore the effect of the probability of seeing a photon very close to T, (4.2), on the

probability of recording one photon, (4.3). In Table 2, we assumed $\delta = 1$, and compare the probabilities of these two cases.

Table 2: Comparison of equation 4.1 with equation 4.2

		T=2	T=4	T=6	T=8	T=10
$\alpha = .5$	for 4.1	.255	.345	.355	.356	.356
	for 4.2	.080	.010	.001	0	0
$\alpha = 1.5$	for 4.1	.503	.556	. 558	.558	.558
	for 4.2	.114	.011	.001	0	0
$\alpha = 2.5$	for 4.1	.559	.513	.502	.501	.501
	for 4.2	.009	0	0	0	0

From this numerical example, we can see that (4.2) is negligible, in which case (4.3) is approximated thus:

$$P(D=1) \simeq \frac{e^{-\alpha(e^{-\delta}-e^{-T})} - e^{-\alpha(1-e^{-(T-\delta)})}}{1-e^{-\delta}}, \tag{4.4}$$

and as $T \to \infty$, we also have (4.4) tends to

$$\frac{\zeta_1}{A_1} - \frac{\zeta_0}{A_1}.$$

The depiction of the event [D=2] is more complicated. We again have two possible ways of recording two photons:

- First case: the second dead period ends before T;
- Second case: the second dead period exceeds T.

In the first situation, we record two photons if no photon emission in $[0,t)$; a photon emission in $[t,t+dt)$; no emission in $[t+\delta,u)$; a photon emission in $[u,u+du)$; no emission in $[u+\delta,T)$. Notice here that $0 < t < T-2\delta$ and $t < u < T-\delta$. By combining the probabilities for these intervals, we get the probability of the first case as

$$\int_{t=0}^{(T-2\delta)} \int_{u=t}^{(T-\delta)} \left(1-e^{-\alpha e^{-t}}\right)\left(e^{-\alpha e^{-(t+\delta)}} - e^{-\alpha e^{-u}}\right)\left(e^{-\alpha e^{-(u+\delta)}} - e^{-\alpha e^{-T}}\right)\alpha^2 e^{-t} e^{-u}\, dt\, du. \tag{4.5}$$

In the second situation, we record two photons if there is no photon emission in $[0, t)$; a photon emission in $[t, t + dt)$; no emission in $[t + \delta, u)$; a photon emission in $[u, u + du)$; and the second dead period exceeds T. Notice here that the second arrival time must be close to T, so this probability will also be small, but not as small as (4.2). However, we do not pursue this calculation because it is quite involved.

4.1.2 Finite time assumption for approach in Section 3.2.2

We start with a Poisson(α) number of photons actually emitted and their arrivals following exponential waiting times with parameter τ. As before, we rescale the dead period $\delta = \Delta/\tau$, and assume $\tau = 1$ without loss of generality. In Section 3.2.2, we used an alternate approach to derive the distribution of D for an infinite observation horizon. Here we derive this for a finite observation time period. Once again, we only do this for $[D = 0]$ and $[D = 1]$ explicitly; we only indicate the steps need to get $P(D = d)$ for larger d because the calculations are very involved.

The case of $P(D = 0)$ is fairly straightforward:

$$\begin{aligned} P(D=0) &= P(N=0) + P(N \geq 1, W_{(1)} > T) \\ &= P(N=0) + \sum_{n=1}^{\infty} P(N=n) P(W_{(1)} > T) \\ &= P(N=0) + \sum_{n=1}^{\infty} P(N=n) e^{-nT} = \sum_{n=0}^{\infty} P(N=n) e^{-nT} \\ &= E(e^{-NT}) = e^{-\alpha(1-e^{-T})}. \end{aligned}$$

However, the case of $P(D = 1)$ already requires the computation of probabilities for many configurations.

$$P(D=1) = \sum_{n=0}^{\infty} P(D=1, N=n) = \sum_{n=0}^{\infty} P_n(D=1) P(N=n)$$

where

$$P_n(D=1) = P(D=1|N=n).$$

The configurations that lead to $[D = 1]$ are the following.

1. When no photon is emitted, $n = 0$, then obviously $P_0(D = 1) = 0$.
2. When a single photon is emitted, $n = 1$, then it is recorded only if it arrives before T. Thus, we have $P_1(D = 1) = P(W_1 \leq T) = (1 - e^{-T})$.
3. When two photons are emitted, $n = 2$, the detector records only one photon if any of the following disjoint events occurs:
   - $[W_{(1)} \in (T - \delta, T)]$ with probability

$$P(T - \delta \leq W_{(1)} \leq T) = P\left(t - \delta \leq \frac{X_1}{2} \leq T\right) = e^{-2(T-\delta)} - e^{-2T} = e^{-2T}(e^{2\delta} - 1);$$

   - $[W_{(1)} \leq T - \delta,\, W_{(2)} \leq W_{(1)} + \delta]$ with probability

$$\begin{aligned} P(W_{(1)} \leq T - \delta, W_{(2)} \leq W_{(1)} + \delta) &= P\left(\frac{X_1}{2} \leq T - \delta, X_2 \leq \delta\right) \\ &= \left(1 - e^{-2(T-\delta)}\right)(1 - e^{-\delta}); \end{aligned}$$

   - $[W_{(1)} \leq T - \delta,\, W_{(2)} > T]$ with probability

$$\begin{aligned} P(W_{(1)} \leq T - \delta, W_{(2)} > T) &= P\left(\frac{X_1}{2} \leq T - \delta, \frac{X_1}{2} + X_2 > T\right) \\ &= \int_0^{T-\delta} 2e^{-2x}e^{-(T-x)}\,dx \\ &= 2e^{-2T}\int_0^{T-\delta} e^{-x}\,dx = 2e^{-T}\left(1 - e^{-(T-\delta)}\right). \end{aligned}$$

   Thus, $P_2(D = 1)$ is the sum of the probabilities of these events:

$$P_2(D = 1) = e^{-2T}(e^{2\delta} - 1) + \left(1 - e^{-2(T-\delta)}\right)(1 - e^{-\delta}) + 2e^{-T}\left(1 - e^{-(T-\delta)}\right).$$

4. When at least three photons are emitted, $n \geq 3$, the detector would record a photon if one of the following disjoint events occurs:
   - $[W_{(1)} \in (T - \delta, T)]$ with probability

$$P(T - \delta \leq W_{(1)} \leq T) = P\left(t - \delta \leq \frac{X_1}{n} \leq T\right) = e^{-nT}(e^{n\delta} - 1);$$

- $[W_{(1)} \le T-\delta$ and $W_{(n)} \le W_{(1)}+\delta]$ with probability

$$\begin{aligned} P(W_{(1)} \le T-\delta, W_{(n)} \le W_{(1)}+\delta) &= P\left(\frac{X_1}{n} \le T-\delta, \sum_{j=1}^{n} \frac{X_{(j)}}{n-j+1} \le \frac{X_1}{n}+\delta\right) \\ &= P[X_1 \le n(T-\delta)] P\left(\sum_{j=2}^{n} \frac{X_{(j)}}{n-j+1} \le \delta\right) \\ &= \left(1-e^{-n(T-\delta)}\right)\left(1-e^{-\delta}\right)^{n-1}; \end{aligned} \tag{4.6}$$

- $[W_{(1)} \le T-\delta$ and $W_{(2)} > T]$ with probability

$$\begin{aligned} P(W_{(1)} \le T-\delta, W_{(2)} > T) &= P\left(\frac{X_1}{n} \le T-\delta, \frac{X_1}{n}+\frac{X_2}{n-1} > T\right) \\ &= \int_0^{T-\delta} n e^{-nx} e^{-(n-1)(T-x)}\, dx \\ &= n e^{-(n-1)T}\left(1-e^{-(T-\delta)}\right); \end{aligned} \tag{4.7}$$

- $[W_{(1)} \le T-\delta$, $W_{(j)} \le W_{(1)}+\delta$, $W_{(j+1)} > T]$ for $2 \le j \le n-1$ with probability

$$P(W_{(1)} \le T-\delta,\ W_{(j)} \le W_{(1)}+\delta,\ W_{(j+1)} > T]) = P(Y_1 \le T-\delta,\ Y_2 < \delta, \sum_{i=1}^{3} Y_i > T)$$

where $Y_1 = X_1/n$ has pdf ne^{-nx}, and

$$Y_2 = \sum_{i=2}^{j} \frac{X_i}{n-i+1}$$

has pdf

$$g_{Y_2}(u) = (j-1)\binom{n-1}{j-1}(1-e^{-u})^{j-2} e^{-(n-j+1)u}$$

because it has the same distribution as that of the $(j-1)$st order statistic from a sample size $(n-1)$, and $Y_3 = X_{j+1}/(n-j)$ has pdf $(n-j)e^{-(n-j)z}$. Thus conditioning on Y_1 and Y_2, we finally get the probability

$$\begin{aligned} &= \int_0^{T-\delta} n e^{-nx} P\left(\sum_{k=j-1}^{n-1} \frac{X_k}{n-k+1} \le \delta, \sum_{k=j-1}^{n-1} \frac{X_k}{n-k+1} + \frac{X_{j+1}}{n-j} > T-x\right) dx \\ &= \int_0^{T-\delta} n e^{-nx} \left(\int_0^{\delta} g_{A_2}(u) e^{-(n-j)(T-u-x)} du\right) dx \\ &= \int_0^{T-\delta} n e^{-nx} \binom{n-1}{j-1} e^{-(n-j)(T-x)} \int_0^{\delta} (j-1)(1-e^{-u})^{j-2} e^{-u} du\, dx \\ &= \binom{n}{j} (1-e^{-\delta})^{j-1} e^{-(n-j)T} (1-e^{-j(T-\delta)}). \end{aligned} \tag{4.8}$$

Note that, in (4.8), if we set $j = n$, we get (4.6); if we set $j = 1$, we get (4.7); if we set $j = 0$, we get 0.

Combining these terms, we get

$$\begin{aligned} P_n(D=1) &= e^{-nT}(e^{n\delta}-1) + \sum_{j=1}^{n} \binom{n}{j} (1-e^{-\delta})^j e^{-(n-j)T}(1-e^{-j(T-\delta)}) \\ &= e^{-nT}(e^{n\delta}-1) + \frac{1}{(1-e^{-\delta})}\left[(1-e^{-\delta}+e^{-T})^n - \left((1-e^{-\delta})e^{-(T-\delta)}+e^{-T}\right)^n\right] \\ &= e^{-nT}(e^{n\delta}-1) + \frac{\left[(1-e^{-\delta}+e^{-T})^n - e^{-n(T-\delta)}\right]}{(1-e^{-\delta})} \end{aligned} \tag{4.9}$$

In fact, the expression (4.9) works for all $n = 0, 1, 2, \ldots$.

Next, we use the generating function of a Poisson random variable

$$E(u^N) = \sum_{n=0}^{\infty} u^n \, P(N=n) = e^{-\alpha(1-u)}$$

to sum the series in (4.9) to get $P(D = 1)$.

$$\begin{aligned} P(D=1) &= \sum_{n=1}^{\infty} P_n(D=1)P(N=n) \\ &= \frac{e^{-\alpha(e^{-\delta}-e^{-T})} - e^{-\delta}e^{-\alpha(1-e^{-(T-\delta)})} - (1-e^{-\delta})e^{-\alpha(1-e^{-T})}}{1-e^{-\delta}} \end{aligned} \tag{4.10}$$

The expression in (4.10) tends to the expression derived by using the infinite observation time period:

$$\frac{e^{-\alpha e^{-\delta}} - e^{-\alpha}}{1-e^{-\delta}} = \frac{\zeta_1}{A_1} - \frac{\zeta_0}{A_1}$$

as $T \to \infty$.

Now we outline the approach to $[D = 2]$. When $n \geq 2$, we record the first and k^{th} photons $(k = 2, \ldots, n)$ under the following configurations.

- $T - 2\delta < W_{(1)} < T - \delta$, $W_{(1)} \leq W_{(k-1)} \leq W_{(1)} + \delta$, and $W_{(1)} + \delta \leq W_{(k)} \leq T$;
- $W_{(1)} < T - 2\delta$, $W_{(1)} \leq W_{(k-1)} \leq W_{(1)} + \delta$, and $T - \delta \leq W_{(k)} \leq T$;
- $W_{(1)} < T - 2\delta$, $W_{(1)} \leq W_{(k-1)} \leq W_{(1)} + \delta$, $W_{(k-1)} + \delta \leq W_{(k)} \leq T - \delta$, and L_k where $L_n = \Omega, \quad L_{n-1} = [W_{(1)} > T] \cup [W_{(n)} \leq W_{(n-1)} + \delta]$, and for $k = 2, 3, \ldots, n-2$, $L_k = [W_{(k+1)} > T] \cup [W_{(n)} \leq W_{(k)} + \delta] \cup \left(\cup_{j=k+1}^{n-1} [W_{(j)} \leq W_{(k)} + \delta] \cap [W_{(j+1)} > T]\right)$.

It is clear that the number of configurations increases considerably for calculating the probabilities of higher photon counts. Since the computations of $P(D = d)$ for $d \geq 2$ are much more cumbersome, we do not pursue these steps.

4.1.3 How large T should be?

We now assess how large T should be in order to use infinite observation time period as an approximation for the finite horizon. We compare the probabilities $P(D = 0)$ and $P(D = 1)$ under both assumptions. Recall that for the finite case,

$$P(D = 0) = e^{-\alpha(1-e^{-T})},$$

$$P(D = 1) = \frac{e^{-\alpha(e^{-\delta}-e^{-T})} - e^{-\delta}e^{-\alpha(1-e^{-(T-\delta)})} - (1 - e^{-\delta})e^{-\alpha(1-e^{-T})}}{1 - e^{-\delta}},$$

and for the infinite case

$$P(D = 0) = e^{-\alpha} \quad \text{and} \quad P(D = 1) = \frac{e^{-\alpha e^{-\delta}} - e^{-\alpha}}{1 - e^{-\delta}}.$$

Table 3: Value $P(D = 0)$ under different model assumption

	T=1	T=2	T=3	T=4	T=5	T=6	T=7	T=8	T=∞
$\alpha = .5$	.729	.649	.622	.612	.609	.607	.607	.607	.607
$\alpha = 1$	.532	.421	.387	.375	.370	.369	.368	.368	.368
$\alpha = 1.5$	.387	.273	.240	.229	.225	.224	.223	.223	.223
$\alpha = 2$	.282	.177	.150	.140	.137	.136	.136	.135	.135
$\alpha = 2.5$	.205	.115	.093	.086	.084	.083	.082	.082	.082
$\alpha = 3$	.150	.075	.058	.053	.051	.050	.050	.050	.050

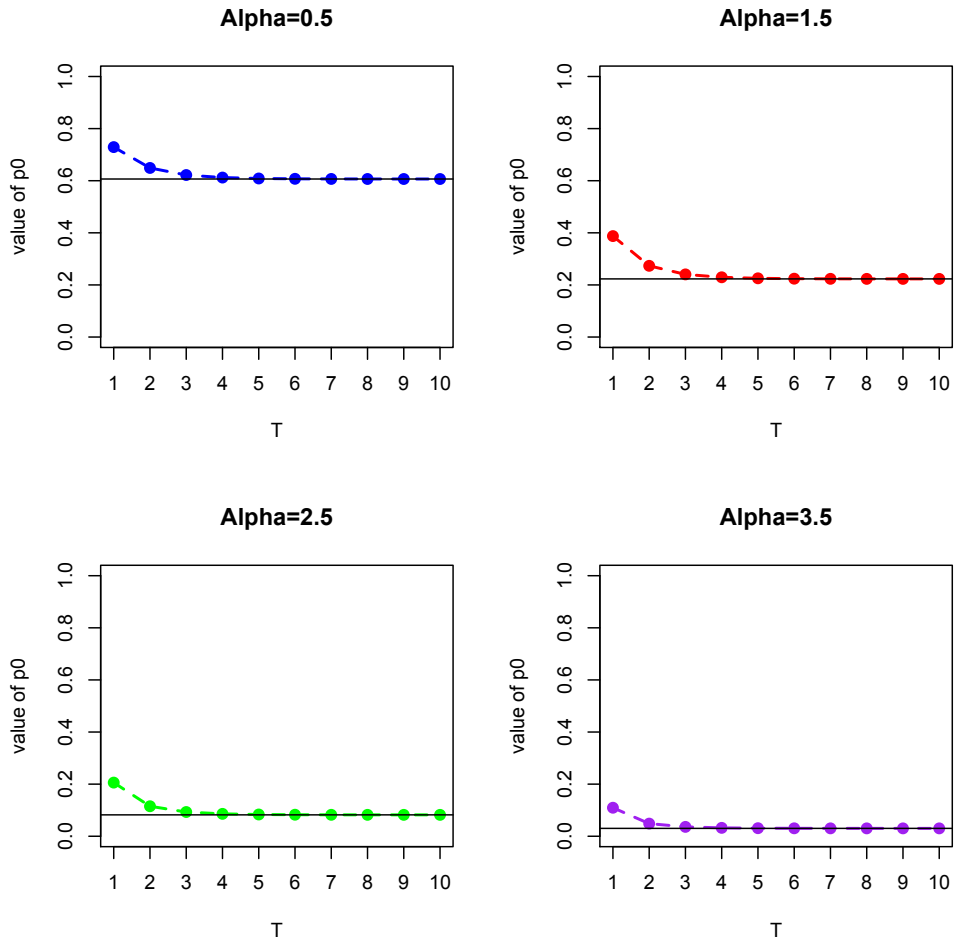

Figure 15: Comparison of values of $P(D = 0)$ under finite observation period assumption (colored line) and infinite observation period (black line).

In Table 3, we compare the value $P(D = 0)$ under finite and infinite observation period assumption. From this table, we can see that as the observation period T grows, the probability of seeing nothing shrinks because the longer we wait the more likely we are to see a photon. In addition to this, as T gets larger, $P(D = 0)$ for the finite case tends to that for the infinite case. And as α gets larger the value of $P(D = 0)$ converges faster as $T \to \infty$. Figure 15 depicts the Table 3.

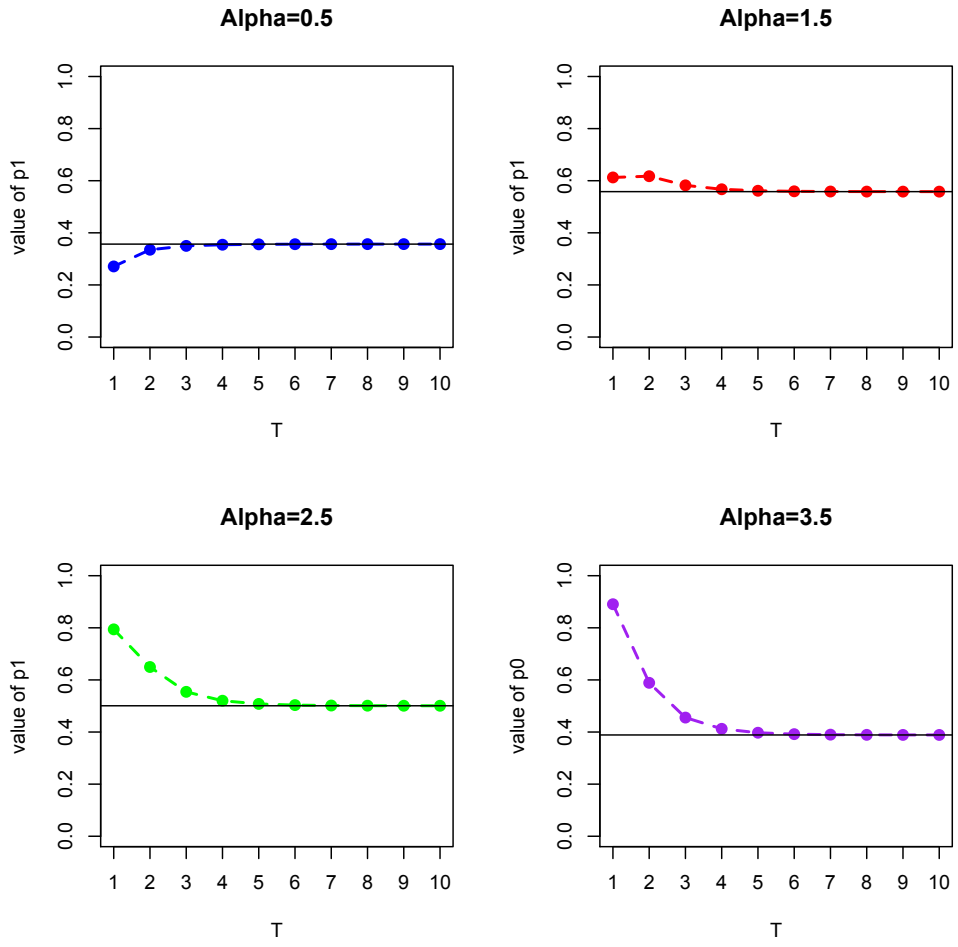

Figure 16: Comparison of values of $P(D = 1)$ under finite observation period (colored line) and infinite observation period (black line).

Now, we do same study for values of $P(D = 1)$; the results are summarized in Table 4. We can conclude that when $\alpha \leq 1$ and observation period is small, we underestimate the probability of seeing a photon: see Figure 16 for $\alpha = 0.5$. This probability tends to that for the the infinite observation case. For $\alpha > 1$, the finite case over-estimates the value of $P(D = 1)$ if the observation period is small. However, its value gets closer to that of from infinite case as we observe longer period of time. To sum up, from these tables, we can see that it is reasonable to assume an infinite observation period because our collaborators use $T \geq 8$ ns.

Table 4: When $\delta = 1$, the value p_1 under different model assumption

	T=1	T=2	T=3	T=4	T=5	T=6	T=7	T=8	T=∞
$\alpha = .5$	.271	.335	.350	.354	.356	.357	.357	.357	.357
$\alpha = 1$	.469	.523	.519	.516	.514	.513	.513	.513	.513
$\alpha = 1.5$	.613	.617	.582	.567	.562	.559	.559	.558	.558
$\alpha = 2$	.718	.1652	.585	.559	.549	.546	.545	.544	.544
$\alpha = 2.5$	.794	.650	.554	.520	.508	.503	.502	.501	.501
$\alpha = 3$	.850	.625	.508	.468	.454	.449	.447	.446	.446

4.2 GAMMA WAITING TIMES ASSUMPTION

When $N = n \geq 1$, the waiting times $W_1, \ldots, W_n$ for the emissions of the photons are assumed to be W_i are independent exponential random variables with time constant τ nanoseconds (ns). We then we use the rescaled dead period $\delta = \Delta/\tau$, so we could assume that $\tau = 1$ without loss of generality. We now change the assumption on the waiting times: suppose that W_i are independent Gamma$(k, 1)$.

To demonstrate the general approach we calculate the probabilities $P(D = d)$ for $d = 0, 1, 2, 3$ under this new assumption using only our first approach given in Section 3.2.1. Under this model setting, we have an inhomogeneous Poisson process $N(t)$ with intensity

$$\lambda(t) = \alpha f(t) = \alpha \frac{t^{k-1}\, e^{-t}}{\Gamma(k)}.$$

We also assume that the observation period is infinite. We use the same notation that as in section 3.1. We also introduce the incomplete gamma function

$$M(t) = \int_t^\infty x^{k-1}\, e^{-x}\, dx.$$

We also introduce new notation for the component probabilities depicted in Figure 6 and 7 in Section 3.2.1: p_{ij} is the probability of the jth interval for the case $[D = i]$.

Since $[D = 0] = [N = 0]$, we have

$$P[D = 0] = P[N = 0] = e^{-\alpha} = \zeta_0.$$

The event $[D = 1]$ is depicted in Figure 6. The probability of no emission in $[0, t)$ is

$$\begin{aligned} p_{11}(t, \alpha, \delta) &= \exp\left(-\int_0^t \alpha \frac{x^{k-1}\, e^{-x}}{\Gamma(k)} dx\right) \\ &= \exp\left(-\int_0^\infty \frac{x^{k-1}\, e^{-x}}{\Gamma(k)} dx + \int_t^\infty \alpha \frac{x^{k-1}\, e^{-x}}{\Gamma(k)} dx\right) \\ &= \exp\left(-\alpha + \frac{\alpha}{\Gamma(k)} M(t)\right); \end{aligned}$$

the probability of an emission in $[t, t + dt)$ is

$$p_{12}(t, \alpha, \delta) = \alpha f(t) = \alpha \frac{t^{k-1}\, e^{-t}}{\Gamma(k)}\, dt;$$

and the probability of no emission in $[t + \delta, \infty)$ is

$$p_{13}(t, \alpha, \delta) = \exp\left(-\int_{t+\delta}^\infty \alpha \frac{x^{k-1}\, e^{-x}}{\Gamma(k)} dx\right) = \exp\left(-\frac{\alpha}{\Gamma(k)} M(t + \delta)\right).$$

By combining these terms, we get

$$P(D = 1) = \int_0^\infty p_{11}(t, \alpha, \delta)\, p_{12}(t, \alpha, \delta)\, p_{13}(t, \alpha, \delta) dt.$$

Similarly, we obtain the probability of $[D = 2]$ by multiplying the probabilities of each interval in Figure 7:

$$p_{21}(t, \alpha, \delta) = p_{11}(t, \alpha, \delta), \quad p_{22}(t, \alpha, \delta) = p_{12}(t, \alpha, \delta),$$

$$p_{23}(t, u, \alpha, \delta) = \exp\left(-\int_{t+\delta}^u \alpha \frac{x^{k-1}\, e^{-x}}{\Gamma(k)} dx\right) = \exp\left(\frac{\alpha}{\Gamma(k)}[M(u) - M(t + \delta)]\right),$$

$$p_{24}(u, \alpha, \delta) = \alpha \frac{u^{k-1}\, e^{-u}}{\Gamma(k)} d_u,$$

and

$$p_{25}(u, \alpha, \delta) = \exp\left(-\int_{u+\delta}^\infty \alpha \frac{x^{k-1}\, e^{-x}}{\Gamma(k)} dx\right) = \exp\left(-\frac{\alpha}{\Gamma(k)} M(u + \delta)\right).$$

Thus,

$$P(D=2)=\int_0^{\infty}\int_{t+\delta}^{\infty} p_{21}(t,\alpha,\delta)\,p_{22}(t,\alpha,\delta)\,p_{23}(t,u,\alpha,\delta)\,p_{24}(u,\alpha,\delta)\,p_{25}(u,\alpha,\delta)dudt.$$

Since these probabilities cannot be written in a compact form, we do not to include the derivation of the $[D=d]$ for $d>2$. However, for $d=0,1,2,3$, we evaluated the probabilities numerically for $k=1,2,3,4$ and $(\alpha,\delta)=(2,1)$: see Table 5 below. We include $k=1$, the exponential case, to verify our computation and to compare with various gammas.

Table 5: First four probabilities based on gamma assumption

	gamma(1,1)	gamma(2,1)	gamma(3,1)	gamma(4,1)
$P(D=0)$	.1353	.1353	.1353	.1353
$P(D=1)$	.4116	.3468	.3266	.3167
$P(D=2)$	.3362	.3273	.3167	.3101
$P(D=3)$	.0908	.1490	.1623	.1678

The first row of this table is independent of waiting time because when there are no photons emitted, waiting times are irrelevant. The probability of recording a one or two photons under gamma assumption is smaller than that for the exponential. On the other hand, the probability of recording three or more than photons (which can be seen clearly from table 6) under gamma waiting times is larger than that of exponential waiting times.

Next, we study the effect of the shape parameter of the gamma on $P(D=d)$. For this purpose, we used simulation to estimate the probabilities. Recall that as the shape parameter k increases, the waiting time for a new photon to be emitted has longer expectation. Hence, a new photon arrival to the detector has higher probability of being outside of the detector's dead period. That is why we are more likely to see higher photon counts as k gets larger. This makes the probability of recording more than two photons increases with the increase in k.

To illustrate this we did a simulation, the details of which are given in the Appendix. The results for the Poisson-Gamma model with $\alpha = 2$, $\delta = 1$, and $k = .5, 1, 2, 4, 8, 16$ are summarized in Table 6.

Table 6: Probabilities for Poisson-gamma model

	Gamma(k,1)						
	k=.5	k=1	k=2	k=4	k=8	k=16	Poisson(2)
$P(D=0)$	.1353	.1353	.1353	.1353	.1353	.1353	.1353
$P(D=1)$	.5317	.4181	.3561	.3244	.3087	.2957	.2707
$P(D=2)$	.2809	.3315	.3230	.3066	.2964	.2884	.1805
$P(D=3)$	.0473	.0993	.1432	.1632	.1660	.1725	.0902
$P(D=4)$	.0027	.0132	.0355	.0554	.0704	.0779	.0361
$P(D=5)$	.0001	.0006	.0045	.0118	.0177	.0229	.1203
$P(D=6)$	0	0	.0004	.0012	.0032	.0046	.0034
$P(D=7)$	0	0	0	.0001	.0003	.0006	.0009

From Table 6, we can see that as k increases, the waiting time for a new photon arrival gets larger. We have a higher chance of detecting more photons because they are more likely to arrive the detector after its dead period. Thus, we can conclude that the dead period becomes less important as k increases because $P(\text{Gamma}(k,1) \leq \delta) \to 0$. Thus, as the shape parameter gets larger, the limiting distribution tends to the original Poisson(α) counts, which are given in the last column of Table 6.

Based on this simulation study, we see how the waiting times of the photon arrivals plays a crucial role in this counting process. If the photons arrive too soon to the detector, they have a high probability to be within the dead period. Longer waiting times make the effect of the dead period on the counting process less important. In particular, longer waiting times lead to smaller undercounts, which in turn allows us to capture the underlying counting process more accurately.

4.3 CHAPTER SUMMARY

In this chapter, we consider several extensions of our model assumptions and derive the distribution of D under them. We first address the finite horizon assumption using our two approaches from Chapter 3. In both cases, we only calculate explicitly $P(D = 0)$ and $P(D = 1)$, and sketch the higher count cases because they are too involved. We also give an approximation for $P(D = 1)$ for the finite case, and assess its accuracy. We then compute the probabilities for various values of T: our numerical work shows that when the observation period is at least 8 ns, we can safely assume an infinite horizon instead of a finite one, leading to simpler expressions for the probabilities. We next propose the gamma assumption for the waiting times. We compare its performance with both exponential times and underlying Poisson process. We conclude that as the shape parameter of the gamma gets larger, we have actual Poisson counts as the limiting distribution.

5.0 APPROXIMATING THE CONWAY-MAXWELL-POISSON NORMALIZING CONSTANT

The Conway-Maxwell-Poisson distribution is a generalization of the Poisson distribution; it is used to model both under-dispersed and over-dispersed data. This distribution involves a normalizing constant which appears also in the moments and MLE calculations of the parameters. Thus, there is no close form for the moments and MLEs, and accurate approximations of the normalizing constant are necessary. There have been some studies that have described the computational schemes and handy approximations for the normalizing constant. However, these approximations perform well only in some certain parameter ranges. Therefore, in this chapter, we propose several new approximations for different parameter ranges where the earlier approximations perform poorly.

We first describe the Conway-Maxwell-Poisson distribution, we then summarize the earlier approximation. Next, we demonstrate the details of our approach. We end this chapter by illustrating the performance of our approximation with a comparison of our approximations with the existing approximation [57].

5.1 CONWAY-MAXWELL-POISSON DISTRIBUTION

For modeling count data the Poisson distribution is widely used. However, in many applications the equality of the mean and variance of the Poisson is often too restrictive. Overdispersion (underdispersion), where the variance is greater (less) than the mean, are commonly encountered in count data [8, 48]; in such cases, the Poisson model can be a poor fit. The Conway-Maxwell-Poisson family is a useful alternative in such cases. Conway and

Maxwell [9] introduced it for studying queuing systems with state-dependent service rates. The Conway-Maxwell-Poisson is a two-parameter family of distributions on the nonnegative integers. Its parameters λ and ν model the intensity and the dispersion, respectively. Since it has two parameters, it is flexible enough to fit a wide range of count data than the Poisson.

For integer $k \geq 0$, the Conway-Maxwell-Poisson variate Y has the probabilities

$$P(Y = k|\lambda, \nu) = C(\lambda, \nu)^{-1} \frac{\lambda^k}{(k!)^\nu}, \quad \text{where} \quad C(\lambda, \nu) = \sum_{k=0}^{\infty} \frac{\lambda^k}{(k!)^\nu}.$$

The parameter space is

$$\Theta = \{(\lambda, \nu) : \lambda > 0, 0 < \nu \leq \infty\} \cup \{(\lambda, \nu) : 0 < \lambda < 1, \nu = 0\}.$$

It can be seen that the series $\frac{\lambda^j}{(j!)^\nu}$ converges for any $\lambda > 0$ and $\nu > 0$ because the ratio of two sunsequent terms of the series $\frac{\lambda}{j^\nu}$ goes to 0 as $j \to \infty$.

There were no systematic studies of the probabilistic or statistical properties of the Conway-Maxwell-Poisson family until the series of papers by Shmueli and her colleagues [39, 54, 50]. In their work, they studied this family in great detail. They showed that the Conway-Maxwell-Poisson is an exponential family, can fit overdispersed and underdispersed data, and can be modified to account for zero-inflated data. They also provide the details of standard approaches to maximum likelihood and Bayes estimation of the parameters, and gave an approximation to the normalizing constant of this distribution. Below, we give a brief summary of the Conway-Maxwell-Poisson distribution.

5.1.1 Special cases of Conway-Maxwell-Poisson

When $\nu = 0$ and $0 < \lambda < 1$, the normalizing constant of Conway-Maxwell-Poisson distribution is just a geometric sum:

$$C(\lambda, \nu) = \sum_{k=0}^{\infty} \lambda^k = \frac{1}{1 - \lambda},$$

and the distribution itself reduces to

$$P(X = x|\lambda, \nu) = \lambda^x (1 - \lambda) \quad \text{for} \quad x = 0, 1, 2, \ldots$$

When $\nu = 0$ and $\lambda \geq 1$, $C(\lambda, \nu)$ diverge; thus, the distribution is undefined. When $\nu = 1$, the normalizing constant of Conway-Maxwell-Poisson distribution is

$$C(\lambda, \nu) = e^{\lambda},$$

so the distribution reduces to

$$P(X = x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!} \quad \text{for} \quad x = 0, 1, 2, \ldots$$

As $\nu \to \infty$, $C(\lambda, \nu) \to 1 + \lambda$; hence the distribution itself approaches to Bernoulli distribution with probability of succsess

$$P(X = 1) = \frac{\lambda}{1 + \lambda}.$$

To sum up, the Conway-Maxwell-Poisson includes three well-known distributions as special cases: geometric when $\nu = 0$ and $0 < \lambda < 1$; Poisson when $\nu = 1$; and Bernoulli with parameter $p = \lambda/(\lambda + 1)$ when $\nu \to \infty$. Hence, it acts as a bridge between these distributions. This also holds for the sum of n independent Conway-Maxwell-Poisson random variables: when $\nu = 0$ and $0 < \lambda < 1$, the distribution of sums reduces to negative binomial distribution with parameters n and $1 - \lambda$; when $\nu = 1$, the sum has a Poisson distribution with parameter $n\lambda$; when $\nu \to \infty$, the sum of Conway-Maxwell-Poisson random variables is binomial with parameters n and $p = \lambda/(1 + \lambda)$.

Imoto [28] proposed a three-parameter generalized Conway-Maxwell-Poisson distribution so that it includes the negative binomial distribution as a special case. We omit the details of this generalization.

5.1.2 Moments

The Conway-Maxwell-Poisson distribution has an exponential family form given by

$$P(X = x|\lambda, \nu) = \exp\left[-\log(C(\lambda, \nu)) + x\log(\lambda) - \nu\log(x!)\right].$$

The moments of the sufficient statistics of the Conway-Maxwell-Poisson distribution,

$$E(X) = \lambda\frac{d\log(C(\lambda, \nu)}{d\lambda} \quad \text{and} \quad E(\log(X!)) = -\frac{d\log(C(\lambda, \nu)}{d\nu}$$

both involve the normalizing constant $C(\lambda, \nu)$. Since these equations cannot solved analytically, we need good approximations for the normalizing constant in order to approximate moments well. Shmueli, et al. [54] approximated this by

$$E(X) \sim \lambda^{1/\nu} - \frac{\nu - 1}{2\nu}. \tag{5.1}$$

More generally, the moments of Conway-Maxwell-Poisson distribution can be expressed recursively because this distribution belongs to the family of two-parameter power series distributions [26, 54].

$$E(X^{r+1}) = \begin{cases} \lambda\, E(X+1)^{1-\nu} & \text{if } r = 0, \\ \lambda\frac{d\,E(X^r)}{d\lambda} + E(X)E(X^r) & \text{if } r > 0. \end{cases}$$

Because the calculation of the moments involve the normalizing constant, for integer values of ν, Nadarajah [41] represented this by rewriting $C(\lambda, \nu)$ in terms of the generalized hypergeometric function. She used the following representation of the generalized hypergeometric function:

$${}_pF_q(a_1, \ldots, a_p; b_1, \ldots, b_p; x) = \sum_{k=0}^{\infty} \frac{(a_1)_k(a_2)_k \ldots (a_p)_k}{(b_1)_k(b_2)_k \ldots (b_p)_k}\frac{x^k}{k!}$$

where $(c)_k = c(c+1)\ldots(c+k+1)$ is the ascending factorial. Later we will use a slightly different representation of the generalized hypergeometric function than this one.

It is clear that

$${}_0F_{\nu-1}(;1, \ldots, 1; \lambda) = C(\lambda, \nu)$$

Thus, the probability generating function of Conway-Maxwell-Poisson distribution is

$$G(s) = E(s^X) = \frac{{}_0F_{\nu-1}(;1,\ldots,1;\lambda s)}{{}_0F_{\nu-1}(;1,\ldots,1;\lambda)}.$$

Then, the mean of Conway-Maxwell-Poisson distribution is

$$E(X) = \left[\frac{dG(s)}{ds}\right]_{s=0} = \lambda\frac{{}_0F_{\nu-1}(;2,\ldots,2;\lambda)}{{}_0F_{\nu-1}(;1,\ldots,1;\lambda)}.$$

In general, she obtained the rth factorial moment of the Conway-Maxwell-Poisson random variate X:

$$E(X(X-1)\ldots(X-k+1)) = \frac{\lambda^k}{(k!)^{\nu-1}}\frac{{}_0F_{\nu-1}(;k+1,\ldots,k+1;\lambda)}{{}_0F_{\nu-1}(;1,\ldots,1;\lambda)}.$$

From this, we can easily get the variance of X as

$$E(X) = \frac{\lambda^2}{2^{\nu-1}}\frac{{}_0F_{\nu-1}(;3,\ldots,3;\lambda)}{{}_0F_{\nu-1}(;1,\ldots,1;\lambda)} + E(X) - [E(X)]^2. \tag{5.2}$$

Nadarajah also compared equation (5.2) with equation (5.1). However, ν has to be an integer in order to be able to use the formulation above.

By using the generalized hypergeometric function representation of the normalizing constant, Nadarajah also obtained the cumulative distribution function (cdf) of the Conway-Maxwell-Poisson distribution; we do not include the cdf in this section since our derivations do not deal with that.

5.1.3 MLEs

The likelihood function of the Conway-Maxwell-Poisson distribution is

$$\begin{aligned} L(k_1,\ldots,k_n|\lambda,\nu) &= C(\alpha,\nu)^{-n} \left(\frac{\prod_{i=1}^{n} \lambda^{k_i}}{\left(\prod_{i=1}^{n} k_i!\right)^{\nu}} \right) \\ &= C(\alpha,\nu)^{-n}\, \lambda^{\sum_{i=1}^{n} k_i} \exp\left(-\nu \sum_{i=1}^{n} \log(k_i!) \right). \end{aligned} \tag{5.3}$$

Since the Conway-Maxwell-Poisson distribution follows an exponential family form, the natural parameter for this is $[\log(\lambda), -\nu]$, and the sufficient statistics are

$$\left[\sum_{i=1}^{n} k_i, \sum_{i=1}^{n} \log(k_i!) \right].$$

It is clear that MLEs cannot be computed analytically because of the complicated normalizing constant. Thus, we need iterative methods to solve for MLEs.

5.2 APPROXIMATION OF THE NORMALIZING CONSTANT

From Section 5.1, we can see that the normalizing constant, $C(\alpha,\nu)$, plays an important role in the computations commonly used for fitting models to data: for computing probabilities, moments, and maximum likelihood estimates and their standard errors. However, it is not always easy to compute, so good approximations are needed. This section describes several ways to approximate the normalizing constant of Conway-Maxwell-Poisson distribution.

Henceforth, we use a reparametrization of Conway-Maxwell-Poisson distribution by letting $\alpha = \lambda^{1/\nu}$ because it gives a simpler centering parameter. This new parametrization is

$$P(X=k|\alpha,\nu) = C(\alpha,\nu)^{-1} \left(\frac{\alpha^k}{k!} \right)^{\nu}, \quad \text{where} \quad C(\alpha,\nu) = \sum_{k=0}^{\infty} \left(\frac{\alpha^k}{k!} \right)^{\nu}.$$

For the case of integer ν, Shmueli, et al. [54] expressed $C(\lambda,\nu)$ as a $(\nu-1)$-dimensional integral involving complex exponentials and used Laplace's method to derive an approximation. Their numerical calculations showed that this approximation was good when $\lambda > 10^{\nu}$,

or $\alpha > 10$. Since their approximation performs well in some parameter ranges but not in others, it is of interest to get an alternate approximation to $C(\alpha, \nu)$ for these parameter ranges.

Our aims in this section are first to show that this approximation is in principle valid for all $\nu > 0$, not just integers; second, to use correction terms based on certain expansions to improve upon this approximation; and third, for ν near 0, to propose the use of the geometric approximation with correction terms. First, we sketch the approach of Shmueli, et al. [54] and discuss its limitations. Next, we give a new derivation of the same approximation of $C(\alpha, \nu)$ by using well known statistical methods: in particular, we express $C(\alpha, \nu)$ as an expectation of a function of a Poisson variate, and use a Gaussian approximation to the square root of a Poisson for large α and all $\nu > 0$. Then, we connect the normalizing constant to two special functions: the generalized hypergeometric function ${}_0F_{\nu-1}$ for integer ν, and the modified Bessel function of the first kind, I_0 for $\nu = 2$. The well-known asymptotic expansion for $I_0(z)$ for $|z| \to \infty$ points to the need for correction terms to the leading term. Thus, finally, we compute the first two correction terms and assess them numerically. We show that we have an improvement over the approximation due to Shmueli, et al. [54], considerably so where it does poorly. We end this section with a discussion and propose future directions.

5.3 EARLIER APPROACH: LAPLACE APPROXIMATION FOR INTEGER ν

Shmueli and her colleagues [54, 39] derived an asymptotic approximation and upper bound for integer ν using Laplace's method on a $(\nu-1)$-dimensional integral that represents $C(\lambda, \nu)$. Below is a brief account of their steps and approximation.

First, since

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{e^{ix}}\, e^{-ixk}\, dx = \frac{1}{k!},$$

we have

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{e^{ix}}\, C(\lambda e^{-ix}, \nu)\, dx = \sum_{k=0}^{\infty} \frac{\lambda^k}{(k!)^\nu} \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{e^{ix}}\, e^{-ixk}\, dx = C(\lambda, \nu+1).$$

This expresses $C(\lambda, \nu + 1)$ as an integral of $C(\lambda, \nu)$. Iterating this process starting from $C(\lambda, 1) = e^{\lambda}$ leads to a representation of $C(\lambda, \nu)$ for integer $\nu > 0$, as a multiple integral:

$$C(\lambda, \nu) = \frac{1}{(2\pi)^{\nu-1}} \int_{-\pi}^{\pi} \dots \int_{-\pi}^{\pi} \exp\left(\sum_{j=1}^{\nu-1} e^{ix_j} + \lambda\, e^{-\sum_{j=1}^{\nu-1} ix_j} \right) dx_1 \dots dx_{\nu-1}.$$

Then, Laplace's method [5] applied to the multiple integral yields

$$C(\lambda, \nu) = \sum_{k=0}^{\infty} \frac{\lambda^k}{(k!)^{\nu}} = \frac{\exp\left(\nu \lambda^{1/\nu}\right)}{\lambda^{(\nu-1)/2\nu} (2\pi)^{(\nu-1)/2} \sqrt{\nu}} \left[1 + \mathcal{O}(\lambda^{-1/\nu})\right], \tag{5.4}$$

which in our parametrization is

$$C(\alpha, \nu) = \frac{e^{\nu\alpha}}{(2\pi\alpha)^{\frac{\nu-1}{2}} \sqrt{\nu}} \left[1 + \mathcal{O}(\alpha^{-1})\right] \tag{5.5}$$

as $\alpha = \lambda^{1/\nu} \to \infty$. Notice that this approximation requires $\nu > 0$; we will see at the end of next section below that when $\nu < 1$ and $\lambda < 1$ an approximation based on the geometric distribution with correction terms is a better approximation.

They also present results of numerical examples to check the relative error of their approximation. Their numerical studies show that the leading term in the asymptotic expression in (5.4) is a good approximation for certain parameter ranges; however for $\nu > 1$ it consistently underestimates the true value of the $C(\alpha, \nu)$. In addition, their results for non-integer values of ν suggests that this approximation also applies to all real ν. An investigation into other aspects of this family by Nadarajah [41] also assumes that ν is an integer, and connects this series to the generalized hypergeometric functions, which we discuss in Section 4.

In 2015 Gillispie and Green [20] extended Shmueli et al.'s [54, 39] derivation of the approximation of $C(\lambda, \nu)$. In their work they used the same framework. After using rather involved complex variable methods including contour integrals, they showed that this approximation holds for all ν. However, they did not address the issue of the improving the performance of this approximation when $\lambda < 10^{\nu}$.

In the next section, we use a statistical approach to derive the asymptotic approximation with the same leading term: our argument works for all $\nu > 0$; it also provides explicit expressions for the first two terms of an asymptotic expansion which can be used to attempt to improve underestimation by the leading term.

5.4 OUR APPROACH

We begin this section by proposing more statistical approach to derive the approximation of the normalizing constant of Conway-Maxwell-Poisson distribution. We then connect this normalizing term with the generalized hypergeometric function. Then, obtain some correction terms to improve the approximation and to address the underestimation. We end this section by proposing a new approximation to the normalizing term for both ν and λ are small.

5.4.1 A statistical approach for all $\nu > 0$

In this section, we describe the steps of our approach. We write the normalizing constant as the expectation of a function of X, a Poisson(α) variate, and use Stirling's formula to approximate that function. Next, we approximate $2\sqrt{X}$ by a normal distribution for large values of α; we then use elementary expansions for the log gamma and log functions, and end by taking expectations of the resulting expression.

5.4.1.1 Express normalizing constant as an expectation

Notice that we can rewrite the normalzing as an expectation of a function of Poisson variate

$$C(\alpha,\nu)=\sum_{k=0}^{\infty}\left(\frac{\alpha^k}{k!}\right)^{\nu}=e^{\alpha}\sum_{k=0}^{\infty}e^{-\alpha}\frac{\alpha^k}{k!}\left(\frac{\alpha^k}{k!}\right)^{\nu-1}=e^{\alpha}E_{\alpha}\left(\frac{\alpha^X}{X!}\right)^{\nu-1}. \tag{5.6}$$

When α is large, X will be large with high probability, with the most likely value near α. Stirling's approximation [1] says that

$$\Gamma(\alpha+1)\sim\alpha^{\alpha}e^{-\alpha}\sqrt{2\pi\alpha}\quad\text{as }\alpha\to\infty.$$

Using Stirling's approximation in (5.6), we get

$$\begin{aligned} C(\alpha,\nu) &= e^{\alpha}E_{\alpha}\left(\frac{\alpha^{X}}{\alpha^{\alpha}e^{-\alpha}\sqrt{2\pi\alpha}}\frac{\alpha^{\alpha}e^{-\alpha}\sqrt{2\pi\alpha}}{X!}\right)^{\nu-1} \\ &= \frac{e^{\nu\alpha}}{(2\pi\alpha)^{\frac{\nu-1}{2}}}E_{\alpha}\left(\frac{\alpha^{X}e^{-\alpha}\sqrt{2\pi\alpha}}{\Gamma(X+1)}\right)^{\nu-1} \\ &= \frac{e^{\nu\alpha}}{(2\pi\alpha)^{\frac{\nu-1}{2}}}E_{\alpha}[U(\alpha,X)]. \end{aligned} \tag{5.7}$$

Note that the constant term in (5.7) is the same as the asymptotic expression in (5.4) above, except for the $\sqrt{\nu}$ term in the denominator. To show that $E_{\alpha}[U(\alpha,X)]$ is approximately $\nu^{-1/2}$ and assess the order of magnitude of the error, we study $U(\alpha,X)$ or equivalently

$$\ln[U(\alpha,X)] = (\nu-1)\left[\left(X+\frac{1}{2}\right)\ln\alpha - \alpha + \frac{\ln(2\pi)}{2} - \ln\Gamma(X+1)\right]. \tag{5.8}$$

5.4.1.2 Use Normal approximation to $2\sqrt{X}$ and expansions

Because X has a Poisson(α) distribution, for large α, the distribution of $2\sqrt{X}$ is approximately normal with mean $2\sqrt{\alpha}$ and variance 1. We first pursue our approximation based on the $2\sqrt{X}$ transformation, which uses the simpler expression than the transformations that are better interms of stabilizing the variance. In the future, we will use Anscombe's $2\sqrt{X+3/8}$ (see [32]), which is a better variance stabilizing transformation than the $2\sqrt{X}$ transformation, to get an approximation to the normalizing term.

Let Z be a standard normal variate; for $2\sqrt{X}$ transformation, we have

$$2\sqrt{X} \sim (Z+2\sqrt{\alpha}) \quad \text{so that} \quad X \sim \frac{(Z+2\sqrt{\alpha})^2}{4} = \alpha + Z\sqrt{\alpha} + \frac{Z^2}{4}.$$

Using this normal approximation, we obtain

$$\begin{aligned} \frac{\ln[U(\alpha,X)]}{(\nu-1)} &\simeq \left[\left(\alpha + Z\sqrt{\alpha} + \frac{Z^2}{4} + \frac{1}{2}\right)\ln(\alpha) - \alpha + \frac{\ln(2\pi)}{2}\right] \\ &- \ln\Gamma\left(\alpha + Z\sqrt{\alpha} + \frac{Z^2}{4} + 1\right) \\ &= A - B. \end{aligned} \tag{5.9}$$

In (5.9), we leave A as is; for B, we use the first three terms of the following expansions

$$\begin{aligned}\ln\Gamma(x+1) &= \left(x+\frac{1}{2}\right)\ln(x+1)-(x+1)+\frac{\ln(2\pi)}{2}+\mathcal{O}(x^{-1}) \text{ for } x\to\infty,\\ \ln(1+t) &= t-\frac{t^2}{2}+\frac{t^3}{3}+\ldots \text{ for } |t|<1.\end{aligned}$$

to obtain

$$\begin{aligned}B=\ln\Gamma\left(\alpha+Z\sqrt{\alpha}+\frac{Z^2}{4}+1\right) &= \left(\alpha+Z\sqrt{\alpha}+\frac{Z^2}{4}+\frac{1}{2}\right)\ln\left(\alpha+Z\sqrt{\alpha}+\frac{Z^2}{4}+1\right)\\ &\quad-\left(\alpha+Z\sqrt{\alpha}+\frac{Z^2}{4}+1\right)+\frac{\ln(2\pi)}{2}\\ &= \left(\alpha+Z\sqrt{\alpha}+\frac{Z^2}{4}+\frac{1}{2}\right)\ln(\alpha) \qquad (5.10)\\ &\quad+\left(\alpha+Z\sqrt{\alpha}+\frac{Z^2}{4}+\frac{1}{2}\right)\ln\left(1+\frac{Z}{\sqrt{\alpha}}+\frac{Z^2}{4\alpha}+\frac{1}{\alpha}\right)\\ &\quad-\left(\alpha+Z\sqrt{\alpha}+\frac{Z^2}{4}+1\right)+\frac{\ln(2\pi)}{2}.\end{aligned}$$

By substituting (5.10) in (5.9), we get

$$\frac{\ln[U(\alpha,X)]}{\nu-1}=A-B\simeq-\frac{Z^2}{2}-\frac{6Z+Z^3}{12\sqrt{\alpha}}+\mathcal{O}(\alpha^{-1})$$

Then, using $e^x\simeq 1+x$ for small x, we have

$$\begin{aligned}U(\alpha,X) &\simeq \exp\left[-(\nu-1)\frac{Z^2}{2}\right]\exp\left[-(\nu-1)\frac{6Z+Z^3}{12\sqrt{\alpha}}\right]\exp\left[\mathcal{O}(\alpha^{-1})\right]\\ &= \exp\left[-(\nu-1)\frac{Z^2}{2}\right]\left[1-(\nu-1)\frac{6Z+Z^3}{12\sqrt{\alpha}}\right]\left[1+\mathcal{O}(\alpha^{-1})\right] \qquad (5.11)\end{aligned}$$

5.4.1.3 Take expectations

Finally, subsitute (5.11) in (5.7), and notice that $6Z+Z^3$ is an odd function, so its expectation is 0, to get

$$\begin{aligned} E_\alpha[U(\alpha, X)] &\simeq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(\nu-1)z^2/2} \left[1 - (\nu-1)\frac{6z+z^3}{12\sqrt{\alpha}}\right] e^{-z^2/2}[1+\mathcal{O}(\alpha^{-1})]\, dz \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(\nu-1)z^2/2} e^{-z^2/2}[1+\mathcal{O}(\alpha^{-1})]\, dz \\ &= \frac{1}{\sqrt{\nu}}[1+\mathcal{O}(\alpha^{-1})], \end{aligned}$$

so that

$$C(\alpha, \nu) = e^\alpha E_\alpha \left(\frac{\alpha^X}{X!}\right)^{\nu-1} = \frac{e^{\nu\alpha}}{(2\pi\alpha)^{\frac{\nu-1}{2}}\sqrt{\nu}}[1+\mathcal{O}(\alpha^{-1})] \text{ as } \alpha \to \infty, \tag{5.12}$$

which is the same asymptotic expression as in (5.5).

Table 7: Percentage errors when approximating $C(\lambda, \nu)$ by using (5.12)

	values of ν									
λ	0.1	0.3	0.5	0.7	0.9	1.1	1.3	1.5	1.7	1.9
0.1	-100	-79	-36	-11	-1	0	-2	-5	-9	-13
0.3	-98	-38	-7	1	1	-1	-4	-7	-10	-13
0.5	-83	-12	3	4	1	-1	-4	-7	-9	-11
0.7	-49	2	6	4	1	-1	-4	-6	-8	-10
0.9	-9	8	7	4	1	-1	-3	-5	-6	-7
1.1	10	9	6	3	1	-1	-3	-5	-6	-7
1.3	5	7	5	3	1	-1	-3	-4	-6	-7
1.5	1	5	4	3	1	-1	-2	-4	-5	-6
1.7	0	3	3	2	1	-1	-2	-3	-5	-6
1.9	0	2	3	2	1	-1	-2	-3	-4	-5

In summary, we have used a standard statistical approach and elementary expansions to show that the expression is valid for all real $\nu > 0$. However, as we pointed out earlier, this approximation does not perform well across all the parameter ranges, which can be seen in

Table 7 summarized in [54]. Negative numbers in this table means that the approximated value is lower than the exact value, so we have an underestimate of the normalizing constant when we use this approximation. Our next aim is to improve the performance of this approximation for the parameter ranges where (5.12) does poorly.

5.4.2 GENERALIZED HYPERGEOMETRIC FUNCTION APPROACH

Nadarajah [41] has connected the Conway-Maxwell-Poisson family with the generalized hypergeometric (Fox-Wright) functions and studied the moments and cdfs; however, she did not deal specifically with the problem of approximating the normalizing constant. Even though ν must be an integer for this connection to hold, we pursue it for illustration because the special case $\nu = 2$ motivates the use of correction terms. The generalized hypergeometric functions are

$$ {}_pF_q\left[\begin{array}{cccc}(a_1A_1) & (a_2A_2) & \dots & (a_pA_p)\\ (b_1B_1) & (b_2B_2) & \dots & (b_qB_q)\end{array};z\right] = \sum_{n=0}^{\infty}\frac{\prod_{i=1}^{p}\Gamma(a_i+A_in)}{\prod_{j=1}^{q}\Gamma(b_j+B_jn)}\frac{z^n}{n!},$$

where an empty product is defined as 1. When $p = 0$, $q = \nu - 1$, and $a_i = A_i = b_j = B_j = 1$ and set $z = \lambda$, we get

$$C(\lambda,\nu) = {}_0F_{\nu-1}[-;1,\cdots,1;\lambda] \quad \text{or} \quad C(\alpha,\nu) = {}_0F_{\nu-1}[-;1,\cdots,1;\alpha^{\nu}].$$

When $\nu = 2$, the generalized hypergeometric function reduces to the modified Bessel function of the first kind: $C(\alpha, 2) = I_0(2\alpha)$. Using the known asymptotic expansion [1]

$$I_\eta(z) \sim \frac{e^z}{\sqrt{2\pi z}}\left(1 - \frac{4\eta^2-1}{8z} + \frac{(4\eta^2-1)(4\eta^2-9)}{2!(8z)^2} + \mathcal{O}(z^{-3})\right) \quad \text{as } |z|\to\infty,$$

we have

$$C(\alpha,2) = I_0(2\alpha) \sim \frac{e^{2\alpha}}{\sqrt{\alpha}\sqrt{4\pi}}\left(1 + \frac{1}{16\alpha} + \frac{9}{512\alpha^2} + \mathcal{O}(\alpha^{-3})\right) \quad \text{as } \alpha\to\infty. \tag{5.13}$$

We see from this expansion that the estimate in (5.4) is most likely an underestimate of the normalizing constant because the first two correction terms of order α^{-1} and α^{-2} are both positive.

As mentioned before, all relative errors in this chapter are expressed in percents. Table 8 presents the relative errors of the two of the approximations in (5.13): the leading term only, and the leading term with first-order correction. We note that Shmueli et al. [54] presented the relative errors for the leading term only for λ and ν values from 0.1 to 1.9. Based on their table, they stated that the approximation is not good when $\alpha = \lambda^{1/\nu} < 10$. Our illustration in Table 8 (using the λ parametrization) demonstrates that even in this range including a correction term improves the approximation considerably. Including the second correction term will in this case lessen the underestimate; however, that correction is typically much smaller than the first-order correction, so we omit it.

Table 8: Relative errors in (5.13) for leading term and for first-order correction; $\nu = 2$.

values of λ	leading term	first correction
0.1	-14.4	2.6
0.3	-13.9	-4.0
0.5	-11.9	-4.1
0.7	-10.3	-3.6
0.9	-9.1	-3.1
1.1	-8.1	-2.6
1.3	-7.4	-2.3
1.5	-6.7	-2.0
1.7	-6.2	-1.7
1.9	-5.8	-1.5

This illustration shows the usefulness of correction terms. The literature [7] on the asymptotics of the generalized hypergeometric functions ${}_pF_q$ require integer values of p and q, which in our case restricts ν to be an integer. Therefore, in the next section we extend our statistical approach by including more terms in the expansions for $\ln(1+t)$ and $\ln\Gamma(x+1)$ to get correction terms of order α^{-1} and α^{-2} for all $\nu > 0$.

5.4.3 CORRECTION TERMS

For our statistical approach to get an approximation to the normalizing term above, we first write the normalizing constant as an expectation as in equation (5.8). We then used the normal approximation, which yielded (5.9). Next, we used the first three terms of the expansions for $\ln(1+t)$ and $\ln\Gamma(x+1)$. However, the estimate in (5.12) underestimates the normalizing constant and it performs very poorly in some certain parameter ranges. Thus, in this section, we allow more terms in these expansions in order to get correction terms, which we will demonstrate will improve the performance of this approximation.

5.4.3.1 Use the expansions for $\ln(1+t)$ and $\ln\Gamma(x+1)$ with more terms

To derive the correction terms, we use more terms in the expansions below to further study B in (5.9): for $x \to \infty$,

$$\begin{aligned}
\ln\Gamma(x+1) &= \left(x+\frac{1}{2}\right)\ln(x+1) - (x+1) + \frac{\ln(2\pi)}{2} + \frac{1}{12(x+1)} + \mathcal{O}(x^{-3}), \text{ and} \\
\ln(1+t) &= \sum_{n=1}^{6}(-1)^{n+1}\frac{t^n}{n} + \mathcal{O}(t^7) \text{ for } |t| \to 0.
\end{aligned}$$

Following the same steps as in Section 3, we get

$$\begin{aligned}
\frac{\ln U(\alpha,\nu)}{(\nu-1)}) &\simeq -\frac{Z^2}{2} - \frac{Z^3+6Z}{12\sqrt{\alpha}} + \frac{Z^4+12Z^2+8}{96\alpha} - \frac{Z^5+2Z^3+4Z}{48\alpha\sqrt{\alpha}} \\
&\quad + \frac{Z^6+30Z^4+120Z^2}{1920\alpha^2} + \mathcal{O}(\alpha^{-5/2}).
\end{aligned} \tag{5.14}$$

As in (5.11), we first exponentiate this expression, and expand all but the $e^{-(\nu-1)z^2/2}$ term in a two-term Taylor series: $e^x \simeq 1+x$ for small x, we have

$$\begin{aligned}
U(\alpha,X) &\simeq \exp\left[-(\nu-1)\frac{Z^2}{2}\right]\exp\left[-\frac{Z^3+6Z}{12\sqrt{\alpha}} - \frac{Z^5+2Z^3+4Z}{48\alpha\sqrt{\alpha}}\right] \\
&\quad \exp\left[\frac{Z^4+12Z^2+8}{96\alpha} + \frac{Z^6+30Z^4+120Z^2}{1920\alpha^2}\right]\left[+\mathcal{O}(\alpha^{-5/2})\right] \\
&= \exp\left[-(\nu-1)\frac{Z^2}{2}\right]\left[1-(\nu-1)\left(\frac{Z^3+6Z}{12\sqrt{\alpha}} + \frac{Z^5+2Z^3+4Z}{48\alpha\sqrt{\alpha}}\right)\right] \\
&\quad \left[1+(\nu-1)\left(\frac{Z^4+12Z^2+8}{96\alpha} + \frac{Z^6+30Z^4+120Z^2}{1920\alpha^2}\right)\right][1+\mathcal{O}(\alpha^{-5/2})]
\end{aligned} \tag{5.15}$$

5.4.3.2 Take expectations

Next, we take the expectation of the (5.15). First, we note that the terms with odd powers of Z integrate to 0, so

$$E\left\{e^{-(\nu-1)Z^2/2}\left[(\nu-1)\left(\frac{Z^3+6Z}{12\sqrt{\alpha}}+\frac{Z^5+2Z^3+4Z}{48\alpha\sqrt{\alpha}}\right)\right]\right\}=0.$$

Thus, as we did before in (5.15), we omit the odd terms when taking the expectation of $U(\alpha, X)$. Because odd powers of Z integrate to zero, the error term here becomes $\mathcal{O}(\alpha^{-3})$ rather than $\mathcal{O}(\alpha^{-5/2})$. The resulting expansion is

$$E\left\{e^{-(\nu-1)Z^2/2}\left[1+(\nu-1)\left(\frac{Z^4+12Z^2+8}{96\alpha}+\frac{Z^6+30Z^4+120Z^2}{1920\alpha^2}\right)\right]\left[1+\mathcal{O}(\alpha^{-3})\right]\right\},$$

so we obtain $E_\alpha[U(\alpha, X)]$ as

$$E_\alpha[U(\alpha,X)]=\frac{1}{\sqrt{\nu}}\left[1+\frac{\nu-1}{12\alpha}\left(\frac{3}{8\nu^2}+\frac{3}{2\nu}+1\right)+\frac{\nu-1}{16\alpha^2}\left(\frac{1}{8\nu^3}+\frac{3}{4\nu^2}+\frac{1}{\nu}\right)+\mathcal{O}(\alpha^{-3})\right].$$

As a result, the normalizing constant with correction terms of order up to α^{-2} is

$$C(\alpha,\nu)=\frac{e^{\nu\alpha}}{(2\pi\alpha)^{\frac{\nu-1}{2}}\sqrt{\nu}}\left[1+(\nu-1)\left(\frac{8\nu^2+12\nu+3}{96\alpha\nu^2}+\frac{1+6\nu}{144\alpha^2\nu^3}\right)+\mathcal{O}(\alpha^{-3})\right] \tag{5.16}$$

as $\alpha\to\infty$.

Before turning to the details of the numerical illustration, we return to the special case $\nu=2$: note that the leading term is the same as in (5.13); however the two correction terms in (5.16) are

$$\frac{1}{\alpha}\frac{59}{16(24)}\simeq\frac{0.1536}{\alpha}\quad\text{and}\quad\frac{1}{\alpha^2}\frac{13}{2(512)}\simeq\frac{0.0127}{\alpha^2}$$

instead of

$$\frac{1}{16\alpha}=\frac{0.0625}{\alpha}\quad\text{and}\quad\frac{9}{512\alpha^2}=\frac{0.0177}{\alpha^2}$$

in (5.13). We attribute this difference to the fact that our statistical approach to the case for all $\nu>0$ is different from the Laplace method approach for the modified Bessel function $I_\eta(z)$.

We now have correction terms to the approximation of normalizing constant. Since these correction terms have positive sign, it will reduce the underestimate. We present several numerical examples to study the effect of these correction terms on the normalizing term.

5.4.4 NUMERICAL EXAMPLES

Our aim in this section is to conduct numerical examples to illustrate how the correction terms improve the approximation of the normalizing constant. In order to assess the performance of the approximations, we need to know the exact values of the normalizing constant. When α is small, we simply sum a finite number of terms in the series because it converges quickly. When α is very large, we sum the series around the modal value after factoring it out. In particular, it is easy to see that $a_k = (\alpha^k/k!)^\nu$, then $a_{k+1}/a_k < 1$ if and only if $\alpha < k+1$, so that the modal probability is k_α is the integer nearest to $\alpha - 1$. Thus, we write

$$\begin{aligned} C(\alpha,\nu) &= \sum_{k=0}^{\infty}\left(\frac{\alpha^k}{k!}\right)^\nu = \left(\frac{\alpha^{k_\alpha}}{k_\alpha!}\right)^\nu \sum_{k=0}^{\infty}\left(\frac{\alpha^{k-k_\alpha}k_\alpha!}{k!}\right)^\nu \\ &\simeq \left(\frac{\alpha^{k_\alpha}}{k_\alpha!}\right)^\nu \sum_{k_\alpha-l}^{k_\alpha+h}\left(\frac{\alpha^{k-k_\alpha}k_\alpha!}{k!}\right)^\nu, \end{aligned}$$

where l and h can be determined by examining the successive differences, and stopping when those differences are sufficiently small.

Let $\tilde{C}_0(\alpha,\nu)$ be the approximated value for the normalizing constant. Then, the approximated value for the equation (5.4) is just the leading term, so we have

$$\tilde{C}(\lambda,\nu) = \sum_{k=0}^{\infty}\frac{\lambda^k}{(k!)^\nu} = \frac{\exp\left(\nu\lambda^{1/\nu}\right)}{\lambda^{(\nu-1)/2\nu}(2\pi)^{(\nu-1)/2}\sqrt{\nu}} = \frac{e^{\nu\alpha}}{(2\pi\alpha)^{(\nu-1)/2}\sqrt{\nu}}. \tag{5.17}$$

Table 9 summarized the relative errors made by using (5.17) which is based on the leading term only.

Our aim is to compare the performance of (5.17) with the approximation that uses correction terms. Thus, we pursue the approximated value for the equation (5.16) which introduces correction terms to (5.17), so we have

$$\tilde{C}_1(\alpha,\nu) = \frac{e^{\nu\alpha}}{(2\pi\alpha)^{\frac{\nu-1}{2}}\sqrt{\nu}}\left[1 + (\nu-1)\frac{8\nu^2+12\nu+3}{96\alpha\nu^2}\right]. \tag{5.18}$$

We did not include the correction term of order α^{-2} because it is much smaller, hence less consequential. The relative error for the approximation which uses the first correction term is summarized in Tables 10.

Notice that, for $\nu > 1$, our correction terms in (5.16) as well as in (5.18) are all positive. From the table in Shmueli, et al. [54]: see Table 7, we know that for $\nu > 1$ the leading term underestimates the true value. Thus, we focus on $\nu > 1$. In addition, from their work, we know that the leading term is a good approximation for $\alpha > 10$; hence we consider $\alpha \leq 10$, and assess the accuracy of the leading term and the first-order correction in Tables 9 and 10, respectively. We see considerable improvement when we use the first-order correction, unless λ is very small, in which case a direct summation of the series is easy and accurate.

Table 9: Relative error using the leading term in (5.17)

	values of ν							
λ	1.1	1.3	1.5	1.7	1.9	2.1	2.3	2.5
0.1	0.1	-1.9	-5.2	-8.9	-12.6	-16.0	-19.1	-21.9
0.3	-1.4	-4.4	-7.4	-10.2	-12.7	-14.9	-16.9	-18.6
0.5	-1.5	-4.2	-6.8	-9.0	-11.0	-12.7	-14.3	-15.6
0.7	-1.3	-3.8	-6.0	-7.9	-9.5	-11.0	-12.3	-13.4
0.9	-1.2	-3.3	-5.2	-6.9	-8.4	-9.7	-10.9	-11.9
1.1	-1.0	-2.9	-4.6	-6.2	-7.5	-8.7	-9.8	-10.7
1.3	-0.9	-2.6	-4.1	-5.5	-6.8	-7.9	-8.9	-9.8
1.5	-0.8	-2.3	-3.7	-5.0	-6.2	-7.3	-8.2	-9.1
1.7	-0.7	-2.1	-3.4	-4.6	-5.7	-6.7	-7.7	-8.5
1.9	-0.6	-1.9	-3.1	-4.3	-5.3	-6.3	-7.2	-8.1
2.1	-0.6	-1.7	-2.9	-3.9	-4.9	-5.9	-6.8	-7.6
2.3	-0.5	-1.6	-2.7	-3.7	-4.7	-5.6	-6.5	-7.3
2.5	-0.5	-1.5	-2.5	-3.5	-4.4	-5.3	-6.2	-7.0

Table 10: Relative errors using first correction term in (5.18)

	values of ν							
λ	1.1	1.3	1.5	1.7	1.9	2.1	2.3	2.5
0.1	9.1	15.2	14.6	11.7	8.2	4.7	1.4	-1.6
0.3	1.9	2.8	1.9	0.5	-1.0	-2.5	-3.8	-4.9
0.5	0.6	0.6	-0.1	-1.0	-1.9	-2.7	-3.4	-4.0
0.7	0.19	-0.02	-0.6	-1.2	-1.8	-2.3	-2.8	-3.1
0.9	0.03	-0.2	-0.6	-1.1	-1.5	-1.9	-2.2	-2.4
1.1	-0.03	-0.3	-0.6	-0.9	-1.3	-1.5	-1.7	-1.8
1.3	-0.04	-0.3	-0.5	-0.8	-1.0	-1.2	-1.3	-1.4
1.5	-0.04	-0.2	-0.4	-0.6	-0.8	-0.9	-1.1	-1.1
1.7	-0.04	-0.2	-0.3	-0.5	-0.6	-0.8	-0.8	-0.9
1.9	-0.03	-0.1	-0.3	-0.4	-0.5	-0.6	-0.7	-0.7
2.1	-0.01	-0.1	-0.2	-0.3	-0.4	-0.5	-0.5	-0.5
2.3	-0.0	-0.1	-0.1	-0.2	-0.3	-0.4	-0.4	-0.4
2.5	0.0	-0.03	-0.1	-0.2	-0.2	-0.3	-0.3	-0.3

5.4.5 Approximation of $C(\lambda, \nu)$ when λ and ν small

In this section, we turn to the case where both ν and λ are small because the leading term in (5.4) can be very poor in that region. For small ν it is more natural to consider an expansion around the geometric distribution than the Poisson ($\nu = 1$) because the Conway-Maxwell-Poisson distribution reduces to the geometric when $\nu = 0$. In this case, the expansion is the much simpler Taylor expansion of $C(\lambda, \nu)$ about $\nu = 0$:

$$C(\lambda, \nu) = \frac{1}{1-\lambda} - \nu \sum_{k=0}^{\infty} \ln(k!)\lambda^k + \mathcal{O}(\nu^2). \tag{5.19}$$

In Table 11, we compare leading term in (5.4) with the geometric (G0), and its linear (G1) correction term. The linear correction term already gives very accurate results, so we do not consider a quadratic correction term in (5.19).

Table 11: Comparison of relative errors for (5.4) and (5.19)

(λ, ν)	Eq (5.4)	G0	G1
(.1,.1)	-100.0	0.1	-0.0
(.3,.1)	-97.7	1.0	-0.1
(.5,.1)	-83.3	4.4	-0.9
(.7,.1)	-49.1	16.4	-8.4
(.1,.3)	-78.7	0.2	-4.0
(.3,.3)	-38.4	2.6	-0.8
(.1,.5)	-35.8	0.3	-0.1
(.3,.5)	-6.8	3.8	-1.9
(.1,.7)	-10.8	0.4	-0.2

Finally, when ν is very big, the denominator $(k!)^\nu$ grows rapidly, so the defining series for $C(\lambda, \nu)$ converges rapidly, and will approximate the limiting Bernoulli distribution.

5.5 CHAPTER SUMMARY

This chapter has proposed a statistical approach to derive the approximation of the normalizing constant of the Conway-Maxwell-Poisson distribution. The approximation can be used for all ν, not only for integers. This approximation is good in certain parameter ranges, namely $\lambda > 10^\nu$. Thus we extend this approximation by including correction terms which improve its performance. We also proposed new approximations for $\nu < 1$ and ν very large.

Our numerical results are encouraging, because they clearly demonstrate that our approach gives reasonably accurate results. The statistical approach above does have one loose

end that must be tied up: we must rigorously justify the taking of the expectation after using the normal approximation to the square root of a Poisson variate. Our contributions are the following: our method applies for all $\nu > 0$, not just integer ν. Next, we provide correction terms and give numerical examples that help guide us in determining when to use them. We also give an alternate approximation for very small ν, when the primary approximation (5.4) is poor.

6.0 APPLICATION OF CONWAY-MAXWELL-POISSON GLM FOR MRNA COUNTS

Previous studies [22, 17, 68] have shown that mRNA transcription occurs in short rapid bursts which lead to non-Poisson counts. Because of that, the mRNA count data are overdispersed. Thus, this stochastic process cannot be modeled by a Poisson process. In this chapter, one of our aims is to propose the use of the Conway-Maxwell-Poisson family of distributions for modeling mRNA count data. This family has the form

$$P(X = k|\lambda, \nu) \propto \frac{\lambda^k}{(k!)^\nu}, \quad k = 0, 1, 2, \ldots.$$

and was introduced by Conway and Maxwell [9] initially for modeling queues and state-dependent service rates. Since then, it has been shown to be useful in other applications, especially by Shmueli and her colleagues [54, 50]. Hoewever, it has not used before for modeling mRNA counts. Thus, we propose the use of Conway-Maxwell-Poisson distribution as an alternative modeling tool for mRNA counts.

This chapter is organized as follows: first we summarize the literature on the application of Conway-Maxwell-Poisson distribution, we then give a brief account of possible modeling tools for mRNA count data. Next, we introduce the generalized linear model (GLM) version of the Conway-Maxwell-Poisson and negative binomial to incorporate covariate information from different experimental conditions. We also consider zero inflation to model excess zero counts. We then apply the proposed models to E. coli bacteria and mammalian cells to illustrate our proposed methods.

6.1 LITERATURE REVIEW

In this section, we summarize applications of the Conway-Maxwell-Poisson distribution and provide a brief overview of the literature. We then present the literature on mRNA count data in order to show what types of models are used.

6.1.1 On Conway-Maxwell-Poisson distribution

Conway and Maxwell [9] introduced the Conway-Maxwell-Poisson family for modeling queues and service rates in 1960s. Until about a decade ago, there were not many studies about it. Then, around mid-2005, Shumeli and her colleagues [29, 39, 50, 52, 54] published a systematic series of papers. They studied the probabilistic and statistical properties of this family: see previous chapter for more details.

Guikema and Goffelt [24] developed a Conway-Maxwell-Poisson generalized linear model (GLM) using a different parametrization for it because the (λ, ν) parametrization does not give a clear centering parameter. This GLM setting has been used to analyze motor vehicle crashes [33]. Later, Barriga and Francisco Louzada [3] constructed a zero-inflated Conway-Maxwell-Poisson regression model and applied it to apple cultivar data.

Most applications of this family are to model motor vehicle crashes [24, 33, 34, 35, 54, 70]. The reason for this is that car crash data are usually overdispersed and sometimes underdispersed. In the situations where the negative binomial distribution is not a good modeling tool, they proposed the Conway-Maxwell-Poisson family as an alternative. We note that Conway-Maxwell-Poisson family does not always give the best fit; most of the time it performs as well as the negative binomial [33]. In additon, there are also computational difficulties when using the Conway-Maxwell-Poisson family because of its normalizing constant. In the previous chapter, we have addressed that issue.

Guikema et al. [24] also reached a similar conclusion for overdispersion. However, they studied the use of the Conway-Maxwell-Poisson GLM on underdispersed data. In particular, they compared the GLM setting of this family with the negative binomial GLM. They conclude that Conway-Maxwell-Poisson GLM can fit data better negative binomial GLM

for underdispersed data sets, and it performs as well as the negative binomial GLM for overdispersed data sets.

In biology, Ridout and Besbeas [48] used the Conway-Maxwell-Poisson for modeling clutch size, which is the number of eggs per nest, for a species of bird. This data is underdispersed, and they used different forms of a weighted Poisson distribution, one of which is Conway-Maxwell-Poisson. They concluded that Conway-Maxwell-Poisson performs poorly.

Imoto [28] proposed a three-parameter generalized Conway-Maxwell-Poisson distribution which includes the negative binomial distribution as a special case. Cordeiro at el. [10] also introduced a new parameter to Conway-Maxwell-Poisson distribution to obtain an exponential-Conway-Maxwell-Poisson distribution. This new three-parameter distribution includes the exponential-geometric and exponential-Poisson distributions as special cases. By using Markov Chain Monte Carlo (MCMC) methods, Vicente et al. [66] proposed a Bayesian analysis of the Conway-Maxwell-Poisson family on right-censored survival data. Rodrigues et al. [49] used this Conway-Maxwell-Poisson cure rate survival model to model cancer recurrence. These two studies demonstrated that the Conway-Maxwell-Poisson family provides a practical tool for modeling survival data with cure rates.

From this brief literature review, the consensus from these studies is that the more complex Conway-Maxwell-Poisson GLM does not provide a significant benefit over the traditional negative binomial GLM for overdispersed data. However, it is useful for underdispersed data since the negative binomial GLM does not do a good job in that case. Even if Conway-Maxwell-Poisson GLM does not outperforms over the traditional methods, it can still address problems with wide range of dispersion [51]. Because the consensus is based on limited applications, in this chapter, we introduce the use of the Conway-Maxwell-Poisson GLM for mRNA count data to see how it performs relative to other models for dispersed data.

6.1.2 On mRNA counts

Direct detection of gene activity is often not possible because new proteins from individual activation events are masked by proteins remaining from previous events [47]. Since mRNA

transcription is a product of gene activity, it can provide valuable information about gene activation and inactivation patterns [23, 36, 60]. In addition, mRNA transcription can also be used to infer information about protein synthesis because it too is typically more difficult than the measurement of its precursor mRNA. Thus, researchers determine the gene activation or inactivation by observing mRNA production, as seen in Figure 17: see [62].

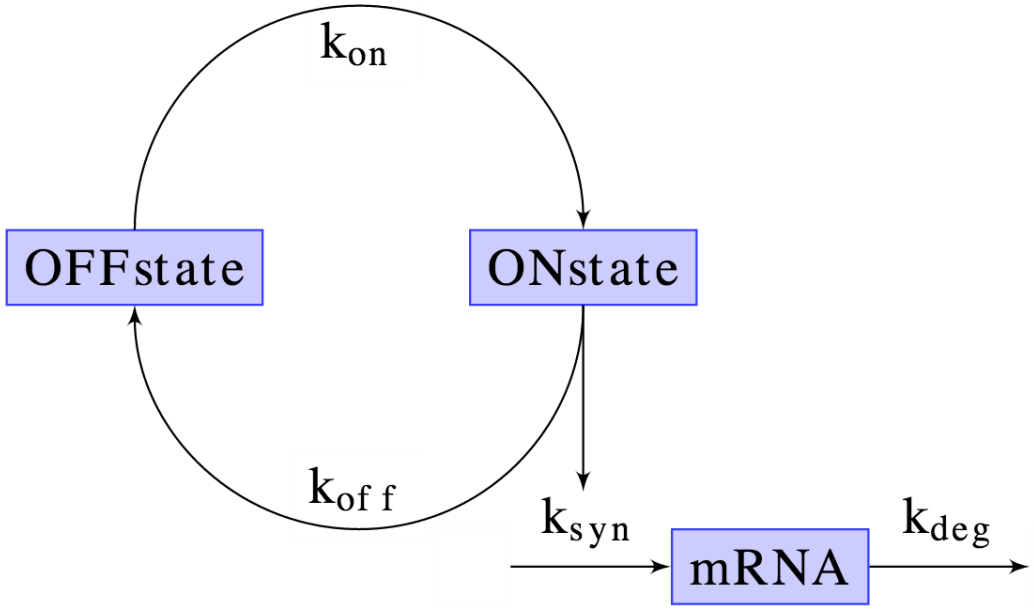

Figure 17: Gene activation and inactivation.

When the gene is in its off-state, mRNA is not transcribed; when it switches to its on-state, mRNA is transcribed at a high rate [44]. Thus, mRNA transcription occurs in short rapid bursts, which is depicted in Figure 18: see [62]. The first part of this figure, Figure 18A, depicts the gene activity, and the second part of it, Figure 18B, depicts the corresponding mRNA production. We can see that the mRNA produced at high rate when the gene is in on state, and no mRNA production in the off state of the gene. The burstiness of mRNA production is not well modeled by a Poisson process: we expect that it is better to use a two-state model for mRNA production [60]. Pendar et al. [45] used partioning of Poisson process to address this issue.

Previous studies [22, 31, 68] have shown that mRNA transcription occurs in short rapid

bursts which lead to non-Poisson counts [42]. Because of this burstiness, the mean number of mRNA transcriptions is typically smaller than its variance, hence it is overdispersed. Instead, earlier studies [60, 53] model the count distribution using a negative binomial distribution.

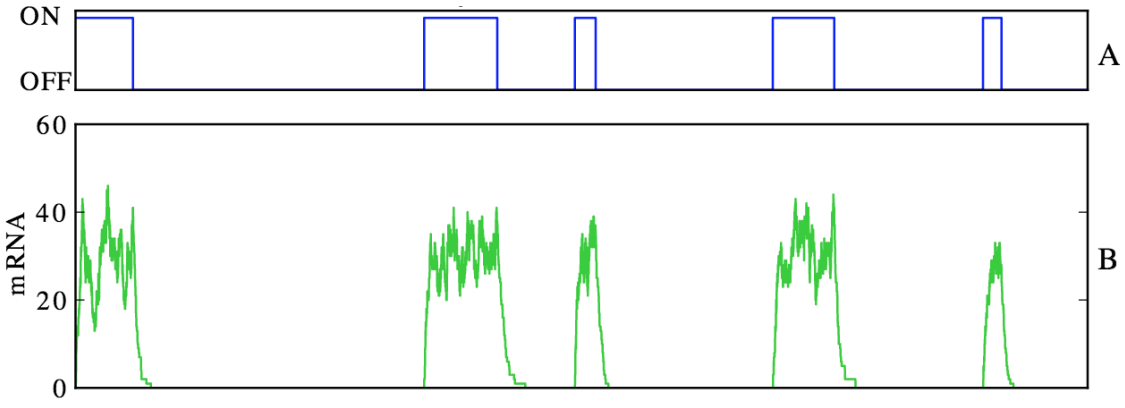

Figure 18: Bursty behavior of mRNA.

GLM methods do not appear to have been used to systematically used to study mRNA counts, but were suggested by [69]. Thus, in this chapter, our aim is to apply GLM methods to mRNA data. We specifically propose the Conway-Maxwell-Poisson (COM-Poisson) GLM as a potential alternative to the more common negative binomial GLM.

6.2 CONWAY-MAXWELL-POISSON GLM FOR MRNA COUNTS

Because of bursty behavior of mRNA, the mean number of mRNA transcriptions is typically smaller than its variance, hence it is overdispersed. Thus, this stochastic process cannot be modeled by a Poisson process. Elgart, et al. [17] illustrate the use of Little's law from queueing theory to stochastic gene expression models: specifically, they relate the mean mRNA burst size, the mean arrival rate of the mRNAs, and the average decay time. They emphasize that Little's law only depends on the means of these three quantities, with no further distributional assumption.

In this chapter, we propose an empirical model for mRNA counts, so we will first give our

motivation for introducing the Conway-Maxwell-Poisson. We then give brief introduction on Conway-Maxwell-Poisson GLM, and apply the empirical model and negative binomial and biophysical models on two example datasets to compare the performance of the empirical distribution versus the others.

6.2.1 Why use Conway-Maxwell-Poisson GLM for mRNA counts?

The Conway-Maxwell-Poisson family has not yet been used to model mRNA count data; however, there are several reasons that suggest that it could be a good candidate.

- They include the bursty behavior of mRNA, which makes it non-Poisson, for which Conway-Maxwell-Poisson is a good alternative tool.
- This bursty behavior leads to overdispersion, and Conway-Maxwell-Poisson is known to be good for handling a variety of dispersed data.
- It contains the common geometric, Poisson, and binomial as special cases, which have been widely studied in the mRNA count modeling literature, so it is a potentially useful umbrella that could detect deviations from these cases.

6.2.2 Conway-Maxwell-Poisson GLM and its zero-inflated version

The general form of the Conway-Maxwell-Poisson GLM is then (see [24])

$$\ln(\alpha) = \beta_0 + \sum_{i=1}^{p} \beta_i x_i \quad \text{and} \quad \ln(\nu) = \gamma_0 + \sum_{j=1}^{q} \gamma_j z_j \tag{6.1}$$

where x_i and z_j are covariates for $i = 1, \dots p$ and $j = 1, \dots q$: p covariates are related to the centering link function and the other q covariates are for the dispersion link function. In previous studies, except [51], no covariates were assigned to the dispersion link function in which case (6.1) reduces to

$$\ln(\alpha) = \beta_0 + \sum_{i=1}^{p} \beta_i x_i \quad \text{and} \quad \ln(\nu) = \gamma_0. \tag{6.2}$$

In this study, we assign covariates to both link functions and compare its performance with the reduced version of it given in (6.2). The resulting likelihood function of the Conway-Maxwell-Poisson GLM is complicated, so we need iterative numerical procedures to compute

maximum likelihood estimates of the parameters and their standard errors. Bayesian methods using Markov chain Monte Carlo have also been used as an alternative: see [24, 33].

The zero-inflated Conway-Maxwell-Poisson is constructed in a standard way as follows:

$$P(X=k) = \begin{cases} p + (1-p)C(\alpha,\nu)^{-1} & \text{if } k = 0 \\ (1-p)C(\alpha,\nu)^{-1}\left(\frac{\alpha^k}{k!}\right)^{\nu} & \text{if } k = 1, 2, \ldots \end{cases} \tag{6.3}$$

where $0 \leq p \leq 1$, $\mu \geq 0$, and $\nu > 0$. For the GLM case the parameters α, ν, p depend on the covariates. The zero-inflated Conway-Maxwell-Poisson model with logistic link function and normal link function have also been studied [3] and implemented in PROC COUNTREG in SAS.

6.2.3 The other models

We first establish our notation. The negative binomial distribution is derived as a gamma mixture of Poisson random variables by assigning a gamma$(1/\alpha, 1/\alpha)$ prior on the parameter of Poisson model.

$$f(y_i|x_i) = \frac{\Gamma(y_i+\alpha)}{\Gamma(y_i+1)\Gamma(\alpha)}\left(\frac{\alpha}{\alpha+\mu_i}\right)^{\alpha}\left(\frac{\mu_i}{\alpha+\mu_i}\right)^{y_i} \quad \text{for} \quad y_i = 0, 1, 2, \ldots$$

where $\alpha > 0$. The zero-inflated Poisson and negative binomial can be constructed as in (6.3). Once again, we use the log link just as for the Conway-Maxwell-Poisson.

Although the gamma distribution is a model for continuous data, we introduce it here because it is an approximation to the probability distributions that arise from certain chemical master equations that describe the mRNA dynamics: see the supplement to [47]. Bokes also conducted a study where he used chemical master equations and its gamma approximation [6]

$$f(y_i|x_i) = \frac{1}{\Gamma(\nu)y_i}\left(\frac{\nu y_i}{\mu_i}\right)^{\nu} e^{-\left(\frac{\nu y_i}{\mu_i}\right)} \quad \text{for} \quad 0 < y_i < \infty.$$

This time, we use the reciprocal link: $\mu_i^{-1} = \mathbf{x_i'}\beta$.

6.3 MRNA COUNT DATA SETS

In this section we apply the models above to mRNA count data from So, et al. [59] and Raj, et al. [47].

6.3.1 Comparing E. coli expression levels for different promoters

We now briefly describe the experimental conditions and resulting mRNA data. So, et al. [59] measured the expression levels of different promoters and different conditions by conducting 20 experiments, and recording the mRNA counts produced under each condition. For illustration, we use mRNA counts of five of these experiments, in which different expression levels from different promoters are used to obtain the counts. In these experiments, bacterial strain TK310, whose relevant genotype was ΔcyaA ΔcpdA ΔlacY, was grown with 0 to 1 mM of IPTG (isopropyl β-D-1-thiogalactopyranoside) and 0 to 10 mM of cAMP (adenosine 3',5'-cyclic monophosphate).

The aim of these experiments is to achieve different expression levels from the P_{lac} promoter, and use that to compare mRNA lifetimes for the same transcript at different expression levels and at different growth rates. The details of the expression levels can be summarized as follows:

- In the first experiment, bacterial strain TK310 was grown with .1mM of cAMP and seven different levels of IPTG: 0μM, 3μM, 10μM, 30μM, 100μM, 300μM, 1000μM.
- In the second experiment, strain TK310 was grown with 1mM of IPTG and six different levels of cAMP: 0μM, 3μM, 10μM, 30μM, 100μM, 300μM.
- The third experiment is similar to the second experiment, but it allows higher levels of cAMP. Strain TK310 was grown with 1mM of IPTG and nine different levels of cAMP: 0μM, 3μM, 10μM, 30μM, 100μM, 300μM, 1000μM, 3000μM, 10000μM.
- In the fifth experiment, strain TK310 was grown with 1mM of IPTG and with seven different levels of cAMP: 0μM, 30μM, 100μM, 300μM, 1000μM, 3000μM, 10000μM.
- The ninth experiment is similar to the first experiment; it uses 10mM of cAMP with the same levels of IPTG as in experiment 1.

For each experimental condition, mRNA count data were collected. More biophysical details of these experiments can be found in the supplementary material of [59].

For experiments 1 and 9, IPTG is the covariate because the cAMP is fixed; in experiments 2,3, and 5, cAMP level is the covariate because IPTG is fixed. We first fit the Conway-Maxwell-Poisson with no covariate assigned to the dispersion link as in (6.2). Next, we fit the Conway-Maxwell-Poisson regression by assigning the same covariate to both link functions in (6.1). For each of the data sets, we compare these fits with those of the negative binomial regression models by using BIC, which are summarized in Table 12 (NB stands for negative binomial). In this table, we abbreviate Conway-Maxwell-Poisson and negative binomial as CMP and NB, respectively. The first row of this table is the fit of model (6.2), and the second row of it is the fit of model (6.1). It is clear that assigning a covariate to the dispersion link function improves the fit of the Conway-Maxwell-Poisson model considerably. In all but one case, the Conway-Maxwell-Poisson GLM with a covariate assigned to both link functions performs better than the negative binomial.

Table 12: BIC of each model fit for E. coli data

	CMP with one-link	CMP with two-link	NB
experiment 1	13064	12131	11853
experiment 2	40833	37609	38060
experiment 3	72265	65885	66789
experiment 5	42504	38736	38869
experiment 9	57424	47442	47770

Next, because of the excessive number of zeroes, we consider zero-inflated versions of the negative binomial and Conway-Maxwell-Poisson, with links described in (6.2) and (6.1). The analyses are summarized in Table 13, and we use the abbreviations ZICMP and ZINB are used for zero-inflated versions of Conway-Maxwell-Poisson and negative binomial, respectively. It is clear that the zero-inflation models are better than the regular models. Among the zero-inflated models, the Conway-Maxwell-Poisson model with covariate assigned to the both link functions is at least as good the negative binomial in two of five cases. The use

of covariates for both links in both the zero-inflated Conway-Maxwell-Poisson and Conway-Maxwell-Poisson models is substantially better than not using covariates for the dispersion link.

Table 13: BIC for each zero-inflated model fit

	ZICMP with one-link	ZICMP with two-link	ZINB
experiment 1	11318	11323	11268
experiment 2	36038	35877	35651
experiment 3	67274	65467	66158
experiment 5	40081	38610	38887
experiment 9	57443	47448	45101

In the tables above, we summarized BIC values of the five experiments for illustrative purposes. We have not presented a summary of the model parameter estimates because our main purpose here is to do model comparison. However, almost the estimates of the model parameters are significant at the .05 level. We also did similar model comparisons for the experiments that we have not included in Tables 12 or 13. A brief summary of the comparison of the models for all 20 experiments follows:

- In 18 experiments, the Conway-Maxwell-Poisson with a covariate assigned to both link functions performs considerably better than the Conway-Maxwell-Poisson with with no covariate assigned to dispersion link. A similar conclusion holds for the zero-inflated case.
- In 9 experiments, the Conway-Maxwell-Poisson with a covariate assigned to both link functions performs better than the negative binomial, and in 2 experiments, their performance is almost the same, that is, the BIC difference between them is less than 10. Here again, the same conclusion holds for the zero-inflated cases of these models.
- In the majority of the experiments, the performance of the Conway-Maxwell-Poisson with with no covariate assigned to dispersion link and its zero-inflated version are much poorer than the negative binomial and its zero-inflated version, respectively.

Overall, the Conway-Maxwell-Poisson with a covariate assigned to both link functions performs as well as the negaive binomial. However, computational difficulties for the normalizing constant of Conway-Maxwell-Poisson in certain parameter ranges leads to slower run times than that of the negative binomial. In addition, the SAS implementation had occasional difficulty in computing the Hessian matrix for both the negative binomial and the Conway-Maxwell-Poisson, more so for the Conway-Maxwell-Poisson.

6.3.2 Comparing different doxycycline levels in mammalian cells

Cells from a homogeneous population can express different numbers of molecules of specific proteins. Raj et al. [47] has studied these variations by counting individual molecules of mRNA produced from a reporter gene. They considered two cell lines: E-YFP-M1-1x (gene-line 1 below) from a 1x-tetO construct, and E- YFP-M1-7x (gene-line 7 below) from 7x-tetO construct. They found that the variability across the population remained constant for all doxycycline concentration levels for the 1x-tetO construct, but that it varied non-monotonically for the 7x-tetO construct.

In their experiments, they varied the doxycycline concentration levels thus: 0 ng/ml, 0.02 ng/ml, 0.04 ng/ml, 0.08 ng/ml, 0.16 ng/ml, 0.32 ng/ml. Then they measured the number of mRNA molecules per cell. From biophysical considerations, they derived a detailed model for the count distribution, which they then approximated by a gamma distribution [47]. Here, we propose different regression models for these data sets by treating doxycycline concentration levels as the covariate, and extending their approximation to a gamma GLM. The fits are summarized in Table 14. In this table the abbreviations CMP and NB are used for Conway-Maxwell-Poisson and negative binomial, respectively.

From Table 14, it can be seen that the fit of Conway-Maxwell-Poisson improves when we introduce in the dispersion link function; however, neither Conway-Maxwell-Poisson or negative binomial is as good as the gamma fit. We also ran the regression models with two covariates: the doxycycline levels and gene lines for these models. The fits are summarized in the last column of Table 14. As in the above case, gamma is the best fit among all proposed models.

Table 14: BIC of each model fit for mammalian cells

	CMP with one-link	CMP with two-link	NB	Gamma
gene-line1	5287	5284	5292	5197.33
gene-line7	5386	5331	5006	4559.53
combined case	10680	10609	10554	9946.88

Once again, we do not summarize the parameter estimates here; for each model fit, we found that the mRNA counts are significantly different for doxycycline concentration levels. We include this example in this study to emphasize the importance of using a model that is derived from biophysical considerations, which yields a continuous approximation to count data. It can be seen that it outperforms other models that are commonly used for count data. We can conclude that when there is a biophysically derived distribution, it performs well; we also show that in the absence of such biophysical knowledge the Conway-Maxwell-Poisson is competitive with the more commonly used negative binomial. Because both the Conway-Maxwell-Poisson and negative binomial arise in queueing theory, we suggest further inquiry into that theory to study mRNA dynamics: see [69].

6.4 CHAPTER SUMMARY

Modeling biological phenomena can be complicated because there are many factors that affect outcomes and are hard to control. In the experimental data that we study here, the variation between genes that are from the same population makes it unreasonable to conclude that a single model is best universally. This chapter has introduced an empirical model for modeling mRNA counts.

Moreover, this chapter serves as a new research direction in the use of Conway-Mawell-Poisson family. Our aims in this section are to introduce the Conway-Maxwell-Poisson family, model zero inflation, and to consider regression methods to study mRNA count data under different experimental conditions. In particular, we show that the use of covariates in both

link functions for the Conway-Maxwell-Poisson or zero-inflated Conway-Maxwell-Poisson GLM is much better than assigning no covariate to the dispersion link. We compare this family with the more commonly used negative binomial and zero-inflated negative binomial GLM; for E. coli data, we see that the zero-inflated Conway-Maxwell-Poisson GLM is as good as the zero-inflated negative binomial GLM.

In the absence of detailed biophysical knowledge, the Conway-Maxwell-Poisson family is a potentially good candidate for fitting over dispersed mRNA count data. And when we do know more about the biophysics, this analysis can help to confirm the derived approximations to probability distributions from master equations, as we demonstrated for the mammalian cell data. The Conway-Maxwell-Poisson model was first proposed in the queueing theory context [9]. Queueing models have also been considered for gene expression and mRNA transcription [17]. Given the good fits of the Conway-Maxwell-Poisson family and its variants above, we suggest that it may be worthwhile to pursue this connection with queueing theory to describe mRNA dynamics.

7.0 CONCLUSION

This thesis addresses two research problems that arise in biological settings, both involving counting processes. We summarize our conclusions on these problems in the following sections.

7.1 PHOTON COUNTING WITH DEAD TIMES

In earlier studies, the investigators were not using all available data to reconstruct images because they did not know how to estimate key parameters in when there are dead periods in the TPLSM detector. To avoid this problem, they only used low laser powers in order to reduce the probability that there is a photon emission during the dead period. They also did not know how to calculate the probability of recording d photons for $d \geq 1$; thus, they grouped the data as seeing nothing (recording 0 photons) and everything else (recording at least one photon). As a result of using only low laser powers and grouping the data (not using all the information that they have), their constructed image quality was very poor. This showed the need for a mathematical model for the number of photons recorded, D, and the main purpose of this study was to derive the distribution of D, from which we derived improved parameter estimates, which led to improved image quality.

When deriving the distribution of D, we first assumed an infinite observation horizon and a fixed dead period. Under this model assumptions, we derived the entire distribution of D by using two equivalent approaches, which in addition gave a family of interesting polynomial identities. There are two important advantages of knowing the entire distribution of D, which are:

- higher laser powers are now feasible, unlike the earlier studies which were restricted to low laser powers;
- we can use all available photon counts without grouping the data unlike the earlier studies which grouped the data as seeing nothing (0 counts) and everything else (at least one photon).

Thus, we can use all the available data to construct the likelihood and make inferences for the model parameters. Using this model, our primary interest is to estimate the underlying intensity parameter because its estimate determines the intensity of a single pixel in an image, and it also gives a measure of the SNR. After making inference on the model parameters, we relate our results to the more general principle of loss of information due to grouping. This generalization also emphasizes the importance in using all available data instead of working with grouped data. We initially used our model on simulated data to assess the magnitude of the loss of information due to grouping.

Next, our collaborators extended earlier work to allow higher laser powers which yield higher emission intensities, or higher Poisson rates α. As the laser power gets larger, we are more likely to see higher photon counts, so we can demonstrate the effectiveness of our model. Hence, our colleagues modified the digital counting electronics used in Driscoll, et al. [14]; they initially collected the data on paper fiber for us to get the resconstructed images based on our method. The reconstructed images show that we sharper images even when we use higher laser powers. Our collaborators are working on getting us data for other applications, so our work continues.

Later, we pursued several extensions of our model: first, we considered a finite time horizon, and second, we considered the gamma waiting time distribution. For the finite horizon, we derived the probabilities of observing either zero or one photon and outlined the method for larger counts. Due to the computational difficulties, we have not pursued the entire distribution of D for larger counts. Using numerical calculations we investigated how large the time horizon must be in order for the infinite case to be a good approximation. This numerical work showed that when the observation period is at least 8 ns, we can safely assume an infinite observation period the infinite observation period approximates the finite one well. We then changed our exponential waiting time assumption to gamma waiting

times. We showed numerically that as the shape parameter of the gamma increases, we approach the original Poisson probabilities. The intuition is clear: as the shape parameter increases, the emission waiting time larger, lessening the effect of the dead period.

7.2 APPROXIMATION OF THE NORMALIZING CONSTANT OF CONWAY-MAXWELL-POISSON

Shmueli and her colleagues [54, 50, 39] brought attention to the Conway-Maxwell-Poisson distribution, which was introduced by Conway and Maxwell [9]. The normalizing constant of this distribution is hard to compute, and so Shmueli and her colleagues proposed an approximation for it. However, their approximation was valid only for integer ν, and the approximation did well only in some certain parameter ranges. In 2015, Gillispie and Green [20] generalized their approximation to all ν. These two works were computationally very involved, and they did not provide correction terms to improve the performance of the approximation.

Thus, we proposed a simpler statistical approach to derive the approximation of of the normalizing constant. It also holds for all ν, not only integers. This approximation is good in some certain parameter ranges, namely $\lambda > 10^{\nu}$. Numerical work showed that it underestimates the normalizing constant for $\nu > 1$. We therefore extended this approximation by introducing correction terms which improved the performance of this approximation for $\nu > 1$. We also studied other parameter ranges where the basic approximation does poorly: namely, for ν very small and very large. For $\nu = 0$, we have the geometric distribution; thus, for small ν we proposed a Taylor series approximation about $nu = 0$. Our numerical example demonstrated a huge improvement. Similarly, for ν very large, we used an approximation based on the Bernoulli distribution.

7.3 CONWAY-MAXWELL-POISSON GLM FOR MRNA COUNTS

In this study, we proposed the use of the Conway-Maxwell-Poisson family for fitting overdispersed mRNA count data in the absence of detailed biophysical knowledge. This work serves as a new research direction in the use of the Conway-Maxwell-Poisson family. We introduced the GLM for this family, modeled zero inflation, and considered regression methods to study mRNA count data under different experimental conditions. We assigned covariates in both link functions for the Conway-Maxwell-Poisson. It turned out that the performance of Conway-Maxwell-Poisson GLM with covariates assigned to both link function is much better than that of with no covariate assigned to the dispersion link. When we compared this family with the more common negative binomial, we see that GLM setting of our empirical fit performed as well as the negative binomial GLM. Given the good fits of the Conway-Maxwell-Poisson family and its variants above, we suggest that it is worthwhile to pursue its connection with queueing theory to describe mRNA dynamics.

8.0 FUTURE WORK

Our work so far has focused on the two different type of counting problems. There remain several interesting open problems described below that may be worthwhile to pursue in the future. We describe them below.

8.1 FOR PHOTON COUNTING PROBLEM

So far, we have worked on the photon counting problem assuming a Type I counter. We obtained the distribution of the number of photons detected by using two different approaches, did inference about the underlying intensity parameter, evaluated the effect of grouping on the Fisher Information, and assessed the fit of the model on the simulated data and images of paper fiber data. We then considered several extensions of our model to investigate different model assumptions. The remaining open problems of interest for this study are given below.

- **Consider Type II counters for this problem:** When recording photons with existence of dead time, the dead period may occur after each pulse (whether it is recorded or not). For such processes, the detector is a Type II counter, so the model assumption should be changed accordingly.
- **Use gamma waiting times for the approach in Section 3.2.2:** In Chapter 4, we considered extensions of our model, and proposed the use of gamma waiting times instead of exponential waiting times. We only pursued this based on our approach in Section 3.2.1, and it could be interesting to do that with under our second approach given in Section 3.2.2.

- **Apply our model on image data from living cells:** Given the wide range of usage of TPLSM, applying our results could yield better image quality for many different applications.

8.2 FOR APPROXIMATION OF THE NORMALIZING CONSTANT OF THE CONWAY-MAXWELL-POISSON

In Chapter 5 we investigated how to improve the performance of the normalizing constant and extended the approximation by including correction terms for $\nu > 1$, and we proposed new approximations for $\nu < 1$ and ν very large.

- **One open problem:** We must rigorously justify the taking the expectation after using normalizing approximation to the square root of a Poisson variate.

- **Use of different transformation:** For X a Poisson(α) random variate, we use the normal approximation to $2\sqrt{X}$ for large α in Chapter 5. There is another transformations that is better in terms of stabilizing the variance: Anscombe's $2\sqrt{X+3/8}$ (see [32]). It would be useful to pursue a similar calculations for the normal approximation to $2\sqrt{X+3/8}$, in order to derive another approximation for the normalizing constant; it would be of interest to see if the correction terms using $2\sqrt{X+3/8}$ transformation will be better than that of the $2\sqrt{X}$ transformation.

- **Extension of the generalized hypergeometric function:** In Chapter 5, when we investigated the correction term, we represented the normalizing constant as generalized hypergeometric function only for the case of integer ν. Possible extension of generalized hypergeometric function could give a good approximations for he normalizing constant for all ν, and lead to other families of count distributions.

8.3 FOR APPLICATION OF THE CONWAY-MAXWELL-POISSON TO MRNA PRODUCTION

In this dissertation we have presented only a limited number of examples of the use of the Conway-Maxwell-Poisson and its variants to mRNA production data. It would be of interest to apply these to a much larger range of experimental data to see how well this model performs relative to both other common empirical models for dispersion like the negative binomial and biophysically derived models.

In addition, queueing models have also been considered for gene expression and mRNA transcription [17, 27]. Given the good fits of the Conway-Maxwell-Poisson family and its variants above, we suggest that it may be worthwhile to pursue this connection with queueing theory to describe mRNA dynamics. In particular, what is known about mRNA production justifies the use of Little's law, which is only moment based. What is needed is a more thorough description using distributional properties that can perhaps bring other queueing theory methods. The CMP, which arose from queueing applications, can be a good candidate for fitting other mRNA production data sets in order to help capture the characteristics of possible underlying biophysical processes.

9.0 APPENDIX: FOR PHOTON COUNTING PROBLEM

9.1 PROOFS

Proof of Lemma 1. Suppose that $d = 2c$ is even. To have a concise presentation, let $\theta = e^{-\delta}$ so that $A_k = A_k(\theta)$, and equation (3.1) becomes

$$\sum_{k=0}^{d-1}(-1)^{k+1}\frac{\theta^{\{k\}}}{\prod_{i=1}^{k}A_i\prod_{i=1}^{d-k}A_i} = \frac{\theta^{\{d\}}}{\prod_{i=1}^{d}A_i}. \tag{9.1}$$

Multiply both sides of (9.1) by the least common denominator

$$\prod_{i=1}^{d}A_i\prod_{i=1}^{c}A_i$$

to get the polynomials $\{f_i : 1 \le i \le d+1\}$ in the first column of Table 15. The cumulative sums $F_3 = f_1 + f_2$ and $\{F_k = F_{k-1} + f_{k-1} : 4 \le k \le d\}$ are in the second column there. Using the fact that for $i < j$, $A_i - A_j = -\theta^i A_{j-i}$, we start with

$$\begin{aligned} F_3 &= f_1 + f_2 = -A_1A_2\cdots A_c + A_2A_3\cdots A_cA_d \\ &= A_2A_3\cdots A_c(-A_1 + A_d) = \theta A_2A_3\cdots A_cA_{d-1}; \end{aligned}$$

the other cumulative sums follow similarly, ending with $F_{d+1} = -f_{d+1}$.

Table 15: For even case $c = d/2$.

$f_1 = -A_1 A_2 \cdots A_c \ \theta^{\{0\}}$	
$f_2 = A_2 A_3 \cdots A_c A_d \ \theta^{\{1\}}$	
$f_3 = -A_3 \cdots A_c Ad - 1 A_d \ \theta^{\{2\}}$	$F_3 = A_2 A_3 \cdots A_c A_{d-1} \ \theta^{\{2\}}$
$f_4 = A_4 \cdots A_c A_{d-2} A_{d-1} A_d \ \theta^{\{3\}}$	$F_4 = -A_3 \cdots A_c A_{d-2} A_{d-1} \ \theta^{\{3\}}$
$\vdots$	$\vdots$
$f_{c-1} = (-1)^{c-1} A_{c-1} A_c A_{c+3} \cdots A_d \ \theta^{\{c-2\}}$	$F_{c-1} = (-1)^c A_{c-2} A_{c-1} A_c A_{c+3} \cdots A_{d-1} \ \theta^{\{c-2\}}$
$f_c = (-1)^c A_c A_{c+2} \cdots A_d \ \theta^{\{c-1\}}$	$F_c = (-1)^{c+1} A_{c-1} A_c A_{c+2} \cdots A_{d-1} \ \theta^{\{c-1\}}$
$f_{c+1} = (-1)^{c+1} A_{c+1} A_{c+2} \cdots A_d \ \theta^{\{c\}}$	$F_{c+1} = (-1)^{c+2} A_c A_{c+1} \cdots A_{d-1} \ \theta^{\{c\}}$
$f_{c+2} = (-1)^{c+2} A_c A_{c+2} \cdots A_d \ \theta^{\{c+1\}}$	$F_{c+2} = (-1)^{c+3} A_c A_{c+1} \ldots A_{d-1} \ \theta^{\{c+1\}}$
$\vdots$	$\vdots$
$f_{d-2} = A_4 \cdots A_c A_{d-2} A_{d-1} A_d \ \theta^{\{d-3\}}$	$F_{d-2} = -A_4 \cdots A_c A_{d-3} A_{d-2} A_{d-1} \ \theta^{\{d-3\}}$
$f_{d-1} = -A_3 \cdots A_c A_{d-1} A_d \ \theta^{\{d-2\}}$	$F_{d-1} = A_3 \cdots A_c A_{d-2} A_{d-1} \ \theta^{\{d-2\}}$
$f_d = A_2 \cdots A_c A_d \ \theta^{\{d-1\}}$	$F_d = -A_2 \cdots A_c A_{d-1} \ \theta^{\{d-1\}}$
	$F_{d+1} = A_1 A_2 \cdots A_c \theta^{\{d\}}$

We omit the details for the case of odd d, given in Table 16 because it the proof for odd d is similar to that of for even d. □

Proof of Theorem 1. Define the functions

$$U_m(t) = e^{-\alpha e^{-(t+m\delta)}},$$

and note that

$$\int_{t_{k-1}+\delta}^{\infty} U_m(t_k) e^{-\alpha[e^{-(t_{k-1}+\delta)} - e^{-t_k}]} \alpha e^{-t_k} dt_k = \frac{U_{m+1}(t_{k-1}) - U_1(t_{k-1})}{A_m}. \tag{9.2}$$

Thus, in equation (3.2), we integrate out t_d to get

$$\int_{t_{d-1}+\delta}^{\infty} e^{-\alpha[e^{-(t_{d-1}+\delta)} - e^{-t_d}]} e^{-\alpha e^{-(t_d+\delta)}} \alpha e^{-t_d} dt_d = \frac{U_2(t_{d-1}) - U_1(t_{d-1})}{A_1}.$$

Table 16: For odd case $c = \frac{d-1}{2}$.

$f_1 = -A_1 A_2 \cdots A_c \ \theta^{\{0\}}$	
$f_2 = A_2 A_3 \cdots A_c A_d \ \theta^{\{1\}}$	
$f_3 = -A_3 \cdots A_c Ad - 1 A_d \ \theta^{\{2\}}$	$F_3 = A_2 A_3 \cdots A_c A_{d-1} \ \theta^{\{2\}}$
$f_4 = A_4 \cdots A_c A_{d-2} A_{d-1} A_d \ \theta^{\{3\}}$	$F_4 = -A_3 \cdots A_c A_{d-2} A_{d-1} \ \theta^{\{3\}}$
$\vdots$	$\vdots$
$f_c = (-1)^c \, A_c A_{c+3} \cdots A_{d-1} A_d \ \theta^{\{c-1\}}$	$F_c = (-1)^{c-1} A_{c-1} A_c A_{c+3} \cdots A_{d-1} \ \theta^{\{c-1\}}$
$f_{c+1} = (-1)^{c+1} \, A_{c+2} \cdots A_{d-1} A_d \ \theta^{\{c\}}$	$F_{c+1} = (-1)^c A_c A_{c+2} A_{c+3} \cdots A_{d-1} \ \theta^{\{c\}}$
$f_{c+2} = (-1)^{c+2} \, A_{c+2} \cdots A_{d-1} A_d \ \theta^{\{c+1\}}$	$F_{c+2} = (-1)^{c+1} A_{c+1} A_{c+2} \cdots A_{d-1} \ \theta^{\{c+1\}}$
$f_{c+3} = (-1)^{c+3} \, A_c A_{c+3} \cdots A_{d-1} A_d \ \theta^{\{c+2\}}$	$F_{c+3} = (-1)^{c+2} A_c A_{c+2} \cdots A_{d-1} \ \theta^{\{c+2\}}$
$\vdots$	$\vdots$
$f_{d-2} = (-1)^{d-2} A_4 \cdots A_c A_{d-2} A_{d-1} A_d \ \theta^{\{d-3\}}$	$F_{d-2} = (-1)^{d-3} A_4 \cdots A_c A_{d-3} A_{d-2} A_{d-1} \ \theta^{\{d-3\}}$
$f_{d-1} = (-1)^{d-1} A_3 \cdots A_c A_{d-2} A_{d-1} \ \theta^{\{d-2\}}$	$F_{d-1} = (-1)^{d-2} A_3 \cdots A_c A_{d-2} A_{d-1} \ \theta^{\{d-2\}}$
$f_d = (-1)^d A_2 \cdots A_c A_d \ \theta^{\{d-1\}}$	$F_d = (-1)^{d-1} A_2 \cdots A_c A_{d-1} \ \theta^{\{d-1\}}$
	$F_{d+1} = -A_1 A_2 \cdots A_c \theta^{\{d\}}$

Next, integrate out t_{d-1} in (3.2) to get

$$\int_{t_{d-2}+\delta}^{\infty} \frac{U_2(t_{d-1}) - U_1(t_{d-1})}{A_1} e^{-\alpha[e^{-(t_{d-2}+\delta)} - e^{-t_{d-1}}]} \alpha e^{-t_{d-1}} dt_{d-1}$$

$$\begin{aligned} &= \frac{U_3(t_{d-2}) - U_1(t_{d-2})}{A_1 A_2} - \frac{U_2(t_{d-2}) - U_1(t_{d-2})}{A_1^2} \\ &= \frac{U_3(t_{d-2})}{A_1 A_2} - \frac{U_2(t_{d-2})}{A_1^2} + \frac{e^{-\{2\}\delta} U_1(t_{d-2})}{A_1 A_2}. \end{aligned}$$

The last expression is obtained by using Lemma 1, to combine the coefficients of the U_1 terms. Next, integrate out t_{d-2} down to t_2 — each time applying Lemma 1 — to get

$$V(t_1) = \sum_{k=0}^{d-1} (-1)^k \frac{e^{-\{k\}\delta} U_{d-k}(t_1)}{\prod_{i=1}^{k} A_i \prod_{j=1}^{d-1-k} A_j}.$$

Finally, integrate out t_1 to get

$$\int_0^\infty e^{-\alpha(1-e^{-t_1})} V(t_1) \alpha e^{-t_1} dt_1 = P(D = d).$$

□

Proof of Corollary 1. Define the vectors $\boldsymbol{p}_d = (p_0, \ldots, p_d)'$ and $\boldsymbol{\zeta}_d = (\zeta_0, \ldots, \zeta_d)'$. We can restate the result in Theorem 1 as

$$\boldsymbol{p} = Q_d \boldsymbol{\zeta},$$

and the result in the Corollary as

$$\boldsymbol{\zeta} = R_d \boldsymbol{p};$$

thus, we must show that $Q_d R_d = I$. Note that for $0 \leq a, b \leq d$,

$$Q_d = (q_{ab}), \text{ where } q_{ab} = \begin{cases} 0 & \text{if } b > a \\ \frac{(-1)^{a-b}\theta^{\{a-b\}}}{\prod_{i=1}^{b} A_i \prod_{j=1}^{a-b} A_j} & \text{if } b \leq a \end{cases}$$

and

$$R_d = (r_{ab}), \text{ where } r_{ab} = \begin{cases} 0 & \text{if } b > a \\ \prod_{i=a-b+1}^{a} A_i & \text{if } b \leq a \end{cases}.$$

Therefore, their product $S_d = (s_{ab})$, where

$$\begin{aligned} s_{ab} &= \sum_{k=0}^{d} q_{ak} r_{kb} = \sum_{k=b}^{a} q_{ak} r_{kb} = \sum_{k=b}^{a} \frac{(-1)^{a-k}\theta^{\{a-k\}}}{\prod_{i=1}^{k} A_i \prod_{j=1}^{a-k} A_j} \prod_{i=k-b+1}^{k} A_i \\ &= \sum_{k=b}^{a} \frac{(-1)^{a-k}\theta^{\{a-k\}}}{\prod_{i=1}^{k-b} A_i \prod_{j=1}^{a-k} A_j}. \end{aligned}$$

By inspection, $s_{ab} = 0$ for $b > a$, and $s_{aa} = 1$. Next, let $j = k - b$ in this expression and use Lemma 1 with $d = a - b$ to get $s_{ab} = 0$ for $b < a$. □

Proof of Lemma 2. In the calculation of the $P_n(D=d)$ several complicated terms appear as a coefficients of $\binom{n-k_{d-i-1}}{d-j+1}$ for $j=1,\cdots,d-2$; we show that these coefficients sum to zero. Lemma 2 gives a general formulation for those coefficients. Before proving this, we first state the relationship between the coefficients for $j=1$ and the coefficients for $j=2,\cdots,d-2$. We can use the coefficient of $\binom{n-k_{d-i-1}}{1}$ as our baseline group since the coefficient of $\binom{n-k_{d-i-1}}{d-j+1}$ for $j=2,\cdots,d-2$ can be derived from the coefficient of $\binom{n-k_{d-i-1}}{1}$. We demonstrate several cases to clarify how we use the coefficients for $j=1$ as baseline group. To calculate $P_n(D=5)$, we have the following identities as the coefficients of $\binom{n-k_{d-4}}{j}$ for $j=0,1,2,3$:

$$c_0 = -\frac{e^{\delta(4n+10)}B_3^0}{B_1B_2B_3} + \frac{e^{\delta(4n+7)}B_2^0}{B_1B_2} - \frac{e^{\delta(4n+5)}B_1^0}{B_1^2} + \frac{e^{\delta(4n+4)}}{B_1B_2} = \frac{e^{\delta(4n+4)}}{B_1B_2B_3},$$

$$\begin{aligned}
c_1 &= -\frac{e^{\delta(4n+6)}B_3^1}{B_1B_2B_3} + \frac{e^{\delta(4n+4)}B_2^1}{B_1B_2} - \frac{e^{\delta(4n+3)}B_1^1}{B_1^2} + \frac{e^{\delta(4n+3)}}{B_1B_2} = 0,\\
c_2 &= -\frac{e^{\delta(4n+2)}B_3^2}{B_1B_2B_3} + \frac{e^{\delta(4n+1)}B_2^2}{B_1B_2} - \frac{e^{\delta(4n+1)}B_1^2}{B_1^2} + \frac{e^{\delta(4n+2)}}{B_1B_2} = 0,\\
c_3 &= -\frac{e^{\delta(4n-2)}B_3^3}{B_1B_2B_3} + \frac{e^{\delta(4n-2)}B_2^3}{B_1B_2} - \frac{e^{\delta(4n-1)}B_1^3}{B_1^2} + \frac{e^{\delta(4n+1)}}{B_1B_2} = 0.
\end{aligned}$$

c_3 and c_2 can be derived from c_1 as follows:

- multiply the first term of c_1 by $e^{-(d-1)(j-1)\delta} = e^{-4\delta(j-1)}$ for $j=2,3$ to get the first terms of c_2 and c_3, respectively.
- multiply the second term of c_1 by $e^{-(d-2)(j-1)\delta} = e^{-3\delta(j-1)}$ for $j=2,3$ to get the second terms of c_2 and c_3, respectively.
- multiply the third term of c_1 by $e^{-(d-3)(j-1)\delta} = e^{-2\delta(j-1)}$ for $j=2,3$ to get the third terms of c_2 and c_3, respectively.
- multiply the last term of c_1 by $e^{-(d-4)(j-1)\delta} = e^{-\delta(j-1)}$ for $j=2,3$ to get the last terms of c_2 and c_3, respectively.

A similar pattern follows for all d. For this reason, the proof of Lemma 2 is only given for $j=1$.

Next, the relationship between the polynomials A_i and B_i is

$$B_i = \frac{A_{i+1}}{e^{-\delta}A_1} \quad \text{for} \quad i=1,2,\ldots \tag{9.3}$$

Multiply both sides of (3.7) by the least common denominator

$$\prod_{i=1}^{d-k-2} B_i \prod_{i=1}^{k-2} B_i$$

and then substitute (9.3) in the resulting equation. Let the polynomials be $\{g_i : 1 \leq i \leq d{+}1\}$ and each cumulative sums be $G_3 = g_1 + g_2$ and $\{G_k = G_{k-1} + g_{k-1} : 4 \leq k \leq d\}$. We illustrate this for $d = 4$. From (3.7), we know that the coefficient of $\binom{n-k_{d-3}}{1}$, c_1, is

$$\begin{aligned} c_1 &= e^{(3n+2)\delta}\left[-\frac{e^{\delta}}{B_1} + 1 - \frac{1}{B_1}\right] = e^{(3n+2)\delta}\frac{(-e^{\delta} + B_1 - 1)}{B_1} \\ &= \frac{e^{(3n+2)\delta}}{B_1}\left[-e^{\delta} + \frac{A_2}{e^{-\delta}A_1} - 1\right] = \frac{e^{(3n+2)\delta}}{B_1}\left[-A_1 + A_2 - e^{-\delta}A_1\right] \end{aligned}$$

The polynomial term that appears in above equation is the same as the polynomial that appears in Lemma 1 for $d = 2$, and $c_1 = 0$ since Lemma 1 can be rewritten as

$$\sum_{k=0}^{d}(-1)^{k+1}\frac{e^{-\{k\}\delta}}{\prod_{i=1}^{k} A_i \prod_{i=1}^{d-k} A_i} = 0.$$

In general, the coefficient of $\binom{n-k_{d-3}}{1}$ terms for $d = k$ yields same polynomial that appear in Lemma 1 for $d = k - 2$ – $g_i = f_{i-2}$.

For equation (3.8), we will use the same argument. After multiplying both sides of (3.8) by the least common denominator

$$\prod_{i=1}^{d-k-2} B_i \prod_{i=1}^{k-2} B_i$$

and then substitute (9.3) in the resulting equation. Let each polynomial be $\{h_i : 1 \leq i \leq d+1\}$. For $d = 4$, the coefficient of $\binom{n-k_{d-3}}{0}$ from (3.8) is,

$$e^{(3n+3)\delta}\left[-\frac{e^{3\delta}}{B_1B_2} + \frac{e^{\delta}}{B_1} - \frac{1}{B_1}\right] = -\frac{e^{(3n+3)\delta}}{B_1B_2}$$

After multiplying both sides by B_1B_2 term, substitute (9.3), we get

$$-A_1 + A_3 - e^{\delta}A_3 = -e^{3\delta}A_1.$$

This is equivalent to the polynomials that appear in Lemma 1 for $d = 3$. Thus, once again the coefficient of $\binom{n-k_{d-3}}{0}$ terms for $d = k$ are exactly the same polynomials that appears in Lemma 1 for $d = (k-1)$: $g_i = f_{i-1}$. Since the polynomials in Lemma 2 are just a shifted version of those in Lemma 1, the proof of Lemma 2 follows from Lemma 1. □

Proof of Lemma 3. In the second approach

$$\begin{aligned} P(D=d) &= \sum_{n=d}^{\infty} P_n(D=d)P(N=n) \\ &= \sum_{n=0}^{\infty}\sum_{k=1}^{d} h_{nk}\frac{e^{-\alpha}\alpha^n}{n!} - \left[\sum_{n=1}^{d-1}\sum_{k=1}^{d} h_{nk}\frac{e^{-\alpha}\alpha^n}{n!}\right] - \sum_{k=1}^{d} h_{0k}\frac{e^{-\alpha}\alpha^0}{0!} \\ &= \sum_{n=0}^{\infty}\sum_{k=1}^{d} h_{nk}\frac{e^{-\alpha}\alpha^n}{n!} - e^{-\alpha}\sum_{k=1}^{d} h_{0k} \end{aligned} \tag{9.4}$$

In equation (9.4), h_{0k} is just the polynomials that appears in Lemma 1, and we know that the terms appears bracket has to be zero by Theorem 1:

$$e^{-\alpha}\sum_{n=1}^{d-1}\sum_{k=1}^{d} h_{nk}\frac{\alpha^n}{n!} = 0$$

In order for this equation to hold, the coefficient of the polynomials, $\alpha^i/i!$, must be zero, completing the proof of Lemma 3.

□

9.2 SIMULATION STEPS

9.2.1 For infinite observation period

- Generate N $\sim$ Poisson (α) for $\alpha = 0.5, 1.0, \ldots, 3.0$.
- If $N = 0$, then $D = 0$ and stop.
- If $N \geq 1$, generate $W_1, \ldots, W_N$ which follows iid Exponential(1). Let $W_{(1)}, \ldots, W_{(N)}$ be the corresponding order statistics.
- $W_{(1)}$ is always recorded. Then the detector has its first dead period $[W_{(1)}, (W_{(1)} + \delta)]$.
- Next, count the number of photons that arrive in the first dead period $[W_{(1)}, (W_{(1)} + \delta)]$.
  - If all $W_{(2)}, \ldots, W_{(N)}$ arrive the detector in the detector, then set $D = 1$.
  - If one of $W_{(2)}, \ldots, W_{(N)}$ arrive the detector after $(W_{(1)} + \delta)$, then detector records it, so increment: $D \leftarrow D + 1$. We now have the second dead period $[W_{(i)}, (W_{(i)} + \delta)]$ and the subscript i here is the second recorded photon. Repeat this step until $W_{(N)}$.

9.2.2 For finite observation period

- Generate N $\sim$ Poisson (α) for $\alpha = 0.5, 1.0, \ldots, 3.0$.
- If $N = 0$, then record $[D = 0]$.
- If $N \geq 1$, generate $W_1, \ldots, W_N$ which follows iid Exponential(1). Let $W_{(1)}, \ldots, W_{(N)}$ be the corresponding order statistics.
- If $W_1 > T$, then just record $[D = 0]$.
- If $W_1 \leq T$, $[W_{(1)}, (W_{(1)} + \delta) \wedge T]$ is the first dead period. Now, count the number of photons that arrive in the dead period. If there is any photon arrival after this first dead period $[W_{(1)}, (W_{(1)} + \delta) \wedge T]$, then the detector would record it and so $[W_{(2)}, (W_{(2)} + \delta) \wedge T]$ is the second dead period. Check for the second dead period to calculate the number of photon arrivals in this dead period. Repeat this until $W_{(N)}$.

BIBLIOGRAPHY

[1] Abramowitz M, Stegun I. (1965) *Handbook of Mathematical Functions.* Dover, New York.

[2] Albert GE, Nelson L. (1953) Contributions to the statistical theory of counter data. *The Annals of Mathematical Statistics*, 24, 9-22.

[3] Barriga GDC, Louzada F. (2014) The zero-inflated Conway-Maxwell-Poisson distribution: Bayesian inference, regression modeling and influence diagnostic. *Statistical Methodology.* 21, 23–34.

[4] Bentley DL. (1963) A contribution to counter theory. *Journal of the Royal Statistical Society. Series B.* 25, 169-178.

[5] Bleistein N, Handelsman R. (1986) *Asymptotic Expansion of Integrals.* Dover, New York.

[6] Bokes P, King JR. (2012) Exact and approximate distributions of protein and mRNA levels in the low-copy regime of gene expression. *Journal of Mathematical Biology.* 64, 829–854.

[7] Braaksma BLJ. (1964) Asymptotic expansions and analytic continuations for a class of Barnes' integrals. *Compositio Mathematica.* 15, 239–341.

[8] Breslow N. (1990) Tests of hypotheses in overdispersed Poisson regression and other quasi-likelihood models. *Journal of American Statistical Association.* 85, 565–571.

[9] Conway RW, Maxwell WL. (1962) A queuing model with state dependent service rates. *Journal of Industrial Engineering.* 12, 132–136.

[10] Cordeiro GM, Castro M.(2012) The exponential COM-Poisson distribution. *Statistical Papers.* 53, 653–664.

[11] Cox DR, Hinkley DV. (1979) *Theoretical Statistics.* Chapman and Hall.

[12] Daley DJ, Vere-Jones D. (1988) *An Introduction to the Theory of Point Processes*. Springer.

[13] Denk W, Strickler JH, Webb WW. (1990) Two-photon laser scanning fluorescence microscopy. *Science*. 248, 73-76.

[14] Driscoll, JD, Shih AY, Iyengar S, Field JJ, White GA, Squier JA, Cauwenberghs G, Kleinfeld D. (2011) Photon counting, censor corrections, and lifetime imaging for improved detection in two-photon microscopy. *Journal of Neurophysiology* 105, 3106-3113.

[15] Driscoll J, Shih A, Iyengar S, Field J, White G, Squier JA, Cauwenberghs G, Kleinfeld D. (2011) Supplementary information on estimating the true emission count for "Photon counting, censor corrections, and lifetime imaging for improved detection in two-photon microscopy."

[16] Efron BN, Hinkley DV. (1978) Assessing the accuracy of the maximum likelihood estimator: observed versus expected Fisher information. *Biometrika*, 65, 457 - 487.

[17] Elgart V, Jia T, Kulkarni RV. (2010) Applications of Little's law to stochastic models of gene expression. *Physical Review E.* 82, 021901(1–6).

[18] Ericson MB, Simonsson C, Ljungblad C, Smedh M. (2008) Two-photon laser-scanning fluorescence microscopy applied for studies of human skin. *Journal of Biophotonics*. 1, 320–330.

[19] Feller W. (1948) On probability problems in the theory of counters. *Courant Anniversary.* 8, 105-115.

[20] Gillispie SB, Green CG. (2015) Approximating the Conway-Maxwell-Poisson distribution normalization constant. *Statistics.* 49, 1062-1073.

[21] Golding I, Cox EC. (2006) Eukaryotic transcription: What does it mean for a gene to be 'on'?. *Current Biology.* 16, R371–R373.

[22] Golding I, Paulsson J, Zawilski SM, Cox EC. (2005) Real-time kinetics of gene activity in individual bacteria. *Cell.* 123, 1025–1036.

[23] Gorissen M, Vanderzande C. (2011) Semi-Markov models of mRNA-translation. *arXiv:1104.0131.*

[24] Guikema SD, Goffelt JP. (2008) A flexible count data regression model for risk analysis. *Risk Analysis.* 28, 213–223.

[25] Helmchen F, Denk, W. (2005) Deep tissue two-photon microscopy. *Nature Methods*, 2, 932 – 940.

[26] Johnson NLKS, Kemp AW (1992) *Univariate Discrete Distributions.* Wiley, New York.

[27] Josic K, Lopez JM, Ott W, Shiau L, Bennett MR. (2011) Stochastic delay accelerates signaling in gene networks. *PLOS Computational Biology:* 1002264e.

[28] Imoto K. (2014) A generalized Conway-Maxwell-Poisson distribution which includes the negative binomial distribution. *Applied Mathematics and Computation.* 247, 824–834.

[29] Kadane JB, Shmueli G, Minka TP, Borle S, Boatwright P. (2006) Conjugate analysis of the Conway-Maxwell-Poisson distribution. *Bayesian Analysis.* 1, 363374.

[30] Kao EPC, Marion SS. (1993) On renewal processes relating to counter models: the case of phase-type interarrival times. *Journal of Applied Probability.* 30, 175-183.

[31] Lee J, Shoudan Y, Liang S, Cai G, Muller P. (2015) Bayesian hierarchical model for differential gene expression using RNA-Seq data. *Statistics in Biosciences.* 7, 48–67.

[32] Lehmann EL. (1986) *Testing Statistical Hypotheses, 2nd edition.* Wiley.

[33] Lord D, Guikema SD, Geedipally SR. (2008) Application of the Conway-Maxwell-Poisson generalized linear model for analyzing motor vehicle crashes. *Accident Analysis & Prevention.* 40, 1123–1134.

[34] Lord D, Guikema SD. (2012) The Conway-Maxwell-Poisson model for analyzing crash data. *Applied Statistical Models in Business and Industry.* 28, 122–127.

[35] Lord D, Mannering F. (2010) The statistical analysis of crash-frequency data: Areview and assesmnet of methodological alternatives. *Transpotation Research Part A: Policy and Practice.* 44, 291–305.

[36] Maier T, Guell M, Serrano L. (2009) Correlation of mRNA and protein in complex biological samples. *FEBS Letters.* 583, 3966–3973.

[37] McConnell G, Erling R. (2004) Two-photon laser scanning fluoresence microscopy using photonic crystal fiber. *Journal of Biomedical Optics.* 9, 922–927.

[38] McLachlan GJ, Krishnan T. (1997) *The EM algorithm and extensions.* Wiley.

[39] Minka T, Shmueli G, Kadane JB, Borle S, Boadwright P. (2003) Computing with the COM-Poisson distribution. *Carnegie Mellon Statistics Department Technical Report.* 776.

[40] Mulligan SJ, MacVicar BA. (2007) Two-photon fluorescence microscopy: Basic principles, advantages and risks. *Modern Research and Educational Topics in Microscopy.* 881–889.

[41] Nadarajah S. (2009) Useful moment and CDF formulations for the COM-Poisson distribution. *Statistics Papers.* 50, 617–622.

[42] Oberg AL, Bot BM, Grill DE, Polan GA, Therneau TM. (2012) Technical and biological variance structure in mRNA-data:life in the real world.*BMC Genomics.* DOI: http://dx.doi.org/10.1186/1471-2164-13-304.

[43] Oheim M, Michael DJ, Geisbauer M, Madsen D, Chow RH. (2006) Principles of two-photon excitation fluorescence microscopy and other nonlinear imaging approaches. *Advanced Drug Delivery Reviews.* 58, 788-808.

[44] Peccoud J, Ycart B. (1995) Markovian modeling of gene-product synthesis. *Theoretical Population Biology.* 48, 222234.

[45] Pendar H, Platini T, Kulkarni RV. (2013) Exact protein distributions for stochastic models of gene expressions using partitioning of Poisson processes. *Physocal Review E.* 87, 042720.

[46] Pyke R. (1958) On renewal processes related to type I and type II counter models. *The Annals of Mathematical Statistics.* 29, 737-754.

[47] Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S. (2006) Stochastic mRNA Synthesis in Mammalian Cells. *PLOS Biology.* 4, 1707–1719.

[48] Ridout MA, Besbeas P. (2004) An empirical model for underdispersed count data. *Statistical Modelling.* 4, 77–89.

[49] Rodrigues J, Castro M, Vicente CG, Balakrishnan N. (2009) COM-Poisson cure rate survival models and application to cutaneous melanoma data. *Jornal of Statistical Planning and Inference.* 139, 3605–3611.

[50] Sellers KF, Borle S, Shmueli G. (2012) The COM-Poisson model for count data: a survey of methods and applications. *Applied Stochastic Models in Business and Industry.* 28, 104–116.

[51] Sellers KF, Shmueli G. (2008) A flexible regression model for count data. *Robert H. Smith School Research Paper No. RHS.* 06–060.

[52] Sellers KF, Shmueli G. (2013) Data dispersion: now you see it, now you dont. *Communications in Statistics Theory and Methods.* 42, 3134-3147.

[53] Shahrezaei V and Swain PS. (2008) Analytical distributions for stochastic gene expression. *Proceedings of the National Academy of Sciences.* 45, 17256–17261.

[54] Shmueli G, Minka T, Kadane JB, Borle S, Boatwright P. (2005) A useful distribution for fitting discrete data: revival of the Conway-Maxwell-Poisson distribution. *Journal of Royal Statistical Society, C.* 54, 127–142.

[55] Simsek B, Iyengar S. (2016) On the distribution of photon counts with censoring in two-photon laser scanning microscopy. Submitted.

[56] Simsek B, Iyengar S. (2016) On distributions arising from photon counting statistics. Submitted.

[57] Simsek B, Iyengar S. (2016) On the approximation of the Conway-Maxwell-Poisson normalizing constant. *To appeared in Filomat.*

[58] So PTC. (2002) Two-photon fluorescence light microscopy. *Encyclopedia of Life Sciences Macmillan Publishers Ltd, Nature Publishing Group.* www.els.net

[59] So PTC, Dong CY, Masters BR, Berland KM. (2000) Two-photon excitation fluorescence microscopy. *Annual Review of Biomedical Engineering.* 2, 399-429.

[60] So L, Ghosh A, Zong C, Sepulveda LA, Segev R, Golding I. (2011) General properties of transcriptional time series in Escherichia coli. *Nature Genetics.* 43, 554560.

[61] Source: http://mcb.berkeley.edu/labs2/robey/content/2-photon-imaging

[62] Source: http://stochpy.sourceforge.net/examples.html

[63] Svoboda K, Yasuda R. (2006) Principles of two-photon excitation microscopy and its applications to neuroscience. *Neuron.* 50, 823-839.

[64] Takàcs, L. (1956) On a probability problem arising in the theory of counters. *Proceedings of the Cambridge Philosophical Society.* 52, 488-498.

[65] Tauer U. (2002) Advantages and risks of multiphoton microscopy in physiology. *Experimental Physiology.* 87, 709–714.

[66] Vicente CG, Castro M, Rodrigues J. (2010) A bayesian analysis of the Conway-Maxwell-Poisson cure rate model. *Statistical Papers.* 53, 165–176.

[67] Yu DF, Fessler JA. (2000) Mean and variance of single photon counting with deadtime. *Physics in Medicine and Biology*. 45, 2043-2056.

[68] Zenklusen D, Larson DR, Singer RH. (2008) Single-RNA counting reveals alternative modes of gene expression in yeast. *Natural Structural & Molecular Biology*. 15, 1263–1271.

[69] Zhang H, Pounds SB, Tang L. (2013) Statistical methods for overdispersion in mRNA-Seq count data. *The Open Bioinformatics Journal*. 7, 34 – 40.

[70] Zou Y, Lord D, Greedipally SR. (2012) Over and under dispersed data: Comparing Conway-Maxwell-Poisson and Double-Poisson Distributions. *Transportation Research Board of the National Academies*. Washington, DC.