ASYMPTOTIC NORMALITY OF HIGHER ORDER TURING FORMULAE

by

Jie Chang

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Applied Mathematics

Charlotte

2022

Approved by:

____

Dr. Michael Grabchak

____

Dr. Jiancheng Jiang

____

Dr. Eliana Christou

____

Dr. Donald Jacobs

ABSTRACT

JIE CHANG. Asymptotic Normality of Higher Order Turing Formulae. (Under the direction of DR. MICHAEL GRABCHAK)

Higher order Turing formulae, denoted as T_r for $r \in \mathbf{Z}^+$, are a powerful result allowing one to estimate the total probability associated with words from a random piece of writing, which have been observed exactly r times in a random sample. In particular T_0 estimates the probability of seeing words not appearing in the sample. To perform statistical inference, e.g., constructing the asymptotic confidence intervals, the asymptotic properties of the higher Turing formulae need to be studied.

In this thesis we extend the validity of the asymptotic normality beyond the previously proven cases by establishing a sufficient and necessary condition for the asymptotic normality of higher order Turing formulae when the underlying distribution is both fixed and changing. We then conduct simulation studies with the complete works of William Shakespeare and data generated from different underlying distributions to check the finite sample performance of the derived asymptotic confidence interval.

Based on our theoretical results we also develop two methodologies for authorship detection with real twitter data analysis.

DEDICATION

To my lovely daughter, Ke, who is my everlasting light whenever I am in a dark life tunnel.

ACKNOWLEDGEMENTS

I must start by expressing my sincere and deepest thanks to my advisor, Dr. Michael Grabchak, for his guidance, constant support and encouragement throughout those four years. I learned from him not only knowledge but also science attitudes, which I shall carry forever. I will never forget in those weekly meetings he demonstrated repeatedly and patiently to clear my thoughts whenever I got stuck in these proofs. He put faith in me and gave me flexibility when I was having a hard time after I gave birth to my daughter. It was from him that I learned the beauty of mathematics and I am proud of choosing him as my academic advisor and life mentor.

I would also like to extend my thanks to Dr. Jiancheng Jiang, Dr. Eliana Christou, and Dr. Donald Jacobs for serving on my committee. I also want to thank Dr. Zhiyi Zhang for the talk on Turing formula in the UNC Charlotte Probability Seminar, which inspired me to study further on this topic.

Thanks also go to Dr. Shaozhong Deng and Dr. Mohammad A. Kazemi for the academic and program advice. My appreciation also goes to my friends, Masoumeh Sheikhi, Yunfei Xia, Xiaoyun Chen, Xueji Wang, and Yuehan Shao for their encouragement and companionship throughout my Ph.D study.

As a first generation college student, I must thank my parents for not setting limits but being unconditionally supportive during my study. I thank my grandma for raising me up. I also thank my husband for his encouragements and supports. Without their love, I could not have completed this dissertation.

Finally, I would like to thank the UNC Charlotte Department of Mathematics and Statistics for the teaching assistant position, the graduate school for the financial support from the GASP grants, and the ISSO for all the international student services.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

"Those who can imagine anything, can create the impossible." – *Alan Turing*

Given a random piece of an author's work, how can we estimate the probability of the author using a word that has not been used before or the probability of using a word that has been used exactly r times in the sample piece? This problem can be generalized to many other practical situations where data has no natural ordering and is categorical in nature, for example, in ecology the words may represent the species in an ecosystem, in biomedical applications they may represent different types of cancer cells in a tumor. Statistical properties of the probability when $r = 0$, which corresponds to the probability of seeing something that has not been seen before and is called the missing mass, have been studied in e.g. [1, 2, 3, 4]. Applications of estimating these probabilities arise in many fields including: ecology [5] [6] [7], genomics [8], natural language processing [9] [10], authorship attribution [11] [12] [13], and computer networks [14].

It has been long recognized that the usual maximum likelihood estimator does not work well for estimating such probabilities. However, Alan Turing developed an alternate approach by giving a mind-bending nonparametric estimator when he was working to decode the Enigma cipher during World War II. It was first introduced by his assistant I.J. Good in [15], and has come to be called Turing's formula or the Good-Turing formula. Turing's intuitive explanation of this formula was claimed to be given to Good, but has been lost, see [15]. Nevertheless, use of the estimator is justified by its many statistical properties.

One of the earliest studies of the statistical properties of Turing's formula is [16], where it is shown that the estimator is not unbiased, but that it would be if we had an additional observation. Detailed formulas for the bias can be found in [17] and [13]. Conditions for consistency are given in [18] and a simulation study focused on the rate of convergence is given in [19]. The problem of asymptotic normality has primarily been studied in the case when $r = 0$. In this case, sufficient conditions are given in [20], [21], and [22] and a necessary and sufficient condition is given in [23]. When $r > 0$, sufficient conditions are given in [24] and [25]. These results, along with a wealth of additional information, are summarized in the recent monograph on Turing's formula [26].

In this thesis we extend the validity of the asymptotic normality beyond the previously proven cases by giving necessary and sufficient conditions for the asymptotic normality of Turing's formula for any $r \geq 0$ when the underlying distribution is both fixed and changing.

The rest of this thesis is organized as follows. In Chapter 2 we present our theoretical results. First we introduce our mathematical framework, next we prove the case where the number of observations follows a Poisson distribution, then we extend our results to a general deterministic case by approximation and introduce a formula to construct the asymptotic confidence interval only with knowledge of the sample, and last we give two examples of distributions where our conditions are satisfied. In Chapter 3 we conduct simulation studies with the complete works of William Shakespeare and data generated from different underlying distributions to check the finite sample performance of the derived asymptotic confidence interval. In Chapter 4 we use our theoretical results to develop two methodologies for authorship detection. We further apply them to analyze real twitter data and present the results. In Chapter 5 we first briefly revisit our main results by pointing out their importance and significance, and then discuss our future work for the improvement of the data application. We postpone proofs to Chapter 6, where details of proofs for the results in Chapter 2 can be found along with several lemmas that may be of independent interest.

Before proceeding we introduce some notation. For two sequences of real numbers $\{a_n\}$ and $\{b_n\}$, we write $a_n \sim b_n$ if $a_n/b_n \to 1$. We write $\lfloor \cdot \rfloor$ and $\lceil \cdot \rceil$ to denote the floor and ceiling functions, respectively. We write $N(0, 1)$ to denote the standard normal distribution and $\text{Pois}(\lambda)$ to denote a Poisson distribution with mean λ. We write $1_{[\cdots]}$ to denote the indicator function on event $[\cdots]$.

CHAPTER 2: NECESSARY AND SUFFICIENT CONDITIONS FOR ASYMPTOTIC NORMALITY OF HIGHER ORDER TURING FORMULAE

2.1 Introduction

Turing formulae are estimators of the total probability/mass of letters observed exactly r times in a random sample. It is not a conventional estimator as it estimates a quantity that depends not only on the population, but on the random sample as well. Though Turing's intuitive explanation for the Good-Turing formula has been lost, see Good [15], attempts to justify its use has never stopped and many applications have been inspired in different fields. In this chapter we study the asymptotic behavior of one modification of Turing formulae for any order and give the necessary and sufficient conditions for it to enrich the literature. Our results allow for many situations that were not covered by previously available sufficient conditions.

We begin by introducing our mathematical framework. Then the main theoretical results are presented for the Poisson case and the Deterministic case, respectively, and their definitions and schemes will be discussed in the following subsections. Two theoretical distribution examples are given in the last subsection to demonstrate how our asymptotic normality conditions can be satisfied.

The Poisson case is studied first as a foundation for proving the deterministic case, nevertheless, it contains results of independent interest. Then, the Deterministic case is approximated by the Poisson case.

2.2 Mathematical Framework

Now we formulate our mathematical scheme in an alphabet context for a simple fixed case.

Let the alphabet $\mathcal{A} = \{a_1, a_2, a_3, ...\}$ be a finite or countably infinite alphabet with associated probability measure $\mathcal{P}_m = \{p_{a,m} : a \in \mathcal{A}\}$ for $m = 1, 2, 3, \ldots$. If there is a distribution $\mathcal{P}$ with $\mathcal{P}_m = \mathcal{P}$ for every m, we say that the distribution is fixed. Otherwise, we say that it is changing. In particular applications the letters of $\mathcal{A}$ may correspond to species in an ecosystem, words in the English language, types of cancer cells in a tumor, or another quantity of interest.

Let $X_1, X_2, ..., X_n$ be a random sample on alphabet $\mathcal{A}$ with distribution $\mathcal{P}$. For each $a \in \mathcal{A}$, let $y'_a = \sum_{i=1}^{n} 1_{[X_i = a]}$ be the sample count of letter a and let $\hat{p}_a = \frac{\sum_{i=1}^{n} 1_{[X_i=a]}}{n} = \frac{y'_a}{n}$ be the sample proportion of letter a. For $r = 0, 1, 2, \ldots, n$, let

$$N'_r = \sum_{a \in \mathcal{A}} 1_{[y'_a = r]}$$

be the number of letters observed exactly r times in the sample, and let

$$\pi'_r = \sum_{a \in \mathcal{A}} p_a 1_{[y'_a = r]}$$

be the total mass of all letters observed exactly r times in the sample. Define further, for $r = 0, 1, 2, ..., (n-1)$,

$$T'_r = \frac{N'_{r+1}}{n} (r+1) .$$

We call T'_r the rth order Turing formula. It is an estimator of π'_r. We notice that there are slightly different versions for Turing's formula used in [24] and [25] for $r \geq 1$. Specifically, they use $T^*_r = \frac{N'_{r+1}}{n-r}(r+1)$. Asymptotically there is no difference and we use T'_r for convenience. We note that T'_r is the form that was originally introduced in [15].

Our ultimate goal is to find conditions for asymptotic normality, specifically when there exists a function g such that

$$g(n)(T'_r - \pi'_r) \xrightarrow{d} N(0,1).$$

2.3 Poisson Case

In this section we discuss two cases where the sample size is random and follows a Poisson distribution. One case is when there is one Poisson distribution $\mathcal{P}$ on the alphabet $\mathcal{A}$, which we say that the distribution is fixed. The other case is when there is a sequence of Poisson distributions $\mathcal{P}_n$ on the alphabet $\mathcal{A}$, which we say that the distribution is changing.

2.3.1 Poisson Case with Fixed Distribution

We begin with the Poisson case with a fixed distribution, where the sample size $N \sim \text{Pois}(\lambda)$ and $\lambda \to \infty$. Let $y_a(\lambda)$ be the number of times that we saw letter a in the sample. By Poisson thinning, these are independent Poisson random variables with

$$\mathrm{E}[y_a(\lambda)] = \lambda p_a.$$

For $r = 0, 1, 2, \ldots, n$, let

$$N_r = N_r(\lambda) = \sum_{a \in \mathcal{A}} 1_{[y_a(\lambda)=r]}$$

be the number of letters observed exactly r times in the sample, and let

$$\pi_r = \pi_r(\lambda) = \sum_{a \in \mathcal{A}} p_a 1_{[y_a(\lambda)=r]}$$

be the total mass of all letters observed exactly r times in the sample. Define further, for r = 0, 1, 2,...,$(n-1)$,

$$T_r = T_r(\lambda) = \frac{N_{r+1}}{\lambda}(r+1).$$

We call T_r the rth order Turing formula. It is an estimator of π_r.

Our goal is to find conditions for asymptotic normality.

Note that

$$\mathrm{E}[N_r] = \sum_{a\in\mathcal{A}} e^{-\lambda p_a}\frac{\lambda (p_a)^r}{r!}$$

and

$$\mathrm{E}[\lambda\pi_r] = \sum_{a\in\mathcal{A}} e^{-\lambda p_a}\frac{(\lambda p_a)^{r+1}}{r!}.$$

Now set

$$\lambda(T_r(\lambda) - \pi_r(\lambda)) = \sum_{a\in\mathcal{A}} Y_a$$

where

$$Y_a = (r+1)\,1_{[y_a(\lambda)=r+1]} - \lambda p_a 1_{[y_a(\lambda)=r]}.$$

$Y_a's$ are independent random variables, because Y_a is a function only of y_a and they are independent random variables. Since $[y_a(\lambda)=r]\cap[y_a(\lambda)=r+1]=\emptyset$, we have

$$Y_a^2 = (r+1)^2\,1_{[y_a(\lambda)=r+1]} + \lambda^2 p_a^2 1_{[y_a(\lambda)=r]}.$$

It follows that

$$\mathrm{E}\left[Y_a\right] = e^{-\lambda p_a} \frac{(\lambda p_a)^{r+1}}{r!} - \lambda p_a e^{-\lambda p_a} \frac{(\lambda p_a)^r}{r!} = 0$$

and

$$\begin{aligned}\sigma_{a,\lambda}^2 &= \mathrm{Var}\left(Y_a\right) \\ &= \mathrm{E}\left[Y_a^2\right] \\ &= (r+1)\, e^{-\lambda p_a} \frac{(\lambda p_a)^{r+1}}{r!} + \lambda p_a e^{-\lambda p_a} \frac{(\lambda p_a)^{r+1}}{r!} \\ &= (r+1+\lambda p_a)\, e^{-\lambda p_a} \frac{(\lambda p_a)^{r+1}}{r!}.\end{aligned}$$

Let

$$s_\lambda^2 = \sum_{a\in\mathcal{A}} \sigma_{a,\lambda}^2 = \sum_{a\in\mathcal{A}} (r+1+\lambda p_a)\, e^{-\lambda p_a} \frac{(\lambda p_a)^{r+1}}{r!},$$

and note that

$$s_\lambda^2 = (r+1)^2\, \mathrm{E}\left[N_{r+1}\right] + (r+2)\,(r+1)\, \mathrm{E}\left[N_{r+2}\right].$$

Now we give main results for this case.

Theorem 1. *Assume that $s_\lambda \to \infty$ as $\lambda \to \infty$. We have*

$$\lim_{\lambda\to\infty} s_\lambda^{-2} \sum_{a\in\mathcal{A}} e^{-\lambda p_a} (\lambda p_a)^{(r+2)} 1_{[\lambda p_a \geq \epsilon s_\lambda]} = 0 \quad \forall \epsilon > 0 \tag{2.1}$$

if and only if

$$\frac{\lambda}{s_\lambda}\left(T_r(\lambda) - \pi_r(\lambda)\right) \xrightarrow[\lambda\to\infty]{d} N(0,1).$$

Corollary 1. *If the conditions in Theorem 1 and* (2.1) *hold, then*

$$P\left(\left|\frac{T_r(\lambda)}{\pi_r(\lambda)} - 1\right| > \epsilon\right) \to 0 \quad \forall \epsilon > 0,$$

(i.e., $\frac{T_r(\lambda)}{\pi_r(\lambda)} \xrightarrow{p} 1$*).*

Corollary 2. *Let*

$$(\hat{s}_\lambda)^2 = (r+1)^2 N_{r+1} + (r+2)(r+1)N_{r+2}.$$

If the conditions in Theorem 1 hold, then $(\hat{s}_\lambda)^2$ *is a consistent estimator of* s_λ^2*, i.e., as* $\lambda \to \infty$*, for all* $\epsilon > 0$

$$P\left(\left|\frac{(\hat{s}_\lambda)^2}{s_\lambda^2} - 1\right| > \epsilon\right) \to 0.$$

2.3.2 Poisson Case with Changing Distribution

Now we come to the Poisson case with a changing distribution. Consider a sequence of positive real numbers $\{\lambda_n\}$ such that $\lambda_n \to \infty$. We assume the sample size $N_n \sim \text{Pois}(\lambda_n)$, and the corresponding underlying distribution $\mathcal{P}_n = \{p_{a,n} : a \in \mathcal{A}\}$ where $\mathcal{A} = \{a_1, a_2, a_3, ...\}$ is a countably infinite alphabet. We are given a random sample of size N_n on alphabet $\mathcal{A}$ with distribution $\mathcal{P}_n$.

Let $y_{a,n}(\lambda_n)$ be the number of times that we see letter a in the sample. These are independent Poisson random variables with

$$\text{E}[y_{a,n}(\lambda_n)] = \lambda_n p_{a,n}.$$

For r= 0, 1, 2,..., let

$$N_{r,n} = N_{r,n}(\lambda_n) = \sum_{a \in \mathcal{A}} 1_{[y_{a,n}(\lambda_n) = r]}$$

be the number of letters observed exactly r times in the sample, and let

$$\pi_{r,n} = \pi_{r,n}(\lambda_n) = \sum_{a\in\mathcal{A}} p_{a,n} 1_{[y_{a,n}(\lambda_n)=r]}$$

be the total mass of all letters observed exactly r times in the sample. Define further, for r = 0, 1, 2,...,(n-1),

$$T_{r,n} = T_{r,n}(\lambda_n) = \frac{N_{r+1,n}}{\lambda_n}(r+1).$$

We call $T_{r,n}$ the rth order Turing formula. It is an estimator of $\pi_{r,n}$. Our goal is to find conditions for asymptotic normality.

Note that

$$\mathrm{E}[N_{r,n}] = \sum_{a\in\mathcal{A}} e^{-\lambda_n p_{a,n}} \frac{(\lambda_n p_{a,n})^r}{r!}$$

and

$$\mathrm{E}[\lambda_n \pi_{r,n}] = \sum_{a\in\mathcal{A}} e^{-\lambda_n p_{a,n}} \frac{(\lambda_n p_{a,n})^{r+1}}{r!}.$$

Now set

$$\lambda_n(T_{r,n}(\lambda_n) - \pi_{r,n}(\lambda_n)) = \sum_{a\in\mathcal{A}} Y_{a,n}$$

where

$$Y_{a,n} = (r+1)\,1_{[y_{a,n}(\lambda_n)=r+1]} - \lambda_n p_{a,n} 1_{[y_{a,n}(\lambda_n)=r]}$$

are independent random variables. Since $[y_{a,n}(\lambda_n)=r] \cap [y_{a,n}(\lambda_n)=r+1] = \emptyset$, we have

$$Y_{a,n}^2 = (r+1)^2 1_{[y_{a,n}(\lambda_n)=r+1]} + \lambda_n^2 p_{a,n}^2 1_{[y_{a,n}(\lambda_n)=r]}.$$

It follows that

$$\mathrm{E}[Y_{a,n}] = e^{-\lambda_n p_{a,n}} \frac{(\lambda_n p_{a,n})^{r+1}}{r!} - \lambda_n p_{a,n} e^{-\lambda_n p_{a,n}} \frac{(\lambda_n p_{a,n})^r}{r!} = 0$$

and

$$\begin{aligned}
\sigma_{a,\lambda_n}^2 &= \mathrm{Var}(Y_{a,n}) \\
&= \mathrm{E}\left[Y_{a,n}^2\right] \\
&= (r+1)\, e^{-\lambda_n p_{a,n}} \frac{(\lambda_n p_{a,n})^{r+1}}{r!} + \lambda_n p_{a,n} e^{-\lambda_n p_{a,n}} \frac{(\lambda_n p_{a,n})^{r+1}}{r!} \\
&= (r+1+\lambda_n p_{a,n})\, e^{-\lambda_n p_{a,n}} \frac{(\lambda_n p_{a,n})^{r+1}}{r!}.
\end{aligned}$$

Let

$$s_{\lambda_n}^2 = \sum_{a\in\mathcal{A}} \sigma_{a,\lambda_n}^2 = \sum_{a\in\mathcal{A}} (r+1+\lambda_n p_{a,n})\, e^{-\lambda_n p_{a,n}} \frac{(\lambda_n p_{a,n})^{r+1}}{r!},$$

and note that

$$s_{\lambda_n}^2 = (r+1)^2\, \mathrm{E}[N_{r+1,n}] + (r+2)(r+1)\, \mathrm{E}[N_{r+2,n}].$$

Note further that, in this case, Turing's formula is unbiased and we have

$$\mathrm{E}[T_{r,n}(\lambda_n)] = \frac{1}{\lambda_n} \sum_{a\in\mathcal{A}} e^{-\lambda_n p_{a,n}} \frac{(\lambda_n p_{a,n})^{r+1}}{r!} = \mathrm{E}[\pi_{r,n}(\lambda_n)]. \tag{2.2}$$

The main results for this case are given as follows.

Theorem 2. *Assume that* $s_{\lambda_n} \to \infty$ *as* $\lambda_n \to \infty$ *and* $\lambda_n \to \infty$ *as* $n \to \infty$. *We have*

$$\lim_{n\to\infty} s_{\lambda_n}^{-2} \sum_{a\in\mathcal{A}} e^{-\lambda_n p_{a,n}} (\lambda_n p_{a,n})^{(r+2)} 1_{[\lambda_n p_{a,n} \geq \epsilon s_{\lambda_n}]} = 0 \quad \forall\epsilon > 0 \tag{2.3}$$

if and only if

$$\frac{\lambda_n}{s_{\lambda_n}} \left(T_{r,n}(\lambda_n) - \pi_{r,n}(\lambda_n)\right) \xrightarrow[n\to\infty]{d} N(0,1).$$

Corollary 3. *If the conditions in Theorem 2 and* (2.3) *hold, then*

$$P\left(\left|\frac{T_{r,n}(\lambda_n)}{\pi_{r,n}(\lambda_n)} - 1\right| > \epsilon\right) \to 0 \quad \forall\epsilon > 0,$$

(i.e., $\frac{T_{r,n}(\lambda_n)}{\pi_{r,n}(\lambda_n)} \xrightarrow{p} 1$*).*

Corollary 4. *Let*

$$(\hat{s}_{\lambda_n})^2 = (r+1)^2 N_{r+1} + (r+2)(r+1) N_{r+2}.$$

If the conditions in Theorem 2 hold, then $(\hat{s}_{\lambda_n})^2$ *is a consistent estimator of* $s^2_{\lambda_n}$, *i.e., as* $\lambda_n \to \infty$, *for all* $\epsilon > 0$

$$P\left(\left|\frac{(\hat{s}_{\lambda_n})^2}{s^2_{\lambda_n}} - 1\right| > \epsilon\right) \to 0.$$

Proof. Since $(r+1)^2 > 0$ and $(r+2)(r+1) > 0$, the result is an application of Lemma 7. □

When s_{λ_n} does not approach infinity we do not get asymptotic normality. Instead, we get a Poisson distribution in the limit. We now give conditions for the Poisson approximation in this case.

Theorem 3. *Fix* $r \in \{0, 1, 2, ...\}$. *Assume that* $s_{\lambda_n,n} \to c \in (0, \infty)$ *and set* $c^* = c^2/(r+1)^2$. *If* (2.3) *holds, then* $\mathrm{E}[N_{r+1,n}] \to c^*$, $\mathrm{E}[N_{r+2,n}] \to 0$,

$$\mathrm{E}\left(\frac{\lambda_n}{r+1}\pi_{r,n}(\lambda_n) - c^*\right)^2 \to 0, \ \frac{\lambda_n}{r+1}\pi_{r,n}(\lambda_n) \xrightarrow{p} c^*, \tag{2.4}$$

and

$$\frac{\lambda_n}{r+1}T_{r,n}(\lambda_n) \xrightarrow{d} \mathrm{Pois}(c^*).$$

2.4 Deterministic Case

In this section we move to the case where the sample size is deterministic, hereafter called the Deterministic case. We consider situations when the underlying distribution is both fixed and changing. Here when we say that the distribution is fixed, it means that there is one $\mathcal{P}_m$ for all $m = 1, 2, 3, \ldots$. Otherwise, we say that the distribution is changing.

2.4.1 Deterministic Case with Fixed Distribution

First we introduce the deterministic case when the underlying distribution is fixed.

Now consider the case of a deterministic sample size n. Without loss of generality let $C = \{C_\lambda : \lambda \geq 0\}$, which is a Poisson process with rate 1, thus $\mathrm{E}[C_\lambda] = \lambda$. Let $y'_a(n)$ be the counts in the first n observations and let $y_a(\lambda)$ be the counts in the first C_λ observations. For $r = 0, 1, 2, ...$, let

$$N'_r(n) = \sum_{a \in \mathcal{A}} 1_{[y'_a(n)=r]}$$

$$\pi'_r(n) = \sum_{a \in \mathcal{A}} p_a 1_{[y'_a(n)=r]}.$$

For $r = 0, 1, 2, ..., (n-1)$, let

$$T'_r(n) = \frac{N'_{r+1}(n)}{n}(r+1).$$

It is readily checked that

$$\mathrm{E}[\pi'_r] = \binom{n}{r} \sum_{a\in\mathcal{A}} p_a^{r+1}(1-p_a)^{n-r} \text{ and } \mathrm{E}[N'_r] = \binom{n}{r} \sum_{a\in\mathcal{A}} p_a^{r}(1-p_a)^{n-r}.$$

Its bias is given by

$$\mathrm{E}\left[T'_r - \pi'_r\right] = \binom{n}{r} \sum_{a\in\mathcal{A}} p_a^{r+1}(1-p_a)^{n-r-1}\left(p_a - \frac{r}{n}\right).$$

We now give our main results for this case.

Theorem 4. *Fix $r \in \{0, 1, 2, ...\}$. Assume that $s_n \to \infty$ as $n \to \infty$ and*

$$\frac{s_n}{\sqrt{n}} \to 0.$$

In this case

$$\lim_{n\to\infty} s_n^{-2} \sum_{a\in\mathcal{A}} e^{-np_a}(np_a)^{(r+2)} 1_{[np_a \geq \epsilon s_n]} = 0 \quad \forall \epsilon > 0 \tag{2.5}$$

if and only if

$$\frac{n}{s_n}\left(T'_r(n) - \pi'_r(n)\right) \xrightarrow{d} N(0,1).$$

Corollary 5. *In the Poissonized case*

$$s_n^2 = (r+1)^2 \,\mathrm{E}\left[N_{r+1}\right] + (r+2)(r+1)\,\mathrm{E}\left[N_{r+2}\right],$$

and in the deterministic case

$$(s'_n)^2 = (r+1)^2\,\mathrm{E}\left[N'_{r+1}\right] + (r+2)\,(r+1)\,\mathrm{E}\left[N'_{r+2}\right].$$

Fix $r \in \{0, 1, 2, ...\}$. *If* $s_n \to \infty$ *as* $n \to \infty$, *then* $(s'_n)^2 \sim s_n^2$, *i.e.*

$$\frac{(s'_n)^2}{s_n^2} \xrightarrow{p} 1.$$

Corollary 6. *If the conditions in Theorem 4 hold, then* (2.5) *holds if and only if*

$$\frac{n}{s'_n}\left(T'_r(n) - \pi'_r(n)\right) \xrightarrow{d} N(0,1). \tag{2.6}$$

Corollary 7. *If the conditions in Theorem 4 and* (2.6) *hold, then*

$$P\left(\left|\frac{T'_r(n)}{\pi'_r(n)} - 1\right| > \epsilon\right) \to 0 \quad \forall \epsilon > 0,$$

(i.e., $\frac{T'_r(n)}{\pi'_r(n)} \xrightarrow{p} 1$*).*

Corollary 8. *For the deterministic case let*

$$s_n^2 = (r+1)^2\mathrm{E}[N'_{r+1}] + (r+2)(r+1)\mathrm{E}[N'_{r+2}]$$
$$(\hat{s}_n)^2 = (r+1)^2 N'_{r+1} + (r+2)(r+1)N'_{r+2}.$$

If the conditions in Theorem 4 hold, $(\hat{s}_n)^2$ *is a consistent estimator of* s_n^2, *i.e., as* $n \to \infty$, *for all* $\epsilon > 0$

$$P\left(\left|\frac{(\hat{s}_n)^2}{s_n^2} - 1\right| > \epsilon\right) \to 0.$$

2.4.2 Deterministic Case with Changing Distribution

In this subsection we show results for Deterministic case when the underlying distribution is changing.

Consider the deterministic case of a fixed sample of size n. $C = \{C_\lambda : \lambda \geq 0\}$ is a Poisson Process with rate 1, where $\mathrm{E}_n[C_\lambda] = \lambda$. Let $y'_{a,n}(n)$ be the counts in the first n observations and let $y_{a,n}(\lambda)$ be the counts in the first C_λ observations.

For $r = 0, 1, 2, ...,$ let

$$N'_{r,n}(n) = \sum_{a \in \mathcal{A}} 1_{[y'_{a,n}(n)=r]}$$
$$\pi'_{r,n}(n) = \sum_{a \in \mathcal{A}} p_{a,n} 1_{[y'_{a,n}(n)=r]}.$$

For $r = 0, 1, 2, ..., (n-1)$, let

$$T'_{r,n}(n) = \frac{N'_{r+1,n}(n)}{n}(r+1).$$

It is readily checked that

$$\mathrm{E}[\pi'_{r,n}] = \binom{n}{r} \sum_{a \in \mathcal{A}} p_{a,n}^{r+1}(1-p_{a,n})^{n-r} \text{ and } \mathrm{E}[N'_{r,n}] = \binom{n}{r} \sum_{a \in \mathcal{A}} p_{a,n}^{r}(1-p_{a,n})^{n-r}.$$

Its bias is given by

$$\mathrm{E}\left[T'_{r,n} - \pi'_{r,n}\right] = \binom{n}{r} \sum_{a \in \mathcal{A}} p_{a,n}^{r+1}(1-p_{a,n})^{n-r-1}\left(p_{a,n} - \frac{r}{n}\right). \tag{2.7}$$

We now give our main results for asymptotic normality in the Deterministic case with a changing distribution.

Theorem 5. *Fix $r \in \{0,1,2,...\}$. Assume that $s_{n,n} \to \infty$ as $n \to \infty$ and*

$$\frac{s_{n,n}}{\sqrt{n}} \to 0.$$

In this case

$$\lim_{n\to\infty} s_{n,n}^{-2} \sum_{a\in\mathcal{A}} e^{-np_{a,n}} (np_{a,n})^{(r+2)} 1_{[np_{a,n} \geq \epsilon s_{n,n}]} = 0 \quad \forall \epsilon > 0 \tag{2.8}$$

if and only if

$$\frac{n}{s_{n,n}} (T'_{r,n}(n) - \pi'_{r,n}(n)) \xrightarrow{d} N(0,1). \tag{2.9}$$

Corollary 9. *In the Poissonized case*

$$s_{n,n}^2 = (r+1)^2 \operatorname{E}[N_{r+1,n}] + (r+2)(r+1) \operatorname{E}[N_{r+2,n}],$$

and in the deterministic case

$$(s'_{n,n})^2 = (r+1)^2 \operatorname{E}\left[N'_{r+1,n}\right] + (r+2)(r+1) \operatorname{E}\left[N'_{r+2,n}\right].$$

If the conditions in Theorem 5 hold, then $(s'_{n,n})^2 \sim s_{n,n}^2$, i.e.

$$\frac{(s'_{n,n})^2}{s_{n,n}^2} \xrightarrow{p} 1.$$

Corollary 10. *If the conditions in Theorem 5 hold, then (2.8) holds if and only if*

$$\frac{n\left(T'_{r,n}(n) - \pi'_{r,n}(n)\right)}{s'_{n,n}} \xrightarrow{d} N(0,1). \tag{2.10}$$

Corollary 11. *If the conditions in Theorem 5 hold, then* (2.10) *holds if and only if*

$$P\left(\left|\frac{T'_{r,n}(n)}{\pi'_{r,n}(n)} - 1\right| > \epsilon\right) \to 0 \quad \forall \epsilon > 0,$$

(i.e., $\frac{T'_{r,n}(n)}{\pi'_{r,n}(n)} \xrightarrow{p} 1$*).*

Corollary 12. *For the deterministic case let*

$$(s'_{n,n})^2 = (r+1)^2\mathrm{E}[N'_{r+1,n}] + (r+2)(r+1)\mathrm{E}[N'_{r+2,n}]$$
$$(\hat{s}'_{n,n})^2 = (r+1)^2 N'_{r+1,n} + (r+2)(r+1)N'_{r+2,n}.$$

If the conditions in Theorem 5 hold, then $(\hat{s}'_{n,n})^2$ *is a consistent estimator of* $(s'_{n,n})^2$*, i.e., as* $n \to \infty$*, for all* $\epsilon > 0$

$$P\left(\left|\frac{(\hat{s}'_{n,n})^2}{(s'_{n,n})^2} - 1\right| > \epsilon\right) \to 0.$$

In practical applications it is most useful to take $\hat{s}'_{r,n}$ in (2.9) as this can be done without any knowledge of $\mathcal{P}_n$. This leads to the following asymptotic confidence interval

$$\left(T'_{r,n} - z_{\alpha/2}\frac{\hat{s}'_{r,n}}{n}, T'_{r,n} + z_{\alpha/2}\frac{\hat{s}'_{r,n}}{n}\right), \tag{2.11}$$

where $z_{\alpha/2}$ is a number such that $P(Z > z_{\alpha/2}) = \alpha/2$ with $Z \sim N(0,1)$. We conduct simulation studies in Chapter 3 mainly based on this result.

Theorem 6. *Fix* $r \in \{0,1,2,...\}$*. Assume that* $s_{r,n} \to c \in (0,\infty)$ *and set* $c^* = c^2/(r+1)^2$*. If* (2.8) *holds, then* $\mathrm{E}[N'_{r+1,n}] \to c^*$, $\mathrm{E}[N'_{r+2,n}] \to 0$,

$$\mathrm{E}\left(\frac{n-r}{r+1}\pi'_{r,n} - c^*\right)^2 \to 0,\ \frac{n}{r+1}\pi'_{r,n} \xrightarrow{p} c^*,\ \textit{and}\ \frac{n}{r+1}T'_{r,n} \xrightarrow{d} \mathrm{Pois}(c^*).$$

For $r = 0$ this is Theorem 2 in [23]. See also [27] for related results in this case. Note that the assumptions of Theorem 6 never hold for fixed distributions. This is because, for such distributions, $s_{r,n} \to c \in (0, \infty)$ implies that (2.8) does not hold.

2.5 Example Distributions

In this section we give two examples to show how conditions of our main theorems can be satisfied when the distribution is both fixed and changing, respectively.

2.5.1 Fixed Discrete Pareto Distributions

Consider that $f(x) = \frac{\beta}{(x+1)^{\alpha+1}}$ where $\beta > 0$, $\alpha > 0$ and $x > 0$. Let $p_k = f(k)$, where $k = 0, 1, 2, ...$

First we show that $s_\lambda \to \infty$.

Proof. Note that $s_\lambda^2 = (r+1)^2 \mathrm{E}[N_{r+1}] + (r+2)(r+1)\mathrm{E}[N_{r+2}]$, so the result can be shown if $\mathrm{E}[N_{r+1}]$ or $\mathrm{E}[N_{r+2}]$ goes to ∞.

Since

$$\begin{aligned}\mathrm{E}[N_{r+1}] &= \sum_{a \in \mathcal{A}} e^{-\lambda p_a} \frac{(\lambda p_a)^{(r+1)}}{(r+1)!} \\ &= \sum_{k=0}^{\infty} e^{-\lambda p_k} \frac{(\lambda p_k)^{(r+1)}}{(r+1)!}.\end{aligned}$$

Let $g_\lambda(x) = e^{-x} x^{r+1}$ for $x > 0$. Since $g'_\lambda(x) = x^r e^{-(r+1)}(r+1-x)$, it follows that

$$\max_{x>0} g_\lambda(x) = g_\lambda(r+1),$$

and $g_\lambda(x)$ is increasing on $(0, r+1]$ and decreasing on $(r+1, \infty)$. Then the summands in $E[N_{r+1}]$ can be expressed by $\frac{1}{(r+1)!} g_\lambda(\lambda f(k))$. Since

$$(g_\lambda(\lambda f(x)))' = \\ \left(\frac{\lambda\beta}{(x+1)^{\alpha+1}}\right)^r exp\left(-\frac{\lambda\beta}{(x+1)^{\alpha+1}}\right)\left(r+1-\frac{\lambda\beta}{(x+1)^{\alpha+1}}\right)(-\alpha-1)\frac{\lambda\beta}{(x+1)^{\alpha+1}},$$

and

$$\max_{x>0} g_\lambda(\lambda f(x)) = g_\lambda(\lambda f(x^*)),$$

where $x^* = \left(\frac{\lambda\beta}{r+1}\right)^{\frac{1}{\alpha+1}} - 1$ for large enough λ. Thus for $x > 0$, $g_\lambda(\lambda f(x))$ is increasing on $(0, x^*)$ and decreasing on $[x^*, \infty)$.

Therefore, by a version of Euler-Maclaurin Lemma, see Lemma 1.6 in [26],

$$\begin{aligned}
&\lim_{\lambda\to\infty} \sum_{k=0}^{\infty} e^{-\lambda p_k} \frac{(\lambda p_k)^{(r+1)}}{(r+1)!} \\
&= \lim_{\lambda\to\infty} \int_0^\infty e^{-\lambda f(x)} \frac{(\lambda f(x))^{(r+1)}}{(r+1)!}\, dx \\
&= \lim_{\lambda\to\infty} \int_0^\infty exp\left(-\frac{\lambda\beta}{(x+1)^{\alpha+1}}\right)\left(\frac{\lambda\beta}{(x+1)^{\alpha+1}}\right)^{r+1} \frac{\lambda^{(r+1)}}{(r+1)!}.
\end{aligned}$$

Changing variable $t = \lambda f(x) = \frac{\lambda\beta}{(x+1)^{\alpha+1}}$ gives

$$\begin{aligned}
\int_0^\infty e^{-\lambda f(x)} \frac{(\lambda f(x))^{(r+1)}}{(r+1)!}\, dx &= \frac{1}{(r+1)!} \int_0^\infty e^{-\lambda f(x)} (\lambda f(x))^{(r+1)}\, dx \\
&= -\frac{(\lambda\beta)^{\frac{1}{\alpha+1}}}{(r+1)!} \int_{\lambda\beta}^0 e^{-t} t^{(r+1)}\, d(t^{-\frac{1}{\alpha+1}}) \\
&= \frac{(\lambda\beta)^{\frac{1}{\alpha+1}}}{(r+1)!} \int_0^{\lambda\beta} e^{-t} t^{(r+1)}\, d(t^{-\frac{1}{\alpha+1}}) \\
&= \frac{(\lambda\beta)^{\frac{1}{\alpha+1}}}{(r+1)!(\alpha+1)} \int_0^{\lambda\beta} e^{-t} t^{(r-\frac{1}{\alpha+1})}\, dt.
\end{aligned}$$

Since $\alpha > 0, r \geq 0$, we have $r > \frac{1}{\alpha+1} - 1$, i.e. $r - \frac{1}{\alpha+1} + 1 > 0$. It follows that

$$\lim_{\lambda\to\infty} \int_0^{\lambda\beta} e^{-t} t^{(r-\frac{1}{\alpha+1})}\, dt = \int_0^\infty e^{-t} t^{(r-\frac{1}{\alpha+1})}\, dt = \Gamma(r - \frac{1}{\alpha+1} + 1),$$

which is well defined. Hence,

$$\lim_{\lambda\to\infty} \int_0^\infty e^{-\lambda f(x)} \frac{(\lambda f(x))^{(r+1)}}{(r+1)!}\, dx = \lim_{\lambda\to\infty} \frac{(\lambda\beta)^{\frac{1}{\alpha+1}}}{(r+1)!(\alpha+1)} \int_0^{\lambda\beta} e^{-t} t^{(r-\frac{1}{\alpha+1})}\, dt$$

$$= \lim_{\lambda\to\infty} \frac{(\lambda\beta)^{\frac{1}{\alpha+1}}}{(r+1)!(\alpha+1)} \Gamma(r - \frac{1}{\alpha+1} + 1),$$

and

$$\lim_{\lambda\to\infty} \mathrm{E}[N_{r+1}] \propto \lim_{\lambda\to\infty} \lambda^{\frac{1}{\alpha+1}} = \infty. \tag{2.12}$$

Similarly,

$$\begin{aligned} \mathrm{E}[N_{r+2}] &= \sum_{a\in\mathcal{A}} e^{-\lambda p_a} \frac{(\lambda p_a)^{(r+2)}}{(r+2)!} \\ &= \sum_{k=0}^{\infty} e^{-\lambda p_k} \frac{(\lambda p_k)^{(r+2)}}{(r+2)!}, \end{aligned}$$

and since $r - \frac{1}{\alpha+1} + 2 > 0$,

$$\begin{aligned} &\lim_{\lambda\to\infty} \sum_{k=0}^{\infty} e^{-\lambda p_k} \frac{(\lambda p_k)^{(r+2)}}{(r+2)!} \\ &= \lim_{\lambda\to\infty} \int_0^{\infty} e^{-\lambda f(x)} \frac{(\lambda f(x))^{(r+2)}}{(r+2)!} \, dx \\ &= \lim_{\lambda\to\infty} \frac{(\lambda\beta)^{\frac{1}{\alpha+1}}}{(r+2)!(\alpha+1)} \Gamma(r - \frac{1}{\alpha+1} + 2), \end{aligned}$$

and thus,

$$\lim_{\lambda\to\infty} \mathrm{E}[N_{r+2}] \propto \lim_{\lambda\to\infty} \lambda^{\frac{1}{\alpha+1}} = \infty. \tag{2.13}$$

Therefore,

$$s_\lambda^2 \xrightarrow{\lambda\to\infty} \infty,$$

and

$$s_\lambda \to \infty.$$

□

Then we show that $\frac{s_\lambda}{\sqrt{\lambda}} \to 0$.

Proof. Since

$$\frac{s_\lambda^2}{\lambda} = \frac{(r+1)^2 \mathrm{E}[N_{r+1}]}{\lambda} + \frac{(r+2)(r+1)\mathrm{E}[N_{r+2}]}{\lambda},$$

it follows from (2.12) and (2.13) that

$$\begin{aligned} \lim_{\lambda\to\infty} \frac{s_\lambda^2}{\lambda} &= \lim_{\lambda\to\infty} \frac{(r+1)^2 \mathrm{E}[N_{r+1}]}{\lambda} + \lim_{\lambda\to\infty} \frac{(r+2)(r+1)\mathrm{E}[N_{r+2}]}{\lambda} \\ &\propto \lim_{\lambda\to\infty} \lambda^{\frac{1}{\alpha+1}-1} + \lim_{\lambda\to\infty} \lambda^{\frac{1}{\alpha+1}-1}, \end{aligned}$$

where $\frac{1}{\alpha+1} - 1 < 0$, so that

$$\lim_{\lambda\to\infty} \frac{s_\lambda^2}{\lambda} = 0,$$

and therefore

$$\frac{s_\lambda}{\sqrt{\lambda}} \to 0.$$

□

By (2.12) and (2.13) we can let

$$s_n = c\sqrt{(r+1)^2 n^{\frac{1}{\alpha+1}} + (r+2)(r+1)n^{\frac{1}{\alpha+1}}},$$

where c is a constant. Then

$$s_n \geq c\sqrt{(r+1)^2 n^{\frac{1}{\alpha+1}}},$$

and

$$\frac{s_n}{\ln n} \geq c\frac{\sqrt{(r+1)^2 n^{\frac{1}{\alpha+1}}}}{\ln n},$$

where by L'Hospital's rule and $\frac{1}{2(\alpha+1)} > 0$

$$\lim_{n\to\infty} c\frac{\sqrt{(r+1)^2 n^{\frac{1}{\alpha+1}}}}{\ln n} = \lim_{n\to\infty} c(r+1)n^{\frac{1}{2(\alpha+1)}} = \infty,$$

thus,

$$\frac{s_n}{\ln n} \to \infty.$$

Last, we show that when $\lambda = n$, if $s_n/\ln n \to \infty$, then

$$\lim_{n\to\infty} s_n^{-2} \sum_{a\in\mathcal{A}} e^{-np_a}(np_a)^{(r+2)} 1_{[np_a \geq \epsilon s_n]} = 0 \quad \forall \epsilon > 0.$$

Proof.

$$\begin{aligned}
&\sum_{a\in\mathcal{A}} e^{-np_a}(np_a)^{(r+2)} 1_{[np_a \geq \epsilon s_n]} \\
=&\sum_{k=0}^{\infty} e^{-np_k}(np_k)^{(r+2)} 1_{[np_k \geq \epsilon s_n]} \\
=&\sum_{k=0}^{\infty} e^{-np_k}(np_k)^{(r+2)} 1_{[np_k \geq M]} \qquad (M = \epsilon s_n) \\
\leq&\sum_{k=0}^{\infty} e^{-np_k}(np_k)^{(r+2)} \sum_{j=0}^{\infty} 1_{[2^j M \leq np_k < 2^{j+1} M]}
\end{aligned}$$

$$
\begin{aligned}
&\leq \sum_{j=0}^{\infty} e^{-2^j M} (2^{j+1} M)^{(r+1)} \sum_{k=0}^{\infty} np_k 1_{[2^j M \leq np_k < 2^{j+1} M]} \\
&\leq \sum_{j=0}^{\infty} e^{-2^j M} (2^{j+1} M)^{(r+1)} n \\
&= nM^{r+1} \sum_{j=0}^{\infty} e^{-2^j M} (2^{j+1})^{(r+1)} \\
&= nM^{r+1} \sum_{j=0}^{\infty} e^{-2^j M + M - M} (2^{j+1})^{(r+1)} \\
&= nM^{r+1} e^{-M} \sum_{j=0}^{\infty} e^{-(2^j - 1)M} (2^{j+1})^{(r+1)} \\
&\leq nM^{r+1} e^{-M} \sum_{j=0}^{\infty} e^{-(2^j - 1)} (2^{j+1})^{(r+1)} \\
&= nM^{r+1} e^{-M} \sum_{j=0}^{\infty} e 2^{r+1} e^{-2^j} (2^j)^{(r+1)} \\
&\leq nM^{r+1} e^{-M} \sum_{j=0}^{\infty} e 2^{r+1} e^{-(r+1)} (r+1)^{(r+1)},
\end{aligned}
$$

where the last equality holds because for $x > 0$, $e^{-x} x^{r+1}$ takes the maximal at $x = r + 1$. Then, let $C = \sum_{j=0}^{\infty} e 2^{r+1} e^{-(r+1)} (r+1)^{(r+1)}$,

$$
\begin{aligned}
& s_n^{-2} \sum_{a \in \mathcal{A}} e^{-np_a} (np_a)^{(r+2)} 1_{[np_a \geq \epsilon s_n]} \\
\leq & s_n^{-2} n e^{-M} M^{(r+1)} C \\
= & C \epsilon^2 n e^{-M/2} e^{-M/2} M^{(r-1)},
\end{aligned}
$$

where

$$
ne^{-M/2} = ne^{-\epsilon s_n/2} = e^{\ln n - \epsilon s_n/2} = e^{\ln n (1 - \frac{\epsilon s_n}{2 \ln n})} \to 0,
$$

if $\frac{s_n}{\ln n} \to \infty$; and

$$e^{-M/2}M^{(r-1)} \to 0,$$

because $M = \epsilon s_n \to \infty$.

Therefore,

$$\lim_{n\to\infty} s_n^{-2} \sum_{a\in\mathcal{A}} e^{-np_a}(np_a)^{(r+2)} 1_{[np_a \geq \epsilon s_n]} = 0 \quad \forall \epsilon > 0.$$

□

2.5.2 Changing Geometric Distributions

Now we consider a sequence of positive real numbers a_n such that $a_n \to \infty$ and $a_n/n \to 0$. Let $f_n(x) = (e^{1/a_n} - 1)e^{-x/a_n}$ for $x > 0$ and $p_{k,n} = f_n(k) = (e^{1/a_n} - 1)e^{-k/a_n}$ where $k = 1, 2....$

First we show that $s_{r,n} \to \infty$.

Proof. Note that $s_{r,n} = \sqrt{(r+1)^2\mathrm{E}[N_{r+1,n}] + (r+2)(r+1)\mathrm{E}[N_{r+2,n}]}$, so the result can be shown if $E[N_{r+1,n}]$ or $E[N_{r+2,n}]$ goes to ∞. Since

$$\begin{aligned}
\mathrm{E}[N_{r+1,n}] &= \sum_{a\in\mathcal{A}} e^{-np_{a,n}} \frac{(np_{a,n})^{(r+1)}}{(r+1)!} \\
&= \sum_{k=1}^{\infty} e^{-np_{k,n}} \frac{(np_{k,n})^{(r+1)}}{(r+1)!}.
\end{aligned}$$

Let $g(x) = e^{-x}x^{r+1}$ for $x > 1$. Since $g'(x) = x^r e^{-(r+1)}(r + 1 - x)$, it follows that $\max_{x>1} g(x) = g(r+1)$, and $g(x)$ is increasing on $(1, r+1]$ and decreasing on $(r+1, \infty)$. Then the summands in $E[N_{r+1,n}]$ can be expressed by $\frac{1}{(r+1)!}g(nf_n(k))$.

Since $f_n(x)$ is monotone decreasing, $\max_{x>1} g(nf_n(x)) = g(nf_n(x_n^*))$, where $x_n^* = -a_n[\ln(r + 1) - n\ln(e^{1/a_n} - 1)]$ by solving $r + 1 = n(e^{1/a_n} - 1)e^{-x/a_n} = nf_n(x)$. And

$x_n^* \to \infty$. Then

$$
\begin{aligned}
\lim_{n\to\infty} \mathrm{E}[N_{r+1,n}] &= \lim_{n\to\infty} \sum_{k=1}^{\infty} e^{-np_{k,n}} \frac{(np_{k,n})^{(r+1)}}{(r+1)!} \\
&\geq \lim_{n\to\infty} \sum_{k=1}^{k_n^*-1} e^{-np_{k,n}} \frac{(np_{k,n})^{(r+1)}}{(r+1)!} \quad \text{with } k_n^* = \lfloor x_n^* \rfloor \\
&\geq \lim_{n\to\infty} \int_1^{k_n^*} e^{-nf_n(x)} \frac{(nf_n(x))^{(r+1)}}{(r+1)!}\, dx \\
&= \lim_{n\to\infty} \frac{n^{r+1}}{(r+1)!} \int_1^{k_n^*} e^{-nf_n(x)} (f_n(x))^{r+1}\, dx \\
&= \lim_{n\to\infty} \frac{a_n}{(r+1)!} \int_{n(e^{1/a_n}-1)e^{-k_n^*/a_n}}^{n(e^{1/a_n}-1)e^{-1/a_n}} t^r e^{-t}\, dt \quad \text{with } t = nf_n(x) \\
&= \lim_{n\to\infty} \frac{a_n}{(r+1)!} \lim_{n\to\infty} \int_{nf_n(k_n^*)}^{n(e^{1/a_n}-1)e^{-1/a_n}} t^r e^{-t}\, dt \\
&> \lim_{n\to\infty} \frac{a_n}{(r+1)!} \int_{e^{1/a_n}(r+1)}^{n(e^{1/a_n}-1)e^{-1/a_n}} t^r e^{-t}\, dt \qquad (2.14) \\
&= \lim_{n\to\infty} \frac{a_n}{(r+1)!} \lim_{n\to\infty} \int_{e^{1/a_n}(r+1)}^{n(e^{1/a_n}-1)e^{-1/a_n}} t^r e^{-t}\, dt \\
&= \lim_{n\to\infty} \frac{a_n}{(r+1)!} \int_{r+1}^{\infty} t^r e^{-t}\, dt \\
&= \lim_{n\to\infty} \frac{a_n}{(r+1)!} \Gamma(r+1, r+1) \\
&= \lim_{n\to\infty} a_n c_1 \qquad (2.15) \\
&= \infty,
\end{aligned}
$$

where (2.14) holds because $n(e^{1/a_n} - 1)e^{-1/a_n} = n(1 - e^{-1/a_n}) \to \infty$ with $n/a_n \to 0$; and with $nf_n(x_n^*) = r + 1$ and $0 \leq R_n < 1$,

$$
\begin{aligned}
nf_n(x_n^*) \leq nf_n(k_n^*) &= nf_n(x_n^* - R_n) \\
&= n(e^{1/a_n} - 1)(e^{-x_n^*/a_n} e^{R_n/a_n}) \\
&= e^{R_n/a_n}(nf_n(x_n^*)) \\
&< e^{1/a_n}(nf_n(x_n^*))
\end{aligned}
$$

$$= e^{1/a_n}(r+1).$$

Similarly,

$$\lim_{n\to\infty} \mathrm{E}[N_{r+2,n}] \geq \lim_{n\to\infty} \frac{a_n}{(r+2)!}\Gamma(r+2, r+2) = \lim_{n\to\infty} a_n c_2 = \infty. \tag{2.16}$$

Therefore, $s_{r,n}^2 \to \infty$ and $s_{r,n} \to \infty$. □

Then we show that $s_{r,n}/\sqrt{n} \to 0$.

Proof. Since $\lim_{n\to\infty} \frac{s_{r,n}^2}{n} = \lim_{n\to\infty} \frac{(r+1)^2 \mathrm{E}[N_{r+1,n}]}{n} + \lim_{n\to\infty} \frac{(r+2)(r+1)\mathrm{E}[N_{r+2,n}]}{n}$, we can show each limit piece goes to 0. First, let $h(k) = e^{-np_{k,n}} \frac{\left(np_{k,n}\right)^{(r+1)}}{(r+1)!}$, then

$$\begin{aligned}
\lim_{n\to\infty} \frac{\mathrm{E}[N_{r+1,n}]}{n} &= \lim_{n\to\infty} \frac{\sum_{k=1}^{\infty} e^{-np_{k,n}} \frac{\left(np_{k,n}\right)^{(r+1)}}{(r+1)!}}{n} \\
&= \lim_{n\to\infty} \frac{\sum_{k=1}^{\infty} h(k)}{n} \\
&\leq \lim_{n\to\infty} \frac{\sum_{k=1}^{\lfloor x_n^* - 1\rfloor} h(k) + \sum_{k=\lceil x_n^*+1\rceil}^{\infty} h(k) + h(\lfloor x_n^*\rfloor) + h(\lceil x_n^*\rceil)}{n} \\
&\leq \lim_{n\to\infty} \frac{\int_1^{\lfloor x_n^*\rfloor} h(x)\,dx + \int_{\lceil x_n^*\rceil}^{\infty} h(x)\,dx + 2h(x_n^*)}{n} \\
&\leq \lim_{n\to\infty} \frac{\int_1^{\infty} h(x)\,dx + 2h(x_n^*)}{n} \\
&= \lim_{n\to\infty} \frac{\int_1^{\infty} e^{-nf_n(x)} \frac{(nf_n(x))^{(r+1)}}{(r+1)!}\,dx}{n} + \lim_{n\to\infty} \frac{2e^{-nf_n(x_n^*)} \frac{\left(nf_n(x_n^*)\right)^{(r+1)}}{(r+1)!}}{n} \\
&= \lim_{n\to\infty} \frac{a_n}{n(r+1)!} \int_0^{n(e^{1/a_n}-1)e^{-1/a_n}} t^r e^{-t}\,dt \\
&\quad + \lim_{n\to\infty} \frac{(r+1)^{(r+1)}}{(r+1)!} \frac{2e^{-n(r+1)}}{n} \text{ with } t = nf_n(x) \text{ and } nf_n(x_n^*) = r+1 \\
&=0.
\end{aligned}$$

Now we show how the last line holds. By the assumption that $a_n/n \to 0$,

$$\begin{aligned}
&\lim_{n\to\infty} \frac{a_n}{n(r+1)!} \int_0^{n(e^{1/a_n}-1)e^{-1/a_n}} t^r e^{-t}\,dt \\
=&\frac{1}{(r+1)!} \lim_{n\to\infty} \frac{a_n}{n} \lim_{n\to\infty} \int_0^{n(e^{1/a_n}-1)e^{-1/a_n}} t^r e^{-t}\,dt \\
=&\frac{1}{(r+1)!} \lim_{n\to\infty} \frac{a_n}{n} \int_0^{\infty} t^r e^{-t}\,dt \\
=&\frac{\Gamma(r+1)}{(r+1)!} \lim_{n\to\infty} \frac{a_n}{n} \\
=&0,
\end{aligned}$$

and

$$\lim_{n\to\infty} \frac{(r+1)^(r+1)}{(r+1)!} \frac{2e^{-n(r+1)}}{n} = 0.$$

Similarly, we can obtain that $\lim_{n\to\infty} \frac{E[N_{r+2,n}]}{n} = 0$.

Therefore, $s_{r,n}/\sqrt{n} \to 0$. □

Last we show that

$$\lim_{n\to\infty} s_{r,n}^{-2} \sum_{k=1}^{\infty} e^{-np_{k,n}} (np_{k,n})^{(r+2)} 1_{[np_{k,n} \geq \epsilon s_{r,n}]} = 0 \quad \forall \epsilon > 0, \tag{2.17}$$

if and only if the sequence a_n satisfies the following conditions: $0 < a_n < n/(r+1)$, $a_n \to \infty$ and $a_n/n \to 0$.

Proof. For all $\epsilon > 0$ and $x > 1$ let $h(x) = e^{-x} x^{r+2} 1_{[x \geq \epsilon s_{r,n}]}$. Since $\max_{x>1} h(x) = h(r+2)$, and $h(x)$ is increasing on $(1, r+2]$ and decreasing on $(r+2, \infty)$. Then the summands in (2.17) can be expressed by $h(nf_n(k))$.

Since $f_n(x)$ is monotone decreasing, $\max_{x>1} h(nf_n(x)) = h(nf_n(x'_n))$, where $x'_n = -a_n \ln((r+2)a_n/n)$ with $0 < a_n < n/(r+2)$ by solving $r+2 = n(a_n^{-1} e^{-x/a_n}) =$

$nf_n(x)$. As $\lim_{n\to\infty} nf_n(x) = \lim_{n\to\infty} \frac{n}{a_n} e^{-x/a_n} = \infty$ for fixed x, we have

$$\lim_{n\to\infty} h(nf_n(x)) = \lim_{n\to\infty} e^{-nf_n(x)}(nf_n(x))^{(r+2)} 1_{[nf_n(x)\geq \epsilon s_{r,n}]} = 0.$$

Meanwhile, with $nf_n(x'_n) = r+2$ and $s_{r,n} \to \infty$, $\lim_{n\to\infty} 1_{[r+2\geq \epsilon s_{r,n}]} = 0$, thus,

$$\lim_{n\to\infty} h(nf_n(x'_n)) = \lim_{n\to\infty} e^{-(r+2)}(r+2)^{(r+2)} 1_{[r+2\geq \epsilon s_{r,n}]} = 0.$$

Then by Euler-Maclaurin lemma, see Lemma 1.6 in [26], $\forall \epsilon > 0$

$$\begin{aligned} &\lim_{n\to\infty} s_{r,n}^{-2} \sum_{k=1}^{\infty} e^{-np_{k,n}}(np_{k,n})^{(r+2)} 1_{[np_{k,n}\geq \epsilon s_{r,n}]} \\ &= \lim_{n\to\infty} s_{r,n}^{-2} \int_1^{\infty} e^{-nf_n(x)} \left(nf_n(x)\right)^{(r+1)} 1_{[nf_n(x)\geq \epsilon s_{r,n}]} \, dx \\ &= \lim_{n\to\infty} a_n s_{r,n}^{-2} \int_0^{n(e^{1/a_n}-1)e^{-1/a_n}} e^{-t} t^{(r+1)} 1_{[t\geq \epsilon s_{r,n}]} \, dt \quad \text{with } t = nf_n(x). \end{aligned} \tag{2.18}$$

Here we need to consider two cases as follows. If $n(e^{1/a_n}-1)e^{-1/a_n} < \epsilon s_{r,n}$, (2.18) =0 with $1_{[t\geq \epsilon s_{r,n}]} = 0$. If $n(e^{1/a_n}-1)e^{-1/a_n} \geq \epsilon s_{r,n}$, (2.18) is equivalent to

$$\lim_{n\to\infty} \frac{a_n}{s_{r,n}^2} \int_{\epsilon s_{r,n}}^{n(e^{1/a_n}-1)e^{-1/a_n}} e^{-t} t^{(r+1)} \, dt.$$

As in (2.15) and (2.16) we have

$$\begin{aligned} \lim_{n\to\infty} \frac{s_{r,n}^2}{a_n} &= \lim_{n\to\infty} \frac{(r+1)^2 \mathrm{E}[N_{r+1,n}] + (r+2)(r+1)\mathrm{E}[N_{r+2,n}]}{a_n} \\ &\geq \lim_{n\to\infty} \frac{(r+1)^2 a_n c_1}{a_n} + \lim_{n\to\infty} \frac{(r+2)(r+1)a_n c_2}{a_n} \\ &= (r+1)^2 c_1 + (r+2)(r+1)c_2 = c_3, \end{aligned}$$

i.e., $\lim_{n\to\infty} \frac{a_n}{s_{r,n}^2} \leq \frac{1}{c_3}$. Now

$$\lim_{n\to\infty} \frac{a_n}{s_{r,n}^2} \int_{\epsilon s_{r,n}}^{n(e^{1/a_n}-1)e^{-1/a_n}} e^{-t} t^{(r+1)}\, dt$$
$$\leq \lim_{n\to\infty} \frac{1}{c_3} \int_{\epsilon s_{r,n}}^{\infty} e^{-t} t^{(r+1)}\, dt = 0, \tag{2.19}$$

where (2.19) holds by the following lemma. □

(Note: **Euler-Maclaurin Lemma:** Let c_n be a sequence of positive real numbers and $c_n \to \infty$. If $f(x)$ is an integrable function, then

$$\lim_{c_n\to\infty} \int_{c_n}^{\infty} f(x)\, dx = 0.$$

Proof. Since $\int_{c_n}^{\infty} f(x)\, dx = \int_{\mathbb{R}} f(x) 1_{[c_n,\infty)}(x)\, dx$, let $f_n(x) = f(x)1_{[c_n,\infty)}(x)$, where $c_n > 0$ and $c_n \to \infty$. Also we have $f_n(x) \leq |f(x)|$ for all x and $f(x)$ is integrable, so does $|f(x)|$. Now by Lebesgue's dominated convergence theorem for

$$\lim_{n\to\infty} \int_{\mathbb{R}} f_n(x)\, dx = \int_{\mathbb{R}} \lim_{n\to\infty} f_n(x)\, dx = \int_{\mathbb{R}} \lim_{n\to\infty} f(x) 1_{[c_n,\infty)}(x)\, dx = 0,$$

where the last equality holds because $1_{[\infty,\infty)}(x) = 0$ for large enough n and any fixed x.) □

CHAPTER 3: SIMULATION STUDY

In this section we perform two simulation studies to check the finite sample performance of the confidence intervals in (2.11), one with data generated from the theoretical distributions and another with the real data as the theoretical population.

3.1 Theoretical Data Simulation Methodology and Results

To better understand how asymptotic normality for Turing formulae works, we perform simulation studies under a variety of distributions and for a variety of sample sizes. Three different types of distribution are considered: the Poisson distribution, the geometric distribution and the discrete Pareto distribution. For each distribution and each choice of the parameters, we simulate samples of size n from 1 to 1000 with increments of 20. After 2000 iterations we calculate the accuracy ratio for the estimator falling inside the 95% confidence interval with results given in Figure 3.1. The accuracy ratio should be close to 0.95 if the asymptotic normality works well.

The results are shown in Figure 3.1. Plots of the accuracy ratio of the higher order Turing Formulae at $r = 0, 3, 5$ are presented. The x-axis is the sample size and the y-axis is the accuracy ratio calculated. The top line is the distribution name and the legends give values of different parameter assigned. The horizontal line at 0.95 is for comparison.

In those three distributions considered, we first consider the Poisson distribution. The probability mass function of a discrete Poisson random variable X is $P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$ for $k = 0, 1, 2, \ldots$ with parameter $\lambda > 0$. The Poisson distribution has the lightest tail among those three distributions because the moment generating function of any Poisson random variables is finite for all $t > 0$. We choose parameter

$\lambda = 1, 5, 10$. By the fact that $\lambda = \mathrm{E}[X] = \mathrm{Var}[X]$, the larger λ is, the heavier the tails are.

Next is the geometric distribution. The probability mass function of a geometrically distributed discrete random variable X is $P(X = k) = (1-p)^k p$ for $k = 0, 1, 2, \ldots$ with parameter $0 < p \leq 1$. By the fact that the moment generating function of the geometric distribution is finite for $t < -\ln(1-p)$ and infinite otherwise, the geometric distribution has intermediate exponential tails. We choose $p = 0.1, 0.25, 0.5, 0.75, 0.9$, and smaller parameter p indicates heavier tails.

Last, we consider the discrete Pareto distribution of the random variable $X = \lfloor Y \rfloor$, where Y has the Pareto probability density function $f(y) = \frac{\alpha}{y^{\alpha+1}}$ for $y > 1$ and with parameter $\alpha > 0$. The discrete Pareto distribution with finite number of finite moments has polynomial heavy tails, which is heavier than the exponential tails. We choose $\alpha = 0.5, 1.5, 2$, and smaller values of α imply heavier tails.

The plots show that the simulation performs better for discrete Pareto distributions, which suggests that the asymptotic normality seems to work better for heavy-tailed distributions.

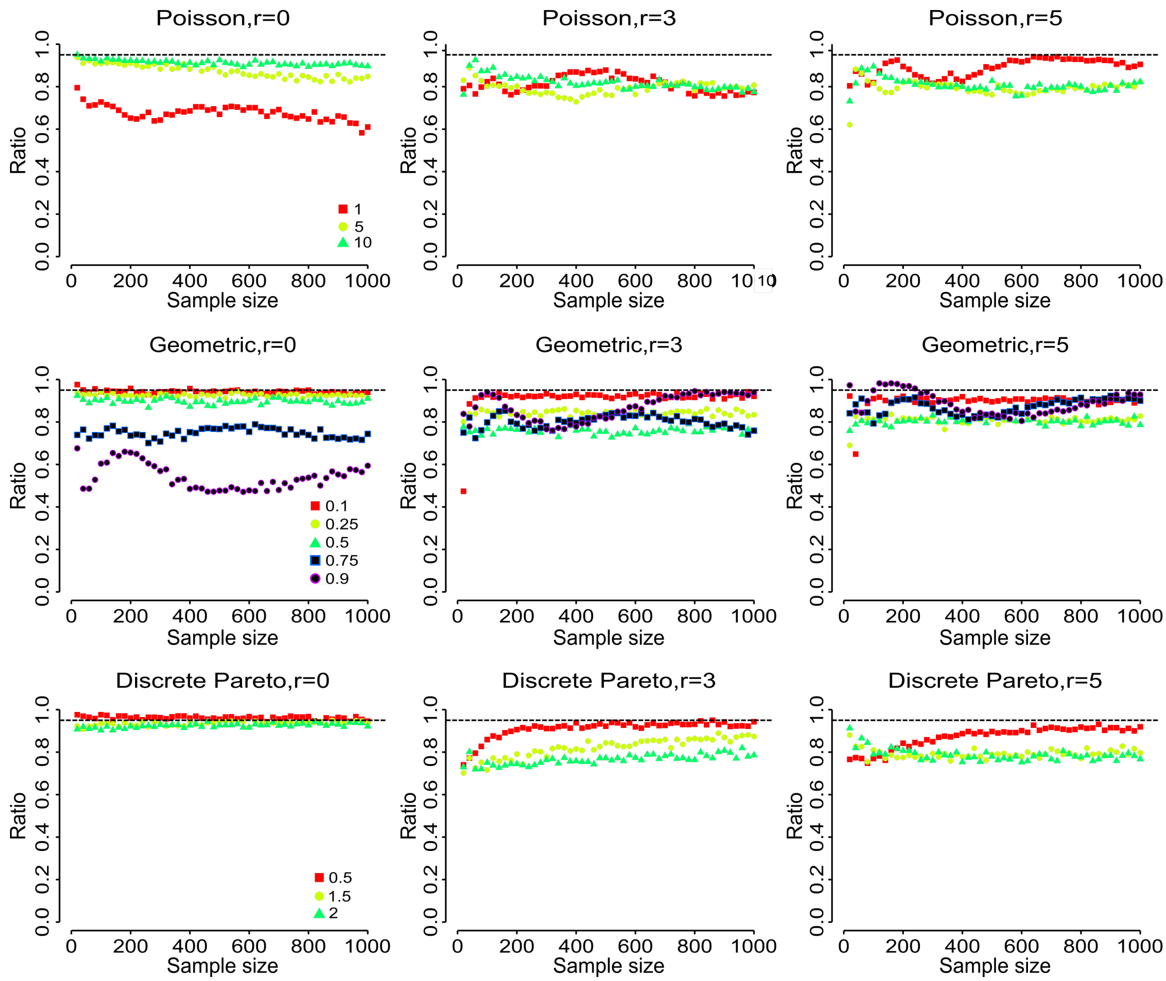

Figure 3.1: Plots of simulation study for data generated from theoretical distributions.

3.2 Literature Work Simulation Methodology and Results

The data for this simulation study is downloaded from http://shakespeare.mit.edu. We took all of the words in the complete works of William Shakespeare as our population. We include the titles of the works, since we believe that the titles also contain the word usage information for the population. In total there are 930,593 words. Ignoring repetition, it has 28,857 unique words, where we consider that words with and without contractions are two different words.

Our alphabet $\mathcal{A}$ is comprised of each of these unique words. For a word $a \in \mathcal{A}$, the probability p_a is the number of times that it appears in the population divided by the size of the population. The most frequent word is "*the*", which has a probability of 0.031812. There are 12667 words that appear only once. They have a probability of $1/930593 \approx 10^{-6}$. For a given order r and sample size n, we sampled $N = 1000$ samples of size n. All sampling was done with replacement. For each sample, we calculated the confidence interval in (2.11) at level $\alpha = 0.05$ and the true value of $\pi'_{r,n}$. We then found the proportion of the samples for which the true value is contained in the confidence interval.

Plots of these proportions for several choices of n and r are given in Figure 3.2a and Figure 3.2b. In the plots the x-axis represents the sample size, where sample size increases from 100 to 1000 with increments of 100, and sample size increases from 1000 to 3000 with increments of 250. Plot (a) shows results for $r = 0, 1$ and 2, and plot (b) shows results for $r = 4, 5$ and 6. These plots should be close to the horizontal line at 0.95. We can see that they are generally close to this value. However, for larger values of r, we typically need larger sample sizes.

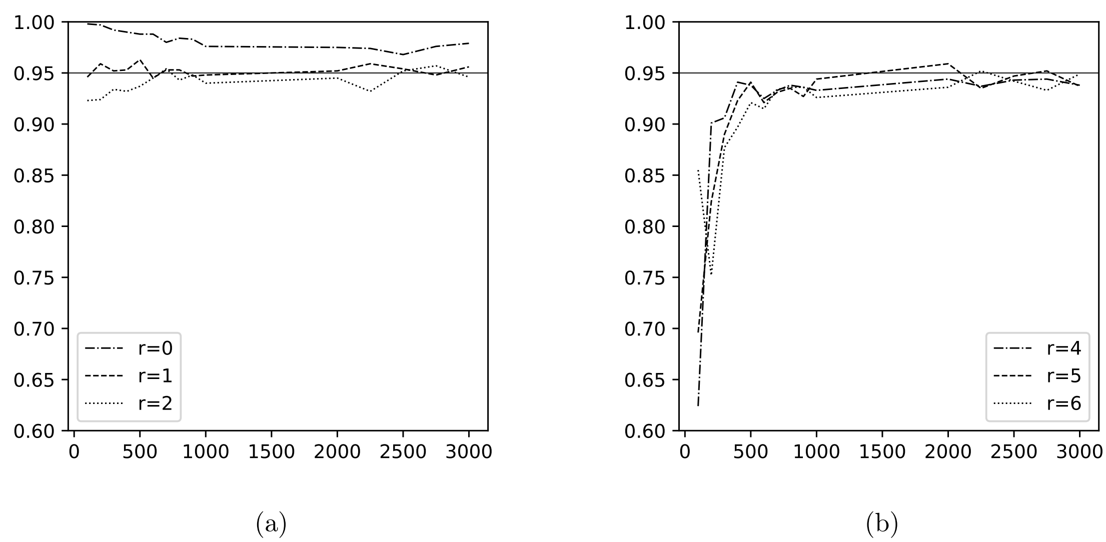

Figure 3.2: Plots of simulation study with data from the complete works of William Shakespeare.

CHAPTER 4: DATA APPLICATIONS

One of the main applications of our results is authorship attribution, i.e., whether we can detect the difference between writing samples from two different authors. We propose two methodologies based on our theoretical results, and illustrate them with tweet data from [28]. This dataset contains tweets of the top 20 popular twitter users with the most followers in 2017. We randomly select and analyze tweets from two users to show preliminary results and then analyze tweets from the top 5 users.

4.1 Data Application Methodology

In the first methodology we begin by constructing 95% asymptotic confidence intervals for $\pi'_{r,n}$ for a fixed n and different choices of r with tweet samples from two authors separately, and for all results we let $r = 1, 2, 3, ..., 7$. Then we check the overlaps of two plotted asymptotic confidence intervals: a lot of overlap suggests that the datasets are from the same author, while little overlap suggests that the datasets are from different authors.

In the second methodology we perform a statistical test to check if two tweet samples come from the same author. The first dataset is treated as the 'corpus set' to construct an asymptotic confidence interval, and the second dataset is treated as the 'testing set' to calculate detecting values, denoted by D_r, for different choices of r, where for $r = 0, 1, 2, ..., (n-1)$,

$$D_r = \frac{\text{sample count of words that are observed r times in corpus set}}{\text{sample size of testing set}},$$

and repetition in the sample count is included. When $r = 0$, the numerator in D_0 is just the number of new words that are not observed in the corpus set. Then the detecting values are compared with the asymptotic confidence interval bounds. If most of the test points fall inside the confidence interval, it suggests that the datasets are from the same author; while if most of the test points fall outside the confidence interval, it suggests that the datasets are from different authors.

4.2 Data Application Results

First, we analyze tweets from Ariana Grande and Jimmy Fallon. For both datasets we put tweets together from each author ignoring punctuation, capitalization and URLs. In total, the dataset for Ariana Grande contains 52647 words and the dataset for Jimmy Fallon contains 36365 words.

We begin by randomly dividing each dataset into two parts and comparing the asymptotic confidence intervals constructed from the two random parts from the same author. The results are shown in Figure 4.1. A and B are comparisons of the asymptotic confidence intervals constructed from two random parts of tweets of Ariana Grande and Jimmy Fallon respectively; C is a comparison of the asymptotic confidence intervals constructed from full tweet datasets of Ariana Grande and Jimmy Fallon. A and B with a lot of overlap for the asymptotic confidence intervals. Then we compare the asymptotic confidence interval from the full datasets from those two different authors in 4.1 C with only little overlap.

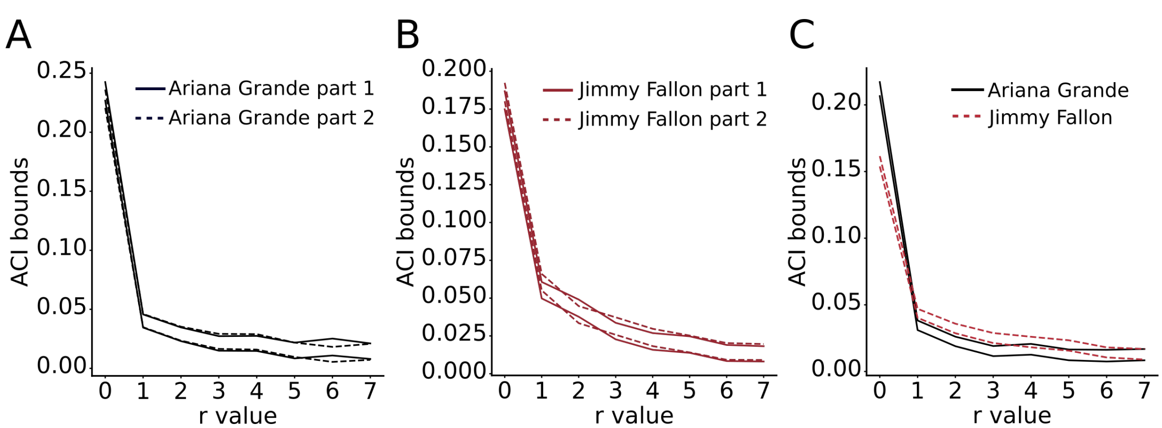

Figure 4.1: Plots of interval comparison between two twitter users

Then we use one of the random parts from each author as the corpus dataset to construct the asymptotic confidence intervals and the other as the testing dataset to draw the detecting points. Then we use the full dataset from each author as the corpus dataset to construct the asymptotic confidence intervals and the full dataset from the other author as the testing dataset to draw the detecting points. From the first and fourth plot in Figure 4.2 we can see that most of the detecting points fall inside or on the boundary of the asymptotic confidence interval, which indicates the testing dataset is from the same author; while some or most of the detecting points fall outside of the asymptotic confidence interval in the second and third plot in Figure 4.2, which indicates the testing dataset is from a different author. We also notice that at $r = 0$ all detecting points are outside of the asymptotic confidence interval, which does not give enough information to tell the author of the dataset, however, authorship can be attributed if we consider r with higher values.

We would also like to have more datasets to see how our methodology performs, so we analyze tweets from the top 5 twitter account users by then, including Katy Perry, Justin Bieber, Rihanna, Barack Obama and Taylor Swift.

First, we randomly divide dataset from each author into two parts with the same number of words. We treat one random part from one author as the corpus to construct the asymptotic confidence intervals for r valued from 0 to 7, shown as the blue solid lines in Figure 4.3 and as the black dashed lines in Figure 4.4.

Then we use the other random part from the same author to construct another asymptotic confidence intervals for different values of r, shown in Figure 4.3 as the green dashed line in the diagonal plots. And we use the full datasets from other authors to construct asymptotic confidence intervals shown in Figure 4.3 as the green dashed line in the off diagonal plots. From the plots we can see compared with the diagonal plots, a majority of the off diagonal plots have less overlaps, indicating data for the diagonal plots are from the same author and the off diagonal ones are not.

Next we use the other random part from the same author and the full datasets

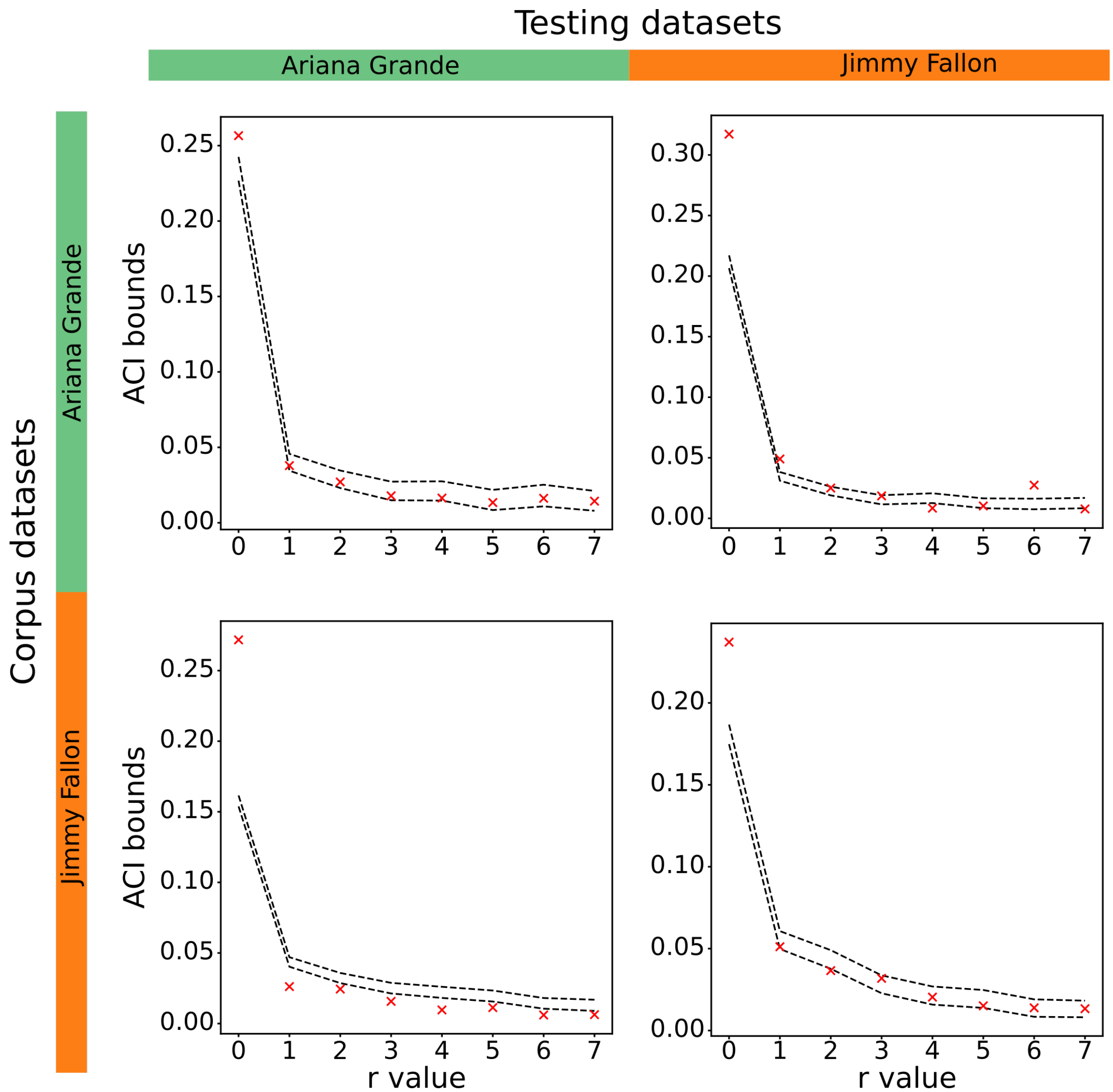

Figure 4.2: Plots of statistical test for two twitter users.

from other authors to calculate the detecting points, shown in Figure 4.4. Similarly, in the diagonal plots most of the detecting points are inside the asymptotic confidence intervals, indicating the data are from the same author; while in the off diagonal plots there are more detecting points outside the asymptotic confidence intervals, indicating the data are from different authors. Again we notice that at $r = 0$ all detecting points are not shown in the plots due to range deduction of y-axis but they are all outside of the asymptotic confidence intervals, which does not give enough information to tell the author of the dataset, however, authorship can be attributed if we consider r

with higher values.

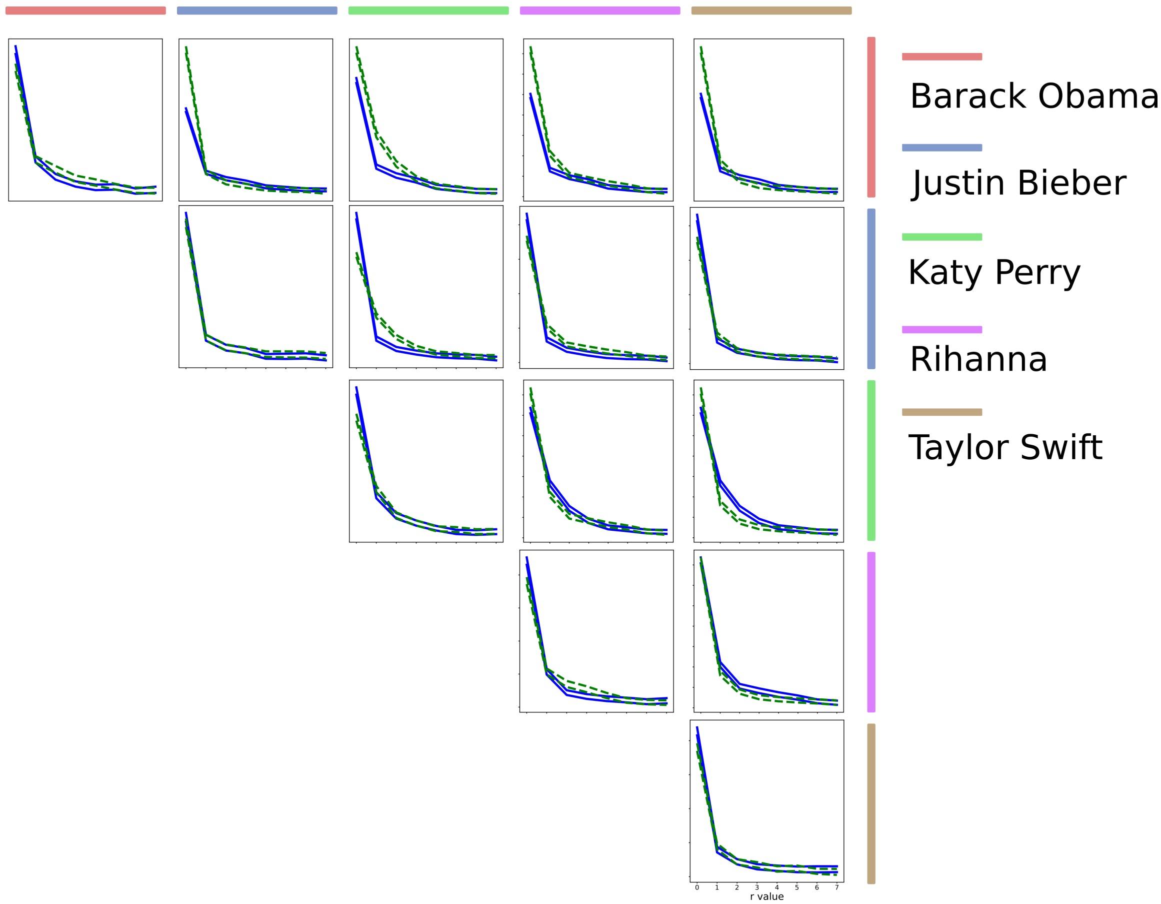

Figure 4.3: Plots of interval comparison between top 5 twitter users.

4.3 Discussion

The data application results suggest that we can distinguish when two datasets are from the same author or different authors by the area of the overlaps of the asymptotic confidence intervals and the number of detecting points falling inside of the corpus asymptotic confidence intervals. We also notice that, for all datasets we considered, the detecting points at $r = 0$ do not fall inside the corpus asymptotic confidence intervals, which can not provide enough information for authorship attribution for our data application. However, considering the detecting points at higher values of r enables us to distinguish the difference between the number of detecting points falling inside and outside the corpus asymptotic confidence intervals, so that we can conduct

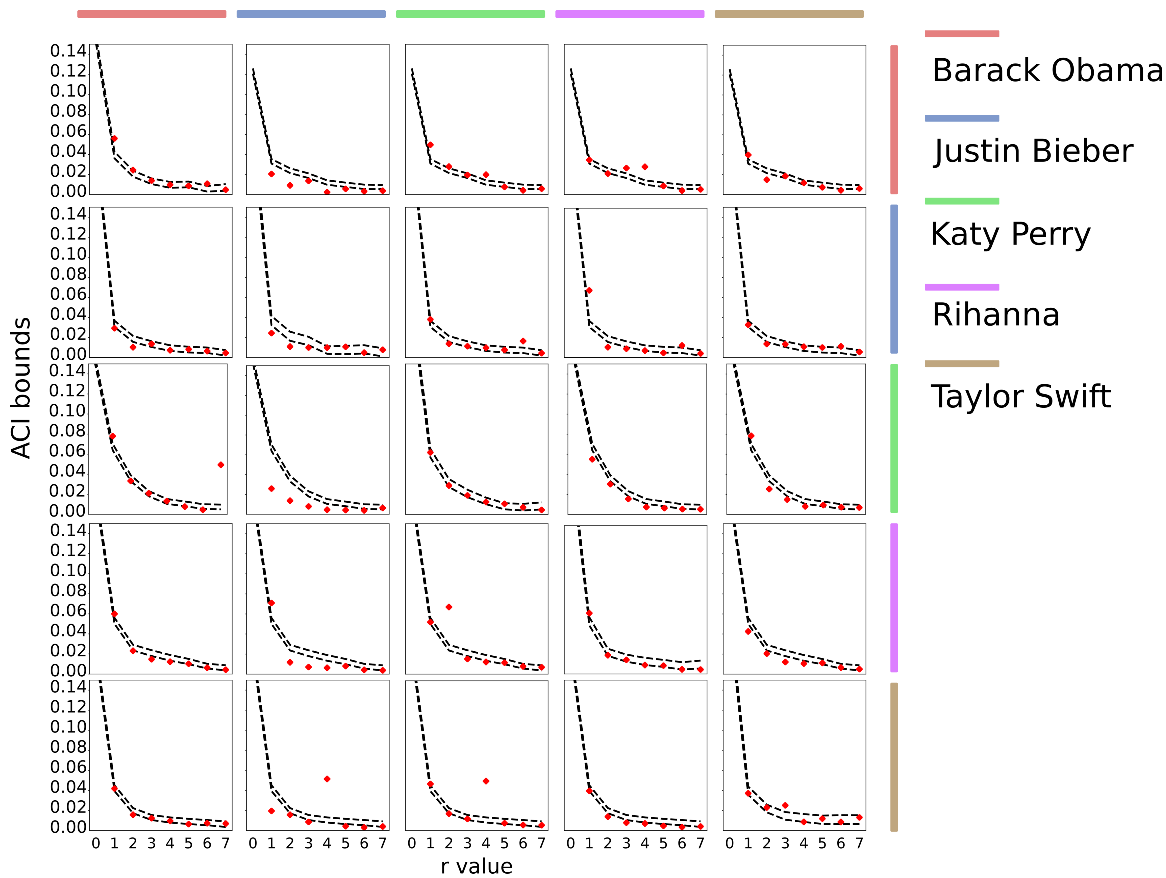

Figure 4.4: Plots of statistical test for top 5 twitter users.

authorship attribution. So far we can only have a intuitive way to explain the plots and we have not yet developed a cutoff point or a threshold for the overlapping area and the number of detecting points falling inside the asymptotic confidence intervals. We will consider this as future work.

CHAPTER 5: CONCLUSIONS

Necessary and sufficient conditions for the asymptotic normality of Turing's formulae have been given for any $r \geq 0$ and under both fixed and changing distributions. These lead to easy to calculate asymptotic confidence intervals. Our results allow for many situations that are not covered by previously available sufficient conditions. Further, in the case where $r = 0$, we correct an error in the conditions given in [23]. We have also studied the case where the sample size is random and follows a Poisson distribution. This case may be of independent interest, and is important for proving our main results. A general version of the Lindeberg-Feller central limit theorem and a number of lemmas are given in the proof. We give several explicit examples where our conditions hold. These include both cases when the underlying distributions are fixed and when they are changing. It should be noted that Turing formula for $r = 0$ is not consistent or asymptotic normal for some fixed Geometric distributions as shown in [18]. However, our conditions are proved to be applicable for the example with changing Geometric distributions in Section 2.5.2.

For the finite performance of the derived asymptotic confidence intervals, the theoretical simulation study indicates that these intervals seem to work better for heavy-tailed distributions; and the simulations creating synthetic poems based on the works of Shakespeare have shown that they can accurately capture the true value of $\pi'_{r,n}$ at least for $r \leq 6$. Larger sample sizes are needed for further study of larger values of r.

In the Data Application example comparing the authorship of different Twitter datasets, preliminary results indicate that we can distinguish whether the author/authors of two datasets are the same or different. These methods are currently based on plots. We leave the question of how to best quantify this for future work.

CHAPTER 6: PROOFS

In this chapter we give our proofs. These require several lemmas, which give interesting results about the limit theorems for infinite sums, the Turing formulae and related quantities in the alphabet scheme. These lemmas may be of independent interest.

6.1 Limit Theorems for Infinite Sums

We begin with an extension of the classical Lindeberg-Feller central limit theorem to the case of infinite triangular arrays.

Proposition 1. *Suppose that for each $n \in \mathbb{N}$, $X_{n1}, X_{n2}, X_{n3}, \ldots$ is a sequence of independent random variables each having a finite variance and satisfying*

$$\mathrm{E}\left[X_{ni}\right] = 0, \qquad \mathrm{Var}\left[X_{ni}\right] = \sigma_{ni}^2 < \infty, \qquad s_n^2 = \sum_{i=1}^{\infty} \sigma_{ni}^2 < \infty,$$

with $\liminf s_n > 0$. *We have*

$$\lim_{n\to\infty} s_n^{-2} \sum_{i=1}^{\infty} \int_{|X_{ni}|\geq \epsilon s_n} X_{ni}^2 \mathrm{d}P = 0 \qquad \forall \epsilon > 0$$

if and only if both

$$\frac{\sum_{i=1}^{\infty} X_{ni}}{s_n} \xrightarrow{d} N(0,1) \quad \text{and} \quad \sup_i \frac{\sigma_{ni}^2}{s_n^2} \xrightarrow{n\to\infty} 0.$$

Proof. Since $s_n^2 < \infty$, there exists a r_n such that $\sum_{i=r_n}^{\infty} \sigma_{ni}^2 < \frac{1}{n}$. Let $(s_n^*)^2 = \sum_{i=1}^{r_n} \sigma_{ni}^2$. Note that $(s_n^*)^2 \leq s_n^2 \leq (s_n^*)^2 + 1/n$ and thus $\liminf s_n > 0$.

By the usual Lindeberg-Feller Central Limit Theorem (see Theorem 27.2 and the discussions on page 361 in [29]),

$$\lim_{n\to\infty}(s_n^*)^{-2}\sum_{i=1}^{r_n}\int_{|X_{ni}|\geq\epsilon s_n^*}X_{ni}^2\mathrm{d}P=0\quad\forall\epsilon>0 \tag{6.1}$$

if and only if

$$\frac{\sum_{i=1}^{r_n}X_{ni}}{s_n^*}\xrightarrow{d}N(0,1) \tag{6.2}$$

and

$$\max_{i\leq r_n}\frac{\sigma_{ni}^2}{s_n^2}\xrightarrow{n\to\infty}0. \tag{6.3}$$

First we claim that (6.3) holds if and only if

$$\sup_i\frac{\sigma_{ni}^2}{s_n^2}\xrightarrow{n\to\infty}0. \tag{6.4}$$

It is clear that (6.3) follows from (6.4). Now assume that (6.3) holds. Since

$$\begin{aligned}\sup_i\frac{\sigma_{ni}^2}{s_n^2}&=\max\{\max_{i\leq r_n}\frac{\sigma_{ni}^2}{s_n^2},\sup_{i>r_n}\frac{\sigma_{ni}^2}{s_n^2}\}\\&\leq\max\{\max_{i\leq r_n}\frac{\sigma_{ni}^2}{s_n^2},\sup_{i>r_n}\frac{\frac{1}{n}}{s_n^2}\},\end{aligned}$$

and

$$\lim_{n\to\infty}\sup_{i>r_n}\frac{1}{ns_n^2}=\lim_{n\to\infty}\frac{1}{ns_n^2}=0,$$

then by the assumption that $s_n\to\infty$ as $n\to\infty$, it follows that

$$\lim_{n\to\infty}\sup_i\frac{\sigma_{ni}^2}{s_n^2}=0.$$

By Chebyshev's inequality

$$P\left(\left|\frac{\sum_{i=r_n+1}^{\infty} X_{ni}}{s_n^*}\right| > \epsilon\right) \leq \frac{\mathrm{Var}(\sum_{i=r_n+1}^{\infty} X_{ni})}{\epsilon^2 (s_n^*)^2} \leq \frac{\frac{1}{n}}{\epsilon^2 (s_n^*)^2}.$$

Since

$$s_n^2 = \sum_{i=1}^{r_n} \sigma_{ni}^2 + \sum_{i=r_n}^{\infty} \sigma_{ni}^2 \leq (s_n^*)^2 + \frac{1}{n},$$

then

$$(s_n^*)^2 \geq s_n^2 - \frac{1}{n}.$$

Since as $n \to \infty$, $s_n^2 \to \infty$ and $\frac{1}{n} \to 0$, it follows that $(s_n^*)^2 \xrightarrow{n\to\infty} \infty$, and

$$\frac{\frac{1}{n}}{\epsilon^2 (s_n^*)^2} \xrightarrow{n\to\infty} 0.$$

Thus

$$\frac{\sum_{i=r_n+1}^{\infty} X_{ni}}{s_n^*} \xrightarrow{p} 0.$$

Since

$$\frac{\sum_{i=1}^{\infty} X_{ni}}{s_n^*} = \frac{\sum_{i=1}^{r_n} X_{ni}}{s_n^*} + \frac{\sum_{i=r_n+1}^{\infty} X_{ni}}{s_n^*},$$

by Slutsky's Theorem, (6.2) holds if and only if

$$\frac{\sum_{i=1}^{\infty} X_{ni}}{s_n^*} \xrightarrow{d} N(0,1). \tag{6.5}$$

Since

$$\frac{s_n^2}{(s_n^*)^2} = \frac{(s_n^*)^2}{(s_n^*)^2} + \frac{\sum_{i=r_n+1}^{\infty} \sigma_{ni}^2}{(s_n^*)^2} \leq 1 + \frac{\frac{1}{n}}{(s_n^*)^2} \xrightarrow{n\to\infty} 1,$$

and

$$\frac{s_n^2}{(s_n^*)^2} \geq 1,$$

we have

$$\frac{s_n^2}{(s_n^*)^2} \xrightarrow{n\to\infty} 1,$$

and

$$\frac{s_n}{s_n^*} \xrightarrow{n\to\infty} 1.$$

Again by Slutsky's theorem, (6.5) is equivalent to

$$\frac{\sum_{i=1}^{\infty} X_{ni}}{s_n} \xrightarrow{d} N(0,1).$$

Similarly, (6.1) holds if and only if

$$\lim_{n\to\infty} s_n^{-2} \sum_{i=1}^{r_n} \int_{|X_{ni}|\geq \epsilon s_n^*} X_{ni}^2 \mathrm{d}P = 0 \quad \forall \epsilon > 0. \tag{6.6}$$

Next we claim that for any $\epsilon > 0$ and large enough n, $|X_{ni}| \geq \epsilon s_n^*$ if and only if there exists an ϵ' such that $|X_{ni}| \geq \epsilon' s_n$. Since $s_n^2 \geq (s_n^*)^2$, if $|X_{ni}| \geq \epsilon s_n$, then $|X_{ni}| \geq \epsilon' s_n^*$, where $\epsilon' = \epsilon$. On the other hand, if $|X_{ni}| \geq \epsilon s_n^*$, let $s_n^2 - (s_n^*)^2 = d_n$, then for a large enough n

$$|X_{ni}| \geq \epsilon\sqrt{s_n^2 - d_n} \geq \epsilon\sqrt{s_n^2 - \frac{1}{n}} \geq \epsilon\sqrt{s_n^2 - \frac{s_n^2}{2}} = \sqrt{\frac{1}{2}}\epsilon s_n.$$

The third inequality holds because we assume $s_n^2 \xrightarrow{n\to\infty} \infty$, which means that $\frac{1}{n} < \frac{s_n^2}{2}$ for a large enough n. So we can choose an $\epsilon' = \sqrt{\frac{1}{2}}\epsilon$. Hence, (6.6) is equivalent to

$$\lim_{n\to\infty} s_n^{-2} \sum_{i=1}^{r_n} \int_{|X_{ni}|\geq \epsilon s_n} X_{ni}^2 \mathrm{d}P = 0 \quad \forall \epsilon > 0. \tag{6.7}$$

Further, (6.7) is also equivalent to

$$\lim_{n\to\infty} s_n^{-2} \sum_{i=1}^{\infty} \int_{|X_{ni}|\geq \epsilon s_n} X_{ni}^2 \mathrm{d}P = 0 \quad \forall \epsilon > 0,$$

because

$$\begin{aligned}
&\lim_{n\to\infty} s_n^{-2} \sum_{i=1}^{\infty} \int_{|X_{ni}|\geq \epsilon s_n} X_{ni}^2 \mathrm{d}P \\
&= \lim_{n\to\infty} \left(s_n^{-2} \sum_{i=1}^{r_n} \int_{|X_{ni}|\geq \epsilon s_n} X_{ni}^2 \mathrm{d}P + s_n^{-2} \sum_{i=r_n+1}^{\infty} \int_{|X_{ni}|\geq \epsilon s_n} X_{ni}^2 \mathrm{d}P \right) \\
&\leq \lim_{n\to\infty} \left(s_n^{-2} \sum_{i=1}^{r_n} \int_{|X_{ni}|\geq \epsilon s_n} X_{ni}^2 \mathrm{d}P + s_n^{-2} \sum_{i=r_n+1}^{\infty} \sigma_{n_i}^2 \right) \\
&\leq \lim_{n\to\infty} \left(s_n^{-2} \sum_{i=1}^{r_n} \int_{|X_{ni}|\geq \epsilon s_n} X_{ni}^2 \mathrm{d}P + s_n^{-2} \frac{1}{n} \right) \\
&= \lim_{n\to\infty} \left(s_n^{-2} \sum_{i=1}^{r_n} \int_{|X_{ni}|\geq \epsilon s_n} X_{ni}^2 \mathrm{d}P + \lim_{n\to\infty} (s_n^{-2} \frac{1}{n}) \right) \\
&= \lim_{n\to\infty} \left(s_n^{-2} \sum_{i=1}^{r_n} \int_{|X_{ni}|\geq \epsilon s_n} X_{ni}^2 \mathrm{d}P \right) + 0 \\
&= \lim_{n\to\infty} \left(s_n^{-2} \sum_{i=1}^{r_n} \int_{|X_{ni}|\geq \epsilon s_n} X_{ni}^2 \mathrm{d}P \right).
\end{aligned}$$

This completes the proof. □

We will also need a Poisson approximation for sums of infinitely many independent Bernoulli random variables.

Proposition 2. *Suppose that for each $n \in \mathbb{N}$, $X_{n1}, X_{n2}, X_{n3}, \ldots$ is a sequence of independent random variables such that $P(X_{nk} = 1) = 1 - P(X_{nk} = 1) = p_{nk}$. If $\sup_k p_{nk} \to 0$ and $\sum_{k=1}^{\infty} p_{nk} \to \lambda \in (0, \infty)$, then*

$$S_n = \sum_{k=1}^{\infty} X_{nk} \xrightarrow{d} \mathrm{Pois}(\lambda).$$

Proof. First, note that for large enough n, the infinite sum converges almost surely by Theorem 22.6 in [29].Next, note that the moment generating function of S_n is given by

$$M_n(t) = \exp\left\{\sum_{k=1}^{\infty} \log\left(1 + (e^t - 1)p_{nk}\right)\right\}.$$

For fixed t and large enough n, $(e^t - 1)\sup_k p_{nk} < 1$, thus by the Taylor expansion of the logarithm (see e.g. 4.1.24 in [30]) and the remainder theorem for alternating series, it follows that

$$M_n(t) \leq \exp\left\{(e^t - 1)\sum_{k=1}^{\infty} p_{nk}\right\} \rightarrow \exp\{\lambda(e^t - 1)\}.$$

Similarly

$$\begin{aligned} M_n(t) &\geq \exp\left\{(e^t - 1)\sum_{k=1}^{\infty} p_{nk} - .5(e^t - 1)^2 \sum_{k=1}^{\infty} p_{nk}^2\right\} \\ &\geq \exp\left\{(e^t - 1)\sum_{k=1}^{\infty} p_{nk} - .5(e^t - 1)^2 \sup_k(p_{nk}) \sum_{k=1}^{\infty} p_{nk}\right\} \\ &\rightarrow \exp\{\lambda(e^t - 1)\}, \end{aligned}$$

and the result follows. □

6.2 Proofs for Section 2.3

6.2.1 Proofs for Section 2.3.1

Lemma 1. *Let X_n and Y_n be two sequences of random variables. If $X_n Y_n$ converges to a distribution and $X_n \xrightarrow{p} \infty$, then $Y_n \xrightarrow{p} 0$.*

Proof. By continuous mapping theorem $X_n \xrightarrow{p} \infty$ implies $\frac{1}{X_n} \xrightarrow{p} 0$.Now noting that $Y_n = \frac{1}{X_n} X_n Y_n$ and applying Slutsky's theorem gives $Y_n \xrightarrow{p} 0$. □

Lemma 2. *Assume that* $s_\lambda \to \infty$ *as* $\lambda \to \infty$*. If*

$$\frac{\lambda}{s_\lambda}\left(T_r(\lambda) - \pi_r(\lambda)\right) \xrightarrow[\lambda\to\infty]{d} N(0,1), \tag{6.8}$$

then

$$\frac{\mathrm{E}[N_{r+1}]}{s_\lambda} \to \infty.$$

Proof. Since

$$\begin{aligned}
\mathrm{E}[N_{r+2}] &= \sum_{a\in\mathcal{A}} \mathrm{E}\left[1_{[y_a(\lambda)=r+2]}\right] \\
&= \sum_{a\in\mathcal{A}} P\left(y_a(\lambda) = r+2\right) \\
&= \sum_{a\in\mathcal{A}} e^{-\lambda p_a}\frac{(\lambda p_a)^{r+2}}{(r+2)!},
\end{aligned}$$

plugging $\mathrm{E}[N_{r+2}]$ into s_λ^2 gives that

$$\begin{aligned}
s_\lambda^2 &= (r+1)^2\,\mathrm{E}\left[N_{r+1}\right] + (r+2)\,(r+1)\,\mathrm{E}\left[N_{r+2}\right] \\
&= (r+1)^2\,\mathrm{E}\left[N_{r+1}\right] + (r+2)\,(r+1)\sum_{a\in\mathcal{A}} e^{-\lambda p_a}\frac{(\lambda p_a)^{r+2}}{(r+2)!}.
\end{aligned}$$

Then for all $\epsilon > 0$,

$$\begin{aligned}
s_\lambda^2 &= (r+1)^2\,E\left[N_{r+1}\right] + (r+1)\sum_{a\in\mathcal{A}} e^{-\lambda p_a}\frac{(\lambda p_a)^{r+2}}{(r+1)!}1_{[\lambda p_a < \epsilon s_\lambda]} \\
&\quad + (r+1)\sum_{a\in\mathcal{A}} e^{-\lambda p_a}\frac{(\lambda p_a)^{r+2}}{(r+1)!}1_{[\lambda p_a \geq \epsilon s_\lambda]} \\
&\leq (r+1)^2\,\mathrm{E}[N_{r+1}] + (r+1)\epsilon s_\lambda \mathrm{E}[N_{r+1}] + \sum_{a\in\mathcal{A}} e^{-\lambda p_a}\frac{(\lambda p_a)^{r+2}}{(r+1)!}1_{[\lambda p_a \geq \epsilon s_\lambda]},
\end{aligned}$$

and dividing s_λ^2 on both sides gives that

$$\begin{aligned}1 &\le s_\lambda^{-2}(r+1)^2\,\mathrm{E}\left[N_{r+1}\right] + s_\lambda^{-1}(r+1)\epsilon E\left[N_{r+1}\right] + s_\lambda^{-2}\sum_{a\in\mathcal{A}} e^{-\lambda p_a}\frac{(\lambda p_a)^{r+2}}{(r+1)!}1_{[\lambda p_a \ge \epsilon s_\lambda]}\\ &= (r+1)\frac{\mathrm{E}[N_{r+1}]}{s_\lambda}\left(\frac{r+1}{s_\lambda}+\epsilon\right) + s_\lambda^{-2}\sum_{a\in\mathcal{A}} e^{-\lambda p_a}\frac{(\lambda p_a)^{r+2}}{(r+1)!}1_{[\lambda p_a \ge \epsilon s_\lambda]}.\end{aligned}$$

Assume that $s_\lambda \to \infty$ as $n \to \infty$. If (6.8) holds, it follows by Theorem 1 that for all $\epsilon > 0$

$$\begin{aligned}&\lim_{\lambda\to\infty}\left[(r+1)\frac{\mathrm{E}[N_{r+1}]}{s_\lambda}\left(\frac{r+1}{s_\lambda}+\epsilon\right) + s_\lambda^{-2}\sum_{a\in\mathcal{A}} e^{-\lambda p_a}\frac{(\lambda p_a)^{r+2}}{(r+1)!}1_{[\lambda p_a \ge \epsilon s_\lambda]}\right]\\ &= \lim_{\lambda\to\infty}(r+1)\frac{\mathrm{E}[N_{r+1}]}{s_\lambda}\left(\frac{r+1}{s_\lambda}+\epsilon\right)\\ &\ge 1.\end{aligned}$$

Since $r+1 \in (0,\infty)$, $\frac{r+1}{s_\lambda} \to 0$. So we argue by contradiction to show that $\frac{E[N_{r+1}]}{s_\lambda} \to \infty$. Suppose that

$$\liminf_{\lambda} \frac{\mathrm{E}[N_{r+1}]}{s_\lambda} = c \in [0,\infty).$$

Then for all $\epsilon > 0$ and some $c \in [0,\infty)$,

$$\liminf_{\lambda}(r+1)\frac{\mathrm{E}[N_{r+1}]}{s_\lambda}\left(\frac{r+1}{s_\lambda}+\epsilon\right) = (r+1)\epsilon c.$$

Taking $0 < \epsilon < \frac{1}{c(r+1)}$ gives that $(r+1)\epsilon c < 1$. This is a contradiction. Thus, this completes the proof of the lemma. □

Lemma 3. *For* $t = 0, 1, 2, \ldots$

$$\mathrm{Var}[N_t] \le \mathrm{E}[N_t].$$

Proof. Since for $t = 0, 1, 2, ...$

$$N_t = \sum_{a \in \mathcal{A}} 1_{[y_a(\lambda)=t]},$$

and the indicator function is only of independent random variables $y_a(\lambda)$'s, it follows that

$$\begin{aligned}
\mathrm{Var}[N_t] &= \mathrm{Var}\left[\sum_{a \in \mathcal{A}} 1_{[y_a(\lambda)=t]}\right] \\
&= \sum_{a \in \mathcal{A}} \mathrm{Var}\left[1_{[y_a(\lambda)=t]}\right] \\
&\leq \sum_{a \in \mathcal{A}} \mathrm{E}\left[1^2_{[y_a(\lambda)=t]}\right] \\
&= \sum_{a \in \mathcal{A}} \mathrm{E}\left[1_{[y_a(\lambda)=t]}\right] \\
&= \mathrm{E}\left[\sum_{a \in \mathcal{A}} 1_{[y_a(\lambda)=t]}\right] \\
&= \mathrm{E}[N_t],
\end{aligned}$$

and this completes the proof of this lemma. □

Lemma 4. *For $c, d \geq 0$ and $c + d > 0$. Let*

$$\begin{aligned}
V &= c\mathrm{E}[N_{r+1}] + d\mathrm{E}[N_{r+2}] \\
\hat{V} &= cN_{r+1} + dN_{r+2}.
\end{aligned}$$

We have

$$\mathrm{Var}[\hat{V}] \leq 2(c + d)\mathrm{E}[\hat{V}].$$

Further, if $V \to \infty$ as $\lambda \to \infty$, $\hat{V}$ is a consistent estimator of V, i.e.,

$$\frac{\hat{V}}{V} \xrightarrow{p} 1.$$

Proof. Set $\kappa = 2(c+d)$. By plugging in $\hat{V}$ and V in the left hand side and the right hand side we get

$$\begin{aligned}
\mathrm{Var}[\hat{V}] = &\mathrm{Var}[cN_{r+1} + dN_{r+2}] \\
=&c^2\mathrm{Var}[N_{r+1}] + d^2\mathrm{Var}[N_{r+2}] + 2cd\mathrm{Cov}[N_{r+1}, N_{r+2}] \\
\leq&2c^2\mathrm{Var}[N_{r+1}] + 2d^2\mathrm{Var}[N_{r+2}] + 2cd\mathrm{Cov}[N_{r+1}, N_{r+2}] \\
\leq&2c^2\mathrm{Var}[N_{r+1}] + 2d^2\mathrm{Var}[N_{r+2}] + 2cd(\mathrm{Var}[N_{r+1}] + \mathrm{Var}[N_{r+2}]) \\
=&2(c+d)c\mathrm{Var}[N_{r+1}] + 2(c+d)d\mathrm{Var}[N_{r+2}] \\
=&\kappa c\mathrm{Var}[N_{r+1}] + \kappa d\mathrm{Var}[N_{r+2}] \\
\leq&\kappa c\mathrm{E}[N_{r+1}] + \kappa d\mathrm{E}[N_{r+2}] \\
=&\kappa E[\hat{V}],
\end{aligned}$$

where the last inequality follows by Lemma 3 and the fourth line holds by the fact that

$$\mathrm{Cov}(X, Y) \leq \mathrm{Var}(X) + \mathrm{Var}(Y),$$

because

$$\mathrm{Var}(X - Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) - 2\mathrm{Cov}(X, Y) \geq 0.$$

Now by Chebyshev's inequality, for all $\epsilon > 0$

$$\begin{aligned} P\left(\left|\frac{\hat{V}}{V} - 1\right| > \epsilon\right) &\le \frac{\mathrm{Var}\left[\frac{\hat{V}}{V}\right]}{\epsilon^2} \\ &= \frac{\mathrm{Var}[\hat{V}]}{\epsilon^2 V^2} \\ &\le \frac{\kappa \mathrm{E}[\hat{V}]}{\epsilon^2 V^2} \\ &\le \frac{\kappa}{\epsilon^2 V^2} \to 0, \end{aligned}$$

where the last inequality holds because $c, d, N_{r+1}, N_{r+2} \ge 0$.

The proof of this lemma is completed. □

Proof of Theorem 1. For any $k > 0$, let $f(x) = x^k e^{-x}$ for $x > 0$. Since

$$f^{'}(x) = (kx^{-1} - 1)x^k e^{-x},$$

it follows that

$$\max_{x \ge 0} f(x) = f(k) = k^k e^{-k}.$$

Hence,

$$\begin{aligned} 0 < \sigma^2_{a,\lambda} &= (r + 1 + \lambda p_a) e^{-\lambda p_a} \frac{(\lambda p_a)^{r+1}}{r!} \\ &\le (r + 1 + \lambda p_a)^{r+2} e^{-(r+1+\lambda p_a)} e^{r+1} \\ &\le (r+2)^{r+2} e^{-(r+2)} e^{r+1} \\ &= (r+2)^{r+2} e^{-1}. \end{aligned}$$

It follows that since

$$\lim_{\lambda \to \infty} s_\lambda = \infty,$$

we have

$$\lim_{\lambda\to\infty} \sup_{a\in\mathcal{A}} \frac{\sigma^2_{a,\lambda}}{s^2_\lambda} = 0.$$

From here Proposition 1 implies that asymptotic normality is equivalent to

$$\lim_{\lambda\to\infty} s_\lambda^{-2} \sum_{a\in\mathcal{A}} \mathrm{E}\left[Y_a^2 1_{[|Y_a|\geq \epsilon s_\lambda]}\right] = 0 \quad \forall \epsilon > 0.$$

We now show that this is equivalent to our condition (2.1). Since $s_\lambda \to \infty$, we can take λ large enough that $\epsilon s_\lambda > (r+1)$. Recall that

$$Y_a = \begin{cases} -\lambda p_a & \text{if } y_a(\lambda) = r \\ r+1 & \text{if } y_a(\lambda) = r+1 \\ 0 & \text{otherwise} \end{cases}$$

Thus, for such λ, if $|Y_a| \geq \epsilon s_\lambda$, then $Y_a = -\lambda p_a$, $y_a(\lambda) = r$, and $Y_a^2 = \lambda^2 p_a^2$. We have

$$[|Y_a| \geq \epsilon s_\lambda] = [Y_a = -\lambda p_a] \cap [\lambda p_a \geq \epsilon s_\lambda] = [y_a(\lambda) = r] \cap [\lambda p_a \geq \epsilon s_\lambda].$$

It follows that

$$\mathrm{E}\left[Y_a^2 1_{[|Y_a|\geq \epsilon s_\lambda]}\right] = \lambda^2 p_a^2 1_{[\lambda p_a \geq \epsilon s_\lambda]} P(y_a = r) = e^{-\lambda p_a} \frac{(\lambda p_a)^{r+2}}{r!} 1_{[\lambda p_a \geq \epsilon s_\lambda]}.$$

□

Proof of Corollary 1. Note that

$$N_r = N_r(\lambda) = \sum_{a\in\mathcal{A}} 1_{[y_a(\lambda)=r]}$$

$$\pi_r = \pi_r(\lambda) = \sum_{a\in\mathcal{A}} p_a 1_{[y_a(\lambda)=r]}$$

$$T_r = T_r\left(\lambda\right) = \frac{N_{r+1}}{\lambda}\left(r+1\right)$$
$$s_\lambda^2 = \left(r+1\right)^2 \mathrm{E}[N_{r+1}] + \left(r+2\right)\left(r+1\right)\mathrm{E}[N_{r+2}].$$

Since we assume that $s_\lambda \xrightarrow{\lambda\to\infty} \infty$ and by Lemma 2,

$$\mathrm{E}[N_{r+1}] \to \infty. \tag{6.9}$$

Now for all $\epsilon > 0$

$$P\left(\left|\frac{N_{r+1}}{\mathrm{E}[N_{r+1}]} - 1\right| > \epsilon\right) = P\left(\left|\frac{N_{r+1}}{\mathrm{E}[N_{r+1}]} - \mathrm{E}\left[\frac{N_{r+1}}{\mathrm{E}[N_{r+1}]}\right]\right| > \epsilon\right),$$

and by Chebyshev's inequality

$$P\left(\left|\frac{N_{r+1}}{\mathrm{E}[N_{r+1}]} - \mathrm{E}\left[\frac{N_{r+1}}{\mathrm{E}[N_{r+1}]}\right]\right| > \epsilon\right) \le \frac{\mathrm{Var}\left[\frac{N_{r+1}}{\mathrm{E}[N_{r+1}]}\right]}{\epsilon^2} = \frac{\mathrm{Var}[N_{r+1}]}{\epsilon^2\left(\mathrm{E}[N_{r+1}]\right)^2}. \tag{6.10}$$

It follows from (6.10) and Lemma 3 that for all $\epsilon > 0$

$$P\left(\left|\frac{N_{r+1}}{\mathrm{E}[N_{r+1}]} - 1\right| > \epsilon\right) \le \frac{\mathrm{E}[N_{r+1}]}{\epsilon^2\left(\mathrm{E}[N_{r+1}]\right)^2} = \frac{1}{\epsilon^2\mathrm{E}[N_{r+1}]},$$

and together with $E\left[N_{r+1}\right] \to \infty$,

$$\lim_{\lambda\to\infty} P\left(\left|\frac{N_{r+1}}{\mathrm{E}[N_{r+1}]} - 1\right| > \epsilon\right) = 0,$$

i.e.,

$$\frac{N_{r+1}}{\mathrm{E}[N_{r+1}]} \xrightarrow{p} 1. \tag{6.11}$$

Since

$$\frac{N_{r+1}}{s_\lambda} = \frac{\mathrm{E}[N_{r+1}]}{s_\lambda}\frac{N_{r+1}}{\mathrm{E}[N_{r+1}]},$$

by continuous mapping theorem Lemma 2 and (6.11) implies that

$$\frac{N_{r+1}}{s_\lambda} \xrightarrow{p} \infty.$$

Since $r+1 \in (0,\infty)$,

$$\frac{N_{r+1}(r+1)}{s_\lambda} \xrightarrow{p} \infty.$$

Now plugging in

$$T_r\left(\lambda\right) = \frac{N_{r+1}}{\lambda}\left(r+1\right),$$

$$\frac{\lambda T_r(\lambda)}{s_\lambda} = \frac{N_{r+1}(r+1)}{s_\lambda} \xrightarrow{p} \infty.$$

By the symmetry of Normal distribution (6.8) implies that

$$\frac{\lambda}{s_\lambda}\left(\pi_r(\lambda) - T_r(\lambda)\right) \xrightarrow[\lambda\to\infty]{d} N(0,1),$$

and so

$$\frac{\lambda T_r(\lambda)}{s_\lambda}\left(\frac{\pi_r(\lambda)}{T_r(\lambda)} - 1\right) \xrightarrow[\lambda\to\infty]{d} N(0,1).$$

Since

$$\frac{\lambda T_r(\lambda)}{s_\lambda} \xrightarrow{p} \infty,$$

it follows from Lemma 1 that

$$\frac{\pi_r(\lambda)}{T_r(\lambda)} - 1 \xrightarrow{p} 0.$$

Therefore,

$$\frac{T_r(\lambda)}{\pi_r(\lambda)} - 1 \xrightarrow{p} 0.$$

□

Proof of Corollary 2. Since $(r+1)^2 > 0$ and $(r+2)(r+1) > 0$, the result is an application of Lemma 4. □

6.2.2 Proofs for Section 2.3.2

Lemma 5. *Assume that $s_\lambda \to \infty$ as $\lambda \to \infty$. If*

$$\frac{\lambda_n}{s_{\lambda_n}} \left(T_{r,n}(\lambda_n) - \pi_{r,n}(\lambda_n)\right) \xrightarrow[\lambda_n \to \infty]{d} N(0,1), \tag{6.12}$$

then

$$\frac{\mathrm{E}[N_{r+1,n}]}{s_{\lambda_n}} \to \infty.$$

Proof. Since

$$\mathrm{E}[N_{r+2,n}] = \sum_{a \in \mathcal{A}} \mathrm{E}\left[1_{[y_{a,n}(\lambda_n)=r+2]}\right]$$

$$= \sum_{a \in \mathcal{A}} P\left(y_{a,n}(\lambda_n) = r + 2\right)$$
$$= \sum_{a \in \mathcal{A}} e^{-\lambda_n p_{a,n}} \frac{(\lambda_n p_{a,n})^{r+2}}{(r+2)!},$$

plugging $\mathrm{E}[N_{r+2,n}]$ into $s^2_{\lambda_n}$ gives that

$$\begin{aligned} s^2_{\lambda_n} &= (r+1)^2 \mathrm{E}\left[N_{r+1,n}\right] + (r+2)(r+1)\mathrm{E}\left[N_{r+2,n}\right] \\ &= (r+1)^2 \mathrm{E}\left[N_{r+1,n}\right] + (r+2)(r+1)\sum_{a \in \mathcal{A}} e^{-\lambda_n p_{a,n}} \frac{(\lambda_n p_{a,n})^{r+2}}{(r+2)!}. \end{aligned}$$

Then for all $\epsilon > 0$

$$\begin{aligned} s^2_{\lambda_n} &= (r+1)^2 \mathrm{E}\left[N_{r+1,n}\right] + (r+1)\sum_{a \in \mathcal{A}} e^{-\lambda_n p_{a,n}} \frac{(\lambda_n p_{a,n})^{r+2}}{(r+1)!} 1_{[\lambda_n p_{a,n} < \epsilon s_{\lambda_n}]} \\ &\quad + (r+1)\sum_{a \in \mathcal{A}} e^{-\lambda_n p_{a,n}} \frac{(\lambda_n p_{a,n})^{r+2}}{(r+1)!} 1_{[\lambda_n p_{a,n} \geq \epsilon s_{\lambda_n}]} \\ &\leq (r+1)^2 \mathrm{E}[N_{r+1,n}] + (r+1)\epsilon s_{\lambda_n} \mathrm{E}[N_{r+1,n}] + \sum_{a \in \mathcal{A}} e^{-\lambda_n p_{a,n}} \frac{(\lambda_n p_{a,n})^{r+2}}{(r+1)!} 1_{[\lambda_n p_{a,n} \geq \epsilon s_{\lambda_n}]}, \end{aligned}$$

and dividing $s^2_{\lambda_n}$ on both sides gives that

$$\begin{aligned} 1 \leq& s^{-2}_{\lambda_n} (r+1)^2 \mathrm{E}\left[N_{r+1,n}\right] + s^{-1}_{\lambda_n}(r+1)\epsilon \mathrm{E}\left[N_{r+1,n}\right] \\ &+ s^{-2}_{\lambda_n} \sum_{a \in \mathcal{A}} e^{-\lambda_n p_{a,n}} \frac{(\lambda_n p_{a,n})^{r+2}}{(r+1)!} 1_{[\lambda_n p_{a,n} \geq \epsilon s_{\lambda_n}]} \\ =&(r+1)\frac{\mathrm{E}[N_{r+1,n}]}{s_{\lambda_n}} \left(\frac{r+1}{s_{\lambda_n}} + \epsilon\right) + s^{-2}_{\lambda_n} \sum_{a \in \mathcal{A}} e^{-\lambda_n p_{a,n}} \frac{(\lambda_n p_{a,n})^{r+2}}{(r+1)!} 1_{[\lambda_n p_{a,n} \geq \epsilon s_{\lambda_n}]}. \end{aligned}$$

Assume that $s_{\lambda_n} \to \infty$ as $n \to \infty$. If (6.12) holds, it follows by Theorem 2 that for all $\epsilon > 0$

$$\begin{aligned} &\lim_{\lambda_n \to \infty} \left[(r+1)\frac{\mathrm{E}[N_{r+1,n}]}{s_{\lambda_n}} \left(\frac{r+1}{s_{\lambda_n}} + \epsilon\right) + s^{-2}_{\lambda_n} \sum_{a \in \mathcal{A}} e^{-\lambda_n p_{a,n}} \frac{(\lambda_n p_{a,n})^{r+2}}{(r+1)!} 1_{[\lambda_n p_{a,n} \geq \epsilon s_{\lambda_n}]}\right] \\ &= \lim_{\lambda_n \to \infty} (r+1)\frac{\mathrm{E}[N_{r+1,n}]}{s_{\lambda_n}} \left(\frac{r+1}{s_{\lambda_n}} + \epsilon\right) \end{aligned}$$

$\geq 1.$

Since $r+1 \in (0,\infty)$, $\frac{r+1}{s_{\lambda_n}} \to 0$, and so we argue by contradiction to show that $\frac{\mathrm{E}[N_{r+1,n}]}{s_{\lambda_n}} \to \infty$. Suppose that

$$\liminf_{\lambda_n} \frac{\mathrm{E}[N_{r+1,n}]}{s_{\lambda_n}} = c \in [0,\infty).$$

Then for all $\epsilon > 0$ and some $c \in [0,\infty)$

$$\liminf_{\lambda_n} (r+1)\frac{\mathrm{E}[N_{r+1,n}]}{s_{\lambda_n}} \left(\frac{r+1}{s_{\lambda_n}} + \epsilon\right) = (r+1)\epsilon c.$$

Taking $0 < \epsilon < \frac{1}{c(r+1)}$ gives that $(r+1)\epsilon c < 1$. This is a contradiction. Thus, this completes the proof. □

Lemma 6. *For* $t = 0, 1, 2, \ldots$

$$\mathrm{Var}[N_{t,n}] \leq \mathrm{E}[N_{t,n}].$$

Proof. Since for $t = 0, 1, 2, \ldots$

$$N_{t,n} = \sum_{a \in \mathcal{A}} 1_{[y_{a,n}(\lambda_n)=t]},$$

and the indicator function is only of independent random variables $y_{a,n}(\lambda_n)$'s, it follows that

$$\begin{aligned}
\mathrm{Var}[N_{t,n}] &= \mathrm{Var}\left[\sum_{a \in \mathcal{A}} 1_{[y_{a,n}(\lambda_n)=t]}\right] \\
&= \sum_{a \in \mathcal{A}} \mathrm{Var}\left[1_{[y_{a,n}(\lambda_n)=t]}\right] \\
&\leq \sum_{a \in \mathcal{A}} \mathrm{E}\left[1^2_{[y_{a,n}(\lambda_n)=t]}\right]
\end{aligned}$$

$$= \sum_{a\in\mathcal{A}} \mathrm{E}\left[1_{[y_a(\lambda_n)=t]}\right]$$
$$= \mathrm{E}\left[\sum_{a\in\mathcal{A}} 1_{[y_{a,n}(\lambda_n)=t]}\right]$$
$$= \mathrm{E}[N_{t,n}],$$

and this completes the proof of the lemma. □

Lemma 7. *For $c, d \geq 0$ and $c + d > 0$. Let*

$$M_n = c\mathrm{E}[N_{r+1,n}] + d\mathrm{E}[N_{r+2,n}]$$
$$\hat{M}_n = cN_{r+1,n} + dN_{r+2,n}.$$

We have

$$\mathrm{Var}[\hat{M}_n] \leq 2(c+d)\mathrm{E}[\hat{M}_n].$$

Further, if $M_n \to \infty$ as $\lambda_n \to \infty$, $\hat{M}_n$ is a consistent estimator of M_n, i.e.,

$$\frac{\hat{M}_n}{M_n} \xrightarrow{p} 1.$$

Proof. Set $\kappa = 2(c+d)$. By plugging $\hat{M}_n$ in the left hand side and the right hand side, we have

$$\begin{aligned}
\mathrm{Var}[\hat{M}_n] =& \mathrm{Var}[cN_{r+1,n} + dN_{r+2,n}] \\
=& c^2\mathrm{Var}[N_{r+1,n}] + d^2\mathrm{Var}[N_{r+2,n}] + 2cd\mathrm{Cov}[N_{r+1,n}, N_{r+2,n}] \\
\leq& 2c^2\mathrm{Var}[N_{r+1,n}] + 2d^2\mathrm{Var}[N_{r+2,n}] + 2cd\mathrm{Cov}[N_{r+1,n}, N_{r+2,n}] \\
\leq& 2c^2\mathrm{Var}[N_{r+1,n}] + 2d^2\mathrm{Var}[N_{r+2,n}] + 2cd(\mathrm{Var}[N_{r+1,n}] + \mathrm{Var}[N_{r+2,n}]) \\
=& 2(c+d)c\mathrm{Var}[N_{r+1,n}] + 2(c+d)d\mathrm{Var}[N_{r+2,n}]
\end{aligned}$$

$$
\begin{aligned}
&=\kappa c \mathrm{Var}[N_{r+1,n}] + \kappa d \mathrm{Var}[N_{r+2,n}] \\
&\leq \kappa c \mathrm{E}[N_{r+1,n}] + \kappa d \mathrm{E}[N_{r+2,n}] \\
&= \kappa E[\hat{M}_n],
\end{aligned}
$$

where the last inequality follows by Lemma 6 and the fourth line holds by the fact that

$$\mathrm{Cov}(X, Y) \leq \mathrm{Var}(X) + \mathrm{Var}(Y),$$

because

$$\mathrm{Var}(X - Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) - 2\mathrm{Cov}(X, Y) \geq 0.$$

Now by Chebyshev's inequality, for all $\epsilon > 0$

$$
\begin{aligned}
P\left(\left|\frac{\hat{M}_n}{M_n} - 1\right| > \epsilon\right) &\leq \frac{\mathrm{Var}\left[\frac{\hat{M}_n}{M_n}\right]}{\epsilon^2} \\
&= \frac{\mathrm{Var}[\hat{M}_n]}{\epsilon^2 M_n^2} \\
&\leq \frac{\kappa \mathrm{E}[\hat{M}_n]}{\epsilon^2 M_n^2} \\
&\leq \frac{\kappa}{\epsilon^2 M_n^2} \to 0,
\end{aligned}
$$

where the last inequality holds because $c, d, N_{r+1,n}, N_{r+2,n} \geq 0$.

This completes the proof. □

Proof of Theorem 2. For any $k > 0$, let $f(x) = x^k e^{-x}$ for $x > 0$. Since

$$f^{'}(x) = (kx^{-1} - 1)x^k e^{-x},$$

it follows that

$$\max_{x \geq 0} f(x) = f(k) = k^k e^{-k}.$$

Hence,

$$\begin{aligned}
0 < \sigma^2_{a,\lambda_n} &= (r+1+\lambda_n) e^{-\lambda_n p_{a,n}} \frac{(\lambda_n p_{a,n})^{r+1}}{r!} \\
&\leq (r+1+\lambda_n p_{a,n})^{r+2} e^{-(r+1+\lambda_n p_{a,n})} e^{r+1} \\
&\leq (r+2)^{r+2} e^{-(r+2)} e^{r+1} \\
&= (r+2)^{r+2} e^{-1}.
\end{aligned}$$

It follows that since

$$\lim_{n \to \infty} s_{\lambda_n} = \infty,$$

we have

$$\lim_{n \to \infty} \sup_{a \in \mathcal{A}} \frac{\sigma^2_{a,\lambda_n}}{s^2_{\lambda_n}} = 0.$$

From here Proposition 1 implies that asymptotic normality is equivalent to

$$\lim_{n \to \infty} s^{-2}_{\lambda_n} \sum_{a \in \mathcal{A}} \mathrm{E}\left[Y^2_{a,n} 1_{[|Y_{a,n}| \geq \epsilon s_{\lambda_n}]} \right] = 0 \quad \forall \epsilon > 0.$$

We now show that this is equivalent to (2.3). Since $s_{\lambda_n} \to \infty$, we can take λ_n large enough that $\epsilon s_{\lambda_n} > (r+1)$. Recall that

$$Y_{a,n} = \begin{cases} -\lambda_n p_{a,n} & \text{if } y_{a,n}(\lambda_n) = r \\ r+1 & \text{if } y_{a,n}(\lambda_n) = r+1 \\ 0 & \text{otherwise} \end{cases}$$

Thus, for such λ_n, if $|Y_a| \geq \epsilon s_{\lambda_n}$, then $Y_{a,n} = -\lambda_n p_{a,n}$, $y_{a,n}(\lambda_n) = r$, and $Y_{a,n}^2 = \lambda_n^2 p_{a,n}^2$.

We have

$$[|Y_{a,n}| \geq \epsilon s_{\lambda_n}] = [Y_{a,n} = -\lambda_n p_{a,n}] \cap [\lambda_n p_{a,n} \geq \epsilon s_{\lambda_n}] = [y_{a,n}(\lambda_n) = r] \cap [\lambda_n p_{a,n} \geq \epsilon s_{\lambda_n}].$$

It follows that

$$\begin{aligned} \mathrm{E}\left[Y_{a,n}^2 1_{[|Y_{a,n}| \geq \epsilon s_{\lambda_n}]}\right] &= \lambda_n^2 p_{a,n}^2 1_{[\lambda_n p_{a,n} \geq \epsilon s_{\lambda_n}]} P(y_{a,n} = r) \\ &= e^{-\lambda_n p_{a,n}} \frac{(\lambda_n p_{a,n})^{r+2}}{r!} 1_{[\lambda_n p_{a,n} \geq \epsilon s_{\lambda_n}]}. \end{aligned}$$

□

Proof of Corollary 3. Note that

$$\begin{aligned} N_{r,n} &= \sum_{a \in \mathcal{A}} 1_{[y_{a,n}(\lambda_n) = r]} \\ \pi_{r,n} &= \sum_{a \in \mathcal{A}} p_{a,n} 1_{[y_{a,n}(\lambda_n) = r]} \\ T_{r,n} &= \frac{N_{r+1,n}}{\lambda_n} (r+1) \\ s_{\lambda_n}^2 &= (r+1)^2 \mathrm{E}[N_{r+1,n}] + (r+2)(r+1) \mathrm{E}[N_{r+2,n}]. \end{aligned}$$

Since we assume that $s_{\lambda_n} \xrightarrow{\lambda_n \to \infty} \infty$ and by Lemma 5,

$$\mathrm{E}[N_{r+1,n}] \to \infty. \tag{6.13}$$

Now for all $\epsilon > 0$

$$P\left(\left|\frac{N_{r+1,n}}{\mathrm{E}[N_{r+1,n}]} - 1\right| > \epsilon\right) = P\left(\left|\frac{N_{r+1,n}}{\mathrm{E}[N_{r+1,n}]} - \mathrm{E}\left[\frac{N_{r+1,n}}{\mathrm{E}[N_{r+1,n}]}\right]\right| > \epsilon\right),$$

and by Chebyshev's inequality

$$P\left(\left|\frac{N_{r+1,n}}{\mathrm{E}[N_{r+1,n}]} - E\left[\frac{N_{r+1,n}}{\mathrm{E}[N_{r+1,n}]}\right]\right| > \epsilon\right) \leq \frac{\mathrm{Var}\left[\frac{N_{r+1,n}}{\mathrm{E}[N_{r+1,n}]}\right]}{\epsilon^2} = \frac{\mathrm{Var}[N_{r+1,n}]}{\epsilon^2\left(\mathrm{E}[N_{r+1,n}]\right)^2}. \quad (6.14)$$

It follows from (6.14) and Lemma 6 that for all $\epsilon > 0$

$$P\left(\left|\frac{N_{r+1,n}}{\mathrm{E}[N_{r+1,n}]} - 1\right| > \epsilon\right) \leq \frac{\mathrm{E}[N_{r+1,n}]}{\epsilon^2\left(\mathrm{E}[N_{r+1,n}]\right)^2} = \frac{1}{\epsilon^2 \mathrm{E}[N_{r+1,n}]},$$

and together with $\mathrm{E}\left[N_{r+1,n}\right] \to \infty$,

$$\lim_{\lambda_n \to \infty} P\left(\left|\frac{N_{r+1,n}}{\mathrm{E}[N_{r+1,n}]} - 1\right| > \epsilon\right) = 0,$$

i.e.,

$$\frac{N_{r+1,n}}{\mathrm{E}[N_{r+1,n}]} \xrightarrow{p} 1 \quad (6.15)$$

Since

$$\frac{N_{r+1,n}}{s_{\lambda_n}} = \frac{\mathrm{E}[N_{r+1,n}]}{s_{\lambda_n}} \frac{N_{r+1,n}}{\mathrm{E}[N_{r+1,n}]},$$

by continuous mapping theorem Lemma 5 and (6.15) imply that

$$\frac{N_{r+1,n}}{s_{\lambda_n}} \xrightarrow{p} \infty.$$

Since $r + 1 \in (0, \infty)$,

$$\frac{N_{r+1,n}(r+1)}{s_{\lambda_n}} \xrightarrow{p} \infty.$$

Now plugging in

$$T_{r,n}\left(\lambda_n\right) = \frac{N_{r+1,n}}{\lambda_n}\left(r+1\right),$$

$$\frac{\lambda_n T_{r,n}(\lambda_n)}{s_{\lambda_n}} = \frac{N_{r+1,n}(r+1)}{s_{\lambda_n}} \xrightarrow{p} \infty.$$

By the symmetry of Normal distribution (6.12) implies that

$$\frac{\lambda_n}{s_{\lambda_n}}\left(\pi_{r,n}(\lambda_n) - T_{r,n}(\lambda_n)\right) \xrightarrow[\lambda_n \to \infty]{d} N(0,1),$$

and so

$$\frac{\lambda_n T_{r,n}(\lambda_n)}{s_{\lambda_n}}\left(\frac{\pi_{r,n}(\lambda_n)}{T_{r,n}(\lambda_n)} - 1\right) \xrightarrow[\lambda_n \to \infty]{d} N(0,1).$$

Since

$$\frac{\lambda T_{r,n}(\lambda_n)}{s_{\lambda_n}} \xrightarrow{p} \infty,$$

it follows from Lemma 1 that

$$\frac{\pi_{r,n}(\lambda_n)}{T_{r,n}(\lambda_n)} - 1 \xrightarrow{p} 0.$$

Therefore,

$$\frac{T_{r,n}(\lambda_n)}{\pi_{r,n}(\lambda_n)} - 1 \xrightarrow{p} 0.$$

□

Proof of Theorem 3. First, for any $\epsilon > 0$, the fact that (2.3) holds gives

$$\begin{aligned} \mathrm{E}[N_{r+2,n}] &= \sum_{a\in\mathcal{A}}(\lambda_n p_{a,n})^{r+2}\frac{e^{-\lambda_n p_{a,n}}}{(r+2)!}1_{[\lambda_n p_{a,n}\le\epsilon s_{r,\lambda_n,n}]} \\ &\quad + \sum_{a\in\mathcal{A}}(\lambda_n p_{a,n})^{r+2}\frac{e^{-\lambda_n p_{a,n}}}{(r+2)!}1_{[\lambda_n p_{a,n}>\epsilon s_{r,\lambda_n,n}]} \\ &\le s_{r,\lambda_n,n}\epsilon\mathrm{E}[N_{r+1,n}] + \sum_{a\in\mathcal{A}}(\lambda_n p_{a,n})^{r+2}e^{-\lambda_n p_{a,n}}1_{[\lambda_n p_{a,n}>\epsilon s_{r,\lambda_n,n}]} \\ &\le s^2_{r,\lambda_n,n}\epsilon + \sum_{a\in\mathcal{A}}(\lambda_n p_{a,n})^{r+2}e^{-\lambda_n p_{a,n}}1_{[\lambda_n p_{a,n}>\epsilon s_{r,\lambda_n,n}]} \to c^2\epsilon, \end{aligned}$$

which implies that $\mathrm{E}[N_{r+2,n}] \to 0$ and hence that $\mathrm{E}[N_{r+1,n}] \to c^*$.

Next, note that $\frac{\lambda_n}{r+1}\mathrm{E}[\pi_{r,n}] \to c^*$ by (2.2) and that

$$\begin{aligned} \mathrm{Var}\left(\frac{\lambda_n}{r+1}\pi_{r,n}(\lambda_n)\right) &= \frac{1}{(r+1)^2}\sum_{a\in\mathcal{A}}(\lambda_n p_{a,n})^2\mathrm{Var}\left(1_{[y_{a,n}(\lambda_n)=r]}\right) \\ &\le \frac{1}{(r+1)^2}\sum_{a\in\mathcal{A}}(\lambda_n p_{a,n})^2 P\left(y_{a,n}\left(\lambda_n\right)=r\right) \\ &= \frac{r+2}{r+1}\mathrm{E}[N_{r+2,n}] \to 0. \end{aligned}$$

From here the first convergence in (2.4) follows by the well known presentation of mean square error as the sum of the variance and the square of the bias. The second convergence follows from the first and Markov's inequality.

Finally, note that

$$\frac{\lambda_n}{r+1}T_{r,n}(\lambda_n) = N_{r+1,n}(\lambda_n) = \sum_{a\in\mathcal{A}}1_{[y_{a,n}(\lambda_n)=r+1]}$$

is the sum of independent Bernoulli random variables. We just need to check that the Poisson approximation to the binomial holds. By Proposition 2 this holds so long as $\sup_{a\in\mathcal{A}} P(y_{a,n}\left(\lambda_n\right) = r+1) \to 0$. Note that

$$P(y_{a,n}\left(\lambda_n\right) = r+1) \;=\; e^{-p_{a,n}\lambda_n}\frac{(p_{a,n}\lambda_n)^{r+1}}{(r+1)!}$$

$$
\begin{aligned}
&= \frac{1}{(r+1)!}\left(e^{-p_{a,n}\lambda_n(r+2)/(r+1)}(p_{a,n}\lambda_n)^{r+2}\right)^{(r+1)/(r+2)} \\
&\le \frac{1}{(r+1)!}\left(\sum_{a\in\mathcal{A}} e^{-p_{a,n}\lambda_n}(p_{a,n}\lambda_n)^{r+2}\right)^{(r+1)/(r+2)} \\
&= \frac{((r+2)!)^{(r+1)/(r+2)}}{(r+1)!}\left(\mathrm{E}[N_{r+2,n}]\right)^{(r+1)/(r+2)} \to 0,
\end{aligned}
$$

and the result follows. □

6.3 Proofs for Section 2.4

The proof of main results in Section 2.4.2 is based on approximating the distribution in the Deterministic case with the distribution in the Poisson case, and we call this process "depoissonization". Toward this end, we introduce a model that contains both of these with both the fixed and changing distributions. Details of the model are given in the following sections of proofs.

6.3.1 Proofs for Section 2.4.1

First, we explain our model. Assume that we are sampling observations following a Poisson Process with rate 1, denoted as $C = \{C_\lambda : \lambda \geq 0\}$.

For $n = 1, 2, 3, \ldots$, let $t_n = min\{\lambda \geq 0 : C_\lambda = n\}$ be the arrival time on the nth observation. If we stop sampling at time t_n, then the sample is of size n and we have the deterministic model studied in Section 2.4.1. Whereas, if we consider the sample taken at time λ, then the sample size is C_λ and we have the Poisson model studied in Section 2.3.1 with $\lambda = n$. Observe that $\mathrm{E}[C_n] = n = C_{t_n}$. Thus, we expect to have the same sample sizes in those two sampling schemes. When the sample size is the deterministic n at a random sampling time t_n, we use notations defined in Section 2.4.1; while, when the sample size is a random C_λ at the deterministic sampling time λ, we use notations defined in Section 2.3.1. Further, let

$$\xi_n = n(T'_r(n) - \pi'_r(n))$$

be the Deterministic version, and

$$\zeta_\lambda = \lambda(T_r(\lambda) - \pi_r(\lambda))$$

be the Poissonized version. Observe that $y'_a(n) = y_a(t_n)$, and t_n follows a gamma distribution with both mean and variance n. Note that for ξ_n we have a deterministic sample size n at a random time t_n, whereas for ζ_λ we have a random sample size N at a fixed time λ. Also note that

$$\zeta_{t_n} = \frac{t_n}{n}\xi_n.$$

To find a necessary and sufficient condition for asymptotic normality of ξ_n, we use the asymptotic normality of ζ_λ and show that $\xi_n - \zeta_\lambda \xrightarrow{p} 0$, specifically when $\lambda = n$.

Before giving the proof of Theorem 4 for the Deterministic case with fixed distribution, we prepare several lemmas.

Lemma 8. *For any $\lambda > 0$ and $\Delta \in (0, \lambda)$, we have*

$$\mathrm{E}\left[\sup_{\lambda < t < \lambda+\Delta} |\zeta_t - \zeta_\lambda|\right] \leq H(\lambda, \Delta)$$

and

$$\mathrm{E}\left[\sup_{\lambda-\frac{\Delta}{2} < t < \lambda+\frac{\Delta}{2}} |\zeta_t - \zeta_\lambda|\right] \leq 2H(\lambda - \frac{\Delta}{2}, \Delta),$$

where for some constant $C > 0$,

$$H(\lambda, \Delta) = C\frac{\Delta}{\lambda}s_\lambda^2.$$

Proof. Recall that for any $\lambda > 0$ we have

$$\zeta_\lambda = \lambda \left(T_r(\lambda) - \pi_r(\lambda) \right) = \sum_{a \in \mathcal{A}} Y_a(\lambda),$$

and

$$Y_a = (r+1)\, 1_{[y_a(\lambda)=r+1]} - \lambda p_a 1_{[y_a(\lambda)=r]}.$$

Fix $t > \lambda$ and note that $y_a(t) \geq y_a(\lambda)$ because greater arrival time yields more or equal arrivals in a Poisson process and

$$\begin{aligned} Y_a(t) - Y_a(\lambda) =& 1_{[y_a(\lambda)<r]} Y_a(t) + 1_{[y_a(\lambda)] \geq r]} Y_a(t) \\ & - Y_a(\lambda) 1_{[y_a(t) > y_a(\lambda)]} - Y_a(\lambda) 1_{[y_a(t) = y_a(\lambda)]} \\ =& - Y_a(\lambda) 1_{[y_a(t) > y_a(\lambda)]} + 1_{[y_a(\lambda)<r]} Y_a(t) \\ & - Y_a(\lambda) 1_{[y_a(t) = y_a(\lambda)]} + 1_{[y_a(\lambda)] \geq r]} Y_a(t). \end{aligned}$$

Since

$$\begin{aligned} & - Y_a(\lambda) 1_{[y_a(t) = y_a(\lambda)]} + 1_{[y_a(\lambda)] \geq r]} Y_a(t) \\ =& - (r+1) 1_{[y_a(\lambda)=r+1]} 1_{[y_a(t)=y_a(\lambda)]} + \lambda p_a 1_{[y_a(\lambda)=r]} 1_{[y_a(t)=y_a(\lambda)]} \\ & + (r+1) 1_{[y_a(t)=r+1]} 1_{[y_a(\lambda) \geq r]} - t p_a 1_{[y_a(t)=r]} 1_{[y_a(\lambda) \geq r]} \\ =& - (r+1) 1_{[y_a(\lambda)=r+1]} 1_{[y_a(t)=r+1]} + \lambda p_a 1_{[y_a(\lambda)=r]} 1_{[y_a(t)=r]} \\ & + (r+1) 1_{[y_a(t)=r+1]} 1_{[y_a(\lambda)=r+1]} + (r+1) 1_{[y_a(t)=r+1]} 1_{[y_a(\lambda)=r]} - t p_a 1_{[y_a(t)=r]} 1_{[y_a(\lambda)=r]} \\ =& (r+1) 1_{[y_a(t)=r+1]} 1_{[y_a(\lambda)=r]} - (t-\lambda) p_a 1_{[y_a(t)=r]} 1_{[y_a(\lambda)=r]} \\ =& 1_{[y_a(\lambda)=r]} \left((r+1) 1_{[y_a(t)=r+1]} - (t-\lambda) p_a 1_{[y_a(t)=r]} \right), \end{aligned}$$

then

$$
\begin{aligned}
Y_a(t) - Y_a(\lambda) = &- Y_a(\lambda) 1_{[y_a(t) > y_a(\lambda)]} + 1_{[y_a(\lambda) < r]} Y_a(t) \\
&+ 1_{[y_a(\lambda) = r]} ((r+1) 1_{[y_a(t) = r+1]} - (t - \lambda) p_a 1_{[y_a(t) = r]}).
\end{aligned}
$$

Now note that

$$
\begin{aligned}
&|\sum_{a \in \mathcal{A}} (Y_a(t) - Y_a(\lambda))| \\
\leq &|\sum_{a \in \mathcal{A}} Y_a(\lambda) 1_{[y_a(t) > y_a(\lambda)]}| + |\sum_{a \in \mathcal{A}} 1_{[y_a(\lambda) < r]} Y_a(t)| \\
&+ |\sum_{a \in \mathcal{A}} 1_{[y_a(\lambda) = r]} ((r+1) 1_{[y_a(t) = r+1]} - (t - \lambda) p_a 1_{[y_a(t) = r]})| \\
\leq &(r+1) \sum_{a \in \mathcal{A}} 1_{[y_a(\lambda) = r+1]} 1_{[y_a(t) > y_a(\lambda)]} + \lambda \sum_{a \in \mathcal{A}} p_a 1_{[y_a(\lambda) = r]} 1_{[y_a(t) > y_a(\lambda)]} \\
&+ (r+1) \sum_{a \in \mathcal{A}} 1_{[y_a(t) = r+1]} 1_{[y_a(\lambda) < r]} + t \sum_{a \in \mathcal{A}} p_a 1_{[y_a(t) = r]} 1_{[y_a(\lambda) < r]} \\
&+ (r+1) \sum_{a \in \mathcal{A}} 1_{[y_a(\lambda) = r]} 1_{[y_a(t) = r+1]} + \sum_{a \in \mathcal{A}} |t - \lambda| p_a 1_{[y_a(\lambda) = r]} 1_{[y_a(t) = r]}.
\end{aligned}
$$

Now set

$$
\begin{aligned}
A_t^1 &= \sum_{a \in \mathcal{A}} 1_{[y_a(\lambda) = r+1]} 1_{[y_a(t) > y_a(\lambda)]} \\
A_t^2 &= \lambda \sum_{a \in \mathcal{A}} p_a 1_{[y_a(\lambda) = r]} 1_{[y_a(t) > y_a(\lambda)]} \\
B_t^1 &= \sum_{a \in \mathcal{A}} 1_{[y_a(t) = r+1]} 1_{[y_a(\lambda) < r]} \\
B_t^2 &= t \sum_{a \in \mathcal{A}} p_a 1_{[y_a(t) = r]} 1_{[y_a(\lambda) < r]} \\
C_t &= \sum_{a \in \mathcal{A}} 1_{[y_a(\lambda) = r]} 1_{[y_a(t) = r+1]} \\
D_t &= \sum_{a \in \mathcal{A}} |t - \lambda| p_a 1_{[y_a(\lambda) = r]} 1_{[y_a(t) = r]},
\end{aligned}
$$

then

$$
\begin{aligned}
|\zeta_t - \zeta_\lambda| &= |\sum_{a\in\mathcal{A}}(Y_a(t) - Y_a(\lambda))| \\
&\le (r+1)A_t^1 + A_t^2 + (r+1)B_t^1 + B_t^2 + (r+1)C_t + D_t.
\end{aligned}
$$

We are going to find the bounds for each element.

Bounds for C_t and D_t:

$$
\begin{aligned}
C_t &= \sum_{a\in\mathcal{A}} 1_{[y_a(\lambda)=r]} 1_{[y_a(t)=r+1]} \\
D_t &= \sum_{a\in\mathcal{A}} |t-\lambda| p_a 1_{[y_a(\lambda)=r]} 1_{[y_a(t)=r]}
\end{aligned}
$$

By Fubini's Theorem and the fact that Poisson processes have independent increments,

$$
\begin{aligned}
\mathrm{E}\left[\sup_{\lambda<t<\lambda+\Delta} C_t\right] &\le \left[\mathrm{E}\sup_{\lambda<t<\lambda+\Delta}\sum_{a\in\mathcal{A}} 1_{[y_a(\lambda)=r]} 1_{[y_a(t)>y_a(\lambda)]}\right] \\
&\le \mathrm{E}\left[\sum_{a\in\mathcal{A}} 1_{[y_a(\lambda)=r]} 1_{[y_a(\lambda+\Delta)>y_a(\lambda)]}\right] (\text{Note} : t < \lambda+\Delta) \\
&= \sum_{a\in\mathcal{A}} P(y_a(\lambda)=r) P(y_a(\lambda+\Delta) > y_a(\lambda)) \\
&= \sum_{a\in\mathcal{A}} \frac{\lambda^r}{r!} e^{-\lambda p_a} p_a^r (1 - e^{-\Delta p_a}) \\
&= \frac{\lambda^r}{r!} \sum_{a\in\mathcal{A}} e^{-\lambda p_a} p_a^r (1 - e^{-\Delta p_a}) \\
&\le \lambda^r \sum_{a\in\mathcal{A}} e^{-\lambda p_a} p_a^r (1 - e^{-\Delta p_a}) \\
&\le \lambda^r \sum_{a\in\mathcal{A}} e^{-\lambda p_a} p_a^r \Delta p_a \\
&= \Delta\lambda^r \sum_{a\in\mathcal{A}} e^{-\lambda p_a} p_a^{r+1},
\end{aligned}
$$

where the last inequality follows by the fact that $1 - e^{-x} \leq x$ for $x > 0$.

By similar arguments,

$$\begin{aligned}
\mathrm{E}\left[\sup_{\lambda<t<\lambda+\Delta} D_t\right] &\leq \mathrm{E}\left[\sum_{a\in\mathcal{A}} \Delta p_a 1_[y_a(\lambda) = r]\right] (\text{Note} : \Delta > t - \lambda > 0) \\
&= \Delta \sum_{a\in\mathcal{A}} p_a P(y_a(\lambda) = r) \\
&= \Delta \frac{\lambda^r}{r!} \sum_{a\in\mathcal{A}} e^{-\lambda p_a} p_a^{r+1} \\
&\leq \Delta \lambda^r \sum_{a\in\mathcal{A}} e^{-\lambda p_a} p_a^{r+1}.
\end{aligned}$$

Bound for B_t^1 and B_t^2:

$$\begin{aligned}
B_t^1 &= \sum_{a\in\mathcal{A}} 1_{[y_a(t)=r+1]} 1_{[y_a(\lambda)<r]} \\
B_t^2 &= t \sum_{a\in\mathcal{A}} p_a 1_{[y_a(t)=r]} 1_{[y_a(\lambda)<r]}
\end{aligned}$$

Clearly, if $r = 0$, then

$$\mathrm{E}\left[\sup_{\lambda<t<\lambda+\Delta} B_t^1\right] = \mathrm{E}\left[\sup_{\lambda<t<\lambda+\Delta} B_t^2\right] = 0.$$

Now, assume that $r \geq 1$. Note that by independent and stationary increments

$$\begin{aligned}
\mathrm{E}\left[\sup_{\lambda<t<\lambda+\Delta} B_t^1\right] &\leq \mathrm{E}\left[\sup_{\lambda<t<\lambda+\Delta} \sum_{a\in\mathcal{A}} \sum_{i=0}^{r-1} 1_{[y_a(t)>r]} 1_{[y_a(\lambda)=i]}\right] \\
&= \mathrm{E}\left[\sup_{\lambda<t<\lambda+\Delta} \sum_{a\in\mathcal{A}} \sum_{i=0}^{r-1} 1_{[y_a(t)-y_a(\lambda)>r-i]} 1_{[y_a(\lambda)=i]}\right] \\
&\leq \sum_{a\in\mathcal{A}} \sum_{i=0}^{r-1} \mathrm{E}\left[1_{[y_a(\lambda+\Delta)-y_a(\lambda)>r-i]} 1_{[y_a(\lambda)=i]}\right] \\
&= \sum_{a\in\mathcal{A}} \sum_{i=0}^{r-1} P(y_a(\Delta) > r - i) P(y_a(\lambda) = i)
\end{aligned}$$

$$\leq \sum_{a\in\mathcal{A}}\sum_{i=0}^{r-1}\frac{(\Delta p_a)^{r-i+1}}{(r-i+1)!}e^{-\lambda p_a}\frac{(p_a\lambda)^i}{i!}$$

$$\leq \Delta\lambda^r\sum_{a\in\mathcal{A}}\sum_{i=0}^{r-1}p_a^{r+1}e^{-\lambda p_a} = r\Delta\lambda^r\sum_{a\in\mathcal{A}}p_a^{r+1}e^{-\lambda p_a},$$

where we use the fact that for any integer $k \geq 0$

$$P(y_a(\Delta) > k) = 1 - \sum_{j=0}^{k} e^{-\Delta p_a}\frac{(\Delta p_a)^j}{j!} \leq \frac{(\Delta p_a)^{k+1}}{(k+1)!},$$

which follows since for any $x > 0$ we have $1 - e^{-x}\sum_{i=0}^{k} x^j/j! \leq x^{k+1}/(k+1)!$, see e.g. Lemma 1 in [31]. Similarly, for B_t^2 we have

$$\mathrm{E}\left[\sup_{\lambda<t<\lambda+\Delta} B_t^2\right] \leq \mathrm{E}\left[\sup_{\lambda<t<\lambda+\Delta} t\sum_{a\in\mathcal{A}}\sum_{i=0}^{r-1}p_a 1_{[y_a(t)>r-1]}1_{[y_a(\lambda)=i]}\right]$$

$$\leq (\lambda+\Delta)\sum_{a\in\mathcal{A}}\sum_{i=0}^{r-1}p_a\mathrm{E}\left[1_{[y_a(\lambda+\Delta)-y_a(\lambda)>r-1-i]}1_{[y_a(\lambda)=i]}\right]$$

$$= (\lambda+\Delta)\sum_{a\in\mathcal{A}}\sum_{i=0}^{r-1}p_a P(y_a(\Delta) > r-1-i)P(y_a(\lambda)=i)$$

$$\leq 2\lambda\sum_{a\in\mathcal{A}}\sum_{i=0}^{r-1}p_a(\Delta p_a)^{r-i}e^{-\lambda p_a}(\lambda p_a)^i$$

$$\leq 2r\Delta\lambda^r\sum_{a\in\mathcal{A}}p_a^{r+1}e^{-\lambda p_a}.$$

Bound for A_t^1 and A_t^2:

$$A_t^1 = \sum_{a\in\mathcal{A}}1_{[y_a(\lambda)=r+1]}1_{[y_a(t)>y_a(\lambda)]}$$

$$A_t^2 = \lambda\sum_{a\in\mathcal{A}}p_a 1_{[y_a(\lambda)=r]}1_{[y_a(t)>y_a(\lambda)]}$$

The proof for A_t^1 is similar to the proof for C_t. Here

$$\begin{aligned}
\mathrm{E}\left[\sup_{\lambda<t<\lambda+\Delta} A_t^1\right] &\leq \mathrm{E}\left[\sum_{a\in\mathcal{A}} 1_{[y_a(\lambda)=r+1]} 1_{[y_a(\lambda+\Delta)>y_a(\lambda)]}\right] \\
&= \sum_{a\in\mathcal{A}} P(y_a(\lambda)=r+1)P((y_a(\lambda+\Delta)-y_a(\lambda))>0) \\
&= \frac{\lambda^{r+1}}{(r+1)!}\sum_{a\in\mathcal{A}} p_a^{r+1}e^{-\lambda p_a}(1-e^{-\Delta p_a}) \\
&\leq \frac{\lambda^{r+1}}{(r+1)}\Delta\sum_{a\in\mathcal{A}} p_a^{r+2}e^{-\lambda p_a}.
\end{aligned}$$

Next, by Fubini's theorem and independent increments we have

$$\begin{aligned}
\mathrm{E}\left[\sup_{\lambda<t<\lambda+\Delta} A_t^2\right] &\leq \mathrm{E}\left[\lambda\sum_{a\in\mathcal{A}} p_a 1_{[y_a(\lambda)=r]} 1_{[y_a(\lambda+\Delta)>y_a(\lambda)]}\right] \\
&= \lambda\sum_{a\in\mathcal{A}} p_a P(y_a(\lambda)=r)P(y_a(\lambda+\Delta)>y_a(\lambda)) \\
&= \frac{\lambda^{r+1}}{r!}\sum_{a\in\mathcal{A}} p_a^{r+1}e^{-\lambda p_a}(1-e^{-\Delta p_a}) \\
&\leq \frac{\lambda^{r+1}}{r!}\Delta\sum_{a\in\mathcal{A}} p_a^{r+2}e^{-\lambda p_a} \\
&\leq \lambda^{r+1}\Delta\sum_{a\in\mathcal{A}} p_a^{r+2}e^{-\lambda p_a},
\end{aligned}$$

which completes the proof of this part. Now putting everything together gives the first bound:

$$\begin{aligned}
&\mathrm{E}\left[\sup_{\lambda<t<\lambda+\Delta} |\zeta_t-\zeta_\lambda|\right] \\
=&\mathrm{E}\left[\sup_{\lambda<t<\lambda+\Delta} \big((r+1)A_t^1+A_t^2+(r+1)B_t^1+B_t^2+(r+1)C_t+D_t\big)\right] \\
\leq&(r+1)\frac{\lambda^{r+1}}{(r+1)}\Delta\sum_{a\in\mathcal{A}} p_a^{r+2}e^{-\lambda p_a}+\lambda^{r+1}\Delta\sum_{a\in\mathcal{A}} p_a^{r+2}e^{-\lambda p_a} \\
&+(r+1)r\Delta\lambda^r\sum_{a\in\mathcal{A}} p_a^{r+1}e^{-\lambda p_a}+2r\Delta\lambda^r\sum_{a\in\mathcal{A}} p_a^{r+1}e^{-\lambda p_a}
\end{aligned}$$

$$
\begin{aligned}
&+ (r+1)\Delta\lambda^r \sum_{a\in\mathcal{A}} e^{-\lambda p_a} p_a^{r+1} \\
&+ \Delta\lambda^r \sum_{a\in\mathcal{A}} e^{-\lambda p_a} p_a^{r+1} \\
=& 2\Delta\frac{\lambda^{r+1}}{r!} \sum_{a\in\mathcal{A}} e^{-\lambda p_a} p_a^{r+2} + (r^2+4r+2)\Delta\lambda^r \sum_{a\in A} e^{-\lambda p_a} p_a^{r+1} \\
=& \frac{\Delta}{\lambda}\left((r^2+4r+2)\lambda^{r+1} \sum_{a\in A} e^{-\lambda p_a} p_a^{r+1} + \frac{2\lambda^{r+2}}{r!} \sum_{a\in\mathcal{A}} e^{-\lambda p_a} p_a^{r+2} \right) \\
=& H(\lambda,\Delta) \\
=& C\frac{\Delta}{\lambda} s_\lambda^2,
\end{aligned}
$$

which can be the upper bounded as required. From here applying the first bound twice gives

$$
\begin{aligned}
\mathrm{E}\left[\sup_{\lambda-\frac{\Delta}{2}<t<\lambda+\frac{\Delta}{2}} |\zeta_{t,n} - \zeta_{\lambda,n}| \right] \leq \mathrm{E}\left[\sup_{\lambda-\frac{\Delta}{2}<t<\lambda+\frac{\Delta}{2}} |\zeta_{t,n} - \zeta_{\lambda-\Delta/2,n}| \right] \\
+ \mathrm{E}\left[|\zeta_{\lambda-\Delta/2,n} - \zeta_{\lambda,n}| \right] \leq 2H\left(\lambda - \frac{\Delta}{2}, \Delta\right).
\end{aligned}
$$

□

Lemma 9. *Let* $0 < \lambda' < \lambda < \infty$*. For any* $\epsilon > 0$*,*

$$
(\frac{\lambda'}{\lambda})^{r+2} s_\lambda^2 \leq (s_{\lambda'})^2 \leq e^\epsilon s_\lambda^2 + (r+1+\lambda)\lambda^{r+1} e^{-\frac{\lambda'\epsilon}{\lambda-\lambda'}}. \tag{6.16}
$$

Further, let λ_n *and* λ'_n *be two sequences of numbers. If* $0 < \lambda'_n < \lambda_n < \infty$*,* $\lambda_n \sim \lambda'_n$*,* $\limsup_n (\frac{\lambda_n}{\lambda'_n} - 1)\lambda_n^\delta < \infty$ *for some* $\delta > 0$*, and* $\liminf_n s_{\lambda_n} > 0$*, then*

$$
s_{\lambda_n} \sim s_{\lambda'_n}.
$$

Proof. Let $0 < \lambda' < \lambda < \infty$, then

$$
\begin{aligned}
(\frac{\lambda'}{\lambda})^{r+2} s_\lambda^2 &= (\frac{\lambda'}{\lambda})^{r+2} \sum_{a\in\mathcal{A}} \left((r+1+\lambda p_a) e^{-\lambda p_a} \frac{(\lambda p_a)^{r+1}}{r!} \right) \\
&= \frac{(\lambda')^{r+2}}{\lambda} \sum_{a\in\mathcal{A}} \left((r+1+\lambda p_a) e^{-\lambda p_a} \frac{p_a^{r+1}}{r!} \right) \\
&= \frac{\lambda'}{\lambda} \sum_{a\in\mathcal{A}} \left((r+1+\lambda p_a) e^{-\lambda p_a} \frac{(\lambda' p_a)^{r+1}}{r!} \right) \\
&= \sum_{a\in\mathcal{A}} \left(\frac{\lambda'}{\lambda}(r+1) e^{-\lambda p_a} \frac{(\lambda' p_a)^{r+1}}{r!} + \lambda' p_a e^{-\lambda p_a} \frac{(\lambda' p_a)^{r+1}}{r!} \right) \\
&\le \sum_{a\in\mathcal{A}} \left((r+1) e^{-\lambda' p_a} \frac{(\lambda' p_a)^{r+1}}{r!} + \lambda' p_a e^{-\lambda p_a} \frac{(\lambda' p_a)^{r+1}}{r!} \right) \\
&= \sum_{a\in\mathcal{A}} \left((r+1+\lambda' p_a) e^{-\lambda' p_a} \frac{(\lambda' p_a)^{r+1})}{r!} \right) \\
&= (s_{\lambda'})^2,
\end{aligned}
$$

and for any $\epsilon > 0$

$$
\begin{aligned}
(s_{\lambda'})^2 \le &\sum_{a\in\mathcal{A}} \left((r+1+\lambda p_a) e^{-\lambda' p_a} \frac{(\lambda p_a)^{r+1}}{r!} \right) \\
= &\sum_{a\in\mathcal{A}} \left((r+1+\lambda p_a) e^{-\lambda' p_a} \frac{(\lambda p_a)^{r+1}}{r!} 1_{[(\lambda-\lambda')p_a \le \epsilon]} \right) \\
&+ \sum_{a\in\mathcal{A}} \left((r+1+\lambda p_a) e^{-\lambda' p_a} \frac{(\lambda p_a)^{r+1}}{r!} 1_{[(\lambda-\lambda')p_a > \epsilon]} \right) \\
= &\sum_{a\in\mathcal{A}} \left((r+1+\lambda p_a) e^{-\lambda' p_a} (e^{(\lambda'-\lambda)p_a} e^{-(\lambda'-\lambda)p_a}) \frac{(\lambda p_a)^{r+1}}{r!} 1_{[(\lambda-\lambda')p_a \le \epsilon]} \right) \\
&+ \sum_{a\in\mathcal{A}} \left((r+1+\lambda p_a) e^{-\lambda' p_a} \frac{(\lambda p_a)^{r+1}}{r!} 1_{[(\lambda-\lambda')p_a > \epsilon]} \right) \\
= &\sum_{a\in\mathcal{A}} \left((r+1+\lambda p_a) e^{-\lambda p_a} e^{(\lambda-\lambda')p_a} \frac{(\lambda p_a)^{r+1}}{r!} 1_{[(\lambda-\lambda')p_a \le \epsilon]} \right) \\
&+ \sum_{a\in\mathcal{A}} \left((r+1+\lambda p_a) e^{-\lambda' p_a} \frac{(\lambda p_a)^{r+1}}{r!} 1_{[(\lambda-\lambda')p_a > \epsilon]} \right) \\
\le &\sum_{a\in\mathcal{A}} \left((r+1+\lambda p_a) e^{-\lambda p_a} e^{\epsilon} \frac{(\lambda p_a)^{r+1}}{r!} 1_{[(\lambda-\lambda')p_a \le \epsilon]} \right)
\end{aligned}
$$

$$
\begin{aligned}
&+\sum_{a\in\mathcal{A}}\left((r+1+\lambda p_a)e^{-\lambda' p_a}\frac{(\lambda p_a)^{r+1}}{r!}1_{[(\lambda-\lambda')p_a>\epsilon]}\right)\\
\leq& e^{\epsilon}\sum_{a\in\mathcal{A}}\left((r+1+\lambda p_a)e^{-\lambda p_a}\frac{(\lambda p_a)^{r+1}}{r!}\right)\\
&+\sum_{a\in\mathcal{A}}\left((r+1+\lambda p_a)e^{-\lambda' p_a}\frac{(\lambda p_a)^{r+1}}{r!}1_{[(\lambda-\lambda')p_a>\epsilon]}\right)\\
=& e^{\epsilon}s_\lambda^2+\sum_{a\in\mathcal{A}}\left((r+1+\lambda p_a)e^{-\lambda' p_a}\frac{(\lambda p_a)^{r+1}}{r!}1_{[(\lambda-\lambda')p_a>\epsilon]}\right)\\
\leq& e^{\epsilon}s_\lambda^2+\sum_{a\in\mathcal{A}}\left((r+1+\lambda p_a)e^{-\frac{\epsilon}{\lambda-\lambda'}\lambda'}(\lambda p_a)^{r+1}1_{[(\lambda-\lambda')p_a>\epsilon]}\right)\ (\text{Note: } p_a>\tfrac{\epsilon}{\lambda-\lambda'})\\
\leq& e^{\epsilon}s_\lambda^2+\sum_{a\in\mathcal{A}}\left((r+1+\lambda)e^{-\frac{\lambda'\epsilon}{\lambda-\lambda'}}(\lambda p_a)^{r+1}1_{[(\lambda-\lambda')p_a>\epsilon]}\right)\\
=& e^{\epsilon}s_\lambda^2+(r+1+\lambda)\lambda^{r+1}e^{-\frac{\lambda'\epsilon}{\lambda-\lambda'}}\sum_{a\in\mathcal{A}}\left((p_a)^{r+1}1_{[(\lambda-\lambda')p_a>\epsilon]}\right)\\
\leq& e^{\epsilon}s_\lambda^2+(r+1+\lambda)\lambda^{r+1}e^{-\frac{\lambda'\epsilon}{\lambda-\lambda'}}\sum_{a\in\mathcal{A}}(p_a)^{r+1}\\
\leq& e^{\epsilon}s_\lambda^2+(r+1+\lambda)\lambda^{r+1}e^{-\frac{\lambda'\epsilon}{\lambda-\lambda'}}\sum_{a\in\mathcal{A}}p_a\\
=& e^{\epsilon}s_\lambda^2+(r+1+\lambda)\lambda^{r+1}e^{-\frac{\lambda'\epsilon}{\lambda-\lambda'}}.
\end{aligned}
$$

This gives (6.16).

By (6.16), we have

$$
(\frac{\lambda'_n}{\lambda_n})^{r+2}s_{\lambda_n}^2\leq s_{\lambda'_n}^2\leq e^{\epsilon}s_{\lambda_n}^2+(r+1+\lambda_n)\lambda_n^{r+1}e^{-\frac{\lambda'_n\epsilon}{\lambda_n-\lambda'_n}}. \tag{6.17}
$$

Since $\liminf_n s_{\lambda_n}>0$, dividing $s_{\lambda_n}^2$ from each side of (6.17) gives

$$
(\frac{\lambda'_n}{\lambda_n})^{r+2}\leq\frac{s_{\lambda'_n}^2}{s_{\lambda_n}^2}\leq e^{\epsilon}+\frac{1}{s_{\lambda_n}^2}(r+1+\lambda_n)\lambda_n^{r+1}e^{-\frac{\epsilon}{\frac{\lambda_n}{\lambda'_n}-1}}\quad\forall\epsilon>0. \tag{6.18}
$$

By assuming that $\lambda_n\sim\lambda'_n$, the first half of (6.18) gets

$$
\lim_{n\to\infty}\frac{s_{\lambda'_n}^2}{s_{\lambda_n}^2}\geq 1. \tag{6.19}
$$

Now we turn to the second half of (6.18).

Fix $\epsilon' > 0$, we can choose an $\epsilon > 0$ such that

$$e^{\epsilon} \leq 1 + \frac{\epsilon'}{2}. \tag{6.20}$$

By assuming that $\limsup_n (\frac{\lambda_n}{\lambda'_n} - 1)\lambda_n^{\delta} < \infty$ for some $\delta > 0$, there exists an $L > 0$ such that for large enough n,

$$e^{-\frac{\epsilon \lambda_n^{\delta}}{(\frac{\lambda_n}{\lambda'_n} - 1)\lambda_n^{\delta}}} \leq e^{-\frac{\epsilon \lambda_n^{\delta}}{2L}}.$$

So we have

$$\begin{aligned} &\frac{1}{s_{\lambda_n}^2}(r + 1 + \lambda_n)\lambda_n^{r+1} e^{-\frac{\epsilon}{\frac{\lambda_n}{\lambda'_n} - 1}} \\ \leq &\frac{1}{s_{\lambda_n}^2}(r + 1 + \lambda_n)\lambda_n^{r+1} e^{-\frac{\epsilon \lambda_n^{\delta}}{2L}}. \end{aligned}$$

Since we assume that $\liminf_n s_{\lambda_n} > 0$,

$$\limsup_n \frac{1}{s_{\lambda_n}^2} < \infty.$$

Then for such ϵ and δ,

$$\lim_{n\to\infty} \frac{1}{s_{\lambda_n}^2}(r + 1 + \lambda_n)\lambda_n^{r+1} e^{-\frac{\epsilon \lambda_n^{\delta}}{2L}} = 0. \tag{6.21}$$

Since (6.21) holds, there exists an $N_{\epsilon,\epsilon'} > 0$ such that if $n \geq N_{\epsilon,\epsilon'}$,

$$\frac{1}{s_{\lambda_n}^2}(r + 1 + \lambda_n)\lambda_n^{r+1} e^{-\frac{\epsilon \lambda_n^{\delta}}{2L}} \leq \frac{\epsilon'}{2}. \tag{6.22}$$

Combining (6.20) and (6.22) gives

$$\begin{aligned}
&\lim_{n\to\infty}\left(e^{\epsilon}+\frac{1}{s_{\lambda_n}^2}(r+1+\lambda_n)\lambda_n^{r+1}e^{-\frac{\epsilon}{\frac{\lambda_n}{\lambda_n'}-1}}\right)\\
=&\lim_{n\to\infty}e^{\epsilon}+\lim_{n\to\infty}\left(\frac{1}{s_{\lambda_n}^2}(r+1+\lambda_n)\lambda_n^{r+1}e^{-\frac{\epsilon}{\frac{\lambda_n}{\lambda_n'}-1}}\right)\\
\leq&(1+\frac{\epsilon'}{2})+\frac{\epsilon'}{2}\\
\leq&1+\epsilon'.
\end{aligned}$$

Since ϵ' is arbitrary, we get

$$\lim_{n\to\infty}\frac{s_{\lambda_n'}^2}{s_{\lambda_n}^2}\leq 1 \quad \forall\epsilon'>0. \tag{6.23}$$

Combining (6.19) and (6.23) gets

$$\lim_{n\to\infty}\frac{s_{\lambda_n'}^2}{s_{\lambda_n}^2}=1 \quad (\text{i.e.,} s_{\lambda_n'}^2\sim s_{\lambda_n}^2),$$

then

$$s_{\lambda_n'}\sim s_{\lambda_n},$$

which completes the proof. □

Lemma 10. *If* $s_n\to\infty$ *and*

$$\frac{s_n}{\sqrt{n}}\to 0,$$

then

$$\frac{|\xi_n-\zeta_n|}{s_n}\xrightarrow{p}0.$$

Proof. Fix $\epsilon, \delta > 0$. We must show that there exists a $K > 0$ such that, if $n \geq K$ then

$$P\left(|\xi_n - \zeta_n| > s_n\epsilon\right) < \delta.$$

Fix $\Delta_n = \sqrt{\frac{8n}{\delta}}$. Let t_n be the nth arrival time of the Poisson process N. Thus $N_{t_n} = n$. Note that $y'_a(n) = y_a(t_n)$. It follows that

$$\begin{aligned}
&\xi_n - \zeta_{t_n} \\
&= \sum_{a\in\mathcal{A}} \left(\left((r+1)1_{[y'_a(n)=r+1]} - np_a 1_{[y'_a(n)=r]}\right) - \left((r+1)1_{[y'_a(n)=r+1]} - t_n p_a 1_{[y'_a(n)=r]}\right)\right) \\
&= (t_n - n)\sum_{a\in\mathcal{A}} p_a 1_{[y'_a(n)=r]}.
\end{aligned}$$

Further, on the event $[|t_n - n| \leq \frac{\Delta_n}{2}]$,

$$\begin{aligned}
|\xi_n - \zeta_n| &\leq |\xi_n - \zeta_{t_n}| + |\zeta_{t_n} - \zeta_n| \\
&= |t_n - n|\sum_{a\in\mathcal{A}} p_a 1_{[y'_a(n)=r]} + |\zeta_{t_n} - \zeta_n| \\
&\leq (0.5)\Delta_n \sum_{a\in\mathcal{A}} p_a 1_{[y'_a(n)=r]} + \sup_{n-\frac{\Delta_n}{2}\leq t\leq n+\frac{\Delta_n}{2}} |\zeta_t - \zeta_n|.
\end{aligned}$$

We have

$$\begin{aligned}
&P\left(|\xi_n - \zeta_n| > s_n\epsilon\right) \\
=&P\left(|\xi_n - \zeta_n| > s_n\epsilon, |t_n - n| > \frac{\Delta_n}{2}\right) \\
&+ P\left(|\xi_n - \zeta_n| > s_n\epsilon, |t_n - n| \leq \frac{\Delta_n}{2}\right) \\
\leq&P\left(|t_n - n| > \frac{\Delta_n}{2}\right) \\
&+ P\left(\left((0.5)\Delta_n \sum_{a\in\mathcal{A}} p_a 1_{[y'_a(n)=r]} + \sup_{n-\frac{\Delta_n}{2}\leq t\leq n+\frac{\Delta_n}{2}} |\zeta_t - \zeta_n|\right) > s_n\epsilon\right).
\end{aligned}$$

Since t_n has a gamma distribution with both mean and variance n, it follows that, by Chebyshev's inequality,

$$P\left(|t_n - n| > .5\Delta_n\right) \leq 4\frac{n}{\Delta_n^2} = \frac{\delta}{2}.$$

By Markov's inequality,

$$\begin{aligned}
&P\left(\left((0.5)\Delta_n \sum_{a\in\mathcal{A}} p_a 1_{[y_a'(n)=r]} + \sup_{n-\frac{\Delta_n}{2}\leq t\leq n+\frac{\Delta_n}{2}} |\zeta_t - \zeta_n|\right) > s_n\epsilon\right)\\
&\leq \epsilon^{-1}s_n^{-1}\mathrm{E}\left[\sup_{n-\frac{\Delta_n}{2}\leq t\leq n+\frac{\Delta_n}{2}} |\zeta_t - \zeta_n| + (0.5)\Delta_n \sum_{a\in\mathcal{A}} p_a 1_{[y_a'(n)=r]}\right]\\
&= \epsilon^{-1}s_n^{-1}\mathrm{E}\left[\sup_{n-\frac{\Delta_n}{2}\leq t\leq n+\frac{\Delta_n}{2}} |\zeta_t - \zeta_n|\right] + \epsilon^{-1}s_n^{-1}\mathrm{E}\left[(0.5)\Delta_n \sum_{a\in\mathcal{A}} p_a 1_{[y_a'(n)=r]}\right].
\end{aligned}$$

Since for large enough n, we have $\Delta \in (0, n)$. It follows from Lemma 8 that

$$\begin{aligned}
&s_n^{-1}\mathrm{E}\left[\sup_{n-\frac{\Delta_n}{2}\leq t_n\leq n+\frac{\Delta_n}{2}} |\zeta_{t_n,n} - \zeta_{n,n}|\right]\\
&\leq s_n^{-1} 2H(\lambda - \frac{\Delta_n}{2}, \Delta_n)\\
&= 2Cs_n^{-1}\frac{\Delta_n}{n - \Delta_n/2} s^2_{n-\Delta_n/2,n}\\
&\sim 2C\sqrt{8/\delta}\frac{1}{\sqrt{n}} s_n \to 0,
\end{aligned}$$

where $s_{n-\Delta_n/2,n} \sim s_n$ by Lemma 9. We just need to verify that the assumptions of that lemma hold.

Let $\lambda_n' = n - \frac{\Delta_n}{2}$, then

$$H(n - \frac{\Delta_n}{2}, \Delta_n) = H(\lambda_n', \Delta_n),$$

and

$$\frac{\lambda'_n}{n} = \frac{n - \frac{\Delta_n}{2}}{n} = 1 - \frac{\frac{\Delta_n}{2}}{n}.$$

Since $\Delta_n = \sqrt{\frac{8n}{\delta}}$,

$$\lim_{n\to\infty} \frac{\frac{\Delta_n}{2}}{n} = \lim_{n\to\infty} \sqrt{\frac{8/\delta}{n}} = 0.$$

Then

$$\lim_{n\to\infty} \frac{\lambda'_n}{n} = 1 - \lim_{n\to\infty} \frac{\frac{\Delta_n}{2}}{n} = 1,$$

(i.e. $\lambda'_n \sim n$). Since

$$\begin{aligned}(\frac{n}{\lambda'_n} - 1)n^{\delta'} &= \frac{n^{\delta'}\Delta_n}{2n - \Delta_n} \\ &= \frac{n^{\delta'}k\sqrt{n}}{n - k\sqrt{n}} \quad \text{(Note: let } k = \sqrt{2/\delta}\text{)} \\ &= \frac{k}{2n^{1/2-\delta'} - kn^{-\delta'}},\end{aligned}$$

if we fix $\delta' \in (0, 1/2)$,

$$\lim_{n\to\infty} (\frac{n}{\lambda'_n} - 1)n^{\delta'} = 0.$$

Thus, there exists an $\delta' > 0$ such that $\limsup_n(\frac{n}{\lambda'_n} - 1)n^{\delta'} < \infty$.

Now $0 < \lambda'_n < n < \infty$, $\lambda'_n \sim n$ and $\limsup_n(\frac{n}{\lambda'_n} - 1)n^{\delta'} < \infty$ for $\delta' \in (0, 1/2)$ satisfy the conditions of Lemma 9.

Since

$$\lim_{n\to\infty} \frac{s_{\lambda'_n}}{\sqrt{n}} = \lim_{n\to\infty} \left(\frac{s_n}{s_n}\frac{s_{\lambda'_n}}{\sqrt{n}}\right) = \lim_{n\to\infty} \frac{s_{\lambda'_n}}{s_n} \lim_{n\to\infty} \frac{s_n}{\sqrt{n}},$$

by Lemma 9 ($s_{\lambda_n'} \sim s_n$) and the assumption $\frac{s_n}{\sqrt{n}} \to 0$ we have

$$\frac{s_{\lambda_n'}}{\sqrt{n}} \to 0.$$

Now, note that

$$\begin{aligned}
& s_n^{-1} \mathrm{E}\left[(0.5)\Delta_n \sum_{a\in\mathcal{A}} p_a 1_{[y_a'(n)=r]}\right] \\
&= (0.5) s_n^{-1} \Delta_n \sum_{a\in\mathcal{A}} \binom{n}{r} p_a^{r+1} (1-p_a)^{n-r} \\
&\sim (0.5) s_n^{-1} \Delta_n \frac{n^r}{r!} \sum_{a\in\mathcal{A}} p_a^{r+1} (1-p_a)^{n-r} \\
&\le (0.5) s_n^{-1} \Delta_n \frac{n^r}{r!} \sum_{a\in\mathcal{A}} p_a^{r+1} e^{-(n-r)p_a} \\
&= (0.5) s_n^{-1} \frac{\Delta_n}{n} \sum_{a\in\mathcal{A}} n p_a \frac{(np_a)^r}{r!} e^{-np_a} e^{rp_a} \\
&\le (0.5) s_n^{-1} \frac{\Delta_n}{n} \sum_{a\in\mathcal{A}} n p_a \frac{(np_a)^r}{r!} e^{-np_a} e^{r} \\
&\le (0.5) e^r \frac{\Delta_n}{n} s_n^{-1} \sum_{a\in\mathcal{A}} (r+1+np_a) e^{-np_a} \frac{(np_a)^{r+1}}{r!} \\
&= (0.5) e^r \frac{\Delta_n}{n} s_n^{-1} s_n^2 \\
&= (0.5) e^r \frac{\Delta_n}{n} s_n \to 0,
\end{aligned}$$

where the third line follows by

$$\frac{\binom{n}{r}}{\frac{n^r}{r!}} = \frac{n!}{(n-r)!n^r} = \frac{n(n-1)...(n-r+1)}{n^r} \to 1$$

(i.e., $\binom{n}{r} \sim \frac{n^r}{r!}$), the fourth line follows by the fact that $(1-x) \le e^{-x}$, and the last line follows by $\Delta_n \sim M_1\sqrt{n}$ and

$$\frac{\Delta_n}{n} s_n = \frac{M_1\sqrt{n}}{M_1\sqrt{n}} \frac{\Delta_n}{n} s_n = \frac{\Delta_n}{M_1\sqrt{n}} \frac{M_1}{\sqrt{n}} s_n \to 0.$$

□

Proof of Theorem 4. Recall that

$$\begin{aligned} s_n^2 &= \sum_{a\in\mathcal{A}} (r+1+np_a)\, e^{-np_a} \frac{(np_a)^{r+1}}{r!} = (r+1)^2 E\left[N_{r+1}\right] + (r+2)\,(r+1)\,E\left[N_{r+2}\right] \\ T'_r(n) &= \frac{N'_{r+1}(n)}{n}(r+1) \\ \pi'_r(n) &= \sum_{a\in\mathcal{A}} p_a 1_{[y'_a(n)=r]} \\ N'_r &= N'_r(n) = \sum_{a\in\mathcal{A}} 1_{[y'_a(n)=r]}(\text{ the deterministic case}) \\ N_r &= N_r(n) = \sum_{a\in\mathcal{A}} 1_{[y_a(n)=r]}(\text{ the Poissonized case}). \end{aligned}$$

Note that

$$\frac{\xi_n}{s_n} = \frac{\xi_n - \zeta_n}{s_n} + \frac{\zeta_n}{s_n},$$

where

$$\begin{aligned} \zeta_n &= n(T_r(n) - \pi_r(n)) \\ T_r(\lambda) &= \frac{N_{r+1}(\lambda)}{\lambda}(r+1) \\ N_r(\lambda) &= \sum_{a\in\mathcal{A}} 1_{[y_a(\lambda)=r]}(\text{the Poissonized case}) \\ \pi_r(\lambda) &= \sum_{a\in\mathcal{A}} p_a 1_{[y_a(\lambda)=r]}. \end{aligned}$$

By Theorem 1, (2.5) holds if and only if

$$\frac{\zeta_n}{s_n} = \frac{\lambda(T_r(n) - \pi_r(n))}{s_n} \xrightarrow[n\to\infty]{d} N(0,1). \tag{6.24}$$

Since $s_n \to \infty$ as $n \to \infty$ and

$$\frac{s_n}{\sqrt{n}} \to 0,$$

Lemma 10 implies that

$$\frac{\xi_n - \zeta_n}{s_n} \xrightarrow{p} 0.$$

Therefore, by Slutsky's theorem, (2.5) if and only if

$$\frac{\xi_n}{s_n} \xrightarrow{d} N(0,1).$$

□

Lemma 11. *For $c, d \geq 0$, let*

$$S = c\mathrm{E}[N_{r+1}] + d\mathrm{E}[N_{r+2}] \tag{6.25}$$

and

$$T = c\mathrm{E}[N'_{r+1}] + d\mathrm{E}[N'_{r+2}]. \tag{6.26}$$

1. For any $\epsilon \in (0, 1/2)$,

$$A_n(S - B_n) \leq T \leq Se^{\epsilon(r+1)} + n^{r+2}(c+d)e^{-\epsilon(n-r-2)},$$

where $0 \leq A_n \to 1$ and $0 \leq B_n \to 0$ as $n \to \infty$ may depend on ϵ.

2. We have $T \to \infty$ if and only if $S \to \infty$.

3. If $S \to \infty$, then $T/S \to 1$.

Proof. Parts 2 and 3 follow immediately from Part 1. We now prove Part 1. Recall that for the Poissonized case

$$\begin{aligned}
\mathrm{E}\left[N_r\right] &= \mathrm{E}\left[\sum_{a\in\mathcal{A}} 1_{[y_a(n)=r]}\right] \\
&= \sum_{a\in\mathcal{A}} \mathrm{E}\left[1_{[y_a(n)=r]}\right] \\
&= \sum_{a\in\mathcal{A}} P\left(y_a(n)=r\right) \\
&= \sum_{a\in\mathcal{A}} e^{-np_a}\frac{(np_a)^r}{r!},
\end{aligned}$$

and for the deterministic case

$$\begin{aligned}
\mathrm{E}\left[N_r'\right] &= \mathrm{E}\left[\sum_{a\in\mathcal{A}} 1_{[y_a'(n)=r]}\right] \\
&= \sum_{a\in\mathcal{A}} \mathrm{E}\left[1_{[y_a'(n)=r]}\right] \\
&= \sum_{a\in\mathcal{A}} P\left(y_a'(n)=r\right) \\
&= \sum_{a\in\mathcal{A}} \binom{n}{r} p_a^r(1-p_a)^{n-r}.
\end{aligned}$$

Now we have

$$\begin{aligned}
S &= c\sum_{a\in\mathcal{A}} e^{-np_a}\frac{(np_a)^{r+1}}{(r+1)!} + d\sum_{a\in\mathcal{A}} e^{-np_a}\frac{(np_a)^{r+2}}{(r+2)!} \\
&= \sum_{a\in\mathcal{A}} \frac{(np_a)^{r+1}}{(r+1)!}e^{-np_a}\left(c + d\frac{np_a}{r+2}\right),
\end{aligned}$$

and

$$\begin{aligned}
T &= \sum_{a\in\mathcal{A}} \left(c\binom{n}{r+1}p_a^{r+1}(1-p_a)^{n-r-1} + d\binom{n}{r+2}p_a^{r+2}(1-p_a)^{n-r-2}\right) \\
&= \sum_{a\in\mathcal{A}} \binom{n}{r+1}p_a^{r+1}(1-p_a)^{n-r-2}\left(c(1-p_a) + d\frac{n-r-1}{r+2}p_a\right)
\end{aligned}$$

$$\begin{aligned}
&\le \sum_{a\in\mathcal{A}} \frac{(np_a)^{r+1}}{(r+1)!} e^{-np_a}\left(c + d\frac{np_a}{r+2}\right) e^{p_a(r+2)} \\
&\le \sum_{a\in\mathcal{A},p_a\le\epsilon} \frac{(np_a)^{r+1}}{(r+1)!} e^{-np_a}\left(c + d\frac{np_a}{r+2}\right) e^{\epsilon(r+2)} \\
&\quad + n^{r+2}\sum_{a\in\mathcal{A},p_a>\epsilon} p_a(c+d)e^{-\epsilon(n-r-2)} \\
&\le S e^{\epsilon(r+2)} + n^{r+2}(c+d)e^{-\epsilon(n-r-2)},
\end{aligned}$$

where we use the facts that $\binom{n}{r} \le \frac{n^r}{r!}$ and $(1-x) \le e^{-x}$. Next, fix $\delta \in (\frac{1}{2}, 1)$. Using the facts that $(1-x) \ge e^{-x/(1-x)}$ for $x > 0$, $\binom{n}{r+2} = \binom{n}{r+1}\frac{n-r-1}{r+2}$ and $\binom{n}{r+1} \sim \frac{n^{r+1}}{(r+1)!}$ we get

$$\begin{aligned}
T =& \sum_{a\in\mathcal{A}}\left(c\binom{n}{r+1}p_a^{r+1}(1-p_a)^{n-r-1} + d\binom{n}{r+2}p_a^{r+2}(1-p_a)^{n-r-2}\right) \\
=& \binom{n}{r+1}\sum_{a\in\mathcal{A}} p_a^{r+1}(1-p_a)^{n-r-2}\left(c(1-p_a) + d\frac{n-r-1}{r+2}p_a\right) \\
\ge& \binom{n}{r+1}\sum_{a\in\mathcal{A},p_a\le\epsilon/n^\delta} p_a^{r+1}e^{-np_a}e^{-\frac{p_a}{1-p_a}(p_a n-r-2)}\left(c + d\frac{n-r-1}{r+2}p_a\right)(1-p_a) \\
\ge& (1-\epsilon/n^\delta)\binom{n}{r+1}e^{-\frac{\epsilon}{n^\delta-\epsilon}(\epsilon n^{1-\delta}-r-2)}\sum_{a\in\mathcal{A},p_a\le\epsilon/n^\delta} p_a^{r+1}e^{-np_a}\left(c + d\frac{n-r-1}{r+2}p_a\right) \\
=& (1-\epsilon/n^\delta)\binom{n}{r+1}e^{-\frac{\epsilon}{n^\delta-\epsilon}(\epsilon n^{1-\delta}-r-2)}\frac{(r+1)!}{n^{r+1}}\left(c\sum_{a\in\mathcal{A},p_a\le\epsilon/n^\delta}\frac{(np_a)^{r+1}}{(r+1)!}e^{-np_a}\right. \\
&\left. + d\frac{(n-r-1)}{n}\sum_{a\in\mathcal{A},p_a\le\epsilon/n^\delta}\frac{(np_a)^{r+2}}{(r+2)!}e^{-np_a}\right) \\
=& A_n(S-B_n).
\end{aligned}$$

$$A_n = e^{-\frac{\epsilon}{n^\delta-\epsilon}(\epsilon n^{1-\delta}-r-2)}(1-\epsilon/n^\delta)\binom{n}{r+1}\frac{(r+1)!}{n^{r+1}} \to 1,$$

and

$$\begin{aligned}B_n =& d\frac{r+1}{n}\sum_{a\in\mathcal{A},p_a\le\epsilon/n^\delta}\frac{(np_a)^{r+2}}{(r+2)!}e^{-np_a}+c\sum_{a\in\mathcal{A},p_a>\epsilon/n^\delta}\frac{(np_a)^{r+1}}{(r+1)!}e^{-np_a}\\&+d\sum_{a\in\mathcal{A},p_a>\epsilon/n^\delta}\frac{(np_a)^{r+2}}{(r+2)!}e^{-np_a}\\=&B_n^{(1)}+B_n^{(2)}+B_n^{(3)}.\end{aligned}$$

We will show that $B_n\to 0$. First, let $M>0$ be a constant with $x^{r+1}e^{-x}\le M$ for $x\ge 0$, then by Dominated Convergence Theorem

$$B_n^{(1)}\le dM\sum_{a\in\mathcal{A},p_a\le\epsilon/n^\delta}p_a\to 0.$$

Next

$$B_n^{(2)}\le ce^{-\epsilon n^{1-\delta}}n^{r+1}\sum_{a\in\mathcal{A}}p_a\to 0,$$

and similarly $B_n^{(3)}\to 0$. □

Proof of Corollary 5. Since $(r+1)^2>0$ and $(r+2)(r+1)>0$, if $s_n\to\infty$ as $n\to\infty$, applying part 3 of Lemma 11 completes the proof. □

Proof of Corollary 6. Since we have Theorem 4 and Corollary 5, the proof can be completed by applying the Slutsky's theorem. □

Proof of Corollary 7. In the proof of Corollary 1 (6.9) we showed that $\mathrm{E}[N_{r+1}]\to\infty$. Now let $c=1$ and $d=0$ in (6.25) and (6.26) of Lemma 11, then applying part 2 of Lemma 11 gives

$$\mathrm{E}[N'_{r+1}]\to\infty,\tag{6.27}$$

and applying part 3 of Lemma 11 gives

$$\mathrm{E}[N'_{r+1}] \sim \mathrm{E}[N_{r+1}]. \tag{6.28}$$

In Lemma 2 we also showed that

$$\frac{\mathrm{E}[N_{r+1}]}{s_\lambda} \to \infty,$$

here we set $\lambda = n$ and get

$$\frac{\mathrm{E}[N_{r+1}]}{s_n} \to \infty. \tag{6.29}$$

By Corollary 5 $(s'_n)^2 \sim s_n^2$, thus together with (6.29) and (6.28) we have

$$\frac{\mathrm{E}[N'_{r+1}]}{s'_n} \xrightarrow{p} \infty. \tag{6.30}$$

Since (6.27) holds, applying part 2 of Lemma 12 gives

$$\frac{N'_{r+1}}{\mathrm{E}[N'_{r+1}]} \xrightarrow{p} 1. \tag{6.31}$$

Since

$$\frac{N'_{r+1}}{s'_n} = \frac{\mathrm{E}[N'_{r+1}]}{s'_n} \frac{N'_{r+1}}{\mathrm{E}[N'_{r+1}]},$$

by continuous mapping theorem (6.30) and (6.31) implies that

$$\frac{N'_{r+1}}{s'_n} \xrightarrow{p} \infty.$$

Since $r+1 \in (0,\infty)$,

$$\frac{N'_{r+1}(r+1)}{s'_n} \xrightarrow{p} \infty.$$

Now plugging in

$$T'_r\left(n\right) = \frac{N'_{r+1}}{n}\left(r+1\right),$$

$$\frac{nT'_r(n)}{s'_n} = \frac{N'_{r+1}(r+1)}{s'_n} \xrightarrow{p} \infty.$$

By the symmetry of Normal distribution, (2.6) implies that

$$\frac{n}{s'_n}\left(\pi'_r(n) - T'_r(n)\right) \xrightarrow[n\to\infty]{d} N(0,1),$$

and so

$$\frac{nT'_r(n)}{s'_n}\left(\frac{\pi'_r(n)}{T'_r(n)} - 1\right) \xrightarrow[n\to\infty]{d} N(0,1).$$

Since

$$\frac{nT'_r(n)}{s'_n} \xrightarrow{p} \infty,$$

it follows from Lemma 1 that

$$\frac{\pi'_r(n)}{T'_r(n)} - 1 \xrightarrow{p} 0.$$

Therefore,

$$\frac{T_r'(n)}{\pi_r'(n)} - 1 \xrightarrow{p} 0.$$

□

Lemma 12. *1. For any $k \leq n/2$, we have*

$$\mathrm{Var}(N_k') \leq A_{k,n}\mathrm{E}[N_k'],$$

where $A_{k,n} = \left(4k^{k+1}\binom{n-k}{k}(n-2k)^{-k} + 1\right) \to \frac{4k^k}{(k-1)!} + 1.$

2. If $\mathrm{E}[N_k'] \to \infty$*, then*

$$\frac{\mathrm{Var}[N_k']}{(\mathrm{E}[N_k'])^2} \xrightarrow{p} 0,$$

and

$$\frac{N_k'}{\mathrm{E}[N_k']} \xrightarrow{p} 1.$$

To show this, we use ideas from the proof of Theorem 3.3 in [24]. Part 2 can also be found without proof in Section 4 of [32].

Proof. First note that, for any $1 \leq k \leq n/2$,

$$(N_k')^2 = \sum_{a\in\mathcal{A}} \sum_{b\in\mathcal{A}, a\neq b} 1_{[y_a'=k]} 1_{[y_b'=k]} + N_k',$$

and

$$\mathrm{E}[(N_k')^2] = \binom{n}{k, k, n-2k} \sum_{a\in\mathcal{A}} \sum_{b\in\mathcal{A}, a\neq b} p_a^k p_b^k (1 - p_a - p_b)^{n-2k} + \mathrm{E}[N_k'].$$

Next, let $B_{k,n} = \binom{n}{k,k,n-2k}/\binom{n}{k}^2 = \binom{n-k}{k}/\binom{n}{k} \leq 1$ and note that $B_{k,n} \to 1$. We have

$$\begin{aligned}\mathrm{Var}(N_k') &= \mathrm{E}[(N_k')^2] - \mathrm{E}[N_k'] - B_{k,n}(\mathrm{E}[N_k'])^2 + (B_{k,n}-1)(\mathrm{E}[N_k'])^2 + \mathrm{E}[N_k'] \\ &\leq \mathrm{E}[(N_k')^2] - \mathrm{E}[N_k'] - B_k(\mathrm{E}[N_k'])^2 + \mathrm{E}[N_k'].\end{aligned}$$

We can upper bound $\mathrm{E}[(N_k')^2] - \mathrm{E}[N_k'] - B_k(\mathrm{E}[N_k'])^2$ by

$$\begin{aligned}&\binom{n}{k,k,n-2k}\sum_{a\in\mathcal{A}}\sum_{b\in\mathcal{A}} p_a^k p_b^k \left((1-p_a-p_b)^{n-2k} - (1-p_a)^{n-k}(1-p_b)^{n-k}\right) \\ &\leq \binom{n}{k,k,n-2k}\sum_{a\in\mathcal{A}}\sum_{b\in\mathcal{A}} p_a^k p_b^k \left((1-p_a)^{n-2k}(1-p_b)^{n-2k} - (1-p_a)^{n-k}(1-p_b)^{n-k}\right) \\ &\leq k\binom{n}{k,k,n-2k}\sum_{a\in\mathcal{A}}\sum_{b\in\mathcal{A}} p_a^k p_b^k (1-p_a)^{n-2k}(1-p_b)^{n-2k}(p_a+p_b) \\ &\leq 2k\binom{n}{k,k,n-2k}\sum_{a\in\mathcal{A}}\sum_{b\in\mathcal{A},p_a\leq p_b} p_a^k p_b^{k+1}(1-p_a)^{n-k}(1-p_b)^{n-3k} \\ &\quad + 2k\binom{n}{k,k,n-2k}\sum_{a\in\mathcal{A}}\sum_{b\in\mathcal{A},p_a>p_b} p_a^{k+1} p_b^k(1-p_a)^{n-3k}(1-p_b)^{n-k} \\ &\leq 4k\binom{n}{k,k,n-2k}\sum_{a\in\mathcal{A}} p_a^k(1-p_a)^{n-k}\sum_{b\in\mathcal{A}} p_b^{k+1}(1-p_b)^{n-3k} \\ &= 4k\binom{n-k}{k}\mathrm{E}[N_k']\sum_{b\in\mathcal{A}} p_b^{k+1}(1-p_b)^{n-3k} \leq 4k^{k+1}\binom{n-k}{k}\mathrm{E}[N_k'](n-2k)^{-k}.\end{aligned}$$

Here the third line uses the facts that $1-p_a-p_b \leq 1-p_a-p_b+p_ap_b = (1-p_a)(1-p_b)$, that $1-(1-p_a)^k(1-p_b)^k \leq 1-(1-p_a-p_b)^k$, and that $1-(1-x)^k \leq kx$ for $x \in [0,1]$, which is easily checked by induction on k. The last inequality follows by the fact that $x^k(1-x)^{n-3k} \leq k^k(n-2k)^{-k}$ for $x \in [0,1]$, which can be shown using standard calculus arguments.

For the second part, by Chebyshev's inequality, it suffices to show that $\frac{\mathrm{Var}(N_k')}{(\mathrm{E}[N_k'])^2} \to 0$. The first part implies that

$$\frac{\mathrm{Var}(N_k')}{(\mathrm{E}[N_k'])^2} \leq A_{k,n}\frac{1}{\mathrm{E}[N_k']} \to 0.$$

This holds since $A_{k,n} \to \frac{4k^k}{(k-1)!} + 1$, which follows by the fact that $\binom{n}{k} \sim \frac{n^k}{k!}$.

$\square$

Lemma 13. *Assume that at least one of* $\mathrm{E}[N'_{r+1}] \to \infty$ *or* $\mathrm{E}[N'_{r+2}] \to \infty$ *holds. In the deterministic case for* $c, d > 0$ *let*

$$T = c\mathrm{E}[N'_{r+1}] + d\mathrm{E}[N'_{r+2}]$$
$$\hat{T} = cN'_{r+1} + dN'_{r+2}.$$

$\hat{T}$ *is a consistent estimator of* T*, i.e., as* $n \to \infty$*, for all* $\epsilon > 0$

$$P\left(\left|\frac{\hat{T}}{T} - 1\right| > \epsilon\right) \to 0.$$

Proof. Note that $\mathrm{E}[\hat{T}] = T$. Chebyshev's inequality implies that for all $\epsilon > 0$

$$\begin{aligned} P\left(\left|\frac{\hat{T}}{T} - 1\right| > \epsilon\right) &\leq \frac{\mathrm{Var}\left[\frac{\hat{T}}{T}\right]}{\epsilon^2} \\ &= \frac{\mathrm{Var}[\hat{T}]}{\epsilon^2 T^2} \\ &= \frac{\mathrm{Var}[\hat{T}]}{\epsilon^2 (\mathrm{E}[\hat{T}])^2}. \end{aligned}$$

By plugging in $\hat{T}$ and T, we obtain

$$\begin{aligned} \frac{\mathrm{Var}[\hat{T}]}{\epsilon^2(\mathrm{E}[\hat{T}])^2} &= \frac{\mathrm{Var}[cN'_{r+1} + dN'_{r+2}]}{\epsilon^2(c\mathrm{E}[N'_{r+1}] + d\mathrm{E}[N'_{r+2}])^2} \\ &= \frac{c^2\mathrm{Var}[N'_{r+1}] + d^2\mathrm{Var}[N'_{r+2}] + 2cd\mathrm{Cov}[N'_{r+1}, N'_{r+2}]}{\epsilon^2(c\mathrm{E}[N'_{r+1}] + d\mathrm{E}[N'_{r+2}])^2} \\ &\leq \frac{c^2\mathrm{Var}[N'_{r+1}] + d^2\mathrm{Var}[N'_{r+2}] + 2cd(\mathrm{Var}[N'_{r+1}] + \mathrm{Var}[N'_{r+2}])}{\epsilon^2(c\mathrm{E}[N'_{r+1}] + d\mathrm{E}[N'_{r+2}])^2} \\ &= \frac{(c^2 + 2cd)\mathrm{Var}[N'_{r+1}] + (d^2 + 2cd)\mathrm{Var}[N'_{r+2}]}{\epsilon^2(c\mathrm{E}[N'_{r+1}] + d\mathrm{E}[N'_{r+2}])^2}, \end{aligned} \tag{6.32}$$

where the third line follows by the fact that

$$\mathrm{Cov}(X, Y) \leq \mathrm{Var}(X) + \mathrm{Var}(Y).$$

Now we consider three cases.

Firstly, if both $\mathrm{E}[N'_{r+1}] \to \infty$ and $\mathrm{E}[N'_{r+2}] \to \infty$, then (6.32) can be expressed by

$$\begin{aligned}
&\frac{1}{\epsilon^2}\left(\frac{c(c+2d)\mathrm{Var}[N'_{r+1}] + d(d+2c)\mathrm{Var}[N'_{r+2}]}{(c\mathrm{E}[N'_{r+1}] + d\mathrm{E}[N'_{r+2}])^2}\right) \\
=&\frac{1}{\epsilon^2}\left(\frac{c(c+2d)\mathrm{Var}[N'_{r+1}]}{(c\mathrm{E}[N'_{r+1}] + d\mathrm{E}[N'_{r+2}])^2} + \frac{d(d+2c)\mathrm{Var}[N'_{r+2}]}{(c\mathrm{E}[N'_{r+1}] + d\mathrm{E}[N'_{r+2}])^2}\right) \\
\leq&\frac{1}{\epsilon^2}\left(\frac{(c+2d)\mathrm{Var}[N'_{r+1}]}{(c\mathrm{E}[N'_{r+1}])^2} + \frac{(d+2c)\mathrm{Var}[N'_{r+2}]}{(d\mathrm{E}[N'_{r+2}])^2}\right) \to 0,
\end{aligned}$$

where $c, d, \mathrm{E}[N'_{r+1}], \mathrm{E}[N'_{r+2}] > 0$ gives the inequality and the convergence follows from Part 2 of Lemma 12.

Secondly, assume that $\mathrm{E}[N'_{r+1}] \to \infty$, but that $\liminf \mathrm{E}[N'_{r+2}] < \infty$. Here, along any subsequence where we have convergence to infinity we can use the above result and along any subsequence were we have convergence to a finite number we have $\lim \mathrm{Var}[N'_{r+2}] < \infty$ by Part 1 of Lemma 12. In this case we can use the bound

$$P\left(\left|\frac{\hat{T}}{T} - 1\right| > \epsilon\right) \leq \frac{1}{\epsilon^2}\left(\frac{(c+2d)\mathrm{Var}[N'_{r+1}]}{(c\mathrm{E}[N'_{r+1}])^2} + \frac{d(d+2c)\mathrm{Var}[N'_{r+2}]}{(d\mathrm{E}[N'_{r+2}])^2}\right) \to 0.$$

The remaining case is similar.

□

Proof of Corollary 8. Since $(r+1)^2 > 0$ and $(r+2)(r+1) > 0$, the result is an application of Lemma 13. □

6.3.2 Proofs for Section 2.4.2

First, we explain our model containing both the Deterministic case and the Poisson case with changing distribution.

Assume that we have a countably infinite number of populations indexed by the natural numbers. Let $C = \{C_\lambda : \lambda \geq 0\}$ be a Poisson process with rate 1. Every time that this process jumps, we sample an observation from each population, where the observation from population m follows distribution $\mathcal{P}_m$.

For $n = 1, 2, \ldots$, let $t_n = min\{\lambda \geq 0 : C_\lambda = n\}$ be the time of the nth jump. If we consider the sequence of samples from population n taken at times t_n, then the size of the nth sample is n and we have the deterministic model studied in Section 2.4.2. On the other hand, if we consider the sequence of samples taken from population n at time n, then the size of the nth sample is C_n and we have the model studied in Section 2.3.2 with $\lambda_n = n$. Note that

$$C_{t_n} = n = \mathrm{E}[C_n].$$

Thus, in the two sampling schemes, we expect to have the same sample sizes, although the actual sizes may be different. When dealing with the sampling scheme with deterministic sample sizes (random sampling times) refered as the Deterministic case we use the notation from Section 2.4.2; and when dealing with the sampling scheme with random sample sizes (deterministic sampling times) refered as the Poissonized case we use the notation from Section 2.3.2. Further, we define

$$\xi_{n,n} = n(T'_{r,n}(n) - \pi'_{r,n}(n))$$

be the Deterministic version, where n letters are observed, and for $\lambda > 0$

$$\zeta_{\lambda,n} = \lambda(T_{r,n}(\lambda) - \pi_{r,n}(\lambda))$$

be the Poissonized version. Let t_n be the arrival time on the nth observation. Note that $y'_{a,n}(n) = y_{a,n}(t_n)$, and t_n follows a gamma distribution with both mean and variance n. We are going to study our estimator at time n and time t_n, and approximate its behavior at time t_n by that at time n. Observe that

$$\zeta_{t_n,n} = \frac{t_n}{n}\xi_{n,n}.$$

The idea of the proof is to transfer the asymptotic properties of $\zeta_{\lambda,n}$ to $\xi_{n,n}$ by showing that $\xi_{n,n} - \zeta_{\lambda,n} \xrightarrow{p} 0$, specifically when $\lambda = n$.

Before giving the proof of Theorem 5 for the Deterministic case with changing distribution, we prepare several lemmas.

Lemma 14. *Fix n and only consider the nth population. For any $\lambda > 0$ and $\Delta \in (0, \lambda)$, we have*

$$\mathrm{E}\left[\sup_{\lambda<t<\lambda+\Delta} |\zeta_{t,n} - \zeta_{\lambda,n}|\right] \leq H(\lambda, \Delta)$$

and

$$\mathrm{E}\left[\sup_{\lambda-\frac{\Delta}{2}<t<\lambda+\frac{\Delta}{2}} |\zeta_{t,n} - \zeta_{\lambda,n}|\right] \leq 2H(\lambda - \frac{\Delta}{2}, \Delta),$$

where for some constant $C > 0$,

$$H(\lambda, \Delta) = C\frac{\Delta}{\lambda}s^2_{\lambda,n}.$$

Proof. Recall that for any $\lambda > 0$ we have

$$\zeta_{\lambda,n} = \lambda\left(T_{r,n}(\lambda) - \pi_{r,n}(\lambda)\right) = \sum_{a\in\mathcal{A}} Y_{a,n}(\lambda),$$

and

$$Y_{a,n} = (r+1)\,1_{[y_{a,n}(\lambda)=r+1]} - \lambda p_{a,n} 1_{[y_{a,n}(\lambda)=r]}.$$

Fix $t > \lambda$ and note that $y_{a,n}(t) \geq y_{a,n}(\lambda)$ because greater arrival time yields more or equal arrivals in a Poisson process and

$$\begin{aligned}Y_{a,n}(t) - Y_{a,n}(\lambda) =& 1_{[y_{a,n}(\lambda)<r]} Y_{a,n}(t) + 1_{[y_{a,n}(\lambda)]\geq r]} Y_{a,n}(t) \\ & - Y_{a,n}(\lambda) 1_{[y_{a,n}(t)>y_{a,n}(\lambda)]} - Y_{a,n}(\lambda) 1_{[y_{a,n}(t)=y_{a,n}(\lambda)]} \\ =& - Y_{a,n}(\lambda) 1_{[y_{a,n}(t)>y_{a,n}(\lambda)]} + 1_{[y_{a,n}(\lambda)<r]} Y_{a,n}(t) \\ & - Y_{a,n}(\lambda) 1_{[y_{a,n}(t)=y_{a,n}(\lambda)]} + 1_{[y_{a,n}(\lambda)]\geq r]} Y_{a,n}(t).\end{aligned}$$

Since

$$\begin{aligned}& - Y_{a,n}(\lambda) 1_{[y_{a,n}(t)=y_{a,n}(\lambda)]} + 1_{[y_{a,n}(\lambda)]\geq r]} Y_{a,n}(t) \\ =& - (r+1) 1_{[y_{a,n}(\lambda)=r+1]} 1_{[y_{a,n}(t)=y_{a,n}(\lambda)]} + \lambda p_{a,n} 1_{[y_{a,n}(\lambda)=r]} 1_{[y_{a,n}(t)=y_{a,n}(\lambda)]} \\ & + (r+1) 1_{[y_{a,n}(t)=r+1]} 1_{[y_{a,n}(\lambda)\geq r]} - t p_{a,n} 1_{[y_{a,n}(t)=r]} 1_{[y_{a,n}(\lambda)\geq r]} \\ =& - (r+1) 1_{[y_{a,n}(\lambda)=r+1]} 1_{[y_{a,n}(t)=r+1]} + \lambda p_{a,n} 1_{[y_{a,n}(\lambda)=r]} 1_{[y_{a,n}(t)=r]} \\ & + (r+1) 1_{[y_{a,n}(t)=r+1]} 1_{[y_{a,n}(\lambda)=r+1]} \\ & + (r+1) 1_{[y_{a,n}(t)=r+1]} 1_{[y_{a,n}(\lambda)=r]} - t p_{a,n} 1_{[y_{a,n}(t)=r]} 1_{[y_{a,n}(\lambda)=r]} \\ =& (r+1) 1_{[y_{a,n}(t)=r+1]} 1_{[y_{a,n}(\lambda)=r]} - (t-\lambda) p_{a,n} 1_{[y_{a,n}(t)=r]} 1_{[y_{a,n}(\lambda)=r]} \\ =& 1_{[y_{a,n}(\lambda)=r]} ((r+1) 1_{[y_{a,n}(t)=r+1]} - (t-\lambda) p_{a,n} 1_{[y_{a,n}(t)=r]}),\end{aligned}$$

then

$$\begin{aligned}& Y_{a,n}(t) - Y_{a,n}(\lambda) \\ =& - Y_{a,n}(\lambda) 1_{[y_{a,n}(t)>y_{a,n}(\lambda)]} + 1_{[y_{a,n}(\lambda)<r]} Y_{a,n}(t) \\ & + 1_{[y_{a,n}(\lambda)=r]} ((r+1) 1_{[y_{a,n}(t)=r+1]} - (t-\lambda) p_{a,n} 1_{[y_{a,n}(t)=r]}).\end{aligned}$$

Note that

$$
\begin{aligned}
&|\sum_{a\in\mathcal{A}}(Y_{a,n}(t)-Y_{a,n}(\lambda))| \\
\leq&|\sum_{a\in\mathcal{A}}Y_{a,n}(\lambda)1_{[y_{a,n}(t)>y_{a,n}(\lambda)]}|+|\sum_{a\in\mathcal{A}}1_{[y_{a,n}(\lambda)<r]}Y_{a,n}(t)| \\
&+|\sum_{a\in\mathcal{A}}1_{[y_{a,n}(\lambda)=r]}((r+1)1_{[y_{a,n}(t)=r+1]}-(t-\lambda)p_{a,n}1_{[y_{a,n}(t)=r]})| \\
\leq&(r+1)\sum_{a\in\mathcal{A}}1_{[y_{a,n}(\lambda)=r+1]}1_{[y_{a,n}(t)>y_{a,n}(\lambda)]} \\
&+\lambda\sum_{a\in\mathcal{A}}p_{a,n}1_{[y_{a,n}(\lambda)=r]}1_{[y_{a,n}(t)>y_{a,n}(\lambda)]} \\
&+(r+1)\sum_{a\in\mathcal{A}}1_{[y_{a,n}(t)=r+1]}1_{[y_{a,n}(\lambda)<r]} \\
&+t\sum_{a\in\mathcal{A}}p_{a,n}1_{[y_{a,n}(t)=r]}1_{[y_{a,n}(\lambda)<r]} \\
&+(r+1)\sum_{a\in\mathcal{A}}1_{[y_{a,n}(\lambda)=r]}1_{[y_{a,n}(t)=r+1]} \\
&+\sum_{a\in\mathcal{A}}|t-\lambda|p_{a,n}1_{[y_{a,n}(\lambda)=r]}1_{[y_{a,n}(t)=r]}.
\end{aligned}
$$

Now set

$$
\begin{aligned}
A_t^1&=\sum_{a\in\mathcal{A}}1_{[y_{a,n}(\lambda)=r+1]}1_{[y_{a,n}(t)>y_{a,n}(\lambda)]} \\
A_t^2&=\lambda\sum_{a\in\mathcal{A}}p_{a,n}1_{[y_{a,n}(\lambda)=r]}1_{[y_{a,n}(t)>y_{a,n}(\lambda)]} \\
B_t^1&=\sum_{a\in\mathcal{A}}1_{[y_{a,n}(t)=r+1]}1_{[y_{a,n}(\lambda)<r]} \\
B_t^2&=t\sum_{a\in\mathcal{A}}p_{a,n}1_{[y_{a,n}(t)=r]}1_{[y_{a,n}(\lambda)<r]} \\
C_t&=\sum_{a\in\mathcal{A}}1_{[y_{a,n}(\lambda)=r]}1_{[y_{a,n}(t)=r+1]} \\
D_t&=\sum_{a\in\mathcal{A}}|t-\lambda|p_{a,n}1_{[y_{a,n}(\lambda)=r]}1_{[y_{a,n}(t)=r]},
\end{aligned}
$$

then

$$
\begin{aligned}
|\zeta_{t,n} - \zeta_{\lambda,n}| &= |\sum_{a\in\mathcal{A}} (Y_{a,n}(t) - Y_{a,n}(\lambda))| \\
&\leq (r+1)A_t^1 + A_t^2 + (r+1)B_t^1 + B_t^2 + (r+1)C_t + D_t.
\end{aligned}
$$

We are going to find the bounds for each element.

Bounds for C_t and D_t:

$$
\begin{aligned}
C_t &= \sum_{a\in\mathcal{A}} 1_{[y_{a,n}(\lambda)=r]} 1_{[y_{a,n}(t)=r+1]} \\
D_t &= \sum_{a\in\mathcal{A}} |t-\lambda| p_{a,n} 1_{[y_{a,n}(\lambda)=r]} 1_{[y_{a,n}(t)=r]}
\end{aligned}
$$

By Fubini's Theorem and the fact that Poisson processes have independent increments,

$$
\begin{aligned}
\mathrm{E}\left[\sup_{\lambda<t<\lambda+\Delta} C_t\right] &\leq \left[\mathrm{E} \sup_{\lambda<t<\lambda+\Delta} \sum_{a\in\mathcal{A}} 1_{[y_{a,n}(\lambda)=r]} 1_{[y_{a,n}(t)>y_{a,n}(\lambda)]}\right] \\
&\leq \mathrm{E}\left[\sum_{a\in\mathcal{A}} 1_{[y_{a,n}(\lambda)=r]} 1_{[y_{a,n}(\lambda+\Delta)>y_{a,n}(\lambda)]}\right] (\text{Note} : t < \lambda+\Delta) \\
&= \sum_{a\in\mathcal{A}} P(y_{a,n}(\lambda)=r) P(y_{a,n}(\lambda+\Delta) > y_{a,n}(\lambda)) \\
&= \sum_{a\in\mathcal{A}} \frac{\lambda^r}{r!} e^{-\lambda p_{a,n}} p_{a,n}^r (1-e^{-\Delta p_{a,n}}) \\
&= \frac{\lambda^r}{r!} \sum_{a\in\mathcal{A}} e^{-\lambda p_{a,n}} p_{a,n}^r (1-e^{-\Delta p_{a,n}}) \\
&\leq \lambda^r \sum_{a\in\mathcal{A}} e^{-\lambda p_{a,n}} p_{a,n}^r (1-e^{-\Delta p_{a,n}}) \\
&\leq \lambda^r \sum_{a\in\mathcal{A}} e^{-\lambda p_{a,n}} p_{a,n}^r \Delta p_{a,n} \\
&= \Delta\lambda^r \sum_{a\in\mathcal{A}} e^{-\lambda p_{a,n}} p_{a,n}^{r+1},
\end{aligned}
$$

where the last inequality follows by the fact that $1-e^{-x} \leq x$ for $x > 0$.

By similar arguments,

$$\begin{aligned}
\mathrm{E}\left[\sup_{\lambda<t<\lambda+\Delta} D_t\right] &\leq \mathrm{E}\left[\sum_{a\in\mathcal{A}} \Delta p_{a,n} 1_{[}y_{a,n}(\lambda)=r]\right] (\text{Note}: \Delta > t-\lambda > 0) \\
&= \Delta \sum_{a\in\mathcal{A}} p_{a,n} P(y_{a,n}(\lambda)=r) \\
&= \Delta \frac{\lambda^r}{r!} \sum_{a\in\mathcal{A}} e^{-\lambda p_{a,n}} p_{a,n}^{r+1} \\
&\leq \Delta \lambda^r \sum_{a\in\mathcal{A}} e^{-\lambda p_{a,n}} p_{a,n}^{r+1}.
\end{aligned}$$

Bound for B_t^1 and B_t^2:

$$\begin{aligned}
B_t^1 &= \sum_{a\in\mathcal{A}} 1_{[y_{a,n}(t)=r+1]} 1_{[y_{a,n}(\lambda)<r]} \\
B_t^2 &= t \sum_{a\in\mathcal{A}} p_{a,n} 1_{[y_{a,n}(t)=r]} 1_{[y_{a,n}(\lambda)<r]}
\end{aligned}$$

Clearly, if $r=0$, then

$$\mathrm{E}\left[\sup_{\lambda<t<\lambda+\Delta} B_t^1\right] = \mathrm{E}\left[\sup_{\lambda<t<\lambda+\Delta} B_t^2\right] = 0.$$

Now, assume that $r \geq 1$. Note that by independent and stationary increments

$$\begin{aligned}
\mathrm{E}\left[\sup_{\lambda<t<\lambda+\Delta} B_t^1\right] &\leq \mathrm{E}\left[\sup_{\lambda<t<\lambda+\Delta} \sum_{a\in\mathcal{A}} \sum_{i=0}^{r-1} 1_{[y_{a,n}(t)>r]} 1_{[y_{a,n}(\lambda)=i]}\right] \\
&= \mathrm{E}\left[\sup_{\lambda<t<\lambda+\Delta} \sum_{a\in\mathcal{A}} \sum_{i=0}^{r-1} 1_{[y_{a,n}(t)-y_{a,n}(\lambda)>r-i]} 1_{[y_{a,n}(\lambda)=i]}\right] \\
&\leq \sum_{a\in\mathcal{A}} \sum_{i=0}^{r-1} \mathrm{E}\left[1_{[y_{a,n}(\lambda+\Delta)-y_{a,n}(\lambda)>r-i]} 1_{[y_{a,n}(\lambda)=i]}\right] \\
&= \sum_{a\in\mathcal{A}} \sum_{i=0}^{r-1} P(y_{a,n}(\Delta) > r-i) P(y_{a,n}(\lambda)=i)
\end{aligned}$$

$$\leq \sum_{a\in\mathcal{A}}\sum_{i=0}^{r-1}\frac{(\Delta p_{a,n})^{r-i+1}}{(r-i+1)!}e^{-\lambda p_{a,n}}\frac{(p_{a,n}\lambda)^i}{i!}$$
$$\leq \Delta\lambda^r\sum_{a\in\mathcal{A}}\sum_{i=0}^{r-1}p_{a,n}^{r+1}e^{-\lambda p_{a,n}} = r\Delta\lambda^r\sum_{a\in\mathcal{A}}p_{a,n}^{r+1}e^{-\lambda p_{a,n}},$$

where we use the fact that for any integer $k \geq 0$

$$P(y_{a,n}(\Delta) > k) = 1 - \sum_{j=0}^{k}e^{-\Delta p_{a,n}}\frac{(\Delta p_{a,n})^j}{j!} \leq \frac{(\Delta p_{a,n})^{k+1}}{(k+1)!},$$

which follows since for any $x > 0$ we have $1 - e^{-x}\sum_{i=0}^{k}x^j/j! \leq x^{k+1}/(k+1)!$, see e.g. Lemma 1 in [31]. Similarly, for B_t^2 we have

$$\mathrm{E}\left[\sup_{\lambda<t<\lambda+\Delta}B_t^2\right] \leq \mathrm{E}\left[\sup_{\lambda<t<\lambda+\Delta}t\sum_{a\in\mathcal{A}}\sum_{i=0}^{r-1}p_{a,n}1_{[y_{a,n}(t)>r-1]}1_{[y_{a,n}(\lambda)=i]}\right]$$
$$\leq (\lambda+\Delta)\sum_{a\in\mathcal{A}}\sum_{i=0}^{r-1}p_{a,n}\mathrm{E}\left[1_{[y_{a,n}(\lambda+\Delta)-y_{a,n}(\lambda)>r-1-i]}1_{[y_{a,n}(\lambda)=i]}\right]$$
$$= (\lambda+\Delta)\sum_{a\in\mathcal{A}}\sum_{i=0}^{r-1}p_{a,n}P(y_{a,n}(\Delta) > r-1-i)P(y_{a,n}(\lambda) = i)$$
$$\leq 2\lambda\sum_{a\in\mathcal{A}}\sum_{i=0}^{r-1}p_{a,n}(\Delta p_{a,n})^{r-i}e^{-\lambda p_{a,n}}(\lambda p_{a,n})^i$$
$$\leq 2r\Delta\lambda^r\sum_{a\in\mathcal{A}}p_{a,n}^{r+1}e^{-\lambda p_{a,n}}.$$

Bound for A_t^1 and A_t^2:

$$A_t^1 = \sum_{a\in\mathcal{A}}1_{[y_{a,n}(\lambda)=r+1]}1_{[y_{a,n}(t)>y_{a,n}(\lambda)]}$$
$$A_t^2 = \lambda\sum_{a\in\mathcal{A}}p_{a,n}1_{[y_{a,n}(\lambda)=r]}1_{[y_{a,n}(t)>y_{a,n}(\lambda)]}$$

The proof for A_t^1 is similar to the proof for C_t. Here

$$\begin{aligned}
\mathrm{E}\left[\sup_{\lambda<t<\lambda+\Delta} A_t^1\right] &\leq \mathrm{E}\left[\sum_{a\in\mathcal{A}} 1_{[y_{a,n}(\lambda)=r+1]} 1_{[y_{a,n}(\lambda+\Delta)>y_{a,n}(\lambda)]}\right] \\
&= \sum_{a\in\mathcal{A}} P(y_{a,n}(\lambda)=r+1) P((y_{a,n}(\lambda+\Delta)-y_{a,n}(\lambda))>0) \\
&= \frac{\lambda^{r+1}}{(r+1)!} \sum_{a\in\mathcal{A}} p_{a,n}^{r+1} e^{-\lambda p_{a,n}} (1-e^{-\Delta p_{a,n}}) \\
&\leq \frac{\lambda^{r+1}}{(r+1)} \Delta \sum_{a\in\mathcal{A}} p_{a,n}^{r+2} e^{-\lambda p_{a,n}}.
\end{aligned}$$

Next, by Fubini's theorem and independent increments we have

$$\begin{aligned}
\mathrm{E}\left[\sup_{\lambda<t<\lambda+\Delta} A_t^2\right] &\leq \mathrm{E}\left[\lambda \sum_{a\in\mathcal{A}} p_{a,n} 1_{[y_{a,n}(\lambda)=r]} 1_{[y_{a,n}(\lambda+\Delta)>y_{a,n}(\lambda)]}\right] \\
&= \lambda \sum_{a\in\mathcal{A}} p_{a,n} P(y_{a,n}(\lambda)=r) P(y_{a,n}(\lambda+\Delta)>y_{a,n}(\lambda)) \\
&= \frac{\lambda^{r+1}}{r!} \sum_{a\in\mathcal{A}} p_{a,n}^{r+1} e^{-\lambda p_{a,n}} (1-e^{-\Delta p_{a,n}}) \\
&\leq \frac{\lambda^{r+1}}{r!} \Delta \sum_{a\in\mathcal{A}} p_{a,n}^{r+2} e^{-\lambda p_{a,n}} \\
&\leq \lambda^{r+1} \Delta \sum_{a\in\mathcal{A}} p_{a,n}^{r+2} e^{-\lambda p_{a,n}},
\end{aligned}$$

which completes the proof of this part. Now putting everything together gives the first bound:

$$\begin{aligned}
&\mathrm{E}\left[\sup_{\lambda<t<\lambda+\Delta} |\zeta_{t,n}-\zeta_{\lambda,n}|\right] \\
=&\mathrm{E}\left[\sup_{\lambda<t<\lambda+\Delta} \big((r+1)A_t^1 + A_t^2 + (r+1)B_t^1 + B_t^2 + (r+1)C_t + D_t\big)\right] \\
\leq&(r+1)\frac{\lambda^{r+1}}{(r+1)} \Delta \sum_{a\in\mathcal{A}} p_{a,n}^{r+2} e^{-\lambda p_{a,n}} + \lambda^{r+1} \Delta \sum_{a\in\mathcal{A}} p_{a,n}^{r+2} e^{-\lambda p_{a,n}} \\
&+ (r+1) r \Delta \lambda^r \sum_{a\in\mathcal{A}} p_{a,n}^{r+1} e^{-\lambda p_{a,n}} + 2r\Delta\lambda^r \sum_{a\in\mathcal{A}} p_{a,n}^{r+1} e^{-\lambda p_{a,n}}
\end{aligned}$$

$$
\begin{aligned}
&+ (r+1)\Delta\lambda^r \sum_{a\in\mathcal{A}} e^{-\lambda p_{a,n}} p_{a,n}^{r+1} \\
&+ \Delta\lambda^r \sum_{a\in\mathcal{A}} e^{-\lambda p_{a,n}} p_{a,n}^{r+1} \\
=&2\Delta\frac{\lambda^{r+1}}{r!} \sum_{a\in\mathcal{A}} e^{-\lambda p_{a,n}} p_{a,n}^{r+2} + (r^2+4r+2)\Delta\lambda^r \sum_{a\in A} e^{-\lambda p_{a,n}} p_{a,n}^{r+1} \\
=&\frac{\Delta}{\lambda}\left((r^2+4r+2)\lambda^{r+1} \sum_{a\in A} e^{-\lambda p_{a,n}} p_{a,n}^{r+1} + \frac{2\lambda^{r+2}}{r!} \sum_{a\in\mathcal{A}} e^{-\lambda p_{a,n}} p_{a,n}^{r+2}\right) \\
=&H(\lambda,\Delta) \\
=&C\frac{\Delta}{\lambda} s_{\lambda,n}^2,
\end{aligned}
$$

which can be upper bounded as required. From here applying the first bound twice gives

$$
\begin{aligned}
\mathrm{E}\left[\sup_{\lambda-\frac{\Delta}{2}<t<\lambda+\frac{\Delta}{2}} |\zeta_{t,n}-\zeta_{\lambda,n}|\right] \le \mathrm{E}\left[\sup_{\lambda-\frac{\Delta}{2}<t<\lambda+\frac{\Delta}{2}} |\zeta_{t,n}-\zeta_{\lambda-\Delta/2,n}|\right] + \\
\mathrm{E}\left[|\zeta_{\lambda-\Delta/2,n}-\zeta_{\lambda,n}|\right] \le 2H\left(\lambda-\frac{\Delta}{2},\Delta\right),
\end{aligned}
$$

which completes the proof. □

Recall that for the Poissonized case

$$
\begin{aligned}
s_{\lambda_n,n}^2 &= (r+1)^2\mathrm{E}[N_{r+1,n}] + (r+2)(r+1)\mathrm{E}[N_{r+2,n}] \\
&= (r+1)^2 \sum_{a\in\mathcal{A}} e^{-\lambda_n p_{a,n}} \frac{(\lambda_n p_{a,n})^{r+1}}{(r+1)!} + (r+2)(r+1) \sum_{a\in\mathcal{A}} e^{-\lambda_n p_{a,n}} \frac{(\lambda_n p_{a,n})^{r+2}}{(r+2)!}.
\end{aligned}
$$

Lemma 15. *Let $0<\lambda'<\lambda<\infty$. For any $\epsilon>0$,*

$$
(\frac{\lambda'}{\lambda})^{r+2} s_{\lambda,n}^2 \le s_{\lambda',n}^2 \le e^{\epsilon} s_{\lambda,n}^2 + (r+1+\lambda)\lambda^{r+1} e^{-\frac{\lambda'\epsilon}{\lambda-\lambda'}}. \tag{6.33}
$$

Further, let λ_n and λ'_n be two sequences of numbers. If $0<\lambda'_n<\lambda_n<\infty$, $\lambda_n \sim \lambda'_n$,

$\limsup_n(\frac{\lambda_n}{\lambda_n'} - 1)\lambda_n^\delta < \infty$ *for some* $\delta > 0$, *and* $\liminf_n s_{\lambda_n,n} > 0$, *then*

$$s_{\lambda_n,n} \sim s_{\lambda_n',n}.$$

Proof. Here we also fix n and only consider the nth population, where the distribution is fixed.

Let $0 < \lambda' < \lambda < \infty$, then

$$\begin{aligned}
(\frac{\lambda'}{\lambda})^{r+2} s_{\lambda,n}^2 &= (\frac{\lambda'}{\lambda})^{r+2} \sum_{a\in\mathcal{A}} \left((r+1+\lambda p_{a,n}) e^{-\lambda p_{a,n}} \frac{(\lambda p_{a,n})^{r+1}}{r!} \right) \\
&= \frac{(\lambda')^{r+2}}{\lambda} \sum_{a\in\mathcal{A}} \left((r+1+\lambda p_{a,n}) e^{-\lambda p_{a,n}} \frac{p_{a,n}^{r+1}}{r!} \right) \\
&= \frac{\lambda'}{\lambda} \sum_{a\in\mathcal{A}} \left((r+1+\lambda p_{a,n}) e^{-\lambda p_{a,n}} \frac{(\lambda' p_{a,n})^{r+1}}{r!} \right) \\
&= \sum_{a\in\mathcal{A}} \left(\frac{\lambda'}{\lambda}(r+1) e^{-\lambda p_{a,n}} \frac{(\lambda' p_{a,n})^{r+1}}{r!} + \lambda' p_{a,n} e^{-\lambda p_{a,n}} \frac{(\lambda' p_{a,n})^{r+1}}{r!} \right) \\
&\leq \sum_{a\in\mathcal{A}} \left((r+1) e^{-\lambda' p_{a,n}} \frac{(\lambda' p_{a,n})^{r+1}}{r!} + \lambda' p_{a,n} e^{-\lambda p_{a,n}} \frac{(\lambda' p_{a,n})^{r+1}}{r!} \right) \\
&= \sum_{a\in\mathcal{A}} \left((r+1+\lambda' p_{a,n}) e^{-\lambda' p_{a,n}} \frac{(\lambda' p_{a,n})^{r+1})}{r!} \right) \\
&= s_{\lambda',n}^2,
\end{aligned}$$

and for any $\epsilon > 0$

$$\begin{aligned}
s_{\lambda',n}^2 \leq& \sum_{a\in\mathcal{A}} \left((r+1+\lambda p_{a,n}) e^{-\lambda' p_{a,n}} \frac{(\lambda p_{a,n})^{r+1}}{r!} \right) \\
=& \sum_{a\in\mathcal{A}} \left((r+1+\lambda p_{a,n}) e^{-\lambda' p_{a,n}} \frac{(\lambda p_{a,n})^{r+1}}{r!} 1_{[(\lambda-\lambda')p_{a,n} \leq \epsilon]} \right) \\
&+ \sum_{a\in\mathcal{A}} \left((r+1+\lambda p_{a,n}) e^{-\lambda' p_{a,n}} \frac{(\lambda p_{a,n})^{r+1}}{r!} 1_{[(\lambda-\lambda')p_{a,n} > \epsilon]} \right) \\
=& \sum_{a\in\mathcal{A}} \left((r+1+\lambda p_{a,n}) e^{-\lambda' p_{a,n}} (e^{(\lambda'-\lambda)p_{a,n}} e^{-(\lambda'-\lambda)p_{a,n}}) \frac{(\lambda p_{a,n})^{r+1}}{r!} 1_{[(\lambda-\lambda')p_{a,n} \leq \epsilon]} \right)
\end{aligned}$$

$$
\begin{aligned}
&+\sum_{a\in\mathcal{A}}\left((r+1+\lambda p_{a,n})e^{-\lambda' p_{a,n}}\frac{(\lambda p_{a,n})^{r+1}}{r!}1_{[(\lambda-\lambda')p_{a,n}>\epsilon]}\right)\\
=&\sum_{a\in\mathcal{A}}\left((r+1+\lambda p_{a,n})e^{-\lambda p_{a,n}}e^{(\lambda-\lambda')p_{a,n}}\frac{(\lambda p_{a,n})^{r+1}}{r!}1_{[(\lambda-\lambda')p_{a,n}\le\epsilon]}\right)\\
&+\sum_{a\in\mathcal{A}}\left((r+1+\lambda p_{a,n})e^{-\lambda' p_{a,n}}\frac{(\lambda p_{a,n})^{r+1}}{r!}1_{[(\lambda-\lambda')p_{a,n}>\epsilon]}\right)\\
\le&\sum_{a\in\mathcal{A}}\left((r+1+\lambda p_{a,n})e^{-\lambda p_{a,n}}e^{\epsilon}\frac{(\lambda p_{a,n})^{r+1}}{r!}1_{[(\lambda-\lambda')p_{a,n}\le\epsilon]}\right)\\
&+\sum_{a\in\mathcal{A}}\left((r+1+\lambda p_{a,n})e^{-\lambda' p_{a,n}}\frac{(\lambda p_{a,n})^{r+1}}{r!}1_{[(\lambda-\lambda')p_{a,n}>\epsilon]}\right)\\
\le&e^{\epsilon}\sum_{a\in\mathcal{A}}\left((r+1+\lambda p_{a,n})e^{-\lambda p_v}\frac{(\lambda p_{a,n})^{r+1}}{r!}\right)\\
&+\sum_{a\in\mathcal{A}}\left((r+1+\lambda p_{a,n})e^{-\lambda' p_{a,n}}\frac{(\lambda p_{a,n})^{r+1}}{r!}1_{[(\lambda-\lambda')p_{a,n}>\epsilon]}\right)\\
=&e^{\epsilon}s^2_{\lambda,n}+\sum_{a\in\mathcal{A}}\left((r+1+\lambda p_{a,n})e^{-\lambda' p_{a,n}}\frac{(\lambda p_{a,n})^{r+1}}{r!}1_{[(\lambda-\lambda')p_{a,n}>\epsilon]}\right)\\
\le&e^{\epsilon}s^2_{\lambda,n}+\sum_{a\in\mathcal{A}}\left((r+1+\lambda p_{a,n})e^{-\frac{\epsilon}{\lambda-\lambda'}\lambda'}(\lambda p_{a,n})^{r+1}1_{[(\lambda-\lambda')p_{a,n}>\epsilon]}\right)\\
\le&e^{\epsilon}s^2_{\lambda,n}+\sum_{a\in\mathcal{A}}\left((r+1+\lambda)e^{-\frac{\lambda'\epsilon}{\lambda-\lambda'}}(\lambda p_{a,n})^{r+1}1_{[(\lambda-\lambda')p_{a,n}>\epsilon]}\right)\\
=&e^{\epsilon}s^2_{\lambda,n}+(r+1+\lambda)\lambda^{r+1}e^{-\frac{\lambda'\epsilon}{\lambda-\lambda'}}\sum_{a\in\mathcal{A}}\left((p_{a,n})^{r+1}1_{[(\lambda-\lambda')p_{a,n}>\epsilon]}\right)\\
\le&e^{\epsilon}s^2_{\lambda,n}+(r+1+\lambda)\lambda^{r+1}e^{-\frac{\lambda'\epsilon}{\lambda-\lambda'}}\sum_{a\in\mathcal{A}}(p_{a,n})^{r+1}\\
\le&e^{\epsilon}s^2_{\lambda,n}+(r+1+\lambda)\lambda^{r+1}e^{-\frac{\lambda'\epsilon}{\lambda-\lambda'}}\sum_{a\in\mathcal{A}}p_{a,n}\\
=&e^{\epsilon}s^2_{\lambda,n}+(r+1+\lambda)\lambda^{r+1}e^{-\frac{\lambda'\epsilon}{\lambda-\lambda'}}.
\end{aligned}
$$

This gives (6.33).

By (6.33), we have

$$
(\frac{\lambda'_n}{\lambda_n})^{r+2}s^2_{\lambda_n,n}\le s^2_{\lambda'_n,n}\le e^{\epsilon}s^2_{\lambda_n,n}+(r+1+\lambda_n)\lambda_n^{r+1}e^{-\frac{\epsilon\lambda'_n}{\lambda_n-\lambda'_n}}. \tag{6.34}
$$

Since $\liminf_n s_{\lambda_n,n} > 0$, by dividing $s^2_{\lambda_n,n}$ from each side of (6.34) we get

$$(\frac{\lambda'_n}{\lambda_n})^{r+2} \le \frac{s^2_{\lambda'_n,n}}{s^2_{\lambda_n,n}} \le e^{\epsilon} + \frac{1}{s^2_{\lambda_n,n}}(r+1+\lambda_n)\lambda_n^{r+1}e^{-\frac{\epsilon}{\frac{\lambda_n}{\lambda'_n}-1}} \quad \forall\epsilon > 0. \tag{6.35}$$

By assuming that $\lambda_n \sim \lambda'_n$, the first half of (6.35) gets

$$\liminf_n \frac{s^2_{\lambda'_n,n}}{s^2_{\lambda_n,n}} \ge 1. \tag{6.36}$$

Now we turn to the second half of (6.35).

Fix $\epsilon' > 0$, we can choose an $\epsilon > 0$ such that

$$e^{\epsilon} \le 1 + \frac{\epsilon'}{2}. \tag{6.37}$$

By assuming that $\limsup_n(\frac{\lambda_n}{\lambda'_n} - 1)\lambda_n^{\delta} < \infty$ for some $\delta > 0$, there exists an $L > 0$ such that for large enough n,

$$e^{-\frac{\epsilon\lambda_n^{\delta}}{(\frac{\lambda_n}{\lambda'_n}-1)\lambda_n^{\delta}}} \le e^{-\frac{\epsilon\lambda_n^{\delta}}{2L}}.$$

So we have

$$\begin{aligned} &\frac{1}{s^2_{\lambda_n,n}}(r+1+\lambda_n)\lambda_n^{r+1}e^{-\frac{\epsilon}{\frac{\lambda_n}{\lambda'_n}-1}} \\ \le&\frac{1}{s^2_{\lambda_n,n}}(r+1+\lambda_n)\lambda_n^{r+1}e^{-\frac{\epsilon\lambda_n^{\delta}}{2L}}. \end{aligned}$$

Since we assume that $\liminf_n s_{\lambda_n,n} > 0$,

$$\limsup_n \frac{1}{s^2_{\lambda_n,n}} < \infty.$$

Then for such ϵ and δ,

$$\lim_{n\to\infty} \frac{1}{s^2_{\lambda_n,n}}(r+1+\lambda_n)\lambda_n^{r+1}e^{-\frac{\epsilon\lambda_n^\delta}{2L}} = 0. \tag{6.38}$$

Since (6.38) holds, there exists an $N_{\epsilon,\epsilon'} > 0$ such that if $n \geq N_{\epsilon,\epsilon'}$,

$$\frac{1}{s^2_{\lambda_n,n}}(r+1+\lambda_n)\lambda_n^{r+1}e^{-\frac{\epsilon\lambda_n^\delta}{2L}} \leq \frac{\epsilon'}{2}. \tag{6.39}$$

By combining (6.37) and (6.39) we get

$$\begin{aligned}
&\lim_{n\to\infty}\left(e^{\epsilon} + \frac{1}{s^2_{\lambda_n,n}}(r+1+\lambda_n)\lambda_n^{r+1}e^{-\frac{\epsilon}{\frac{\lambda_n}{\lambda'_n}-1}}\right) \\
=&\lim_{n\to\infty} e^{\epsilon} + \lim_{n\to\infty}\left(\frac{1}{s^2_{\lambda_n,n}}(r+1+\lambda_n)\lambda_n^{r+1}e^{-\frac{\epsilon}{\frac{\lambda_n}{\lambda'_n}-1}}\right) \\
\leq&(1+\frac{\epsilon'}{2}) + \frac{\epsilon'}{2} \\
\leq&1+\epsilon'.
\end{aligned}$$

Since ϵ' is arbitrary, we get

$$\lim_{n\to\infty} \frac{s^2_{\lambda'_n,n}}{s^2_{\lambda_n,n}} \leq 1 \quad \forall\epsilon' > 0. \tag{6.40}$$

Combining (6.36) and (6.40) gets

$$\lim_{n\to\infty} \frac{s^2_{\lambda'_n,n}}{s^2_{\lambda_n,n}} = 1 \quad (\text{i.e.,} s^2_{\lambda'_n,n} \sim s^2_{\lambda_n,n}),$$

then

$$s_{\lambda'_n,n} \sim s_{\lambda_n,n},$$

which completes the proof. □

Lemma 16. *Let the waiting time* λ *be the same as the number of observations, i.e.,* $\lambda = n$*. If* $s_{\lambda,n} \to \infty$*,* $\liminf s_{\lambda,n} > 0$ *and*

$$\frac{s_{\lambda,n}}{\sqrt{n}} \to 0,$$

then

$$\frac{|\xi_{n,n} - \zeta_{\lambda,n}|}{s_{\lambda,n}} \xrightarrow{p} 0.$$

Proof. Let $\lambda = n$ and then $\zeta_{\lambda,n} = \zeta_{n,n}$. Fix $\epsilon, \delta > 0$. We must show that there exists a $K > 0$ such that, if $n \geq K$ then

$$P\left(|\xi_{n,n} - \zeta_{n,n}| > s_{\lambda,n}\epsilon\right) < \delta$$

Fix $\Delta_n = \sqrt{\frac{8n}{\delta}}$. Let t_n be the nth arrival time of the Poisson process N. Thus $N_{t_n} = n$. Note that $y'_{a,n}(n) = y_{a,n}(t_n)$. It follows that

$$\begin{aligned}
\xi_{n,n} - \zeta_{t_n,n} =& \sum_{a\in\mathcal{A}} \left((r+1)1_{\left[y'_{a,n}(n)=r+1\right]} - np_{a,n}1_{\left[y'_{a,n}(n)=r\right]}\right) \\
&- \sum_{a\in\mathcal{A}} \left((r+1)1_{\left[y'_{a,n}(n)=r+1\right]} - t_np_{a,n}1_{\left[y'_{a,n}(n)=r\right]}\right) \\
=&(t_n - n)\sum_{a\in\mathcal{A}} p_{a,n}1_{\left[y'_{a,n}(n)=r\right]}.
\end{aligned}$$

Further, on the event $[|t_n - n| \leq \frac{\Delta_n}{2}]$,

$$\begin{aligned}
|\xi_{n,n} - \zeta_{n,n}| &\leq |\xi_{n,n} - \zeta_{t_n,n}| + |\zeta_{t_n,n} - \zeta_{n,n}| \\
&= |t_n - n|\sum_{a\in\mathcal{A}} p_{a,n}1_{\left[y'_{a,n}(n)=r\right]} + |\zeta_{t_n,n} - \zeta_{n,n}| \\
&\leq (0.5)\Delta_n \sum_{a\in\mathcal{A}} p_{a,n}1_{\left[y'_{a,n}(n)=r\right]} + \sup_{n-\frac{\Delta_n}{2}\leq t\leq n+\frac{\Delta_n}{2}} |\zeta_{t,n} - \zeta_{n,n}|.
\end{aligned}$$

We have

$$
\begin{aligned}
&P\left(|\xi_{n,n}-\zeta_{n,n}|>s_{\lambda,n}\epsilon\right)\\
=&P\left(|\xi_{n,n}-\zeta_{n,n}|>s_{\lambda,n}\epsilon, |t_n-n|>\frac{\Delta_n}{2}\right)\\
&+P\left(|\xi_{n,n}-\zeta_{n,n}|>s_{\lambda,n}\epsilon, |t_n-n|\leq\frac{\Delta_n}{2}\right)\\
\leq&P\left(|t_n-n|>\frac{\Delta_n}{2}\right)\\
&+P\left(\left((0.5)\Delta_n\sum_{a\in\mathcal{A}}p_{a,n}1_{\left[y'_{a,n}(n)=r\right]}+\sup_{n-\frac{\Delta_n}{2}\leq t\leq n+\frac{\Delta_n}{2}}|\zeta_{t,n}-\zeta_{n,n}|\right)>s_{\lambda,n}\epsilon\right).
\end{aligned}
$$

Since t_n has a gamma distribution with both mean and variance n, it follows that, by Chebyshev's inequality,

$$
P\left(|t_n-n|>.5\Delta_n\right)\leq 4\frac{n}{\Delta_n^2}=\frac{\delta}{2}.
$$

By Markov's inequality,

$$
\begin{aligned}
&P\left(\left((0.5)\Delta_n\sum_{a\in\mathcal{A}}p_{a,n}1_{\left[y'_{a,n}(n)=r\right]}+\sup_{n-\frac{\Delta_n}{2}\leq t\leq n+\frac{\Delta_n}{2}}|\zeta_{t,n}-\zeta_{n,n}|\right)>s_{\lambda,n}\epsilon\right)\\
&\leq \epsilon^{-1}s_{\lambda,n}^{-1}E\left[\sup_{n-\frac{\Delta_n}{2}\leq t\leq n+\frac{\Delta_n}{2}}|\zeta_{t,n}-\zeta_{n,n}|+(0.5)\Delta_n\sum_{a\in\mathcal{A}}p_{a,n}1_{\left[y'_{a,n}(n)=r\right]}\right]\\
&=\epsilon^{-1}s_{\lambda,n}^{-1}\mathrm{E}\left[\sup_{n-\frac{\Delta_n}{2}\leq t\leq n+\frac{\Delta_n}{2}}|\zeta_{t,n}-\zeta_{n,n}|\right]+\epsilon^{-1}s_{\lambda,n}^{-1}\mathrm{E}\left[(0.5)\Delta_n\sum_{a\in\mathcal{A}}p_{a,n}1_{\left[y'_{a,n}(n)=r\right]}\right].
\end{aligned}
$$

Here we have a population with fixed n. Since for large enough n we have $\Delta\in(0,n)$, from Lemma 14 it follows that

$$
\begin{aligned}
&s_{\lambda,n}^{-1}\mathrm{E}\left[\sup_{n-\frac{\Delta_n}{2}\leq t_n\leq n+\frac{\Delta_n}{2}}|\zeta_{t_n,n}-\zeta_{n,n}|\right]\\
&\leq s_{\lambda,n}^{-1}2H(\lambda-\frac{\Delta_n}{2},\Delta_n)
\end{aligned}
$$

$$= 2Cs_n^{-1}\frac{\Delta_n}{n - \Delta_n/2}s^2_{n-\Delta_n/2,n}$$
$$\sim 2C\sqrt{8/\delta}\frac{1}{\sqrt{n}}s_n \to 0,$$

where $s_{n-\Delta_n/2,n} \sim s_n$ by Lemma 15. We just need to verify that the assumptions of that lemma hold.

Let $\lambda'_n = \lambda - \frac{\Delta_n}{2}$, then

$$H(\lambda - \frac{\Delta_n}{2}, \Delta_n) = H(\lambda'_n, \Delta_n)$$

and

$$\frac{\lambda'_n}{\lambda} = \frac{\lambda - \frac{\Delta_n}{2}}{\lambda} = 1 - \frac{\frac{\Delta_n}{2}}{\lambda}.$$

Since $\Delta_n = \sqrt{\frac{8n}{\delta}}$ and $\lambda = n$,

$$\lim_{n\to\infty} \frac{\frac{\Delta_n}{2}}{\lambda} = \lim_{n\to\infty} \sqrt{\frac{8/\delta}{n}} = 0.$$

Then

$$\lim_{n\to\infty} \frac{\lambda'_n}{\lambda} = 1 - \lim_{n\to\infty} \frac{\frac{\Delta_n}{2}}{\lambda} = 1, \tag{6.41}$$

(i.e. $\lambda'_n \sim n = \lambda$). Since

$$\begin{aligned}(\frac{\lambda}{\lambda'_n} - 1)\lambda^{\delta'} &= \frac{\lambda^{\delta'}\Delta_n}{2\lambda - \Delta_n} \\ &= \frac{n^{\delta'}\Delta_n}{2\lambda - \Delta_n} \\ &= \frac{n^{\delta'}k\sqrt{n}}{n - k\sqrt{n}} \quad \text{(Note: let } k = \sqrt{2/\delta}) \\ &= \frac{k}{2n^{1/2-\delta'} - kn^{-\delta'}},\end{aligned}$$

if we fix $\delta' \in (0, 1/2)$,

$$\lim_{n\to\infty} (\frac{\lambda}{\lambda'_n} - 1)\lambda^{\delta'} = 0.$$

Thus, there exists an $\delta' > 0$ such that $\limsup_n (\frac{\lambda}{\lambda'_n} - 1)\lambda^{\delta'} < \infty$.

Now $0 < \lambda'_n < \lambda < \infty$, (6.41) and $\limsup (\frac{\lambda}{\lambda'_n} - 1)\lambda^{\delta'} < \infty$ for $\delta' \in (0, 1/2)$ satisfy the conditions of Lemma 15.

Since

$$\lim_{n\to\infty} \frac{s_{\lambda'_n,n}}{\sqrt{n}} = \lim_{n\to\infty} \left(\frac{s_{\lambda,n}}{s_{\lambda,n}} \frac{s_{\lambda'_n,n}}{\sqrt{n}} \right) = \lim_{n\to\infty} \frac{s_{\lambda'_n,n}}{s_{\lambda,n}} \lim_{n\to\infty} \frac{s_{\lambda,n}}{\sqrt{n}},$$

by Lemma 15 for the changing distribution ($s_{\lambda'_n,n} \sim s_{\lambda,n}$) and the assumption $\frac{s_{\lambda,n}}{\sqrt{n}} \to 0$ we have

$$\frac{s_{\lambda'_n,n}}{\sqrt{n}} \to 0.$$

Now, note that

$$\begin{aligned}
& s_{\lambda,n}^{-1} E\left[(0.5)\Delta_n \sum_{a\in\mathcal{A}} p_{a,n} 1_{\left[y'_{a,n}(n)=r\right]} \right] \\
=& (0.5) s_{\lambda,n}^{-1} \Delta_n \sum_{a\in\mathcal{A}} \binom{n}{r} p_{a,n}^{r+1} \left(1 - p_{a,n}\right)^{n-r} \\
\sim& (0.5) s_{\lambda,n}^{-1} \Delta_n \frac{n^r}{r!} \sum_{a\in\mathcal{A}} p_{a,n}^{r+1} \left(1 - p_{a,n}\right)^{n-r} \\
\leq& (0.5) s_{\lambda,n}^{-1} \Delta_n \frac{n^r}{r!} \sum_{a\in\mathcal{A}} p_{a,n}^{r+1} e^{-(n-r)p_{a,n}} \\
=& (0.5) s_{\lambda,n}^{-1} \frac{\Delta_n}{n} \sum_{a\in\mathcal{A}} np_{a,n} \frac{(np_{a,n})^r}{r!} e^{-np_{a,n}} e^{rp_{a,n}} \\
\leq& (0.5) s_{\lambda,n}^{-1} \frac{\Delta_n}{n} \sum_{a\in\mathcal{A}} np_{a,n} \frac{(np_{a,n})^r}{r!} e^{-np_{a,n}} e^{r} \\
\leq& (0.5) e^r \frac{\Delta_n}{n} s_{\lambda,n}^{-1} \sum_{a\in\mathcal{A}} (r + 1 + np_{a,n}) e^{-np_{a,n}} \frac{(np_{a,n})^{r+1}}{r!} \qquad \text{(Note: } \lambda = n\text{)}
\end{aligned}$$

$$
\begin{aligned}
&=(0.5)e^r \frac{\Delta_n}{n} s_{\lambda,n}^{-1} s_{\lambda,n}^2 \\
&=(0.5)e^r \frac{\Delta_n}{n} s_{\lambda,n} \to 0,
\end{aligned}
$$

where the third line follows by

$$
\frac{\binom{n}{r}}{\frac{n^r}{r!}} = \frac{n!}{(n-r)!n^r} = \frac{n(n-1)...(n-r+1)}{n^r} \to 1
$$

(i.e., $\binom{n}{r} \sim \frac{n^r}{r!}$), the fourth line follows by the fact that $(1-x) \leq e^{-x}$, and the last line follows by $\Delta_n \sim M_1\sqrt{n}$ and

$$
\frac{\Delta_n}{n} s_{\lambda,n} = \frac{M_1\sqrt{n}}{M_1\sqrt{n}} \frac{\Delta_n}{n} s_{\lambda,n} = \frac{\Delta_n}{M_1\sqrt{n}} \frac{M_1}{\sqrt{n}} s_{\lambda,n} \to 0.
$$

□

Proof of Theorem 5. Note that

$$
\frac{\xi_{n,n}}{s_{n,n}} = \frac{\xi_{n,n} - \zeta_{n,n}}{s_{n,n}} + \frac{\zeta_{n,n}}{s_{n,n}},
$$

where

$$
\begin{aligned}
\zeta_{n,n} &= \lambda(T_{r,n}(n) - \pi_{r,n}(n)) \\
T_{r,n}(n) &= \frac{N_{r+1,n}(n)}{n}(r+1) \\
N_{r,n}(n) &= \sum_{a \in \mathcal{A}} 1_{[y_{a,n}(n)=r]} \\
\pi_{r,n}(n) &= \sum_{a \in \mathcal{A}} p_{a,n} 1_{[y_{a,n}(n)=r]}.
\end{aligned}
$$

By Theorem 2, (2.8) holds if and only if

$$
\frac{\zeta_{n,n}}{s_{n,n}} = \frac{\lambda(T_{r,n}(n) - \pi_{r,n}(n))}{s_{n,n}} \xrightarrow[n\to\infty]{d} N(0,1). \tag{6.42}
$$

Since $s_{n,n} \to \infty$ as $n \to \infty$ and

$$\frac{s_{n,n}}{\sqrt{n}} \to 0,$$

Lemma 16 implies that

$$\frac{\xi_{n,n} - \zeta_{n,n}}{s_{n,n}} \xrightarrow{p} 0.$$

Therefore, by Slutsky's theorem, (2.8) if and only if

$$\frac{\xi_{n,n}}{s_{n,n}} \xrightarrow{d} N(0,1).$$

□

Lemma 17. *For $c, d \geq 0$, let*

$$S_n = c\mathrm{E}[N_{r+1,n}] + d\mathrm{E}[N_{r+2,n}] \tag{6.43}$$

and

$$T_n = c\mathrm{E}[N'_{r+1,n}] + d\mathrm{E}[N'_{r+2,n}]. \tag{6.44}$$

1. For any $\epsilon \in (0, \frac{1}{2})$ and $n \geq r+2$

$$A_n(S - B_n) \leq T_n \leq S_n e^{\epsilon(r+1)} + n^{r+2}(c+d)e^{-\epsilon(n-r-2)},$$

for some $0 \leq A_n \to 1$ and $0 \leq B_n \to 0$ as $n \to \infty$, which may depend on ϵ.

2. We have $T_n \to \infty$ if and only if $S_n \to \infty$. And we have and $\liminf S_n = 0$ if and only if $\liminf T_n = 0$.

3. If $S_n \to \infty$, then $T_n/S_n \to 1$.

Proof. Parts 2 and 3 follow immediately from Part 1. We now prove Part 1. Recall that for the Poissonized case

$$\begin{aligned}
\mathrm{E}\left[N_{r,n}\right] &= \mathrm{E}\left[\sum_{a\in\mathcal{A}} 1_{[y_{a,n}(n)=r]}\right] \\
&= \sum_{a\in\mathcal{A}} \mathrm{E}\left[1_{[y_{a,n}(n)=r]}\right] \\
&= \sum_{a\in\mathcal{A}} P\left(y_{a,n}(n)=r\right) \\
&= \sum_{a\in\mathcal{A}} e^{-np_{a,n}} \frac{\left(np_{a,n}\right)^r}{r!},
\end{aligned}$$

and for the deterministic case

$$\begin{aligned}
\mathrm{E}\left[N'_{r,n}\right] &= \mathrm{E}\left[\sum_{a\in\mathcal{A}} 1_{[y'_{a,n}(n)=r]}\right] \\
&= \sum_{a\in\mathcal{A}} \mathrm{E}\left[1_{[y'_{a,n}(n)=r]}\right] \\
&= \sum_{a\in\mathcal{A}} P\left(y'_{a,n}(n)=r\right) \\
&= \sum_{a\in\mathcal{A}} \binom{n}{r} p_{a,n}^r (1-p_{a,n})^{n-r}.
\end{aligned}$$

For $n \geq r+2$

$$\begin{aligned}
S_n &= c\sum_{a\in\mathcal{A}} e^{-np_{a,n}} \frac{\left(np_{a,n}\right)^{r+1}}{(r+1)!} + d\sum_{a\in\mathcal{A}} e^{-np_{a,n}} \frac{\left(np_{a,n}\right)^{r+2}}{(r+2)!} \\
&= \sum_{a\in\mathcal{A}} \frac{\left(np_{a,n}\right)^{r+1}}{(r+1)!} e^{-np_{a,n}} \left(c + d\frac{np_{a,n}}{r+2}\right),
\end{aligned}$$

and using the fact that $\binom{n}{r+2} = \binom{n}{r+1}\frac{n-r-1}{r+2}$

$$T_n = \sum_{a\in\mathcal{A}} \left(c\binom{n}{r+1} p_{a,n}^{r+1}(1-p_{a,n})^{n-r-1} + d\binom{n}{r+2} p_{a,n}^{r+2}(1-p_{a,n})^{n-r-2}\right)$$

$$
\begin{aligned}
&= \sum_{a\in\mathcal{A}} \binom{n}{r+1} p_{a,n}^{r+1}(1-p_{a,n})^{n-r-2} \left(c(1-p_{a,n}) + d\frac{n-r-1}{r+2} p_{a,n} \right) \\
&\leq \sum_{a\in\mathcal{A}} \frac{(np_{a,n})^{r+1}}{(r+1)!} e^{-np_{a,n}} \left(c + d\frac{np_{a,n}}{r+2} \right) e^{p_{a,n}(r+2)} \\
&\leq \sum_{a\in\mathcal{A}, p_{a,n}\leq\epsilon} \frac{(np_{a,n})^{r+1}}{(r+1)!} e^{-np_{a,n}} \left(c + d\frac{np_{a,n}}{r+2} \right) e^{\epsilon(r+2)} \\
&\quad + n^{r+2} \sum_{a\in\mathcal{A}, p_{a,n}>\epsilon} p_{a,n}(c+d)e^{-\epsilon(n-r-2)} \\
&\leq S_n e^{\epsilon(r+2)} + n^{r+2}(c+d)e^{-\epsilon(n-r-2)},
\end{aligned}
$$

where we use the facts that $\binom{n}{r} \leq \frac{n^r}{r!}$ and $(1-x) \leq e^{-x}$. Next, fix $\delta \in (\frac{1}{2}, 1)$. Using the facts that $(1-x) \geq e^{-x/(1-2x^2)}$ for $x \in (0, 1/2)$, see Lemma 2.6 in [24], we get

$$
\begin{aligned}
T_n &= \sum_{a\in\mathcal{A}} \left(c\binom{n}{r+1} p_{a,n}^{r+1}(1-p_{a,n})^{n-r-1} + d\binom{n}{r+2} p_{a,n}^{r+2}(1-p_{a,n})^{n-r-2} \right) \\
&= \binom{n}{r+1} \sum_{a\in\mathcal{A}} p_{a,n}^{r+1}(1-p_{a,n})^{n-r-2} \left(c(1-p_{a,n}) + d\frac{n-r-1}{r+2} p_{a,n} \right) \\
&\geq \binom{n}{r+1} \\
&\quad \times \sum_{a\in\mathcal{A}, p_{a,n}\leq\epsilon/n^\delta} p_{a,n}^{r+1} e^{-np_{a,n}} e^{-\frac{p_{a,n}}{1-p_{a,n}}(p_{a,n}n-r-2)} \left(c + d\frac{n-r-1}{r+2} p_{a,n} \right) (1-p_{a,n}) \\
&\geq (1-\epsilon/n^\delta) \binom{n}{r+1} e^{-\frac{\epsilon}{n^\delta-\epsilon}(\epsilon n^{1-\delta}-r-2)} \\
&\quad \times \sum_{a\in\mathcal{A}, p_{a,n}\leq\epsilon/n^\delta} p_{a,n}^{r+1} e^{-np_{a,n}} \left(c + d\frac{n-r-1}{r+2} p_{a,n} \right) \\
&= (1-\epsilon/n^\delta) \binom{n}{r+1} e^{-\frac{\epsilon}{n^\delta-\epsilon}(\epsilon n^{1-\delta}-r-2)} \frac{(r+1)!}{n^{r+1}} \left(c \sum_{a\in\mathcal{A}, p_{a,n}\leq\epsilon/n^\delta} \frac{(np_{a,n})^{r+1}}{(r+1)!} e^{-np_{a,n}} \right. \\
&\quad \left. + d\frac{(n-r-1)}{n} \sum_{a\in\mathcal{A}, p_{a,n}\leq\epsilon/n^\delta} \frac{(np_{a,n})^{r+2}}{(r+2)!} e^{-np_{a,n}} \right) \\
&= A_n(S - B_n),
\end{aligned}
$$

where we use the fact that $\binom{n}{r+1} \sim \frac{n^{r+1}}{(r+1)!}$.

$$A_n = e^{-\frac{\epsilon}{n^\delta - \epsilon}(\epsilon n^{1-\delta} - r - 2)}(1 - \epsilon/n^\delta)\binom{n}{r+1}\frac{(r+1)!}{n^{r+1}} \to 1,$$

and

$$\begin{aligned} B_n =& d\frac{r+1}{n} \sum_{a \in \mathcal{A}, p_{a,n} \le \epsilon/n^\delta} \frac{(np_{a,n})^{r+2}}{(r+2)!} e^{-np_{a,n}} + c \sum_{a \in \mathcal{A}, p_{a,n} > \epsilon/n^\delta} \frac{(np_{a,n})^{r+1}}{(r+1)!} e^{-np_{a,n}} \\ &+ d \sum_{a \in \mathcal{A}, p_{a,n} > \epsilon/n^\delta} \frac{(np_{a,n})^{r+2}}{(r+2)!} e^{-np_{a,n}} \\ =& B_n^{(1)} + B_n^{(2)} + B_n^{(3)}. \end{aligned}$$

We will show that $B_n \to 0$. First, let $M > 0$ be a constant with $x^{r+1}e^{-x} \le M$ for $x \ge 0$, then by Dominated convergence Theorem

$$B_n^{(1)} \le d(r+1)M \sum_{a \in \mathcal{A}, p_{a,n} \le \epsilon/n^\delta} p_{a,n} \to 0.$$

Next

$$B_n^{(2)} \le c e^{-\epsilon n^{1-\delta}} n^{r+1} \sum_{a \in \mathcal{A}} p_{a,n} \to 0,$$

and similarly $B_n^{(3)} \to 0$. □

Lemma 18. *1. For any $1 \le k \le n/2$, we have*

$$\mathrm{Var}(N'_{k,n}) \le A_{k,n} \mathrm{E}[N'_{k,n}],$$

where $A_{k,n} = \left(4k^{k+1}\binom{n-k}{k}(n-2k)^{-k} + 1\right) \to \frac{4k^k}{(k-1)!} + 1$.

2. If $\mathrm{E}[N'_{k,n}] \to \infty$, *then*

$$\frac{\mathrm{Var}(N'_{k,n})}{(\mathrm{E}[N'_{k,n}])^2} \to 0 \text{ and } \frac{N'_{k,n}}{\mathrm{E}[N'_{k,n}]} \xrightarrow{p} 1.$$

To show this, we use ideas from the proof of Theorem 3.3 in [24]. Part 2 can also be found without proof in Section 4 of [32].

Proof. First note that, for any $1 \leq k \leq n/2$,

$$(N'_{k,n})^2 = \sum_{a\in\mathcal{A}} \sum_{b\in\mathcal{A}, a\neq b} 1_{[y'_{a,n}=k]} 1_{[y'_{b,n}=k]} + N'_{k,n}$$

and

$$\mathrm{E}[(N'_{k,n})^2] = \binom{n}{k, k, n-2k} \sum_{a\in\mathcal{A}} \sum_{b\in\mathcal{A}, a\neq b} p^k_{a,n} p^k_{b,n} (1 - p_{a,n} - p_{b,n})^{n-2k} + \mathrm{E}[N'_{k,n}].$$

Next, let and $B_{k,n} = \binom{n}{k,k,n-2k} / \binom{n}{k}^2 = \binom{n-k}{k} / \binom{n}{k} \leq 1$ and note that $B_{k,n} \to 1$. We have

$$\begin{aligned}
\mathrm{Var}(N'_{k,n}) = \quad & \mathrm{E}[(N'_{k,n})^2] - \mathrm{E}[N'_{k,n}] - B_{k,n}(\mathrm{E}[N'_{k,n}])^2 \\
& + (B_{k,n} - 1)(\mathrm{E}[N'_{k,n}])^2 + \mathrm{E}[N'_{k,n}] \\
\leq \quad & \mathrm{E}[(N'_{k,n})^2] - \mathrm{E}[N'_{k,n}] - B_{k,n}(\mathrm{E}[N'_{k,n}])^2 + \mathrm{E}[N'_{k,n}].
\end{aligned}$$

We can upper bound $\mathrm{E}[(N'_{k,n})^2] - \mathrm{E}[N'_{k,n}] - B_{k,n}(\mathrm{E}[N'_{k,n}])^2$ by

$$\begin{aligned}
& \binom{n}{k, k, n-2k} \sum_{a\in\mathcal{A}} \sum_{b\in\mathcal{A}} p^k_{a,n} p^k_{b,n} \left((1 - p_{a,n} - p_{b,n})^{n-2k} - (1 - p_{a,n})^{n-k} (1 - p_{b,n})^{n-k} \right) \\
& \leq k \binom{n}{k, k, n-2k} \sum_{a\in\mathcal{A}} \sum_{b\in\mathcal{A}} p^k_{a,n} p^k_{b,n} (1 - p_{a,n})^{n-2k} (1 - p_{b,n})^{n-2k} (p_{a,n} + p_{b,n}) \\
& \leq 2k \binom{n}{k, k, n-2k} \sum_{a\in\mathcal{A}} \sum_{b\in\mathcal{A}, p_{a,n} \leq p_{b,n}} p^k_{a,n} p^{k+1}_{b,n} (1 - p_{a,n})^{n-k} (1 - p_{b,n})^{n-3k}
\end{aligned}$$

$$
\begin{aligned}
&+ 2k\binom{n}{k,k,n-2k}\sum_{a\in\mathcal{A}}\sum_{b\in\mathcal{A},p_{a,n}>p_{b,n}} p_{a,n}^{k+1}p_{b,n}^{k}(1-p_{a,n})^{n-3k}(1-p_{b,n})^{n-k} \\
&\leq 4k\binom{n}{k,k,n-2k}\sum_{a\in\mathcal{A}} p_{a,n}^{k}(1-p_{a,n})^{n-k}\sum_{b\in\mathcal{A}} p_{b,n}^{k+1}(1-p_{b,n})^{n-3k} \\
&= 4k\binom{n-k}{k}\mathrm{E}[N'_{k,n}]\sum_{b\in\mathcal{A}} p_{b,n}^{k+1}(1-p_{b,n})^{n-3k} \leq 4k^{k+1}\binom{n-k}{k}\mathrm{E}[N'_{k,n}](n-2k)^{-k}.
\end{aligned}
$$

Here the second line uses the facts that $1-p_{a,n}-p_{b,n}\leq 1-p_{a,n}-p_{b,n}+p_{a,n}p_{b,n} = (1-p_{a,n})(1-p_{b,n})$, that $1-(1-p_{a,n})^k(1-p_{b,n})^k \leq 1-(1-p_{a,n}-p_{b,n})^k$, and that $1-(1-x)^k \leq kx$ for $x\in[0,1]$, which is easily checked by induction on k. The last inequality follows by the fact that $x^k(1-x)^{n-3k}\leq k^k(n-2k)^{-k}$ for $x\in[0,1]$, which can be shown using standard calculus arguments.

For the second part, by Chebyshev's inequality, it suffices to show that $\frac{\mathrm{Var}(N'_k)}{(\mathrm{E}[N'_k])^2}\to 0$. The first part implies that

$$
\frac{\mathrm{Var}(N'_k)}{(\mathrm{E}[N'_k])^2} \leq A_{k,n}\frac{1}{\mathrm{E}[N'_k]} \to 0.
$$

This holds since $A_{k,n}\to \frac{4k^k}{(k-1)!}+1$, which follows by the fact that $\binom{n}{k}\sim\frac{n^k}{k!}$. □

Proof of Corollary 9. Since $(r+1)^2>0$ and $(r+2)(r+1)>0$, if $s_n\to\infty$ as $n\to\infty$, applying part 3 of Lemma 17 completes the proof. □

Proof of Corollary 10. Since we have Theorem 5 and Corollary 9, the proof can be completed by applying the Slutsky's theorem. □

Proof of Corollary 11. In the proof of Corollary 3 (6.13) we showed that $\mathrm{E}[N_{r+1,n}]\to\infty$. Now let $c=1$ and $d=0$ in (6.43) and (6.44) of Lemma 17, then using part 2 of Lemma 17 gives

$$
\mathrm{E}[N'_{r+1,n}]\to\infty, \tag{6.45}
$$

and using part 3 of Lemma 17 gives

$$\mathrm{E}[N'_{r+1,n}] \sim \mathrm{E}[N_{r+1,n}]. \tag{6.46}$$

We also showed in Lemma 5 that

$$\frac{\mathrm{E}[N_{r+1,n}]}{s_\lambda} \to \infty,$$

and here we set $\lambda = n$ and get

$$\frac{\mathrm{E}[N_{r+1,n}]}{s_\lambda} \to \infty. \tag{6.47}$$

Corollary 9 gives $(s'_{n,n})^2 \sim s^2_{n,n}$; thus together with (6.47) and (6.46) we have

$$\frac{\mathrm{E}[N'_{r+1,n}]}{s'_{n,n}} \xrightarrow{p} \infty. \tag{6.48}$$

Since (6.45) holds, using part 2 of Lemma 12 gives

$$\frac{N'_{r+1,n}}{\mathrm{E}[N'_{r+1,n}]} \xrightarrow{p} 1. \tag{6.49}$$

Since

$$\frac{N'_{r+1,n}}{s'_{n,n}} = \frac{\mathrm{E}[N'_{r+1,n}]}{s'_{n,n}} \frac{N'_{r+1,n}}{\mathrm{E}[N'_{r+1,n}]},$$

by continuous mapping theorem (6.48) and (6.49) implies that

$$\frac{N'_{r+1,n}}{s'_{n,n}} \xrightarrow{p} \infty.$$

Since $r+1 \in (0,\infty)$,

$$\frac{N'_{r+1,n}(r+1)}{s'_{n,n}} \xrightarrow{p} \infty.$$

Now plugging in

$$T'_{r,n}\left(n\right) = \frac{N'_{r+1,n}}{n}\left(r+1\right),$$

$$\frac{nT'_{r,n}(n)}{s'_{n,n}} = \frac{N'_{r+1,n}(r+1)}{s'_{n,n}} \xrightarrow{p} \infty.$$

By the symmetry of Normal distribution (2.10) implies that

$$\frac{n}{s'_{n,n}}\left(\pi'_{r,n}(n) - T'_{r,n}(n)\right) \xrightarrow[n\to\infty]{d} N(0,1),$$

and so

$$\frac{nT'_{r,n}(n)}{s'_{n,n}}\left(\frac{\pi'_{r,n}(n)}{T'_{r,n}(n)} - 1\right) \xrightarrow[n\to\infty]{d} N(0,1).$$

Since

$$\frac{nT'_{r,n}(n)}{s'_{n,n}} \xrightarrow{p} \infty,$$

it follows from Lemma 1 that

$$\frac{\pi'_{r,n}(n)}{T'_{r,n}(n)} - 1 \xrightarrow{p} 0.$$

Therefore,

$$\frac{T'_{r,n}(n)}{\pi'_{r,n}(n)} - 1 \xrightarrow{p} 0.$$

□

Lemma 19. *For the deterministic case let*

$$T_n = c\mathrm{E}[N'_{r+1,n}] + d\mathrm{E}[N'_{r+2}]$$
$$\hat{T}_n = cN'_{r+1,n} + dN'_{r+2,n}.$$

$\hat{T}$ *is a consistent estimator of* T*, i.e., as* $n \to \infty$*, for all* $\epsilon > 0$

$$P\left(\left|\frac{\hat{T}_n}{T_n} - 1\right| > \epsilon\right) \to 0.$$

Proof. Here we use similar arguments from the proof of Lemma 13, because Lemma 12 still holds when the distribution is changing. □

Proof of Corollary 12. Since $(r+1)^2 > 0$ and $(r+2)(r+1) > 0$, the result is an application of Lemma 19. □

Lemma 20. *For any* $0 \leq k \leq n/2$*, we have*

$$0 \leq \mathrm{E}\left[N'_{k+1,n} - \frac{n-k}{k+1}\pi'_{k,n}\right] \leq \frac{e^{k+1}}{n}\mathrm{E}[N_{k+2,n}]$$

and

$$\mathrm{Var}(\pi'_{k,n}) \leq n^{-2}e^k\mathrm{E}[N_{k+2,n}] + 2ke^{4k}n^{-3}\mathrm{E}[N_{k+1,n}]\mathrm{E}[N_{k+2,n}].$$

Proof. First note that, for any $0 \le k \le n/2$,

$$(\pi'_{k,n})^2 = \sum_{a\in\mathcal{A}} \sum_{b\in\mathcal{A}, a\neq b} p_{a,n} p_{b,n} 1_{[y'_{a,n}=k]} 1_{[y'_{b,n}=k]} + \sum_{a\in\mathcal{A}} p_{a,n}^2 1_{[y'_{a,n}=k]}$$

and

$$\begin{aligned} \mathrm{E}[(\pi'_{k,n})^2] &= \binom{n}{k,k,n-2k} \sum_{a\in\mathcal{A}} \sum_{b\in\mathcal{A}, a\neq b} p_{a,n}^{k+1} p_{b,n}^{k+1} (1-p_{a,n}-p_{b,n})^{n-2k} \\ &\quad + \binom{n}{k} \sum_{a\in\mathcal{A}} p_{a,n}^{k+2} (1-p_{a,n})^{n-k} =: H_1 + H_2. \end{aligned}$$

We have

$$\begin{aligned} 0 &\le \mathrm{E}\left[N'_{k+1,n} - \frac{n-k}{k+1}\pi'_{k,n}\right] = \binom{n}{k+1} \sum_{a\in\mathcal{A}} p_{a,n}^{k+2}(1-p_{a,n})^{n-k-1} \\ &\le n^{k+1} \sum_{a\in\mathcal{A}} p_{a,n}^{k+2} e^{-p_{a,n}(n-k-1)} \\ &\le \frac{e^{k+1}}{n} \sum_{a\in\mathcal{A}} (np_{a,n})^{k+2} e^{-p_{a,n}n} = \frac{e^{k+1}}{n} \mathrm{E}[N_{k+2,n}], \end{aligned}$$

where we use the facts that $\binom{n}{k+1} = \frac{n-k}{k+1}\binom{n}{k}$, that $\binom{n}{k} \le n^k$, and that $1-x \le e^{-x}$ for $x > 0$. In a similar way we can upper bound H_2 by

$$\begin{aligned} H_2 &\le n^k \sum_{a\in\mathcal{A}} p_{a,n}^{k+2}(1-p_{a,n})^{n-k} \le n^{-2} \sum_{a\in\mathcal{A}} (np_{a,n})^{k+2} e^{-p_{a,n}(n-k)} \\ &\le n^{-2} e^k \sum_{a\in\mathcal{A}} (np_{a,n})^{k+2} e^{-np_{a,n}} = n^{-2} e^k \mathrm{E}[N_{k+2,n}]. \end{aligned}$$

Now, let $B_{k,n} = \binom{n}{k,k,n-2k} / \binom{n}{k}^2 = \binom{n-k}{k} / \binom{n}{k}$ and note that $B_{k,n} \le 1$. We can upper bound $H_1 - B_{k,n}(\mathrm{E}[\pi'_{k,n}])^2$ by

$$\binom{n}{k,k,n-2k} \sum_{a\in\mathcal{A}} \sum_{b\in\mathcal{A}} p_{a,n}^{k+1} p_{b,n}^{k+1} \left((1-p_{a,n}-p_{b,n})^{n-2k} - (1-p_{a,n})^{n-k}(1-p_{b,n})^{n-k} \right)$$

$$
\begin{aligned}
&\le kn^{2k}\sum_{a\in\mathcal{A}}\sum_{b\in\mathcal{A}} p_{a,n}^{k+1}p_{b,n}^{k+1}(1-p_{a,n})^{n-2k}(1-p_{b,n})^{n-2k}(p_{a,n}+p_{b,n}) \\
&= 2kn^{2k}\sum_{a\in\mathcal{A}} p_{a,n}^{k+1}(1-p_{a,n})^{n-2k}\sum_{b\in\mathcal{A}} p_{b,n}^{k+2}(1-p_{b,n})^{n-2k} \\
&\le 2kn^{-3}\sum_{a\in\mathcal{A}}(np_{a,n})^{k+1}e^{-p_{a,n}n+2kp_{a,n}}\sum_{b\in\mathcal{A}}(np_{b,n})^{k+2}e^{-p_{b,n}n+2kp_{a,n}} \\
&\le 2ke^{4k}n^{-3}\mathrm{E}[N_{k+1,n}]\mathrm{E}[N_{k+2,n}],
\end{aligned}
$$

where the second line follows by arguments similar to those in the proof of Lemma 18 and the third by symmetry. From here the result follows. □

Proof of Theorem 6. Theorem 3 and Lemma 17 imply that $\mathrm{E}[N'_{r+1,n}] \to c^*$ and $\mathrm{E}[N'_{r+2,n}] \to 0$. Then Lemma 20 implies that $\mathrm{E}[\frac{n-r}{r+1}\pi_{r,n}] \to c^*$ and that $\mathrm{Var}(\frac{n-r}{r+1}\pi'_{r,n}) \to 0$. Now, the first convergence follows by the well-known representation of the mean square error as the sum of the variance and the square of the bias. From here, Markov's inequality combined with Slutsky's Theorem gives the second convergence. The last convergence follows from Theorem 3, Lemma 16, and Slutsky's Theorem. □

REFERENCES

[1] D. A. McAllester and L. E. Ortiz, "Concentration inequalities for the missing mass and for histogram rule error," *Journal of Machine Learning Research*, vol. 4, pp. 895–911, 2003.

[2] D. Berend and A. Kontorovich, "On the concentration of the missing mass," *Electronic Communication in Probability*, vol. 18(3), pp. 1–7, 2013.

[3] G. Decrouez, M. Grabchak, and Q. Paris, "Finite sample properties of the mean occupancy counts and probabilities," *Bernoulli*, vol. 24, pp. 1910–1941, 2016.

[4] A. Ben-Hamou, S. Boucheron, and M. I. Ohannessian, "Concentration inequalities in the infinite urn scheme for occupancy counts and the missing mass, with applications," *Bernoulli*, vol. 23(1), pp. 249–287, 2017.

[5] I. J. Good and G. H. Toulmin, "The number of new species, and the increase in population coverage, when a sample is increased," *Biometrika*, vol. 43, pp. 45–63, 1956.

[6] A. Chao, "On estimating the probability of discovering a new species," *The Annals of Statistics*, vol. 9, pp. 1339–1342, 1981.

[7] A. Chao, T. C. Hsieh, R. L. Chazdon, R. K. Colwell, and N. J. Gotelli, "Unveiling the species-rank abundance distribution by generalizing the Good-Turing sample coverage theory," *Ecology*, vol. 96, pp. 1189–1201, 2015.

[8] C. X. Mao and B. G. Lindsay, "A Poisson model for the coverage problem with a genomic application," *Biometrika*, vol. 89, pp. 669–681, 2002.

[9] S. F. Chen and J. Goodman, "An empirical study of smoothing techniques for language modeling," *Computer Speech & Language*, vol. 13, pp. 359–394, 1999.

[10] A. B. Wagner, P. Viswanath, and S. R. Kulkarni, "Probability estimation in the rare-events regime," *IEEE Transactions on Information Theory*, vol. 57, pp. 3207–3229, 2011.

[11] B. Efron and R. Thisted, "Estimating the number of unseen species: How many words did Shakespeare know?," *Biometrika*, vol. 63, pp. 435–447, 1976.

[12] R. Thisted and B. Efron, "Did Shakespeare write a newly discovered poem," *Biometrika*, vol. 74, pp. 445–455, 1987.

[13] Z. Zhang and H. Huang, "Turing's formula revisited," *Journal of Quantitative Linguistics*, vol. 14(2-3), pp. 222–241, 2017.

[14] C. H. Zhang, "Estimation of sums of random variables: Examples and information bounds," *The Annals of Statistics*, vol. 33, pp. 2022–2041, 2005.

[15] I. J. Good, "The population frequencies of species and the estimation of population parameters," *Biometrika*, vol. 40, pp. 237–264, 1953.

[16] H. E. Robbins, "Estimating the total probability of the unobserved outcomes of an experiment," *Annals of Mathematical Statistics*, vol. 39(1), pp. 256–257, 1968.

[17] A. Chao, S. M. Lee, and T. C. Chen, "A generalized Good's nonparametric coverage estimator," *Chinese Journal of Mathematics*, vol. 16, pp. 189–199, 1988.

[18] M. I. Ohannessian and M. A. Dahleh, "Rare probability estimation under regularly varying heavy tails," *Proceedings of the 25th Annual Conference on Learning Theory, in Proceedings of Machine Learning Research*, vol. 23, pp. 21.1–21.24, 2012.

[19] M. Grabchak and V. Cosme, "Performance of Turing's formula: A simulation study," *Communications in Statistics – Simulation and Computation*, vol. 46(6), pp. 4199–4209, 2017.

[20] W. W. Esty, "A normal limit law for a nonparametric estimator of the coverage of a random sample," *Annals of Statistics*, vol. 11(3), pp. 905–912, 1983.

[21] Z. Zhang and H. Huang, "A sufficient normality condition for Turing's formula," *Journal of Nonparametric Statistics*, vol. 20(5), pp. 431–446, 2008.

[22] Z. Zhang, "On normal law conditions for Turing's formula," *Wiley StatsRef: Statistics Reference Online*, pp. 1–10, 2018.

[23] C. H. Zhang and Z. Zhang, "Asymptotic normality of a nonparametric estimator of sample coverage," *Annals of Statistics*, vol. 37(5A), pp. 2582–2595, 2009.

[24] Z. Zhang, "A multivariate normal law for Turing's formula," *Sankhya*, vol. 75A(1), pp. 51–73, 2003.

[25] M. Grabchak and Z. Zhang, "Asymptotic properties of Turing's formula in relative error," *Machine Learning*, vol. 106(11), pp. 1771–1785, 2017.

[26] Z. Zhang, *Statistical Implications of Turing's Formula*. Hoboken: Wiley, 2017.

[27] W. W. Esty, "Confidence intervals for the coverage of low coverage samples," *Annals of Statistics*, vol. 10(1), pp. 190–196, 1982.

[28] R. Bin Tareaf, "Tweets Dataset - Top 20 most followed users in Twitter social platform," 2017.

[29] P. Billingsley, *Probability and Measure*. New York: John Wiley & Sons, 3 ed., 1995.

[30] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, vol. 10. New York: Dover Publications, 1972.

[31] M. Grabchak, "On the transition laws of p-tempered α-stable OU-processes," *Computational Statistics*, vol. 36(2), pp. 1415–1436, 2021.

[32] A. Gnedin, B. Hansen, and J. Pitman, "Notes on the occupancy problem with infinitely many boxes: general asymptotics and power laws," *Probability Surveys*, vol. 4, pp. 146–171, 2007.