

# Coincidences, Goodness of Fit Test and Confidence Interval for Poisson Distribution Parameter via Coincidence

Victor Nijimbere\*

School of Mathematics and Statistics, Carleton University, Ottawa, Canada

\*Corresponding author: victornijimbere@gmail.com

**Abstract** The probability of the coincidence of some discrete random variables having a Poisson distribution with parameters  $\lambda_1, \lambda_2, \dots, \lambda_n$ , and moments are expressed in terms of the hypergeometric function  ${}_1F_n$  or the modified Bessel function of the first kind if  $n=2$ . Considering the null hypothesis  $H_0: \lambda_1=\lambda_2=\dots=\lambda_n=\theta$ , where  $\theta$  is some positive constant number, asymptotic approximations of the probability and moments are derived for large  $\theta$  using the asymptotic expansion of the hypergeometric function  ${}_1F_n$  and that of the modified Bessel function of the first kind if  $n=2$ . Further, we show that if the sample mean is a minimum variance unbiased estimator (MVUE) for the parameter  $\lambda_i$ , then the probability that  $H_0$  is true can be approximated by that of a coincidence. In that case, a chi-square  $\chi^2$  goodness of fit test can be established and a  $100(1-\alpha)\%$  confidence interval (CI) for  $\theta$  can be constructed using the variance of the coincidence (or via coincidence) and the Central Limit Theorem (CLT).

**Keywords:** probability, hypergeometric functions, modified bessel functions, asymptotic expansion, MVUE

**Cite This Article:** Victor Nijimbere, "Coincidences, Goodness of Fit Test and Confidence Interval for Poisson Distribution Parameter via Coincidence." *American Journal of Applied Mathematics and Statistics*, vol. 4, no. 6 (2016): 185-193. doi: 10.12691/ajams-4-6-4.

## 1. Introduction

A discrete random variable  $X$  having a Poisson distribution with parameter  $\lambda$  is denoted as  $X \sim Poi(\lambda)$ , and its probability mass function (p.m.f) [9] is

$$P(X = s) = \frac{e^{-\lambda} \lambda^s}{s!}, s = 0, 1, 2, \dots \quad (1.1)$$

Poisson distribution is often used for modeling real-life discrete random phenomena and has many applications in economics, business, health care, science, and engineering [3,5,7,13]. For instance, the number of hungry persons entering MacDonald's restaurant, the number of birth, death, marriages, the number of patients arriving at an emergency room, the number of customers who call in call centers by the end of the month to complain about a service problem, and many more follow a Poisson distribution model [3,5,7,13]. The mathematical and statistical analysis throughout this paper may apply to most of the above examples, however, we mainly focus on scenarios of telephone calls in call centers in order the analysis to be comprehended without much difficulties.

Now, let us consider that  $X_1, X_2, \dots, X_n$  are Poisson distributed random variables, independent and identically distributed (*iid*), and  $X_i \sim Poi(\lambda_i), i = 1, 2, \dots, n$ . Then the joint p.m.f of  $X_1 = s_1, X_2 = s_2, \dots, X_n = s_n$  is obviously given by

$$\begin{aligned} P(X_1 = s_1, X_2 = s_2, \dots, X_n = s_n) &= P(X_1 = s_1)P(X_2 = s_2) \dots P(X_n = s_n) \\ &= \frac{e^{-\lambda_1} \lambda_1^{s_1}}{s_1!} \frac{e^{-\lambda_2} \lambda_2^{s_2}}{s_2!} \dots \frac{e^{-\lambda_n} \lambda_n^{s_n}}{s_n!} = \prod_{i=1}^n \frac{e^{-\lambda_i} \lambda_i^{s_i}}{s_i!} \quad (1.2) \\ &= e^{-\sum_{i=1}^n \lambda_i} \prod_{i=1}^n \frac{\lambda_i^{s_i}}{s_i!}, s_i \geq 0, \forall i \end{aligned}$$

In the context of applications, we may think about  $n$  call centers each receiving an average of  $\lambda_i$  telephone calls in an hour. Let the random variable  $X_{ik}, i = 1, 2, \dots, n$  and  $k = 1, 2, \dots, K_i$  represents the number of calls received by the center  $i$  at the time  $k$ . A coincidence will occur if all the centers receive exactly the same amount of calls

$$X_i = \sum_{k=1}^{K_i} X_{ik}, \quad (1.3)$$

at the time  $K_i = K$  (or in  $K_i$  hours) for all  $i$ . Thus the coincidence the event is defined by

$$C = \bigcup_{s=0}^{\infty} \{X_1 = X_2 = \dots = X_n = s\}. \quad (1.4)$$

And its probability of occurrence is

$$P(C) = P\left(\bigcup_{s=0}^{\infty} \{X_1 = X_2 = \dots = X_n = s\}\right) = e^{-\sum_{i=1}^n \lambda_i} \sum_{s=0}^{\infty} \frac{\left(\prod_{i=1}^n \lambda_i\right)^s}{(s!)^n}, s \geq 0. \tag{1.5}$$

More coincidences can be defined. For instance, let  $m < n$ ,  $r_i, i=1,2,\dots,m \in \{1,2,\dots,n\}$  such that  $r_i \neq r_j$  for all  $i, j$ ; and let  $Q_{r_1,r_2,\dots,r_m}$  denotes the set  $\{1,2,\dots,n\} \setminus \{r_1,r_2,\dots,r_m\}$ . In addition, let

$$B_{r_1,r_2,\dots,r_m} = \{X_{a_1} = X_{a_2} = \dots = X_{a_{n-m}} = s\},$$

where  $\{a_1, a_2, \dots, a_{n-m}\} = Q_{r_1,r_2,\dots,r_m}$ , and consider the event

$$C_{m,l} = \bigcup_{s=0}^{\infty} \bigcup_{r_1,\dots,r_m} \left\{ B_{r_1,\dots,r_m} \cap \{X_{r_1} = \dots = X_{r_m} = s+l\} \right\}, \tag{1.6}$$

where  $l \in \{1,2,\dots,m\}$  and the inner union is over all possible  $(r_1, r_2, \dots, r_m)$  combinations such that  $r_i, i=1,2,\dots,m \in \{1,2,\dots,n\}$  and  $r_i \neq r_j$  for all  $i, j$ . Considering once again the call-center scenario, a near coincidence will occur if one center receives exactly one more call than other centers at the time  $k$ , while the others receive exactly the same number of calls at that time. Setting  $m=1$  so that  $B_{r_1,r_2,\dots,r_m}$  becomes

$$B_{r_1} = \{X_{a_1} = X_{a_2} = \dots = X_{a_{n-1}} = s\},$$

where  $\{a_1, a_2, \dots, a_{n-1}\} = Q_{r_1}$ . Then a near coincidence is the event corresponding to the event  $C_{m,l}$  with  $m=1$  and  $l=1$  and is defined by

$$C_{m=1,l=1} = \bigcup_{s=0}^{\infty} \bigcup_{r_1=1}^n \left\{ B_{r_1} \cap \{X_{r_1} = s+1\} \right\}. \tag{1.7}$$

Further coincidences can be defined using (1.6). Consider that 3 centers receive exactly 2 more calls than other centers in a given hour, while the others receive exactly the same number of calls. In this case,  $m=3$  so that  $B_{r_1,r_2,\dots,r_m}$  becomes

$$B_{r_1,r_2,r_3} = \{X_{a_1} = X_{a_2} = \dots = X_{a_{n-3}} = s\},$$

$a_1, a_2, \dots, a_{n-3} = Q_{r_1,r_2,r_3}$ . Then, the coincidence is the event corresponding to the event  $C_{l,m}$  with  $m=3$  and  $l=2$  and is

$$C_{m=3,l=2} = \bigcup_{s=0}^{\infty} \bigcup_{r_1,r_2,r_3} \left\{ B_{r_1,r_2,r_3} \cap \{X_{r_1} = X_{r_2} = X_{r_3} = s+2\} \right\}. \tag{1.8}$$

We may define as many coincidences using (1.6). Is it important? Yes, it is certainly. For instance, understanding (or interpreting) and using such events (coincidences) for inferential and statistical, and other purposes has attracted researchers in many fields such as brain and cognitive sciences, psychology, law, discrete mathematics, physics and many more, obviously probability and statistics [7,8,10,12]. In this paper, we show (later in section 6) that a coincidence such the event  $C$  given by (1.4) may lead to interesting statistical inference results, as it can be interpreted as the null hypothesis of a statistical hypothesis test.

Next, let us consider the null hypothesis

$$H_0 : \lambda_1 = \lambda_2 = \dots = \lambda_n = \theta, \tag{1.9}$$

against the alternative hypothesis

$$H_a : \text{One of } \lambda_i, i=1,2,\dots,n \text{ is not equal } \theta. \tag{1.10}$$

We observe that  $\lambda_i = \Lambda_i K_i$  is a linear function of the time  $K_i$ . This makes the test hypothesis a linear function of time. Moreover, if  $K_i = K$  hours for all  $i$ , see (1.3), then  $X_{ik} \sim Poi(\Lambda_i)$ , where  $\Lambda_i = \lambda_i / K$ . Alternatively, we may consider another statistics test on slopes  $\Lambda_i$  which is steady in time. In that case, we should consider the null hypothesis

$$H_{00} : \Lambda_1 = \Lambda_2 = \dots = \Lambda_n = \theta / K, \tag{1.11}$$

against the alternative hypothesis

$$H_{aa} : \text{One of } \Lambda_i, i=1,2,\dots,n \text{ is not equal } \theta / K. \tag{1.12}$$

Returning to the call-center scenario, if  $H_0$  is true almost surely (with probability 1), then the centers become undistinguishable. This means that knowing the average number of calls in one center in a hour, for example  $\bar{X}_1 = \left( \sum_{k=1}^K X_{1k} \right) / K$ , implies we have enough

information for all centers. In this case,  $\bar{X}_1$  is a sufficient statistics for  $\theta$  [4], and if for instance  $\theta$  is large ( $\theta \geq 15$ ), then a  $100(1-\alpha)\%$  confidence interval (CI) for  $\theta$  can readily be obtained using the Central Limit Theorem (CLT) [4]. However taking into account the alternative hypothesis  $H_a$ , one has to be careful when constructing a  $100(1-\alpha)\%$  CI for  $\theta$  since  $\bar{X}_1$  may be a biased estimator for  $\theta$ .

Before we proceed to the main objectives of the paper, we shall first give the definition of the generalized hypergeometric function as it is an important mathematical tool that we are going to use in this paper.

**Definition 1.** The generalized hypergeometric function is a special function denoted as  ${}_pF_q(a_1, a_2, \dots, a_p; b_1, b_2, \dots, b_q; x)$  [1,11], where  $a_1, a_2, \dots, a_n$  and  $b_1, b_2, \dots, b_n$  are arbitrary constants. It is given by the power series

$${}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; x) = \sum_{s=0}^{\infty} \frac{(a_1)_s \dots (a_p)_s}{(b_1)_s \dots (b_q)_s} \frac{x^s}{s!}, \tag{1.13}$$

where  $(\beta)_s = \frac{\Gamma(\beta + s)}{\Gamma(\beta)}$  for any complex  $\beta$  and integer  $s$ ,

with  $(\beta)_0 = 1$ , and  $\Gamma(\beta)$  is the gamma function.

This paper has indeed two main goals. As a first goal, the probability and moments of  $C$ , see (1.4), is expressed in terms of the hypergeometric function  ${}_1F_n$ . Since the properties of this function is well known [1,11], its asymptotic expansion is used to derive the asymptotic approximations for the probability and moments of the event  $C$  for large  $\theta$  ( $\theta \geq 15$ ), under the null hypothesis  $H_0$ , see (1.9). And the probability  $P(C_{l,m})$ , where  $C_{l,m}$  is given by (1.6), is also expressed in term of  ${}_1F_n$ , and its asymptotic expression under the null hypothesis is also derived.

The second goal is to establish a chi-square ( $\chi^2$ ) goodness of fit test to examine  $H_0$ , see (1.9), via the variance of the coincidence, and to propose a new method to construct a  $100(1-\alpha)\%$  CI for the parameter  $\theta$  ( $\theta \geq 15$ ) using the variance of the coincidence.

The paper is organized as following. In section 2, formulas for the probabilities of the events  $C$  and  $C_{l,m}$ ,  $P(C)$  and  $P(C_{l,m})$ , are obtained in terms of the hypergeometric function  ${}_1F_n$  and under the null hypothesis  $H_0$  given in (1.9). The moments of  $C$  and its variance are expressed in terms of  ${}_1F_n$  in section 3. Since the asymptotic expansion of  ${}_1F_n$  is well known [11] as already mentioned above, it is used to obtain the asymptotic expressions of the probabilities  $P(C)$  and  $P(C_{l,m})$  in section 4 and that of the variance of  $C$  in section 5. The asymptotic expansions in sections 4 and 5 are valid for  $n \geq 3$ . In the case with  $n = 2$ , the moments are expressed in terms of the modified Bessel functions of the first kind in Appendix A. In section 6, a chi-square test to examine the null hypothesis test in (1.9) is established and a confidence interval for  $\theta$  is constructed using the variance of the coincidence obtained in section 5. Important discussion and conclusions are given in section 7.

## 2. The Probabilities $P(C)$ and $P(C_{l,m})$ in Terms of the Hypergeometric Function ${}_1F_n$

In this section, the probability of the event  $C$  and that of the event  $C_{l,m}$ ,  $P(C)$  and  $P(C_{l,m})$  respectively, are written in terms of the hypergeometric function  ${}_1F_n$ . The results are summarized in form of Theorems 1 and 2.

**Theorem 1.** The probability of the coincidence  $C$ , in (1.4), is given by

$$P(C) = e^{-\sum_{i=1}^n \lambda_i} {}_1F_n \left( 1; 1, 1, \dots, 1; \prod_{i=1}^n \lambda_i \right). \quad (2.14)$$

Then, under the null hypothesis  $H_0$ , see (1.9),

$$P(C | \lambda_i = \theta, i = 1, 2, \dots, n) = e^{-n\theta} {}_1F_n \left( 1; 1, 1, \dots, 1; \theta^n \right). \quad (2.15)$$

**Proof.** From (1.5), we have

$$\begin{aligned} P(C) &= P \left( \bigcup_{s=0}^{\infty} \{X_1 = X_2 = \dots = X_n = s\} \right) \\ &= \sum_{s=0}^{\infty} \frac{\left( \prod_{i=1}^n \lambda_i \right)^s}{(s!)^n} \\ &= e^{-\sum_{i=1}^n \lambda_i} \sum_{s=0}^{\infty} \frac{\Gamma(s+1)}{\Gamma^n(s+1)} \frac{\left( \prod_{i=1}^n \lambda_i \right)^s}{s!} \\ &= e^{-\sum_{i=1}^n \lambda_i} \sum_{s=0}^{\infty} \frac{(1)_s}{[(1)_s]^n} \frac{\left( \prod_{i=1}^n \lambda_i \right)^s}{s!} \\ &= e^{-\sum_{i=1}^n \lambda_i} {}_1F_n \left( 1; 1, 1, \dots, 1; \prod_{i=1}^n \lambda_i \right). \end{aligned} \quad (2.16)$$

Hence, under  $H_0$ ,

$$\begin{aligned} P(C | \lambda_i = \theta, i = 1, 2, \dots, n) &= e^{-\sum_{i=1}^n \lambda_i} {}_1F_n \left( 1; 1, 1, \dots, 1; \prod_{i=1}^n \lambda_i \right) \\ &= e^{-n\theta} {}_1F_n \left( 1; 1, 1, \dots, 1; \theta^n \right). \end{aligned} \quad (2.17)$$

It is known that, for  $n = 2$ ,  $P(C)$  can be written in terms of the Modified Bessel of the first kind [7,14]. Theorem 1 generalizes the results in [7] for any integer  $n \geq 2$ . We also would like to point out that the series that Griffiths [7] calls remodified Bessel functions are in fact hypergeometric functions  ${}_1F_n$  [1,11].

**Theorem 2.** The probability of any event defined using (1.6) is given by

$$P(C_{l,m}) = \binom{n}{m} e^{-\sum_{i=1}^n \lambda_i} \left( \prod_{i=1}^m \lambda_i \right)^l {}_1F_n \left( 1; l+1, \dots, l+1, 1, \dots, 1; \prod_{i=1}^n \lambda_i \right). \quad (2.18)$$

Then, under the null hypothesis  $H_0$ ,

$$\begin{aligned} P(C_{l,m} | \lambda_i = \theta, i = 1, 2, \dots, n) &= \binom{n}{m} e^{-n\theta} \theta^{lm} {}_1F_n \left( 1; l+1, \dots, l+1, 1, \dots, 1; \theta^n \right). \end{aligned} \quad (2.19)$$

**Proof.**

$$\begin{aligned}
 &P(C_{l,m}) \\
 &= P\left(\bigcup_{s=0}^{\infty} \bigcup_{\eta_1, \dots, \eta_m} \left\{ B_{\eta_1, \dots, \eta_m} \cap \left\{ \begin{matrix} X_{\eta_1} = \dots \\ = X_{\eta_m} = s+l \end{matrix} \right\} \right\}\right) \\
 &= \binom{n}{m} \sum_{s=0}^{\infty} \frac{\prod_{i=1}^m e^{-\lambda_i} \lambda_i^{s+l} \prod_{i=m+1}^n e^{-\lambda_i} \lambda_i^s}{[(s+l)!]^m [(s)!]^{n-m}} \\
 &= \binom{n}{m} e^{-\sum_{i=1}^n \lambda_i} \left(\prod_{i=1}^m \lambda_i\right)^l \\
 &\quad \times \sum_{s=0}^{\infty} \frac{s!}{[(s+l)!]^m [(s)!]^{n-m}} \frac{\left(\prod_{i=1}^n \lambda_i\right)^s}{s!} \\
 &= \binom{n}{m} e^{-\sum_{i=1}^n \lambda_i} \left(\prod_{i=1}^m \lambda_i\right)^l \\
 &\quad \times \sum_{s=0}^{\infty} \frac{\Gamma(s+1)}{\Gamma^m(s+l+1) \Gamma^{n-m}(s+1)} \frac{\left(\prod_{i=1}^n \lambda_i\right)^s}{s!} \\
 &= \binom{n}{m} e^{-\sum_{i=1}^n \lambda_i} \left(\prod_{i=1}^m \lambda_i\right)^l \\
 &\quad \times \sum_{s=0}^{\infty} \frac{(1)_s}{[(l+1)_s]^m [(1)_s]^{n-m}} \frac{\left(\prod_{i=1}^n \lambda_i\right)^s}{s!} \\
 &= \binom{n}{m} e^{-\sum_{i=1}^n \lambda_i} \left(\prod_{i=1}^m \lambda_i\right)^l \\
 &\quad \times {}_1F_n\left(1; l+1, \dots, l+1, 1, \dots, 1; \prod_{i=1}^n \lambda_i\right), \tag{2.20}
 \end{aligned}$$

where

$$\binom{n}{m} = \frac{n!}{(n-m)!m!}. \tag{2.21}$$

Hence, under  $H_0$ ,

$$\begin{aligned}
 &P(C_{l,m} \mid \lambda_i = \theta, i = 1, 2, \dots, n) \\
 &= \binom{n}{m} e^{\sum_{i=1}^m \lambda_i} \left(\prod_{i=1}^m \lambda_i\right)^l \\
 &\quad {}_1F_n\left(1; l+1, \dots, l+1, 1, \dots, 1; \prod_{i=1}^n \lambda_i\right) \\
 &= \binom{n}{m} e^{-n\theta} \theta^{lm} {}_1F_n\left(1; l+1, \dots, l+1, 1, \dots, 1; \theta^n\right). \tag{2.22}
 \end{aligned}$$

### 3. Moments and the Variance Associated with the Event $C$ in Terms of the Hypergeometric Function ${}_1F_n$

Here, the moments and the variance associated with the coincidence  $C$  are written in terms of the hypergeometric function  ${}_1F_n$ . We mainly focus on the coincidence  $C$  rather than  $C_{l,m}$ , the reason of doing so will become clear later in section 6. Results are summarized in Lemma 1 and Theorem 2.

**Lemma 1.** The  $\gamma^{th}$  moment  $\mu_\gamma$  of the coincidence  $C$  is given by

$$\mu_\gamma = e^{-\sum_{i=1}^n \lambda_i} \left(\prod_{i=1}^n \lambda_i\right) {}_1F_n\left(1; 1, \dots, 1, 2, \dots, 2; \prod_{i=1}^n \lambda_i\right), \tag{3.23}$$

where  $\gamma = 1, 2, 3, \dots$ .

**Proof.** The  $\gamma^{th}$  moment associated with the coincidence is given by

$$\begin{aligned}
 \mu_\gamma &= \sum_{s=0}^{\infty} s^\gamma \frac{\prod_{i=1}^n e^{-\lambda_i} \lambda_i^s}{(s!)^n} \\
 &= e^{-\sum_{i=1}^n \lambda_i} \sum_{s=1}^{\infty} \frac{s^\gamma \left(\prod_{i=1}^n \lambda_i\right)^s}{(s!)^n} \\
 &= e^{-\sum_{i=1}^n \lambda_i} \left(\prod_{i=1}^n \lambda_i\right) \sum_{s=0}^{\infty} \frac{(s+1)^\gamma \left(\prod_{i=1}^n \lambda_i\right)^s}{[(s+1)!]^n} \\
 &= e^{-\sum_{i=1}^n \lambda_i} \left(\prod_{i=1}^n \lambda_i\right) \sum_{s=0}^{\infty} \frac{\Gamma(s+1)}{\Gamma^\gamma(s+1) \Gamma^{n-\gamma}(s+2)} \frac{\left(\prod_{i=1}^n \lambda_i\right)^s}{s!} \\
 &= e^{-\sum_{i=1}^n \lambda_i} \left(\prod_{i=1}^n \lambda_i\right) \sum_{s=0}^{\infty} \frac{(1)_s}{[(1)_s]^\gamma [(2)_s]^{n-\gamma}} \frac{\left(\prod_{i=1}^n \lambda_i\right)^s}{s!} \\
 &= e^{-\sum_{i=1}^n \lambda_i} \left(\prod_{i=1}^n \lambda_i\right) {}_1F_n\left(1; 1, \dots, 1, 2, \dots, 2; \prod_{i=1}^n \lambda_i\right). \tag{3.24}
 \end{aligned}$$

**Theorem 3.** Under  $H_0$ , the first order moment (mean) associated with the coincidence  $C$  is

$$\mu_1 = e^{-n\theta} \theta^n {}_1F_n\left(1; 1, 2, \dots, 2; \theta^n\right), \tag{3.25}$$

while the second order moment is

$$\mu_2 = e^{-n\theta} \theta^n {}_1F_n\left(1; 1, 1, 2, \dots, 2; \theta^n\right). \tag{3.26}$$

Thus the variance associated with the coincidence is

$$\sigma^2 = \mu_2 - (\mu_1)^2 = e^{-n\theta} \theta^n {}_1F_n(1; 1, 1, 2, \dots, 2; \theta^n) - e^{-2n\theta} \theta^{2n} \left[ {}_1F_n(1; 1, 2, \dots, 2; \theta^n) \right]^2. \tag{3.27}$$

**Proof.** Setting  $\gamma = 1$  in (3.24) in Lemma 1, and using the fact that under  $H_0$ ,  $\lambda_i = \theta, i = 1, 2, \dots, n$ , gives

$$\begin{aligned} \mu_1 &= e^{-n\theta} \theta^n \sum_{s=0}^{\infty} \frac{(1)_s}{(1)_s [(2)_s]^{n-1}} \frac{(\theta^n)^s}{s!} \\ &= e^{-n\theta} \theta^n {}_1F_n(1; 1, 2, \dots, 2; \theta^n), \end{aligned} \tag{3.28}$$

while setting  $\gamma = 2$  and taking into account the null hypothesis gives

$$\begin{aligned} \mu_2 &= e^{-n\theta} \theta^n \sum_{s=0}^{\infty} \frac{(1)_s}{[(1)_s]^2 [(2)_s]^{n-2}} \frac{(\theta^n)^s}{s!} \\ &= e^{-n\theta} \theta^n {}_1F_n(1; 1, 1, 2, \dots, 2; \theta^n) \end{aligned} \tag{3.29}$$

Hence, using the fact that the variance  $\sigma^2 = \mu_2 - (\mu_1)^2$  gives (3.27).

### 4. Asymptotic Evaluation of $P(C)$ and $P(C_{l,m})$ under $H_0$ , and for Large $\theta$

In this section, we consider the case with  $n > 2$ , and derive the asymptotic approximations of the probabilities of  $C$  and  $C_{l,m}$  when  $\theta$  is large. The case with  $n = 2$  is considered later in Appendix A. The main reason is that the formulas used here do not work for  $n = 2$ . The results are summarized in Theorems 4 and 5.

**Theorem 4.** Under  $H_0$ , see (1.9), if  $n > 2$ , then for large  $\theta$  ( $\theta \geq 15$ ),

$$P(C) \sim 2(2\pi\theta)^{1/2-n/2} n^{-1/2}. \tag{4.30}$$

**Proof.** To prove (4.30), we use formulas (16.11.1), (16.11.3), (16.11.4) and (16.11.9) in [11]. Setting  $p = 1$  and  $q = n$  in (16.11.3) yields  $\kappa = n$  and  $\nu = 1 - n + (n - 1)/2 = 1/2 - n/2$ . Substituting  $p = 1$  and  $q = n$ ,  $\kappa = n$  and  $\nu = 1/2 - n/2$  in (16.11.1) gives

$$\begin{aligned} E_{p,q}(z) &= E_{1,n}(z) = (2\pi)^{1/2-n/2} n^{n/2-1/2} n^{-1/2} \\ &\times e^{nz^{1/n}} \sum_{j=0}^{\infty} c_j (nz^{1/n})^{1/2-n/2-lm-j} \\ &= (2\pi)^{1/2-n/2} n^{-1/2} (z^{1/n})^{1/2-n/2-lm} \\ &\times e^{nz^{1/n}} \sum_{j=0}^{\infty} c_j (nz^{1/n})^{-j}, \end{aligned} \tag{4.31}$$

where  $c_j$  is given by formula (16.11.4) in [11]. We now let  $z = -\theta^n$ , use formula (16.11.9) in [11] and obtain

$$\begin{aligned} &{}_1F_n(1; 1, 1, \dots, 1; \theta^n) \\ &= {}_1F_n(1; 1, 1, \dots, 1; -(-\theta^n)) = 2E_{1,n}(\theta^n) \\ &= (2\pi)^{1/2-n/2} n^{-1/2} (\theta)^{1/2-n/2} \\ &\times e^{nz^{1/n}} \sum_{j=0}^{\infty} c_j (n\theta)^{-j}, \theta \gg 1. \end{aligned} \tag{4.32}$$

Hence, dropping the terms corresponding to  $j \geq 1$  while keeping the leading term corresponding to  $j = 0$  yields

$$\begin{aligned} P(C) &= e^{-n\theta} {}_1F_n(1; 1, 1, \dots, 1; \theta^n) \sim 2(2\pi\theta)^{1/2-n/2} n^{-1/2}. \end{aligned} \tag{4.33}$$

**Theorem 5.** Under  $H_0$ , see (1.9), if  $n > 2$ , then for large  $\theta$  ( $\theta \geq 15$ ),

$$P(C_{l,m}) \sim 2 \binom{n}{m} (2\pi\theta)^{1/2-n/2} n^{-1/2}. \tag{4.34}$$

Although, the proof of Theorem 5 is similar to that of Theorem 4, we prefer to show it here. For instance, one should doubt about (4.34) since the subscript  $l$  in (2.19) does not appear in its asymptotic approximation (4.34).

**Proof.** To prove (4.35), we use formulas (16.11.1), (16.11.3), (16.11.4) and (16.11.9) in [11] as before. Setting  $p = 1$  and  $q = n$  in (16.11.3) yields  $\kappa = n$ , and

$$\begin{aligned} \nu &= 1 - [(l+1)m + 1(n-m)] + (n-1)/2 \\ &= 1/2 - lm - n/2. \end{aligned}$$

Substituting  $p = 1$  and  $q = n$ ,  $\kappa = n$  and  $\nu = 1/2 - lm - n/2$  in (16.11.1) gives

$$\begin{aligned} E_{p,q}(z) &= E_{1,n}(z) = (2\pi)^{1/2-n/2} n^{lm+n/2-1/2} \\ &\times n^{-1/2} e^{nz^{1/n}} \sum_{j=0}^{\infty} c_j (nz^{1/n})^{1/2-n/2-lm-j} \\ &= (2\pi)^{1/2-n/2} n^{-1/2} (z^{1/n})^{1/2-n/2-lm} \\ &\times e^{nz^{1/n}} \sum_{j=0}^{\infty} c_j (nz^{1/n})^{-j}, \end{aligned} \tag{4.35}$$

where  $c_j$  is given by formula (16.11.4) in [11].

We now let  $z = -\theta^n$ , use formula (16.11.9) in [11] and obtain

$$\begin{aligned} &\sum_{s=0}^{\infty} \frac{(1)_s}{[(l+1)_s]^m [(1)_s]^{n-m}} \frac{(\theta^n)^s}{s!} \\ &= {}_1F_n(1; l+1, \dots, l+1, 1, \dots, 1; \theta^n) \\ &= {}_1F_n(1; l+1, \dots, l+1, 1, \dots, 1; -(-\theta^n)) = 2E_{1,n}(\theta^n) \\ &= (2\pi)^{1/2-n/2} n^{-1/2} (\theta)^{1/2-n/2-lm} \\ &\times e^{nz^{1/n}} \sum_{j=0}^{\infty} c_j (n\theta)^{-j}, \theta \gg 1. \end{aligned} \tag{4.36}$$

Hence, dropping the terms corresponding to  $j \geq 1$  while keeping the leading term corresponding to  $j = 0$  yields

$$\begin{aligned}
 &P(C_{l,m}) \\
 &= \binom{n}{m} e^{-n\theta} \theta^{lm} {}_1F_n(1; l+1, \dots, l+1, 1, \dots, 1; \theta^n) \\
 &= 2 \binom{n}{m} (2\pi)^{1/2-n/2} n^{-1/2} \theta^{1/2-n/2} \tag{4.37} \\
 &\sim 2 \binom{n}{m} (2\pi\theta)^{1/2-n/2} n^{-1/2}.
 \end{aligned}$$

### 5. Asymptotic Evaluation of Moments of the Coincidence for Large $\theta$ , and under $H_0$

For the same reason as in section 4, we also consider that  $n > 2$ , and as before, we use formulas (16.11.1), (16.11.3), (16.11.4) and (16.11.9) in [11] to obtain asymptotic expressions for the  $\gamma^{th}$  moment  $\mu_\gamma$  and the variance of the coincidence  $C$ ,  $\sigma^2$ , which are valid when  $\theta$  is large. Results are given in Theorem 6.

**Theorem 6.** Under  $H_0$ , see (1.9), if  $n > 2$ , then for large  $\theta$  ( $\theta \geq 15$ ),

$$\mu_\gamma \sim 2(2\pi)^{1/2-n/2} n^{-1/2} \theta^{1/2+\gamma-n/2}. \tag{5.38}$$

And the variance is asymptotically given by

$$\sigma^2 \sim 2(2\pi)^{1/2-n/2} n^{-1/2} \theta^{5/2-n/2} - 4n^{-1} (2\pi)^{1-n} \theta^{3-n}. \tag{5.39}$$

**Proof.** To derive an asymptotic expression for  $\mu_\gamma$  valid for large  $\theta$  ( $\theta \geq 15$ ), we set  $p=1$  and  $q=n$  in (16.11.3) in [11] and obtain  $\kappa = n$ , and

$$\nu = 1 - [\gamma + 2(n - \gamma)] + (n - 1) / 2 = 1 / 2 - \gamma - 3n / 2.$$

Substituting  $p=1$  and  $q=n$ ,  $\kappa = n$  and  $\nu = 1 / 2 - \gamma - (3n) / 2$  in (16.11.1) gives

$$\begin{aligned}
 &E_{p,q}(z) = E_{1,n}(z) \\
 &= (2\pi)^{1/2-n/2} n^{n/2-1/2} n^{-1/2} e^{nz^{1/n}} \\
 &\times \sum_{j=0}^{\infty} c_j (nz^{1/n})^{1/2+\gamma-3n/2-j} \tag{5.40} \\
 &= (2\pi)^{1/2-n/2} n^{-1/2} (z^{1/n})^{1/2+\gamma-3n/2} \\
 &\times e^{nz^{1/n}} \sum_{j=0}^{\infty} c_j (nz^{1/n})^{-j},
 \end{aligned}$$

where  $c_j$  is given by formula (16.11.4) in [11].

Setting  $z = -\theta^n$  in (5.40), taking into account the null hypothesis  $H_0$ , and using formula (16.11.9) in [11] as before yields

$$\begin{aligned}
 &\sum_{s=0}^{\infty} \frac{(1)_s}{[(1)_s]^\gamma [(2)_s]^{n-\gamma}} \frac{(\theta^n)^s}{s!} \\
 &= {}_1F_n(1; 2, \dots, 2, 1, \dots, 1; \theta^n) \\
 &= {}_1F_n(1; 1, \dots, 1, 2, \dots, 2; -(\theta^n)) \tag{5.41} \\
 &= 2E_{1,n}(\theta^n) = (2\pi)^{1/2-n/2} n^{-1/2} \\
 &\times (\theta)^{1/2+\gamma-3n/2} e^{nz^{1/n}} \sum_{j=0}^{\infty} c_j (n\theta)^{-j}, \theta \gg 1.
 \end{aligned}$$

Hence, dropping the terms corresponding to  $j \geq 1$  while keeping the leading term corresponding to  $j = 0$  yields

$$\begin{aligned}
 &\mu_\gamma = e^{-n\theta} \theta^n {}_1F_n(1; 1, \dots, 1, 2, \dots, 2; \theta^n) \\
 &\sim 2(2\pi)^{1/2-n/2} n^{-1/2} \theta^{1/2+\gamma-n/2}. \tag{5.42}
 \end{aligned}$$

And hence,

$$\begin{aligned}
 &\sigma^2 = \mu_2 - (\mu_1)^2 = \mu_{\gamma=2} - (\mu_{\gamma=1})^2 \\
 &\sim 2(2\pi)^{1/2-n/2} n^{-1/2} \theta^{5/2-n/2} \\
 &- \left( 2(2\pi)^{1/2-n/2} n^{-1/2} \theta^{3/2-n/2} \right)^2 \tag{5.43} \\
 &= 2(2\pi)^{1/2-n/2} n^{-1/2} \theta^{5/2-n/2} \\
 &- 4n^{-1} (2\pi)^{1-n} \theta^{3-n}
 \end{aligned}$$

### 6. A Chi-square Goodness of Fit Test and Confidence Interval for $\theta$ via Coincidence

In this section, we use the fact that, under certain conditions that we are going to mention shortly, the coincidence  $C$  can be interpreted as the null hypothesis, see (1.9) or (1.11) and thus establish a chi-square goodness of fit test to examine the null hypothesis  $H_0$  via the variance of the coincidence  $C$ , and consequently construct a  $100(1 - \alpha)\%$  CI for the parameter  $\theta$  if there is no significance evidence to reject  $H_0$ . We also consider that  $\theta$  is large ( $\theta \geq 15$ ) so that we apply the Central Limit Theorem (CLT). This justifies the derivation of the asymptotic approximation of the variance of the coincidence  $C$  for large  $\theta$  in section 3.

For  $n$  centers and each receiving a total of  $K_i$  calls at the time  $K_i$ , we define  $X_i = \sum_{k=1}^{K_i} X_{ik}, i = 1, 2, 3, \dots, n$  the overall standard deviation as

$$s_n^2 = \frac{1}{\left( \sum_{i=1}^n K_i \right) - 1} \sum_{i=1}^n \sum_{k=1}^{K_i} (X_{ik} - GM)^2, \tag{6.44}$$

where the generalized mean  $GM$  is given by



$$GM = \frac{1}{n} \sum_{i=1}^n \frac{1}{K_i} \sum_{k=1}^{K_i} X_{ik} = \frac{1}{n} \sum_{i=1}^n \frac{X_i}{K_i} = \frac{1}{n} \sum_{i=1}^n \overline{X}_i. \quad (6.45)$$

If  $\Lambda_i \geq 15$ , then  $X_{ik}$  approximately follows a normal distribution with mean  $\Lambda_i$  and variance  $\Lambda_i$  by the CLT [4]. Thus,

$$X_{ik} \sim N(\Lambda_i, \Lambda_i). \quad (6.46)$$

After a certain amount of time,  $K_i$  hours, and under the null hypothesis  $H_0$ ,

$$X_{ik} \sim N(K_i \Lambda_i, K_i \Lambda_i). \quad (6.47)$$

**Theorem 7.** If  $\overline{X}_i = \frac{1}{K_i} \sum_{k=1}^{K_i} X_{ik}$  is a minimum variance unbiased estimator (MVUE) [4] for  $\lambda_i$  and  $K_i = K$  for all  $i$ , and  $H_0$  can not be rejected, then the probability that  $H_0$  is true is approximately given by

$$P(H_0 : \text{true}) \approx P(C), \quad (6.48)$$

where  $P(C)$  is given by (2.15) or (4.33) when the parameter  $\theta \gg 1$ .

**Proof.** The proof is straightforward. If  $\overline{X}_i = \frac{1}{K_i} \sum_{k=1}^{K_i} X_{ik}$  is a minimum variance unbiased estimator (MVUE) [4] for  $\lambda_i$  and  $K_i = K$  for all  $i$ , and the null hypothesis can not be rejected, then

$$\begin{aligned} P(H_0 : \text{true}) &= P(\lambda_1 = \lambda_2 = \dots = \lambda_n) \\ &\approx P\left(\frac{X_1}{K} = \frac{X_2}{K} = \dots = \frac{X_n}{K}\right) \\ &= P(X_1 = X_2 = \dots = X_n) = P(C). \end{aligned} \quad (6.49)$$

Moreover, if there is no evidence to reject  $H_0$ , we shall expect a coincidence to occur after  $K$  hours, and hence,

$$\begin{aligned} X_1 &= \sum_{k=1}^K X_{1k}, X_2 = \sum_{k=1}^K X_{2k}, \\ \dots, X_n &= \sum_{k=1}^K X_{nk} \sim N(\theta, \sigma^2), \end{aligned} \quad (6.50)$$

where  $\sigma^2$  is the variance of  $C$  and is approximated by (5.39) if  $\theta$  is large. In that case,

$$\frac{(nK-1)s_n^2}{\sigma^2} \sim \chi_{nK-1}^2, \quad (6.51)$$

where

$$s_n^2 = \frac{1}{nK-1} \sum_{i=1}^n \sum_{k=1}^K (X_{ik} - GM)^2. \quad (6.52)$$

Using (5.39) gives

$$\frac{(nK-1)s_n^2}{2(2\pi)^{1/2-n/2} n^{-1/2} \theta^{5/2-n/2} - 4n^{-1} (2\pi)^{1-n} \theta^{3-n}} \sim \chi_{nK-1}^2 \quad (6.53)$$

Thus a Chi-square ( $\chi^2$ ) goodness of fit test can now be carried out as following. The null hypothesis  $H_0$  will be rejected if

$$\begin{aligned} &\frac{(nK-1)s_n^2}{2(2\pi)^{1/2-n/2} n^{-1/2} \theta^{5/2-n/2} - 4n^{-1} (2\pi)^{1-n} \theta^{3-n}} \\ &< \chi_{nK-1}^2(\alpha/2) \end{aligned} \quad (6.54)$$

or (and)

$$\begin{aligned} &\frac{(nK-1)s_n^2}{2(2\pi)^{1/2-n/2} n^{-1/2} \theta^{5/2-n/2} - 4n^{-1} (2\pi)^{1-n} \theta^{3-n}} \\ &> \chi_{nK-1}^2(1-\alpha/2), \end{aligned} \quad (6.55)$$

where  $\alpha$  is the significance level needed to be achieved by the decision maker.

If that is not the case, there is no significant evidence to reject  $H_0$ , a  $100(1-\alpha)\%$  confidence interval for the variance of  $C$ ,  $\sigma^2$  can eventually be obtained,

$$\frac{(nK-1)s_n^2}{\chi_{nK-1}^2(\alpha/2)} < \sigma^2 < \frac{(nK-1)s_n^2}{\chi_{nK-1}^2(1-\alpha/2)}, \quad (6.56)$$

and leads to a confidence interval for  $\theta$ ,

$$\theta_1 < \theta < \theta_2, \quad (6.57)$$

where  $\theta_1$  and  $\theta_2$  respectively satisfy

$$\begin{aligned} &2(2\pi)^{1/2-n/2} n^{-1/2} (\theta_1)^{5/2-n/2} \\ &- 4n^{-1} (2\pi)^{1-n} (\theta_1)^{3-n} = \frac{(nK-1)s_n^2}{\chi_{nK-1}^2(1-\alpha/2)} \end{aligned} \quad (6.58)$$

and

$$\begin{aligned} &2(2\pi)^{1/2-n/2} n^{-1/2} (\theta_2)^{5/2-n/2} \\ &- 4n^{-1} (2\pi)^{1-n} (\theta_2)^{3-n} = \frac{(nK-1)s_n^2}{\chi_{nK-1}^2(\alpha/2)}. \end{aligned} \quad (6.59)$$

For  $n=3$  and  $n=5$ , (6.58) and (6.59) can readily be solve, while for  $n=2,4$  and  $n>5$ , (6.58) and (6.59) can readily be solved using basic numerical methods.

For  $n=3$ ,

$$\theta_1 = \frac{1}{(3)^{1/2} \pi} + (3)^{1/2} \pi \frac{(3K-1)s_3^2}{\chi_{3K-1}^2(1-\alpha/2)} \quad (6.60)$$

and

$$\theta_2 = \frac{1}{(3)^{1/2} \pi} + (3)^{1/2} \pi \frac{(3K-1)s_3^2}{\chi_{3K-1}^2(\alpha/2)}. \quad (6.61)$$

For  $n=5$ ,

$$\theta_1 = \frac{1}{(2)^{1/2} (5)^{1/4} \pi} \left[ \frac{1}{1 - 2(5)^{1/2} \pi^2 \frac{(5K-1)s_5^2}{\chi_{5K-1}^2(1-\alpha/2)}} \right]^{1/2} \quad (6.62)$$

$$\theta_2 = \frac{1}{(2)^{1/2}(5)^{1/4}\pi} \left[ \frac{1}{1-2(5)^{1/2}\pi^2 \frac{(5K-1)s_5^2}{\chi_{5K-1}^2(\alpha/2)}} \right]^{1/2} \quad (6.63)$$

### 7. Discussion and Conclusion

Having expressed the probability of the coincidence  $C$  given by (1.4) and that of  $C_{l,m}$  given by (1.6) and the moments of  $C$  in terms of the special function  ${}_1F_n$  for  $n > 2$  (Theorem 1, Theorem 2 and Theorem 3) on one hand, and in terms of the modified Bessel function of the first kind of order  $\phi$ ,  $I_\phi$  (Corollary 1) if  $n = 2$  on another, we also derived their corresponding approximations (Theorem 4, Theorem 5 and Theorem 6).

Griffths [7] defined some coincidences, for example the near coincidence (see (1.7)), and expressed their probabilities in terms of some power series that he named re-modified Bessel functions. Here we have shown in Theorem 1 that these series are indeed hypergeometric functions  ${}_1F_n$  for any event that can be defined using (1.6).

We also found that if  $\bar{X}_i$  is a minimum variance unbiased estimator (MVUE), for  $\lambda_i$  [4] and  $K_i = K$  for all  $i$  and there is no evidence to reject the null hypothesis  $H_0 : \lambda_1 = \lambda_2 = \dots = \lambda_n = \theta$ , where  $\theta$  is some positive number, then the probability that  $H_0$  is true can be approximated by that of the coincidence  $C$ , (Theorem 7). One may understand this fact this way. If it happens that  $H_0$  is true almost surely then  $P(C) = 1$ . This scenario is no longer a coincidence because the centers become undistinguishable. But since  $P(C) < 1$  always, then we shall be able to distinguish one center to another.

Moreover, if the probability that  $H_0$  is true almost equals the probability of the coincidence  $C$ , and the null hypothesis  $H_0$  can not be rejected, then the variance of  $X_i, i = 1, 2, \dots, n$  is that of the coincidence  $H_0$ . In that case, with a  $\chi^2$  test as established in section 6, if there is no evidence to reject  $H_0$ , one can readily construct a  $100(1-\alpha)\%$  confidence interval (CI) for the parameter  $\theta$  (section 6 and appendix A) using the variance of the coincidence  $C$  (or via coincidence).

The results obtained in this paper can be used, for example, to achieve better results in telephone traffic measurements [2].

#### A. A Chi-square Goodness of Fit Test and $100(1-\alpha)\%$ CI for $\theta$ via Coincidence for $n = 2$

In sections 4 and 5, we considered that the number of call centers  $n$  is greater or equals  $3 (n \geq 3)$ . As mentioned

before, the main reason is that the asymptotic expansion of the hypergeometric functions  ${}_1F_n$  for  $\theta \gg 1$  in sections 4 and 5 is valid when  $n \geq 3$ . Here, we consider that  $n = 2$ , write the first and second moments  $\mu_1$  and  $\mu_2$  in terms of the modified Bessel function of the first kind, and use the asymptotic expansion of the modified Bessel function [1] to derive the asymptotic approximations for  $\mu_1$  and  $\mu_2$  valid for  $\theta \gg 1$ . A  $\chi^2$  goodness of fit test is established and hence a  $100(1-\alpha)\%$  CI for the parameter  $\theta$  is obtained if the null hypothesis  $H_0$  is not rejected.

**Definition 2.** The modified Bessel function of the first kind of order  $\phi$ ,  $I_\phi$ , is the series (formula 9.6.10 in [1])

$$I_\phi(x) = \left(\frac{1}{2}x\right)^\phi \sum_{s=0}^{\infty} \frac{\left(\frac{1}{4}x^2\right)^s}{\Gamma(\phi+s+1)s!}, \quad (A.64)$$

where  $\phi$  is some number.

**Corollary 1.** For  $n = 2$ , the first and second order moments of the coincidence  $C$ , see Theorem 3, are

$$\mu_1 = e^{-2\theta} \theta^2 {}_0F_1\left(;2;\theta^2\right) = \theta e^{-2\theta} I_1(2\theta) \quad (A.65)$$

and

$$\mu_2 = e^{-2\theta} \theta^2 {}_0F_1\left(;1;\theta^2\right) = \theta^2 e^{-2\theta} I_0(2\theta) \quad (A.66)$$

respectively.

Thus the variance of the coincidence is

$$\begin{aligned} \sigma^2 &= \mu_2 - (\mu_1)^2 \\ &= \theta^2 e^{-2\theta} I_0(2\theta) - \theta^2 e^{-4\theta} [I_1(2\theta)]^2. \end{aligned} \quad (A.67)$$

**Proof.** Setting  $n = 2$  in Theorem 3 yields

$$\begin{aligned} \mu_1 &= e^{-2\theta} \theta^2 {}_1F_2\left(1;1,2;\theta^2\right) = e^{-2\theta} \theta^2 {}_0F_1\left(;2;\theta^2\right) \\ &= e^{-2\theta} \theta^2 \sum_{s=0}^{\infty} \frac{1}{(2)_s} \frac{(\theta^2)^s}{s!}. \end{aligned} \quad (A.68)$$

Rearranging terms, setting  $x = 2\theta$  and  $\phi = 1$  in (A.64) gives

$$\begin{aligned} \mu_1 &= e^{-2\theta} \theta^2 \sum_{s=0}^{\infty} \frac{1}{(2)_s} \frac{(\theta^2)^s}{s!} \\ &= \theta e^{-2\theta} \left(\frac{1}{2}(2\theta)\right)^1 \sum_{s=0}^{\infty} \frac{\left(\frac{1}{4}(2\theta)^2\right)^s}{\Gamma(1+s+1)s!} \\ &= \theta e^{-2\theta} I_1(2\theta). \end{aligned} \quad (A.69)$$

On the other hand,

$$\begin{aligned} \mu_2 &= e^{-2\theta} \theta^2 {}_1F_2\left(1;1,1;\theta^2\right) \\ &= e^{-2\theta} \theta^2 {}_0F_1\left(;1;\theta^2\right) \\ &= e^{-2\theta} \theta^2 \sum_{s=0}^{\infty} \frac{1}{(1)_s} \frac{(\theta^2)^s}{s!}. \end{aligned} \quad (A.70)$$



Rearranging terms, setting  $x = 2\theta$  and  $\phi = 0$  in (A.64) gives

$$\begin{aligned} \mu_2 &= e^{-2\theta} \theta^2 \sum_{s=0}^{\infty} \frac{1}{(2)_s} \frac{(\theta^2)^s}{s!} \\ &= \theta^2 e^{-2\theta} \left(\frac{1}{2}(2\theta)\right)^0 \sum_{s=0}^{\infty} \frac{\left(\frac{1}{4}(2\theta)^2\right)^s}{\Gamma(0+s+1)s!} \quad (\text{A.71}) \\ &= \theta^2 e^{-2\theta} I_0(2\theta). \end{aligned}$$

Hence, the variance of the coincidence is

$$\sigma^2 = \mu_2 - (\mu_1)^2 = \theta^2 e^{-2\theta} I_0(2\theta) - \theta^2 e^{-4\theta} [I_1(2\theta)]^2. \quad (\text{A.72})$$

**Theorem 8.** If  $\theta \gg 1$ , then for  $n = 2$ , the variance of the coincidence  $C$ ,  $\sigma^2$ , is asymptotically given by

$$\sigma^2 = \frac{\theta^{3/2}}{2(\pi)^{1/2}} - \frac{\theta}{4\pi}. \quad (\text{A.73})$$

**Proof.** The asymptotic approximation of  $I_\phi(2\theta), \theta \gg 1$  (formula 9.6.10 in [1]), is

$$I_\phi(2\theta) = \frac{e^{2\theta}}{(4\pi\theta)^{1/2}}. \quad (\text{A.74})$$

for any  $\phi$ . Then,

$$\mu_1 = \theta e^{-2\theta} I_1(2\theta) \sim \theta e^{-2\theta} \frac{e^{2\theta}}{(4\pi\theta)^{1/2}} = \left(\frac{\theta}{4\pi}\right)^{1/2} \quad (\text{A.75})$$

and

$$\begin{aligned} \mu_2 &= \theta^2 e^{-2\theta} I_0(2\theta) \sim \theta^2 e^{-2\theta} \frac{e^{2\theta}}{(4\pi\theta)^{1/2}} \\ &= \frac{\theta^{3/2}}{2(\pi)^{1/2}}. \end{aligned} \quad (\text{A.76})$$

Hence,

$$\sigma^2 = \mu_2 - (\mu_1)^2 = \frac{\theta^{3/2}}{2(\pi)^{1/2}} - \frac{\theta}{4\pi}. \quad (\text{A.77})$$

Following section 6, a chi-square goodness of fit test can be conducted as following.

The null hypothesis  $H_0$  will be rejected if

$$\frac{(2K-1)s_2^2}{\frac{\theta^{3/2}}{2(\pi)^{1/2}} - \frac{\theta}{4\pi}} < \chi_{2K-1}^2(\alpha/2) \quad (\text{A.78})$$

or (and)

$$\frac{(2K-1)s_2^2}{\frac{\theta^{3/2}}{2(\pi)^{1/2}} - \frac{\theta}{4\pi}} > \chi_{2K-1}^2(1-\alpha/2), \quad (\text{A.79})$$

where, as before,  $\alpha$  is the significance level needed to be achieved by the decision maker.

If there is no evidence to reject the null hypothesis  $H_0$ , a  $100(1-\alpha)\%$  CI for  $\theta$  is thus given by

$$\theta_1 < \theta < \theta_2, \quad (\text{A.80})$$

where  $\theta_1$  and  $\theta_2$  respectively satisfy

$$\frac{\theta_1^{3/2}}{2(\pi)^{1/2}} - \frac{\theta_1}{4\pi} = \frac{(2K-1)s_2^2}{\chi_{2K-1}^2(1-\alpha/2)} \quad (\text{A.81})$$

and

$$\frac{\theta_2^{3/2}}{2(\pi)^{1/2}} - \frac{\theta_2}{4\pi} = \frac{(2K-1)s_2^2}{\chi_{2K-1}^2(\alpha/2)}. \quad (\text{A.82})$$

## References

- [1] M. Abramowitz, I.A. Stegun, Handbook of mathematical functions with formulas, graphs and mathematical tables, Nat. Bur. Stands, 1964.
- [2] Andersen, B., Hansen, N.H., Iversen, V.B., Use of minicomputer for telephone traffic measurements, Teletenik (Engl. Ed.), 15, 33-46, 1971.
- [3] D. Bear, Principles of telecommunication traffic engineering, 3rd Edition, Peter Peregrinus Ltd, 1988.
- [4] G. Casella, R.L. Berger, Statistical inference, 2nd Edition, Duxbury, 2001.
- [5] D. Doane, L. Seward, Applied statistics in business and economics, 3rd Edition, Macgraw-Hill, 2010.
- [6] R.A. Donnelly Jr., Business statistics, Pearson Education inc., Upper Saddle River New Jersey, 2012.
- [7] Griffiths, M., Remodified Bessel functions via coincidences and near coincidences, J. Integer Seq., 14, 1-10, 2011.
- [8] Griffiths, T.L., Tenenbaum, J.B., From mere coincidences to meaningful discoveries, Cognition, 103 (2), 180-226. May 2007.
- [9] G. Grimmett, D. Stirzaker, Probability and random processes, 3rd Edition, Oxford University Press, 2001.
- [10] Nemenman, E., Coincidences and estimation of entropies of random variables with large cardinalities, Entropy, 13(12), 2013-2023. Dec. 2011.
- [11] NIST Digital Library of Mathematical Functions. [Online]. Available: <http://dlmf.nist.gov/16.11>. [July15, 2016].
- [12] Sullivan, S.P., Probative inference from phenomenal coincidence demystifying the doctrine of chances, Law, Probability and Risk, 14(1), 27-50, 2015.
- [13] R.M. Weiers, Introduction to business statistics, Cengage Learning, Mason Ohio, 1988.
- [14] Skellam distribution, Wikipedia, The free encyclopedia, 2016. [Online] Available: [http://en.wikipedia.org/wiki/Skellam\\_distribution](http://en.wikipedia.org/wiki/Skellam_distribution). [June 15, 2016]