CHAPTER 5

STATISTICAL INFERENCES ON THE CONWAY-MAXWELL-POISSON DISTRIBUTION

5.0 Introduction

The COM-Poisson distribution was first introduced by Conway and Maxwell (1962) for modelling queue systems and service rates. Recently, this distribution has been revived by Shmueli *et al.* (2005) to model count data which are characterized by over- or under-dispersion. The COM-Poisson distribution which received much attention is a member of the exponential family of distributions. Thus, sufficient statistics and other properties can be derived. Unfortunately, its pmf involves an infinite sum and approximation is required for applications.

The COM-Poisson distribution contains several common distributions as special cases, such as the geometric, Bernoulli and Poisson distributions. This flexibility attracts great interest and it has been widely used in many fields; for example, in modelling word lengths (Wimmer *et al.*, 1994), in the prediction of purchase timing and quantity decisions (Boatwright *et al.*, 2003), internet search electric power system reliability (Guikema and Coffelt, 2008) and motor crashes (Lord *et al.*, 2008).

The conjugate distributions for the COM-Poisson distribution were studied by Kadane *et al.* (2006) and the maximum likelihood estimation of the parameters as a cure rate survival model was discussed. Guikema and Coffelt (2008), Lord *et al.* (2008) and Sellers and Shmueli (2008) examined the COM-Poisson distribution in the context of a generalized linear model. Rodrigues *et al.* (2009) have developed a flexible cure rate survival model by assuming the number of competing causes of the event of interest to follow the COM- Poisson distribution. It is shown that modelling survival data with a cure rate is useful. Cordeiro *et al.* (2011) have compounded an exponential distribution with the COM-Poisson distribution to define a new three parameter distribution referred to as the exponential-Conway-Maxwell Poisson (ECOMP) distribution, and the characterization of the ECOMP distribution was given.

In this thesis, the COM-Poisson distribution is compared with the $GIT_{3,1}$ in the modelling of dispersion as both distributions can handle under-, equi- and over-dispersion. Due to the simple and flexibility form of COM-Poisson, we found that it is worthwhile to explore further properties of the COM-Poisson distribution. The chapter is organized as follows. Section 1 provides some properties and a brief background of the COM-Poisson distribution. The limitation of computer accuracy on the infinite sum is investigated in section 2. Meanwhile in section 3, we examine the accuracy of applying the asymptotic approximation on the infinite sum. Section 4 describes the maximum likelihood parameter estimation and section 5 provides the test of equi-dispersion by using Rao's score test and likelihood ratio test. The last section presents the statistical power analysis with discussion.

5.1 Probability Functions of COM-Poisson Distribution

The pmf of the COM-Poisson distribution is given by

$$f(X=x)=\frac{\lambda^x}{(x!)^v}\frac{1}{Z(\lambda,v)} \tag{5.1}$$

where $Z(\lambda,v)=\sum_{j=0}^{\infty}\frac{\lambda^j}{(j!)^v}$ for $\lambda>0$ and $v\geq 0$.

The COM-Poisson distribution is a generalization of some well-known discrete distributions. For example:

1) When $v=1$, $Z(\lambda,v)=e^{\lambda}$ the distribution is simply the Poisson distribution.

2) When $v\to\infty$, $Z(\lambda,v)\to 1+\lambda$ the distribution approaches a Bernoulli distribution

with $f(X=1)=\dfrac{\lambda}{1+\lambda}$.

3) For $v=0$ and $\lambda<1$, $Z(\lambda,v)=\sum_{j=0}^{\infty}\lambda^{j}=\dfrac{1}{1-\lambda}$ is a geometric sum and the distribution

itself is $f(X=x|\lambda,v)=\lambda^{x}(1-\lambda)$ for $x=0,1,2,...$

Note that when $v=0$ and $\lambda\geq 1$, $Z(\lambda,v)$ does not converge and the distribution is not defined.

The pmf of the COM-Poisson distribution can be easily computed by using the recurrence relation of the pmf given by

$$f(X=x)=\frac{\lambda}{x^{v}}f(x-1)$$

but the initial probability, $f(X=0)=Z(\lambda,v)^{-1}$ is not available in closed form. Shmueli *et al.* (2005) have derived the following asymptotic approximation of $Z(\lambda,v)$:

$$Z(\lambda,v)=\frac{\exp(v\lambda^{1/v})}{\lambda^{\frac{v-1}{2v}}(2\pi)^{\frac{v-1}{2}}\sqrt{v}}(1+O(\lambda^{-\frac{1}{v}})) \tag{5.2}$$

However, in this formula, $Z(\lambda,v)$ grows rapidly as λ increases or v decreases. To overcome the difficulty of computing the infinite sum $Z(\lambda,v)$, Shmueli *et al.* (2005) suggested to truncate the sum and bound the error.

5.2 Limitation of Computer Accuracy

With the present computational power available, it is possible to compute the infinite sum, $Z(\lambda, v)$ by using most of the software available. Shmueli *et al.* (2005) mentioned that $Z(\lambda, v)$ may be computed by truncating the numerical series. Suppose that the ratio of the term *j* and *j*-1 in the series is $\frac{\lambda}{j^{v}}$, $j \geq 1$. Choose a small ε (say, 0.01) and take the smallest k such that $k > \left(\frac{\lambda}{\varepsilon}\right)^{\frac{1}{v}}$. The numerical approximation is then presented as $Z(\lambda, v) \approx \sum_{j=0}^{k} \frac{\lambda^{j}}{(j!)^{v}}$.

In this section, $Z(\lambda, v)$ is calculated by recursion with double-precision accuracy and without truncation of the series. By varying the parameter v under λ=5 and λ=30, the performance of the recursion is examined in the following tables and

$$a(j+1) = \lambda\left(\frac{1}{j+1}\right)^{v} a(j)$$

where $\frac{\lambda}{(j+1)^{v}}$ is the ratio of the subsequent terms of $Z(\lambda, v)$.

Tables 5.1 to 5.4 show that when $v < 1$ and λ is large, convergence is extremely slow and this causes overflow. When λ is small (=5), the bound has to be set for v ($\geq$ 0.3) in order to avoid overflow. However, for larger value of λ (=30), the bound of the parameter v has to be increased to 0.5 due to the limitation of computer's accuracy.

Table 5.1: Calculate $Z(\lambda,\nu)$ with λ=5 (ν=0.1, 0.3 and 0.5)

ν	0.1			0.3			0.5	
j	ratio	$a(j+1)$	j	ratio	$a(j+1)$	j	ratio	$a(j+1)$
1	5.000	5.00E+00	1	5.000	5.00E+00	1	5.000	5.00E+00
2	4.6652	2.33E+01	2	4.0613	2.03E+01	2	3.5355	1.77E+01
3	4.4798	1.04E+02	3	3.5961	7.30E+01	3	2.8868	5.10E+01
4	4.3528	4.55E+02	4	3.2988	2.41E+02	4	2.5000	1.28E+02
5	4.2567	1.94E+03	5	3.0852	7.43E+02	5	2.2361	2.85E+02
6	4.1798	8.09E+03	6	2.921	2.17E+03	6	2.0412	5.82E+02
7	4.1159	3.33E+04	7	2.7889	6.05E+03	7	1.8898	1.10E+03
8	4.0613	1.35E+05	8	2.6794	1.62E+04	8	1.7678	1.95E+03
9	4.0137	5.43E+05	9	2.5864	4.20E+04	9	1.6667	3.24E+03
10	3.9716	2.16E+06	10	2.5059	1.05E+05	10	1.5811	5.13E+03
⋮			⋮			⋮		
300	2.8266	1.75E+148	300	0.9033	2.21E+25	50	0.7071	5.09E+02
301	2.8256	4.94E+148	301	0.9024	2.00E+25	51	0.7001	3.57E+02
302	2.8247	1.39E+149	302	0.9015	1.80E+25	52	0.6934	2.47E+02
⋮			⋮			⋮		
669	2.6087	2.90E+307	642	0.719	1.29E-09	99	0.5025	1.63E-09
670	2.6083	7.58E+307	643	0.7186	9.29E-10	100	0.5000	8.17E-10
671		Overflow						
$Z(\lambda,\nu)$		Error			1.60E+29			1.34E+06

Table 5.2: Calculate $Z(\lambda,\nu)$ with λ=5 (ν=1, 2 and 3)

ν	1			2			3	
j	ratio	$a(j+1)$	j	ratio	$a(j+1)$	j	ratio	$a(j+1)$
1	5.000	5.00E+00	1	5.000	5.00E+00	1	5.0000	5.00E+00
2	2.500	1.25E+01	2	1.250	6.25E+00	2	0.6250	3.13E+00
3	1.6667	2.08E+01	3	0.5556	3.47E+00	3	0.1852	5.79E-01
4	1.2500	2.60E+01	4	0.3125	1.09E+00	4	0.0781	4.52E-02
5	1.000	2.60E+01	5	0.200	2.17E-01	5	0.0400	1.81E-03
6	0.8333	2.17E+01	6	0.1389	3.01E-02	6	0.0231	4.19E-05
7	0.7143	1.55E+01	7	0.102	3.08E-03	7	0.0146	6.10E-07
8	0.625	9.69E+00	8	0.0781	2.40E-04	8	0.0098	5.96E-09
9	0.5556	5.38E+00	9	0.0617	1.48E-05	9	0.0069	4.09E-11
10	0.5000	2.69E+00	10	0.0500	7.42E-07			
⋮			11	0.0413	3.06E-08			
14	0.3571	7.00E-02	12	0.0347	1.06E-09			
15	0.3333	2.33E-02	13	0.0296	3.15E-11			
16	0.3125	7.29E-03						
⋮								
26	0.1923	3.69E-09						
27	0.1852	6.84E-10						
$Z(\lambda,\nu)$		1.48E+02			1.71E+01			9.75E+00

Table 5.3: Calculate $Z(\lambda, v)$ with λ=30 (v=0.1, 0.3 and 0.5)

v	0.1			0.3			0.5	
j	ratio	$a(j+1)$	j	ratio	$a(j+1)$	j	ratio	$a(j+1)$
1	30.0000	3.00E+01	1	30.0000	3.00E+01	1	30.0000	3.00E+01
2	27.9910	8.40E+02	2	24.3676	7.31E+02	2	21.2132	6.36E+02
3	26.8788	2.26E+04	3	21.5767	1.58E+04	3	17.3205	1.10E+04
4	26.1165	5.89E+05	4	19.7926	3.12E+05	4	15.0000	1.65E+05
5	25.5402	1.51E+07	5	18.5110	5.78E+06	5	13.4164	2.22E+06
6	25.0788	3.78E+08	6	17.5257	1.01E+08	6	12.2474	2.72E+07
7	24.6951	9.32E+09	7	16.7337	1.69E+09	7	11.3389	3.08E+08
8	24.3676	2.27E+11	8	16.0766	2.72E+10	8	10.6066	3.27E+09
9	24.0822	5.47E+12	9	15.5185	4.23E+11	9	10.0000	3.27E+10
10	23.8298	1.30E+14	10	15.0356	6.36E+12	10	9.4868	3.10E+11
⋮			⋮			⋮		
171	17.9400	4.77E+221	200	6.1209	9.02E+182	963	0.9667	1.01E+194
172	17.9295	8.55E+222	201	6.1117	5.51E+183	964	0.9662	9.75E+193
173	17.9191	1.53E+224	202	6.1026	3.36E+184	965	0.9657	9.41E+193
⋮			⋮			⋮		
239	17.3493	2.56E+306	367	5.1018	1.11E+307	2482	0.6022	1.50E-09
240	17.3420	4.45E+307	368	5.0976	5.66E+307	2483	0.6021	9.05E-10
241		Overflow	369		Overflow			
$Z(\lambda,v)$		Error			Error			3.32E+196

Table 5.4: Calculate $Z(\lambda, v)$ with λ=30 (v=1, 2 and 3)

v	1			2			3	
j	ratio	$a(j+1)$	j	ratio	$a(j+1)$	j	ratio	$a(j+1)$
1	30.0000	3.00E+01	1	30.0000	3.00E+01	1	30.0000	3.00E+01
2	15.0000	4.50E+02	2	7.5000	2.25E+02	2	3.7500	1.13E+02
3	10.0000	4.50E+03	3	3.3333	7.50E+02	3	1.1111	1.25E+02
4	7.5000	3.38E+04	4	1.8750	1.41E+03	4	0.4688	5.86E+01
5	6.0000	2.03E+05	5	1.2000	1.69E+03	5	0.2400	1.41E+01
6	5.0000	1.01E+06	6	0.8333	1.41E+03	6	0.1389	1.95E+00
7	4.2857	4.34E+06	7	0.6122	8.61E+02	7	0.0875	1.71E-01
8	3.7500	1.63E+07	8	0.4688	4.04E+02	8	0.0586	1.00E-02
9	3.3333	5.42E+07	9	0.3704	1.49E+02	9	0.0412	4.12E-04
10	3.0000	1.63E+08	10	0.3000	4.48E+01	10	0.0300	1.24E-05
⋮			⋮			11	0.0225	2.79E-07
48	0.6250	6.43E+09	15	0.1333	8.39E-03	12	0.0174	4.84E-09
49	0.6122	3.93E+09	16	0.1172	9.83E-04	13	0.0137	6.60E-11
50	0.6000	2.36E+09	17	0.1038	1.02E-04			
⋮			⋮					
97	0.3093	1.98E-09	21	0.0680	4.01E-09			
98	0.3061	6.07E-10	22	0.0620	2.48E-10			
$Z(\lambda,v)$		1.07E+13			6.98E+03			343.291

5.3 Partial Derivatives of the Constant Z(λ, v)

Minka *et al.* (2003) suggested that by truncating the numerical series, the difficulty of computing $Z(\lambda, v)$ can be overcome and numerical approximation is presented as $Z(\lambda, v) \approx \sum_{j=0}^{k} \frac{\lambda^j}{(j!)^v}$. The approximations for the moments are obtained from the asymptotic approximation of (5.2):

$$E[X] = \lambda \frac{\partial \log Z(\lambda, v)}{\partial \lambda} \approx \lambda^{\frac{1}{v}} - \frac{v-1}{2v} \tag{5.3}$$

$$E[\log(X!)] = -\frac{\partial \log Z(\lambda, v)}{\partial v} \approx \frac{1}{2v^2} \log \lambda + \lambda^{\frac{1}{v}} \left(\frac{\log \lambda}{v} - 1 \right) \tag{5.4}$$

These approximations are good for $v \le 1$ or $\lambda > 10^v$.

Here, we present computation of (5.3) and (5.4) based on a direct differentiation of $Z(\lambda, v)$. The accuracy of using the asymptotic approximation and the direct differentiation of $Z(\lambda, v)$) are displayed in Table 5.5. The partial derivatives of $Z(\lambda, v)$ are as follows.

$$\lambda \frac{\partial \log Z(\lambda, v)}{\partial \lambda} = \lambda \frac{\sum_{j=1}^{\infty} \frac{j \lambda^{j-1}}{(j!)^v}}{\sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^v}} \tag{5.5}$$

$$-\frac{\partial \log Z(\lambda, v)}{\partial v} = -\frac{\sum_{j=2}^{\infty} \frac{-\lambda^{j-1}}{(j!)^v} \ln(j!)}{\sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^v}} \tag{5.6}$$

Table 5.5 clearly illustrates the inaccuracy of the asymptotic approximation for $Z(\lambda, v)$. Note that for $E[X]$, the computed partial derivative of $Z(\lambda, v)$ in equation (5.5) is slightly higher than equation (5.3). The difference between equation (5.3) and (5.5) decreases when the value of the v decreases for any λ. However, when $v = 1$, equations (5.3) and (5.5) achieve the same value. Cleary, the accuracy of the asymptotic

Table 5.5: Comparison between the moments obtained from the asymptotic approximation $Z(\lambda,\nu)$ and the partial derivatives of $Z(\lambda,\nu)$

ν	0.7	0.8	0.9	1.0
10^{ν}	5.0119	6.3096	7.9433	10.0000
Equation	$\lambda = 5$			
(5.3)	10.180462	7.601744	6.034621	5.000000
(5.5)	10.185840	7.605837	6.036846	5.000000
(5.3)-(5.5)	*-0.005378*	*-0.004093*	*-0.002225*	*0.000000*
(5.4)	14.590310	8.822323	5.706565	3.851909
(5.6)	16.221346	10.357898	7.166900	5.251585
(5.4)-(5.6)	*-1.631036*	*-1.535575*	*-1.460335*	*-1.399676*
Equation	$\lambda = 10$			
(5.3)	27.041244	17.907794	12.971052	10.000000
(5.5)	27.042957	17.909221	12.971889	10.000000
(5.3)-(5.5)	*-0.001713*	*-0.001427*	*-0.000837*	*0.000000*
(5.4)	63.767409	35.199096	21.549219	14.177143
(5.6)	65.401361	36.740846	23.018248	15.587261
(5.4)-(5.6)	*-1.633952*	*-1.541750*	*-1.469029*	*-1.410118*
Equation	$\lambda = 20$			
(5.3)	72.427101	42.419851	27.954571	20.000000
(5.5)	72.427714	42.420422	27.954937	20.000000
(5.3)-(5.5)	*-0.000613*	*-0.000571*	*-0.000366*	*0.000000*
(5.4)	239.887286	118.425626	66.814626	41.412512
(5.6)	241.521355	119.969133	68.286922	42.827172
(5.4)-(5.6)	*-1.634069*	*-1.543507*	*-1.472296*	*-1.414660*

approximation for $Z(\lambda,\nu)$ depends on the value of ν when ν is smaller than one. On the other hand, the displayed values of $E[\log(X!)]$ appear to be much higher under equation (5.6) than equation (5.4). The differences are high when ν is small, regardless of λ.

Obviously, the use of partial derivatives $Z(\lambda,\nu)$ given by equations (5.4) and (5.6) are favoured if higher accuracy is required.

5.4 Maximum Likelihood Estimation

The application of the commonly used maximum likelihood estimation where the log-likelihood function is maximized is proposed. The log-likelihood function for count frequency data is

$$\ln L = \sum_{x=0}^{\infty} \pi_x \ln f(x)$$

where $\pi_x =$ observed frequency and $\ln f(x) = x \ln \lambda - \nu \ln(x!) - \ln \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^\nu}$

The likelihood score equations of λ and ν are

$$\frac{\partial \ln L}{\partial \lambda} = \sum_{x=0}^{\infty} \pi_x \frac{\partial \ln f(x)}{\partial \lambda} \quad \text{and} \quad \frac{\partial \ln L}{\partial \nu} = \sum_{x=0}^{\infty} \pi_x \frac{\partial \ln f(x)}{\partial \nu}$$

where

$$\frac{\partial \ln f(x)}{\partial \lambda} = \frac{x}{\lambda} - \frac{\sum_{j=1}^{\infty} \frac{j\lambda^{j-1}}{(j!)^\nu}}{\sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^\nu}} = \frac{1}{\lambda}\left(x - E(x)\right) \tag{5.7}$$

$$\frac{\partial \ln f(x)}{\partial \nu} = -\ln(x!) - \frac{\sum_{j=2}^{\infty} \frac{-j\lambda^j}{(j!)^\nu} \ln(j!)}{\sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^\nu}} = -\ln(x!) + E\left(\ln(x!)\right) \tag{5.8}$$

The infinite sum in the COM-Poisson distribution, $Z(\lambda, \nu)$ is calculated recursively with double-precision accuracy and truncation, $Z(\lambda, \nu) \le 1 \times 10^{50}$, by a Fortran program. The simulated annealing (SA) algorithm (Metropolis *et al.*, 1953) is used in the numerical optimization to solve for the parameters which maximize the log-likelihood function.

5.5 Test for Equi-Dispersion

The COM-Poisson distribution reduces to the ordinary Poisson distribution with parameter λ when $\nu = 1$. In order to test for the deviation from the Poisson distribution, we derive the Rao's score test and the likelihood ratio test for testing the null hypothesis $H_0 : \nu = 1$ against the alternative $H_1 : \nu \ne 1$. The study of the power of this statistical hypothesis test is developed and will be provided in the next subsection.

5.5.1 Rao's Score Test

The Rao's score test statistic is given by

$$T = VI^{-1}V^{T}$$

where V is the score vector and I is the information matrix. The score vector and the information matrix are obtained by evaluating the derivative of the log-likelihood function, ln L under the null hypothesis, given as

$$V = \left(\frac{\partial \ln L}{\partial \nu}, \frac{\partial \ln L}{\partial \lambda} \right)$$

$$I = -\begin{bmatrix} E\left[\frac{\partial^2 \ln L}{\partial \nu^2}\right] & E\left[\frac{\partial^2 \ln L}{\partial \lambda \partial \nu}\right] \\ E\left[\frac{\partial^2 \ln L}{\partial \lambda \partial \nu}\right] & E\left[\frac{\partial^2 \ln L}{\partial \lambda^2}\right] \end{bmatrix}$$

The partial derivatives and elements of information matrix for COM-Poisson distribution are given in Appendix B where the second order partial derivatives of the log-probability functions are

$$\frac{\partial^2 \ln f(x)}{\partial \lambda^2} = -\frac{x}{\lambda^2} + \frac{\left(\sum_{j=1}^{\infty} \frac{j\lambda^{j-1}}{(j!)^{\nu}} \right)^2}{\left(\sum_{j=0}^{\infty} \frac{\lambda^{j}}{(j!)^{\nu}} \right)^2} - \frac{\left(\sum_{j=2}^{\infty} \frac{(j-1)j\lambda^{j-2}}{(j!)^{\nu}} \right)}{\sum_{j=0}^{\infty} \frac{\lambda^{j}}{(j!)^{\nu}}}$$

$$\frac{\partial^2 \ln f(x)}{\partial \nu^2} = \frac{\left(\sum_{j=2}^{\infty} \frac{-\lambda^{j}}{(j!)^{\nu}} \ln(j!) \right)^2 - \left(\sum_{j=0}^{\infty} \frac{\lambda^{j}}{(j!)^{\nu}} \right) \sum_{j=2}^{\infty} \frac{\lambda^{j}}{(j!)^{\nu}} \ln(j!)^2}{\left(\sum_{j=0}^{\infty} \frac{\lambda^{j}}{(j!)^{\nu}} \right)^2}$$

$$\frac{\partial^2 \ln f(x)}{\partial \nu \partial \lambda} = \frac{1}{\left(\sum_{j=0}^{\infty} \frac{\lambda^{j}}{(j!)^{\nu}} \right)^2} \left\{ \sum_{j=0}^{\infty} \frac{\lambda^{j}}{(j!)^{\nu}} \left(\sum_{j=2}^{\infty} \frac{j\lambda^{j-1}}{(j!)^{\nu}} \ln(j!) \right) - \sum_{j=1}^{\infty} \frac{j\lambda^{j-1}}{(j!)^{\nu}} \left(\sum_{j=2}^{\infty} \frac{\lambda^{j}}{(j!)^{\nu}} \ln(j!) \right) \right\}$$

for $x = 0,\ 1,\ 2,\ \ldots$

5.5.2 Likelihood Ratio Test

The LRT requires estimation under the null and alternative models. Based on the comparison of the log likelihood scores of the two models, the LRT gives evidence whether the deviation under one model is statistically significant. The LR test statistic is

$$\text{LR} = -2\ln\left(\frac{L(\hat{\boldsymbol{\theta}}^{*};x)}{L(\hat{\boldsymbol{\theta}};x)}\right)$$

where $\hat{\theta}^{*}$ is the restricted ML estimator and $\hat{\theta}$ is the unrestricted ML estimator.

5.6 Statistical Power Analysis of the Rao's Score and Likelihood Ratio Tests

In this section, the tables of the power through a simulation study are displayed. One thousand simulation runs are found to give results of sufficient accuracy. In the simulation study we let the significance level, α to be 5% and 10%, and the sample sizes used are N =100 (small), 500 (moderate) and 1000 (large). On the other hand, the effect size which serves as the index of departure from the null hypothesis, ($|\nu - 1|$) is designed by considering the case of over and under-dispersion. The effect size is set as 0.2, 0.5, 1.0, 3.0 and 4.0.

The results of the power study are presented in Tables 5.6, 5.7, 5.8 for λ = 5 (short tailed data), 10 (moderate tailed data) and 20 (long tailed data). The power is given by the number of rejections divided by number of repetitions. Furthermore, the estimated empirical level for λ =1, 5, 7, 10 and 20 is studied and the result is presented in Table 5.9.

Table 5.6: Simulated power of score test and LRT: COM-Poisson distribution (λ =5)

λ			5	5	5	5	5	5	5
ν			0.5	0.8	1.2	1.5	2	4	5
Effect size			0.5	0.2	0.2	0.5	1.0	3.0	4.0
N	α	Method	Power						
100	0.05	score	0.996	0.351	0.169	0.644	0.976	1.000	1.000
		LR	0.995	0.296	0.211	0.700	0.986	1.000	1.000
	0.10	score	0.998	0.451	0.282	0.762	0.997	1.000	1.000
		LR	0.997	0.417	0.330	0.790	0.997	1.000	1.000
500	0.05	score	1.000	0.917	0.689	1.000	1.000	1.000	1.000
		LR	1.000	0.906	0.713	1.000	1.000	1.000	1.000
	0.10	score	1.000	0.944	0.814	1.000	1.000	1.000	1.000
		LR	1.000	0.938	0.827	1.000	1.000	1.000	1.000
1000	0.05	score	1.000	0.997	0.953	1.000	1.000	1.000	1.000
		LR	1.000	0.997	0.954	1.000	1.000	1.000	1.000
	0.10	score	1.000	0.997	0.967	1.000	1.000	1.000	1.000
		LR	1.000	0.997	0.973	1.000	1.000	1.000	1.000

Table 5.7: Simulated power of score test and LRT: COM-Poisson distribution (λ =10)

λ			10	10	10	10	10	10	10
ν			0.5	0.8	1.2	1.5	2	4	5
Effect size			0.5	0.2	0.2	0.5	1.0	3.0	4.0
N	α	Method	Power						
100	0.05	score	1.000	0.378	0.184	0.719	0.993	1.000	1.000
		LR	1.000	0.330	0.241	0.765	1.000	1.000	1.000
	0.10	score	1.000	0.478	0.307	0.829	0.999	1.000	1.000
		LR	1.000	0.441	0.354	0.858	1.000	1.000	1.000
500	0.05	score	1.000	0.930	0.755	1.000	1.000	1.000	1.000
		LR	1.000	0.921	0.770	1.000	1.000	1.000	1.000
	0.10	score	1.000	0.959	0.857	1.000	1.000	1.000	1.000
		LR	1.000	0.954	0.862	1.000	1.000	1.000	1.000
1000	0.05	score	1.000	0.997	0.967	1.000	1.000	1.000	1.000
		LR	1.000	0.997	0.970	1.000	1.000	1.000	1.000
	0.10	score	1.000	0.999	0.985	1.000	1.000	1.000	1.000
		LR	1.000	0.999	0.987	1.000	1.000	1.000	1.000

Table 5.8: Simulated power of score test and LRT: COM-Poisson distribution (λ =20)

λ			20	20	20	20	20	20	20
ν			0.5	0.8	1.2	1.5	2	4	5
Effect size			0.5	0.2	0.2	0.5	1.0	3.0	4.0
N	α	Method	Power						
100	0.05	score	1.000	0.387	0.194	0.733	0.996	1.000	1.000
		LR	1.000	0.705	0.245	0.787	0.998	1.000	1.000
	0.10	score	1.000	0.479	0.320	0.852	1.000	1.000	1.000
		LR	1.000	0.747	0.361	0.879	1.000	1.000	1.000
500	0.05	score	1.000	0.938	0.769	1.000	1.000	1.000	1.000
		LR	1.000	0.997	0.789	1.000	1.000	1.000	1.000
	0.10	score	1.000	0.964	0.868	1.000	1.000	1.000	1.000
		LR	1.000	0.997	0.875	1.000	1.000	1.000	1.000
1000	0.05	score	1.000	0.997	0.972	1.000	1.000	1.000	1.000
		LR	1.000	0.994	0.974	1.000	1.000	1.000	1.000
	0.10	score	1.000	0.999	0.993	1.000	1.000	1.000	1.000
		LR	1.000	0.994	0.993	1.000	1.000	1.000	1.000

Table 5.9: Estimated empirical level of score test and LRT: COM-Poisson distribution

λ			1	5	7	10	20	30
ν			1	1	1	1	1	1
N	α	Method	Power					
100	0.05	score	0.042	0.060	0.052	0.056	0.059	0.062
		LR	0.052	0.067	0.055	0.060	0.050	0.044
	0.10	score	0.095	0.109	0.105	0.112	0.117	0.121
		LR	0.096	0.114	0.107	0.122	0.104	0.097
500	0.05	score	0.047	0.054	0.051	0.054	0.051	0.059
		LR	0.044	0.057	0.053	0.057	0.050	0.037
	0.10	score	0.095	0.111	0.104	0.108	0.116	0.126
		LR	0.094	0.110	0.101	0.105	0.115	0.111
1000	0.05	score	0.041	0.053	0.055	0.055	0.055	0.059
		LR	0.042	0.049	0.051	0.052	0.056	0.055
	0.10	score	0.087	0.090	0.100	0.104	0.103	0.103
		LR	0.088	0.093	0.103	0.104	0.100	0.970

Overall, Tables 5.6, 5.7 and 5.8 yield similar results. The power of the score test and LRT are very close to each other when the sample size N is large enough ($N \geq 500$), regardless of over-dispersion ($\nu < 1$) and under-dispersion ($\nu > 1$). For the case of equi-dispersion (Table5, $\nu = 1$), both tests have estimated empirical levels close to the specified significance levels of 5% and 10%.

The statistical power as shown in Tables 5.6 to 5.8 greatly depends upon the sample size and the effect size. Note that the larger the sample size, the higher is the statistical power. Likewise, the power increases with increasing effect size. For over-dispersion data, when the effect size is 0.5, an almost 100% detection is achieved even for a small sample size of 100. When the sample size increases ($N \geq 500$), an effect size of 0.2 can be detected with the power close to 1.0. For under-dispersion, the detection is slightly weaker where larger effect size and sample sizes are needed to achieved 100% detection. When N=100, an effect size larger than 1.0 are required to detect such deviation. The model will be strongly rejected if $N \geq 500$ and the effect size equals or is larger than 0.5. For a large sample size of 1000, we have almost 100% detection even for a small effect size of 0.2.