MASTER

Likelihood Ratio Testing of Proportionality for Rapid Micro-Biological Methods (RMM) with Poisson Distributed Data

Keizer, Sandra P.

Award date:
2020

Likelihood Ratio Testing of Proportionality for Rapid Micro-Biological Methods (RMM) with Poisson Distributed Data

Master Thesis

Sandra Keizer
0849077
June 15, 2020

Supervisor 1: E. van den Heuvel
Supervisor 2: M.A. Manju
Supervisor 3: P. IJzerman-Boon

Abstract

Quantitative Rapid Microbiological Methods (RMM's) are methods to determine the number of micro organisms. RMM are faster than the classical methods, but to be allowed to use them, they need to be validated. Of the several validation parameters for quantitative RMM's, one such parameter is linearity which means that the observed values are counted in a linear manner compared to the average number of CFU's (Colony Forming Units) in the measured solution. There are two formulations of linearity that can be tested in our specific model: the linearity and log-linearity with respect to the concentration of microbes, together known as proportionality. The model uses the Mitscherlich equation as the parameter for the Poisson distribution of the observed values. Here, the parameters of the null-hypothesis representing the number of false positives, lies on the boundary of the parameter space. This implies that the distribution of the likelihood ratio test is unknown. In this thesis the distribution of the likelihood ratio test (LRT) for one specific case with one parameter on the boundary is determined. The model will also be simulated to see if there is a difference between joint or sequential hypothesis testing. The simulation will also be used to see the difference between differently chosen spike concentrations. We found that, depending on the inclusion or exclusion of a blank spike concentration, the distribution the LRT differs. If a blank spike is included, the LRT of the joint hypothesis case is chi-squared distributed with one degree of freedom. Here the LRT of the sequential hypothesis cases is deterministic. The LRT of the joint hypothesis, excluding the blank spike, is a mixture of two chi-sqaured random variables, with one and two degrees of freedom, each with ratio 1/2. For the sequential hypothesis, excluding the blank spike, the distribution is a mixture of chi-squared distribution with one degree of freedom and zero, again each with ratio 1/2. We found that there was no clear difference in using either joint or sequential hypothesis cases.

Contents

Chapter 1

Introduction

In the pharmaceutical industry, it is important to determine the presence or number of micro organisms in, for example, the productioin environments and the water systems. For more than a century, the industry relied on classical growth based methods, where colony forming units (CFU's) are being counted or detected. Since the classical methods are slow, new methods have been developed to replace the classical methods, so called Rapid Microbiological Methods (RMM). The European Pharmacopoeia (EP) Chapter 5.1.6 (Council of Europe, 2017) divides the methods in three categories: growth-based methods, direct measurement, and cell-component analysis. The first category, growth-based methods, relies on the growth of the microbes through use of a culture medium. After a period of time, the number of microbes or CFU's will can then be counted. Direct measurement mostly uses fluorescence to directly detect and count microbes. This fluorescence can for example be created by staining the microbes with fluorescent dye or the fluoresence can be induced in the microbes with use of lasers. Lastly, cell-component analysis results in an indirect measure of the presence of microbes. Methods that belong in this category use the expression of specific cell-components to determine the presence of microbes or to identify certain microbes.

Since the RMM's differ from the classical method, the method needs to be tested to see if its performance is adequate compared to the classical method. The EP (Council of Europe, 2017) and USP (USP-NF, 2015) determined several parameters that need to be validated: accuracy, precision, detection limit (or quantitation limit), linearity, specificity, and robustness. The accuracy of a method is the closeness of the method's test results to test results from the pharmacopoeial method. The precision relates to the variability between different test samples of homogeneous suspensions. It looks at the closeness between the different samples. The limit of detection in the USP (or quantitation limit in EP) is the lowest number of CFU's that the system can quantify with high enough precision and accuracy. Both the USP and EP define the robustness of a method as follows "*its capacity to remain unaffected by small but deliberate variations in method parameters*" (Council of Europe, 2017; USP-NF, 2015). The linearity of the method is described as the proportionality between the results produced by the method and the concentration of micro-organisms present in the sample. Here, two quantities, x and y, are proportional if $x/y = k$ for some constant k.

Janssen Pharmaceuticals is planning to test a new system for microbial detection, the IMB-W™, a quantitative RMM application. This method attempts to count the number of microbes in flowing water. A laser is used to stir the naturally occurring fluorescence of the microbes. This is called Light Induced Fluorescence (LIF). As the laser passes through the particles a

particle signal is created. This signal should allow the method to count the number of biological particles passing through the pipe without having to grow colonies. Inert particles would not be counted since they have no fluorescence, and should have a different signal. As this system is a quantitative method, from which the counting unit is different than CFU, one of the validation parameters that is important is the linearity of the system.
As mentioned above, in the USP and EP, linearity is referred to as the proportionality between the test results and the spiked concentration. This means both linearity and log-linearity in the number of organisms in the solution need to be satisfied. If you just want to test linearity, one method is determining the slope between adjacent points and testing if the slope is equal between any set of two adjecent points (Niermann, 2007). This only tests linearity and not log-linearity, meaning that both have to be tested separately. In Xie et al. (2017) data is tested for proportionality, the function they use is the same as the Mitscherlich equation, is what we will use to test for proportionality. The Mitscherlich equation, $f(\xi|\vartheta_1, \vartheta_2, \vartheta_3) = \vartheta_1 + \vartheta_2 \exp(\vartheta_3 \xi)$ (Box & Lucas, 1959), with $\vartheta_i \in \mathbb{R}$ for $i = 1, 2, 3$, is considered proportional in concentration $x = exp(\xi)$ if $\vartheta_1 = 0$ and $\vartheta_3 = 1$. Here x represents the number of organisms present in the sample. Note that here ϑ_1 will represent the false positives, the number of inert particles that the system counted. To test for proportionality using the Mitscherlich equation, you need to test if $\vartheta_1 = 0$ and $\vartheta_3 = 1$.
One of the methods to do this is the Likelihood Ratio Test (LRT). Under certain generality conditions the LRT test statistic follows a chi-squared distribution, where the degrees of freedom is equal to the difference in number of constrained parameters of the null and alternative hypotheses (Wilks, 1938). However, one of these conditions for the LRT to follow this distribution is that the value of the parameter(s) to be tested cannot be on the boundary of the parameter space. This is not the case for the Mitscherlich equation mentioned above, where ϑ_1 represents the non-negative false positive count. In this thesis, we try to determine the distribution of the LRT in our specific case, where one of the parameters is on the boundary. For this, a Theorem by Self & Liang (1987) is used. The regularity conditions that need to hold to be able to use this theorem can still hold when the value of the parameters under null-hypothesis can lie on the boundary of the parameter space.
The LRT distribution is determined for different set-ups of the null-hypothesis: the joint hypothesis, where $\vartheta_1 = 0$ and $\vartheta_3 = 1$ are tested simultaneously, and the two sequential hypotheses, in which $\vartheta_1 = 0$ and $\vartheta_3 = 1$ are tested in two different tests, using two different sequences. We do not know if the different hypotheses set-ups have different consequences for the power of the tests used. That is why we simulate the different test set-ups for our model to compare the different tests. Here we also look at the influence of choosing different spike concentrations for the experiment.
In Chapter 2, the statistical model and the hypotheses of interest will be explained, and the notation and definitions used in the thesis will be given. In Chapter 3, the distribution of the likelihood ratio tests (LRT) for the hypotheses given in Chapter 2 will be determined. This chapter is divided in three parts: the theorem of Self & Liang (1987), the distribution of the LRT when all spike solutions are non-zero, and the distribution of the LRT when one of the spike solutions is blank. In Chapter 4, our simulation study will be described. Here we will look at the differences in power and type-I error rate for the different type of hypotheses and different spike concentrations sets. Lastly, Chapter 5 will give the conclusion.

Chapter 2

Model Description

To be able to determine if the counts of the system are linearly correlated to the number of CFU's in the water, data has to be collected. To do this, several spike solutions will be made. These solutions will each have a different concentration of microbes. In-house isolates or BioBalls™Morgan et al. (2005) will be used to create the solutions. BioBalls are a reference material that is regularly used for testing RMM's. The expected number of CFU's in a BioBall as well as an approximation of the standard deviation is known. This means we know the expected number of CFU's in the spike solutions. These solutions will be divided over several samples, each with equal size. Based on these samples for each spike solution, we will determine how much the system counts. The data retrieved from this can be used to determine if the system satisfies the linearity criterion, given by the USP and EP.

2.1 Statistical Model

This experiment is translated in the following model. There are m different the spiked solutions, where λ_i denotes the spike concentration per sample of solution $i = 1, \ldots, m$. From each of these spiked solutions, n different samples are taken. Testing the samples with the system results in the observed values X_{ij} where i denotes the solution and j the sample, where $i \in \{1, \ldots, m\}$ and $j \in \{1, \ldots, n\}$. We assume the observed values to be independent and Poisson distributed with mean $p\lambda_i^\eta + c$. Here p is the expected number of target particles counted if the spike concentration is 1, η is the parameter used to test for linearity of the observed value in λ_i and c is the parameter denoting the number of non-target particles that are counted by the system. The probability mass function of X_{ij} is as follows:

$$\mathbb{P}(X_{ij} = k) = e^{-(p\lambda_i^\eta + c)} \frac{(p\lambda_i^\eta + c)^k}{k!}.$$

The mean of the observed variables $p\lambda_i^\eta + c$ is related to the Mitscherlich function, $f(\vartheta_1, \vartheta_2, \vartheta_3) = \vartheta_1 + \vartheta_2 \exp(\vartheta_3 \xi)$, where $\exp(\xi) = \lambda_i$ and $(\vartheta_1, \vartheta_2, \vartheta_3) = (c, p, \eta)$. Next we define $\theta := (p, \eta, c)^T$. The parameter θ lies in the parameter space Ω. The subspace Ω_0 contains all parameters for which the null-hypothesis is true. This includes the parameter θ_0, the 'true' parameter if the null-hypothesis is true.

2.2 Maximum Likelihood Estimation

After the data is collected, the parameters can be estimated. This is can be done by finding the supremum of the log-likelihood function, i.e. maximum likelihood estimation. The maximum likelihood estimator $\hat{\theta}$ is the parameter such that $l(\hat{\theta}) = \sup_{\theta \in \Omega} l(\theta)$, with $l(\theta)$ the log-likelihood and Ω the parameter space. The general parameter space in this model is as follows:

$$\Omega = \left\{ \theta = \begin{pmatrix} p \\ \eta \\ c \end{pmatrix} : \begin{matrix} 0 \leq p \\ 0 < \eta \\ 0 \leq c \end{matrix} \right\} = [0, \infty) \times (0, \inf] \times [0, \infty),$$

The log-likelihood, $l_{mn}(\theta)$, for our model, with parameter $\theta = (p, \eta, c)^T$, can be seen in the following equations:

$$l_{mn}(\theta) = \sum_{i=1}^{m} \sum_{j=1}^{n} [-p\lambda_i^\eta - c + X_{ij} \cdot \log(p\lambda_i^\eta + c) - \log(X_{ij}!)] \tag{2.1}$$

The first derivatives of the log-likelihood equal zero at the maximum likelihood estimator, $\hat{\theta}$, the MLE. The equations for the first derivatives can be seen below:

$$\frac{\partial}{\partial p} l(p, \eta, c) = \sum_{i=1}^{m} \lambda_i^\eta \left(\frac{X_{i.}}{p\lambda_i^\eta + c} - n \right) \tag{2.2}$$

$$\frac{\partial}{\partial \eta} l(p, \eta, c) = \sum_{i=1}^{m} p \log(\lambda_i) \lambda_i^\eta \left(\frac{X_{i.}}{p\lambda_i^\eta + c} - n \right) \tag{2.3}$$

$$\frac{\partial}{\partial c} l(p, \eta, c) = \sum_{i=1}^{m} \left(\frac{X_{i.}}{p\lambda_i^\eta + c} - n \right) \tag{2.4}$$

where $X_{i.} = \sum_{j=1}^{n} X_{ij}$. To find the MLE $\hat{\theta} = (\hat{p}, \hat{\eta}, \hat{c})^T$, the following three equations have to be solved: $\frac{\partial}{\partial p} l(\hat{p}, \hat{\eta}, \hat{c}) = 0$, $\frac{\partial}{\partial \eta} l(\hat{p}, \hat{\eta}, \hat{c}) = 0$ and $\frac{\partial}{\partial c} l(\hat{p}, \hat{\eta}, \hat{c}) = 0$. Only if the MLE is at the boundary of the parameter space, then the derivatives might not equal zero. At the left boundary the corresponding derivative should be smaller than or equal to zero and at the right boundary the corresponding derivative should be bigger than or equal to zero. This means, for example, if $\hat{c} = 0$, then Equation 2.4 must be smaller than zero, instead of equal to zero. Below, the Fisher information matrix, $I_f(\theta)$, can be found. This matrix gives information on the variability of the gradient of the log-likelihood.

$$I_f(\theta) = \mathbb{E}[\frac{\partial^2}{\partial \theta^2} l_{mn}(\theta)] = \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{1}{p\lambda_i^\eta + c} \begin{bmatrix} \lambda_i^{2\eta} & p \log(\lambda_i) \lambda_i^{2\eta} & \lambda_i^\eta \\ p \log(\lambda_i) \lambda_i^{2\eta} & (p \log(\lambda_i) \lambda_i^\eta)^2 & p \log(\lambda_i) \lambda_i^\eta \\ \lambda_i^\eta & p \log(\lambda_i) \lambda_i^\eta & 1 \end{bmatrix}$$

2.3 Hypothesis Testing

The model discussed in this chapter is linear in λ_i if η equals 1 and it is log-linear in λ_i if c equals 0. When both the linearity and log-linearity in λ_i hold, the model is proportional. This gives several options for the null-hypotheses that can be tested:

Case 1, Joint Hypothesis: Both linearities are tested at the same time, which results in the following joint null-hypothesis: $H_0 : \eta = 1 \wedge c = 0$ and $H_1 : \eta \neq 1 \vee c > 0$. This null-hypothesis can be tested with the following likelihood ratio test statistic (LRT):

$$LRT = -2\left(\sup_{p} l_{mn}(p,1,0) - \sup_{p,\eta,c} l_{mn}(p,\eta,c)\right)$$

Once the distribution of the LRT is determined, the p-value can be calculated. The p-value is the probability that the value of the found distribution is greater than the LRT value. This can then be compared to the significance level $\alpha = 0.05$.

Case 2, Sequential Hypotheses(η, c): First test the linearity in λ_i then the log-linearity in λ_i, which results in two sequential tests: $H_{01} : \eta = 1$ versus $H_{11} : \eta \neq 1$ and $H_{02} : c = 0$ versus $H_{12} : c > 0$. The second null-hypothesis, H_{02}, only needs to be tested if the first hypothesis, H_{01}, is not rejected. These hypotheses result in two different LRT test statistics, where the distribution of the likelihood ratio test under null-hypothesis for the first hypothesis H_{01} is known to be chi-squared with 1 degree of freedom (Wilks, 1938). LRT_1 is the likelihood ratio test value for the first hypothesis and LRT_2 is the LRT for the second hypothesis.

$$LRT_1 = -2\left(\sup_{p,c} l_{mn}(p,1,c) - \sup_{p,\eta,c} l_{mn}(p,\eta,c)\right) \stackrel{d}{=} \chi^2_1$$
$$LRT_2 = -2\left(\sup_{p} l_{mn}(p,1,0) - \sup_{p,c} l_{mn}(p,1,c)\right)$$

Note $\stackrel{d}{=}$ means equal in distribution. Here, unlike Case 1, there are two separate tests. Generally, when using multiple test, the total significance level, $\alpha = 0,05$, will be equally divided over all tests. This would mean that each of the two LRT's p-value will be compared to $\alpha/2 = 0.025$. However, this might not be optimal division of the significance. In the simulation, we first used this equal division. In the second simulation, we adjusted this division using the results from the first simulation. It is possible that other methods of dividing the significance level improve the model further, but we did not look into this.
Note that for LRT_1, c is a nuisance parameter that might be located on the boundary of the parameter space, which might create a change in the distribution of LRT_1.

Case 3, Sequential Hypotheses(c, η): First the log-linearity in λ_i is tested then the linearity in λ_i is tested. Which results in the following two sequential null-hypotheses: $H_{01} : c = 0$ versus $H_{11} : c > 0$ and $H_{02} : \eta = 1$ versus $H_{12} : \eta \neq 1$. These hypotheses can be tested with the following LRT's:

$$LRT_1 = -2\left(\sup_{p,\eta} l_{mn}(p,\eta,0) - \sup_{p,\eta,c} l_{mn}(p,\eta,c)\right)$$
$$LRT_2 = -2\left(\sup_{p} l_{mn}(p,1,0) - \sup_{p,\eta} l_{mn}(p,\eta,0)\right) \stackrel{d}{=} \chi^2_1$$

Note that the distribution of the second LRT is under the null-hypothesis.The p-values of the two LRT's will be compared to $\alpha/2$, the same as in Case 2. As mentioned in case 2, there might be other ways to divide the significance level, that could yield better results. Here the second test is executed if the first hypothesis $H_{01} : c = 0$ is not rejected. To be able to know when to reject H_{01}, the distribution of LRT_1 is needed.

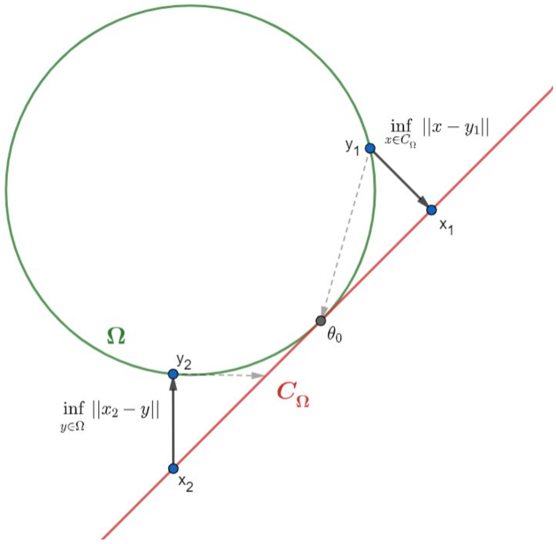


Figure 2.1: Cone Approximation of a Sphere

2.4 Notations and Definitions

Here we will discuss notation and definitions not yet mentioned earlier in the model description. The sum over the samples $X_{i.}$ is defined as follows: $X_{i.} = \sum_{j=1}^{n} X_{ij}$. As mentioned above, the log-likelihood is denoted by $l_{mn}(\theta)$. The first and second derivatives of the log-likelihood are respectively given by $U_{mn}(\theta)$ and $-I_{mn}(\theta)$.
Next we define the approximation of a set Ω by a cone C_Ω, corresponding to Self & Liang (1987):

Definition. The set $\Omega \subset \mathbb{R}^d$ is approximated at $\theta_0 \in \Omega$ by a cone C_Ω with vertex at θ_0 if

$$\inf_{x \in C_\Omega} ||x - y|| = o(||y - \theta_0||) \text{ for all } y \in \Omega \tag{2.5}$$

and

$$\inf_{y \in \Omega} ||x - y|| = o(||x - \theta_0||) \text{ for all } x \in C_\Omega. \tag{2.6}$$

Note that $\inf_{x \in C_\Omega} ||x - y|| = o(||y - \theta_0||)$ if $\frac{\inf_{x \in C_\Omega} ||x-y||}{||y-\theta_0||} \to 0$ for $y \to \theta_0$, i.e. $\inf_{x \in C_\Omega} ||x - y||$ converges to zero faster than $||y - \theta_0||$ for $y \to \theta_0$. Here a cone, C, with vertex at θ_0 is a set of points such that $a(x - \theta_0) + \theta_0 \in C$ for all $a \geq 0$ and $x \in C$. An example of a set that can be approximated by a cone is a sphere, which can be approximated by the tangent line through θ_0. This example can be seen in Figure 2.1. Here Ω is the sphere and C_Ω is the cone used to approximate Ω. As you look at point y_1, we have that $x_1 = \arg\inf_{x \in C_\Omega} ||x - y_1||$, and you can see that $\inf_{x \in C_\Omega} ||x - y_1|| = ||x_1 - y_1|| < ||y_1 - \theta_0||$, since $||y_1 - \theta_0||$ is the long edge of a right triangle with the 90 degree angle at x_1. Similarly, the closest point on Ω to x_2 is y_2. Here $||x_2 - \theta_0||$ is longer than the long edge of the right triangle with the 90 degree angle at y_2. This implies that $\inf_{y \in \Omega} ||x_2 - y|| = ||x_2 - y_2|| < ||x_2 - \theta_0||$.

Chapter 3

Distribution of the Likelihood Ratio Test (LRT)

One of the tests commonly used to compare two models is the likelihood ratio test (LRT). The LRT test statistic follows a chi-square distribution under certain regularity conditions, with the number of degrees of freedom being equal to the difference in the number of model parameters under the null-hypothesis and the alternative hypothesis (Wilks, 1938). One of the conditions is that the parameters of the null-hypothesis do not lie on the boundary of the parameter space. However in our model the value of parameter c under the null-hypothesis lies on the boundary. This means that the distribution of the likelihood ratio test has to be determined for our model.
This chapter is divided in three sections. The first section gives a theorem from Self & Liang (1987) that can be used for determining the distribution of the LRT, when some parameters may be on the boundary of the parameter space. In Section 2, the conditions of this theorem will be verified when the spike concentrations are all non-zero. In the last section, the distribution for the LRT will be determined when one of the spike concentrations is zero will be determined. This has to be done separately since some of the conditions of Self & Liang (1987) do not hold in this case.

3.1 LRT with True Parameters on Boundary

In the model, described in Chapter 2, the value of the parameter c under the null-hypothesis is on the boundary of the parameter space. Self & Liang (1987) looked at the likelihood ratio test where the values of the parameters under the null-hypothesis are allowed to be on the boundary of the parameter space. They found a method to write the LRT so that determining the distribution becomes easier. This is done using the true parameter, which equals θ_0 under the null-hypothesis. When describing the following theorem and conditions we will use θ_0, since we want to determine the distribution of the likelihood ratio test under the null-hypothesis. Self & Liang (1987) assume that the following regularity conditions are satisfied:

1. The almost sure existence of the first three derivatives of the log-likelihood function with respect to θ was assumed on the intersection of neighborhoods of the true parameter θ_0 and Ω. Here Ω is the parameter space and the true parameter θ_0 is the parameter value

under the null-hypothesis. If the true parameter θ_0 is located at the boundary of Ω then the derivatives are taken from the appropriate side.

2. On the same intersection of neighborhoods of θ_0 and Ω, there exists a function $f_{k_1k_2k_3}(X_{11}, \ldots, X_{mn})$, for which $\mathbb{E}[f_{k_1k_2k_3}(X_{11}, \ldots, X_{mn})] < \infty$, such that for all $k_1, k_2, k_3 \in \{1, 2, \ldots, \dim(\theta)\}$ the following holds for the log-likelihood function $l_{mn}(\theta)$:

$$\frac{1}{mn}\left|\frac{\partial^3}{\partial\theta_{k_1}\partial\theta_{k_2}\partial\theta_{k_3}} l_{mn}(\theta)\right| \leq f_{k_1k_2k_3}(X_{11}, \ldots, X_{mn}) \tag{3.1}$$

3. The first and second derivatives of $l_{mn}(\theta)$ with respect to θ are denoted by $U_{mn}(\theta)$ and $-I_{mn}(\theta)$. On neighborhoods of θ_0 the following holds for $I_{mn}(\theta)$:

$$I(\theta) := \mathbb{E}\left[\frac{1}{mn}I_{mn}(\theta)\right] > 0, \tag{3.2}$$

which imlpies the positive-definiteness of $I(\theta)$. Also at θ_0, the following needs to hold:

$$I(\theta_0) = \frac{1}{mn}\operatorname{cov}\left(U_{mn}(\theta_0), U_{mn}(\theta_0)\right), \tag{3.3}$$

where $\operatorname{cov}(V, W)$, for some vectors V and W, is as follows:

$$\operatorname{cov}(V, W) = \mathbb{E}[VW^T] - \mathbb{E}[V]\mathbb{E}[W]^T \tag{3.4}$$

4. Lastly Ω_0 and Ω are assumed to be regular enough to be approximated by cones with vertices at θ_0. Here Ω_0 is the subset of Ω for testing the null-hypothesis that θ_0 lies in the subset Ω_0 of Ω.

To test the null hypothesis $H_0 : \theta \in \Omega_0$ versus the alternative $H_1 : \theta \in \Omega$, the likelihood ratio statistic can be defined as follows:

$$-2\ln\Lambda = -2\left(\sup_{\theta\in\Omega_0} l_{mn}(\theta) - \sup_{\theta\in\Omega} l_{mn}(\theta)\right). \tag{3.5}$$

We now study the asymptotic properties of the likelihood ratio statistic. If these regularity conditions discussed above hold, Theorem 3 from Self & Liang (1987) states the following:
"*Let Z be a random variable with a multivariate Gaussian distribution with mean θ and co-variance matrix $I^{-1}(\theta_0)$, and let C_{Ω_0} and C_{Ω_1} be non-empty cones approximating Ω_0 and Ω_1 at θ_0 respectively, where $\Omega_1 = \Omega\backslash\Omega_0$. Then under the regularity conditions given above, the asymptotic distribution of the likelihood ratio test statistic is the same as the distribution of the likelihood ratio test of $\theta \in C_{\Omega_0}$ versus the alternative $\theta \in C_{\Omega_1}$ based on a single realization of Z when $\theta = \theta_0$.*"

Based on this Theorem, the asymptotic representation of the likelihood ratio statistic $-2\ln\Lambda$ may be written as:

$$\begin{aligned} -2\ln\Lambda &= \sup_{\theta\in C_{\Omega_0}-\theta_0}\{-(\check{Z}-\theta)^T I(\theta_0)(\check{Z}-\theta)\} - \sup_{\theta\in C_{\Omega}-\theta_0}\{-(\check{Z}-\theta)^T I(\theta_0)(\check{Z}-\theta)\} \\ &= \inf_{\theta\in\tilde{C}_0} ||\tilde{Z}-\theta||^2 - \inf_{\theta\in\tilde{C}} ||\tilde{Z}-\theta||^2, \end{aligned} \tag{3.6}$$

where C_{Ω_0} and C_Ω are the cones approximating Ω_0 and Ω at θ_0 respectively. The set $C - \theta_0$ denotes the cone with vertex at the origin that is a result from translating a cone C with θ_0. The random variable $\check{Z} = Z - \theta_0$ has a multivariate Gaussian distribution with mean 0 and covariance matrix $I(\theta_0)^{-1}$. Similarly $\tilde{Z}$ has a multivariate Gaussian distribution with mean 0, but with identity covariance matrix. Let the spectral decomposition of $I(\theta_0)$ be $P\Lambda P^T$, where P is the eigenvector matrix and Λ the diagonal matrix with the eigenvalues of $I(\theta_0)$ as its entries. Lastly $\tilde{C}_0$ and $\tilde{C}$ are defined as follows:

$$\tilde{C}_0 = \{\tilde{\theta} : \tilde{\theta} = \Lambda^{\frac{1}{2}} P^T \theta \text{ for all } \theta \in C_{\Omega_0} - \theta_0\}$$
$$\tilde{C} = \{\tilde{\theta} : \tilde{\theta} = \Lambda^{\frac{1}{2}} P^T \theta \text{ for all } \theta \in C_{\Omega} - \theta_0\}$$

In Appendix C $\tilde{C}$ is calculated for a two-dimensional C, as an example.

3.2 Distribution LRT for non-zero spike concentrations

To be able to use the theorem of Self & Liang (1987), the four conditions, mentioned in Section 3.1 have to be verified. In this section these four regularity conditions will be verified, where the spike concentrations are all non-zero, i.e. $\lambda_i > 0$ for all i.

3.2.1 Verification Regularity Conditions: Case 1, Joint Hypothesis

Next we will show the regularity conditions from Self & Liang (1987) hold for our model in Case 1, the joint hypothesis, if $\lambda_i > 0$ for all i, when the following assumptions on the parameter space hold. First, we assume that p is greater than a value $\varepsilon_p > 0$ and smaller than some value $M_p < \infty$. This means we have $\theta = (\theta_1, \theta_2, \theta_3)^T = (p, \eta, c)^T$ and $\Omega = (\varepsilon_p, M_p) \times (0, \infty) \times [0, \infty)$. The null-hypothesis is $H_0 : \eta = 1 \wedge c = 0$, which gives $\theta_0 = (p, 1, 0)^T$ and $\Omega_0 = (\varepsilon_p, M_p) \times \{1\} \times \{0\}$. Since we can choose ε_p as small as we want and M_p as large as we want, these assumptions are practically not an issue.

Regularity Condition 1

The first condition was the existence of the first three derivatives of $l_{mn}(\theta)$ on the intersection of neighborhoods of θ_0 and Ω, where $l_{mn}(\theta) = \log(L_{mn}(\theta))$. The log-likelihood can be seen in Equation 2.1.
For neighborhoods of θ_0, $N(\theta_0)$, there are $\delta_p > 0$, $\delta_\eta > 0$ and $\delta_c > 0$ such that:

$$N(\theta_0) \subset (\varepsilon_p - \delta_p, M_p + \delta_p) \times (1 - \delta_\eta, 1 + \delta_\eta) \times (-\delta_c, \delta_c)$$

This means that the intersection of $N(\theta_0)$ with Ω is a subset of $(\varepsilon_p, M_p) \times (1 - \delta_\eta, 1 + \delta_\eta) \times [0, \delta_c)$. The derivatives of $l_{mn}(\theta)$ exist if the derivatives of $\log(p\lambda_i^\eta + c)$ exist. The function $\log(x)$ is infinitly differentiable with respect to x, when $x > 0$. This means, as long as $p\lambda_i^\eta + c > 0$, then $\log(p\lambda_i^\eta + c)$ also infinitely differentiable. Since $p > \varepsilon_p > 0$, $\lambda_i > 0$ for all i and $c \geq 0$, $p\lambda_i^\eta + c > 0$. This also holds on the intersection of $N(\theta_0)$ and Ω. Therefore the derivatives of $\log(p\lambda_i^\eta + c)$ exist and thus the first three derivatives of $l_{mn}(\theta)$ exist. The proof of the existence of the derivatives of $\log(p\lambda_i^\eta + c)$ can be seen in Appendix A. As mentioned before, the first two derivatives are denoted by $U_{mn}(\theta)$ and $-I_{mn}(\theta)$ respectively. $U_{mn}(\theta)$ is a vector of order 3 and $I_{mn}(\theta)$ is a 3×3 matrix. If we define $u_i := (\lambda_i^\eta, p\log(\lambda_i)\lambda_i^\eta, 1)^T$, $U_{mn}(\theta)$ and $I_{mn}(\theta)$ can be written as:

$$U_{mn}(\theta) = \begin{bmatrix} \frac{\partial}{\partial p} l_{mn}(\theta) \\ \frac{\partial}{\partial \eta} l_{mn}(\theta) \\ \frac{\partial}{\partial c} l_{mn}(\theta) \end{bmatrix} = \sum_{i=1}^{m}\sum_{j=1}^{n} \left(\frac{X_{ij}}{p\lambda_i^\eta + c} - 1 \right) \begin{bmatrix} \lambda_i^\eta \\ p\log(\lambda_i)\lambda_i^\eta \\ 1 \end{bmatrix} = \sum_{i=1}^{m}\sum_{j=1}^{n} \left(\frac{X_{ij}}{p\lambda_i^\eta + c} - 1 \right) u_i \tag{3.7}$$

$$\begin{aligned} I_{mn}(\theta) &= \sum_{i=1}^{m}\sum_{j=1}^{n} \frac{X_{ij}}{(p\lambda_i^\eta + c)^2} \begin{bmatrix} \lambda_i^{2\eta} & p\log(\lambda_i)\lambda_i^{2\eta} & \lambda_i^\eta \\ p\log(\lambda_i)\lambda_i^{2\eta} & (p\log(\lambda_i)\lambda_i^\eta)^2 & p\log(\lambda_i)\lambda_i^\eta \\ \lambda_i^\eta & p\log(\lambda_i)\lambda_i^\eta & 1 \end{bmatrix} \\ &- \sum_{i=1}^{m}\sum_{j=1}^{n} \log(\lambda_i)\lambda_i^\eta \left(\frac{X_{ij}}{p\lambda_i^\eta + c} - 1 \right) \begin{bmatrix} 0 & 1 & 0 \\ 1 & p\log(\lambda_i) & 0 \\ 0 & 0 & 0 \end{bmatrix} \\ &= \sum_{i=1}^{m}\sum_{j=1}^{n} \frac{X_{ij}}{(p\lambda_i^\eta + c)^2} u_i u_i^T - \log(\lambda_i)\lambda_i^\eta \left(\frac{X_{ij}}{p\lambda_i^\eta + c} - 1 \right) \begin{bmatrix} 0 & 1 & 0 \\ 1 & p\log(\lambda_i) & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{aligned} \tag{3.8}$$

Regularity Condition 2

For the second regularity condition we need to show that all third derivatives are bounded by a function with finite expectation on the intersection of neighborhoods of θ_0 and Ω. This intersection is the same as seen in the verification of the first regularity condition. Without loss of generality take $0 < \lambda_1 < \lambda_2 < \ldots < \lambda_m$, $\lambda_1^{-\eta_s} := \max\{\lambda_1^{-1-\delta_\eta}, \lambda_1^{-1+\delta_\eta}\}$ and $\lambda_m^{\eta_M} := \max\{\lambda_m^{1+\delta_\eta}, \lambda_m^{1-\delta_\eta}\}$, then $\lambda_1^{\eta_s} \le \lambda_1^\eta < \lambda_2^\eta < \lambda_3^\eta < \ldots < \lambda_m^\eta \le \lambda_m^{\eta_M}$. Define $\mathrm{M}_{|\log\lambda|} := \max_{i=1,\ldots,m} |\log(\lambda_i)|$. Next define $\bar{X} := \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} X_{ij}$, for which, on the intersection of $N(\theta_0)$ and Ω, $\mathbb{E}[\bar{X}] \le M_p \max\{\lambda_m^{1-\delta_\eta}, \lambda_m^{1+\delta_\eta}\} + \delta_c = M_p\lambda_m^{\eta_M} + \delta_c$ holds. This means that if the function $f(X_{11}, \ldots, X_{mn}) = a \cdot \bar{X}$, then f has finite expectation. Here $a \in \mathbb{R}$ is a scalar that may be dependent on λ_i, $i = 1, \ldots, m$.
The inequalities $\left|\frac{\lambda_i^\eta}{p\lambda_i^\eta + c}\right| \le \left|\frac{\lambda_i^\eta}{p\lambda_i^\eta}\right| = \frac{1}{p} \le \frac{1}{\varepsilon_p}$ and $|\sum_i a_i| \le \sum_i |a_i|$ (triangle-inequality) will be used in the following to show the third derivatives are bounded. Next we will determine the absolute values of the third derivatives of $l_{mn}(\theta)$ and an upper bound for each.

$$\begin{aligned} \frac{1}{mn}\left|\frac{\partial^3}{\partial p^3} l_{mn}(\theta)\right| &= \frac{1}{mn}\left|\sum_{i=1}^{m}\sum_{j=1}^{n} \frac{2\lambda_i^{3\eta}}{(p\lambda_i^\eta + c)^3} X_{ij}\right| \le \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} \left|\frac{2\lambda_i^{3\eta}}{(p\lambda_i^\eta + c)^3} X_{ij}\right| \\ &\le \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} \left|\frac{\lambda_i^\eta}{p\lambda_i^\eta + c}\right|^3 2X_{ij} \le \frac{2}{\varepsilon_p^3}\bar{X} \end{aligned} \tag{3.9}$$

$$\begin{aligned}\frac{1}{mn}\left|\frac{\partial^3}{\partial p^2\partial\eta}\,l_{mn}(\theta)\right| &= \frac{1}{mn}\left|\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{2\log(\lambda_i)\lambda_i^{2\eta}}{(p\lambda_i^{\eta}+c)^2}\left(\frac{p\lambda_i^{\eta}}{p\lambda_i^{\eta}+c}-1\right)X_{ij}\right| \\ &\le \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\left|\frac{2\log(\lambda_i)\lambda_i^{2\eta}}{(p\lambda_i^{\eta}+c)^2}\right|\left(\left|\frac{p\lambda_i^{\eta}}{p\lambda_i^{\eta}+c}\right|+1\right)X_{ij} \\ &\le \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{4}{\varepsilon_p^2}|\log(\lambda_i)|\,X_{ij}\le\frac{4}{\varepsilon_p^2}\bar{X}\cdot\mathrm{M}_{|\log\lambda|} \qquad (3.10)\end{aligned}$$

$$\begin{aligned}\frac{1}{mn}\left|\frac{\partial^3}{\partial p^2\partial c}\,l_{mn}(\theta)\right| &= \frac{1}{mn}\left|\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{2\lambda_i^{2\eta}X_{ij}}{(p\lambda_i^{\eta}+c)^3}\right|\le\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\left|\frac{2\lambda_i^{2\eta}X_{ij}}{(p\lambda_i^{\eta}+c)^3}\right| \\ &\le \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\left|\frac{2X_{ij}}{p^3\lambda_i^{\eta}}\right|\le\frac{2\bar{X}}{\varepsilon_p^3\,\lambda_1^{\eta_s}} \qquad (3.11)\end{aligned}$$

$$\begin{aligned}\frac{1}{mn}\left|\frac{\partial^3}{\partial\eta^3}\,l_{mn}(\theta)\right| &= \frac{1}{mn}\left|\sum_{i=1}^{m}\sum_{j=1}^{n}(\log(\lambda_i))^3\left(\frac{2(p\lambda_i^{\eta})^3X_{ij}}{(p\lambda_i^{\eta}+c)^3}-\frac{3(p\lambda_i^{\eta})^2X_{ij}}{(p\lambda_i^{\eta}+c)^2}+\frac{p\lambda_i^{\eta}X_{ij}}{p\lambda_i^{\eta}+c}-p\lambda_i^{\eta}\right)\right| \\ &\le \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}|\log(\lambda_i)|^3\left(\left|\frac{2(p\lambda_i^{\eta})^3X_{ij}}{(p\lambda_i^{\eta}+c)^3}\right|+\left|\frac{3(p\lambda_i^{\eta})^2X_{ij}}{(p\lambda_i^{\eta}+c)^2}\right|+\left|\frac{p\lambda_i^{\eta}X_{ij}}{p\lambda_i^{\eta}+c}\right|+|p\lambda_i^{\eta}|\right) \\ &\le \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}|\log(\lambda_i)|^3(6X_{ij}+p\lambda_i^{\eta})\le\mathrm{M}_{|\log\lambda|}^3(6\bar{X}+M_p\lambda_m^{\eta_M}) \qquad (3.12)\end{aligned}$$

$$\begin{aligned}\frac{1}{mn}\left|\frac{\partial^3}{\partial p\partial\eta^2}\,l_{mn}(\theta)\right| &= \frac{1}{mn}\left|\sum_{i=1}^{m}\sum_{j=1}^{n}(\log(\lambda_i))^2\left(\frac{2p^2\lambda_i^{3\eta}X_{ij}}{(p\lambda_i^{\eta}+c)^3}-\frac{3p\lambda_i^{2\eta}X_{ij}}{(p\lambda_i^{\eta}+c)^2}+\frac{\lambda_i^{\eta}X_{ij}}{p\lambda_i^{\eta}+c}-\lambda_i^{\eta}\right)\right| \\ &\le \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}|\log(\lambda_i)|^2\left(\left|\frac{2p^2\lambda_i^{3\eta}X_{ij}}{(p\lambda_i^{\eta}+c)^3}\right|+\left|\frac{3p\lambda_i^{2\eta}X_{ij}}{(p\lambda_i^{\eta}+c)^2}\right|+\left|\frac{\lambda_i^{\eta}X_{ij}}{p\lambda_i^{\eta}+c}\right|+|\lambda_i^{\eta}|\right) \\ &\le \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}|\log(\lambda_i)|^2(\frac{6X_{ij}}{p}+\lambda_i^{\eta})\le\mathrm{M}_{|\log\lambda|}^2\left(\frac{6\bar{X}}{\varepsilon_p}+\lambda_m^{\eta_M}\right) \qquad (3.13)\end{aligned}$$

$$\begin{aligned}\frac{1}{mn}\left|\frac{\partial^3}{\partial\eta^2\partial c}\,l_{mn}(\theta)\right| &= \frac{1}{mn}\left|\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{p(\log(\lambda_i))^2\lambda_i^{\eta}X_{ij}}{(p\lambda_i^{\eta}+c)^2}\left(\frac{2p\lambda_i^{\eta}}{p\lambda_i^{\eta}+c}-1\right)\right| \\ &\le \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\left|\frac{p(\log(\lambda_i))^2\lambda_i^{\eta}X_{ij}}{(p\lambda_i^{\eta}+c)^2}\right|\left(\left|\frac{2p\lambda_i^{\eta}}{p\lambda_i^{\eta}+c}\right|+1\right) \\ &\le \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{3(\log(\lambda_i))^2X_{ij}}{p\lambda_i^{\eta}}\le\frac{3\bar{X}}{\varepsilon_p\,\lambda_1^{\eta_s}}\mathrm{M}_{|\log\lambda|}^2 \qquad (3.14)\end{aligned}$$

$$\frac{1}{mn}\left|\frac{\partial^3}{\partial c^3}\, l_{mn}(\theta)\right| = \frac{1}{mn}\left|\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{2X_{ij}}{(p\lambda_i^{\eta}+c)^3}\right| \leq \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\left|\frac{2X_{ij}}{(p\lambda_i^{\eta}+c)^3}\right|$$

$$\leq \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{2X_{ij}}{p^3\lambda_i^{3\eta}} \leq \frac{2\bar{X}}{\varepsilon_p^3\,\lambda_1^{3\eta_s}} \tag{3.15}$$

$$\frac{1}{mn}\left|\frac{\partial^3}{\partial p\partial c^2}\, l_{mn}(\theta)\right| = \frac{1}{mn}\left|\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{2\lambda_i^{\eta}X_{ij}}{(p\lambda_i^{\eta}+c)^3}\right| \leq \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\left|\frac{2\lambda_i^{\eta}X_{ij}}{(p\lambda_i^{\eta}+c)^3}\right|$$

$$\leq \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{2X_{ij}}{p^3\lambda_i^{2\eta}} \leq \frac{2\bar{X}}{\varepsilon_p^3\,\lambda_1^{2\eta_s}} \tag{3.16}$$

$$\frac{1}{mn}\left|\frac{\partial^3}{\partial \eta\partial c^2}\, l_{mn}(\theta)\right| = \frac{1}{mn}\left|\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{2p\log(\lambda_i)\lambda_i^{\eta}X_{ij}}{(p\lambda_i^{\eta}+c)^3}\right| \leq \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\left|\frac{2p\log(\lambda_i)\lambda_i^{\eta}X_{ij}}{(p\lambda_i^{\eta}+c)^3}\right|$$

$$\leq \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{2\log(\lambda_i)X_{ij}}{p^2\lambda_i^{2\eta}} \leq \frac{2\bar{X}}{\varepsilon_p^2\,\lambda_1^{2\eta_s}}\mathrm{M}_{|\log\lambda|} \tag{3.17}$$

$$\frac{1}{mn}\left|\frac{\partial^3}{\partial p\partial\eta\partial c}\, l_{mn}(\theta)\right| = \frac{1}{mn}\left|\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{\log(\lambda_i)\lambda_i^{\eta}X_{ij}}{(p\lambda_i^{\eta}+c)^2}\left(\frac{2p\lambda_i^{\eta}}{p\lambda_i^{\eta}+c}-1\right)\right|$$

$$\leq \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\left|\frac{\log(\lambda_i)\lambda_i^{\eta}X_{ij}}{(p\lambda_i^{\eta}+c)^2}\right|\left(\left|\frac{2p\lambda_i^{\eta}}{p\lambda_i^{\eta}+c}\right|+1\right)$$

$$\leq \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{|\log(\lambda_i)|X_{ij}}{p^2\lambda_i^{\eta}}(2+1)$$

$$\leq \frac{3\bar{X}}{\varepsilon_p^2\,\lambda_1^{\eta_s}}\mathrm{M}_{|\log\lambda|} \tag{3.18}$$

As mentioned we have $\mathbb{E}[\bar{X}] < \infty$ on the intersection of neighborhoods of θ_0 and Ω, i.e. on $(\varepsilon_p, M_p) \times (1-\delta_\eta, 1+\delta_\eta) \times [0, \delta_c)$. The bounds described in Equations 3.9 to 3.18 define the functions f, such that $\mathbb{E}[f(X_{11}, \dots, X_{mn})] < \infty$, and the third derivatives are therefore bounded, as all these functions are of the form $a \cdot \bar{X}$, $a \in \mathbb{R}$. Note if p is not bounded from above by M_p, then $\mathbb{E}[X]$ not bounded. Also if p is not bounded from below by $\varepsilon_p > 0$, but by 0, then p^{-1} will have no upper bound. This also holds for λ_i, if $\lambda_i \to 0$ then $\lambda_i^{-1} \to \infty$.

Regularity Condition 3

Next we need to show that $I(\theta)$ is positive definite. A matrix A is positive definite if for all vectors v holds that $v^T A v > 0$. From Equation 3.2 and Equation 3.8, we derive $I(\theta)$ as follows:

$$I(\theta) = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\left(\frac{\mathbb{E}[X_{ij}]}{(p\lambda_i^\eta + c)^2}u_iu_i^T - \log(\lambda_i)\lambda_i^\eta\left(\frac{\mathbb{E}[X_{ij}]}{p\lambda_i^\eta + c} - 1\right)\begin{bmatrix}0 & 1 & 0\\ 1 & p\log(\lambda_i) & 0\\ 0 & 0 & 0\end{bmatrix}\right)$$

$$= \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{1}{(p\lambda_i^\eta + c)}\,u_iu_i^T = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{(p\lambda_i^\eta + c)}\,u_iu_i^T \tag{3.19}$$

Since we have three or more spike concentrations, if you pick three different values for i: $i_1, i_2, i_3 \in \{1, 2, \dots, m\}$, the vectors $\{u_{i_1}, u_{i_2}, u_{i_3}\}$ are linearly independent, i.e. there are no constants $a_1, a_2 \in \mathbb{R}$ such that $u_{i_1} = a_1 \cdot u_{i_2} + a_2 \cdot u_{i_3}$. This means that the matrix $I(\theta)$ has full rank 3 and thus has three non-zero eigenvalues. We also know that for all $v \in \mathbb{R}^3$ there is a $k \in \{1, 2, \dots, m\}$ such that $v^T u_k \neq 0$ if $v \neq 0$, which gives:

$$v^T I(\theta) v = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{(p\lambda_i^\eta + c)}\,v^T u_i u_i^T v = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{(p\lambda_i^\eta + c)}\,(u_i^T v)^2$$

$$\geq \frac{1}{(p\lambda_k^\eta + c)}\,(u_k^T v)^2 > 0 \tag{3.20}$$

This shows that $I(\theta)$ is positive definite.
Next we need to show that $I(\theta_0)$ is the variance-covariance matrix of $(mn)^{-1/2}U_{mn}(\theta_0)$. We will first determine $\mathbb{E}[U_{mn}(\theta_0)]$ and $\mathbb{E}[U_{mn}(\theta_0)U_{mn}(\theta_0)^T]$. For this, note that substituting θ_0 in u_i gives: $\tilde{u}_i := u_i|_{\theta_0} = (\lambda_i, p\log(\lambda_i)\lambda_i, 1)^T$.

$$\mathbb{E}[U_{mn}(\theta_0)] = \mathbb{E}\left[\sum_{i=1}^{m}\sum_{j=1}^{n}\left(\frac{X_{ij}}{p\lambda_i} - 1\right)\tilde{u}_i\right]$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{n}\left(\frac{\mathbb{E}[X_{ij}]}{p\lambda_i} - 1\right)\tilde{u}_i = 0 \tag{3.21}$$

$$\mathbb{E}[U_{mn}(\theta_0)U_{mn}(\theta_0)^T] = \mathbb{E}\left[\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{m}\sum_{l=1}^{n}\left(\frac{X_{ij}}{p\lambda_i} - 1\right)\left(\frac{X_{kl}}{p\lambda_k} - 1\right)\tilde{u}_i\tilde{u}_k^T\right]$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{m}\sum_{l=1}^{n}\mathbb{E}\left[\frac{X_{ij} - \mathbb{E}[X_{ij}]}{p\lambda_i}\,\frac{X_{kl} - \mathbb{E}[X_{kl}])]}{p\lambda_k}\right]\tilde{u}_i\tilde{u}_k^T$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{m}\sum_{l=1}^{n}\frac{\operatorname{cov}(X_{ij}, X_{kl})}{p^2\lambda_i\lambda_k}\,\tilde{u}_i\tilde{u}_k^T$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{n}\frac{\operatorname{var}(X_{ij})}{p^2\lambda_i^2}\,\tilde{u}_i\tilde{u}_i^T = \sum_{i=1}^{m}\sum_{j=1}^{n}\frac{1}{p\lambda_i}\,\tilde{u}_i\tilde{u}_i^T$$

$$= n\sum_{i=1}^{m}\frac{1}{p\lambda_i}\,\tilde{u}_i\tilde{u}_i^T \tag{3.22}$$

Note that X_{ij} is independent and Poisson distributed with parameter $p\lambda_i^\eta + c$ as explained in Chapter 2, and thus $\operatorname{cov}(X_{ij}, X_{kl}) = \operatorname{var}(X_{ij})$ if $i, j = k, l$ and $\operatorname{cov}(X_{ij}, X_{kl}) = 0$ otherwise.

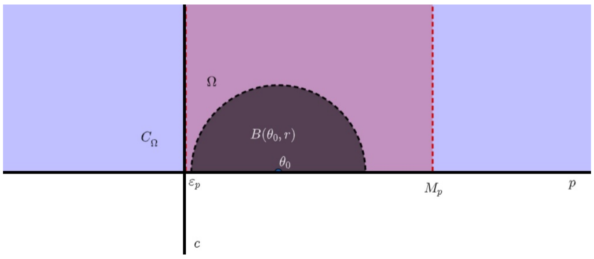


Figure 3.1: Visualisation $B(\theta_0, r)$, with p on the x-axis and c on the y-axis

Also $\mathbb{E}[X_{ij}|\theta_0] = \text{var}(X_{ij}|\theta_0) = p\lambda_i$. Next we substitute Equation 3.21 and 3.22 into the variance-covariance matrix definition:

$$\begin{aligned}\frac{1}{mn}\operatorname{cov}\left(U_{mn}(\theta_0), U_{mn}(\theta_0)\right) &= \frac{1}{mn}\left(\mathbb{E}[U_{mn}(\theta_0)U_{mn}(\theta_0)^T] - \mathbb{E}[U_{mn}(\theta_0)]\mathbb{E}[U_{mn}(\theta_0)]^T\right) \\ &= \frac{1}{m}\sum_{i=1}^{m}\frac{1}{p\lambda_i}\,\tilde{u}_i\tilde{u}_i^T \qquad (3.23)\end{aligned}$$

This is equal to $I(\theta_0)$, since substituting θ_0 in Equation 3.19 gives:

$$I(\theta_0) = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{p\lambda_i}\tilde{u}_i\tilde{u}_i^T$$

Regularity Condition 4

As mentioned in Chapter 2, $\Omega = (\varepsilon_p, M_p) \times (0, \infty] \times [0, \infty)$. Also let $C_\Omega = \mathbb{R}^2 \times [0, \infty)$. The definition of a cone approximation given in Section 2.4 will be used to show that C_Ω approximates Ω. We have that $\Omega \subset C_\Omega$, which means that $\inf_{x \in C_\Omega} ||x - y|| = 0 = o(||y - \theta_0||)$ for all $y \in \Omega$. This holds since for all $y \in \Omega$ we can choose $x \in C_\Omega$ equal to y such that $||x - y|| = 0$. Hence Equation 2.5 is satisfied. For a particular true parameter $\theta_0 = (p_0, 1, 0)^T$, there will be a half-sphere $B(\theta_0, r) := \{(p, \eta, c) : c \geq 0, ||(p, \eta, c) - (p_0, 1, 0)|| < r\}$ around θ_0 with radius $r > 0$ such that $B(\theta_0, r) \subset \Omega$. This means for $x \in B(\theta_0, r)$: $\inf_{y \in \Omega} ||x - y|| = 0$, while, as long as $x \neq \theta_0$, $||x - \theta_0|| > 0$. A visualization of the half-sphere is presented in Figure 3.1. No matter how x converges to θ_0, x will pass through the sphere where $\inf_{y \in \Omega} ||x - y|| = 0$ while $||x - \theta_0||$ is not yet 0. This means that $\frac{\inf_{y \in \Omega} ||x-y||}{||x-\theta_0||} \to 0$ if $x \to \theta_0$, which implies Equation 2.6: $\inf_{y \in \Omega} ||x - y|| = o(||x - \theta_0||)$.

Next, the definition of Section 2.4 will be used with $\Omega_0 = (\varepsilon_p, M_p) \times \{1\} \times \{0\}$ being approximated by the cone $C_{\Omega_0} = \mathbb{R} \times \{1\} \times \{0\}$. We again have that $\Omega_0 \subset C_{\Omega_0}$ and thus $\inf_{x \in C_{\Omega_0}} ||x - y|| = 0 = o(||y - \theta_0||)$ for all $y \in \Omega_0$, implying Equation 2.5. Also we again have a half-sphere: $\tilde{B}(\theta_0, r) := \{(p, 1, 0) : ||p - p_0|| < r\}$ such that $\tilde{B}(\theta_0, r) \subset \Omega_0$. Then again, for all $x \in \tilde{B}(\theta_0, r)$ we have that $\inf_{y \in \Omega_0} ||x - y|| = 0$ for all $x \in \tilde{B}(\theta_0, r)$. This means that here $\inf_{y \in \Omega_0} ||x - y||$ reaches 0 before $||x - \theta_0||$, if $x \to \theta_0$, and thus $\inf_{y \in \Omega_0} ||x - y|| = o(||x - \theta_0||)$. This means that both Ω and Ω_0 are regular enough to be approximated by cones.

Distribution LRT: Case 1, Joint Hypothesis

Since we now know that Theorem 3 of Self & Liang (1987) can be applied to our setting, we know that the likelihood ratio test can be written as shown in Equation 3.6. Now we can apply an orthonormal transformation onto $\tilde{C}$ and $\tilde{C}_0$ such that $\tilde{C} = [0, \infty) \times \mathbb{R}^2$and $\tilde{C}_0 = \{0\}^2 \times \mathbb{R}$. Note $\tilde{C}_0$ is calculated from $C_{\Omega_0} - \theta_0$, and thus the vertex of the cone $\tilde{C}_0$ is at the origin. This gives the following:

$$\inf_{\theta \in \tilde{C}_0} ||\tilde{Z} - \theta||^2 = \inf_{\theta \in \{0\}^2 \times \mathbb{R}} \left|\left| \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \end{bmatrix} - \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} \right|\right| = Z_1^2 + Z_2^2 \tag{3.24}$$

$$\inf_{\theta \in \tilde{C}} ||\tilde{Z} - \theta||^2 = \inf_{\theta \in [0,\infty) \times \mathbb{R}^2} \left|\left| \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \end{bmatrix} - \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} \right|\right| = Z_1^2 \cdot \mathbb{1}\{Z_1 < 0\} \tag{3.25}$$

Here $\theta = (\theta_1, \theta_2, \theta_3)^T$ and Z_i, $i = 1, 2, 3$, is a standard normal random variable. Combining these two gives:

$$\begin{aligned} \text{LRT} &= \inf_{\theta \in \tilde{C}_0} ||\tilde{Z} - \theta||^2 - \inf_{\theta \in \tilde{C}} ||\tilde{Z} - \theta||^2 = Z_1^2 + Z_2^2 - Z_1^2 \cdot \mathbb{1}\{Z_1 < 0\} \\ &= Z_1^2 \cdot \mathbb{1}\{Z_1 \geq 0\} + Z_2^2 \end{aligned} \tag{3.26}$$

Note that a chi-squared distribution with d degrees of freedom, χ_d^2 is the sum of d squared standard normal distributions. This means that if $Z_1 \geq 0$ then the LRT is χ_2^2, but if $Z_1 < 0$, the LRT is χ_1^2. The probability that $Z_1 < 0$ is exactly $\frac{1}{2}$, from which we can conclude that the likelihood ratio test statistic is a mixture of χ_1^2 and χ_2^2, both with ratio $\frac{1}{2}$, i.e. $\frac{1}{2}\chi_1^2 + \frac{1}{2}\chi_2^2$. The p-value for this LRT is as follows:

$$\text{p-val} = \frac{1}{2}\mathbb{P}(\chi_1^2 > LRT) + \frac{1}{2}\mathbb{P}(\chi_2^2 > LRT)$$

3.2.2 Verification Regularity Conditions: Case 2, Sequential Hypotheses(η, c)

Instead of testing both $\eta = 1$ and $c = 0$ at the same time, here the testing will be done sequentially. First $H_{01} : \eta = 1$ will be tested against the alternative hypothesis $H_{11} : \eta \neq 1$. If H_{01} is not rejected, then $H_{02} : c = 0$ will be tested against the alternative: $H_{12} : c > 0$, under the assumption that η is known and equal to 1. For the first test we know that the distribution of the LRT is χ_1^2, because here none of the tested parameters are on the boundary of the parameter space. To find the distribution of the LRT of the second test, we will use the same theorem as in Case 1. To use this theorem we will have to show that the four assumptions hold in the case when testing $c = 0$ after accepting the null-hypothesis $H_{01} : \eta = 1$, if the assumption mentioned below hold.

In this section, $\theta = (\theta_1, \theta_2)^T = (p, c)^T$. Similarly to Case 1, we have again assume that there is a ε_p and M_p such that $0 < \varepsilon_p < p < M_p$, which means the parameter space is as follows: $\Omega = (\varepsilon_p, M_p) \times [0, \infty)$. From the null-hypothesis we get $\theta_0 = (p, 0)^T$ and $\Omega_0 = (\varepsilon_p, M_p) \times \{0\}$. Again choosing the boundary values for parameter$p$ should not create a problem, as they can be chosen as small or as large as you want.

Regularity Condition 1

For the first assumption, we need to show the first three derivatives of the log-likelihood exist on the intersection of neighborhoods of θ_0 and Ω. In this case the the log-likelihood, $l_{mn}(\theta)$, is as follows:

$$l_{mn}(\theta) = \sum_{i=1}^{m}\sum_{j=1}^{n}\left[-p\lambda_i - c + X_{ij}\cdot\log(p\lambda_i + c) - \log(X_{ij}!)\right] \tag{3.27}$$

Next, similarly to Case 1, there are $\delta_p > 0$ and $\delta_c > 0$ such that for neighborhoods $N(\theta_0)$ of θ_0:

$$N(\theta_0) \subset (\varepsilon_p - \delta_p, M_p + \delta_p) \times (-\delta_c, \delta_c)$$

And the intersection of $N(\theta_0)$ and Ω is a subset of $(\varepsilon_p, M_p) \times [0, \delta_c)$.
From Case 1 we already know that the derivatives of $\log(p\lambda_i + c)$ exist if $p\lambda_i + c > 0$, which holds, since $p > 0$, $c \geq 0$, and $\lambda_i > 0$ for all i. This means that the first condition is satisfied. The first two derivatives are denoted by $U_{mn}(\theta)$ and $-I_{mn}(\theta)$ respectively. Here $U_{mn}(\theta)$ is a vector of order 2 and $I_{mn}(\theta)$ is a 2×2 matrix. $U_{mn}(\theta)$ and $I_{mn}(\theta)$ can be written as follows, where $u_i = (\lambda_i, 1)^T$:

$$U_{mn}(\theta) = \begin{bmatrix}\frac{\partial}{\partial p}l_{mn}(\theta)\\ \frac{\partial}{\partial c}l_{mn}(\theta)\end{bmatrix} = \sum_{i=1}^{m}\sum_{j=1}^{n}\left(\frac{X_{ij}}{p\lambda_i + c} - 1\right)\begin{bmatrix}\lambda_i\\ 1\end{bmatrix} = \sum_{i=1}^{m}\sum_{j=1}^{n}\left(\frac{X_{ij}}{p\lambda_i + c} - 1\right)u_i \tag{3.28}$$

$$I_{mn}(\theta) = \sum_{i=1}^{m}\sum_{j=1}^{n}\frac{X_{ij}}{(p\lambda_i + c)^2}\begin{bmatrix}\lambda_i^2 & \lambda_i\\ \lambda_i & 1\end{bmatrix} = \sum_{i=1}^{m}\sum_{j=1}^{n}\frac{X_{ij}}{(p\lambda_i + c)^2}u_i u_i^T \tag{3.29}$$

Note that here u_i is know independent of the parameters and thus $u_i|_{\theta_0} = u_i$.

Regularity Condition 2

For the second condition we need to show that on the intersection of neighborhoods of θ_0 and Ω the third derivatives of the log-likelihood are bounded by a function of $X_{11}, \ldots, X_{mn}$ with finite expectation. Note that, again, $0 < \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_m$, and thus $\left|\frac{\lambda_i}{p\lambda_i + c}\right| \leq \left|\frac{\lambda_i}{p\lambda_i}\right| \leq \frac{1}{p} < \frac{1}{\varepsilon_p}$. Also $\left|\frac{1}{p\lambda_i + c}\right| \leq \left|\frac{1}{p\lambda_i}\right| \leq \frac{1}{\varepsilon_p\lambda_1}$. As for the expectation of $\bar{X} = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}X_{ij}$:$\mathbb{E}[\bar{X}] \leq M_p\lambda_m + \delta_c$. Then the third derivatives and their respective bounds are as follows:

$$\begin{aligned}\frac{1}{mn}\left|\frac{\partial^3}{\partial p^3}l_{mn}(\theta)\right| &= \frac{1}{mn}\left|\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{2\lambda_i^3 X_{ij}}{(p\lambda_i + c)^3}\right| \leq \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\left|\frac{2\lambda_i^3 X_{ij}}{(p\lambda_i + c)^3}\right| \\ &= \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}2\left|\frac{\lambda_i}{p\lambda_i + c}\right|^3 X_{ij} \leq \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{2}{\varepsilon_p^3}{}_{ij} = \frac{2}{\varepsilon_p^3}\bar{X}\end{aligned} \tag{3.30}$$

$$\frac{1}{mn}\left|\frac{\partial^3}{\partial p^2 \partial c} l_{mn}(\theta)\right| = \frac{1}{mn}\left|\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{2\lambda_i^2 X_{ij}}{(p\lambda_i+c)^3}\right| \leq \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\left|\frac{2\lambda_i^2 X_{ij}}{(p\lambda_i+c)^3}\right|$$
$$= \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} 2\left|\frac{\lambda_i}{p\lambda_i+c}\right|^2 \left|\frac{1}{p\lambda_i+c}\right| X_{ij} \leq \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{2}{\varepsilon_p^3\lambda_1} X_{ij} = \frac{2}{\varepsilon_p^3\lambda_1}\bar{X} \tag{3.31}$$

$$\frac{1}{mn}\left|\frac{\partial^3}{\partial p \partial c^2} l_{mn}(\theta)\right| = \frac{1}{mn}\left|\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{2\lambda_i X_{ij}}{(p\lambda_i+c)^3}\right| \leq \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\left|\frac{2\lambda_i X_{ij}}{(p\lambda_i+c)^3}\right|$$
$$= \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} 2\left|\frac{\lambda_i}{p\lambda_i+c}\right| \left|\frac{1}{p\lambda_i+c}\right|^2 X_{ij} \leq \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{2}{\varepsilon_p^3\lambda_1^2} X_{ij} = \frac{2}{\varepsilon_p^3\lambda_1^2}\bar{X} \tag{3.32}$$

$$\frac{1}{mn}\left|\frac{\partial^3}{\partial c^3} l_{mn}(\theta)\right| = \frac{1}{mn}\left|\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{2X_{ij}}{(p\lambda_i+c)^3}\right| \leq \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\left|\frac{2X_{ij}}{(p\lambda_i+c)^3}\right|$$
$$= \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} 2\left|\frac{1}{p\lambda_i+c}\right|^3 X_{ij} \leq \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{2}{\varepsilon_p^3\lambda_1^3} X_{ij} = \frac{2}{\varepsilon_p^3\lambda_1^3}\bar{X} \tag{3.33}$$

Since $\mathbb{E}[\bar{X}] < \infty$, we showed that, on the intersection of neighborhoods of θ_0 and Ω, there are functions f, defined by the bounds found above, for each of the third derivatives of the form $a \cdot \bar{X}$, $a \in \mathbb{R}$, with finite expectation.

Regularity Condition 3

For the third regularity condition two things need to be shown. First we need that $I(\theta)$, from Equation 3.19, is positive definite.

$$I(\theta) = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{\mathbb{E}[X_{ij}]}{(p\lambda_i+c)^2} u_i u_i^T = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{1}{p\lambda_i+c} u_i u_i^T = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{p\lambda_i+c} u_i u_i^T \tag{3.34}$$

Similarly as with Case 1, if you take $i_1, i_2 \in \{1, \dots, m\}$, u_{i_1} and u_{i_2} linearly independent, and $I(\theta)$ is a rank 2 matrix with 2 non-zero eigenvalues. Now if we have a $v \in \mathbb{R}^2$, $k \in \{1, \dots, m\}$ such that $u_k v^t \neq 0$ if $v \neq 0$, which gives:

$$v^T I(\theta) v = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{p\lambda_i+c} v^T u_i u_i^T v = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{(p\lambda_i+c)}(u_i^T v)^2$$
$$\geq \frac{1}{(p\lambda_k+c)}(u_k^T v)^2 > 0 \tag{3.35}$$

Since we now know that $I(\theta)$ is positive definite, next we need to show that $I(\theta_0)$ is the variance-covariance matrix of $\frac{1}{\sqrt{mn}} U_{mn}(\theta_0)$. Next $\mathbb{E}[U_{mn}(\theta_0)]$ and $\mathbb{E}[U_{mn}(\theta_0) U_{mn}(\theta_0)^T]$ are determined:

$$\mathbb{E}[U_{mn}(\theta_0)] = \mathbb{E}\left[\sum_{i=1}^{m}\sum_{j=1}^{n}\left(\frac{X_{ij}}{p\lambda_i}-1\right)u_i\right] = \sum_{i=1}^{m}\sum_{j=1}^{n}\left(\frac{\mathbb{E}[X_{ij}]}{p\lambda_i}-1\right)u_i = 0 \tag{3.36}$$

$$\begin{aligned}
\mathbb{E}[U_{mn}(\theta_0)U_{mn}(\theta_0)^T] &= \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{m}\sum_{l=1}^{n}\left(\frac{X_{ij}}{p\lambda_i}-1\right)\left(\frac{X_{kl}}{p\lambda_k}-1\right)u_i u_k^T \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{m}\sum_{l=1}^{n}\frac{\mathbb{E}[(X_{ij}-\mathbb{E}[X_{ij}])(X_{kl}-\mathbb{E}[X_{kl}])]}{p^2\lambda_i\lambda_k}u_i u_k^T \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{m}\sum_{l=1}^{n}\frac{\mathrm{cov}(X_{ij},X_{kl})}{p^2\lambda_i\lambda_k}u_i u_k^T = \sum_{i=1}^{m}\sum_{j=1}^{n}\frac{\mathrm{var}(X_{ij})}{p^2\lambda_i^2}u_i u_i^T \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n}\frac{1}{p\lambda_i}u_i u_i^T = n\sum_{i=1}^{m}\frac{1}{p\lambda_i}u_i u_i^T
\end{aligned} \tag{3.37}$$

Note $u_i|_{\theta_0} = (\lambda_i, 1)^T = u_i$, and that X_{ij} is independent and Poisson distributed with parameter $p\lambda_i + c$ as explained in Chapter 2, and thus, $\mathrm{cov}(X_{ij}, X_{kl}) = \mathrm{var}(X_{ij})$ if $i, j = k, l$ and $\mathrm{cov}(X_{ij}, X_{kl}) = 0$ otherwise. Also $\mathbb{E}[X_{ij}|\theta_0] = \mathrm{var}(X_{ij}|\theta_0) = p\lambda_i$. Next we substitute Equations 3.363.37, using Equation 3.4, to determine the variance-covariance matrix:

$$\begin{aligned}
\frac{1}{mn}\mathrm{cov}\left(U_{mn}(\theta_0), U_{mn}(\theta_0)\right) &= \frac{1}{mn}\left(\mathbb{E}[U_{mn}(\theta_0)U_{mn}(\theta_0)^T] - \mathbb{E}[U_{mn}(\theta_0)]\mathbb{E}[U_{mn}(\theta_0)]^T\right) \\
&= \frac{1}{m}\sum_{i=1}^{m}\frac{1}{p\lambda_i}u_i u_i^T
\end{aligned} \tag{3.38}$$

It is easy to see this equals $I(\theta_0)$ after substituting $c = 0$ in Equation 3.34. This shows that the variance-covariance matrix of $(mn)^{-1/2}U_{mn}(\theta_0)$ is $I(\theta_0)$. Now we know that the third condition is satisfied for Case 2.

Regularity Condition 4

In Case 2, $\Omega = (\varepsilon_p, M_p) \times [0, \infty)$ and $\Omega_0 = (\varepsilon_p, M_p) \times \{0\}$. We want to show these two sets can be approximated by the cones $C_\Omega = \mathbb{R} \times [0, \infty)$ and $C_{\Omega_0} = \mathbb{R} \times \{0\}$, respectively. Similarly as in Case 1, $\Omega \subset C_\Omega$, which means that $\inf_{x \in C_\Omega} ||x - y|| = 0 = o(||y - \theta_0||)$ for all $y \in \Omega$. If you define $B(\theta_0, r) := \{(p, c) : c \geq 0, ||(p, c) - (p_0, 0)|| < r\}$ such that $B(\theta_0, r) \subset \Omega$, the same way as you did with Case 1, you again have that $\inf_{y \in \Omega} ||x - y|| = 0$ for $x \in B(\theta_0, r)$. This means that if $x \to \theta_0$ then $\inf_{y \in \Omega} ||x - y||$ will equal 0 before $||x - \theta_0|| = 0$ and thus $\inf_{y \in \Omega} ||x - y|| = o(||x - \theta_0||)$. This means that C_Ω approximates Ω, following the definition of a cone approximation in Section 2.4.
We use the same method to show that $C_{\Omega_0} = \mathbb{R} \times \{0\}$ is a cone-approximation of Ω_0. Now $\Omega_0 \subset C_{\Omega_0}$, and thus $\inf_{x \in C_{\Omega_0}} ||x - y|| = 0 = o(||y - \theta_0||)$ for all $y \in \Omega_0$, and Equation 2.5 is satisfied. Next define $\tilde{B}(\theta_0,) = \{(p, 0) : ||p - p_0|| < r\}$ with r such that $\tilde{B}(\theta_0,) \subset \Omega_0$. Here we again have that if $x \in \tilde{B}(\theta_0, r)$ then $\inf_{y \in \Omega_0} ||x - y|| = 0$. Now if $x \to \theta_0$, where $x \in C_{\Omega_0}$ we again have that $\inf_{y \in \Omega_0} ||x - y||$ will equal 0 before $||x - \theta_0|| = 0$ and thus $\inf_{y \in \Omega_0} ||x - y|| = o(||x - \theta_0||)$. Since this means that both Equations Equation 2.5 and

Equation 2.6 are satisfied for both Ω and Ω_0, we know both sets are regular enough to be approximated by cones.

Distribution $\mathbf{LRT_2}$: Case 2, Sequential Hypotheses(η, c)

Now we can apply Theorem 3 of Self & Liang (1987), and the likelihood ratio test can be written as in Equation 3.6. After applying an orthonormal transformation on $\tilde{C}$ and $\tilde{C}_0$ such that $\tilde{C} = [0,\infty) \times \mathbb{R}$ and $\tilde{C}_0 = \{0\} \times \mathbb{R}$, we get the following:

$$\inf_{\theta \in \tilde{C}_0} ||\tilde{Z} - \theta||^2 = \inf_{\theta \in \{0\}\times\mathbb{R}} \left|\left| \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} - \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} \right|\right| = Z_1^2 \tag{3.39}$$

$$\inf_{\theta \in \tilde{C}} ||\tilde{Z} - \theta||^2 = \inf_{\theta \in [0,\infty)\times\mathbb{R}} \left|\left| \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} - \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} \right|\right| = Z_1^2 \cdot \mathbb{1}\{Z_1 < 0\} \tag{3.40}$$

Where $\theta = (\theta_1, \theta_2)^T$ and Z_1 and Z_2 are standard normal random variables. Substituting the above equations in the Equation 3.6 gives the following:

$$\begin{aligned} LRT_2 &= \inf_{\theta \in \tilde{C}_0} ||\tilde{Z} - \theta||^2 - \inf_{\theta \in \tilde{C}} ||\tilde{Z} - \theta||^2 = Z_1^2 - Z_1^2 \cdot \mathbb{1}\{Z_1 < 0\} \\ &= Z_1^2 \cdot \mathbb{1}\{Z_1 \geq 0\} \end{aligned} \tag{3.41}$$

Here, LRT_2 is the LRT of the second hypothesis of Case 2: $H_{02} : c = 0$This means that with a probability $\frac{1}{2}$ the LRT will be equal to 0 and with probability $\frac{1}{2}$ the LRT will be χ_1^2. This, since Z_i^2 is χ_1^2 and $Z_i > 0$ with probability $\frac{1}{2}$. Here, the p-value is as follows:

$$\text{p-val} = \frac{1}{2}\mathbb{P}(0 > LRT_2) + \frac{1}{2}\mathbb{P}(\chi_1^2 > LRT_2) = \frac{1}{2}\mathbb{P}(\chi_1^2 > LRT_2)$$

Note $\mathbb{P}(0 > LRT_2) = 0$ since the likelihood ratio test is always non-negative.

3.2.3 Verification Regularity Conditions: Case 3, Sequential Hypotheses(c, η)

In Case 3, sequential hypotheses(c, η), first $H_{01} : c = 0$ is tested and then $H_{02} : \eta = 1$ is tested if H_{01} is not rejected. When testing the second hypothesis, we assume c to be equal to 0. Here we assume the parameter space for the first null-hypothesis is as follows: $\Omega = (\varepsilon_p, M_p) \times (0, M_\eta] \times [0, \infty)$, where $\varepsilon_p < p < M_p$ and $0 \leq \eta \leq M_\eta$. Here, M_η can be chosen as large as needed, similarly to M_p, and thus will not create a problem in practical settings. Note $\theta = (\theta_1, \theta_2, \theta_3)^T = (p, \eta, c)^T$. Next, for the first null-hypothesis, we have $\theta_0 = (p, \eta, 0)^T$ and $\Omega_0 = (\varepsilon_p, M_p) \times (0, M_\eta] \times \{0\}$. Note that Ω is the same as in Case 1, joint hypothesis (Section 3.2.1), which results in the log-likelihood function and derivatives. For the second null-hypothesis, $H_{02} : \eta = 1$, none of the tested parameters are on the boundary of the parameter space and the distribution of the LRT is thus χ_1^2.

Regularity Condition 1

As mentioned before the parameter space Ω is the same for this case, Case 3, as Case 1, joint hypothesis. This means that the log-likelihood functions, $l_{mn}(\theta)$ is the same as in Equations

?? and 2.1. For the neighborhoods of θ_0, $N(\theta_0)$, there are $\delta_p > 0$, $\delta_\eta > 0$ and $\delta_c > 0$ such that:

$$N(\theta_0) \subset (\varepsilon_p - \delta_p, M_p + \delta_p) \times (0 - \delta_\eta, M_\eta + \delta_\eta) \times (-\delta_c, \delta_c)$$

The intersection of $N(\theta_0)$ with Ω gives the following set:

$$N(\theta_0) \cap \Omega = (\varepsilon_p, M_p) \times (0, M_\eta] \times [0, \delta_c)$$

For $(p, \eta, c)^T \in N(\theta_0) \cap \Omega$, we have that $p\lambda_i^\eta + c > 0$. This means that, as in Case 1 (Section 3.2.1), the derivatives of the log-likelihood function $l_{mn}(\theta_0)$ exist on the intersection of neighborhoods of θ_0 and Ω and the first condition holds.
Since the log-likelihood is the same as in Case 1, the second and third derivatives, $U_{mn}(\theta)$ and $-I_{mn}(\theta)$, are the same as Case 1 as well. $U_{mn}(\theta)$ and $I_{mn}(\theta)$ can be seen in Equations 3.7 and 3.8 respectively.

Regularity Condition 2

For this regularity condition, we need to show that the third derivatives of $l_{mn}(\theta)$ are bounded by a function with finite expectation on the intersection of neighborhoods of θ_0 and Ω. Here the third derivatives are equal to the third derivatives of the log-likelihood in Case 1 (see Equation 3.9 till3.18).
Take $0 < \lambda_1 < \lambda_2 < \ldots < \lambda_m$ without loss of generality. also define $\mathrm{M}_{|\log\lambda|} := \max_{i=1,\ldots,m} |\log(\lambda_i)|$ and $\bar{X} := \frac{1}{mn}\sum_{i=1}^m \sum_{j=1}^n X_{ij}$, where $\mathbb{E}[\bar{X}] \leq M_p \max\{\lambda_m^0, \lambda_m^{M_p}\} + \delta_c$ similar to in Case 1. Note here we need the upper-bound on η, otherwise the expectation could become infinitely large if $\lambda_m > 1$.
The only difference between the intersection of neighborhoods of θ_0 and Ω for Case 1, joint hypothesis, and Case 3, sequential hypotheses(c, η), lies in the interval of η. In Case 1 the interval for η is $(1-\delta_\eta, 1+\delta_\eta)$, while, here in Case 3, this interval is $[0, M_\eta]$. We know that the bounds in Case 1 include the constants $\lambda_1^{-\eta_s} := \max\{\lambda_1^{-1-\delta_\eta}, \lambda_1^{-1+\delta_\eta}\}$ and $\lambda_m^{\eta_M} := \max\{\lambda_m^{1+\delta_\eta}, \lambda_m^{1-\delta_\eta}\}$, such that $\lambda_1^{\eta_s} \leq \lambda_2^\eta \leq \ldots \leq \lambda_{m-1}^\eta \leq \lambda_m^{\eta_M}$, where η is in the intersection of neighborhoods of θ_0 and Ω of Case 1. For Case 3, define similarly, $\lambda_1^{-\tilde{\eta}_s} := \max\{\lambda_1^0, \lambda_1^{M_\eta}\}$ and $\lambda_m^{\tilde{\eta}_M} := \max\{\lambda_m^0, \lambda_m^{M_\eta}\}$, such that $\lambda_1^{\tilde{\eta}_s} \leq \lambda_2^{\tilde{\eta}} \leq \ldots \leq \lambda_{m-1}^{\tilde{\eta}} \leq \lambda_m^{\tilde{\eta}_M}$, for $\tilde{\eta}$ in the intersection of neighborhoods of θ_0 and Ω for Case 3. If $\lambda_1^{\eta_s}$ and $\lambda_m^{\eta_M}$ are replaced by $\lambda_1^{\tilde{\eta}_s}$ and $\lambda_m^{\tilde{\eta}_M}$ in Equations 3.9 till 3.18, you get bounds for the third derivatives with finite expectation for Case 3.

Regularity Condition 3

As mentioned, $I_{mn}(\theta)$ in Case 3 is the same as in Case 1. Since $I(\theta)$ is not dependent on the null-hypothesis, $I(\theta)$ is also the same, and can be seen in Equation 3.19. From this we can directly conclude that, as in Case 1, $I(\theta)$ is positive definite. Next we need to show that the variance-covariance matrix of $\frac{1}{\sqrt{mn}} U_{mn}(\theta_0)$ is $I(\theta_0)$, we determine $\mathbb{E}[U_{mn}(\theta_0)]$ and $\mathbb{E}[U_{mn}(\theta_0)U_{mn}(\theta_0)^T]$, where $u_i = (\lambda_i^\eta, p\log(\lambda_i)\lambda_i^\eta, 1)^T$. Note that in this Case, as mentioned before, $u_i|_{\theta_0} = u_i$.

$$\begin{aligned}
\mathbb{E}[U_{mn}(\theta_0)] &= \mathbb{E}\left[\sum_{i=1}^{m}\sum_{j=1}^{n}\left(\frac{X_{ij}}{p\lambda_i^\eta} - 1\right)u_i\right] \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n}\left(\frac{\mathbb{E}[X_{ij}]}{p\lambda_i^\eta} - 1\right)u_i = 0 && (3.42) \\
\mathbb{E}[U_{mn}(\theta_0)U_{mn}(\theta_0)^T] &= \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{m}\sum_{l=1}^{n}\left(\frac{X_{ij}}{p\lambda_i^\eta} - 1\right)\left(\frac{X_{kl}}{p\lambda_k^\eta} - 1\right)u_i u_k^T \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{m}\sum_{l=1}^{n}\frac{\mathbb{E}[(X_{ij} - \mathbb{E}[X_{ij}])(X_{kl} - \mathbb{E}[X_{kl}])]}{p^2\lambda_i^\eta\lambda_k^\eta}u_i u_k^T \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{m}\sum_{l=1}^{n}\frac{\mathrm{cov}(X_{ij}, X_{kl})}{p^2\lambda_i^\eta\lambda_k^\eta}u_i u_k^T \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n}\frac{\mathrm{var}(X_{ij})}{p^2\lambda_i^{2\eta}}u_i u_i^T = \sum_{i=1}^{m}\sum_{j=1}^{n}\frac{1}{p\lambda_i^\eta}u_i u_i^T \\
&= n\sum_{i=1}^{m}\frac{1}{p\lambda_i^\eta}u_i u_i^T && (3.43)
\end{aligned}$$

As mentioned in Chapter 2, X_{ij} is independent and Poisson distributed with parameter $p\lambda_i^\eta + c$. This means that the covariance between X_{ij} and X_{kl} equals the variance of X_{ij} if $(i, j) = (k, l)$ and is 0 otherwise. Note that $\mathbb{E}[X_{ij}|\theta_0] = \mathrm{var}(X_{ij}|\theta_0) = p\lambda_i^\eta$. Substituting the Equations 3.42 and 3.43, using Equation 3.4 into the variance-covariance matrix, gives the following:

$$\begin{aligned}
\frac{1}{mn}\mathrm{cov}\left(U_{mn}(\theta_0), U_{mn}(\theta_0)\right) &= \frac{1}{mn}\left(\mathbb{E}[U_{mn}(\theta_0)U_{mn}(\theta_0)^T] - \mathbb{E}[U_{mn}(\theta_0)]\mathbb{E}[U_{mn}(\theta_0)]^T\right) \\
&= \frac{1}{m}\sum_{i=1}^{m}\frac{1}{p\lambda_i^\eta}u_i u_i^T && (3.44)
\end{aligned}$$

Which equals $I(\theta_0)$, as can be seen if you substitute $\theta = \theta_0$, i.e. $c = 0$, in Equation 3.19. This means that for Case 3, the third regularity condition is satisfied.

Regularity Condition 4

Since Ω is the same as the Ω in Case 1, it can be approximated by the cone $C_\Omega = \mathbb{R}^2 \times [0, \infty)$ as shown for Case 1. Next we need to show that $\Omega_0 = (\varepsilon_p, M_p) \times (0, M_\eta] \times \{0\}$ can be approximated by the cone $C_{\Omega_0} = \mathbb{R}^2 \times \{0\}$. Again, since $\Omega_0 \subset C_{\Omega_0}$ and thus $\inf_{x \in C_{\Omega_0}} ||x - y|| = 0 = o(||y - \theta_0||)$ for all $y \in \Omega_0$, and Equation 2.5 is satisfied. Now define the following half-sphere: $B(\theta_0, r) := \{x = (p, \eta, 0) : ||x - \theta_0|| < r\}$ such that $B(\theta_0, r) \subset \Omega_0$. Then $\inf_{y \in \Omega_0} ||x - y|| = 0$ for all $x \in B(\theta_0, r)$. If $x \to \theta_0$, we now know that $\inf_{y \in \Omega_0} ||x - y||$ reaches 0 before $||x - \theta_0||$, and thus $\inf_{y \in \Omega_0} ||x - y|| = o(||x - \theta_0||)$, implying Equation 2.6 is also satisfied. From this we can conclude that both Ω and Ω_0 are regular enough to be approximated by cones.

Distribution LRT_1: Case 3, Sequential Hypotheses(c, η)

Now we can use Theorem 3 of Self & Liang (1987), which shows that the likelihood ratio test can be written as in Equation 3.6. After an orthonormal transformation on $\tilde{C}$ and $\tilde{C}_0$, we have that $\tilde{C} = [0, \infty) \times \mathbb{R}^2$and $\tilde{C}_0 = \{0\} \times \mathbb{R}^2$. This gives the following:

$$\inf_{\theta \in \tilde{C}_0} ||\tilde{Z} - \theta||^2 = \inf_{\theta \in \{0\} \times \mathbb{R}^2} \left|\left| \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \end{bmatrix} - \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} \right|\right| = Z_1^2 \tag{3.45}$$

$$\inf_{\theta \in \tilde{C}} ||\tilde{Z} - \theta||^2 = \inf_{\theta \in [0,\infty) \times \mathbb{R}^2} \left|\left| \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \end{bmatrix} - \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} \right|\right| = Z_1^2 \cdot \mathbb{1}\{Z_1 < 0\} \tag{3.46}$$

Here $\theta = (\theta_1, \theta_2, \theta_3)^T$ and Z_i, $i = 1, 2, 3$, is a standard normal random variable. After substituting this in Equation 3.6, we have:

$$\begin{aligned} LRT_1 &= \inf_{\theta \in \tilde{C}_0} ||\tilde{Z} - \theta||^2 - \inf_{\theta \in \tilde{C}} ||\tilde{Z} - \theta||^2 = Z_1^2 - Z_1^2 \cdot \mathbb{1}\{Z_1 < 0\} \\ &= Z_1^2 \cdot \mathbb{1}\{Z_1 \geq 0\}, \end{aligned} \tag{3.47}$$

where LRT_1 is the LRT of the first hypothesis in Case 3. This is the same distribution of LRT_2 in Case 2, where the LRT test statistic is 0 with probability $\frac{1}{2}$ and χ_1^2 with probability $\frac{1}{2}$. This gives the p-value p-val $= \frac{1}{2}\mathbb{P}(\chi_1^2 > LRT_1)$.

3.3 Distribution LRT with one blank spike concentration

If one of the spike concentrations is blank ($\lambda_i = 0$), then the theorem of Self & Liang (1987) cannot be used, for the derivatives around $c = 0$ do not exist. Also, in for example Equation 3.11, the function for which the third derivative is bounded depends on $\lambda_1^{-\eta_s}$, which will converge to infinity if $\lambda_1 \to 0$. This means there will be no function such that the third derivatives will be bounded in expectation.
In this section we will first try to determine the distribution of the LRT test statistic when there are exactly three spike concentrations for Case 1, joint Hypothesis. This will not be done for the other two cases, since explicit values only depending on X_{ij} or λ_i cannot be found for p_1, η_1 or p_2, c_2 such that $l_{mn}(p_1, \eta_1, 0) = \sup_{p,\eta} l_{mn}(p, \eta, 0)$ and $l_{mn}(p_2, 1, c_2) = \sup_{p,c} l_{mn}(p, 1, c)$. After this, the distribution of the LRT under null-hypothesis will be determined for more than three spike concentrations.

3.3.1 Three Spike Concentrations: Case 1, Joint Hypothesis

To find the distribution of the likelihood ratio test we will first try to determine the parameter estimators for which the log-likelihood is maximal in a general setting. The log-likelihood, $l_{mn}(\theta)$ can be seen in Equation 2.1, the first derivatives to p, η and c can be found respectively in Equations 2.2, 2.3 and 2.4. Once the maximal likelihood estimators are found, we will use these in the found expression of the likelihood ratio test is Section 2.3 to determine the distribution of the likelihood ratio test statistic for the joint hypothesis case. In this section we take $m = 3$.

Supremum Log-Likelihood over p, η and c

First the supremum of the log-likelihood over p, η and c is determined. If the supremum of the log-likelihood is achieved at $\hat{\theta}$ then the derivatives are equal to 0 at this point. Without loss of generality assume that $0 = \lambda_1 < \lambda_2 < \lambda_3$, and we define $w_i = p\lambda_i^\eta + c$, then:

$$\frac{\partial}{\partial p} l_{3n}(\hat{\theta}) = \lambda_2^{\hat{\eta}}(X_{2.}\hat{w}_2^{-1} - n) + \lambda_3^{\hat{\eta}}(X_{3.}\hat{w}_3^{-1} - n) = 0 \tag{3.48}$$

$$\frac{\partial}{\partial \eta} l_{3n}(\hat{\theta}) = \hat{p}\log(\lambda_2)\lambda_2^{\hat{\eta}}(X_{2.}\hat{w}_2^{-1} - n) + \hat{p}\log(\lambda_3)\lambda_3^{\hat{\eta}}(X_{3.}\hat{w}_3^{-1} - n) = 0 \tag{3.49}$$

$$\frac{\partial}{\partial c} l_{3n}(\hat{\theta}) = (X_{1.}\hat{w}_1^{-1} - n) + (X_{2.}\hat{w}_2^{-1} - n) + (X_{3.}\hat{w}_3^{-1} - n) = 0 \tag{3.50}$$

Since $\lambda_2 \neq \lambda_3$, this system of equations can only be solved when, $X_{i.}/\hat{w}_i - n = 0$. Rearranging this gives $\bar{X}_{i.} = \hat{p}\lambda_i^{\hat{\eta}} + \hat{c}$ for $i = 1, 2, 3$, where $\bar{X}_{i.} = X_{i.}/n$. Since $\lambda_1 = 0$, we get $\hat{c} = \bar{X}_{1.}$ Substituting this gives $\bar{X}_{i.} - \bar{X}_{1.} = \hat{p}\lambda_i^{\hat{\eta}}$ for $i = 2, 3$. Using this, we have:

$$\hat{\eta} = \log\left(\frac{\bar{X}_{2.} - \bar{X}_{1.}}{\bar{X}_{3.} - \bar{X}_{1.}}\right) / \log\left(\frac{\lambda_2}{\lambda_3}\right) \tag{3.51}$$

Note $\hat{\eta}$ will be infinite if either $X_{1.} = X_{2.}$ or $X_{1.} = X_{3.}$. However, once we look at the asymptotic distribution of the LRT, this is no longer a problem, as the probability that either $X_{1.} = X_{2.}$ or $X_{1.} = X_{3.}$ will converge to zero as n converges to infinity.
Next, can be substituted in $\bar{X}_{2.} - \bar{X}_{1.} = \hat{p}\lambda_2^{\hat{\eta}}$ to obtain the expression for the maximum likelihood estimator of p:

$$\hat{p} = \frac{\bar{X}_{2.} - \bar{X}_{1.}}{\lambda_2^{\hat{\eta}}} = (\bar{X}_{2.} - \bar{X}_{1.})\lambda_2^{-\log\left(\frac{\bar{X}_{2.} - \bar{X}_{1.}}{\bar{X}_{3.} - \bar{X}_{1.}}\right)/\log\left(\frac{\lambda_2}{\lambda_3}\right)} \tag{3.52}$$

Next we substitute $\bar{X}_{i.} = \hat{p}\lambda_i^{\hat{\eta}} + \hat{c}$ in the log-likelihood to get the supremum:

$$\begin{aligned} l_{3n}(\hat{\theta}) &= \sum_{i=1}^{3}\sum_{j=1}^{n}\left[-\bar{X}_{i.} + \bar{X}_{ij} \cdot \log(\bar{X}_{i.}) - \log(X_{ij}!)\right] \\ &= -X_{..} + \sum_{i=1}^{3} n\bar{X}_{i.}\log(\bar{X}_{i.}) - \sum_{i=1}^{3}\sum_{j=1}^{n}\log(X_{ij}!) \end{aligned} \tag{3.53}$$

with $X_{..} = \sum_{i=1}^{3}\sum_{j=1}^{n} X_{ij}$ the total count over all three concentrations.

Supremum Log-Likelihood over p

Next we will determine the value of p such that the log-likelihood under the joint null-hypothesis $H_0 : \eta = 1 \wedge c = 0$, $l_{3n}(p, 1, 0)$, is maximal. Here we have the following log-likelihood and its derivative:

$$\begin{aligned}
l_{3n}(p,1,0) &= \sum_{i=1}^{3}\sum_{j=1}^{n} -p\lambda_i + X_{ij}\log(p\lambda_i) - \log(X_{ij}!) \\
&= -np\lambda_. + X_{..}\log(p) + \sum_{i=1}^{3} X_{i.}\log(\lambda_i) - \sum_{i=1}^{3}\sum_{j=1}^{n}\log(X_{ij}!) \qquad (3.54)\\
\frac{\partial}{\partial p} l_{3n}(p,1,0) &= -n\lambda_. + \frac{X_{..}}{p} \qquad (3.55)
\end{aligned}$$

where $\lambda_. = \sum_{i=1}^{3}\lambda_i$. We know that for $\hat{p}_0$ such that $l_{3n}(\hat{p}_0,1,0)$ is the maximal value for the log-likelihood, that $-n\lambda_. + X_{..}/\hat{p}_0 = 0$. Rearranging this gives $\hat{p}_0 = X_{..}/n\lambda_.$. Substitution this in $l_{3n}(\hat{p}_0,1,0)$ gives the following:

$$\begin{aligned}
l_{3n}(\hat{p}_0,1,0) &= -n\lambda_.\frac{X_{..}}{n\lambda_.} + X_{..}\log(\frac{X_{..}}{n\lambda_.}) + \sum_{i=1}^{3} X_{i.}\log(\lambda_i) - \sum_{i=1}^{3}\sum_{j=1}^{n}\log(X_{ij}!) \\
&= -X_{..} + X_{..}\log(\frac{X_{..}}{n}) + \sum_{i=1}^{3} X_{i.}\log(\frac{\lambda_i}{\lambda_.}) - \sum_{i=1}^{3}\sum_{j=1}^{n}\log(X_{ij}!) \qquad (3.56)
\end{aligned}$$

Next we substitute $l_{3n}(\hat{\theta})$ and $l_{3n}(\hat{p}_0,1,0)$ in the likelihood ratio test:

$$\begin{aligned}
\text{LRT} &= -2\left(\sup_{\theta\in\Omega_0} l_{3n}(\theta) - \sup_{\theta\in\Omega} l_{3n}(\theta)\right) = -2\left(\sup_{(p,1,0)^T\in\Omega_0} l_{3n}(p,1,0) - \sup_{(p,\eta,1)^T\in\Omega} l_{3n}(p,\eta,c)\right)\\
&= -2(l_{3n}(\hat{p}_0,1,0) - l_{3n}(\hat{p},\hat{\eta},\hat{c}))\\
&= -2\left(-X_{..} + X_{..}\log(\frac{X_{..}}{n}) + \sum_{i=1}^{3} X_{i.}\log(\frac{\lambda_i}{\lambda_.}) + X_{..} - \sum_{i=1}^{3} X_{i.}\log(\frac{X_{i.}}{n})\right)\\
&= -2\left(\sum_{i=1}^{3} X_{i.}\log(\frac{\lambda_i X_{..}}{\lambda_. X_{i.}})\right) \qquad (3.57)
\end{aligned}$$

LRT under Null-Hypothesis: Case 1, Joint Hypothesis

Next, using the found suprema above, we will look at the distribution of the LRT under the null-hypothesis. In this case we know that $X_{ij} \sim \text{Poisson}(p\lambda_i)$. Since $\lambda_1 = 0$, all counts for the first solution are zero, i.e. $X_{1j} = 0$. Note that Ω and Ω_0 are $(\varepsilon_p, M_p) \times [0, M_\eta] \times [0,\infty)$ and $(\varepsilon_p, M_p) \times \{1\} \times \{0\}$, respectively. Since we found that $\hat{c}$ of the supremum of $l_{3n}(p,\eta,c)$ equals $\bar{X}_{1.}$, under the null-hypothesis, we get $\hat{c} = 0$. Next we substitute this in the LRT:

$$\begin{aligned}
\text{LRT} &= -2\left(\sup_{\theta\in\Omega_0} l_{3n}(\theta) - \sup_{\theta\in\Omega} l_{3n}(\theta)\right) = -2(l_{3n}(\hat{p}_0,1,0) - l_{3n}(\hat{p},\hat{\eta},\hat{c}))\\
&= -2(l_{3n}(\hat{p}_0,1,0) - l_{3n}(\hat{p},\hat{\eta},0))\\
&= -2\left(\sup_{(p,1,0)^T\in\Omega_0} l_{3n}(p,1,0) - \sup_{(p,\eta,0)\in\Omega} l_{3n}(p,\eta,0)\right) \qquad (3.58)
\end{aligned}$$

Under H_0, this is the same likelihood ratio test as the likelihood ratio test of the null-hypothesis $\eta = 1$ after determining $c = 0$, which is χ^2_1 distributed. This means that also the likelihood ratio test of the joint null-hypothesis $H_0 : \eta = 1 \wedge c = 0$ is χ^2_1 distributed under H_0 if $\lambda_1 = 0$ and $m = 3$.

3.3.2 Suprema Log-Likelihood, More Than Three Spike Concentrations

Each likelihood ratio test is composed of the difference of two different suprema. As we want to determine the asymptotic distribution of the likelihood ratio test statistic, in this section, we will try to determine where several suprema converge to as the sample size n converges to infinity. These suprema can then be used to determine the where the log-likelihood converges as $n \to \infty$, which in turn can be used to determine the asymptotic distribution of the likelihood ratio test. In this section, we have $\lambda_1 = 0$ and $m \geq 3$. The found suprema can then be used to find the asymptotic distribution of the likelihood ratio test for the joint hypothesis case as well as the distribution of the LRT of the test for $c = 0$ for the sequential hypotheses cases. The determination of the distributions is done in Section 3.3.3.
The suprema needed are: $\sup_p l_{mn}(p, 1, 0)$, $\sup_{p,\eta,c} l_{mn}(p, \eta, c)$ and $\sup_{p,\eta} l_{mn}(p, \eta, 0)$. Note that in this section all $(p, \eta, c) \in \Omega$, and thus $\sup_p l_{mn}(p, 1, 0) = \sup_{p>0} l_{mn}(p, 1, 0)$. Next, under null-hypothesis, $\eta = 1$ and $c = 0$, we know that X_{ij} is Poisson distributed with parameter $p\lambda_i$ and thus $X_{1j} = 0$ for all $j = 1, \ldots, n$. For this, note that the log-likelihood function and its first derivatives are as follows:

$$\begin{aligned} l_{mn}(\theta) &= \sum_{i=1}^{m}\sum_{j=1}^{n} \left[-p\lambda_i^\eta - c + X_{ij}\log(p\lambda_i^\eta + c) - \log(X_{ij}!)\right] \\ &= -nc + \sum_{i=2}^{m} \left[-n \cdot (p\lambda_i^\eta + c) + X_{i.}\log(p\lambda_i^\eta + c)\right] - \sum_{i=2}^{m}\sum_{j=1}^{n} \log(X_{ij}!) \end{aligned} \tag{3.59}$$

Note, $\log(X_{1j}!) = \log(0!) = 0$.

$$\frac{\partial}{\partial p} l_{mn}(\theta) = \sum_{i=2}^{m} \lambda_i^\eta \left(\frac{X_{i.}}{p\lambda_i^\eta + c} - n\right) = n\sum_{i=2}^{m} \lambda_i^\eta \left(\frac{\bar{X}_{i.}}{p\lambda_i^\eta + c} - 1\right) \tag{3.60}$$

$$\frac{\partial}{\partial \eta} l_{mn}(\theta) = \sum_{i=2}^{m} p\lambda_i^\eta \log(\lambda_i) \left(\frac{X_{i.}}{p\lambda_i^\eta + c} - n\right) = n\sum_{i=2}^{m} p\lambda_i^\eta \log(\lambda_i) \left(\frac{\bar{X}_{i.}}{p\lambda_i^\eta + c} - 1\right) \tag{3.61}$$

$$\frac{\partial}{\partial c} l_{mn}(\theta) = -n + \sum_{i=2}^{m} \left(\frac{X_{i.}}{p\lambda_i^\eta + c} - n\right) = n\left(-1 + \sum_{i=2}^{m} \left(\frac{\bar{X}_{i.}}{p\lambda_i^\eta + c} - 1\right)\right) \tag{3.62}$$

In contrast to Section 3.3.1, where we first determine the maximum likelihood estimator and then fill in $X_{1j} = 0$, here we first fill in $X_{1j} = 0$ and then determine the maximum likelihood estimator. This does not cause a problem, since filling in $X_{1j} = 0$ in the derivatives of log-likelihood, gives the same derivatives as first filling in $X_{1j} = 0$. Since the maximum likelihood estimators are determined from these derivatives, they will also be equal independent of the order of determining the estimator or filling in X_{1j}.

Supremum Log-Likelihood over p, η and c

Usually the supremum of the log-likelihood is attained at the point where the derivatives are equal to zero. This means, for sample size n, to find $\hat{p}_n$, $\hat{\eta}_n$ and $\hat{c}_n$ such that $l_{mn}(\hat{p}_n, \hat{\eta}_n, \hat{c}_n) =$

$\sup_{p,\eta,c} l_{mn}(p,\eta,c)$, the following equation have to be solved:

$$\sum_{i=2}^{m} \lambda_i^{\hat{\eta}_n} \left(\frac{\bar{X}_{i.}}{\hat{p}_n \lambda_i^{\hat{\eta}_n} + \hat{c}_n} - 1 \right) = 0$$
$$\sum_{i=2}^{m} \hat{p}_n \lambda_i^{\hat{\eta}_n} \log(\lambda_i) \left(\frac{\bar{X}_{i.}}{\hat{p}_n \lambda_i^{\hat{\eta}_n} + \hat{c}_n} - 1 \right) = 0$$
$$-1 + \sum_{i=2}^{m} \left(\frac{\bar{X}_{i.}}{\hat{p}_n \lambda_i^{\hat{\eta}_n} + \hat{c}_n} - 1 \right) = 0$$

At the limit $n \to \infty$, solving the equations above also determine $\hat{p}$, $\hat{\eta}$ and $\hat{c}$, such that $(\hat{p}_n, \hat{\eta}_n, \hat{c}_n) \to (\hat{p}, \hat{\eta}, \hat{c})$. Note that at the supremum, as mentioned before, the derivatives might be unequal to zero if the optimum value of the corresponding parameter is at the boundary of the parameter space. If the optimum parameter value is on the left boundary of the parameter space, then the derivative could be smaller than or equal to zero, and if the parameter value is on the right boundary, then it could be greater than or equal to zero.
The weak law of large numbers (Billingsley, 1995) states that $\bar{X}_{i.}$ converges to its mean in probability. Under the null-hypothesis, this mean is equal to $p\lambda_i$, i.e. $(\bar{X}_{i.} \xrightarrow{p} p\lambda_i)$, and thus $\lim_{n\to\infty} \bar{X}_{i.} \overset{p}{=} p\lambda_i$. Substituting this into the equations above we get:

$$0 = \lim_{n\to\infty} \sum_{i=2}^{m} \lambda_i^{\hat{\eta}_n} \left(\frac{\bar{X}_{i.}}{\hat{p}_n \lambda_i^{\hat{\eta}_n} + \hat{c}_n} - 1 \right) \overset{p}{=} \sum_{i=2}^{m} \lambda_i^{\hat{\eta}} \left(\frac{p\lambda_i}{\hat{p}\lambda_i^{\hat{\eta}} + \hat{c}} - 1 \right)$$
$$0 = \lim_{n\to\infty} \sum_{i=2}^{m} \hat{p}_n \lambda_i^{\hat{\eta}_n} \log(\lambda_i) \left(\frac{\bar{X}_{i.}}{\hat{p}_n \lambda_i^{\hat{\eta}_n} + \hat{c}_n} - 1 \right) \overset{p}{=} \sum_{i=2}^{m} \hat{p}\lambda_i^{\hat{\eta}} \log(\lambda_i) \left(\frac{p\lambda_i}{\hat{p}\lambda_i^{\hat{\eta}} + \hat{c}} - 1 \right)$$
$$0 = \lim_{n\to\infty} -1 + \sum_{i=2}^{m} \left(\frac{\bar{X}_{i.}}{\hat{p}_n \lambda_i^{\hat{\eta}_n} + \hat{c}_n} - 1 \right) \overset{p}{=} -1 + \sum_{i=2}^{m} \left(\frac{p\lambda_i}{\hat{p}\lambda_i^{\hat{\eta}} + \hat{c}} - 1 \right)$$

The right-hand sides of the first two equations equal zero if $(\hat{p}, \hat{\eta}, \hat{c}) = (p, 1, 0)$, and for this value of $(\hat{p}, \hat{\eta}, \hat{c})$, the right-hand side of the third equation equals -1, which is smaller than zero. Also note that for any $c \geq 0$, $\hat{p}\lambda_i^{\hat{\eta}} + c > \hat{p}\lambda_i^{\hat{\eta}}$. This implies that the right-hand side of the third equation is smaller than 0 for all $c \geq 0$. Since $\hat{c} = 0$ is on the boundary of the parameter space, this means that, for $n \to \infty$, $(\hat{p}_n, \hat{\eta}_n, \hat{c}_n) \xrightarrow{p} (p, 1, 0)$. Similarly, if $\tilde{p}_n$ and $\tilde{\eta}_n$, such that $l_{mn}(\tilde{p}_n, \tilde{\eta}_n, 0) = \sup_{p,\eta} l_{mn}(p, \eta, 0)$, then you also have that $(\tilde{p}_n, \tilde{\eta}_n, 0) \to (p, 1, 0)$ for $n \to \infty$.
Since we know that the log-likelihood function under the null-hypothesis is differentiable on the parameter space Ω, we know that it is continuous on Ω. From this we know, if we take $\tilde{\theta} \in \Omega$ then $\lim_{\theta\to\tilde{\theta}} l_{mn}(\theta) = l_{mn}(\tilde{\theta})$. Since we also know that $(\hat{p}_n, \hat{\eta}_n, \hat{c}_n)$ and $(\tilde{p}_n, \tilde{\eta}_n, 0)$ both go to $(p, 1, 0)$ if $n \to \infty$, we have:

$$\begin{aligned} \lim_{n\to\infty} l_{mn}(\hat{p}_n, \hat{\eta}_n, \hat{c}_n) &\overset{p}{=} \lim_{n\to\infty} \lim_{(\hat{p}_n,\hat{\eta}_n,\hat{c}_n)\to(p,1,0)} l_{mn}(\hat{p}_n, \hat{\eta}_n, \hat{c}_n) = \lim_{n\to\infty} l_{mn}(p, 1, 0) \\ &= \lim_{n\to\infty} \lim_{(\tilde{p}_n,\tilde{\eta}_n)\to(p,1)} l_{mn}(\tilde{p}_n, \tilde{\eta}_n, 0) \overset{p}{=} \lim_{n\to\infty} l_{mn}(\tilde{p}_n, \tilde{\eta}_n, 0) \end{aligned} \tag{3.63}$$

Supremum Log-Likelihood over p and c for $\eta = 1$

Next we want to determine the location of the supremum of the log-likelihood function where $\eta = 1$. To determine the *LRT* under the second null-hypothesis of Case 2: Sequential

Hypothesis(η, c), the value of $\sup_{p,c} l_{mn}(p, 1, c)$ is needed. Here the log-likelihood and its derivatives to p and c are as follows, using that $X_{1j} = 0$:

$$\begin{aligned} l_{mn}(\theta) &= \sum_{i=1}^{m}\sum_{j=1}^{n} \left[-p\lambda_i - c + X_{ij}\log(p\lambda_i + c) - \log(X_{ij}!)\right] \\ &= -nc + \sum_{i=2}^{m} \left[-n\cdot(p\lambda_i + c) + X_{i.}\log(p\lambda_i + c)\right] - \sum_{i=2}^{m}\sum_{j=1}^{n}\log(X_{ij}!) \end{aligned} \tag{3.64}$$

Since $X_{1j}! = 1$ and $\log(1) = 0$.

$$\frac{\partial}{\partial p} l_{mn}(\theta) = \sum_{i=2}^{m} \lambda_i \left(\frac{X_{i.}}{p\lambda_i + c} - n\right) = n\sum_{i=2}^{m} \lambda_i \left(\frac{\bar{X}_{i.}}{p\lambda_i + c} - 1\right) \tag{3.65}$$

$$\frac{\partial}{\partial c} l_{mn}(\theta) = -n + \sum_{i=2}^{m} \left(\frac{X_{i.}}{p\lambda_i + c} - n\right) = n\left(-1 + \sum_{i=2}^{m}\left(\frac{\bar{X}_{i.}}{p\lambda_i + c} - 1\right)\right) \tag{3.66}$$

Considering that $\bar{X}_{i.}$ converges to $p\lambda_i$ in probability under the null-hypothesis, the following equations have to be solved to determine the supremum:

$$\begin{aligned} 0 &= \lim_{n\to\infty} \sum_{i=2}^{m} \lambda_i \left(\frac{\bar{X}_{i.}}{\hat{p}_n\lambda_i + \hat{c}_n} - 1\right) \overset{p}{=} \sum_{i=2}^{m} \lambda_i \left(\frac{p\lambda_i}{\hat{p}\lambda_i + \hat{c}} - 1\right) \\ 0 &= \lim_{n\to\infty} \left(-1 + \sum_{i=2}^{m}\left(\frac{\bar{X}_{i.}}{\hat{p}_n\lambda_i + \hat{c}_n} - 1\right)\right) \overset{p}{=} \left(-1 + \sum_{i=2}^{m}\left(\frac{p\lambda_i}{\hat{p}\lambda_i + \hat{c}} - 1\right)\right) \end{aligned}$$

The solution to this is, similar as before, $(\hat{p}, \hat{c}) = (p, 0)$. For this value of $(\hat{p}, \hat{c})$, the first equation equals zero, while the second is smaller than zero. Again $\hat{c} = 0$ is on the boundary of the parameter space, and thus we find that $(\hat{p}_n, \hat{c}_n) \xrightarrow{p} (p, 0)$. Now take $\tilde{p}_n$ such that $l_{mn}(\tilde{p}_n, 1, 0) = \sup_p l_{mn}(p, 1, 0)$, note that $\tilde{p}_n \xrightarrow{p} p$. Since both $(\tilde{p}_n, 1, 0)$ and $(\hat{p}_n, 1, \hat{c}_n)$ converge in probability to the same point $(p, 1, 0)$, and, similarly as before, $l_{mn}(\theta)$ is continuous, we have:

$$\begin{aligned} \lim_{n\to\infty} l_{mn}(\hat{p}_n, 1, \hat{c}_n) &\overset{p}{=} \lim_{n\to\infty} \lim_{(\hat{p}_n,\hat{c}_n)\to(p,0)} l_{mn}(\hat{p}_n, 1, \hat{c}_n) = \lim_{n\to\infty} l_{mn}(p, 1, 0) \\ &= \lim_{n\to\infty}\lim_{\tilde{p}_n\to p} l_{mn}(\tilde{p}_n, 1, 0) \overset{p}{=} \lim_{n\to\infty} l_{mn}(\tilde{p}_n, 1, 0) \end{aligned} \tag{3.67}$$

3.3.3 Distribution LRT, More Than Three Spike Concentrations

In this section, the suprema found in Section 3.3.2 will be used to determine the asymptotic distribution of LRT for the three different hypothesis cases.

Distribution LRT for Case 1: Joint Hypothesis

Now, we have that the likelihood ratio test is as follows:

$$LRT = -2\left(\sup_{p} l_{mn}(p, 1, 0) - \sup_{p,\eta,c} l_{mn}(p, \eta, c)\right)$$

Next define $LRT_{\approx}$, an approximation of LRT, as follows:

$$LRT_{\approx} := -2\left(\sup_{p} l_{mn}(p,1,0) - \sup_{p,\eta} l_{mn}(p,\eta,0)\right)$$

For $LRT_{\approx}$ we know that the distribution is χ_1^2 under the null-hypothesis. Since the $LRT_{\approx}$ is equal to the likelihood ratio test of hypothesis $H_{02} : \eta = 1$ of Case 3, where $H_{01} : c = 0$ is approved. Next we look at the difference of LRT and $LRT_{\approx}$ if $n \to \infty$:

$$\begin{aligned}
\lim_{n\to\infty}(LRT - LRT_{\approx}) &= \lim_{n\to\infty} -2\left(\sup_{p} l_{mn}(p,1,0) - \sup_{p,\eta,c} l_{mn}(p,\eta,c)\right) \\
&\quad + \lim_{n\to\infty} 2\left(\sup_{p} l_{mn}(p,1,0) - \sup_{p,\eta} l_{mn}(p,\eta,0)\right) \\
&= 2\left(\lim_{n\to\infty}\sup_{p,\eta,c} l_{mn}(p,\eta,c) - \lim_{n\to\infty}\sup_{p,\eta} l_{mn}(p,\eta,0)\right) \\
&= 2\left(\lim_{n\to\infty} l_{mn}(\hat{p}_n,\hat{\eta}_n,\hat{c}_n) - \lim_{n\to\infty} l_{mn}(\tilde{p}_n,\tilde{\eta}_n,0)\right) \\
&\overset{p}{=} 0
\end{aligned}$$

This means that LRT and $LRT_{\approx}$ converge to the same limit. Since $LRT_{\approx}$ is χ_1^2 distributed, we can conclude that the asymptotic distribution of LRT is also χ_1^2.

Distribution LRT for Case 2: Sequential Hypotheses(η, c)

For this case, the distribution of the LRT for the second hypothesis: $H_{02} : c = 0$ has to be determined. The LRT is as follows:

$$LRT_2 = -2\left(\sup_{p} l_{mn}(p,1,0) - \sup_{p,c} l_{mn}(p,1,c)\right)$$

Using the result from Equation 3.67, the following is found:

$$\begin{aligned}
\lim_{n\to\infty} LRT_2 &= \lim_{n\to\infty} -2\left(\sup_{p} l_{mn}(p,1,0) - \sup_{p,c} l_{mn}(p,1,c)\right) \\
&= -2\left(\lim_{n\to\infty}\sup_{p} l_{mn}(p,1,0) - \lim_{n\to\infty}\sup_{p,c} l_{mn}(p,1,c)\right) \\
&\overset{p}{=} -2\left(\lim_{n\to\infty}\sup_{p} l_{mn}(p,1,0) - \lim_{n\to\infty}\sup_{p} l_{mn}(p,1,0)\right) = 0
\end{aligned}$$

From this we can see that the likelihood ratio test of the second hypothesis H_{02} converges in probability to 0. This means that LRT_2 is deterministic, and if the value of LRT_2 is unequal to 0 then the null-hypothesis $H_{02} : c = 0$ is rejected.

Distribution LRT for Case 3: Sequential Hypotheses(c, η)

Similarly to Case 1: joint hypothesis $H_0 : \eta = 1 \wedge c = 0$, the distribution of the LRT under null-hypothesis for the first hypothesis of Case 3, $H_{01} : c = 0$, can be determined. We know that the LRT for this hypothesis is as follows:

$$LRT_1 = -2\left(\sup_{p,\eta} l_{mn}(p,\eta,0) - \sup_{p,\eta,c} l_{mn}(p,\eta,c)\right)$$

For this, we know, the same as for Case 1, that $\lim_{n\to\infty}\sup_{p,\eta,c} l_{mn}(p,\eta,c) \overset{p}{=} \lim_{n\to\infty}\sup_{p,\eta} l_{mn}(p,\eta,0)$ from Equation 3.63. From this it follows that:

$$\begin{aligned}
\lim_{n\to\infty} LRT_1 &= \lim_{n\to\infty} -2\left(\sup_{p,\eta} l_{mn}(p,\eta,0) - \sup_{p,\eta,c} l_{mn}(p,\eta,c)\right) \\
&= -2\left(\lim_{n\to\infty}\sup_{p,\eta} l_{mn}(p,\eta,0) - \lim_{n\to\infty}\sup_{p,\eta,c} l_{mn}(p,\eta,c)\right) \\
&\overset{p}{=} -2\left(\lim_{n\to\infty}\sup_{p,\eta} l_{mn}(p,\eta,0) - \lim_{n\to\infty}\sup_{p,\eta} l_{mn}(p,\eta,0)\right) = 0
\end{aligned}$$

This means that the distribution of LRT_1 under null-hypothesis H_{01} converges to the value 0 in probability, i.e. LRT_1 is deterministic. If the value of the LRT_1 is unequal to 0, then the null-hypothesis can be rejected.

Chapter 4

Simulation

In this chapter, we will describe the simulation and the found results. The model was simulated in R and the code can be found in Appendix B.

4.1 Simulation Description

In this section, the simulation is described. There are two different reasons as to why a simulation is made. First, the simulation can be used to investigate the performance of the found asymptotic distribution of the LRT to generated data with low number of samples per spike concentration. Secondly, we want to determine there is an advantage to using a certain hypothesis case, or to use a certain chosen set of spike concentrations. The power and type-I error of models with different hypotheses cases and spike concentration sets will be determined for a different number of samples per spike concentration. Once determined, the power can be used to determine the hypothesis and spike concentrations that perform the best. The type-I error rate can be used to check the performance of the found distribution.
In the simulation, several scenarios will be looked at, to determine an overlook on the performance of the model. For each scenario, the type-I error rate and the power will be determined for a specific set of parameters, which can be seen in Table 4.2. Each scenario that is simulated will consist of two parts: one of the hypothesis cases of Section 2.3 and one set of spike concentrations. In Table 4.1 the values for the different spike concentration sets can be seen. They are chosen, such that there are sets with blank concentration and with non-zero concentrations. Set 1 (S_1) is chosen following the values that were found in Heidari (2020). They determined the values for the spike concentrations that minimize the variances of the parameter estimates. The values found where $\lambda_1 = 0$, $\lambda_3 = \lambda_{max}$ and $\lambda_2 \in \{\lambda_3 e^{-2/\eta}, \lambda_3 e^{-1/\eta}\}$. If you take $\lambda_{max} = 100$ and $\eta = 1$ then $\lambda_2 \in \{13.5, 36.8\}$. To be able to compare including a blank solution to no blank solution, Set 3 is only different in λ_1 as compared to Set 1. For Set 2, an equal distribution of the spike concentrations is chosen. Set 4 and 5 have more than 3 spike concentrations. Set 4 is an extension of Set 1 and Set 5 follows a logarithmic scale. To simulate the data, first the observations x_{ij} are created using the Poisson distribution with the parameter $p\lambda_i^{\eta} + c$, where p, η, c and n are values from the Table 4.2 and m is the size of the spike concentration set used. Using this simulated data, the values of the likelihood ratio test statistic is determined for the given hypothesis cases, which in turn can be compared to the found distribution to determine if the null-hypothesis is rejected or not. This will be repeated $N = 10.000$ times, to determine the probability that the null-hypothesis will be rejected, given

Set	Spike Concentrations			
	λ_1	λ_2	λ_3	λ_4
S_1	0	13.5	100	-
S_2	0	50	100	-
S_3	1	13.5	100	-
S_4	0	13.5	25	100
S_5	0.1	1	10	100

Table 4.1: Spike concentration sets used in the simulation

p	0.7	0.9						
η	0.9	1	1.1					
c	0	0.5						
n	3	5	8	10	15	20	25	50

Table 4.2: Parameters used in Simulation

p, η, c and n. If $\eta = 1$ and $c = 0$, the probability of rejection is the type-I error rate, otherwise it is the power.

To determine the value of the LRT given the simulated data for all three hypothesis cases, the supremum of the log-likelihood needs to be determined for four different situations. First with p, η and c free in the parameter space, then with either $\eta = 1$ or $c = 0$, and lastly when both $\eta = 1$ and $c = 0$. Once the suprema are found the different LRT's needed for each hypothesis case can be calculated using the found suprema, using the equations given in Section 2.3. Next we need to determine if the null-hypothesis will be rejected or not. The criteria for rejection of each of the hypothesis cases, with or without blank spike, can be seen in Table 4.3. For the sequential hypothesis, proportionality is rejected if either of the two hypotheses is rejected. The code for the simulation can be seen in Appendix B.

	Case		Found Distribution	Rejected If
non-zero spike	1: joint	LRT	$\frac{1}{2}\chi_1^2 + \frac{1}{2}\chi_2^2$	$\frac{1}{2}\mathbb{P}(\chi_1^2 > LRT) + \frac{1}{2}\mathbb{P}(\chi_2^2 > LRT) < \alpha$
	2: (η, c)	LRT_1	χ_1^2	$\mathbb{P}(\chi_1^2 > LRT_1) < \frac{1}{2}\alpha$
		LRT_2	$\frac{1}{2}\chi_1^2$	$\frac{1}{2}\mathbb{P}(\chi_1^2 > LRT_2) < \frac{1}{2}\alpha$
	3: (c, η)	LRT_1	$\frac{1}{2}\chi_1^2$	$\frac{1}{2}\mathbb{P}(\chi_1^2 > LRT_1) < \frac{1}{2}\alpha$
		LRT_2	χ_1^2	$\mathbb{P}(\chi_1^2 > LRT_2) < \frac{1}{2}\alpha$
blank spike	1: joint	LRT	χ_1^2	$\mathbb{P}(\chi_1^2 > LRT) < \alpha$
	2: (η, c)	LRT_1	χ_1^2	$\mathbb{P}(\chi_1^2 > LRT_1) < \frac{1}{2}\alpha$
		LRT_2	0	$LRT_2 \neq 0$
	3: (c, η)	LRT_1	0	$LRT_1 \neq 0$
		LRT_2	χ_1^2	$\mathbb{P}(\chi_1^2 > LRT_2) < \frac{1}{2}\alpha$

Table 4.3: Distribution of LRT for each hypothesis case and rejection criteria

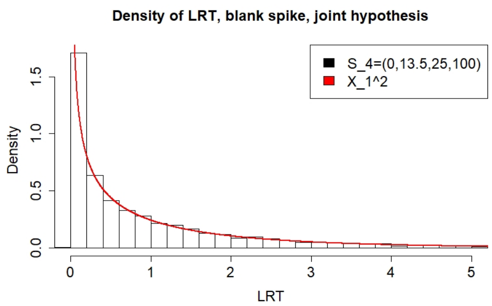
4.2 Results

First we will look at the fit of the found distributions. In the first subsection we will investigate the performance of the found distributions, through use of the histograms of the LRT for each of the hypothesis case, with or without blank spike, as well as the type-I error rate. Then we will look at the statistical power for each scenario explained in the last section, using the rejection criteria as discussed above. In the first sub-section we find that the rejection criteria might need to be adjusted, the results of the adjusted simulation on the power can be seen in the last subsection.

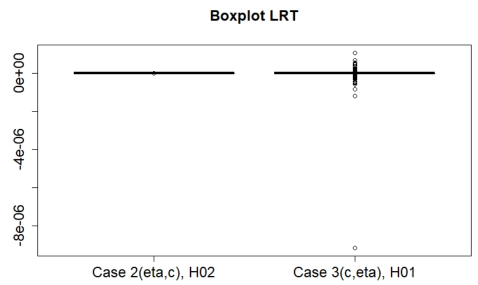
4.2.1 Fit Found Distributions

First the histograms of N simulated LRT values of the three different hypothesis cases will be shown, along with the density line of the corresponding determined asymptotic distribution. For the spike concentrations, S_4 and S_5 are used for the scenario with blank spike and non-zero spike respectively. The simulation here uses ten samples per spike concentration ($n = 10$).

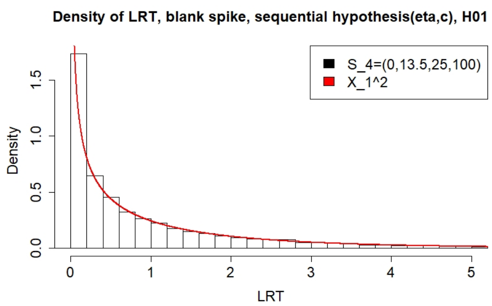
In Figure 4.1 the distribution of the LRT, with a blank spike included, can be seen. In



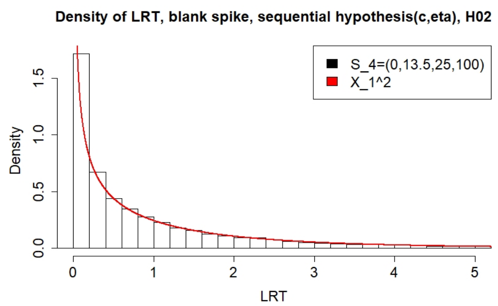
(a) Case 1, Joint hypothesis



(b) Case 2 and 3, Sequential hypotheses



(c) Case 2, Sequential hypotheses(η, c), H_{01}



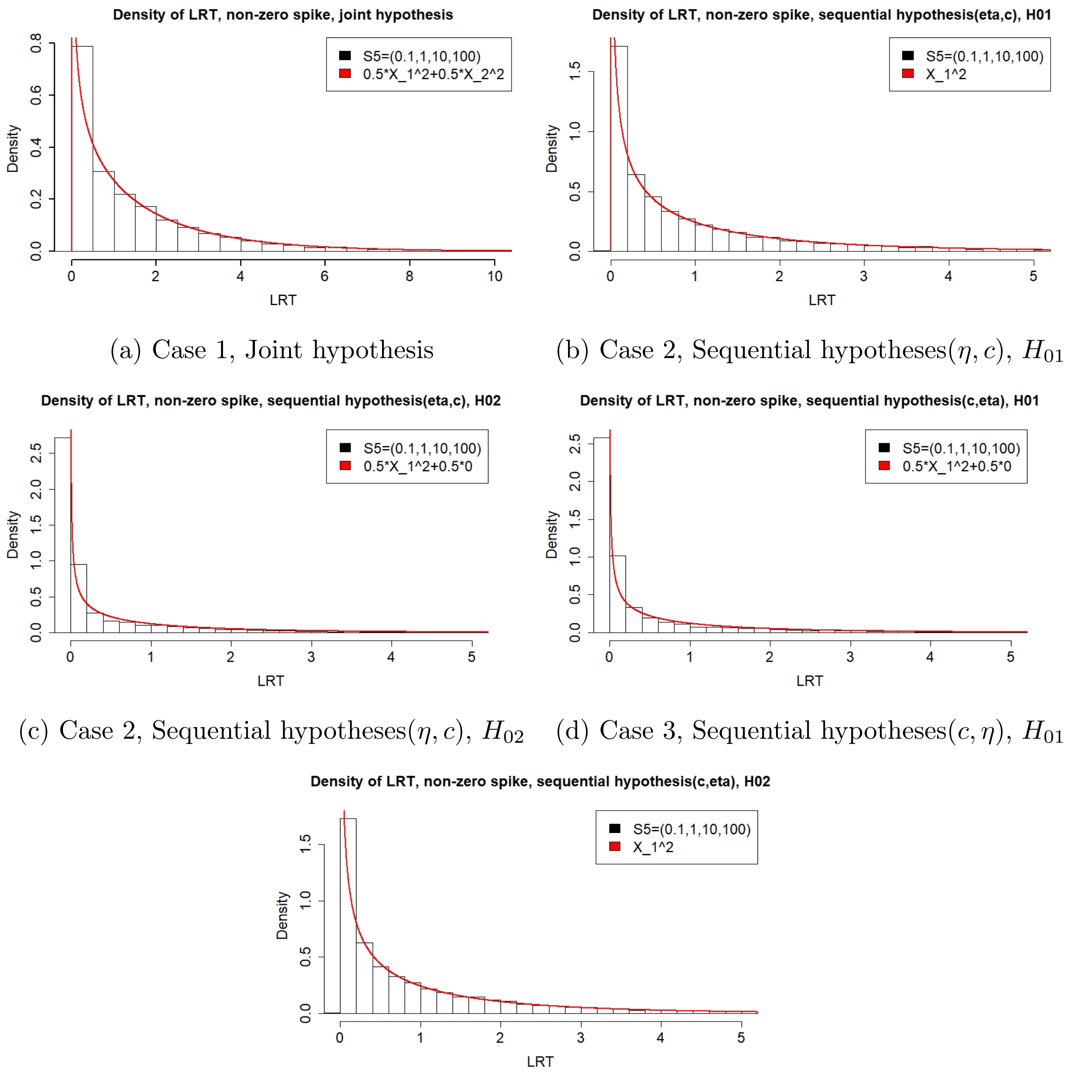
(d) Case 2, Sequential hypotheses(c, η), H_{02}

Figure 4.1: Distribution of the LRT with spike concentrations S_4, icl. blank spike

the first graph (Figure 4.1a), the density is shown when using joint hypothesis, which is very similar to the found distribution χ^2_1. Figure 4.1b shows the box-plots of the LRT for Case 2 and 3, the sequential hypotheses cases. Here you can see that the LRT sometimes, very rarely, deviates from 0. The size of this deviation is smaller than ≈E-08. This difference is most likely because of the small error in calculating the LRT in R, for $p\lambda_i^\eta + c < 1E - 08$. The function sc optim in R used to find the suprema needed for the LRT, cannot be used

if the function it optimizes over has infinite function values in the defined parameter space. If $p\lambda_i^\eta + c$ comes close to zero, $\log(p\lambda_i^\eta + c)$ will move to infinity, which causes problems in OPTIM. These problems is why for small $p\lambda_i^\eta + c$, this value is bounded in the code. This does not cause problems in the values of the estimates, which is why the estimation of $\hat{c}$ can be used in the simulation to determine whether the null-hypothesis of $c = 0$ is accepted for the sequential hypotheses cases. Figures 4.1c and 4.1d show the distribution of the LRT of the sequential hypothesis test for $\eta = 1$, for both cases.
The density of the LRT with non-zero spike concentrations can be seen in Figure 4.2. Similar as with the LRT of blank spikes, the density of the LRT for Case 1 fits the found distribution the mixture of χ_1^2 and χ_2^2 both with ratio $\frac{1}{2}$. Similarly for the sequential hypotheses cases, the distribution found in Section 3.2 seems to fit the simulated data well.

Next we will look at the type-I error rate for the parameters shown in Table 4.2. In Figure



(a) Case 1, Joint hypothesis

(b) Case 2, Sequential hypotheses(η, c), H_{01}

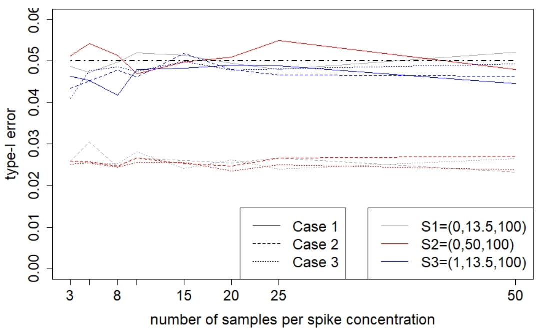
(c) Case 2, Sequential hypotheses(η, c), H_{02}

(d) Case 3, Sequential hypotheses(c, η), H_{01}

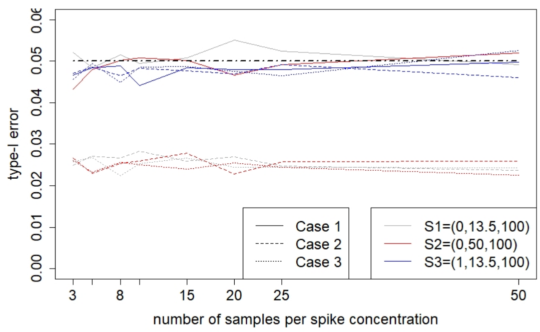
(e) Case 3, Sequential hypotheses(c, η), H_{02}

Figure 4.2: Distribution of the LRT with spike concentrations S_5, non-zero spike

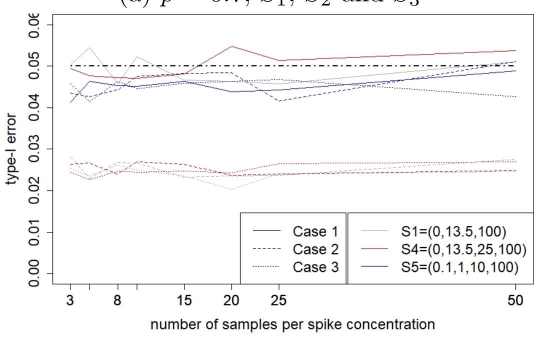
4.3, the type-I error rate is shown for the first simulation. For the type-I error rate, you can see that most of the lines are around the alpha-value, 0.05. Only the lines when testing using the sequential hypotheses, Case 2 and 3, and the spike concentrations set with the blank spike the type-I error rate is around 0.025. In these two cases the significance level for the two tests is set at $\alpha/2 = 0.025$, while one of the sequential test, the test for $c = 0$, is independent of the α value. In this test, in the simulation, we test if $\hat{c} = 0$, which equals the test $LRT = 0$, as can be seen in Section 3.3.3. Testing if $LRT = 0$ is a degenerative test, since it is equal to testing $LRT = D$, with D a distribution where $\mathbb{P}(D = 0) = 1$. As we know that $LRT \overset{p}{=} 0$ if $\eta = 1$ and $c = 0$, as shown in Section 3.3.3, the overall type-I error rate is only $\alpha/2$. This creates a conservative test, and it might be interesting to look at the simulation where the significance level is adjusted. Instead of using $\alpha/2$ as the significance level for testing $\eta = 1$, α will be used for this test, when testing sequentially with a blank spike.
Lastly, one more thing to notice, in the type-I error rate graphs for the first simulation, is that in all four graphs the blue lines, for the spike concentration sets excluding the blank spike (S_3 and S_5), are slightly below the α line. The lines for the spike concentration set S_1 are the closest to the significance line, if you look at all three lines.



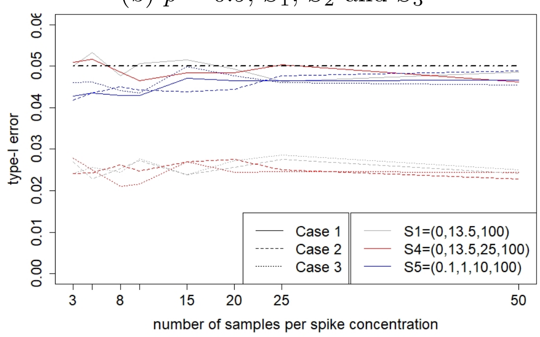
(a) $p = 0.7$, S_1, S_2 and S_3



(b) $p = 0.9$, S_1, S_2 and S_3



(c) $p = 0.7$, S_1, S_4 and S_5



(d) $p = 0.9$, S_1, S_4 and S_5

Figure 4.3: Type-I error rate, for $p \in \{0.7, 0.9\}$, for all spike-concentration sets

In Figure 4.4, the type-I error rate for the adjusted simulation can be seen. As there does not seem to be much difference between $p = 0.7$ and $p = 0.9$, only $p = 0.7$ is used here. As the adjustment only influence the scenarios with a blank spike and sequential hypotheses, we look at spike concentration sets with a blank spike. The joint hypothesis test is included for comparison to the first simulation. The type-I error rate for the adjusted simulation is all around $\alpha = 0.05$, this means that the adjustment of the significance level solved the conservative nature seen before the adjustment of the sequential tests with a blank spike.

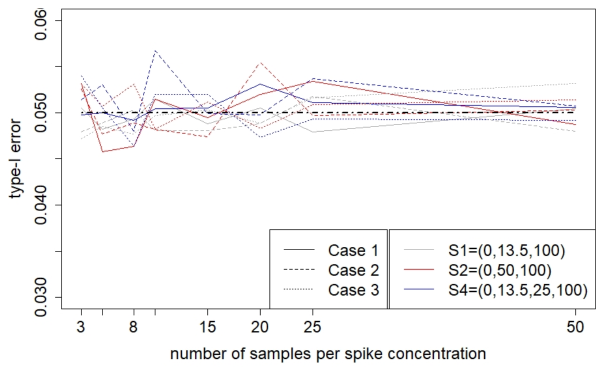
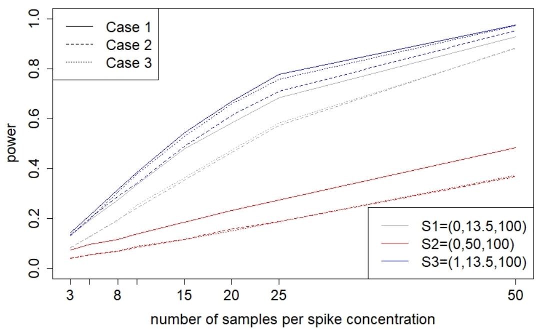


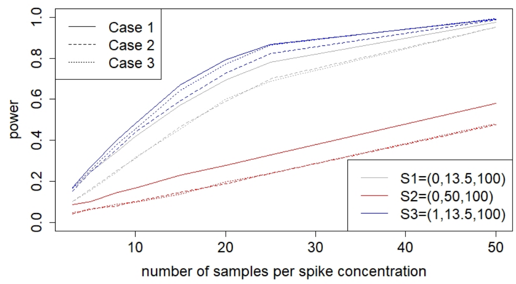
Figure 4.4: Type-I error rate, for $p = 0.7$, spike concentrations S_1, S_2 and S_4

4.2.2 First Simulation

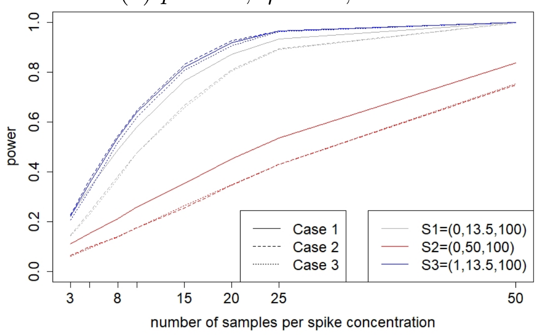
Here we will discuss the statistical power found for the first simulation. In Figures 4.5 to 4.8, the power of the different test situations can be seen. In Figures 4.5 and 4.6 are the power for when c is taken equal to 0. Here the first figure has spike concentration sets S_1, S_2 and S_3, and the second figure shows the results for sets S_1, S_4 and S_5. In the last two figures, Figures 4.7 and 4.8, the power found for simulating with $c = 0.5$ is shown.
First we will take a look at the figures where $c = 0$ (Figures 4.5 and 4.6). Here you can see that if you look per spike concentration set, Case 1, joint hypothesis, has the most power compare to Case 2 and 3, while the sequential hypothesis cases, Case 2 and 3, do not show much difference. Note that this holds only for the spike concentration sets with a blank spike, for the spike concentration sets with non-negative spikes there does not seem to be much difference between the hypothesis cases. If you look at the different spike concentration sets, among the first three, S_1, S_2 and S_3, S_2 performs the worst. Meanwhile S_1 and S_3 are very similar with S_3 performing slightly better three out of four times. For the last three S_4 has to most power, while S_1 has the least.
Now if we look at the results of the simulations where $c = 0.5$, the first thing to notice is that the spike concentration sets with non-zero spikes (S_3 and S_5) have less power compared to the other spike concentration sets (in their respective graphs). In both graphs there is not much difference between the two other spike concentration sets. In these graphs these graphs there is not one of the three cases with clearly more power than another. In most of the graphs either Case 2 or Case 3 hold the most power, but the three lines never seem to differ much from one another.
From the results found in this section, the best spike concentration set, with 3 different concentrations, is $S_1 = (0, 13.5, 100)$. Adding one more spike concentration with value 25 to this gives S_4, which has overall the most power. Considering Case 1, joint hypothesis, has the most power if $c = 0$, while there is not much difference between the cases if $c = 0.5$, we determine Case 1, joint hypothesis, is the best option.



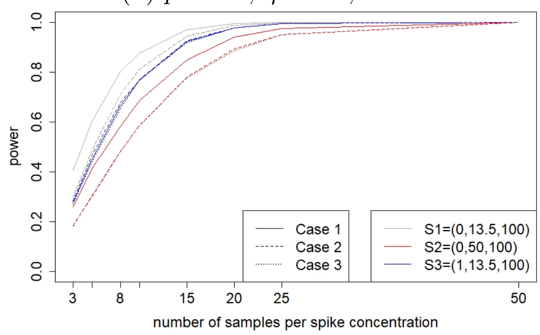
(a) $p = 0.7$, $\eta = 0.9$, $c = 0$



(b) $p = 0.9$, $\eta = 0.9$, $c = 0$



(c) $p = 0.7$, $\eta = 1.1$, $c = 0$

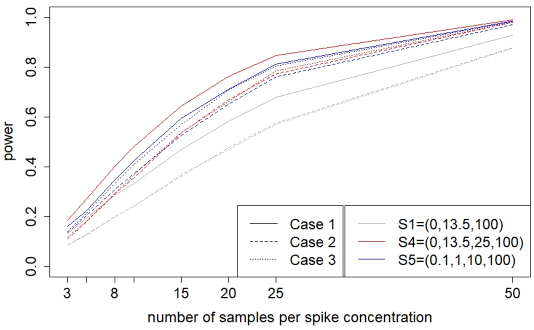


(d) $p = 0.9$, $\eta = 1.1$, $c = 0$

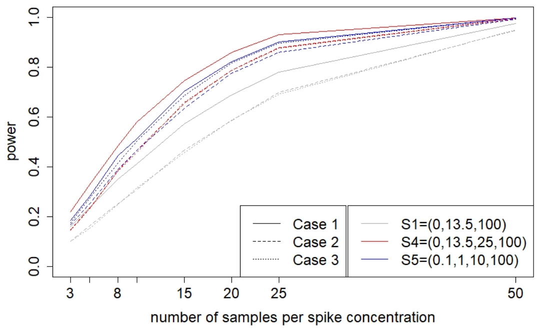
Figure 4.5: Power, for different values of p, η, and $c = 0$, spike concentrations S_1, S_2 and S_3

4.2.3 Simulation with Adjusted Significance Levels

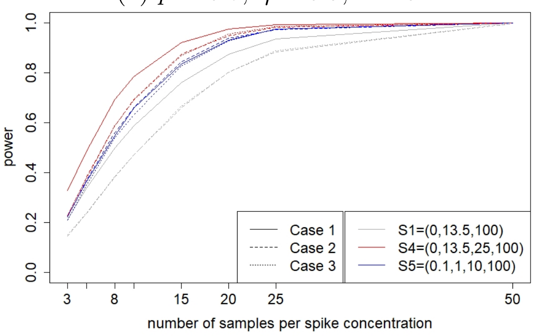
In this section the values for the significance level for the sequential tests, with one blank spike will be adjusted. For the adjustment, instead of dividing the significance level over both test, when testing sequentially, all of the significance level will be used for the test of $\eta = 1$, since the test for $c = 0$ is degenerative. As this difference only applies when one of the spike concentrations is zero, and we test sequentially, only spike concentrations S_1, S_2, and S_3 will be tested in this section. We will test all three Hypothesis cases, such that we are able to compare to the first simulation. We also noticed that there is not much difference between the two different values of p that were used in the first simulation, and we decided to only use $p = 0.7$ in this simulation.
In Figures 4.9 the result from the simulation can be seen. The difference in the power, as compared to the first simulation, seems to be that there is no longer a difference between the hypothesis cases. When comparing the difference spike concentration sets, the results are the same as in Section4.2.2. The spike concentration set with the most power is S_4, followed by S_1. The only difference between these two sets is that set S_4 also has spike concentration 25. From the results in this section, we found that, when including a blank spike in the spike concentrations used, the significance level needs to be adjusted for the sequential cases to make them perform the same as the joint hypothesis case. Once adjusted, it does not matter which hypothesis case is used. Depending on the goal and the preference to either log-linearity or linearity, the different cases can be used. For the spike concentration, we found that the result is the same as found in the previous section, S_4 has the most power, but if you need or want to limit yourself to 3 spike concentrations, S_1 should be chosen.



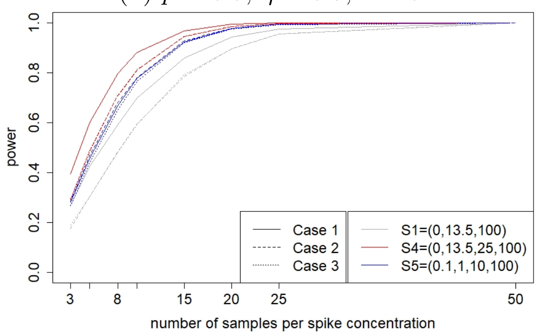
(a) $p = 0.7$, $\eta = 0.9$, $c = 0$



(b) $p = 0.9$, $\eta = 0.9$, $c = 0$

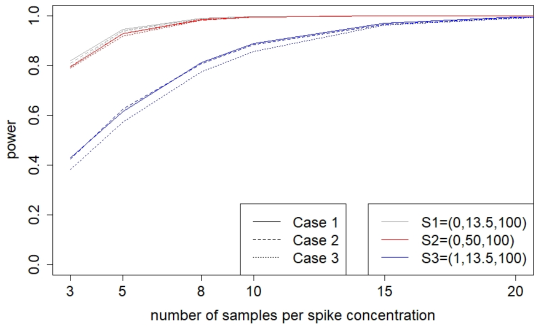


(c) $p = 0.7$, $\eta = 1.1$, $c = 0$

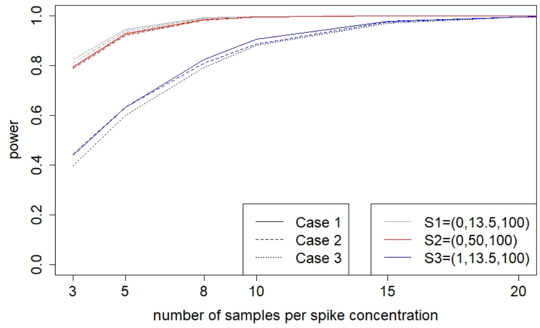


(d) $p = 0.9$, $\eta = 1.1$, $c = 0$

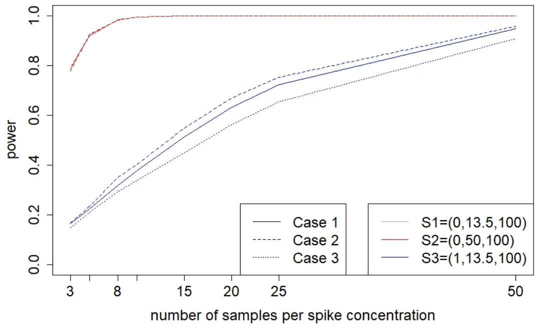
Figure 4.6: Power, for different values of p, η, and $c = 0$, spike concentrations S_1, S_4 and S_5



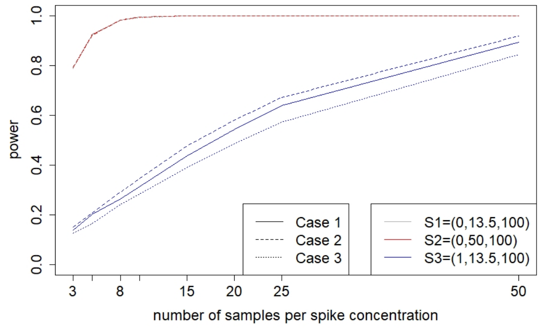
(a) $p = 0.7$, $\eta = 0.9$, $c = 0.5$



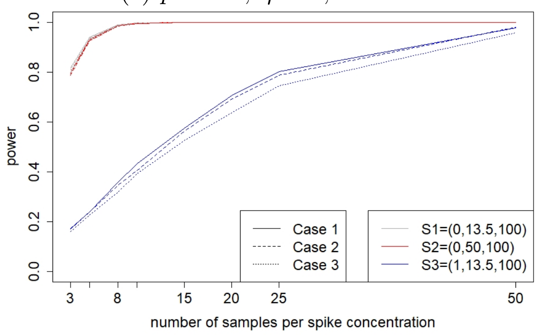
(b) $p = 0.9$, $\eta = 0.9$, $c = 0.5$



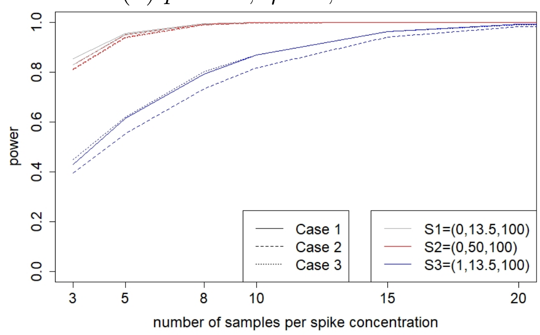
(c) $p = 0.7$, $\eta = 1$, $c = 0.5$



(d) $p = 0.9$, $\eta = 1$, $c = 0.5$

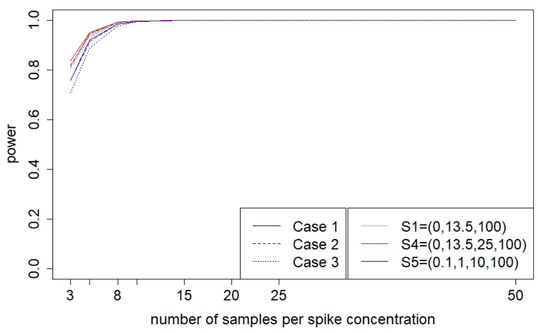


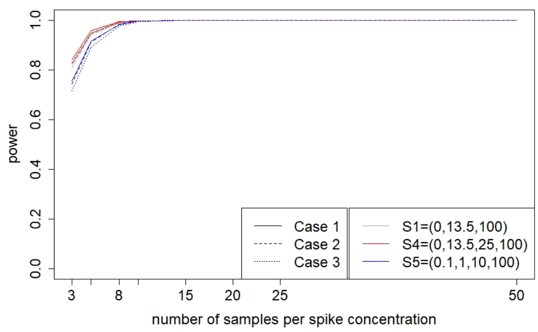
(e) $p = 0.7$, $\eta = 1.1$, $c = 0.5$



(f) $p = 0.9$, $\eta = 1.1$, $c = 0.5$

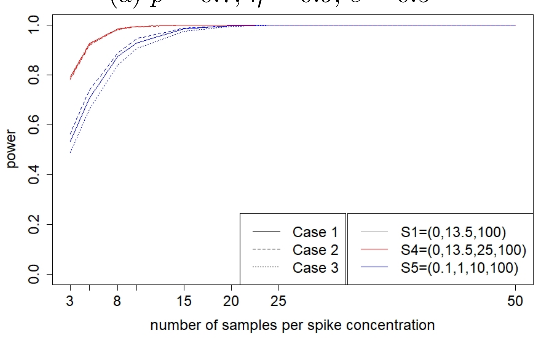
Figure 4.7: Power, for different values of p, η, and $c = 0.5$, spike concentrations S_1, S_2 and S_3

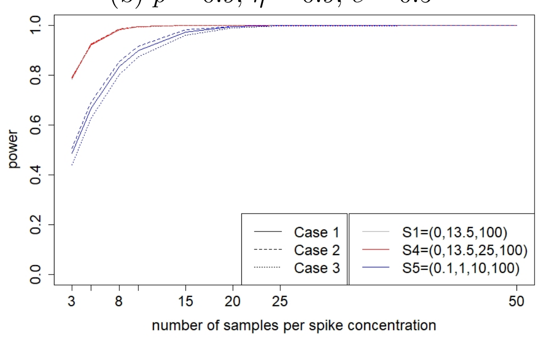




(a) $p = 0.7$, $\eta = 0.9$, $c = 0.5$

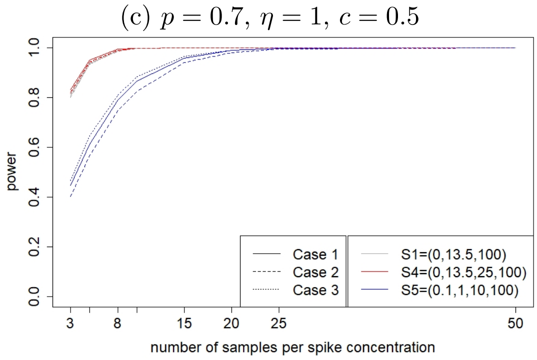
(b) $p = 0.9$, $\eta = 0.9$, $c = 0.5$

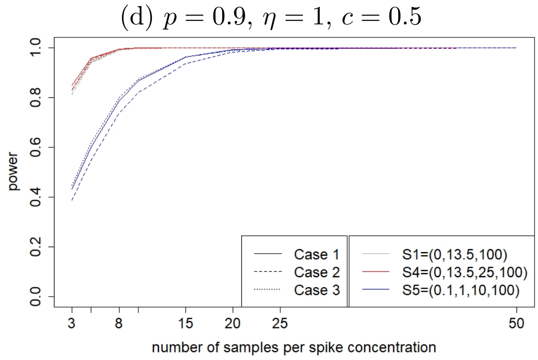




(c) $p = 0.7$, $\eta = 1$, $c = 0.5$

(d) $p = 0.9$, $\eta = 1$, $c = 0.5$

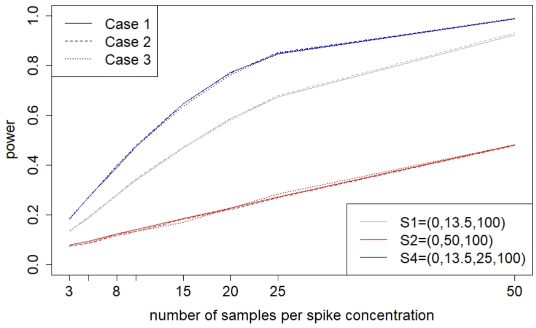




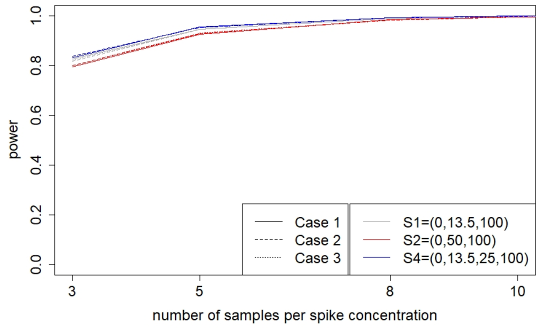
(e) $p = 0.7$, $\eta = 1.1$, $c = 0.5$

(f) $p = 0.9$, $\eta = 1.1$, $c = 0.5$

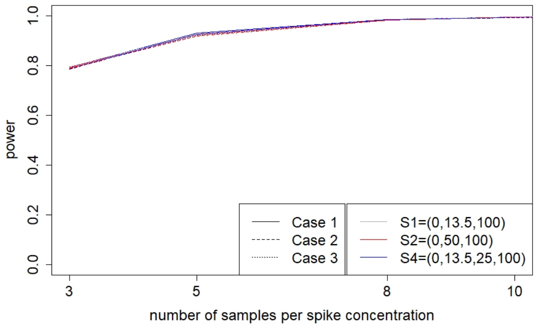
Figure 4.8: Power, for different values of p, η, and $c = 0.5$, spike concentrations S_1, S_4 and S_5



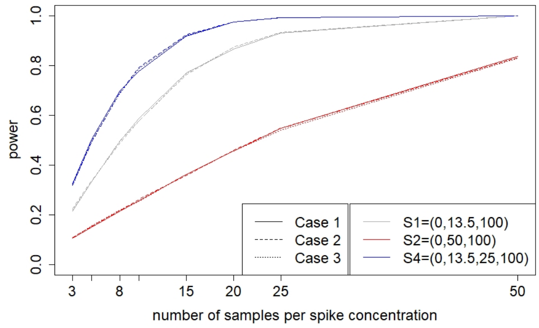
(a) $\eta = 0.9$, $c = 0$



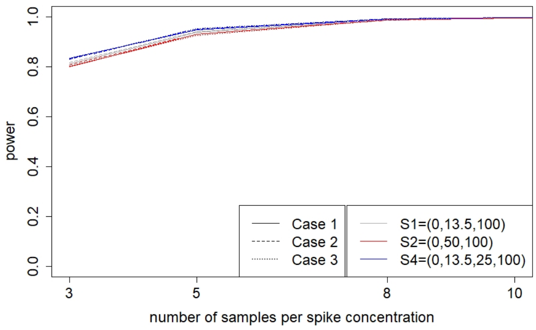
(b) $\eta = 0.9$, $c = 0.5$



(c) $\eta = 1$, $c = 0.5$



(d) $\eta = 1.1$, $c = 0$



(e) $\eta = 1.1$, $c = 0.5$

Figure 4.9: Power, for different values of η, and c, for $p = 0.7$, spike concentrations S_1, S_2 and S_4

Chapter 5

Conclusion

In this thesis, we determined the distribution of the LRT, when a blank spike is either included or excluded. This was needed for the model used in this thesis, since on of the parameters was on the boundary of the parameter space. We assumed that the data is poisson distributed with Mitscherlich mean $(p\lambda_i^\eta + c)$. The Mitscherlich function is linear and log-linear under specific conditions, namely $\eta = 1$ and $c = 0$. The distribution of the LRT has been determined for when both linearity and log-linearity are tested at once, and for when one is tested before the other. For the joint hypothesis testing, the distribution of the LRT with a blank is χ_1^2, while the distribution for the LRT with non-zero spikes is $\frac{1}{2}\chi_1^2 + \frac{1}{2}\chi_2^2$. For sequential hypotheses testing, when testing if $\eta = 1$, the distribution of the LRT is χ_1^2. The LRT of the test for $c = 0$ is $\frac{1}{2}\chi_1^2$ if the spike concentrations are non-zero. When a blank spike is included the test becomes deterministic, and the null-hypothesis will be rejected if LRT$\neq 0$. The distribution of the LRT when no blank was used, was determined using a theorem from Self & Liang (1987). For this several regularity conditions were proven.
After the distribution of the LRT was determined, the model was simulated to find the best option for null-hypotheses and spike concentrations. From this was found that the joint hypothesis test was better than either of the sequential tests, but only if, for the sequential test with blank spike, the significance level was equally divided over the two test. If the significance level would be adjusted for the sequential test, if a blank spike is included, by only using it for the η test, there is no difference between the different hypothesis cases. For the spike concentrations was found that taking the spike concentrations as $(\lambda_1, \lambda_2, \lambda_3) = (0, 13.5, 100)$ was the best option if you only wanted the minimum number of spike concentrations. If one more concentration can be added, adding the spike concentration 25 to the above mentioned spike concentrations would be the best option. This set of spike concentrations was the best set among the tested sets for both the first and adjusted simulation.

Bibliography

Billingsley, P. (1995). *Probability and measure* (3rd ed.). New York: John Wiley & Sons, Inc.

Box, G. E. P. & Lucas, H. L. (1959). Design of Experiments in Non-Linear Situations. *Biometrika*, *46*(1), 77–90.

Council of Europe (2017). 5.1.6. Alternative Methods for Control of Microbiological Quality. In *European Pharmacopoeia* (9.2. ed.). (pp. 4339–4348). Strasbourg: EDQM.

Heidari, M. (2020). D-optimal design for the "mitscherlich" regression function: Assessing linearity in measurement systems. unpublished.

Morgan, C. A., Bigeni, P., Herman, N., Gauci, M., White, P., & Vesey, G. (2005). Production of precise microbiology standarts using flow cytometry and freeze drying.

Niermann, S. (2007). Testing for linearity in simple regression models. *AStA Advances in Statistical Analysis*, *91*(2), 129–139.

Self, S. G. & Liang, K. Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, *82*(398), 605–610.

USP-NF (2015). <1223> validation of alternative microbiological methods. In *United States Pharmacopeia and National Formulary* (UPS40-NF35 ed.). Rockville, MD: United States Pharmacopeial Convention.

Wilks, S. S. (1938). The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. *The Annals of Mathematical Statistics*, *9*(1), 60–62.

Xie, J., Nie, Y., & Liu, X. (2017). Testing the proportionality condition with taxi trajectory data. *Transportation Research Part B: Methodological*, *104*, 583–601.

Appendix A

Existence of derivatives of $\log(p\lambda_i^\eta + c)$

Here we will show that the derivatives of $f(p, \eta, c) = \log(p\lambda_i^\eta + c)$ exist if $p\lambda_i^\eta + c > 0$. For this, note that the derivative of a function $f(x_1, x_2, x_3)$ exists if the partial derivatives $\frac{\partial f}{\partial x_i}$ exist for all i. These partial derivatives exist if the following limit exists:

$$\lim_{h \to 0} \frac{f(x + h \cdot e_i) - f(x)}{h}$$

Here $x = (x_1, x_2, x_3)^T$ and e_i is the i-th standard basis vector, thus $e_1 = (1, 0, 0)^T$. Note that $\lim_{h \to 0}(1 + hx)^{1/h} = e^x$, also note that $\lambda_i^h - 1 = h\log(\lambda_i) + O(h^2)$ is the Taylor expansion of $\lambda_i^h - 1$ around $h = 0$. Then we have the following:

$$\begin{aligned}
\lim_{h \to 0} \frac{f(p + h, \eta, c) - f(p, \eta, c)}{h} &= \lim_{h \to 0} \frac{\log((p + h)\lambda_i^\eta + c) - \log(p\lambda_i^\eta + c)}{h} \\
&= \lim_{h \to 0} \frac{1}{h} \log\left(\frac{(p + h)\lambda_i^\eta + c}{p\lambda_i^\eta + c}\right) = \lim_{h \to 0} \frac{1}{h} \log\left(1 + \frac{h\lambda_i^\eta}{p\lambda_i^\eta + c}\right) \\
&= \lim_{h \to 0} \log\left(\left(1 + \frac{h\lambda_i^\eta}{p\lambda_i^\eta + c}\right)^{1/h}\right) = \log\left(e^{\lambda_i^\eta/(p\lambda_i^\eta + c)}\right) \\
&= \frac{\lambda_i^\eta}{p\lambda_i^\eta + c}
\end{aligned}$$

$$\begin{aligned}
\lim_{h \to 0} \frac{f(p, \eta + h, c) - f(p, \eta, c)}{h} &= \lim_{h \to 0} \frac{\log(p\lambda_i^{\eta + h} + c) - \log(p\lambda_i^\eta + c)}{h} = \lim_{h \to 0} \frac{1}{h} \log\left(\frac{p\lambda_i^{\eta + h} + c}{p\lambda_i^\eta + c}\right) \\
&= \lim_{h \to 0} \frac{1}{h} \log\left(1 + \frac{p\lambda_i^\eta(\lambda_i^h - 1)}{p\lambda_i^\eta + c}\right) \\
&= \lim_{h \to 0} \log\left(\left(1 + \frac{p\lambda_i^\eta(h\log(\lambda_i) + O(h^2))}{p\lambda_i^\eta + c}\right)^{1/h}\right) \\
&= \log\left(e^{p\lambda_i^\eta \log(\lambda_i)/(p\lambda_i^\eta + c)}\right) = \frac{p\lambda_i^\eta \log(\lambda_i)}{p\lambda_i^\eta + c}
\end{aligned}$$

$$
\begin{aligned}
\lim_{h\to 0}\frac{f(p,\eta,c+h)-f(p,\eta,c)}{h} &= \lim_{h\to 0}\frac{\log(p\lambda_i^\eta+c+h)-\log(p\lambda_i^\eta+c)}{h}\\
&= \lim_{h\to 0}\frac{1}{h}\log\left(\frac{p\lambda_i^\eta+c+h}{p\lambda_i^\eta+c}\right) = \lim_{h\to 0}\frac{1}{h}\log\left(1+\frac{h}{p\lambda_i^\eta+c}\right)\\
&= \lim_{h\to 0}\log\left(\left(1+\frac{h}{p\lambda_i^\eta+c}\right)^{1/h}\right) = \log\left(e^{1/(p\lambda_i^\eta+c)}\right)\\
&= \frac{1}{p\lambda_i^\eta+c}
\end{aligned}
$$

This shows that the partial derivatives $\frac{\partial}{\partial p}\log(p\lambda_i^\eta+c)$, $\frac{\partial}{\partial \eta}\log(p\lambda_i^\eta+c)$ and $\frac{\partial}{\partial c}\log(p\lambda_i^\eta+c)$ exists if $p\lambda_i^\eta+c>0$, which means the derivatives of $\log(p\lambda_i^\eta+c)$ exists if $p\lambda_i^\eta+c>0$.

Appendix B

Simulation Code

```
ll<-function(x,l,p,eta,c){ #loglikelhood for 1 lambda
  #x observations for spikeconcentration l, p, eta, c parameters
  par<-ifelse(p*l^eta+c<=1E-08,(1e-08),p*l^eta+c)
  par_log<-log(par)
  return(-length(x)*par+sum(x*par_log-lfactorial(x)))
}

loglike_pnc<-function(theta,x, lambda){
  #get l(theta) for observations x, spikes lambda, where theta=(p,eta,c)
  p<-theta[1]
  eta<-theta[2]
  c<-theta[3]
  y<-rep(NA,length(lambda))

  for(i in 1:length(lambda)){
    y[i]<-ll(x[,i],lambda[i],p,eta,c)
  }
  return(sum(y))
}

loglike_pn<-function(theta,x, lambda){
  #get l(p,eta,0) for observations x, spikes lambda, where theta=(p,eta), c=0
  p<-theta[1]
  eta<-theta[2]
  y<-rep(NA,length(lambda))

  for(i in 1:length(lambda)){
    y[i]<-ll(x[,i],lambda[i],p,eta,0)
  }
  return(sum(y))
}

loglike_pc<-function(theta,x, lambda){
  #get l(p,1,c) for observations x, spikes lambda, where theta=(p,c), eta=1
  p<-theta[1]
  c<-theta[2]

  y<-rep(NA,length(lambda))
  for(i in 1:length(lambda)){
    y[i]<-ll(x[,i],lambda[i],p,1,c)
  }
```

```r
  return(sum(y))
}

loglike_p<-function(p,x,lambda){
  # l(p,1,0) for observations x, spikes lambda
  y<-rep(NA,length(lambda))

  for(i in 1:length(lambda)){
    y[i]<-ll(x[,i],lambda[i],p,1,0)
  }
  return(sum(y))
}
```

Listing B.1: Likelihood and Loglikelihood function

```r
run<-function(p,eta,c,lambda, m,n, h0_type="both"){#single run model
  alpha<-0.05

  # create observations
  x<-c()
  for(l in lambda){
    x<-cbind(x,rpois(n,p*l^eta+c))
  }


  #find optimum over p eta c
  opt_pnc<-optim(par=c(p,eta,c),fn=loglike_pnc,x=x,lambda=lambda,
                 lower=c(0,0,0),upper=c(Inf,Inf,Inf),
                 method="L-BFGS-B",control = list(maxit=200000,fnscale=-1),
                 hessian=TRUE)
  # find optimum over p if eta=1,c=0
  opt_p<-optim(par=c(p),fn=loglike_p,x=x,lambda=lambda,
               lower=0,upper=Inf, method="L-BFGS-B",
               control = list(maxit=200000,fnscale=-1),hessian=TRUE)


  #sequential or joint hypothesis
  if(h0_type=="ceta"){
    # h01:c=0, h02: eta=1

    # find optimum over p,eta if c=0
    opt_pn<-optim(par=c(p,eta),fn=loglike_pn,x=x,lambda=lambda,
                  lower=c(0,0),upper=c(Inf,Inf),
                  method="L-BFGS-B",control = list(maxit=200000,fnscale=-1),
                  hessian=TRUE)

    #lrt for h01: c=0
    lrt_1<- -2*(opt_pn$value -opt_pnc$value)

    #lrt for h02: eta=1
    lrt_2<- -2*(opt_p$value -opt_pn$value)


    #if min(lambda)=0,  lrt1= 0, here h0 gets rejected if lrt_1!=0
    #if min(lambda)!=0,  lrt1 = 1/2*0+1/2*chi_1^2, lrt=chi_1^2

```

```
      if(min(lambda)==0){
        reject_h01<-opt_pnc$par[3]!=0

      # non-corrected version
      # reject_h02<-pchisq(lrt_2,1,lower.tail = FALSE)<alpha/2

        #corrected version
      reject_h02<-pchisq(lrt_2,1,lower.tail = FALSE)<alpha
    }else{
      reject_h01<-0.5*pchisq(lrt_1,1,lower.tail = FALSE)<alpha/2

        reject_h02<-pchisq(lrt_2,1,lower.tail=FALSE)<alpha/2
    }

    #does h01 or h02 get rejected?
    reject<-reject_h01||reject_h02


  }else if(h0_type=="etac"){
    #h01:eta=1, h01: c=0

    # find optimum over p,c if eta=1
    opt_pc<-optim(par=c(p,c),fn=loglike_pc,x=x,lambda=lambda,
                  lower=c(0,0),upper=c(Inf,Inf),
                  method="L-BFGS-B",control = list(maxit=200000,fnscale=-1),
                  hessian=TRUE)

    #lrt for h01: eta=1
    lrt_1<- -2*(opt_pc$value -opt_pnc$value)

    #lrt for h02: c=0
    lrt_2<- -2*(opt_p$value -opt_pc$value)


    #if min(lambda)=0,  lrt1= chi_1^2, lrt2=0
    #if min(lambda)!=0,  lrt1 = chi_1^2, lrt=1/2*0+1/2*chi_1^2

    if(min(lambda)==0){
      # non-corrected version
      # reject_h01<-pchisq(lrt_1,1,lower.tail = FALSE)<alpha/2

        #corrected version
      reject_h01<-pchisq(lrt_1,1,lower.tail = FALSE)<alpha

        reject_h02<-opt_pc$par[2]!=0
    }else{
      reject_h01<-pchisq(lrt_1,1,lower.tail = FALSE)<alpha/2

        reject_h02<-0.5*pchisq(lrt_2,1,lower.tail=FALSE)<alpha/2
    }

    #does h01 or h02 get rejected?
    reject<-reject_h01||reject_h02

  }else{#h0_type="both"

    #lrt for h0:(eta=1 and c=0)
```

```
    lrt_1<- -2*(opt_p$value -opt_pnc$value)
    lrt_2<- -2 #restvalue


   #if min(lambda)=0,  lrt1= chi_1^2
   #if min(lambda)!=0,  lrt1=1/2*chi_2^2+1/2*chi_1^2
    pval<-(0.5*pchisq(lrt_1,1,lower.tail=FALSE)
           +0.5*pchisq(lrt_1,2,lower.tail=FALSE))
    reject<-ifelse(min(lambda)==0, pchisq(lrt_1,1,lower.tail = FALSE)<alpha,
                   pval<alpha)

  }
  return(c(lrt_1,lrt_2,reject))
}

runsim<-function(p, eta, c, lambda, N, m, n, h0_type="both"){
# repeats run N times, to get N LRT values
  lrt1<-c()
  lrt2<-c()
  reject<-c()
  for(i in 1:N){
    res<-c()
    res<-run(p,eta,c,lambda,m,n,h0_type)

    lrt1<-append(lrt1,res[1])
    lrt2<-append(lrt2,res[2])
    reject<-append(reject,res[3])
  }
  return(cbind(lrt1,lrt2,reject))
}
```

Listing B.2: Experiment simulation

```

det_p_reject<-function(p, eta, c, lambda, N, m, h0_type="both"){
  prob<-c()
  n_values<-c(3,5,8,10,15,20,25,50)
  for(n in n_values){
    res<-c()
    res<-runsim(p,eta,c,lambda,N,m,n, h0_type)
    prob<-append(prob,mean(res[,3]))
  }
  return(prob)
}
```

Listing B.3: Determination reject probability

Appendix C

Example Cone Transformation

In this chapter, we will give an example of how $\tilde{C}$ from Section 3.1 is calculated from a two-dimensional $C = [0, \infty) \times \mathbb{R} = \{a \cdot e_1 + b \cdot e_2 : a \geq 0 \wedge b \in \mathbb{R}\}$ for the standard base $e_1 = (1, 0)^T$ and $e_2 = (0, 1)^T$. We take $\theta_0 = (0, 0)^T$ and $I(\theta_0)$ is as follows:

$$I(\theta_0) = \begin{pmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{pmatrix}$$

The spectral decomposition of $I(\theta_0)$, $P\Lambda P^T$, is as follows:

$$I(\theta_0) = P\Lambda P^T = \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \frac{3}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix}$$

Note:

$$\Lambda^{1/2} P^T = \begin{pmatrix} \sqrt{3/2} & 0 \\ 0 & \sqrt{1/2} \end{pmatrix} \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} -\sqrt{3/2} & \sqrt{3/2} \\ \sqrt{1/2} & \sqrt{1/2} \end{pmatrix}$$

Using this to calculate $\tilde{C}$ gives:

$$\begin{aligned} \tilde{C} &= \left\{ \tilde{\theta} : \tilde{\theta} = \Lambda^{1/2} P^T \theta \text{ for all } \theta \in C \right\} \\ &= \left\{ \begin{pmatrix} \tilde{\theta}_1 \\ \tilde{\theta}_2 \end{pmatrix} : \begin{pmatrix} \tilde{\theta}_1 \\ \tilde{\theta}_2 \end{pmatrix} = \begin{pmatrix} \sqrt{3/2}(-\theta_1 + \theta_2) \\ \sqrt{1/2}(\theta_1 + \theta_2) \end{pmatrix} \text{ for all } \begin{matrix} \theta_1 \geq 0 \\ \theta_2 \in \mathbb{R} \end{matrix} \right\} \\ &= \left\{ \begin{pmatrix} \tilde{\theta}_1 \\ \tilde{\theta}_2 \end{pmatrix} : \begin{pmatrix} \tilde{\theta}_1 \\ \tilde{\theta}_2 \end{pmatrix} = \sqrt{\frac{3}{2}}(-a + b) \cdot e_1 + \sqrt{\frac{1}{2}}(a + b) \cdot e_2 \text{ for all } \begin{matrix} a \geq 0 \\ b \in \mathbb{R} \end{matrix} \right\} \end{aligned}$$

In Figure C.1 both C and $\tilde{C}$ are shown. Now we want to find an orthonormal transformation so that $\tilde{C} = [0, \infty) \times \mathbb{R}$ in terms of the transformed base. For this new base, $\{\tilde{e}_1, \tilde{e}_2\}$, we have that $\tilde{e}_2 = (\sqrt{3/2}e_1 + \sqrt{1/2}e_2)/\sqrt{2} = \sqrt{3/4}e_1 + \sqrt{1/4}e_2$. For $\tilde{e}_1$, we know $\tilde{e_1} \perp \tilde{e}_2$, and thus $\tilde{e}_1 = -\sqrt{1/4}e_1 + \sqrt{3/4}e_2$. Note that both $\tilde{e}_1$ and $\tilde{e}_2$ have to be the same length as e_1 and e_2. Now $\tilde{C} = \{a \cdot \tilde{e}_1 + b \cdot \tilde{e}_2 : a \geq 0 \wedge b \in \mathbb{R}\}$.

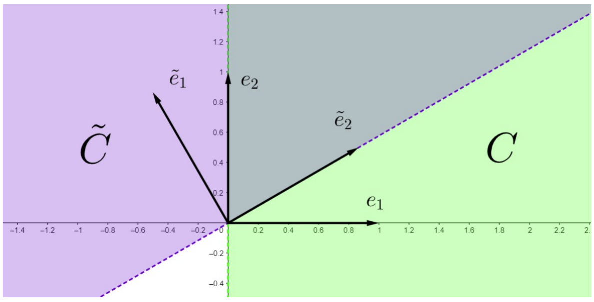


Figure C.1: C and $\tilde{C}$, and the standard and transformed bases