Asymptotic Normality of Goodness–of–Fit Statistics for Sparse Poisson and Case–Control Data

Ursula Müller
Department of Mathematics and Computer Sciences
University of Bremen

July 1997

Referees: Prof. Dr. Gerhard Osius (University of Bremen)
Prof. Dr. Hans–Peter Kinder (University of Bremen)

Degree day: May 29, 1997

In contrast to the original dissertation from April 1997, this revised version additionally contains some "Final Remarks", which have been appended in order to survey the derived results — in particular under the aspect of applicability — and to give a general view on advisable improvements and open questions. More detailed supplementary information concerning this subject was also included in chapter 2, sec. 2.2. For reasons of clarity and in order to facilitate the reading of the proofs, several slight modifications were carried out in chapter 5 and 6, as, for example, the separate formulation of frequently used auxiliary results. Moreover, in chapter 4 and 6 some auxiliary calculations were added. Finally, all chapters were revised and minor mistakes were removed.

Bremen, July 1997 Ursula Müller

My special thanks go to Prof. Dr. Gerhard Osius, who suggested the theme of this thesis and supported its development with valuable ideas and advice.

Contents

1. Introduction

The subject of this thesis are goodness–of–fit tests for discrete data with parameters to be estimated. For this purpose, observed and expected counts for a given parametric model will be compared applying a certain "distance measure", which should be small if the model is true and large if it is not. Of course, the distribution of the distance under the nullhypothesis, i.e. the case that the model holds, is needed in order to check the goodness–of–fit. For those tests, commonly Pearson's $\mathcal{X}^2$ and the Likelihood Ratio Statistic are used. Cressie and Read (1984) have embedded them in a family of "Power–Divergence Statistics" SD_λ ($\lambda \in \mathbf{R}$), where each member SD_λ is a sum over all deviations between observed and expected counts:

$$SD_\lambda = \sum_{cells} a_\lambda(obs, exp)$$

with distance function

$$a_\lambda(obs, exp) = \frac{2 \cdot obs}{\lambda(\lambda+1)} \cdot \left(\left(\frac{obs}{exp} \right)^\lambda - 1 \right) - \frac{2}{\lambda+1}(obs - exp) \geq 0.$$

The values $\lambda = 0$, where a_0 is defined by continuity, and $\lambda = 1$ yield the Likelihood Ratio and Pearson's $\mathcal{X}^2$ Statistic respectively. To allow zero observations, which are typical when data are sparse, only values $\lambda \in (-1, \infty)$ will be considered.
The observed data consist of a $J \times K$ contingency table:

	categories D					
groups	1	$\cdots$	k	$\cdots$	K	sum
1	X_{11}	$\cdots$	X_{1k}	$\cdots$	X_{1K}	X_{1+}
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
j	X_{j1}	$\cdots$	X_{jk}	$\cdots$	X_{jK}	X_{j+}
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
J	X_{J1}	$\cdots$	X_{Jk}	$\cdots$	X_{JK}	X_{J+}
sum	X_{+1}	$\cdots$	X_{+k}	$\cdots$	X_{+K}	X_{++}

with J groups usually represented by covariables $Z \in \mathbf{R}^R$ ($R \in \mathbf{N}$), K categories $D \in \{1, \ldots, K\}$ and observed counts $X_{jk} \in \mathbf{N}_0$ of objects (Z, D) belonging to group

j and category k.

The sampling schemes considered here are Poisson and product–multinomial. Poisson distribution, i.e. when $X_{11}, \ldots, X_{JK}$ are independent Poisson distributed random variables, typically occurs when all available data (Z, D) are collected within a fixed period. The product–multinomial distribution here is column–multinomial, that is to say that the surveyed columns are K independent multinomials. An important application are case–control studies where D is usually an indicator for a disease with $K = 2$ values 1 (disease) or 0 (no disease).

The main interest in the investigation of contingency tables lies in the description of associations within a table rather than in the marginal distribution of covariables and categories. Thus, the models to be tested are association models, where dependencies between covariables and categories are specified by a finite–dimensional parametrized model and where the marginal distribution may vary.

It is commonly held that for increasing sample size, SD_λ is asymptotically χ^2 distributed under the common sampling schemes, which are Poisson or conditional Poisson, such as (product–) multinomial sampling. For this approach, especially the number of cells $J \times K$ is assumed to be fixed, and hence the number of parameters is finite. There are, however, additional assumptions which are often not given, especially not when the data are sparse, for instance the increase of all expected values. The aim of this dissertation is to meet this situation using an "increasing–cells" approach, i.e. the number of categories K is fixed and the number of groups J increases, and to derive a limiting normal distribution of SD_λ. In particular, the expected values of each cell may be small but need not be. One difficulty in proving such a limiting result is that the models considered do not specify the marginal distribution of the covariable groups. So when the number of groups tends towards infinity, one has to deal with an asymptotically *infinite* number of nuisance parameters.

A number of authors have also discussed the possibility of taking a normal rather than a χ^2 approximation using the increasing–cells approach, e.g. Peter McCullagh (1986), Gerhard Osius and Dieter Rojek (1992) and Carl Morris (1975). McCullagh considers Pearson's χ^2 and the Likelihood Ratio Statistic for Poisson and binomial sampling with all expected values and hence the distribution of the table specified by a *finite*–dimensional parametric model. Osius and Rojek derive the asymptotic normality of SD_λ for "row–multinomial" sampling, which is also product–multinomial but with the *rows* being J *independent* multinomials. In terms of expected values, they examine the same models as this thesis. Because of the underlying sampling scheme, though — the group sizes are given, they do not have to deal with an increasing number of nuisance parameters. Although Morris is not concerned with parameter estimation but with given expected values, his paper also must be mentioned. Morris proves the asymptotic normality of Pearson's χ^2 and the Likelihood

Ratio Statistic for multinomial sampling, i.e. for a J–dimensional multinomial vector where J increases to infinity, by making use of the fact that the multinomial distribution is a special conditional Poisson distribution. Thus, he can dispense with the stochastic dependencies within the components and apply the central limit theorem. His work and the work of Osius and Rojek provide important ideas for the general concept of this thesis, in particular valuable elements of proofs, which could partially be adopted. Especially the results for calculating asymptotic orders and certain approximation steps could be deduced in a way similar to Osius and Rojek's. Morris' idea is of central importance for the asymptotics in the column–multinomial case.

The general background will be provided in chapter 2. In the first section, the sampling schemes will be given, which yield the distribution models considered here, Poisson– and column–multinomial. After this, the "increasing–cells" approach will be explained in detail. The following section then presents the parametric models in question, which, in order to describe dependencies within a table, formulate ratios of expectations as functions of a finite–dimensional parameter vector. Since only ratios within a table are modelled and hence the marginal distribution of the covariable groups is not specified, this will in particular lead to additional nuisance–parameters, namely the expectations of the row–sums. For applications usually generalized linear models are taken, in particular logistic regression models. These will be sketched as an example and it will be outlined in how far they fit into the modell class considered here and which points are in need of improvement. With the models of interest being characterized, the nullhypothesis to test, which will later be assumed for the derivation of the asymptotic results, can be formulated. Finally the model fit will be described, which will be done using maximum likelihood or asymptotic equivalent estimators, and supplementary information like the formulae for information matrix and scores will be given. In the last section, the Power–Divergence Family SD_λ, respectively the distance function a_λ, is defined and some characteristic properties and derivatives are summarized. With all necessary background provided, then in anticipation of the last chapter, the final standardization terms for a goodness–of–fit test for both distribution models and the decision rules are given. Further, all assumptions will be listed and explained in brief.

Chapter 3 will treat Morris' approach (1975). Morris not only studied the concrete statistics mentioned, but above all introduced a general method to achieve a limiting normal distribution for arbitrary multinomial sums. This approach will be used later to derive the asymptotic normality of the goodness–of–fit statistic in the case of column–multinomial sampling. In this chapter now, the simple generalization from the multinomial to the column–multinomial model will be considered. The first section illustrates the method, the theoretical statements will then be formulated in the following section. Since this approach makes use of the fact that the multinomial is a particular conditional Poisson distribution, it will turn out that asymptotic results for column–multinomial sums can be deduced to the Poisson case and hence that Poisson and column–multinomial statistics can partially be treated together. Additionally, however, some requirements concerning the convergence of the conditional distribution will have to be fulfilled. These can be checked using a criterion from Steck (1957), which is essential for Morris' approach and will also be given. Of special importance for the later application case will be the generalization of a fundamental lemma from Morris (1975), which summarizes the approach formulating conditions for the asymptotic normality of arbitrary column–multinomial sums.

In chapter 4, some general results for the Poisson distribution will be presented, which, applied to a_λ, will be very useful for further proofs concerning the weak convergence of the test statistic. The first section deals with expected values of functions under Poisson distribution. By considering these expectations as functions of the parameter μ, conditions for existence, continuity and differentiability in μ are given. Further, a very helpful Taylor approximation for expectations of the form $E(\mu^r(H(\frac{X}{\mu}))$ is stated for the asymptotics $\mu \to \infty$, where X is $Poisson(\mu)$ distributed, H is a real valued function and r a nonnegative integer. Thus, a valuable tool for calculating asymptotic orders is given, which can in particular be applied to the distance function a_λ. The idea for this theorem goes back to Osius (1984), who showed a similar result for the binomial distribution $B(n,p)$ for the asymptotics $n \to \infty$. Decisive for the proof of this theorem is the fundamental property of the central Poisson moments to be polynomials in μ with the degree not exceeding half of the order, which will also be given. The statements following in the second section are applications of these results to moments of a_λ and will lead to important bounding statements, which will be used throughout the proofs of the later chapters.

Subject of the actual main chapters of this thesis, chapter 5 and chapter 6, will be the derivation of a limiting normal distribution for SD_λ under the nullhypothesis. In order to eliminate the correlations caused by estimating, chapter 5 deals with a gradual approximation of the centered goodness–of–fit statistic. This will be done formulating the single steps for both sampling schemes together, and thus leads to analytically identical approximations for both distribution models. Especially because of the underlying stochastic dependencies in case of column–multinomial sampling, at certain points, however, differentiated argumentation will be necessary. In particular, one auxiliary result concerning the column–multinomial model will be shown using Morris' approach, which requires rather comprehensive argumentation. For reasons of clarity, it therefore will be treated separately in section 5.1, prior to the actual approximation steps. In there, some arguments from Morris' proof concerning the asymptotic normality of the Likelihood Ratio Statistic (1975) could be directly adopted. Referring to Osius and Rojek (1992), who also considered the Power–Divergence Family, though in the case of row–multinomial sampling, it should be mentioned that although their approximation is similarly structured and the proceeding could partially be adopted, here additional arguments become necessary in order to handle the nuisance parameters as well as the stochastic dependencies in the column–multinomial case.

In chapter 6, the limiting normality of the approximated sum will be proved. When Poisson sampling is considered, the approximation is, in contrast to the column–multinomial case, a sum of independent variables and the asymptotic normality can, using the the auxiliary results from chapter 4, thus be easily derived applying the central limit theorem. In the case of column–multinomial sampling, the desired limiting normality of the approximated statistic will be shown choosing Morris' approach, which uses results from the proof concerning the Poisson approximation. Similar to the proof of the auxiliary result concerning the approximation in section 5.1, where Morris' method was applied, for this proof some of Morris' results could be adopted. With the asymptotic normality of the approximated statistic, scaled with the "true" asymptotic variance, given in section 6.1, the next section establishes as the last step needed the consistency of the variance estimation. The following theorems then present as a conclusion of the preceding statements the main results of this thesis, namely the asymptotic normality of the test statistic under the nullhypothesis for Poisson (Theorem 6.4) and column–multinomial–sampling (Theorem 6.5). The result for the column–multinomial statistic is only verified for the subclass with $\lambda \in (-1, 1]$, which, however, includes all important goodness–of–fit statistics.

In conclusion to the actual chapters follow some "Final Remarks", giving an outlook on advisable improvements and formulating open questions. Finally, the appendix lists the first central moments of the Poisson distribution, which are especially necessary for the bounding results in chapter 4. Further, an inequality concerning the distance function and two technical results, which do not deal directly with the goodness–of–fit statistic, but are needed in several proofs, will be given.

2. Model and Goodness–of–Fit

2.1 Stochastic Model and Asymptotics

Starting off with the sampling scheme, in this section contingency tables, models for their distributions and the asymptotic approach will be explained.

Starting point for the **Poisson distribution model** is a sample of independent, identically distributed observations (Z_i, D_i), $i = 1, \ldots, N$, with $Z \in \mathbf{R}^R$ being a vector of covariables and D a categorial random variable with values in $\{1, \ldots, K\}$. For the image space of Z, let a disjoint decomposition be considered: $ImZ = \bigcup_{j=1}^{J} I_j$. If the observations (Z_i, D_i), $i = 1, \ldots, N$, come in by chance within a fixed period t_0 — that is to say that the total size N is a random variable — and nearby assumptions are fulfilled (see for example Billingsley, 1986, section 23), then the distribution of the counts $X_{jk} = X_{jk}(t_0)$ for the event $(Z \in I_j, D = k)$ is given by $J \cdot K$ independent Poisson processes with intensity $\lambda_{jk} > 0$, i.e. for every $j \in \{1, \ldots, J\}$ and $k \in \{1, \ldots, K\}$ X_{jk} is Poisson distributed with the expected value $\mu_{jk} = t_0 \cdot \lambda_{jk} > 0$, and the variables $X_{11}, \ldots, X_{JK}$ are stochastically independent. Since sums of stochastically independent Poisson variables are Poisson distributed themselves, this also holds for the size N, i.e. $N = \sum_{j=1}^{J} \sum_{k=1}^{K} X_{jk} =: X_{++} \sim Pois(\mu_{++})$ with $\mu_{++} := \sum_{j=1}^{J} \sum_{k=1}^{K} \mu_{jk}$. Hence the Poisson distribution model is of a very general kind and typically occurs if all available data are collected within a fixed period.

After these preliminary remarks, to see the Poisson model come into being, in the following a Poisson distribution will be assumed. In order to study an asymptotic approach, let this from now on be indicated by a running index n. Now for each $n \in \mathbf{N}$ a disjoint decomposition of the image space of Z is considered:

$$ImZ = \bigcup_{j=1}^{J^n} I_j^n, \quad I_1^n, \ldots, I_{J^n}^n \text{ pairwise disjoint.}$$

Hence, with N^n as the stochastic total sample size, for the counts X_{jk}^n will be supposed $(j \in \{1, \ldots, J^n\}, k \in \{1, \ldots, K\}, n \in \mathbf{N})$:

$$X_{jk}^n = |\{1 \leq i \leq N^n | Z_i \in I_j^n, D_i = k\}| \sim Pois(\mu_{jk}^n),$$
$$X_{11}^n, X_{12}^n, \ldots, X_{J^n K}^n \quad \text{stochastically independent.}$$

If the event $Z \in I_j^n$ is interpreted as the belonging to a covariable group, indicated through the code $C = j$, this yields for the counts the following representation as a contingency table $X^n = (X_{jk}^n)_{j,k}$:

group/ code	categories 1	$\cdots$	k	$\cdots$	K	sum
1	X_{11}^n	$\cdots$	X_{1k}^n	$\cdots$	X_{1K}^n	X_{1+}^n
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
j	X_{j1}^n	$\cdots$	X_{jk}^n	$\cdots$	X_{jK}^n	X_{j+}^n
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
J^n	$X_{J^n1}^n$	$\cdots$	$X_{J^nk}^n$	$\cdots$	$X_{J^nK}^n$	$X_{J^n+}^n$
sum	X_{+1}^n	$\cdots$	X_{+k}^n	$\cdots$	X_{+K}^n	$X_{++}^n = N^n$

The subscript "$\cdot$" will always denote a vector and "+" a summation over the corresponding index (for example is $X_{\cdot k} = (X_{1k}, \ldots, X_{J^nk})^T$ the k–th column and X_{+k} the sum over its components).

The relevant sampling scheme to achieve a **column–multinomial distribution** are K independent samples, each of size n_k, from the conditional distribution of Z given $D = k$ $(k = 1, \ldots, K)$. Notably, n_k and hence the total size $n = n_+ = \sum_{k=1}^K n_k$ is fixed. If the k–th sample $(Z_i|D = k)$, $i \in \{1, \ldots, n_k\}$, is considered, the vector $Y_{\cdot k}^n = (Y_{1k}^n, \ldots, Y_{J^nk}^n)^T$ of counts

$$Y_{jk}^n = |\{1 \le i \le n_k | Z_i \in I_j^n\}|$$

has a multinomial distribution of size n_k with J^n classes and probability vector $\pi_{\cdot k|D}^n = (\pi_{1k|D}^n, \ldots, \pi_{J^nk|D}^n)$. Here each component $\pi_{jk|D}^n$ of $\pi_{\cdot k|D}^n$ equals the probability $P(Z \in I_j^n | D = k)$ and with $E(Y_{jk}^n) = \mu_{jk}^n$ especially holds $\pi_{jk|D}^n = \mu_{jk}^n / n_k$.

group/ code	categories 1	$\cdots$	k	$\cdots$	K	sum
1	Y_{11}^n	$\cdots$	Y_{1k}^n	$\cdots$	Y_{1K}^n	Y_{1+}^n
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
j	Y_{j1}^n	$\cdots$	Y_{jk}^n	$\cdots$	Y_{jK}^n	Y_{j+}^n
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
J^n	$Y_{J^n1}^n$	$\cdots$	$Y_{J^nk}^n$	$\cdots$	$Y_{J^nK}^n$	$Y_{J^n+}^n$
sum	$Y_{+1}^n = n_1$	$\cdots$	$Y_{+k}^n = n_k$	$\cdots$	$Y_{+K}^n = n_K$	$Y_{++}^n = n$

The independence of the samples thus yields for $Y^n = (Y^n_{jk})_{j,k}$ a product–multinomial distribution:

$$Y^n_{\cdot 1}, \ldots, Y^n_{\cdot k}, \ldots, Y^n_{\cdot K} \quad \text{stochastically independent},$$
$$Y^n_{\cdot k} \sim Multi_{J^n}(n_k, \pi^n_{\cdot k|D}) \text{ with } \pi^n_{\cdot k|D} = (\pi^n_{1k|D}, \ldots, \pi^n_{J^n k|D}),\ \pi^n_{jk|D} = \frac{\mu^n_{jk}}{n_k}.$$

Because of the conventional arrangement of a contingency table, this distribution model is often also called "column–multinomial".

Particularly useful for theoretical considerations is the fact that the column–multinomial is a conditional Poisson distribution, obtained when the column sums are fixed in the Poisson model. This will be essential for the derivation of the limiting distribution for the column–multinomial statistic and will be discussed in detail in the next chapter. Further conditional Poisson models, which will not be studied here but should nevertheless be mentioned, are the multinomial, the row–multinomial and the hypergeometric distribution. The underlying sampling scheme of a multinomial distribution is a sample of size n from the common distribution of Z and D, i.e. the underlying distribution table consists of the probabilities $P(Z \in I^n_j, D = k)$. It can be derived from the Poisson distribution through fixing of the total size. The row–multinomial distribution is the analogue to the column–multinomial distribution — here J^n independent samples of size n_j ($j \in \{1, \ldots, J^n\}$) are taken out of the conditional distribution $\mathcal{L}(D|Z \in I^n_j)$. Hence the rows are stochastically independent multinomials and the components of each probability vector equal the conditional probabilities $P(D = k|Z \in I^n_j)$. Fixing of all marginal sums in the Poisson distribution model finally yields the hypergeometric distribution, which, however, is merely of theoretical interest.

In the following, for Poisson and column–multinomial sampling the **asymptotics**, which have in the preceding considerations already been indicated through the index n, will be explained in detail. Just like for the commonly considered "fixed–cells approach", here will also be assumed:

- The expected total sample size tends towards infinity, $\mu^n_{++} \longrightarrow \infty$,
- the dimension R of the covariable vector is fixed,
- the number K of categories is fixed.

The meaning of n for the column–multinomial distribution is clear: Because the expected sum $\sum_{j=1}^{J^n} \sum_{k=1}^{K} \mu^n_{jk} = \mu^n_{++}$ equals the total sample size Y^n_{++}, an increase of $Y^n_{++} = \sum_{k=1}^{K} n_k = n$ will be considered. In the case of Poisson sampling, n will in practice certainly be identified with the *realized* sample size. Because this quantity is in fact a random variable, it is more useful for theoretical investigations to choose n

as a formal index, which increases proportionally to μ^n_{++}, i.e. $\mu^n_{++} = nc + o(1)$ with constant $c > 0$. Additionally will now be supposed:

- The number of groups J^n increases, $J^n \longrightarrow \infty$.

This approach allows an application of asymptotic results to tables with sparse data, since in contrast to the "fixed–cells asymptotics", an increase of all expected values μ^n_{jk} needs not be assumed. Because in both distribution models, Poisson and column–multinomial, the marginal distribution concerning the covariables will here be left arbitrary, additional conditions concerning grouping resp. the underlying probabilities will have to be met (see sec. 2.3). One basic requirement to accomplish the increasing–cells approach is the existence of a sequence of decompositions $\bigcup_{j=1}^{J^n} I^n_j$ increasing in number, where for each partition the probability of getting filled must be positive, i.e. $P(Z \in I^n_j) > 0$ for all j, n. This is in particular not given if the distribution of the covariables is discrete with finite domain. As a stronger condition one will even have to demand that asymptotically all groups be filled with probability one (see cond. LC0, sec. 2.3).

2.2 Parametric Modelling and Model Fitting

Keeping the notation of the last section, now the actual models of interest, for which the goodness–of–fit shall be tested, and the estimation will be described. At first let not a concrete distribution, but merely the table of expectations $\mu^n = (\mu^n_{jk})_{j,k} \in (0,\infty)^{J^n \times K}$ be considered. Of interest are associations within a table, thus between covariable groups and categories, which are described by parametric models. For $j = 1, \ldots J^n$ and $k = 1, \ldots, K$ the expectations are modelled as follows:

$$\mu^n_{jk} = \mu^n_{jk}(\theta) = \mu^n_{j+}\pi^n_{jk|C}(\theta) \quad \text{with } \pi^n_{jk|C}(\theta) = \frac{\mu^n_{jk}(\theta)}{\mu^n_{j+}}, \theta \in \mathbf{R}^S. \tag{2.1}$$

Because the expectations $\mu^n_{1+}, \ldots, \mu^n_{J^n+}$ may vary, there is no reduction of dimension in the marginal distribution $\mathcal{L}(Z)$, which is usually of minor interest.
The hypothesis to check with a goodness–of–fit test now says that the model is true:

$$H_0 : \quad \exists\, \theta_0 \in \Theta : \mu^n_{jk} = \mu^n_{j+}\pi^n_{jk|C}(\theta_0) \quad \text{for all } j, k, n$$

with $\Theta \subset \mathbf{R}^S$ being an open parameter space. Because of the given sampling scheme in case of Poisson or *row*–multinomial distribution, the ratio $\pi^n_{jk|C} = \mu^n_{jk}/\mu^n_{j+} = \mu^n_{jk}/n_j$ equals the conditional probability $P(D = k | Z \in I^n_j)$, which is of primary interest in the study of contingency tables. This concrete interpretation as a probability does not apply if *column*–multinomial distribution is considered. Even though by sampling holds $\pi^n_{jk|D} = \mu^n_{jk}/\mu^n_{+k} = \mu^n_{jk}/n_k = P(Z \in I^n_j | D = k)$, through the column sums being fixed, a marginal distribution of D is already given. Hence neither the true common distribution $\mathcal{L}(Z, D)$ nor the conditional distribution $\mathcal{L}(D|Z)$ can be derived.

By (2.1) the ratios $\pi^n_{jk|D} = \mu^n_{jk}/\mu^n_{+k}$, which are conditional probabilities in the case of *column*–multinomial sampling, depend on the same parameter vector θ as $\pi^n_{jk|C}$:

$$\mu^n_{jk} = \mu^n_{jk}(\theta) = \mu^n_{+k}\pi^n_{jk|D}(\theta) = \mu^n_{j+}\pi^n_{jk|C}(\theta) \quad \Leftrightarrow \quad \pi^n_{jk|D}(\theta) = \frac{\mu^n_{j+}}{\mu^n_{+k}}\pi^n_{jk|C}(\theta),\ \theta \in \mathbf{R}^S,$$

with $\mu^n_{+k} = n_k$ for all k, n if column–multinomial sampling is considered. In this distribution model, however, and in contrast to Poisson sampling, additional marginal conditions have to be fullfilled in order to guarantee

$$\sum_{j=1}^{J^n} \pi^n_{jk|D}(\theta) = \sum_{j=1}^{J^n} \frac{\mu^n_{j+}}{n_k} \cdot \pi^n_{jk|C}(\theta) = 1. \tag{2.2}$$

This leads to evident restrictions concerning the class of models $\pi^n_{jk|C}(\theta)$ described in (2.1), which are, in view of applications, in general not met (see also (2.6)). For the case of column–multinomial sampling, the investigation of parametric models as described in (2.1) thus turns out to be merely of theoretical interest.

In applications, the single groups are usually represented by covariable vectors z^n_j, and $\pi^n_{jk|C}(\theta) = F_k(z^n_j, \theta)$ is modelled with $F_1, \ldots, F_K$ being given functions. In generalized linear models, these frequencies depend on the covariables only through a linear combination, e.g.:

$$\pi^n_{jk|C}(\theta) = G_k(\langle z^n_j, \theta\rangle) \quad \text{for } j = 1, \ldots, J^n, k = 1, \ldots, K, n \in \mathbf{N}.$$

Models for dependencies within contingency tables are of special interest in epidemiology, where the $\pi^n_{jk|C}$ are typically disease risks and the categorial variable D indicates different stages of a disease. For those investigations, usually a more specific parametrization is chosen, which will now be briefly described. Starting point is a *log linear model*, $\log \mu^n_{jk} = \eta^n_{jk}$, with linear predictor η^n_{jk} and the following parametrization of the complete model:

$$\log \mu^n_{jk} = \eta^n_{jk} = \alpha^n + \rho^n_j + \gamma^n_k + \psi^n_{jk}.$$

Interpretability and uniqueness of the parameters are given through suitable marginal conditions. Choosing especially $\rho^n_1 = \gamma^n_1 = 0$, $\psi^n_{j1} = 0$, $\psi^n_{1k} = 0$ for all j, k, n, the parameters ψ^n_{jk} turn out to be the "log–odds–ratios":

$$\psi^n_{jk} = \eta^n_{jk} + \eta^n_{11} - \eta^n_{j1} - \eta^n_{1k} = \log \frac{\mu^n_{jk} \cdot \mu^n_{11}}{\mu^n_{j1} \cdot \mu^n_{1k}}.$$

A suitable parametric model in order to accomplish (2.1) is now as follows:

$$\eta^n_{jk} = \alpha^n + \rho^n_j + \gamma_k + \psi^n_{jk}(\beta) \quad \text{for all } j, k, n \tag{2.3}$$

with the commonly considered special case

$$\eta^n_{jk} = \alpha^n + \rho^n_j + \gamma_k + \langle z^n_j, \beta_k\rangle \quad \text{for all } j,k,n \quad (\beta_k \in \mathbf{R}^R).$$

Here $\rho^n_1 = \gamma_1 = 0$, $\beta_1 = 0$ and further (without loss of generality) $z^n_1 = 0$ are set in order to meet the marginal conditions. With the special choice $\psi^n_{jk} = \langle z^n_j, \beta_k\rangle$ this model assumes linear dependencies between the covariables and the different stages of a disease. An equivalent rewriting now reveals (2.3) to be a *multivariate logit model*:

$$\text{logit}_k \pi^n_{j\cdot|C} := \log \pi^n_{jk|C} - \log \pi^n_{j1|C} = \gamma_k + \psi^n_{jk}(\beta).$$

In particular the ratio $\pi^n_{jk|C} = \mu^n_{jk}/\mu^n_{j+}$ states in accordance with (2.1) explicitly as follows:

$$\pi^n_{jk|C} = \pi^n_{jk|C}(\theta) = \frac{\exp(\gamma_k + \psi^n_{jk}(\beta))}{\sum_{l=1}^K \exp(\gamma_l + \psi^n_{lk}(\beta))}. \tag{2.4}$$

In the concrete example this clearly results in

$$\pi^n_{jk|C} = \pi^n_{jk|C}(\theta) = \frac{\exp(\gamma_k + \langle z^n_j, \beta_k\rangle)}{\sum_{l=1}^K \exp(\gamma_l + \langle z^n_j, \beta_l\rangle)} \tag{2.5}$$

with $\theta = (\gamma_2, \ldots, \gamma_K, \beta_2, \ldots, \beta_K) \in \mathbf{R}^S$ $(S = (K-1)\cdot(R+1))$ being the parameter vector of interest.

By definition, model (2.4) fits well into the more general model (2.1). In the case of Poisson sampling, the finite–dimensional parametrization of the odds–ratios and the marginal distribution $\mathcal{L}(D)$ considered there is reasonable — in particular, since in this distribution model all parameters may vary independently of each other. As already announced, the adequacy of model (2.4) for column–multinomial sampling is, however, not given. Instead of (2.4), in this case more generally

$$\pi^n_{jk|C} = \pi^n_{jk|C}(\theta^n) = \frac{\exp(\gamma^n_k + \psi^n_{jk}(\beta))}{\sum_{l=1}^K \exp(\gamma^n_l + \psi^n_{lk}(\beta))} \tag{2.6}$$

should be considered with especially the sequences of parameters (γ^n_k) not being constant $(k = 2, \ldots, K)$. As can be seen by simple calculations, in order to accomplish the marginal conditions (2.2) each γ^n_k is already uniquely determined through *all* other parameters. Beside the fact that these clearly vary with n, this additionally entails each $\pi^n_{jk|C}$ to be dependent on the vector of row sums $\mu^n_{\cdot+} = (\mu^n_{j+})_j$. Hence the appropriate model for column–multinomial sampling (2.6) does not fit into the assumed model class (2.1), which considers a decomposition of the expectations μ^n_{jk} in two *separate* parts, μ^n_{j+} and $\pi^n_{jk|C}(\theta)$.

For the estimation of θ_0 in the following, the maximum likelihood estimator $\hat{\theta}^n$, or some equivalent estimation function in regard to the approximability through information matrix and scores, will be taken. The **loglikelihood function** $l^n(\theta|X^n, \mu^n_{\cdot+})$

and the **score vector** $U^n(\theta|X^n)$ are in the *Poisson distribution model* given as follows:

$$\begin{aligned} l^n(\theta|X^n,\mu^n_{\cdot+}) &= \sum_{j=1}^{J^n}\sum_{k=1}^{K}(X^n_{jk}\log\mu^n_{jk}-\mu^n_{jk}-\log X^n_{jk}!) \\ &= \sum_{j=1}^{J^n}\sum_{k=1}^{K}(X^n_{jk}\log\mu^n_{j+}\pi^n_{jk|C}(\theta)-\mu^n_{j+}\pi^n_{jk|C}(\theta)-\log X^n_{jk}!) \\ &= \sum_{j=1}^{J^n}\Big(\sum_{k=1}^{K}X^n_{jk}\log\mu^n_{j+}+\sum_{k=1}^{K}X^n_{jk}\log\pi^n_{jk|C}(\theta)-\mu^n_{j+}-\sum_{k=1}^{K}\log X^n_{jk}!\Big) \end{aligned}$$

with loglikelihood kernel $\sum_{j=1}^{J^n}\sum_{k=1}^{K}X^n_{jk}\log\pi^n_{jk|C}(\theta)$,

$$\begin{aligned} U^n(\theta|X^n) &= \sum_{j=1}^{J^n}U^n_j(\theta|X^n_{j\cdot}) \\ &= \sum_{j=1}^{J^n}\sum_{k=1}^{K}X^n_{jk}D^T_\theta\log\pi^n_{jk|C}(\theta) \\ &= D^T_\theta l^n(\theta|X^n,\mu^n_{\cdot+}). \end{aligned}$$

The ML estimator $\hat\theta^n$ there clearly has to fulfil $U^n(\hat\theta^n|X^n)=0$. Since multinomial densities are particular conditional Poisson densities, for *column–multinomial distribution* holds

$$\begin{aligned} l^n(\theta|Y^n,\mu^n_{\cdot+}) &= \sum_{k=1}^{K}\log\left(\binom{n_k}{Y^n_{\cdot k}}(\pi^n_{\cdot k|D}(\theta))^{Y^n_{\cdot k}}\right) \\ &= \sum_{k=1}^{K}\log\binom{n_k}{Y^n_{\cdot k}}+\sum_{k=1}^{K}\log\prod_{j=1}^{J^n}(\pi^n_{jk|D}(\theta))^{Y^n_{jk}} \\ &= \sum_{k=1}^{K}\log\binom{n_k}{Y^n_{\cdot k}}+\sum_{k=1}^{K}\sum_{j=1}^{J^n}Y^n_{jk}\log\pi^n_{jk|D}(\theta) \\ &= \sum_{k=1}^{K}\log\binom{n_k}{Y^n_{\cdot k}}+\sum_{k=1}^{K}\sum_{j=1}^{J^n}Y^n_{jk}\log\frac{\mu^n_{j+}\pi^n_{jk|C}(\theta)}{n_k} \end{aligned}$$

with loglikelihood kernel $\sum_{j=1}^{J^n}\sum_{k=1}^{K}Y^n_{jk}\log\pi^n_{jk|C}(\theta)$,

$$\begin{aligned} U^n(\theta|Y^n) &= \sum_{j=1}^{J^n}U^n_j(\theta|Y^n_{j\cdot}) \\ &= \sum_{j=1}^{J^n}\sum_{k=1}^{K}Y^n_{jk}D^T_\theta\log\pi^n_{jk|C}(\theta) \\ &= D^T_\theta l^n(\theta|Y^n,\mu^n_{\cdot+}), \end{aligned}$$

i.e. in both distribution models the scores and thus the estimators $\hat{\theta}^n$ are analytically identical – as functions of X^n resp. Y^n, not, however, their distributions.

Since in later chapters many arguments will hold for both sampling schemes and the score vector will be repeatedly used, let in the following $U^n(\theta)$ denote the score vector for Poisson as well as for column–multinomial distribution. Specific statements will be indicated writing $U^n(\theta|X^n)$ or $U^n(\theta|Y^n)$. The same notational convention will in the following also be used for other stochastic terms, which are analytically identical in both distribution models.

If μ^n denotes the table of expectations for Poisson as well as for column–multinomial sampling, then in both cases the **information matrix** under H_0 is the same. If $\mu^n_{jk} = \mu^n_{j+}\pi^n_{jk|C}(\theta_0)$ holds for all j, k, n, for some $\theta_0 \in \Theta$, simple calculations give in the case of *Poisson sampling*

$$\begin{aligned}
I^n(\mu^n_{\cdot+}, \theta_0) &= Cov(U^n(\theta_0|X^n)) \\
&= \sum_{j=1}^{J^n}\sum_{k=1}^{K} Cov(X^n_{jk} D^T_\theta \log \pi^n_{jk|C}(\theta_0)) \\
&= \sum_{j=1}^{J^n}\sum_{k=1}^{K} D^T_\theta \log \pi^n_{jk|C}(\theta_0) \cdot \mu^n_{jk} \cdot D_\theta \log \pi^n_{jk|C}(\theta_0) \\
&= \sum_{j=1}^{J^n} \mu^n_{j+} \sum_{k=1}^{K} \pi^n_{jk|C}(\theta_0) D^T_\theta \log \pi^n_{jk|C}(\theta_0) \cdot D_\theta \log \pi^n_{jk|C}(\theta_0) \\
&= \sum_{j=1}^{J^n} \mu^n_{j+} \sum_{k=1}^{K} \frac{1}{\pi^n_{jk|C}(\theta_0)} D^T_\theta \pi^n_{jk|C}(\theta_0) \cdot D_\theta \pi^n_{jk|C}(\theta_0)
\end{aligned}$$

and in the *column–multinomial model*

$$\begin{aligned}
I^n(\mu^n_{\cdot+}, \theta_0) &= Cov(U^n(\theta_0|Y^n)) \\
&= Cov(\sum_{k=1}^{K}\sum_{j=1}^{J^n} Y^n_{jk} D^T_\theta \log \pi^n_{jk|C}(\theta_0)) \\
&= \sum_{k=1}^{K} Cov(\sum_{j=1}^{J^n} Y^n_{jk} D^T_\theta \log \pi^n_{jk|C}(\theta_0)) \\
&= \sum_{k=1}^{K} \left(Cov(\sum_{j=1}^{J^n} Y^n_{jk} D_{\theta_s} \log \pi^n_{jk|C}(\theta_0), \sum_{j=1}^{J^n} Y^n_{jk} D_{\theta_r} \log \pi^n_{jk|C}(\theta_0)) \right)_{r,s=1,\ldots,S} \\
&= \sum_{k=1}^{K} \left(\sum_{j=1}^{J^n} \mu^n_{j+} \frac{1}{\pi^n_{jk|C}(\theta_0)} D_{\theta_s} \pi^n_{jk|C}(\theta_0) \cdot D_{\theta_r} \pi^n_{jk|C}(\theta_0) \right)_{r,s=1,\ldots,S}
\end{aligned}$$

$$= \sum_{j=1}^{J^n} \mu_{j+}^n \sum_{k=1}^{K} \frac{1}{\pi_{jk|C}^n(\theta_0)} D_\theta^T \pi_{jk|C}^n(\theta_0) \cdot D_\theta \pi_{jk|C}^n(\theta_0).$$

In order to estimate the vector of expectations $\mu_{\cdot+}^n$, the row sums X_{j+}^n resp. Y_{j+}^n $(j = 1, \ldots, J^n)$ will be used. Since the ML estimator — if it exists — is uniquely determined through the normal equations, which require conformity of observed and estimated marginal sums, the row sums are ML estimators for μ_{j+}^n.
In conclusion, the following **estimators** will be considered:

- $\hat{\theta}^n$ maximum likelihood or some asymptotic equivalent estimator for θ_0 with
  - $\sqrt{n}(\hat{\theta}^n - \theta_0) = O_p(1)$,
  - $\hat{\theta}^n - \theta_0 = (I^n(\mu_{\cdot+}, \theta_0))^{-1} \cdot U^n(\theta_0) + O_p(n^{-1})$,
- $\hat{\mu}_{j+}^n$ $(j \in \{1, \ldots, J^n\})$ maximum likelihood estimator for μ_{j+}^n:
  - $\hat{\mu}_{j+}^n = X_{j+}^n$ (Poisson),
  - $\hat{\mu}_{j+}^n = Y_{j+}^n$ (column–multinomial).

The estimators for the modelled expectations thus will be $\hat{\mu}_{j+}^n \pi_{jk|C}^n(\hat{\theta}^n)$ for all $j \in \{1, \ldots, J^n\}, k \in \{1, \ldots, K\}, n \in \mathbf{N}$.

2.3 Goodness–of–Fit Statistic and Test

The best known statistics usually taken for goodness–of–fit tests are Pearson's X^2 and the Likelihood Ratio Statistic ("Deviance"). Cressie und Read (1984) have embedded them in the more general class of the so–called "Power–Divergence Statistics" SD_λ, whose members are characterized through the subscript $\lambda \in \mathbf{R}$ and which will be considered in the following. Each representative SD_λ is a sum over all deviations between observed and expected values,

$$SD_\lambda = \sum_{cells} a_\lambda(observed, expected),$$

in the situation considered here hence between observed data and fitted modelled expectations. The deviation is measured by the distance function a_λ, which will in the following only be defined for $\lambda \in (-1, \infty)$ to allow zero observations.

Definition 2.1 *For $\lambda \in (-1, \infty)$, the distance function*

$$\begin{array}{rcl} a_\lambda : \mathbf{R}_0^+ \times \mathbf{R}^+ & \longrightarrow & \mathbf{R}_0^+ \\ (x, \mu) & \longmapsto & a_\lambda(x, \mu) \end{array}$$

with $\mathbf{R}_0^+$ *denoting* $[0,\infty)$ *and* $\mathbf{R}^+ = (0,\infty)$*, is defined as follows:*

$$\begin{aligned} a_\lambda(x,\mu) &= \frac{2}{\lambda(\lambda+1)} \cdot x \cdot \left(\left(\frac{x}{\mu}\right)^\lambda - 1 \right) - \frac{2}{\lambda+1}(x-\mu) \quad \textit{for } \lambda \neq 0, \\ a_0(x,\mu) &= \lim_{\lambda\to 0} a_\lambda(x,\mu) = 2\Big(x \log\frac{x}{\mu} - (x-\mu)\Big). \end{aligned}$$

□

Certain values of λ indicate known goodness–of–fit statistics:

$$\begin{aligned} a_{-1/2}(x,\mu) &= 4(\sqrt{x}-\sqrt{\mu})^2 && \text{(Freeman–Tukey)}, \\ a_0(x,\mu) &= 2\Big(x\log x/\mu - (x-\mu)\Big) && \text{(Likelihood Ratio)}, \\ a_1(x,\mu) &= (x-\mu)^2/\mu && \text{(Pearson's } X^2\text{)}. \end{aligned}$$

Contingent on λ, the deviation between observation x and comparative value μ is weighted differently. The statistic of Pearson ($\lambda = 1$) is the only one which does not distinguish whether observations with the same absolute distance lie above or below μ. In so far, distance functions with $\lambda \neq 1$ are asymmetrical; for values $\lambda < 1$ observations below μ are weighted stronger, for $\lambda > 1$ it is just the other way round.

Before test statistic and decision rule will be given, first some important properties and derivatives of the distance function are presented in the following lemma, which can immediately be deduced from the definition. Beside the positive homogenity of $a_\lambda(x,\mu)$, in the following chapters especially the fact will be used that the first partial derivatives equal zero if x and μ coincide. Let also already be pointed out that the derivatives of $a_\lambda(x,\mu)$ in x do not for all λ exist in $x = 0$, which will in later Taylor expansions require a separate study of the zero.

Lemma 2.2 *For the distance function* $a_\lambda(x,\mu)$, $\lambda \in (-1,\infty)$*, holds*

$$\begin{aligned} a_\lambda(\mu,\mu) &= 0, \\ c \cdot a_\lambda(x,\mu) &= a_\lambda(cx,c\mu) \quad \textit{for all } c \in \mathbf{R}^+ \quad \textit{(positive homogenity)}. \end{aligned}$$

Differentiation in x *gives the following derivatives:*

$$D_1 a_\lambda(x,\mu) = \begin{cases} \frac{2}{\lambda}\left(\left(\frac{x}{\mu}\right)^\lambda - 1\right) & \textit{for } \lambda \in (-1,\infty)\backslash\{0\} \\ 2\log\frac{x}{\mu} & \textit{for } \lambda = 0, \end{cases}$$

$$D_1 a_\lambda(0,\mu) = \lim_{x\to 0} D_1 a_\lambda(x,\mu) = -\infty \textit{ for } \lambda \in (-1,0],$$

$$D_1^k a_\lambda(x,\mu) = 2 \cdot \frac{x^{\lambda-(k-1)}}{\mu^\lambda} \cdot \prod_{i=1}^{k-2}(\lambda - i) \quad \textit{for } k \geq 2, \lambda \in (-1,\infty),$$

$$|D_1^k a_\lambda(0,\mu)| = |\lim_{x\to 0} D_1^k a_\lambda(x,\mu)| = \infty \textit{ for } k \geq 2, \lambda \in (-1,k-1)\backslash\{0,1,\ldots,k-2\}.$$

For $\lambda \in (-1, \infty)$ in particular holds:

$$D_1 a_\lambda(\mu,\mu) = 0, \quad D_1^2 a_\lambda(\mu,\mu) = \frac{2}{\mu}, \quad D_1^k a_\lambda(\mu,\mu) = 2\mu^{1-k} \cdot \prod_{i=1}^{k-2} (\lambda - i) \ \textit{for}\ k \geq 3.$$

Differentiation in μ provides

$$D_2 a_\lambda(x,\mu) = \frac{2}{\lambda+1}\Big(1 - (\frac{x}{\mu})^{\lambda+1}\Big), \quad D_2 a_\lambda(\mu,\mu) = 0,$$

$$D_2^2 a_\lambda(x,\mu) = \frac{2}{\mu}(\frac{x}{\mu})^{\lambda+1}, \quad D_2^2 a_\lambda(\mu,\mu) = \frac{2}{\mu},$$

and application of the functional equation

$$a_\lambda(x,\mu) = x \cdot D_1 a_\lambda(x,\mu) + \mu \cdot D_2 a_\lambda(x,\mu)$$

further yields as a conclusion:

$$D_2 a_\lambda(x,\mu) = \frac{1}{\mu}(a_\lambda(x,\mu) - x \cdot D_1 a_\lambda(x,\mu)),$$

$$D_1 D_2 a_\lambda(x,\mu) = -\frac{x}{\mu} D_1^2 a_\lambda(x,\mu) = -\frac{2}{\mu}(\frac{x}{\mu})^\lambda, \quad D_1 D_2 a_\lambda(\mu,\mu) = -\frac{2}{\mu}.$$

□

As can immediately be seen considering the monotonous behavior of the distance function — $a_\lambda(\cdot,\mu)$ is strictly decreasing on $[0,\mu)$ and increasing on (μ,∞) — large deviations between observed and fitted expectations and hence large values of a_λ resp. SD_λ obviously speak against the nullhypothesis.

Using the notation from section 2.1 and 2.2, now the test statistic $SD_\lambda^n(\hat{\mu}_{\cdot+}^n, \hat{\theta}^n)$ is complete. Considering both distribution models separately gives the statistics

$$SD_\lambda^n(\hat{\mu}_{\cdot+}^n, \hat{\theta}^n | X^n) = \sum_{j=1}^{J^n}\sum_{k=1}^{K} a_\lambda(X_{jk}^n, \hat{\mu}_{j+}^n \pi_{jk|C}^n(\hat{\theta}^n)) \quad \text{(Poisson)},$$

$$SD_\lambda^n(\hat{\mu}_{\cdot+}^n, \hat{\theta}^n | Y^n) = \sum_{j=1}^{J^n}\sum_{k=1}^{K} a_\lambda(Y_{jk}^n, \hat{\mu}_{j+}^n \pi_{jk|C}^n(\hat{\theta}^n)) \quad \text{(column–multinomial)},$$

whose asymptotic normality under H_0 will be derived in chapters 5 and 6. Writing

$$m_\lambda^n(\mu_{\cdot+}^n, \theta_0) = E(SD_\lambda^n(\mu_{\cdot+}^n, \theta_0 | X^n)) = \sum_{j=1}^{J^n}\sum_{k=1}^{K} E(a_\lambda(X_{jk}^n, \mu_{jk}^n)),$$

$$c_\lambda^n(\mu_{\cdot+}^n, \theta_0) = \sum_{j=1}^{J^n}\sum_{k=1}^{K} D_\theta \log \pi_{jk|C}^n(\theta_0) \cdot Cov(a_\lambda(X_{jk}^n, \mu_{jk}^n), X_{jk}^n),$$

$$\begin{aligned}
\sigma_\lambda^{n2}(\mu_{\cdot+}^n,\theta_0) &= \sum_{j=1}^{J^n}\sum_{k=1}^{K} Var(a_\lambda(X_{jk}^n,\mu_{jk}^n)) + 2J^n + \sum_{j=1}^{J^n}\frac{1}{\mu_{j+}^n} \\
&\quad -2\sum_{j=1}^{J^n}\frac{1}{\mu_{j+}^n}\sum_{k=1}^{K} Cov(a_\lambda(X_{jk}^n,\mu_{jk}^n),(X_{jk}^n)^2) \\
&\quad +4\sum_{j=1}^{J^n}\sum_{k=1}^{K}\pi_{jk|C}^n(\theta_0^n) Cov(a_\lambda(X_{jk}^n,\mu_{jk}^n),X_{jk}^n) \\
&\quad -c_\lambda^n(\mu_{\cdot+}^n,\theta_0)(I^n(\mu_{\cdot+}^n,\theta_0))^{-1}(c_\lambda^n(\mu_{\cdot+}^n,\theta_0))^T, \\
s_\lambda^{n2}(\mu_{\cdot+}^n,\theta_0) &= \sigma_\lambda^{n2}(\mu_{\cdot+}^n,\theta_0) - \sum_{k=1}^{K}\mu_{+k}^n(\gamma_{\lambda k}^n(\mu_{\cdot+}^n,\theta_0))^2, \\
\gamma_{\lambda k}^n(\mu_{\cdot+}^n,\theta_0) &= \frac{1}{\mu_{+k}^n}\sum_{j=1}^{J^n}(Cov(a_\lambda(X_{jk}^n,\mu_{jk}^n),X_{jk}^n) - \pi_{jk|C}^n(\theta_0)) \qquad (k=1,\ldots,K),
\end{aligned}$$

where in case of column–multinomial distribution for each $k \in \{1,\ldots,K\}$ the column sizes $\mu_{+k}^n = n_k$ are known, then the following standardizations and in particular decision rules for an asymptotic level α test will be obtained:

$$\text{rejection of } H_0 \quad \Leftrightarrow \quad \frac{SD_\lambda^n(\hat\mu_{\cdot+}^n,\hat\theta^n|X^n) - m_\lambda^n(\hat\mu_{\cdot+}^n,\hat\theta^n) + J^n}{\sigma_\lambda^n(\hat\mu_{\cdot+}^n,\hat\theta^n)} > z_\alpha$$

if Poisson sampling and

$$\text{rejection of } H_0 \quad \Leftrightarrow \quad \frac{SD_\lambda^n(\hat\mu_{\cdot+}^n,\hat\theta^n|Y^n) - m_\lambda^n(\hat\mu_{\cdot+}^n,\hat\theta^n) + J^n}{s_\lambda^n(\hat\mu_{\cdot+}^n,\hat\theta^n)} > z_\alpha$$

if column–multinomial sampling is considered (z_α is the upper α–quantile of the standard normal distribution $N(0,1)$). The main results stating the asymptotic normality of the test statistics are given in Theorem 6.4 for Poisson and in Theorem 6.5 for column–multinomial distribution as conclusions of the preceding results. The asymptotic normality will in the column–multinomial case only be shown for statistics with $\lambda \in (-1,1]$, which, however, include all important representatives of the power divergence family. In the case of Poisson sampling, the result is proved for arbitrary $\lambda > -1$.

In order to achieve a normal limit now for both distribution models, the following conditions are required to hold:

(RC1) $\pi_{jk|C}^n(\theta)$ is continuously differentiable twice in θ for all j,k,n,

(RC2) $\exists \epsilon > 0 : \pi_{jk|C}^n(\theta) \ge \epsilon$ for all $j,k,n,\theta \in \bar W$,

(RC3) $\exists M > 0 :$ a) $\|D_\theta \pi_{jk|C}^n(\theta)\| < M$ for all $j,k,n,\theta \in \bar W$,
b) $\|D_\theta^2 \pi_{jk|C}^n(\theta)\| < M$ for all $j,k,n,\theta \in \bar W$,

(LC0) $P(\hat{\mu}^n_{j+} > 0 \, \forall \, j \in \{1, \dots, J^n\}) \longrightarrow 1,$

where $\bar{W} \subset \Theta$ is a convex compact neighbourhood of the true parameter θ_0. Condition (LC0) concerning the estimators $\hat{\mu}^n_{j+} = X^n_{j+}$ resp. Y^n_{j+} $(j = 1, \dots, J^n)$ for the expectations of the row sums requires that asymptotically with probability 1 all groups are filled. Hence it is actually directed towards the distribution of the covariables resp. the way of grouping them. The first three assumptions (RC1) – (RC3) are **r**egularity **c**onditions for the modelled ratios, which, in the presence of covariables $z^n_1, \dots, z^n_{J^n}$, are usually met if the covariables are bounded. In this case generally $\pi^n_{jk|C}(\theta) = F_k(z^n_j, \theta)$ with given functions $F_1, \dots, F_K$ (cp. p. 14) is assumed. These functions and their derivatives are typically continuous in z^n_j and — in accordance with (RC1) — continuous in θ. Hence (RC3) is clearly fulfilled if the covariables are bounded (For an illustration consider the multivariate logit model (2.5), for which the derivatives can be determined by simple calculations).

Writing $\hat{\theta}^n$ and $U^n(\theta_0)$ for the parameter estimators and the scores in both distribution models in accordance with the notational convention provided last section, then the other required conditions for Poisson and column–multinomial sampling are as follows:

(LC1) $\frac{1}{n} I^n(\mu^n_{\cdot+}, \theta_0) \longrightarrow I_\infty$ positive definite,

(LC2) $\sqrt{n}(\hat{\theta}^n - \theta_0) = O_p(1),$

(LC3) $(\hat{\theta}^n - \theta_0) = (I^n(\mu^n_{\cdot+}, \theta_0))^{-1} U^n(\theta_0) + O_p(\frac{1}{n}),$

(BC) $\exists \, \epsilon > 0 : \mu^n_{jk} \geq \epsilon$ for all $j, k, n,$

(VCP) $\frac{J^n}{\sigma^{n2}_\lambda(\mu^n_{\cdot+}, \theta_0)} = O(1)$ (Poisson),

(VCC) $\frac{J^n}{s^{n2}_\lambda(\mu^n_{\cdot+}, \theta_0)} = O(1)$ (column–multinomial),

(MD1) $\frac{1}{\sqrt{J^n} n} \sum_{j=1}^{J^n} \sqrt{\mu^n_{j+}} \longrightarrow 0,$

(MD2) $\frac{1}{\sqrt{J^n}} \sum_{j=1}^{J^n} \frac{1}{\sqrt{\mu^n_{j+}}} \longrightarrow 0,$

(MD3) $\max_{1 \leq j \leq J^n} \pi^n_{jk|D}(\theta_0) \longrightarrow 0$ for all k (column–multinomial).

Assumptions (LC1) – (LC3) (“limiting **c**onditions”) are standard conditions; (LC2) and (LC3) are generally met by the maximum–likelihood estimator. The existence of a sequence of ML estimators will not be a prerequisite here — merely a sequence of estimators $\hat{\theta}^n$ is needed, which is consistent with the rate $1/\sqrt{n}$ (LC2) and approximable through information matrix and score vector (LC3). Beside condition (BC) ("**b**ounding **c**ondition") concerning the expected values, a **v**ariance **c**ondition (VCP) for the **P**oisson and a **v**ariance **c**ondition (VCC) for the **c**olumn–multinomial statistic will also be necessary. Here σ_λ^{n2} and s_λ^{n2} are the variances of the test statistics for Poisson respectively column–multinomial sampling approximative. The assumptions (MD1) to (MD3) are conditions concerning the (not modelled) **m**arginal **d**istributions respectively the way of grouping.

Except for their different stochastic behaviour, the conditions for both distribution models almost coincide. (MD3) will only be required for column multinomial sampling, which can, due to (RC2), also be regarded as a condition concerning the marginal distribution ($\pi^n_{jk|D}(\theta_0) = \frac{\mu^n_{jk}}{n_k} = \frac{\mu^n_{j+}}{n_k} \cdot \pi^n_{jk|C}(\theta_0)$ with $\pi^n_{jk|C}(\theta_0) \geq \epsilon > 0$).

Let now finally the assumptions (MD1) and (MD2) be illustrated, which can also be expressed in terms of sample means. For this purpose, consider the p–th mean

$$M_p(x.) = \Big(\frac{1}{J^n}\sum_{j=1}^{J^n} x_j^p\Big)^{\frac{1}{p}} \quad \text{with } x. = (x_1, \ldots, x_J)^T \in \mathbf{R}^J,\ p \in \mathbf{R}.$$

For example, the case $p = -1$ denotes the harmonic, $p = 1$ the arithmetic and $p = 2$ the quadratic mean. Using this notation, condition (MD2), i.e. $\frac{1}{\sqrt{J^n}}\sum_{j=1}^{J^n}\frac{1}{\sqrt{\mu^n_{j+}}} \longrightarrow 0$, can be equivalently stated as follows

$$\Big(\frac{1}{\sqrt{J^n}}\sum_{j=1}^{J^n}\frac{1}{\sqrt{\mu^n_{j+}}}\Big)^{-2} = \frac{1}{J^n}\Big(\frac{1}{J^n}\sum_{j=1}^{J^n}(\mu^n_{j+})^{-\frac{1}{2}}\Big)^{-2} = \frac{1}{J^n}M_{-\frac{1}{2}}(\mu^n_{\cdot+}) \longrightarrow \infty. \tag{2.7}$$

In order to meet (MD2), thus a sufficiently large increase of the $-\frac{1}{2}$–th mean is required. If the terms of the sum are positive, the p–th mean M_p is increasing in p (see Kendall/Stewart (1969), chapter 2). This yields

$$\frac{\mu^n_{++}}{(J^n)^2} = \frac{1}{J^n}\cdot\frac{1}{J^n}\sum_{j=1}^{J^n}\mu^n_{j+} = \frac{1}{J^n}M_1(\mu^n_{\cdot+}) \geq \frac{1}{J^n}M_p(\mu^n_{\cdot+}) \geq \frac{1}{J^n}M_{-\frac{1}{2}}(\mu^n_{\cdot+}) \tag{2.8}$$

for all $p \in (-\frac{1}{2}, 1)$. (MD2) thus requires a fast increase of all p–th means with $p \geq -\frac{1}{2}$, and due to the choice of n especially $\frac{(J^n)^2}{n} \longrightarrow 0$.

In regard to condition (MD1), i.e. $\frac{1}{\sqrt{J^n n}}\sum_{j=1}^{J^n}\sqrt{\mu^n_{j+}} \longrightarrow 0$, the particular property $M_{\frac{1}{2}}(\mu^n_{\cdot+}) \leq M_1(\mu^n_{\cdot+})$ gives already the boundedness of the term considered there:

$$\begin{aligned}\Big(\frac{1}{\sqrt{J^n n}}\sum_{j=1}^{J^n}\sqrt{\mu^n_{j+}}\Big)^2 &= \frac{J^n}{\mu^n_{++}}\cdot\Big(\frac{1}{J^n}\sum_{j=1}^{J^n}(\mu^n_{j+})^{\frac{1}{2}}\Big)^2 \\ &= (M_1(\mu^n_{\cdot+}))^{-1}\cdot M_{\frac{1}{2}}(\mu^n_{\cdot+})\end{aligned}$$

$$
\begin{aligned}
&= \left(\frac{1}{J^n} M_1(\mu^n_{\cdot+})\right)^{-1} \cdot \frac{1}{J^n} M_{\frac{1}{2}}(\mu^n_{\cdot+}) \\
&\leq 1.
\end{aligned}
$$

(MD1), which requires convergence to zero, thus demands the increase of the arithmetic mean $M_1(\mu^n_{\cdot+})$ to be faster than that of the $\frac{1}{2}$–th mean. Hence it just strengthens the requirements $(\frac{1}{J^n} M_1(\mu^n_{\cdot+}))^{-1} \to 0$ and $\frac{1}{J^n} M_{\frac{1}{2}}(\mu^n_{\cdot+}) \to \infty$ implied by (MD2) (see (2.7) and (2.8)).

In view of applications, (MD1) and (MD2) are approximatively met, if — first of all — the total sample size is much larger than the number of groups. Further, the scaled $-\frac{1}{2}$–th mean $\frac{1}{J^n} M_{-\frac{1}{2}}(\mu^n_{\cdot+}) = \frac{1}{J^n}\left(\frac{1}{J^n}\sum_{j=1}^{J^n}(\mu^n_{j+})^{-\frac{1}{2}}\right)^{-2}$ has to be large, which can in the application case be checked inserting the estimated expectations of the row sums Y^n_{j+} respectively X^n_{j+}. Finally, the arithmetic mean $M_1(\mu^n_{\cdot+}) = \frac{1}{J^n}\sum_{j=1}^{J^n}\mu^n_{j+} = \frac{\mu^n_{++}}{J^n}$ respectively $\frac{n}{J^n}$ should be notably larger than the $\frac{1}{2}$–th mean $M_{\frac{1}{2}}(\mu^n_{\cdot+}) = \left(\frac{1}{J^n}\sum_{j=1}^{J^n}(\mu^n_{j+})^{\frac{1}{2}}\right)^2$.

3. Weak Convergence for Column–Multinomial Sums

In this chapter, the limiting normal distribution of column multinomial sums for the increasing cells approach is in the centre of interest. Statements on this matter are required, for example, in order to prove the asymptotic normality of the approximated goodness–of–fit statistic and will here be studied in general. The considerations of this chapter are merely theoretical and not concerned with modelling or model fitting at all.
The main difficulty in the derivation of such limiting results now lies in the stochastic dependencies within each column, which prevent an application of the central limit theorem. To fulfil these conditions an indirect approach is chosen. It goes back to Morris (1975), who proved a "fundamental lemma", giving conditions for the asymptotic normality of multinomial sums. It is essential for his approach that the basic property of the multinomial distribution be conditional Poisson, i.e. that it coincides with the distribution of a vector of stochastic independent Poisson variables conditional on the sum equalling the multinomial sample size.
This particular method will be illustrated in the first section. The second section then will discuss the approach more precisely, formulating statements to provide the theoretical background. Finally, a generalization of Morris' lemma for the column–multinomial model will be given.

3.1 Illustration of the Approach

Let in the following be $Y = (Y_{jk})_{j,k}$ ($j = 1, \ldots, J$, $k = 1, \ldots, K$) a $J \times K$ column–multinomial distributed contingency table and X the analogously defined Poisson table. For these considerations, a simpler notation than elsewhere can be chosen: For each $k \in \{1, \ldots, K\}$ consider $Y_{\cdot k} = (Y_{1k}, \ldots, Y_{Jk})^T \sim Multi_J(n_k, p_{\cdot k})$ with probability vector $p_{\cdot k} = (p_{1k}, \ldots, p_{Jk})^T$ and stochastically independent columns $Y_{\cdot 1}, \ldots, Y_{\cdot K}$. In the Poisson distribution model let the entries X_{jk}, $j = 1, \ldots, J$, $k = 1, \ldots, K$ be stochastically independent $Poisson(\mu_{jk})$ distributed random variables having the same expectations μ_{jk} as the corresponding column–multinomial variables, i.e. $\mu_{jk} := n_k p_{jk}$ for all j, k.

The object of interest is the asymptotic distribution of column–multinomial sums

$\sum_{j=1}^J \sum_{k=1}^K f_{jk}(Y_{jk})$, $f_{jk} : \mathbf{N}_0 \to \mathbf{R}$ or more generally of sums of the type $\sum_{j=1}^J f_j(Y_{j\cdot})$, $f_j : \mathbf{N}_0^K \to \mathbf{R}$ if the number of classes J increases. These will be considered in the following, when Morris' method is illustrated without going into mathematical detail. In contrast to Morris, this will already be done for the simple generalization from the multinomial to the column–multinomial distribution.

Starting point is the fact that the column–multinomial distribution model is conditional Poisson, i.e.

$$\mathcal{L}(Y) = \mathcal{L}\Big(X\Big|\sum_{j=1}^J X_{j1} = n_1, \ldots, \sum_{j=1}^J X_{jK} = n_K\Big).$$

Hence, instead of studying the distribution of $\sum_{j=1}^J f_j(Y_{j\cdot})$ directly, it suffices to consider the conditional distribution $\mathcal{L}(\sum_{j=1}^J f_j(X_{j\cdot})| \sum_{j=1}^J X_{j1} = n_1, \ldots, \sum_{j=1}^J X_{jK} = n_K)$, with the advantage that under Poisson distribution only sums of stochastically independent random variables appear.

Using $s_J^2 := Var(\sum_{j=1}^J f_j(X_{j\cdot})) = \sum_{j=1}^J Var(f_j(X_{j\cdot}))$ and

$$\sum_{j=1}^J X_{jk} = n_k \Leftrightarrow \sum_{j=1}^J X_{jk} = \sum_{j=1}^J \mu_{jk} \Leftrightarrow \frac{1}{\sqrt{n_k}} \sum_{j=1}^J (X_{jk} - \mu_{jk}) = 0$$

for $k = 1, \ldots, K$ gives

$$\mathcal{L}\Big(\frac{1}{s_J} \sum_{j=1}^J f_j(Y_{j\cdot})\Big) = \mathcal{L}\Big(\frac{1}{s_J} \sum_{j=1}^J f_j(X_{j\cdot})\Big|\frac{1}{\sqrt{n_k}} \sum_{j=1}^J (X_{jk} - \mu_{jk}) = 0 \;\forall\, k = 1, \ldots, K\Big). \quad (3.1)$$

If for reasons of clarity the notation

$$U_J(Y) := \frac{1}{s_J} \sum_{j=1}^J f_j(Y_{j\cdot}), \qquad U_J(X) = \frac{1}{s_J} \sum_{j=1}^J f_j(X_{j\cdot}),$$

$$V_{kJ}(X) := \frac{1}{\sqrt{n_k}} \sum_{j=1}^J (X_{jk} - \mu_{jk}) \qquad (k = 1, \ldots, K),$$

is chosen, (3.1) can be equivalently stated:

$$\mathcal{L}\Big(U_J(Y)\Big) = \mathcal{L}\Big(U_J(X)\Big|V_{kJ}(X) = 0 \quad \forall\, k = 1, \ldots, K\Big).$$

The stochastic independence of the Poisson variables yields for the conditioning sums

$$E(V_{kJ}(X)) = 0, \quad Var(V_{kJ}(X)) = 1 \quad \text{for all } k \in \{1, \ldots, K\},$$

$$Cov(V_{kJ}(X), V_{k'J}(X)) = 0 \quad \text{for all } k, k' \in \{1, \ldots, K\}, k \neq k',$$

and by definition of $U_J(X)$ and s_J holds $Var(U_J(X)) = 1$. If further

$$E(U_J(X)) = 0, \quad Cov(U_J(X), V_{kJ}(X)) = 0 \; \forall \, k \in \{1, \ldots, K\} \tag{3.2}$$

is fulfilled, then under known assumptions (e.g. if each variable meets the Lindeberg Condition for the central limit theorem) follows:

$$(U_J(X), V_{1J}(X), \ldots, V_{KJ}(X)) \xrightarrow{\mathcal{L}} N_{K+1}(0, I_{K+1}) \qquad (J \to \infty) \tag{3.3}$$

with I_{K+1} being the $(K+1) \times (K+1)$ unity matrix. If now the conditional distribution $\mathcal{L}(U_J(X)|V_{kJ}(X) = 0 \; \forall \, k = 1, \ldots, K)$ converges to the conditional limiting distribution, the statistic of interest, scaled with the Poisson variance, is asymptotically normal:

$$\begin{aligned} \mathcal{L}\Big(\frac{1}{s_J}\sum_{j=1}^{J} f_j(Y_{j\cdot})\Big) &= \mathcal{L}\big(U_J(Y)\big) \\ &= \mathcal{L}\Big(U_J(X)\Big|V_{kJ}(X) = 0 \; \forall \, k = 1, \ldots, K\Big) \stackrel{J\to\infty}{\longrightarrow} N(0,1). \end{aligned} \tag{3.4}$$

The whole approach will be studied more precisely in the next section. Apart from its general theoretical background, a generalization of Morris' fundamental lemma will be given which is based on this method and formulates conditions for the asymptotic multivariate normality (cp. (3.3)) and the convergence of the conditional distribution to the conditional limit. There the validity of (3.2), which is a condition concerning the functions f_j, will be assumed. This does not hold in general, it can, however, always be accomplished by a suitable transformation of the given functions. Because such a transformation will be necessary in the later application case, the method will now be shortly described. For this purpose let the given functions (which do not fulfil (3.2)) be denoted by g_j $(j = 1, \ldots, J)$. Let be $z = (z_{jk})_{j,k}$ an arbitrary $J \times K$ table with nonnegative entries and for $j \in \{1, \ldots, J\}$ let the function $f_j : \mathbf{N}_0^K \to \mathbf{R}$ be defined as follows:

$$\begin{aligned} f_j(z_{j\cdot}) &:= g_j(z_{j\cdot}) - E(g_j(X_{j\cdot})) - \sum_{k=1}^{K} \gamma_{kJ}(z_{jk} - \mu_{jk}), \\ \gamma_{kJ} &:= \frac{1}{\mu_{+k}} \sum_{j=1}^{J} Cov(g_j(X_{j\cdot}), X_{jk}). \end{aligned}$$

Next, let the notation be as before, i.e. $s_J^2 := Var(\sum_{j=1}^{J} f_j(X_{j\cdot}))$ etc.. With such chosen f_j, all considerations apply, especially (3.2) holds, e.g. $(k = 1, \ldots, K)$:

$$\begin{aligned} &Cov(U_J(X), V_{kJ}(X)) \\ = \; &\frac{1}{s_J}\frac{1}{\sqrt{\mu_{+k}}} Cov\Big(\sum_{j=1}^{J} g_j(X_{j\cdot}) - \sum_{j=1}^{J}\sum_{k=1}^{K} \gamma_{kJ} X_{jk}, \sum_{j=1}^{J} X_{jk}\Big) \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{s_J}\frac{1}{\sqrt{\mu_{+k}}}\Big(\sum_{j=1}^{J} Cov(g_j(X_{j\cdot}), X_{jk}) - \sum_{j=1}^{J} Cov\Big(\sum_{k=1}^{K}\gamma_{kJ}X_{jk}, X_{jk}\Big)\Big)\\
&= \frac{1}{s_J}\frac{1}{\sqrt{\mu_{+k}}}\Big(\sum_{j=1}^{J} Cov(g_j(X_{j\cdot}), X_{jk}) - \gamma_{kJ}\mu_{+k}\Big)\\
&= \frac{1}{s_J}\frac{1}{\sqrt{\mu_{+k}}}\Big(\sum_{j=1}^{J} Cov(g_j(X_{j\cdot}), X_{jk}) - \mu_{+k}\frac{1}{\mu_{+k}}\sum_{j=1}^{J} Cov(g_j(X_{j\cdot}), X_{jk})\Big)\\
&= 0. \qquad (3.5)
\end{aligned}$$

Using (3.4) and

$$\begin{aligned}
\sum_{j=1}^{J} f_j(Y_{j\cdot}) &= \sum_{j=1}^{J} g_j(Y_{j\cdot}) - \sum_{j=1}^{J} E(g_j(X_{j\cdot})) - \sum_{k=1}^{K}\gamma_{kJ}\Big(\sum_{j=1}^{J} Y_{jk} - \sum_{j=1}^{J}\mu_{jk}\Big)\\
&= \sum_{j=1}^{J} g_j(Y_{j\cdot}) - \sum_{j=1}^{J} E(g_j(X_{j\cdot}))
\end{aligned}$$

gives the asymptotic normality for the column–multinomial sum in question:

$$\mathcal{L}\Big(\frac{1}{s_J}\Big(\sum_{j=1}^{J} g_j(Y_{j\cdot}) - E\Big(\sum_{j=1}^{J} g_j(X_{j\cdot})\Big)\Big)\Big) = \mathcal{L}\Big(\frac{1}{s_J}\sum_{j=1}^{J} f_j(Y_{j\cdot})\Big) \overset{J\to\infty}{\longrightarrow} N(0,1). \quad (3.6)$$

Hence the standardization terms are the expected value of the original and the variance of the corrected sum under Poisson distribution. This variance can be specified more precisely:

$$\begin{aligned}
s_J^2 &= Var\Big(\sum_{j=1}^{J} f_j(X_{j\cdot})\Big)\\
&= Var\Big(\sum_{j=1}^{J} g_j(X_{j\cdot}) - \sum_{j=1}^{J}\sum_{k=1}^{K}\gamma_{kJ}(X_{jk}-\mu_{jk})\Big)\\
&= Var\Big(\sum_{j=1}^{J} g_j(X_{j\cdot})\Big) + Var\Big(\sum_{j=1}^{J}\sum_{k=1}^{K}\gamma_{kJ}(X_{jk}-\mu_{jk})\Big)\\
&\quad -2Cov\Big(\sum_{j=1}^{J} g_j(X_{j\cdot}), \sum_{j=1}^{J}\sum_{k=1}^{K}\gamma_{kJ}(X_{jk}-\mu_{jk})\Big)\\
&= Var\Big(\sum_{j=1}^{J} g_j(X_{j\cdot})\Big) + \sum_{j=1}^{J}\sum_{k=1}^{K}\gamma_{kJ}^2 Var(X_{jk}) - 2\sum_{j=1}^{J}\sum_{k=1}^{K}\gamma_{kJ} Cov(g_j(X_{j\cdot}), X_{jk})\\
&= Var\Big(\sum_{j=1}^{J} g_j(X_{j\cdot})\Big) + \sum_{k=1}^{K}\gamma_{kJ}^2\sum_{j=1}^{J}\mu_{jk} - 2\sum_{k=1}^{K}\gamma_{kJ}\sum_{j=1}^{J} Cov(g_j(X_{j\cdot}), X_{jk})
\end{aligned}$$

$$= \quad Var\Big(\sum_{j=1}^{J} g_j(X_{j\cdot})\Big) + \sum_{k=1}^{K} \gamma_{kJ}^2 \mu_{+k} - 2\sum_{k=1}^{K} \gamma_{kJ}^2 \mu_{+k}$$

$$= \quad Var\Big(\sum_{j=1}^{J} g_j(X_{j\cdot})\Big) - \sum_{k=1}^{K} \gamma_{kJ}^2 \mu_{+k}. \tag{3.7}$$

3.2 Theoretical Results

To provide the theoretical background, a theorem from Steck (1957) concerning the convergence of conditional distributions in general will be given. Since Steck proves his result considering conditional characteristic functions and in particular requires a certain equicontinuity condition to hold, this will be defined first. After Steck's theorem, a simple lemma for asymptotic multivariate normality will be presented.

Definition 3.1 *Consider metric spaces M_1 and M_2. The family of functions $\psi^n : M_1 \to M_2$, $v \mapsto \psi^n(v)$ ($n \in \mathbf{N}$) is called* **uniformly equicontinuous on bounded sets**, *if for every $\epsilon > 0$ and every bounded set $C \subset M_1$ there exists a $\delta > 0$ such that for all $n \in \mathbf{N}$ holds:*

$$v, v' \in C, \|v, v'\| < \delta \Rightarrow \|\psi^n(v), \psi^n(v')\| < \epsilon. \tag{3.8}$$

□

Considering the special case $\psi^n : \mathbf{R}^K \to \mathbf{R}$ (3.8) reduces to

$$v, v + h \in C \subset \mathbf{R}^K, \|h\| < \delta \Rightarrow |\psi^n(v + h) - \psi^n(v)| < \epsilon.$$

Theorem 3.2 (Steck (1957), Thm. 2.1) *Let (U^n, V^n) be a sequence of random vectors ($U^n \in \mathbf{R}, V^n \in \mathbf{R}^K$) with $(U^n, V^n) \xrightarrow{\mathcal{L}} (U, V)$ ($n \to \infty$) and*

$$\begin{aligned} \psi^n : \mathbf{R}^K \times \mathbf{R} &\to \mathbf{R} \\ (v, t) &\mapsto \psi^n(v, t) := E(e^{itU^n} | V^n = v) \end{aligned}$$

a version of the conditional characteristic function. If for every fixed $t \in \mathbf{R}$ the family of functions

$$\{\psi^n(-, t) | n \in \mathbf{N}\}$$

is uniformly equicontinuous on bounded sets, then it holds

$$\mathcal{L}(U^n | V^n = v) \to \mathcal{L}(U | V = v) \qquad (n \to \infty).$$

□

Lemma 3.3 *Assume S^n is a M–dimensional random vector, $(S^n)^T = (S_1^n, \ldots, S_M^n)$ $= \sum_{i=1}^n X_i^n$ with $X_i^n = (X_{i1}^n, \ldots, X_{iM}^n)$ and the vectors $X_1^n, \ldots, X_M^n$ being stochastically independent. For all $i \in \{1, \ldots, n\}$, $m \in \{1, \ldots, M\}$, $n \in \mathbf{N}$ resp. $m, m' \in$*

*$\{1,\ldots,M\}$, $n \in \mathbf{N}$ let be $E(X^n_{im}) = 0$ and $E(S^n_m \cdot S^n_{m'}) = \delta_{mm'}$ with $\delta_{mm'}$ being the Kronecker-delta. Suppose further that all components of S^n, i.e. $S^n_m = \sum_{i=1}^n X^n_{im}$, fulfil the Feller Condition for the central limit theorem ("**u**niformly **a**symptotically **n**egligible" condition) and are asymptotically normal distributed, i.e.*

$$\max_{1\leq i\leq n} Var(X^n_{im}) \overset{n\to\infty}{\longrightarrow} 0 \quad \text{and} \quad \mathcal{L}(S^n_m) \overset{n\to\infty}{\longrightarrow} N(0,1) \quad \text{for all } m \in \{1,\ldots,M\}.$$

Then it holds (I_M is the $M \times M$ unity matrix):

$$S^n \xrightarrow{\mathcal{L}} N(0, I_M) \quad \text{with} \quad n \to \infty.$$

□

The simple proof can be found in Morris (1975, Lemma 2.1). It is based upon the method of Cramèr–Wold (reduction of dimension) and makes use of the fact that each component of S^n meets the Lindeberg Condition for the central limit theorem, which is implied by the Feller Condition and the asymptotic normality.

With the necessary theoretical statements given, the following lemma, which is a simple generalization of Morris' "fundamental lemma" (1975, Lemma 2.2) from the multinomial to the column–multinomial distribution, states the exact conditions needed for the approach described in section 3.1. The conditions (3.9) to (3.11) assure the assumptions for Lemma 3.3 and hence the common distribution of the statistic of interest under Poisson distribution and the conditioning sums to be asymptotically normal (cp. 3.3). Condition (3.12) implies Steck's criterion for the convergence of the conditional distribution in the special case considered here.

To illustrate the method without exact formulation of the assumptions (sec. 3.1), it was sufficient to consider an increasing cells approach ($J \to \infty$). For asymptotic normality though, an increase of the total sample size will definitely be necessary. With regard to the later application, and in contrast to Morris this will now be the primary asymptotics, indicated by the running index n, entailing an increase of the number of classes J^n.

Lemma 3.4 *Consider a column–multinomial distributed $J^n \times K$ contingency table $Y^n = (Y^n_{jk})_{j,k}$, i.e. the columns $Y^n_{\cdot 1},\ldots,Y^n_{\cdot K}$ are stochastically independent and multinomial distributed, $Y^n_{\cdot k} = (Y^n_{1k},\ldots,Y^n_{J^n k})^T \sim Multi_{J^n}(n_k, p^n_{\cdot k})$ with probability vector $p^n_{\cdot k} = (p^n_{1k},\ldots,p^n_{J^n k})^T$ for $k = 1,\ldots,K$. Analogously, let $X^n = (X^n_{jk})_{j,k}$ be a Poisson distributed table with stochastically independent entries X^n_{jk} ($j = 1,\ldots,J^n$, $k = 1,\ldots,K$) and especially $E(X^n_{jk}) = \mu^n_{jk} = n_k p^n_{jk}$.*
For every $n := \sum_{k=1}^K n_k$, $j \in \{1,\ldots,J^n\}$ let functions $f^n_j : \mathbf{N}_0^K \to \mathbf{R}$ be given with $E(f^n_j(X^n_{j\cdot})) = 0$, $Cov(\sum_{j=1}^{J^n} f^n_j(X^n_{j\cdot}), \sum_{j=1}^{J^n} X^n_{jk}) = \sum_{j=1}^{J^n} Cov(f^n_j(X^n_{j\cdot}), X^n_{jk}) = 0$ for all $k = 1,\ldots,K$.

Besides the asymptotics $n \to \infty$, consider $J^n \to \infty$ and $n_k \to \infty$ for each $k = 1, \ldots, K$ and assume the following conditions to hold:

$$\lim_{n\to\infty} \max_{1\le j\le J^n} p^n_{jk} = 0 \qquad \textit{for all } k = 1, \ldots, K, \tag{3.9}$$

$$\frac{1}{\left(\sum_{j=1}^{J^n} Var(f_j^n(X_{j\cdot}^n))\right)^{1/2}} \sum_{j=1}^{J^n} f_j^n(X_{j\cdot}^n) \xrightarrow{\mathcal{L}} N(0,1) \qquad (n \to \infty), \tag{3.10}$$

$$\lim_{n\to\infty} \frac{1}{\sum_{j=1}^{J^n} Var(f_j^n(X_{j\cdot}^n))} \cdot \max_{1\le j\le J^n} Var(f_j^n(X_{j\cdot}^n)) = 0 \quad \textit{(Feller Condition)}. \tag{3.11}$$

Further suppose

$$\lim_{h\to 0} \sup_n \sup_{v^n} \frac{1}{\sum_{j=1}^{J^n} Var(f_j^n(X_{j\cdot}^n))} E\Big(\Big(\sum_{j=1}^{J} f_j^n(L_{j\cdot}^n + M_{j\cdot}^n) - f_j^n(L_{j\cdot}^n)\Big)^2\Big) = 0 \tag{3.12}$$

to be true, with $v^n = (v_1^n, \ldots, v_K^n)^T$ being a bounded sequence and $L^n = (L^n_{jk})_{j,k}$, $M^n = (M^n_{jk})_{j,k}$ column–multinomial distributed contingency tables, for which holds $L^n_{\cdot k} = (L^n_{1k}, \ldots, L^n_{J^n k})^T \sim Multi_{J^n}(n_k + v^n_k\sqrt{n_k}, p^n_{\cdot k})$ and $M^n_{\cdot k} \sim Multi_{J^n}(h_k\sqrt{n_k}, p^n_{\cdot k})$ ($h = (h_1, \ldots, h_K)^T \in \mathbf{R}^K$) with stochastically independent columns $L^n_{\cdot 1}, \ldots, L^n_{\cdot K}$, $M^n_{\cdot 1}, \ldots, M^n_{\cdot K}$. In particular, let v^n and h be such that the sizes $l_k := n_k + v^n_k\sqrt{n_k}$ and $m_k := h_k\sqrt{n_k}$ are nonnegative integers for every $k \in \{1, \ldots, K\}$. Then it holds:

$$\frac{1}{\left(\sum_{j=1}^{J^n} Var(f_j^n(X_{j\cdot}^n))\right)^{1/2}} \sum_{j=1}^{J^n} f_j^n(Y_{j\cdot}^n) \xrightarrow{\mathcal{L}} N(0,1) \qquad (n \to \infty).$$

Proof:
Defining $V^n_k := 1/\sqrt{n_k}\sum_{j=1}^{J^n}(X^n_{jk} - \mu^n_{jk})$ for each $k \in \{1, \ldots, K\}$ clearly gives $E(V^n_k) = 0$, $Var(V^n_k) = 1$. Further condition (3.9) and $n_k \to \infty$ assure for V^n_k Ljapounov's condition for the central limit theorem

$$\sum_{j=1}^{J^n} E\Big(\Big(\frac{X^n_{jk} - \mu^n_{jk}}{\sqrt{n_k}}\Big)^4\Big) = \sum_{j=1}^{J^n} \frac{\mu^n_{jk} + 3(\mu^n_{jk})^2}{n_k^2} = \frac{1}{n_k} + 3\sum_{j=1}^{J^n}(p^n_{jk})^2 \le \frac{1}{n_k} + 3 \max_{1\le j\le J^n} p^n_{jk} = o(1)$$

for $n \to \infty$. Since the Ljapounov implies the Lindeberg condition, which is equivalent to the validity of the Feller Condition and the asymptotic normality of the sum, it follows

$$\max_{1\le j\le J^n} Var\Big(\frac{X^n_{jk} - \mu^n_{jk}}{\sqrt{n_k}}\Big) = \max_{1\le j\le J^n} p^n_{jk} \overset{n\to\infty}{\longrightarrow} 0 \quad \text{and} \quad \mathcal{L}(V^n_k) \overset{n\to\infty}{\longrightarrow} N(0,1).$$

Hence, condition (3.9) is the Feller Condition directed to the terms of each sum V^n_k. The variable

$$U^n := \frac{1}{s^n}\sum_{j=1}^{J^n} f_j^n(X_{j\cdot}^n), \quad s^{n2} := \sum_{j=1}^{J^n} Var(f_j^n(X_{j\cdot}^n)),$$

with $E(U^n) = 0$ and $Var(U^n) = 1$, is also by assumption asymptotically normal and meets the Feller Condition (cond. (3.9),(3.10)). The stochastic independence of the columns gives further $Cov(V_k^n, V_{k'}^n) = 0$ for all $k, k' \in \{1, \ldots, K\}, k \neq k'$, and assumption $Cov(\sum_{j=1}^{J^n} f_j^n(X_{j\cdot}^n), \sum_{j=1}^{J^n} X_{jk}^n) = 0$ yields $Cov(U^n, V_k^n) = 0$ for all $k \in \{1, \ldots, K\}$. Hence all conditions of Lemma 3.3 are met, thus giving

$$(U^n, V_1^n, \ldots, V_K^n)^T \xrightarrow{\mathcal{L}} N_{K+1}(0, I_{K+1}) \quad (n \to \infty).$$

If now Steck's equicontinuity condition from Theorem 3.2 holds, which is sufficient for the convergence of the conditional distribution to the conditional limiting distribution, i.e.

$$\mathcal{L}(U^n | V_k^n = 0 \;\forall\, k = 1, \ldots, K) \overset{n\to\infty}{\longrightarrow} N(0,1),$$

then the result follows from

$$\begin{aligned}
&\mathcal{L}(U^n | V_k^n = 0 \;\forall\, k = 1, \ldots, K) \\
= \;& \mathcal{L}\Big(\frac{1}{s^{n2}} \sum_{j=1}^{J^n} f_j^n(X_{j\cdot}^n) \Big| \sum_{j=1}^{J^n} X_{jk}^n = n_k \;\forall\, k = 1, \ldots, K\Big) \\
= \;& \mathcal{L}\Big(\frac{1}{s^{n2}} \sum_{j=1}^{J^n} f_j^n(Y_{j\cdot}^n)\Big).
\end{aligned}$$

Now it holds $V_k^n = v_k^n \Leftrightarrow \frac{1}{\sqrt{n_k}} \sum_{j=1}^{J^n} (X_{jk}^n - \mu_{jk}^n) = v_k^n \Leftrightarrow \sum_{j=1}^{J^n} X_{jk}^n = n_k + v_k^n \sqrt{n_k}$ and hence by definition of L^n

$$\begin{aligned}
&\mathcal{L}(U^n | V_k^n = v_k^n \;\forall\, k = 1, \ldots, K) \\
= \;& \mathcal{L}\Big(\frac{1}{s^{n2}} \sum_{j=1}^{J^n} f_j^n(X_{j\cdot}^n) \Big| \sum_{j=1}^{J^n} X_{jk}^n = n_k + v_k^n \sqrt{n_k} \;\forall\, k = 1, \ldots, K\Big) \\
= \;& \mathcal{L}\Big(\frac{1}{s^{n2}} \sum_{j=1}^{J^n} f_j^n(L_{j\cdot}^n)\Big).
\end{aligned}$$

Analogous argumentation gives

$$\mathcal{L}(U^n | V_k^n = v_k^n + h_k \;\forall\, k = 1, \ldots, K) = \mathcal{L}\Big(\frac{1}{s^{n2}} \sum_{j=1}^{J^n} f_j^n(L_{j\cdot}^n + M_{j\cdot}^n)\Big).$$

Let now, for reasons of brevity, be defined $S_1^n := \frac{1}{s^{n2}} \sum_{j=1}^{J^n} f_j^n(L_{j\cdot}^n + M_{j\cdot}^n)$, $S_2^n := \frac{1}{s^{n2}} \sum_{j=1}^{J^n} f_j^n(L_{j\cdot}^n)$ and let ψ^n denote a version of the characteristic function of the conditional distribution of $U^n | V_\cdot^n$. Then it holds

$$\begin{aligned}
&|\psi^n(v_1^n + h_1, \ldots, v_K^n + h_K, t) - \psi^n(v_1^n, \ldots, v_K^n, t)| \\
= \;& |E(e^{itS_1^n}) - E(e^{itS_2^n})| \\
= \;& |E(e^{itS_1^n} - e^{itS_2^n})|
\end{aligned}$$

$$
\begin{aligned}
&\leq \ E(|e^{itS_1^n} - e^{itS_2^n}|) \\
&= \ E(|e^{it(S_1^n - S_2^n)} - 1|) \\
&\leq \ E(|t||S_1^n - S_2^n|) && (3.13) \\
&= \ \frac{|t|}{s^n} E\Big(\Big|\sum_{j=1}^{J^n} f_{j\cdot}^n(L_{j\cdot}^n + M_{j\cdot}^n) - \sum_{j=1}^{J^n} f_{j\cdot}^n(L_{j\cdot}^n)\Big|\Big) \\
&\leq \ |t| \sqrt{\frac{1}{s^{n2}} E\Big(\Big(\sum_{j=1}^{J^n} f_{j\cdot}^n(L_{j\cdot}^n + M_{j\cdot}^n) - \sum_{j=1}^{J^n} f_{j\cdot}^n(L_{j\cdot}^n)\Big)^2\Big)}, && (3.14)
\end{aligned}
$$

with the last inequality following from Cauchy–Schwarz and (3.13) from Billingsley (1986), p. 353, (26.4_0):

$$|e^{it(S_1^n - S_2^n)} - 1| \leq \min\{|t(S_1^n - S_2^n)|, 2\} \leq |t(S_1^n - S_2^n)| \leq |t||S_1^n - S_2^n|.$$

(3.14) and assumption (3.12) finally yield

$$\lim_{h \to 0} \sup_n \sup_{v^n} |\psi^n(v_1^n + h_1, \ldots, v_K^n + h_K, t) - \psi^n(v_1^n, \ldots, v_K^n, t)| = 0,$$

i.e. Steck's equicontinuity condition concerning the family of conditional characteristic functions is met. □

4. Properties of the Poisson Distribution

In this chapter, some preliminary results for later proofs will be provided. The first section treats the central Poisson moments and the expected values of numerical functions under Poisson distribution. In particular, useful statements concerning continuity and differentiability are given and the asymptotic behaviour is investigated. The application of these results to expectations of the distance function a_λ in the following section, will provide some bounding statements, which will be used throughout the proofs of the following chapters.

4.1 Expected Values

Theorem 4.1 *Consider a Poisson distributed random variable X with expected value $\mu \in \mathbf{R}^+$ and a function $f : \mathbf{N_0} \times \mathbf{R}^+ \to \mathbf{R}$, $(x, \mu) \mapsto f(x, \mu)$.*

a) *Suppose that f is continuous in μ and for every $[a, b] \subset \mathbf{R}^+$ there exist constants $\alpha, t \in \mathbf{R}^+$, so that for all $\mu \in [a, b], x \in \mathbf{R}_0^+$ holds*

$$|f(x, \mu)| \leq \alpha e^{tx}, \tag{4.1}$$

then $F(\mu) := E(f(X, \mu))$ exists and is continuous in μ.

b) *If additional to the requirements in a) the following assumptions are true, i.e.*

$$\text{for each } x \in \mathbf{N}_0 \text{ exists } \frac{\partial}{\partial \mu} f(x, \mu) \text{ for all } \mu \in \mathbf{R}^+, \tag{4.2}$$

for each $[a, b] \subset \mathbf{R}^+$ there exist constants $\alpha, t \in \mathbf{R}^+$,
so that for all $\mu \in [a, b], x \in \mathbf{R}_0^+$ holds

$$|\frac{\partial}{\partial \mu} f(x, \mu)| \leq \alpha e^{tx}, \tag{4.3}$$

then $E(f(X, \mu))$ is differentiable in $\mu \in \mathbf{R}^+$ with derivative

$$\frac{\partial}{\partial \mu} E(f(X, \mu)) = E(\frac{\partial}{\partial \mu} f(X, \mu)) + \frac{1}{\mu} Cov(f(X, \mu), X).$$

c) *Suppose that the assumptions in a) and b) hold and that further* $\frac{\partial}{\partial\mu}f(x,\mu)$ *is continuous in* μ*. Then* $E(f(X,\mu))$ *is continuously differentiable in* μ *on* $\mathbf{R}^+$.

Proof:

a) Let ϵ be the counting measure and $p(x|\mu)$ the density of the Poisson distribution, i.e. $p(x|\mu) = e^{-\mu}\mu^x/x!$ for $x \in \mathbf{N}_0$. Now existence and continuity in μ of

$$F(\mu) = E(f(X,\mu)) = \int f(x,\mu)p(x|\mu)\epsilon dx$$

have to be shown. Therefore let $[a,b]$ be an arbitrary compact subset of $\mathbf{R}^+$. Referring to Billingsley (1986), Thm. 16.8, $F(\mu)$ is continuous in $\mu \in [a,b]$, if $f(x,\mu)\cdot p(x|\mu)$ is continuous in μ, which is given by assumption, and if

$$|f(x,\mu)\cdot p(x|\mu)| \leq g(x) \quad \epsilon\text{–integrable}$$

holds. The latter in particular guarantees the existence of $F(\mu)$. Using condition (4.1), it is verified as follows:

$$\begin{aligned} |f(x,\mu)\cdot p(x|\mu)| &\leq \alpha e^{tx}\cdot p(x|\mu) \\ &\leq \sup_{\mu\in[a,b]} \alpha e^{tx}\cdot p(x|\mu) \\ &:= g(x), \end{aligned}$$

with $g(x)$ ϵ–integrable, because $\alpha e^{tx}p(x|\mu) > 0$ and the monotony of the integral yield

$$\begin{aligned} \int g(x)\epsilon dx &= \int \sup_{\mu\in[a,b]} \alpha e^{tx}p(x|\mu)\epsilon dx \\ &= \sup_{\mu\in[a,b]} \int \alpha e^{tx}p(x|\mu)\epsilon dx \\ &= \sup_{\mu\in[a,b]} \alpha e^{\mu(e^t-1)} \\ &< \infty. \end{aligned}$$

Thus existence and continuity of $F(\mu)$ are shown for every $\mu \in [a,b]$. Since $[a,b] \subset \mathbf{R}^+$ was chosen arbitrarily, the result holds for all $\mu \in \mathbf{R}^+$.

b) Keeping the notation of the proof of a) now for $\mu \in \mathbf{R}^+$

$$\frac{\partial}{\partial\mu}\int f(x,\mu)p(x|\mu)\epsilon dx = \int \frac{\partial}{\partial\mu}(f(x,\mu)p(x|\mu))\epsilon dx. \tag{4.4}$$

has to be verified. This result combined with

$$\begin{aligned} &\frac{\partial}{\partial\mu}(f(x,\mu)p(x|\mu)) \\ &= \frac{\partial}{\partial\mu}f(x,\mu)\cdot p(x|\mu) + f(x,\mu)\cdot\frac{\partial}{\partial\mu}p(x|\mu) \end{aligned}$$

$$
\begin{aligned}
&= \frac{\partial}{\partial\mu} f(x,\mu)\cdot \frac{e^{-\mu}\mu^x}{x!} + f(x,\mu)\cdot \frac{e^{-\mu}x\mu^{x-1} - e^{-\mu}\mu^x}{x!} \\
&= \frac{\partial}{\partial\mu} f(x,\mu)\cdot \frac{e^{-\mu}\mu^x}{x!} + f(x,\mu)\cdot \Big(\frac{x}{\mu}\cdot\frac{e^{-\mu}\mu^x}{x!} - \frac{e^{-\mu}\mu^x}{x!}\Big) \\
&= \frac{\partial}{\partial\mu} f(x,\mu)\cdot p(x|\mu) + f(x,\mu)\cdot\frac{x}{\mu}\cdot p(x|\mu) - f(x,\mu)\cdot p(x|\mu) \qquad (4.5)
\end{aligned}
$$

then yields the stated formula:

$$
\begin{aligned}
\frac{\partial}{\partial\mu}E(f(X,\mu)) &= E(\frac{\partial}{\partial\mu}f(X,\mu)) + E(f(X,\mu)\cdot\frac{X}{\mu}) - E(f(X,\mu) \\
&= E(\frac{\partial}{\partial\mu}f(X,\mu)) + Cov(f(X,\mu),\frac{X}{\mu}). \qquad (4.6)
\end{aligned}
$$

For the proof of (4.4), let an arbitrary subset $[a,b]\subset \mathbf{R}^+$ be given. Referring to Billingsley (1986), Thm. 16.8 b), for $\mu\in[a,b]$ a change of differentiation und integration is admissible, if the following asumptions are met:

$$\text{for each } x\in\mathbf{N}_0 \text{ exists } \frac{\partial}{\partial\mu}(f(x,\mu)p(x|\mu)) \text{ for all } \mu\in[a,b], \qquad (4.7)$$

$$\Big|\frac{\partial}{\partial\mu}(f(x,\mu)p(x|\mu))\Big| \le g(x) \quad \epsilon\text{–integrable for all } x\in\mathbf{N}_0 \text{ and } \mu\in[a,b]. \qquad (4.8)$$

Using (4.5) and the differentiability of $f(x,\mu)$ in μ on R^+ (condition (4.2)), (4.7) obviously holds. For the proof of (4.8) consider

$$
\begin{aligned}
\Big|\frac{\partial}{\partial\mu}(f(x,\mu)p(x|\mu))\Big| &= \Big|\frac{\partial}{\partial\mu}f(x,\mu) + f(x,\mu)\cdot(\frac{x}{\mu}-1)\Big|\cdot p(x|\mu) \\
&\le \Big(|\frac{\partial}{\partial\mu}f(x,\mu)| + |f(x,\mu)|\cdot|\frac{x}{\mu}-1|\Big)\cdot p(x|\mu) \\
&\le \beta e^{sx}\cdot p(x|\mu) \qquad (4.9)
\end{aligned}
$$

with suitable positive constants β and s. The last inequality holds, because for $\mu\in[a,b]$ $|\frac{\partial}{\partial\mu}f(x,\mu)|$ and $|f(x,\mu)|$ are by assumption (4.1) and (4.3) dominated by an exponential function. An integrable majorizing function for (4.9) can be obtained analogously to the proof of a), thus showing (4.8). Hence (4.7) and (4.8) yield (4.4) for $\mu\in[a,b]$. Since $[a,b]\subset\mathbf{R}^+$ was chosen arbitrarily, (4.4) holds for all $\mu\in\mathbf{R}^+$.

c) Using the results of b) only the continuity of $\frac{\partial}{\partial\mu}E(f(X,\mu))$ in μ on $\mathbf{R}^+$ remains to be shown. Considering formula (4.6), i.e.

$$\frac{\partial}{\partial\mu}E(f(X,\mu)) = E\Big(\frac{\partial}{\partial\mu}f(X,\mu) + f(X,\mu)\cdot\frac{X}{\mu} - f(X,\mu)\Big),$$

the desired continuity follows with a) if $\frac{\partial}{\partial\mu}f(x,\mu) + f(x,\mu)\cdot\frac{x}{\mu} - f(x,\mu)$ is continuous in μ and dominated by an exponential function. This however holds by assumption respectively has already been shown in the proof of b). □

Corollary 4.2 *The assumptions (4.1) and (4.3) of Theorem 4.1 concerning f resp.* $\frac{\partial}{\partial\mu}f$ *are in particular fulfilled if f resp.* $\frac{\partial}{\partial\mu}f$ *is on every compact set including* μ *majorized by a power function as follows:*

$$|\frac{\partial}{\partial\mu}f(x,\mu)|,\ |f(x,\mu)| \leq \beta + \gamma x^m, \quad \beta,\gamma \in \mathbf{R}^+, m \in \mathbf{N} \text{ constant.}$$

Proof: Using $x \geq 0$ this immediately follows from

$$\begin{aligned} \beta + \gamma x^m &\leq \beta + \gamma \cdot m! e^x \\ &\leq (\beta + \gamma \cdot m!) e^x \\ &= \alpha \cdot e^x \end{aligned}$$

with $\alpha = \beta + \gamma m! \in \mathbf{R}^+$ constant. □

The recursion formula for the central moments in part a) of the following theorem (see Johnson/Kotz (1992)) is proved using the results of Theorem 4.1. It immediately yields b), i.e. the central moments are polynomials in the expected value μ, and an easy to compute recursion formula for the coefficients (see also Appendix 8.1).

Theorem 4.3 *For each* $r \in \mathbf{N}$ *let* $\nu_r(\mu)$ *be the r–th central moment of the Poisson distribution with parameter* μ*. Then the following statements hold:*
a)

$$\nu_{r+1}(\mu) = r\mu\nu_{r-1}(\mu) + \mu\frac{d\nu_r(\mu)}{d\mu} \quad \text{for each} \quad r \in \mathbf{N},$$

b) $\nu_{2s}(\mu)$ *and* $\nu_{2s+1}(\mu)$ *are polynomials in* μ *of degree s for each* $s \in \mathbf{N}_0$*. More precisely, for each* $r \in \mathbf{N}_0$ *holds*

$$\nu_r(\mu) = \sum_{0\leq i\leq r/2} a_{r,i}\mu^i$$

with $a_{0,0} = 1$*,* $a_{r,0} = 0$ *for* $r \geq 1$*. Putting formally* $a_{r,i} = 0$ *for* $i > r/2$*, the following recursion formula for the coefficients is obtained* ($1 \leq i \leq r/2$, $r \geq 2$)*:*

$$a_{r,i} = (r-1)\cdot a_{r-2,i-1} + i\cdot a_{r-1,i} \in \mathbf{N}.$$

Proof:
a) For $x \in \mathbf{N}_0$, let $p(x|\mu)$ denote the density of the Poisson distribution with parameter μ. Then for $r \in \mathbf{N}$ holds:

$$\begin{aligned} &\nu_{r+1}(\mu) \\ = \ &r\mu\nu_{r-1}(\mu) - r\mu\nu_{r-1}(\mu) + \sum_{x=0}^{\infty}(x-\mu)^{r+1}p(x|\mu) \end{aligned}$$

$$\begin{aligned}
&= r\mu\nu_{r-1}(\mu) - r\mu\sum_{x=0}^{\infty}(x-\mu)^{r-1}p(x|\mu) + \sum_{x=0}^{\infty}\Big((x-\mu)^r x - (x-\mu)^r\mu\Big)p(x|\mu) \\
&= r\mu\nu_{r-1}(\mu) + \mu\sum_{x=0}^{\infty}\frac{\partial(x-\mu)^r}{\partial\mu}\cdot p(x|\mu) + \mu\cdot\frac{1}{\mu}\sum_{x=0}^{\infty}\Big((x-\mu)^r x - (x-\mu)^r\mu\Big)p(x|\mu) \\
&= r\mu\nu_{r-1}(\mu) + \mu\Big(\sum_{x=0}^{\infty}\frac{\partial(x-\mu)^r}{\partial\mu}\cdot p(x|\mu) + \frac{1}{\mu}\sum_{x=0}^{\infty}\Big((x-\mu)^r x - (x-\mu)^r\mu\Big)p(x|\mu)\Big) \\
&= r\mu\nu_{r-1}(\mu) + \mu\,\frac{\partial\sum_{x=0}^{\infty}(x-\mu)^r p(x|\mu)}{\partial\mu} \\
&= r\mu\nu_{r-1}(\mu) + \mu\frac{\partial\nu_r(\mu)}{\partial\mu}.
\end{aligned}$$

The change of differentiation and integration in the equation before last is granted by Theorem 4.1 b), because $(x-\mu)^r$ is differentiable in μ and for $\mu \in [a,b] \subset \mathbf{R}^+$ obviously dominated by an exponential function not depending on μ. The same applies to the derivative $\partial(x-\mu)^r/\partial\mu$.

b) The cases $r=0$ und $r=1$ ($r=2s$ and $r=2s+1$ with $s=0$) are trivial: $\nu_0(\mu) = a_{0,0}\mu^0 = a_{0,0} = 1$, $\nu_1(\mu) = a_{1,0}\mu^0 = a_{1,0} = 0$. Considering $r=2s$ and $r=2s+1$, $s=1$ part a) gives

$$\begin{aligned}
\nu_2(\mu) &= \mu\nu_0(\mu) + \mu\frac{\partial\nu_1(\mu)}{\partial\mu} = \mu, \\
\nu_3(\mu) &= 2\mu\nu_1(\mu) + \mu\frac{\partial\nu_2(\mu)}{\partial\mu} = \mu,
\end{aligned}$$

hence $\nu_2(\mu) = \sum_{i=0}^{1} a_{2,i}\mu^i$ with $a_{2,0} = 0, a_{2,1} = 1$ and $\nu_3(\mu) = \sum_{i=0}^{1} a_{3,i}\mu^i$ with $a_{3,0} = 0, a_{3,1} = 1$. In particular, the stated recursion holds ($a_{1,1} = 0$ because $a_{r,i} = 0$ for $i > r/2$):

$$\begin{aligned}
a_{2,1} &= 1\cdot a_{0,0} + 1\cdot a_{1,1} = a_{0,0} = 1, \\
a_{3,1} &= 2\cdot a_{1,0} + 1\cdot a_{2,1} = a_{2,1} = 1.
\end{aligned}$$

For the proof of the cases $r=2s$ und $r=2s+1$, $s\geq 2$, let the statement be true for $r=2(s-1)$ and $r=2(s-1)+1$. Then by a) for $r=2s$ holds:

$$\begin{aligned}
\nu_{2s}(\mu) &= (2s-1)\mu\nu_{2(s-1)}(\mu) + \mu\frac{\partial\nu_{2(s-1)+1}(\mu)}{\partial\mu} \\
&= (2s-1)\mu\sum_{i=0}^{s-1} a_{2(s-1),i}\mu^i + \mu\frac{\partial}{\partial\mu}\sum_{i=0}^{s-1} a_{2(s-1)+1,i}\mu^i \\
&= (2s-1)\sum_{i=1}^{s} a_{2(s-1),i-1}\mu^i + \sum_{i=1}^{s-1} i a_{2(s-1)+1,i}\mu^i
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{s-1} \Big((2s-1) a_{2(s-1),i-1} + i a_{2(s-1)+1,i} \Big) \mu^i + (2s-1) a_{2(s-1),s-1} \mu^s \\
&= \sum_{i=1}^{s} \Big((2s-1) a_{2(s-1),i-1} + i a_{2(s-1)+1,i} \Big) \mu^i ,
\end{aligned}
$$

with $a_{2(s-1)+1,s} = 0$ by assumption, because $\nu_{2(s-1)+1}$ is a polynomial of degree $(s-1)$ and $a_{2(s-1)+1,s-1}$ is the highest non zero coefficient. Hence for $r = 2s$ the desired formula is established:

$$
\begin{aligned}
\nu_r(\mu) &= \sum_{i=0}^{r/2} a_{r,i} \mu^i \\
\text{with} \quad a_{r,i} &= (r-1) \cdot a_{r-2,i-1} + i \cdot a_{r-1,i} \in \mathbf{N} \quad \text{for } i = 1, \ldots, r/2,\ a_{r,0} = 0.
\end{aligned}
$$

If r is odd, $r = 2s + 1$ $(s \geq 2)$, analogous argumentation yields

$$
\begin{aligned}
\nu_{2s+1}(\mu) &= 2s\mu \sum_{i=0}^{s-1} a_{2s-1,i} \mu^i + \mu \frac{\partial}{\partial \mu} \sum_{i=0}^{s} a_{2s,i} \mu^i \\
&= \sum_{i=1}^{s} 2s a_{2s-1,i-1} \mu^i + \sum_{i=1}^{s} i a_{2s,i} \mu^i \\
&= \sum_{i=1}^{s} \Big(2s a_{2s-1,i-1} + i a_{2s,i} \Big) \mu^i \\
&= \sum_{i=1}^{\frac{r-1}{2}} \Big((r-1) a_{r-2,i-1} + i a_{r-1,i} \Big) \mu^i \\
&= \sum_{i=1}^{\frac{r-1}{2}} a_{r,i} \mu^i
\end{aligned}
$$

with coefficients $a_{r,i}$ $(i = 1, \ldots, \frac{r-1}{2})$, $a_{r,0} = 0$ as before. This completes the proof. □

The following theorem will provide a useful Taylor approximation for expected values of functions of $Pois(\mu)$ distributed random variables for the asymptotics $\mu \to \infty$. Thus valuable information concerning the asymptotic order will be given, illustrated by two examples. The idea to this theorem goes back to Osius (1984), who showed a similar result for the binomial distribution $B(n, p)$ for the asymptotics $n \to \infty$ and from whom essential elements of proof could be adopted. Decisive for Theorem 4.4 is the fundamental property of the Poisson distribution stated in the preceding theorem, showing that the r–th central moments are polynomials in μ of degree $r/2$ or $r/2 - 1/2$ — the same applies in the binomial case to the highest degree of n.

Theorem 4.4 *Let X be a Poisson distributed random variable with expected value μ and $\bar{X} := \frac{X}{\mu}$ a suitable scaling. Further let a function $H : [0, \infty) \to \mathbf{R}$, $x \mapsto H(x)$*

be given, which is $(2r+2)$ times continuously differentiable on $(0,\infty)$ ($r \in \mathbf{N}_0$) and dominated by a power function: $|H(x)| \leq c(1+x^m)$ with $c \in \mathbf{R}^+$ and $m \in \mathbf{N}_0$ constant. Then for $h_r(\mu) := E(\mu^r H(\bar{X}))$ holds:

$$\Big|h_r(\mu) - \sum_{k=0}^{2r} \frac{1}{k!} H^{(k)}(1)\mu^{r-k}\nu_k(\mu)\Big| = O(\frac{1}{\mu}) \quad \textit{for} \quad \mu \to \infty.$$

Here $\nu_k(\mu)$ is the k–th central moment of the Poisson distribution with parameter μ, regarded as a function of μ.

Proof:
The main component of this proof is a Taylor expansion of H around 1. Since the expansion is only admissible on $(0,\infty)$, and in order to handle the error term, first a modification of the statement concerning the restricted expected value $h_r^\epsilon(\mu) := E(\mathbf{1}_{K_\epsilon}(\bar{X}) \cdot \mu^r H(\bar{X}))$ with $K_\epsilon = [1-\epsilon, 1+\epsilon] \subset (0,\infty)$, $\epsilon \in (0,1)$ will be proved. Thereafter, the difference between restricted and unrestricted expectation will be shown to disappear asymptotically. The modified result to begin with is

$$\Big|h_r^\epsilon(\mu) - \sum_{k=0}^{2r+1} \frac{1}{k!} H^{(k)}(1)\mu^{r-k}\nu_k^\epsilon(\mu)\Big| = O(\frac{1}{\mu}) \quad (\mu \to \infty). \tag{4.10}$$

Here

$$\nu_k^\epsilon(\mu) = E(\mathbf{1}_{K_\epsilon}(\bar{X}) \cdot (X - E(X))^k) = E(\mathbf{1}_{K_\epsilon}(\bar{X}) \cdot \mu^k(\bar{X}-1)^k)$$

is defined in accordance with $h_k^\epsilon(\mu) = E(\mathbf{1}_{K_\epsilon}(\bar{X}) \cdot \mu^k H(\bar{X}))$ choosing $H(\bar{X}) = (\bar{X}-1)^k$. Now by assumption, H is at least $(2r+2)$ times continuously differentiable on $(0,\infty)$. Thus, Taylor expansion of H around 1 on $(0,\infty)$ gives for $n = 2r+1$ (see also Dieudonné (1960), Th. 8.14.3):

$$H(\bar{X}) = \sum_{k=0}^{n} \frac{1}{k!} H^{(k)}(1)(\bar{X}-1)^k + R_n(\bar{X}) \tag{4.11}$$

$$\text{with } R_n(\bar{X}) = \frac{1}{n!}\Big(\int_0^1 (1-z)^n H^{(n+1)}(1+z(\bar{X}-1))dz\Big)(\bar{X}-1)^{n+1}, \; n \leq 2r+1.$$

Restricted on K_ϵ the first $2r+2$ derivatives of H are bounded. Hence for $\bar{X} \in K_\epsilon$, i.e. $|\bar{X}-1| \leq \epsilon$, the error term is dominated as follows:

$$|R_n(\bar{X})| \leq s_n(\epsilon) \cdot |\bar{X}-1|^{n+1} \tag{4.12}$$

$$\text{with} \quad s_n(\epsilon) = \frac{1}{n!} \sup_{|x-1|\leq\epsilon} |H^{(n+1)}(x)| < \infty.$$

Multiplying (4.11) with μ^r gives

$$\mu^r H(\bar{X}) = \sum_{k=0}^{2r+1} \frac{1}{k!} H^{(k)}(1)\mu^{r-k}(\mu\bar{X}-\mu)^k + \mu^r R_{2r+1}(\bar{X})$$

$$= \sum_{k=0}^{2r+1} \frac{1}{k!} H^{(k)}(1)\mu^{r-k}(X-\mu)^k + \mu^r R_{2r+1}(\bar{X}).$$

Taking the restricted expectation on both sides yields

$$h_r^\epsilon(\mu) - \sum_{k=0}^{2r+1} \frac{1}{k!} H^{(k)}(1)\mu^{r-k}\nu_k^\epsilon(\mu) = \mu^r E\left(\mathbf{1}_{K_\epsilon}(\bar{X}) \cdot R_{2r+1}(\bar{X})\right). \tag{4.13}$$

The bounding result for the error term (4.12) and $\nu_{2r+2}^\epsilon < \nu_{2r+2}$ now give for the right–hand side of (4.13):

$$\begin{aligned}
\left|\mu^r E\left(\mathbf{1}_{K_\epsilon}(\bar{X}) \cdot R_{2r+1}(\bar{X})\right)\right| &\leq \mu^r E\left(\mathbf{1}_{K_\epsilon}(\bar{X}) \cdot |(R_{2r+1}(\bar{X}))|\right) \\
&\leq \mu^r s_{2r+1}(\epsilon) E\left(\mathbf{1}_{K_\epsilon}(\bar{X}) \cdot (\bar{X}-1)^{2r+2}\right) \\
&= \mu^r s_{2r+1}(\epsilon) E\Big(\mathbf{1}_{K_\epsilon}(\bar{X}) \cdot \left(\frac{X}{\mu} - \frac{\mu}{\mu}\right)^{2r+2}\Big) \\
&= \mu^r s_{2r+1}(\epsilon)\mu^{-2r-2} \cdot E\left(\mathbf{1}_{K_\epsilon}(\bar{X}) \cdot (X-\mu)^{2r+2}\right) \\
&= s_{2r+1}(\epsilon)\mu^{-r-2}\nu_{2r+2}^\epsilon(\mu) \\
&\leq s_{2r+1}(\epsilon)\mu^{-r-2}\nu_{2r+2}(\mu) \\
&= s_{2r+1}(\epsilon)\mu^{-1}\mu^{-(r+1)}\nu_{2r+2}(\mu).
\end{aligned}$$

For the asymptotics $\mu \to \infty$ the last term of the chain tends towards zero with rate $O(1/\mu)$ since by Theorem 4.3 $\nu_{2(r+1)}(\mu)$ is a polynomial in μ of degree $r+1$, and hence bounded if divided by μ^{r+1}. This and (4.13) thus yield the modified statement (4.10):

$$\left|\mu^r E\Big(\mathbf{1}_{K_\epsilon}(\bar{X}) \cdot (R_{2r+1}(\bar{X})\Big)\right| = \left|h_r^\epsilon(\mu) - \sum_{k=0}^{2r+1} \frac{1}{k!} H^{(k)}(1)\mu^{r-k}\nu_k^\epsilon(\mu)\right| = O(\frac{1}{\mu}).$$

In order to prove the actual result, it remains to be shown that $h_r^\epsilon(\mu)$ and $h_r(\mu)$ have the same asymptotic order for $\mu \to \infty$:

$$\mu^s |h_r(\mu) - h_r^\epsilon(\mu)| = O(1) \quad \text{for all} \quad s \in \mathbf{R}. \tag{4.14}$$

For this part of the proof, the assumption $|H(\bar{X})| \leq c(1+\bar{X}^m)$ with $c \in \mathbf{R}^+$, $m \in \mathbf{N}$ constant and further in particular m sufficiently large is needed. Thus for any given $l \in \mathbf{N}$, let without loss of generality be $m \geq r+l+1$.
For the difference of interest now the following holds:

$$\begin{aligned}
& |h_r(\mu) - h_r^\epsilon(\mu)| \\
= \; & |E(\mu^r H(\bar{X})) - E(\mathbf{1}_{K_\epsilon}(\bar{X}) \cdot \mu^r H(\bar{X}))| \\
= \; & |\mu^r E(\mathbf{1}_{[0,\infty)\setminus K_\epsilon}(\bar{X}) \cdot H(\bar{X}))| \\
\leq \; & \mu^r E(\mathbf{1}_{[0,\infty)\setminus K_\epsilon}(\bar{X}) \cdot |H(\bar{X})|) \\
= \; & \mu^r E(\mathbf{1}_{[0,1-\epsilon)}(\bar{X}) \cdot |H(\bar{X})| + \mathbf{1}_{(1+\epsilon,\infty)}(\bar{X}) \cdot |H(\bar{X})|)
\end{aligned}$$

$$
\begin{aligned}
&\le \mu^r \left(2cE(\mathbf{1}_{[0,1-\epsilon)}(\bar{X})) + cE(\mathbf{1}_{(1+\epsilon,\infty)}(\bar{X}) \cdot (1+\bar{X}^m))\right) \\
&\le \mu^r c \left(2E\left(\mathbf{1}_{[0,1-\epsilon)}(\bar{X}) + \mathbf{1}_{(1+\epsilon,\infty)}(\bar{X})\right) + E(\mathbf{1}_{(1+\epsilon,\infty)}(\bar{X}) \cdot \bar{X}^m)\right) \\
&= \mu^r c \left(2P(|\bar{X}-1| > \epsilon) + E(\mathbf{1}_{((1+\epsilon)^m,\infty)}(\bar{X}^m) \cdot \bar{X}^m)\right) \\
&\le \mu^r c \Big(2P(|\bar{X}-1| > \epsilon) + (1+\epsilon)^m P(\bar{X}^m \ge (1+\epsilon)^m)\Big) \\
&\quad + \mu^r c \int_{(1+\epsilon)^m}^{\infty} P(\bar{X}^m \ge s) ds \qquad (4.15) \\
&= \mu^r c \Big(2P(|\bar{X}-1| > \epsilon) + (1+\epsilon)^m P(\bar{X} \ge 1+\epsilon)\Big) \\
&\quad + \mu^r c \int_{(1+\epsilon)^m}^{\infty} P(\bar{X}^m \ge s) ds \\
&\le \mu^r c \Big(2P(|\bar{X}-1| > \epsilon) + (1+\epsilon)^m P(|\bar{X}-1| \ge \epsilon)\Big) \\
&\quad + \mu^r c \int_{(1+\epsilon)^m}^{\infty} P(\bar{X}^m \ge s) ds \\
&\le \mu^r c (2 + (1+\epsilon)^m) P(|\bar{X}-1| > \epsilon) + \mu^r c \int_{(1+\epsilon)^m}^{\infty} P(\bar{X}^m \ge s) ds \qquad (4.16)
\end{aligned}
$$

(for (4.15) see Billingsley, 1968, page 223 (3)). Application of Markov's inequality to $P(|\bar{X}-1| \ge \epsilon)$ yields for the first term of the sum

$$
\begin{aligned}
&\mu^r c (2 + (1+\epsilon)^m) P(|\bar{X}-1| \ge \epsilon) \\
&\le \mu^r c (2 + (1+\epsilon)^m) \frac{1}{\epsilon^{2(r+l)}} E(|\bar{X}-1|^{2(r+l)}) \\
&= \mu^r c (2 + (1+\epsilon)^m) \frac{1}{\epsilon^{2(r+l)}} E((\bar{X}-1)^{2(r+l)}) \\
&= \mu^r c (2 + (1+\epsilon)^m) \frac{1}{\epsilon^{2(r+l)}} \frac{E((X-\mu)^{2(r+l)})}{\mu^{2(r+l)}} \\
&= \mu^r c (2 + (1+\epsilon)^m) \frac{1}{\epsilon^{2(r+l)}} \frac{\nu_{2(r+l)}(\mu)}{\mu^{2(r+l)}} \\
&= c_1 \frac{\mu^r \nu_{2(r+l)}(\mu)}{\mu^{2(r+l)}} \\
&= c_1 \frac{1}{\mu^l} \frac{\nu_{2(r+l)}(\mu)}{\mu^{r+l}}
\end{aligned}
$$

with $c_1 = c(2 + (1+\epsilon)^m)\epsilon^{-2(r+1)}$ constant. For the second term follows

$$
\begin{aligned}
&\mu^r c \int_{(1+\epsilon)^m}^{\infty} P(\bar{X}^m \ge s) ds \\
&= \mu^r c \int_{(1+\epsilon)^m}^{\infty} P(\bar{X} \ge s^{\frac{1}{m}}) ds
\end{aligned}
$$

$$
\begin{aligned}
&= \mu^r c \int_{1+\epsilon}^{\infty} P(\bar{X} \geq t) m t^{m-1} dt \\
&= \mu^r c \int_{1+\epsilon}^{\infty} P(\bar{X} - 1 \geq t - 1) m t^{m-1} dt \\
&\leq \mu^r c \int_{1+\epsilon}^{\infty} \frac{E((\bar{X}-1)^{2(m-1)})}{(t-1)^{2(m-1)}} m t^{m-1} dt \qquad (4.17) \\
&= \mu^r c \int_{1+\epsilon}^{\infty} \frac{1}{\mu^{2(m-1)}} \frac{E((X-\mu)^{2(m-1)})}{(t-1)^{2(m-1)}} m t^{m-1} dt \\
&= \mu^r c m \frac{\nu_{2(m-1)}(\mu)}{\mu^{2(m-1)}} \int_{1+\epsilon}^{\infty} \frac{t^{m-1}}{(t-1)^{2(m-1)}} dt \\
&= \mu^r c m \frac{\nu_{2(m-1)}(\mu)}{\mu^{2(m-1)}} \int_{1+\epsilon}^{\infty} \Big(\frac{t}{t-1}\Big)^{m-1} \frac{1}{(t-1)^{m-1}} dt \\
&\leq \mu^r c m \frac{\nu_{2(m-1)}(\mu)}{\mu^{2(m-1)}} (1+\frac{1}{\epsilon})^{m-1} \int_{1+\epsilon}^{\infty} \frac{1}{(t-1)^{m-1}} dt \qquad (4.18) \\
&= \mu^r c m \frac{\nu_{2(m-1)}(\mu)}{\mu^{2(m-1)}} (1+\frac{1}{\epsilon})^{m-1} \int_{\epsilon}^{\infty} \frac{1}{u^{m-1}} du \\
&= \mu^r c m \frac{\nu_{2(m-1)}(\mu)}{\mu^{2(m-1)}} (1+\frac{1}{\epsilon})^{m-1} \Big[u^{-m+2} \frac{1}{-m+2} \Big]_{\epsilon}^{\infty} \\
&= \mu^r c m \frac{\nu_{2(m-1)}(\mu)}{\mu^{2(m-1)}} (1+\frac{1}{\epsilon})^{m-1} \frac{1}{(m-2)\epsilon^{m-2}} \\
&= c_2 \frac{\mu^r \nu_{2(m-1)}(\mu)}{\mu^{2(m-1)}} \\
&= c_2 \frac{\mu^r}{\mu^{m-1}} \frac{\nu_{2(m-1)}(\mu)}{\mu^{m-1}}
\end{aligned}
$$

with constant $c_2 = c \cdot (1+\epsilon^{-1})^{m-1} \cdot m \cdot (m-2)^{-1} \cdot \epsilon^{-m+2}$. Statement (4.17) is obtained applying Markov's inequality, and (4.18) holds because of

$$
t \geq 1+\epsilon \Leftrightarrow \frac{t}{t-1} \leq 1 + \frac{1}{\epsilon} \Leftrightarrow \Big(\frac{t}{t-1}\Big)^{m-1} \leq \Big(1+\frac{1}{\epsilon}\Big)^{m-1}.
$$

Using these inequalities for both terms of (4.16) now gives

$$
\begin{aligned}
|h_r(\mu) - h_r^{\epsilon}(\mu)| &\leq \frac{1}{\mu^l} c_1 \frac{\nu_{2(r+l)}(\mu)}{\mu^{r+l}} + \frac{\mu^r}{\mu^{m-1}} c_2 \frac{\nu_{2(m-1)}(\mu)}{\mu^{m-1}} \\
&= \frac{1}{\mu^l} \Big(c_1 \frac{\nu_{2(r+l)}(\mu)}{\mu^{r+l}} + \frac{\mu^{r+l+1}}{\mu^m} c_2 \frac{\nu_{2(m-1)}(\mu)}{\mu^{m-1}} \Big) \\
&= \frac{1}{\mu^l} \cdot O(1) \quad (l \in \mathbf{N}).
\end{aligned}
$$

The last equation holds since $m \geq r+l+1$ for arbitrarily chosen $l \in \mathbf{N}$ and because $\mu^{-(r+1)} \cdot \nu_{2(r+1)}(\mu)$ resp. $\mu^{-(m-1)} \cdot \nu_{2(m-1)}(\mu)$ are bounded as μ tends towards infinity.

The statement given above immediately yields (4.14), i.e.

$$\mu^s |h_r(\mu) - h_r^\epsilon(\mu)| = O(1) \quad \text{for all } s \in \mathbf{R}.$$

Since $\nu_r(\mu)$ and $\nu_r^\epsilon(\mu)$ are special cases of $h_r(\mu)$ und $h_r^\epsilon(\mu)$ $(H(\bar{X}) = (\bar{X} - E(\bar{X}))^r = (\bar{X}-1)^r$ with $|H(\bar{X})| \leq 1 + \bar{X}^r)$ and hence in particular

$$\mu^s |\nu_r(\mu) - \nu_r^\epsilon(\mu)| = O(1) \quad \text{for all } s \in \mathbf{R}$$

holds, it is possible to replace $h_r^\epsilon(\mu)$ and $\nu_k^\epsilon(\mu)$ through $h_r(\mu)$ and $\nu_k(\mu)$ in the modified statement (4.10) thus getting

$$\Big| h_r(\mu) - \sum_{k=0}^{2r+1} \frac{1}{k!} H^{(k)}(1) \mu^{r-k} \nu_k(\mu) \Big| = O(\frac{1}{\mu}) \quad (\mu \to \infty).$$

Since by Theorem 4.3, $\mu^{-r}\nu_{2r+1}(\mu)$ is bounded for the asymptotics $\mu \to \infty$, the last term of the sum reduces to

$$\frac{1}{(2r+1)!} H^{(2r+1)}(1) \mu^{r-(2r+1)} \nu_{2r+1}(\mu) = \frac{1}{(2r+1)!} H^{(2r+1)}(1) \mu^{-1} \mu^{-r} \nu_{2r+1}(\mu) = O(\frac{1}{\mu}),$$

thus giving the result. □

Examples 4.5 If in addition to the assumptions in Theorem 4.4 $H(1) = 0$ holds, then for $r = 0$ and $r = 1$ $\lim_{\mu\to\infty} h_r(\mu) = \lim_{\mu\to\infty} E(\mu^r H(\bar{X}))$ can be computed as follows (tabulation of the central Poisson moments see Appendix 8.1, $\nu_1(\mu) = 0$):

$$\lim_{\mu\to\infty} h_0(\mu) = \lim_{\mu\to\infty} E(H(\bar{X})) = H^{(0)}(1)\mu^0 \nu_0(\mu) = H(1) = 0,$$

$$\begin{aligned}
\lim_{\mu\to\infty} h_1(\mu) &= \lim_{\mu\to\infty} E(\mu H(\bar{X})) \\
&= \lim_{\mu\to\infty} \sum_{k=0}^{2} \frac{1}{k!} H^{(k)}(1) \mu^{1-k} \nu_k(\mu) \\
&= \lim_{\mu\to\infty} \frac{1}{2!} H^{(2)}(1) \mu^{-1} \nu_2(\mu) \\
&= \lim_{\mu\to\infty} \frac{1}{2} H^{(2)}(1) \mu^{-1} \mu \\
&= \frac{1}{2} H^{(2)}(1).
\end{aligned}$$

□

4.2 Results concerning the Distance Function

To begin with, in this section the distance function a_λ will be shown to be dominated by a power function and hence accomplishes the requirements of Theorem 4.1 (continuity/differentiability of the expected value) and Theorem 4.4 (Taylor approximation). Application of these results then yields, by rather long but simple calculations, the auxiliary results concerning the moments of a_λ needed in the following chapters.

Lemma 4.6 *Consider the distance function* $a_\lambda : \mathbf{R}_0^+ \to \mathbf{R}_0^+$, $x \mapsto a_\lambda(x,1)$ *from Definition 2.1. Then for each* $\lambda \in (-1,\infty)$, *there exist constants* $c \in \mathbf{R}^+$ *and* $m \in \mathbf{N}$ *such that*

$$|a_\lambda(x,1)| \le c(1+x^m).$$

Proof:
Consider $\lambda \in (-1,\infty)\backslash\{0\}$ first. Since $x \ge 0$ one gets

$$\begin{aligned}|a_\lambda(x,1)| &= \Big|\frac{2}{\lambda(\lambda+1)} \cdot x \cdot (x^\lambda - 1) - \frac{2}{\lambda+1}(x-1)\Big| \\ &\le \frac{2}{|\lambda|(\lambda+1)} \cdot (x^{\lambda+1} + x) + \frac{2}{\lambda+1}(x+1),\end{aligned}$$

and for the case $x \in [0,1]$ the upper bound

$$c = \frac{2}{|\lambda|(\lambda+1)} \cdot 2 + \frac{2}{\lambda+1} \cdot 2 = \frac{4}{\lambda+1}\Big(1 + \frac{1}{|\lambda|}\Big).$$

If $x \ge 1$ obviously $|a_\lambda(x,1)| \le c \cdot x^{\lambda+2}$ holds and hence

$$|a_\lambda(x,1)| \le c(1+x^{\lambda+2}) \text{ for each } \lambda \in (-1,\infty)\backslash\{0\}$$

is shown. For the Likelihood Ratio Statistic $a_0(x,1)$, i.e. $\lambda = 0$, it suffices to investigate the cases $0 < x \le 1$ and $x > 1$, since for $x = 0$ by definition $a_0(0,1) = 2$ holds. Considering

$$|a_0(x,1)| = |2(x \log x - (x-1))| \le 2x|\log x| + 2(x+1),$$

and applying the inequality $\log x \le x - 1$ yields for the case $x \in (0,1]$:

$$|a_0(x,1)| \le 2x(-\log x) + 2(x+1) = 2x \log\frac{1}{x} + 2(x+1) \le 2x(\frac{1}{x} - 1) + 2(x+1) = 4.$$

For $x > 1$ analogous argumentation leads to

$$|a_0(x,1)| \le 2x \log x + 2(x+1) \le 2x(x-1) + 2(x+1) = 2(x^2+1),$$

and hence also for $\lambda = 0$ ($c = 4$) the desired inequality

$$|a_\lambda(x,1)| \le c(1+x^{\lambda+2})$$

is established. Thus it remains to show

$$|a_\lambda(x,1)| \le c(1+x^m), \quad m \in \mathbf{N}.$$

This, however, is clear, because the majorization through a power function only concerns $x \geq 1$ and in this case holds

$$|a_\lambda(x,1)| \leq cx^{\lambda+2} \leq cx^m \quad \text{for arbitrary} \quad m \geq \lambda + 2 > 1, m \in \mathbf{N}.$$

□

Example 4.7 Using the positive homogenity of a_λ, the preceding Lemma 4.6 especially yields

$$|a_\lambda(x,\mu)| = \mu|a_\lambda(\frac{x}{\mu},1)| \leq \mu c\Big(1 + (\frac{x}{\mu})^m\Big) = c\mu + c\frac{1}{\mu^{m-1}}x^m$$

with $c \in \mathbf{R}^+$ and $m \in \mathbf{N}$. Hence for $\mu \in [a,b]$, $[a,b]$ being an arbitrary subset of $(0,\infty)$, $a_\lambda(x,\mu)$ is dominated by

$$c \cdot \sup_{\mu\in[a,b]} \mu + c \cdot \sup_{\mu\in[a,b]} \frac{1}{\mu^{m-1}} \cdot x^m = \beta + \gamma x^m$$

with $\beta = cb$ and $\gamma = c/a^{m-1}$. Since a_λ is continuous in μ, Theorem 4.1 a) resp. Corollary 4.2 yields $E(a_\lambda(X,\mu))$ being continuous in μ.
Further, $E(a_\lambda(X,\mu))$ is even continuously differentiable in μ (Theorem 4.1 c)), because $\frac{\partial}{\partial\mu}a_\lambda(x,\mu) = \frac{2}{\lambda+1}\Big(1 - (\frac{x}{\mu})^{\lambda+1}\Big)$ is continuous in μ and for $\mu \in [a,b] \subset \mathbf{R}^+$ obviously dominated by a power function as seen before.

□

Lemma 4.8 *Let be given the distance function a_λ from Definition 2.1, a positive constant ϵ and a Poisson distributed random variable X with expected value μ. Then for every $\lambda \in (-1,\infty)$ there exists a constant $c \in \mathbf{R}^+$, such that for all $\mu \in [\epsilon,\infty) \subset \mathbf{R}^+$ holds:*

a) $\Big|E\Big(a_\lambda(X,\mu)\Big)\Big| \leq c,$

b) $\Big|E\Big(a_\lambda^4(X,\mu)\Big)\Big| \leq c,$

c) $\Big|Var\Big(a_\lambda(X,\mu)\Big)\Big| \leq c,$

d) $\Big|Cov\Big(a_\lambda(X,\mu),X\Big)\Big| \leq c,$

e) $\Big|\mu E\Big((\frac{\partial}{\partial\mu}a_\lambda(X,\mu))^2\Big)\Big| \leq c,$

f) $\Big|\frac{1}{\mu}Cov\Big(a_\lambda(X,\mu),X^2\Big)\Big| \leq c,$

g) $\Big|\mu^2\frac{\partial}{\partial\mu}E\Big(a_\lambda(X,\mu)\Big)\Big| \leq c,$

h) $\Big|\mu\frac{\partial^2}{(\partial\mu)^2}E\Big(a_\lambda(X,\mu)\Big)\Big| \leq c,$

i) $\Big|\mu\frac{\partial}{\partial\mu}Var\Big(a_\lambda(X,\mu)\Big)\Big| \le c,$

j) $\Big|\mu\frac{\partial}{\partial\mu}Cov\Big(a_\lambda(X,\mu),X\Big)\Big| \le c,$

k) $\Big|\frac{\partial}{\partial\mu}Cov\Big(a_\lambda(X,\mu),X^2\Big)\Big| \le c,$

l) $\Big|\mu\Big(Cov\Big(\frac{\partial}{\partial\mu}a_\lambda(X,\mu),X\Big)+2\Big)\Big| \le c,$

m) $\Big|\mu\Big(\mu Var(\frac{\partial}{\partial\mu}a_\lambda(X,\mu))-4\Big)\Big| \le c,$

n) $\Big|\mu^2 E\Big(\Big(\frac{\partial}{\partial\mu}a_\lambda(X,\mu)-E(\frac{\partial}{\partial\mu}a_\lambda(X,\mu))\Big)^4\Big)\Big| \le c.$

Here the constant $c \in \mathbf{R}^+$ (which depends on λ) is a global supremum.

Proof:
The expressions in a) – n) are all of the kind $E(g(X,\mu))$ with $g : \mathbf{N_0} \times \mathbf{R}^+ \to \mathbf{R}$. This of course also includes the derived expectations, since by Theorem 4.1 b) for any function $f : \mathbf{N_0} \times \mathbf{R}^+ \to \mathbf{R}$

$$\begin{aligned}\frac{\partial}{\partial\mu}E(f(X,\mu)) &= E\Big(\frac{\partial}{\partial\mu}f(X,\mu)\Big)+\frac{1}{\mu}Cov\Big(f(X,\mu),X\Big)\\ &= E\Big(\frac{\partial}{\partial\mu}f(X,\mu)+f(X,\mu)\cdot\frac{X}{\mu}-f(X,\mu)\Big)\\ &= E(g(X,\mu))\end{aligned}$$

holds, if $\frac{\partial}{\partial\mu}f(X,\mu)$ exists and both $f(X,\mu)$ and $\frac{\partial}{\partial\mu}f(X,\mu)$ are dominated by an exponential function. Since the functions g resp. f considered here are simple products of a_λ and the random variable X, this obviously holds (cp. Example 4.7). With the functions g further being continuous in μ, Th. 4.1 in particular gives the continuity in μ for all expectations considered in a) – n). Now if for the asymptotics $\mu \to \infty$ additionally $E(g(X,\mu)) = O(1)$ holds, this together with the assumption $\mu \in [\epsilon,\infty) \subset \mathbf{R}^+$ gives the stated boundedness. Thus only the arguments concerning the limiting behaviour are left to be proved. They can be easily derived applying Theorem 4.4. Now, first of all, the following statements will be shown, which will be repeatedly used in the actual proofs ($\bar{X} := X/\mu$):

$$\Big|E\Big(\mu a_\lambda(\bar{X},1)-1\Big| = O(\frac{1}{\mu}), \tag{4.19}$$

$$\Big|E\Big(\mu^2 a_\lambda(\bar{X},1)(\bar{X}-1)\Big)-\lambda\Big| = O(\frac{1}{\mu}), \tag{4.20}$$

$$\Big|E\Big(\mu^2 a_\lambda(\bar{X},1)\bar{X}(\bar{X}-1)\Big)-\lambda-3\Big| = O(\frac{1}{\mu}), \tag{4.21}$$

$$\left|E\Big(\mu D_2 a_\lambda(\bar{X},1)\Big)+\lambda\right| = O(\frac{1}{\mu}), \quad (4.22)$$

$$\left|E\Big(\mu D_2 a_\lambda(\bar{X},1)(\bar{X}-1)\Big)+2\right| = O(\frac{1}{\mu}). \quad (4.23)$$

By Theorem 4.4, for functions $H(\bar{X})$ as defined there, which further fulfill $H(1)=0$, generally holds:

$$\left|E(\mu H(\bar{X}))-\frac{1}{2}H^{(2)}(1)\right| = O(\frac{1}{\mu}). \quad (4.24)$$

Choosing $H(\bar{X}) := a_\lambda(\bar{X},1)$, for which Lemma 2.2 gives $H(1)=a_\lambda(1,1)=0$ and $H^{(2)}(1)=D_1^2 a_\lambda(1,1)=2$, statement (4.19) immediately follows:

$$\left|E(\mu a_\lambda(\bar{X},1))-\frac{1}{2}\cdot 2\right| = O(\frac{1}{\mu}).$$

In order to establish (4.20), consider $H(\bar{X}) := a_\lambda(\bar{X},1)(\bar{X}-1)$ and the general statement deduced from Theorem 4.4,

$$\left|E(\mu^2 H(\bar{X}))-\frac{\mu}{2}H^{(2)}(1)-\frac{1}{3!}H^{(3)}(1)-\frac{3}{4!}H^{(4)}(1)\right| = O(\frac{1}{\mu}), \quad (4.25)$$

which presumes $H(1)=0$ as in the given case. The first derivatives of products $H=f\cdot g$ state as follows:

$$\begin{aligned}
H^{(1)} &= f^{(1)}g+fg^{(1)},\\
H^{(2)} &= f^{(2)}g+2f^{(1)}g^{(1)}+fg^{(2)}\\
H^{(3)} &= f^{(3)}g+3f^{(2)}g^{(1)}+3f^{(1)}g^{(2)}+fg^{(3)},\\
H^{(4)} &= f^{(4)}g+4f^{(3)}g^{(1)}+6f^{(2)}g^{(2)}+4f^{(1)}g^{(3)}+fg^{(4)}.
\end{aligned}$$

Hence, for the derivatives of $H(\bar{X})=f(\bar{X})\cdot g(\bar{X})$ with $f(\bar{X})=a_\lambda(\bar{X},1), g(\bar{X})=\bar{X}-1$ and especially $f(1)=a_\lambda(1,1)=0$, $f^{(1)}(1)=D_1^1 a_\lambda(1,1)=0$, $g(1)=0$, $g^{(k)}(1)=0$ $(k\geq 2)$ clearly holds

$$\begin{aligned}
H(1) &= H^{(1)}(1)=H^{(2)}(1)=0,\\
H^{(3)}(1) &= 3f^{(2)}(1)g^{(1)}(1),\\
H^{(4)}(1) &= 4f^{(3)}(1)g^{(1)}(1).
\end{aligned}$$

Further using $f^{(2)}(1)=D_1^2 a_\lambda(1,1)=2$, $f^{(3)}(1)=2(\lambda-1)$ and $g^{(1)}(1)=1$, which was also provided in Lemma 2.2, one gets $H^{(3)}(1)=3\cdot 2\cdot 1=6$, $H^{(4)}(1)=4\cdot 2(\lambda-1)\cdot 1=8(\lambda-1)$ and, applying (4.25), thus (4.20):

$$\begin{aligned}
&\left|E(\mu^2 H(\bar{X}))-\frac{\mu}{2}H^{(2)}(1)-\frac{1}{3!}H^{(3)}(1)-\frac{3}{4!}H^{(4)}(1)\right|\\
=\ &\left|E\Big(\mu^2 a_\lambda(\bar{X},1)(\bar{X}-1)\Big)-\frac{1}{3!}\cdot 6-\frac{3}{4!}\cdot 8(\lambda-1)\right|
\end{aligned}$$

$$= \left|E\Big(\mu^2 a_\lambda(\bar X,1)(\bar X-1)\Big)-\lambda\right|$$
$$= O(\frac{1}{\mu}).$$

The proof of (4.21) will be done with analogous arguments, since again a product $H(\bar X)=f(\bar X)\cdot g(\bar X)$ is studied with here $f(\bar X)=a_\lambda(\bar X,1)$ as before and, differently, $g(\bar X)=\bar X(\bar X-1)$. Using $f(1)=f^{(1)}(1)=0$ and $g(1)=0$, $g^{(1)}(1)=1$, $g^{(2)}(1)=2$, the general formulae for the derivatives of $H=f\cdot g$ just stated immediately entail

$$\begin{aligned}
H(1) &= H^{(1)}(1)=H^{(2)}(1)=0,\\
H^{(3)}(1) &= 3f^{(2)}(1)\cdot g^{(1)}(1)=3D_1^2a_\lambda(1,1)\cdot g^{(1)}(1)=3\cdot 2\cdot 1=6,\\
H^{(4)}(1) &= 4f^{(3)}(1)\cdot g^{(1)}(1)+6f^{(2)}(1)\cdot g^{(2)}(1)\\
&= 4D_1^3a_\lambda(1,1)\cdot g^{(1)}(1)+6D_1^2a_\lambda(1,1)\cdot g^{(2)}(1)\\
&= 4\cdot 2(\lambda-1)\cdot 1+6\cdot 2\cdot 2\\
&= 8(\lambda-1)+24
\end{aligned}$$

und thus the result

$$\begin{aligned}
&\left|E(\mu^2 H(\bar X))-\frac{\mu}{2}H^{(2)}(1)-\frac{1}{3!}H^{(3)}(1)-\frac{3}{4!}H^{(4)}(1)\right|\\
=\ &\left|E\Big(\mu^2 a_\lambda(\bar X,1)\bar X(\bar X-1)\Big)-\frac{1}{3!}6-\frac{3}{4!}(8(\lambda-1)+24)\right|\\
=\ &\left|E\Big(\mu^2 a_\lambda(\bar X,1)\bar X(\bar X-1)\Big)-3-\lambda\right|\\
=\ &O(\frac{1}{\mu}).
\end{aligned}$$

For the proof of (4.22) define $H(\bar X):=\frac{2}{\lambda+1}(1-\bar X^{\lambda+1})=D_2a_\lambda(\bar X,1)$, for which obviously $H(1)=0$ holds. Hence, similar to the proof of (4.19), the result follows using (4.24), i.e.

$$\left|E(\mu H(\bar X))-\frac{1}{2}H^{(2)}(1)\right|=O(\frac{1}{\mu}),$$

since $H^{(1)}(\bar X)=-2\bar X^\lambda$ and $H^{(2)}(\bar X)=-2\lambda\bar X^{\lambda-1}$, which gives $\frac{1}{2}H^{(2)}(1)=-\lambda$.

The same formula, i.e. (4.24), will now be applied in order to prove (4.23). For this purpose consider $H(\bar X):=\frac{2}{\lambda+1}(1-\bar X^{\lambda+1})(\bar X-1)=D_2a_\lambda(\bar X,1)(\bar X-1)$, which will again be written as a product as follows: $H(\bar X)=\frac{2}{\lambda+1}f(\bar X)\cdot g(\bar X)$ with $f(\bar X)=(1-\bar X^{\lambda+1})$ and $g(\bar X)=(\bar X-1)$. For these functions clearly holds $f(1)=g(1)=0$, $f^{(1)}(\bar X)=-(\lambda+1)\bar X^\lambda$, $f^{(1)}(1)=-(\lambda+1)$ and $g^{(1)}(\bar X)=1$. Similar arguments as before thus give

$$H^{(1)}(1) = \frac{2}{\lambda+1}(f^{(1)}(1)\cdot g(1)+f(1)\cdot g^{(1)}(1))=0,$$

$$\begin{aligned} H^{(2)}(1) &= \frac{2}{\lambda+1}(f^{(2)}(1)\cdot g(1) + 2f^{(1)}(1)\cdot g^{(1)}(1) + f(1)\cdot g^{(2)}(1)) \\ &= \frac{2}{\lambda+1}(2f^{(1)}(1)\cdot g^{(1)}(1)) \\ &= \frac{2}{\lambda+1}(2(-(\lambda+1))\cdot 1) \\ &= -4. \end{aligned}$$

Hence (4.23) is established:

$$\begin{aligned} \left|E(\mu H(\bar{X})) - \frac{1}{2}H^{(2)}(1)\right| &= \left|E(\mu D_2 a_\lambda(\bar{X},1)(\bar{X}-1)) - \frac{1}{2}(-4)\right| \\ &= |E(\mu D_2 a_\lambda(\bar{X},1)(\bar{X}-1)) + 2| \\ &= O(\frac{1}{\mu}). \end{aligned}$$

With the auxiliary results (4.19) – (4.23) being proved, now the actual statements a) – n) will be verified.

a) $E(a_\lambda(X,\mu)) = \mu \cdot E(a_\lambda(\bar{X},1)) = O(1)$ is an immediate consequence of the stronger result (4.19).

b) Application of Theorem 4.4 to $H(\bar{X}) := a_\lambda^4(\bar{X},1)$ and use of $H^{(k)}(1) = 0$ for $k = 1,\ldots,6$, which is verified through simple calculations especially using $a_\lambda(1,1) = D_1^1 a_\lambda(1,1) = 0$, yields

$$\begin{aligned} |E(a_\lambda^4(X,\mu))| &= \mu\cdot|E(\mu^3 a_\lambda(\bar{X},1))| \\ &= \mu\cdot|E(\mu^3 H(\bar{X}))| \\ &= \mu\cdot\left|E(\mu^3 H(\bar{X})) - \sum_{k=0}^{6}\frac{1}{k!}\mu^{3-k}\nu_k(\mu)H^{(k)}(1)\right| \\ &= \mu\cdot O(\frac{1}{\mu}) \\ &= O(1). \end{aligned}$$

c) The boundedness of the variance $Var(a_\lambda(X,\mu)) = E(a_\lambda^2(X,\mu)) - E^2(a_\lambda(X,\mu))$ is immediately deduced from b).

d) Here the desired order is already given by statement (4.20):

$$\begin{aligned} O(1) &= |E(\mu^2 a_\lambda(\bar{X},1)(\bar{X}-1))| \\ &= |E(a_\lambda(X,\mu)(X-\mu))| \\ &= |Cov(a_\lambda(X,\mu),X)|. \end{aligned}$$

e) Define $H(\bar{X}) = (\frac{2}{\lambda+1}(1 - \bar{X}^{\lambda+1}))^2$ with $H(1) = 0$. Then Theorem 4.4 (see also Example 4.5) entails

$$|E(\mu^0 H(\bar{X})) - H(1)| = |E(H(\bar{X}))| = O(\frac{1}{\mu}) \quad (\mu \to \infty),$$

and, with regard to

$$D_2 a_\lambda(X, \mu) = \frac{2}{\lambda+1}(1 - (\frac{X}{\mu})^{\lambda+1}) = \frac{2}{\lambda+1}(1 - \bar{X}^{\lambda+1}) = D_2 a_\lambda(\bar{X}, 1), \qquad (4.26)$$

thus the needed boundedness

$$|E(\mu(D_2 a_\lambda(X, \mu))^2| = |E(\mu H(\bar{X})| = O(1).$$

f) In order to prove $|\frac{1}{\mu} Cov(a_\lambda(X, \mu), X^2)| = O(1)$ consider

$$\begin{aligned}
& \left|\frac{1}{\mu} Cov(a_\lambda(X, \mu), X^2)\right| \\
= & \left|\frac{1}{\mu}\Big(E(a_\lambda(X, \mu) \cdot X^2) - E(a_\lambda(X, \mu)) \cdot E(X^2)\Big)\right| \\
= & \left|\frac{1}{\mu}\Big(E(a_\lambda(X, \mu) \cdot X^2) - E(a_\lambda(X, \mu)) \cdot (\mu + \mu^2)\Big)\right| \\
\leq & \left|\frac{1}{\mu}\Big(E(a_\lambda(X, \mu) \cdot X^2) - \mu^2 E(a_\lambda(X, \mu))\Big)\right| + \left|\frac{1}{\mu} \cdot \mu E(a_\lambda(X, \mu))\right| \\
= & \left|\frac{1}{\mu}\Big(E(\mu^3 a_\lambda(\bar{X}, 1) \cdot \bar{X}^2) - E(\mu^3 a_\lambda(\bar{X}, 1))\Big)\right| + |E(a_\lambda(X, \mu))| \\
= & \mu |E(\mu a_\lambda(\bar{X}, 1)(\bar{X}^2 - 1))| + |E(a_\lambda(X, \mu))|. \qquad (4.27)
\end{aligned}$$

The second term is bounded by a). The order of the first term is determined analogously to preceding proofs applying Theorem 4.4 resp. (4.24), namely

$$\left|E(\mu H(\bar{X})) - \frac{1}{2} H^{(2)}(1)\right| = O(\frac{1}{\mu})$$

with here $H(\bar{X}) = a_\lambda(\bar{X}, 1)(\bar{X}^2 - 1)$, $H(1) = 0$ and especially $H^{(2)}(1) = 0$ using $a_\lambda(1, 1) = D_1^1 a_\lambda(1, 1) = 0$. This gives

$$\left|E\Big(\mu a_\lambda(\bar{X}, 1)(\bar{X}^2 - 1)\Big)\right| = \left|E(\mu H(\bar{X}))\right| = \left|E(\mu H(\bar{X})) - \frac{1}{2} H^{(2)}(1)\right| = O(\frac{1}{\mu})$$

and thus for the first term of (4.27) the order $O(1)$, too.

g) Using the formula given in Theorem 4.1 b), $D_2 a_\lambda(X, \mu) = D_2 a_\lambda(\bar{X}, 1)$ as stated in (4.26), and finally statement (4.20) and (4.22), the result is obtained as follows:

$$\left|\mu^2 \cdot \frac{\partial}{\partial \mu} E(a_\lambda(X, \mu))\right|$$

$$
\begin{aligned}
&= \mu^2 \cdot \Big|E(D_2 a_\lambda(X,\mu)) + \frac{1}{\mu} Cov(a_\lambda(X,\mu), X)\Big| \\
&= \mu \cdot \Big|E(\mu D_2 a_\lambda(X,\mu)) + E(a_\lambda(X,\mu)(X-\mu))\Big| \\
&= \mu \cdot \Big|E(\mu D_2 a_\lambda(\bar{X},1)) + \lambda - \lambda + E(\mu^2 a_\lambda(\bar{X},1)(\bar{X}-1))\Big| \\
&\leq \mu \cdot \Big|E(\mu D_2 a_\lambda(\bar{X},1)) + \lambda\Big| + \mu\Big|E(\mu^2 a_\lambda(\bar{X},1)(\bar{X}-1)) - \lambda\Big| \\
&= \mu \cdot O(\frac{1}{\mu}).
\end{aligned}
$$

h) For the proof of

$$
\Big|\mu \frac{\partial^2}{(\partial\mu)^2} E(a_\lambda(X,\mu))\Big| = O(1),
$$

which is again done applying the formula for derived Poisson expectations given in Theorem 4.1, consider the following decomposition:

$$
\begin{aligned}
&\Big|\mu \frac{\partial^2}{(\partial\mu)^2} E(a_\lambda(X,\mu))\Big| \\
&= \mu\Big|\frac{\partial}{\partial\mu}\Big(E\Big(\frac{\partial}{\partial\mu} a_\lambda(X,\mu)\Big) + E\Big(a_\lambda(X,\mu)\cdot\frac{X}{\mu}\Big) - E(a_\lambda(X,\mu))\Big)\Big| \\
&= \mu\Big|E\Big(\frac{\partial^2}{(\partial\mu)^2} a_\lambda(X,\mu)\Big) + Cov\Big(\frac{\partial}{\partial\mu} a_\lambda(X,\mu), \frac{X}{\mu}\Big) + E\Big(\frac{\partial}{\partial\mu}\Big(a_\lambda(X,\mu)\cdot\frac{X}{\mu}\Big)\Big) \\
&\qquad + Cov\Big(a_\lambda(X,\mu)\cdot\frac{X}{\mu}, \frac{X}{\mu}\Big) - \frac{\partial}{\partial\mu} E(a_\lambda(X,\mu))\Big|
\end{aligned}
$$

$$
\leq \mu\Big|E\Big(\frac{\partial^2}{(\partial\mu)^2} a_\lambda(X,\mu)\Big)\Big| \tag{4.28}
$$

$$
+\mu\Big|E\Big(\frac{\partial}{\partial\mu} a_\lambda(X,\mu)\cdot(\frac{X}{\mu}-1)\Big)\Big| \tag{4.29}
$$

$$
+\mu\Big|E\Big(\frac{\partial}{\partial\mu}\Big(a_\lambda(X,\mu)\cdot\frac{X}{\mu}\Big)\Big)\Big| \tag{4.30}
$$

$$
+\mu\Big|E\Big(a_\lambda(X,\mu)\cdot\frac{X}{\mu}\cdot(\frac{X}{\mu}-1)\Big)\Big| \tag{4.31}
$$

$$
+\mu\Big|\frac{\partial}{\partial\mu} E(a_\lambda(X,\mu))\Big|. \tag{4.32}
$$

Using (4.26), namely $\frac{\partial}{\partial\mu} a_\lambda(X,\mu) = D_2 a_\lambda(\bar{X},1)$, respectively $\frac{\partial}{\partial\mu} a_\lambda(X,\mu)(\frac{X}{\mu}-1) = D_2 a_\lambda(\bar{X},1)(\bar{X}-1)$, the second term of the expression, i.e. (4.29), is bounded by (4.23). The same applies to the fourth term (4.31) by statement (4.21),

$$
\mu\Big|E(a_\lambda(X,\mu)\cdot\frac{X}{\mu}\cdot(\frac{X}{\mu}-1))\Big| = \Big|E(\mu^2 a_\lambda(\bar{X},1)\cdot\bar{X}\cdot(\bar{X}-1))\Big| = O(1),
$$

and to the last term (4.32) by g). In order to study (4.28), consider $\frac{\partial^2}{(\partial\mu)^2}a_\lambda(X,\mu) = \frac{\partial}{\partial\mu}\Big(\frac{2}{\lambda+1}\Big(1-(\frac{X}{\mu})^{\lambda+1}\Big)\Big) = \frac{2}{\mu}\cdot(\frac{X}{\mu})^{\lambda+1}$. The expected value $\mu E(\frac{\partial^2}{(\partial\mu)^2}a_\lambda(X,\mu)) = E(2\bar{X}^{\lambda+1})$ is obviously bounded. This can also be seen applying Theorem 4.4, which gives for $H(\bar{X}) = 2\bar{X}^{\lambda+1}$ with $H(1)=2$:

$$O(\frac{1}{\mu}) = |E(H(\bar{X}) - H(1))| = |E(2\bar{X}^{\lambda+1}-2)| = \Big|E(\mu\frac{\partial^2}{(\partial\mu)^2}a_\lambda(X,\mu)-2)\Big|.$$

The boundedness of (4.30) follows immediately using

$$\mu\cdot\Big|E(H(\bar{X}))-H(1)\Big| = O(1)$$

(Theorem 4.4) with here

$$\begin{aligned} H(\frac{X}{\mu}) &:= \frac{\partial}{\partial\mu}\Big(a_\lambda(X,\mu)\cdot\frac{X}{\mu}\Big) \\ &= \frac{\partial}{\partial\mu}a_\lambda(X,\mu)\cdot\frac{X}{\mu} - a_\lambda(X,\mu)\cdot\frac{X}{\mu^2} \\ &= \frac{\partial}{\partial\mu}a_\lambda(X,\mu)\cdot\frac{X}{\mu} - a_\lambda(\frac{X}{\mu},1)\cdot\frac{X}{\mu} \end{aligned}$$

and $H(1)=0$.

i) Using the formula from Theorem 4.1 b) gives for the derived variance

$$\begin{aligned} &\mu\Big|\frac{\partial}{\partial\mu}Var(a_\lambda(X,\mu))\Big| \\ =\ &\mu\Big|\frac{\partial}{\partial\mu}\Big(E(a_\lambda^2(X,\mu)) - E^2(a_\lambda(X,\mu))\Big)\Big| \\ =\ &\mu\Big|E(\frac{\partial}{\partial\mu}a_\lambda^2(X,\mu)) + Cov(a_\lambda^2(X,\mu),\frac{X}{\mu}) - 2E(a_\lambda(X,\mu))\cdot\frac{\partial}{\partial\mu}E(a_\lambda(X,\mu))\Big|. \end{aligned} \tag{4.33}$$

The boundedness of the last term has already been shown in a) and g), which assure

$$E(a_\lambda(X,\mu))\cdot\frac{\partial}{\partial\mu}E(a_\lambda(X,\mu)) = O(1)\cdot O(\frac{1}{\mu^2}). \tag{4.34}$$

In order to investigate the first term of (4.33) consider

$$\frac{\partial}{\partial\mu}a_\lambda^2(X,\mu) = 2a_\lambda(X,\mu)\cdot\frac{\partial}{\partial\mu}a_\lambda(X,\mu) = 2\mu a_\lambda(\bar{X},1)\cdot D_2a_\lambda(\bar{X},1) = \mu H(\bar{X})$$

($D_2a_\lambda(X,\mu) = D_2a_\lambda(\bar{X},1)$ see (4.26)) with $H(\bar{X}) = 2a_\lambda(\bar{X},1)\cdot D_2a_\lambda(\bar{X},1)$ and $H(1)=0$. Thus formula (4.24) deduced from Theorem 4.4 can again be applied, which, provided $H^{(2)}(1)=0$ holds, gives

$$|E(\frac{\partial}{\partial\mu}a_\lambda^2(X,\mu))| = |E(\mu H(\bar{X}))| = |E(\mu H(\bar{X})) - \frac{1}{2}H^{(2)}(1)| = O(\frac{1}{\mu}). \tag{4.35}$$

$H^{(2)}(1) = 0$ is proved with standard argumentation (cp. proof of (4.22)), just using $a_\lambda(1,1) = D_1^1 a_\lambda(1,1) = D_2^1 a_\lambda(1,1) = 0$. The order of the second term of (4.33) will be determined applying formula (4.25) again. For this purpose consider

$$\begin{aligned} Cov(a_\lambda^2(X,\mu), \frac{X}{\mu}) &= E(a_\lambda^2(X,\mu)\cdot(\frac{X}{\mu}-1)) \\ &= E(\mu^2 a_\lambda^2(\bar{X},1)\cdot(\bar{X}-1)) \\ &= E(\mu^2 H(\bar{X})) \end{aligned}$$

with $H(\bar{X}) = a_\lambda^2(\bar{X},1)\cdot(\bar{X}-1)$, $H(1)=0$. Simple calculations analogously to the proof of (4.20) give $H^{(k)}(1) = 0$ for $k = 1,\ldots,4$ and thus

$$|Cov(a_\lambda^2(X,\mu), \frac{X}{\mu})| = |E(\mu^2 H(\bar{X})) - \frac{\mu}{2}H^{(2)}(1) - \frac{1}{3!}H^{(3)}(1) - \frac{3}{4!}H^{(4)}(1)| = O(\frac{1}{\mu}). \quad (4.36)$$

Statement (4.34), (4.35) and (4.36) finally establish the boundedness of the expression in (4.33) and hence

$$\mu\Big|\frac{\partial}{\partial\mu}Var(a_\lambda(X,\mu))\Big| = O(1).$$

j) The stated boundedness of the derived covariance resp. $\frac{\partial}{\partial\mu}Cov(a_\lambda(X,\mu),X) = O(\frac{1}{\mu})$, can be checked with known arguments. Application of Theorem 4.1 b) and further calculations give

$$\begin{aligned} &\frac{\partial}{\partial\mu}Cov(a_\lambda(X,\mu),X) \\ =\ &\frac{\partial}{\partial\mu}E(a_\lambda(X,\mu)(X-\mu)) \\ =\ &E\Big(\frac{\partial}{\partial\mu}\Big(a_\lambda(X,\mu)(X-\mu)\Big)\Big) + \frac{1}{\mu}Cov\Big(a_\lambda(X,\mu)(X-\mu),X\Big) \\ =\ &E\Big(D_2 a_\lambda(X,\mu)\cdot(X-\mu)\Big) - E(a_\lambda(X,\mu)) + \frac{1}{\mu}E(a_\lambda(X,\mu)(X-\mu)^2) \\ =\ &E\Big(\mu D_2 a_\lambda(X,\mu)\cdot(\frac{X}{\mu}-1)\Big) - E(\mu a_\lambda(\frac{X}{\mu},1)) + E(\mu^2 a_\lambda(\frac{X}{\mu},1)\cdot(\frac{X}{\mu}-1)^2) \\ =\ &E\Big(\mu D_2 a_\lambda(\bar{X},1)\cdot(\bar{X}-1)\Big) + 2 \\ &-E(\mu a_\lambda(\bar{X},1)) + 1 \\ &+E(\mu^2 a_\lambda(\bar{X},1)\cdot\bar{X}\cdot(\bar{X}-1)) - \lambda - 3 \\ &-E(\mu^2 a_\lambda(\bar{X},1)\cdot(\bar{X}-1)) + \lambda \end{aligned}$$

with each of the last four terms being of the order $O(\frac{1}{\mu})$, which follows using the auxiliary statements (4.19) – (4.21) and (4.23).

k) In order to prove $\frac{\partial}{\partial\mu}Cov(a_\lambda(X,\mu),X^2) = O(1)$, consider the following transformation, which is again done applying Theorem 4.1 b):

$$\begin{aligned}
&\frac{\partial}{\partial\mu}Cov(a_\lambda(X,\mu),X^2)\\
=\;& \frac{\partial}{\partial\mu}E(a_\lambda(X,\mu)\cdot X^2)-\frac{\partial}{\partial\mu}\Big(E(a_\lambda(X,\mu))E(X^2)\Big)\\
=\;& E\Big(\frac{\partial}{\partial\mu}\Big(a_\lambda(X,\mu)\cdot X^2\Big)\Big)+\frac{1}{\mu}Cov\Big(a_\lambda(X,\mu)\cdot X^2,X\Big)-\frac{\partial}{\partial\mu}\Big(E(a_\lambda(X,\mu))(\mu+\mu^2)\Big)\\
=\;& E\Big(X^2\frac{\partial}{\partial\mu}(a_\lambda(X,\mu)\Big)+\frac{1}{\mu}E\Big(a_\lambda(X,\mu)X^2(X-\mu)\Big)\\
&-\frac{\partial}{\partial\mu}E(a_\lambda(X,\mu))\cdot(\mu+\mu^2)-E(a_\lambda(X,\mu))\cdot(1+2\mu). \qquad (4.37)
\end{aligned}$$

In the following $\mu\frac{\partial}{\partial\mu}E(a_\lambda(X,\mu))$ and $E(a_\lambda(X,\mu))$ can be omitted, since the boundedness of these terms was already shown in g) respectively a). Use of $\mu^2\frac{\partial}{\partial\mu}E(a_\lambda(X,\mu)) = \mu^2E(\frac{\partial}{\partial\mu}a_\lambda(X,\mu))+\mu E(a_\lambda(X,\mu)(X-\mu))$ (cp. proof of g)), then gives for the remaining terms in (4.37):

$$\begin{aligned}
&E\Big(X^2\frac{\partial}{\partial\mu}(a_\lambda(X,\mu)\Big)+\frac{1}{\mu}E\Big(a_\lambda(X,\mu)X^2(X-\mu)\Big)\\
&-\mu^2\frac{\partial}{\partial\mu}E(a_\lambda(X,\mu))-2\mu E(a_\lambda(X,\mu)).\\
=\;& E\Big(\mu^2\bar{X}^2\frac{\partial}{\partial\mu}(a_\lambda(X,\mu)\Big)+E\Big(\mu^3a_\lambda(\bar{X},1)\bar{X}^2(\bar{X}-1)\Big)\\
&-E\Big(\mu^2\frac{\partial}{\partial\mu}a_\lambda(X,\mu)\Big)-E\Big(\mu^3a_\lambda(\bar{X},1)(\bar{X}-1)\Big)-2E\Big(\mu^2a_\lambda(\bar{X},1)\Big)\\
=\;& \mu\Big(E\Big(\mu^2a_\lambda(\bar{X},1)(\bar{X}-1)(\bar{X}^2-1)\Big)+E\Big(\mu\frac{\partial}{\partial\mu}a_\lambda(X,\mu)\cdot(\bar{X}^2-1)\Big)\\
&\quad-2E(\mu a_\lambda(\bar{X},1))\Big)\\
=\;& \mu\Big(E(\mu^2H_1(\bar{X}))+E(\mu H_2(\bar{X}))-2E(\mu H_3(\bar{X}))\Big) \qquad (4.38)
\end{aligned}$$

with (cp. (4.26))

$$\begin{aligned}
H_1(\bar{X}) &= a_\lambda(\bar{X},1)(\bar{X}-1)(\bar{X}^2-1),\\
H_2(\bar{X}) &= \frac{2}{\lambda+1}(1-\bar{X}^{\lambda+1})(\bar{X}^2-1),\\
H_3(\bar{X}) &= a_\lambda(\bar{X},1).
\end{aligned}$$

In order to establish the boundedness of (4.38) and hence of the derived covariance in question, it obviouly suffices to show

$$|E(\mu^2H_1(\bar{X})-6)| = O(\frac{1}{\mu}), \qquad (4.39)$$

$$|E(\mu H_2(\bar{X})+4)| \;=\; O(\frac{1}{\mu}), \tag{4.40}$$

$$|E(\mu H_3(\bar{X})-1)| \;=\; O(\frac{1}{\mu}). \tag{4.41}$$

Now, since (4.41) equals statement (4.19), which was already proved in the beginning, only (4.39) and (4.40) remain to be checked. These statements can be verified using formula (4.24) and (4.25) again, namely

$$\Big|E(\mu^2 H_1(\bar{X})) - \big(\frac{\mu}{2}H_1^{(2)}(1) + \frac{1}{3!}H_1^{(3)}(1) + \frac{3}{4!}H_1^{(4)}(1)\big)\Big| \;=\; O(\frac{1}{\mu}),$$

$$\Big|E(\mu H_2(\bar{X})) - \frac{1}{2}H_2^{(2)}(1)\Big| \;=\; O(\frac{1}{\mu}).$$

Simple calculations give for the derivatives of the first function $H_1^{(k)}(1) = 0$ for $k = 0, \ldots, 3$ and $H_1^{(4)}(1) = 48$. For H_2 one gets $H_2^{(k)}(1) = 0$ for $k = 0, 1$ and $H_2^{(2)}(1) = -8$, which establishes

$$\frac{\mu}{2}H_1^{(2)}(1) + \frac{1}{3!}H_1^{(3)}(1) + \frac{3}{4!}H_1^{(4)}(1) \;=\; \frac{3}{4!}48 \;=\; 6,$$

$$\frac{1}{2}H_2^{(2)}(1) \;=\; \frac{1}{2}\cdot(-8) \;=\; -4,$$

and hence (4.39) and (4.40).

l) The desired bounding result $\Big|\mu\Big(Cov(D_2a_\lambda(X,\mu),X)+2\Big)\Big| = O(1)$ or equivalently

$$\Big|Cov\Big(D_2a_\lambda(X,\mu),X\Big)+2\Big| = O(\frac{1}{\mu})$$

was already shown in (4.23) since (cp. (4.26): $D_2a_\lambda(X,\mu) = D_2a_\lambda(\bar{X},1)$)

$$Cov\Big(D_2a_\lambda(X,\mu),X\Big) = E\Big(D_2a_\lambda(X,\mu)(X-\mu)\Big) = E\Big(\mu D_2a_\lambda(\bar{X},1)(\bar{X}-1)\Big).$$

m) For the proof of $\Big|\mu\Big(\mu Var(D_2a_\lambda(X,\mu))-4\Big)\Big| = O(1)$ or equivalently

$$|\mu Var(D_2a_\lambda(X,\mu))-4| = O(\frac{1}{\mu}), \tag{4.42}$$

let the following decomposition be considered:

$$\begin{aligned} & |\mu Var(D_2a_\lambda(X,\mu))-4| \\ = \;& |\mu Var(D_2a_\lambda(\bar{X},1))-4| \\ \leq \;& |E(\mu(D_2a_\lambda(\bar{X},1))^2)-4| + \mu|E^2(D_2a_\lambda(\bar{X},1))| \\ = \;& |E(\mu H_1(\bar{X}))-4| + \mu|E^2(H_2(\bar{X}))| \end{aligned} \tag{4.43}$$

with

$$H_1(\bar{X}) = (D_2a_\lambda(\bar{X},1))^2, \quad H_2(\bar{X}) = D_2a_\lambda(\bar{X},1).$$

Since $H_2(1) = 0$, Theorem 4.4 entails for the second term

$$O(\frac{1}{\mu}) = |E(H_2(\bar{X}))| = |E(D_2 a_\lambda(\bar{X}, 1))| = |E(D_2 a_\lambda(X, \mu))| \tag{4.44}$$

and thus $\mu |E^2(H_2(\bar{X}))| = \mu \cdot O(\frac{1}{\mu^2}) = O(\frac{1}{\mu})$. Application of formula (4.24) (Th. 4.4) and use of $H_1(1) = 0$, $H_1^{(2)}(1) = 8$, gives for the first term:

$$|E(\mu H_1(\bar{X})) - \frac{1}{2} H_1^{(2)}(1)| = |E(\mu H_1(\bar{X})) - 4| = O(\frac{1}{\mu}).$$

Both statements assert (4.43) to be of the order $O(\frac{1}{\mu})$, which establishes (4.42).

n) In order to prove

$$\left| E\Big(\big(D_2 a_\lambda(X, \mu) - E(D_2 a_\lambda(X, \mu))\big)^4\Big)\right| = O(\frac{1}{\mu^2}),$$

it suffices to show

$$|E((D_2 a_\lambda(X, \mu))^4)| = O(\frac{1}{\mu^2}), \tag{4.45}$$

$$|E^4(D_2 a_\lambda(X, \mu))| = O(\frac{1}{\mu^2}), \tag{4.46}$$

because the simple inequality $(x - y)^2 \leq 2x^2 + 2y^2$ $(x, y \in \mathbf{R})$, which gives $(x - y)^4 \leq 8x^4 + 8y^4$, can be applied to the term in question. Since (4.46) respectively the stronger result $\left|E^4\big(D_2 a_\lambda(X, \mu)\big)\right| = O(\frac{1}{\mu^4})$ was already shown in the preceding proof (see (4.44)), only (4.45) needs to be checked. This result is immediately verified applying formula (4.24), since for $H(\bar{X}) := (\frac{1}{\lambda+1}(1 - \bar{X}^{\lambda+1}))^4 = (D_2 a_\lambda(X, \mu))^4$ with especially $H(1) = H^{(1)}(1) = H^{(2)}(1) = 0$ holds

$$|E((D_2 a_\lambda(X, \mu))^4)| = |E(\mu H(\bar{X}))| = |E(\mu H(\bar{X})) - \frac{1}{2} H^{(2)}(1)| = O(\frac{1}{\mu}).$$

Hence the proof is complete. □

It should be noted that the statements in the preceding lemma also hold for the expressions multiplied with $1/\mu^r$ $(r \in \mathbf{R}^+)$, because μ is bounded away from zero (for example $\frac{1}{\mu}|E(a_\lambda(X, \mu))| \leq \frac{c}{\mu} \leq \frac{c}{\epsilon}$).

5. Approximation of the Test Statistic

In this chapter, a suitable approximation of the goodness–of–fit statistic for both distribution models, namely column–multinomial and Poisson, will be derived, whose asymptotic normality will be shown in the chapter following. The asymptotics considered are the "increasing–cells asymptotics" already illustrated in chapter 2, i.e. the expected total sample size $\mu^n_{++} = \sum_{j=1}^{J^n} \sum_{k=1}^{K} \mu^n_{jk}$ and the number of groups J^n tend towards infinity. This is indicated by the running index n, which was chosen to increase proportionally to the expected sum μ^n_{++} and in the case of column–multinomial sampling, where μ^n_{++} and the total sample size Y^n_{++} coincide, in particular to equal this quantity. Apart from that, let the notation be as in chapter 2, where the explicit formulae for information matrix and score vector can also be found, which, as seen there, are analytically identical. Hence $X^n = (X^n_{jk})_{j,k}$ is a $J^n \times K$ table with independent entries $X^n_{jk} \sim Pois(\mu^n_{jk})$ for $j = 1, \ldots, J^n$, $k = 1, \ldots, K$, $n \in \mathbf{N}$ and, analogously, Y^n is a contingency table with the columns being independent multinomials. Further, let both distribution models have the same underlying tables of expectations $\mu^n = (\mu^n_{jk})_{j,k}$.
In order to derive the normal distribution under the nullhypothesis

$$H_0 : \exists\, \theta_0 : \frac{\mu^n_{jk}}{\mu^n_{j+}} = \pi^n_{jk|C}(\theta_0) \quad \forall\, j,k,n \quad (\theta_0 \in \Theta \subset \mathbf{R}^S),$$

let, if not differently stated, for the single expectations $\mu^n_{jk} = \mu^n_{jk}(\theta_0) = \mu^n_{j+}\pi^n_{jk|C}(\theta_0)$ be assumed.
For the further considerations of this chapter let now the following assumptions be *generally* fulfilled:

(RC1) $\pi^n_{jk|C}(\theta)$ is continuously differentiable twice in θ for all j,k,n,

(RC2) $\exists\, \epsilon > 0 : \pi^n_{jk|C}(\theta) \geq \epsilon$ for all $j,k,n,\theta \in \bar{W}$,

(RC3) $\exists\, M > 0 :$ a) $\|D_\theta \pi^n_{jk|C}(\theta)\| < M$ for all $j,k,n,\theta \in \bar{W}$,
b) $\|D^2_\theta \pi^n_{jk|C}(\theta)\| < M$ for all $j,k,n,\theta \in \bar{W}$,

(LC0) $P(\hat{\mu}^n_{j+} > 0\, \forall\, j \in \{1,\ldots,J^n\}) \longrightarrow 1$

with $\hat{\mu}^n_{j+} = Y^n_{j+}$ if column multinomial and $\hat{\mu}^n_{j+} = X^n_{j+}$ if Poisson sampling is considered. The other conditions required in this chapter, which are not supposed to hold throughout, but will be stated every time needed, are as follows:

$$\begin{array}{ll} \text{(LC1)} & \frac{1}{n} I^n(\mu^n_{\cdot+}, \theta_0) \longrightarrow I_\infty \text{ positive definite,} \\ \text{(LC2)} & \sqrt{n}(\hat{\theta}^n - \theta_0) = O_p(1), \\ \text{(LC3)} & (\hat{\theta}^n - \theta_0) = (I^n(\mu^n_{\cdot+}, \theta_0))^{-1} U^n(\theta_0) + O_p(\frac{1}{n}), \\ \text{(BC)} & \exists\, \epsilon > 0 : \mu^n_{jk} \geq \epsilon \quad \text{for all } j, k, n, \\ \text{(MD3)} & \max_{1 \leq j \leq J^n} \pi^n_{jk|D}(\theta_0) \longrightarrow 0 \quad \text{for all } k. \end{array}$$

(LC1) and (MD3) will only be used for proofs concerning column–multinomial sampling. Moreover, for this distribution $\frac{J^n}{n} \longrightarrow 0$ will be required, which is a condition needed in the next chapter to accomplish the assumptions concerning the marginal distribution, especially (MD2) (cp. section 2.3).

For further considerations and proofs let now the following notation be introduced:

$$\begin{aligned} e_\lambda(\mu^n_{j+}, \pi^n_{jk|C}(\theta_0)) &:= E(a_\lambda(X^n_{jk}, \mu^n_{jk})), \quad \mu^n_{jk} = \mu^n_{j+}\pi^n_{jk|C}(\theta_0) && (5.1) \\ &\quad\; (j = 1, \ldots, J^n, k = 1, \ldots, K, n \in \mathbf{N}), \\ m^n_\lambda(\mu^n_{\cdot+}, \theta_0) &:= \sum_{j=1}^{J^n} \sum_{k=1}^{K} e_\lambda(\mu^n_{j+}, \pi^n_{jk|C}(\theta_0)) && (5.2) \\ &= E(SD^n_\lambda(\mu^n_{\cdot+}, \theta_0 | X^n)), \\ Z^n_\lambda(\mu^n_{\cdot+}, \theta_0) &:= SD^n_\lambda(\mu^n_{\cdot+}, \theta_0) - m^n_\lambda(\mu^n_{\cdot+}, \theta_0). && (5.3) \end{aligned}$$

The notational convention provided in chapter 2 let also be maintained, so that, for example, $Z^n_\lambda(\mu^n_{\cdot+}, \theta_0)$ will denote the centered statistic for both sampling schemes and $Z^n_\lambda(\mu^n_{\cdot+}, \theta_0 | X^n)$ the centered Poisson statistic.
For the following results a vector of Poisson expectations c^n_λ will repeatedly be used (cp. Lemma 5.5):

$$\begin{aligned} c^n_\lambda(\mu^n_{\cdot+}, \theta_0) &:= \sum_{j=1}^{J^n} \sum_{k=1}^{K} D_\theta \log \pi^n_{jk|C}(\theta_0) \cdot Cov(a_\lambda(X^n_{jk}, \mu^n_{jk}), X^n_{jk}) && (5.4) \\ &= -E\Big(D_\theta\Big(SD^n_\lambda(\mu^n_{\cdot+}, \theta_0|X^n) - m^n_\lambda(\mu^n_{\cdot+}, \theta_0)\Big)\Big) \\ &= Cov\Big(SD^n_\lambda(\mu^n_{\cdot+}, \theta_0|X^n), U^n(\theta_0|X^n)^T\Big). \end{aligned}$$

The last formula, stating c^n_λ to be the covariance between SD^n_λ and the score vector under Poisson distribution, is obvious by definition and will especially be needed in the next chapter.

To eliminate the correlations caused by estimating, the further considerations of this chapter will deal with a gradual derivation of an approximation for the test statistic, which does not depend on the estimators anymore. Although in case of Poisson distribution the approximation will yield a sum of independent random variables and thus allows application of the central limit theorem, the same does not hold if column–multinomial sampling is considered because of the underlying stochastic dependencies. This problem, however, will be circumvented applying Morris' method (1975), described in chapter 3. Since this approach provides a centering with the Poisson expectation m_λ^n, the approximation steps will, for both distribution models, be started by the statistic $SD_\lambda^n(\hat{\mu}_{\cdot +}^n, \hat{\theta}^n)$ centered with $m_\lambda^n(\hat{\mu}_{\cdot +}^n, \hat{\theta}^n)$, which has already been introduced as "centered statistic Z_λ^n". Moreover, it will be seen that the single reduction steps are invariant under the considered sampling schemes, hence both distributions can be treated together and yield, analytically, the same final approximation. For the proofs, however, a somehow different argumentation will be necessary. In particular, an auxiliary result concerning column–multinomial sampling is needed, which requires rather lengthy calculations and hence is given in section 5.1 prior to the actual approximation (section 5.2).

5.1 An Auxiliary Result for Column–Multinomial Sampling

The result given in the next lemma is only needed for one step of the approximation treated in section 5.2 and for the column–multinomial case only. What will actually be required is information concerning the asymptotic order of the difference between the gradient $D_\theta SD_\lambda^n$ and its Poisson expectation. If Poisson sampling and general λ or column–multinomial–sampling with λ being integer valued is considered, the order is easily determined using Chebyshev's inequality respectively elementary calculations. To treat the column–multinomial statistic centered with its Poisson expectation in the difficult case of λ being noninteger valued, Morris' approach (see chapter 3), which provides just this centering, can here also be applied to derive a normal limit and thus the asymptotic order of the difference.

For the following proof, several arguments could be adopted from Morris (1975), who not only stated basic results in his article, but also derived the asymptotic normality of the Likelihood Ratio Statistic in the simpler case of multinomial sampling and without parameter estimation (Morris, 1975, Theorem 5.2).

Lemma 5.1 *Consider the asymptotics $n \to \infty$ and suppose that the assumptions $\frac{1}{n} I^n(\mu_{\cdot +}^n, \theta_0) \longrightarrow I_\infty$ positive definite (LC1), $\mu_{jk}^n \geq \epsilon$ ($\epsilon > 0$ constant) for all j, k, n (BC) and $\max_{1 \leq j \leq J^n} \pi_{jk|D}^n(\theta_0) \longrightarrow 0$ for all k (MD3) as well as $\frac{J^n}{n} \longrightarrow 0$ hold. For any given $s \in \{1, \ldots, S\}$, let be v_λ^{n2} the corrected variance of $\frac{\partial}{\partial \theta_s} SD_\lambda^n(\mu_{\cdot +}^n, \theta_0 | X^n)$:*

$$v_\lambda^{n2} = Var\Big(\frac{\partial}{\partial \theta_s} SD_\lambda^n(\mu_{\cdot +}^n, \theta_0 | X^n)\Big) - \sum_{k=1}^{K} \mu_{+k}^n (\gamma_{\lambda k}^n)^2$$

$$= \quad Var\Big(\sum_{j=1}^{J^n}\sum_{k=1}^{K}\frac{\partial}{\partial\theta_s}a_\lambda(X_{jk}^n,\mu_{jk}^n(\theta_0))\Big) - \sum_{k=1}^{K}\mu_{+k}^n(\gamma_{\lambda k}^n)^2$$

with

$$\gamma_{\lambda k}^n \quad = \quad \frac{1}{\mu_{+k}^n}\sum_{j=1}^{J^n}Cov\Big(\frac{\partial}{\partial\theta_s}\sum_{k=1}^{K}a_\lambda(X_{jk}^n,\mu_{jk}^n(\theta_0)),X_{jk}^n\Big),$$

where for column–multinomial sampling the column sizes $\mu_{+k}^n = n_k$ are known for all k. Then for $\lambda \in (-1,1]$ follows

$$\begin{aligned}
&\frac{1}{v_\lambda^n}\Big(\frac{\partial}{\partial\theta_s}SD_\lambda^n(\mu_{\cdot+}^n,\theta_0|Y^n) - E\Big(\frac{\partial}{\partial\theta_s}SD_\lambda^n(\mu_{\cdot+}^n,\theta_0|X^n)\Big)\Big)\\
= \; &\frac{1}{v_\lambda^n}\Big(\sum_{j=1}^{J^n}\sum_{k=1}^{K}\frac{\partial}{\partial\theta_s}a_\lambda(Y_{jk}^n,\mu_{jk}^n(\theta_0)) - E\Big(\sum_{j=1}^{J^n}\sum_{k=1}^{K}\frac{\partial}{\partial\theta_s}a_\lambda(X_{jk}^n,\mu_{jk}^n(\theta_0))\Big)\Big)\\
&\xrightarrow{\mathcal{L}} N(0,1),
\end{aligned}$$

in particular, $\frac{1}{v_\lambda^{n2}} = O_e(n)$ holds, i.e. $v_\lambda^{n2} = O(n)$ and $n = O(v_\lambda^{n2})$.

Proof:
The result will be proved using the generalized version of Morris' (1975) "fundamental lemma" (Lemma 3.4). In order to apply the method described in section 3.1 and for the sake of notational clarity, let in the following the explicit representation of SD_λ^n as a sum be considered. For $j \in \{1,\ldots,J^n\}$, $n \in \mathbf{N}$, let now the given functions be denoted by $g_{\lambda j}^n$ as follows ($z = (z_{jk})_{j,k}$ is an arbitrary nonnegative $J^n \times K$ table):

$$\begin{aligned}
g_{\lambda j}^n : \mathbf{N}_0^K \quad &\to \quad \mathbf{R}\\
z_{j\cdot} \quad &\mapsto \quad g_{\lambda j}^n(z_{j\cdot}) = \sum_{k=1}^{K}\frac{\partial}{\partial\theta_s}a_\lambda(z_{jk},\mu_{jk}^n(\theta_0)).
\end{aligned}$$

Obviously, these functions do not fulfil the conditions concerning expected value and covariance required for Lemma 3.4, but as already performed in section 3.1, this can be accomplished by a suitable transformation choosing

$$\begin{aligned}
f_{\lambda j}^n(z_{j\cdot}) \quad &= \quad g_{\lambda j}^n(z_{j\cdot}) - E(g_{\lambda j}^n(X_{j\cdot})) - \sum_{k=1}^{K}\gamma_{\lambda k}^n(z_{jk}-\mu_{jk}^n),\\
\gamma_{\lambda k}^n \quad &= \quad \frac{1}{\mu_{+k}^n}\sum_{j=1}^{J^n}Cov(g_{\lambda j}^n(X_{j\cdot}^n),X_{jk}^n)\\
&= \quad \frac{1}{\mu_{+k}^n}\sum_{j=1}^{J^n}Cov\Big(\sum_{k=1}^{K}\frac{\partial}{\partial\theta_s}a_\lambda(X_{jk}^n,\mu_{jk}^n(\theta_0)),X_{jk}^n\Big)
\end{aligned}$$

$$= \frac{1}{\mu_{+k}^n} \sum_{j=1}^{J^n} Cov\Big(\frac{\partial}{\partial\theta_s} a_\lambda(X_{jk}^n, \mu_{jk}^n(\theta_0)), X_{jk}^n\Big),$$

thus getting $E(f_{\lambda j}^n(X_{j\cdot}^n)) = 0$ and $Cov(\sum_{j=1}^{J^n} f_{\lambda j}^n(X_{j\cdot}^n), \sum_{j=1}^{J^n} X_{jk}^n) = 0$ for each $k \in \{1,\ldots,K\}$ (see (3.5)). The stated variance v_λ^{n2} equals the variance of the transformation under Poisson sampling:

$$\begin{aligned} Var\Big(\sum_{j=1}^{J^n} f_{\lambda j}^n(X_{j\cdot}^n)\Big) &= Var\Big(\sum_{j=1}^{J^n} g_{\lambda j}^n(X_{j\cdot}^n) - \sum_{j=1}^{J^n}\sum_{k=1}^{K} \gamma_{\lambda k}^n (X_{jk}^n - \mu_{jk}^n)\Big) \\ &= Var\Big(\sum_{j=1}^{J^n} g_{\lambda j}^n(X_{j\cdot}^n)\Big) - \sum_{k=1}^{K} (\gamma_{\lambda k}^n)^2 \mu_{+k}^n \\ &= \sum_{j=1}^{J^n}\sum_{k=1}^{K} Var\Big(\frac{\partial}{\partial\theta_s} a_\lambda(X_{jk}^n, \mu_{jk}^n(\theta_0))\Big) - \sum_{k=1}^{K} (\gamma_{\lambda k}^n)^2 \mu_{+k}^n \\ &= v_\lambda^{n2} \end{aligned}$$

(compare (3.7)). Since the asymptotics $n \to \infty$, $J^n \to \infty$ are considered, condition (BC) entails an increase of the column sizes n_k $(k = 1,\ldots,K)$: $n_k = \mu_{+k}^n = \sum_{j=1}^{J^n} \mu_{jk}^n \geq \sum_{j=1}^{J^n} \epsilon = J^n\epsilon \longrightarrow \infty$. These considerations, in particular the choice of $f_{\lambda j}^n$ $(j = 1,\ldots,J^n,\ n \in \mathbf{N})$ and the assumptions made, assure that the situation of Lemma 3.4 is met. Since further the probabilities p_{jk}^n from Lemma 3.4 agree with the $\pi_{jk|D}^n(\theta_0)$ considered here, requirement (3.9) from Lemma 3.4 equals (MD3) and hence holds already by assumption. Thus for every $\lambda \in (-1, 1]$, the conditions (3.10) – (3.12) remain to be checked in order to give the desired result:

$$\begin{aligned} &\frac{1}{\Big(\sum_{j=1}^{J^n} Var(f_{\lambda j}^n(X_{j\cdot}^n))\Big)^{\frac{1}{2}}} \sum_{j=1}^{J^n} f_{\lambda j}^n(Y_{j\cdot}^n) \\ = &\frac{1}{v_\lambda^n}\Big(\sum_{j=1}^{J^n}\sum_{k=1}^{K} \frac{\partial}{\partial\theta_s} a_\lambda(Y_{jk}^n, \mu_{jk}^n(\theta_0)) - E\Big(\sum_{j=1}^{J^n}\sum_{k=1}^{K} \frac{\partial}{\partial\theta_s} a_\lambda(X_{jk}^n, \mu_{jk}^n(\theta_0))\Big)\Big) \xrightarrow{\mathcal{L}} N(0,1) \end{aligned}$$

for all $\lambda \in (-1, 1]$. It remains to be shown that the standardization of the Poisson statistic $\sum_{j=1}^{J^n} f_{\lambda j}^n(X_{j\cdot}^n)$ meets the Ljapounov Condition for the central limit theorem, thus yielding asymptotic normality (3.10) and the Feller Condition (3.11). Secondly (3.12), which grants convergence of the conditional distribution, must be checked. Since both proofs expect the variance to fulfill $n/v_\lambda^{n2} = O(1)$, this will be proved first. Therefore more generally $v_\lambda^{n2} = O_e(n)$ will be shown, which additionally requires the easily verified inverse condition $v_\lambda^{n2}/n = O(1)$. Hence the rest of this proof divides into three parts:

(*i*) $v_\lambda^{n2} = O_e(n)$,

$$(ii) \quad \frac{1}{v_\lambda^{n4}} \sum_{j=1}^{J^n} E\Big((f_{\lambda j}^n(X_{j\cdot}^n))^4 \Big) \longrightarrow 0 \quad \text{(Ljapounov Condition)},$$

$$(iii) \quad \lim_{h \to 0} \sup_n \sup_{v^n} \frac{1}{v_\lambda^{n2}} E\bigg(\Big(\sum_{j=1}^{J^n} (f_{\lambda j}^n(L_{j\cdot}^n + M_{j\cdot}^n) - f_{\lambda j}^n(L_{j\cdot}^n)) \Big)^2 \bigg) = 0 \quad (\lambda \in (-1,1]).$$

Here the column multinomials L^n and M^n as well as the sequence v^n and $h \in \mathbf{R}^K$ are defined as in Lemma 3.4. Condition (iii) will only be checked for $\lambda \in (-1,1]$, whereas (i) and (ii) are easily derived for arbitrary $\lambda > -1$.

(i) To determine the exact order of $v_\lambda^{n2} = \sum_{j=1}^{J^n} \sum_{k=1}^K Var\Big(\frac{\partial}{\partial \theta_s} a_\lambda(X_{jk}^n, \mu_{jk}^n(\theta_0))\Big) - \sum_{k=1}^K (\gamma_{\lambda k}^n)^2 \mu_{+k}^n$ let $\gamma_{\lambda k}^n$ be considered first. Simple calculations using $\frac{\partial}{\partial \theta_s}\mu_{jk}^n(\theta_0) = \mu_{j+}^n \frac{\partial}{\partial \theta_s} \pi_{jk|C}^n(\theta_0) = \mu_{j+}^n \pi_{jk|C}^n(\theta_0) \frac{\partial}{\partial \theta_s} \log \pi_{jk|C}^n(\theta_0) = \mu_{jk}^n \cdot \frac{\partial}{\partial \theta_s} \log \pi_{jk|C}^n(\theta_0)$ give

$$\begin{aligned}
\gamma_{\lambda k}^n &= \frac{1}{\mu_{+k}^n} \sum_{j=1}^{J^n} Cov\Big(\frac{\partial}{\partial \theta_s} a_\lambda(X_{jk}^n, \mu_{jk}^n(\theta_0)), X_{jk}^n \Big) \\
&= \frac{1}{\mu_{+k}^n} \sum_{j=1}^{J^n} \frac{\partial}{\partial \theta_s} \mu_{jk}^n(\theta_0) \cdot \Big(Cov\Big(D_2 a_\lambda(X_{jk}^n, \mu_{jk}^n), X_{jk}^n \Big) + 2 - 2 \Big) \\
&= -\frac{2}{\mu_{+k}^n} \sum_{j=1}^{J^n} \frac{\partial}{\partial \theta_s} \mu_{jk}^n(\theta_0) + \frac{1}{\mu_{+k}^n} \sum_{j=1}^{J^n} \frac{\partial}{\partial \theta_s} \mu_{jk}^n(\theta_0) \cdot \Big(Cov\Big(D_2 a_\lambda(X_{jk}^n, \mu_{jk}^n), X_{jk}^n \Big) + 2 \Big) \\
&= -\frac{2}{\mu_{+k}^n} \sum_{j=1}^{J^n} \mu_{+k}^n \frac{\partial}{\partial \theta_s} \pi_{jk|D}^n(\theta_0) \\
&\quad + \frac{1}{\mu_{+k}^n} \sum_{j=1}^{J^n} (\frac{\partial}{\partial \theta_s} \log \pi_{jk|C}^n(\theta_0)) \cdot \mu_{jk}^n \Big(Cov\Big(D_2 a_\lambda(X_{jk}^n, \mu_{jk}^n), X_{jk}^n \Big) + 2 \Big) \\
&= \frac{1}{\mu_{+k}^n} \sum_{j=1}^{J^n} (\frac{\partial}{\partial \theta_s} \log \pi_{jk|C}^n(\theta_0)) \cdot \mu_{jk}^n \Big(Cov\Big(D_2 a_\lambda(X_{jk}^n, \mu_{jk}^n), X_{jk}^n \Big) + 2 \Big).
\end{aligned}$$

Since $\mu_{jk}^n(Cov(D_2 a_\lambda(X_{jk}^n, \mu_{jk}^n), X_{jk}^n) + 2)$ and $\frac{\partial}{\partial \theta_s} \log \pi_{jk|C}^n(\theta_0) = \frac{1}{\pi_{jk|C}^n(\theta_0)} \cdot \frac{\partial}{\partial \theta_s} \pi_{jk|C}^n(\theta_0)$ are for all j, k, n bounded by Lemma 4.8 1), respectively (RC2) and (RC3), there exists a positive constant c such that

$$\begin{aligned}
\frac{\mu_{+k}^n}{J^n} \cdot |\gamma_{\lambda k}^n| &\leq \frac{\mu_{+k}^n}{J^n} \cdot \frac{1}{\mu_{+k}^n} \sum_{j=1}^{J^n} \Big| \frac{\partial}{\partial \theta_s} \log \pi_{jk|C}^n(\theta_0) \Big| \cdot \Big| \mu_{jk}^n \Big(Cov(D_2 a_\lambda(X_{jk}^n, \mu_{jk}^n), X_{jk}^n) + 2 \Big) \Big| \\
&\leq \frac{1}{J^n} \sum_{j=1}^{J^n} c \\
&= c
\end{aligned}$$

for all k, n holds, thus giving

$$\gamma_{\lambda k}^n = O(\frac{J^n}{\mu_{+k}^n}) \quad \text{for all } k. \tag{5.5}$$

The boundedness of $\frac{J^n}{n_k} = \frac{J^n}{\mu_{+k}^n}$ and hence $\frac{J^n}{n} = \frac{J^n}{\mu_{++}^n}$, which is implied by condition (BC), e.g. since $n = \sum_{j=1}^{J^n}\sum_{k=1}^{K}\mu_{jk}^n \geq \sum_{j=1}^{J^n}\sum_{k=1}^{K}\epsilon = J^n K\epsilon$, already yields for the second term of $v_\lambda^{n2} = \sum_{j=1}^{J^n}\sum_{k=1}^{K} Var\Big(\frac{\partial}{\partial\theta_s}a_\lambda(X_{jk}^n, \mu_{jk}^n(\theta_0))\Big) - \sum_{k=1}^{K}(\gamma_{\lambda k}^n)^2\mu_{+k}^n$ the order $O(J^n)$ and hence $O(n)$:

$$\sum_{k=1}^{K}(\gamma_{\lambda k}^n)^2\mu_{+k}^n = O(1)\sum_{k=1}^{K}(\frac{J^n}{\mu_{+k}^n})^2\mu_{+k}^n = O(1)J^n\sum_{k=1}^{K}\frac{J^n}{\mu_{+k}^n} = O(J^n).$$

To obtain $v_\lambda^{n2} = O(n)$, it thus suffices to show $\sum_{j=1}^{J^n}\sum_{k=1}^{K} Var\Big(\frac{\partial}{\partial\theta_s}a_\lambda(X_{jk}^n, \mu_{jk}^n(\theta_0))\Big)$ $= O(n)$. In the opposite case, where

$$\frac{n}{v_\lambda^{n2}} = \frac{1}{\frac{1}{n}\sum_{j=1}^{J^n}\sum_{k=1}^{K} Var(\frac{\partial}{\partial\theta_s}a_\lambda(X_{jk}^n, \mu_{jk}^n(\theta_0))) - \frac{1}{n}\sum_{k=1}^{K}(\gamma_{\lambda k}^n)^2\mu_{+k}^n} = O(1)$$

has to be checked, the stronger condition $\frac{J^n}{n} \longrightarrow 0$, which is used only at this point of the proof, immediately yields $\frac{1}{n}\sum_{k=1}^{K}(\gamma_{\lambda k}^n)^2\mu_{+k}^n = o(1)$. Consequently, the proof of $v_\lambda^{n2} = O_e(n)$ thus reduces to

$$\frac{1}{n}\sum_{j=1}^{J^n}\sum_{k=1}^{K} Var\Big(\frac{\partial}{\partial\theta_s}a_\lambda(X_{jk}^n, \mu_{jk}^n(\theta_0))\Big) = O_e(1). \tag{5.6}$$

Simple calculations now give

$$\begin{aligned}
&\frac{1}{n}\sum_{j=1}^{J^n}\sum_{k=1}^{K} Var\Big(\frac{\partial}{\partial\theta_s}a_\lambda(X_{jk}^n, \mu_{jk}^n(\theta_0))\Big)\\
= \;&\frac{1}{n}\sum_{j=1}^{J^n}\sum_{k=1}^{K} Var\Big(\frac{\partial}{\partial\theta_s}\mu_{jk}^n(\theta_0)\cdot D_2a_\lambda(X_{jk}^n, \mu_{jk}^n)\Big)\\
= \;&\frac{1}{n}\sum_{j=1}^{J^n}\sum_{k=1}^{K}(\mu_{j+}^n)^2(\frac{\partial}{\partial\theta_s}\pi_{jk|C}^n(\theta_0))^2 Var\Big(D_2a_\lambda(X_{jk}^n, \mu_{jk}^n)\Big)\\
= \;&\frac{1}{n}\sum_{j=1}^{J^n}\mu_{j+}^n\sum_{k=1}^{K}(\frac{\partial}{\partial\theta_s}\pi_{jk|C}^n(\theta_0))^2\frac{1}{\pi_{jk|C}^n(\theta_0)}\mu_{jk}^n Var\Big(D_2a_\lambda(X_{jk}^n, \mu_{jk}^n)\Big).
\end{aligned}$$

Provided for all j, k, n holds $\mu_{jk}^n Var(D_2a_\lambda(X_{jk}^n, \mu_{jk}^n)) \in [c_1, c_2]$ with c_1 and c_2 being positive constants, the lower bound can be obtained as follows:

$$\frac{1}{n}\sum_{j=1}^{J^n}\mu_{j+}^n\sum_{k=1}^{K}(\frac{\partial}{\partial\theta_s}\pi_{jk|C}^n(\theta_0))^2\frac{1}{\pi_{jk|C}^n(\theta_0)}\mu_{jk}^n Var\Big(D_2a_\lambda(X_{jk}^n, \mu_{jk}^n)\Big)$$

$$\begin{aligned} &\geq \frac{1}{n}\sum_{j=1}^{J^n}\mu_{j+}^n\sum_{k=1}^{K}(\frac{\partial}{\partial\theta_s}\pi_{jk|C}^n(\theta_0))^2\frac{1}{\pi_{jk|C}^n(\theta_0)}c_1 \\ &= c_1\frac{1}{n}(I^n(\mu_{\cdot+}^n,\theta_0))_{s,s} \end{aligned}$$

with $(I^n(\mu_{\cdot+}^n,\theta_0))_{s,s}$ denoting the s–th diagonal element of the information matrix. $\frac{1}{n}(I^n(\mu_{\cdot+}^n,\theta_0))_{s,s}$ is positive and has by assumption (LC1), i.e. $\frac{1}{n}I^n(\mu_{\cdot+}^n,\theta_0)\longrightarrow I_\infty$ positive definite, a positive limit thus giving

$$c_1\frac{1}{n}(I^n(\mu_{\cdot+}^n,\theta_0))_{s,s}\geq c_1\inf_n\frac{1}{n}(I^n(\mu_{\cdot+}^n,\theta_0))_{s,s}>0.$$

For the upper bound follows analogously

$$\begin{aligned} &\frac{1}{n}\sum_{j=1}^{J^n}\mu_{j+}^n\sum_{k=1}^{K}(\frac{\partial}{\partial\theta_s}\pi_{jk|C}^n(\theta_0))^2\frac{1}{\pi_{jk|C}^n(\theta_0)}\mu_{jk}^n Var\Big(D_2a_\lambda(X_{jk}^n,\mu_{jk}^n)\Big) \\ &\leq c_2\sup_n\frac{1}{n}(I^n(\mu_{\cdot+}^n,\theta_0))_{s,s}<\infty \end{aligned}$$

and hence (5.6). To see the remaining argument

$$\mu_{jk}^n Var\Big(D_2a_\lambda(X_{jk}^n,\mu_{jk}^n)\Big)\in[c_1,c_2]\subset\mathbf{R}^+\quad\text{for all } j,k,n, \tag{5.7}$$

consider a Poisson distributed random variable X with expected value $\mu\geq\epsilon>0$. If for the variance term regarded as a function of μ holds

$$\mu Var\Big(D_2a_\lambda(X,\mu)\Big)\in[c_1,c_2]\subset\mathbf{R}^+\quad\text{for all }\mu\in[\epsilon,\infty),$$

(5.7) immediately follows since $\mu_{jk}^n\geq\epsilon>0$ for all j,k,n (BC). Now Lemma 4.8 m) guarantees the existence of a constant $c>0$ such that

$$\Big|\mu Var\Big(D_2a_\lambda(X,\mu)\Big)-4\Big|\leq\frac{c}{\mu}\leq\frac{c}{\epsilon}\quad\text{for all }\mu\geq\epsilon, \tag{5.8}$$

hence $\mu Var(D_2a_\lambda(X,\mu))\leq c_2$ for all $\mu\geq\epsilon$ is given. The other inequality follows if $\mu Var(D_2a_\lambda(X,\mu))$ is continuous in μ, has a positive limit for increasing μ, and does not equal zero, thus yielding

$$\mu Var\Big(D_2a_\lambda(X,\mu)\Big)\geq\inf_{\mu\in[\epsilon,\infty)}\mu Var\Big(D_2a_\lambda(X,\mu)\Big)=:c_1>0.$$

Now $\mu Var(D_2a_\lambda(X,\mu))$ is continuous in μ by Th. 4.1 since this holds for $D_2a_\lambda(X,\mu)$ which is further dominated by an exponential function. $\lim_{\mu\to\infty}\mu Var(D_2a_\lambda(X,\mu))=4$ is given by (5.8). To prove $\mu Var(D_2a_\lambda(X,\mu))>0$, suppose $Var(D_2a_\lambda(X,\mu))=0$. Then by definition holds

$$Var\Big(D_2a_\lambda(X,\mu)\Big)\;=\;Var\Big(\frac{2}{\lambda+1}(1-(\frac{X}{\mu})^{\lambda+1})\Big)$$

$$
\begin{aligned}
&= (\frac{2}{\lambda+1})^2 E\Big(\big((\frac{X}{\mu})^{\lambda+1} - E\big((\frac{X}{\mu})^{\lambda+1}\big)\big)^2\Big) \\
&= 0.
\end{aligned}
$$

This yields $P\Big((\frac{X}{\mu})^{\lambda+1} = E((\frac{X}{\mu})^{\lambda+1})\Big) = 1$ and hence the distribution of $(\frac{X}{\mu})^{\lambda+1}$ is degenerated. Since $\lambda > -1$, this contradicts $X \sim Pois(\mu)$ with $\mu \geq \epsilon > 0$.

(ii) Using $v_\lambda^{n2} = O_e(n)$, which especially means $\frac{1}{v_\lambda^{n2}} = \frac{n}{v_\lambda^{n2}} \cdot \frac{1}{n} = O(1) \cdot \frac{1}{n}$, it suffices to show

$$
\sum_{j=1}^{J^n} E\Big((f_{\lambda j}^n(X_{j\cdot}^n))^4\Big) = o(n^2) \qquad (n \to \infty)
$$

to assure the validity of the Ljapounov condition. Applying the simple inequality $(x-y)^2 \leq 2x^2 + 2y^2$ $(x, y \in \mathbf{R})$ twice, splits the sum into two parts,

$$
\begin{aligned}
&\sum_{j=1}^{J^n} E\Big((f_{\lambda j}^n(X_{j\cdot}^n))^4\Big) \\
= \ &\sum_{j=1}^{J^n} E\Big(\Big(g_{\lambda j}^n(X_{j\cdot}^n) - E(g_{\lambda j}^n(X_{j\cdot}^n)) - \sum_{k=1}^{K} \gamma_{\lambda k}^n (X_{jk}^n - \mu_{jk}^n)\Big)^4\Big) \\
\leq \ & 8\sum_{j=1}^{J^n} E\Big(\Big(g_{\lambda j}^n(X_{j\cdot}^n) - E(g_{\lambda j}^n(X_{j\cdot}^n))\Big)^4\Big) + 8\sum_{j=1}^{J^n} E\Big(\Big(\sum_{k=1}^{K} \gamma_{\lambda k}^n (X_{jk}^n - \mu_{jk}^n)\Big)^4\Big), \qquad (5.9)
\end{aligned}
$$

for which the asymptotic order $o(n^2)$ will be determined separately. Now for real valued variables $x_1, \ldots, x_K$ generally holds

$$
\begin{aligned}
(\sum_{k=1}^{K} x_k)^2 &= \sum_{k=1}^{K}\sum_{i=1}^{K} x_k x_i = \sum_{k=1}^{K} x_k^2 + \sum_{k<i} 2x_k x_i \\
&\leq \sum_{k=1}^{K} x_k^2 + \sum_{k<i}(x_k^2 + x_i^2) = \sum_{k=1}^{K} x_k^2 + \sum_{k=1}^{K}(K-1)x_k^2 = \sum_{k=1}^{K} K \cdot x_k^2 \qquad (5.10)
\end{aligned}
$$

ensuing

$$
(\sum_{k=1}^{K} x_k)^4 = \Big((\sum_{k=1}^{K} x_k)^2\Big)^2 \leq \Big(K\sum_{k=1}^{K} x_k^2\Big)^2 \leq K^3 \sum_{k=1}^{K} x_k^4, \qquad (5.11)
$$

hence for the first term of (5.9) follows

$$
\begin{aligned}
&\sum_{j=1}^{J^n} E\Big(\Big(g_{\lambda j}^n(X_{j\cdot}^n) - E(g_{\lambda j}^n(X_{j\cdot}^n))\Big)^4\Big) \\
= \ &\sum_{j=1}^{J^n} E\Big(\Big(\sum_{k=1}^{K}\Big(\frac{\partial}{\partial\theta_s} a_\lambda(X_{jk}^n, \mu_{jk}^n(\theta_0)) - E(\frac{\partial}{\partial\theta_s} a_\lambda(X_{jk}^n, \mu_{jk}^n(\theta_0)))\Big)\Big)^4\Big)
\end{aligned}
$$

$$
\begin{aligned}
&\leq K^3 \sum_{j=1}^{J^n} E\Big(\sum_{k=1}^{K} \big(\frac{\partial}{\partial \theta_s} a_\lambda(X_{jk}^n, \mu_{jk}^n(\theta_0)) - E(\frac{\partial}{\partial \theta_s} a_\lambda(X_{jk}^n, \mu_{jk}^n(\theta_0)))\big)^4 \Big) \\
&= K^3 \sum_{j=1}^{J^n} \sum_{k=1}^{K} (\frac{\partial}{\partial \theta_s} \log \pi_{jk|C}^n(\theta_0))^4 (\mu_{jk}^n)^4 \\
&\qquad \cdot E\Big(\big(D_2 a_\lambda(X_{jk}^n, \mu_{jk}^n) - E(D_2 a_\lambda(X_{jk}^n, \mu_{jk}^n))\big)^4 \Big).
\end{aligned}
$$

Since $\frac{\partial}{\partial \theta_s} \log \pi_{jk|C}^n(\theta_0)$ and $(\mu_{jk}^n)^2 E\Big(\big(D_2 a_\lambda(X_{jk}^n, \mu_{jk}^n) - E(D_2 a_\lambda(X_{jk}^n, \mu_{jk}^n))\big)^4 \Big)$ are for all j, k, n bounded by assumption (RC2) and (RC3) respectively Lemma 4.8 n), the terms of the last sum have the order $O((\mu_{jk}^n)^2)$. This, $\frac{\mu_{+k}^n}{n} = \frac{n_k}{n} < 1$ and condition (MD3) then yield

$$
\begin{aligned}
\sum_{j=1}^{J^n} E\Big(\big(g_{\lambda j}^n(X_{j\cdot}^n) - E(g_{\lambda j}^n(X_{j\cdot}^n))\big)^4 \Big) &= O(1) \sum_{j=1}^{J^n} \sum_{k=1}^{K} (\mu_{jk}^n)^2 \\
&= O(1) \sum_{k=1}^{K} (\mu_{+k}^n)^2 \sum_{j=1}^{J^n} (\pi_{jk|D}^n)^2 \\
&\leq O(1) \sum_{k=1}^{K} (\mu_{+k}^n)^2 \cdot \max_{1 \leq j \leq J^n} \pi_{jk|D}^n \\
&= o(n^2).
\end{aligned}
$$

For the second term of (5.9) similarly follows

$$
\begin{aligned}
\sum_{j=1}^{J^n} E\Big(\big(\sum_{k=1}^{K} \gamma_{\lambda k}^n (X_{jk}^n - \mu_{jk}^n)\big)^4 \Big) &\leq K^3 \sum_{j=1}^{J^n} E\Big(\sum_{k=1}^{K} (\gamma_{\lambda k}^n)^4 (X_{jk}^n - \mu_{jk}^n)^4 \Big) \\
&= K^3 \sum_{k=1}^{K} (\gamma_{\lambda k}^n)^4 \sum_{j=1}^{J^n} E\Big((X_{jk}^n - \mu_{jk}^n)^4 \Big) \\
&= K^3 \sum_{k=1}^{K} (\gamma_{\lambda k}^n)^4 \sum_{j=1}^{J^n} (3(\mu_{jk}^n)^2 + \mu_{jk}^n) \\
&= O(1) \sum_{k=1}^{K} \sum_{j=1}^{J^n} (\mu_{jk}^n)^2,
\end{aligned}
$$

where the last equality holds using (5.5) and (BC), in particular $\gamma_{\lambda k}^n = O(\frac{J^n}{\mu_{+k}^n}) = O(1)$ and $\frac{1}{\mu_{jk}^n} \leq \frac{1}{\epsilon}$ for all j, k, n. Argumentation as before finally gives for the second term of (5.9), too, the asymptotic order $o(n^2)$. Hence the Ljapounov Condition is verified.

(iii) For the proof of (iii) let the notation be as in Lemma 3.4: $v^n = (v_1^n, \ldots, v_K^n)^T$ is a bounded sequence, L^n and M^n are product–multinomial distributed $J^n \times K$

tables with columns $L^n_{\cdot k} = (L^n_{1k}, \ldots, L^n_{J^n k})^T \sim Multi_{J^n}(n_k + v^n_k\sqrt{n_k}, \pi^n_{\cdot k|D})$ and $M^n_{\cdot k} \sim Multi_{J^n}(h_k\sqrt{n_k}, \pi^n_{\cdot k|D})$ having the same underlying probability vector $\pi^n_{\cdot k|D} = \pi^n_{\cdot k|D}(\theta_0)$ for each $k = 1, \ldots, K$. The sequence v^n und $h = (h_1, \ldots, h_K)^T \in \mathbf{R}^K$ are assumed to be such that for each k the sizes $l^n_k = n_k + v^n_k\sqrt{n_k}$ and $m^n_k = h_k\sqrt{n_k}$ are nonnegative integers. Further, all columns $L^n_{\cdot 1}, \ldots, L^n_{\cdot K}, M^n_{\cdot 1}, \ldots, M^n_{\cdot K}$ are supposed to be stochastically independent.
Using $\frac{n}{v^{n2}_\lambda} = O(1)$, as shown in (i), the result follows, if for $\lambda \in (-1, 1]$

$$\frac{1}{n}E\Big(\Big(\sum_{j=1}^{J^n}\Big(f^n_{\lambda j}(L^n_{j\cdot} + M^n_{j\cdot}) - f^n_{\lambda j}(L^n_{j\cdot})\Big)\Big)^2\Big) = \|h\|_{max} \cdot O(1)$$

holds. Inserting the definition of the functions $f^n_{\lambda j}$ yields for the expected value

$$\begin{aligned}
& E\Big(\Big(\sum_{j=1}^{J^n}\Big(f^n_{\lambda j}(L^n_{j\cdot} + M^n_{j\cdot}) - f^n_{\lambda j}(L^n_{j\cdot})\Big)\Big)^2\Big) \\
= \;& E\Big(\Big(\sum_{j=1}^{J^n}\Big(g^n_{\lambda j}(L^n_{j\cdot} + M^n_{j\cdot}) - g^n_{\lambda j}(L^n_{j\cdot})\Big) - \sum_{j=1}^{J^n}\sum_{k=1}^{K}\gamma^n_{\lambda k}M^n_{jk}\Big)^2\Big) \\
\leq \;& 2E\Big(\Big(\sum_{j=1}^{J^n}\Big(g^n_{\lambda j}(L^n_{j\cdot} + M^n_{j\cdot}) - g^n_{\lambda j}(L^n_{j\cdot})\Big)\Big)^2\Big) + 2E\Big(\Big(\sum_{j=1}^{J^n}\sum_{k=1}^{K}\gamma^n_{\lambda k}M^n_{jk}\Big)^2\Big) \\
= \;& 2E\Big(\Big(\sum_{j=1}^{J^n}\sum_{k=1}^{K}\Big(\frac{\partial}{\partial\theta_s}a_\lambda(L^n_{jk} + M^n_{jk}, \mu^n_{jk}(\theta_0)) - \frac{\partial}{\partial\theta_s}a_\lambda(L^n_{jk}, \mu^n_{jk}(\theta_0))\Big)\Big)^2\Big) \\
& +2E\Big(\Big(\sum_{k=1}^{K}\gamma^n_{\lambda k}\sum_{j=1}^{J^n}M^n_{jk}\Big)^2\Big).
\end{aligned}$$

The general inequality (5.10), (5.5), i.e. $\gamma^n_{\lambda k} = O(\frac{J^n}{\mu^n_{+k}})$, and $\frac{J^n}{\mu^n_{+k}} = \frac{J^n}{n_k} = O(1)$ for all k, gives immediately for the second term

$$\begin{aligned}
E\Big((\sum_{k=1}^{K}\gamma^n_{\lambda k}\sum_{j=1}^{J^n}M^n_{jk})^2\Big) &= (\sum_{k=1}^{K}\gamma^n_{\lambda k}m^n_k)^2 \\
&\leq K\sum_{k=1}^{K}(\gamma^n_{\lambda k}m^n_k)^2 \\
&= K\sum_{k=1}^{K}(\gamma^n_{\lambda k})^2 h^2_k n_k \\
&\leq K \cdot \|h\|^2_{max} \cdot \sum_{k=1}^{K}(\gamma^n_{\lambda k})^2 n_k
\end{aligned}$$

$$
\begin{aligned}
&= \|h\|^2_{max} \cdot O(1) \sum_{k=1}^{K} (\frac{J^n}{n_k})^2 n_k \\
&= \|h\|^2_{max} \cdot O(J^n)
\end{aligned}
$$

and since $\frac{J^n}{n}$ is bounded the order $\|h\|^2_{max} \cdot O(n)$. Hence it remains to establish

$$
E\Big(\big(\sum_{j=1}^{J^n}\sum_{k=1}^{K}\big(\frac{\partial}{\partial\theta_s}a_\lambda(L^n_{jk}+M^n_{jk},\mu^n_{jk}(\theta_0))-\frac{\partial}{\partial\theta_s}a_\lambda(L^n_{jk},\mu^n_{jk}(\theta_0))\big)\big)^2\Big)=\|h\|_{max}\cdot O(n). \tag{5.12}
$$

Obvious inequalities, the assumed regularity conditions, which imply $|\frac{\partial}{\partial\theta_s}\log\pi^n_{jk|C}(\theta_0)| \leq c_1$ constant for all j,k,n and Lemma 8.1 b) now give

$$
\begin{aligned}
& E\Big(\big(\sum_{k=1}^{K}\sum_{j=1}^{J^n}\big(\frac{\partial}{\partial\theta_s}a_\lambda(L^n_{jk}+M^n_{jk},\mu^n_{jk}(\theta_0))-\frac{\partial}{\partial\theta_s}a_\lambda(L^n_{jk},\mu^n_{jk}(\theta_0))\big)\big)^2\Big) \\
\leq\;& K\sum_{k=1}^{K}E\Big(\big(\sum_{j=1}^{J^n}\big(\frac{\partial}{\partial\theta_s}a_\lambda(L^n_{jk}+M^n_{jk},\mu^n_{jk}(\theta_0))-\frac{\partial}{\partial\theta_s}a_\lambda(L^n_{jk},\mu^n_{jk}(\theta_0))\big)\big)^2\Big) \\
\leq\;& K\sum_{k=1}^{K}E\Big(\big(\sum_{j=1}^{J^n}|\frac{\partial}{\partial\theta_s}\log\pi^n_{jk|C}(\theta_0)| \\
& \qquad \cdot\mu^n_{jk}\cdot\Big|D_2a_\lambda(L^n_{jk}+M^n_{jk},\mu^n_{jk})-D_2a_\lambda(L^n_{jk},\mu^n_{jk})\Big|\big)^2\Big) \\
\leq\;& K\sum_{k=1}^{K}E\Big(\big(\sum_{j=1}^{J^n}c_1\mu^n_{jk}\Big|\frac{2}{\lambda+1}\Big(1-(\frac{L^n_{jk}+M^n_{jk}}{\mu^n_{jk}})^{\lambda+1}\Big)-\frac{2}{\lambda+1}\Big(1-(\frac{L^n_{jk}}{\mu^n_{jk}})^{\lambda+1}\Big)\Big|\big)^2\Big) \\
=\;& K\sum_{k=1}^{K}E\Big(\big(\sum_{j=1}^{J^n}c_1\mu^n_{jk}\frac{2}{\lambda+1}\Big|-(\frac{L^n_{jk}+M^n_{jk}}{\mu^n_{jk}})^{\lambda+1}+(\frac{L^n_{jk}}{\mu^n_{jk}})^{\lambda+1}\Big|\big)^2\Big) \\
=\;& Kc_1^2\sum_{k=1}^{K}E\Big(\big(\sum_{j=1}^{J^n}\mu^n_{jk}\frac{2}{\lambda+1}\big((\frac{L^n_{jk}+M^n_{jk}}{\mu^n_{jk}})^{\lambda+1}-(\frac{L^n_{jk}}{\mu^n_{jk}})^{\lambda+1}\big)\big)^2\Big) \\
\leq\;& Kc_1^2\sum_{k=1}^{K}E\Big(\big(\sum_{j=1}^{J^n}c\big(\frac{(M^n_{jk})^2}{\mu^n_{jk}}+M^n_{jk}h(L^n_{jk},\mu^n_{jk})\big)\big)^2\Big) \\
\leq\;& Kc_1^2c^2\sum_{k=1}^{K}2E\Big(\big(\sum_{j=1}^{J^n}\frac{(M^n_{jk})^2}{\mu^n_{jk}}\big)^2+\big(\sum_{j=1}^{J^n}M^n_{jk}h(L^n_{jk},\mu^n_{jk})\big)^2\Big)
\end{aligned}
$$

with $h(L^n_{jk},\mu^n_{jk}) = 1+\frac{L^n_{jk}}{\mu^n_{jk}}+\frac{2\mu^n_{jk}}{L^n_{jk}+1}$ and $c=\max\{2,\frac{2}{\lambda+1}\}$. For the proof of (5.12) it thus suffices to show

$$
E\Big(\big(\sum_{j=1}^{J^n}\frac{(M^n_{jk})^2}{\mu^n_{jk}}\big)^2\Big)+E\Big(\big(\sum_{j=1}^{J^n}M^n_{jk}h(L^n_{jk},\mu^n_{jk})\big)^2\Big)=h_k\cdot O(n_k) \quad \text{for all } k. \tag{5.13}
$$

Now let in the following any $k \in \{1, \ldots, K\}$ be given. Using the notation $x^{(j)} := \prod_{i=1}^{j}(x - i + 1)$ for $i, j \in \mathbf{N}$ and the Kronecker–delta δ_{ij}, for the first expectation holds

$$
\begin{aligned}
& E\Big(\big(\sum_{j=1}^{J^n} \frac{(M^n_{jk})^2}{\mu^n_{jk}}\big)^2\Big) \\
= \;& E\Big(\big(\sum_{j=1}^{J^n} \big(\frac{(M^n_{jk})^{(2)}}{\mu^n_{jk}} + \frac{M^n_{jk}}{\mu^n_{jk}}\big)\big)^2\Big) \\
\leq \;& E\Big(2\big(\sum_{j=1}^{J^n} \frac{(M^n_{jk})^{(2)}}{\mu^n_{jk}}\big)^2\Big) + E\Big(2\big(\sum_{j=1}^{J^n} \frac{M^n_{jk}}{\mu^n_{jk}}\big)^2\Big) \\
= \;& 2E\Big(\sum_{j=1}^{J^n}\sum_{i=1}^{J^n} \frac{(M^n_{jk})^{(2)}(M^n_{ik})^{(2)}}{\mu^n_{jk}\mu^n_{ik}}\Big) + 2E\Big(\sum_{j=1}^{J^n}\sum_{i=1}^{J^n} \frac{M^n_{jk}M^n_{ik}}{\mu^n_{jk}\mu^n_{ik}}\Big) \\
= \;& 2\sum_{j=1}^{J^n}\sum_{i=1}^{J^n} \frac{1}{\mu^n_{jk}\mu^n_{ik}} \Big((m^n_k)^{(4)}(\pi^n_{jk|D})^2(\pi^n_{ik|D})^2 \\
& +\delta_{ij}\Big(4(m^n_k)^{(3)}(\pi^n_{jk|D})^3 + 2(m^n_k)^{(2)}(\pi^n_{jk|D})^2\Big)\Big) + 2\sum_{j=1}^{J^n} E\Big(\frac{M^n_{jk}}{\mu^n_{jk}} \sum_{i=1}^{J^n} \frac{M^n_{ik}}{\mu^n_{ik}}\Big) \qquad (5.14) \\
\leq \;& 2\sum_{j=1}^{J^n}\sum_{i=1}^{J^n} \frac{1}{\mu^n_{jk}\mu^n_{ik}} \Big((m^n_k)^4(\pi^n_{jk|D})^2(\pi^n_{ik|D})^2 \\
& +\delta_{ij}\Big(4(m^n_k)^3(\pi^n_{jk|D})^3 + 2(m^n_k)^2(\pi^n_{jk|D})^2\Big)\Big) + 2\sum_{j=1}^{J^n} \frac{E(M^n_{jk})}{\mu^n_{jk}} \cdot \frac{m^n_k}{\epsilon} \qquad (5.15) \\
= \;& 2\Big(\sum_{j=1}^{J^n}\sum_{i=1}^{J^n} \frac{h^4_k n^2_k (\pi^n_{jk|D})^2(\pi^n_{ik|D})^2}{n^2_k \pi^n_{jk|D}\pi^n_{ik|D}} + 4\sum_{j=1}^{J^n} \frac{h^3_k n^{\frac{3}{2}}_k (\pi^n_{jk|D})^3}{n^2_k(\pi^n_{jk|D})^2} \\
& +2\sum_{j=1}^{J^n} \frac{h^2_k n_k (\pi^n_{jk|D})^2}{n^2_k(\pi^n_{jk|D})^2} + \sum_{j=1}^{J^n} \frac{h^2_k n_k \pi^n_{jk|D}}{n_k \pi^n_{jk|D}\epsilon}\Big) \\
= \;& 2\Big(h^4_k + h^3_k \cdot \frac{4}{n^{1/2}_k} + h^2_k \cdot \frac{2J^n}{n_k} + h^2_k \cdot \frac{J^n}{\epsilon}\Big) \\
= \;& h_k \cdot O(J^n). \qquad (5.16)
\end{aligned}
$$

In (5.14) the factorial moments of the multinomial distribution (see Johnson/Kotz, 1969) have been inserted (in the case $i = j$ through simple transformations), and (5.15) and (5.16) deduce from assumption (BC), which implies $\frac{1}{\mu^n_{jk}} \leq \frac{1}{\epsilon}$ for all j, k, n and $\frac{J^n}{\mu^n_{+k}} = \frac{J^n}{n_k} = O(1)$. Using the independence of $M^n_{\cdot k}$ and $L^n_{\cdot k}$, for the second term of (5.13) follows:

$$
\begin{aligned}
& E\Big(\Big(\sum_{j=1}^{J^n} M_{jk}^n \cdot h(L_{jk}^n, \mu_{jk}^n)\Big)^2\Big) \\
= \;& E\Big(\sum_{j=1}^{J^n}\sum_{i=1}^{J^n} M_{jk}^n M_{ik}^n h(L_{jk}^n, \mu_{jk}^n) h(L_{ik}^n, \mu_{ik}^n)\Big) \\
= \;& \sum_{j=1}^{J^n} E((M_{jk}^n)^2) E(h^2(L_{jk}^n, \mu_{jk}^n)) + \sum_{j=1}^{J^n}\sum_{\substack{i=1\\ i<j}}^{J^n} E(M_{jk}^n M_{ik}^n) E(2h(L_{jk}^n, \mu_{jk}^n) h(L_{ik}^n, \mu_{ik}^n)) \\
\le \;& \sum_{j=1}^{J^n} E((M_{jk}^n)^2) E(h^2(L_{jk}^n, \mu_{jk}^n)) \\
& + \sum_{j=1}^{J^n}\sum_{\substack{i=1\\ i<j}}^{J^n} E(M_{jk}^n M_{ik}^n) E(h^2(L_{jk}^n, \mu_{jk}^n) + h^2(L_{ik}^n, \mu_{ik}^n)) \\
= \;& \sum_{j=1}^{J^n} E((M_{jk}^n)^2) E(h^2(L_{jk}^n, \mu_{jk}^n)) + \sum_{j=1}^{J^n}\sum_{\substack{i=1\\ i<j}}^{J^n} E(M_{jk}^n M_{ik}^n) E(h^2(L_{jk}^n, \mu_{jk}^n)) \\
& + \sum_{j=1}^{J^n}\sum_{\substack{i=1\\ i>j}}^{J^n} E(M_{jk}^n M_{ik}^n) E(h^2(L_{jk}^n, \mu_{jk}^n)) \\
= \;& \sum_{j=1}^{J^n}\sum_{i=1}^{J^n} E(M_{jk}^n M_{ik}^n) E(h^2(L_{jk}^n, \mu_{jk}^n)) \\
= \;& \sum_{j=1}^{J^n} E(h^2(L_{jk}^n, \mu_{jk}^n)) E(M_{jk}^n \sum_{i=1}^{J^n} M_{ik}^n) \\
= \;& \sum_{j=1}^{J^n} m_k^n E(M_{jk}^n) E(h^2(L_{jk}^n, \mu_{jk}^n)). \qquad (5.17)
\end{aligned}
$$

Now provided that

$$E(h^2(L_{jk}^n, \mu_{jk}^n)) = O(1) \text{ for all } j,k \qquad (5.18)$$

holds, which will be proved in the following, the second term of (5.13) also has the stated order:

$$
\begin{aligned}
\sum_{j=1}^{J^n} m_k^n E(M_{jk}^n) E(h^2(L_{jk}^n, \mu_{jk}^n)) &= \sum_{j=1}^{J^n} (m_k^n)^2 \pi_{jk|D}^n \cdot O(1) \\
&= \sum_{j=1}^{J^n} h_k^2 n_k \pi_{jk|D}^n \cdot O(1)
\end{aligned}
$$

$$= \quad h_k \cdot O(n_k).$$

Thus only (5.18) remains to be checked. Applying (5.10) now yields ($k \in \{1, \dots, K\}$):

$$E(h^2(L^n_{jk}, \mu^n_{jk})) = E\Big((1 + \frac{L^n_{jk}}{\mu^n_{jk}} + \frac{2\mu^n_{jk}}{L^n_{jk}+1})^2\Big) \le 3 + 3E\Big((\frac{L^n_{jk}}{\mu^n_{jk}})^2\Big) + 3E\Big((\frac{2\mu^n_{jk}}{L^n_{jk}+1})^2\Big).$$

Since (BC) entails $\frac{1}{\mu^n_{jk}} = \frac{1}{n_k \pi^n_{jk|D}} = O(1)$ for all j, k and further

$$\frac{l^n_k}{n_k} = \frac{n_k + v^n_k \sqrt{n_k}}{n_k} = O(1), \quad \frac{n_k}{l^n_k} = O(1) \quad \text{for all } k$$

holds, the first expectation is bounded:

$$\begin{aligned} E\Big((\frac{L^n_{jk}}{\mu^n_{jk}})^2\Big) &= \frac{1}{(\mu^n_{jk})^2}\Big(l^n_k \pi^n_{jk|D}(1-\pi^n_{jk|D}) + (l^n_k \pi^n_{jk|D})^2\Big) \\ &= \frac{l^n_k \pi^n_{jk|D} - l^n_k (\pi^n_{jk|D})^2 + (l^n_k \pi^n_{jk|D})^2}{(n_k \pi^n_{jk|D})^2} \\ &= \frac{l^n_k}{n_k} \cdot \frac{1}{n_k \pi^n_{jk|D}} - \frac{l^n_k}{n_k} \cdot \frac{1}{n_k} + (\frac{l^n_k}{n_k})^2. \end{aligned}$$

Using $L^n_{jk} \ge 0 \Leftrightarrow \frac{1}{L^n_{jk}+1} \le \frac{2}{L^n_{jk}+2}$ and the following formula for the inverse factorial moments of a binomial distributed random variable $x \sim B(n,p)$ (Johnson/Kotz, 1992),

$$E\left(((x+i)^{(i)})^{-1}\right) = \left(((n+i)^{(i)} p^i)^{-1}\right) \cdot \left(1 - \sum_{j=0}^{i-1} \binom{n+i}{j} p^j (1-p)^{n+i-j}\right),$$

the second expectation is also bounded:

$$\begin{aligned} (\mu^n_{jk})^2 E\Big((\frac{1}{L^n_{jk}+1})^2\Big) &\le (\mu^n_{jk})^2 E\Big(\frac{2}{(L^n_{jk}+1)(L^n_{jk}+2)}\Big) \\ &= 2(\mu^n_{jk})^2 \frac{1 - (1-\pi^n_{jk|D})^{l^n_k+2} - (l^n_k+2)\pi^n_{jk|D}(1-\pi^n_{jk|D})^{l^n_k+1}}{(l^n_k+2)(l^n_k+1)(\pi^n_{jk|D})^2} \\ &\le 2(\mu^n_{jk})^2 \frac{1}{(l^n_k+2)(l^n_k+1)(\pi^n_{jk|D})^2} \\ &\le \frac{2n_k^2(\pi^n_{jk|D})^2}{(l^n_k \pi^n_{jk|D})^2} \\ &= 2(\frac{n_k}{l^n_k})^2 \\ &= O(1). \end{aligned}$$

Hence (5.18) is shown, thus completing the proof. □

5.2 Approximation Steps

In the following, i.e. in Lemma 5.3 – 5.6 and 5.8, the centered goodness–of–fit statistic $Z^n_\lambda(\hat{\mu}^n_{\cdot+}, \hat{\theta}^n) = SD^n_\lambda(\hat{\mu}^n_{\cdot+}, \hat{\theta}^n) - m^n_\lambda(\hat{\mu}^n_{\cdot+}, \hat{\theta}^n)$ will be gradually approximated through the "true" statistic $SD^n_\lambda(\mu^n_{\cdot+}, \theta_0) - m^n_\lambda(\mu^n_{\cdot+}, \theta_0)$ and additional correction terms. As explained in the beginning, the centering term $m^n_\lambda(\mu^n_{\cdot+}, \theta_0)$ is the same Poisson expectation for both distribution models (see (5.2)). In Corollary 5.9 stated thereafter, all steps will be summarized. This corollary thus in particular gives an informative overview of the proceeding in this section.

The now following Lemma 5.2 is an auxiliary result needed in the next chapter for the variance estimation and only stated here because the first approximation (Lemma 5.3) is proved by analogous argumentation.

Lemma 5.2 *Consider the asymptotics $n \to \infty$ and suppose that $\sqrt{n}(\hat{\theta}^n - \theta_0) = O_p(1)$ (LC2) holds. Let be given a compact and convex neighbourhood $\bar{W} \subset \Theta \subset \mathbf{R}^S$ of θ_0 and a function*

$$\begin{aligned} H^n : \mathbf{R}^S \supset \Theta &\to \mathbf{R} \\ \theta &\mapsto H^n(\hat{\mu}^n_{\cdot+}, \theta) \\ &= \sum_{j=1}^{J^n} \sum_{k=1}^{K} h(\hat{\mu}^n_{j+}, f^n_{jk}(\theta)), \end{aligned}$$

with h and f^n_{jk} defined as follows:

$$\begin{aligned} f^n_{jk} : \mathbf{R}^S \supset \Theta &\to \mathbf{R}^m \quad (m \in \mathbf{N}) \\ \theta &\mapsto f^n_{jk}(\theta), \\ h : \mathbf{R}^+_0 \times \mathbf{R}^m &\to \mathbf{R}, \\ (x, f^n_{jk}(\theta)) &\mapsto h(x, f^n_{jk}(\theta)). \end{aligned}$$

For $x > 0$ let be $h(x, f^n_{jk}(\theta))$ continuously differentiable in θ.

a) *If there exists a constant $c > 0$ such that for every j, k and n holds $\sup_{\theta \in \bar{W}} \|D_\theta h(x, f^n_{jk}(\theta))\| \leq x \cdot c$ for all $x > 0$, then follows $\frac{1}{\sqrt{n}}(H^n(\hat{\mu}^n_{\cdot+}, \hat{\theta}^n) - H^n(\hat{\mu}^n_{\cdot+}, \theta_0)) = O_p(1) \quad (n \to \infty)$.*

b) *If there exist constants $c > 0$ and $\epsilon \in (0,1)$ such that for every j, k and n holds $\sup_{\theta \in \bar{W}} \|D_\theta h(x, f^n_{jk}(\theta))\| \leq c$ for all $x \geq \epsilon$, then follows $\frac{\sqrt{n}}{J^n}(H^n(\hat{\mu}^n_{\cdot+}, \hat{\theta}^n) - H^n(\hat{\mu}^n_{\cdot+}, \theta_0)) = O_p(1) \quad (n \to \infty)$.*

Proof:
a) Let any $\delta \in (0,1)$ be given. To be shown now is the existence of a constant M_δ such that for almost all $n \in \mathbf{N}$ holds

$$P(\frac{1}{\sqrt{n}}||H^n(\hat{\mu}^n_{\cdot+}, \hat{\theta}^n) - H^n(\hat{\mu}^n_{\cdot+}, \theta_0)|| > M_\delta) \leq \delta$$

with $\hat{\mu}^n_{\cdot+}$ being the vector of row sums $Y^n_{\cdot+}$ resp. $X^n_{\cdot+}$. To enable Taylor expansion in θ, which requires $\hat{\mu}^n_{\cdot+}$ being positive, consider the following decomposition:

$$\begin{aligned}
& P(\frac{1}{\sqrt{n}}||H^n(\hat{\mu}^n_{\cdot+}, \hat{\theta}^n) - H^n(\hat{\mu}^n_{\cdot+}, \theta_0)|| > M_\delta) \\
\leq \; & P(\frac{1}{\sqrt{n}}||H^n(\hat{\mu}^n_{\cdot+}, \hat{\theta}^n) - H^n(\hat{\mu}^n_{\cdot+}, \theta_0)|| > M_\delta \wedge \hat{\theta}^n \in \bar{W} \wedge \hat{\mu}^n_{j+} > 0 \, \forall j) \\
+ \; & P(\hat{\theta}^n \notin \bar{W} \vee \exists j : \hat{\mu}^n_{j+} = 0). \qquad (5.19)
\end{aligned}$$

Assumption (LC0) $P(\hat{\mu}^n_{j+} > 0 \, \forall j) \longrightarrow 1$ and the convergence of $\hat{\theta}^n$ to $\theta_0 \in \bar{W}$ in probability implied by (LC2), guarantee the existence of a constant $n_0 \in \mathbf{N}$ such that for all $n > n_0$

$$P(\hat{\theta}^n \notin \bar{W}) \leq \frac{\delta}{3}, \quad P(\exists j : \hat{\mu}^n_{j+} = 0) \leq \frac{\delta}{3}$$

holds. Hence for the second probability in (5.19) follows

$$\begin{aligned}
P(\hat{\theta}^n \notin \bar{W} \vee \exists j : \hat{\mu}^n_{j+} = 0) \; & \leq \; P(\hat{\theta}^n \notin \bar{W}) + P(\exists j : \hat{\mu}^n_{j+} = 0) \\
& \leq \; \frac{2}{3}\delta \qquad (5.20)
\end{aligned}$$

for almost all n. To study the first term of (5.19), let $\theta^n_z = \theta_0 + z(\hat{\theta}^n - \theta_0)$, $z \in [0,1]$, be a value between $\hat{\theta}^n$ and the true parameter θ_0. For $\hat{\theta}^n \in \bar{W}$ and hence $\theta^n_z \in \bar{W}$ and $\hat{\mu}^n_{j+} > 0$ for all j, application of the mean value theorem (0-th order Taylor expansion) gives:

$$\begin{aligned}
& \frac{1}{\sqrt{n}}||H^n(\hat{\mu}^n_{\cdot+}, \hat{\theta}^n) - H^n(\hat{\mu}^n_{\cdot+}, \theta_0)|| \\
= \; & \frac{1}{\sqrt{n}}|| \int_0^1 D_\theta H^n(\hat{\mu}^n_{\cdot+}, \theta^n_z) dz \cdot (\hat{\theta}^n - \theta_0)|| \\
\leq \; & \frac{1}{\sqrt{n}}||\hat{\theta}^n - \theta_0|| \cdot || \int_0^1 \sum_{j=1}^{J^n} \sum_{k=1}^{K} D_\theta h(\hat{\mu}^n_{j+}, f^n_{jk}(\theta^n_z)) dz|| \\
\leq \; & \frac{1}{\sqrt{n}}||\hat{\theta}^n - \theta_0|| \cdot \sum_{j=1}^{J^n} \sum_{k=1}^{K} || \int_0^1 D_\theta h(\hat{\mu}^n_{j+}, f^n_{jk}(\theta^n_z)) dz|| \\
\leq \; & \frac{1}{\sqrt{n}}||\hat{\theta}^n - \theta_0|| \cdot \sum_{j=1}^{J^n} \sum_{k=1}^{K} \sup_{\theta \in \bar{W}} ||D_\theta h(\hat{\mu}^n_{j+}, f^n_{jk}(\theta))||
\end{aligned}$$

$$\begin{aligned}
&\leq \frac{1}{\sqrt{n}}\|\hat{\theta}^n - \theta_0\| \cdot \sum_{j=1}^{J^n}\sum_{k=1}^{K} \hat{\mu}^n_{j+} \cdot c \\
&= \|(\hat{\theta}^n - \theta_0)\cdot\sqrt{n}\| \cdot \frac{\hat{\mu}^n_{++}}{n} \cdot c.
\end{aligned}$$

In case of Poisson distribution, the Poisson limit theorem immediately yields $\frac{\hat{\mu}^n_{++}}{\mu^n_{++}} = \frac{X^n_{++}}{\mu^n_{++}} \xrightarrow{P} 1$, which together with the assumed proportional increase of n and μ^n_{++} gives $\frac{\hat{\mu}^n_{++}}{n} = \frac{X^n_{++}}{\mu^n_{++}} \cdot \frac{\mu^n_{++}}{n} = O_p(1)$. If column–multinomial sampling is considered, $\hat{\mu}^n_{++} = Y^n_{++} = \mu^n_{++} = n$ holds and hence $\frac{\hat{\mu}^n_{++}}{n} = 1$. These arguments together with assumption $\sqrt{n}(\hat{\theta}^n - \theta_0) = O_p(1)$ (LC2) now guarantee the existence of a bound M_δ such that for almost all n holds

$$\begin{aligned}
&P\Big(\frac{1}{\sqrt{n}}\|H^n(\hat{\mu}^n_{\cdot+}, \hat{\theta}^n) - H^n(\hat{\mu}^n_{\cdot+}, \theta_0)\| > M_\delta \wedge \hat{\theta}^n \in \bar{W} \wedge \hat{\mu}^n_{j+} > 0\,\forall\, j\Big) \\
&\leq P\Big(\|(\hat{\theta}^n - \theta_0)\cdot\sqrt{n}\| \cdot \frac{\hat{\mu}^n_{++}}{n} \cdot c > M_\delta\Big) \\
&\leq \frac{\delta}{3}. \qquad (5.21)
\end{aligned}$$

Inserting (5.20) and (5.21) in (5.19) finally establishes the result.

b) Taylor expansion as in a) and use of $\sup_{\theta\in\bar{W}} \|D_\theta h(\hat{\mu}^n_{j+}, f^n_{jk}(\theta))\| \leq c$ for all j, k, n, gives for $\hat{\theta}^n \in \bar{W}$ and $\hat{\mu}^n_{j+} \geq 1 > \epsilon$ for all j the inequality

$$\begin{aligned}
\|H^n(\hat{\mu}^n_{\cdot+}, \hat{\theta}^n) - H^n(\hat{\mu}^n_{\cdot+}, \theta_0)\| &\leq \|\hat{\theta}^n - \theta_0\| \cdot \sum_{j=1}^{J^n}\sum_{k=1}^{K} \sup_{\theta\in\bar{W}} \|D_\theta h(\hat{\mu}^n_{j+}, f^n_{jk}(\theta))\| \\
&\leq \|(\hat{\theta}^n - \theta_0)\cdot\sqrt{n}\| \cdot \frac{J^n}{\sqrt{n}} \cdot K \cdot c
\end{aligned}$$

and hence $\frac{\sqrt{n}}{J^n}(H^n(\hat{\mu}^n_{\cdot+}, \hat{\theta}^n) - H^n(\hat{\mu}^n_{\cdot+}, \theta_0))$ being stochastically bounded. □

In the following lemma, a first order Taylor expansion in $\hat{\theta}^n$ around θ_0 gives a linear approximation of the centered goodness–of–fit statistic with an error bounded in probability.

Lemma 5.3 *Suppose the assumptions (BC) and (LC2) hold, i.e.* $\mu^n_{jk} \geq \epsilon > 0$ *for all* j, k, n *and* $\sqrt{n}(\hat{\theta}^n - \theta_0) = O_p(1)$. *Then for the centered statistic* Z^n_λ *defined in (5.3) follows*

$$Z^n_\lambda(\hat{\mu}^n_{\cdot+}, \hat{\theta}^n) - Z^n_\lambda(\hat{\mu}^n_{\cdot+}, \theta_0) - D_\theta Z^n_\lambda(\hat{\mu}^n_{\cdot+}, \theta_0)\cdot(\hat{\theta}^n - \theta_0) = O_p(1) \quad (n \to \infty).$$

Proof:
The proof of the statement will be structured similarly to the proof of the preceding lemma, thus considering a Taylor expansion for positive row sums $\hat{\mu}^n_{1+}, \ldots, \hat{\mu}^n_{J^n+}$ and using $\hat{\mu}^n_{++}/n = O_p(1)$. Since in contrast to Lemma 5.2 a concrete statistic is considered, here a more detailed investigation of the error term will be necessary, requiring additional analytical arguments such as $\frac{Y^n_{jk}}{Y^n_{j+}}$ resp. $\frac{X^n_{jk}}{X^n_{j+}} \in [0,1]$, which obviously hold for both sampling schemes. Hence it suffices to consider Y^n ($\hat{\mu}^n_{\cdot+} = Y^n_{\cdot+}$), representing both column multinomial and Poisson distribution (the following arguments also hold writing X^n instead of Y^n with $\hat{\mu}^n_{\cdot+} = X^n_{\cdot+}$).
Let now be $\theta^n_z := \theta_0 + z(\hat{\theta}^n - \theta_0)$, $z \in [0,1]$, a value between $\hat{\theta}^n$ and the true parameter θ_0. Further let be $\bar{W} \subset \Theta$ a convex compact neighborhood with $\theta_0 \in \bar{W} \subset \Theta$. For positive $\hat{\mu}^n_{\cdot+}$ ($\hat{\mu}^n_{j+} > 0$ for all j), $\hat{\theta}^n \in \bar{W}$ and hence $\theta^n_z \in \bar{W}$, first order Taylor expansion of $Z^n_\lambda(\hat{\mu}^n_{\cdot+}, \hat{\theta}^n)$ in $\hat{\theta}^n$ around θ_0 yields (see also Lemma 5.2):

$$\begin{aligned}
& \|Z^n_\lambda(\hat{\mu}^n_{\cdot+}, \hat{\theta}^n) - Z^n_\lambda(\hat{\mu}^n_{\cdot+}, \theta_0) - D_\theta Z^n_\lambda(\hat{\mu}^n_{\cdot+}, \theta_0) \cdot (\hat{\theta}^n - \theta_0)\| \\
= \;& \|(\hat{\theta}^n - \theta_0)^T \int_0^1 (1-z) D^2_\theta Z^n_\lambda(\hat{\mu}^n_{\cdot+}, \theta^n_z) dz \, (\hat{\theta}^n - \theta_0)\| \\
\leq \;& \|\hat{\theta}^n - \theta_0\|^2 \\
& \cdot \int_0^1 (1-z) \sum_{j=1}^{J^n} \sum_{k=1}^{K} \left\| D^2_\theta \Big(a_\lambda(Y^n_{jk}, \hat{\mu}^n_{j+} \pi^n_{jk|C}(\theta^n_z)) - e_\lambda(\hat{\mu}^n_{j+}, \pi^n_{jk|C}(\theta^n_z)) \Big) \right\| dz \\
\leq \;& \|\hat{\theta}^n - \theta_0\|^2 \\
& \cdot \sum_{j=1}^{J^n} \sum_{k=1}^{K} \sup_{\theta \in \bar{W}} \left\| D^2_\theta \Big(a_\lambda(Y^n_{jk}, \hat{\mu}^n_{j+} \pi^n_{jk|C}(\theta)) - e_\lambda(\hat{\mu}^n_{j+}, \pi^n_{jk|C}(\theta)) \Big) \right\|
\end{aligned}$$

with e_λ being the Poisson expectation of a_λ (see (5.1) respectively (5.1) – (5.3)). If now for all j, k, n holds

$$\sup_{\theta \in \bar{W}} \left\| D^2_\theta \Big(a_\lambda(Y^n_{jk}, \hat{\mu}^n_{j+} \pi^n_{jk|C}(\theta)) - e_\lambda(\hat{\mu}^n_{j+}, \pi^n_{jk|C}(\theta)) \Big) \right\| \leq \hat{\mu}^n_{j+} \cdot c, \tag{5.22}$$

with $c \in \mathbf{R}^+$ constant, then for the considered case, $\hat{\theta}^n \in \bar{W}$ and $\hat{\mu}^n_{j+} \geq 1$ for all j, further follows

$$\begin{aligned}
& \|Z^n_\lambda(\hat{\mu}^n_{\cdot+}, \hat{\theta}^n) - Z^n_\lambda(\hat{\mu}^n_{\cdot+}, \theta_0) - D_\theta Z^n_\lambda(\hat{\mu}^n_{\cdot+}, \theta_0) \cdot (\hat{\theta}^n - \theta_0)\| \\
\leq \;& \|\hat{\theta}^n - \theta_0\|^2 \cdot \sum_{j=1}^{J^n} \sum_{k=1}^{K} \hat{\mu}^n_{j+} \cdot c \\
= \;& \|\hat{\theta}^n - \theta_0\|^2 \cdot \hat{\mu}^n_{++} \cdot K \cdot c \\
= \;& \|(\hat{\theta}^n - \theta_0)\sqrt{n}\|^2 \cdot \frac{\hat{\mu}^n_{++}}{n} \cdot K \cdot c.
\end{aligned}$$

Analogous arguments as in the proof of Lemma 5.2 a) then yield the result.

Now it remains to establish (5.22) for $\hat{\mu}^n_{j+} \geq 1$. As already mentioned, from here on only analytical arguments which hold for both distribution models apply. Considering the distance function first, the existence of a constant $c \in \mathbf{R}^+$ will be shown with

$$\sup_{\theta \in \bar{W}} \left\| D^2_\theta \left(a_\lambda(Y^n_{jk}, \hat{\mu}^n_{j+}\pi^n_{jk|C}(\theta)) \right) \right\| \leq \hat{\mu}^n_{j+} \cdot c \quad \text{for all } j,k,n \quad (\hat{\mu}^n_{j+} \geq 1). \quad (5.23)$$

The term in norm brackets can now be stated as follows

$$\begin{aligned} & D^2_\theta a_\lambda(Y^n_{jk}, \hat{\mu}^n_{j+}\pi^n_{jk|C}(\theta)) \\ = \; & \hat{\mu}^n_{j+} \cdot D^2_\theta a_\lambda(\frac{Y^n_{jk}}{\hat{\mu}^n_{j+}}, \pi^n_{jk|C}(\theta)) \\ = \; & \hat{\mu}^n_{j+} \cdot \Big(D^T_\theta \pi^n_{jk|C}(\theta) \cdot D^2_2 a_\lambda(\frac{Y^n_{jk}}{\hat{\mu}^n_{j+}}, \pi^n_{jk|C}(\theta)) \cdot D_\theta \pi^n_{jk|C}(\theta) \\ & + D_2 a_\lambda(\frac{Y^n_{jk}}{\hat{\mu}^n_{j+}}, \pi^n_{jk|C}(\theta)) \cdot D^2_\theta \pi^n_{jk|C}(\theta) \Big). \end{aligned} \quad (5.24)$$

For $\theta \in \bar{W}$ the generally assumed regularity condition (RC3) assures $D_\theta \pi^n_{jk|C}(\theta)$ and $D^2_\theta \pi^n_{jk|C}(\theta)$ being bounded for all j,k,n. The derivatives

$$D_2 a_\lambda(\frac{Y^n_{jk}}{\hat{\mu}^n_{j+}}, \pi^n_{jk|C}(\theta)) = \frac{2}{\lambda+1}\Big(1 - (\frac{Y^n_{jk}}{\hat{\mu}^n_{j+}} \cdot \frac{1}{\pi^n_{jk|C}(\theta)})^{\lambda+1}\Big)$$

and

$$D^2_2 a_\lambda(\frac{Y^n_{jk}}{\hat{\mu}^n_{j+}}, \pi^n_{jk|C}(\theta)) = 2(\frac{Y^n_{jk}}{\hat{\mu}^n_{j+}})^{\lambda+1}(\frac{1}{\pi^n_{jk|C}(\theta)})^{\lambda+2}$$

are obviously also bounded, since it holds $\frac{Y^n_{jk}}{\hat{\mu}^n_{j+}} \in [0,1]$ and $\pi^n_{jk|C}(\theta) \geq \epsilon > 0$ for every j,k,n and $\theta \in \bar{W}$ (RC2). These bounding results for the terms in (5.24) thus give (5.23).

The inequality for the second term of (5.22),

$$\sup_{\theta \in \bar{W}} \| D^2_\theta e_\lambda(\hat{\mu}^n_{j+}, \pi^n_{jk|C}(\theta)) \| \leq \hat{\mu}^n_{j+} \cdot c \quad \text{for all } j,k,n \quad (c \in \mathbf{R}^+ \text{ constant}), \quad (5.25)$$

with $\hat{\mu}^n_{j+} \geq 1$ being the row sums of a column multinomial resp. Poisson contingency table, will be verified by showing that for each j,k,n holds

$$\sup_{\theta \in \bar{W}} \| D^2_\theta e_\lambda(\mu^n_{j+}, \pi^n_{jk|C}(\theta)) \| \leq \mu^n_{j+} \cdot c \quad \text{for all } \mu^n_{j+} \in [K\epsilon, \infty) \quad (5.26)$$

($\epsilon \leq \frac{1}{K}$ without loss of generality). This result combined with $\hat{\mu}^n_{j+} \geq 1$ then establishes (5.25) for both distribution models. Using $\mu^n_{jk}(\theta) = \mu^n_{j+} \cdot \pi^n_{jk|C}(\theta)$ and e_λ being a Poisson expectation by definition, $e_\lambda(\mu^n_{j+}, \pi^n_{jk|C}(\theta)) = E_{\mu^n_{jk}(\theta)}(a_\lambda(X^n_{jk}, \mu^n_{jk}(\theta))) =:$

$E(a_\lambda(X^n_{jk},\mu^n_{jk}(\theta)))$ yields

$$\begin{aligned}
& D^2_\theta e_\lambda(\mu^n_{j+},\pi^n_{jk|C}(\theta)) \\
= \; & D^T_\theta \mu^n_{jk}(\theta)\cdot\frac{\partial^2}{(\partial\mu^n_{jk}(\theta))^2}E(a_\lambda(X^n_{jk},\mu^n_{jk}(\theta)))\cdot D_\theta\mu^n_{jk}(\theta) \\
& +\frac{\partial}{\partial\mu^n_{jk}(\theta)}E(a_\lambda(X^n_{jk},\mu^n_{jk}(\theta)))\cdot D^2_\theta\mu^n_{jk}(\theta) \\
= \; & (\mu^n_{j+})^2\cdot D^T_\theta\pi^n_{jk|C}(\theta)\cdot\frac{\partial^2}{(\partial\mu^n_{jk}(\theta))^2}E(a_\lambda(X^n_{jk},\mu^n_{jk}(\theta)))\cdot D_\theta\pi^n_{jk|C}(\theta) \\
& +\mu^n_{j+}\cdot\frac{\partial}{\partial\mu^n_{jk}(\theta)}E(a_\lambda(X^n_{jk},\mu^n_{jk}(\theta)))\cdot D^2_\theta\pi^n_{jk|C}(\theta) \\
= \; & \mu^n_{j+}\Big(\frac{\mu^n_{j+}}{\mu^n_{jk}(\theta)}\cdot D^T_\theta\pi^n_{jk|C}(\theta)\cdot\mu^n_{jk}(\theta)\frac{\partial^2}{(\partial\mu^n_{jk}(\theta))^2}E(a_\lambda(X^n_{jk},\mu^n_{jk}(\theta)))\cdot D_\theta\pi^n_{jk|C}(\theta)\Big) \\
& +\mu^n_{j+}\Big(\frac{\partial}{\partial\mu^n_{jk}(\theta)}E(a_\lambda(X^n_{jk},\mu^n_{jk}(\theta)))\cdot D^2_\theta\pi^n_{jk|C}(\theta)\Big).
\end{aligned}$$

Since for $\theta\in\bar W$ $\mu^n_{jk}(\theta)=\mu^n_{j+}\pi^n_{jk|C}(\theta)$ is bounded away from zero for all j,k,n using regularity condition (RC2) and $\mu^n_{j+}\geq K\epsilon$ (BC), Lemma 4.8 shows the boundedness of $\frac{\partial}{\partial\mu^n_{jk}(\theta)}E(a_\lambda(X^n_{jk},\mu^n_{jk}(\theta)))$ and $\mu^n_{jk}(\theta)\frac{\partial^2}{(\partial\mu^n_{jk}(\theta))^2}E(a_\lambda(X^n_{jk},\mu^n_{jk}(\theta)))$, since it holds (statement g) and h))

$$\begin{aligned}
\Big|\mu\frac{\partial}{\partial\mu}E\Big(a_\lambda(X,\mu)\Big)\Big| &\leq c, \\
\Big|\mu\frac{\partial^2}{(\partial\mu)^2}E\Big(a_\lambda(X,\mu)\Big)\Big| &\leq c
\end{aligned}$$

($c\in\mathbf{R}^+$ constant) for $X\sim Pois(\mu)$ with $\mu\geq\epsilon>0$. The boundedness $\frac{\mu^n_{j+}}{\mu^n_{jk}(\theta)}=\frac{1}{\pi^n_{jk|C}(\theta)}$, $D_\theta(\pi^n_{jk|C}(\theta))$ and $D^2_\theta(\pi^n_{jk|C}(\theta))$ for $\theta\in\bar W$ follows from the generally assumed regularity conditions (RC2) and (RC3).

□

Since after the first approximation step the derivative still depends on the estimator for the nuisance parameters, now in the second step $\hat\mu^n_{\cdot+}$ will be substituted by $\mu^n_{\cdot+}$, using Taylor expansion again.

Lemma 5.4 *If all expectations are bounded away from zero (BC), it holds*

$$D_\theta Z^n_\lambda(\hat\mu^n_{\cdot+},\theta_0) \;=\; D_\theta Z^n_\lambda(\mu^n_{\cdot+},\theta_0)+\sum_{j=1}^{J^n}O_p(\sqrt{\mu^n_{j+}})\quad(n\to\infty).$$

Proof:
The decisive arguments in the following proof apply to column–multinomial as well as Poisson distribution, thus consider a column–multinomial table Y^n ($\hat{\mu}^n_{\cdot+} = Y^n_{\cdot+}$) representing both sampling schemes. Since by definition $D_\theta Z^n_\lambda(\hat{\mu}^n_{\cdot+}, \theta_0) = D_\theta SD^n_\lambda(\hat{\mu}^n_{\cdot+}, \theta_0) - D_\theta m^n_\lambda(\hat{\mu}^n_{\cdot+}, \theta_0)$ holds, and the result can be proved treating the two terms separately, let the derivative of the goodness–of–fit statistic be considered first. Hence for given $\delta \in (0,1)$, there is to be shown the existence of a constant M_δ, such that for almost all n holds

$$P\Big(\Big\|\frac{1}{\sqrt{\mu^n_{j+}}} D_\theta \sum_{k=1}^{K} \Big(a_\lambda(Y^n_{jk}, \hat{\mu}^n_{j+} \cdot \pi^n_{jk|C}(\theta_0)) - a_\lambda(Y^n_{jk}, \mu^n_{j+} \cdot \pi^n_{jk|C}(\theta_0))\Big)\Big\| > M_\delta\Big) \leq \delta \quad (5.27)$$

for all $j \in \{1, \ldots, J^n\}$. Then the difference is stochastically bounded giving

$$D_\theta SD^n_\lambda(\hat{\mu}^n_{\cdot+}, \theta_0) - D_\theta SD^n_\lambda(\mu^n_{\cdot+}, \theta_0) = \sum_{j=1}^{J^n} O_p(\sqrt{\mu^n_{j+}}) \quad (n \to \infty).$$

To prove 5.27, let in the following any $j \in \{1, \ldots, J^n\}$ be given. In the case of $\hat{\mu}^n_{j+} > 0$, now for the term in question holds

$$\begin{aligned}
&\Big\|\frac{1}{\sqrt{\mu^n_{j+}}} D_\theta \sum_{k=1}^{K} \Big(a_\lambda(Y^n_{jk}, \hat{\mu}^n_{j+} \cdot \pi^n_{jk|C}(\theta_0)) - a_\lambda(Y^n_{jk}, \mu^n_{j+} \cdot \pi^n_{jk|C}(\theta_0))\Big)\Big\| \\
\leq\ & \sqrt{\mu^n_{j+}} \sum_{k=1}^{K} \Big\| D_\theta a_\lambda\Big(\frac{Y^n_{jk}}{\mu^n_{j+}}, \frac{\hat{\mu}^n_{j+}}{\mu^n_{j+}} \cdot \pi^n_{jk|C}(\theta_0)\Big) - D_\theta a_\lambda\Big(\frac{Y^n_{jk}}{\mu^n_{j+}}, \pi^n_{jk|C}(\theta_0)\Big)\Big\| \\
=\ & \sqrt{\mu^n_{j+}} \sum_{k=1}^{K} \Big\| D_\theta \frac{\hat{\mu}^n_{j+}}{\mu^n_{j+}} \pi^n_{jk|C}(\theta_0) \cdot D_2 a_\lambda\Big(\frac{Y^n_{jk}}{\mu^n_{j+}}, \frac{\hat{\mu}^n_{j+}}{\mu^n_{j+}} \cdot \pi^n_{jk|C}(\theta_0)\Big) \\
& \qquad - D_\theta \pi^n_{jk|C}(\theta_0) \cdot D_2 a_\lambda\Big(\frac{Y^n_{jk}}{\mu^n_{j+}}, \pi^n_{jk|C}(\theta_0)\Big)\Big\| \\
\leq\ & \sqrt{\mu^n_{j+}} \sum_{k=1}^{K} \|D_\theta \pi^n_{jk|C}(\theta_0)\| \qquad\qquad (5.28) \\
& \qquad \cdot \Big|\frac{\hat{\mu}^n_{j+}}{\mu^n_{j+}} D_2 a_\lambda\Big(\frac{Y^n_{jk}}{\mu^n_{j+}}, \frac{\hat{\mu}^n_{j+}}{\mu^n_{j+}} \cdot \pi^n_{jk|C}(\theta_0)\Big) - D_2 a_\lambda\Big(\frac{Y^n_{jk}}{\mu^n_{j+}}, \pi^n_{jk|C}(\theta_0)\Big)\Big|.
\end{aligned}$$

a_λ is in the second component continuously differentiable arbitrarily often on $(0, \infty)$. Hence for $\frac{\hat{\mu}_{j+}}{\mu_{j+}} > 0$, Taylor expansion of $\frac{\hat{\mu}^n_{j+}}{\mu^n_{j+}} \cdot D_2 a_\lambda\Big(\frac{Y^n_{jk}}{\mu^n_{j+}}, \frac{\hat{\mu}^n_{j+}}{\mu^n_{j+}} \cdot \pi^n_{jk|C}(\theta_0)\Big)$ in $\frac{\hat{\mu}^n_{j+}}{\mu^n_{j+}}$ around 1 gives $D_2 a_\lambda\Big(\frac{Y^n_{jk}}{\mu^n_{j+}}, \pi^n_{jk|C}(\theta_0)\Big)$ plus an error term (for clarity of notation let the explicit formula of the derivative be considered):

$$\frac{\hat{\mu}^n_{j+}}{\mu^n_{j+}} \cdot D_2 a_\lambda\Big(\frac{Y^n_{jk}}{\mu^n_{j+}}, \frac{\hat{\mu}^n_{j+}}{\mu^n_{j+}} \cdot \pi^n_{jk|C}(\theta_0)\Big)$$

$$
\begin{aligned}
&= \frac{\hat{\mu}^n_{j+}}{\mu^n_{j+}} \cdot \frac{2}{\lambda+1}\Big(1 - \big(\frac{Y^n_{jk}}{\mu^n_{j+}} \cdot \frac{\mu^n_{j+}}{\hat{\mu}^n_{j+}} \cdot \frac{1}{\pi^n_{jk|C}(\theta_0)}\big)^{\lambda+1}\Big) \\
&= \frac{2}{\lambda+1}\Big(\frac{\hat{\mu}^n_{j+}}{\mu^n_{j+}} - \big(\frac{\hat{\mu}^n_{j+}}{\mu^n_{j+}}\big)^{-\lambda}\big(\frac{Y^n_{jk}}{\mu^n_{j+}\pi^n_{jk|C}(\theta_0)}\big)^{\lambda+1}\Big) \\
&= \frac{2}{\lambda+1}\Big(1 - \big(\frac{Y^n_{jk}}{\mu^n_{j+}\pi^n_{jk|C}(\theta_0)}\big)^{\lambda+1}\Big) \\
&\quad + \int_0^1 \frac{2}{\lambda+1}\Big(1 + \lambda(1 + z(\frac{\hat{\mu}^n_{j+}}{\mu^n_{j+}} - 1))^{-(\lambda+1)}\big(\frac{Y^n_{jk}}{\mu^n_{j+}\pi^n_{jk|C}(\theta_0)}\big)^{\lambda+1}\Big)dz \\
&\quad \cdot \big(\frac{\hat{\mu}^n_{j+}}{\mu^n_{j+}} - 1\big)
\end{aligned} \tag{5.29}
$$

with $\frac{2}{\lambda+1}\Big(1 - \big(\frac{Y^n_{jk}}{\mu^n_{j+}\pi^n_{jk|C}(\theta_0)}\big)^{\lambda+1}\Big) = D_2 a_\lambda\big(\frac{Y^n_{jk}}{\mu^n_{j+}}, \pi^n_{jk|C}(\theta_0)\big)$. The error is now dominated as follows

$$
\begin{aligned}
&\Big|\int_0^1 \frac{2}{\lambda+1}\Big(1 + \lambda(1 + z(\frac{\hat{\mu}^n_{j+}}{\mu^n_{j+}} - 1))^{-(\lambda+1)}\big(\frac{Y^n_{jk}}{\mu^n_{j+}\pi^n_{jk|C}(\theta_0)}\big)^{\lambda+1}\Big)dz \cdot \big(\frac{\hat{\mu}^n_{j+}}{\mu^n_{j+}} - 1\big)\Big| \\
\leq\; & H\big(\frac{Y^n_{jk}}{\mu^n_{j+}}, \frac{\hat{\mu}^n_{j+}}{\mu^n_{j+}}, \pi^n_{jk|C}(\theta_0)\big) \cdot \Big|\frac{\hat{\mu}^n_{j+}}{\mu^n_{j+}} - 1\Big|
\end{aligned}
$$

with

$$
\begin{aligned}
&H\big(\frac{Y^n_{jk}}{\mu^n_{j+}}, \frac{\hat{\mu}^n_{j+}}{\mu^n_{j+}}, \pi^n_{jk|C}(\theta_0)\big) \\
=\; & \sup_{z\in[0,1]} \Big|\frac{2}{\lambda+1}\Big(1 + \lambda(1 + z(\frac{\hat{\mu}^n_{j+}}{\mu^n_{j+}} - 1))^{-(\lambda+1)}\big(\frac{Y^n_{jk}}{\mu^n_{j+}\pi^n_{jk|C}(\theta_0)}\big)^{\lambda+1}\Big)\Big|
\end{aligned}
$$

being continuous on $[0,\infty) \times (0,\infty) \times (0,1)$. This inequality for the error term in (5.29) gives for (5.28)

$$
\begin{aligned}
&\Big\|\frac{1}{\sqrt{\mu^n_{j+}}}\Big(D_\theta \sum_{k=1}^K a_\lambda(Y^n_{jk}, \hat{\mu}^n_{j+} \cdot \pi^n_{jk|C}(\theta_0)) - D_\theta \sum_{k=1}^K a_\lambda(Y^n_{jk}, \mu^n_{j+} \cdot \pi^n_{jk|C}(\theta_0))\Big)\Big\| \\
\leq\; & \sum_{k=1}^K \|D_\theta \pi^n_{jk|C}(\theta_0)\| \cdot H\big(\frac{Y^n_{jk}}{\mu^n_{j+}}, \frac{\hat{\mu}^n_{j+}}{\mu^n_{j+}}, \pi^n_{jk|C}(\theta_0)\big) \cdot \Big|\frac{\hat{\mu}^n_{j+} - \mu^n_{j+}}{\sqrt{\mu^n_{j+}}}\Big|
\end{aligned}
$$

and hence for the probability in (5.27):

$$
P\Big(\Big\|\frac{1}{\sqrt{\mu^n_{j+}}} D_\theta \sum_{k=1}^K \Big(a_\lambda(Y^n_{jk}, \hat{\mu}^n_{j+}\pi^n_{jk|C}(\theta_0)) - a_\lambda(Y^n_{jk}, \mu^n_{j+}\pi^n_{jk|C}(\theta_0))\Big)\Big\| > M_\delta\Big)
$$

$$\leq \quad P\Big(\Big\|\frac{1}{\sqrt{\mu^n_{j+}}}D_\theta \sum_{k=1}^{K}\Big(a_\lambda(Y^n_{jk},\hat{\mu}^n_{j+}\pi^n_{jk|C}(\theta_0)) - a_\lambda(Y^n_{jk},\mu^n_{j+}\pi^n_{jk|C}(\theta_0))\Big)\Big\| > M_\delta$$

$$\wedge\, \hat{\mu}^n_{j+} > 0\,\forall\, j\Big) + P(\exists\, j : \hat{\mu}^n_{j+} = 0) \tag{5.30}$$

$$\leq \quad P\Big(\sum_{k=1}^{K}\|D_\theta\pi^n_{jk|C}(\theta_0)\| \cdot H(\frac{Y^n_{jk}}{\mu^n_{j+}}, \frac{\hat{\mu}^n_{j+}}{\mu^n_{j+}}, \pi^n_{jk|C}(\theta_0)) \cdot |\frac{\hat{\mu}^n_{j+} - \mu^n_{j+}}{\sqrt{\mu^n_{j+}}}| > M_\delta \wedge \hat{\mu}^n_{j+} > 0\Big)$$

$$+ \quad P(\exists\, j : \hat{\mu}^n_{j+} = 0)$$

$$= \quad P\Big(\sum_{k=1}^{K}\|D_\theta\pi^n_{jk|C}(\theta_0)\| \cdot \mathbf{1}_{\mathbf{N}}(\hat{\mu}^n_{j+}) \cdot H(\frac{Y^n_{jk}}{\mu^n_{j+}}, \frac{\hat{\mu}^n_{j+}}{\mu^n_{j+}}, \pi^n_{jk|C}(\theta_0)) \cdot |\frac{\hat{\mu}^n_{j+} - \mu^n_{j+}}{\sqrt{\mu^n_{j+}}}| > M_\delta\Big)$$

$$+ \quad P(\exists\, j : \hat{\mu}^n_{j+} = 0). \tag{5.31}$$

Considering the second probability, assumption $P(\hat{\mu}^n_{j+} > 0\,\forall\, j \in \{1,\ldots,J^n\}) \longrightarrow 1$ immediately yields for sufficiently large n

$$P(\exists\, j : \hat{\mu}^n_{j+} = 0) \leq \frac{\delta}{2}, \tag{5.32}$$

hence the first term remains to be studied. Apart from assumptions holding in both distribution models, up to here only analytical arguments and basic properties of probability measures have been used. In order to obtain the stochastic bounding result for the first probability in (5.31), information about the underlying distribution, in particular $Var(Y^n_{jk}) \leq \mu^n_{jk}$ for all j,k,n, is necessary now. This clearly holds for column multinomial (where each component is binomial) and Poisson sampling. The assumed independence of the columns in both distribution models further yields another required condition, namely $Var(Y^n_{j+}) = Var(\hat{\mu}^n_{j+}) \leq \mu^n_{j+}$ for all j,k.
Application of the Chebyshev Inequality now gives

$$P\Big(\Big|\frac{\hat{\mu}^n_{j+} - \mu^n_{j+}}{\sqrt{\mu^n_{j+}}}\Big| > \frac{1}{\sqrt{\delta}}\Big) \leq \delta \cdot Var\Big(\frac{\hat{\mu}^n_{j+}}{\sqrt{\mu^n_{j+}}}\Big) \leq \delta$$

and hence

$$|\frac{\hat{\mu}^n_{j+}}{\mu^n_{j+}} - 1| = O_p(\frac{1}{\sqrt{\mu^n_{j+}}}) \quad \text{for all } j. \tag{5.33}$$

This further asserts $\frac{\hat{\mu}^n_{j+}}{\mu^n_{j+}} = O_p(1)$ for all j and $\frac{Y^n_{jk}}{\mu^n_{j+}} = O_p(1)$ for all j,k using $\frac{Y^n_{jk}}{\mu^n_{j+}} \leq \frac{Y^n_{j+}}{\mu^n_{j+}} = \frac{\hat{\mu}^n_{j+}}{\mu^n_{j+}}$. For the function H now holds by Lemma 8.2 for fixed j

$$\mathbf{1}_{\mathbf{N}}(\hat{\mu}^n_{j+}) \cdot H(\frac{Y^n_{jk}}{\mu^n_{j+}}, \frac{\hat{\mu}^n_{j+}}{\mu^n_{j+}}, \pi^n_{jk|C}(\theta_0)) = O_p(1) \quad \text{for all } k. \tag{5.34}$$

Here $\frac{\hat{\mu}^n_{j+}}{\mu^n_{j+}}$ corresponds with the sequence $\bar{X}^n$ and $\frac{Y^n_{jk}}{\mu^n_{j+}}, \pi^n_{jk|C}(\theta_0)$ with the vector Z^n from Lemma 8.2. Further assumptions as the continuity of H and the variance condition are met as seen before. The boundedness of $D_\theta \pi^n_{jk|C}(\theta_0)$ (condition (RC3)) as well as statements (5.33) and (5.34) now guarantee the existence of a constant M_δ, so that for the first probability of (5.31) holds

$$P\Big(\sum_{k=1}^{K} \|D_\theta \pi^n_{jk|C}(\theta_0)\| \cdot \mathbf{1}_{\mathbf{N}}(\hat{\mu}^n_{j+}) \cdot H(\frac{Y^n_{jk}}{\mu^n_{j+}}, \frac{\hat{\mu}^n_{j+}}{\mu^n_{j+}}, \pi^n_{jk|C}(\theta_0)) \cdot |\frac{\hat{\mu}^n_{j+} - \mu^n_{j+}}{\sqrt{\mu^n_{j+}}}| > M_\delta\Big) \leq \frac{\delta}{2}$$

for almost all $n \in \mathbf{N}$. This result, combined with (5.32) (inserting in (5.31)), finally establishes (5.27) for both column multinomial and Poisson sampling. Now it remains to be shown

$$D_\theta m^n_\lambda(\hat{\mu}^n_{\cdot+}, \theta_0) = D_\theta m^n_\lambda(\mu^n_{\cdot+}, \theta_0) + O_p(\sum_{j=1}^{J^n} \sqrt{\mu^n_{j+}}) \tag{5.35}$$

for both distribution models. Since for the Poisson expectation m^n_λ of SD^n_λ by definition holds $m^n_\lambda(\mu^n_{\cdot+}, \theta_0) = \sum_{j=1}^{J^n} \sum_{k=1}^{K} e_\lambda(\mu^n_{j+}, \pi^n_{jk|C}(\theta_0))$ with $e_\lambda(\mu^n_{j+}, \pi^n_{jk|C}(\theta_0)) = E(a_\lambda(X^n_{jk}, \mu^n_{jk}))$ (see (5.1), (5.2)), this can obviously be proved by establishing

$$\frac{1}{\sqrt{\mu^n_{j+}}}\Big(\frac{\partial}{\partial \theta_s} e_\lambda(\hat{\mu}^n_{j+}, \pi^n_{jk|C}(\theta_0)) - \frac{\partial}{\partial \theta_s} e_\lambda(\mu^n_{j+}, \pi^n_{jk|C}(\theta_0))\Big) = O_p(1)$$

for all j, k and $s \in \{1, \ldots S\}$. Using analogous arguments as in the proof of (5.27), it suffices (cp. (5.30)) to verify that for every $\delta \in (0,1)$, there exists a constant M_δ, such that for all j, k and almost all n the inequality

$$\begin{aligned} & P\Big(\frac{1}{\sqrt{\mu^n_{j+}}}\Big|\frac{\partial}{\partial \theta_s} e_\lambda(\hat{\mu}^n_{j+}, \pi^n_{jk|C}(\theta_0)) - \frac{\partial}{\partial \theta_s} e_\lambda(\mu^n_{j+}, \pi^n_{jk|C}(\theta_0))\Big| > M_\delta, \hat{\mu}^n_{j+} > 0 \,\forall\, j\Big) \\ \leq\ & P\Big(\frac{1}{\sqrt{\mu^n_{j+}}} \cdot \mathbf{1}_{\mathbf{N}}(\hat{\mu}^n_{j+}) \cdot \Big|\frac{\partial}{\partial \theta_s} e_\lambda(\hat{\mu}^n_{j+}, \pi^n_{jk|C}(\theta_0)) - \frac{\partial}{\partial \theta_s} e_\lambda(\mu^n_{j+}, \pi^n_{jk|C}(\theta_0))\Big| > M_\delta\Big) \\ \leq\ & \frac{\delta}{2} \end{aligned}$$

holds. Since $(\hat{\mu}^n_{j+})_{n \in \mathbf{N}}$ is for every j a sequence of random variables with $E(\hat{\mu}^n_{j+}) = \mu^n_{j+} \in [K\epsilon, \infty)$, this follows from Lemma 8.3, if for all j, k, n

$$|\frac{\partial}{\partial \mu}\frac{\partial}{\partial \theta_s} e_\lambda(\mu, \pi^n_{jk|C}(\theta_0))| \leq c \quad \text{for all } \mu \geq K\epsilon \quad (c \in \mathbf{R}^+ \text{ constant}) \tag{5.36}$$

can be shown. To see this, one has to prove for every j, k, n

$$|\frac{\partial}{\partial \mu^n_{j+}}\frac{\partial}{\partial \theta_s} e_\lambda(\mu^n_{j+}, \pi^n_{jk|C}(\theta_0))| \leq c \in \mathbf{R}^+ \quad \text{for all } \mu^n_{j+} \geq K\epsilon,$$

which yields (5.36) by definition of e_λ. Now it holds

$$
\begin{aligned}
& |\frac{\partial}{\partial \mu_{j+}^n} \frac{\partial}{\partial \theta_s} e_\lambda(\mu_{j+}^n, \pi_{jk|C}^n(\theta_0))| \\
= \; & |\frac{\partial}{\partial \theta_s} \pi_{jk|C}^n(\theta_0)| \cdot \Big| \frac{\partial}{\partial \pi_{jk|C}^n} \Big(\frac{\partial}{\mu_{j+}^n} E(a_\lambda(X_{jk}^n, \mu_{j+}^n \pi_{jk|C}^n(\theta_0))) \Big) \Big| \\
= \; & |\frac{\partial}{\partial \theta_s} \pi_{jk|C}^n(\theta_0)| \\
& \cdot \Big| \frac{\partial}{\partial \pi_{jk|C}^n} \Big((\frac{\partial}{\partial \mu_{j+}^n} \mu_{j+}^n \pi_{jk|C}^n(\theta_0)) \frac{\partial}{\partial \mu_{j+}^n \pi_{jk|C}^n} E(a_\lambda(X_{jk}^n, \mu_{j+}^n \pi_{jk|C}^n(\theta_0))) \Big) \Big| \\
= \; & |\frac{\partial}{\partial \theta_s} \pi_{jk|C}^n(\theta_0)| \cdot \Big| \frac{\partial}{\partial \pi_{jk|C}^n} \Big(\pi_{jk|C}^n(\theta_0) \frac{\partial}{\partial \mu_{j+}^n \pi_{jk|C}^n} E(a_\lambda(X_{jk}^n, \mu_{j+}^n \pi_{jk|C}^n(\theta_0))) \Big) \Big| \\
= \; & |\frac{\partial}{\partial \theta_s} \pi_{jk|C}^n(\theta_0)| \\
& \cdot \Big| \frac{\partial}{\partial \mu_{j+}^n \pi_{jk|C}^n} E(a_\lambda(X_{jk}^n, \mu_{j+}^n \pi_{jk|C}^n(\theta_0))) \\
& \quad + \pi_{jk|C}^n(\theta_0) \cdot \frac{\partial}{\partial \pi_{jk|C}^n} \mu_{j+}^n \pi_{jk|C}^n(\theta_0) \cdot \frac{\partial^2}{(\partial \mu_{j+}^n \pi_{jk|C}^n)^2} E(a_\lambda(X_{jk}^n, \mu_{j+}^n \pi_{jk|C}^n(\theta_0))) \Big| \\
= \; & |\frac{\partial}{\partial \theta_s} \pi_{jk|C}^n(\theta_0)| \\
& \cdot \Big| \frac{\partial}{\partial \mu_{j+}^n \pi_{jk|C}^n} E(a_\lambda(X_{jk}^n, \mu_{j+}^n \pi_{jk|C}^n(\theta_0))) \\
& \quad + \mu_{j+}^n \pi_{jk|C}^n(\theta_0) \cdot \frac{\partial^2}{(\partial \mu_{j+}^n \pi_{jk|C}^n)^2} E(a_\lambda(X_{jk}^n, \mu_{j+}^n \pi_{jk|C}^n(\theta_0))) \Big| \\
= \; & |\frac{\partial}{\partial \theta_s} \pi_{jk|C}^n(\theta_0)| \cdot \Big| \frac{\partial}{\partial \mu_{jk}^n} E(a_\lambda(X_{jk}^n, \mu_{jk}^n)) + \mu_{jk}^n \cdot \frac{\partial^2}{(\partial \mu_{jk}^n)^2} E(a_\lambda(X_{jk}^n, \mu_{jk}^n)) \Big| \\
\leq \; & c
\end{aligned}
$$

for some constant $c \in \mathbf{R}^+$, using the generally assumed regularity condition (RC3) and in particular $\frac{\partial}{\partial \mu_{jk}^n} E(a_\lambda(X_{jk}^n, \mu_{jk}^n))$ and $\mu_{jk}^n \frac{\partial^2}{(\partial \mu_{jk}^n)^2} E(a_\lambda(X_{jk}^n, \mu_{jk}^n))$ being bounded for all $\mu_{jk}^n = \mu_{j+}^n \pi_{jk|C}^n(\theta_0) \geq \epsilon$ (see Lemma 4.8 g), h)). Since by assumption (RC2) the probabilities $\pi_{jk|C}^n(\theta_0)$ are bounded away from zero for all j, k, n, the expectations e_λ, regarded as a function in μ_{j+}^n, are bounded for all $\mu_{j+}^n \in [K\epsilon, \infty)$. This establishes (5.36) and hence (5.35). □

In Lemma 5.5, the gradient can now be replaced by its Poisson expectation. This is immediately obtained in case of Poisson sampling applying Chebyshev's inequality. In the column–multinomial case, where two distributions have to be compared, the result will be proved using the auxiliary result from section 5.1.

Lemma 5.5 *For both distribution models assume* $\sqrt{n}(\hat{\theta}^n - \theta_0) = O_p(1)$ *(LC2) and* $\mu^n_{jk} \geq \epsilon > 0$ *for all* j, k, n *(BC). In case of column–multinomial sampling, suppose that additionally* $\frac{J^n}{n} \longrightarrow 0$ *and the conditions (LC1) and (MD3) hold, i.e.* $\frac{1}{n} I^n(\mu^n_{\cdot+}, \theta_0) \longrightarrow I_\infty$ *and* $\max_{1 \leq j \leq J^n} \pi^n_{jk|D}(\theta_0) \longrightarrow 0$ *for all* k*. Then for* $\lambda \in (-1, 1]$ *if column multinomial and* $\lambda \in (-1, \infty)$*, if Poisson sampling is considered, follows*

$$\Big(D_\theta Z^n_\lambda(\mu^n_{\cdot+}, \theta_0) + c^n_\lambda(\mu^n_{\cdot+}, \theta_0)\Big)(\hat{\theta}^n - \theta_0) \quad = \quad O_p(1) \quad (n \to \infty)$$

with $c^n_\lambda(\mu^n_{\cdot+}, \theta_0)$ *defined in (5.4), for which in particular holds*

$$c^n_\lambda(\mu^n_{\cdot+}, \theta_0) \quad = \quad -E\Big(D_\theta Z^n_\lambda(\mu^n_{\cdot+}, \theta_0 | X^n)\Big).$$

Proof:
Using $c^n_\lambda(\mu^n_{\cdot+}, \theta_0) = \sum_{j=1}^{J^n} \sum_{k=1}^{K} D_\theta \log \pi^n_{jk|C}(\theta_0) Cov(a_\lambda(X^n_{jk}, \mu^n_{jk}), X^n_{jk})$ as defined in (5.4) at the beginning of this chapter, let first be verified that c^n_λ is the stated Poisson expectation. This is immediately shown applying the formula for derivatives of Poisson expectations given in Theorem 4.1 (for notational reference see (5.1) – (5.3)):

$$\begin{aligned}
& E\left(D_\theta Z^n_\lambda(\mu^n_{\cdot+}, \theta_0 | X^n)\right) \\
= \; & \sum_{j=1}^{J^n} \sum_{k=1}^{K} \Big(E\Big(D_\theta a_\lambda(X^n_{jk}, \mu^n_{j+} \pi^n_{jk|C}(\theta_0)) - D_\theta e_\lambda(\mu^n_{j+}, \pi^n_{jk|C}(\theta_0))\Big)\Big) \\
= \; & \sum_{j=1}^{J^n} \sum_{k=1}^{K} D_\theta \mu^n_{jk}(\theta_0) \cdot \Big(E\Big(\frac{\partial}{\partial \mu^n_{jk}} a_\lambda(X^n_{jk}, \mu^n_{jk})\Big) - \frac{\partial}{\partial \mu^n_{jk}} E\Big(a_\lambda(X^n_{jk}, \mu^n_{jk})\Big)\Big) \\
= \; & \sum_{j=1}^{J^n} \sum_{k=1}^{K} D_\theta \mu^n_{jk}(\theta_0) \cdot \Big(-\frac{1}{\mu^n_{jk}} \cdot Cov\Big(a_\lambda(X^n_{jk}, \mu^n_{jk}), X^n_{jk}\Big)\Big) \\
= \; & -\sum_{j=1}^{J^n} \sum_{k=1}^{K} D_\theta \log \pi^n_{jk|C}(\theta_0) \cdot Cov\Big(a_\lambda(X^n_{jk}, \mu^n_{jk}), X^n_{jk}\Big) \\
= \; & -c^n_\lambda(\mu^n_{\cdot+}, \theta_0).
\end{aligned}$$

Since now $\sqrt{n}(\hat{\theta}^n - \theta_0) = O_p(1)$ and $\mu^n_{++} - nc \to 0$ $(c > 0)$ is presumed (in particular $n = \mu^n_{++} = Y^n_{++}$ in the column–multinomial case), it suffices to determine the order of the difference between derived statistic and Poisson expectation,

$$\frac{1}{\sqrt{\mu^n_{++}}}\Big(D_\theta Z^n_\lambda(\mu^n_{\cdot+}, \theta_0) - E(D_\theta Z^n_\lambda(\mu^n_{\cdot+}, \theta_0 | X^n))\Big) = O_p(1),$$

with $\lambda \in (-1, 1]$ in case of column multinomial and $\lambda \in (-1, \infty)$ in case of Poisson sampling. Now by definition holds

$$\begin{aligned}
& D_\theta Z^n_\lambda(\mu^n_{\cdot+}, \theta_0) - E\Big(D_\theta Z^n_\lambda(\mu^n_{\cdot+}, \theta_0 | X^n)\Big) \\
= \; & D_\theta\Big(SD^n_\lambda(\mu^n_{\cdot+}, \theta_0) - m^n_\lambda(\mu^n_{\cdot+}, \theta_0)\Big) - E\Big(D_\theta SD^n_\lambda(\mu^n_{\cdot+}, \theta_0 | X^n) - D_\theta m^n_\lambda(\mu^n_{\cdot+}, \theta_0)\Big)
\end{aligned}$$

$$= D_\theta\Big(SD^n_\lambda(\mu^n_{\cdot +}, \theta_0)\Big) - E\Big(D_\theta SD^n_\lambda(\mu^n_{\cdot +}, \theta_0|X^n)\Big),$$

hence the result follows, if for each component, i.e. for every $s \in \{1, \ldots, S\}$,

$$\frac{1}{\sqrt{\mu^n_{++}}}\Big(\frac{\partial}{\partial\theta_s}SD^n_\lambda(\mu^n_{\cdot +}, \theta_0) - E\Big(\frac{\partial}{\partial\theta_s}SD^n_\lambda(\mu^n_{\cdot +}, \theta_0|X^n)\Big)\Big) = O_p(1) \tag{5.37}$$

can be established with $\lambda \in (-1, 1]$ (column–multinomial) resp. $\lambda \in (-1, \infty)$ (Poisson). In case of Poisson sampling, where especially $SD^n_\lambda(\mu^n_{\cdot +}, \theta_0|X^n)$ is considered, (5.37) immediately follows if for every $s \in \{1, \ldots, S\}$ and $\lambda \in (-1, \infty)$ holds

$$Var\Big(\frac{1}{\sqrt{\mu^n_{++}}}\frac{\partial}{\partial\theta_s}SD^n_\lambda(\mu^n_{\cdot +}, \theta_0|X^n)\Big) \le c \quad \text{for some constant } c \in \mathbf{R}^+, \tag{5.38}$$

since then the Chebyshev Inequality can be directly applied giving

$$\begin{aligned}
&P\Big(\frac{1}{\sqrt{\mu^n_{++}}}\Big|\frac{\partial}{\partial\theta_s}SD^n_\lambda(\mu^n_{\cdot +}, \theta_0|X^n) - E\Big(\frac{\partial}{\partial\theta_s}SD^n_\lambda(\mu^n_{\cdot +}, \theta_0|X^n)\Big)\Big| \ge \sqrt{\frac{c}{\delta}}\Big)\\
&\le \frac{\delta}{c}Var\Big(\frac{1}{\sqrt{\mu^n_{++}}}\frac{\partial}{\partial\theta_s}SD^n_\lambda(\mu^n_{\cdot +}, \theta_0|X^n)\Big)\\
&\le \delta
\end{aligned}$$

($\delta \in (0, 1)$ constant). (5.38) can now be seen as follows. It holds

$$\begin{aligned}
&Var\Big(\frac{1}{\sqrt{\mu^n_{++}}}\frac{\partial}{\partial\theta_s}SD^n_\lambda(\mu^n_{\cdot +}, \theta_0|X^n)\Big)\\
= \;& \frac{1}{\mu^n_{++}}\sum_{j=1}^{J^n}\sum_{k=1}^{K} Var\Big(\frac{\partial}{\partial\theta_s}a_\lambda(X^n_{jk}, \mu^n_{jk}(\theta_0))\Big)\\
&\le \frac{1}{\mu^n_{++}}\sum_{j=1}^{J^n}\sum_{k=1}^{K} E\Big(\Big(\frac{\partial}{\partial\theta_s}a_\lambda(X^n_{jk}, \mu^n_{jk}(\theta_0))\Big)^2\Big)\\
= \;& \frac{1}{\mu^n_{++}}\sum_{j=1}^{J^n}\sum_{k=1}^{K}(\frac{\partial}{\partial\theta_s}\mu^n_{jk}(\theta_0))^2 \cdot E\Big(\Big(D_2 a_\lambda(X^n_{jk}, \mu^n_{jk})\Big)^2\Big)\\
= \;& \frac{1}{\mu^n_{++}}\sum_{j=1}^{J^n}\sum_{k=1}^{K}(\mu^n_{j+})^2(\frac{\partial}{\partial\theta_s}\pi^n_{jk|C}(\theta_0))^2 \cdot E\Big(\Big(D_2 a_\lambda(X^n_{jk}, \mu^n_{jk})\Big)^2\Big)\\
= \;& \frac{1}{\mu^n_{++}}\sum_{j=1}^{J^n}\sum_{k=1}^{K}(\frac{\partial}{\partial\theta_s}\log\pi^n_{jk|C}(\theta_0))^2 \cdot \mu^n_{jk}E\Big(\mu^n_{jk}\Big(D_2 a_\lambda(X^n_{jk}, \mu^n_{j+}\pi^n_{jk|C}(\theta_0))\Big)^2\Big)\\
\le \;& \frac{1}{\mu^n_{++}}\sum_{j=1}^{J^n}\sum_{k=1}^{K}\mu^n_{jk} \cdot c\\
= \;& c
\end{aligned}$$

for some $c \in \mathbf{R}^+$ constant, since $\frac{1}{\pi^n_{jk|C}(\theta_0)}$ and $\frac{\partial}{\partial\theta_s}\pi^n_{jk|C}(\theta_0)$ are bounded by assumption (RC2), (RC3) and $E\Big(\mu^n_{jk}\Big(D_2 a_\lambda(X^n_{jk}, \mu^n_{jk})\Big)^2\Big)$ by Lemma 4.8 e) using $\mu^n_{jk} \geq \epsilon > 0$ for all j, k, n (BC).

To prove (5.37) for the much more difficult case of column multinomial sampling (cp. preliminaries to sec. 5.1), Lemma 5.1 in section 5.1 has been stated. Based on the method introduced by Morris (1975), this lemma provides under the given conditions the asymptotic normality of the column multinomial statistic $\frac{\partial}{\partial\theta_s} SD^n_\lambda(\mu^n_{\cdot+}, \theta_0|Y^n)$ centered with its Poisson expectation $E\Big(\frac{\partial}{\partial\theta_s} SD^n_\lambda(\mu^n_{\cdot+}, \theta_0|X^n)\Big)$:

$$\frac{1}{v^n_\lambda}\Big(\frac{\partial}{\partial\theta_s} SD^n_\lambda(\mu^n_{\cdot+}, \theta_0|Y^n) - E\Big(\frac{\partial}{\partial\theta_s} SD^n_\lambda(\mu^n_{\cdot+}, \theta_0|X^n)\Big)\Big) \xrightarrow{\mathcal{L}} N(0,1)$$

($\lambda \in (-1, 1]$) with v^{n2}_λ being a certain Poisson variance with the order $O_e(n)$. Since the standard deviation is of the order $O_e(\sqrt{n})$, this result and $n = \mu^n_{++}$ immediately give (5.37) for this distribution model as well:

$$\frac{1}{\sqrt{\mu^n_{++}}}\Big(\frac{\partial}{\partial\theta_s} SD^n_\lambda(\mu^n_{\cdot+}, \theta_0|Y^n) - E\Big(\frac{\partial}{\partial\theta_s} SD^n_\lambda(\mu^n_{\cdot+}, \theta_0|X^n)\Big)\Big) = O_p(1) \quad (\lambda \in (-1, 1]).$$

□

Using the assumed approximability of the parameter difference through information matrix and scores and the immediately determined order of c^n_λ, it is now possible to pass on to an expression without depending on $\hat{\theta}^n$ anymore. Resuming the preceding approximation steps, thus for the correction term concerning the parameter estimation $\hat{\theta}^n$ derived in the first step (Lemma 5.3), an approximation through a sum of independent variables is given.

Lemma 5.6 *Suppose all expectations are bounded away from zero (BC), then for the asymptotics $n \to \infty$ holds*

$$c^n_\lambda(\mu^n_{\cdot+}, \theta_0) \;=\; O(J^n)$$

with c^n_λ defined in (5.4). If further the assumption $\hat{\theta}^n - \theta_0 = (I^n(\mu^n_{\cdot+}, \theta_0))^{-1}U^n(\theta_0) + O_p(\frac{1}{n})$ (LC3) is met, this immediately establishes

$$c^n_\lambda(\mu^n_{\cdot+}, \theta_0)(\hat{\theta}^n - \theta_0) \;=\; c^n_\lambda(\mu^n_{\cdot+}, \theta_0)(I^n(\mu^n_{\cdot+}, \theta_0))^{-1}U^n(\theta_0) + O_p(1) \quad (n \to \infty).$$

Proof:
Since all expected values are bounded away from zero, Lemma 4.8 d) yields for all j, k, n $Cov(a_\lambda(X^n_{jk}, \mu^n_{jk}), X^n_{jk})$ being bounded. This, together with the presumed re-

gularity conditions thus gives

$$c_\lambda^n(\mu_{\cdot+}^n, \theta_0) = -\sum_{j=1}^{J^n}\sum_{k=1}^{K} D_\theta \log \pi_{jk|C}^n(\theta_0)) \cdot Cov\Big(a_\lambda(X_{jk}^n, \mu_{jk}^n), X_{jk}^n\Big) = O(J^n).$$

Further using the approximability of the estimator through information matrix and scores yields

$$\begin{aligned} c_\lambda^n(\mu_{\cdot+}^n, \theta_0)(\hat{\theta}^n - \theta_0) &= c_\lambda^n(\mu_{\cdot+}^n, \theta_0)\Big((I^n(\mu_{\cdot+}^n, \theta_0))^{-1}U^n(\theta_0) + O_p(\frac{1}{n})\Big) \\ &= c_\lambda^n(\mu_{\cdot+}^n, \theta_0)(I^n(\mu_{\cdot+}^n, \theta_0))^{-1}U^n(\theta_0) + O_p(\frac{J^n}{n}) \end{aligned}$$

and $\frac{J^n}{n} = O(1)$, which follows using (BC) and $\frac{\mu_{++}^n}{n} = O(1)$, gives the result. □

Example 5.7
Considering Pearson's X^2 Statistic ($\lambda = 1$), the expectation m_1^n does not depend on θ since for every j, k, n holds $e_1(\mu_{j+}^n, \pi_{jk|C}^n(\theta_0)) = E(a_1(X_{jk}^n, \mu_{jk}^n)) = E\Big(\frac{(X_{jk}^n - \mu_{jk}^n)^2}{\mu_{jk}^n}\Big) = 1$. Using

$$D_\theta SD_1^n(\mu_{\cdot+}^n, \theta|X^n) = \sum_{j=1}^{J^n}\sum_{k=1}^{K} \mu_{jk}^n(\theta)\Big(1 - (\frac{X_{jk}^n}{\mu_{jk}^n(\theta)})^2\Big) \cdot D_\theta \log \pi_{jk|C}^n(\theta),$$

gives

$$\begin{aligned} c_1^n(\mu_{\cdot+}^n, \theta_0) &= -E\Big(D_\theta Z_1^n(\mu_{\cdot+}^n, \theta_0|X^n)\Big) \\ &= -E\Big(D_\theta SD_1^n(\mu_{\cdot+}^n, \theta_0|X^n)\Big) \\ &= \sum_{j=1}^{J^n}\sum_{k=1}^{K} D_\theta \log \pi_{jk|C}^n(\theta_0). \end{aligned}$$

□

The last missing step from $Z_\lambda^n(\hat{\mu}_{\cdot+}^n, \theta_0)$ to $Z_\lambda^n(\mu_{\cdot+}^n, \theta_0)$ (cp. Lemma 5.3) is probably the most difficult one of the whole approximation. To get the error small, a second order Taylor expansion of a_λ in both components is necessary, which yields $\sum_{j=1}^{J^n}(\hat{\mu}_{j+}^n - \mu_{j+}^n)^2/\mu_{j+}^n$ as an additional correction term with $\hat{\mu}_{j+}^n = Y_{j+}^n$ resp. $\hat{\mu}_{j+}^n = X_{j+}^n$ (column–multinomial resp. Poisson). This term equals Pearson's X^2 Statistic ($\lambda = 1$) for the row sums: $\sum_{j=1}^{J^n}(\hat{\mu}_{j+}^n - \mu_{j+}^n)^2/\mu_{j+}^n = \sum_{j=1}^{J^n} a_1(\hat{\mu}_{j+}^n, \mu_{j+}^n)$.

Lemma 5.8 *When all expectations are bounded away from zero (BC), i.e.* $\mu_{jk}^n \geq \epsilon > 0$ *for all* j, k, n*, then for the asymptotics* $n \to \infty$ *holds*

$$Z_\lambda^n(\hat{\mu}_{\cdot+}^n, \theta_0) = Z_\lambda^n(\mu_{\cdot+}^n, \theta_0) - \sum_{j=1}^{J^n} \frac{(\hat{\mu}_{j+}^n - \mu_{j+}^n)^2}{\mu_{j+}^n} + \sum_{j=1}^{J^n} O_p\Big(\sqrt{\frac{1}{\mu_{j+}^n}}\Big).$$

Proof:
The statement will be proved in two parts, separately for goodness–of–fit statistic and expectation:

$$SD^n_\lambda(\hat{\mu}^n_{\cdot+}, \theta_0) = SD^n_\lambda(\mu^n_{\cdot+}, \theta_0) - \sum_{j=1}^{J^n} \frac{(\hat{\mu}^n_{j+} - \mu^n_{j+})^2}{\mu^n_{j+}} + O_p\Big(\sum_{j=1}^{J^n} \sqrt{\frac{1}{\mu^n_{j+}}}\Big), \tag{5.39}$$

$$m^n_\lambda(\hat{\mu}^n_{\cdot+}, \theta_0) = m^n_\lambda(\mu^n_{\cdot+}, \theta_0) + O_p\Big(\sum_{j=1}^{J^n} \sqrt{\frac{1}{\mu^n_{j+}}}\Big). \tag{5.40}$$

Since in this step of the approximation the nuisance parameters are treated just like in Lemma 5.4, this proof will be similar and particularly the essential arguments will be identical. Hence consider, just as in the proof of Lemma 5.4, a table Y^n representing column multinomial and Poisson sampling ($\hat{\mu}^n_{\cdot+} = Y^n_{\cdot+}$). For the proof of (5.39) now for every $\delta \in (0,1)$, the existence of a constant $M_\delta \in \mathbf{R}^+$ will be shown with M_δ chosen such that for every $j \in \{1, \ldots, J^n\}$ holds

$$\begin{aligned} & P\Big(\sqrt{\mu^n_{j+}}\Big| \sum_{k=1}^{K} \Big(a_\lambda(Y^n_{jk}, \hat{\mu}^n_{j+} \cdot \pi^n_{jk|C}(\theta_0)) - a_\lambda(Y^n_{jk}, \mu^n_{jk})\Big) + \frac{(\hat{\mu}^n_{j+} - \mu^n_{j+})^2}{\mu^n_{j+}}\Big| > M_\delta \\ & \qquad \wedge \hat{\mu}^n_{j+} > 0\Big) \\ + \quad & P\Big(\exists j : \hat{\mu}^n_{j+} = 0\Big) \\ \le \quad & \delta \end{aligned} \tag{5.41}$$

for almost all $n \in \mathbf{N}$. As seen in the proof of Lemma 5.4, it suffices to study the first probability. Therefore let any $j \in \{1, \ldots, J^n\}$ be given and the difference $\sum_{k=1}^{K} \Big(a_\lambda(Y^n_{jk}, \hat{\mu}^n_{j+} \cdot \pi^n_{jk|C}(\theta_0)) - a_\lambda(Y^n_{jk}, \mu^n_{jk})\Big)$ be considered, which can be written as follows for $\hat{\mu}^n_{j+} > 0$:

$$\begin{aligned} & \sum_{k=1}^{K} \Big(a_\lambda(Y^n_{jk}, \hat{\mu}^n_{j+} \cdot \pi^n_{jk|C}(\theta_0)) - a_\lambda(Y^n_{jk}, \mu^n_{jk})\Big) \\ = \quad & \sum_{k=1}^{K} \Big(a_\lambda(Y^n_{jk}, \hat{\mu}^n_{j+} \cdot \frac{\mu^n_{jk}}{\mu^n_{j+}}) - a_\lambda(Y^n_{jk}, \mu^n_{jk})\Big) \\ = \quad & \sum_{k=1}^{K} \mu^n_{jk}\Big(a_\lambda(\frac{Y^n_{jk}}{\mu^n_{jk}}, \frac{\hat{\mu}^n_{j+}}{\mu^n_{j+}}) - a_\lambda(\frac{Y^n_{jk}}{\mu^n_{jk}}, 1)\Big). \end{aligned} \tag{5.42}$$

The distance function a_λ is differentiable arbitrarily often on $\mathbf{R}^+ \times \mathbf{R}^+$ and hence can be expanded for $\tilde{\mu}_j > 0$ (consider especially $\tilde{\mu}_j = \mu^n_{j+} > 0$, $\tilde{\mu}_j = \hat{\mu}^n_{j+} > 0$) and

$Y^n_{jk} > 0$ in a second order Taylor Series around $(1,1)$:

$$\begin{aligned} a_\lambda(\frac{Y^n_{jk}}{\mu^n_{jk}}, \frac{\tilde\mu_j}{\mu^n_{j+}}) &= \frac{1}{2}(\frac{Y^n_{jk}}{\mu^n_{jk}} - 1, \frac{\tilde\mu_j}{\mu^n_{j+}} - 1) \cdot D^2 a_\lambda(1,1) \cdot (\frac{Y^n_{jk}}{\mu^n_{jk}} - 1, \frac{\tilde\mu_j}{\mu^n_{j+}} - 1)^T + R^n_{jk}(\tilde\mu_j) \\ &= (\frac{Y^n_{jk}}{\mu^n_{jk}} - \frac{\tilde\mu_j}{\mu^n_{j+}})^2 + R^n_{jk}(\tilde\mu_j) \end{aligned}$$

with $D^2 a_\lambda(1,1) = \begin{pmatrix} 2 & -2 \\ -2 & 2 \end{pmatrix}$ and $a_\lambda(1,1) = D_1 a_\lambda(1,1) = D_2 a_\lambda(1,1) = 0$ (derivatives of a_λ see Lemma 2.2). Let α denote a tupel $(\alpha_1, \alpha_2) \in \mathbf{N}^2_\mathbf{0}$, then the error term states as follows

$$\begin{aligned} R^n_{jk}(\tilde\mu_j) &= \int_0^1 \frac{(1-z)^2}{2!} \sum_{|\alpha|=3} \frac{3!}{\alpha!} D^\alpha a_\lambda\Big(1 + z(\frac{Y^n_{jk}}{\mu^n_{jk}} - 1), 1 + z(\frac{\tilde\mu_j}{\mu^n_{j+}} - 1)\Big) \\ &\quad \cdot(\frac{Y^n_{jk}}{\mu^n_{jk}} - 1)^{\alpha_1} (\frac{\tilde\mu_j}{\mu^n_{j+}} - 1)^{\alpha_2} dz. \end{aligned}$$

For $\tilde\mu_j > 0$ and $Y^n_{jk} = 0$ holds

$$a_\lambda(0, \frac{\tilde\mu_j}{\mu^n_{j+}}) = \frac{2}{\lambda+1} \cdot \frac{\tilde\mu_j}{\mu^n_{j+}} = (0 - \frac{\tilde\mu_j}{\mu^n_{j+}})^2 + (\frac{2}{\lambda+1} \cdot \frac{\tilde\mu_j}{\mu^n_{j+}} - (\frac{\tilde\mu_j}{\mu^n_{j+}})^2).$$

This result combined with the Taylor expansion now gives for the difference in the case $\hat\mu^n_{j+} > 0$ (inserting in (5.42)):

$$\begin{aligned} &\sum_{k=1}^K \Big(a_\lambda(Y^n_{jk}, \hat\mu^n_{j+} \cdot \pi^n_{jk|C}(\theta_0)) - a_\lambda(Y^n_{jk}, \mu^n_{jk})\Big) \\ =\ &\sum_{k=1}^K \mu^n_{jk} \Big(a_\lambda(\frac{Y^n_{jk}}{\mu^n_{jk}}, \frac{\hat\mu^n_{j+}}{\mu^n_{j+}}) - a_\lambda(\frac{Y^n_{jk}}{\mu^n_{jk}}, 1)\Big) \\ =\ &\sum_{k=1}^K \mu^n_{jk} \Bigg((\frac{Y^n_{jk}}{\mu^n_{jk}} - \frac{\hat\mu^n_{j+}}{\mu^n_{j+}})^2 + \mathbf{1}_{\{0\}}(Y^n_{jk})(\frac{2}{\lambda+1} \cdot \frac{\hat\mu^n_{j+}}{\mu^n_{j+}} - (\frac{\hat\mu^n_{j+}}{\mu^n_{j+}})^2) + \mathbf{1}_\mathbf{N}(Y^n_{jk}) R^n_{jk}(\hat\mu^n_{j+}) \\ &\quad -(\frac{Y^n_{jk}}{\mu^n_{jk}} - 1)^2 - \mathbf{1}_{\{0\}}(Y^n_{jk})(\frac{2}{\lambda+1} - 1) - \mathbf{1}_\mathbf{N}(Y^n_{jk}) R^n_{jk}(\mu^n_{j+})\Bigg) \\ =\ & -\frac{(\hat\mu^n_{j+} - \mu^n_{j+})^2}{\mu^n_{j+}} + \bar R^n_{jk}(\hat\mu^n_{j+}) - \bar R^n_{jk}(\mu^n_{j+}) \end{aligned}$$

with

$$\bar R^n_{jk}(\tilde\mu_j) = \sum_{k=1}^K \mu^n_{jk} \Big(\mathbf{1}_{\{0\}}(Y^n_{jk})(\frac{2}{\lambda+1} \cdot \frac{\tilde\mu_j}{\mu^n_{j+}} - (\frac{\tilde\mu_j}{\mu^n_{j+}})^2) + \mathbf{1}_\mathbf{N}(Y^n_{jk}) R^n_{jk}(\tilde\mu_j)\Big).$$

This and further inequalities yield for the first probability in (5.41)

$$
\begin{aligned}
& P\Big(\sqrt{\mu^n_{j+}}\Big|\sum_{k=1}^{K}\Big(a_\lambda(Y^n_{jk},\hat{\mu}^n_{j+}\cdot\pi^n_{jk|C}(\theta_0)) - a_\lambda(Y^n_{jk},\mu^n_{jk})\Big) + \frac{(\hat{\mu}^n_{j+}-\mu^n_{j+})^2}{\mu^n_{j+}}\Big| > M_\delta \\
& \quad \wedge \hat{\mu}^n_{j+} > 0\Big) \\
= \ & P\Big(\sqrt{\mu^n_{j+}}\Big|\sum_{k=1}^{K}\Big(\bar{R}^n_{jk}(\hat{\mu}^n_{j+}) - \bar{R}^n_{jk}(\mu^n_{j+})\Big)\Big| > M_\delta \ \wedge \ \hat{\mu}^n_{j+} > 0\Big) \\
= \ & P\Big(\sqrt{\mu^n_{j+}}\cdot \mathbf{1}_{\mathbf{N}}(\hat{\mu}^n_{j+})\cdot\Big|\sum_{k=1}^{K}\Big(\bar{R}^n_{jk}(\hat{\mu}^n_{j+}) - \bar{R}^n_{jk}(\mu^n_{j+})\Big)\Big| > M_\delta\Big) \\
\leq \ & P\Big(\sqrt{\mu^n_{j+}}\cdot \mathbf{1}_{\mathbf{N}}(\hat{\mu}^n_{j+})\cdot|\sum_{k=1}^{K}\bar{R}^n_{jk}(\hat{\mu}^n_{j+})| + \sqrt{\mu^n_{j+}}\cdot \mathbf{1}_{\mathbf{N}}(\hat{\mu}^n_{j+})\cdot|\sum_{k=1}^{K}\bar{R}^n_{jk}(\mu^n_{j+})| > M_\delta\Big),
\end{aligned}
$$

hence it suffices by definition of $\bar{R}^n_{jk}$ to show

$$
\begin{aligned}
& \mathbf{1}_{\mathbf{N}}(\hat{\mu}^n_{j+})\sum_{k=1}^{K}\mu^n_{jk}\Big(\mathbf{1}_{\{0\}}(Y^n_{jk})(\frac{2}{\lambda+1}\frac{\hat{\mu}^n_{j+}}{\mu^n_{j+}} - (\frac{\hat{\mu}^n_{j+}}{\mu^n_{j+}})^2) + \mathbf{1}_{\mathbf{N}}(Y^n_{jk})R^n_{jk}(\hat{\mu}^n_{\cdot+})\Big) \\
= \ & O_p\Big(\frac{1}{\sqrt{\mu^n_{j+}}}\Big)
\end{aligned}
\tag{5.43}
$$

and

$$
\mathbf{1}_{\mathbf{N}}(\hat{\mu}^n_{j+})\sum_{k=1}^{K}\mu^n_{jk}\Big(\mathbf{1}_{\{0\}}(Y^n_{jk})(\frac{2}{\lambda+1}-1) + \mathbf{1}_{\mathbf{N}}(Y^n_{jk})R^n_{jk}(\mu^n_{\cdot+})\Big) = O_p\Big(\frac{1}{\sqrt{\mu^n_{j+}}}\Big). \tag{5.44}
$$

Considering (5.43), the following inequality holds

$$
\begin{aligned}
& \Big|\mathbf{1}_{\mathbf{N}}(\hat{\mu}^n_{j+})\cdot\sum_{k=1}^{K}\mu^n_{jk}\Big(\mathbf{1}_{\{0\}}(Y^n_{jk})(\frac{2}{\lambda+1}\cdot\frac{\hat{\mu}^n_{j+}}{\mu^n_{j+}} - (\frac{\hat{\mu}^n_{j+}}{\mu^n_{j+}})^2) + \mathbf{1}_{\mathbf{N}}(Y^n_{jk})R^n_{jk}(\hat{\mu}^n_{\cdot+})\Big)\Big| \\
\leq \ & \sum_{k=1}^{K}\mu^n_{jk} \\
& \cdot\Big|\mathbf{1}_{\mathbf{N}}(\hat{\mu}^n_{j+})\cdot\mathbf{1}_{\{0\}}(Y^n_{jk})(\frac{2}{\lambda+1}\cdot\frac{\hat{\mu}^n_{j+}}{\mu^n_{j+}} - (\frac{\hat{\mu}^n_{j+}}{\mu^n_{j+}})^2) \\
& +\mathbf{1}_{\mathbf{N}}(Y^n_{jk})\int_0^1\frac{(1-z)^2}{2!}\sum_{|\alpha|=3}\frac{3!}{\alpha!}D^\alpha a_\lambda\Big(1+z(\frac{Y^n_{jk}}{\mu^n_{jk}}-1),1+z(\frac{\hat{\mu}^n_{j+}}{\mu^n_{j+}}-1)\Big) \\
& \qquad \cdot(\frac{Y^n_{jk}}{\mu^n_{jk}}-1)^{\alpha_1}\cdot(\frac{\hat{\mu}^n_{j+}}{\mu^n_{j+}}-1)^{\alpha_2}dz\Big|.
\end{aligned}
\tag{5.45}
$$

The asymptotic order of the term in absolute brackets can now be obtained using obvious conversions and inequalities first, and explaining the final arguments afterwards:

(consider any $k \in \{1,\ldots,K\}$):

$$\Big|\Big(\mathbf{1}_{\mathbf{N}}(\hat\mu^n_{j+})\cdot \mathbf{1}_{\{0\}}(Y^n_{jk})\cdot\big((\frac{\hat\mu^n_{j+}}{\mu^n_{j+}})^2-\frac{2}{\lambda+1}\cdot\frac{\hat\mu^n_{j+}}{\mu^n_{j+}}\big)\cdot(-1)^3$$

$$+\mathbf{1}_{\mathbf{N}}(Y^n_{jk})\cdot\int_0^1\frac{(1-z)^2}{2!}\frac{3!}{3!0!}D_1^3a_\lambda\Big(1+z(\frac{Y^n_{jk}}{\mu^n_{jk}}-1),1+z(\frac{\hat\mu^n_{j+}}{\mu^n_{j+}}-1)\Big)dz\Big)$$

$$\cdot(\frac{Y^n_{jk}}{\mu^n_{jk}}-1)^3$$

$$+\mathbf{1}_{\mathbf{N}}(Y^n_{jk})\cdot\sum_{\substack{|\alpha|=3\\ \alpha\neq(3,0)}}\int_0^1\frac{(1-z)^2}{2!}\frac{3!}{\alpha!}D^\alpha a_\lambda\Big(1+z(\frac{Y^n_{jk}}{\mu^n_{jk}}-1),1+z(\frac{\hat\mu^n_{j+}}{\mu^n_{j+}}-1)\Big)dz$$

$$\cdot(\frac{Y^n_{jk}}{\mu^n_{jk}}-1)^{\alpha_1}\cdot(\frac{\hat\mu^n_{j+}}{\mu^n_{j+}}-1)^{\alpha_2}\Big|$$

$$\leq\ \Big(\mathbf{1}_{\mathbf{N}}(\hat\mu^n_{j+})\cdot\mathbf{1}_{\{0\}}(Y^n_{jk})\cdot\Big|(\frac{\hat\mu^n_{j+}}{\mu^n_{j+}})^2-\frac{2}{\lambda+1}\cdot\frac{\hat\mu^n_{j+}}{\mu^n_{j+}}\Big|$$

$$+\mathbf{1}_{\mathbf{N}}(Y^n_{jk})\cdot\Big|\int_0^1\frac{(1-z)^2}{2!}D_1^3a_\lambda\Big(1+z(\frac{Y^n_{jk}}{\mu^n_{jk}}-1),1+z(\frac{\hat\mu^n_{j+}}{\mu^n_{j+}}-1)\Big)dz\Big|\Big)$$

$$\cdot\frac{1}{\sqrt{\mu^n_{jk}}^3}\cdot\Big|\frac{Y^n_{jk}-\mu^n_{jk}}{\sqrt{\mu^n_{jk}}}\Big|^3$$

$$+\mathbf{1}_{\mathbf{N}}(Y^n_{jk})\cdot\sum_{\substack{|\alpha|=3\\ \alpha\neq(3,0)}}\Big|\int_0^1\frac{(1-z)^2}{2!}\frac{3!}{\alpha!}D^\alpha a_\lambda\Big(1+z(\frac{Y^n_{jk}}{\mu^n_{jk}}-1),1+z(\frac{\hat\mu^n_{j+}}{\mu^n_{j+}}-1)\Big)dz\Big|$$

$$\cdot\frac{1}{\sqrt{\mu^n_{jk}}^{\alpha_1}}\Big|\frac{Y^n_{jk}-\mu^n_{jk}}{\sqrt{\mu^n_{jk}}}\Big|^{\alpha_1}\cdot\frac{1}{\sqrt{\mu^n_{j+}}^{\alpha_2}}\Big|\frac{\hat\mu^n_{j+}-\mu^n_{j+}}{\sqrt{\mu^n_{j+}}}\Big|^{\alpha_2}$$

$$\leq\ \Big(\Big|(\frac{\hat\mu^n_{j+}}{\mu^n_{j+}})^2-\frac{2}{\lambda+1}\cdot\frac{\hat\mu^n_{j+}}{\mu^n_{j+}}\Big|$$

$$+\mathbf{1}_{\mathbf{N}}(Y^n_{jk})\cdot\sup_{z\in[0,1]}\Big|D_1^3a_\lambda\Big(1+z(\frac{Y^n_{jk}}{\mu^n_{jk}}-1),1+z(\frac{\hat\mu^n_{j+}}{\mu^n_{j+}}-1)\Big)\Big|\Big)$$

$$\cdot\frac{1}{\sqrt{\mu^n_{jk}}^3}\cdot\Big|\frac{Y^n_{jk}-\mu^n_{jk}}{\sqrt{\mu^n_{jk}}}\Big|^3$$

$$+\mathbf{1}_{\mathbf{N}}(Y^n_{jk})\cdot\sum_{\substack{|\alpha|=3\\ \alpha\neq(3,0)}}\sup_{z\in[0,1]}\Big|D^\alpha a_\lambda\Big(1+z(\frac{Y^n_{jk}}{\mu^n_{jk}}-1),1+z(\frac{\hat\mu^n_{j+}}{\mu^n_{j+}}-1)\Big)\Big|$$

$$\cdot\frac{1}{\sqrt{\mu^n_{jk}}^3}\Big|\frac{Y^n_{jk}-\mu^n_{jk}}{\sqrt{\mu^n_{jk}}}\Big|^{\alpha_1}\cdot\Big|\frac{\hat\mu^n_{j+}-\mu^n_{j+}}{\sqrt{\mu^n_{j+}}}\Big|^{\alpha_2}$$

$$= \quad O_p(1) \cdot O_p\Big(\frac{1}{\sqrt{\mu^n_{jk}}^3}\Big). \tag{5.46}$$

Except for the last statement, similar to the proof of Lemma 5.4, up to here only arguments were applied, which do not use specific information about the considered distribution models. To see the validity of the last equation sign, now additionally $Var(Y^n_{jk}) \leq \mu^n_{jk}$ for all j, k, n is needed, which implies $Var(\hat{\mu}^n_{j+}) \leq \mu^n_{j+}$ for all j, n and holds for column multinomial as well as Poisson sampling. This gives on the one hand $\frac{\hat{\mu}^n_{j+} - \mu^n_{j+}}{\sqrt{\mu^n_{j+}}} = O_p(1)$ for all j (see Lemma 5.4, (5.33)), and in particular ($K = 1$) $\frac{Y^n_{jk} - \mu^n_{jk}}{\sqrt{\mu^n_{jk}}} = O_p(1)$. On the other hand, this variance property is necessary to apply Lemma 8.2, which yields for every j, k

$$\begin{aligned}
& \mathbf{1}_{\mathbf{N}}(Y^n_{jk}) \cdot \sup_{z \in [0,1]} \Big| D^\alpha a_\lambda\Big(1 + z(\frac{Y^n_{jk}}{\mu^n_{jk}} - 1), 1 + z(\frac{\hat{\mu}^n_{j+}}{\mu^n_{j+}} - 1)\Big)\Big| \\
= \quad & \mathbf{1}_{(0,\infty)^2}(\frac{Y^n_{jk}}{\mu^n_{jk}}, \frac{\hat{\mu}^n_{j+}}{\mu^n_{j+}}) \cdot \sup_{z \in [0,1]} \Big| D^\alpha a_\lambda\Big(1 + z(\frac{Y^n_{jk}}{\mu^n_{jk}} - 1), 1 + z(\frac{\hat{\mu}^n_{j+}}{\mu^n_{j+}} - 1)\Big)\Big| \\
=: \quad & \mathbf{1}_{(0,\infty)^2}(\frac{Y^n_{jk}}{\mu^n_{jk}}, \frac{\hat{\mu}^n_{j+}}{\mu^n_{j+}}) \cdot f(\frac{Y^n_{jk}}{\mu^n_{jk}}, \frac{\hat{\mu}^n_{j+}}{\mu^n_{j+}})
\end{aligned}$$

being stochastically bounded for $\alpha \in \mathbf{N}^2_{\mathbf{0}}$, $|\alpha| = 3$. Further, the continuity of f on $\mathbf{R}^+ \times \mathbf{R}^+$ was used, and $(\frac{Y^n_{jk}}{\mu^n_{jk}}, \frac{\hat{\mu}^n_{j+}}{\mu^n_{j+}})$ takes the role of the random vector $\bar{X}^n$ in Lemma 8.2.

Statement (5.46) concerning the asymptotic order of the term in absolute brackets in (5.45) thus leads to (5.43), since it holds

$$\begin{aligned}
& \Big|\sum_{k=1}^{K} \mu^n_{jk}\Big(\mathbf{1}_{\{0\}}(Y^n_{jk})(\frac{2}{\lambda+1} \cdot \frac{\hat{\mu}^n_{j+}}{\mu^n_{j+}} - (\frac{\hat{\mu}^n_{j+}}{\mu^n_{j+}})^2) + \mathbf{1}_{\mathbf{N}}(Y^n_{jk}) R^n_{jk}(\hat{\mu}^n_{\cdot+})\Big)\Big| \\
= \quad & \sum_{k=1}^{K} \mu^n_{jk} \cdot O_p\Big(\frac{1}{\sqrt{\mu^n_{jk}}^3}\Big) \\
= \quad & O_p\Big(\frac{1}{\sqrt{\mu^n_{j+}}}\Big).
\end{aligned}$$

The last equality follows because each $\pi^n_{jk|C}(\theta_0)$ is bounded away from zero (condition (RC2)) thus giving $\frac{\mu^n_{j+}}{\mu^n_{jk}} = \frac{1}{\pi^n_{jk|C}(\theta_0)} = O(1)$ for all j, k. Analogous argumentation as in the proof of (5.43) yields for the more simple statement (5.44) the same asymptotic order and hence (5.39) is shown.

To prove (5.40) for $m^n_\lambda(\mu^n_{\cdot+}, \theta_0) = \sum_{j=1}^{J^n} \sum_{k=1}^{K} e_\lambda(\mu^n_{j+}, \pi^n_{jk|C}(\theta_0))$ as defined in (5.1) and (5.2), it obviously suffices to show

$$e_\lambda(\hat{\mu}^n_{j+}, \pi^n_{jk|C}(\theta_0)) = e_\lambda(\mu^n_{j+}, \pi^n_{jk|C}(\theta_0)) + O_p\Big(\sqrt{\frac{1}{\mu^n_{j+}}}\Big) \quad \text{for all } j,k$$

respectively (cp. Lemma 5.4, proof of (5.35) and (5.27))

$$\mathbf{1_N}(\hat{\mu}^n_{j+})\Big(e_\lambda(\hat{\mu}^n_{j+}, \pi^n_{jk|C}(\theta_0)) - e_\lambda(\mu^n_{j+}, \pi^n_{jk|C}(\theta_0))\Big) = O_p\Big(\sqrt{\frac{1}{\mu^n_{j+}}}\Big) \quad \text{for all } j,k.$$

This follows from Lemma 8.3, since for both distribution models the required conditions are met (in particular $Var(\hat{\mu}^n_{j+}) \leq \mu^n_{j+}$ holds and the expectations μ^n_{j+} of the row sums are bounded away from zero), if for every j,k,n holds

$$\mu^n_{j+}\Big|\frac{\partial}{\partial\mu^n_{j+}} e_\lambda(\mu^n_{j+}, \pi^n_{jk|C}(\theta_0))\Big| \leq c \in \mathbf{R}^+ \text{ for all } \mu^n_{j+} \geq \epsilon \quad (c \in \mathbf{R}^+ \text{ constant}). \tag{5.47}$$

Since e_λ is a Poisson expectation, this term can now be explicitly stated as follows

$$\begin{aligned}
&\mu^n_{j+}\Big|\frac{\partial}{\partial\mu^n_{j+}} e_\lambda(\mu^n_{j+}, \pi^n_{jk|C}(\theta_0))\Big| \\
= \;& \mu^n_{j+}\Big|\frac{\partial}{\partial\mu^n_{j+}} E\Big(a_\lambda(Y^n_{jk}, \mu^n_{j+}\pi^n_{jk|C}(\theta_0))\Big)\Big| \\
= \;& \mu^n_{j+}\Big|\frac{\partial}{\partial\mu^n_{j+}} (\mu^n_{j+}\pi^n_{jk|C}(\theta_0)) \cdot \frac{\partial}{\partial\mu^n_{j+}\pi^n_{jk|C}(\theta_0)} E\Big(a_\lambda(Y^n_{jk}, \mu^n_{j+}\pi^n_{jk|C}(\theta_0))\Big)\Big| \\
= \;& \mu^n_{j+}\pi^n_{jk|C}(\theta_0) \cdot \Big|\frac{\partial}{\partial\mu^n_{j+}\pi^n_{jk|C}(\theta_0)} E\Big(a_\lambda(Y^n_{jk}, \mu^n_{j+}\pi^n_{jk|C}(\theta_0))\Big)\Big|.
\end{aligned}$$

Using $\mu^n_{jk} = \mu^n_{j+}\pi^n_{jk|C}(\theta_0)$ being bounded away from zero for each j,k,n, the required inequality for (5.47) immediately follows from the auxiliary results given in Lemma 4.8 (Lemma 4.8 g)) thus completing the proof. □

As a conclusion of the preceding lemmata, the following corollary summarizes all steps of the approximation. Above all, it gives the final approximated statistic, which then will be further studied in the next chapter.

Corollary 5.9 *Consider $\lambda \in (-1, 1]$ if column–multinomial and $\lambda > -1$ if Poisson sampling is given. Further suppose that all assumptions from Lemma 5.3 to 5.8 are met, i.e. $\sqrt{n}(\hat{\theta}^n - \theta_0) = O_p(1)$ (LC2), $\hat{\theta}^n - \theta_0 = (I^n(\mu^n_{\cdot+}, \theta_0))^{-1}U^n(\theta_0) + O_p(\frac{1}{n})$ (LC3), $\mu^n_{jk} \geq \epsilon > 0$ for all j, k, n (BC), and, in case of column–multinomial sampling, additionally $\frac{J^n}{n} \longrightarrow 0$, $\frac{1}{n}I^n(\mu^n_{\cdot+}, \theta_0) \longrightarrow I_\infty$ (LC1) and $\max_{1\leq j\leq J^n} \pi^n_{jk|D}(\theta_0) \longrightarrow 0$ for all k (MD3). Then all results together yield for $Z^n_\lambda(\hat{\mu}^n_{\cdot+}, \hat{\theta}^n) = SD^n_\lambda(\hat{\mu}^n_{\cdot+}, \hat{\theta}^n) - m^n_\lambda(\hat{\mu}^n_{\cdot+}, \hat{\theta}^n)$ the approximation*

$$SD^n_\lambda(\hat{\mu}^n_{\cdot+}, \hat{\theta}^n) - m^n_\lambda(\hat{\mu}^n_{\cdot+}, \hat{\theta}^n) = \Psi^n_{\lambda+} - m^n_\lambda(\mu^n_{\cdot+}, \theta^n_0) + O_p(1) + \sum_{j=1}^{J^n} O_p(1)\Big((\frac{\mu^n_{j+}}{n})^{\frac{1}{2}} + (\frac{1}{\mu^n_{j+}})^{\frac{1}{2}}\Big)$$

with $\Psi^n_{\lambda+} = SD^n_\lambda(\mu^n_{\cdot+}, \theta_0) - \sum_{j=1}^{J^n} \frac{(\hat{\mu}^n_{j+} - \mu^n_{j+})^2}{\mu^n_{j+}} - c^n_\lambda(\mu^n_{\cdot+}, \theta_0)(I^n(\mu^n_{\cdot+}, \theta_0))^{-1}U^n(\theta_0)$ (c^n_λ as defined in (5.4)) and Poisson expectation $m^n_\lambda(\mu^n_{\cdot+}, \theta^n_0) = E\Big(SD^n_\lambda(\mu^n_{\cdot+}, \theta_0|X^n)\Big)$. Stating both distribution models separately gives

$$\Psi^n_{\lambda+}(Y^n) = SD^n_\lambda(\mu^n_{\cdot+}, \theta_0|Y^n) - \sum_{j=1}^{J^n} \frac{(Y^n_{j+} - \mu^n_{j+})^2}{\mu^n_{j+}} - c^n_\lambda(\mu^n_{\cdot+}, \theta_0)(I^n(\mu^n_{\cdot+}, \theta_0))^{-1}U^n(\theta_0|Y^n),$$

$$\Psi^n_{\lambda+}(X^n) = SD^n_\lambda(\mu^n_{\cdot+}, \theta_0|X^n) - \sum_{j=1}^{J^n} \frac{(X^n_{j+} - \mu^n_{j+})^2}{\mu^n_{j+}} - c^n_\lambda(\mu^n_{\cdot+}, \theta_0)(I^n(\mu^n_{\cdot+}, \theta_0))^{-1}U^n(\theta_0|X^n).$$

Proof:
Combining the single approximation steps from the preceding lemmata yields

$$SD^n_\lambda(\hat{\mu}^n_{\cdot+}, \hat{\theta}^n) - m^n_\lambda(\hat{\mu}^n_{\cdot+}, \hat{\theta}^n)$$
$$= Z^n_\lambda(\hat{\mu}^n_{\cdot+}, \hat{\theta}^n)$$
$$= Z^n_\lambda(\hat{\mu}^n_{\cdot+}, \theta_0) + D_\theta Z^n_\lambda(\hat{\mu}^n_{\cdot+}, \theta_0)(\hat{\theta}^n - \theta_0) + O_p(1) \tag{5.48}$$
$$= Z^n_\lambda(\hat{\mu}^n_{\cdot+}, \theta_0) + D_\theta Z^n_\lambda(\mu^n_{\cdot+}, \theta_0)(\hat{\theta}^n - \theta_0) + O_p\Big(\sum_{j=1}^{J^n}(\frac{\mu^n_{j+}}{n})^{\frac{1}{2}}\Big) + O_p(1) \tag{5.49}$$
$$= Z^n_\lambda(\hat{\mu}^n_{\cdot+}, \theta_0) - c^n_\lambda(\mu^n_{\cdot+}, \theta_0)(\hat{\theta}^n - \theta_0) + O_p\Big(\sum_{j=1}^{J^n}(\frac{\mu^n_{j+}}{n})^{\frac{1}{2}}\Big) + O_p(1) \tag{5.50}$$
$$= Z^n_\lambda(\hat{\mu}^n_{\cdot+}, \theta_0) - c^n_\lambda(\mu^n_{\cdot+}, \theta_0)(I^n(\mu^n_{\cdot+}, \theta_0))^{-1}U^n(\theta_0) + O_p\Big(\sum_{j=1}^{J^n}(\frac{\mu^n_{j+}}{n})^{\frac{1}{2}}\Big) + O_p(1) \tag{5.51}$$
$$= Z^n_\lambda(\mu^n_{\cdot+}, \theta_0) - \sum_{j=1}^{J^n} a_1(\hat{\mu}^n_{j+}, \mu^n_{j+}) - c^n_\lambda(\mu^n_{\cdot+}, \theta_0)(I^n(\mu^n_{\cdot+}, \theta_0))^{-1}U^n(\theta_0)$$

$$+O_p\Big(\sum_{j=1}^{J^n}(\frac{1}{\mu_{j+}^n})^{\frac{1}{2}}\Big)+O_p\Big(\sum_{j=1}^{J^n}(\frac{\mu_{j+}^n}{n})^{\frac{1}{2}}\Big)+O_p(1) \tag{5.52}$$

$$= SD_\lambda^n(\mu_{\cdot+}^n,\theta_0)-m_\lambda^n(\mu_{\cdot+}^n,\theta_0)-\sum_{j=1}^{J^n}a_1(\hat{\mu}_{j+}^n,\mu_{j+}^n)-c_\lambda^n(\mu_{\cdot+}^n,\theta_0)(I^n(\mu_{\cdot+}^n,\theta_0))^{-1}U^n(\theta_0)$$

$$+O_p\Big(\sum_{j=1}^{J^n}(\frac{1}{\mu_{j+}^n})^{\frac{1}{2}}\Big)+O_p\Big(\sum_{j=1}^{J^n}(\frac{\mu_{j+}^n}{n})^{\frac{1}{2}}\Big)+O_p(1).$$

In (5.48), (5.50) – (5.52) the results given in Lemma 5.3 and Lemma 5.5 to 5.8 were adopted. In (5.50), which is the only step not holding for arbitrary $\lambda > -1$ since it has not been verified for $\lambda > 1$ in the column–multinomial case, $D_\theta Z_\lambda^n(\mu_{\cdot+},\theta_0)$ was replaced by its (Poisson) expectation $E(D_\theta Z_\lambda^n(\mu_{\cdot+},\theta_0|X^n)) = -c_\lambda^n(\mu_{\cdot+},\theta_0)$. Assumption $\sqrt{n}(\hat{\theta}^n-\theta_0)=O_p(1)$ (LC2) and Lemma 5.4 stating

$$D_\theta Z_\lambda^n(\hat{\mu}_{\cdot+}^n,\theta_0) \;=\; D_\theta Z_\lambda^n(\mu_{\cdot+}^n,\theta_0)+O_p\Big(\sum_{j=1}^{J^n}\sqrt{\mu_{j+}^n}\Big)$$

finally yield (5.49). □

6. Limiting Results for the Goodness–of–Fit Statistic

In this chapter, for both distribution models, Poisson and column–multinomial, the asymptotic normality of the test statistic under the nullhypothesis will be proved. Starting point is the approximation $\Psi^n_{\lambda+}$ derived in chapter 5 (see Cor. 5.9). This statistic, which is, as seen, analytically identical for both distribution models, will be shown in section 6.1 to have a normal limit. In the case of Poisson sampling, this will be done applying the central limit theorem. When column–multinomial sampling is considered, the approximated statistic is, unlike the Poisson model, not a sum of independent variables; hence the central limit theorem cannot be applied directly. Instead, the asymptotic normality will be derived using Morris' method (1975) as presented in chapter 3. Since for both approaches the "true" and hence unknown standard deviations will be taken, which for a concrete application certainly will have to be estimated, in section 6.2 the consistency of the variance estimation will be shown. With this final statement, as a conclusion to the preceding results, the main theorems of this thesis can be given, namely the asymptotic normality of the goodness–of–fit statistic with estimated mean and standard deviation.
Since this chapter is a continuation of chapter 5, the situation described in its beginning is restored. In particular, the regularity conditions (RC1) – (RC3) and (LC0) are again assumed to hold throughout the chapter, other requirements to be explicitly stated each time needed. As already announced in section 2.3, where the final test statistic was introduced, now additional assumptions concerning the variance and the marginal distributions will be required. The latter ones will only be necessary for the main results in section 6.2, and will be given there together with the other conditions.

6.1 Asymptotic Normality of the Approximated Statistic

The asymptotic normality for the approximation $\Psi^n_{\lambda+} - m_\lambda(\mu^n_{\cdot+}, \theta_0)$ of the difference $SD^n_\lambda(\hat{\mu}^n_{\cdot+}, \hat{\theta}^n) - m^n_\lambda(\hat{\mu}^n_{\cdot+}, \hat{\theta}^n)$ derived in chapter 5, where m^n_λ is the Poisson expectation of SD^n_λ as defined in (5.2), will, as already mentioned, be proved using the central limit theorem in case of Poisson, and Morris' method (Lemma 3.4) in case of column–

multinomial distribution. Considering the Poisson expectation of $\Psi^n_{\lambda+}$ it holds

$$E\Big(\sum_{j=1}^{J^n} a_1(X^n_{j+},\mu^n_{j+})\Big) = E\Big(\sum_{j=1}^{J^n} \frac{(X^n_{j+}-\mu^n_{j+})^2}{\mu^n_{j+}}\Big) = \sum_{j=1}^{J^n} 1 = J^n$$

and

$$E(U^n(\theta_0|X^n)) = \sum_{j=1}^{J^n} E(U^n_j(\theta_0|X^n_{j\cdot})) = \sum_{j=1}^{J^n}\sum_{k=1}^{K} \mu^n_{jk} D^T_\theta \log \pi^n_{jk|C}(\theta_0) = 0$$

thus giving

$$\begin{aligned} & E(\Psi^n_{\lambda+}(X^n)) \\ = \; & E\Big(SD^n_\lambda(\mu^n_{j+},\theta_0|X^n) - \sum_{j=1}^{J^n} a_1(X^n_{j+},\mu^n_{j+}) - c^n_\lambda(\mu^n_{\cdot+},\theta_0)(I^n(\mu^n_{\cdot+},\theta_0))^{-1}U^n(\theta_0|X^n)\Big) \\ = \; & m^n_\lambda(\mu^n_{\cdot+},\theta_0) - J^n \end{aligned}$$

with c^n_λ being the covariance between SD^n_λ and the score vector U^n in the Poisson distribution model as defined in (5.4). Since both approaches require a centering with the Poisson expectation $E(\Psi^n_{\lambda+}(X^n))$, the expected value of the correction term derived in Lemma 5.8, i.e. $E(a_1(X^n_{j+},\mu^n_{j+})) = J^n$, must be incorporated. In the following, the recentered statistic $\Psi^n_{\lambda+} - m_\lambda(\mu^n_{\cdot+},\theta_0) + J^n$ will be considered.
Since the approach for Poisson sampling is clear, and in order to clarify the proceedings and theorems to come, let Morris' construction method concerning column–multinomial sampling from section 3.1 be recapitulated for the special case considered here. As already explained in section 3.1, an application of this method is necessary to meet the covariance condition required for Lemma 3.4. Starting with

$$\begin{aligned} \Psi^n_{\lambda+}(z) &= \sum_{j=1}^{J^n} \Psi^n_{\lambda j}(z_{j\cdot}), \\ \Psi^n_{\lambda j}(z_{j\cdot}) &= \sum_{k=1}^{K} a_\lambda(z_{jk},\mu^n_{jk}) - a_1(z_{j+},\mu^n_{j+}) - c^n_\lambda(\mu^n_{\cdot+},\theta_0)(I^n(\mu^n_{\cdot+},\theta_0))^{-1}U^n_j(z_{j\cdot},\theta_0) \end{aligned}$$

($z = (z_{jk})_{j,k}$ is a $J^n \times K$ table with nonnegative entries), let for $j \in \{1,\ldots,J^n\}$ now functions $f^n_{\lambda j}$ be defined as follows:

$$f^n_{\lambda j}(z_{j\cdot}) = \Psi^n_{\lambda j}(z_{j\cdot}) - E(\Psi^n_{\lambda j}(X^n_{j\cdot})) - \sum_{k=1}^{K} \gamma^n_{\lambda k}(z_{jk} - \mu^n_{jk})$$

with a correction term

$$\gamma^n_{\lambda k} = \frac{1}{\mu^n_{+k}} \sum_{j=1}^{J^n} Cov\Big(\Psi^n_{\lambda j}(X^n_{j\cdot}), X^n_{jk}\Big)$$

$$
\begin{aligned}
&= \frac{1}{\mu^n_{+k}} \sum_{j=1}^{J^n} Cov\Big(\sum_{k'=1}^{K} a_\lambda(X^n_{jk'}, \mu^n_{jk'}), X^n_{jk}\Big) \\
&\quad -\frac{1}{\mu^n_{+k}} \sum_{j=1}^{J^n} Cov\Big(a_1(X^n_{j+}, \mu^n_{j+}), X^n_{jk}\Big) \\
&\quad -\frac{1}{\mu^n_{+k}} \sum_{j=1}^{J^n} Cov\Big(c^n_\lambda(\mu^n_{\cdot +}, \theta_0)(I^n(\mu^n_{\cdot +}, \theta_0))^{-1} U^n_j(\theta_0, X^n_{j\cdot}), X^n_{jk}\Big) \\
&= \frac{1}{\mu^n_{+k}} \sum_{j=1}^{J^n} \Big(Cov(a_\lambda(X^n_{jk}, \mu^n_{jk}), X^n_{jk}) - \pi^n_{jk|C}(\theta_0)\Big).
\end{aligned}
$$

The last formula for $\gamma^n_{\lambda k}$ is generated through simple calculations. With such chosen $f^n_{\lambda j}$, the covariance condition for Lemma 3.4 is met (cp. sec. 3.1 (3.5)). This Lemma will now be applied to prove the limiting normality of the column–multinomial sum $\sum_{j=1}^{J^n} f^n_{\lambda j}(Y^n_{j\cdot}) = \Psi^n_{\lambda +}(Y^n) - E(\Psi^n_{\lambda +}(X^n))$ in question, scaled with the Poisson variance $Var(\sum_{j=1}^{J^n} f^n_{\lambda j}(X^n_{j\cdot}))$. The proof consists of two major parts: The condition for the convergence of the conditional distribution as well as the Lindeberg Condition for the central limit theorem for the transformed Poisson approximation

$$\sum_{j=1}^{J^n} f^n_{\lambda j}(X^n_{j\cdot}) \Big/ Var(\sum_{j=1}^{J^n} f^n_{\lambda j}(X^n_{j\cdot}))$$

with

$$\sum_{j=1}^{J^n} f^n_{\lambda j}(X^n_{j\cdot}) = \Psi^n_{\lambda +}(X^n) - \sum_{j=1}^{J^n} \sum_{k=1}^{K} \gamma^n_{\lambda k}(X_{jk} - \mu^n_{jk}) - E(\Psi^n_{\lambda +}(X^n)$$

have to be verified. Since this Poisson statistic, up to the additional correction term, coincides with the pure Poisson approximation, both statistics will be treated together in Theorem 6.1, where the validity of the Ljapounov Condition will be checked for both. In Theorem 6.2, the asymptotic normality of the approximation for column–multinomial sampling applying Lemma 3.4 is shown by proving the remaining condition concerning the convergence of the conditional distribution.

Theorem 6.1 *Suppose that the assumptions $\mu^n_{jk} \geq \epsilon > 0$ for all j, k, n (BC), $\frac{1}{n} I^n(\mu^n_{\cdot +}, \theta_0) \longrightarrow I_\infty$ positive definite (LC1) and*

$$(J^n)^2 \cdot \sum_{j=1}^{J^n} \sum_{k=1}^{K} (\mu^n_{jk})^2 / (\mu^n_{++})^4 = o(1)$$

hold for the asymptotics $n \to \infty$. For each $j \in \{1, \ldots, J^n\}$, $n \in \mathbf{N}$, let the function $\Psi^n_{\lambda j} : [0, \infty)^K \to \mathbf{R}$, $z_{j\cdot} = (z_{j1}, \ldots, z_{jK}) \mapsto \Psi^n_{\lambda j}(z_{j\cdot})$ ($z = (z_{jk})_{j,k}$ is an arbitrary

$J^n \times K$ table with nonnegative entries) be defined as follows:

$$\Psi^n_{\lambda j}(z_{j\cdot}) = \sum_{k=1}^{K} a_\lambda(z_{jk}, \mu^n_{jk}) - a_1(z_{j+}, \mu^n_{j+}) - c^n_\lambda(\mu^n_{\cdot+}, \theta_0)(I^n(\mu^n_{\cdot+}, \theta_0))^{-1} U^n_j(z_{j\cdot}, \theta_0)$$

with especially $c^n_\lambda(\mu^n_{\cdot+}, \theta_0) = \sum_{j=1}^{J^n} \sum_{k=1}^{K} D_\theta \log \pi^n_{jk|C}(\theta_0) Cov(a_\lambda(X^n_{jk}, \mu^n_{jk}), X^n_{jk})$ *(see (5.4)). Hence* $\Psi^n_{\lambda+}(z) = \sum_{j=1}^{J^n} \Psi^n_{\lambda j}(z_{j\cdot})$ *coincides with the analytical definition of the approximation derived in chapter 5, Cor. 5.9. Now consider the Poisson statistic* $\Psi^n_{\lambda+}(X^n)$ *and the transformation* $\Psi^n_{\lambda+}(X^n) - \sum_{j=1}^{J^n} \sum_{k=1}^{K} \gamma^n_{\lambda k}(X^n_{jk} - \mu^n_{jk})$*, which have the variances*

$$\begin{aligned}
\sigma^{n2}_\lambda(\mu^n_{\cdot+}, \theta_0) &:= Var(\Psi^n_{\lambda+}(X^n)),\\
s^{n2}_\lambda(\mu^n_{\cdot+}, \theta_0) &:= Var\Big(\Psi^n_{\lambda+}(X^n) - \sum_{j=1}^{J^n} \sum_{k=1}^{K} \gamma^n_{\lambda k}(X_{jk} - \mu^n_{jk})\Big)\\
&= \sigma^{n2}_\lambda(\mu^n_{\cdot+}, \theta_0) - \sum_{k=1}^{K} \mu^n_{+k}(\gamma^n_{\lambda k})^2
\end{aligned}$$

with $\gamma^n_{\lambda k} = 1/\mu^n_{+k} \sum_{j=1}^{J^n} \Big(Cov(a_\lambda(X^n_{jk}, \mu^n_{jk}), X^n_{jk}) - \pi^n_{jk|C}(\theta_0)\Big) = O(J^n/\mu^n_{+k})$*. For these let the variance condition* $J^n/\sigma^{n2}_\lambda(\mu^n_{\cdot+}, \theta_0) = O(1)$ *(VCP) and* $J^n/s^{n2}_\lambda(\mu^n_{\cdot+}, \theta_0) = O(1)$ *(VCC) hold. Then the standardizations of both statistics fulfil Ljapounov's condition for the central limit theorem. Using* $E(\Psi^n_{\lambda+}(X^n)) = E\Big(\Psi^n_{\lambda+}(X^n) - \sum_{j=1}^{J^n} \sum_{k=1}^{K} \gamma^n_{\lambda k} \cdot (X_{jk} - \mu^n_{jk})\Big) = m^n_\lambda(\mu^n_{\cdot+}, \theta_0) - J^n$ *(m^n_λ is the Poisson expectation of SD^n_λ) in particular holds*

$$\frac{\Psi^n_{\lambda+}(X^n) - m^n_\lambda(\mu^n_{\cdot+}, \theta_0) + J^n}{\sigma^n_\lambda(\mu^n_{\cdot+}, \theta_0)} \xrightarrow{\mathcal{L}} N(0,1),$$

$$\frac{\Psi^n_{\lambda+}(X^n) - \sum_{j=1}^{J^n} \sum_{k=1}^{K} \gamma^n_{\lambda k}(X_{jk} - \mu^n_{jk}) - m^n_\lambda(\mu^n_{\cdot+}, \theta_0) + J^n}{s^n_\lambda(\mu^n_{\cdot+}, \theta_0)} \xrightarrow{\mathcal{L}} N(0,1).$$

Exhaustive calculation of the variance further gives

$$\begin{aligned}
\sigma^{n2}_\lambda(\mu^n_{\cdot+}, \theta_0) &= Var\Big(\sum_{j=1}^{J^n} \sum_{k=1}^{K} a_\lambda(X^n_{jk}, \mu^n_{j+}\pi^n_{jk|C}(\theta_0))\Big) + 2J^n\\
&+ \sum_{j=1}^{J^n} \frac{1}{\mu^n_{j+}} - 2\sum_{j=1}^{J^n} \frac{1}{\mu^n_{j+}} \sum_{k=1}^{K} Cov(a_\lambda(X^n_{jk}, \mu^n_{jk}), (X^n_{jk})^2)\\
&+ 4\sum_{j=1}^{J^n} \sum_{k=1}^{K} \pi^n_{jk|C}(\theta^n_0) Cov(a_\lambda(X^n_{jk}, \mu^n_{jk}), X^n_{jk})\\
&- c^n_\lambda(\mu^n_{\cdot+}, \theta_0)(I^n(\mu^n_{\cdot+}, \theta_0))^{-1}(c^n_\lambda(\mu^n_{\cdot+}, \theta_0))^T.
\end{aligned}$$

Proof:
The stated standardization terms for both statistics are immediately checked: Using $E(U^n(\theta_0|X^n)) = 0$ and $E(a_1(X^n_{j+}, \mu^n_{j+})) = 1$ gives (cp. preliminaries)

$$E\Big(\Psi^n_{\lambda+}(X^n) - \sum_{j=1}^{J^n}\sum_{k=1}^{K} \gamma^n_{\lambda k}(X_{jk} - \mu^n_{jk})\Big) = E\Big(\Psi^n_{\lambda+}(X^n)\Big) = m^n_\lambda(\mu^n_{\cdot+}, \theta_0) - J^n.$$

$\sigma^{n2}_\lambda(\mu^n_{\cdot+}, \theta_0)$ is by definition the variance of $\Psi^n_{\lambda+}(X^n)$, and for the variance of the transformed statistic holds $s^{n2}_\lambda(\mu^n_{\cdot+}, \theta_0) = \sigma^{n2}_\lambda(\mu^n_{\cdot+}, \theta_0) - \sum_{k=1}^{K} \mu^n_{+k}(\gamma^n_{\lambda k})^2$ by construction (see chapter 3, (3.7)). In particular $\gamma^n_{\lambda k} = O(J^n/\mu^n_{+k})$ holds, since (BC) — i.e. $\mu^n_{jk} \geq \epsilon$ for all j, k, n — implies $\frac{J^n}{\mu^n_{+k}} = O(1)$ and $Cov(a_\lambda(X^n_{jk}, \mu^n_{jk}), X^n_{jk}) = O(1)$ (Lemma 4.8 d)), hence

$$\mu^n_{+k}\gamma^n_{\lambda k} = \sum_{j=1}^{J^n} \Big(Cov(a_\lambda(X^n_{jk}, \mu^n_{jk}), X^n_{jk}) - \pi^n_{jk|C}(\theta_0)\Big) = O(J^n). \tag{6.1}$$

The detailed formula for $\sigma^{n2}_\lambda(\mu^n_{\cdot+}, \theta_0)$ will be verified at the end of the proof.

For the approximation and the transformed approximation, both being sums of independent random variables, Ljapounov's inequality will now be shown in the form

$$\frac{1}{\sigma^{n4}_\lambda(\mu^n_{\cdot+}, \theta_0)} \sum_{j=1}^{J^n} E\Big(\big(\Psi^n_{\lambda j}(X^n_{j\cdot}) - E(\Psi^n_{\lambda j}(X^n_{j\cdot}))\big)^4\Big) \longrightarrow 0$$

respectively

$$\frac{1}{s^{n4}_\lambda(\mu^n_{\cdot+}, \theta_0)} \sum_{j=1}^{J^n} E\Big(\big(\Psi^n_{\lambda j}(X^n_{j\cdot}) - \sum_{k=1}^{K} \gamma^n_{\lambda k}(X^n_{jk} - \mu^n_{jk}) - E(\Psi^n_{\lambda j}(X^n_{j\cdot}))\big)^4\Big) \longrightarrow 0.$$

Since $J^n/\sigma^{n2}_\lambda(\mu^n_{\cdot+}, \theta_0) = O(1)$ (VCP) and $J^n/s^{n2}_\lambda(\mu^n_{\cdot+}, \theta_0) = O(1)$ (VCC) is presumed, it obviously suffices to show

$$\sum_{j=1}^{J^n} E\Big(\big(\Psi^n_{\lambda j}(X^n_{j\cdot}) - E(\Psi^n_{\lambda j}(X^n_{j\cdot}))\big)^4\Big) = o((J^n)^2), \tag{6.2}$$

$$\sum_{j=1}^{J^n} E\Big(\big(\Psi^n_{\lambda j}(X^n_{j\cdot}) - \sum_{k=1}^{K} \gamma^n_{\lambda k}(X^n_{jk} - \mu^n_{jk}) - E(\Psi^n_{\lambda j}(X^n_{j\cdot}))\big)^4\Big) = o((J^n)^2). \tag{6.3}$$

Considering the statistic in (6.3) and the simple inequality

$$\begin{aligned} &\sum_{j=1}^{J^n} E\Big(\big(\Psi^n_{\lambda j}(X^n_{j\cdot}) - E(\Psi^n_{\lambda j}(X^n_{j\cdot})) - \sum_{k=1}^{K} \gamma^n_{\lambda k}(X^n_{jk} - \mu^n_{jk})\big)^4\Big) \\ \leq\ & 8 \cdot \sum_{j=1}^{J^n} E\Big(\big(\Psi^n_{\lambda j}(X^n_{j\cdot}) - E(\Psi^n_{\lambda j}(X^n_{j\cdot}))\big)^4\Big) + 8 \cdot \sum_{j=1}^{J^n} E\Big(\big(\sum_{k=1}^{K} \gamma^n_{\lambda k}(X^n_{jk} - \mu^n_{jk})\big)^4\Big), \end{aligned}$$

the proof of (6.3) can be traced back to that of (6.2) concerning the pure approximation. Additionally, only

$$\sum_{j=1}^{J^n} E\Big(\Big(\sum_{k=1}^{K} \gamma_{\lambda k}^n (X_{jk}^n - \mu_{jk}^n)\Big)^4\Big) = o((J^n)^2)$$

has to be shown, which will be done first. Using $(\sum_{k=1}^K x_k)^4 \leq \sum_{k=1}^K K^3 \cdot x_k^4$ (see (5.11)), $\gamma_{\lambda k}^n = O(J^n/\mu_{+k}^n)$ (6.1) and further $E((X_{jk}^n - \mu_{jk}^n)^4) = 3(\mu_{jk}^n)^2 + \mu_{jk}^n = O((\mu_{jk}^n)^2)$, this is verified as follows:

$$\begin{aligned}
& \sum_{j=1}^{J^n} E\Big(\Big(\sum_{k=1}^{K} \gamma_{\lambda k}^n (X_{jk}^n - \mu_{jk}^n)\Big)^4\Big) \\
\leq\ & \sum_{j=1}^{J^n} E\Big(\sum_{k=1}^{K} K^3 (\gamma_{\lambda k}^n)^4 (X_{jk}^n - \mu_{jk}^n)^4\Big) \\
=\ & \sum_{k=1}^{K} K^3 (\gamma_{\lambda k}^n)^4 \sum_{j=1}^{J^n} E((X_{jk}^n - \mu_{jk}^n)^4) \\
=\ & \sum_{k=1}^{K} O(1) \cdot (\frac{J^n}{\mu_{+k}^n})^4 \cdot \sum_{j=1}^{J^n} (\mu_{jk}^n)^2 \\
=\ & \sum_{k=1}^{K} O(1) \cdot (J^n)^2 \cdot (\frac{\mu_{++}^n}{\mu_{+k}^n})^4 \cdot \frac{(J^n)^2}{(\mu_{++}^n)^4} \cdot \sum_{j=1}^{J^n} (\mu_{jk}^n)^2 \\
=\ & \sum_{k=1}^{K} O(1) \cdot (J^n)^2 \cdot \frac{(J^n)^2}{(\mu_{++}^n)^4} \cdot \sum_{j=1}^{J^n} (\mu_{jk}^n)^2 && (6.4) \\
=\ & o((J^n)^2). && (6.5)
\end{aligned}$$

Equation (6.4) follows with condition (RC2), which assumes $\frac{\mu_{jk}^n}{\mu_{j+}^n} = \pi_{jk|C}^n(\theta_0) \geq \epsilon > 0$ for all j, k, n and hence

$$\frac{\mu_{jk}^n}{\mu_{j+}^n} \geq \epsilon \quad \Rightarrow \quad \sum_{j=1}^{J^n} \mu_{jk}^n = \mu_{+k}^n \geq \epsilon \sum_{j=1}^{J^n} \mu_{j+}^n = \epsilon \cdot \mu_{++}^n \quad \Rightarrow \quad \frac{\mu_{++}^n}{\mu_{+k}^n} \leq \frac{1}{\epsilon}. \qquad (6.6)$$

Finally (6.5) is the presumed condition concerning the marginal distribution.

For the proof of (6.2), consider now $\Psi_{\lambda +}^n(X^n) - E(\Psi_{\lambda +}^n(X^n))$ written as follows

$$\begin{aligned}
& \Psi_{\lambda +}^n(X^n) - E(\Psi_{\lambda +}^n(X^n)) \\
=\ & \sum_{j=1}^{J^n} (\Psi_{\lambda j}^n(X_{j\cdot}^n) - E(\Psi_{\lambda j}^n(X_{j\cdot}^n)))
\end{aligned}$$

$$= \sum_{j=1}^{J^n} \Big(\sum_{k=1}^{K} a_\lambda(X_{jk}^n, \mu_{jk}^n) - a_1(X_{j+}^n, \mu_{j+}^n) - \Big(\sum_{k=1}^{K} E(a_\lambda(X_{jk}^n, \mu_{jk}^n)) - 1 \Big)$$
$$- c_\lambda^n(\mu_{\cdot+}^n, \theta_0) \cdot (I^n(\mu_{\cdot+}^n, \theta_0))^{-1} \cdot (U_j^n(\theta_0 | X_{j\cdot}^n) - U_j^n(\theta_0 | \mu_{j\cdot}^n)) \Big).$$

Using this representation, application of the Cauchy–Schwarz Inequality splits the expectation in (6.2) into three parts:

$$\begin{aligned}
& E\Big(\big(\Psi_{\lambda j}^n(X_{j\cdot}^n) - E(\Psi_{\lambda j}^n(X_{j\cdot}^n)) \big)^4 \Big) \\
\leq\ & 8 \cdot E\Big(\Big(\sum_{k=1}^{K} a_\lambda(X_{jk}^n, \mu_{jk}^n) - a_1(X_{j+}^n, \mu_{j+}^n) - \Big(\sum_{k=1}^{K} E(a_\lambda(X_{jk}^n, \mu_{jk}^n)) - 1 \Big) \Big)^4 \Big) \\
+\ & 8 \cdot E\Big(\Big(c_\lambda^n(\mu_{\cdot+}^n, \theta_0) \cdot (I^n(\mu_{\cdot+}^n, \theta_0))^{-1} \cdot (U_j^n(\theta_0 | X_{j\cdot}^n) - U_j^n(\theta_0 | \mu_{j\cdot}^n)) \Big)^4 \Big) \\
\leq\ & 8 \cdot E\Big(\Big(\sum_{k=1}^{K} a_\lambda(X_{jk}^n, \mu_{jk}^n) - a_1(X_{j+}^n, \mu_{j+}^n) - \Big(\sum_{k=1}^{K} E(a_\lambda(X_{jk}^n, \mu_{jk}^n)) - 1 \Big) \Big)^4 \Big) \\
+\ & 8 \cdot \| c_\lambda^n(\mu_{\cdot+}^n, \theta_0) \cdot (I^n(\mu_{\cdot+}^n, \theta_0))^{-1} \|^4 \cdot E\Big(\| U_j^n(\theta_0 | X_{j\cdot}^n) - U_j^n(\theta_0 | \mu_{j\cdot}^n)) \|^4 \Big).
\end{aligned} \tag{6.7}$$

The single terms can in the following be studied separately. For the first term holds

$$\begin{aligned}
& E\Big(\Big(\sum_{k=1}^{K} a_\lambda(X_{jk}^n, \mu_{jk}^n) - a_1(X_{j+}^n, \mu_{j+}^n) - \Big(\sum_{k=1}^{K} E(a_\lambda(X_{jk}^n, \mu_{jk}^n)) - 1 \Big) \Big)^4 \Big) \\
\leq\ & E\Big(\Big(\sum_{k=1}^{K} a_\lambda(X_{jk}^n, \mu_{jk}^n) + 1 \Big)^4 \Big) \\
\leq\ & E\Big(8 \Big(\sum_{k=1}^{K} a_\lambda(X_{jk}^n, \mu_{jk}^n) \Big)^4 + 8 \Big) \\
\leq\ & 8K^3 \sum_{k=1}^{K} E(a_\lambda^4(X_{jk}^n, \mu_{jk}^n)) + 8 \\
\leq\ & 8K^3 \sum_{k=1}^{K} c + 8 \\
=\ & 8K^4 c + 8
\end{aligned}$$

for some constant $c \in \mathbf{R}^+$. The general inequality (5.11), i.e. $(\sum_{k=1}^{K} x_k)^4 \leq \sum_{k=1}^{K} K^3 \cdot x_k^4$, and the boundedness of the fourth moments of a_λ was used here, the latter holding due to Lemma 4.8 b) since $\mu_{jk}^n \geq \epsilon$ for all j, k, n (BC). Hence the first term of (6.7) is bounded:

$$E\Big(\Big(\sum_{k=1}^{K} a_\lambda(X_{jk}^n, \mu_{jk}^n) - a_1(X_{j+}^n, \mu_{j+}^n) - \Big(\sum_{k=1}^{K} E(a_\lambda(X_{jk}^n, \mu_{jk}^n)) - 1 \Big) \Big)^4 \Big) = O(1). \tag{6.8}$$

For the second term, condition (LC1), which immediately yields $(I^n(\mu^n_{\cdot+}, \theta_0))^{-1} = O(\frac{1}{n})$ and $c^n_\lambda(\mu^n_{\cdot+}, \theta_0) = O(J^n)$, which has been shown in Lemma 5.6 using (BC), is needed. This, together with $\mu^n_{++}/n = O(1)$, gives

$$\|c^n_\lambda(\mu^n_{\cdot+}, \theta_0) \cdot (I^n(\mu^n_{\cdot+}, \theta_0))^{-1}\|^4 = O\Big((\frac{J^n}{n})^4\Big) = O\Big((\frac{J^n}{\mu^n_{++}})^4\Big). \tag{6.9}$$

Known arguments and especially condition (RC2) and (RC3) finally yield for the last expectation in (6.7):

$$\begin{aligned}
& E\Big(\|U^n_j(\theta_0|X^n_{j\cdot}) - U^n_j(\theta_0|\mu^n_{j\cdot})\|^4\Big) \\
= \;& E\Big(\|\sum_{k=1}^{K}(X^n_{jk} - \mu^n_{jk})D^T_\theta \log \pi^n_{jk|C}(\theta_0)\|^4\Big) \\
\leq \;& E\Big(\Big(\sum_{k=1}^{K}\|(X^n_{jk} - \mu^n_{jk})D^T_\theta \log \pi^n_{jk|C}(\theta_0)\|\Big)^4\Big) \\
= \;& O(1) \cdot E\Big(\sum_{k=1}^{K}(\mu^n_{jk})^2 \cdot \Big|\frac{X^n_{jk} - \mu^n_{jk}}{\sqrt{\mu^n_{jk}}}\Big|^4\Big) \\
= \;& O\Big(\sum_{k=1}^{K}(\mu^n_{jk})^2\Big).
\end{aligned} \tag{6.10}$$

The statements (6.8), (6.9) and (6.10) concerning the asymptotic order of the terms in (6.7) thus show $E\Big(\Big(\Psi^n_{\lambda j}(X^n_{j\cdot}) - E(\Psi^n_{\lambda j}(X^n_{j\cdot}))\Big)^4\Big) = O(1) \cdot \Big(1 + (\frac{J^n}{\mu^n_{++}})^4 \cdot \sum_{k=1}^{K}(\mu^n_{jk})^2\Big)$ and combined with $\frac{(J^n)^2 \cdot \sum_{j=1}^{J^n}\sum_{k=1}^{K}(\mu^n_{jk})^2}{(\mu^n_{++})^4} \longrightarrow 0$ for the sum in (6.2) follows

$$\begin{aligned}
& \sum_{j=1}^{J^n} E\Big(\Big(\Psi^n_{\lambda j}(X^n_{j\cdot}) - E(\Psi^n_{\lambda j}(X^n_{j\cdot}))\Big)^4\Big) \\
& = O(J^n) + (\frac{J^n}{\mu^n_{++}})^4 \cdot \sum_{j=1}^{J^n}\sum_{k=1}^{K}(\mu^n_{jk})^2 \cdot O(1) \\
& = O((J^n)^2) \cdot \Big(O(\frac{1}{J^n}) + \frac{(J^n)^2 \cdot \sum_{j=1}^{J^n}\sum_{k=1}^{K}(\mu^n_{jk})^2}{(\mu^n_{++})^4}\Big) \\
& = O((J^n)^2) \cdot o(1) \\
& = o((J^n)^2).
\end{aligned}$$

This establishes the result.

Let now the additionally stated formula for the variance $\sigma^{n2}_\lambda(\mu^n_{\cdot+}, \theta_0)$ of the approximated statistic be verified, which is calculated as follows:

$$
\begin{aligned}
& Var(\Psi^n_{\lambda+}(X^n)) \\
= \; & Var\Big(SD^n_\lambda(\mu^n_{\cdot+},\theta_0|X^n) - \sum_{j=1}^{J^n} a_1(X^n_{j+},\mu^n_{j+}) - c^n_\lambda(\mu^n_{\cdot+},\theta_0)(I^n(\mu^n_{\cdot+},\theta_0))^{-1}U^n(\theta_0|X^n)\Big) \\
= \; & Var\Big(SD^n_\lambda(\mu^n_{\cdot+},\theta_0|X^n) - \sum_{j=1}^{J^n} a_1(X^n_{j+},\mu^n_{j+})\Big) \\
+ \; & c^n_\lambda(\mu^n_{\cdot+},\theta_0)(I^n(\mu^n_{\cdot+},\theta_0))^{-1}\cdot Cov(U^n(\theta_0|X^n))\cdot\Big(c^n_\lambda(\mu^n_{\cdot+},\theta_0)(I^n(\mu^n_{\cdot+},\theta_0))^{-1}\Big)^T \\
- \; & 2Cov\Big(SD^n_\lambda(\mu^n_{\cdot+},\theta_0|X^n), c^n_\lambda(\mu^n_{\cdot+},\theta_0)(I^n(\mu^n_{\cdot+},\theta_0))^{-1}U^n(\theta_0|X^n)\Big) \\
+ \; & 2Cov\Big(\sum_{j=1}^{J^n} a_1(X^n_{j+},\mu^n_{j+}), c^n_\lambda(\mu^n_{\cdot+},\theta_0)(I^n(\mu^n_{\cdot+},\theta_0))^{-1}U^n(\theta_0|X^n)\Big) \\
= \; & Var\Big(SD^n_\lambda(\mu^n_{\cdot+},\theta_0|X^n) - \sum_{j=1}^{J^n} a_1(X^n_{j+},\mu^n_{j+})\Big) \\
+ \; & c^n_\lambda(\mu^n_{\cdot+},\theta_0)(I^n(\mu^n_{\cdot+},\theta_0))^{-1}(c^n_\lambda(\mu^n_{\cdot+},\theta_0))^T \\
- \; & 2c^n_\lambda(\mu^n_{\cdot+},\theta_0)(I^n(\mu^n_{\cdot+},\theta_0))^{-1}Cov\Big(SD^n_\lambda(\mu^n_{\cdot+},\theta_0|X^n), U^n(\theta_0|X^n)\Big) \\
+ \; & 2c^n_\lambda(\mu^n_{\cdot+},\theta_0)(I^n(\mu^n_{\cdot+},\theta_0))^{-1}Cov\Big(\sum_{j=1}^{J^n} a_1(X^n_{j+},\mu^n_{j+}), U^n(\theta_0|X^n)\Big) \\
= \; & Var\Big(SD^n_\lambda(\mu^n_{\cdot+},\theta_0|X^n) - \sum_{j=1}^{J^n} a_1(X^n_{j+},\mu^n_{j+})\Big) \\
- \; & c^n_\lambda(\mu^n_{\cdot+},\theta_0)(I^n(\mu^n_{\cdot+},\theta_0))^{-1}(c^n_\lambda(\mu^n_{\cdot+},\theta_0))^T \qquad (6.11) \\
= \; & Var\Big(SD^n_\lambda(\mu^n_{\cdot+},\theta_0|X^n)\Big) + Var\Big(\sum_{j=1}^{J^n} a_1(X^n_{j+},\mu^n_{j+})\Big) \\
- \; & 2Cov\Big(SD^n_\lambda(\mu^n_{\cdot+},\theta_0|X^n), \sum_{j=1}^{J^n} a_1(X^n_{j+},\mu^n_{j+})\Big) \\
- \; & c^n_\lambda(\mu^n_{\cdot+},\theta_0)(I^n(\mu^n_{\cdot+},\theta_0))^{-1}(c^n_\lambda(\mu^n_{\cdot+},\theta_0))^T \\
= \; & Var\Big(SD^n_\lambda(\mu^n_{\cdot+},\theta_0|X^n)\Big) + \sum_{j=1}^{J^n} Var\Big(\frac{(X^n_{j+}-\mu^n_{j+})^2}{\mu^n_{j+}}\Big) \\
- \; & 2\sum_{j=1}^{J^n} Cov\Big(\sum_{k=1}^{K} a_\lambda(X^n_{jk},\mu^n_{jk}), a_1(X^n_{j+},\mu^n_{j+})\Big) \\
- \; & c^n_\lambda(\mu^n_{\cdot+},\theta_0)(I^n(\mu^n_{\cdot+},\theta_0))^{-1}(c^n_\lambda(\mu^n_{\cdot+},\theta_0))^T
\end{aligned}
$$

$$
\begin{aligned}
&= Var\Big(SD^n_\lambda(\mu^n_{\cdot+},\theta_0|X^n)\Big) + \sum_{j=1}^{J^n}\Big(2+\frac{1}{\mu^n_{j+}}\Big) \\
&- 2\sum_{j=1}^{J^n}\sum_{k=1}^{K} Cov\Big(a_\lambda(X^n_{jk},\mu^n_{jk}), \frac{(X^n_{j+}-\mu^n_{j+})^2}{\mu^n_{j+}}\Big) \\
&- c^n_\lambda(\mu^n_{\cdot+},\theta_0)(I^n(\mu^n_{\cdot+},\theta_0))^{-1}(c^n_\lambda(\mu^n_{\cdot+},\theta_0))^T \\
&= Var\Big(SD^n_\lambda(\mu^n_{\cdot+},\theta_0|X^n)\Big) + 2J^n + \sum_{j=1}^{J^n}\frac{1}{\mu^n_{j+}} \\
&- 2\sum_{j=1}^{J^n}\sum_{k=1}^{K}\Big(\frac{1}{\mu^n_{j+}}Cov\Big(a_\lambda(X^n_{jk},\mu^n_{jk}),(X^n_{jk})^2\Big) - 2\pi^n_{jk|C}(\theta_0)Cov\Big(a_\lambda(X^n_{jk},\mu^n_{jk}),X^n_{jk}\Big)\Big) \\
&- c^n_\lambda(\mu^n_{\cdot+},\theta_0)(I^n(\mu^n_{\cdot+},\theta_0))^{-1}(c^n_\lambda(\mu^n_{\cdot+},\theta_0))^T \qquad (6.12) \\
&= Var\Big(\sum_{j=1}^{J^n}\sum_{k=1}^{K} a^n_\lambda(X^n_{jk},\mu^n_{jk})\Big) + 2J^n + \sum_{j=1}^{J^n}\frac{1}{\mu^n_{j+}} \\
&- 2\sum_{j=1}^{J^n}\frac{1}{\mu^n_{j+}}\sum_{k=1}^{K} Cov\Big(a_\lambda(X^n_{jk},\mu^n_{jk}),(X^n_{jk})^2\Big) \\
&+ 4\sum_{j=1}^{J^n}\sum_{k=1}^{K}\pi^n_{jk|C}(\theta_0)Cov\Big(a_\lambda(X^n_{jk},\mu^n_{jk}),X^n_{jk}\Big) \\
&- c^n_\lambda(\mu^n_{\cdot+},\theta_0)\cdot(I^n(\mu^n_{\cdot+},\theta_0))^{-1}(c^n_\lambda(\mu^n_{\cdot+},\theta_0))^T.
\end{aligned}
$$

Here (6.11) and (6.12) are obtained through simple calculations, which give

$$Cov\Big(\sum_{j=1}^{J^n} a_1(X^n_{j+},\mu^n_{j+}), U^n(\theta_0|X^n)\Big) = 0$$

and

$$
\begin{aligned}
Cov\Big(a_\lambda(X^n_{jk},\mu^n_{jk}),\frac{(X^n_{j+}-\mu^n_{j+})^2}{\mu^n_{j+}}\Big) &= \frac{1}{\mu^n_{j+}}Cov\Big(a_\lambda(X^n_{jk},\mu^n_{jk}),(X^n_{jk})^2\Big) \\
&\quad -2\pi^n_{jk|C}(\theta_0)Cov\Big(a_\lambda(X^n_{jk},\mu^n_{jk}),X^n_{jk}\Big).
\end{aligned}
$$

Hence the proof of this theorem is complete. □

In the following theorem, the limiting normality of the approximation under column–multinomial distribution will be shown using Morris' method. As already mentioned in the beginning, the major part of the proof will treat the required convergence of the conditional distribution, since the validity of the Lindeberg Condition for the transformed Poisson approximation could easily be derived together with the pure approximation in the preceding theorem.

Theorem 6.2 *Consider the asymptotics* $n \to \infty$ *and let* $\Psi^n_{\lambda+} = \sum_{j=1}^{J^n} \Psi^n_{\lambda j}$ *be defined as in Theorem 6.1, i.e.*

$$\Psi^n_{\lambda j}(z_{j\cdot}) = \sum_{k=1}^{K} a_\lambda(z_{jk}, \mu^n_{jk}) - a_1(z_{j+}, \mu^n_{j+}) - c^n_\lambda(\mu^n_{\cdot+}, \theta_0)(I^n(\mu^n_{\cdot+}, \theta_0))^{-1} U^n_j(z_{j\cdot}, \theta_0)$$

($z = (z_{jk})_{j,k}$ *is an arbitrary* $J^n \times K$ *table with nonnegative entries). Just as in Theorem 6.1, let* $\sigma^{n2}_\lambda(\mu^n_{\cdot+}, \theta_0)$ *denote the variance of the approximation under Poisson distribution,* $\sigma^{n2}_\lambda(\mu^n_{\cdot+}, \theta_0) = Var(\Psi^n_{\lambda+}(X^n))$*, and* $s^{n2}_\lambda(\mu^n_{\cdot+}, \theta_0)$ *the corrected variance:*

$$s^{n2}_\lambda(\mu^n_{\cdot+}, \theta_0) \quad = \quad \sigma^{n2}_\lambda(\mu^n_{\cdot+}, \theta_0) - \sum_{k=1}^{K} \mu^n_{+k}(\gamma^n_{\lambda k})^2$$

with

$$\gamma^n_{\lambda k} \quad = \quad \frac{1}{\mu^n_{+k}} \sum_{j=1}^{J^n} \Big(Cov(a_\lambda(X^n_{jk}, \mu^n_{jk}), X^n_{jk}) - \pi^n_{jk|C}(\theta_0) \Big).$$

If the assumptions $\frac{1}{n} I^n(\mu^n_{\cdot+}, \theta_0) \longrightarrow I_\infty$ *positive definite (LC1),* $\mu^n_{jk} \geq \epsilon > 0$ *for all* j, k, n *(BC),* $\max_{1 \leq j \leq J^n} \pi^n_{jk|D}(\theta_0) \longrightarrow 0$ *for all* k *(MD3) and* $\frac{J^n}{s^{n2}_\lambda(\mu^n_{\cdot+}, \theta_0)} = O(1)$ *(VCC) hold, then under column–multinomial sampling the approximated goodness–of–fit statistic is asymptotically normal as follows:*

$$\frac{\Psi^n_{\lambda+}(Y^n) - m^n_\lambda(\mu^n_{\cdot+}, \theta_0) + J^n}{s^n_\lambda(\mu^n_{\cdot+}, \theta_0)} \xrightarrow{\mathcal{L}} N(0,1) \quad \text{for } \lambda \in (-1, 1].$$

Proof:
In order to prove this theorem using Lemma 3.4, which is based on Morris (1975), and applying the construction method outlined at the beginning of this section resp. explained in chapter 3, sec. 3.1, let for $j \in \{1, \ldots, J^n\}$, $n \in \mathbf{N}$, functions $f^n_{\lambda j} : \mathbf{R}^K \to \mathbf{R}$ be defined as follows (z is a $J^n \times K$ table):

$$f^n_{\lambda j}(z_{j\cdot}) \quad := \quad \Psi^n_{\lambda j}(z_{j\cdot}) - E(\Psi^n_{\lambda j}(X^n_{j\cdot})) - \sum_{k=1}^{K} \gamma^n_{\lambda k}(z_{jk} - \mu^n_{jk}).$$

By construction holds $Cov(\sum_{j=1}^{J^n} f^n_{\lambda j}(X^n_{j\cdot}), \sum_{j=1}^{J^n} X^n_{jk}) = \sum_{j=1}^{J^n} Cov(f^n_{\lambda j}(X^n_{j\cdot}), X^n_{jk}) = 0$ and $E(f^n_{\lambda j}(X^n_{j\cdot})) = 0$ for all j, k. With such chosen $f^n_{\lambda j}$, the situation of Lemma 3.4 is met ($f^n_{\lambda j}$ corresponds with f^n_j from Lemma 3.4). Moreover, except for the probability vectors, which here are called $\pi^n_{\cdot k|D}$, the notation is the same. Finally, condition (3.9) coincides with (MD3) and hence holds already by assumption.
If now the remaining conditions (3.10) – (3.12) for $\lambda \in (-1, 1]$ hold, then because of

$$\sum_{j=1}^{J^n} f^n_{\lambda j}(Y^n_{j\cdot}) \quad = \quad \sum_{j=1}^{J^n} \Psi^n_{\lambda j}(Y^n_{j\cdot}) - \sum_{j=1}^{J^n} E(\Psi^n_{\lambda j}(X^n_{j\cdot}))$$

$$
\begin{aligned}
&= \Psi^n_{\lambda+}(Y^n) - E(\Psi^n_{\lambda+}(X^n)) \\
&= \Psi^n_{\lambda+}(Y^n) - m^n_\lambda(\mu^n_{\cdot+}, \theta_0) + J^n
\end{aligned}
$$

(cp. p. 97) and

$$
\begin{aligned}
&\sum_{j=1}^{J^n} Var(f^n_{\lambda j}(X^n_{j\cdot})) \\
&= Var\Big(\sum_{j=1}^{J^n} \Psi^n_{\lambda j}(X^n_{j\cdot}) - \sum_{j=1}^{J^n}\sum_{k=1}^{K} \gamma^n_{\lambda k}(X^n_{jk} - \mu^n_{jk})\Big) \\
&= Var\Big(\sum_{j=1}^{J^n} \Psi^n_{\lambda j}(X^n_{j\cdot})\Big) - \sum_{k=1}^{K} (\gamma^n_{\lambda k})^2 \mu^n_{+k} \\
&= \sigma^{n2}_\lambda(\mu^n_{\cdot+}, \theta_0) - \sum_{k=1}^{K} (\gamma^n_{\lambda k})^2 \mu^n_{+k}
\end{aligned}
$$

(cp. sec. 3.1, (3.7)) follows the result:

$$
\frac{1}{(\sum_{j=1}^{J^n} Var(f^n_{\lambda j}(X^n_{j\cdot})))^{1/2}} \sum_{j=1}^{J^n} f^n_{\lambda j}(Y^n_{j\cdot}) = \frac{\Psi^n_{\lambda+}(Y^n) - m^n_\lambda(\mu^n_{\cdot+}, \theta_0) + J^n}{s^n_\lambda(\mu^n_{\cdot+}, \theta_0)} \xrightarrow{\mathcal{L}} N(0,1)
$$

for $\lambda \in (-1, 1]$. To check the missing conditions (3.10) to (3.12), it suffices to verify (3.12), because the validity of the Ljapounov Condition for the statistic

$$
\begin{aligned}
\frac{\sum_{j=1}^{J^n} f^n_{\lambda j}(X^n_{j\cdot})}{s^n_\lambda(\mu^n_{\cdot+}, \theta_0)} &= \frac{\sum_{j=1}^{J^n} \Big(\Psi^n_{\lambda j}(X^n_{j\cdot}) - E(\Psi^n_{\lambda j}(X^n_{j\cdot})) - \sum_{k=1}^{K} \gamma^n_{\lambda k}(X^n_{jk} - \mu^n_{jk})\Big)}{s^n_\lambda(\mu^n_{\cdot+}, \theta_0)} \\
&= \frac{\sum_{j=1}^{J^n} \Psi^n_{\lambda j}(X^n_{j\cdot}) - m^n_\lambda(\mu^n_{\cdot+}, \theta_0) + J^n}{s^n_\lambda(\mu^n_{\cdot+}, \theta_0)},
\end{aligned}
$$

which implies the Feller Condition and the asymptotic normality (3.10) and (3.11), has already been shown in Theorem (6.1) for arbitrary $\lambda > -1$. The condition concerning the marginal distribution required for Theorem (6.1), is here fulfilled because of the stronger condition (MD3):

$$
(J^n)^2 \sum_{j=1}^{J^n}\sum_{k=1}^{K} \frac{(\mu^n_{jk})^2}{(\mu^n_{++})^4} = \frac{(J^n)^2}{n^2} \sum_{k=1}^{K} \frac{n_k^2}{n^2} \sum_{j=1}^{J^n} (\pi^n_{jk|D}(\theta_0))^2 \le \frac{(J^n)^2}{n^2} \sum_{k=1}^{K} \max_j \pi^n_{jk|D}(\theta_0) = o(1).
$$

Now it remains to establish (3.12) for $\lambda \in (-1, 1]$:

$$
\lim_{h\to 0} \sup_n \sup_{v^n} \frac{1}{s^{n2}_\lambda(\mu^n_{\cdot+}, \theta_0)} E\Big(\Big(\sum_{j=1}^{J^n} f^n_{\lambda j}(L^n_{j\cdot} + M^n_{j\cdot}) - f^n_{\lambda j}(L^n_{j\cdot})\Big)^2\Big) = 0 \qquad (6.13)
$$

with $v^n = (v^n_1, \ldots, v^n_K)^T$ being a bounded sequence and L^n, M^n column–multinomial distributed $J^n \times K$ contingency tables, for which holds $L^n_{\cdot k} = (L^n_{1k}, \ldots, L^n_{J^n k})^T \sim$

$Multi_{J^n}(n_k + v_k^n\sqrt{n_k}, \pi^n_{\cdot k|D})$ and $M^n_{\cdot k} \sim Multi_{J^n}(h_k\sqrt{n_k}, \pi^n_{\cdot k|D})$ $(h = (h_1, \ldots, h_K)^T \in \mathbf{R}^K,\ k = 1, \ldots, K)$. All columns $L^n_{\cdot 1}, \ldots, L^n_{\cdot K}, M^n_{\cdot 1}, \ldots, M^n_{\cdot K}$ are supposed to be stochastically independent and the sizes are abbreviated by $l^n_k = n_k + v^n_k\sqrt{n_k}$ and $m^n_k = h_k\sqrt{n_k}$ for every $k \in \{1, \ldots, K\}$.

Since $\frac{J^n}{s_\lambda^{n2}(\mu^n_{\cdot +}, \theta_0)} = O(1)$ (VCC) is presumed, instead of proving (6.13), it suffices to show:

$$E\Big(\Big(\sum_{j=1}^{J^n} f^n_{\lambda j}(L^n_{j\cdot} + M^n_{j\cdot}) - f^n_{\lambda j}(L^n_{j\cdot})\Big)^2\Big) = \|h\|_{max} \cdot O(J^n) \quad (n \to \infty). \tag{6.14}$$

Before verifying (6.14), it should be mentioned that in the following proof it will be possible to adopt some results from the proof of Lemma 5.1 respectively from Morris (1975, Theorem 5.2). Lemma 5.1 has also been derived using Morris' method, i.e. Lemma 3.4, and hence considers the same situation, especially the same tables L^n and M^n.
Stating now the statistic in question explicitly and using the general inequality (5.10) yields

$$\begin{aligned}
& E\Big(\Big(\sum_{j=1}^{J^n} f^n_{\lambda j}(L^n_{j\cdot} + M^n_{j\cdot}) - f^n_{\lambda j}(L^n_{j\cdot})\Big)^2\Big) \\
= \; & E\Big(\Big(\sum_{j=1}^{J^n} \Big(\Psi^n_{\lambda j}(L^n_{j\cdot} + M^n_{j\cdot}) - \Psi^n_{\lambda j}(L^n_{j\cdot}) - \sum_{k=1}^{K} \gamma^n_{\lambda k} M^n_{jk}\Big)\Big)^2\Big) \\
= \; & E\Big(\Big(\sum_{j=1}^{J^n}\sum_{k=1}^{K}(a_\lambda(L^n_{jk} + M^n_{jk}, \mu^n_{jk}) - a_\lambda(L^n_{jk}, \mu^n_{jk})) \\
& \quad - \sum_{j=1}^{J^n}(a_1(L^n_{j+} + M^n_{j+}, \mu^n_{j+}) - a_1(L^n_{j+}, \mu^n_{j+})) \\
& \quad - \sum_{j=1}^{J^n} \Big(c^n_\lambda(\mu^n_{\cdot +}, \theta_0)(I^n(\mu^n_{\cdot +}, \theta_0))^{-1} U^n_j(L^n_{j\cdot} + M^n_{j\cdot}, \theta_0) \\
& \qquad\qquad - c^n_\lambda(\mu^n_{\cdot +}, \theta_0)(I^n(\mu^n_{\cdot +}, \theta_0))^{-1} U^n_j(L^n_{j\cdot}, \theta_0)\Big) \\
& \quad - \sum_{j=1}^{J^n}\sum_{k=1}^{K} \gamma^n_{\lambda k} M^n_{jk}\Big)^2\Big) \\
\leq \; & 4E\Big(\Big(\sum_{j=1}^{J^n}\sum_{k=1}^{K}(a_\lambda(L^n_{jk} + M^n_{jk}, \mu^n_{jk}) - a_\lambda(L^n_{jk}, \mu^n_{jk}))\Big)^2\Big) \\
& + 4E\Big(\Big(\sum_{j=1}^{J^n}(a_1(L^n_{j+} + M^n_{j+}, \mu^n_{j+}) - a_1(L^n_{j+}, \mu^n_{j+}))\Big)^2\Big)
\end{aligned}$$

$$+4E\Big(\Big(\sum_{j=1}^{J^n}\Big(c_\lambda^n(\mu_{\cdot+}^n,\theta_0)(I^n(\mu_{\cdot+}^n,\theta_0))^{-1}U_j^n(L_{j\cdot}^n+M_{j\cdot}^n,\theta_0)$$
$$-c_\lambda^n(\mu_{\cdot+}^n,\theta_0)(I^n(\mu_{\cdot+}^n,\theta_0))^{-1}U_j^n(L_{j\cdot}^n,\theta_0)\Big)\Big)^2\Big)$$
$$+4\Big(\sum_{k=1}^{K}\gamma_{\lambda k}^n m_k^n\Big)^2.$$

To prove (6.14) now for each of the four terms of the dominating sum, the order $||h||_{max}\cdot O(J^n)$ will be determined. Hence one has to show

$$E\Big(\Big(\sum_{j=1}^{J^n}\Big(c_\lambda^n(\mu_{\cdot+}^n,\theta_0)(I^n(\mu_{\cdot+}^n,\theta_0))^{-1}U_j^n(L_{j\cdot}^n+M_{j\cdot}^n,\theta_0)$$
$$-c_\lambda^n(\mu_{\cdot+}^n,\theta_0)(I^n(\mu_{\cdot+}^n,\theta_0))^{-1}U_j^n(L_{j\cdot}^n,\theta_0)\Big)\Big)^2\Big) = ||h||_{max}\cdot O(J^n), \quad (6.15)$$
$$\Big(\sum_{k=1}^{K}\gamma_{\lambda k}^n m_k^n\Big)^2 = ||h||_{max}\cdot O(J^n), \quad (6.16)$$
$$E\Big(\Big(\sum_{j=1}^{J^n}\sum_{k=1}^{K}(a_\lambda(L_{jk}^n+M_{jk}^n,\mu_{jk}^n)-a_\lambda(L_{jk}^n,\mu_{jk}^n))\Big)^2\Big) = ||h||_{max}\cdot O(J^n), \quad (6.17)$$
$$E\Big(\Big(\sum_{j=1}^{J^n}(a_1(L_{j+}^n+M_{j+}^n,\mu_{j+}^n)-a_1(L_{j+}^n,\mu_{j+}^n))\Big)^2\Big) = ||h||_{max}\cdot O(J^n), \quad (6.18)$$

where (6.15) and (6.16) are briefly verified and will be given prior to the proof of (6.17), which requires lengthy calculations.
For the proof of (6.15) now $c_\lambda^n(\mu_{\cdot+}^n,\theta_0)(I^n(\mu_{\cdot+}^n,\theta_0))^{-1}=O(\frac{J^n}{n})$, which follows from (BC), (LC1) and (LC3) (cp. proof of Theorem 6.1, (6.9)), the generally assumed regularity conditions (RC2), (RC3) and $\frac{J^n}{n}=O(1)$ are needed to obtain the desired order:

$$E\Big(\Big(\sum_{j=1}^{J^n}\Big(c_\lambda^n(\mu_{\cdot+}^n,\theta_0)(I^n(\mu_{\cdot+}^n,\theta_0))^{-1}U_j^n(L_{j\cdot}^n+M_{j\cdot}^n,\theta_0)$$
$$-c_\lambda^n(\mu_{\cdot+}^n,\theta_0)(I^n(\mu_{\cdot+}^n,\theta_0))^{-1}U_j^n(L_{j\cdot}^n,\theta_0)\Big)\Big)^2\Big)$$
$$= E\Big(\Big(c_\lambda^n(\mu_{\cdot+}^n,\theta_0)(I^n(\mu_{\cdot+}^n,\theta_0))^{-1}\sum_{j=1}^{J^n}\Big(U_j^n(L_{j\cdot}^n+M_{j\cdot}^n,\theta_0)-U_j^n(L_{j\cdot}^n,\theta_0)\Big)\Big)^2\Big)$$
$$= E\Big(\Big(c_\lambda^n(\mu_{\cdot+}^n,\theta_0)(I^n(\mu_{\cdot+}^n,\theta_0))^{-1}\sum_{j=1}^{J^n}\sum_{k=1}^{K}M_{jk}^n D_\theta^T\log\pi_{jk|C}^n(\theta_0)\Big)^2\Big)$$

$$\begin{aligned}
&\leq K\sum_{k=1}^{K} E\Big(\Big(c_\lambda^n(\mu_{\cdot+}^n,\theta_0)(I^n(\mu_{\cdot+}^n,\theta_0))^{-1}\sum_{j=1}^{J^n} M_{jk}^n D_\theta^T \log\pi_{jk|C}^n(\theta_0)\Big)^2\Big)\\
&\leq K\sum_{k=1}^{K} \|c_\lambda^n(\mu_{\cdot+}^n,\theta_0)(I^n(\mu_{\cdot+}^n,\theta_0))^{-1}\|^2\cdot E\Big(\|\sum_{j=1}^{J^n} M_{jk}^n D_\theta^T \log\pi_{jk|C}^n(\theta_0)\|^2\Big)\\
&= K\sum_{k=1}^{K} O\Big((\frac{J^n}{n})^2\Big)\cdot O(1)\cdot E\Big(\Big(\sum_{j=1}^{J^n} M_{jk}^n\Big)^2\Big)\\
&= K\sum_{k=1}^{K} O\Big((\frac{J^n}{n})^2\Big)\cdot (m_k^n)^2\\
&= K\sum_{k=1}^{K} O(\frac{(J^n)^2}{n^2})\cdot h_k^2 n_k\\
&= \|h\|_{max}^2\cdot O(\frac{(J^n)^2}{n})\\
&= \|h\|_{max}\cdot O(J^n).
\end{aligned}$$

Using $\mu_{+k}^n\gamma_{\lambda k}^n = n_k\gamma_{\lambda k}^n = O(J^n)$, which follows from (BC) and has already been shown in the proof of Theorem 6.1 (6.1), now for (6.16), similar to the proof of Lemma 5.1, follows:

$$\begin{aligned}
\Big(\sum_{k=1}^{K}\gamma_{\lambda k}^n m_k^n\Big)^2 &\leq K\sum_{k=1}^{K}(\gamma_{\lambda k}^n m_k^n)^2\\
&= K\sum_{k=1}^{K} O\Big((\frac{J^n}{n_k})^2\Big)h_k^2 n_k\\
&= O(1)\cdot\|h\|_{max}^2\cdot J^n\cdot\sum_{k=1}^{K}\frac{J^n}{n_k}\\
&= \|h\|_{max}\cdot O(J^n).
\end{aligned}$$

For (6.17) one now has to check

$$E\Big(\Big(\sum_{j=1}^{J^n}\sum_{k=1}^{K}(a_\lambda(L_{jk}^n+M_{jk}^n,\mu_{jk}^n)-a_\lambda(L_{jk}^n,\mu_{jk}^n))\Big)^2\Big) = \|h\|_{max}\cdot O(J^n).$$

Since change of the summation signs and inequality (5.10) give

$$\begin{aligned}
&E\Big(\Big(\sum_{k=1}^{K}\sum_{j=1}^{J^n}(a_\lambda(L_{jk}^n+M_{jk}^n,\mu_{jk}^n)-a_\lambda(L_{jk}^n,\mu_{jk}^n))\Big)^2\Big)\\
&\leq K\cdot\sum_{k=1}^{K} E\Big(\Big(\sum_{j=1}^{J^n}(a_\lambda(L_{jk}^n+M_{jk}^n,\mu_{jk}^n)-a_\lambda(L_{jk}^n,\mu_{jk}^n))\Big)^2\Big),
\end{aligned}$$

it suffices to show for each $k \in \{1, \ldots, K\}$:

$$E\Big(\Big(\sum_{j=1}^{J^n}(a_\lambda(L_{jk}^n + M_{jk}^n, \mu_{jk}^n) - a_\lambda(L_{jk}^n, \mu_{jk}^n))\Big)^2\Big) = h_k \cdot O(J^n). \tag{6.19}$$

For the proof of (6.19) (for $\lambda \in (1, 1]$), let in the following any $k \in \{1, \ldots, K\}$ be considered. Using the inequality given in Lemma 8.1 yields

$$\begin{aligned}
& E\Big(\Big(\sum_{j=1}^{J^n}\Big(a_\lambda(L_{jk}^n + M_{jk}^n, \mu_{jk}^n) - a_\lambda(L_{jk}^n, \mu_{jk}^n)\Big)\Big)^2\Big) \\
\leq\ & E\Big(\Big(\sum_{j=1}^{J^n}|a_\lambda(L_{jk}^n + M_{jk}^n, \mu_{jk}^n) - a_\lambda(L_{jk}^n, \mu_{jk}^n)|\Big)^2\Big) \\
\leq\ & 2E\Big(\Big(\sum_{j=1}^{J^n}\Big(\frac{(M_{jk}^n)^2}{\mu_{jk}^n} + M_{jk}^n \cdot h(L_{jk}^n, \mu_{jk}^n)\Big)\Big)^2\Big) \\
\leq\ & 4E\Big(\Big(\sum_{j=1}^{J^n}\frac{(M_{jk}^n)^2}{\mu_{jk}^n}\Big)^2\Big) + 4E\Big(\Big(\sum_{j=1}^{J^n} M_{jk}^n \cdot h(L_{jk}^n, \mu_{jk}^n)\Big)^2\Big)
\end{aligned} \tag{6.20}$$

with $h(L_{jk}^n, \mu_{jk}^n) = |L_{jk}^n - \mu_{jk}^n|(\frac{1}{\mu_{jk}^n} + \frac{c}{L_{jk}^n+1})$, $c = \max\{2, \frac{1}{\lambda+1}\}$. For the first expectation, the desired order is already given by the proof of Lemma 5.1, where the same variables are considered, and only (BC) is used to assert (5.16), namely

$$E\Big(\Big(\sum_{j=1}^{J^n}\frac{(M_{jk}^n)^2}{\mu_{jk}^n}\Big)^2\Big) = h_k \cdot O(J^n).$$

In (5.17) (Lemma 5.1) further

$$E\Big(\Big(\sum_{j=1}^{J^n} M_{jk}^n \cdot h(L_{jk}^n, \mu_{jk}^n)\Big)^2\Big) \leq \sum_{j=1}^{J^n} m_k^n E(M_{jk}^n) E(h^2(L_{jk}^n, \mu_{jk}^n))$$

has also already been shown, just using the independence of $L_{\cdot k}$ and $M_{\cdot k}$ and not the concrete definition of the function $h(L_{jk}^n, \mu_{jk}^n)$. Now provided that

$$E(h^2(L_{jk}^n, \mu_{jk}^n)) = O(\frac{1}{\mu_{jk}^n}) \text{ for all } j, k, n \tag{6.21}$$

holds, the second term of (6.20) will have the desired order as well, since then follows $(\pi_{jk|D}^n = \pi_{jk|D}^n(\theta_0))$

$$\sum_{j=1}^{J^n} m_k^n E(M_{jk}^n) E(h^2(L_{jk}^n, \mu_{jk}^n)) \;=\; \sum_{j=1}^{J^n}(m_k^n)^2 \pi_{jk|D}^n \cdot O(\frac{1}{\mu_{jk}^n})$$

$$
\begin{aligned}
&= \sum_{j=1}^{J^n} h_k^2 n_k \pi_{jk|D}^n \cdot O(\frac{1}{n_k \pi_{jk|D}^n}) \\
&= h_k \cdot O(J^n),
\end{aligned}
$$

thus establishing (6.19). Hence only (6.21) remains to be shown. Now it holds

$$
\begin{aligned}
& h^2(L_{jk}^n, \mu_{jk}^n) \\
=\ & \Big(|L_{jk}^n - \mu_{jk}^n|(\frac{1}{\mu_{jk}^n} + \frac{c}{L_{jk}^n + 1})\Big)^2 \\
\leq\ & 2\frac{(L_{jk}^n - \mu_{jk}^n)^2}{(\mu_{jk}^n)^2} + 2c^2 \frac{(L_{jk}^n - \mu_{jk}^n)^2}{(L_{jk}^n + 1)^2} \\
\leq\ & 2\frac{(L_{jk}^n - \mu_{jk}^n)^2}{(\mu_{jk}^n)^2} + 4c^2 \frac{(L_{jk}^n - \mu_{jk}^n)^2}{(L_{jk}^n + 1)(L_{jk}^n + 2)} \qquad (6.22) \\
=\ & 2\frac{(L_{jk}^n - \mu_{jk}^n)^2}{(\mu_{jk}^n)^2} + 4c^2 \Big(\frac{(L_{jk}^n - \mu_{jk}^n)^2 - (2 + \mu_{jk}^n)^2}{(L_{jk}^n + 1)(L_{jk}^n + 2)} + \frac{(2 + \mu_{jk}^n)^2}{(L_{jk}^n + 1)(L_{jk}^n + 2)}\Big) \\
=\ & 2\frac{(L_{jk}^n - \mu_{jk}^n)^2}{(\mu_{jk}^n)^2} \\
& +4c^2 \Big(\frac{(L_{jk}^n + 1)(L_{jk}^n + 2) - (3 + 2\mu_{jk}^n)(L_{jk}^n + 2)}{(L_{jk}^n + 1)(L_{jk}^n + 2)} + \frac{(2 + \mu_{jk}^n)^2}{(L_{jk}^n + 1)(L_{jk}^n + 2)}\Big) \\
=\ & 2\frac{(L_{jk}^n - \mu_{jk}^n)^2}{(\mu_{jk}^n)^2} + 4c^2 \Big(1 - \frac{3 + 2\mu_{jk}^n}{L_{jk}^n + 1} + \frac{(2 + \mu_{jk}^n)^2}{(L_{jk}^n + 1)(L_{jk}^n + 2)}\Big), \qquad (6.23)
\end{aligned}
$$

where in (6.22) $\frac{1}{L_{jk}^n + 1} \leq \frac{2}{L_{jk}^n + 2} \Leftrightarrow L_{jk}^n \geq 0$ has been used. Computing the expected value of the expression in (6.23) gives for the first term

$$
\begin{aligned}
& E\Big(\frac{(L_{jk}^n - \mu_{jk}^n)^2}{(\mu_{jk}^n)^2}\Big) \\
=\ & \frac{1}{(\mu_{jk}^n)^2} E\left((L_{jk}^n - n_k \pi_{jk|D}^n - v_k^n \sqrt{n_k} \pi_{jk|D}^n + v_k^n \sqrt{n_k} \pi_{jk|D}^n)^2\right) \\
=\ & \frac{1}{(\mu_{jk}^n)^2} E\left((L_{jk}^n - l_k^n \pi_{jk|D}^n)^2 + (v_k^n \sqrt{n_k} \pi_{jk|D}^n)^2 + 2(L_{jk}^n - l_k^n \pi_{jk|D}^n) v_k^n \sqrt{n_k} \pi_{jk|D}^n\right) \\
=\ & \frac{1}{(\mu_{jk}^n)^2} \left(l_k^n \pi_{jk|D}^n (1 - \pi_{jk|D}^n) + (v_k^n \sqrt{n_k} \pi_{jk|D}^n)^2\right). \qquad (6.24)
\end{aligned}
$$

Using the notation $x^{(j)} := \prod_{i=1}^{j} (x - i + 1)$ with $i, j \in \mathbf{N}$ and the formula for the inverse factorial moments of a binomial distributed random variable $x \sim B(n, p)$ like in Lemma 5.1,

$$E\Big(((x+i)^{(i)})^{-1}\Big) = \Big(((n+i)^{(i)}p^i)^{-1}\Big)\cdot\Big(1-\sum_{j=0}^{i-1}\binom{n+i}{j}p^j(1-p)^{n+i-j}\Big),$$

gives for the expected value of the second term of the sum

$$\begin{aligned} & E\Big(1-\frac{3+2\mu^n_{jk}}{L^n_{jk}+1}+\frac{(2+\mu^n_{jk})^2}{(L^n_{jk}+1)(L^n_{jk}+2)}\Big) \\ = \; & 1-(3+2\mu^n_{jk})\frac{1-(1-\pi^n_{jk|D})^{l^n_k+1}}{(l^n_k+1)\pi^n_{jk|D}} \\ & +(2+\mu^n_{jk})^2\frac{1-(1-\pi^n_{jk|D})^{l^n_k+2}-(l^n_k+2)\pi^n_{jk|D}(1-\pi^n_{jk|D})^{l^n_k+1}}{(l^n_k+2)(l^n_k+1)(\pi^n_{jk|D})^2}. \end{aligned} \tag{6.25}$$

Because of

$$3+2\mu^n_{jk}\le 4+4\mu^n_{jk}+(\mu^n_{jk})^2=(2+\mu^n_{jk})^2$$

and

$$\begin{aligned} 1 & \le \frac{1-\pi^n_{jk|D}}{(l^n_k+2)\pi^n_{jk|D}}+1=\frac{1-\pi^n_{jk|D}+(l^n_k+2)\pi^n_{jk|D}}{(l^n_k+2)\pi^n_{jk|D}} \\ \iff \quad \frac{(1-\pi^n_{jk|D})^{l^n_k+1}}{(l^n_k+1)\pi^n_{jk|D}} & \le \frac{(1-\pi^n_{jk|D})^{l^n_k+1}}{(l^n_k+1)\pi^n_{jk|D}}\cdot\frac{1-\pi^n_{jk|D}+(l^n_k+2)\pi^n_{jk|D}}{(l^n_k+2)\pi^n_{jk|D}} \\ & = \frac{(1-\pi^n_{jk|D})^{l^n_k+2}+(l^n_k+2)\pi^n_{jk|D}(1-\pi^n_{jk|D})^{l^n_k+1}}{(l^n_k+1)(l^n_k+2)(\pi^n_{jk|D})^2} \end{aligned}$$

holds

$$(3+2\mu^n_{jk})\frac{(1-\pi^n_{jk|D})^{l^n_k+1}}{(l^n_k+1)\pi^n_{jk|D}}\le(2+\mu^n_{jk})^2\frac{(1-\pi^n_{jk|D})^{l^n_k+2}+(l^n_k+2)\pi^n_{jk|D}(1-\pi^n_{jk|D})^{l^n_k+1}}{(l^n_k+1)(l^n_k+2)(\pi^n_{jk|D})^2}$$

thus giving

$$\begin{aligned} & -(3+2\mu^n_{jk})\frac{1-(1-\pi^n_{jk|D})^{l^n_k+1}}{(l^n_k+1)\pi^n_{jk|D}} \\ & +(2+\mu^n_{jk})^2\frac{1-(1-\pi^n_{jk|D})^{l^n_k+2}-(l^n_k+2)\pi^n_{jk|D}(1-\pi^n_{jk|D})^{l^n_k+1}}{(l^n_k+1)(l^n_k+2)(\pi^n_{jk|D})^2} \\ \le \; & -(3+2\mu^n_{jk})\frac{1}{(l^n_k+1)\pi^n_{jk|D}}+(2+\mu^n_{jk})^2\frac{1}{(l^n_k+1)(l^n_k+2)(\pi^n_{jk|D})^2}, \end{aligned}$$

i.e. (6.25) is majorized as follows:

$$\begin{aligned} & E\Big(1-\frac{3+2\mu^n_{jk}}{L^n_{jk}+1}+\frac{(2+\mu^n_{jk})^2}{(L^n_{jk}+1)(L^n_{jk}+2)}\Big) \\ \le\ & 1-\frac{3+2\mu^n_{jk}}{(l^n_k+1)\pi^n_{jk|D}}+\frac{(2+\mu^n_{jk})^2}{(l^n_k+1)(l^n_k+2)(\pi^n_{jk|D})^2}. \end{aligned} \tag{6.26}$$

For n_k sufficiently large holds

$$E\Big(1-\frac{3+2\mu^n_{jk}}{L^n_{jk}+1}+\frac{(2+\mu^n_{jk})^2}{(L^n_{jk}+1)(L^n_{jk}+2)}\Big)\le 1-\frac{3+2\mu^n_{jk}}{l^n_k\pi^n_{jk|D}}+\frac{(2+\mu^n_{jk})^2}{(l^n_k)^2(\pi^n_{jk|D})^2}, \tag{6.27}$$

since in this case for the right–hand side of (6.26) the following equivalences hold:

$$\begin{aligned} & -\frac{3+2\mu^n_{jk}}{(l^n_k+1)\pi^n_{jk|D}}+\frac{(2+\mu^n_{jk})^2}{(l^n_k+1)(l^n_k+2)(\pi^n_{jk|D})^2} \quad \le \quad -\frac{3+2\mu^n_{jk}}{l^n_k\pi^n_{jk|D}}+\frac{(2+\mu^n_{jk})^2}{(l^n_k)^2(\pi^n_{jk|D})^2} \\ \Longleftrightarrow\quad & (3+2\mu^n_{jk})\Big(\frac{1}{l^n_k\pi^n_{jk|D}}-\frac{1}{(l^n_k+1)\pi^n_{jk|D}}\Big) \\ \le\quad & (2+\mu^n_{jk})^2\cdot\Big(\frac{1}{(l^n_k)^2(\pi^n_{jk|D})^2}-\frac{1}{(l^n_k+1)(l^n_k+2)(\pi^n_{jk|D})^2}\Big) \\ \Longleftrightarrow\quad & (3+2\mu^n_{jk})\frac{1}{l^n_k(l^n_k+1)} \quad \le \quad (2+\mu^n_{jk})^2\frac{(l^n_k+1)(l^n_k+2)-(l^n_k)^2}{(l^n_k)^2(l^n_k+1)(l^n_k+2)\pi^n_{jk|D}} \\ \Longleftrightarrow\quad & (3+2n_k\pi^n_{jk|D})l^n_k(l^n_k+2)\pi^n_{jk|D} \quad \le \quad (2+n_k\pi^n_{jk|D})^2(3l^n_k+2) \\ \Longleftrightarrow\quad & 2(l^n_k)^2n_k(\pi^n_{jk|D})^2+3(l^n_k)^2\pi^n_{jk|D}+4l^n_kn_k(\pi^n_{jk|D})^2+6l^n_k\pi^n_{jk|D} \\ \le\quad & 3l^n_kn_k^2(\pi^n_{jk|D})^2+12l^n_kn_k\pi^n_{jk|D}+12l^n_k+2n_k^2(\pi^n_{jk|D})^2+8n_k\pi^n_{jk|D}+8 \end{aligned}$$

(e.g. holds $2(l^n_k)^2n_k(\pi^n_{jk|D})^2\le 3l^n_kn_k^2(\pi^n_{jk|D})^2 \Leftrightarrow 2(n_k+v^n_k\sqrt{n_k})\le 3n_k \Leftrightarrow 2v^n_k\le\sqrt{n_k}$). For n_k sufficiently large, statement (6.24) and (6.27) thus yield for the expectations of the terms of (6.23):

$$\begin{aligned} E\left(h^2(L^n_{jk},\mu^n_{jk})\right) \ \le\ & \frac{2}{(\mu^n_{jk})^2}\left(l^n_k\pi^n_{jk|D}(1-\pi^n_{jk|D})+(v^n_k\sqrt{n_k}\pi^n_{jk|D})^2\right) \\ & +4c^2\Big(1-\frac{3+2\mu^n_{jk}}{l^n_k\pi^n_{jk|D}}+\frac{4+4\mu^n_{jk}+(\mu^n_{jk})^2}{(l^n_k)^2(\pi^n_{jk|D})^2}\Big). \end{aligned} \tag{6.28}$$

Since the presumed boundedness of v^n_k gives $\frac{l^n_k}{n_k}=O(1)$ and $\frac{n_k}{l^n_k}=O(1)$, multiplication of (6.28) with μ^n_{jk} now yields for the terms of the right–hand side

$$\mu^n_{jk}\cdot\left(\frac{2}{(\mu^n_{jk})^2}\left(l^n_k\pi^n_{jk|D}(1-\pi^n_{jk|D})+(v^n_k\sqrt{n_k}\pi^n_{jk|D})^2\right)\right)$$

$$
\begin{aligned}
&= 2\Big(\frac{(n_k\pi^n_{jk|D} + v^n_k\sqrt{n_k}\pi^n_{jk|D})(1-\pi^n_{jk|D})}{n_k\pi^n_{jk|D}} + \frac{(v^n_k)^2 n_k(\pi^n_{jk|D})^2}{n_k\pi^n_{jk|D}}\Big) \\
&= O(1) \qquad (6.29)
\end{aligned}
$$

and

$$
\mu^n_{jk}\cdot 4c^2\Big(1 - \frac{3+2\mu^n_{jk}}{l^n_k\pi^n_{jk|D}} + \frac{4+4\mu^n_{jk}+(\mu^n_{jk})^2}{(l^n_k)^2(\pi^n_{jk|D})^2}\Big) = O(1), \qquad (6.30)
$$

where the last equality follows from

$$
\begin{aligned}
\mu^n_{jk}\cdot\Big(1 - \frac{2\mu^n_{jk}}{l^n_k\pi^n_{jk|D}} + \frac{(\mu^n_{jk})^2}{(l^n_k)^2(\pi^n_{jk|D})^2}\Big) &= n_k\pi^n_{jk|D}\cdot\Big(1 - \frac{2n_k}{l^n_k} + \frac{n_k^2}{(l^n_k)^2}\Big) \\
&= n_k\pi^n_{jk|D}\cdot\frac{(l^n_k - n_k)^2}{(l^n_k)^2} \\
&= \frac{n_k}{l^n_k}\cdot\pi^n_{jk|D}\cdot\frac{(v^n_k)^2 n_k}{l^n_k} \\
&= O(1).
\end{aligned}
$$

Hence statements (6.29) and (6.30) concerning the terms of (6.28) give

$$
\mu^n_{jk}\cdot E\Big(h^2(L^n_{jk},\mu^n_{jk})\Big) = O(1),
$$

which establishes (6.21).

The finally missing statement (6.18) can now for the major part be verified using the preceding results. First, just as in the proof of (6.19), holds (cp. (6.20))

$$
\begin{aligned}
&E\Big(\Big(\sum_{j=1}^{J^n}\Big(a_1(L^n_{j+}+M^n_{j+},\mu^n_{j+}) - a_1(L^n_{j+},\mu^n_{j+})\Big)\Big)^2\Big) \\
\le\ & E\Big(\Big(\sum_{j=1}^{J^n}|a_1(L^n_{j+}+M^n_{j+},\mu^n_{j+}) - a_1(L^n_{j+},\mu^n_{j+})|\Big)^2\Big) \\
\le\ & 2E\Big(\Big(\sum_{j=1}^{J^n}(\frac{(M^n_{j+})^2}{\mu^n_{j+}} + M^n_{j+}\cdot g(L^n_{j+},\mu^n_{j+}))\Big)^2\Big) \\
\le\ & 4E\Big(\Big(\sum_{j=1}^{J^n}\frac{(M^n_{j+})^2}{\mu^n_{j+}}\Big)^2\Big) + 4E\Big(\Big(\sum_{j=1}^{J^n}M^n_{j+}\cdot g(L^n_{j+},\mu^n_{j+})\Big)^2\Big) \qquad (6.31)
\end{aligned}
$$

with $g(L^n_{j+},\mu^n_{j+}) = |L^n_{j+} - \mu^n_{j+}|\cdot\frac{1}{\mu^n_{j+}}$, since in this case, $\lambda = 1$, the absolute value can be studied using the concrete definition:

$$
|a_1(L^n_{j+}+M^n_{j+},\mu^n_{j+}) - a_1(L^n_{j+},\mu^n_{j+})| = \Big|\frac{(L^n_{j+}+M^n_{j+}-\mu^n_{j+})^2}{\mu^n_{j+}} - \frac{(L^n_{j+}-\mu^n_{j+})^2}{\mu^n_{j+}}\Big|
$$

$$
\begin{aligned}
&= \Big|\frac{(M_{j+}^n)^2}{\mu_{j+}^n} - \frac{2M_{j+}^n(L_{j+}^n-\mu_{j+}^n)}{\mu_{j+}^n}\Big| \\
&\le 2\Big(\frac{(M_{j+}^n)^2}{\mu_{j+}^n} + M_{j+}^n\cdot\frac{|L_{j+}^n-\mu_{j+}^n|}{\mu_{j+}^n}\Big).
\end{aligned}
$$

For the first term of (6.31) one gets applying inequality (5.10) twice

$$
\begin{aligned}
E\Big(\big(\sum_{j=1}^{J^n}\frac{(M_{j+}^n)^2}{\mu_{j+}^n}\big)^2\Big) &= E\Big(\big(\sum_{j=1}^{J^n}\frac{1}{\mu_{j+}^n}(\sum_{k=1}^{K}M_{jk}^n)^2\big)^2\Big) \\
&\le E\Big(\big(\sum_{j=1}^{J^n}\frac{1}{\mu_{j+}^n}\cdot K\cdot\sum_{k=1}^{K}(M_{jk}^n)^2\big)^2\Big) \\
&\le K^3\cdot\sum_{k=1}^{K}E\Big(\big(\sum_{j=1}^{J^n}\frac{1}{\mu_{jk}^n}\cdot(M_{jk}^n)^2\big)^2\Big) \\
&= \|h\|_{max}\cdot O(J^n).
\end{aligned}
$$

The last equation sign holds since already in Lemma 5.1 (5.16) for each $k\in\{1,\ldots,K\}$ $E\Big((\sum_{j=1}^{J^n}\frac{1}{\mu_{jk}^n}\cdot(M_{jk}^n)^2)^2\Big) = h_k\cdot O(J^n)$ has been shown. The second term of the sum also has the desired order:

$$
\begin{aligned}
& E\Big(\big(\sum_{j=1}^{J^n}M_{j+}^n\cdot g(L_{j+}^n,\mu_{j+}^n)\big)^2\Big) \\
=\;& E\Big(\big(\sum_{k=1}^{K}\sum_{j=1}^{J^n}M_{jk}^n\cdot g(L_{j+}^n,\mu_{j+}^n)\big)^2\Big) \\
\le\;& K\cdot\sum_{k=1}^{K}E\Big(\big(\sum_{j=1}^{J^n}M_{jk}^n\cdot g(L_{j+}^n,\mu_{j+}^n)\big)^2\Big) \\
\le\;& K\cdot\sum_{k=1}^{K}\sum_{j=1}^{J^n}m_k^nE(M_{jk}^n)\cdot E(g^2(L_{j+}^n,\mu_{j+}^n)) \qquad (6.32) \\
=\;& K\cdot\sum_{k=1}^{K}\sum_{j=1}^{J^n}(m_k^n)^2\pi_{jk|D}^n\cdot E\Big(\big(\sum_{k'=1}^{K}\frac{L_{jk'}^n-\mu_{jk'}^n}{\mu_{j+}^n}\big)^2\Big) \\
\le\;& K\cdot\sum_{k=1}^{K}\sum_{j=1}^{J^n}h_k^2n_k\pi_{jk|D}^n\cdot K\cdot\sum_{k'=1}^{K}E\Big(\big(\frac{L_{jk'}^n-\mu_{jk'}^n}{\mu_{j+}^n}\big)^2\Big) \\
\le\;& \|h\|_{max}^2\cdot K^2\cdot\sum_{j=1}^{J^n}\sum_{k=1}^{K}\sum_{k'=1}^{K}\mu_{jk}^n\cdot E\Big(\big(\frac{L_{jk'}^n-\mu_{jk'}^n}{\mu_{jk'}^n}\big)^2\Big) \\
=\;& \|h\|_{max}^2\cdot K^2\cdot\sum_{j=1}^{J^n}\sum_{k=1}^{K}\sum_{k'=1}^{K}\mu_{jk}^n\cdot O\Big(\frac{1}{\mu_{jk'}^n}\Big) \qquad (6.33)
\end{aligned}
$$

$$= \quad \|h\|^2_{max} \cdot K^2 \cdot \sum_{j=1}^{J^n} \sum_{k=1}^{K} \sum_{k'=1}^{K} O(1) \tag{6.34}$$

$$= \quad \|h\|_{max} \cdot O(J^n).$$

Inequality (6.32) follows again from the proof of Lemma 5.1 (see (5.17)), where $E\Big((\sum_{j=1}^{J^n} M^n_{jk} \cdot h(L^n_{jk}, \mu^n_{jk}))^2\Big) \leq \sum_{j=1}^{J^n} m^n_k E(M^n_{jk}) E(h^2(L^n_{jk}, \mu^n_{jk}))$ was established using the presumed stochastic independence of the columns of L^n und M^n and not the specific shape of the function h. Statement (6.33) has already been proved as well (see (6.24) and (6.29)) and (6.34) holds since (RC2) assures $\pi^n_{jk|C} = \pi^n_{jk|C}(\theta_0)$ being bounded away from zero for every j,k,n thus giving $\frac{\mu^n_{jk}}{\mu^n_{jk'}} = \frac{\mu^n_{j+}\pi^n_{jk|C}}{\mu^n_{j+}\pi^n_{jk'|C}} = \frac{\pi^n_{jk|C}}{\pi^n_{jk'|C}} = O(1)$.

□

6.2 Asymptotic Normality of the Test Statistic

In the preceding section, for both distribution models the limiting normality of the approximated test statistic, scaled with the true unknown variance, has been derived; this section now will start with the proof of the consistency of the variance estimation. With the Poisson variance s^{n2}_λ of the transformation being just the variance σ^{n2}_λ of the Poisson approximation plus an additional correction term, and since the proof needs not distinguish between the sampling schemes, it suggests itself to show the consistency of the variance estimations $\sigma^{n2}_\lambda(\hat{\mu}^n_{\cdot+}, \hat{\theta}^n)$ and $s^{n2}_\lambda(\hat{\mu}^n_{\cdot+}, \hat{\theta}^n)$ together. The proof will require rather technical results such as those given in Lemma 5.2 and the appendix; comprehensive calculation will be necessary. Thus a major part of the bounding results summarized in Lemma 4.8 is solely needed for this proof.

Lemma 6.3 *For the asymptotics $n \to \infty$, let in both distribution models, column-multinomial and Poisson, the conditions*

$$\sum_{j=1}^{J^n} \frac{1}{\sqrt{\mu^n_{j+}}} = o(J^n) \quad \textit{and} \quad \sum_{j=1}^{J^n} \sqrt{\mu^n_{j+}} = o(\sqrt{J^n n}) \ \textit{(MD1)}$$

as well as (BC), (LC1) and (LC2) be fulfilled, i.e. $\mu^n_{jk} \geq \epsilon > 0$ for all j,k,n, $\frac{1}{n} I^n(\mu^n_{\cdot+}, \theta_0) \to I_\infty$ positive definite and $\sqrt{n}(\hat{\theta}^n - \theta_0) = O_p(1)$.
Consider now the variance $\sigma^{n2}_\lambda(\mu^n_{\cdot+}, \theta_0)$ of the Poisson approximation and the variance $s^{n2}_\lambda(\mu^n_{\cdot+}, \theta_0)$ of the corrected approximation, both given in Theorem 6.1. For these let be assumed $\frac{J^n}{\sigma^{n2}_\lambda(\mu^n_{\cdot+}, \theta_0)} = O(1)$ (VCP) and $\frac{J^n}{s^{n2}_\lambda(\mu^n_{\cdot+}, \theta_0)} = O(1)$ (VCC). Then the variance estimation is consistent:

$$\frac{\sigma^{n2}_\lambda(\mu^n_{\cdot+}, \theta_0)}{\sigma^{n2}_\lambda(\hat{\mu}^n_{\cdot+}, \hat{\theta}^n)} \xrightarrow{P} 1 \quad \textit{and} \quad \frac{s^{n2}_\lambda(\mu^n_{\cdot+}, \theta_0)}{s^{n2}_\lambda(\hat{\mu}^n_{\cdot+}, \hat{\theta}^n)} \xrightarrow{P} 1 \qquad (n \to \infty).$$

Proof:
In order to prove the result, respectively

$$\frac{\sigma_\lambda^{n2}(\hat\mu_{\cdot+}^n,\hat\theta^n)-\sigma_\lambda^{n2}(\mu_{\cdot+}^n,\theta_0)}{\sigma_\lambda^{n2}(\mu_{\cdot+}^n,\theta_0)}=o_p(1)\quad\text{and}\quad\frac{s_\lambda^{n2}(\hat\mu_{\cdot+}^n,\hat\theta^n)-s_\lambda^{n2}(\mu_{\cdot+}^n,\theta_0)}{s_\lambda^{n2}(\mu_{\cdot+}^n,\theta_0)}=o_p(1),$$

and since $J^n/\sigma_\lambda^{n2}(\mu_{\cdot+}^n,\theta_0)=O(1)$ and $J^n/s_\lambda^{n2}(\mu_{\cdot+}^n,\theta_0)=O(1)$ was presumed, it obviously suffices to show

$$\sigma_\lambda^{n2}(\hat\mu_{\cdot+}^n,\hat\theta^n)-\sigma_\lambda^{n2}(\mu_{\cdot+}^n,\theta_0)=o_p(J^n)\quad\text{and}\quad s_\lambda^{n2}(\hat\mu_{\cdot+}^n,\hat\theta^n)-s_\lambda^{n2}(\mu_{\cdot+}^n,\theta_0)=o_p(J^n).$$

Writing $\gamma_{\lambda k}^n=\gamma_{\lambda k}^n(\mu_{\cdot+}^n,\theta_0)$ for $k=1,\ldots,K$ and using the definition of s_λ^{n2},

$$s_\lambda^{n2}(\mu_{\cdot+}^n,\theta_0)=\sigma_\lambda^{n2}(\mu_{\cdot+}^n,\theta_0)-\sum_{k=1}^K\mu_{+k}^n(\gamma_{\lambda k}^n(\mu_{\cdot+}^n,\theta_0))^2,$$

the proof now divides into two parts, i.e. one has to establish:

$$\begin{aligned}\sigma_\lambda^{n2}(\hat\mu_{\cdot+}^n,\hat\theta^n)-\sigma_\lambda^{n2}(\mu_{\cdot+}^n,\theta_0)&=o_p(J^n), &(6.35)\\ \sum_{k=1}^K\mu_{+k}^n\Big((\gamma_{\lambda k}^n(\hat\mu_{\cdot+}^n,\hat\theta^n))^2-(\gamma_{\lambda k}^n(\mu_{\cdot+}^n,\theta_0))^2\Big)&=o_p(J^n). &(6.36)\end{aligned}$$

Although the first expression concerns both sampling schemes and the second only column–multinomial sampling with especially $\mu_{+k}^n=n_k$ being the known sample size of the k–th column, the proofs of the statements hold, as preliminarily mentioned, for both distribution models and need no differentiated argumentation. Considering (6.35) first, let for reasons of clarity and brevity $v_\lambda^{n2}(\mu_{\cdot+}^n,\theta_0)$ denote the Poisson variance of SD_λ^n and the correction term concerning the nuisance parameters:

$$\begin{aligned}v_\lambda^{n2}(\mu_{\cdot+}^n,\theta_0)&:=Var\Big(SD_\lambda^n(\mu_{\cdot+}^n,\theta_0|X^n)-\sum_{j=1}^{J^n}a_1(\hat\mu_{j+}^n,\mu_{j+}^n)\Big)\\&=Var\Big(\sum_{j=1}^{J^n}\sum_{k=1}^{K}a_\lambda^n(X_{jk}^n,\mu_{jk}^n)\Big)+2J^n+\sum_{j=1}^{J^n}\frac{1}{\mu_{j+}^n}\\&\quad-2\sum_{j=1}^{J^n}\frac{1}{\mu_{j+}^n}\sum_{k=1}^{K}Cov\Big(a_\lambda(X_{jk}^n,\mu_{jk}^n),(X_{jk}^n)^2\Big)\\&\quad+4\sum_{j=1}^{J^n}\sum_{k=1}^{K}\pi_{jk|C}^n(\theta_0)Cov\Big(a_\lambda(X_{jk}^n,\mu_{jk}^n),X_{jk}^n\Big).\end{aligned}$$

For the proof of (6.35), now the following 6 statements have to be checked:

$$\begin{aligned}\frac{1}{n}(I^n(\hat\mu_{\cdot+}^n,\hat\theta^n)-I^n(\hat\mu_{\cdot+}^n,\theta_0))&=o_p(1), &(6.37)\\ \frac{1}{n}(I^n(\hat\mu_{\cdot+}^n,\theta_0)-I^n(\mu_{\cdot+}^n,\theta_0))&=o_p(1), &(6.38)\end{aligned}$$

$$\frac{1}{J^n}(c_\lambda^n(\hat\mu_{\cdot+}^n, \hat\theta^n) - c_\lambda^n(\hat\mu_{\cdot+}^n, \theta_0)) = o_p(1), \tag{6.39}$$
$$\frac{1}{J^n}(c_\lambda^n(\hat\mu_{\cdot+}^n, \theta_0) - c_\lambda^n(\mu_{\cdot+}^n, \theta_0)) = o_p(1), \tag{6.40}$$
$$\frac{1}{J^n}(v_\lambda^{n2}(\hat\mu_{\cdot+}^n, \hat\theta^n) - v_\lambda^{n2}(\hat\mu_{\cdot+}^n, \theta_0)) = o_p(1), \tag{6.41}$$
$$\frac{1}{J^n}(v_\lambda^{n2}(\hat\mu_{\cdot+}^n, \theta_0) - v_\lambda^{n2}(\mu_{\cdot+}^n, \theta_0)) = o_p(1). \tag{6.42}$$

Provided these hold, (6.37) and (6.38) assert

$$\frac{1}{n}(I^n(\hat\mu_{\cdot+}^n, \hat\theta^n) - I^n(\mu_{\cdot+}^n, \theta_0)) = o_p(1), \tag{6.43}$$

which combined with assumption (LC1), $\frac{1}{n}I^n(\mu_{\cdot+}^n, \theta_0) \longrightarrow I_\infty$ positive definite, immediately gives the (stochastic) convergence of $n(I^n(\mu_{\cdot+}^n, \theta_0))^{-1}$ and $n(I^n(\hat\mu_{\cdot+}^n, \hat\theta^n))^{-1}$ to I_∞^{-1}, thus in particular

$$\begin{aligned} n(I^n(\mu_{\cdot+}^n, \theta_0))^{-1} &= O(1), \\ n(I^n(\hat\mu_{\cdot+}^n, \hat\theta^n))^{-1} &= O_p(1). \end{aligned}$$

Multiplication of both terms to (6.43) further yields

$$(I^n(\hat\mu_{\cdot+}^n, \hat\theta^n))^{-1} = (I^n(\mu_{\cdot+}^n, \theta_0))^{-1} + o_p(\frac{1}{n}).$$

This, the trivial conclusions from (6.39) to (6.42),

$$\begin{aligned} c_\lambda^n(\hat\mu_{\cdot+}^n, \hat\theta^n) &= c_\lambda^n(\mu_{\cdot+}^n, \theta_0) + o_p(J^n), \\ v_\lambda^{n2}(\hat\mu_{\cdot+}^n, \hat\theta^n) &= v_\lambda^{n2}(\mu_{\cdot+}^n, \theta_0) + o_p(J^n), \end{aligned}$$

and (BC), which provides $c_\lambda^n(\mu_{\cdot+}^n, \theta_0) = O(J^n)$ (Lemma 5.6) and, together with $\frac{\mu_{++}^n}{n} = O(1)$, further $\frac{J^n}{n} = O(1)$, then establish statement (6.35):

$$\begin{aligned} \sigma_\lambda^{n2}(\hat\mu_{\cdot+}^n, \hat\theta^n) - \sigma_\lambda^{n2}(\mu_{\cdot+}^n, \theta_0) &= v_\lambda^{n2}(\hat\mu_{\cdot+}^n, \hat\theta^n) - v_\lambda^{n2}(\mu_{\cdot+}^n, \theta_0) \\ &\quad - c_\lambda^n(\hat\mu_{\cdot+}^n, \hat\theta^n)(I^n(\hat\mu_{\cdot+}^n, \hat\theta^n))^{-1}(c_\lambda^n(\hat\mu_{\cdot+}^n, \hat\theta^n))^T \\ &\quad + c_\lambda^n(\mu_{\cdot+}^n, \theta_0)(I^n(\mu_{\cdot+}^n, \theta_0))^{-1}(c_\lambda^n(\mu_{\cdot+}^n, \theta_0))^T \\ &= o_p(J^n) + o_p\Big(\frac{(J^n)^2}{n}\Big) \\ &= o_p(J^n). \end{aligned}$$

In order to verify the statements (6.37) to (6.42) now, let first those concerning the estimation of the finite–dimensional model parameters be examined, i.e. (6.37), (6.39) and (6.41), which will be proved applying Lemma 5.2. Considering an arbitrary component of

$$I^n(\hat{\mu}^n_{\cdot +}, \theta) \;=\; \sum_{j=1}^{J^n} \hat{\mu}^n_{j+} \sum_{k=1}^{K} \frac{1}{\pi^n_{jk|C}(\theta)} D^T_\theta \pi^n_{jk|C}(\theta) D_\theta \pi^n_{jk|C}(\theta),$$

the comprehensive assumptions (RC2) and (RC3) concerning $\pi^n_{jk|C}$ assure the conditions of Lemma 5.2 a) to be met, since it holds $(s, r \in \{1, \ldots, S\})$

$$\begin{aligned} & \sup_{\theta \in \bar{W}} \Big\| D_\theta \Big(\hat{\mu}^n_{j+} \frac{1}{\pi^n_{jk|C}(\theta)} \frac{\partial}{\partial \theta_s} \pi^n_{jk|C}(\theta) \frac{\partial}{\partial \theta_r} \pi^n_{jk|C}(\theta) \Big) \Big\| \\ = \;& \hat{\mu}^n_{j+} \cdot \sup_{\theta \in \bar{W}} \Big\| D_\theta \Big(\frac{1}{\pi^n_{jk|C}(\theta)} \frac{\partial}{\partial \theta_s} \pi^n_{jk|C}(\theta) \frac{\partial}{\partial \theta_r} \pi^n_{jk|C}(\theta) \Big) \Big\| \\ \le \;& \hat{\mu}^n_{j+} \cdot c, \quad c \text{ constant.} \end{aligned}$$

Lemma 5.2 a) thus gives

$$\frac{1}{\sqrt{n}} (I^n(\hat{\mu}^n_{\cdot +}, \hat{\theta}^n) - I^n(\hat{\mu}^n_{\cdot +}, \theta_0)) = O_p(1),$$

and hence (6.37). To prove (6.39), let the s–th component $(s \in \{1, \ldots, S\})$ of c^n_λ (for the definition see also (5.4)) be considered,

$$(c^n_\lambda(\mu^n_{\cdot +}, \theta))_s \;=\; \sum_{j=1}^{J^n} \sum_{k=1}^{K} f\Big(\mu^n_{j+}, \pi^n_{jk|C}(\theta), \frac{\partial}{\partial \theta_s} \pi^n_{jk|C}(\theta) \Big)$$

with

$$\begin{aligned} & f\Big(\mu^n_{j+}, \pi^n_{jk|C}(\theta), \frac{\partial}{\partial \theta_s} \pi^n_{jk|C}(\theta) \Big) \\ = \;& \frac{\partial}{\partial \theta_s} \log \pi^n_{jk|C}(\theta) \cdot Cov_{(\mu^n_{j+}, \pi^n_{jk|C}(\theta))} \Big(a_\lambda(X^n_{jk}, \mu^n_{j+} \pi^n_{jk|C}(\theta)), X^n_{jk} \Big). \end{aligned}$$

Let the subscript of the covariance term in the following be neglected. If $\bar{W}$ is a convex and compact neighbourhood of θ_0 in the $\mathbf{R}^S$ and if for every j, k, n holds

$$\sup_{\theta \in \bar{W}} \Big\| D_\theta f\Big(\mu^n_{j+}, \pi^n_{jk|C}(\theta), \frac{\partial}{\partial \theta_s} \pi^n_{jk|C}(\theta) \Big) \Big\| \le c \in \mathbf{R}^+ \text{ constant} \tag{6.44}$$

for all $\mu^n_{j+} \in [K\epsilon, \infty)$, $K\epsilon \le 1$, the requirements of Lemma 5.2 b) are fulfilled. For each component and hence for the whole vector then follows

$$\frac{\sqrt{n}}{J^n} (c^n_\lambda(\hat{\mu}^n_{\cdot +}, \hat{\theta}^n) - c^n_\lambda(\hat{\mu}^n_{\cdot +}, \theta_0)) = O_p(1),$$

in particular (6.39). In order to establish statement (6.44) consider

$$D_\theta f\Big(\mu^n_{j+}, \pi^n_{jk|C}(\theta), \frac{\partial}{\partial \theta_s} \pi^n_{jk|C}(\theta) \Big)$$

$$
\begin{aligned}
&= D_\theta\Big(\frac{\partial}{\partial\theta_s}\log\pi^n_{jk|C}(\theta)\cdot Cov\Big(a_\lambda(X^n_{jk},\mu^n_{j+}\pi^n_{jk|C}(\theta)),X^n_{jk}\Big)\Big)\\
&= D_\theta\frac{\partial}{\partial\theta_s}\log\pi^n_{jk|C}(\theta)\cdot Cov\Big(a_\lambda(X^n_{jk},\mu^n_{j+}\pi^n_{jk|C}(\theta)),X^n_{jk}\Big)\\
&\quad+\frac{\partial}{\partial\theta_s}\log\pi^n_{jk|C}(\theta)\cdot D_\theta Cov\Big(a_\lambda(X^n_{jk},\mu^n_{j+}\pi^n_{jk|C}(\theta)),X^n_{jk}\Big)\\
&= D_\theta\frac{\partial}{\partial\theta_s}\log\pi^n_{jk|C}(\theta)\cdot Cov\Big(a_\lambda(X^n_{jk},\mu^n_{j+}\pi^n_{jk|C}(\theta)),X^n_{jk}\Big)\\
&\quad+\frac{\partial}{\partial\theta_s}\log\pi^n_{jk|C}(\theta)\cdot D_\theta\log\pi^n_{jk|C}(\theta)\\
&\quad\cdot\mu^n_{j+}\pi^n_{jk|C}(\theta)\frac{\partial}{\partial\mu^n_{j+}\pi^n_{jk|C}(\theta)}Cov\Big(a_\lambda(X^n_{jk},\mu^n_{j+}\pi^n_{jk|C}(\theta)),X^n_{jk}\Big).
\end{aligned}
$$

With $\mu^n_{j+}\geq K\epsilon$ and $\theta\in\bar W$ the expectations $\mu^n_{j+}\pi^n_{jk|C}(\theta)$ are bounded away from zero and Lemma 4.8 d), j) thus gives for all j,k,n

$$
\begin{aligned}
\Big|Cov\Big(a_\lambda(X^n_{jk},\mu^n_{j+}\pi^n_{jk|C}(\theta)),X^n_{jk}\Big)\Big| &\leq c \qquad (6.45)\\
\Big|\mu^n_{j+}\pi^n_{jk|C}(\theta)\frac{\partial}{\partial\mu^n_{j+}\pi^n_{jk|C}(\theta)}Cov\Big(a_\lambda(X^n_{jk},\mu^n_{j+}\pi^n_{jk|C}(\theta)),X^n_{jk}\Big)\Big| &\leq c \qquad (6.46)
\end{aligned}
$$

for some constant c. The boundedness stated in (6.44) now follows using these results and the generally assumed conditions (RC2) and (RC3), which assure $1/\pi^n_{jk|C}(\theta)$, $D_\theta\pi^n_{jk|C}(\theta)$, $D^2_\theta\pi^n_{jk|C}(\theta)$ and hence especially $D_\theta\log\pi^n_{jk|C}(\theta)$ being bounded for all j,k,n, $\theta\in\bar W$. The latter arguments will in the following be continuously of evidence and not be stated explicitly each time needed.

The proof of (6.41),

$$
\frac{1}{J^n}(v^{n2}_\lambda(\hat\mu^n_{\cdot+},\hat\theta^n)-v^{n2}_\lambda(\hat\mu^n_{\cdot+},\theta_0))=o_p(1),
$$

will also be done applying Lemma 5.2 b). Hence the condition corresponding to (6.44) for the inner terms of the sum $v^{n2}_\lambda(\mu^n_{\cdot+},\theta)$ has to be checked, which is again done using the results given in Lemma 4.8. For the proof consider the representation

$$
\begin{aligned}
v^{n2}_\lambda(\mu^n_{\cdot+},\theta) &= \sum_{j=1}^{J^n}\sum_{k=1}^{K}f_1(\mu^n_{j+},\pi^n_{jk|C}(\theta))-2\sum_{j=1}^{J^n}\sum_{k=1}^{K}f_2(\mu^n_{j+},\pi^n_{jk|C}(\theta))\\
&\quad+4\sum_{j=1}^{J^n}\sum_{k=1}^{K}f_3(\mu^n_{j+},\pi^n_{jk|C}(\theta))+2J^n+\sum_{j=1}^{J^n}\frac{1}{\mu^n_{j+}}
\end{aligned}
$$

with the last two terms not depending on the model parameters and

$$
f_1(\mu^n_{j+},\pi^n_{jk|C}(\theta)) = Var\Big(a_\lambda(X^n_{jk},\mu^n_{j+}\pi^n_{jk|C}(\theta))\Big),
$$

$$\begin{aligned}
f_2(\mu^n_{j+}, \pi^n_{jk|C}(\theta)) &= \frac{1}{\mu^n_{j+}} Cov\Big(a_\lambda(X^n_{jk}, \mu^n_{j+}\pi^n_{jk|C}(\theta)), (X^n_{jk})^2\Big),\\
f_3(\mu^n_{j+}, \pi^n_{jk|C}(\theta)) &= \pi^n_{jk|C}(\theta) Cov\Big(a_\lambda(X^n_{jk}, \mu^n_{j+}\pi^n_{jk|C}(\theta)), X^n_{jk}\Big).
\end{aligned}$$

Now there is to show the existence of a constant $c \in \mathbf{R}^+$ such that for every j, k, n and $i = 1, 2, 3$ holds

$$\sup_{\theta \in \bar{W}} \|D_\theta f_i(\mu^n_{j+}, \pi^n_{jk|C}(\theta))\| \leq c \quad \text{for all } \mu^n_{j+} \in [K\epsilon, \infty)$$

(cp. (6.44)). Lemma 5.2 b) then immediately yields (6.41). Since for $\mu^n_{j+} \geq K\epsilon$ and $\theta \in \bar{W}$ the expectations $\mu^n_{j+}\pi^n_{jk|C}(\theta)$ are bounded away from zero, Lemma 4.8 i) gives for the first term

$$\begin{aligned}
&\sup_{\theta \in \bar{W}} \|D_\theta f_1(\mu^n_{j+}, \pi^n_{jk|C}(\theta))\|\\
=\ &\sup_{\theta \in \bar{W}} \Big\|D_\theta \mu^n_{j+}\pi^n_{jk|C}(\theta) \cdot \frac{\partial}{\partial \mu^n_{j+}\pi^n_{jk|C}(\theta)} Var\Big(a_\lambda(X^n_{jk}, \mu^n_{j+}\pi^n_{jk|C}(\theta))\Big)\Big\|\\
=\ &\sup_{\theta \in \bar{W}} \Big\|D_\theta \log \pi^n_{jk|C}(\theta) \cdot \mu^n_{jk}(\theta) \cdot \frac{\partial}{\partial \mu^n_{jk}(\theta)} Var\Big(a_\lambda(X^n_{jk}, \mu^n_{jk}(\theta)\Big)\Big\|\\
\leq\ &\sup_{\theta \in \bar{W}} \|D_\theta \log \pi^n_{jk|C}(\theta)\| \cdot \Big|\mu^n_{jk}(\theta) \cdot \frac{\partial}{\partial \mu^n_{jk}(\theta)} Var\Big(a_\lambda(X^n_{jk}, \mu^n_{jk}(\theta)\Big)\Big|\\
\leq\ &c, \quad c \in \mathbf{R}^+ \text{ constant},
\end{aligned}$$

with the variance and the following expectations being $Pois(\mu^n_{j+}\pi^n_{jk|C}(\theta))$. Analogously Lemma 4.8 j), k) provides the boundedness of

$$\begin{aligned}
&\sup_{\theta \in \bar{W}} \|D_\theta f_2(\mu^n_{j+}, \pi^n_{jk|C}(\theta))\|\\
=\ &\sup_{\theta \in \bar{W}} \Big\|\frac{1}{\mu^n_{j+}} D_\theta Cov\Big(a_\lambda(X^n_{jk}, \mu^n_{j+}\pi^n_{jk|C}(\theta)), (X^n_{jk})^2\Big)\Big\|\\
=\ &\sup_{\theta \in \bar{W}} \Big\|\frac{1}{\mu^n_{j+}} \cdot D_\theta \mu^n_{jk}(\theta) \cdot \frac{\partial}{\partial \mu^n_{jk}(\theta)} Cov\Big(a_\lambda(X^n_{jk}, \mu^n_{jk}(\theta)), (X^n_{jk})^2\Big)\Big\|\\
=\ &\sup_{\theta \in \bar{W}} \Big\|D_\theta \pi^n_{jk|C}(\theta) \cdot \frac{\partial}{\partial \mu^n_{jk}(\theta)} Cov\Big(a_\lambda(X^n_{jk}, \mu^n_{jk}(\theta)), (X^n_{jk})^2\Big)\Big\|\\
\leq\ &\sup_{\theta \in \bar{W}} \|D_\theta \pi^n_{jk|C}(\theta)\| \cdot \Big|\frac{\partial}{\partial \mu^n_{jk}(\theta)} Cov\Big(a_\lambda(X^n_{jk}, \mu^n_{jk}(\theta)), (X^n_{jk})^2\Big)\Big|
\end{aligned}$$

and

$$\begin{aligned}
&\sup_{\theta \in \bar{W}} \|D_\theta f_3(\mu^n_{j+}, \pi^n_{jk|C}(\theta))\|\\
=\ &\sup_{\theta \in \bar{W}} \Big\|D_\theta\Big(\pi^n_{jk|C}(\theta) \cdot Cov\Big(a_\lambda(X^n_{jk}, \mu^n_{j+}\pi^n_{jk|C}(\theta)), X^n_{jk}\Big)\Big)\Big\|
\end{aligned}$$

$$\leq \sup_{\theta\in\bar{W}} \|D_\theta \pi^n_{jk|C}(\theta)\| \cdot \Big|Cov\Big(a_\lambda(X^n_{jk}, \mu^n_{jk}(\theta)), X^n_{jk}\Big)\Big|$$
$$+ \sup_{\theta\in\bar{W}} \|D_\theta \pi^n_{jk|C}(\theta)\| \cdot \Big|\mu^n_{jk}(\theta) \cdot \frac{\partial}{\partial \mu^n_{jk}(\theta)} Cov\Big(a_\lambda(X^n_{jk}, \mu^n_{jk}(\theta)), X^n_{jk}\Big)\Big|,$$

which completes the proof of (6.41).

The statements (6.38), (6.40) and (6.42) concerning the estimation of the nuisance parameters will now be considered. (6.38) immediately follows using $(\hat{\mu}^n_{j+} - \mu^n_{j+})/\sqrt{\mu^n_{j+}} = O_p(1)$, which has already been used several times and can for example be shown applying Chebyshev's inequality. Then it holds

$$\begin{aligned}
&\frac{1}{n}(I^n(\hat{\mu}^n_{\cdot+}, \theta_0) - I^n(\mu^n_{\cdot+}, \theta_0)) \\
= &\frac{1}{n}\sum_{j=1}^{J^n}(\hat{\mu}^n_{j+} - \mu^n_{j+}) \sum_{k=1}^{K} \frac{1}{\pi^n_{jk|C}(\theta_0)} D^T_\theta \pi^n_{jk|C}(\theta_0) D_\theta \pi^n_{jk|C}(\theta_0) \\
= &\frac{1}{n}\sum_{j=1}^{J^n} \sqrt{\mu^n_{j+}} \frac{(\hat{\mu}^n_{j+} - \mu^n_{j+})}{\sqrt{\mu^n_{j+}}} \cdot O(1) \\
= &\sqrt{\frac{J^n}{n}} \cdot \sqrt{\frac{1}{J^n}} \sum_{j=1}^{J^n} \sqrt{\frac{\mu^n_{j+}}{n}} \cdot O_p(1) \\
= &o_p(1),
\end{aligned}$$

with the last equation following from $\frac{J^n}{n} = O(1)$ and assumption (MD1). In order to prove (6.40), one has to show for each component of c^n_λ, i.e. for every $s \in \{1, \ldots, S\}$:

$$\frac{1}{J^n}\Big(\sum_{j=1}^{J^n}\sum_{k=1}^{K} \frac{\partial}{\partial \theta_s} \log \pi^n_{jk|C}(\theta_0) \cdot \Big(f(\hat{\mu}^n_{j+}, \pi^n_{jk|C}(\theta_0)) - f(\mu^n_{j+}, \pi^n_{jk|C}(\theta_0))\Big)\Big) = o_p(1) \quad (6.47)$$

with

$$f(\mu^n_{j+}, \pi^n_{jk|C}(\theta_0)) := Cov_{(\mu^n_{j+}, \pi^n_{jk|C}(\theta_0))}\Big(a_\lambda(X^n_{jk}, \mu^n_{j+}\pi^n_{jk|C}(\theta_0)), X^n_{jk}\Big).$$

The subscript of the covariance will in the following again be omitted. Lemma 4.8 j) now gives for all j, k, n the boundedness of

$$\begin{aligned}
&|\mu^n_{j+} \cdot \frac{\partial}{\partial \mu^n_{j+}} f(\mu^n_{j+}, \pi^n_{jk|C}(\theta_0))| \\
= &\Big|\mu^n_{j+} \cdot \frac{\partial}{\partial \mu^n_{j+}} Cov\Big(a_\lambda(X^n_{jk}, \mu^n_{j+}\pi^n_{jk|C}(\theta_0)), X^n_{jk}\Big)\Big| \\
= &\Big|\mu^n_{j+} \cdot \frac{\partial}{\partial \mu^n_{j+}} \mu^n_{j+}\pi^n_{jk|C} \cdot \frac{\partial}{\partial \mu^n_{j+}\pi^n_{jk|C}} Cov\Big(a_\lambda(X^n_{jk}, \mu^n_{j+}\pi^n_{jk|C}(\theta_0)), X^n_{jk}\Big)\Big|
\end{aligned}$$

$$= \left|\mu^n_{j+}\pi^n_{jk|C} \cdot \frac{\partial}{\partial \mu^n_{j+}\pi^n_{jk|C}} Cov\Big(a_\lambda(X^n_{jk}, \mu^n_{j+}\pi^n_{jk|C}(\theta_0)), X^n_{jk}\Big)\right| \tag{6.48}$$

(cp. (6.46)). Hence the assumptions of Lemma 8.3 are fulfilled and it follows

$$\mathbf{1_N}(\hat{\mu}^n_{j+})\Big(f(\hat{\mu}^n_{j+}, \pi^n_{jk|C}(\theta_0)) - f(\mu^n_{j+}, \pi^n_{jk|C}(\theta_0))\Big) = O_p(\frac{1}{\sqrt{\mu^n_{j+}}}) \quad \text{for all } \ j,k,$$

i.e. for each $\delta \in (0,1)$ and j there exists a constant M_δ, so that for every k holds

$$P\Big(\sqrt{\mu^n_{j+}} \cdot \mathbf{1_N}(\hat{\mu}^n_{j+}) \cdot \Big|f(\hat{\mu}^n_{j+}, \pi^n_{jk|C}(\theta_0)) - f(\mu^n_{j+}, \pi^n_{jk|C}(\theta_0))\Big| > M_\delta\Big) \leq \frac{\delta}{2}$$

for almost all n. This combined with assumption (LC0) resp.

$$P(\exists\, j : \hat{\mu}^n_{j+} = 0) \leq \frac{\delta}{2} \quad \text{for } n \text{ sufficiently large,}$$

gives the inequality

$$\begin{aligned}
& P\Big(\sqrt{\mu^n_{j+}}\Big|f(\hat{\mu}^n_{j+}, \pi^n_{jk|C}(\theta_0)) - f(\mu^n_{j+}, \pi^n_{jk|C}(\theta_0))\Big| > M_\delta\Big) \\
\leq\ & P\Big(\sqrt{\mu^n_{j+}}\Big|f(\hat{\mu}^n_{j+}, \pi^n_{jk|C}(\theta_0)) - f(\mu^n_{j+}, \pi^n_{jk|C}(\theta_0))\Big| > M_\delta, \hat{\mu}^n_{j+} > 0\Big) \\
& + P(\exists\, j : \hat{\mu}^n_{j+} = 0) \\
=\ & P\Big(\sqrt{\mu^n_{j+}} \cdot \mathbf{1_N}(\hat{\mu}^n_{j+}) \cdot \Big|f(\hat{\mu}^n_{j+}, \pi^n_{jk|C}(\theta_0)) - f(\mu^n_{j+}, \pi^n_{jk|C}(\theta_0))\Big| > M_\delta\Big) \\
& + P(\exists\, j : \hat{\mu}^n_{j+} = 0) \\
\leq\ & \frac{\delta}{2} + \frac{\delta}{2}
\end{aligned}$$

and hence the stochastic boundedness of the expression:

$$f(\hat{\mu}^n_{j+}, \pi^n_{jk|C}(\theta_0)) - f(\mu^n_{j+}, \pi^n_{jk|C}(\theta_0)) = O_p(\frac{1}{\sqrt{\mu^n_{j+}}}) \quad \text{for all } \ j,k. \tag{6.49}$$

Inserting in (6.47) and use of condition $\frac{1}{J^n}\sum_{j=1}^{J^n} \frac{1}{\sqrt{\mu^n_{j+}}} = o(1)$ finally yields (6.40):

$$\begin{aligned}
\frac{1}{J^n}(c^n_\lambda(\hat{\mu}^n_{\cdot+}, \theta_0) - c^n_\lambda(\mu^n_{\cdot+}, \theta_0)) &= \frac{1}{J^n}\sum_{j=1}^{J^n}\sum_{k=1}^{K} D_\theta \log \pi^n_{jk|C}(\theta_0) \cdot O_p(\frac{1}{\sqrt{\mu^n_{j+}}}) \\
&= \frac{1}{J^n} \cdot O_p(1) \sum_{j=1}^{J^n} \frac{1}{\sqrt{\mu^n_{j+}}} \\
&= o_p(1).
\end{aligned}$$

In order to verify (6.42),

$$\frac{1}{J^n}(v^{n2}_\lambda(\hat{\mu}^n_{\cdot+}, \theta_0) - v^{n2}_\lambda(\mu^n_{\cdot+}, \theta_0)) = o_p(1),$$

which can again be done using arguments analogous to the preceding proof of the statement concerning c_λ^n, i.e. (6.40), consider

$$
\begin{aligned}
& v_\lambda^{n2}(\hat{\mu}_{\cdot+}^n, \theta_0) - v_\lambda^{n2}(\mu_{\cdot+}^n, \theta_0) \\
&= \sum_{j=1}^{J^n}\sum_{k=1}^{K} \Big(f_1(\hat{\mu}_{j+}^n, \pi_{jk|C}^n(\theta_0)) - f_1(\mu_{j+}^n, \pi_{jk|C}^n(\theta_0)) \Big) \\
&+ \sum_{j=1}^{J^n} \Big(f_2(\hat{\mu}_{j+}^n, \pi_{jk|C}^n(\theta_0)) - f_2(\mu_{j+}^n, \pi_{jk|C}^n(\theta_0)) \Big) \\
&-2 \sum_{j=1}^{J^n}\sum_{k=1}^{K} \Big(f_3(\hat{\mu}_{j+}^n, \pi_{jk|C}^n(\theta_0)) - f_3(\mu_{j+}^n, \pi_{jk|C}^n(\theta_0)) \Big) \\
&+4 \sum_{j=1}^{J^n}\sum_{k=1}^{K} \Big(f_4(\hat{\mu}_{j+}^n, \pi_{jk|C}^n(\theta_0)) - f_4(\mu_{j+}^n, \pi_{jk|C}^n(\theta_0)) \Big)
\end{aligned}
$$

with

$$
\begin{aligned}
f_1(\mu_{j+}^n, \pi_{jk|C}^n(\theta_0)) &= Var\Big(a_\lambda(X_{jk}^n, \mu_{j+}^n \pi_{jk|C}^n(\theta_0))\Big), \\
f_2(\mu_{j+}^n, \pi_{jk|C}^n(\theta_0)) &= f_2(\mu_{j+}^n) = \frac{1}{\mu_{j+}^n} \\
f_3(\mu_{j+}^n, \pi_{jk|C}^n(\theta_0)) &= \frac{1}{\mu_{j+}^n} Cov\Big(a_\lambda(X_{jk}^n, \mu_{j+}^n \pi_{jk|C}^n(\theta_0)), (X_{jk}^n)^2\Big), \\
f_4(\mu_{j+}^n, \pi_{jk|C}^n(\theta_0)) &= \pi_{jk|C}^n(\theta_0) Cov\Big(a_\lambda(X_{jk}^n, \mu_{j+}^n \pi_{jk|C}^n(\theta_0)), X_{jk}^n\Big).
\end{aligned}
$$

What will be shown is a bounding condition corresponding to (6.48), namely ($1 \leq i \leq 4$):

$$
|\mu_{j+}^n \frac{\partial}{\partial \mu_{j+}^n} f_i(\mu_{j+}^n, \pi_{jk|C}^n(\theta_0))| \leq c \in \mathbf{R}^+ \text{ constant} \quad \text{for all } j, k, n. \tag{6.50}
$$

Application of Lemma 8.3 and analogous arguments then entail each term to be of the order $\sum_{j=1}^{J^n} O_p(\frac{1}{\sqrt{\mu_{j+}^n}})$ and the assumption concerning the marginal distribution, $\frac{1}{J^n} \cdot \sum_{j=1}^{J^n} O_p(\frac{1}{\sqrt{\mu_{j+}^n}}) = o_p(1)$, gives (6.42). Use of condition (BC), which implies $\frac{1}{\mu_{j+}^n} \leq \frac{1}{K\epsilon}$, and Lemma 4.8 f), i) – k) establish (6.50) as follows:

$$
\begin{aligned}
& |\mu_{j+}^n \frac{\partial}{\partial \mu_{j+}^n} f_1(\mu_{j+}^n, \pi_{jk|C}^n(\theta_0))| \\
= \; & \Big|\mu_{j+}^n \frac{\partial}{\partial \mu_{j+}^n} Var\Big(a_\lambda(X_{jk}^n, \mu_{j+}^n \pi_{jk|C}^n(\theta_0))\Big)\Big| \\
= \; & \Big|\mu_{j+}^n \cdot \frac{\partial}{\partial \mu_{j+}^n} \mu_{j+}^n \pi_{jk|C}^n(\theta_0) \cdot \frac{\partial}{\partial \mu_{j+}^n \pi_{jk|C}^n(\theta_0)} Var\Big(a_\lambda(X_{jk}^n, \mu_{j+}^n \pi_{jk|C}^n(\theta_0))\Big)\Big|
\end{aligned}
$$

$$
\begin{aligned}
&= \Big|\mu^n_{jk} \cdot \frac{\partial}{\partial \mu^n_{jk}} Var\Big(a_\lambda(X^n_{jk}, \mu^n_{jk})\Big)\Big| \\
&\leq c \quad \text{for all } \; j, k, n,
\end{aligned}
$$

$$
\begin{aligned}
&|\mu^n_{j+} \frac{\partial}{\partial \mu^n_{j+}} f_2(\mu^n_{j+}, \pi^n_{jk|C}(\theta_0))| \\
&= |\mu^n_{j+} \cdot \frac{\partial}{\partial \mu^n_{j+}} \frac{1}{\mu^n_{j+}}| \\
&= \frac{1}{\mu^n_{j+}} \\
&\leq \frac{1}{K\epsilon} \quad \text{for all } \; j, k, n,
\end{aligned}
$$

$$
\begin{aligned}
&|\mu^n_{j+} \frac{\partial}{\partial \mu^n_{j+}} f_3(\mu^n_{j+}, \pi^n_{jk|C}(\theta_0))| \\
&= \Big|\mu^n_{j+} \cdot \frac{\partial}{\partial \mu^n_{j+}} \Big(\frac{1}{\mu^n_{j+}} Cov\Big(a_\lambda(X^n_{jk}, \mu^n_{j+}\pi^n_{jk|C}(\theta_0)), (X^n_{jk})^2\Big)\Big)\Big| \\
&= \Big|\mu^n_{j+} \cdot \Big(- (\frac{1}{\mu^n_{j+}})^2 Cov\Big(a_\lambda(X^n_{jk}, \mu^n_{j+}\pi^n_{jk|C}(\theta_0)), (X^n_{jk})^2\Big)\Big) \\
&\quad + \frac{\partial}{\partial \mu^n_{j+}} \mu^n_{j+}\pi^n_{jk|C}(\theta_0) \cdot \frac{\partial}{\partial \mu^n_{j+}\pi^n_{jk|C}(\theta_0)} Cov\Big(a_\lambda(X^n_{jk}, \mu^n_{j+}\pi^n_{jk|C}(\theta_0)), (X^n_{jk})^2\Big)\Big| \\
&\leq \Big|\frac{1}{\mu^n_{j+}} Cov\Big(a_\lambda(X^n_{jk}, \mu^n_{jk}), (X^n_{jk})^2\Big)\Big| \\
&\qquad + \Big|\pi^n_{jk|C}(\theta_0) \cdot \frac{\partial}{\partial \mu^n_{jk}} Cov\Big(a_\lambda(X^n_{jk}, \mu^n_{jk}), (X^n_{jk})^2\Big)\Big| \\
&\leq c \quad \text{for all } \; j, k, n,
\end{aligned}
$$

$$
\begin{aligned}
&|\mu^n_{j+} \frac{\partial}{\partial \mu^n_{j+}} f_4(\mu^n_{j+}, \pi^n_{jk|C}(\theta_0))| \\
&= \Big|\mu^n_{j+} \cdot \frac{\partial}{\partial \mu^n_{j+}} \Big(\pi^n_{jk|C}(\theta_0) \cdot Cov\Big(a_\lambda(X^n_{jk}, \mu^n_{j+}\pi^n_{jk|C}(\theta_0)), X^n_{jk}\Big)\Big)\Big| \\
&= \Big|\mu^n_{j+}\pi^n_{jk|C}(\theta_0) \cdot \frac{\partial}{\partial \mu^n_{j+}} \mu^n_{j+}\pi^n_{jk|C}(\theta_0) \\
&\quad \cdot \frac{\partial}{\partial \mu^n_{j+}\pi^n_{jk|C}(\theta_0)} Cov\Big(a_\lambda(X^n_{jk}, \mu^n_{j+}\pi^n_{jk|C}(\theta_0)), X^n_{jk}\Big)\Big| \\
&= \Big|\pi^n_{jk|C}(\theta_0) \cdot \mu^n_{jk} \cdot \frac{\partial}{\partial \mu^n_{jk}} Cov\Big(a_\lambda(X^n_{jk}, \mu^n_{jk}), X^n_{jk}\Big)\Big| \\
&\leq c \quad \text{for all } \; j, k, n,
\end{aligned}
$$

With (6.42) and hence (6.35) now being proved, statement (6.36),

$$\sum_{k=1}^{K} \mu_{+k}^n \Big((\gamma_{\lambda k}^n(\hat{\mu}_{\cdot+}^n, \hat{\theta}^n))^2 - (\gamma_{\lambda k}^n(\mu_{\cdot+}^n, \theta_0))^2 \Big) = o_p(J^n),$$

remains to be verified. Writing $f(\mu_{j+}^n, \pi_{jk|C}^n(\theta_0)) := Cov(a_\lambda(X_{jk}^n, \mu_{jk}^n), X_{jk}^n) - \pi_{jk|C}^n(\theta_0)$ then by definition of $\gamma_{\lambda k}^n$ holds (see Th. 6.1 and p. 97)

$$\gamma_{\lambda k}^n(\mu_{\cdot+}, \theta_0) = \frac{1}{\mu_{+k}^n} \sum_{j=1}^{J^n} f(\mu_{j+}^n, \pi_{jk|C}^n(\theta_0)),$$

so that the term in question in (6.36) can be rewritten as follows:

$$\begin{aligned}
&\sum_{k=1}^{K} \mu_{+k}^n \Big((\gamma_{\lambda k}^n(\hat{\mu}_{\cdot+}^n, \hat{\theta}^n))^2 - (\gamma_{\lambda k}^n(\mu_{\cdot+}^n, \theta_0))^2 \Big) \\
= &\sum_{k=1}^{K} \frac{1}{\mu_{+k}^n} \bigg(\Big(\sum_{j=1}^{J^n} f(\hat{\mu}_{j+}^n, \pi_{jk|C}^n(\hat{\theta})) \Big)^2 - \Big(\sum_{j=1}^{J^n} f(\mu_{j+}^n, \pi_{jk|C}^n(\theta_0)) \Big)^2 \bigg) \\
= &\sum_{k=1}^{K} \frac{1}{\mu_{+k}^n} \bigg(\sum_{j=1}^{J^n} \sum_{l=1}^{J^n} \Big(f(\hat{\mu}_{j+}^n, \pi_{jk|C}^n(\hat{\theta})) f(\hat{\mu}_{l+}^n, \pi_{lk|C}^n(\hat{\theta})) \\
&\qquad - f(\mu_{j+}^n, \pi_{jk|C}^n(\theta_0)) f(\mu_{l+}^n, \pi_{lk|C}^n(\theta_0)) \Big) \bigg).
\end{aligned}$$

Because of the double sum, the determination of the order will be done using slightly different arguments as before. Since, however, essentially the same approach will be considered, and this modified argumentation will only be needed at this point of the proof, just a short outline will be given. In order to show the result, again two steps are considered:

$$\begin{aligned}
\frac{1}{J^n} \sum_{k=1}^{K} \frac{1}{\mu_{+k}^n} \bigg(\sum_{j=1}^{J^n} \sum_{l=1}^{J^n} \Big(f(\hat{\mu}_{j+}^n, \pi_{jk|C}^n(\hat{\theta})) f(\hat{\mu}_{l+}^n, \pi_{lk|C}^n(\hat{\theta})) \\
- f(\hat{\mu}_{j+}^n, \pi_{jk|C}^n(\theta_0)) f(\hat{\mu}_{l+}^n, \pi_{lk|C}^n(\theta_0)) \Big) \bigg) &= o_p(1), &(6.51) \\
\frac{1}{J^n} \sum_{k=1}^{K} \frac{1}{\mu_{+k}^n} \bigg(\sum_{j=1}^{J^n} \sum_{l=1}^{J^n} \Big(f(\hat{\mu}_{j+}^n, \pi_{jk|C}^n(\theta_0)) f(\hat{\mu}_{l+}^n, \pi_{lk|C}^n(\theta_0)) \\
- f(\mu_{j+}^n, \pi_{jk|C}^n(\theta_0)) f(\mu_{l+}^n, \pi_{lk|C}^n(\theta_0)) \Big) \bigg) &= o_p(1). &(6.52)
\end{aligned}$$

The proceeding for the proof of (6.51), is the same as in Lemma 5.2: using a Taylor expansion in $\hat{\theta}^n$ on a convex compact set $\bar{W} \subset \mathbf{R}^S$ with $\hat{\mu}_{j+}^n > 0$, which yields the inequality

$$
\begin{aligned}
&\sqrt{n}\Big\| \frac{1}{J^n} \sum_{k=1}^{K} \frac{1}{\mu^n_{+k}} \Big(\sum_{j=1}^{J^n} \sum_{l=1}^{J^n} \Big(f(\hat{\mu}^n_{j+}, \pi^n_{jk|C}(\hat{\theta})) f(\hat{\mu}^n_{l+}, \pi^n_{lk|C}(\hat{\theta})) \\
&\qquad\qquad - f(\hat{\mu}^n_{j+}, \pi^n_{jk|C}(\theta_0)) f(\hat{\mu}^n_{l+}, \pi^n_{lk|C}(\theta_0)) \Big) \Big) \Big\| \\
&\le \|\hat{\theta}^n - \theta_0\| \cdot \frac{\sqrt{n}}{J^n} \cdot \sum_{k=1}^{K} \frac{1}{\mu^n_{+k}} \sum_{j=1}^{J^n} \sum_{l=1}^{J^n} \sup_{\theta \in \bar{W}} \Big\| D_\theta \Big(f(\hat{\mu}^n_{j+}, \pi^n_{jk|C}(\theta)) f(\hat{\mu}^n_{l+}, \pi^n_{lk|C}(\theta)) \Big) \Big\|. \quad (6.53)
\end{aligned}
$$

If now holds

$$
\sup_{\theta \in \bar{W}} \Big\| D_\theta \Big(f(\hat{\mu}^n_{j+}, \pi^n_{jk|C}(\theta)) f(\hat{\mu}^n_{l+}, \pi^n_{lk|C}(\theta)) \Big) \Big\| \le c \quad \text{for some } c \in \mathbf{R}^+ \text{ constant}, \quad (6.54)
$$

(6.53) is dominated by

$$
\|(\hat{\theta}^n - \theta_0)\sqrt{n}\| \cdot \frac{1}{\sqrt{n}} \cdot \frac{\sqrt{n}}{J^n} \cdot \sum_{k=1}^{K} \frac{1}{\mu^n_{+k}} \sum_{j=1}^{J^n} \sum_{l=1}^{J^n} c = c \cdot \|(\hat{\theta}^n - \theta_0)\sqrt{n}\| \cdot \sum_{k=1}^{K} \frac{J^n}{\mu^n_{+k}}.
$$

The presumed (stochastic) boundedness of $(\hat{\theta}^n - \theta_0)\sqrt{n}$ (LC2), $J^n/\mu^n_{+k} = O(1)$, which follows from (BC), and known arguments, then give for the expression in (6.51) the order $O_p(\frac{1}{\sqrt{n}})$, and hence the stochastic convergence to zero. Similar to prior argumentation, (6.54) follows using the results from Lemma 4.8.

Statement (6.52) can now be verified similarly to the previous proofs concerning the nuisance parameters, where especially the condition given in Lemma 8.3 was to be checked. The approach of Lemma 8.3, a one–dimensional Taylor expansion, will now be transferred to the case considered here, which, because of the double sum, requires an evaluation in two variables. The essential arguments are the same, however. Let now g denote the product on the inner terms of the sum, i.e.

$$
g(\hat{\mu}^n_{j+}, \hat{\mu}^n_{l+}, \pi^n_{jk|C}(\theta_0), \pi^n_{lk|C}(\theta_0)) = f(\hat{\mu}^n_{j+}, \pi^n_{jk|C}(\theta_0)) \cdot f(\hat{\mu}^n_{l+}, \pi^n_{lk|C}(\theta_0)),
$$

and for $j \neq l$ now consider the following Taylor expansion:

$$
\begin{aligned}
& \mathbf{1}_{\mathbf{N}}(\hat{\mu}^n_{j+} \cdot \hat{\mu}^n_{l+}) \\
& \cdot |g(\hat{\mu}^n_{j+}, \hat{\mu}^n_{l+}, \pi^n_{jk|C}(\theta_0), \pi^n_{lk|C}(\theta_0)) - g(\mu^n_{j+}, \mu^n_{l+}, \pi^n_{jk|C}(\theta_0), \pi^n_{lk|C}(\theta_0))| \\
= \; & \mathbf{1}_{\mathbf{N}}(\hat{\mu}^n_{j+} \cdot \hat{\mu}^n_{l+}) \\
& \cdot \Big| \int_0^1 \Big(D_1 g(\mu^n_{j+} + z(\hat{\mu}^n_{j+} - \mu^n_{j+}), \mu^n_{l+} + z(\hat{\mu}^n_{l+} - \mu^n_{l+}), \pi^n_{jk|C}(\theta_0), \pi^n_{lk|C}(\theta_0)) \\
& \quad \cdot (\hat{\mu}^n_{j+} - \mu^n_{j+}) \\
& \quad + D_2 g(\mu^n_{j+} + z(\hat{\mu}^n_{j+} - \mu^n_{j+}), \mu^n_{l+} + z(\hat{\mu}^n_{l+} - \mu^n_{l+}), \pi^n_{jk|C}(\theta_0), \pi^n_{lk|C}(\theta_0)) \\
& \quad \cdot (\hat{\mu}^n_{l+} - \mu^n_{l+}) \Big) dz \Big|.
\end{aligned}
$$

Splitting the error term and examining the first derivative yields for it the following majorization:

$$
\begin{aligned}
& \mathbf{1}_{\mathbf{N}}(\hat{\mu}^n_{j+} \cdot \hat{\mu}^n_{l+}) \cdot |\hat{\mu}^n_{j+} - \mu^n_{j+}| \\
& \cdot \sup_{z \in [0,1]} |D_1 g(\mu^n_{j+} + z(\hat{\mu}^n_{j+} - \mu^n_{j+}), \mu^n_{l+} + z(\hat{\mu}^n_{l+} - \mu^n_{l+}), \pi^n_{jk|C}(\theta_0), \pi^n_{lk|C}(\theta_0))| \\
= \; & \mathbf{1}_{\mathbf{N}}(\hat{\mu}^n_{j+} \cdot \hat{\mu}^n_{l+}) \cdot \frac{|\hat{\mu}^n_{j+} - \mu^n_{j+}|}{\mu^n_{j+}} \\
& \cdot \sup_{z \in [0,1]} \Big| \frac{\mu^n_{j+}}{\mu^n_{j+} + z(\hat{\mu}^n_{j+} - \mu^n_{j+})} \cdot (\mu^n_{j+} + z(\hat{\mu}^n_{j+} - \mu^n_{j+})) \\
& \qquad \cdot D_1 g(\mu^n_{j+} + z(\hat{\mu}^n_{j+} - \mu^n_{j+}), \mu^n_{l+} + z(\hat{\mu}^n_{l+} - \mu^n_{l+}), \pi^n_{jk|C}(\theta_0), \pi^n_{lk|C}(\theta_0)) \Big| \\
= \; & O_p(\frac{1}{\sqrt{\mu^n_{j+}}}).
\end{aligned}
$$

The last inequality follows from $\mathbf{1}_{\mathbf{N}}(\hat{\mu}^n_{j+}) \sup_{z \in [0,1]} |\frac{\mu^n_{j+}}{\mu^n_{j+} + z(\hat{\mu}^n_{j+} - \mu^n_{j+})}| = O_p(1)$ (compare Lemma 8.3), $(\hat{\mu}^n_{j+} - \mu^n_{j+})/\sqrt{\mu^n_{j+}} = O_p(1)$ and the boundedness of

$$
\begin{aligned}
& (\mu^n_{j+} + z(\hat{\mu}^n_{j+} - \mu^n_{j+})) \\
& \cdot D_1 g(\mu^n_{j+} + z(\hat{\mu}^n_{j+} - \mu^n_{j+}), \mu^n_{l+} + z(\hat{\mu}^n_{l+} - \mu^n_{l+}), \pi^n_{jk|C}(\theta_0), \pi^n_{lk|C}(\theta_0)) \\
= \; & (\mu^n_{j+} + z(\hat{\mu}^n_{j+} - \mu^n_{j+})) \\
& \cdot f(\mu^n_{l+} + z(\hat{\mu}^n_{l+} - \mu^n_{l+}), \pi^n_{lk|C}(\theta_0)) \cdot D_1 f(\mu^n_{j+} + z(\hat{\mu}^n_{j+} - \mu^n_{j+}), \pi^n_{jk|C}(\theta_0))
\end{aligned}
$$

for positive $\hat{\mu}^n_{j+}$ and $\hat{\mu}^n_{l+}$, which holds using the same arguments as in Lemma 8.3 and additionally, since a concrete function is considered, the bounding statements (6.45) and (6.46) adopted from Lemma 4.8. Analogously for the second derivative of the error term, the order $O_p(\frac{1}{\sqrt{\mu^n_{l+}}})$ is obtained and argumentation as in the proof of (6.40) resp. (6.49) yields

$$
\begin{aligned}
& g(\hat{\mu}^n_{j+}, \hat{\mu}^n_{l+}, \pi^n_{jk|C}(\theta_0), \pi^n_{lk|C}(\theta_0)) - g(\mu^n_{j+}, \mu^n_{l+}, \pi^n_{jk|C}(\theta_0), \pi^n_{lk|C}(\theta_0)) \\
= \; & O_p(\frac{1}{\sqrt{\mu^n_{j+}}}) + O_p(\frac{1}{\sqrt{\mu^n_{l+}}})
\end{aligned}
$$

for all j, l, k $(j \neq l)$. Further, including the case $j = l$ and using the assumptions concerning the marginal distribution, gives in conclusion 6.51 as follows:

$$
\begin{aligned}
& \frac{1}{J^n} \sum_{k=1}^{K} \frac{1}{\mu^n_{+k}} \Big(\sum_{j=1}^{J^n} \sum_{l=1}^{J^n} \big(f(\hat{\mu}^n_{j+}, \pi^n_{jk|C}(\theta_0)) f(\hat{\mu}^n_{l+}, \pi^n_{lk|C}(\theta_0)) \\
& \qquad\qquad - f(\mu^n_{j+}, \pi^n_{jk|C}(\theta_0)) f(\mu^n_{l+}, \pi^n_{lk|C}(\theta_0)) \big) \Big)
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{J^n}\sum_{k=1}^{K}\frac{1}{\mu^n_{+k}}\Big(\sum_{j=1}^{J^n}\sum_{\substack{l=1\\l\neq j}}^{J^n}\big(O_p(\frac{1}{\sqrt{\mu^n_{j+}}})+O_p(\frac{1}{\sqrt{\mu^n_{l+}}})\big)+\sum_{j=1}^{J^n}O_p(\frac{1}{\sqrt{\mu^n_{j+}}})\Big)\\
&= \sum_{k=1}^{K}\frac{1}{\mu^n_{+k}}\sum_{j=1}^{J^n}O_p(\frac{1}{\sqrt{\mu^n_{j+}}})\\
&= \frac{1}{J^n}\sum_{j=1}^{J^n}O_p(\frac{1}{\sqrt{\mu^n_{j+}}})\sum_{k=1}^{K}\frac{J^n}{\mu^n_{+k}}\\
&= o_p(1).
\end{aligned}
$$

□

All lemmata and theorems stated in this and the last chapter, in particular all conditions listed in section 2.3, now finally yield the following **main results** of this thesis, namely the asymptotic normality of the goodness–of–fit statistic under the nullhypothesis. In the next theorem, the result for Poisson and in the adjacent theorem that for column–multinomial sampling will be given. For these statements additionally the conditions concerning the marginal distribution will be needed, which have not been used so far. These are necessary to handle the error terms caused by the approximation steps concerning the nuisance parameters, i.e. by the transition from $\hat{\mu}_{\cdot+}$ to $\mu_{\cdot+}$.

Theorem 6.4 *Consider the Poisson distribution model and the asymptotics $n \to \infty$. Assume that the conditions (BC) and (VCP) hold, i.e. $\mu^n_{jk} \geq \epsilon > 0$ for all j, k, n and $\frac{J^n}{\sigma^{n2}_\lambda(\mu^n_{\cdot+},\theta_0)} = O(1)$. $\sigma^{n2}_\lambda(\mu^n_{\cdot+},\theta_0)$ is the variance given in Theorem 6.1:*

$$
\begin{aligned}
\sigma^{n2}_\lambda(\mu^n_{\cdot+},\theta_0) &= Var\Big(\sum_{j=1}^{J^n}\sum_{k=1}^{K}a_\lambda(X^n_{jk},\mu^n_{j+}\pi^n_{jk|C}(\theta_0))\Big)+2J^n\\
&+ \sum_{j=1}^{J^n}\frac{1}{\mu^n_{j+}}-2\sum_{j=1}^{J^n}\frac{1}{\mu^n_{j+}}\sum_{k=1}^{K}Cov(a_\lambda(X^n_{jk},\mu^n_{jk}),(X^n_{jk})^2)\\
&+ 4\sum_{j=1}^{J^n}\sum_{k=1}^{K}\pi^n_{jk|C}(\theta_0)Cov(a_\lambda(X^n_{jk},\mu^n_{jk}),X^n_{jk})\\
&- c^n_\lambda(\mu^n_{\cdot+},\theta_0)(I^n(\mu^n_{\cdot+},\theta_0))^{-1}(c^n_\lambda(\mu^n_{\cdot+},\theta_0))^T
\end{aligned}
$$

with $c^n_\lambda(\mu^n_{\cdot+},\theta_0) = \sum_{j=1}^{J^n}\sum_{k=1}^{K}D_\theta\log\pi^n_{jk|C}(\theta_0)Cov(a_\lambda(X^n_{jk},\mu^n_{jk}),X^n_{jk})$ as defined in (5.4). Suppose the sequence of estimators $\hat{\theta}^n$ accomplishes the conditions $\sqrt{n}(\hat{\theta}^n - \theta_0) = O_p(1)$ (LC2), $(\hat{\theta}^n-\theta_0) = (I^n(\mu^n_{\cdot+},\theta_0))^{-1}U^n(\theta_0|X^n)+O_p(\frac{1}{n})$ (LC3) with $\frac{1}{n}I^n(\mu^n_{\cdot+},\theta_0)\to I_\infty$ positive definite (LC1). Further assume that the conditions (MD1) *and (MD2) concerning the marginal distribution are met:*

$$\sum_{j=1}^{J^n} \sqrt{\mu_{j+}^n} = o(\sqrt{J^n n}), \qquad \sum_{j=1}^{J^n} \frac{1}{\sqrt{\mu_{j+}^n}} = o(\sqrt{J^n}).$$

With m_λ^n being the Poisson expectation of SD_λ^n, $m_\lambda^n(\mu_{\cdot+}^n, \theta_0) = E(SD_\lambda^n(\mu_{\cdot+}^n, \theta_0|X^n))$, then under the nullhypothesis H_0, i.e. $\mu_{jk}^n = \mu_{j+}^n \pi_{jk|C}^n(\theta_0)$ for all j, k, n, holds

$$\frac{SD_\lambda^n(\hat\mu_{\cdot+}^n, \hat\theta^n|X^n) - m_\lambda^n(\hat\mu_{\cdot+}^n, \hat\theta^n) + J^n}{\sigma_\lambda^n(\hat\mu_{\cdot+}^n, \hat\theta^n)} \xrightarrow{\mathcal{L}} N(0,1).$$

Proof:
Using condition (LC2), (LC3) and (BC), Corollary 5.9, which summarizes the single reduction steps of chapter 5, gives for the centered test statistic the approximation

$$\begin{aligned}
& SD_\lambda^n(\hat\mu_{\cdot+}^n, \hat\theta^n|X^n) - m_\lambda^n(\hat\mu_{\cdot+}^n, \hat\theta^n) + J^n \\
= \; & SD_\lambda^n(\mu_{\cdot+}^n, \theta_0|X^n) - \sum_{j=1}^{J^n} \frac{(X_{j+}^n - \mu_{j+}^n)^2}{\mu_{j+}^n} - c_\lambda^n(\mu_{\cdot+}^n, \theta_0)(I^n(\mu_{\cdot+}^n, \theta_0))^{-1} U^n(\theta_0|X^n) \\
& - m_\lambda^n(\mu_{\cdot+}^n, \theta_0) + J^n + O_p\Big(\sum_{j=1}^{J^n} \sqrt{\frac{1}{\mu_{j+}^n}}\Big) + O_p\Big(\sum_{j=1}^{J^n} \sqrt{\frac{\mu_{j+}^n}{n}}\Big) + O_p(1) \\
= \; & \Psi_{\lambda+}^n(X^n) - E(\Psi_{\lambda+}^n(X^n)) + O_p\Big(\sum_{j=1}^{J^n} \sqrt{\frac{1}{\mu_{j+}^n}}\Big) + O_p\Big(\sum_{j=1}^{J^n} \sqrt{\frac{\mu_{j+}^n}{n}}\Big) + O_p(1)
\end{aligned}$$

with

$$\Psi_{\lambda+}^n(X^n) = SD_\lambda^n(\mu_{\cdot+}^n, \theta_0|X^n) - \sum_{j=1}^{J^n} \frac{(X_{j+}^n - \mu_{j+}^n)^2}{\mu_{j+}^n} - c_\lambda^n(\mu_{\cdot+}^n, \theta_0)(I^n(\mu_{\cdot+}^n, \theta_0))^{-1} U^n(\theta_0|X^n).$$

In order to meet assumption (MD2), now $J^n/n \to 0$ or, even stronger, $(J^n)^2/n \to 0$ is required, as seen in section 2.3. Condition $J^n/n \to 0$ can also be deduced from (BC) and (MD1), since it holds

$$\sqrt{\frac{J^n}{n} \cdot K\epsilon} = \sqrt{\frac{1}{J^n \cdot n}} \sum_{j=1}^{J^n} \sqrt{K\epsilon} \le \sqrt{\frac{1}{J^n \cdot n}} \sum_{j=1}^{J^n} \sqrt{\mu_{j+}^n} = \frac{1}{\sqrt{J^n}} \sum_{j=1}^{J^n} \sqrt{\frac{\mu_{j+}^n}{n}} \longrightarrow 0. \qquad (6.55)$$

This, i.e. $\frac{J^n}{n} \to 0$, gives $\frac{J^n}{\mu_{++}^n} \to 0$, which, combined with $\sum_{j=1}^{J^n} \sum_{k=1}^{K} (\mu_{jk}^n)^2 \le (\mu_{++}^n)^2$, then asserts the condition concerning the marginal distribution required for Theorem 6.1, namely

$$\frac{(J^n)^2 \cdot \sum_{j=1}^{J^n} \sum_{k=1}^{K} (\mu_{jk}^n)^2}{(\mu_{++}^n)^4} = \Big(\frac{J^n}{\mu_{++}^n}\Big)^2 \cdot \frac{\sum_{j=1}^{J^n} \sum_{k=1}^{K} (\mu_{jk}^n)^2}{(\mu_{++}^n)^2} = o(1).$$

Further using (LC1), Theorem 6.1 thus gives the asymptotic normality of the standardization:

$$\frac{\left(\Psi^n_{\lambda+}(X^n) - E(\Psi^n_{\lambda+}(X^n))\right)}{\sqrt{Var(\Psi^n_{\lambda+}(X^n))}} \xrightarrow{\mathcal{L}} N(0,1) \quad \text{with} \quad Var(\Psi^n_{\lambda+}(X^n)) = \sigma^{n2}_\lambda(\mu^n_{\cdot+}, \theta_0).$$

The consistency of the variance estimation $\sigma^{n2}_\lambda(\hat\mu^n_{\cdot+}, \hat\theta^n)$ is provided in Lemma 6.3, which uses (LC1), (BC), (VCP) and $\frac{1}{J^n}\sum_{j=1}^{J^n} \frac{1}{\sqrt{\mu^n_{j+}}} \longrightarrow 0$. The last condition is fulfilled because of the stronger assumption (MD2).
These statements, combined with the variance condition $J^n/\sigma^{n2}_\lambda(\mu^n_{\cdot+}, \theta_0) = O(1)$ and the assumptions concerning the marginal distribution (MD1) and (MD2), finally give the result as follows:

$$\begin{aligned}
&\frac{SD^n_\lambda(\hat\mu^n_{\cdot+}, \hat\theta^n | X^n) - m^n_\lambda(\hat\mu^n_{\cdot+}, \hat\theta^n) + J^n}{\sigma^n_\lambda(\hat\mu^n_{\cdot+}, \hat\theta^n)} \\
=\; & \frac{\Psi^n_{\lambda+}(X^n) - E(\Psi^n_{\lambda+}(X^n))}{\sigma^n_\lambda(\mu^n_{\cdot+}, \theta_0)} \cdot \frac{\sigma^n_\lambda(\mu^n_{\cdot+}, \theta_0)}{\sigma^n_\lambda(\hat\mu^n_{\cdot+}, \hat\theta^n)} \\
& + \frac{O_p(\sum_{j=1}^{J^n}(\frac{\mu^n_{j+}}{n})^{1/2}) + O_p(\sum_{j=1}^{J^n}(\frac{1}{\mu^n_{j+}})^{1/2}) + O_p(1)}{\sqrt{J^n}} \cdot \frac{\sqrt{J^n}}{\sigma^n_\lambda(\mu^n_{\cdot+}, \theta_0)} \cdot \frac{\sigma^n_\lambda(\mu^n_{\cdot+}, \theta_0)}{\sigma^n_\lambda(\hat\mu^n_{\cdot+}, \hat\theta^n)} \\
=\; & \left(\frac{\Psi^n_{\lambda+}(X^n) - E(\Psi^n_{\lambda+}(X^n))}{\sigma^n_\lambda(\mu^n_{\cdot+}, \theta_0)} + o_p(1)\cdot O(1)\right) \cdot \frac{\sigma^n_\lambda(\mu^n_{\cdot+}, \theta_0)}{\sigma^n_\lambda(\hat\mu^n_{\cdot+}, \hat\theta^n)} \\
& \xrightarrow{\mathcal{L}} N(0,1).
\end{aligned}$$

□

As a second conclusion from the preceding results, and as the main result for column–multinomial distribution, now the asymptotic normality of the test statistic under the nullhypothesis is given. Apart from the approximation, the most important statement for the following theorem is the asymptotic normality of the approximated statistic, given in Theorem 6.2, which was derived using Morris' approach and hence, of course, leads to standardization terms being Poisson.

Theorem 6.5 *Now assume column–multinomial sampling and for the asymptotics $n \to \infty$ the conditions (BC) and (VCC), i.e. $\mu^n_{jk} \geq \epsilon$ for all j, k, n ($\epsilon > 0$ constant) and $J^n/s^{n2}_\lambda(\mu^n_{\cdot+}, \theta_0) = O(1)$. The Poisson variance s^{n2}_λ is defined as in Theorem 6.1:*

$$s^{n2}_\lambda(\mu^n_{\cdot+}, \theta_0) = \sigma^{n2}_\lambda(\mu^n_{\cdot+}, \theta_0) - \sum_{k=1}^{K} \mu^n_{+k}(\gamma^n_{\lambda k}(\mu^n_{\cdot+}, \theta_0))^2$$

with correction terms $\gamma^n_{\lambda k}(\mu^n_{\cdot+}, \theta_0) = \frac{1}{\mu^n_{+k}} \sum_{j=1}^{J^n}(Cov(a_\lambda(X^n_{jk}, \mu^n_{jk}), X^n_{jk}) - \pi^n_{jk|C}(\theta_0))$ and known column sizes $n_k = \mu^n_{+k}$ for $k = 1, \ldots, K$. The approximative variance

$\sigma_\lambda^{n2}(\mu_{\cdot+}^n, \theta_0)$ *of the Poisson statistic is just as in the preceding theorem (Theorem 6.4) resp. Theorem 6.1. For the parameter estimators* $\hat{\theta}^n$ *assume (LC2) and (LC3), i.e.* $\sqrt{n}(\hat{\theta}^n - \theta_0) = O_p(1)$, $\hat{\theta}^n - \theta_0 = (I^n(\mu_{\cdot+}^n, \theta_0))^{-1} U^n(\theta_0|Y^n) + O_p(\frac{1}{n})$, *with (LC1)* $\frac{1}{n} I^n(\mu_{\cdot+}^n, \theta_0) \to I_\infty$ *positive definite. Further let (MD1), (MD2) and in particular (MD3) be fulfilled:*

$$\sum_{j=1}^{J^n} \sqrt{\mu_{j+}^n} = o(\sqrt{J^n n}), \quad \sum_{j=1}^{J^n} \sqrt{\frac{1}{\mu_{j+}^n}} = o(\sqrt{J^n}), \quad \max_{1 \le j \le J^n} \pi_{jk|D}^n(\theta_0) \longrightarrow 0 \text{ for every } k.$$

Then the test statistic has a normal limit under the nullhypothesis, i.e. $\mu_{jk}^n = \mu_{j+}^n \cdot \pi_{jk|C}^n(\theta_0)$ *for all* j, k, n, *as follows* (m_λ^n *is the Poisson expectation of* SD_λ^n):

$$\frac{SD_\lambda^n(\hat{\mu}_{\cdot+}^n, \hat{\theta}^n|Y^n) - m_\lambda^n(\hat{\mu}_{\cdot+}^n, \hat{\theta}^n) + J^n}{s_\lambda^n(\hat{\mu}_{\cdot+}^n, \hat{\theta}^n)} \xrightarrow{\mathcal{L}} N(0,1).$$

Proof:
The essential arguments to deduct the limiting normality of the test statistic from the preceding theorems are clearly completely analogous to the proof of Theorem 6.4 concerning Poisson sampling. Corollary 5.9 provides the approximation as follows

$$\begin{aligned} & SD_\lambda^n(\hat{\mu}_{\cdot+}^n, \hat{\theta}^n|Y^n) - m_\lambda^n(\hat{\mu}_{\cdot+}^n, \hat{\theta}^n) + J^n \\ = \; & SD_\lambda^n(\mu_{\cdot+}^n, \theta_0|Y^n) - \sum_{j=1}^{J^n} \frac{(Y_{j+}^n - \mu_{j+}^n)^2}{\mu_{j+}^n} - c_\lambda^n(\mu_{\cdot+}^n, \theta_0)(I^n(\mu_{\cdot+}^n, \theta_0))^{-1} U^n(\theta_0|Y^n) \\ & - m_\lambda^n(\mu_{\cdot+}^n, \theta_0) + J^n + O_p\Big(\sum_{j=1}^{J^n} \sqrt{\frac{1}{\mu_{j+}^n}}\Big) + O_p\Big(\sum_{j=1}^{J^n} \sqrt{\frac{\mu_{j+}^n}{n}}\Big) + O_p(1) \\ = \; & \Psi_{\lambda+}^n(Y^n) - E(\Psi_{\lambda+}^n(X^n)) + O_p\Big(\sum_{j=1}^{J^n} \sqrt{\frac{1}{\mu_{j+}^n}}\Big) + O_p\Big(\sum_{j=1}^{J^n} \sqrt{\frac{\mu_{j+}^n}{n}}\Big) + O_p(1) \end{aligned}$$

with $E(\Psi_{\lambda+}^n(X^n)) = m_\lambda^n(\mu_{\cdot+}^n, \theta_0) - J^n$. For the centered approximation $\Psi_{\lambda+}^n(Y^n) - E(\Psi_{\lambda+}^n(X^n))$, scaled with the (Poisson–) standard deviation $s_\lambda^n(\mu_{\cdot+}^n, \theta_0)$, in Theorem 6.2 the limiting normality has been shown using Morris' method (1975), and, in contrast to the preceding theorem, especially condition (MD3). Lemma 6.3 gives the consistency of the variance estimation and (MD1) and (MD2) guarantee the error terms to be of the order $o_p(\sqrt{J^n})$. These arguments combined with variance condition $J^n / s_\lambda^{n2}(\mu_{\cdot+}^n, \theta_0) = O(1)$ (VCC) now give the result:

$$\begin{aligned} & \frac{SD_\lambda^n(\hat{\mu}_{\cdot+}^n, \hat{\theta}^n|Y^n) - m_\lambda^n(\hat{\mu}_{\cdot+}^n, \hat{\theta}^n) + J^n}{s_\lambda^n(\hat{\mu}_{\cdot+}^n, \hat{\theta}^n)} \\ = \; & \left(\frac{\Psi_{\lambda+}^n(Y^n) - E(\Psi_{\lambda+}^n(X^n))}{s_\lambda^n(\mu_{\cdot+}^n, \theta_0)} + \frac{o_p(\sqrt{J^n})}{s_\lambda^n(\mu_{\cdot+}^n, \theta_0)} \right) \cdot \frac{s_\lambda^n(\mu_{\cdot+}^n, \theta_0)}{s_\lambda^n(\hat{\mu}_{\cdot+}^n, \hat{\theta}^n)} \end{aligned}$$

$$= \left(\frac{\Psi^n_{\lambda+}(Y^n) - E(\Psi^n_{\lambda+}(X^n))}{s^n_\lambda(\mu^n_{\cdot+}, \theta_0)} + o_p(1)\right) \cdot \frac{s^n_\lambda(\mu^n_{\cdot+}, \theta_0)}{s^n_\lambda(\hat{\mu}^n_{\cdot+}, \hat{\theta}^n)}$$
$$\xrightarrow{\mathcal{L}} N(0,1).$$

□

Finally, it should be mentioned that the variance conditions (VCP) and (VCC), which, for reasons of clarity, have been presumed for the main theorems, Th. 6.4 and Th. 6.5, can certainly be relaxed. Considering the derived variance for column–multinomial sampling first,

$$s^{n2}_\lambda(\mu^n_{\cdot+}, \theta_0) = \sigma^{n2}_\lambda(\mu^n_{\cdot+}, \theta_0) - \sum_{k=1}^{K} \mu^n_{+k}(\gamma^n_{\lambda k}(\mu^n_{\cdot+}, \theta_0))^2,$$

it is easily seen that the second term disappears in the limit. Theorem 6.1 gives $\gamma^n_{\lambda k}(\mu^n_{\cdot+}, \theta_0) = O(\frac{J^n}{\mu^n_{+k}})$ for each k and hence

$$\sum_{k=1}^{K} \mu^n_{+k}(\gamma^n_{\lambda k}(\mu^n_{\cdot+}, \theta_0))^2 = \sum_{k=1}^{K} O(\frac{(J^n)^2}{\mu^n_{+k}}).$$

Since condition (MD2), i.e. $\sum_{j=1}^{J^n} \sqrt{\frac{1}{\mu^n_{j+}}} = o(\sqrt{J^n})$, requires $\frac{(J^n)^2}{n} = \frac{(J^n)^2}{\mu^n_{++}} \to 0$ as seen in section 2.3, and the regularity condition (RC2) further implies $\frac{\mu^n_{++}}{\mu^n_{+k}} = O(1)$ for each k (see (6.6), proof of Theorem 6.1), it follows

$$\frac{(J^n)^2}{\mu^n_{+k}} = \frac{(J^n)^2}{\mu^n_{++}} \cdot \frac{\mu^n_{++}}{\mu^n_{+k}} = o(1),$$

and hence the stated zero convergence. Condition (VCC) might thus be replaced by condition (VCP) concerning σ^{n2}_λ.

Referring to the variance $\sigma^{n2}_\lambda(\mu^n_{\cdot+}, \theta_0)$ concerning both distribution models and using $\frac{(J^n)^2}{n} \to 0$ again, the correction term $c^n_\lambda(\mu^n_{\cdot+}, \theta_0) \cdot (I^n(\mu^n_{\cdot+}, \theta_0))^{-1} \cdot (c^n_\lambda(\mu^n_{\cdot+}, \theta_0))^T$ which has the order $O(\frac{(J^n)^2}{n})$ (see e.g. proof of Lemma 6.3) turns out to disappear asymptotically, too. Hence, it would also be possible to formulate a variance condition directly for the first part of σ^{n2}_λ, which merely concerns the variance of $SD^n_\lambda(\mu^n_{\cdot+}, \theta_0|X^n) - \sum_{j=1}^{J^n} a_1(X^n_{j+}, \mu^n_{j+})$.

Now, as a final example for Theorem 6.4 and Theorem 6.5, the derived formulae for Pearson's X^2 Statistic will be given.

Example 6.6
For Pearson's X^2 Statistic $SD^n_1(\mu^n_{\cdot+}, \theta_0)$ $(\lambda = 1)$ with especially $SD^n_1(\mu^n_{\cdot+}, \theta_0|X^n) =$

$\sum_{j=1}^{J^n}\sum_{k=1}^{K}\frac{(X_{jk}^n-\mu_{jk}^n)^2}{\mu_{jk}^n}$ in the Poisson and $SD_1^n(\mu_{\cdot+}^n,\theta_0|Y^n)=\sum_{j=1}^{J^n}\sum_{k=1}^{K}\frac{(Y_{jk}^n-\mu_{jk}^n)^2}{\mu_{jk}^n}$ in the column–multinomial distribution model, the standardization terms can be stated explicitly:

$$
\begin{aligned}
m_1^n(\mu_{\cdot+}^n,\theta_0)-J^n &= J^nK-J^n,\\
\sigma_1^{n2}(\mu_{\cdot+}^n,\theta_0) &= 2J^nK+\sum_{j=1}^{J^n}\sum_{k=1}^{K}\frac{1}{\mu_{jk}^n}+2J^n+\sum_{j=1}^{J^n}\frac{1}{\mu_{j+}^n}-2(K\sum_{j=1}^{J^n}\frac{1}{\mu_{j+}^n}+2J^n)\\
&\quad -c_1^n(\theta_0)(I^n(\mu_{\cdot+}^n,\theta_0))^{-1}(c_1^n(\theta_0))^T\\
&= 2J^n(K-1)+\sum_{j=1}^{J^n}\frac{1}{\mu_{j+}^n}(\sum_{k=1}^{K}\frac{1}{\pi_{jk|C}^n(\theta_0)}+1-2K)\\
&\quad -c_1^n(\theta_0)(I^n(\mu_{\cdot+}^n,\theta_0))^{-1}(c_1^n(\theta_0))^T,\\
s_1^{n2}(\mu_{\cdot+}^n,\theta_0) &= \sigma_1^{n2}(\mu_{\cdot+}^n,\theta_0)-\sum_{k=1}^{K}\mu_{+k}^n(\gamma_{\lambda k}^n(\mu_{\cdot+}^n,\theta_0))^2
\end{aligned}
$$

with

$$
\begin{aligned}
c_1^n(\theta_0) &= \sum_{j=1}^{J^n}\sum_{k=1}^{K}D_\theta\log\pi_{jk|C}^n(\theta_0),\\
\gamma_{\lambda k}^n(\mu_{\cdot+}^n,\theta_0) &= \frac{1}{\mu_{+k}^n}(J-\sum_{j=1}^{J^n}\pi_{jk|C}^n(\theta_0))\quad\text{for each } k.
\end{aligned}
$$

Hence the relevant test statistic for Poisson sampling is

$$
\frac{1}{\sigma_1^n(X_{\cdot+}^n,\theta_0)}\Big(\sum_{j=1}^{J^n}\sum_{k=1}^{K}\frac{(X_{jk}^n-X_{j+}^n\pi_{jk|C}^n(\hat\theta^n))^2}{X_{j+}^n\pi_{jk|C}^n(\hat\theta^n)}-J^n(K-1)\Big)
$$

with

$$
\begin{aligned}
\sigma_1^n(X_{\cdot+}^n,\theta_0) &= \Big(2J^n(K-1)+\sum_{j=1}^{J^n}\frac{1}{X_{j+}^n}(\sum_{k=1}^{K}\frac{1}{\pi_{jk|C}^n(\hat\theta^n)}+1-2K)\\
&\quad -c_1^n(\hat\theta^n)(I^n(X_{\cdot+}^n,\hat\theta^n))^{-1}(c_1^n(\hat\theta^n))^T\Big)^{\frac{1}{2}}
\end{aligned}
$$

and

$$
c_1^n(\hat\theta^n)=\sum_{j=1}^{J^n}\sum_{k=1}^{K}D_\theta\log\pi_{jk|C}^n(\hat\theta^n).
$$

Similarly for column–multinomial sampling, where especially $\mu^n_{+k} = n_k$ is the known column size for every k, the derived statistics states as follows:

$$\frac{1}{s^n_1(Y^n_{\cdot+},\theta_0)}\Big(\sum_{j=1}^{J^n}\sum_{k=1}^{K}\frac{(Y^n_{jk}-Y^n_{j+}\pi^n_{jk|C}(\hat{\theta}^n))^2}{Y^n_{j+}\pi^n_{jk|C}(\hat{\theta}^n)} - J^n(K-1)\Big)$$

using

$$\begin{aligned}
s^n_1(Y^n_{\cdot+},\theta_0) &= \Big(\sigma^n_1(Y^n_{\cdot+},\theta_0) - \sum_{k=1}^{K}\mu^n_{+k}(\gamma^n_{\lambda k}(Y^n_{\cdot+},\theta_0))^2\Big)^{\frac{1}{2}}\\
&= \Big(\sigma^n_1(Y^n_{\cdot+},\theta_0) - \sum_{k=1}^{K}\frac{1}{\mu^n_{+k}}(J-\sum_{j=1}^{J^n}\pi^n_{jk|C}(\theta_0))^2\Big)^{\frac{1}{2}},\\
\sigma^n_1(Y^n_{\cdot+},\theta_0) &= \Big(2J^n(K-1)+\sum_{j=1}^{J^n}\frac{1}{Y^n_{j+}}(\sum_{k=1}^{K}\frac{1}{\pi^n_{jk|C}(\hat{\theta}^n)}+1-2K)\\
&\quad -c^n_1(\hat{\theta}^n)(I^n(Y^n_{\cdot+},\hat{\theta}^n))^{-1}(c^n_1(\hat{\theta}^n))^T\Big)^{\frac{1}{2}}.
\end{aligned}$$

Here the centering term $m_1(\mu^n_{\cdot+},\theta_0) - J^n = J^n(K-1)$ coincides with the column–multinomial expectation $E(SD^n_1(\mu^n_{\cdot+},\theta_0|Y^n))$ of the pure Pearson Statistic.

Referring to the preceding discussion of the variance, and since the first part of σ^{n2}_1, i.e. $2J^n(K-1)+\sum_{j=1}^{J^n}\frac{1}{\mu^n_{j+}}(\sum_{k=1}^{K}\frac{1}{\pi^n_{jk|C}(\theta_0)}+1-2K)$, has obviously the exact order J^n, in this case ($\lambda = 1$) the variance conditions (VCC) respectively (VCP) are met.

□

7. Final Remarks

In conclusion, it may be said that in this dissertation, through the consideration of an "increasing cells" approach, goodness–of–fit tests for Poisson and column–multinomial sampling were derived, which meet the common situation when data are sparse. Since in these distribution models the marginal distribution of the (covariable) groups is not given, nor is it fixed by modelling, here the case of asymptotically infinite many nuisance was treated for the first time in this context. However, as already discussed in chapter 2, the deduced tests suffer in view of applications. This especially affects column–multinomial distributed contingency tables (case–control studies), since, as the parametric models in question, only quite generally the ratios

$$\pi^n_{jk|C}(\theta) = \frac{\mu^n_{jk}(\theta)}{\mu^n_{j+}} \quad \text{respectively} \quad \mu^n_{jk}(\theta) = \mu^n_{j+}\pi^n_{jk|C}(\theta),\ \theta \in \mathbf{R}^S$$

($j = 1, \ldots, J^n$, $k = 1, \ldots, K$, $n \in \mathbf{N}$) are considered, which have turned out to be inadequate for this sampling scheme. As seen by the discussion of a different parametrization, the so–called "odds–ratio" models, which for example contain the well–known logistic regression models and only model the dependencies within contingency tables, do not fit, in the case of column–multinomial sampling, into the model class considered here. An exception can only be made if further restrictive assumptions are fulfilled, which guarantee that the row expectations μ^n_{j+} and the modelled ratios $\pi^n_{jk|C}(\theta)$ do not depend on each other. This property is required for the approach considered here; in the case of case–control studies, because of the given column sizes, it is in general not met. In regard of this important application and in order to allow utmost variability for the marginal distribution — which is usually of no interest — it is reasonable for further efforts on this subject to consider the parametrization and models suggested in chapter 2 (see (2.6), sec. 2.2), namely

$$\pi^n_{jk|C} = \pi^n_{jk|C}(\theta^n) = \frac{\exp(\gamma^n_k + \psi^n_{jk}(\beta))}{\sum_{l=1}^K \exp(\gamma^n_l + \psi^n_{lk}(\beta))}.$$

Here only the odds–ratios $\psi^n_{jk} = \psi^n_{jk}(\beta)$ are modelled, and, in particular, no assumptions concerning the marginal distribution are made. An approach to derive the limiting distribution of the family of test statistics SD_λ ($\lambda > -1$) for this model class in analogy to this thesis, would especially require a modification of the approximation steps in chapter 5. The results from chapter 6, where the asymptotic normality of the approximated statistics was shown — for column–multinomial sampling with Morris' (1975) method — then could be adopted without major changes. Since this chapter treats the true and not the fitted expectations, the essential arguments are just the same. Hence, although the tests deduced for column–multinomial sampling are of rather theoretical interest, the proceeding of this thesis nevertheless provides a method for further proofs concerning this distribution model.

Beside the advisable extension of the results for column–multinomials to the consideration of pure odds–ratio models, it would certainly also be desirable to tackle this systematically for conditional Poisson models, as there are Poisson, multinomial, row–multinomial, column–multinomial and hypergeometric distribution as the best known representatives. These distribution models underly most epidemiological studies and hence are of major interest. An approach with odds–ratio models would in particular also affect the results for (unconditional) Poisson sampling derived here, where some slight generalizations would have to be taken up. Hence, as continuation and generalization of this thesis, the next problem to tackle is the derivation of the limiting distribution of SD_λ for models which specify the odds–ratios *only* and for sampling schemes being conditional Poisson. In order to obtain information about the power of the deduced tests, the limiting distribution should not only be derived under the nullhypothesis, i.e. the model holds, but also under the alternative. Statements concerning the distribution of SD_λ for the increasing cells approach under the alternative hypothesis are hitherto available only for row–multinomial sampling (Dieter Rojek (1989)). In contrast to conditional Poisson models in general, this special case is relatively simple to handle, since the marginal distribution of the covariable groups is fixed. Hence Rojek needs not deal with an infinite number of nuisance parameters. Moreover, he investigates this distribution model directly and does not touch the aspect "conditional Poisson" at all.

Subsequently, when the distribution of SD_λ under both null– and alternative hypothesis is given, the goodness of the tests deduced should be discussed. For this purpose, the power has to be studied and simulations in general are necessary. In this context, it would certainly also be of interest to investigate the role of the parameter λ, especially which λ should be preferred — perhaps contingent on possible situations — in order to obtain the most reliable results. To improve the goodness of the tests, it would further also be desirable to derive higher order approximations such as edgeworth– and saddlepoint–approximations as carried out by Osius (1994) for the row–multinomial case.

Finally, as well as for the results deduced here as for the intended results concerning pure odds–ratio models, the aspect of applicability has to be discussed in greater detail. Here several open questions arise, which require clarification: How can concrete designs be embedded in the asymptotics considered here, so that the mathematical assumptions are met? In which situations should the tests not be used? Are in these cases alternatives available and, if yes, which?

For an illustration of the theoretical results, the derived goodness–of–fit tests should, of course, also be applied to concrete data sets. Most interesting for this purpose are doubtless case–control studies, which are frequently performed in practice and often meet situations where data are sparse. However, since for this sampling scheme and the models of interest the mathematical theory is still incomplete and not trivial at all, one will probably have to wait another while until a suitable test can be realized for this important application.

8. Appendix

8.1 Central Moments of the Poisson Distribution

Consider the r–th central moments of the Poisson distribution with parameter μ

$$\nu_r(\mu) = \sum_{1 \le i \le r/2} a_{r,i}\mu^i \quad (r \in \mathbf{N}_0).$$

By Theorem 4.3 then, for the coefficients $a_{r,i}$ holds $a_{0,0} = 1$, $a_{r,0} = 0$ $(r \ge 1)$ and the following recursion formula

$$a_{r,i} = (r-1) \cdot a_{r-2,i-1} + i \cdot a_{r-1,i} \in \mathbf{N}$$

for $r \ge 2$, $1 \le i \le r/2$ ($a_{r,i} := 0$ for $i > r/2$).

Tabulation of the coefficients for $2 \le r \le 10$ ($a_{r,0} = 0$):

r	$a_{r,1}$	$a_{r,2}$	$a_{r,3}$	$a_{r,4}$	$a_{r,5}$
2	1				
3	1				
4	1	3			
5	1	10			
6	1	25	15		
7	1	56	105		
8	1	119	490	105	
9	1	246	1918	1260	
10	1	501	6825	9450	945.

8.2 Technical Results

Lemma 8.1 *Consider $x, y \in \mathbf{N}_0$, $\mu \in \mathbf{R}^+$ and the distance function a_λ defined in Def. 2.1. Then for $\lambda \in (-1,1]$ holds:*

a)

$$|a_\lambda(x+y,\mu) - a_\lambda(x,\mu)| \le 2\Big(\frac{y^2}{\mu} + y \cdot h(x,\mu)\Big)$$

with $h(x,\mu) := |x-\mu|\left(\frac{1}{\mu} + \frac{c}{x+1}\right)$, $c := \max\{2, \frac{1}{\lambda+1}\}$,

b)

$$\frac{2}{\lambda+1}\mu\Big(\Big(\frac{x+y}{\mu}\Big)^{\lambda+1} - \Big(\frac{x}{\mu}\Big)^{\lambda+1}\Big) \le c\Big(\frac{y^2}{\mu} + y \cdot h(x,\mu)\Big)$$

with $h(x,\mu) := 1 + \frac{x}{\mu} + \frac{2\mu}{x+1}$, $c := \max\{2, \frac{2}{\lambda+1}\}$.

Proof:

a) The result obviously holds for $x = y = 0$ and $x \ge 1, y = 0$, so only the cases $x \ge 1, y \ge 1$ and $x = 0, y \ge 1$ have to be studied. The following simple inequalities will be used repeatedly:

$$\log z \le z - 1 \text{ for all } z > 0, \tag{8.1}$$

$$\frac{1}{\lambda}(z^\lambda - 1) \le z - 1 \text{ for all } \lambda \in (0,1) \text{ and } z > 0. \tag{8.2}$$

(8.1) is well known and (8.2) resp. $z^\lambda \le \lambda(z-1)+1$ holds because z^λ is concave for $\lambda \in (0,1)$ and hence every value of z^λ lies beneath the tangent $\lambda z + 1 - \lambda = \lambda(z-1)+1$ of z^λ in $z = 1$.

Considering $\mathbf{x \ge 1, y \ge 1}$, application of the mean value theorem to the term of interest yields

$$\frac{1}{2}|a_\lambda(x+y,\mu) - a_\lambda(x,\mu)| = \frac{1}{2}y \cdot |D_1 a_\lambda(x+ty,\mu)| = \begin{cases} y\Big|\frac{1}{\lambda}\Big(\Big(\frac{x+ty}{\mu}\Big)^\lambda - 1\Big)\Big| & (\lambda \ne 0) \\ y\Big|\log\frac{x+ty}{\mu}\Big| & (\lambda = 0) \end{cases} \tag{8.3}$$

for some $t \in (0,1)$ (derivatives of a_λ see Lemma 2.2).
If $x + ty \ge \mu$, the mean values are dominated by the term belonging to $\lambda = 1$, i.e. $y(\frac{x+ty}{\mu} - 1)$. Using (8.1) resp. the equivalent statement $-z^{-\lambda} \le -\log z^{-\lambda} - 1 \Leftrightarrow \log z^{-\lambda} \le z^{-\lambda} - 1$, this can, for $\lambda \in (-1,0)$, be seen as follows

$$\begin{aligned} y \cdot \frac{1}{\lambda}\Big(\Big(\frac{x+ty}{\mu}\Big)^\lambda - 1\Big) &= y \cdot \frac{1}{-\lambda}\Big(1 - \Big(\frac{\mu}{x+ty}\Big)^{-\lambda}\Big) \\ &\le y \cdot \frac{1}{-\lambda}\Big(1 - \log\Big(\frac{\mu}{x+ty}\Big)^{-\lambda} - 1\Big) \end{aligned}$$

$$\begin{aligned} &= y \cdot \frac{1}{-\lambda} \cdot \lambda \log \frac{\mu}{x+ty} \\ &= y \log \frac{x+ty}{\mu} \\ &\le y\Big(\frac{x+ty}{\mu} - 1\Big). \end{aligned}$$

The penultimate term, i.e. $y \log \frac{x+ty}{\mu}$, equals the case $\lambda = 0$, thus only $\lambda \in (0,1)$ remains to be considered. For these values, however, (8.2) immediately implies

$$y \cdot \frac{1}{\lambda}\Big(\Big(\frac{x+ty}{\mu}\Big)^{\lambda} - 1\Big) \le y\Big(\frac{x+ty}{\mu} - 1\Big).$$

For $y\Big(\frac{x+ty}{\mu} - 1\Big)$ now the desired inequality is given:

$$y\Big(\frac{x+ty}{\mu} - 1\Big) \le y\Big(\frac{x+y}{\mu} - \frac{\mu}{\mu}\Big) = y\Big(\frac{y}{\mu} + (x-\mu)\frac{1}{\mu}\Big) \le \frac{y^2}{\mu} + y|x-\mu|\Big(\frac{1}{\mu} + \frac{c}{x+1}\Big).$$

In case of $x + ty < \mu$ the expressions in (8.3) are as follows:

$$\begin{aligned} y\Big|\frac{1}{\lambda}\Big(\Big(\frac{x+ty}{\mu}\Big)^{\lambda} - 1\Big)\Big| &= y\frac{1}{\lambda}\Big(1 - \Big(\frac{x+ty}{\mu}\Big)^{\lambda}\Big) && (\lambda \neq 0), \\ y\Big|\log \frac{x+ty}{\mu}\Big| &= y \log \frac{\mu}{x+ty} && (\lambda = 0). \end{aligned}$$

At first, the case $\lambda \in (0,1]$ can be traced back to $\lambda = 0$, since by (8.1) holds $\log z^{\lambda} \le z^{\lambda} - 1 \Leftrightarrow 1 - z^{\lambda} \le -\log z^{\lambda} = \lambda \log \frac{1}{z} \Leftrightarrow \frac{1}{\lambda}(1 - z^{\lambda}) \le \log \frac{1}{z}$ and hence,

$$y\frac{1}{\lambda}\Big(1 - \Big(\frac{x+ty}{\mu}\Big)^{\lambda}\Big) \le y \log \frac{\mu}{x+ty}$$

for $\lambda \in (0,1]$. Applying (8.1) again further yields

$$y \log \frac{\mu}{x+ty} \le y\Big(\frac{\mu}{x+ty} - 1\Big).$$

The same term also dominates the mean value for $\lambda \in (-1, 0)$:

$$\begin{aligned} y\frac{1}{\lambda}\Big(1 - \Big(\frac{x+ty}{\mu}\Big)^{\lambda}\Big) &= y\frac{1}{-\lambda}\Big(\Big(\frac{x+ty}{\mu}\Big)^{\lambda} - 1\Big) \\ &= y\frac{1}{-\lambda}\Big(\Big(\frac{\mu}{x+ty}\Big)^{-\lambda} - 1\Big) \\ &\le y\Big(\frac{\mu}{x+ty} - 1\Big). \end{aligned}$$

Here inequality (8.2), i.e. $\frac{1}{-\lambda}(z^{-\lambda} - 1) \le z - 1$ $(-\lambda \in (0,1), z > 0)$ was applied. Hence for $x + ty < \mu$, $\lambda \in (-1, 1]$, follows

$$\frac{1}{2}y \cdot |D_1 a_{\lambda}(x+ty, \mu)| = -\frac{1}{2}y \cdot D_1 a_{\lambda}(x+ty, \mu) \le y\Big(\frac{\mu}{x+ty} - 1\Big).$$

Simple inequalities as $x \geq 1 \Leftrightarrow \frac{x+1}{x} \leq 2 \Leftrightarrow \frac{\mu}{x} - 1 = \frac{(\mu-x)}{x} \leq 2\frac{(\mu-x)}{x+1}$ finally give the required result:

$$y\Big(\frac{\mu}{x+ty} - 1\Big) \leq y\Big(\frac{\mu}{x} - 1\Big) \leq 2y\frac{(\mu - x)}{x+1} \leq \frac{y^2}{\mu} + y|x-\mu|\Big(\frac{1}{\mu} + \frac{2}{x+1}\Big).$$

If $\mathbf{x} = \mathbf{0}, \mathbf{y} \geq \mathbf{1}$ one has to show

$$\frac{1}{2}|a_\lambda(y,\mu) - a_\lambda(0,\mu)| \leq \frac{y^2}{\mu} + y \cdot h(0,\mu) = \frac{y^2}{\mu} + y + cy\mu$$

with

$$\frac{1}{2}|a_\lambda(y,\mu) - a_\lambda(0,\mu)| = \begin{cases} \Big|\frac{1}{\lambda}y\Big(\Big(\frac{y}{\mu}\Big)^\lambda \frac{1}{\lambda+1} - 1\Big)\Big| & \text{for } \lambda \neq 0 \\ \Big|y\Big(\log\frac{y}{\mu} - 1\Big)\Big| & \text{for } \lambda = 0. \end{cases}$$

Considering $a_\lambda(y,\mu) - a_\lambda(0,\mu) \geq 0$, the difference is dominated by $y \log\frac{y}{\mu}$ if $\lambda \in (-1, 0]$. This is obvious for $\lambda = 0$ and for $\lambda \in (-1, 0)$ it results from (8.1) and $\log(\lambda + 1) < 0$:

$$\begin{aligned}
\frac{1}{\lambda}y\Big(\Big(\frac{y}{\mu}\Big)^\lambda \frac{1}{\lambda+1} - 1\Big) &= \frac{1}{-\lambda}y\Big(1 - \Big(\frac{\mu}{y}\Big)^{-\lambda}\frac{1}{\lambda+1}\Big) \\
&\leq \frac{1}{-\lambda}y\Big(1 - \log\Big(\Big(\frac{\mu}{y}\Big)^{-\lambda}\frac{1}{\lambda+1}\Big) - 1\Big) \\
&= \frac{1}{-\lambda}y\Big(\lambda\log\frac{\mu}{y} + \log(\lambda+1)\Big) \\
&\leq \frac{1}{-\lambda}y\lambda\log\frac{\mu}{y} \\
&= y\log\frac{y}{\mu}.
\end{aligned}$$

Using (8.1) gives

$$y\log\frac{y}{\mu} \leq y\Big(\frac{y}{\mu} - 1\Big) \leq \frac{y^2}{\mu} \leq \frac{y^2}{\mu} + y + cy\mu \quad \text{for } \lambda \in (-1, 0],$$

thus only $\lambda \in (0, 1]$ remains to be studied. However, in this case (8.2) immediately implies the desired result since

$$y\frac{1}{\lambda}\Big(\Big(\frac{y}{\mu}\Big)^\lambda \frac{1}{\lambda+1} - 1\Big) \leq y\frac{1}{\lambda}\Big(\Big(\frac{y}{\mu}\Big)^\lambda - 1\Big) \leq y\Big(\frac{y}{\mu} - 1\Big) \leq \frac{y^2}{\mu}.$$

If $a_\lambda(y,\mu) - a_\lambda(0,\mu) < 0$ (8.2) yields for the term ($\lambda \in (-1, 0)$):

$$\begin{aligned}
-y\frac{1}{\lambda}\Big(\Big(\frac{y}{\mu}\Big)^\lambda \frac{1}{\lambda+1} - 1\Big) &= y\frac{1}{-\lambda}\Big(\Big(\frac{\mu}{y}\Big)^{-\lambda}\frac{1}{\lambda+1} - 1\Big) \\
&= y\frac{1}{\lambda+1}\frac{1}{-\lambda}\Big(\Big(\frac{\mu}{y}\Big)^{-\lambda} - (\lambda+1)\Big)
\end{aligned}$$

$$
\begin{aligned}
&= y\frac{1}{\lambda+1}\Big(\frac{1}{-\lambda}\Big(\Big(\frac{\mu}{y}\Big)^{-\lambda}-1\Big)-\frac{1}{-\lambda}\lambda\Big)\\
&\le y\frac{1}{\lambda+1}\Big(\frac{\mu}{y}-1+1\Big)\\
&= \frac{1}{\lambda+1}\mu\\
&\le c\mu\\
&\le \frac{y^2}{\mu}+y+cy\mu
\end{aligned}
$$

with $c=\max\{2,\frac{1}{\lambda+1}\}$. If $\lambda=0$ application of (8.1) immediately gives

$$
-y\Big(\log\frac{y}{\mu}-1\Big)=y\Big(\log\frac{\mu}{y}+1\Big)\le y\Big(\frac{\mu}{y}-1+1\Big)=\mu\le\frac{y^2}{\mu}+y+cy\mu.
$$

The same follows for $\lambda\in(0,1]$ by repeated use of (8.1) resp. $-z\le-\log z-1$:

$$
\begin{aligned}
-y\frac{1}{\lambda}\Big(\Big(\frac{y}{\mu}\Big)^{\lambda}\frac{1}{\lambda+1}-1\Big) &= y\frac{1}{\lambda}\Big(1-\Big(\frac{y}{\mu}\Big)^{\lambda}\frac{1}{\lambda+1}\Big)\\
&\le y\frac{1}{\lambda}\Big(1-\log\Big(\Big(\frac{y}{\mu}\Big)^{\lambda}\frac{1}{\lambda+1}\Big)-1\Big)\\
&= y\frac{1}{\lambda}\Big(\lambda\log\frac{\mu}{y}+\log(\lambda+1)\Big)\\
&\le y\Big(\log\frac{\mu}{y}+\frac{1}{\lambda}(\lambda+1-1)\Big)\\
&\le y\Big(\frac{\mu}{y}-1+1)\Big)\\
&= \mu.
\end{aligned}
$$

b) If $y=0$, $x\ge 0$, the inequality obviously holds. In case of both variables x and y being positive ($x\ge 1$, $y\ge 1$), application of the mean value theorem yields

$$
\begin{aligned}
\frac{2}{\lambda+1}\mu\Big(\Big(\frac{x+y}{\mu}\Big)^{\lambda+1}-\Big(\frac{x}{\mu}\Big)^{\lambda+1}\Big) &= \frac{2}{\lambda+1}\mu y\Big(\frac{x+ty}{\mu}\Big)^{\lambda}(\lambda+1)\frac{1}{\mu}\\
&= 2y\Big(\frac{x+ty}{\mu}\Big)^{\lambda}
\end{aligned}
$$

for some $t\in(0,1)$. Considering $(\frac{x+ty}{\mu})^{\lambda}\le 1$ and $(\frac{x+ty}{\mu})^{\lambda}>1$, gives for $\lambda\in[0,1]$:

$$
\begin{aligned}
2y\Big(\frac{x+ty}{\mu}\Big)^{\lambda} &\le 2y\Big(1+\frac{x+ty}{\mu}\Big)\\
&\le 2y\Big(1+\frac{x+y}{\mu}\Big)\\
&= 2\Big(\frac{y^2}{\mu}+y\Big(\frac{x}{\mu}+1\Big)\Big)
\end{aligned}
$$

$$\leq \max\{2, \frac{2}{\lambda+1}\}\Big(\frac{y^2}{\mu} + y\Big(1 + \frac{x}{\mu} + \frac{2\mu}{x+1}\Big)\Big).$$

Analogous argumentation and $x \geq 1 \Leftrightarrow \frac{1}{x} \leq \frac{2}{x+1}$ leads to the same result for the case $\lambda \in (-1, 0)$:

$$\begin{aligned} 2y\Big(\frac{x+ty}{\mu}\Big)^{\lambda} &= 2y\Big(\frac{\mu}{x+ty}\Big)^{-\lambda} \\ &\leq 2y\Big(1 + \frac{\mu}{x+ty}\Big) \\ &\leq 2y\Big(1 + \frac{\mu}{x}\Big) \\ &\leq 2y\Big(1 + \frac{2\mu}{x+1}\Big) \\ &\leq \max\{2, \frac{2}{\lambda+1}\}\Big(\frac{y^2}{\mu} + y\Big(1 + \frac{x}{\mu} + \frac{2\mu}{x+1}\Big)\Big). \end{aligned}$$

For $y \geq 1,\ x = 0$ finally holds

$$\begin{aligned} \frac{2}{\lambda+1}\mu\Big(\frac{y}{\mu}\Big)^{\lambda+1} &\leq \frac{2}{\lambda+1}\mu\Big(1 + \Big(\frac{y}{\mu}\Big)^2\Big) \\ &\leq \max\{2, \frac{2}{\lambda+1}\}\Big(\frac{y^2}{\mu} + y(2\mu+1)\Big) \\ &= c\Big(\frac{y^2}{\mu} + yh(0,\mu)\Big). \end{aligned}$$

□

Lemma 8.2 *Consider a sequence of discrete random vectors* $(X^n)_{n\in\mathbf{N}}$ *with values in* $\mathbf{N}_\mathbf{0}^p$, $X^n = (X_1^n, \ldots, X_p^n)^T$. *For each* $i \in \{1,\ldots,p\}$ *and* $n \in \mathbf{N}$ *assume* $E(X_i^n) = \mu_i^n \in \mathbf{R}^+$ *and* $Var(X_i^n) \leq \mu_i^n$. *Let* $\bar{X}^n$ *denote a scaling of* X^n, $\bar{X}_i^n := \frac{X_i^n}{\mu_i^n}$ *for* $i \in \{1,\ldots,p\}$. *Further let be given a stochastically bounded sequence* Z^n *with values in* $I \subset \mathbf{R}^m$ *(*$m \in \mathbf{N}$*) and a function* $f : \mathbf{R}^p \times \mathbf{R}^m \to \mathbf{R}$, $(X, z) \mapsto f(X, z)$, *which is continuous on* $(0,\infty)^p \times I$. *Then for the asymptotics* $n \to \infty$ *holds*

$$\mathbf{1}_{(0,\infty)^p}(\bar{X}^n) \cdot f(\bar{X}^n, Z^n) = O_p(1).$$

Proof:
Let an arbitrary but fixed $\delta \in (0,1)$ be given. It will be shown at the end of the proof that there exists a positive compact set $K_\delta^1 \subset (0,\infty)^p$ with

$$P(\bar{X}^n \notin K_\delta^1) \leq P\Big(\exists\, i \in \{1,\ldots,p\} : \bar{X}_i^n = 0\Big) + \frac{\delta}{2} \quad \text{for all } n \in \mathbf{N}. \tag{8.4}$$

Since $Z^n = O_p(1)$ holds by assumption, there further exists a compact set $K_\delta^2 \subset I$ with

$$P(Z^n \notin K_\delta^2) \leq \frac{\delta}{2} \quad \text{for almost all } n \in \mathbf{N}. \tag{8.5}$$

The implication

$$(\bar{X}^n, Z^n) \in K_\delta^1 \times K_\delta^2 \quad \Rightarrow \quad f(\bar{X}^n, Z^n) \in f(K_\delta^1 \times K_\delta^2)$$

moreover entails

$$P\Big((\bar{X}^n, Z^n) \in K_\delta^1 \times K_\delta^2\Big) \leq P\Big(f(\bar{X}^n, Z^n) \in f(K_\delta^1 \times K_\delta^2)\Big) \tag{8.6}$$

with $f(K_\delta^1 \times K_\delta^2)$ being compact since f is continuous on $(0,\infty)^p \times I \supset K_\delta^1 \times K_\delta^2$. Assuming now $0 \in f(K_\delta^1 \times K_\delta^2)$, which can, in accordance with the previous arguments, be done without loss of generality, and using (8.4) – (8.6), $f(K_\delta^1 \times K_\delta^2)$ fulfils the required demands since for almost all $n \in \mathbf{N}$ holds

$$\begin{aligned}
& P\Big(\mathbf{1}_{(0,\infty)^p}(\bar{X}^n) \cdot f(\bar{X}^n, Z^n) \in f(K_\delta^1 \times K_\delta^2)\Big) \\
= \; & P\Big(\mathbf{1}_{(0,\infty)^p}(\bar{X}^n) \cdot f(\bar{X}^n, Z^n) \in f(K_\delta^1 \times K_\delta^2), \exists\, i \in \{1,\ldots,p\} : \bar{X}_i^n = 0\Big) \\
& + P\Big(\mathbf{1}_{(0,\infty)^p}(\bar{X}^n) \cdot f(\bar{X}^n, Z^n) \in f(K_\delta^1 \times K_\delta^2), \bar{X}_i^n \neq 0 \,\forall\, i \in \{1,\ldots,p\}\Big) \\
= \; & P\Big(\exists\, i \in \{1,\ldots,p\} : \bar{X}_i^n = 0\Big) + P\Big(f(\bar{X}^n, Z^n) \in f(K_\delta^1 \times K_\delta^2), \bar{X}_i^n \neq 0 \,\forall\, i\Big) \\
\geq \; & P\Big(\exists\, i \in \{1,\ldots,p\} : \bar{X}_i^n = 0\Big) + P\Big((\bar{X}^n, Z^n) \in K_\delta^1 \times K_\delta^2\Big) \\
= \; & P\Big(\exists\, i \in \{1,\ldots,p\} : \bar{X}_i^n = 0\Big) + 1 - P\Big((\bar{X}^n, Z^n) \notin K_\delta^1 \times K_\delta^2\Big) \\
= \; & P\Big(\exists\, i \in \{1,\ldots,p\} : \bar{X}_i^n = 0\Big) + 1 - P\Big(\bar{X}^n \notin K_\delta^1 \vee Z^n \notin K_\delta^2\Big) \\
\geq \; & P\Big(\exists\, i \in \{1,\ldots,p\} : \bar{X}_i^n = 0\Big) + 1 - P\Big(\bar{X}^n \notin K_\delta^1\Big) - P\Big(Z^n \notin K_\delta^2\Big) \\
\geq \; & P\Big(\exists\, i \in \{1,\ldots,p\} : \bar{X}_i^n = 0\Big) + 1 - P\Big(\exists\, i \in \{1,\ldots,p\} : \bar{X}_i^n = 0\Big) - \frac{\delta}{2} - \frac{\delta}{2} \\
= \; & 1 - \delta.
\end{aligned}$$

Now it remains to be shown (8.4), i.e. that for every given $\delta \in (0,1)$ there exists a compact set $K_\delta^1 \subset (0,\infty)^p \subset \mathbf{R}^p$ with $P(\bar{X}^n \notin K_\delta^1) \leq P(\exists\, i \in \{1,\ldots,p\} : \bar{X}_i^n = 0) + \frac{\delta}{2}$ for all $n \in \mathbf{N}$. To see this, consider a constant c_δ chosen sufficiently large, so that for one–dimensional intervals $\tilde{K}_\delta$ and $\tilde{K}_\delta^1$ the following inclusion holds:

$$\tilde{K}_\delta := \left[1 - \sqrt{\frac{2p}{c_\delta \cdot \delta}}, 1 + \sqrt{\frac{2p}{c_\delta \cdot \delta}}\right] \subset \left[\frac{1}{c_\delta}, \frac{2p}{\delta}\right] := \tilde{K}_\delta^1.$$

Putting $K_\delta^1 := (\tilde{K}_\delta^1)^p$ leads to (8.4) as follows:

$$\begin{aligned}
& P\Big(\bar{X}^n \notin K_\delta^1\Big) \\
= \; & P\Big(\exists\, i \in \{1,\ldots,p\} : \bar{X}_i^n \notin \tilde{K}_\delta^1\Big) \\
= \; & P\Big(\exists\, i \in \{1,\ldots,p\} : \bar{X}_i^n = 0 \vee \exists\, i \in \{1,\ldots,p\} : \bar{X}_i^n \in (0, \frac{1}{c_\delta}) \cup (\frac{2p}{\delta}, \infty)\Big)
\end{aligned}$$

$$\leq \quad P\Big(\exists\, i \in \{1,\ldots,p\} : \bar{X}_i^n = 0\Big) + \sum_{i=1}^{p} P\Big(\bar{X}_i^n \in (0, \frac{1}{c_\delta}) \cup (\frac{2p}{\delta}, \infty)\Big)$$
$$\leq \quad P\Big(\exists\, i \in \{1,\ldots,p\} : \bar{X}_i^n = 0\Big) + \frac{\delta}{2}.$$

The last inequality is valid since for every $i = 1, \ldots, p$ holds

$$P\Big(\bar{X}_i^n \in (0, \frac{1}{c_\delta}) \cup (\frac{2p}{\delta}, \infty)\Big) \leq \frac{\delta}{2p} \quad \text{for all } n \in \mathbf{N}, \tag{8.7}$$

which can be seen if the two cases $\mu_i^n > c_\delta$ and $\mu_i^n \in (0, c_\delta]$ are treated separately. Considering $\mu_i^n > c_\delta$, the Chebyshev Inequality, $Var(\bar{X_i}^n) \leq \frac{1}{\mu_i^n}$ and $\tilde{K}_\delta \subset \tilde{K}_\delta^1$ give

$$\begin{aligned}
P\Big(\bar{X}_i^n \in (0, \frac{1}{c_\delta}) \cup (\frac{2p}{\delta}, \infty)\Big) &\leq P\Big(\bar{X}_i^n \in [0, \frac{1}{c_\delta}) \cup (\frac{2p}{\delta}, \infty)\Big) \\
&= P\Big(\bar{X}_i^n \notin \tilde{K}_\delta^1\Big) \\
&\leq P\Big(\bar{X}_i^n \notin \tilde{K}_\delta\Big) \\
&= P\Big(|\bar{X}_i^n - 1| \geq \sqrt{\frac{2p}{c_\delta \cdot \delta}}\Big) \\
&\leq \frac{c_\delta \cdot \delta}{2p} Var(\bar{X}_i^n) \\
&\leq c_\delta \cdot \frac{\delta}{2p} \cdot \frac{1}{\mu_i^n} \\
&\leq \frac{\delta}{2p}.
\end{aligned} \tag{8.8}$$

If $\mu_i^n \in (0, c_\delta]$ holds, $\bar{X}_i^n > 0$ implies $\bar{X}_i^n = \frac{X_i^n}{\mu_i^n} \geq \frac{1}{\mu_i^n} \geq \frac{1}{c_\delta}$. This yields $P(\bar{X}_i^n \in (0, \frac{1}{c_\delta})) = 0$ and hence

$$P\Big(\bar{X}_i^n \in (0, \frac{1}{c_\delta}) \cup (\frac{2p}{\delta}, \infty)\Big) = P\Big(\bar{X}_i^n > \frac{2p}{\delta}\Big) \leq \frac{\delta}{2p} E(\bar{X}_i^n) = \frac{\delta}{2p}. \tag{8.9}$$

Since $\mu_i^n > 0$ for all $n \in \mathbf{N}$ by assumption, (8.8) and (8.9) finally establish (8.7). □

Lemma 8.3 *Let $(X^n)_{n\in\mathbf{N}}$ be a sequence of nonnegative integer valued random variables and $(u^n)_{n\in\mathbf{N}}$ an arbitrary vector valued sequence ($u^n \in \mathbf{R}^m, m \in \mathbf{N}$). Assume $E(X^n) = \mu^n \in [\epsilon, \infty) \subset \mathbf{R}^+$ ($0 < \epsilon \leq 1$) and $Var(X^n) \leq \mu^n$ for all $n \in \mathbf{N}$. Further let be given a function $g : \mathbf{R_0^+} \times \mathbf{R^m} \to \mathbf{R}$, $(X, u) \mapsto g(X, u)$, which is continuously differentiable in the first component on $\mathbf{R}^+$. If now for any given $k \in \mathbf{N}_0$ holds*

$$\mu^k \cdot \Big|\frac{\partial}{\partial \mu} g(\mu, u^n)\Big| \leq c \quad \textit{for all } n \in \mathbf{N} \textit{ and } \mu \geq \epsilon \quad (c \in \mathbf{R}^+),$$

then for the asymptotics $n \to \infty$ holds

$$\mathbf{1}_{\mathbf{N}}(X^n) \cdot \Big(g(X^n, u^n) - g(\mu^n, u^n)\Big) = O_p(\frac{1}{\mu^{k-1/2}}).$$

Proof:
Since g is continuously differentiable in the first component on $\mathbf{R}^+$, Taylor expansion and obvious arguments yield:

$$\begin{aligned}
& \mathbf{1}_{\mathbf{N}}(X^n) \cdot |g(X^n, u^n) - g(\mu^n, u^n)| \\
= \; & \mathbf{1}_{\mathbf{N}}(X^n) \cdot \Big|(X^n - \mu^n) \cdot \int_0^1 \frac{\partial}{\partial \mu} g(\mu, u^n)|_{\mu=\mu^n+z(X^n-\mu^n)} dz\Big| \\
\leq \; & \mathbf{1}_{\mathbf{N}}(X^n) \cdot |X^n - \mu^n| \cdot \sup_{z\in[0,1]} \Big|\frac{\partial}{\partial \mu} g(\mu, u^n)|_{\mu=\mu^n+z(X^n-\mu^n)}\Big| \\
= \; & \mathbf{1}_{\mathbf{N}}(X^n) \cdot \frac{|X^n - \mu^n|}{(\mu^n)^k} \\
& \cdot \sup_{z\in[0,1]} \Big(\Big(\frac{\mu^n}{\mu^n + z(X^n - \mu^n)}\Big)^k \cdot (\mu^n + z(X^n - \mu^n))^k \cdot \Big|\frac{\partial}{\partial \mu} g(\mu, u^n)|_{\mu=\mu^n+z(X^n-\mu^n)}\Big|\Big) \\
\leq \; & c \cdot \frac{|X^n - \mu^n|}{(\mu^n)^k} \cdot \mathbf{1}_{\mathbf{N}}(X^n) \cdot \sup_{z\in[0,1]} \Big(\frac{\mu^n}{\mu^n + z(X^n - \mu^n)}\Big)^k. \qquad (8.10)
\end{aligned}$$

The last inequality is trivial for $X^n = 0$. If $X^n \geq 1$ it holds since for $\mu \geq \epsilon$ ($\epsilon \in (0,1]$) $\mu^k|\frac{\partial}{\partial \mu} g(\mu, u^n)| \leq c$ is presumed, which is in particular met for $\mu = \mu^n + z(X^n - \mu^n) \geq \epsilon$. Using Chebyshev's inequality and the assumption $Var(X^n) \leq \mu^n$ now yields

$$P\Big(\frac{|X^n - \mu^n|}{\sqrt{\mu^n}} \geq \frac{1}{\sqrt{\delta}}\Big) \leq \delta \cdot Var\Big(\frac{X^n - \mu^n}{\sqrt{\mu^n}}\Big) = \delta \cdot \frac{1}{\mu^n} \cdot Var(X^n) \leq \delta$$

for all $n \in \mathbf{N}$ and hence

$$\frac{|X^n - \mu^n|}{(\mu^n)^k} = \frac{|X^n - \mu^n|}{\sqrt{\mu^n}} \cdot (\mu^n)^{1/2-k} = O_p\Big(\frac{1}{(\mu^n)^{k-1/2}}\Big). \qquad (8.11)$$

Writing $\bar{X}^n = \frac{X^n}{\mu^n}$ and $f(\bar{X}^n) = \sup_{z\in[0,1]} \Big(\frac{1}{1+z(\bar{X}^n-1)}\Big)^k$ Lemma 8.2 further gives

$$\begin{aligned}
& \mathbf{1}_{\mathbf{N}}(X^n) \cdot \sup_{z\in[0,1]} \Big(\frac{\mu^n}{\mu^n + z(X^n - \mu^n)}\Big)^k \\
= \; & \mathbf{1}_{(0,\infty)}(\frac{X^n}{\mu^n}) \cdot \sup_{z\in[0,1]} \Big(\frac{1}{1 + z(\frac{X^n}{\mu^n} - 1)}\Big)^k \\
= \; & \mathbf{1}_{(0,\infty)}(\bar{X}^n) \cdot f(\bar{X}^n) \\
= \; & O_p(1).
\end{aligned}$$

This statement and (8.11) concerning the asymptotic order of the terms in (8.10) finally yield the result. □

Bibliography

[1] Billingsley, Patrick (1968): *Convergence of Probability Measures*, John Wiley & Sons, New York.

[2] Billingsley, Patrick (1986): *Probability and Measure*, John Wiley & Sons, New York.

[3] Cressie, N.A.C. and Read T.R.C. (1984): "Multinomial Goodness–of Fit Tests", *Journal of the Royal Statistical Society*, Ser. B, 46, No. 3, 440 – 464.

[4] Cressie, N.A.C. and Read T.R.C. (1988): *Goodness–of Fit Statistics for Discrete Multivariate Data*, Springer, New York.

[5] Dieudonné, J. (1960): *Foundations of Modern Analysis*, Academic Press, New York.

[6] Habermann, S.J. (1974): *The Analysis of Frequency Data*, The University of Chicago Press, Chicago and London.

[7] Johnson, N.L. and Kotz S. (1969): *Discrete Distributions*, John Wiley & Sons, New York.

[8] Johnson, N.L., Kotz S. and Kemp, A.W. (1992): *Univariate Discrete Distributions*, John Wiley & Sons, New York.

[9] Kendall, M.G. and Stewart A. (1969): *The Advanced Theory of Statistics*, Vol. 1, Third Edition, Charles Griffin & Company Limited, London & High Wycombe.

[10] McCullagh, Peter (1985): "Sparse Data and Conditional Tests", *Bulletin of the International Statistics Institute*, Proceedings of the 45th Session of ISI (Amsterdam), Invited Paper 28.3, 1 – 10.

[11] McCullagh, Peter (1986): "The Conditional Distibution of Goodness–of Fit Statistics for Discrete Data", *Journal of the American Statistical Association*, Vol. 81, No. 393, 104 – 107.

[12] Morris, Carl (1975): "Central Limit Theorems for Multinomial Sums", *The Annals of Statistics*, Vol. 3, No. 1, 165 – 188.

[13] Osius, Gerhard (1984): "Binomialverteilung: Momente", unpublished manuscript, University of Bremen.

[14] Osius, Gerhard and Rojek, Dieter (1992): "Normal Goodness–of–Fit Tests for Multinomial Models with Large Degrees of Freedom", *Journal of the American Statistical Association*, Vol. 87, No. 420, 1145 – 1152.

[15] Osius, Gerhard (1994): "Evaluating the Significance Level of Goodness-of-Fit Statistics for Large Discrete Data". In: Dirschedl, P. and Ostermann, R. (Eds.), *Computational Statistics*, Physica-Verlag Heidelberg, p. 393-417.

[16] Osius, Gerhard (1996): "Statistische Methoden in der Epidemiologie", unpublished script, University of Bremen.

[17] Prentice, R.L. and Pyke, R. (1979): "Logistic Disease Incidence Models and Case–Control Studies, *Biometrika*, Vol. 66, No. 3, 403 – 411.

[18] Rojek, Dieter (1989): "Asymptotik für Anpassungstests in Produkt–Multinomialmodellen bei wachsendem Freiheitsgrad", unpublished Ph.D., University of Bremen.

[19] Steck, George P. (1957): "Limit Theorems for Conditional Distributions", *University of California Publications in Statistics*, Vol. 2, No. 12, 237 – 284.