Decompounding: An Estimation Problem for the Compound Poisson Distribution

Vom Fachbereich Mathematik und Informatik
der Universität Hannover
zur Erlangung des Grades
Doktor der Naturwissenschaften
Dr. rer. nat.

genehmigte Dissertation

von

Dipl.-Math. Boris Buchmann
geboren am 22. Oktober 1969 in Göttingen

2001

Referent: Prof. Dr. Rudolf Grübel, Universität Hannover
Korreferentin: Dr. Susan M. Pitts, University of Cambridge
Tag der Promotion: 27.04.2001
Datum der Veröffentlichung: Juli 2001

in memory of Gustav Finger

and dedicated to my father

Acknowledgment

First of all I want to express my deepest gratitude to my supervisor, Prof. Dr. R. Grübel, for excellent guidance and long discussions - in both mathematics and music. His mathematical advice has fostered my research. Also, he helped to secure an opportunity to visit Cambridge University and the "Mathematisches Forschungsinstitut Oberwolfach".

Special thanks also go to Dr. Susan Pitts for serving as co-examiner for my thesis, and especially for her warm hospitality as well as scholarly discussions during my visit at the "Statistical Laboratory" at Cambridge University.

Like most scholars, I have accumulated many debts in the course of my investigation. First of all, I owe many institutional debts. The burdens of research and writing have been lightened by the aid of the members of the "Institut für Mathematische Stochastik" at the University of Hannover. In particular, I want to thank Prof. Dr. L. Baringhaus, Ms. R. Middelmann, Ms. U. Pätzold, and Prof. Dr. D. Morgenstern. You hopefully realized how much I enjoyed the very supportive atmosphere at the institute.

I would also like to thank my office mate, Dr. Anke Reimers, for her encouragement and advice over the previous years. The conversations with Dr. Christian Alpert and associate professor PD Dr. Jeroen Spandaw at my office were also beneficial to this work. I enjoyed our discussions about pure mathematics. Nevertheless, I am indebted to them for sharing their knowlegde of the Gauss Bonnet theorem with me. I am also grateful to Dipl. Math. Johannes Emmrich and "the Big W." Dipl. Math. Stefan Weber for long, helpful discussions. In particularly, I want to thank Dr. Bernd Buchwald for sharing his friendship, and never-ending conservations about life, music and of course mathematics with me.

And finally, I owe thanks to all my friends, especially Arnd and André, for their patience when I was neglecting them while writing this thesis. Last but not least, I recognize as my heaviest creditors my family; my two sisters Sonja and Anne-Kathrin, and my parents Ulrike und Klaus Buchmann, who patiently put faith in me and my work, as well as especially Kerstin Fielauf for her very great patience.

Zusammenfassung

Zusammengesetzte Poisson-Verteilungen spielen in der Versicherungsmathematik und in der Warteschlangentheorie eine große Rolle. Ist $(X_i)_{i\in\mathbb{N}}$ eine Folge von unabhängigen und identisch mit P verteilten Zufallsvariablen und τ eine von $(X_i)_{i\in\mathbb{N}}$ unabhängige mit Parameter $\lambda > 0$ Poisson-verteilte Zufallsvariable, so heißt die Verteilung Q der zufälligen Summe $Y = \sum_{k=1}^{\tau} X_k$ zusammengesetzten Poisson-Verteilung mit Intensität $\lambda > 0$ und Basisverteilung P. Q läßt sich als Faltungsreihe in λ und P schreiben, es gilt $Q = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} P^{*k}$.

Ausgehend von einem Datensatz aus unabhängigen und je mit gleicher Intensität und Basisverteilung zusammengesetzt Poisson-verteilten Zufallsvariablen soll Intensität und Basisverteilung nichtparametrisch geschätzt werden.

Im Falle diskreter Beobachtungen läßt sich die Massenfunktion der zusammengesetzten Verteilung mit Hilfe der Panjer-Rekursion leicht aus der Intensität und der Massenfunktion der Basisverteilung gewinnen. Diese Rekursion kann umgekehrt werden und führt auf eine Rekursionsformel für λ und Massenfunktion von P bei gegebener zusammengesetzter Verteilung Q, der Panjer-Inversion. Schätzt man die Massenfunktion der zusammengesetzten Verteilung durch die relativen Häufigkeiten, so liefert die Panjer-Inversion angewendet auf die relativen Häufigkeiten einen Schätzer für die Intensität und die Massenfunktion. Für diesen Schätzer wird starke Konsistenz und asymptotische Normalität in der Banach-Algebra ℓ^1 der absolut summierbaren Folgen hergeleitet. Da die Folge der Panjer-invertierten relativen Häufigkeiten stets negative Einträge aufweist, ist eine Projektion unumgänglich. Indem man alle negativen Einträge auf Null setzt und die Folge wieder normiert, erhält man eine Wahrscheinlichkeitsmassenfunktion. Eine Variante dieser Methode ist die Inversion nur bis zu einer datenabhängigen Oberschranke auszuführen und das so gewonnene Anfangssegment analog zu behandeln. Beide Methoden führen auf in ℓ^1 stark konsistente Schätzer, und Verteilungskonvergenz in ℓ^1 gegen einen u.U. nicht Gaußschen Grenzwert wird gezeigt. Für die Variante werden Bedingungen an das statistische Verhalten der zufälligen Oberschranke gegeben, die bei Wahl des Maximums der vorliegenden Daten als Oberschranke erfüllt werden.

Abschneiden der Daten an einer festen Schranke führt auf ein endlichdimensionales parametrisches Modell unter Ordnungsrestriktionen. Es wird gezeigt das sich der Maximumlikelihoodschätzer in diesem Modell asymptotisch wie eine

Kegelprojektion verhält. Dies motiviert eine Projektionsmethode mit datenabhängiger Projektionsmatrix. Alle Methoden werden am Beispiel der Bortkiewicz-Daten illustriert. Als Nebenprodukt ergibt sich außerdem der Verteilungsgrenzwert des Likelihoodquotiententests für die Hypothese über das Vorliegen von Poisson-verteilten Daten gegenüber der generelleren Annahme, daß die Daten zusammengesetzt Poisson-verteilt sind. Ohne Abschneiden der Daten kann ein nichtparametrischer Maximumlikelihoodschätzer definiert werden. Es wird eine hinreichende Bedingung an die Basisverteilung gestellt, die seine Konsistenz in ℓ^1 gewährleistet.

Besitzt P eine Dichte bezüglich des Lebesgue-Maßes, so kann die Dichte des absolut stetigen Anteils von Q durch ein Histogramm geschätzt werden. Es wird eine Panjer Inversionsformel für Histogramme gegeben und starke Konsistenz des so gewonnenen Schätzers im Raum der Lebesgue-integrierbaren Funktionen L^1 bewiesen.

Für den Fall, daß P eine beliebige auf den positiven reellen Zahlen konzentrierte Verteilung ist, wird ein auf einer Faltungsreihe und der empirischen Verteilungsfunktion basierender Einsetzschätzer konstruiert und seine starke Konsistenz und asymptotische Normalität in einem Funktionenraum mit gewichteter Supremumsnorm hergeleitet.

Schlagwörter: Zusammengesetzte Poisson-Verteilung, Nichtparametrische Schätzung, Kegelprojektionen

Abstract

Given an iid-sample from a compound Poisson distribution Q, we consider the estimation of the corresponding rate parameter $\lambda > 0$ and base distribution P. This has applications in insurance mathematics and queueing theory. The ingredients λ, P and Q are connected by a convolution power series, i.e. $Q = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} P^{*k}$. If P is concentrated on the positive real numbers then the probability mass function of Q can be calculated from λ and the probability mass function of P using the Panjer recursion formula. This formula can be inverted leading to a recursion formula for λ and the probability mass function of P based on the probability mass function of Q, the Panjer inversion. This suggests a simple plug-in estimator for λ and P based on the relative frequencies measured from the compound Poisson sample. Strong consistency and asymptotic normality is shown in the Banach algebra of absolutely summable sequences ℓ^1. Although the sequence, which comes out of the Panjer inversion of the relative frequencies is an absolutely summable sequence, for large enough sample sizes, it must have negative entries. Therefore a projection procedure is necessary. Two methods are under investigation. The first method is to put all negative entries to zero and then norm to one. A variant is also discussed: Compute the Panjer inversion up to some data driven end point, i.g. the sample maximum, then put the negative entries of this finite segment to zero and norm to one. Strong consistency and a distributional limit result is proved for both methods in ℓ^1 under suitable conditions on P and the statistical behaviour of the end point. A possible choice for the end point is the maximum of the data. The limit turns out to be not necessarily Gaussian. The second approach is based on the ideas of maximum likelihood estimation. Truncation of the data leads to a finitely dimensional parametric model under cone restrictions. It is shown that the maximum likelihood estimator behaves asymptotically like a cone projection. This motivates a second data driven projection estimator. The methods are illustrated using the famous Bortkiewicz data. Furthermore, we derive the limit law that rules the log likelihood ratio test statistic for testing the hypothesis of Poissonity within the class of compound Poisson distributions. Without truncation of the data, a nonparametric maximum likelihood estimator can be defined and is consistent in ℓ^1 under suitable conditions on the underlying basis distribution.

If P is absolutely continuous with respect to the Lebesgue measure, then the density of the absolutely continuous part of Q can be estimated by a histogram. An analogue Panjer inversion for histograms is given and strong consistency of the estimator in the space of Lebesgue integrable functions is established.

If P is just some probability measure concentrated on the positive reals, then we propose an estimator based on a convolution power series. Strong consistency and asymptotic normality is proved for this estimator in a Banach space of functions topologized with a weighted sup norm.

Keywords: Compound Poisson distributions, Nonparametric Estimation, Cone Projections

***

All used notations in the thesis are standard writings in probability theory and statistics or given during the text.

Introduction

The importance of compound Poisson distributions in both probability theory as important subclass of infinitely divisible distributions and its applications is well known. We just indicate two of them.

Consider the standard risk model in actuarial mathematics (see for example [Bu70], p.35, [Pa92], p.165). Suppose that N_t is the number of damages or claims that occur until time t and $X_1, X_2, \ldots$ are their amounts. Then $S_t = \sum_{l=1}^{N_t} X_l$ with $S_0 \equiv 0$ is the total amount of damages accumulated in the time interval $[0, t]$. Of course, everything is random. The general assumptions of the standard risk model are the following: $(N_t)_{t\geq 0}$ is supposed to be a homogeneous Poisson process with constant rate λ. Furthermore, the single claims X_i, $i \in \mathbb{N}$, form a sequence of independent and identically distributed random variables, themselves independent from N_t and each of them distributed according to a probability measure P. If we observe the total damage process at times kT with $k \in \mathbb{N}_0$ and $T > 0$ fixed, then the random variable $Y_k = S_{kT} - S_{(k-1)T}$ measures the total claim accumulated in the time interval $((k-1)T, kT]$.

One can imagine the same situation in the context of queueing models. Customers arrive at a service system in groups, e.g. touring busses at the zoo. Once again, the number of groups arrived at the system until time t is modelled by a homogeneous Poisson process with rate λ. The numbers of the single groups are given by a sequence of independent and identically distributed random variables with distribution P. The total number of customers is S_t. The random variable Y_k represents the total number of costumers who arrived during the time interval $((k-1)T, kT]$.

Within these two models the Y-variables themselves form a sequence of independent and identically distributed random variables. The distribution Q of Y_i is given by a compound Poisson distribution with intensity parameter λT and claim distribution P that can be written as a convolution power series, i.e.

$$Q = \sum_{k=0}^{\infty} e^{-\lambda T} \frac{(\lambda T)^k}{k!} P^{*k}.$$

In the following, we assume $T = 1$.

This thesis investigates the problem of estimating λ and P nonparametrically from a given sample of independent random variables $Y_1, \ldots, Y_n$ with distribution

Q. Equivalently, we do not directly observe the single claims, but want to construct their distribution P from a sample of total claims. The methods developed here can be used to restore information about lost data or compressed data.

We have to deal with a nonlinear deconvolution problem. It can be seen as the inverse of the compounding mapping

$$(\lambda, P) \longmapsto \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} P^{*k}.$$

Since the inverse procedure of convolution is called „deconvolution“ in the literature, it was chosen the analogous term „decompounding“ as a name for the inverse compounding and as title for this thesis.

The deconvolution problem, i.e. to estimate P from a sample $Y_1, \ldots, Y_n$ with $Y_i = X_i + \varepsilon_i$, is the linear variant. The random variables X_i and ε_i, $i = 1, \ldots, n$, are supposed to be mutually independent. The distribution P of the X-variables is unknown and has to be estimated, the distribution of the ε-variables is known. This problem has been widely studied in the literature from various aspects. The literature can be grouped into two main classes. The assumption that both the ε-variables and the X-variables are absolutely continuously distributed leads to a density estimator for the density of P, based on Fourier transforms. We refer to Fan, who has discussed rates of convergence and other asymptotic properties (see [Fa91], [Fa97] and the references given there). Under the assumption that ε is again absolutely continuously distributed with some continuous and monotone decreasing density, nonparametric maximum likelihood estimation for the distribution function of P can be performed. This is considered by Van Es, Groeneboom and Jongbloed (see [Es91], [Jo95], [Gr92] and the references given there).

If $P = \delta_1$, δ_1 is the Dirac measure concentrated in 1, then we have the important special case of Poissonity, i.e. Y_i is Poisson distributed with parameter $\lambda > 0$. Testing the hypothesis of Poissonity within the class of compound Poisson distributions has been investigated by Puri (see [Pu85], see also [Ne79] and the references given there).

Some work is also available for parametric estimation (see [Hu90], [Pa92]).

The direct compounding, i.e. the nonparametric estimation of Q from a given sample of claims $X_1, \ldots, X_n$, was considered by S.M. Pitts (see [Pi94]) using a plug-in-estimation procedure.

The nonparametric decompounding has not been studied in literature yet, in spite of its obvious usefulness.

This thesis starts with two simple ideas. Firstly, Panjer (see [Pa81]) has given a simple recursion formula for the case of discrete data. Given the intensity λ and the counting density p, the compound counting density q can easily be calculated. This formula can be inverted, as has been remarked by [Hu90]. This inversion will be called Panjer inversion in the sequel. In fact, the Panjer inversion can be used to generate an intensity λ and a sequence of real numbers $p_1, p_2, \ldots$ for every (not necessarily compound Poisson) counting density q with mass at zero. Of course, p needs to be neither a probability density function nor a summable sequence. In spite of this, it is a natural procedure to estimate q by the relative frequencies q^n and then λ and p by the Panjer inversion.

Another starting point (see Chapter 1) is to deal with the convolution series. Up to an affine transformation the compounding with Poisson weights can be viewed as the exponential function. Hence it is natural to estimate $\lambda(P - \delta_0)$ via a logarithm in an appropriate Banach algebra. The first chapter provides a necessary and sufficient condition for the existence of a logarithm based on the Gelfand transform in the setting of commutative Banach algebras. It is proved with elementary methods at the cost of some density assumption on the space of Gelfand transforms. The density assumption is fulfilled in both cases, discrete and absolutely continuous distributed data. Chapter 2 discusses the Panjer inversion based estimator. If the sample size is large enough this estimator coincides with the real logarithm of the relative frequencies in the space of absolutely summable sequences. We prove strong consistency in this space and give sufficient and necessary conditions for asymptotic normality. The Panjer inversion estimator has the big disadvantage that it is not a probability density; it has negative entries. We investigate the following naive method: Replace the negative entries by zero and normalize the sequence to one. A variant of this method is studied also: Calculate the Panjer inversion up to a data driven end point S_n and apply the naive normalization to it. Again, we proof strong consistency and asymptotic normality under suitable conditions on S_n. The limit distributions of both methods coincide and are not necessarily Gaussian.

Panjer inversion of relative frequencies does not result in a probability counting measure. We have to perform a suitable projection onto the probability simplex. Projections are driven by inner products. Under truncation of the data, we analyse the maximum likelihood method that is known to have good statistical properties. In contrast to the standard parametric situation, we discuss the asymptotic behaviour of the maximum likelihood estimator (MLE) at boundary points of the parameter set. The MLE can locally be viewed as a cone projection driven by the Fisher information matrix. This is quite similar to the situation in the isotonic regression (see [Ro88], a small example is in the introduction of the Chapter 3). However, we have to deal with the difficulties of both nonlinearity and nonconvexity of the parameter space. In contrast to the isotonic regression localizing reasoning is necessary. To avoid the problem of the direct calculation of the MLE, we construct an analogue of a one-step-Newton-iteration that turns out to be efficient. We apply the method to the famous Bortkiewicz data describing the number of men kicked to death by horses in the Prussian army. As a second application we derive the asymptotic distribution of the log likelihood ratio tests for testing the Poissonity hypothesis within the class of compound Poisson distributions. We can profit from both our localization and the analogue test situations in order restricted inference. Similar to the situation there the limit distribution of the test statistics turns out to be a mixture of χ^2-distributions (see again [Ro88]). The mixing coefficients depend on the unknown parameter, but studentization is possible. Since the limit distribution cannot be calculated for higher dimensions, we propose a Monte Carlo method for an approximation.

The next chapter, Chapter 4, returns to the estimation of untruncated data. The existence of nonparametrical maximum likelihood estimation is under investigation. Also a consistency proof for more general estimators including the nonparametric MLE is given.

If P is absolutely continuous with density p, then Q can be written as $q_0\delta_0+q$. A Panjer inversion formula is proved to calculate p from a histogram estimating q in Chapter 5. The general case of decompounding, i.e. to estimate the distribution function of P, is discussed in the last chapter. We construct a plug-in-estimator based on a convolution power series and show strong consistency and asymptotic normality in a weighted Banach space.

Due to the limit laws in effect at the boundary points of the parameter set of λ there are two principal difficulties in the decompounding problem. If λ gets smaller and smaller, or in terms of the insurance risk model, if less and less damages will be observed, their amount cannot be measured anymore, of course. There is a loss of information. In more mathematical terms, if P is fixed then the compound Poisson distribution will tend in total variation norm to the Dirac measure concentrated in zero for λ converging to zero.

On the other side, increasing the intensity will produce a larger and larger number of claims in a fixed interval. They will be lumped together in the sum Y_i. This also means a loss of information that can be formalized using a central limit theorem for random sums (see [Fe66], p.265). If X_i has finite second moment, and Y_λ is compound Poisson distributed with intensity λ and claim distribution P, then it holds that

$$\frac{Y_\lambda - \lambda E X_i}{\sqrt{\lambda E X_i^2}} \xrightarrow{\mathcal{D}} Z$$

for a standard normally distributed random variable Z. Hence the only information that remains available for the distribution P of the claims X_i is contained in its expectation value and its second moment.

For completeness, let us also note that we have to make some assumptions on λ and P. We will always assume that P has no mass in zero. Otherwise, the true parameter (λ, P) could not be specified anymore. This is easily seen by a rescaling argument in the exponent of the exponential function:

$$\exp\left(\lambda(P-\delta_0)\right) = \exp\left(\lambda(1-P(\{0\}))\left(\frac{1}{1-P(\{0\})}P(\cdot\cap(0,\infty)) - \delta_0\right)\right).$$

Hence we have two corresponding pairs λ, P and $\lambda(1-P(\{0\})), P(\cdot|(0,\infty))$ generating the same compound Poisson distribution.

Finally, a technical remark: For the rest of the thesis Y_i is an iid-sequence of random variables on a common probability space $(\Omega, \mathcal{A}, \mathbb{P})$. The distribution of Y_i is given by a compound Poisson distribution with intensity $\lambda > 0$ and claim distribution P. We will always use $Q, q, q^n, \ldots$ for the compound distributions and $P, p, p^p, \ldots$ for the claim distributions to simplify the notation.

Contents

Chapter 1

Logarithms in Banach Algebras

1.1 Motivation

The true Q is in the range of the convolution series of the exponential function i.e. $Q = \exp(\lambda(P - \delta_0))$. We will also use the notation $e^{\lambda(P-\delta_0)}$. Natural domains of convolution are special Banach algebras. One straightforward approach of estimating Q therefore is considering Q as an element of a Banach algebra, then to estimate it by an estimator $\hat{Q}$ with $\hat{Q}$ taken from the same algebra. Solving $\hat{Q} = \exp(\lambda(\hat{P} - \delta_0))$ provides an estimator $\hat{P}$ of P. Hence we need a criterion for $\hat{Q}$ to be in the range of the exponential function. Obviously, the set of probability measures is very complicated, so this is a very ambitious task. A criterion based on the Fourier transform would be much more convenient. Roughly spoken, the standard textbooks (see [Ru91]) provide the following basic logarithm theorem: If the spectrum of Q does not separate zero and infinity then there is a logarithm.

The first example illustrates that this theorem does not cover the whole situtation. Consider the family $(\mathcal{P}_\lambda)_{\lambda>0} = (\exp(\lambda(\delta_1 - \delta_0)))_\lambda$ of Poisson distributions. Obviously, every $\mathcal{P}_\lambda$ is in the range of the exponential function. If we consider $\mathcal{P}_\lambda$ as an element of the Banach algebra of two-sided absolutely summable complex sequences then the spectrum of $\mathcal{P}_\lambda$ turns out to be the same as the range of the Fourier transform $\hat{\mathcal{P}}_\lambda(\theta) = \exp(\lambda(e^{i\theta} - 1))$. For each $\lambda > 0$ the latter is a parametrization of a curve in the complex plane (see fig.1.1 on p.23). Indeed, for $\lambda \geq \pi$ the Fourier transform separates zero and infinity. For the same reason the approach to use a one dimensional functional calculus to define the logarithm via a Cauchy formula

$$\log a := \frac{1}{2\pi i} \int_\gamma (ze - a)^{-1} \log z \, dz$$

with some γ surrounding the spectrum fails as well. An analytic version of a logarithm demands some path in the complement of its domain connecting zero and infinity[1].

[1] Indeed this is the proof of the basic logarithm theorem.

On the other hand, however complicated the Fourier transform $\hat{\mathcal{P}}_\lambda$ winds around zero, the picture shows that all curves can be contracted continuously to point 1 without touching zero. They are null-homotopic in $\mathbb{C}^* := \mathbb{C}\backslash\{0\}$. The next section will deal rigorously with this situation.

To see the practical limitation of our approach consider the space of complex measures on the line $\mathbb{R}$ (a nice survey for the situation of measure algebras is [Ta73]). That is the canonical domain for probability measures. Here the situation is much more complicated. The range of the Fourier transform is a subset of the spectrum. The Gelfand transforms needed to calculate the spectrum are not satisfactorly specified to use them for practical purposes. Note that $\exp(\lambda(P - \delta_0))$ is an invertible element in every commutative algebra with unit element. The inverse is given by $\exp(-\lambda(P - \delta_0))$. Hence a necessary condition for Q to be in the range of the exponential function is invertibility. As an extreme example we should mention the existence of a probability measure on the real line with the following properties: The range of its Fourier transform is contained in the real numbers and its spectrum contains the whole unit circle. Furthermore, there is a measure for which the Fourier transform is bounded away from zero, but which is not invertible (this can be found in [Ta73]). These two facts indicate that it is not enough to consider Fourier transforms on their own. We have to deal with Gelfand transforms.

In the next section we will derive a criterion using elementary methods for the existence of a logarithm under some additional assumption on the space of Gelfand transforms. We show that this condition holds in some important examples. Example 1.5 and a much deeper result on logarithms based on multivariate complex function theory can be used to remove this assumption.

1.1.1 A Logarithm Criterion

Our reference here is [Ru91] (chapter 10, chapter 11). First let us recall the basic facts about commutative Banach algebras and Gelfand theory.

Let $(A, \|\cdot\|)$ be a commutative Banach algebra with identity e, i.e. A is a Banach space over the field of complex numbers with an additional binary operation $*$, a multiplication, that makes A into a commutative algebra with identity e and

$$\|a * b\| \leq \|a\|\|b\|$$

holds for all $a, b \in A$. We will see examples in the next chapter. We will restrict ourselves to commutative Banach algebras with some identity e.

Entire functions can be defined in Banach algebras in a very natural way using their power series representation, e.g. $\exp(a) := \sum_{l=0}^{\infty} \frac{1}{k!} a^{*k}$, a^{*k} the k-th power of a, $a^0 = e$. There is no need for a functional calculus. Recall the functional equation $\exp(a + b) = \exp(a) * \exp(b)$, which is true in this setting too.

Let Δ be the maximal ideal space. This is defined to be the space of all nontrivial homomorphism h, i.e. all nontrivial linear functionals h that are multiplicative ($h(a * b) = h(a)h(b)$). The name „maximal ideal space“ comes from the canonical identification of kernels of homomorphism with maximal ideals of A. The Gelfand transform of an $a \in A$ is defined to be the mapping $\Delta \ni h \to \hat{a}(h) := h(a)$. This is a mapping from Δ to the field of complex numbers $\mathbb{C}$. The Gelfand topology is defined to be the weakest topology on Δ that makes every $\hat{a}$ continuous. Equipped with this topology, Δ turns out to be a compact Hausdorff space. Defining $\hat{A}$ to be the space of all Gelfand transforms, we have the set theoretic inequality $\hat{A} \subset C(\Delta)$.

The spectrum of an $a \in A$ is defined to be the set $\sigma(a)$ of all $\lambda \in \mathbb{C}$ such that $a - \lambda e$ is not invertible. $\sigma(a)$ can be characterized using the Gelfand transform of a in a very simple manner, i.e. $\sigma(a) = \hat{a}(\Delta)$. If $\hat{a}(h) \neq 0$ holds for all $h \in \Delta$, then a is invertible.

A Banach algebra is called *semisimple* iff the intersection of all maximal ideals is trivial. Obviously, this is equivalent to the fact that an element of a semisimple Banach algebra is determined uniquely by its Gelfand transform $\hat{a}$, i.e.

$$\hat{a}(h) = 0 \text{ for all } h \in \Delta \Rightarrow a = 0.$$

Define $G(A)$ to be the group of invertible elements. Then $G(A)$ is an open subset of A. Therefore $G(A)$ is the union of disjoint maximal connected open subsets of A, the components of $G(A)$. One of them, G_1, contains e. G_1 is called the principal component; it turns out to be the image of A under the exponential function, i.e. $G_1 = \exp(A)$. This will be the key ingredient for our elementary proof.

We now return to our purpose. Consider a b in A, then $[0,1] \ni t \longmapsto \exp(tb)$ is a continuous path connecting e and $a = \exp(b)$. This path lies entirely in $G_1 = \exp(A)$. Let us define

$$[0,1] \times \Delta \longmapsto H(t,h) := \exp(t\hat{b}(h)),$$

then H defines a homotopy between $\hat{a}$ and 1, i.e. H is a continuous function on $[0,1] \times \Delta \to \mathbb{C}$ with $H(0,h) = 1$, $H(1,h) = \hat{a}(h)$ for all $h \in \Delta$. Since $t \to \exp(ta)$ is a path contained in $G_1 \subset G(A)$, the group of invertible elements, the range of H will not contain zero, i.e. $H([0,1] \times \Delta) \subset \mathbb{C}^*$. The next theorem shows that the other direction is also true. We define $\|f\|_K := \sup_{k \in K} f(k)$ for a complex valued function on some compact set K. For a set $B \subset C(K)$ let $\overline{B}^{\|\cdot\|_K}$ denote the usual topological closure of B in $C(K)$, if the topology is induced by the norm $\|\cdot\|_K$.

Theorem 1.1 *Let $a \in A$. Suppose that*

$$(D) \qquad \overline{\hat{A}}^{\|\cdot\|} = C(\Delta).$$

There exists some $b \in A$ with $a = \exp(b)$, iff $\hat{a}$ is null-homotopic in $\mathbb{C}^*$, i.e. there is a continuous mapping $H : [0,1] \times \Delta \to \mathbb{C}^*$ with $H(0,\cdot) \equiv 1$ and $H(1,\cdot) = \hat{a}(\cdot)$.

Proof: Only the backward direction remains for a proof. We have already mentioned that $\exp(A)$ equals the component of $G(A)$ containing e. Therefore for proving a to have a logarithm, it is enough to construct a continuous path taking values in the group of invertible elements and connecting a and e. We will take H and replace it by an appropriate approximation.

Consider the two following subalgebras of $C([0,1]\times\Delta)$

$$\begin{aligned}
\mathcal{A} &:= \Big\{[0,1]\times\Delta \ni (t,h) \longmapsto \sum_{k=1}^{N} f_k(t)\hat{a}_k(h) : \\
&\qquad N\in\mathbb{N},\ f_k\in C([0,1]),\ a_k\in A\ , k=1,\dots,N\Big\}\ , \\
\mathcal{A}_C &:= \Big\{[0,1]\times\Delta \ni (t,h) \longmapsto \sum_{k=1}^{N} f_k(t)g_k(h) : \\
&\qquad N\in\mathbb{N},\ f_k\in C([0,1]),\ g_k\in C(\Delta),\ k=1,\dots,N\Big\}\ .
\end{aligned}$$

Since Δ and $[0,1]$ are compact spaces, they are also regular. Therefore points in $[0,1]\times\Delta$ can be separated using functions $f\in C([0,1])$ and $g\in C(\Delta)$ (Urysohn's lemma). Furthermore, $\mathcal{A}_C$ is closed under conjugation and contains the constant functions. The Stone-Weierstrass' approximation theorem (see [La93], theorem 1.4.) implies

$$C([0,1]\times\Delta) = \overline{\mathcal{A}_C}^{\|\cdot\|_{[0,1]\times\Delta}}.$$

Consider some $\sum_{k=1}^{N} f_k(t)\hat{a}_k(h)\in\mathcal{A}$ and some $\sum_{k=1}^{N} f_k(t)g_k(h)\in\mathcal{A}_C$. The simple inequality

$$\left\|\sum_{k=1}^{N} f_k g_k - \sum_{k=1}^{N} f_k(t)\hat{a}_k\right\|_{I\times\Delta} \le \sum_{k=1}^{N}\|f_k\|_{[0,1]}\|g_k-\hat{a}_k\|_{\Delta}$$

and the assumption (D) forces $\mathcal{A}$ to be dense in $\mathcal{A}_C$. Since we have already proved that $\mathcal{A}_C$ is dense in $C([0,1]\times\Delta)$, we conclude that $\mathcal{A}$ is dense in $C([0,1]\times\Delta)$.

Let $\epsilon := \inf\{|H(t,h)| :\ t\in[0,1], h\in\Delta\}$. The compactness of $[0,1]\times\Delta$ implies $\epsilon>0$. We find some $N\in\mathbb{N}$, $a_1,\dots,a_N\in A$, $f_1,\dots,f_N$, such that for $\tilde{H} := \sum_{k=1}^{N} f_k a_k\in\mathcal{A}$

$$\|\tilde{H}-H\|_{[0,1]\times\Delta} < \frac{\epsilon}{3}\ .$$

$\tilde{H}([0,1]\times\Delta)\subset\mathbb{C}^*$ is true: if not there would be some $(t,h)\in[0,1]\times\Delta$ with $\tilde{H}(t,h)=0$. This would imply $|H(t,h)|\le|\tilde{H}(t,h)|+\|H-\tilde{H}\|_{[0,1]\times\Delta}<\frac{\epsilon}{3}$, a contradiction.

Furthermore,

$$\begin{aligned}
\Big\{x(h) :\ h\in\Delta,\ x\in C(\Delta),\ \|x-\tilde{H}(0,\cdot)\|_{\Delta}<\frac{\epsilon}{3}\Big\} &\subset \mathbb{C}^*, \\
\Big\{x(h) :\ h\in\Delta,\ x\in C(\Delta),\ \|x-\tilde{H}(1,\cdot)\|_{\Delta}<\frac{\epsilon}{3}\Big\} &\subset \mathbb{C}^*,
\end{aligned}$$

again, otherwise there would be some $h \in \Delta$ and some $x \in C(\Delta)$ with $x(h) = 0$ and $\|x - \tilde{H}(1,\cdot)\|_\Delta < \frac{\epsilon}{3}$. This would imply

$$|H(1,h)| \leq \|H - \tilde{H}\|_{[0,1]\times\Delta} + \|\tilde{H}(1,\cdot) - x\|_\Delta + |x(h)| < \frac{2\epsilon}{3},$$

a contradiction.

We define the following two mappings

$$\hat{H} : \left\{ \begin{array}{lll} [0,3]\times\Delta & \rightarrow & \mathbb{C}^* \\ (\alpha,h) & \mapsto & \left\{ \begin{array}{ll} (1-\alpha)\hat{e}(h) + \alpha\sum_{k=1}^N f(0)\hat{a}_k(h)\ , & \text{if} \quad \alpha \in [0,1), \\ \sum_{k=1}^N f(\alpha-1)\hat{a}_k(h), & \text{if} \quad \alpha \in [1,2), \\ (3-\alpha)\sum_{k=1}^N f(1)\hat{a}_k(h) + (\alpha-2)\hat{a}\ , & \text{if} \quad \alpha \in [2,3]\ . \end{array} \right. \end{array} \right. ,$$

$$H_A : \left\{ \begin{array}{lll} [0,3] & \rightarrow & A \\ \alpha & \mapsto & \left\{ \begin{array}{ll} (1-\alpha)e + \alpha\sum_{k=1}^N f(0)a_k\ , & \text{if} \quad \alpha \in [0,1), \\ \sum_{k=1}^N f(\alpha-1)a_k & \text{if}\ , \quad \alpha \in [1,2), \\ (3-\alpha)\sum_{k=1}^N f(1)a_k + (\alpha-2)a\ , & \text{if} \quad \alpha \in [2,3]\ . \end{array} \right. \end{array} \right. .$$

Obviously, $H_A : [0,3] \rightarrow (A, \|\cdot\|)$ is continuous. Furthermore, $\widehat{H_A(\alpha)}(h) = \hat{H}(\alpha,h) \neq 0$. So H_A is the desired continuous path in the group of invertible elements connecting e and a.□

The following theorem provides the uniqueness of „real"-valued elements and establishes a Banach space analogue for logarithmic sheets. We need the concept of an involution (see [Ru91], p.287f). This is a mapping $A \ni a \longmapsto a^* \in A$ with the following properties

$$(x+y)^* = x^* + y^*, \quad (\lambda x)^* = \bar{\lambda}x^*, \quad (xy)^* = y^*x^*, \quad x^{**} = x, \quad \forall\lambda \in \mathbb{C}\, \forall x,y \in A.$$

Note that this definition is made for Banach algebras which are not necessarily commutative. We have $(xy)^* = x^*y^*$ for commutative algebras.

An element $a \in A$ is called *hermitian*iff $x^* = x$. Recall the two following facts: If A is a Banach algebra with some involution then every $a \in A$ has the unique representation $a = u + iv$ with some hermitian $u, v \in A$. If A is commutative and semisimple then every involution is continuous. Let us state the theorem that holds for commutative Banach algebras with unit element e.

Theorem 1.2 *i) If $a = \exp(b)$ with $a, b \in A$ then $a = \exp(b + 2\pi ike)$ for all $k \in \mathbb{Z}$.*
ii) Assume A to be semisimple and Δ to be a connected set. Consider $b, c \in A$ with $\exp(b) = \exp(c)$. Then there is a unique $k \in \mathbb{Z}$ with $b = c + 2\pi ike$. If there is an involution on A and if $a = \exp(b)$ holds for some hermitian a then there exists a unique hermitian $\tilde{b}$ with $a = \exp(\tilde{b})$.

Proof: i) This is an easy consequence of $\exp(2\pi i k e) = e$ for all $k \in \mathbb{Z}$.

ii) Assume A to be semisimple. Consider $b, c \in A$ with $\exp(b) = \exp(c)$. This implies $\exp(b-c) = e$. Then for all $h \in \Delta$ the equality $\exp(h(b-c)) = 1$ holds, hence $h(b-c) \in 2\pi i\mathbb{Z}$. Since Δ is connected and the Gelfand transform of $b-c$ is a continuous function on Δ , there can only exist one $k_0 \in \mathbb{Z}$ with $h(b-c) = 2\pi k_0 i$ for all $h \in \Delta$. The semisimplicity implies $b - c = 2\pi i k_0 e$.

Now let us assume A to have an involution. Consider a hermitian $a \in A$ and $b \in A$ with $\exp(b) = a$. b has an unique representation $b = b_1 + ib_2$ with $b_i \in A$ hermitian, $i = 1, 2$. If A is commutative and semisimple then every involution is continuous, therefore $\exp(b_1 - ib_2) = a^* = a = \exp(b_1 + b_2 i)$ holds. The above reasoning shows that $b_1 - ib_2 + 2\pi i k_0 e = b_1 + ib_2$ for some $k_0 \in \mathbb{Z}$. The uniqueness of the representation yields $b_2 = \pi k_0 e$ and the statement follows with $\tilde{b} := b_1$.□

Let us summarize the considerations made above. Let A be a commutative semisimple Banach algebra with unit e and some involution. If we consider A_h to be the set of hermitian elements, then A_h is a Banach algebra over the field of real numbers. We have proved then that $a = \exp(b)$ for $a, b \in A_h$ iff the $\hat{a}$ is null-homotopic in $\mathbb{C}^*$. b is unique. If we define $U = \{a \in A_h :\ a(h)$ null-homotopic in $\mathbb{C}^*\}$ then we have a well defined mapping $\log : U \to A_h$ with $\exp(\log(a)) = a$.

Some remarks on the smoothness: in later applications we are interested in the continuity and differentiability properties of log. Consider the exponential mapping exp. This is an analytic mapping on A_h, its Fréchet derivative at $a \in A$ is given by the bounded linear operator $b \longmapsto \exp(a) * b$. This operator is invertible. The inverse mapping (see [La93], p.361) is given by $b \longmapsto \exp(-a) * b$. Hence the inverse mapping theorem ([La93], p.361), provides a local inversion of exp that is at least C^∞. The local inversion of exp defines a real valued logarithm in an open neighbourhood of $\exp(a)$ with respect to A_h. The uniqueness of the logarithm implies by one stroke that log must be at least C^∞ and U is an open subset of A_h.

1.1.2 Examples

Details can again be found in [Ru91]. We give only a short survey and show how to apply the results of the last sections, especially how to refine them from the two-sided case to the one-sided one.

Example 1.3 Consider the space of two-sided absolutely summable sequences

$$\ell^1_{\mathbb{C}} := \left\{ (z_k) \subset \mathbb{C}^{\mathbb{Z}} : \sum_{l=-\infty}^{\infty} |z_k| < \infty \right\}$$

with the usual ℓ^1 norm. Equipped with the convolution

$$(a * b)_k = \sum_{l \in \mathbb{Z}} a_l b_{k-l},$$

$\ell^1_{\mathbb{C}}$ turns out to be a commutative Banach algebra with unit $\delta_0 = (\delta_{0j})_j$ (δ_{ij} is the Kronecker symbol). The Gelfand space Δ can be identified with the unit circle $S_1 = \{|z| = 1\}$. The Gelfand transforms are the usual Fourier transforms, i.e. $\hat{a}(h) = \sum_{k\in\mathbb{Z}} a_k h^k$. Note that (D) is satisfied here. $\ell^1_{\mathbb{C}}(\mathbb{Z})$ is semisimple. If we consider the mapping $a \longmapsto \bar{a} = (\bar{a}_k)$ then this defines an involution. We can use the results for the algebra of hermitian elements that is in fact the space of two-sided real valued sequences $\ell^1_{\mathbb{R}}(\mathbb{Z})$. A subspace useful for applications is the space of one sided real valued sequences $\ell^1 := \ell^1_{\mathbb{R}}(\mathbb{N}_0)$. Suppose that we have an element $a \in \ell^1$ lying in the set U described in the last section. Then there is a unique $b \in \ell^1_{\mathbb{R}}(\mathbb{Z})$ with $\exp(b) = a$. Is this b also an element of ℓ^1? This is true. Obviously, $\ell^1_{\mathbb{R}}(\mathbb{Z}) = \ell^1_{\mathbb{R}}(-\mathbb{N}) \oplus \ell^1$. Hence there exists $b_1 \in \ell^1(-\mathbb{N}), b_2 \in \ell^1$ with $b = b_1 + b_2$. Therefore $a = \exp(b_1) * \exp(b_2)$ and hence

$$\underbrace{\exp(-b_2) * a}_{\in\ell^1} = \exp(b_1) = e + \underbrace{\sum_{l=1}^{\infty} \frac{1}{l!} b_1^{*l}}_{\in\ell^1(-\mathbb{N})}.$$

This shows that $\exp(-b_2) * a = e$. Hence a has a logarithm in ℓ^1. From the uniqueness it follows that $b_1 = 0$. We will show that this logarithm is given by the Panjer inversion. Assume that $a \in \ell^1$ with $\exp(b) = a$. Write $b = \lambda(x - \delta_0)$ for some $\lambda > 0$ and $x \in \ell^1$ with $x_0 = 0$. The Fourier transforms are leading to the equation $\hat{a}(z) = \exp(\lambda(\hat{x}(z)-1))$ for all $|z| = 1$. Since a and x are one-sided we can also consider the power series or generating functions, i.e. $\hat{a}(z) = \exp(\lambda(\hat{x}(z)-1))$ holds for all $|z| \leq 1$. The usual calculations can be made (see [Pa92], p 171). We state them here for completeness. Differentiating both sides yields

$$\begin{aligned}\sum_{k=0}^{\infty}(k+1)a_{k+1}z^k &= \lambda \exp(\lambda(\hat{x}(z)-1)) \sum_{k=1}^{\infty} k x_k z^{k-1} \\ &= \lambda \sum_{k=0}^{\infty} a_k z^k \sum_{k=0}^{\infty}(k+1)x_{k+1}z^k = \lambda \sum_{k=0}^{\infty} z^k \sum_{l=1}^{k+1} l x_l a_{k+1-l}.\end{aligned}$$

Comparing the coefficients we derive

$$a_{k+1} = \frac{\lambda}{k+1} \sum_{l=1}^{k+1} l x_l a_{l+1-l}.$$

This is the Panjer recursion formula. Furthermore, we have

$$a_0 = \hat{a}(0) = \exp(-\lambda)$$

Hence $\lambda = -\log(a_0)$.

Example 1.4 Consider the space of complex valued functions $L^1_{\mathbb{C}}(\mathbb{R})$ on the real line that are integrable with respect to the Lebesgue measure. Adjungate the Dirac measure δ_0 to it. Again equipped with the usual norm

$$\|\alpha\delta_0 + f\| := |\alpha| + \int |f| \, dx$$

and the convolution

$$(\alpha\delta_0 + f) * (\beta\delta_0 + g) = \alpha\beta\delta_0 + \alpha g + \beta f + \int f(x)g(\cdot - x) \, dx,$$

we have a commutative Banach algebra with unit element δ_0. The maximal ideal space can be identified with the one point compactification $\hat{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$. The Gelfand transforms are given by

$$h_t(\alpha\delta_0 + f) = \alpha + \int \exp(itx)f(x) \, dx \quad \forall t \in \mathbb{R}, \qquad h_\infty(\alpha\delta_0 + f) = \alpha.$$

Note that again Δ is connected and that the space of Gelfand transforms is dense in $C(\Delta)$ as can be shown using the Stone-Weierstrass Theorem. The same considerations are true when going from the two-sided case to the one sided one. We show in a later chapter that recursion formulas analogue to Panjer's in the counting density case can also be found for histograms.

Example 1.5 Consider a compact Hausdorff space K and the space of continuous functions $C(K)$. $C(K)$ with the usual pointwise operations is a commutative Banach algebra with unit $K \ni k \longmapsto 1$. Δ can be identified with K. The Gelfand transforms are the pointwise evaluations $t \longmapsto f(t)$. Hence (D) is trivial here. Hence a continuous function f has a continuous logarithm g, i.e. $f = \exp(g)$, iff f is null-homotopic in $\mathbb{C}^*$.

We cite here a theorem that is taken from [Ga69] (theorem and corollary, p 86f). Its proof is based on multivariate complex function theory. Again, A is a commutative Banach algebra with an identity. It can be viewed as an implicit mapping theorem.

Theorem 1.6 *Let $a_0, \dots, a_n \in A$. Let $g \in C(\Delta)$ and $\sigma(g, a_0, \dots, a_n)$ be the set of $(n+2)$-tuples $(g(h), \hat{a}_0(h), \dots, \hat{a}_n(h))$, $h \in \Delta$. Let $F(\omega, z_0, \dots, z_n)$ be a function analytic in a neighbourhood of $\sigma(g, a_0, \dots, a_n)$, such that $F(g, \hat{a}_0, \dots, \hat{a}_n) = 0$, while $\partial F/\partial w$ does not vanish on $\sigma(g, a_0, \dots, a_n)$. Then there exists a unique element $b \in A$ such that $g = \hat{b}$ and $F(b, a_0, \dots, a_n) = 0$.*

We consider the mapping $(w, z) \longmapsto F(w, z) := \exp(w) - z$. Obviously, F is analytic on $\mathbb{C}^2$ and $\partial F/\partial w = \exp(w) \neq 0$ for all $w \in \mathbb{C}$. Assume now that $a \in A$ has a null-homotopic Gelfand transform in $\mathbb{C}^*$. Example 1.5 provides a continuous function $g \in C(\Delta)$ with $\hat{a} = \exp(g)$, i.e. $F(g(h), \hat{a}(h)) = 0$ for all $h \in \Delta$. The theorem then yields an unique $b \in A$ with $\hat{b} = g$ and $F(b, a) = 0$, i.e. $a = \exp(b)$. This reasoning shows that the asumption (D) can be removed and theorem 1.1 holds in full generality.

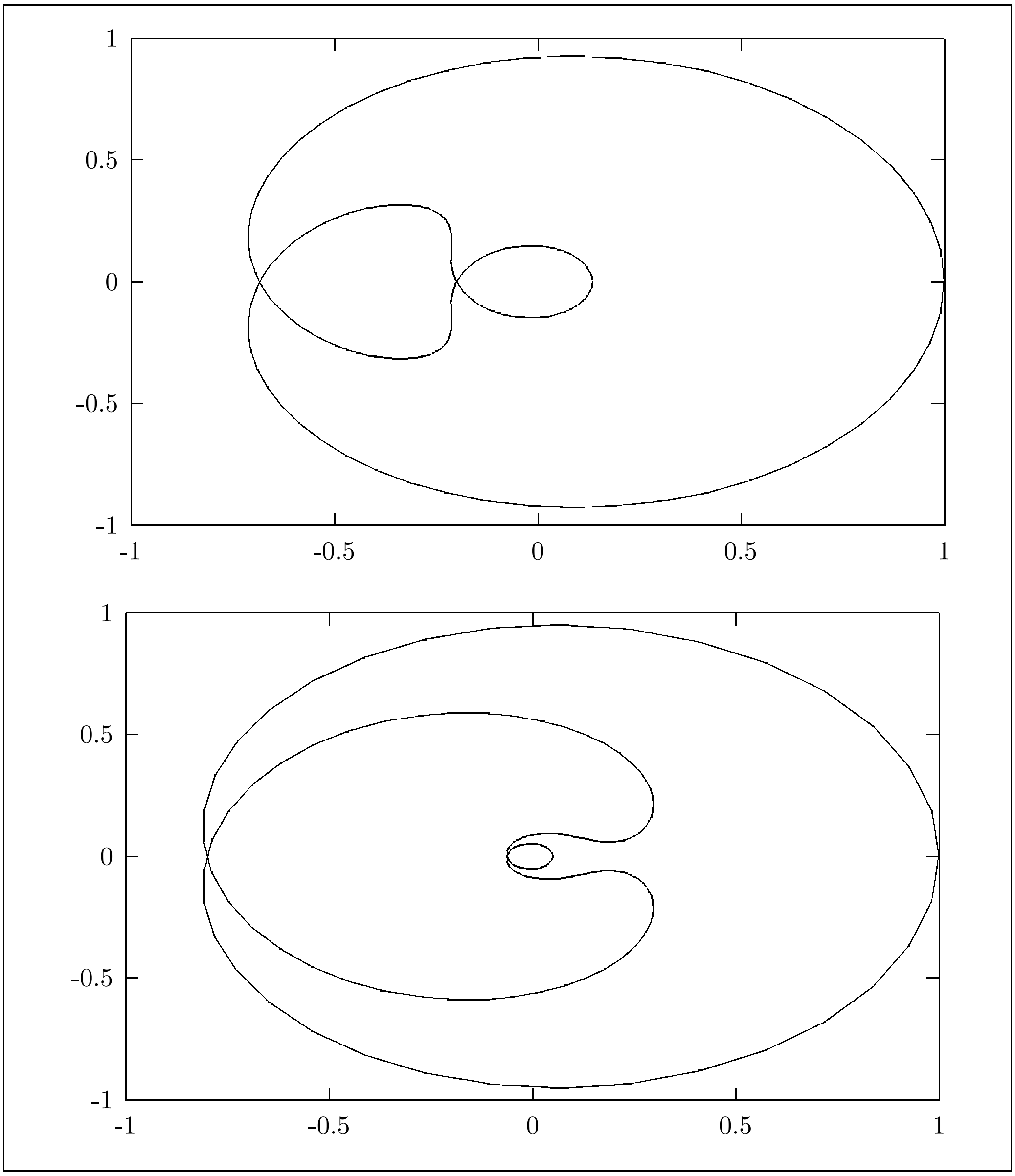

Figure 1.1: The Fourier transforms of the Poisson distributions with parameter $\lambda = 4, 6$. The fourth root was applied to their moduli.

Chapter 2

Panjer Inversion based Estimators

2.1 Introduction

Let us assume that the claim distribution is concentrated on $\mathbb{N}$. We want to estimate λ and P.

We identify P and Q with their counting density p and q, respectively. Recall that p, λ and q are connected via the Panjer recursion formula

$$q_0 = e^{-\lambda}, \qquad q_k = \frac{\lambda}{k} \sum_{l=1}^{k} l p_l q_{k-l}.$$

This recursion can be inverted, leading to an inverse Panjer recursion

$$\begin{aligned} \lambda &= -\log(q_0), \\ \lambda e^{-\lambda} p_k &= q_k - \frac{\lambda}{k} \sum_{l=1}^{k-1} l p_l q_{k-l}. \end{aligned}$$

This leads to a simple plug-in estimator: Estimate q_k using the relative frequencies $q_k^n := \frac{1}{n} \sum_{l=1}^{n} 1_{Y_l = k}$, $k \in \mathbb{N}_0$. Then calculate an estimator λ^n and an estimator p^n for λ and p using the inverse Panjer recursion formula with q_k^n instead of q_k. We have seen in the last chapter, example 1.3, that if the Fourier transform of q^n is null-homotopic, then p^n is an absolut summable sequence. The Fourier transform is the empirical characteristic function $\hat{q}^n(\theta) = \frac{1}{n} \sum_{l=1}^{n} e^{i\theta Y_k}$. Section 2.2 discusses the Panjer inversion again. It is shown that the plug-in estimator will be in the range of the exponential function with probability one if the sample size n is large enough. Some notation will be given. A large deviation upper bound will be given for the non-null-homotopy of the empirical characteristic function. Section 2.3 gives some calculations of the derivatives of the underlying mappings concerning the plug-in estimator and shows that they have a very simple structure and can easily be computed. Note the analogue in the univariate case $\exp(x)' = \exp(x)$ and $\log(x)' = 1/x$. Section 2.4 shows strong consistency and asymptotic normality of the estimator in ℓ^1. Section 2.5 investigates the naive projection estimator if only a finite segment of p^n is calculated. The end point is data driven.

2.2 A Plug-In Estimator

We want to apply the results of the last chapter. Consider (λ, p) and q as elements of the space $\ell^1 := \ell^1_{\mathbb{R}}(\mathbb{N}_0)$. Then with exp denoting the exponential function in ℓ^1 we have the equation

$$q = \exp(\lambda(p - \delta_0)).$$

We have shown the existence of an open subset U of ℓ^1 consisting of those elements whose Fourier transforms are null-homotopic in $\mathbb{C}^*$ and a mapping, the unique real logarithm log, such that

$$a = \exp(\log(a)) \qquad \forall a \in U.$$

Obviously, $\lambda(p - \delta_0)$ is real valued, hence $\log q = \lambda(p - \delta_0)$. Let $T_k : \ell^1 \to \mathbb{R}$ be the projection on the kth coordinate $T_k(x) := x_k$, $k \in \mathbb{N}_0$ and $T_0^\perp : \ell^1 \to \ell^1$, $T_0^\perp(x) := (0, x_1, x_2 \dots) \in \ell_1$. Then we have the following equalities

$$\begin{aligned} \lambda &= -T_0 \log(q) \\ p &= \frac{1}{\lambda} T_0^\perp \log q = -\frac{1}{T_0 \log(q)} T_0^\perp \log q. \end{aligned}$$

Again note that this is only a compact way to write down the inverse Panjer recursion.

We want to give ℓ^1 a measurable structure. ℓ^1 is a separable Banach space. Let $\mathcal{B}$ be the Borel σ-algebra, i.e. the σ-algebra generated by the open balls in ℓ^1. Since ℓ^1 is separable, a mapping f from a measurable space into $(\ell^1, \mathcal{B})$ is measurable iff $T_k \circ f$ is measurable (see [Va87], p 17). Hence the sequence $q^n = (q^n_k)$ of relative frequencies is a random variable taking values in ℓ^1.

Let us define an estimator $\theta^n = (\lambda^n, p^n)$ taking values in ℓ^1. Define

$$\theta^n := -T_0 \log(q^n)\delta_0 - \frac{1}{T_0 \log(q^n)} T_0^\perp \log q, \qquad \text{if } q^n \in U \cap \{T_0 \neq 1\},$$

and $\theta^n := \delta_0 + \delta_1$, otherwise. This is a measurable mapping taking values in ℓ^1.

Note that the first component returns an estimator for λ and the sequence $(T_1\theta^n, T_2\theta^n, \dots)$ an estimator for p.

First we have a look at the Fourier transform $\hat{q}^n$ of q^n. It is the empirical characteristic function, i.e.

$$\hat{q}^n(\theta) = \frac{1}{n}\sum_{l=1}^{n} e^{i\theta Y_l}.$$

The law of large numbers implies that $\hat{q}^n(\theta)$ tends to $\hat{q}(\theta)$ a.s.. The pointwise convergence of characteristic functions can be strengthened to local uniform convergence (see [Lu70], p.50). Especially, $\hat{q}^n$ tends uniformly to $\hat{q}$ on the compact interval $[0, 2\pi]$ a.s.. The next lemma shows how uniform convergence is connected to null-homotopy.

Lemma 2.1 *Consider $C(K)$ for a compact set K.*
i) Consider some $f \in C(K)$ that is null-homotopic in $\mathbb{C}^*$. Let $g \in C(K)$. If

$$\|f - g\|_K < \inf_K |f|$$

then g is null-homotopic in $\mathbb{C}^*$.
ii) Let $x \in U$ and $y \in \ell_1$. If

$$\|\hat{x} - \hat{y}\|_{[0,2\pi]} < \inf_{[0,2\pi]} |\hat{x}|$$

then $y \in U$.

iii) Consider $f_n, f \in C(K)$, $n \in \mathbb{N}$, with $\|f_n - f\|_K \to 0$. Suppose f to be null-homotopic in $\mathbb{C}^*$. Then there exists some n_0, such that f_n is null-homotopic in $\mathbb{C}^*$ for $n > n_0$.

Proof: Obviously, $\epsilon := \inf_{t\in K} |f| > 0$, since K is a compact set.
i) Let H_1 be a homotopy between f and 1 in $\mathbb{C}^*$. Define

$$H(\alpha, t) := (1-\alpha)g(t) + \alpha f(t).$$

Then H is a continuous mapping taking values in $\mathbb{C}$. We show that $H \in C([0,1] \times K, \mathbb{C}^*)$. If there is some $\alpha_0 \in [0,1], t_0 \in K$ with $H(\alpha_0, t_0) = 0$, then we would have the following contradiction

$$\inf_{t\in K} |f| \le |f(t_0) - H(\alpha_0, t_0)| = (1-\alpha_0)|f(t_0) - g(t_0)| < \inf_{t\in K} |f| \ .$$

H is a homotopy between g and f in $\mathbb{C}^*$. We therefore can define a homotopy between g and 1 in $\mathbb{C}^*$ using the homotopy $\tilde{H}$ with $\tilde{H}(\alpha, t) := H(2\alpha, t)$, $\alpha \in [0, 1/2)$, $\tilde{H}(\alpha, t) := H_1(2\alpha - 1, t)$, $\alpha \in [1/2, 1]$, $t \in K$.
ii) is obvious.
iii) Since $\epsilon > 0$ there is an n_0 such that $\|f_n - f\| < \epsilon$ for all $n > n_0$. Applying the first part of the lemma we conclude that all f_n, $n > n_0$, are null-homotopic in $\mathbb{C}^*$.$\square$

This lemma shows that the empirical characteristic function $\hat{q}^n$ is null-homotopic for n large enough with probability one and θ^n will be given by the Panjer inversion formula. Especially, the unbounded oscillations noted by [Hu90] will vanish for n large enough, since our estimator is ℓ^1-valued.

We use the lemma to bound the probability that $q^n \notin U$, i.e. the empirical characteristic function $\hat{q}^n$ is not null-homotopic. We will show that this event is a rare event in the sense of the theory of large deviations (see [De93]). The probability of the event $q^n \notin U$ decreases exponentially fast. $\Re z$ denotes the real part of a complex number z.

Theorem 2.2

$$\limsup_{n\to\infty} \frac{1}{n} \log P(q^n \in U^C) \leq -\frac{1}{8} \exp\left(2\lambda\left(\inf_{t\in T} \Re\hat{p}(t) - 1\right)\right) .$$

Proof: We have the inequality $\|\hat{q}^n - \hat{q}\|_{[0,2\pi]} \leq \|q^n - q\|_1$. Hence the following set theoretic inclusion holds:

$$\{\|\hat{q}^n - \hat{q}\|_{[0,2\pi]} \geq \epsilon\} \subset \{\|q^n - q\|_1 \geq \epsilon\} =: B.$$

We want to apply the upper bound of Sanov's theorem (see [De93], corollary 6.2.3). We state it here in a simplified form:

$$\limsup_{n\to\infty} \frac{1}{n} \log P(q^n \in A) \leq -\inf_{x\in A} H(x|q)$$

for all A, closed in the weak topology on the space of probability measures on $\mathbb{N}_0$. $H(x|q)$ denotes the Kullback-Leibler divergence, defined by

$$H(x|q) = \sum_{l=0}^{\infty} x_k \log \frac{x_k}{q_k},$$

if the support of x is a subset of the support of q (i.e. the Radon-Nikodym derivative dx/dq exists), and $H(x|q) = \infty$, otherwise.

Obviously, the weak topology equals the topology of pointwise convergence of the probability densities. In view of Scheffé's theorem the topology of pointwise convergence of probability densities is the same as the $\|\cdot\|_1$-norm topology restricted to the set of probabilty densities. Hence the large deviation upper bound can be used with B as defined above.

We now want to estimate the upper bound $-\inf_{x\in B} H(x|q)$. The $\|\cdot\|_1$-norm is a lower bound for the Kullback-Leibler distance. It holds that

$$\frac{1}{8}\|x - q\|_1^2 \leq H(x|q).$$

Hence we have the inequalities

$$-\inf_B H(x|q) \leq -\frac{1}{8} \inf_B \|x - q\|_1^2 \leq -\frac{1}{8}\epsilon^2.$$

If we choose $\epsilon := \inf_{[0,2\pi]} |\hat{q}|$ then lemma 2.1 ii) provides the upper bound

$$\limsup_{n\to\infty} \frac{1}{n} \log P(q^n \notin U) \leq \limsup_{n\to\infty} \frac{1}{n} \log P(\|q^n - q\|_1 \geq \epsilon) \leq -\frac{1}{8} \inf_{[0,2\pi]} |\hat{q}|^2.$$

Since $|\hat{q}|^2 = \exp(2\lambda(\Re\hat{p}-1))$, the assertion of the theorem follows by the monotony of the exponential function.□

The upper bound is of course a function of λ. As p is not the Dirac measure concentrated in zero, we have $\inf \Re\hat{p} < 1$. The upper bound tends to 0 exponentially fast with increasing λ. This again is a hint that our methods are not well suited for large λ.

2.3 The mappings Ψ and L

The linear function $T_0 : \ell^1 \to \mathbb{R}$ can be regarded as an element of the dual of ℓ^1, so we write $\langle T_0, x\rangle$ instead of $T_0 x$. Define the function

$$L(x) := -\langle T_0, \log(x)\rangle\delta_0 - \frac{1}{\langle T_0, \log(x)\rangle} T_0^{\perp} \log x, \qquad \text{if } x \in U,\ x_0 < 1,$$

and $L(x) := \delta_0 + \delta_1$, else. For shortness, let us denote $U' = U \cap \{x \in \ell^1 : \langle T_0, x\rangle < 1\}$. This is obviously an open subset of ℓ^1. Then L maps ℓ^1 to ℓ^1. We also have by definition $\theta^n = L(q^n)$ and $L(q) = \lambda\delta_0 + p$. Let us also define the compounding mapping

$$\ell^1 \ni x \to \Psi(x) := e^{\langle T_0, x\rangle)(T_0^{\perp} x - \delta_0)} \in \ell^1.$$

Of course, these mappings are inverse to each other. To be more specific, for each $u \in U'$ we have $\Psi \circ L(u) = u$ and for every $v \in V := \Psi^{-1}(U')$ we have $L \circ \Psi(v) = v$. Obviously, Ψ is continuous, therefore V is an open subset of ℓ^1. Ψ is even a $C^\infty(\ell^1, \ell^1)$ mapping. We will see in the next lemma that its derivative $A := \Psi'_{\lambda\delta_0+p}$ is bijective, hence Ψ is a local C^∞-diffeomorphism. Its local inverse Ψ^{-1}, i.e. L, is therefore a C^∞-mapping, too. The derivative of L in $\Psi(\lambda\delta_0 + p)$ is the inverse of A. The next two lemmas will be devoted to the calculation of A and A^{-1}.

Lemma 2.3 *i) Ψ's derivative at $\lambda\delta_0 + p$ is given by*

$$Ah := \Psi'_{\lambda\delta_0+p} h = \langle T_0, h\rangle(q * p - (1+\lambda)q) + \lambda q * h.$$

A is an isomorphism mapping ℓ^1 onto ℓ^1.

ii) Let $L_{U'}$ be the restriction of L to the set U'. Then $L_{U'}$ is continuous and Fréchet-differentiable. If $q = exp(\lambda(p - \delta_0))$, then the derivative is given by the operator

$$L'_q h = e^{\lambda}\langle T_0, h\rangle \left(\frac{1}{\lambda} p - \left(1 + \frac{1}{\lambda}\right) q\right) + \frac{1}{\lambda} q^{-1} * h.$$

iii) $A = (L'_q)^{-1}$.

Proof: Ψ is a composition of two C^∞-mappings, $x \to g(x) := \langle T_0, x\rangle(T_0^{\perp} x - \delta_0)$ and the exponential function. Both are Fréchet-differentiable. The derivative of Ψ at $\lambda\delta_0 + p$ can be calculated from the chain rule. We have

$$\begin{aligned}
g'_{\lambda\delta_0+p} h &= \langle T_0, h\rangle(T_0^{\perp}(\lambda\delta_0 + p) - \delta_0) + \langle T_0, \lambda\delta_0 + p\rangle T_0^{\perp} h \\
&= \langle T_0, h\rangle(p - \delta_0) + \lambda T_0^{\perp} h, \\
\exp'_{\lambda(p-\delta_0)} h &= e^{\lambda(p-\delta_0)} * h = q * h, \\
Ah = \Psi'_{\lambda\delta_0+p} h &= \exp'_{\lambda(p-\delta_0)} g'_{\lambda\delta_0+p} h \\
&= \langle T_0, h\rangle q * (p - \delta_0) + \lambda \underbrace{q * T_0^{\perp} h}_{= q*(h - \langle T_0, h\rangle\delta_0)} \\
&= \langle T_0, h\rangle(q * p - (1+\lambda)q) + \lambda q * h.
\end{aligned}$$

ii) Note that $L_{U'} = g \circ \log_{U'}$ with

$$g : \{T_0 \neq 0\} \to \ell^1, \qquad g(x) = -\langle T_0, x\rangle \delta_0 - \frac{1}{\langle T_0, x\rangle} T_0^{\perp} x.$$

Since both g and $\log_{U'}$ are differentiable, we can use the chain rule again. As seen before (see the remark at the end of section 1.1.1)

$$(\log)'_q h = q^{-1} * h.$$

The derivative of g is

$$\begin{aligned} g'_{\lambda(p-\delta_0)} h &= -\langle T_0, h\rangle \delta_0 + \frac{1}{\langle T_0, \lambda(p-\delta_0)\rangle^2} \langle T_0, h\rangle T_0^{\perp} \lambda(p-\delta_0) - \frac{1}{\langle T_0, \lambda(p-\delta_0\rangle} T_0^{\perp} h \\ &= -\langle T_0, h\rangle \delta_0 + \frac{1}{\lambda}\langle T_0, h\rangle p + \frac{1}{\lambda} T_0^{\perp} h \end{aligned}$$

Note that

$$\langle T_0, q^{-1}\rangle = \langle T_0, e^{\lambda(\delta_0 - p)}\rangle = e^{\lambda}$$

and

$$\langle T_0, q^{-1} * h\rangle = \langle T_0, e^{\lambda(\delta_0-p)} * h\rangle = e^{\lambda}\langle T_0, h\rangle.$$

Therefore

$$\begin{aligned} L'_q h &= g'_{\lambda(p-\delta_0)}(q^{-1} * h) \\ &= -\langle T_0, q^{-1} * h\rangle \delta_0 + \frac{1}{\lambda}\langle T_0, q^{-1} * h\rangle p + \frac{1}{\lambda} T_0^{\perp}(q^{-1} * h) \\ &= -e^{\lambda}\langle T_0, h\rangle \delta_0 + \frac{e^{\lambda}}{\lambda}\langle T_0, h\rangle p + \frac{1}{\lambda}\left(q^{-1} * h - \langle T_0, q^{-1} * h\rangle \delta_0\right) \\ &= e^{\lambda}\langle T_0, h\rangle \left(\frac{1}{\lambda} p - \left(1 + \frac{1}{\lambda}\right)\delta_0\right) + \frac{1}{\lambda} q^{-1} * h. \end{aligned}$$

iii) It is straightforward to prove that

$$L'_q \Psi'_{\lambda\delta_0+p} = \Psi'_{\lambda\delta_0+p} L'_q = \mathrm{id}_{\ell^1,\ell^1}$$

□

For later applications we derive the partial derivatives for Ψ and L too. For $0 < l \leq k$ we have for example

$$\begin{aligned} T_k \Psi'_{\lambda\delta_0+p} \delta_l &= 0 + \lambda T_k q * \delta_l = \lambda q_{k-l}, \\ T_k \Psi'_{\lambda\delta_0+p} \delta_0 &= T_k q * p - (1+\lambda) q_k + \lambda q_k = \sum_{l=1}^{k} p_l q_{k-l} - q_k. \end{aligned}$$

The components of the directional derivatives of $\Psi'_{\lambda\delta_0+p}\delta_l$ define an infinite triangular matrix of the following simple form:

$$(T_k\Psi'_{\lambda\delta_0+p}\delta_l)_{\substack{0\le k<\infty\\0\le l<\infty}} = \left(\begin{array}{ccccc} -q_0 & 0 & \cdots & & 0 \\ p_1q_0-q_1 & \lambda q_0 & \ddots & & \\ p_1q_1+p_2q_0-q_2 & \lambda q_1 & \ddots & & \\ p_1q_2+p_2q_1+p_3q_0-q_3 & \lambda q_2 & \ddots & \ddots & \\ \vdots & \vdots & \ddots & \ddots & \lambda q_0 \\ \sum_{l=1}^k p_lq_{k-l}-q_k & \lambda q_{k-1} & \lambda q_{k-2} & \cdots & \ddots \\ \vdots & \vdots & & & \ddots \quad \ddots \end{array}\right)$$

If $p=\delta_1$ (equivalently $q\in\mathcal{P}$) then we have an even simpler form

$$(T_k\Psi'_{\lambda\delta_0+\delta_1}\delta_l)_{\substack{0\le k<\infty\\0\le l<\infty}} = \left(\begin{array}{cccccc} -e^{-\lambda} & 0 & & \cdots & & 0 \\ e^{-\lambda}\lambda\left(\frac{1}{\lambda}-1\right) & \lambda e^{-\lambda} & \ddots & & & 0 \\ e^{-\lambda}\frac{\lambda^2}{2!}\left(\frac{2}{\lambda}-1\right) & \lambda^2e^{-\lambda} & \ddots & & & 0 \\ e^{-\lambda}\frac{\lambda^3}{3!}\left(\frac{3}{\lambda}-1\right) & \frac{\lambda^3}{2!}e^{-\lambda} & \ddots & \ddots & & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & 0 \\ e^{-\lambda}\frac{\lambda^S}{S!}\left(\frac{S}{\lambda}-1\right) & \frac{\lambda^S}{(S-1)!}e^{-\lambda} & \frac{\lambda^{S-1}}{(S-2)!}e^{-\lambda} & \cdots & & \lambda e^{-\lambda} \\ \vdots & \vdots & & & & \quad \ddots \end{array}\right).$$

Analogously, the matrix associated with L can be computed as follows

$$T_kL'_q\delta_l = e^\lambda\langle T_0,\delta_l\rangle\left(\frac{1}{\lambda}\langle T_k,p\rangle-\left(1+\frac{1}{\lambda}\right)\langle T_k,\delta_0\rangle\right)+\frac{1}{\lambda}\langle T_k,q^{-1}*\delta_l\rangle.$$

The matrix has the following form:

$$(T_kL'_q\delta_l)_{\substack{0\le k\le\infty\\0\le l\le\infty}} = \left(\begin{array}{ccccccc} -e^\lambda & 0 & 0 & 0 & 0 & \cdots & \\ \frac{e^\lambda}{\lambda}p_1+\frac{1}{\lambda}q_1^{-1} & \frac{e^\lambda}{\lambda} & 0 & 0 & 0 & \cdots & \\ \frac{e^\lambda}{\lambda}p_2+\frac{1}{\lambda}q_2^{-1} & \frac{1}{\lambda}q_1^{-1} & \frac{e^\lambda}{\lambda} & 0 & 0 & \cdots & \\ \frac{e^\lambda}{\lambda}p_3+\frac{1}{\lambda}q_3^{-1} & \frac{1}{\lambda}q_2^{-1} & \frac{1}{\lambda}q_1^{-1} & \frac{e^\lambda}{\lambda} & 0 & \cdots & \\ \vdots & \vdots & & & & & \\ \frac{e^\lambda}{\lambda}p_k+\frac{1}{\lambda}q_k^{-1} & \frac{1}{\lambda}q_{k-1}^{-1} & \cdots & \cdots & \frac{1}{\lambda}q_1^{-1} & \frac{e^\lambda}{\lambda} & \\ \vdots & \vdots & & & & & \ddots \end{array}\right).$$

If $q\in\mathcal{P}$ then $q^{-1}=e^{\lambda(\delta_0-\delta_1)}=\sum_{k=0}^\infty\frac{(-\lambda)^k}{k!}e^\lambda\delta_k$. Therefore the components of q^{-1} are

$$q_k^{-1}=(-1)^ke^\lambda\frac{\lambda^k}{k!}.$$

The matrix then has the form

$$(T_k L'_q \delta_l)_{\substack{0\le k\le\infty \\ 0\le l\le\infty}} = \begin{pmatrix} -e^\lambda & 0 & 0 & 0 & 0d & \cdots \\ \frac{e^\lambda}{\lambda}(1-\lambda) & \frac{e^\lambda}{\lambda} & 0 & 0 & 0 & \cdots \\ e^\lambda \frac{\lambda}{2} & -e^\lambda & \frac{e^\lambda}{\lambda} & 0 & 0 & \cdots \\ -e^\lambda \frac{\lambda^2}{3!} & \frac{e^\lambda \lambda}{2} & -e^\lambda & \frac{e^\lambda}{\lambda} & 0 & \cdots \\ \vdots & & & \ddots & \ddots & \ddots \end{pmatrix}.$$

2.4 Consistency and Asymptotic Normality

We will use the continuity and differentiability properties of L to establish consistency and asymptotic normality.

Let us first prove strong consistency.

Theorem 2.4 *The following is true:*

$$\begin{aligned} (i) \qquad & \|q^n - q\|_1 \xrightarrow{a.s.} 0, \\ (ii) \qquad & \|\theta^n - (\lambda, p)\|_1 \xrightarrow{a.s.} 0. \end{aligned}$$

Proof: i) This is a direct corollary from Scheffé's theorem and the a.s. pointwise convergence of the densities $k \longmapsto q_k^n$ to the density $k \longmapsto q_k$.

ii) Fix an ω such that $q^n(\omega) \to q$ in ℓ^1. Since $q \in U'$ and U' is open in ℓ^1, we have $q^n(\omega) \in U'$ for n large enough. The continuity of $L_{U'}$ guarantees that

$$\lim_{n\to\infty} L(q^n(\omega)) = \lim_{n\to\infty} L_{U'}(q^n(\omega)) = L_{U'}(q) = L(q) = \lambda\delta_0 + p.$$

$\square$

To obtain the asymptotic normality of θ^n we establish asymptotic normality for q^n and apply the delta method. The next theorems will establish asymptotic normality for the relative frequencies.

First we want to give a heuristic argument that there must be some condition on the decay of the sequence (q_k). Let us consider the empirical distribution function $F^n(t) := \frac{1}{n}\sum_{l=1}^n 1_{[0,t]}(Y_l)$. Let F be the distribution function of Y. In particular, we have weak convergence of the family

$$(\sqrt{n}(F^n(t) - F(t)))_{t\in T} \xrightarrow{\mathcal{D}} (B_{F(t)})_{t\in T} = (W_{F(t)} - F(t)W_1)_{t\in T}$$

for a finite subset T of $[0,1]$, W denoting the Wiener process and B denoting a Brownian bridge process.

The relative frequencies can be calculated from F^n via $q_k^n = F^n(k) - F^n(k-1)$, $k \in \mathbb{N}_0$, $F^n(-1) := 0$. The same is true for q: $q_k = F(k) - F(k-1)$, $k \in \mathbb{N}_0$, $F(-1) := 0$.

Applying the continuous mapping theorem we derive weak convergence of the following finite dimensional families

$$(\sqrt{n}(q_k^n - q_k)_{k\in I} \xrightarrow{\mathcal{D}} (B_{F(k)} - B_{F(k-1)})_{k\in I}.$$

Hence $(B_{F(k)} - B_{F(k-1)})_{k\in\mathbb{N}_0}$ is the only candidate as a limit with respect to weak convergence of $\sqrt{n}(q^n - q)$.

However, there must be made some assumptions on the decay of q for $k \to \infty$ if we do not want to have a „mass defect“. Consider the sequence

$$\begin{aligned}(B_{F(k)} - B_{F(k-1)})_{k\in\mathbb{N}_0} &= (W_{F(k)} - W_{F(k-1)} - q_k W_1)_{k\in\mathbb{N}_0},\\ W_{F(-1)} &= 0.\end{aligned}$$

We need (Z_k) to be an element of ℓ^1 a.s.. The term $q_k W_1$ behaves nicely, because $(W_1 q_k)_{k\in\mathbb{N}_0}$ is an element of ℓ^1. The first one can be estimated: The path of W is locally Hölder continuous of order $\gamma < 1/2$ at 1 with probability one. Assume that W is realized over some underlying probability space $(\Omega', \mathcal{A}', \mathbb{P}')$. Fix an $\omega \in \Omega'$, such that $t \longmapsto W_t(\omega)$ is locally Hölder continuous of order $\gamma < 1/2$. Then for $\epsilon > 0$ there is some constant C, such that for all $|1-s|, |1-t| < \epsilon$ $|W_s(\omega) - W_t(\omega)| < C|t-s|^\gamma$. Therefore

$$\begin{aligned}&\sum_{k=0}^{\infty} |W_{F(k)}(\omega) - W_{F(k-1)}(\omega)|\\ &\leq C \sum_{\substack{k=0\\ 1-F(k-1)<\epsilon}}^{\infty} |F(k) - F(k-1)|^\gamma + \text{ some large, but finite number}\\ &= C \sum_{\substack{k=0\\ 1-F(k-1)<\epsilon}}^{\infty} q_k^\gamma + \text{ some large, but finite number.}\end{aligned}$$

Hence if $\sum q_k^\gamma < \infty$ for some $\gamma < 1/2$, then (Z_k) is an element of ℓ^1 a.s..

The next theorem provides an elegant tool to prove weak convergence in Banach spaces. It is a combination of two corollaries quoted from [Va87] (p.29, p. 229).

Theorem 2.5 *Let B be a separable Banach space and Λ be a separating subspace of B^*. Let (B_k) be an ascending sequence of finite dimensional subspaces of B with $\overline{\bigcup_k B_k} = B$. Let P, P_k, $k \in \mathbb{N}$, be probability measures defined on the Borel σ-algebra of B. If*

$$\begin{aligned}&(i) \quad \lim_{n\to\infty} \int_X e^{i\langle\lambda,x\rangle} dP_n(x) = \int_X e^{i\langle\lambda,x\rangle} dP(x) \ \forall \lambda \in \Lambda,\\ &(ii) \quad \lim_{m\to\infty} \sup_{k\in\mathbb{N}} P_k(x \in X : \inf_{y\in B_m} \|x-y\| > \epsilon) = 0 \ \forall \epsilon > 0\end{aligned}$$

then $P_k \xrightarrow{w} P$.

From this we derive the theorem for the relative frequencies. For the notation of centered Gaussian random variables and covariance operator see [Li95] (p 76f).

Theorem 2.6 *Let $(W_t)_{t\in[0,1]}$ be the Wiener process. Let $F(x) := \sum_{l\leq x} q_l$ be the distribution function associated with q. Let $Z_k := W_{F(k)} - W_{F(k-1)} - q_k W_1$, $k \in \mathbb{N}_0$.*
i) Then the following equivalence holds:

$$(Z_k)_{k\in\mathbb{N}_0} \in \ell^1(\mathbb{N}_0) \quad a.s \qquad \Leftrightarrow \qquad \sum_{k\in\mathbb{Z}} \sqrt{q_k} < \infty.$$

ii) Suppose one of the conditions of the equivalence in i) to be fulfilled. Then Z is a centered Gaussian random variable with covariance operator

$$K : \ell^\infty \to \ell^1, \qquad K\lambda = M_q\lambda - \langle\lambda, q\rangle q.$$

$M_q : \ell^\infty \to \ell^1$ denotes the multiplication operator, i.e. $T_i M_q \lambda := \lambda_i q_i$.
iii) If one of the conditions in i) is fulfilled, then in $\ell^1(\mathbb{N}_0)$

$$\sqrt{n}(q_k^n - q) \xrightarrow{\mathcal{D}} Z = (Z_k)_{k\in\mathbb{N}_0}.$$

Proof: i) As mentioned above,

$$(Z_k)_{k\in\mathbb{N}_0} \text{ a.s} \qquad \Leftrightarrow \qquad \sum_{k\in\mathbb{N}_0} |W_{F(k)} - W_{F(k-1)}| < \infty \text{ a.s.},$$

since $W_1 q \in \ell^1(\mathbb{Z})$. Hence we can restrict ourselves to the analysis of the sum $\sum_{k\in\mathbb{N}_0} |W_{F(k)} - W_{F(k-1)}|$.

First suppose that $\sum_{k\in\mathbb{Z}} \sqrt{q_k} < \infty$ holds. We use the elementary inequality $E|Y| \leq \sqrt{EY^2}$ for a random variable Y. With Fubini's theorem we obtain

$$\begin{aligned} E\sum_{k\in\mathbb{N}_0} |W_{F(k)} - W_{F(k-1)}| &= \sum_{k\in\mathbb{N}_0} E|W_{F(k)} - W_{F(k-1)}| \\ &\leq \sum_{k\in\mathbb{N}_0} \sqrt{E(W_{F(k)} - W_{F(k-1)})^2} = \sum_{k\in\mathbb{N}_0} \sqrt{q_k} < \infty. \end{aligned}$$

This shows that $(W_{F(k)} - W_{F(k-1)})_{k\in\mathbb{Z}} \in \ell^1$ with probability 1. Therefore $(Z_k)_{k\in\mathbb{N}_0} \in \ell^1$.

Suppose now that we have $\sum_k |W_{F(k)} - W_{F(k-1)}| < \infty$ with probability 1. Without loss of generality we may assume that $q_k > 0$ for all $k \in \mathbb{Z}$. The random variables $(|W_{F(k)} - W_{F(k-1)}|)_{k\in\mathbb{N}_0}$ are independent. We can apply Kolmogorov's three-series-theorem (see [Fe66], p.317): In particular,

$$\sum_{k\in\mathbb{N}_0} E|W_{F(k)} - W_{F(k-1)}| 1_{|W_{F(k)} - W_{F(k-1)}|\leq c} < \infty \qquad \text{for all } c > 0.$$

Fix some c, say $c = 1$. We have

$$\begin{aligned}
\sum_{k\in\mathbb{N}_0} E|W_{F(k)} - W_{F(k-1)}|1_{|W_{F(k)}-W_{F(k-1)}|\le 1} &= \sum_{k\in\mathbb{N}_0} \sqrt{q_k}\frac{1}{\sqrt{2\pi}}\int_{-\frac{1}{\sqrt{q_k}}}^{\frac{1}{\sqrt{q_k}}} |x|e^{-\frac{x^2}{2}}\,dx \\
&= \sum_{k\in\mathbb{N}_0} \sqrt{q_k}\frac{2}{\sqrt{2\pi}}\Big(1 - e^{-\frac{1}{2q_k}}\Big) \\
&\ge \sum_{k\in\mathbb{N}_0} \sqrt{q_k}\frac{2}{\sqrt{2\pi}}(1 - e^{-\frac{1}{2}})\ .
\end{aligned}$$

This shows that $\sum_{k\in\mathbb{N}_0}\sqrt{q_k} < \infty$. Hence the first part of the assertion is established.

For the proof of ii) and iii) let

$$\Lambda := \Big\{\sum_{k=0}^{N}\alpha_k T_k :\ N\in\mathbb{N},\ \alpha_0,\dots,\alpha_N\in\mathbb{R}\Big\}.$$

ii) Let $\sum_{k\in\mathbb{N}_0}\sqrt{q_k} < \infty$. Assume that W is defined on some probability space $(\Omega', \mathcal{A}', \mathbb{P}')$. As we have proved above, the distribution of Z defines a probability measure on ℓ^1 (with the usual Borel σ-algebra). We want to prove that Z is a Gaussian random variable on ℓ^1. By definition this is equivalent to the statement that $\langle\lambda, Z\rangle$ is Gaussian for every $\lambda\in\ell^\infty$. As is well known the finite dimensional distributions of Z are those of a multidimensional Gaussian random variable. Moreover, $\langle\lambda, Z\rangle$ is Gaussian for all $\lambda\in\Lambda$ by definition. Now consider an arbitrary $\lambda = (\lambda_n)\in\ell^\infty$. Define $\lambda^N := \sum_{l=0}^{N}\lambda_k T_k$. Then we have $\langle\lambda^N, x\rangle \to \langle\lambda, x\rangle$ for every $x\in\ell^1$. In particular, $\langle\lambda^N, Z\rangle \to \langle\lambda, Z\rangle$ a.s.. Therefore $\langle\lambda^N, Z\rangle \xrightarrow{\mathcal{D}} \langle\lambda, Z\rangle$. Since the class of one-dimensional Gaussian distributions is closed with respect to weak convergence (of probability measures), $\langle\lambda, Z\rangle$ is Gaussian too. This shows that Z is Gaussian.

The barycenter a of Z is zero: let $\lambda = (\lambda_k)_{k\in\mathbb{N}_0}\in\ell^\infty$. Note that

$$\sum_{k=0}^{\infty}|\lambda_k||Z_k| \le \|\lambda\|_\infty\|Z\|_1 < \infty \quad \mathbb{P}'-\text{a.s.} \qquad \text{and} \qquad \left|\sum_{k=0}^{\infty}\lambda_k Z_k\right| \le \sum_{k=0}^{\infty}|\lambda_k||Z_k|.$$

Hence by dominated convergence

$$\begin{aligned}
E\langle\lambda, Z\rangle &= E\sum_{k=0}^{\infty}\lambda_k(W_{F(k)} - W_{F(k-1)} - q_k W_1) \\
&= \sum_{k=0}^{\infty}\lambda_k E(W_{F(k)} - W_{F(k-1)} - q_k W_1) = 0 = \langle\lambda, 0\rangle.
\end{aligned}$$

Since $\lambda\in\ell^\infty$ was arbitrarily chosen, the barycenter of Z is zero by definition. This shows that Z is centered.

Let us calculate the covariance operator of Z. This is an operator $K : \ell^\infty \to \ell^1$ fulfilling the equation

$$\langle \mu, K\lambda\rangle = E\langle \mu, Z\rangle\langle \lambda, Z\rangle.$$

for all $\lambda, \mu \in \ell^\infty$. Such a K exists (see [Li95], p. 77) and is continuous, if we topologize ℓ^1, ℓ^∞ with the weak topologies, i.e. the topology on ℓ^∞ is the one induced by the seminorms $\ell^\infty \ni \lambda \longmapsto q_f(\lambda) := |\langle \lambda, f\rangle|$, $f \in \ell^1$, and the topology on ℓ^1 is the one induced by the seminorms $\ell^1 \ni f \longmapsto q_\lambda(f) := |\langle \lambda, f\rangle|$, $\lambda \in \ell^\infty$ (see [Li95], p. 70). Suppose M_q to be the multiplication operator as defined in the assertion of the theorem. It holds

$$\langle T_k, M_q T_l - \langle T_l, q\rangle q\rangle = q_{kl}\delta_k - q_k q_l = EZ_kZ_l = E\langle T_k, Z\rangle\langle T_l, Z\rangle = \langle T_l, KT_k\rangle.$$

for all $k, l \in \mathbb{N}_0$. Therefore by linearity

$$\langle \lambda, M_q\mu - \langle \mu, q\rangle q\rangle = \langle \lambda, K\mu\rangle \quad \forall \lambda \in \Lambda.$$

The set Λ is dense in ℓ^∞ equipped with the weak topology. Since both mappings $\mu \longmapsto \langle \lambda, M_q\mu - \langle \mu, q\rangle q\rangle$ and $\mu \longmapsto \langle \lambda, K\mu\rangle$ are continuous with respect to this topology, we have

$$\langle \lambda, M_q\mu - \langle \mu, q\rangle q\rangle = \langle \lambda, K\mu\rangle \text{ for all } \lambda \in \Lambda, \mu \in \ell^\infty.$$

Since Λ is seperating, we can conclude $K\mu = M_q\mu - \langle \mu, q\rangle q$ for all $\mu \in \ell^\infty$, hence $K = M_q \cdot - \langle \cdot, q\rangle q$.

iii) Now we want to apply theorem 2.5. The multidimensional central limit theorem provides weak convergence of $\sqrt{n}(q_k^n - q_k)_{0\le k\le N}$ to a centered $N+1$-dimensional Gaussian random variable with covariance $\Sigma = (q_k\delta_{kl} - q_l q_k)_{0\le l,k\le N}$. Therefore:

$$Ee^{i\langle \sum_{k=0}^N \alpha_k T_k, \sqrt{n}(q^n - q)\rangle} = Ee^{i\sum_{k=1}^N \alpha_k \sqrt{n}(q_k^n - q_k)} \longrightarrow e^{-\frac{1}{2}(\alpha_0,\ldots,\alpha_N)^T\Sigma(\alpha_0,\ldots,\alpha_N)}.$$

It is easy to show that

$$\mathrm{Cov}(W_{F(k)} - W_{F(k-1)} - q_kW_1, W_{F(l)} - W_{F(l-1)} - q_lW_1) = q_k\delta_{kl} - q_kq_l.$$

Hence

$$e^{-\frac{1}{2}(\alpha_0,\ldots,\alpha_N)^T\Sigma(\alpha_0,\ldots,\alpha_N)} = Ee^{i\langle \sum_{k=0}^N \alpha_k T_k, W_{F(\cdot)} - W_{F(\cdot-1)} + qW_1\rangle}.$$

Therefore $\widehat{P^{\sqrt{n}(q^n - q)}}(\lambda) \to \widehat{P^{W_{F(\cdot)} - W_{F(\cdot-1)} - qB_1}}(\lambda)$ for all $\lambda \in \Lambda$.

Let us construct a sequence of of ascending subspaces

$$B_m := \{(\alpha_k)_{k\in\mathbb{Z}} : \ \alpha_k = 0, k > m\}.$$

Obviously, (B_k) satisfies the condition of theorem 2.5. We have the following inequality, using Beppo Levi's theorem

$$\begin{aligned} P\left(\sum_{|k|>m} |\sqrt{n}(q_k^n - q_k)| > \varepsilon\right) &\leq \frac{1}{\varepsilon} E \sum_{|k|>m} |\sqrt{n}(q_k^n - q_k)| \\ &= \frac{1}{\varepsilon} \sum_{|k|>m} E|\sqrt{n}(q_k^n - q_k)| \\ &\leq \frac{1}{\varepsilon} \sum_{|k|>m} \sqrt{E\left(\sqrt{n}(q_k^n - q_k)\right)^2} \\ &= \frac{1}{\varepsilon} \sum_{|k|>m} \sqrt{q_k - q_k^2} \leq \frac{1}{\varepsilon} \sum_{|k|>m} \sqrt{q_k} \end{aligned}$$

Since $\sum_k \sqrt{q_k} < \infty$, we have

$$\lim_{m\to\infty} \sup_n P\left(\sum_{|k|>m} |\sqrt{n}(q_k^n - q_k)| > \varepsilon\right) = 0 \quad \text{for all } \varepsilon > 0.$$

Hence both conditions of theorem 2.5 are established. This shows the second part of the assertion. □

Note that $M_q - \langle \cdot, q\rangle q$ is a bounded operator, hence continuous in the stronger norm-topology. Furthermore, it is a compact operator. Indeed, the multiplication operator can be approximated by finite rank operators with respect to the norm topology, e.g. define $M_q^n \lambda$ with $\langle T_k, M_q^n \lambda\rangle = q_k \lambda_k$ for $0 \leq k \leq n$ and $\langle T_k, M_q^n \lambda\rangle = 0$, else. Therefore M_q is a compact operator (see [La93], p. 416). Since the covariance operator is the sum of a compact operator and a finite rank operator, it is compact as well.

Now we will return to the question whether there is asymptotic normality for θ^n. This is proved by applying the delta method.

We want to use a generalisation of the Skohorod representation theorem, that can easily be derived from [Po84] (see p. 71, Representation Theorem). We state it here as a lemma:

Lemma 2.7 *Consider random variables X_n, X, $n \in \mathbb{N}$, taking values in a separable metric space D with $X_n \xrightarrow{\mathcal{D}} X$ and distributions $\mathcal{L}(X_n)$,$\mathcal{L}(X)$ $n \in \mathbb{N}$. Then there is a probability space $(\Omega', \mathcal{A}', \mathbb{P}')$, random variables $X_n', X' : \Omega' \to D$, $n \in \mathbb{N}$ such that $X' \stackrel{\mathcal{D}}{=} X$, $X_n' \stackrel{\mathcal{D}}{=} X_n$ and $X_n' \to X'$ a.s..*

Remark: As usual, $X_n \xrightarrow{\mathcal{D}} X$ iff $E(f(X_n)) \to Ef(X)$ for all bounded continuous functions.

Theorem 2.8 *If* $\sum_{k=0}^{\infty} \sqrt{q_k} < \infty$*, then*

$$\sqrt{n}(\theta^n - (\lambda\delta_0 + p)) \xrightarrow{\mathcal{D}} G := A^{-1}Z.$$

G is a centered Gaussian process with covariance operator $A^{-1} \circ (M_q - \langle\cdot, q\rangle q) \circ A^{-T}$ with $A^{-1} = L'_q$ as in lemma 2.3.

Proof: The proof is standard. From theorem 2.6 we have $\sqrt{n}(q^n - q) \xrightarrow{\mathcal{D}} Z$. Since ℓ^1 is a separable metric space with respect to the norm topology, we can apply lemma 2.7 above. Hence there are a probability space $(\Omega', \mathcal{A}', \mathbb{P}')$ and ℓ^1-valued random variables q'_n and Z' defined on it with

$$\sqrt{n}(q'_n - q) \to Z' \qquad \mathbb{P}' - \text{a.s.}, \qquad q^n \stackrel{\mathcal{D}}{=} q'_n,\ n \in \mathbb{N}, \qquad Z \stackrel{\mathcal{D}}{=} Z'.$$

Since we have Fréchet-differentiability of L in q, we have for every $\omega' \in \Omega'$ in the complement of a set of $\mathbb{P}'$-probability zero

$$\begin{aligned}&\sqrt{n}(L(q'_n(\omega')) - L(q)) \\ &\quad = L'_q\sqrt{n}(q'_n(\omega') - q) + o(\sqrt{n}(q'_n(\omega') - q)) = L'_q Z'(\omega') + o(\sqrt{n}(q'_n(\omega') - q)).\end{aligned}$$

This shows that

$$\sqrt{n}(L(q'_n) - L(q)) \to L'_q Z' \quad \mathbb{P}' - \text{a.s.}.$$

Hence

$$\sqrt{n}(L(q'_n) - L(q)) \xrightarrow{\mathcal{D}} L'_q Z'.$$

Since $q^n \stackrel{\mathcal{D}}{=} q'_n$, $n \in \mathbb{N}$, and $Z \stackrel{\mathcal{D}}{=} Z'$, we have

$$\sqrt{n}(L(q^n) - L(q)) \stackrel{\mathcal{D}}{=} \sqrt{n}(L(q'_n) - L(q)) \xrightarrow{\mathcal{D}} L'_q Z' \stackrel{\mathcal{D}}{=} L'_q Z.$$

If we define $G = L'_q Z$, then G turns out to be the distributional limit of $\sqrt{n}(L(q'^n) - L(q))$. The first assertion is proved.

Obviously, G is the image of an Gaussian random variable under a continuous linear mapping. Hence it is a Gaussian random variable. The representations of the covariance operator and the barycenter are calculated directly from the definition of G. □

Of course it is of some interest how the decay condition of q is connected to an appropriate condition on the claim distribution p. This will be answered by the next lemma. The main statement is more general than we need in the moment, but it turns out to be useful later on, when we are looking for a proof of consistency for the maximum likelihood estimator.

Lemma 2.9 *The following holds for all $\lambda > 0$. Let $\gamma \in (0,1)$. Then*

$$\sum_{k=1}^{\infty} p_k^{\gamma} < \infty \qquad \Leftrightarrow \qquad \sum_{k=0}^{\infty} q_k^{\gamma} < \infty\ .$$

Remark: at one stroke, we see that for all p concentrated on finite subsets of $\mathbb{N}_0$ our decay condition is fulfilled.

Proof: „$\Leftarrow$“: This is immediately clear. Since for $k \in \mathbb{N}_0$ fixed the event that the total claim is k includes the event that the number of single claims is one and the only claim is k, in probabilities: $q_k \geq e^{-\lambda}\lambda p_k$.

„$\Rightarrow$“: Now consider the space $\ell^\gamma_{\mathbb{R}}(\mathbb{N}_0)$, including the sequences $(x_k)_{k\in\mathbb{N}_0}$ with $\sum_{k\in\mathbb{N}_0} |x_k|^\gamma < \infty$. $\|(x)_{k\in\mathbb{N}_0}\|_\gamma = \sum_{k=0}^\infty |x|^\gamma$ defines a socalled quasinorm on $\ell^\gamma_{\mathbb{R}}(\mathbb{N}_0)$. In particular, $\|\cdot\|_\gamma$ is continuous and subadditive and $(\ell^\gamma_{\mathbb{R}}(\mathbb{N}_0), \|\cdot\|_\gamma$ is a complete quasinormed space (see [Sw92], Example 17, p.22). Obviously, $\ell^\gamma_{\mathbb{R}}(\mathbb{N}_0) \subset \ell^1_{\mathbb{R}}(\mathbb{N}_0)$. Therefore the convolution $x * y$ of $x, y \in \ell^\gamma_{\mathbb{R}}(\mathbb{N}_0)$ is well defined and the sequence $x * y$ is at least an element of ℓ^1. We show even more, namely $z = x * y \in \ell^\gamma_{\mathbb{R}}(\mathbb{N}_0)$:

$$\sum_{k=0}^\infty |z_k|^\gamma = \sum_{k=0}^\infty |\sum_{m=0}^k x_m y_{k-m}|^\gamma \leq \sum_{k=0}^\infty \sum_{m=0}^k |x_m y_{k-m}|^\gamma = \sum_{k=0}^\infty |x_k|^\gamma \sum_{m=0}^\infty |y_m|^\gamma$$

Hence $\|x * y\|_\gamma \leq \|x\|_\gamma \|y\|_\gamma$. We have proved the submultiplicity. If $p \in \ell^\gamma_{\mathbb{R}}(\mathbb{N}_0)$ and $q = \exp(\lambda(p - \delta_0))$, we therefore have

$$\|q\|_\gamma \leq \sum_{k=0}^\infty \left(\frac{\lambda^k}{k!}\right)^\gamma e^{-\lambda\gamma} \|p\|^k_\gamma.$$

It can easily be seen that the right hand is finite. This proves „$\Rightarrow$“.$\Box$

2.5 Two Naive Projection Estimators

We have investigated the estimator θ^n and have proved strong consistency and asymptotic normality. These are beautiful properties for an estimator. We should pay some attention to the question whether the range of our estimator is a subset of our natural parameter space $\mathbb{R}^+ \oplus M_1(\mathbb{N})$, $M_1(\mathbb{N})$ denoting the set of counting densities on $\mathbb{N}$. By definition the estimator $T_0\theta^n$ for λ takes only positive values, hence it is an element of the natural parameter space for λ. Furthermore, summing up the components, we have $\sum_{l=1}^\infty \theta^n_k = 1$. Indeed, for $q^n \notin U'$ this is trivial by definition of $\theta^n = \delta_0 + \delta_1$. If $q^n \in U'$, then

$$\begin{aligned} 1 &= \sum_{k=0}^\infty q^n_k = \widehat{q^n}(0) = \exp(\widehat{\log q^n}(0)) \\ &= \exp(\langle T_0, \theta^n \rangle \widehat{(T_0^\perp \theta^n - \delta_0)}(0)) = \exp(\langle T_0, \theta^n \rangle (\sum_{l=1}^\infty \theta^n_k - 1)). \end{aligned}$$

Hence $\sum_{l=1}^\infty \theta^n_k = 1$.

However, it turns out that $T_0^{\perp}\theta^n$ is never nonnegative, hence not a probability measure. The reason for this is that q^n has finite support, but nontrivial compound distributions always have unbounded support: Consider some probability measure p not concentrated on $\{0\}$, say $p_k > 0$ for some $k > 0$. Then the event that the total claim is lk, $l \geq 0$, includes the event that all single claims take the value k and the number of claims is l, in probabilities $q_{lk} \geq p_k^l e^{-\lambda}\frac{\lambda^l}{l!} > 0$. The support of q includes all multiplies of k.

How can we turn θ^n into a probability density? A simple method is to replace all negative entries of $T_0^{\perp}\theta^n$ by zero. This gives us a new sequence, say (y_k). Since the sequence does not sum up to one anymore, we normalize it by dividing every entry by the total mass $\sum_{l=1}^{\infty} y_k$. Note that

$$1 = \sum_{l=1}^{\infty}\langle T_k, T_0^{\perp}\theta^n\rangle \leq \sum_{l=1}^{\infty} y_k \leq \|T_0^{\perp}\theta^n\|_1 < \infty.$$

This justifies our procedure.

To be more precise, define the mapping

$$\pi : \ell^1 \to \ell^1, \qquad \pi((x_k)) := T_0^{\perp}((x_k^+)).$$

This mapping is well defined. Let us abbreviate the summations in ℓ_1. Define summation operators $\sum_k^m x := \sum_{l=k}^m x_l$ for $x = (x_k)_k \in \ell^1$, $k \leq m \leq \infty$. These are bounded linear mappings, hence continuous.

Our new estimator can be described in a compact form as follows

$$\eta^n = L^2(\theta^n) := \langle T_0, \theta^n\rangle\delta_0 + \frac{1}{\sum_1^{\infty}\pi(\theta^n)}\pi(\theta^n).$$

For consistency we look at the underlying mappings: From the simple inequality $|x^+ - y^+| \leq |x - y|$, $x, y \in \mathbb{R}$, we conclude that π is a continuous mapping, even more: $\|\pi(x) - \pi(y)\|_1 \leq \|x - y\|_1$ holds for all $x, y \in \ell^1$. Since we have already proved strong consistency of θ^n in ℓ^1, we obtain

$$\pi(\theta^n) \to \pi(\lambda\delta_0 + p) = p \qquad \mathbb{P} - a.s..$$

The summation operator involved in the definition of η^n is continuous too. Hence $L^2(x)$ is pointwise continuous for all x with $\sum_1^{\infty}\pi(x) \neq 0$. As already seen we have $\sum_1^{\infty}\pi(\theta^n) \geq 1$ with probability one. This shows strong consistency, i.e.

$$\eta^n \to \lambda\delta_0 + p \qquad \mathbb{P} - \text{a.s.}.$$

What is the impact of our normalization procedure on the distributional limit? Before investigating this, let us define the following mapping. Of course, we could try to apply the delta-method on the asymptotic normality of $\sqrt{n}(\theta^n - (\lambda\delta_0 + p))$ and the mapping L^2. However, this fails here since π is not differentiable anymore.

This and some refinements on the estimation procedure considered later on forces us to go a more direct way.

For $x^0 \in \ell_1$ define

$$\pi_{x^0} : \begin{cases} \ell^1(\mathbb{N}_0) & \to & \ell^1(\mathbb{N}_0) \\ (x_k) & \longmapsto & T_0^{\perp}(x_k 1_{x_k^0>0} + x_k^+ 1_{x_k^0=0}). \end{cases}$$

Again, π_{x^0} is well defined for all $x^0 \in \ell^1$. Furthermore, it is continuous, indeed

$$\|\pi_{x^0}(x) - \pi_{x^0}(y)\|_1 = \sum_{x_k^0>0} |x_k - y_k| + \sum_{x_k^0=0} |x_k^+ - y_k^+| \le \|x-y\|_1 \qquad \forall x, y \in \ell^1.$$

Lemma 2.10 *i) Consider $\xi^n, g \in \ell^1$ with $\|\sqrt{n}(\xi^n - (\lambda\delta_0 + p)) - g\|_1 \to 0$. Then*

$$\left\|\sqrt{n}\left(\pi(\xi^n) - p\right) - \pi_p(g)\right\|_1 \to 0, \qquad \left|\sqrt{n}\left(\sum_1^\infty \pi(\xi^n) - 1\right) - \sum_1^\infty \pi_p(g)\right| \to 0.$$

ii) Consider $p^n, g \in \ell^1$, $\lambda_n, g_\lambda, \gamma_n, g_\gamma$ with

$$\|\sqrt{n}(y^n - p) - g\|_1 \to 0, \quad |\sqrt{n}(\lambda_n - \lambda) - g_\lambda| \to 0, \quad |\sqrt{n}(\gamma_n - 1) - g_\gamma| \to 0.$$

Then

$$\left\|\sqrt{n}\left(\left(\lambda_n\delta_0 + \frac{1}{\gamma_n}p^n\right) - (\lambda\delta_0 + p)\right) - (g_\lambda\delta_0 - g_\gamma p + g)\right\|_1 \to 0.$$

Proof: i) Again note that $|x^+ - y^+| \le |x-y|$. Consider for $S \in \mathbb{N}$

$$\begin{aligned}
&\|\pi_p(\sqrt{n}(\xi^n - (\lambda\delta_0 + p))) - \sqrt{n}(\pi(\xi^n) - p)\|_1 \\
\le\ & \sum_{\substack{l=1\\p_l=0}}^S |\underbrace{(\sqrt{n}\xi_l^n)^+ - \sqrt{n}(\xi_l^n)^+}_{=0}| + \sum_{\substack{l=1\\p_l>0}}^S |\sqrt{n}(\xi_l^n - p_l) - \sqrt{n}((\xi_l^n)^+ - p_l)| \\
&+ 2 \underbrace{\sqrt{n}\sum_{l=S+1}^\infty |\xi_l^n - p_l|}_{\le \sum_{l=S+1}^\infty |\sqrt{n}(\xi_l^n - p_l) - g_l| + \sum_{l=S+1}^\infty |g_l|} \\
\le\ & \sum_{\substack{l=1\\p_l>0}}^S |\sqrt{n}(\xi_l^n - p_l) - \sqrt{n}((\xi_l^n)^+ - p_l)| + 2\|\sqrt{n}(\xi^n - (\lambda\delta_0 + p) - g\|_1 + 2\sum_{l=S+1}^\infty |g_l|
\end{aligned}$$

Let $\epsilon > 0$. Then there is an S with $2\sum_{l=S+1}^\infty |g_l| < \epsilon$. Since $\|\cdot\|_1$-convergence implies pointwise convergence, we have $\xi_k^n > 0$ for all $k = 1, \ldots, S$ with $p_k > 0$ for n big enough. Then

$$\sum_{\substack{l=1\\p_l>0}}^S |\sqrt{n}(\xi_l^n - p_l) - \sqrt{n}((\xi_l^n)^+ - p_l)| = 0.$$

Therefore

$$\limsup_{n\to\infty} \|\pi_p(\sqrt{n}(\xi^n - (\lambda\delta_0 + p))) - \sqrt{n}(\pi(\xi^n) - p)\|_1 \leq \epsilon.$$

Since ϵ was arbitrary, we have proved $\|\pi_p(\sqrt{n}(\xi^n-(\lambda\delta_0+p)))-\sqrt{n}(\pi(\xi^n)-p)\|_1 \to 0$. The assertion follows from

$$\begin{aligned}
&\|\sqrt{n}(\pi(\xi^n) - p) - \pi_p(g)\|_1 \\
&\leq \|\sqrt{n}(\pi(\xi^n) - p) - \pi_p(\sqrt{n}(\xi^n - (\lambda\delta_0 + p)))\|_1 + \|\pi_p(\sqrt{n}(\xi^n - (\lambda\delta_0 + p))) - \pi_p(g)\|_1 \\
&\leq \|\sqrt{n}(\pi(\xi^n) - p) - \pi_p(\sqrt{n}(\xi^n - (\lambda\delta_0 + p)))\|_1 + \|\sqrt{n}(\xi^n - (\lambda\delta_0 + p)) - g\|_1.
\end{aligned}$$

The second assertion is proved by

$$|\sqrt{n}(\sum_1^\infty \pi(\xi^n) - \underbrace{1}_{=\sum_1^\infty p}) - \sum_1^\infty \pi_p(g)| \leq \|\sqrt{n}(\pi(\xi^n) - p) - \pi_p(g)\|_1 \to 0.$$

ii) This is a consequence of the following inequality

$$\begin{aligned}
&\|\sqrt{n}\left\|\left(\left(\lambda_n\delta_0 + \frac{1}{\gamma_n}\,p_n\right) - (\lambda\delta_0 + p)\right) - (g_\lambda\delta_0 - g_\gamma p + g)\right\|_1 \\
&\leq |\sqrt{n}(\lambda_n - \lambda) - g_\lambda| + \left\|\sqrt{n}\left(\frac{1}{\gamma_n}p_n - p\right) + g_\gamma p - g\right\|_1 \\
&\leq o(1) + \frac{1}{\gamma_n}\|\sqrt{n}(p_n - p) - g\|_1 + \left\|\sqrt{n}\left(\frac{1}{\gamma_n} - 1 + g_\gamma\right)p - \left(1 - \frac{1}{\gamma_n}\right)g\right\|_1 \\
&\leq o(1) + \|p\|_1\left|\sqrt{n}\left(\frac{1}{\gamma_n} - 1\right) + g_\gamma\right| + \|g\|_1\left|1 - \frac{1}{\gamma_n}\right| \\
&\leq o(1) + \underbrace{\|p\|_1}_{=1}\frac{1}{|\gamma_n|}\left|\sqrt{n}(1 - \gamma_n) + \gamma_n g_\gamma\right| \\
&\leq o(1) + \frac{1}{|\gamma_n|}\left(|\sqrt{n}(1 - \gamma_n) - g_\gamma| + |\gamma_n - 1||g_\gamma|\right) \\
&= o(1). \square
\end{aligned}$$

The following theorem then is a consequence of the representation theorem (lemma 2.7), the lemma above and the limit $\sqrt{n}(\theta^n - (\lambda\delta_0 + p)) \xrightarrow{\mathcal{D}} G$.

Theorem 2.11 *If $\sum_k \sqrt{q_k} < \infty$ holds then in ℓ^1*

$$\sqrt{n}(\eta^n - (\lambda\delta_0 + p)) \xrightarrow{\mathcal{D}} \langle T_0, G\rangle\delta_0 - \big(\sum_1^\infty \pi_p(G)\big)p + \pi_p(G).$$

We now derive a second estimation procedure. Recall that θ^n is calculated via an inverse Panjer recursion formula (at least for n big enough). First we state the following lemma, that is of its own interest in investigations later on:

Lemma 2.12 *Consider $q_k \geq 0$, $k > 1$, and $1 > q_0 > 0$. Suppose that $\lambda, x_1, \ldots$ are calculated using the inverse Panjer recursion formula with input q_k, $k \in \mathbb{N}_0$. Let $S \in \mathbb{N}$. Then*

$$(x_i \leq 0 \, \forall i = 1 \ldots, S) \qquad \Leftrightarrow \qquad (q_i = 0 \, \forall i = 0, \ldots, S)\,.$$

If one of the two statements above is true then it holds that

$$x_i = 0 \qquad \forall i = 1, \ldots, S.$$

Proof: This is proved by induction: Note that $\lambda = -\log q_0 \in (0, \infty)$. If $x_1 \leq 0$ then $q_1 = \lambda x_1 \leq 0$, hence $q_1 = 0$, and vice versa. Now consider the induction step $S - 1 \to S$. If $x_1, \ldots, x_S \leq 0$, then our induction assumption yields $q_1 = \cdots = q_{S-1} = 0$. Therefore

$$q_S = \frac{\lambda}{S} \sum_{l=1}^{S} l x_l q_{k-l} = q_0 x_S \leq 0.$$

Hence, $q_S = 0$. The same is true for the other direction.

The conclusion that $x_1 = \ldots x_S = 0$ is then trivial. □

Now back to the new estimator: It is based on a Panjer inversion just considering a finite initial segment of q^n. Its end point is data driven. Let (S_n) be a sequence of $\mathbb{N}$-valued random variables. If $q_0^n = 0$ or $q_0^n = 1$ or $q_i^n = 0$ for all $i = 1, \ldots, S_n$ we put $\eta^{n,S} := \delta_0 + \delta_1$. If not we can calculate λ^n, $p_1^n \ldots, p_{S_n}^n$ using the Panjer inversion. Then we define $\eta^{n,S} := L^2(\lambda^n \delta_0 + \sum_{l=1}^{S_n} p_l^n \delta_l)$, i.e. put negative entries to zero and norm to one. Because of the fact that at least for one $i \in \{1, \ldots, S_n\}$ $q_i > 0$, we can apply the lemma above to obtain

$$\sum_{1}^{\infty} \pi \left(\lambda^n \delta_0 + \sum_{l=1}^{S_n} p_l^n \delta_l \right) > 0.$$

Hence, $\eta^{n,S}$ is a well defined element of ℓ^1. It is easy to show that it is also a measurable mapping from $(\Omega, \mathcal{A}, \mathbb{P})$ to ℓ^1.

The idea is that if $S_n \to \infty$ fast enough then we can expect both strong consistency of T_n and the same distributional limit for $\sqrt{n}(T_n - (\lambda\delta_0 + p))$ as in the case of η^n.

Theorem 2.13 *i) If $\liminf_{n\to\infty} S_n \geq \sup\{k : \; p_k > 0\}$ a.s. then $\eta^{n,S} \to \lambda\delta_0 + p$ a.s.*

ii) Suppose that (u_n) is a sequence of natural numbers tending to infinity and that the following conditions hold:

$$\lim_{n\to\infty} \sqrt{n} \sum_{l=u_n}^{\infty} p_l = 0, \quad \textit{and} \quad P(S_n \leq u_n \textit{ infinitely often }) = 0.$$

If $\sum_k \sqrt{q}_k < \infty$ *then in* ℓ^1

$$\sqrt{n}(\eta^{n,S} - (\lambda\delta_0 + p)) \xrightarrow{\mathcal{D}} \langle T_0, G\rangle\delta_0 - \sum_1^\infty \pi_p(G)p + \pi_p(G).$$

Proof: Define $\xi^{n,S}$ to be $\delta_0 + \delta_1$, $q_0^n = 0$ or $q_0^n = 1$ or $q_i^n = 0$ for all $i = 1, \dots, S_n$ and $\xi^{n,S}$ to be the sequence generated by the Panjer inversion of q^n up to S_n. For the entries with index greater than S_n put $\xi_k^{n,S} = 0$.
i) If we look at the definition of L^2 we see that it is again the composition of two mappings

$$\theta \mapsto \langle T_0, \theta\rangle\delta_0 + \pi(\theta) \mapsto \langle T_0, \theta\rangle\delta_0 + \frac{1}{\sum_1^\infty \pi(\theta)}\pi(\eta^n).$$

The second one is continuous in $\delta_0\lambda + p$. We already know that $\xi_0^{n,S} \to \lambda$ a.s.. Therefore it is sufficient to show that $\|\pi(\xi^{n,S}) - p\|_1 \to 0$ a.s.. Again, with probability one and for n large enough $\xi_k^{n,S} = \theta_k^n$, $k = 1, \dots, S_n$. Therefore

$$\|\pi(\xi^{S,n}) - p\|_1 = \sum_{l=1}^{S_n} |(\eta_k^n)^+ - p_k| + \sum_{l=S_n+1}^\infty |p_k| \le \|T_0^\perp(\eta^n) - p\|_1 + \sum_{l=S_n+1}^\infty |p_k|.$$

Since we have already proved that $\|T_0^\perp(\eta^n) - p\|_1 \to 0$ a.s. the assertion follows from the condition on $\liminf S_n$.
ii) First we want to show that $\sqrt{n}\|\theta^n - \xi^{n,S}\|_1 \to 0$ in probability. Define the sets $A_n := \{q_0^n \in \{0,1\}\}$, $B_n := \{q_1^n = \dots = q_{S_n}^n = 0\}$, $C_n = \{q^n \in U'\}$. We again have $p_k > 0$ for some k. Since S_n tends to infinity we have $\{q_k^n = 0\} \supset B_n$ for n large enough. This shows for arbitrary sequences (h_n^i), $i = 1, 2, 3$ that

$$\sqrt{n}1_{A_n}h_n^1 \to 0, \qquad \sqrt{n}1_{B_n}h_n^2 \to 0, \qquad \sqrt{n}1_{C_n^C}h_n^3 \to 0.$$

Hence

$$\begin{aligned}
\|\sqrt{n}(\xi^{S,n} - \theta^n)\|_1 &= 1_{A_n^C \cap B_n^C \cap C_n}\sqrt{n}\|\xi^{S,n} - \theta^n\| + o_{\text{a.s.}}(1) \\
&\le 1_{A_n^C \cap B_n^C \cap C_n}\sqrt{n}\sum_{k=S_n+1}^\infty |\theta_k^n| + o_{\text{a.s.}}(1) \\
&\le 1_{A_n^C \cap B_n^C \cap C_n}\sqrt{n}\sum_{k=u_n}^\infty |\theta_k^n| + o_{\text{a.s.}}(1) \\
&\le 1_{A_n^C \cap B_n^C \cap C_n}\left(\sqrt{n}\sum_{k=u_n}^\infty |\theta_k^n - p_k| + \sqrt{n}\sum_{l=u_n}^\infty p_k\right) + o_{\text{a.s.}}(1) \\
&= 1_{A_n^C \cap B_n^C \cap C_n}\sqrt{n}\sum_{k=u_n}^\infty |\theta_k^n - p_k| + o_{\text{a.s.}}(1).
\end{aligned}$$

Let $(\Omega', \mathcal{A}', \mathbb{P}')$ be the probability space in the Skorhorod representation theorem. Let $\theta'_n \stackrel{\mathcal{D}}{=} \theta^n$, $n \in \mathbb{N}$, $G' \stackrel{\mathcal{D}}{=} G$ with $\theta^{n\prime}$, G' being the random variables with $\sqrt{n}(\theta'_n - (\lambda\delta_0 + p)) \to G'$ $\mathbb{P}'$-a.s.. Then for $\epsilon > 0$

$$\begin{aligned}\mathbb{P}(\sqrt{n}\sum_{k=u_n}^{\infty}|\theta_k^n - p_k| > \epsilon) &= \mathbb{P}'(\sqrt{n}\sum_{k=u_n}^{\infty}|\theta'_{n,k} - p_k| > \epsilon)\\ &\le \mathbb{P}'(\|\sqrt{n}(\theta'_{n,k} - p) - G')\|_1 > \epsilon/2) + \mathbb{P}'(\sum_{k=u_n}^{\infty}|G'_k| > \epsilon/2).\end{aligned}$$

Since $\|\sqrt{n}(\theta'_{n,k} - p) - G')\|_1 \to 0$ $\mathbb{P}'$-a.s., we have $\|\sqrt{n}(\theta'_{n,k} - p) - G')\|_1 \to 0$ in probability with respect to $\mathbb{P}'$. Since $\sum_{k=0}^{\infty}|G'_k| < \infty$ $\mathbb{P}'$-a.s., we have $\sum_{k=u_n}^{\infty}|G'_k| \to 0$ $\mathbb{P}'$-a.s., therefore in $\mathbb{P}'$-probability. This implies that

$$\mathbb{P}'(\|\sqrt{n}(\theta'_{n,k} - p) - G')\|_1 > \epsilon/2) + \mathbb{P}'\left(\sum_{k=u_n}^{\infty}|G'_k|)\| > \epsilon/2\right) \to 0.$$

Hence $\sqrt{n}\sum_{k=u_n}^{\infty}|\theta_k^n - p_k| \to 0$ in probability. This shows

$$\sqrt{n}\|\theta^n - \xi^{S,n}\| \to 0$$

in probability with respect to $\mathbb{P}$. Hence $\xi^{S,n}$ and θ^n are asymptotically equivalent, i.e. $\sqrt{n}(\xi^{S,n} - (\lambda\delta_0 + p)) \to G$. The assertion is established by applying lemma 2.10 on the Skohorod representation of the distributional limit for $\xi^{S,n}$.□

We now want to show that $S_n := Y_{(n)} = \max\{Y_1, \ldots, Y_n\}$ is a possible choice for S_n. We recall the following theorem about almost sure convergence of the n-th order statistic ([Ga87],p 252)

Theorem 2.14 *Let (X_i) be some iid-sequence of random variables with continuous distribution function F. Assume that $(u_n)_{n\in\mathbb{N}}$ is a sequence such that $n(1 - F(u_n))$ is nondecreasing and that $u_n \le \sup\{x : F(x) < 1\}$ $\forall n$. Then*

$$P(\max_{l=1,\ldots,n} X_l \le u_n \ i.o.) = 0 \quad or \qquad 1$$

according as $\sum_{j=1}^{\infty}(1 - F(u_j))e^{-j(1-F(u_j))} < \infty$ or $= \infty$.

Let (U_i) be some iid-sequence of random variables, uniformly distributed on $(0, 1)$ and independent of (Y_i). Define $\tilde{Y}_i := Y_i - U_i$. Then $(\tilde{Y}_i)$ is an iid-sequence of random variables with a continuous distribution function $\tilde{F}$ that is strictly monotone increasing on $(-1, \infty)$. Suppose that $\tilde{S}_n := \max_{l=1\ldots n} Y_l$. Define v_n through the equation

$$\sqrt{n}(1 - \tilde{F}(v_n)) = n^{-1/4}.$$

Then

i) $n(1-\tilde{F}(v_n))$ is nondecreasing.

ii) Since
$$\int_a^\infty x^{-3/4}e^{-x^{1/4}}\,dx = -4e^{-x^{1/4}}|_a^\infty < \infty,$$
we have $\sum_{j=1}^\infty (1-F(v_j))e^{-j(1-F(v_j))}\,dx \approx \int^\infty x^{-3/4}e^{-x^{1/4}}\,dx < \infty$.

Hence $P(\tilde{S}_n < v_n \text{ i.o.}) = 0$. Let u_n be the smallest integer greater than v_n. Then $P(\tilde{S}_n \le v_n \text{ i.o.}) = P(S_n \le u_n \text{ i.o.})$. We also have

$$1-\tilde{F}(v_n) = P(\tilde{Y}_1 > v_n) = P(Y_1 > u_n) = \sum_{l=u_n+1}^\infty q_l.$$

Since
$$\sqrt{n}\sum_{l=u_n+1}^\infty q_l \ge \lambda e^{-\lambda}\sqrt{n}\sum_{l=u_n+1}^\infty p_l$$
and $\sqrt{n}p_{u_n} \to 0$ for every sequence $(u_n)_{n\in\mathbb{N}}$ with $u_n \to \infty$ as $n \to \infty$, we have

$$\sqrt{n}\sum_{l=u_n}^\infty p_l \to 0.$$

This shows that the n-th order statistic, i.e. the largest observation, is a possible choice for S_n.

Chapter 3

Cones

3.1 Truncated Decompounding and Order Restricted Inference

This chapter is devoted to the investigation of the maximum likelihood estimation for discrete data. We truncate the data at a fixed threshold $S+1$. This leads to a parametric estimation problem for multinomial distributions. We want to use the maximum likelihood method. To do so, we have to maximize the log likelihood function over the set of truncated compound Poisson distributions. If we consider the closure of this set then we have to maximize an upper semicontinuous or concave function over a compact set, hence there exists at least one maximum likelihood estimator. Note that the set of truncated compound Poisson distributions is not convex for dimensions higher then 6[1]. We have to maximize over a nonconvex set. However, the situation is not completely hopeless, as we will see. The reason for is the smoothness of the compounding mapping. Furthermore, the underlying parameter space for λ and p is a simplex. The maximization can be performed over the underlying parameter space or over the set of truncated compound Poisson distributions. A simplex looks locally like a cone. Hence locally, the maximization is performed over a convex set. This indicates that the set of maximum likelihood estimators will become a singleton. Locally, the log likelihood function can be approximated by a quadratic form. This quadratic form has to be maximized, or up to a sign to be minimized. Minimizing a quadratic form over a cone is a cone projection with respect to the inner product defined by this

[1]Indeed take two claim distributions $p_1 = \delta_2$ and $p_3 = \delta_3$ and fix some $\lambda > 0$. The asssociated compound Poisson distributions, say q_1 and q_2, are concentrated on the sets $2\mathbb{N}_0$ and $3\mathbb{N}_0$ respectively. There is a gap at 5 in the union of their supports. Now take some proper convex combination of q_1 and q_2. The support is then the union of the supports $2\mathbb{N}_0$ and $3\mathbb{N}_0$. From the Panjer recursion formula we see that if the convex combination is a compound Poisson distribution then the numbers $2, 3$ must be members of its support. On the other side, the support of a compound Poisson distribution must be a semigroup. Hence 5 must be an element of the support, a contradiction.

quadratic form. This is the underlying idea of this chapter. Maximum likelihood is known to have good statistical properties in regular parametric models. Hence this is a motivation to use an approximation of it, its approximating quadratic form, as the basis for projection estimators. This will be explained in section 3.4.

This chapter uses many ideas from the theory of order restricted inference, that deals with special cones. To give an instructive example (see [Ro88], p.6, for a similar example and details, for binomials see p.32), estimate the heights of children grouped in different age classes

$$a_1 < \cdots < a_d$$

from a given sample

$$Y_{kl}, \qquad l = 1, \ldots, n_k, \qquad k = 1, \ldots, d.$$

The numbers n_i are the sample sizes for the different groups. If we assume that the samples Y_{kl} are mutually independent and normally distributed with expectation μ_k and a common variance, a common sense assumption is

$$\mu_1 \leq \cdots \leq \mu_d,$$

i.e. $i \longmapsto \mu_i$ is an isotonic function. Maximization of the likelihood function under the assumption that the expected values μ_i should be an isotonic function turns out to be equivalent to the minimization of the equation

$$\sum_{k=1}^{d} (\bar{Y}_k - \mu_k)^2 n_k, \qquad \bar{Y}_k = \frac{1}{n_k} \sum_{l=1}^{n_k} Y_{kl}$$

over the set

$$K := \{(\mu_1, \ldots, \mu_d)^T \in \mathbb{R}^d : \ \mu_1 \leq \cdots \leq \mu_d\}.$$

K is called the cone of isotonic functions. The more or less explicit solution is to project the vector $(\bar{Y}_1, \ldots, \bar{Y}_d)^T$ onto the cone K, i.e. find the closest point g^* on K with respect to the distance that is given by the weighted inner product $\langle x, y \rangle = \sum_{k=1}^{d} x_i y_i n_i$. The projected vector can be found using the greatest convex minorant (see [Ro88], p.7): Plot the diagram with points $P_i = (\sum_{k=1}^{i} n_k, \sum_{k=1}^{i} n_k \bar{Y}_k)_i$, $i = 0, \ldots, d$, with $P_0 = 0$. Connecting the points P_i and P_{i+1}, $i = 1, \ldots, d-1$, with line segments leads to a function G on $[0, \sum_{k=1}^{d} n_i]$. If G^* is the greatest convex minorant (GCM) of G and $\bar{\mu}_i$ is the left hand derivative of G^* at $\sum_{l=1}^{i} n_i$, $i = 1, \ldots, d$, then it holds that $g^* = (\bar{\mu}_1, \ldots, \bar{\mu}_d)$. The pool-adjacent-violator-algorithm gives a method to calculate the projection g^* in a finite number of steps (see [Ro88], p.9).

Our problem is similar to isotonic regression, since both problems lead to cone projections. The main difference between the truncated decompounding and the isotonic regression is that the maximum likelihood estimator in the decompounding problem cannot be identified explicitly with a cone projection. Hence some „localizing“ step is necessary. We will do this in a slightly more general framework in section 3.3. This is then easily applied to the decompounding case and leads to a simple efficient estimation procedure. Section 3.5 shows how to derive tests of Poissonity within the class of compound Poisson distributions. The test requires a procedure to derive cone projections. Lemma 3.2 below indicates that the simultaneous transformation from our „main cone“ to the isotonic cone projection and our projection matrix to a diagonal matrix is not possible. We therefore propose a new method. But first let us recall the basic facts about cone projections in the next section.

3.2 Cone Projections

The facts listed here are either quoted or slightly generalized from [Ro88] except where indicated. Consider the d-dimensional Euclidean space V with some inner product $\langle \cdot, \cdot \rangle$.

A closed convex set $C \subset V$ is called a *cone*iff for all $x \in C$ and for all $\alpha \geq 0$ $\alpha x \subset C$ (cone property). Most authors define a cone to be a convex set that fulfills the cone property. However, since we are mainly interested in closed ones and some interesting properties hold only for closed ones, we have included closedness in the definition of a cone.

If a convex set C fulfills only the weak cone property, i.e. $\alpha C \subset C$ for all $\alpha > 0$, then we call C a *semicone*. (Semicones need not be closed.)

For a set $M \subset C$, write $\mathcal{C}(M)$ for the smallest cone containing M, i.e.

$$\mathcal{C}(M) := \bigcap_{\substack{M \subset C \\ C \text{cone}}} C.$$

It is easy to see that this really is a cone.

Our main interest is in finitely generated cones, i.e. cones that are generated by a set $\{\nu_1, \ldots, \nu_m\} \subset V$. We use the short hand notation $\mathcal{C}(\nu_1, \ldots, \nu_m) = \mathcal{C}(\{\nu_1, \ldots, \nu_m\})$. Note that

$$\mathcal{C}(\{\nu_1, \ldots, \nu_m\}) = \left\{ \sum_{k=1}^{m} \alpha_k \nu_k : \ \alpha_1, \ldots, \alpha_m \geq 0. \right\}.$$

Consider a convex closed set $C \subset V$. We have the *closest point property*, i.e. for every $x \in V$ there is a unique $y \in C$ with

$$\|x - y\| = \inf_{z \in C} \|x - z\|,$$

with $\|\cdot\|$ being the usual norm induced by the inner product. We use the notation $\pi(x|C) := y$. The mapping $\pi(\cdot|C)$ is called *projection onto*C with respect to $\langle \cdot, \cdot \rangle$.

Recall that for a vector space $W \subset V$ $\pi(\cdot|W)$ is the usual orthogonal projection with respect to $\langle\cdot,\cdot\rangle$.

If C is a cone then the following conditions are necessary and sufficent for a $y \in V$ to be $\pi(x|C)$:

$$(NSC) \qquad y \in C \qquad \text{and} \qquad \langle x-y, z-y\rangle \leq 0 \qquad \forall z \in C.$$

The geometric interpretation is that the angle between $x-\pi(x|C)$ and $z-\pi(x|C)$ has to be obtuse for every $z \in C$.

Every cone has a dual cone, denoted by C^*, that consists of all vectors with obtuse angle between themselves and all elements of C, i.e.

$$C^* := \{x \in V : \langle x, c\rangle \leq 0 \; \forall c \in C\}.$$

We have

$$\pi(x|C) = x - \pi(x|C^*)$$

([Ro88] p. 17), which can be regarded as a generalization of a corresponding property of orthogonal projections. Moreover, $\langle\pi(x|C), \pi(x|C^*)\rangle = 0$, i.e. $\pi(x|C)$ and $\pi(x|C^*)$ are perpendicular to each other. Note also that $(C^*)^* = C$ holds for closed cones (our cones are defined to be closed).

If $C = \mathcal{C}(\nu_1, \ldots, \nu_m)$ is finitely generated then we can identify $\pi(x|C^*)$ locally with an orthogonal projection onto a suitable subspace of V. Which subspace to use depends on a decomposition of V into semicones, as explained in the next lemma. Note that $W^\perp$ is defined to be the orthogonal complement of W with respect to the inner product $\langle\cdot,\cdot\rangle$. $\text{lin}\{M\}$ denotes the linear hull of M. If $M = \emptyset$, we set $\text{lin}\{M\} := \{0\}$. If $M \subset W$ and $N \subset W^\perp$ we use the notation $M \oplus N = \{m + n : \; m \in M; n \in N\}$. If N is empty we define $M \oplus N := M$, analogously $M \oplus N := N$ for $M = \emptyset$.

Lemma 3.1 *Let $\nu_1, \ldots, \nu_m \in V$ be linearly independent vectors. Suppose $C = \mathcal{C}(\nu_1, \ldots, \nu_m)^*$. Define $\{x \in V : \langle\nu_i, x\rangle < 0, \; i \in \emptyset\} := V$. Then the following is true:*

i) Let $I \subset \{1, \ldots, m\}$. Then

$$\Theta_I := \mathcal{C}(\nu_i, i \in I) \oplus (\text{lin}\{\nu_i : i \in I\}^\perp \cap \{x \in V : \langle\nu_i, x\rangle < 0, \; i \in I^C\})$$

is a semicone.

(ii) $\bigcup_{I\subset\{1,\ldots,m\}} \Theta_I$ is a decomposition of V into disjoint sets.

(iii) If $x \in \Theta_I$ then $\pi(x|C) = \pi(x|\text{lin}\{\nu_i : i \in I\}^\perp)$.

Proof: i) This is straightforward. For $I = \{1, \ldots, m\}$ we have $\Theta_I = \mathcal{C}(\nu_1, \ldots, \nu_m) \oplus \text{lin}\{\nu_1, \ldots, \nu_m\}^\perp$. This is even a proper cone. If $I = \emptyset$ we have $\Theta_I = \{x \in V : \; \langle\nu_i, x\rangle < 0, \; \forall i = 1, \ldots, m\}$. This is the relative interior of C. The third case dealing with a nonempty proper subset I of $\{1, \ldots, m\}$ leads

to a semicone too.
ii) It is easy to see that

$$\bigcup_{I\subset\{1,\dots,m\}} (\mathrm{lin}\{\nu_i : i \in I\}^\perp \cap \{x \in V : \langle \nu_i, x\rangle < 0,\ i \in I^C\})$$

is a decomposion of C into disjoint sets. Therefore

$$\bigcup_{I\subset\{1,\dots,m\}} \pi^{-1}\left((\mathrm{lin}\{\nu_i : i \in I\}^\perp \cap \{x \in V : \langle \nu_i, x\rangle < 0,\ i \in I^C\})|C\right)$$

is a decomposition of V into disjoint sets. Hence it is sufficient to show that

$$\pi^{-1}\left((\mathrm{lin}\{\nu_i : i \in I\}^\perp \cap \{x \in V : \langle \nu_i, x\rangle < 0,\ i \in I^C\})|C\right) = \Theta_I .$$

If $y \in \Theta_I$ then $y = \sum_{i\in I} y_i\nu_i + \tilde{y}$ with $y_i \geq 0,\ i \in I$, and $\tilde{y} \in \mathrm{lin}\{\nu_i : i \in I\}^\perp \cap \{x \in V : \langle \nu_i, x\rangle < 0,\ i \in I^C\}$. From our criterion (NSC) we derive $\pi(y|C) = \tilde{y}$. Indeed, $\tilde{y} \in C$ and for all $k \in C$

$$\langle y - \tilde{y}, k - \tilde{y}\rangle = \sum_{i\in I} y_i \left(\underbrace{\langle \nu_i, k\rangle}_{\leq 0} - \underbrace{\langle \nu_i, \tilde{y}\rangle}_{=0} \right) \leq 0 \ .$$

This shows that $y \in \pi^{-1}\left((\mathrm{lin}\{\nu_i : i \in I\}^\perp \cap \{x \in V : \langle \nu_i, x\rangle < 0,\ i \in I^C\})|C\right)$. Moreover, we can prove iii): obviously, $\tilde{y} = \pi(y|\mathrm{lin}\{\nu_i : i \in I\}^\perp)$. Therefore

$$\pi(y|\mathrm{lin}\{\nu_i : i \in I\}^\perp) = \pi(y|C)$$

for $y \in \Theta_I$.

Suppose that $y \in \pi^{-1}\left((\mathrm{lin}\{\nu_i : i \in I\}^\perp \cap \{x \in V : \langle \nu_i, x\rangle < 0,\ i \in I^C\})|C\right)$. We have $y = \pi(y|C^*) + \pi(y|C)$. Recall that $\langle \pi(y|C), \pi(y|C^*)\rangle = 0$. In particular, $\pi(y|C) \in (\mathrm{lin}\{\nu_i : i \in I\})^\perp$. Hence $\pi(y|C^*) \in \mathrm{lin}\{\nu_i : i \in I\}$. Therefore $\pi(y|C^*) = \sum_{i\in I} y_i\nu_i$. Recall that $C^* = ((\mathcal{C}(\nu_1, \dots, \nu_m))^*)^* = \mathcal{C}(\nu_1, \dots, \nu_m)$. Since the $\nu_1, \dots, \nu_m$ are linearly independent, we have $y_i \geq 0,\ i \in I$. To summarize, $y = \sum_{i\in I} y_i\nu_i + \tilde{y}$ with $\tilde{y} = \pi(y|C) \in (\mathrm{lin}\{\nu_i : i \in I\}^\perp \cap \{x \in V : \langle \nu_i, x\rangle < 0,\ i \in I^C\}$, hence $y \in \Theta_I$. □

The following lemma shows how cone projections behave under linear transformations. Before stating it, assume that we have another vector space V_1 that is isomorphic to V via a linear mapping $A : V_1 \to V$. Define an inner product on V_1 via $\langle x, y\rangle_A := \langle Ax, Ay\rangle$. $(V_1, \langle\cdot,\cdot\rangle_A)$ is an finitely dimensional Euclidean space too. If C is a cone in V_1, then AC turns out to be a cone in V. The next lemma discusses how the corresponding cone projections are related to each other. With the obvious definitions for $\pi_{\langle\cdot,\cdot\rangle}$ and $\pi_{\langle\cdot,\cdot\rangle_A}$ we have

Lemma 3.2 *Let $(V, \langle\cdot,\cdot\rangle)$ be a finite dimensional Euclidean space and $A : V_1 \to V$ an isomorphism from a vector space V_1 to V. Let $x \in V_1$ and C be some cone in V_1. Then*

$$A^{-1}\pi_{\langle\cdot,\cdot\rangle}(Ax|AC) = \pi_{\langle\cdot,\cdot\rangle_A}(x|C).$$

Proof: This is derived with our condition (NSC). Obviously, $A^{-1}\pi_{\langle\cdot,\cdot\rangle}(Ax|AC) \in C$. Let $k \in C$. Therefore $Ak \in AC$. Hence

$$\begin{aligned}&\langle x - A^{-1}\pi_{\langle\cdot,\cdot\rangle}(Ax|AC), k - A^{-1}\pi_{\langle\cdot,\cdot\rangle}(Ax|AC)\rangle_A \\ &\quad = \langle Ax - \pi_{\langle\cdot,\cdot\rangle}(Ax|AC), Ak - \pi_{\langle\cdot,\cdot\rangle}(Ax|AC)\rangle \le 0.\end{aligned}$$

□

We usually work with inner products that are defined by some positive definite symmetric matrix, say B. Positive definite always means strictly positive definite. Then $\langle x, y\rangle_B := x^T By$ defines an inner product on $\mathbb{R}^d$. We use the notation $\pi_B(x|C)$, $\|\cdot\|_B$, C^{*B} etc., if we want to stress the dependence on the matrix B. For instance, we have from the lemma above

$$\pi_{A^TBA}(x|C) = A^{-1}\pi_B(Ax|AC).$$

We need the distribution of $\|\pi(Z|C) - \pi(Z|W)\|^2$ and $\|\pi(Z|C) - Z\|^2$ for the analysis of the likelihood ratio tests later on. C and W are then approximations of some sets $\Psi(\Theta)$ and $\Psi(\mathrm{lin}\{v_0\})$ that are denoting alternative and hypothesis in a decision problem. W is a vector space that is contained in C. It turns out that the distributions are mixtures of χ^2-distributions with different degrees of freedom. The result is similar to [Ro88] (see theorem 2.3.1).

Theorem 3.3 *Let B be a positive definite symmetric matrix defining the inner product $\langle\cdot,\cdot\rangle := \langle\cdot,\cdot\rangle_B$. Suppose $\nu_1, \ldots, \nu_m \in \mathbb{R}^d$ to be linearly independent, $C = \mathcal{C}(\nu_1, \ldots, \nu_m)^{*B}$ and Θ_I to be as in the lemma 3.1 above. Assume that $W \subset \mathbb{R}^d$ is a subspace contained in C and that Z is a d-dimensional normally distributed centered random variable with covariance matrix B^{-1}.*

Then for all Borel sets $A_1, A_2 \in \mathbb{R}$,

$$\begin{aligned}&P(\|\pi_B(Z|C) - Z\|_B^2 \in A_1,\ \|\pi_B(Z|C) - \pi_B(Z|W)\|_B^2 \in A_2) \\ &\quad = \sum_{I\subset\{1,\ldots,m\}} P\left(Z \in \Theta_I\right)\chi^2_{\#I}(A_1)\chi^2_{d-\dim W-\#I}(A_2)\ .\end{aligned}$$

Remark: We write $\kappa_i := \sum_{\#I=i} P(Z \in \Theta_I)$ for the mixing coefficients.

Proof: Lemma 3.1 above yields

$$\begin{aligned}&P(\|\pi_B(Z|C) - Z\|_B^2 \in A_1,\ \|\pi_B(Z|C) - \pi_B(Z|W)\|_B^2 \in A_2) \\ &\quad = \sum_{I\subset\{1,\ldots,m\}} P\left(Z \in \Theta_I,\ \|\pi_B(Z|C) - Z\|_B^2 \in A_1,\ \|\pi_B(Z|C) - \pi_B(Z|W)\|_B^2 \in A_2\right)\ .\end{aligned}$$

Therefore it is sufficient to show

$$\begin{aligned} &P\left(Z \in \Theta_I,\ \|\pi_B(Z|C) - Z\|_B^2 \in A_1,\ \|\pi_B(Z|C) - \pi_B(Z|W)\|_B^2 \in A_2\right) \\ &\qquad\qquad\qquad = \quad P\left(Z \in \Theta_I\right)\ \chi_{\#I}(A_1)\ \chi_{d-\dim W-\#I}(A_2). \end{aligned}$$

There exists an invertible matrix $L \in \mathbb{R}^{d\times d}$ with $B = L^T L$ (Cholesky decomposition). We fix some $I \subset \{1,\ldots,m\}$. It is easy to see that $W \subset \mathrm{lin}\{\nu_i : i = 1,\ldots,m\}^{\perp_B}$, indeed: if $v \in W$ then $v \in C$, hence $\langle \nu_i, v\rangle_B \leq 0$ for all $i = 1,\ldots,m$. On the other hand, if $v \in W$ then $-v \in W$, hence $\langle \nu_i, -v\rangle_B \leq 0$ for all $i = 1,\ldots,m$. Therefore $\langle \nu_i, v\rangle_B = 0$, hence $v \in \{\nu_1,\ldots,\nu_m\}^{\perp_B}$.

Since $L(W^{\perp_B}) = (LW)^{\perp_{\mathrm{id}}}$ holds, there is an orthonormal basis (ONB) $w_1,\ldots,w_d$ of $\mathbb{R}^d$ with respect to the inner product $\langle\cdot,\cdot\rangle_{\mathrm{id}}$, such that $w_1,\ldots,w_{\#I}$ is an ONB of $\mathrm{lin}\{L\nu_i : i \in I\}$, and $w_1,\ldots,w_{d-\dim W}$ is an ONB of $(LW)^{\perp_{\mathrm{id}}}$ and $w_{d-\dim W+1},\ldots,w_d$ is an ONB of LW. Suppose Q to be the orthogonal matrix with columns $w_1,\ldots,w_d$. If $Z \in \Theta_I$, then

$$\begin{aligned} \|\pi_B(Z|C) - Z\|_B^2 &= \|\pi_B(Z|\mathrm{lin}\{\nu_i : i \in I\}^{\perp_B}) - Z\|_B^2 \\ &= \|\pi_{L^TL}(Z|\mathrm{lin}\{\nu_i : i \in I\})\|_B^2 \\ &= \|L^{-1}\pi_{\mathrm{id}}(LZ|\mathrm{lin}\{L\nu_i : i \in I\})\|_B^2 \\ &= \|\pi_{QQ^T}(LZ|\mathrm{lin}\{L\nu_i : i \in I\})\|_{\mathrm{id}}^2 \\ &= \|Q^T\pi_{\mathrm{id}}(Q^TLZ|\mathrm{lin}\{Q^TL\nu_i : i \in I\})\|_{\mathrm{id}}^2 \\ &= \|\sum_{i=1}^{\#I}(Q^TLZ)_i e_i\|_{\mathrm{id}}^2 \\ &= \sum_{i=1}^{\#I}(Q^TLZ)_i^2\ . \\ \|\pi_B(Z|C) - \pi_B(Z|W)\|_B^2 &= \|\pi_B(Z|\mathrm{lin}\{\nu_i : i \in I\}^{\perp_B}) - \pi_B(Z|W)\|_B^2 \\ &= \|\pi_B(Z|\mathrm{lin}\{\nu_i : i \in I\}) - \pi_B(Z|W^{\perp_B})\|_B^2 \\ &= \|\pi_{\mathrm{id}}(LZ|\mathrm{lin}\{L\nu_i : i \in I\}) - \pi_{\mathrm{id}}(LZ|(LW)^{\perp_{\mathrm{id}}})\|_{\mathrm{id}}^2 \\ &= \|\pi_{\mathrm{id}}(Q^TLZ|\mathrm{lin}\{Q^TL\nu_i : i \in I\}) - \pi_{\mathrm{id}}(Q^TLZ|Q^T(LW^{\perp_{\mathrm{id}}}))\|_{\mathrm{id}}^2 \\ &= \sum_{k=\#I+1}^{d-\dim W}(Q^TLZ)_k^2. \end{aligned}$$

Again note that $LW^{\perp_B} = (LW)^{\perp_{\mathrm{id}}}$ for a subspace $W \subset \mathbb{R}^d$. Hence

$$\begin{aligned}
& Q^T L\Big(\mathcal{C}(\nu_i|i\in I) \\
& \quad \oplus \big(\mathrm{lin}\{\nu_i : i \in I\}^{\perp_B} \cap \{x \in \mathbb{R}^d : \langle \nu_i, x\rangle_B < 0,\ i \in I^C\}\big)\Big) \\
= {} & Q^T\Big(\mathcal{C}(L\nu_i|i\in I) \\
& \quad \oplus \big(\mathrm{lin}\{L\nu_i : i \in I\}^{\perp_{\mathrm{id}}} \cap \{Lx \in \mathbb{R}^d : \langle \nu_i, x\rangle_B < 0,\ i \in I^C\}\big)\Big) \\
= {} & \mathcal{C}(Q^T L\nu_i|i\in I) \\
& \oplus \big(\underbrace{\mathrm{lin}\{Q^T L\nu_i : i \in I\}^{\perp_{\mathrm{id}}}}_{=\mathrm{lin}\{e_{\#I+1},\dots,e_{d-\dim W}\}\oplus\mathrm{lin}\{e_{d-\dim W+1},\dots,e_d\}} \\
& \quad \cap \{y \in \mathbb{R}^d : (Q^T L\nu_i)^T y < 0,\ i \in I^C\}\big) \\
= {} & \mathcal{C}(Q^T L\nu_i|i\in I) \\
& \oplus \big(\underbrace{\mathrm{lin}\{Q^T L\nu_i : i \in I\}^{\perp_{\mathrm{id}}}}_{\mathrm{lin}\{e_{\#I+1},\dots,e_{d-\dim W}\}} \oplus Q^T LW\} \cap \{y \in \mathbb{R}^d : (Q^T L\nu_i)^T y < 0,\ i \in I^C\}\big) \\
= {} & \underbrace{\mathcal{C}(Q^T L\nu_i|i\in I)}_{\subset\mathrm{lin}\{e_1,\dots,e_{\#I}\}} \\
& \oplus \big(\big(\mathrm{lin}\{e_{\#I+1},\dots,e_{d-\dim W}\} \cap \{y \in \mathbb{R}^d : (Q^T L\nu_i)^T y < 0,\ i \in I^C\}\big) \oplus Q^T LW\big)\ .
\end{aligned}$$

Note that $\mathrm{Cov}(Q_I^T LZ, Q_I^T LZ) = \mathrm{id}$. Therefore the random variables $\tilde{Z}_i = (Q^T LZ)_i$ are independent and standard normally distributed. We introduce polar coordinates

$$\begin{aligned}
\tilde{Z}_1 &= R_1 \sin\theta_1, \\
\tilde{Z}_2 &= R_1 \cos\theta_1 \sin\theta_2,\ ,\dots \\
\tilde{Z}_i &= R_1 \cos\theta_1 \cdot \dots \cdot \cos\theta_{i-1} \sin\theta_i, \dots \\
\tilde{Z}_{\#I} &= R_1 \cos\theta_1 \cdot \dots \cdot \cos\theta_{\#I-1}, \\
\tilde{Z}_{\#I+1} &= R_2 \sin\theta_{\#I+1}, \\
\tilde{Z}_{\#I+2} &= R_2 \cos\theta_{\#I+1} \sin\theta_{\#I+2},\ \dots \\
\tilde{Z}_i &= R_2 \cos\theta_{\#I+1} \cdot \dots \cdot \cos\theta_{i-1} \sin\theta_i,\ \dots \\
\tilde{Z}_{d-\dim W} &= R_2 \cos\theta_{\#I+1} \cdot \dots \cdot \cos\theta_{d-\dim W-1}\ .
\end{aligned}$$

The random variables $R_1, R_2, \theta_1, \dots, \theta_{\#I-1}, \theta_{\#I+1}, \dots, \theta_{d-\dim W-1}, \tilde{Z}_{d-\dim W+1}, \dots, \tilde{Z}_d$ are independent (see [Ro88], p. 71, or apply a density transformation argument). In particular,

$$R_1^2 = \sum_{i=1}^{\#I} \tilde{Z}_i^2 \sim \chi_{\#I} \qquad \text{and} \quad R_2^2 \sim \sum_{i=\#I+1}^{d-\dim W} \tilde{Z}_i^2 \chi_{d-\dim W-\#I}.$$

Define

$$\begin{aligned}
M_i &:= \left(\sin\theta_i \prod_{l=1}^{i-1}\cos\theta_l\right) e_i, \quad i = 1,\ldots \#I-1,\\
M_{\#I} &:= \left(\prod_{l=1}^{\#I-1}\cos\theta_l\right) e_{\#I},\\
M_i &:= \left(\sin\theta_i \prod_{l=\#I+1}^{i}\cos(\theta_l)\right) e_i, \quad i = \#I+1,\ldots,d-\dim W-1,\\
M_{d-\dim W} &:= \left(\prod_{l=\#I+1}^{d-\dim W-1}\cos(\theta_l)\right) e_{d-\dim W}.
\end{aligned}$$

Note that

$$\mathrm{lin}\{e_{\#I+1},\ldots,e_{d-\dim W+1}\} \cap \{y \in \mathbb{R}^d : (Q^T L\nu_i)^T y < 0,\ i \in I^C\}.$$

is a semicone. Then the following holds

$$\begin{aligned}
&P\left(Z \in \Theta_I,\ \|\pi_B(Z|C) - Z\|_B^2 \in A_1,\ \|\pi_B(Z|C) - \pi_B(Z|W)\|_B^2 \in A_2)\right)\\
&= P\Bigg(\sum_{i=1}^{\#I} M_i \in \mathcal{C}(Q^T L\nu_i | i \in I),\\
&\qquad \sum_{i=\#I+1}^{d-\dim W} M_i \in \mathrm{lin}\{e_{\#I+1},\ldots,e_{d-\dim W}\} \cap \{y \in \mathbb{R}^d : (Q^T L\nu_i)^T y < 0,\ i \in I^C\},\\
&\qquad \sum_{i=d-\dim W+1}^{d} \tilde{Z}_i e_i \in Q^T LW,\ R_1^2 \in A_1,\ R_2^2 \in A_2)\Bigg)\\
&= P(\sum_{i=1}^{\#I} M_i \in \mathcal{C}(Q^T L\nu_i | i \in I))\\
&\quad \cdot P(\sum_{i=\#I+1}^{d-\dim W} M_i \in \mathrm{lin}\{e_{\#I+1},\ldots,e_{d-\dim W}\} \cap \{y \in \mathbb{R}^d : (Q^T L\nu_i)^T y < 0,\ i \in I^C\},)\\
&\quad \underbrace{\cdot P(\sum_{d-\dim W+1}^{d} \tilde{Z}_i e_i \in Q^T LW)}_{=1} P(R_1^2 \in A_1) P(R_2^2 \in A_2)\\
&= P(Z \in \Theta_I)\chi^2_{\#I}(A_1)\chi^2_{d-\dim W-\#I}(A_2)\ .
\end{aligned}$$

This proves the assertion. □

3.3 Maximum Likelihood and Local Cones

Consider the simplex $M_1^d := \{p \in \mathbb{R}^d : \; p_i \geq 0; \sum_{i=1}^d p_i = 1\}$. This set can be regarded as the set of probability measures on the finite set $\{1, \ldots, d\}$. We will use the notation $P(\{k\}) = P(k) = P_k$ for $P \in M_1^d$ and $k \in \{1, \ldots, d\}$.

Let K be some nonempty compact subset of M_1^d. Suppose (Z_i) to be an iid-sequence of discrete random variables defined on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ taking values in the set $\{1, \ldots, d\}$. Assume that the distribution of Z is some unknown $P \in K$. We want to use the maximum likelihood method to estimate P. The likelihood of P given $i_1, \ldots, i_d \in \{1, \ldots, d\}$ is

$$\mathbb{P}_P(Z_1 = i_1, \ldots, Z_k = i_d) = P(\{i_1\}) \cdot \cdots \cdot P(\{i_d\}).$$

If we write down the vector of relative frequencies $\bar{r}^n$ with entries

$$\bar{r}_i^n = \frac{1}{n} \sum_{l=1}^{n} 1_{Z_l = i},$$

then it is easy to see that the maximization of the likelihood is equivalent to the maximization of the function

$$P \mapsto L(P|\bar{r}^n)$$

with $L(x|y) := \sum_{l=1}^d y_k \log x_k$. There is always a maximizing $P \in K$, since $L(\cdot|\bar{r}^n)$ is upper semicontinuous. This P might not be unique. We therefore define a mapping $M_1^d \ni r \mapsto \Theta(r) := \{P \in K : L(P|r) = \max_{Q \in K} L(Q|r)\}$. Note that $\Theta(r) = \{L(\cdot|r) \geq \max_{Q \in K} L(Q|r)\}$ is closed, hence compact, because of the upper semicontinuity of $L(\cdot|r)$. Then $\Theta(r)$ is the set of maximum likelihood estimators given the frequency vector r. If $r \in K$, then it is well known that r itself maximizes $L(\cdot|r)$ over K. Moreover, $\Theta(r) = \{r\}$ in this case.

Define also

$$\delta(r_1, r_2) := \sup\{|P_1 - P_2| : P_1 \in \Theta(r_1), \; P_2 \in \Theta(r_2)\}.$$

Then we have the following continuity property:

Lemma 3.4 *Let $r \in M_1^d$. Assume that $r_i > 0$, $i = 1, \ldots, d$. Assume that $\Theta(r) = \{P\}$, $P \in K$, $P(i) > 0$, $i = 1, \ldots, d$. Then* $\lim_{\substack{r' \to r \\ r' \in M_1^d}} \delta(r^n, r) = 0$.

Remark: This is a continuity property of Θ holding in „amenable“ points r. From this property the strong consistency is derived in the following way: Assume that the true distribution $P \in K$ of Z_i has the property $P(k) > 0$, $k = 1, \ldots, d$. Then $\Theta(P) = \{P\}$ holds, as noted before. Choose $N^C := \{\omega \in \Omega : \lim_{n \to \infty} \bar{r}^n(\omega) = P\}$. It is clear from the law of large numbers that $\mathbb{P}(N) = 0$ and $\delta(\bar{r}^n(\omega), P) \to 0$ for all $\omega \in N^C$ from the lemma.

Proof: Assume that $\limsup_{\substack{r'\to r\\ r'\in M_1^d}} \delta(r',r) \geq \epsilon > 0$. Then there exists a sequence $r^n \in M_1^d$ with $\limsup \delta(r^n,r) \geq \epsilon$. Since every $\Theta(r^n)$ is compact we can choose a sequence $P^n \in K$ with $P^n \in \Theta(r^n)$ and $\delta(r^n,r) = |P^n - P|$. Since K is compact, there is some subsequence n' and a $Q \in K$ and $\lim_{n'\to\infty} P^{n'} = Q$. Furthermore, we can assume that $Q \neq P$.

Note that $Q(k) > 0$, $k = 1,\ldots,d$. Otherwise we would have $L(P^{n'}|r^{n'}) \to -\infty$, since $r^{n'} \to r$ and $r_i > 0$, $i = 1,\ldots,d$. Because of $L(P|r^{n'}) > \infty$ for all n', we would have $L(r|r^{n'}) > L(P^{n'}|r^{n'})$ for n' big enough. This is a contradiction, because P^n maximizes $L(\cdot|r^n)$ on K.

Since $Q(k) > 0$, we have $L(P^n|r^n) \to L(Q|r)$. However:

$$L(Q|r) = \lim_{n'\to\infty} L(P^{n'}|r^{n'}) \geq \lim_{n'\to\infty} L(P|r^{n'}) = \max_{R\in K} L(R|r).$$

Hence $Q \in \Theta(r) = \{P\}$, a contradiction. This proves the assertion. $\square$

Lemma 3.5 *Suppose that $r^n, r \in M_1^d$ and $r_i > 0$, $i = 1,\ldots,d$, $r \in K$. Assume, that $\lim_{n\to\infty} \sqrt{n}(r^n - r) = g$. Then for n big enough the set $\sqrt{n}(\Theta(r^n) - r)$ is included in the set*

$$\left\{ y \in \mathbb{R}^d : \sum_k y_k = 0,\ \|y\|_2 \leq \frac{2}{1-\log(2)} \max_k r_k (1 + \|g/r\|_2). \right\},$$

with $\|\cdot\|_2$ denoting the usual euclidean norm $\|x\|_2 := \sqrt{\sum_{l=1}^d x_k^2}$. ($x/r$ denotes the vector with components x_k/r_k). Hence $\bigcup_n \sqrt{n}(\Theta(r^n) - r)$ is a bounded set.

Proof: Note that $\|\frac{\sqrt{n}(r^n-r)}{r}\|_2 \leq 1 + \|g/r\|_2$ for n big enough. Note also that $r_k^n > r_k/2$ for n big enough. Define

$$A_n := \{x \in M_1^d :\ L(x|r^n) \geq L(r|r^n),\ 0 < x_k < 2r_k,\ k = 1,\ldots,d\}.$$

Let n be big enough such that $r_i^n > 0$, $i = 1,\ldots,d$. Then every $x \in \Theta(r^n)$ fulfills $x_k > 0$, $k = 1,\ldots,d$. Note that $\Theta(r) = \{r\}$. Since $\lim_{n\to\infty} \delta(r^n,r) = 0$ we have $x_k < 2r_k$, $k = 1,\ldots,d$, for n big enough. Hence $\Theta(r^n) \subset A_n$ for n big enough. We now assume that n is big enough, such that all statements above hold.

Fix some $x \in A_n$. Then $0 < x_k < 2r_k$ holds for $k = 1,\ldots,d$. This implies

$$\left|\frac{x_k - r_k}{r_k}\right| < 1.$$

Consider the function

$$g(h) := \frac{\log(1+h) - h}{h^2} \text{ for all } h \in (-1,1]\backslash\{0\}, \qquad g(0) := -1/2.$$

It is easy to check that g is continuous and nondecreasing. Hence $g(h) \leq g(1) = \log(2) - 1 < 0$. This gives us the inequality

$$\log(1+h) \leq h - (1-\log(2))h^2.$$

Hence

$$\log\left(\frac{x_k - r_k}{r_k} + 1\right) \leq \frac{x_k - r_k}{r_k} - (1-\log(2))\frac{(x_k - r_k)^2}{r_k^2}.$$

This yields

$$0 \leq \sum_k r_k^n \log\frac{x_k}{r_k} \leq \sum_k r_k^n \frac{x_k - r_k}{r_k} - (1-\log(2))\sum_k r_k^n \frac{(x_k - r_k)^2}{r_k^2}.$$

Now consider the sets $B_n := \sqrt{n}(A_n - r)$. If $y \in B_n$ then there exists an $x \in A_n$ with $x = n^{-1/2}y + r$. Hence y satisfies

$$0 \leq \frac{1}{\sqrt{n}}\sum_k r_k^n \frac{y_k}{r_k} - n^{-1}(1-\log(2))\sum_k r_k^n \frac{(y_k)^2}{r_k^2}.$$

Equivalently,

$$0 \leq \sqrt{n}\sum_k r_k^n \frac{y_k}{r_k} - (1-\log(2))\sum_k r_k^n \frac{(y_k)^2}{r_k^2}.$$

Note that $\sum_k y_k$ must be zero, hence

$$0 \leq \sum_k \sqrt{n}(r_k^n - r_k)\frac{y_k}{r_k} - (1-\log(2))\sum_k r_k^n \frac{(y_k)^2}{r_k^2}.$$

The Cauchy-Schwarz inequality implies

$$\sum_k \sqrt{n}(r_k^n - r_k)\frac{y_k}{r_k} \leq \|y\|_2 \left\|\frac{\sqrt{n}(r^n - r)}{r}\right\|_2 \leq \|y\|_2(\|g/r\|_2 + 1).$$

We also have

$$\sum_k r_k^n \frac{(y_k)^2}{r_k^2} \geq \frac{1}{2}\sum_k r_k \frac{(y_k)^2}{r_k^2} = \frac{1}{2}\sum_k \frac{(y_k)^2}{r_k} \geq \frac{1}{2}\|y\|_2^2 \min_k 1/r_k.$$

To summarize, we have proved for $y \in \sqrt{n}(A_n - r)$ the inequality

$$0 \leq \|y\|_2(1 + \|g/r\|_2) - \frac{1}{2}(1-\log(2))\min_k \frac{1}{r_k}\|y\|_2^2.$$

Hence

$$\|y\|_2 \leq \frac{2}{1-\log(2)}\max_k r_k(1 + \|g/r\|_2).$$

This proves the assertion.
□

We will assume that K can be approximated in some neighbourhood of $P \in K$ using some cone in the following sense:

Suppose that there is an open convex set $O_1 \subset \mathbb{R}^{d-1}$ containing 0 and an open set $O_2 \subset P + \{x \in \mathbb{R}^d : \sum_{l=1}^d x_l = 0\}$ containing P and a C^2-diffeomorphism $\Phi : O_1 \to O_2$, i.e. $\Phi = (\Phi_1, \ldots \Phi_d)^T$ is a C^2-mapping from $O_1 \to \mathbb{R}^d$ taking values in $r + \{x \in \mathbb{R}^d : \sum_{l=1}^d x_l = 0\}$ and the derivative Φ'_x, $x \in O_1$, is an isomorphism from $\mathbb{R}^{d-1}$ onto $\{x \in \mathbb{R}^d : \sum_{l=1}^d x_l = 0\}$. Let C be a cone $C \subset \mathbb{R}^{d-1}$. If

$$\begin{aligned} \Phi(0) &= P \\ \Phi(C \cap O_1) &= K \cap O_2 \end{aligned}$$

holds then we say that K is *approximated by a cone in P with representation tuple* (O_1, O_2, Φ, C).

Remarks: we assume $P(i) > 0$, $i = 1, \ldots, d$, since otherwise we could reduce the dimension of the problem. At P the family of probability measures $(P)_{P \in K}$ is locally embedded in a regular parametric model $\{\Phi(\theta) : \theta \in O_1\}$. The Fisher infomation matrix $I = -(L(\Phi)|P)''_0$ is of some interest here. Compute the derivative of $x \mapsto L(\Phi(x)|s)$ for $s \in M_1^d$ and $x \in O_1$. This is a linear mapping from $\mathbb{R}^{d-1} \to \mathbb{R}$, i.e.

$$L(\Phi|s)'_x h = \left(\frac{s_1}{\Phi_1(x)}, \ldots, \frac{s_d}{\Phi_d(x)} \right) \Phi'_x h, \qquad h \in \mathbb{R}^{d-1}.$$

The second derivative is a symmetric bilinear form on $\mathbb{R}^{d-1} \times \mathbb{R}^{d-1}$.

$$L(\Phi|s))''_x(u, v) = (\frac{s_1}{\Phi_1(x)}, \ldots, \frac{s_d}{\Phi_d(x)})\Phi''_x(u, v) + u^T(\Phi'_x)^T \mathrm{diag}(-\frac{s_1}{\Phi_1^2}, \ldots, -\frac{s_d}{\Phi_d^2})\Phi'_x v.$$

Note that $\Phi(0) = P$, hence

$$L(\Phi|P))''_0(u, v) = (1, \ldots, 1)\Phi''_0(u, v) + u^T(\Phi'_0)^T \mathrm{diag}\left(-\frac{1}{P(1)}, \ldots, -\frac{1}{P(d)} \right) \Phi'_0 v.$$

Φ takes only values in $P + \{x \in \mathbb{R}^d : \sum_i x_i = 0\}$, hence $(1, \ldots 1)\Phi(x) = 1$ for all x in O_1. Therefore $(1, \ldots 1)\Phi'_x = 0$ and $(1, \ldots 1)\Phi''_x = 0$ for all x in O_1. Hence we have

$$L(\Phi|P))''_0(u, v) = -u^T(\Phi'_0)^T \mathrm{diag}\left(\frac{1}{P(1)}, \ldots, \frac{1}{P(d)} \right) \Phi'_0 v = -u^T I v.$$

We should be aware of the special form of Φ. The components are given by

$\Phi = (\Phi_1, \dots, \Phi_d)^T = (\Phi_1, \dots, \Phi_{d-1}, -\sum_{l=1}^{d-1} \Phi_l)^T$. Therefore

$$\Phi_0' = \begin{pmatrix} (\nabla\Phi_1)_0 \\ \vdots \\ (\nabla\Phi_{d-1})_0 \\ -\sum_{l=1}^{d-1}(\nabla\Phi_l)_0 \end{pmatrix} = \begin{pmatrix} \mathrm{id}_{d-1\times d-1} \\ -1 \cdots -1 \end{pmatrix} \begin{pmatrix} (\nabla\Phi_1)_0 \\ \vdots \\ (\nabla\Phi_{d-1})_0 \end{pmatrix} =: \begin{pmatrix} \mathrm{id}_{d-1\times d-1} \\ -1 \cdots -1 \end{pmatrix} A.$$

Note that the first matrix defines an isomorphism from $\mathbb{R}^{d-1}$ to $\{x \in \mathbb{R}^d : \sum_i x_i = 0\}$. Since Φ_0' is an isomorphism from $\mathbb{R}^{d-1}$ to $\{x \in \mathbb{R}^d : \sum_i x_i = 0\}$ by definition, the matrix A must be an isomorphism from $\mathbb{R}^{d-1}$ onto $\mathbb{R}^{d-1}$. Let us have a look at the inverse Φ_0^{-1} defining an isomorphism form $\{x \in \mathbb{R}^d : \sum_k x_k = 0\}$ onto $\mathbb{R}^{d-1}$. It is now easy to calculate

$$\Phi_0^{-1} = A^{-1} \begin{pmatrix} \mathrm{id}_{d-1\times d-1} \\ -1 \cdots -1 \end{pmatrix}^{-1} = A^{-1} \begin{pmatrix} & 0 \\ \mathrm{id}_{d-1\times d-1} & \vdots \\ & 0 \end{pmatrix}.$$

The Fisher information can be written as

$$\begin{aligned} I &= A^T \begin{pmatrix} & -1 \\ \mathrm{id}_{d-1\times d-1} & \vdots \\ & -1 \end{pmatrix} \mathrm{diag}\left(\frac{1}{P(1)}, \dots, \frac{1}{P(d)}\right) \begin{pmatrix} \mathrm{id}_{d-1\times d-1} \\ -1 \cdots -1 \end{pmatrix} A \\ &= A^T \left(\mathrm{diag}\left(\frac{1}{P(1)}, \dots, \frac{1}{P(d-1)}\right) + \frac{1}{P(d)}\mathbf{1}\mathbf{1}^T\right) A \end{aligned}$$

with $\mathbf{1} = (1, \dots, 1)^T \in \mathbb{R}^{d-1}$. I is the product of three invertible $(d-1) \times (d-1)$-matrices. The inverse of the matrix in the middle is given by

$$\left(\mathrm{diag}\left(\frac{1}{P(1)}, \dots, \frac{1}{P(d-1)}\right) + \frac{1}{P(d)}\mathbf{1}\mathbf{1}^T\right)^{-1} = (\delta_{ij}P(i) - P(i)P(j))_{\substack{1\le i\le d-1 \\ 1\le j\le d-1}}.$$

The latter is the covariance matrix of $\sqrt{n}(\bar{r}^n - P)$ if we consider the first $d-1$ components of the vector only. The Fisher information depends on the underlying choice of our regular model $\Phi(O_1)$. Obviously, I is positive definite symmetric matrix. We can therefore compute the Cholesky decomposition $I = \tilde{L}^T\tilde{L}$ with $\tilde{L}$ invertible. Defining $\tilde{\Phi}(x) := \Phi(\tilde{L}^{-1})(x)$, $\tilde{O}_1 := \tilde{L}^{-1}O_1$ and $\tilde{C} := \tilde{L}^{-1}C$ will give us a another approximation tuple $(\tilde{O}_1, O_2, \tilde{\Phi}, \tilde{C})$. The Fisher information is then $\tilde{I} = \mathrm{id}_{r-1\times r-1}$. Its form is much simpler at the cost of a perhaps more complicated cone.

Lemma 3.6 *Assume that K is approximated by a cone in P with approximation tuple (O_1, O_2, Φ, C). Suppose that $P(i) > 0$, $i = 1, \ldots, d$.*
Then there exists an open set $O \subset M_1^d$ containing P, such that $\Theta(r') \subset \Phi(C \cap O_1)$ for all $r' \in O$ and $\Theta(r')$ is a singleton.

Proof: Consider the representation tuple (O_1, O_2, Φ, C) for the approximation of K in $P \in K$. Note that $L(\Phi|P)_0'' = -I$ is a strictly negative definite symmetric bilinear form on $\mathbb{R}^{d-1}$. Indeed, the bilinear form given by $(u, v) \mapsto B(u, v) := \sum_i \frac{u_i v_i}{r_i}$ is strictly positive definite. Its restriction on $\{x \in \mathbb{R}^d : \sum_l x_l = 0\}^2$ is strictly positive definite too. Since Φ_0' is an isomorphism from $\mathbb{R}^{d-1}$ onto $\{x \in \mathbb{R}^d : \sum_l x_l = 0\}$, $B(\Phi_0', \Phi_0')$ is strictly positive definite.

The set of strictly negative bilinear forms on $\mathbb{R}^{d-1}$ is an open subset in the set of symmetric bilinear forms. Note that $L(\Phi|s)_x''$ depends continuously on $s \in \{u \in M_1 : u_i > 0\}$ and $x \in O_1$, provided that $\Phi_k(x) > 0$. Since Φ is continuous there is an open set $O_3 \subset O_1$ with $\Phi(O_3) \subset \{u \in M_1 : u_i > 0\}$ and an open set $O_4 \subset \{x \in M_1 : x_i > 0\}$ containing P such that $L(\Phi|s)_x''$ is a stricty negative definite symmetric bilinear form for all $(x, s) \in O_3 \times O_4$. Without loss of generality we may assume that O_3 is convex (choose a smaller O_3 if necessary). Then the mappings $O_3 \ni x \mapsto L(\Phi(x)|s)$ are strictly concave for every $s \in O_4$.

We have already proved that $\lim_{\substack{r' \to r \\ r' \in M_1^d}} \delta(r', r) = 0$. Hence there is an $\epsilon > 0$ such that $\Theta(r') \subset \Phi(O_3 \cap C)$ for all $r' \in M_1^d$ with $|r' - P| < \epsilon$. Choose this ϵ small enough such that $B_\epsilon = \{r' \in M_1^d : |r' - P| < \epsilon\} \subset O_4$.

Now consider an $r' \in B_\epsilon$. Then $\Theta(r') \subset \Phi(C \cap O_3)$. Then for all $x \in \Phi^{-1}(\Theta(r'))$

$$L(\Phi(x)|r') = \max_K L(\cdot|r') \geq \sup_{C \cap O_3} L(\Phi(\cdot)|r').$$

Note that the function $L(\Phi(\cdot)|r')$ is strictly concave on the convex set $O_3 \cap C$. Hence the argmax of $L(\Phi(\cdot)|r')$ taken over $O_3 \cap C$ is unique. Therefore $\Phi^{-1}(\Theta(r'))$ is a singleton. The same is true for $\Theta(r')$, of course. This proves the assertion setting $O := B_\epsilon$.□

This shows that Θ is a well defined mapping on O. We write $\Theta(s) = r$ iff $\Theta(s) = \{r\}$. The lemma 3.4 shows that Θ is continuous on O. If we define $\theta^n := 1_{\bar{r}^n \in O} \Theta(\bar{r}^n)$ for the relative frequencies $\bar{r}^n$ then we have a measurable mapping that coincides with the maximum likelihood estimator $\Theta(\bar{r}^n)$ for n large enough. Therefore it makes sense to discuss the distributional limit behaviour for $\Theta(\bar{r}^n)$. The next lemma shows that the maximum likelihood estimator behaves locally like a cone projection with respect to some suitable inner product that is given by the Fisher information. We will state the following theorem.

Theorem 3.7 *Assume that K can be approximated by a cone in P with representation triple by (O_1, O_2, Φ, C). Assume that Z is some $d-1$-dimensional*

Gaussian random variable with expectation zero and covariance matrix I^{-1} with I^{-1} denoting the inverse of the Fisher information

$$I = (\Phi_0')^T \mathrm{diag}(1/P(1), \ldots, 1/P(d))\Phi_0'.$$

Assume that P is the distribution of Z_i. Then

$$\sqrt{n}\Phi^{-1}(\Theta(\bar{r}^n)) \to \pi_I(Z|C).$$

Proof: Apply Skohorods representation theorem. Without loss of generality we may assume that $\sqrt{n}(\bar{r}^n - P) \to W'$ a.s. with a Gaussian centered random variable W' with covariance matrix $(P(i)\delta_{ij} - P(i)P(j))_{\substack{1\le i\le d\\1\le j\le d}}$. Note that $W' \in \{y \in \mathbb{R}^d : \sum_i y_i = 0\}$. From the definition we have that Φ can be inverted locally at $\Phi(0)$ as a mapping from O_1 to $P+\{y \in \mathbb{R}^d : \sum_i y_i = 0\}$. Therefore applying the delta method, we have $\sqrt{n}(\Phi^{-1}(\bar{r}^n)) = (\Phi_0')^{-1}W'$ a.s.. Note that $\Phi^{-1}(\bar{r}^n)$ is well defined if n is large enough. Clearly, $(\Phi_0')^{-1}W'$ is a centered Gaussian random variable. Let us compute the covariance of $(\Phi_0')^{-1}W'$, i.e.

$$\begin{aligned}
&E(\Phi_0')^{-1}W'((\Phi_0')^{-1}W')^T \\
&= A^{-1}\begin{pmatrix} & 0 \\ \mathrm{id}_{d-1\times d-1} & \vdots \\ & 0 \end{pmatrix} (P(i)\delta_{ij} - P(i)P(j))_{\substack{1\le i\le d\\1\le j\le d}} \begin{pmatrix} & 0 \\ \mathrm{id}_{d-1\times d-1} & \vdots \\ & 0 \end{pmatrix}^T A^{-T} \\
&= A^{-1}\left(P(i)\delta_{ij} - P(i)P(j)\right)_{\substack{1\le i\le d-1\\1\le j\le d-1}} A^{-T} \\
&= I^{-1}.
\end{aligned}$$

This shows that $(\Phi_0')^{-1}W' \stackrel{\mathcal{D}}{=} Z$. Hence $\pi_I((\Phi_0')^{-1}W'|C) \stackrel{\mathcal{D}}{=} \pi_I(Z|C)$.

Fix some $\omega \in \{\omega \in \Omega : \lim_{n\to\infty} \sqrt{n}(\bar{r}^n(\omega) - P) = W'(\omega)\}$ We have already proved that $\sqrt{n}(\Theta(\bar{r}^n(\omega)) - r)$ stays bounded (lemma 3.5). Since Φ^{-1} is a C^1-mapping, it is easy to show that also $\sqrt{n}\Phi^{-1}(\Theta(\bar{r}^n(\omega)))$ is bounded. Hence there are subsequences n' such that $\sqrt{n'}\Phi^{-1}(\Theta(\bar{r}^{n'}(\omega)))$ converges to a limit $h = h(n', \omega)$. h must be an element of the cone C, indeed: $\Phi^{-1}(\Theta(\bar{r}^{n'}(\omega))) \in C$ for n' large enough. Then $\sqrt{n'}\Phi^{-1}(\Theta(\bar{r}^{n'}(\omega))) \in C$ from the cone property. Since C is closed, we have $h = \lim_{n'\to\infty} \sqrt{n'}\Phi^{-1}(\Theta(\bar{r}^{n'}(\omega))) \in C$.

Let $c \in C$ be arbitrary. We want to show that

$$(*) \qquad \langle(\Phi_0')^{-1}W'(\omega) - h, c - h\rangle_I \le 0.$$

This and the already shown fact that $h \in C$ will imply that $h = \pi_I((\Phi_0')^{-1}W'(\omega))$ by (NSC). Note that the right-hand side of the latter equality, i.e $\pi_I((\Phi_0')^{-1}W'(\omega))$, does not depend on the subsequence n'. The boundness of the sequence $(\sqrt{n}\Phi^{-1}(\Theta(\bar{r}^n(\omega))))$ implies that the limit $\lim_{n\to\infty} \sqrt{n}\Phi^{-1}(\Theta(\bar{r}^n(\omega)))$ exists and has to be equal to $\pi_I((\Phi'_0)^{-1}W'(\omega)|C)$. Hence $\sqrt{n}(\Phi^{-1}(\Theta(\bar{r}^n))) \xrightarrow{\mathcal{D}} \pi_I(W|C)$, showing our assertion.

Let us prove $(*)$. Denote $c_{n'} := n'^{-1/2}c$ and $\xi_{n'} := \Phi^{-1}(\Theta(\bar{r}^{n'}(\omega)))$. Then $c_{n'}, \xi_{n'} \in O_3 \cap C$ for n' big enough. The convexity forces $\xi_{n'} + \epsilon(c_{n'} - \xi_{n'}) \in O_1 \cap C$ for $\epsilon \in [0,1]$. Since $\xi_{n'}$ maximizes $L(\Phi|\bar{r}^{n'}(\omega))$ on $C \cap O_1$, the derivative of $\epsilon \mapsto L(\Phi(\xi_{n'} + \epsilon(c_{n'} - \xi_{n'}))|\bar{r}^{n'}(\omega))$ must be smaller then zero, hence

$$L(\Phi|\bar{r}^{n'}(\omega))'_{\xi_{n'}}(c_{n'} - \xi_{n'}) \leq 0.$$

Note that $\eta^{n'} := \Phi^{-1}(\bar{r}^{n'}(\omega))$ maximizes $L(\Phi(\cdot)|\bar{r}^{n'}(\omega))$ on the open set O_1 for n' big enough, hence

$$L(\Phi(\cdot|\bar{r}^{n'}(\omega))'_{\eta_{n'}} = 0.$$

Since $L(\Phi(\cdot|\bar{r}^{n'}(\omega))$ is twice continuously differentiable on O_1, we can find $\theta^{n'}$ lying on the segment $\{\alpha\xi_{n'} + (1-\alpha)\eta_{n'} : \alpha \in [0,1]\}$ with

$$L(\Phi|\bar{r}^{n'}(\omega))'_{\xi_{n'}}(c_{n'} - \xi_{n'}) = \underbrace{L(\Phi(\cdot|\bar{r}^{n'}(\omega))'_{\eta_{n'}}(c_{n'} - \xi_{n'})}_{=0} + (\xi_{n'} - \eta_{n'})^T L(\Phi(\cdot)|\bar{r}^{n'}(\omega))''_{\theta_{n'}}(c_{n'} - \xi_{n'}).$$

Note that $\theta_{n'} \to 0$, since $\xi_{n'} \to 0$ and $\eta_{n'} \to 0$. Note also that $L(\Phi(\cdot)|\bar{r}^{n'}(\omega))''_{\theta_{n'}} \to -I$, since Φ is twice continuously differentiable. Blowing up with n', we have

$$\begin{aligned} 0 &\leq n'(\xi_{n'} - \eta_{n'})^T L(\Phi(\cdot)|\bar{r}^{n'}(\omega))''_{\theta_{n'}}(c_{n'} - \xi_{n'}) \\ &= (\sqrt{n'}\xi_{n'} - \sqrt{n'}\eta_{n'})^T L(\Phi(\cdot)|\bar{r}^{n'}(\omega))''_{\theta_{n'}}(c - \sqrt{n'}\xi_{n'}), \end{aligned}$$

hence

$$\begin{aligned} 0 &\leq \lim_{n' \to \infty} (\sqrt{n'}\xi_{n'} - \sqrt{n'}\eta_{n'})^T L(\Phi(\cdot)|\bar{r}^{n'}(\omega))''_{\theta_{n'}}(c - \sqrt{n'}\xi_{n'}) \\ &= -\langle h - (\Phi'_0)^{-1}W'(\omega), c - h\rangle_I = \langle (\Phi'_0)^{-1}W'(\omega) - h, c - h\rangle_I. \end{aligned}$$

This shows $(*)$.
□

As a corollary we note

Corollary 3.8 *Let the same conditions hold as in theorem 3.7 above. If W is some centered Gaussian random variable with covariance $(P(i)\delta_{i,j} - P(i)P(j))_{\substack{1\leq i\leq d \\ 1\leq j\leq d}}$, then*

$$\sqrt{n}(\Theta(\bar{r}^n) - P) \xrightarrow{\mathcal{D}} \pi_D(W|\Phi'_0 C)$$

with $D = \mathrm{diag}\left(\frac{1}{P(1)}, \ldots, \frac{1}{P(d)}\right)$.

Proof: From the theorem above and the delta method we have

$$\sqrt{n}\left(\Phi\left(\Phi^{-1}(\Theta(\bar{r}^n))\right) - \Phi(0)\right) = \sqrt{n}\left(\Phi\left(\Phi^{-1}(\Theta(\bar{r}^n))\right) - P\right) \xrightarrow{\mathcal{D}} \Phi'_0\pi_I(Z|C).$$

We have seen in the proof of the last theorem that $Z \stackrel{\mathcal{D}}{=} (\Phi_0')^{-1}W$. Therefore applying lemma 3.2 yields

$$\begin{aligned}\Phi_0'\pi_I(Z|C) &\stackrel{\mathcal{D}}{=} \Phi_0'\pi_I((\Phi_0')^{-1}W|C)\\ &= \Phi_0'\pi_{(\Phi_0')^T\mathrm{diag}(1/P(1),\ldots,1/P(d))\Phi_0}(\Phi_0')^{-1}W|C) = \pi_{\mathrm{diag}(1/P(1),\ldots,1/P(d))}(W|\Phi_0'C).\end{aligned}$$

□

Remark: The theorem and the corollary can be generalized in the following direction dealing with local alternatives. Consider triangular arrays $(Z_k^n)_{\substack{1\le k\le n\\ n\in\mathbb{N}}}$ with $Z_1^n,\ldots,Z_n^n$ independent and identically distributed according to $P_n = P + n^{-1/2}\Phi_0'c + o(n^{-1/2})$, $c \in C$. Consider the relative frequencies $\tilde{r}^n$ with entries

$$\tilde{r}_k^n = \frac{1}{n}\sum_{l=1}^{n} 1_{Z_l^n=k}.$$

It is easy to see that $\sqrt{n}(\bar{r}^n - P) \xrightarrow{\mathcal{D}} \Phi_0'c + Z$ with Z as in the proof of the theorem, i.e. Z is a centered Gaussian random variable with covariance $(P(i)\delta_{ij} - P(i)P(j))_{\substack{1\le i\le d\\ 1\le j\le d}}$. Going through the proofs again we see that

$$\sqrt{n}\Phi^{-1}(\Theta(\tilde{r}_n)) \xrightarrow{\mathcal{D}} \pi_I(c+Z)$$

and

$$\sqrt{n}(\Theta(\tilde{r}_n) - P) \xrightarrow{\mathcal{D}} \pi_{\mathrm{diag}(1/P(1),\ldots,1/P(d))}(\Phi_0'c + W|\Phi_0'C).$$

3.4 Application to the Decompounding Problem

We apply the results of the last section to the estimation of λ and a finite initial segment of p.

Fix some threshold $S \in \mathbb{N}$. Consider the random variables $Z_i := Y_i \wedge (S+1)$. (Z_i) is again an iid-sequence of integer valued random variables. Let $\bar{q}^n$ be the vector of the relative frequencies.

We should pay some attention to the distribution of Z_i. Let us define $\Psi_k^S(x) := T_k\Psi(x)$, $k = 0,\ldots,S$, and $\Psi_{S+1}^S(x) := 1 - \sum_{l=0}^{S} T_k\Psi(x)$. Ψ^S is a mapping from ℓ^1 to $\mathbb{R}^{S+2}$. If $\mathbb{R}^{S+1}$ is embedded canonically into ℓ^1 then Ψ^S is a mapping from $\mathbb{R}^{S+1}$ to $\mathbb{R}^{S+2}$. The components up to S of $\Psi^S(\lambda,p)$ are computed by the Panjer recursion. Analogously, if $x \in \{x \in \mathbb{R}^{S+2}, \sum_{l=0}^{S+1} y_l = 1\}$ and $x_0 > 0$ then the inverse mapping $(\Psi^S)^{-1}(x)$ can be calculated using the Panjer inversion.

If $Y_1 \sim \Psi(\lambda\delta_0 + p)$ for some $\lambda > 0$ and $p \in M_1(\mathbb{N})$ then $Z_i \sim \Psi^S(\lambda, p_1,\ldots,p_S)$.

The possible distributions of Z_i are elements of $\Psi^S(\Delta_S)$ with

$$\Delta_S := \left\{(\lambda, p_1\ldots,p_S)^T \in \mathbb{R}^{S+1} :\ \lambda > 0,\ \sum_{l=1}^{S} p_l \le 1,\ p_l \ge 0,\ l = 1,\ldots,S\right\}.$$

Define the compactification of Δ_S, i.e.

$$K_1 := \{(\lambda, p_1 \ldots, p_S)^T \in \mathbb{R}^{S+1} :\ \lambda \in [0,\infty],\ \sum_{l=1}^{S} p_l \leq 1,\ p_l \geq 0,\ l = 1, \ldots, S\}.$$

If we define $\Psi_l^S(0,p) := \delta_{0l}$ and $\Psi_l^S(\infty,p) := \delta_{S+1}$, $l = 0, \ldots, S+1$, then Ψ is continuously extended to a mapping from K_1 to M_1^{S+2}. Hence $K := \Psi^S(K_1)$ is compact. This K is a candidate that allows approximations by cones. Note that Ψ^S as a mapping from $\mathbb{R}^{S+1}$ to $\mathbb{R}^{S+2}$ is C^∞.

Lemma 3.9 *Let $(\lambda, p)^T \in \Delta_S$.*
i) If $p_1 > 0$ then $r = \Psi^S(\lambda, p)$ fulfills $r_i > 0$, $i = 0, \ldots, S+1$.
ii) Let $e_i := (\delta_{ij})_j$ be the ith unit vector. There exists some open convex set O containing zero such that for all $B \subset O$

$$B \cap \mathcal{C}(\Delta_S - (\lambda, p)) = B \cap (\Delta_S - (\lambda, p)).$$

Furthermore,

$$\mathcal{C}(\Delta_S - (\lambda, p)) = \{x \in \mathbb{R}^S :\ v^T x \leq 0\ \forall v \in V\}$$

with

$$V = V(\lambda, p) := \begin{cases} \{(0,1,1,\ldots,1)^T\} \cup \{-e_i :\ i \in \{j :\ p_j = 0\}\}, & \text{if } \sum_{i=1}^S p_i = 1, \\ \{-e_i :\ i \in \{j :\ p_j = 0\}\}, & \text{else.} \end{cases}$$

iii) If $p_1 > 0$ and $\lambda \in (0,\infty)$ then $\Psi^S(K)$ is approximated by an cone in $\Psi^S(\lambda, p)$ with approximation tuple

$$(O_1, O_2, \Psi^S(\cdot + (\lambda, p)^T), C(\Delta_S - (\lambda, p))).$$

Proof: i) This is a consequence of lemma 2.12 and the fact that the support of a compound Poisson distribution is unbounded, hence there will be some mass at a point greater than $S+1$.
ii) Let V be the set of vectors given in the assertion. If $I = \{i : p_i = 0\}$ then

$$\Delta_S - (\lambda, p) = \left\{ (\lambda' - \lambda) e_0 + \sum_{i \in I} p_i' + \sum_{i \in I^C} (p_i' - p_i) : (\lambda', p') \in \Delta_S \right\}.$$

It is easy to check that $\Delta_S - (\lambda, p) \subset \{x \in \mathbb{R}^S :\ v^T x \leq 0\ \forall v \in V\} =: C$. Note that C is a cone too. Hence $\mathcal{C}(\Delta_S - (\lambda, p)^T) \subset C$.

Now assume $x \in C$. First assume that $\sum_i p_i < 1$. Then there is some $\gamma \geq 0$ with

$$\gamma x_i > -p_i \quad \forall i \in I^C, \qquad \gamma x_0 > -\lambda, \qquad \gamma \sum_{i=1}^{S} x_i + \sum_{i=1}^{S} p_i \leq 1.$$

Obviously, $(\lambda, p)^T + \gamma x \in \Delta_S$, hence $x \in C(\Delta_S - (\lambda, p)^T)$ because of the cone property. The case $\sum_i p_i < 1$ works analogously. Let

$$O := \begin{cases} \{x \in \mathbb{R}^{S+1} : |x_0 - \lambda| < \lambda, |x_i - p_i| < p_i, i \in I^C\} \text{ if } \sum_i p_i = 1, \\ \{x \in \mathbb{R}^{S+1} : |x_0 - \lambda| < \lambda, |x_i - p_i| < p_i, i \in I^C, |\sum_{i=1}^S x_i - 1| < 1\} \quad \text{else.} \end{cases}$$

O is an open and convex neighbourhood of zero. Furthermore, $O \cap C = O \cap (\Delta_S - (\lambda, p)^T)$. The latter is true for every subset of O.

iii) If $p_1 > 0$ and $\lambda \in (0, \infty)$ then $\Psi_k^S(\lambda, p) > 0$, $k = 0, \ldots, S$, and $\sum_{l=0}^S \Psi_k^S(\lambda, p) < 1$. Since Ψ_k^S is continuous on $\mathbb{R}^{S+2}$, we find an $\epsilon > 0$ such that $B_\epsilon(\lambda, p)$ contains $(\lambda, p)^T$ and $\Psi_k^S > 0$ and $\sum_{k=0}^S \Phi_k^S < 1$ holds on B_ϵ. Hence Ψ^S is a C^∞-mapping from B_ϵ to $\Psi^S(\lambda, p) + \{x \in \mathbb{R}^{S+2} : \sum_{l=0}^{S+2} x_i = 0\}$. The derivative of Ψ^S is given by

$$(\mathrm{id}_{S+1\times S+1}, -\mathbf{1})(T_k \Psi \delta_l)_{\substack{k=0\ldots,S \\ l=0,\ldots,S}}.$$

This is an isomorphism from $\mathbb{R}^{S+1}$ onto $\{x \in \mathbb{R}^{S+2} : \sum_{l=0}^{S+2} x_i = 0\}$, because $(T_k \Psi \delta_l)_{\substack{k=0\ldots,S \\ l=0,\ldots,S}}$ is an triangular matrix with nonzero entries on the diagonal and $(\mathrm{id}_{S+1\times S+1}, -\mathbf{1})^T$ is an isomorphism from $\mathbb{R}^{S+1}$ to $\{x \in \mathbb{R}^{S+2} : \sum_{l=0}^{S+2} x_k = 0\}$. Obviously, we can find an ϵ such that

$$B_\epsilon(0) \cap ((\Delta_S - (\lambda, p)^T) = B_\epsilon(0) \cap \mathcal{C}\left(\Delta_S - (\lambda, p)^T\right).$$

The assertion is established.

□

The results of the last chapter therefore can be applied here. For the rest of the section assume that $Z_i \sim q = \Psi^S(\lambda, p)$ with some $\lambda > 0$ and $p \in M_1^S$, $p_1 > 0$.

We use the notions of the last section, i.e. $\Phi = \Psi(\cdot + (\lambda, p))$. Obviously, $\Phi^{-1} = (\Psi^S)^{-1} - (\lambda, p)^T$, hence we have

$$(\Psi^S)^{-1}(\Theta(\bar{q}^n)) \to (\lambda, p) \qquad \text{a.s..}$$

If A is the derivative at $(\lambda, p)^T$ of the mapping $(\Psi_0^S, \ldots, \Psi_S^S)^T$ and

$$I := A^T(\mathrm{diag}(1/q_0, \ldots, 1/q_S) + 1/\tilde{q}_{S+1}\mathbf{1}\mathbf{1}^T)A$$

with $\tilde{q}_{S+1} := 1 - \sum_{l=0}^S q_l$ then

$$\sqrt{n}\big((\Psi^S)^{-1}(\Theta(\bar{q}^n)) - (\lambda, p)^T\big) \xrightarrow{\mathcal{D}} \pi_I(Z|\mathcal{C}(\Theta_S - (\lambda, p)))$$

with $Z \sim N_{S+1}(0, I^{-1})$. If we want to stress the dependence of λ and p we use the notation $I_{\lambda,p}, Z_{\lambda,p}$.

If $(\lambda, p) \in \Delta_S$ lies in the interior of Δ_S, i.e.

$$p_i > 0 \text{ for all } i = 1, \ldots, S, \qquad \sum_{i=1}^S p_i < 1,$$

it turns out that $\mathcal{C}(\Delta_S - (\lambda, p)^T) = \mathbb{R}^{S+1}$. Hence $\pi_I(Z|\mathcal{C}(\Delta_S - (\lambda, p)^T)) = Z$. Apply the delta method to see that

$$\sqrt{n}(\Theta(\bar{q}^n) - q) \xrightarrow{\mathcal{D}} Z' = (\mathrm{id}_{S+1\times+1}\mathbf{1})AZ.$$

The calculations of the last section show that $EZ'Z'^T = (q_i\delta_{ij} - q_iq_j)_{\substack{0\le i\le S+1\\ 0\le j\le S+1}}$. This is the same distributional limit as for the relative frequencies. Hence the knowledge of having measured some truncated compound Poisson variables does not have any impact on the estimation of q if (λ, p) lies in the interior of Δ_S.

The maximum likelihood estimator $\Theta(\bar{q}^n)$ is tedious to calculate, since we have to maximize a nonconvex function over a set with constraints (see [Lu89], p. 330). The next theorem shows that one projection is enough to have the same efficiency as the maximum likelihood estimator, i.e. the same distributional limit behaviour. It can be regarded as an analogon to the one-step Newton iteration in the regular parametric situation. Define $\Psi^{-1}((0, x))_k := \delta_{k0}$, $x \in M_1^{S+1}$, and

$$\Delta'_S := \left\{ (\lambda, p_1 \ldots, p_S)^T \in \mathbb{R}^{S+1} : \ \lambda \ge 0, \ \sum_{l=1}^{S} p_l \le 1, \ p_l \ge 0, l = 1, \ldots, S \right\}.$$

Note that Δ'_S is a closed convex set, hence $\pi_C(x|\Delta'_S)$ is well defined for every positive definite symmetric matrix $C \in \mathbb{R}^{S+1\times S+1}$.

Theorem 3.10 *Assume that ϕ is a mapping on M_1^{S+1} into the set of positive definite symmetric matrices in $\mathbb{R}^{S+1\times S+1}$. Suppose that $Z_i \sim q = \Psi^S(\lambda, p)$ with $(\lambda, p)^T \in \Delta_S$ and $p_1 > 0$.*

If ϕ is continuous in $\Psi^S(\lambda, p)$ and $\phi(\Psi^S(\lambda, p)) = I_{\lambda,p}$ then

$$\sqrt{n}\left(\pi_{\phi(\bar{q}^n)}((\Psi^S)^{-1}(\bar{q}_n)|\Delta'_S) - (\lambda, p)^T\right) \xrightarrow{\mathcal{D}} \pi_{I_{\lambda,p}}(Z_{\lambda,p}|\mathcal{C}(\Delta_S - (\lambda, p)^T)).$$

Proof: Without loss of generality we may assume that

$$\sqrt{n}\left((\Psi^S)^{-1}(\bar{q}^n) - (\lambda, p)^T\right) \to Z_{\lambda,p}$$

with probability one as an application of the Skohorod representation. Fix some $\omega \in \Omega$ with

$$\sqrt{n}\left((\Psi^S)^{-1}(\bar{q}^n(\omega)) - (\lambda, p)^T\right) \to Z_{\lambda,p}(\omega).$$

For shortness, write $\xi^n := (\Psi^S)^{-1}(\bar{q}^n(\omega))$, $\xi = (\lambda, p)^T$, $z = Z_{\lambda,p}(\omega)$, $I_n := \phi(\bar{q}^n(\omega))$, $I := \phi(q)$ and $C := \mathcal{C}(\Delta_S - (\lambda p)^T)$.

We want to show that

$$\pi_{I_n}(\xi^n - \xi|C) + \xi = \pi_{I_n}(\xi^n|\Delta'_S)$$

for n large enough. First note that $\pi_{I_n}(\xi^n - \xi|C) \to 0$ for $n \to \infty$, indeed: Since $0 \in C$ is a worse approximation we have

$$\|\xi^n - \xi - \pi_{I_n}(\xi^n - \xi|C)\|_{I_n} \le \|\xi^n - \xi - 0\|_{I_n}.$$

Hence for $n \to \infty$

$$\|\pi_{I_n}(\xi^n - \xi|C)\|_{I_n} \le \|\xi^n - \xi - \pi_{I_n}(\xi^n - \xi|C)\|_{I_n} + \|\xi^n - \xi\|_{I_n} \le 2\|\xi^n - \xi\|_{I_n} \to 0.$$

Since $I_n \to I$, we have $\|\pi_{I_n}(\xi^n - \xi|C)\|_2 \to 0$. Hence for n large enough

$$\pi_{I_n}(\xi^n - \xi|C) \in (\Delta_S - \xi) \cap B_\epsilon \subset \Delta'_S - \xi.$$

Note that

$$\max_{\theta \in \Delta'_S - \xi} \|\xi^n - \xi - \theta\|_{I_n} \le \max_{\theta \in C} \|\xi^n - \xi - \theta\|_{I_n},$$

Therefore

$$\begin{aligned}\pi_{I_n}(\xi^n - \xi|C) &= \mathrm{argmax}_{\theta \in \Delta'_S - \xi}\|\xi^n - \xi - \theta\| \\ &= \mathrm{argmax}_{\theta \in \Delta'_S}\|\xi^n - \theta\| - \xi = \pi_{I_n}(\xi^n|\Delta'_S) - \xi.\end{aligned}$$

It is easy to show that $\alpha\pi_{I_n}(x|C) = \pi_{I_n}(\alpha x|C)$ for all $\alpha \ge 0$ (use (NSC)). Hence

$$\sqrt{n}(\pi_{I_n}(\xi^n|\Delta'_S) - \xi) = \pi_{I_n}(\sqrt{n}(\xi^n - \xi)|C).$$

We have to show that

$$\pi_{I_n}(\sqrt{n}(\xi^n - \xi)|C) \to \pi(z|C).$$

First

$$\|\pi_{I_n}(\sqrt{n}(\xi^n - \xi)|C)\|^2_{I_n} \le \|\sqrt{n}(\xi^n - \xi)\|^2_{I_n} \le (z^t I z + 1).$$

for n large enough. Again the limit result $I_n \to I$ forces $\sqrt{n}\pi_{I_n}(\xi^n - \xi|C)$ to be bounded. Note that if $\lim_{n\to\infty} \pi_{I_n}(\sqrt{n}(\xi^n - \xi)|C) =: L$ exists then we have $L \in C$, since C is closed. Hence for every $c \in C$ we have

$$\left(\sqrt{n}(\xi^n - \xi) - \pi_{I_n}(\sqrt{n}(\xi^n - \xi)|C)\right)^T I_n \left(c - \pi_{I_n}(\sqrt{n}(\xi^n - \xi)|C)\right) \le 0$$

because of (NSC). For $n \to \infty$ we obtain

$$(z - L)^T I(c - L) \le 0.$$

(NCS) impies that $L = \pi_I(z|C)$. An argument that uses boundedness and subsequences yields the assertion (as in the proof of the last chapter's theorem).□

This theorem motivates an estimation procedure. Calculate λ, p from $\bar{q}^n$ using the Panjer inversion and project onto the parameter set using some positive definite matrix that estimates the Fisher information. The projection can carried out using the active set method in the appendix (see [Lu89], p. 423, a Maple routine is given in the appendix).

We illustrate this method using the data given by Bortkiewicz. They describe the number of soldiers that died by horse kicks in the Prussian army and have been observed from ten corps over twenty years (see [Qu87], [Cs89] for further references). The data and the relative frequencies $\bar{q}$ resulting from them are collected in the next table

no. deaths	0	1	2	3	4	≥ 5
abs. freq.	109	65	22	3	1	0
rel. freq. $\bar{q}$	0.545	0.325	0.110	0.015	0.005	0

In spite of the fact that the interpretation within the compound Poisson model may appear to be a little bit odd, they are a(n) „(in)famous “example ([Cs89]) for the law of small numbers. We therefore expect our projection estimators to give a point at the boundary of Δ'_S. We discuss two examples. Fix $S = 3$. Compute the Panjer inversion $\hat{\lambda}, \hat{p}_1, \ldots, \hat{p}_3$. Then approximate the matrix A via a matrix $\tilde{A}$. For that use a truncated version of the matrix given in section 2.3. Just plug $\bar{q}$, $\hat{\lambda}$ and $\hat{p}$ into $(T_k \Psi' \delta_l)_{\substack{0 \leq k \leq 3 \\ 0 \leq l \leq 3}}$. Then estimate

$$I \approx \tilde{I}_3 := \tilde{A}^T \left(\operatorname{diag}\left(\frac{1}{\bar{q}_0}, \ldots, \frac{1}{\bar{q}_3} \right) + \frac{1}{1 - \sum_{l=0}^{3} \bar{q}_l} \mathbf{1}\mathbf{1}^T \right) \tilde{A}.$$

We have calculated

$$(\tilde{\lambda}, \tilde{p}_3)^T = \pi_{\tilde{I}}((\hat{\lambda}, \hat{p}) | \Delta'_3), \qquad (\bar{\lambda}, \bar{p}_3) = \pi_{\mathrm{id}}((\hat{\lambda}, \hat{p}) | \Delta'_3).$$

The naive estimators $(\hat{\lambda}, \hat{p}_m)$, calculated for $m = 3, 4, 10$, have the components

$$\hat{p}_{m,i} = \frac{\max(0, \hat{p}_i)}{\sum_{l=1}^{m} \max(0, \hat{p}_l)}, \quad \text{for } i = 1, \ldots, m, \quad p_{m,i} = 0, \quad \text{for } i > m.$$

The corresponding λ is $\hat{\lambda}$. Note that $\hat{p}_4$ is the consistent naive estimator with end point driven by the maximum of the data (section 2.5).

How to get rid of gaps? Consider the choice $S = 10$. We then have to use another approximation to I, since the relative frequencies with respect to the number of deaths higher then 5 are zero. We therefore do not use the relative frequencies themselves. First compute the Panjer inversion up to 10, i.e. $\hat{\lambda}, \hat{p}_1, \ldots, \hat{p}_{10}$, then compute the naive estimator $\hat{p}_{10}$. We plug the naive estimator into the compound Poisson functional, i.e. perform the Panjer recursion formula for the naive estimator $(\hat{\lambda}, \hat{p}_{10})$ up to 10. This gives us a new sequence $\tilde{q}_0, \ldots, \tilde{q}_{10}$. In contrast to the relative frequencies this sequence has no gaps. We estimate $\tilde{A} \approx A$ again, but now plugging in $\tilde{q}$, $\hat{\lambda}$ and $\tilde{p}_{10}$ into the truncated matrix $(T_k \Psi' \delta_l)_{\substack{0 \leq k \leq 10 \\ 0 \leq l \leq 10}}$. Then estimate I via

$$\tilde{I}_{10} := \tilde{A}^T \left(\operatorname{diag}\left(\frac{1}{\tilde{q}_0}, \ldots, \frac{1}{\tilde{q}_{10}} \right) + \frac{1}{1 - \sum_{l=0}^{10} \tilde{q}_l} \mathbf{1}\mathbf{1}^T \right) \tilde{A}.$$

Again we have computed

$$(\tilde\lambda_{10}, \tilde p_{10}) = \pi_{\tilde I_{10}}((\hat\lambda, \hat p)|\Delta'_{10}), \qquad (\bar\lambda_{10}, \bar p_{10}) = \pi_{\text{id}}((\hat\lambda, \hat p)|\Delta'_{10}).$$

The next table summarizes the computational results. The first row includes the Poisson approximation $\mathcal{P}$

$\mathcal{P}$	$(\hat\lambda, \hat p)$	$(\hat\lambda, \hat p_{10})$	$\pi_{\tilde I_{10}}$	π_{id}	$(\hat\lambda, \hat p_4)$
0.61	0.6070	0.6070	0.6064	0.6070	0.6070
1	0.9825	0.9406	0.9977	0.9682	0.9422
0	0.0396	0.0379	0.0000	0.0253	0.0380
0	-0.0365	0	0.000	0.0000	0
0	0.0207	0.0198	0.0023	0.0064	0.0198
0	-0.0077	0	0.0000	0.0000	0
0	0.0014	0.0014	0.0000	0.0000	
0	0.0002	0.0002	0.0000	0.0000	
0	-0.0003	0	0.0000	0.0000	
0	0.0001	0.0001	0.0000	0.0000	
0	0.0000	0.0000	0.0000	0.0000	

$\hat\lambda$	0.6070
$\hat p$	0.9825
	0.0396
	-0.0365

$\hat\lambda$	0.6070
$\hat p_3$	0.9613
	0.0387
	0

$\tilde\lambda_3$	0.5999
$\tilde p_3$	0.9971
	0.0000
	0.0000

$\bar\lambda_3$	0.6070
$\bar p_3$	0.9714
	0.0286
	0

3.5 Likelihood Ratio Tests

In this chapter we apply our results to the analysis of a class of likelihood ratio tests that can be used for testing the hypothesis that $Y = (Y_1 \ldots, Y_n)$ is a vector of Poisson distributed variables within the general assumption that Y is a vector compound Poisson distributed entries. This is done considering truncated data. Fix some S. As in the last section, let (Z_n) be an iid-sequence of $\{0, \ldots, S+1\}$-valued random variables with $Z_i \sim q$. Let

$$\begin{aligned} Z &\sim N_{S+2}(0, (q_i\delta_{ij} - q_i q_j)_{\substack{0\le i\le S+1\\ 0\le j\le S+1}}), \\ D &:= \operatorname{diag}(1/q_0, \ldots 1/q_S) \\ C_q &:= (\Psi^S)'_{\lambda,p}\mathcal{C}(\Delta_S - (\lambda, p)), \end{aligned}$$

if $q = \Psi^S(\lambda, p)$. Denote $\mathcal{P}^S = \{\Psi^S(\lambda, e_1) : \ \lambda > 0\} \subset \mathbb{R}^{S+2}$. This is the set of truncated Poisson distributions.

We want to use the results of the last section for the performance of the likelihood ratio tests in the following decision problem:

$$H : \ q \in \mathcal{P}^S \quad \text{versus} \quad K : \ q \in \Psi^S(\Delta_S)\backslash\mathcal{P}^S.$$

Define

$$\begin{aligned} q_H^n &= \operatorname{argmax}_{q\in\Psi^S([0,\infty]\times\{(1,0,\dots,0)\})} L(q|\bar{q}^n), \\ q_K^n &= \operatorname{argmax}_{q\in\Psi(K_1)} L(q|\bar{q}^n), \end{aligned}$$

Remarks: i) Define $K := \Psi^S([0,\infty] \times \{(1,0,\dots,0)\})$. If $\lambda > 0$, then K is approximated by a cone in $\Psi^S(\lambda,1,0,\dots,0)$ with representation tuple

$$(O_1, O_2, \Psi^S(\cdot + (\lambda,1,0,\dots,0)^T), \operatorname{lin}\{e_0\})$$

with $O_1 := \{x \in \mathbb{R}^{S+1}, |x_0| < \lambda\}$ and $O_2 := \Psi^S(O_1)$. The approximating cone is then a linear space. Hence, if $q = \Psi^S(\lambda,1,0,\dots,0)$ then

$$\sqrt{n}(q_H^n - q) \xrightarrow{\mathcal{D}} \pi_D(Z|(\Psi^S)'_{(\lambda,1,0,\dots,0)}\operatorname{lin}\{e_0\}).$$

Write $V := (\Psi^S)'_{(\lambda,1,0,\dots,0)}\operatorname{lin}\{e_0\})$.

ii) Going through the proof of theorem 3.7 again we see that everything is based on purely analytical considerations regarding the relative frequencies. The consequence is that we have joint distributional limit laws, i.e. if $q = \Psi^S(\lambda,1,0,\dots,0)$ then

$$\sqrt{n}\left(\begin{pmatrix} \bar{q}^n \\ q_K^n \\ q_H^n \end{pmatrix} - \begin{pmatrix} q \\ q \\ q \end{pmatrix}\right) \xrightarrow{\mathcal{D}} \begin{pmatrix} Z \\ \pi_D(Z|C_q) \\ \pi_D(Z|V) \end{pmatrix}.$$

If $z_1,\dots,z_n$ are the realisations of Z_i, $i = 1,\dots,n$ then the likelihood ratio test for H versus K rejects H for large values of the statistic

$$\frac{\max_{(\lambda,p)^T\in K_1}\prod_{i=1}^n \Psi_{z_i}^S(\lambda,p)}{\max_{\lambda\in[0,\infty]}\prod_{i=1}^n \Psi_{z_i}^S(\lambda,1,0,\dots,0)}.$$

Note that the usual suprema over the nonclosed sets of hypothesis and alternative are the same as the maxima over the closures regarding H and K because of the continuity of Ψ^S. Furthermore, applying a monotone function to the statistic does not change the test, hence the same test is performed if we reject H for large values of

$$T_1^n := n\log\frac{\max_{(\lambda,p)^T\in K_1}\prod_{i=1}^n \Psi_{z_i}^S(\lambda,p)}{\max_{\lambda\in[0,\infty]}\prod_{i=1}^n \Psi_{z_i}^S(\lambda,1,0,\dots,0)} = -n\big(L(q_H^n|\bar{q}^n) - L(q_K^n|\bar{q}^n)\big).$$

The further analysis is similar to [Ro88] (see p.61). Assume that we are on the hypothesis, i.e. $q = \Psi^S(\lambda,1,0,\dots,0)$. A Taylor expansion of $L(\cdot|\bar{q}^n)$ about $\bar{q}^n$ yields the representation

$$\begin{aligned} T_1^n &= -n(L(q_H^n|\bar{q}^n) - L(\bar{q}^n|\bar{q}^n) + L(\bar{q}^n|\bar{q}^n) - L(q_K^n|\bar{q}^n)) \\ &= -n((q_H^n - \bar{q}^n)^T C_n^1 (q_H^n - \bar{q}^n) - (q_K^n - \bar{q}^n)^T C_n^2 (q_K^n - \bar{q}^n)), \end{aligned}$$

which is true for some matrices C_n^1, C_n^2 at least if $\bar{q}_k^n > 0$, indeed: since $\bar{q}^n$ maximizes $L(\cdot|\bar{q}^n)$, the directional derivative vanishes in every direction $q - \bar{q}^n$, $q \in M_1^{S+2}$, provided that $\bar{q}^n + \epsilon(q - \bar{q}^n) \in M_1$ for all ϵ with $|\epsilon|$ small enough. In this case $\nabla L(\cdot|\bar{q}^n)_{\bar{q}^n}(q - \bar{q}^n) = 0$. Since we are on the hypothesis, this is true for $\bar{q}_k^n$ for n large enough with probability one. The matrices are given by $C_n^i = D^2 L(\cdot|\bar{q}^n)_{\theta_n^i}$ for θ_n^1 in the convex hull spanned by $\bar{q}^n$ and q_H^n and θ_n^2 in the convex hull spanned by $\bar{q}^n$ and q_K^n.

Since we are on the hypothesis, we have $q_H^n \to q$, $q_K^n \to q$ and $\bar{q}^n \to q$ a.s.. The log likelihood function is smooth enough near q to provide $C_n^i \to -D$.

Because of remark ii) we have

$$T_1^n \xrightarrow{\mathcal{D}} \|\pi_D(Z|V) - Z\|_D^2 - \|\pi_D(Z|C_q) - Z\|_D^2.$$

Since $V = -V \subset C_q$ we have for all v in V by (NSC)

$$\langle v - \pi_D(Z|C_q), Z - \pi_D(Z|C_q)\rangle_D \leq 0,$$

replacing $-v$ in the inequality gives us the equality

$$\langle v - \pi_D(Z|C_q), Z - \pi_D(Z|C_q)\rangle_D = 0$$

for all $v \in V$. By Phythagoras' law the following identity holds

$$\|Z - \pi_D(Z|C_q)\|_D^2 + \|\pi_D(Z|V) - \pi_D(Z|C_q)\|_D^2 = \|Z - \pi_D(Z|V))\|_D^2.$$

Hence

$$T_n^1 \xrightarrow{\mathcal{D}} \|\pi_D(Z|V) - \pi_D(Z|C_q)\|_D^2.$$

This is quite similar to the expression in theorem 3.3. We make the usual linear transformation from $\{y \in \mathbb{R}^{S+2} : \sum_{i=0}^{S+1} y_i = 0\}$ to $\mathbb{R}^{S+1}$ via the isomorphism given by the matrix $T := (\mathrm{id}_{S+1\times S+1}, 0)$ (i.e. the projection onto the first $S+1$ coordinates). Note again that the inverse mapping is given by $T^{-1} = (\mathrm{id}_{S+1\times S+1}, -\mathbf{1})^T$. Lemma 3.2 then yields

$$\begin{aligned}\|\pi_D(Z|V) - \pi_D(Z|C_q)\|_D^2 &= \|T^{-1}\pi_{T^{-T}DT^{-1}}(TZ|TV) - T^{-1}\pi_{T^{-T}DT^{-1}}(TZ|TC_q)\|_D^2 \\ &= \|\pi_{\tilde{D}}(\tilde{Z}|\tilde{V}) - \pi_{\tilde{D}}(\tilde{Z}|\tilde{C})\|_{\tilde{D}}^2,\end{aligned}$$

with A denoting the derivative of the first $S+1$ coordinates and $\tilde{D} = T^{-T}DT^{-1} = \mathrm{diag}(1/q_0, \ldots, 1/q_S) + 1/\tilde{q}_{S+1}\mathbf{1}\mathbf{1}^T$. Furthermore, we have used the abbreviations $\tilde{V} = A\mathrm{lin}\{e_0\}$ and $\tilde{C} := A\mathcal{C}(\Delta_S - (\lambda, 1, 0, \ldots, 0)^T))$. We have $\tilde{Z} \sim N_{S+1}(0, (q_i\delta_{ij} - q_iq_j)_{\substack{0\leq i\leq S\\ 0\leq j\leq S}}) = N_{S+1}(0, \tilde{D}^{-1})$.

We have a closer look at the underlying cone. An application of lemma 3.9 yields

$$\mathcal{C}(\Delta_S - (\lambda, 1, 0\ldots, 0)^T) = \{x \in \mathbb{R}^{S+1} : N^T x \leq 0\}$$

with

$$N = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 1 & \\ \vdots & \mathrm{diag}(-1,\ldots,-1)_{S\times S} \\ 1 & \end{pmatrix} \in \mathbb{R}^{(S+1)\times S}.$$

($x \leq 0$ is understood componentwise). Hence

$$\begin{aligned} \tilde{C} = \{y \in \mathbb{R}^{S+1} :\ & N^T A^{-1} y \leq 0\} \\ = \ & \{y \in \mathbb{R}^{S+1} : \ N^T A^{-1} \tilde{D}^{-1} \tilde{D} y \leq 0\} \\ = \ & \{y \in \mathbb{R}^{S+1} : \ (\underbrace{\tilde{D}^{-T}}_{=\tilde{D}^{-1}} A^{-T} N)^T \tilde{D} y \leq 0\}. \end{aligned}$$

If we denote the columns of $\mathcal{N} := \tilde{D}^{-1} A^{-T} N$ by $\nu_1, \ldots, \nu_S$ then we have found that

$$\tilde{C} = \mathcal{C}(\nu_1, \ldots, \nu_S)^{*\tilde{D}},$$

i.e. $\tilde{C}$ is the dual cone of a finitely generated one. Hence we are in the situation of theorem 3.3. Therefore

$$\|\pi_D(Z|V) - \pi_D(Z|C_q)\|_D^2 = \|\pi_{\tilde{D}}(\tilde{Z}|\tilde{V}) - \pi_{\tilde{D}}(\tilde{Z}|\tilde{C})\|_{\tilde{D}}^2 \sim \sum_{i=0}^{S} \kappa_i \chi^2_{S-i}.$$

The κ_i are given by

$$\kappa_i = \sum_{\#I=i} P\left(\tilde{Z} \in \mathcal{C}(\nu_i | i \in I) \oplus \left(\mathrm{lin}\{\nu_i : \ i \in I\}^{\perp_{\tilde{D}}} \cap \{x \in \mathbb{R}^{S+1} : \ \langle \nu_i, x\rangle_{\tilde{D}} < 0; i \in I^C\}\right)\right).$$

Obviously, we can replace the $<$ by $\leq$ in the formula above, since the covariance of $\tilde{Z}$ is nonsingular.

We want to make more explicit calculations for the cone probabilities which will be shown to coincide with orthant probabilities of some appropriate centered Gaussian random variable.

We need a matrix representation for the underlying orthogonal projections. Assume that W is spanned by some linearly independent vectors $w_1, \ldots, w_m$. Write M_W for the matrix with columns $w_1, \ldots, w_m$. Then $W = M_W \mathbb{R}^m$. Then

$$\begin{aligned} \pi_{\tilde{D}}(z|W) &= M_W (M_W^T \tilde{D} M_W)^{-1} M_W^T \tilde{D} z, \\ \pi_{\tilde{D}}(z|W^{\perp_{\tilde{D}}}) &= z - M_W (M_W^T \tilde{D} M_W)^{-1} M_W^T \tilde{D} z. \end{aligned}$$

This is a well known fact from linear regression. Note that the coefficients of the representations of $\pi_{\tilde{D}}(z|W) = \sum_{l=1}^{m} \alpha_i w_i$ are given by

$$(\alpha_1, \ldots, \alpha_m)^T = (M_W^T \tilde{D} M_W)^{-1} M_W^T \tilde{D} z.$$

Fix some $I \subset \{1, \ldots, S\}$ with $I \neq \emptyset$ and $I^C \neq \emptyset$. Let E_I be the $\mathbb{R}^{S \times \#I}$-matrix with columns e_i, $i \in I$, where again $e_i \in \mathbb{R}^S$ denotes the i-th unit vector. Then $\mathrm{lin}\{\nu_i : i \in I\} = \mathcal{N} E_I \mathbb{R}^{\#I}$. Therefore

$$\tilde{Z} \in \mathcal{C}(\nu_i | i \in I) \oplus \Big(\mathrm{lin}\{\nu_i : \ i \in I\}^{\perp_{\tilde{D}}} \cap \mathcal{C}^{*\tilde{D}}(\nu_i | i \in I^C\} \Big)$$

$$\begin{aligned}
\Leftrightarrow \quad & \pi_{\tilde{D}}(\tilde{Z} | \mathrm{lin}\{\nu_i, i \in I\}) \in \mathcal{C}(\nu_i | i \in I)) \\
\wedge \quad & \pi_{\tilde{D}}(\tilde{Z} | \mathrm{lin}\{\nu_i, i \in I\}^{\perp_{\tilde{D}}}) \in \mathcal{C}^{*\tilde{D}}(\nu_i | i \in I^C)
\end{aligned}$$

$$\begin{aligned}
\Leftrightarrow \quad & ((\mathcal{N} E_I)^T \tilde{D} \mathcal{N} E_I)^{-1} (\mathcal{N} E_I)^T \tilde{D} \tilde{Z} \geq 0 \\
\wedge \quad & (\mathcal{N} E_{I^C})^T \tilde{D} (\tilde{Z} - \mathcal{N} E_I ((\mathcal{N} E_I)^T \tilde{D} \mathcal{N} E_I)^{-1} (\mathcal{N} E_I)^T \tilde{D} \tilde{Z}) \leq 0
\end{aligned}$$

$$\begin{aligned}
\Leftrightarrow \quad & (E_I^T N^T \tilde{D}^{-1} A^{-T} N E_I)^{-1} E_I^T N^T A^{-1} \tilde{Z}) \geq 0 \\
\wedge \quad & E_{I^C}^T N^T A^{-1} (\tilde{Z} - \tilde{D}^{-1} A^{-T} N E_I (E_I^T N^T A^{-1} \tilde{D}^{-1} A^{-T} N E_I)^{-1} E_I^T N^T A^{-1} \tilde{Z}) \leq 0
\end{aligned}$$

$$\Leftrightarrow \quad (E_I^T \bar{Q} E_I)^{-1} E_I^T \bar{Q} \bar{Z} \geq 0 \quad \wedge \quad E_{I^C}^T \bar{Q} (\bar{Z} - E_I (E_I^T \bar{Q} E_I)^{-1} E_I^T \bar{Q} \bar{Z}) \leq 0$$

with $\bar{Q} = N^T A^{-1} \tilde{D}^{-1} A^{-T} N$ and $\bar{Z} = \bar{Q}^{-1} N^T A^{-1} \tilde{Z}$. Note that

$$\mathrm{Cov}(\bar{Z}, \bar{Z}) = \bar{Q}^{-1} N^T A^{-1} \mathrm{Cov}(\tilde{Z}, \tilde{Z}) A^T N \bar{Q}^{-1} = \bar{Q}^{-1}.$$

For $I = \emptyset$ we have

$$\tilde{Z} \in \mathcal{C}^{*\tilde{D}}(\nu_i | i = 1, \ldots, S) \qquad \Leftrightarrow \qquad \bar{Q} \bar{Z} \leq 0$$

and for $I = \{1, \ldots, S\}$

$$\tilde{Z} \in \mathcal{C}(\nu_i | i \in I) \oplus \mathrm{lin}\{\nu_i | i \in I\}^{\perp_{\tilde{D}}}) \qquad \Leftrightarrow \qquad \bar{Z} \geq 0.$$

A closer look at the formulas shows that we have reduced the dimension of the problem of finding the semicone probabilities. They are the same as in the corresponding projection $\pi_{\bar{Q}}(\bar{Z} | \mathcal{C}(e_i, i = 1, \ldots, S)^{*\bar{Q}})$. We should note that the following pairs of Gaussian random variables are independent:

$$\begin{array}{lrcl}
\text{First pair :} & \pi_{\bar{Q}}(\bar{Z} | \mathrm{lin}\{e_i, i \in I\}) & = & E_I (E_I^T \bar{Q} E_I)^{-1} E_I^T \bar{Q} \bar{Z}, \\
& \pi_{\bar{Q}}(\bar{Z} | \mathrm{lin}\{e_i, i \in I\}^{\perp_{\bar{Q}}}) & = & \bar{Z} - E_I (E_I^T Q E_I)^{-1} E_I^T Q \bar{Z}, \\
\text{Second pair :} & \multicolumn{3}{l}{(E_I^T \bar{Q} E_I)^{-1} E_I^T \bar{Q} \bar{Z}, \quad E_{I^C}^T \bar{Q} (\bar{Z} - E_I (E_I^T Q E_I)^{-1} E_I^T Q \bar{Z}).}
\end{array}$$

This is easy to see from the covariance structure of the first pair.

For S small ($S \leq 3$) we can give explicit formulas for the cone probabilities using some geometric reasoning.

We have a look at three examples:

Example 3.11 If $S = 1$ then we have $\mathcal{C}(e_1) = [0, \infty)$ and $\mathcal{C}(e_1)^{*\bar{Q}} = (-\infty, 0]$, of course. Therefore $\kappa_1 = \kappa_0 = 1/2$.

Example 3.12 If $S = 2$ we have three mixing coefficients with the following corresponding cones

$$\begin{aligned} \kappa_2 : &\ \mathcal{C}(e_1, e_2), \\ \kappa_1 : &\ \mathcal{C}(e_1) \oplus \left(\{e_1\}^{\perp_{\bar{Q}}} \cap \mathcal{C}^{*\bar{Q}}(e_2)\right), \ \ \mathcal{C}(e_2) \oplus \left(\{e_2\}^{\perp_{\bar{Q}}} \cap \mathcal{C}^{*\bar{Q}}(e_1)\right), \\ \kappa_0 : &\ \mathcal{C}^{*\bar{Q}}(e_1, e_2) \end{aligned}$$

First look at κ_1. Here, the derivation of the cone probabilities is reduced again to one dimensional problems using the independence of the underlying Gaussian random variables:

$$\begin{aligned} &P\left(\bar{Z} \in \mathcal{C}(e_1) \oplus \left(\{e_1\}^{\perp_{\bar{Q}}} \cap \mathcal{C}^{*\bar{Q}}(e_2)\right)\right) \\ &\quad = P(\pi_{\bar{Q}}(\bar{Z}|\mathrm{lin}\{e_i\}) \in \mathcal{C}(e_1))) \times P(e_2^T \bar{Q} \pi_{\bar{Q}}(\bar{Z}|\mathrm{lin}\{e_i\}^{\perp_{\bar{Q}}}) \le 0) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}. \end{aligned}$$

Analogously, $P(\bar{Z} \in \mathcal{C}(e_1) \oplus \left(\{e_1\}^{\perp_{\bar{Q}}} \cap \{x \in \mathbb{R}^2 :\ e_2^T \bar{Q} x \le 0\}\right) = \frac{1}{4}$.
Hence $\kappa_1 = \frac{1}{2}$.

It is enough to derive the probability $P(\bar{Z} \ge 0) = P(\bar{Z} \in \mathcal{C}(e_1, e_2))$, since $P(e_i^T \bar{Q} \bar{Z} \le 0,\ i = 1, 2) = 1/2 - P(\bar{Z} \ge 0)$ holds.

If $LL^T = \bar{Q}$ is the Cholesky decomposition then we have $L^T \bar{Z} \sim N(0, \mathrm{id})$, i.e. the components of $L^T \bar{Z}$ are independent and $N(0,1)$-distributed. Then $P(\bar{Z} \ge 0) = P(L^T \bar{Z} \in \mathcal{C}(L^T e_1, L^T e_2))$. The latter probability is given by the angle between the vectors $L^T e_1$ and $L^T e_2$ if the angle is normed by dividing by 2π, i.e.

$$\begin{aligned} \kappa_2 = P(\bar{Z} \ge 0) &= P(L^T \bar{Z} \in \mathcal{C}(L^T e_1, L^T e_2)) \\ &= \frac{1}{2\pi} \arccos\left(\frac{(L^T e_1)^T L^T e_2}{\sqrt{(L^T e_1)^T L^T e_1}\sqrt{(L^T e_2)^T L^T e_2}}\right) \\ &= \frac{1}{2\pi} \arccos\left(\frac{e_1^T \bar{Q} e_2}{\sqrt{e_1^T \bar{Q} e_1}\sqrt{e_2^T \bar{Q} e_2}}\right) \\ &= \frac{1}{2\pi} \arccos\left(\frac{\langle e_1, e_2 \rangle_{\bar{Q}}}{\|e_1\|_{\bar{Q}} \|e_2\|_{\bar{Q}}}\right). \end{aligned}$$

Hence the cone probability turns out to be the angle in the corresponding inner product $\langle \cdot, \cdot \rangle_{\bar{Q}}$. Recall that

$$N = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 1 & -1 \end{pmatrix}, \quad A^{-1} = \begin{pmatrix} -e^{\lambda} & 0 & 0 \\ \frac{e^{\lambda}}{\lambda}(1 - \lambda) & \frac{e^{\lambda}}{\lambda} & 0 \\ \frac{1}{2} e^{\lambda} \lambda & -e^{\lambda} & \frac{e^{\lambda}}{\lambda} \end{pmatrix}$$

and

$$\tilde{D}^{-1} = \begin{pmatrix} e^{-\lambda} & 0 & 0 \\ 0 & e^{-\lambda}\lambda & 0 \\ 0 & 0 & \frac{1}{2}\,e^{-\lambda}\lambda^2 \end{pmatrix} - \begin{pmatrix} e^{-2\,\lambda} & e^{-2\,\lambda}\lambda & \frac{1}{2}\,e^{-2\,\lambda}\lambda^2 \\ e^{-2\,\lambda}\lambda & e^{-2\,\lambda}\lambda^2 & \frac{1}{2}\,e^{-2\,\lambda}\lambda^3 \\ \frac{1}{2}\,e^{-2\,\lambda}\lambda^2 & \frac{1}{2}\,e^{-2\,\lambda}\lambda^3 & \frac{1}{4}\,e^{-2\,\lambda}\lambda^4 \end{pmatrix}.$$

The next calculations have been made with Maple Version V/Release 5 using the `linalg` package.

$\bar{Q}$ has the following entries

$$\begin{aligned} \bar{Q}_{1,1} &= \frac{1}{4}\,\frac{4\,e^{\lambda} - 4\,e^{\lambda}\lambda - 4 + 2\,\lambda^2 e^{\lambda} + \lambda^4 e^{\lambda}}{\lambda^2} \\ \bar{Q}_{1,2} = \bar{Q}_{2,1} &= -\frac{1}{4}\,\lambda\,e^{\lambda}\,(2+\lambda) \\ \bar{Q}_{2,2} &= \frac{1}{4}\,\lambda^2 e^{\lambda} + e^{\lambda}\lambda + \frac{1}{2}\,e^{\lambda}. \end{aligned}$$

If

$$\begin{aligned} z(\lambda) &:= -\frac{1}{4}\,\lambda\,e^{\lambda}\,(2+\lambda) \\ n^2(\lambda) &:= \frac{1}{16}\,e^{2\lambda}\lambda^4 + \frac{1}{4}\,e^{2\lambda}\lambda^3 + \frac{1}{4}\,e^{2\lambda}\lambda^2 \\ &\quad + \frac{1}{4}\,e^{2\lambda}\lambda - \frac{1}{2}\,e^{2\lambda} + \frac{1}{2}\,\frac{e^{2\lambda}}{\lambda} \\ &\quad - \frac{1}{4}\,e^{\lambda} + \frac{1}{2}\,\frac{e^{2\lambda}}{\lambda^2} - \frac{e^{\lambda}}{\lambda} - \frac{1}{2}\,\frac{e^{\lambda}}{\lambda^2} \end{aligned}$$

then $\kappa_2 = \frac{1}{2\pi}\arccos(z(\lambda)/n(\lambda))$. Hence the mixing coefficients depend on the unknown parameter λ. We therefore need a studentization procedure as explained below.

Example 3.13 $S = 3$. The κ_i correspond to the following cones

$$\begin{aligned} &\kappa_3: \ \mathcal{C}(e_1, e_2, e_3), \\ &\kappa_2: \ \mathcal{C}(e_k, e_l) \oplus \mathrm{lin}\{e_k, e_l\}^{\perp_{\bar{Q}}} \cap \mathcal{C}^{*\bar{Q}}(e_m), \quad \{k,l,m\} = \{1,2,3\} \\ &\kappa_1: \ \mathcal{C}(e_k) \oplus \left(\mathrm{lin}\{e_k\}^{\perp_{\bar{Q}}} \cap \mathcal{C}^{*\bar{Q}}(e_l, e_m)\right), \quad \{k,l,m\} = \{1,2,3\} \\ &\kappa_0: \ \mathcal{C}^{*\bar{Q}}(e_1, e_2, e_3) \qquad . \end{aligned}$$

In addition to the already derived formulas above we need $P(\pi_{\bar{Q}}(\bar{Z}|\mathrm{lin}\{e_1\}^{\perp_{\bar{Q}}}) \in \mathcal{C}^{*\bar{Q}}(e_2, e_3))$. We have

$$\pi_{\bar{Q}}(\bar{Z}|\mathrm{lin}\{e_1\}^{\perp_{\bar{Q}}}) \in \mathcal{C}^{*\bar{Q}}(e_2, e_3) \Leftrightarrow Y := E^T_{\{2,3\}}\bar{Q}\left(e_1(e_1^T\bar{Q}e_1)^{-1}e_1^T\bar{Q}\bar{Z} - \bar{Z}\right) \geq 0.$$

It is easy to calculate that the covariance matrix of Y is given by

$$EYY^T = E^T_{\{2,3\}}\bar{Q}E_{\{2,3\}} - \frac{1}{\|e_1\|^2_{\bar{Q}}} E^T_{\{2,3\}}\bar{Q}e_1e_1^T\bar{Q}E_{\{2,3\}}.$$

We are in the two-dimensional case again. To derive the probability we have to calculate the quotient in the formula of example 1. It is given by

$$\begin{aligned} & \frac{e_2^T\bar{Q}e_3 - e_2^T\bar{Q}e_1\|e_1\|^{-2}_{\bar{Q}}e_1^T\bar{Q}e_3}{\sqrt{e_2^T\bar{Q}e_2 - e_2^T\bar{Q}e_1\|e_1\|^{-2}_{\bar{Q}}e_1^T\bar{Q}e_2}\sqrt{e_3^T\bar{Q}e_3 - e_3^T\bar{Q}e_1\|e_1\|^{-2}_{\bar{Q}}e_1^T\bar{Q}e_3}} \\ = & \frac{\langle e_2, e_3\rangle_{\bar{Q}}\|e_1\|^2_{\bar{Q}} - \langle e_1, e_2\rangle_{\bar{Q}}\langle e_1, e_3\rangle_{\bar{Q}}}{\sqrt{\|e_1\|^2_{\bar{Q}}\|e_2\|^2_{\bar{Q}} - \langle e_1, e_2\rangle^2_{\bar{Q}}}\sqrt{\|e_1\|^2_{\bar{Q}}\|e_3\|^2_{\bar{Q}} - \langle e_1, e_3\rangle^2_{\bar{Q}}}}. \end{aligned}$$

Hence

$$\begin{aligned} & P(\pi_{\bar{Q}}(\bar{Z}|\mathrm{line}_1{}^{\perp_{\bar{Q}}}) \in \mathcal{C}^{*\bar{Q}}(e_2, e_3)) \\ & = \frac{1}{2\pi}\arccos \frac{\langle e_2, e_3\rangle_{\bar{Q}}\|e_1\|^2_{\bar{Q}} - \langle e_1, e_2\rangle_{\bar{Q}}\langle e_1, e_3\rangle_{\bar{Q}}}{\sqrt{\|e_1\|^2_{\bar{Q}}\|e_2\|^2_{\bar{Q}} - \langle e_1, e_2\rangle^2_{\bar{Q}}}\sqrt{\|e_1\|^2_{\bar{Q}}\|e_3\|^2_{\bar{Q}} - \langle e_1, e_3\rangle^2_{\bar{Q}}}}. \end{aligned}$$

The orthant probability $P(\bar{Z} \in \mathcal{C}(e_1, e_2, e_3))$ can be calculated with the Gauss-Bonnet Theorem ([Kl78], p.141, Gauss' theorema elegantissimum). Assume that S_2 is the unit sphere in $\mathbb{R}^3$ and Δ is a geodesic triangle on S_2. If β_j are the interior angles at the three corners of Δ then the volume $V(\Delta)$ is given by

$$V(\Delta) = \beta_1 + \beta_2 + \beta_3 - \pi.$$

Define $C := \mathcal{C}(\nu_1, \nu_2, \nu_3)$ for $\nu_1, \nu_2, \nu_3 \in \mathbb{R}^3$ linearly independent. Suppose that $P = C \cap S_2$. Then P is a geodesic triangle. To compute β_1, we have to project the difference vectors $\nu_2 - \nu_1$ and $\nu_2 - \nu_1$ onto the tangential space at the corner $\|\nu_1\|^{-1}\nu_1$ of the triangle. The projected vectors are tangential vectors of the triangle at corner $\|\nu_1\|^{-1}\nu_1$. Hence the inner angle is given by

$$\begin{aligned} \beta_1 &= \arccos \frac{(\nu_2 - \nu_1 - \frac{\nu_1^T(\nu_2-\nu_1)}{\nu_1^T\nu_1}\nu_1)^T(\nu_3 - \nu_1 - \frac{\nu_1^T(\nu_3-\nu_1)}{\nu_1^T\nu_1}\nu_1)}{\|\nu_2 - \nu_1 - \frac{\nu_1^T(\nu_2-\nu_1)}{\nu_1^T\nu_1}\nu_1\|_{\mathrm{id}}\|\nu_3 - \nu_1 - \frac{\nu_1^T(\nu_3-\nu_1)}{\nu_1^T\nu_1}\nu_1\|_{\mathrm{id}}} \\ &= \arccos \frac{\nu_2^T\nu_3\|\nu_1\|^2_{\mathrm{id}} - \nu_1^T\nu_2\nu_1^T\nu_3}{\sqrt{(\|\nu_1\|^2_{\mathrm{id}}\|\nu_2\|^2_{\mathrm{id}} - (\nu_1^T\nu_2)^2)(\|\nu_1\|^2_{\mathrm{id}}\|\nu_3\|^2_{\mathrm{id}} - (\nu_1^T\nu_3)^2)}}. \end{aligned}$$

Hence the probability for a three dimensional standard normal random variable to be in a cone $\mathcal{C}(\nu_1,\nu_2,\nu_3)$ is given by

$$\begin{aligned}
&\frac{1}{V(S_2)}\arccos\frac{\nu_2^T\nu_3\|\nu_1\|^2_{\text{id}}-\nu_1^T\nu_2\nu_1^T\nu_3}{\sqrt{(\|\nu_1\|^2_{\text{id}}\|\nu_2\|^2_{\text{id}}-(\nu_1^T\nu_2)^2)(\|\nu_1\|^2_{\text{id}}\|\nu_3\|^2_{\text{id}}-(\nu_1^T\nu_3)^2)}}\\
&+\frac{1}{V(S_2)}\arccos\frac{\nu_1^T\nu_3\|\nu_2\|^2_{\text{id}}-\nu_2^T\nu_1\nu_2^T\nu_3}{\sqrt{(\|\nu_2\|^2_{\text{id}}\|\nu_1\|^2_{\text{id}}-(\nu_2^T\nu_1)^2)(\|\nu_2\|^2_{\text{id}}\|\nu_3\|^2_{\text{id}}-(\nu_2^T\nu_3)^2)}}\\
&+\frac{1}{V(S_2)}\arccos\frac{\nu_1^T\nu_2\|\nu_3\|^2_{\text{id}}-\nu_3^T\nu_1\nu_3^T\nu_2}{\sqrt{(\|\nu_3\|^2_{\text{id}}\|\nu_1\|^2_{\text{id}}-(\nu_3^T\nu_1)^2)(\|\nu_3\|^2_{\text{id}}\|\nu_2\|^2_{\text{id}}-(\nu_3^T\nu_2)^2)}}\\
&-\frac{\pi}{V(S_2)}.
\end{aligned}$$

For $LL^T=\bar{Q}$ we have again $L^T\bar{Z}$ and with $V(S_2)=4\pi$

$$\begin{aligned}
&P(\bar{Z}\in\mathcal{C}(e_1,e_2,e_3))\\
&\quad= P(L^T\bar{Z}\in\mathcal{C}(L^Te_1,L^Te_2,L^Te_3))\\
&\quad= \frac{1}{4\pi}\cdot\frac{1}{2}\sum_{\{k,l,m\}=\{1,2,3\}}\arccos\frac{\langle e_l,e_m\rangle_{\bar{Q}}\|e_k\|^2_{\bar{Q}}-\langle e_k,e_l\rangle_{\bar{Q}}\langle e_k,e_m\rangle_{\bar{Q}}}{\sqrt{(\|e_k\|^2_{\bar{Q}}\|e_l\|^2_{\bar{Q}}-\langle e_k,e_m\rangle^2_{\bar{Q}})(\|e_k\|^2_{\bar{Q}}\|e_l\|^2_{\bar{Q}}-\langle e_k,e_l\rangle^2_{\bar{Q}})}}\\
&\qquad-\frac{1}{4}.
\end{aligned}$$

For higher dimensions closed forms do not exist (see [Ro88], p.75, for an exhaustive history of the derivation of orthant probabilities see [To90], p.188). More recent papers are [Ni00] or [No98]. Both papers discuss the orthant probabilities $P(Y\geq 0)$ for some multidimensional normally distributed random variable with covariance Σ. However, even at this stage of research the methods are restricted to special Σ. The first paper discusses the equicorrelated case, i.e. $\Sigma=(1-\tau)\text{id}+\tau\mathbf{1}\mathbf{1}^T$, the second one is about tridiagonal Σ. Neither of these conditions is satisfied here. Moreover, the formulas for $S=3$ are complicated enough to convince us that it would be more useful to give another method to derive the κ_i. We will propose a Monte Carlo method below.

We have seen that the mixing coefficients $\kappa_i=\kappa_i^S(\lambda)$ are depending on the unknown parameter λ. We will prove that studentization is possible.

Lemma 3.14 *The mapping $(0,\infty)\ni\lambda\longmapsto\kappa_i(S,\lambda)$ is continuously differentiable for every $S\geq 1$ and $0\leq i\leq S$.*

Proof: Recall the definition of $\bar{Q}=N^TA^{-1}\tilde{D}^{-1}A^{-T}N$. The matrices $A^{-1}=A^{-1}(\lambda,S)$ are given as the truncated versions of the derivative of L (section 2.3). They depend continuously differentiable on λ (consider some matrix norm). Moreover, $\tilde{D}^{-1}$ is a continuous differentiable function of λ. Therefore, $\bar{Q}$ and also

$\bar{Q}^{-1}$ are continuously differentiable functions. We can rewrite the cone probabilities in the following integral forms

$$\begin{aligned}
P(\bar{Z} \in \mathcal{C}(e_1, \ldots, e_S)) &= \frac{\sqrt{\det \bar{Q}}}{(2\pi)^{S/2}} \int_{z \geq 0} e^{-\frac{1}{2} z^T \bar{Q} z} \, dz \\
P(\bar{Z} \in \mathcal{C}(e_1, \ldots, e_S)^{*\bar{Q}}) &= P(\bar{Q}\bar{Z} \leq 0) \\
&= \frac{1}{(2\pi^{S/2})\sqrt{\det \bar{Q}}} \int_{z \leq 0} e^{-\frac{1}{2} z^T \bar{Q}^{-1} z} \, dz
\end{aligned}$$

and

$$\begin{aligned}
&P\left(\bar{Z} \in C(e_i | i \in I) \oplus \left(\mathrm{lin}\{e_i | i \in I\}^{\perp_{\bar{Q}}} \cap C(e_i | i \in I^C)^{*\bar{Q}}\right)\right) \\
&= \frac{\sqrt{\det E_I^T \bar{Q} E_I}}{(2\pi)^{S/2} \sqrt{\det\left(E_{I^C}^T \bar{Q} E_{I^C} - E_{I^C}^T \bar{Q} E_I (E_I^T \bar{Q} E_I)^{-1} E_I^T \bar{Q} E_{I^C}\right)}} \\
&\times \int_{z_1 \geq 0} e^{-\frac{1}{2} z_1^T (E_I^T \bar{Q} E_I) z_1} \, dz_1 \times \\
&\times \int_{z_2 \leq 0} e^{-\frac{1}{2} z_2^T (E_{I^C}^T \bar{Q} E_{I^C} - E_{I^C}^T \bar{Q} E_I (E_I \bar{Q} E_I)^{-1} E_I^T \bar{Q} E_{I^C})^{-1} z_2} \, dz_2.
\end{aligned}$$

The latter integral is calculated using the independence and the covariance structure of $(E_I^T \bar{Q} E_I)^{-1} E_I^T \bar{Q} \bar{Z}$ and $E_{I^C} \bar{Q}(\bar{Z} - E_I(E_I^T \bar{Q} E_I)^{-1} E_I^T \bar{Q} \bar{Z})$. Therefore z_1 is $\#I$-dimensional, and z_2 is $(S - \#I)$-dimensional. The corresponding covariance matrices and their inverses are again continuously differentiable functions of λ. Since the determinant is a continuously differentiable function and is not zero in a neighbourhood of the true parameter, everything is proved by considering a lemma discussing the continuously differentiability of parameter integrals in this case. This is no problem here, since we can differentiate under the integral sign.□

Let $\xi^S_{\lambda,1-\alpha}$ be the $(1-\alpha)$-quantile of the distribution $\sum_{l=0}^S \kappa_i(\lambda) \chi^2_{S-i}$.

Theorem 3.15 *Suppose that $\lambda_0 > 0$. Assume that H is satisfied. Let $\hat{\lambda}_n$ be a consistent estimator for λ. Let $1 - \alpha > \kappa^S_S(\lambda_0)$. Then*

$$\lim_{n \to \infty} P_{H,\lambda_0}(T_n \geq \xi_{\hat{\lambda}_n, 1-\alpha}) = \alpha \, .$$

Proof: Assume that F_λ is the distribution function corresponding to $\sum_{l=0}^S \kappa_i(\lambda) \chi^2_{S-i}$. Assume that λ_0 ist the true paramater, hence $\lambda_n \to \lambda_0$ in probability.

We want to show that $\lambda \longmapsto \xi_{1-\alpha}(\lambda)$ is continuous. We therefore solve the equation $1 - \alpha = F_\lambda(\xi)$ with the implicite mapping theorem. Note that $\frac{d}{dx} F_\lambda(x) > 0$ for all $x > 0$. Furthermore, for all $1 - \alpha > F_\lambda(0) = \kappa_0(\lambda)$ there exists a $\xi_{1-\alpha}(\lambda)$ such that

$$1 - \alpha = F_\lambda(\xi_{1-\alpha}(\lambda)).$$

Applying the implicit mapping theorem we have that $\xi_{1-\alpha}(\lambda)$ as the unique solution must be continuously differentiable mapping.

Hence $\xi_{(1-\alpha)}(\hat{\lambda}_n)$ converges to $\xi_{1-\alpha}(\lambda_0)$ in probability. The distributional limit law is established

$$T_n^1 - \xi_{(1-\alpha)}(\hat{\lambda}_n) \xrightarrow{\mathcal{D}} \|\pi(\bar{Z}|C^{*\bar{Q}}(e_1, \ldots, e_S))\|_{\bar{Q}}^2 - \xi_{\lambda_0, 1-\alpha}.$$

Under the condition that $1-\alpha > \kappa_0(\lambda)$, 0 is a continuity point of this distribution. Hence the the theorem has been proved. □

Some remarks about Monte Carlo approximations of the κ_i: We have seen that the calculation of the κ_i gets more and more tedious increasing threshold S. We have to calculate 2^S cone probabilities.

Hence we propose the following Monte-Carlo algorithm. We assume that $(\bar{Z}_1, \ldots, \bar{Z}_n)$ is an iid sample of S-dimensional centered Gaussian random variables with covariance $\bar{Q}$. This is easy realized on a computer using a random generator that produces nS uniform independent (pseudo) random variables. They can transformed via the Box-Muller method ([To90], p 12) to nS standard normally distributed ones. Put them in n vectors Y_i of dimension S. If $LL^T = \bar{Q}$ is the Cholesky decomposition then $\bar{Z}_i = L^{-T} Y_i$ is a simulation of an iid-sample from the desired normal distribution.

We estimate the κ_i via the relative frequency of the event that $\bar{Z}_j$ falls into the corresponding union of semicones. Again, the problem is to determine the semicone Θ_I with $z \in \Theta_I$ for a vector $z \in \mathbb{R}$. One way is to calculate its projection onto the cone via the active set method described in the appendix. If I is the active set of the solution then $z \in \Theta_I$. We should remark that the usual GCM-algorithm fails here for the following reasons: The GCM-Algorithm is made for the special cone of monotone functions, i.e. $K = \{x \in \mathbb{R}^S : \ x_0 \leq \cdot \leq x_S\}$ and $\pi_m(z|K)$ for some diagonal matrix. If we think of our basic cone $C := \mathcal{C}(\Delta_S - (\lambda, 1, 0, \ldots, 0)^T)$ then we have the same geometric structure. Hence we could make a linear transformation. But tranforming cones to cones will induce a transformation of the underlying matrix (see lemma 3.2). The author's conjecture is that the right transformation $C \longmapsto K$ and $A^T \tilde{D} A \to$ some diagonal matrix is possible for $S = 3$ and fails for higher dimensions. Hence at the moment there is a need for another method.

We propose here another algorithm. Note that it is not of primary interest to calculate the projection, but to determine the semicone. We again neglect the boundaries of the semicones here, since the union of the boundaries forms a set of Lebesgue measure 0.

Recall the matrix representation for the underlying cones, i.e.

$$\begin{array}{ll} (*) & \bar{Z}_j \in \mathcal{C}(e_i : i \in I) \oplus \Big(\mathrm{lin}\{e_i : i \in I\}^{\perp_{\bar{Q}}} \cap \mathcal{C}^{*\bar{Q}}(e_i | i \in I^C)\Big) \\ \Leftrightarrow & (E_I^T \bar{Q} E_I)^{-1} E_I^T \bar{Q} \bar{Z}_j \geq 0 \quad \wedge \quad E_{I^C}^T \bar{Q}(\bar{Z}_j - E_I (E_I^T \bar{Q} E_I)^{-1} E_I^T \bar{Q} \bar{Z}_j) \leq 0 \end{array}$$

for $0 < \#I < S$. For such I we define matrices $D_I \in \mathbb{R}^{S\times S}$ via the conditions

$$\begin{aligned} e_i^T D_I &= (E_I^T \bar{Q} E_I)^{-1} E_I^T \bar{Q} \text{ for } i \in I, \\ e_i^T D_I &= -e_i^T (\bar{Q} - \bar{Q} E_I (E_I^T \bar{Q} E_I)^{-1} E_I^T \bar{Q}) \text{ for } i \in I^C. \end{aligned}$$

Then $(*)$ can be written in a compact form as $D_I \bar{Z}_j \geq 0$. Write $I(\bar{Z}_i)$ for the corresponding set I.

The matrices D_I, once calculated, can be used again and again for the determination of other $I(\bar{Z}_m)$ because the geometric situation of cone and $\bar{Q}$ does not change.

We propose the following heuristic rule to find $I(\bar{Z}_j)$. It is very similar to the one used in the active set method explained in the appendix:

(1) If $\bar{Z}_j \geq 0$ then $I(\bar{Z}_j) = \{1 \leq i \leq S\}$.

(2) If not, calculate $-\bar{Q}\bar{Z}_j$. If $-\bar{Q}\bar{Z}_j \geq 0$, then $I(\bar{Z}_j) = \emptyset$.

(3) If not, then define $i_{\min}$ to be the index with

$$(-\bar{Q}\bar{Z}_j)_{i_{\min}} = \min\{(-QY)_m :\ m = 1\ldots, S\}.$$

Define $I := \{i_{\min}\}$.

(4) If D_I is not calculated yet, calculate it and save it.

(5) Calculate $D_I Y$. If $D_I \bar{Z}_j \geq 0$ then $I(\bar{Z}_j) = I$. Else define $i_{\min}$ to be the index with $(D_I \bar{Z}_j)_{i_{\min}} = \min\{(D_I Y)_m :\ m = 1\ldots, S\}$. If $i_{\min} \in I$, then define $I := I \backslash \{i_{\min}\}$, else define $I := I \cup \{i_{\min}\}$. Go back to (4).

This algorithm seems to converge quite quickly in practical applications. However, we should mentioned that we have no convergence proof for it. Cyclic behaviour might be possible, but it was not observed.

Chapter 4

Nonparametric Maximum Likelihood Estimation

The last chapter was devoted to the analysis of maximum likelihood estimation under truncation of the data. The upper truncation threshold S is somewhat artificial. So there is an urgent need to consider the untruncated case. Again we assume that the $Y's$ are concentrated on the nonnegative real numbers. We give some existence conditions for the nonparametric maximum likelihood estimator (NPMLE). Also a consistency proof is given that holds for the two-sided case as well, but it is not clear whether an NPMLE exist in this case.

The idea to prove the existence theorem is similar to that of Simar (see [Si76]). In contrast to the title of his paper he considers the NPMLE in the case of mixed Poisson and not compound Poisson distributions (in fact there is a big confusion in the literature concerning the definitions of compound and mixed Poisson distributions). A random variable Z has a mixed Poisson distribution if there is some probability measure μ on $[0, \infty)$ such that

$$P_\mu(Z = k) = \int e^{-\lambda} \frac{\lambda^k}{k!}\, \mu(d\lambda) \quad \text{for all } k \in \mathbb{N}_0$$

The generation of a mixed Poisson distributed random variable works as follows: Take some random variable Λ with distribution μ. This random variable generates some intensity λ. Then sample a Poisson distributed random variable X with parameter λ. X is mixed Poisson distributed according to the measure μ.

Simar has used Helly's theorem to prove the existence of some $\hat{\mu}$ maximizing the log likelihood function

$$\mu \longmapsto \sum_k r_k \log P_\mu(Z = k).$$

with r_k denoting the relative frequencies based on the observations. He also showed uniqueness and consistency of the NPMLE. Furthermore, he was able to specify the support of the NPMLE. The support is finite and given by an algebraic (to be more precise, a transcendent) equation. To derive the support of the NPMLE seems rather hopeless in the context of compound Poisson distributions, but if we restrict ourselves to the nonnegative integers then we can compute the estimator as a finitely dimensional maximization problem under linear constraints.

Roughly spoken we use the upper semicontinuity of the likelihood function defined as a function on some compact space to prove the existence of a NPMLE. Assume that $y = (y_1, \dots, y_n)$ are observed, so that we have the likelihood function

$$(\lambda, P) \longmapsto l(\lambda, P|y) := \prod_{l=1}^{n} e^{\lambda(P-\delta_0)}(\{y_l\}).$$

First note the following lemma that discusses the semicontinuity of the likelihood function. As usual we include the Dirac measure δ_0 into our reasoning about compound Poisson distributions according to the equation $\delta_0 = \exp(0(P - \delta_0))$. A Poisson distribution with parameter $\lambda = 0$ is also identified with the Dirac measure δ_0.

Lemma 4.1 *Let $\lambda, \lambda_k \geq 0$, $k \in \mathbb{N}$, and $P_k, P \in M_1(\mathbb{R})$. Suppose that*

$$\lim_{k\to\infty} \lambda_k = \lambda, \qquad P_k \xrightarrow{w} P \qquad \textit{for } k \to \infty,$$

then

$$\limsup_{k\to\infty} l(\lambda_k, P_k|y) \leq l(\lambda, P|y).$$

Proof: We immediately see from the corresponding characteristic functions that $\exp(\lambda_k(P_k - \delta_0)) \xrightarrow{w} \exp(\lambda(P - \delta_0))$. If $Q_k \xrightarrow{w} Q$ for some $Q_k, Q \in M_1(\mathbb{R})$, then $Q_k^{\otimes n} \xrightarrow{w} Q^{\otimes n}$ for $k \to \infty$, as easily seen from the characteristic functions. This shows that for all $n \in \mathbb{N}$

$$\left(e^{\lambda_k(P_k-\delta_0)}\right)^{\otimes n} \xrightarrow{w} \left(e^{\lambda(P-\delta_0)}\right)^{\otimes n}.$$

Obviously the set $\{(y_1, \dots, y_n)\}$ is closed. Therefore the assertion follows from the Portmanteau theorem (see [Ba92], p.227).□

Since the observation $y_1 = \cdots = y_n = 0$ would make the estimation problem trivial (estimate λ by 0), we exclude this case in the next theorem. We restrict ourselves to the case that the claim distribution is known to be concentrated on some closed subsemigroup H of the nonnegative real numbers, e.g. $H = \mathbb{N}, [2, \infty), \dots$.

Theorem 4.2 *Let $0 \leq y_1 \leq \cdots \leq y_n$ be given with $y_n > 0$. Suppose that $y_i \in H \cup \{0\}$, $i = 1, \dots, n$, for some closed subsemigroup H of $[0, \infty)$.*

i) Let S be a compact subset of the nonnegative real numbers with $S\backslash\{0\} \neq \emptyset$. Then there exists a $\hat{\lambda} \in S\backslash\{0\}, \hat{P} \in M_1(H \cap [0, y_n])$ with

$$l(\hat{\lambda}, \hat{P}|y) = \max_{\lambda \in S\backslash\{0\}, P \in M_1(H)} l(\lambda, P|y) \; .$$

ii) If $(0, \delta)^C \cap H = H$ for some $\delta > 0$ then there exists a $\hat{\lambda} > 0$ and a $\hat{P} \in M_1(H \cap (0, y_n])$ such that

$$l(\hat{\lambda}, \hat{P}|y) = \max_{\substack{P \in M_1(H \cap (0,\infty)) \\ \lambda > 0}} l(\lambda, P|y) \; .$$

Proof: An application of the inequality already used in the proof of part i) of lemma 2.9 yields

$$\mathbb{P}\left(\sum_{l=1}^{\tau} X_l = m\right) \geq \mathbb{P}(\tau = 1)\, \mathbb{P}(X_1 = m).$$

We therefore see that the likelihood function takes positive values. Indeed

$$\sup_{(\lambda,P)\in\Theta} l(\lambda, P|y) \geq l\left(s, \frac{1}{\#\{y_i > 0\}} \sum_{\substack{l=1 \\ y_l > 0}}^{n} \delta_{y_l} \middle| \, y \right) \geq e^{-ns} s^n \prod_{\substack{l=1 \\ y_l > 0}}^{n} \frac{1}{\#\{y_i > 0\}} > 0,$$

for both $\Theta = S\backslash\{0\} \times H$ and $\Theta = (0, \infty) \times M_1(H \cap (0, \infty))$ and $s \in S\backslash\{0\}$.

i) Let $P \in M_1(H)$ and $\lambda > 0$. Suppose $(\xi_i)_i$ to be some iid-sequence of random variables with $\xi_i \sim P$, defined on some common probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Suppose $\tau \sim \mathcal{P}(\lambda)$ to be defined on the same probability space and independent from (ξ_i). Define $\eta_i := \xi_i \wedge y_n$. Then $(\eta_i)_i$ is itself an iid-sequence of random variables. Suppose P' to be the distribution of η_i. Obviously, $P'(K) = 1$ with $K := H \cap [0, y_n]$. Furthermore we have

$$\mathbb{P}\left(\sum_{l=1}^{\tau} \xi_l = y_m\right) = \mathbb{P}\left(\sum_{l=1}^{\tau} \eta_l = y_m\right)$$

for all $y_m < y_n$ and

$$\begin{aligned}
\mathbb{P}\left(\sum_{l=1}^{\tau} \xi_l = y_n\right) &= \mathbb{P}\left(\sum_{l=1}^{\tau} \xi_l \leq y_n\right) - \underbrace{\mathbb{P}\left(\sum_{l=1}^{\tau} \xi_l < y_n\right)}_{=\mathbb{P}\left(\sum_{l=1}^{\tau} \eta_l < y_n\right)} \\
&\leq \mathbb{P}\left(\sum_{l=1}^{\tau} \eta_l \leq y_n\right) - \mathbb{P}\left(\sum_{l=1}^{\tau} \eta_l < y_n\right) = \mathbb{P}\left(\sum_{l=1}^{\tau} \eta_l = y_n\right).
\end{aligned}$$

Hence for $\lambda > 0$ fixed, the maximizing P is an element of $M_1(K)$, in symbols

$$\sup_{P \in M_1(H)} l(\lambda, P|y) \leq \sup_{P \in M_1(K)} l(\lambda, P|y).$$

K is compact. Therefore $M_1(K)$ is a compact subset of $M_1([0,\infty))$ if we equip $M_1([0,\infty))$ with the weak topology. The compactness of the set $S \times M_1(K)$ and the upper semicontinuity of l imply the existence of some maximizing $(\hat{\lambda}, \hat{P}) \in S \times M_1(K)$. Note that $\hat{\lambda}$ must be greater zero. Otherwise $l(\hat{\lambda}, \hat{P}|y) = 0$ would hold. Since $(\hat{\lambda}, \hat{P})$ maximizes $l(\cdot|y)$ even on the larger set $S \times M_1(K)$, hence on the smaller set $S \backslash \{0\} \times M_1(K)$, the assertion is proved.

ii) Define $L := H \cap (0, y_n]$. L is not empty and compact, since H is closed and zero can be dropped because of $H \cap (0, \delta) = H$ for some $\delta > 0$.

With the same argument as in i) we can reduce the maximization of the likelihood to the set $(0,\infty) \times M_1(L)$. Hence there exists some sequence $(\lambda_m, P_m) \in (0,\infty) \times M_1(L)$ with

$$l(\lambda_m, P_m) \geq \sup_{\substack{P \in M_1(H \cap (0,\infty)) \\ \lambda > 0}} l(\lambda, P|y) - \frac{1}{m}.$$

We show that λ_m has a bounded subsequence. If not, then $\lim_{m\to\infty} \lambda_m = \infty$. Since $M_1(L)$ is compact, we find some subsequence P_{m_k} and some $P \in M_1(L)$ with $P_{m_k} \xrightarrow{w} P$. Again consider independent random variables $X_{m,l}, \tau_m$ with $X_{m,l} \sim P_m$ and $\tau_m \sim \mathcal{P}(\lambda_m)$ defined on some underlying probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Note that $E(X_{m,l}^N)$ exists for all $N \in \mathbb{N}$ and that the following inequalities hold:

$$\delta^N \leq E(X_{m,l}^N) \leq y_n^N.$$

Hence Wald's identity $E\sum_{l=1}^{\tau_m} X_{m,l} = E\tau_m E X_{m,l}$ leads to

$$\lim_{m\to\infty} E \sum_{l=1}^{\tau_m} X_{m,l} = \lim_{m\to\infty} \lambda_m \underbrace{E X_{m,1}}_{\geq \delta} = \infty.$$

Note that $\operatorname{Var} \sum_{l=1}^{\tau_m} X_{m,l} = \lambda_m E(X_{m,1}^2)$. By Chebychev's inequality and Wald's identity

$$\begin{aligned} \mathbb{P}\left(\sum_{l=1}^{\tau_m} X_{m,l} \leq y_n\right) &= \mathbb{P}\left(\sum_{l=1}^{\tau_m} X_{m,l} - \lambda_m E X_{m,1} \leq y_n - \lambda_m E X_{m,1}\right) \\ &\leq \frac{\lambda_m E(X_{m,1}^2)}{(y_n - \lambda_m E X_{m,1})^2}. \end{aligned}$$

The right-hand side tends to zero. This implies that $l(\lambda_{m_k}, P_{m_k}|y) \to 0$ for $k \to \infty$, which is a contradiction, since the supremum of the likelihood must be greater than zero. Hence there exists a bounded subsequence of (λ_m). Without loss of generality we may assume that $\lambda_{m_k} \to \lambda$ for some $\lambda \geq 0$ and $P_{m_k} \xrightarrow{w} P$ for some $P \in M_1(L)$. Note that λ must be greater than zero, since otherwise

$$\limsup_{k\to\infty} l(\lambda_{m_k}, P_{m_k}|y) \leq l(\lambda, P|y) = 0.$$

The choice of λ_m, P_m shows that this is not possible. The assertion is established setting $\hat{\lambda} := \lambda$ and $\hat{P} = P$. $\Box$

Remarks: If there is no restriction on H or the parameter set of λ then the following pathologic situation can occur. Let $K := [0, y_n] \cap H$ as in the proof. Now fix some μ, λ with $\mu > \lambda > 0$. According to the first assertion of the theorem, there exist $P_\lambda, P_\mu \in M_1(K)$ with

$$\sup_{P\in M_1(K)} l(\lambda, P|y) = l(\lambda, P_\lambda), \qquad \sup_{P\in M_1(K)} l(\mu, P|y) = l(\mu, P_\mu).$$

Then define P_1 to be the probability measure

$$P_1 := \frac{\lambda}{\mu} P_\lambda + \left(1 - \frac{\lambda}{\mu}\right) \delta_0.$$

We obtain $\exp(\mu(P_1 - \delta_0)) = \exp(\lambda(P_\lambda - \delta_0))$, hence

$$l(\lambda, P_\lambda|y) = l(\mu, P_1|y) \leq l(\mu, P_\mu|y).$$

Obviously there are two cases. The first is that there exists some $\lambda > 0$ such that

$$l(\lambda, P_\lambda|y) = l(\mu, P_\mu|y)$$

holds for all $\mu > \lambda$. Then there are uncountable many maximizers for $l(\cdot|y)$ on $(0, \infty) \times M_1(H)$. The second is that for all $\lambda > 0$ there is some $\mu > \lambda$

$$l(\lambda, P_\lambda) < l(\mu, P_\mu|y).$$

Then a maximizer does not exist.

Under the assumption that the claim distribution is concentrated on the positive integers we want to investigate the consistency of the NPMLE for the case $H = \mathbb{N}_0$. We return to speak of probability densities instead of measures. Again, suppose (Y_i) to be an iid-sequence of random variables with $Y_i \sim q = \exp(\lambda(p - \delta_0))$ for some $\lambda > 0$ and $p \in M_1(\mathbb{N})$, defined on a common probability space $(\Omega, \mathbb{P}, \mathcal{A})$. As in Chapter 2 let q^n be the sequence of relative frequencies. Let $L(x|y) := \sum x_i \log y_i$ be the log likelihood function. Instead of a direct study of the NPMLEs we discuss the consistency of estimators T^n with the following property

$$(*) \qquad L(T^n|q^n) \geq L(q|q^n) + o_{\mathbb{P}}(1).$$

If $(\hat{\lambda}^n, \hat{p}^n)$ is a measurable selection of the NPMLE, then $(*)$ holds for $T^n := \exp(\hat{\lambda}^n(\hat{p}^n - \delta_0))$. If T^n is consistent, i.e. T^n converges to q in probability, then the continuity of the logarithm implies the consistency of $(\hat{\lambda}^n, \hat{p}^n)$.

First we state the theorem. It is based on two lemmas which are given after the proof of the theorem. The main ingredient of the proof is the following simple observation that uses the special feature of countability in the discrete case:

$$(**) \qquad q \otimes \mathbb{P}\left(\left\{(k,\omega) : \lim_{n\to\infty} \frac{q_k^n(\omega)}{q_k} = 1\right\}\right) \quad \geq$$
$$q \otimes \mathbb{P}\left(\mathbb{N}_0 \times \bigcap_k \left\{\omega : \lim_{n\to\infty} \frac{q_k^n(\omega)}{q^k} = 1\right\}\right) \; = \; 1 \, .$$

Theorem 4.3 *Suppose that $\sum q_k^{1-\gamma} < \infty$ for some $\gamma > 0$. If $(T_n)_n$ is a sequence of ℓ^1-valued estimators that satisfies $(*)$ then $\|T_n - q^n\|_1 \xrightarrow{\mathbb{P}} 0$. Hence $T_n \xrightarrow{\mathbb{P}} q$.*

Remark: Recall that $\sum q_k^{1-\gamma} < \infty$ iff $\sum p_k^{1-\gamma} < \infty$ because of lemma 2.9.

Proof: We have the following identity for the Kullback-Leibler divergence

$$H(x|y) = L(x|x) - L(y|x).$$

According to lemma 4.5, $((k,\omega) \to \frac{q_k^n(\omega)}{q_k} \log \frac{q_k^n(\omega)}{q_k})_n$ is uniformly $q \otimes \mathbb{P}$-integrable. From $(**)$ it follows that

$$\lim_{n\to\infty} \frac{q_k^n(\omega)}{q_k} \log \frac{q_k^n(\omega)}{q_k} = 0 \qquad q \otimes \mathbb{P} - \text{a.s.},$$

hence

$$\lim_{n\to\infty} \sum_k q_k E \left| \frac{q_k^n}{q_k} \log \frac{q_k^n}{q_k} \right| = 0.$$

With Chebychev's inequality and Fubini we obtain

$$\mathbb{P}(H(q^n|q) > \epsilon) \leq \frac{1}{\epsilon} E|H(q^n|q)| = \frac{1}{\epsilon} \int \left| \frac{q_k^n(\omega)}{q_k} \log \frac{q_k^n(\omega)}{q_k} \right| q \otimes \mathbb{P}(dk,\omega).$$

This yields $H(q^n|q) \xrightarrow{\mathbb{P}} 0$, hence the assertion is established by the following inequality

$$\begin{aligned} \frac{1}{8}\|T_n - q^n\|_1^2 &\leq H(q^n|T_n) = L(q^n|q^n) - L(T_n|q^n) \\ &\leq L(q^n|q^n) - L(q|q^n) + o_{\mathbb{P}}(1) = H(q^n|q) + o_{\mathbb{P}}(1). \end{aligned}$$

□

The next two lemmas discuss the uniform integrability.

Lemma 4.4 *Suppose that* $\sum_k q_k^{1-\gamma} < \infty$ *for some* $\gamma > 0$. *Then the family*

$$\left((k,\omega) \mapsto \left(\frac{q_k^n(\omega)}{q_k} \right)^{1+\epsilon} \right)_{n \in \mathbb{N}_0}$$

is uniformly $q \otimes \mathbb{P}$*-integrable for every* $\epsilon \in [0, \gamma)$.

Remark: We have the simple inequality $E(q_k^n)^{1+\epsilon} \leq E q_k^n = q_k$. Therefore

$$\sum_k q_k E \left(\frac{q_k^n}{q_k} \right)^{1+\epsilon} \leq \sum_k q_k^{1-\epsilon}.$$

Fubini's theorem implies the $q \otimes \mathbb{P}$-integrability of the family.

Proof: If $\epsilon = 0$ then it holds that

$$\sum_k q_k E \frac{q_k^n}{q_k} = 1.$$

This implies uniform $q \otimes \mathbb{P}$-integrability (see theorem 21.7 (ii) in [Ba92]).

Let $\epsilon \in (0, \gamma)$. Choose some $\delta > 0$ such that $1 - \delta(1+\epsilon)/2 \geq 1 - \gamma$.

Define $\gamma_k := (1 + q_k^{-\delta/2})^{1+\epsilon}$. The following inequality is true for every $k \in \mathbb{N}_0$

$$\left(1 + q_k^{-\delta/2}\right)^{1+\epsilon} \leq 2^{1+\epsilon} \left(1 + q_k^{-\delta(1+\epsilon)/2}\right).$$

Therefore

$$\sum_k \gamma_k q_k \leq 2^{1+\epsilon} \left(1 + \sum_k q_k^{1-\delta(1+\epsilon)/2} \right) \leq 2^{1+\epsilon} \left(1 + \sum_k q_k^{1-\gamma} \right) < \infty.$$

The random variable $Z(k, \omega) := \gamma_k$ is therefore $q \otimes \mathbb{P}$-integrable.

Applying Chebychev's inequality we conclude that

$$\mathbb{P}\left(\frac{q_k^n}{q_k} > \gamma_k^{\frac{1}{1+\epsilon}} \right) = \mathbb{P}\left(q_k^n - q_k > q_k \left(\gamma_k^{\frac{1}{1+\epsilon}} - 1 \right) \right) \leq \frac{1}{n} \frac{1}{q_k^2 \left(\gamma_k^{\frac{1}{1+\epsilon}} - 1 \right)^2} q_k = \frac{1}{n} q_k^{\delta - 1}.$$

Suppose $r > 1$ to be chosen large enough to assure $1 - \epsilon + \frac{1}{r}(\delta - 2) > 1 - \gamma$. Recall $E(q_k^n)^{s(1+\epsilon)} \leq q_k$. With $s^{-1} = 1 - r^{-1}$ Hölder's inequality yields

$$\int_{\left(\frac{q_k^n(\omega)}{q_k}\right)^{1+\epsilon} > \gamma_k} \left(\frac{q_k^n(\omega)}{q_k} \right)^{1+\epsilon} q \otimes \mathbb{P}(dk, d\omega) = \sum_k q_k^{-\epsilon} E 1_{\frac{q_k^n}{q_k} > \gamma_k^{\frac{1}{1+\epsilon}}} (q_k^n)^{1+\epsilon}$$

$$\leq \sum_k q_k^{-\epsilon} \sqrt[s]{E(q_k^n)^{s(1+\epsilon)}} \sqrt[r]{\mathbb{P}\left(\frac{q_k^n}{q_k} > \gamma_k^{\frac{1}{1+\epsilon}} \right)} \leq \frac{1}{\sqrt[r]{n}} \sum q_k^{1-\epsilon+\frac{1}{r}(\delta-2)} \leq \frac{1}{\sqrt[r]{n}} \sum_k q_k^{1-\gamma}.$$

This shows

$$\lim_{n\to\infty}\int_{\left(\frac{q_k^n(\omega)}{q_k}\right)^{1+\epsilon}>Z(k,\omega)}\left(\frac{q_k^n(\omega)}{q_k}\right)^{1+\epsilon} q\otimes\mathbb{P}(dk,d\omega)=0.$$

Lebesgue's dominated convergence theorem yields

$$\lim_{n\to\infty}\int\left(\frac{q_k^n}{q_k}\right)^{1+\epsilon} dq\otimes\mathbb{P}=\lim_{n\to\infty}\int 1_{\left(\frac{q_k^n}{q_k}\right)^{1+\epsilon}\leq Z}\left(\frac{q_k^n}{q_k}\right)^{1+\epsilon} dq\otimes\mathbb{P}=1\ .$$

This is equivalent to uniform integrability (see [Ba92], 21.7).□

Lemma 4.5 *Suppose that* $\sum q_k^{1-\gamma}<\infty$ *for some* $\gamma>0$. *Then the family*

$$\left((k,\omega)\mapsto\frac{q_k^n}{q_k}(\omega)\log\frac{q_k^n}{q_k}(\omega)\right)_{n\in\mathbb{N}}$$

is uniformly $q\otimes\mathbb{P}$*-integrable.*

Proof: Define $\epsilon:=\gamma/2$. There is some $S>0$ with $|x\log x|\leq S\vee x^{1+\epsilon}$ for all $x\geq 0$. Since

$$\left((k,\omega)\mapsto\left(\frac{q_k^n(\omega)}{q_k}\right)^{1+\epsilon}\right)_{n\in\mathbb{N}}$$

is uniformly integrable, we also have uniform integrability for

$$\left((k,\omega)\mapsto\left(\frac{q_k^n(\omega)}{q_k}\right)^{1+\epsilon}\vee S\right)_{n\in\mathbb{N}}.$$

Hence our family is dominated by a uniformly integrable family, hence itself uniformly integrable.□

Chapter 5

An Inverse Panjer Formula for Histograms

We have seen that the inverse Panjer formula turns up very naturally if we consider the counting densities as elements of an appropriate Banach algebra. Up to some affine transformation the Panjer inversion can be identified with the logarithm in a Banach algebra. Our goal is to find an analogue for the Panjer inversion in the case of claims with an absolutely continuous distribution. Very popular density estimators are histograms (see [Si86], [De85]), so it seems natural to look for a formula there.

We use the setting of the Banach algebra $\delta_0 \oplus L^1$ with $L_1 := L^1_{\mathbb{R}}([0,\infty))$. Suppose that $q_0\delta_0 + q = \exp(\lambda(p - \delta_0))$ for some $q, p \in L^1, q_0, \lambda > 0$.

For completeness we derive the Volterra integral equation for p and q which can be found in the book of [Pa92] (see p.222). It is stated there in a much more general setting and has some remarkable similarity to the Panjer recursion formula.

Writing down the Laplace transforms we obtain

$$q_0 + \int q(t)e^{-\mu t}\, dt = \exp\left(\lambda\left(\int p(t)e^{-\mu t} dt - 1\right)\right) \qquad \text{for } \mu \geq 0.$$

Both sides are differentiable for $\mu > 0$ and can be differentiated under the integral sign. Therefore

$$\begin{aligned}
\int t\exp(-\mu t)q(t)dt &= \exp\left(\lambda\left(\int p(t)\exp(-\mu t)dt - 1\right)\right)\lambda\int tp(t)\exp(-\mu t)dt \\
&= \left(q_0 + \int q(t)\exp(-\mu t)\, dt\right)\lambda\int tp(t)\exp(-\mu t)dt \\
&= q_0\lambda\int tp(t)\exp(-\mu t)dt + \lambda\int\exp(-\mu t)\int_0^t sp(s)q(t-s)dsdt.
\end{aligned}$$

For fixed μ_0 define

$f_{\mu_0}(t) := t\exp(-\mu_0 t)q(t)$,$g_{\mu_0}(t) := \lambda\exp(-\mu_0 t)(q_0 tp(t) + tp(t)\int_0^t sp(s)q(t-s)ds)$.
Then $g_{\mu_0}, f_{\mu_0} \in L^1$ and their Laplace transforms coincide. Hence $g_{\mu_0} = f_{\mu_0}$ Lebesgue-a.e.. Multiplying both sides with $e^{\mu_0 t}$ yields the Volterra integral equation

$$(VI) \qquad tq(t) = \lambda q_0 tp(t) + \lambda\int_0^t sp(s)q(t-s)ds,\ t>0.$$

Now consider histograms: Let $h>0$ and

$$\begin{aligned} a(t) &= q_0\delta_0 + \sum_{k=0}^{\infty}\alpha_k 1_{I_k}(t), \\ I_k &= [kh,(k+1)h),\ k\geq 0 \\ \lambda &= -\log(q_0). \end{aligned}$$

A function p can be calculated such that (VI) holds with a in place of q. It does not matter whether a has a logarithm in $\delta_0 \oplus L^1_{\mathbb{R}}(\mathbb{R})$. Otherwise p is just some piecewise continuous function, not necessarily in L^1. This is similar to the Panjer inversion.

Theorem 5.1 *Let $\alpha_0 > 0$. Let $\gamma_0 := \alpha_0/q_0$. Then*

$$p(t) = \frac{1}{\lambda t}\sum_{k\geq 0}\phi_k(t)1_{I_k}(t)$$

with

$$\phi_k(t) = 1 + e^{-\gamma_0(t-kh)}\sum_{l=0}^{k} D_{k,l}(t-kh)^l.$$

The coefficients can be computed via the following recursion formulas:

$$\begin{aligned} D_{0,0} &= -1, \\ D_{k,0} &= \frac{kh(\alpha_k - \alpha_{k-1})}{q_0} + e^{-\gamma_0 h}\sum_{j=0}^{k-1} D_{k-1,j}h^j \quad \forall k\geq 1, \\ D_{k,j} &= \frac{1}{jq_0}\Bigg(\alpha_0 D_{k-1,j-1} + \sum_{l=1}^{k-j}\alpha_l(D_{k-l-1,j-1} - D_{k-l,j-1}) \\ &\qquad - \alpha_{k-j+1}D_{j-1,j-1}\Bigg) \quad \forall k\geq 1,\ 0\leq j\leq k. \end{aligned}$$

Proof: Define $\phi(s) := \lambda sp(s)$. Then p satisfies the integral equation

$$(I) \qquad ta(t) = q_0\phi(t) + \int_0^t \phi(s)a(t-s)\,ds.$$

Define ϕ_k to be the restriction of ϕ onto the interval I_k. It is sufficient to solve (I). This is done by induction over the k's. Assume $k = 0$, then (I) simplifies for $t \in I_0$ to

$$t\alpha_0 = q_0\phi(t) + \alpha_0 \int_0^t \phi(s)\,ds \qquad \forall t \in I_0.$$

We solve the equivalent initial value problem

$$(0) \qquad \alpha_0 = q_0\phi'(t) + \alpha_0\phi(t), \quad \phi(0) = 0.$$

This is a linear differential equation that can be tackled by the method of separation of variables. The solution of the homogeneous equation is given by $\phi_0^H(t) = e^{-\gamma_0 t}$, i.e.

$$0 = q_0{\phi_0^H}'(t) + \alpha_0\phi_0^H(t).$$

Consider $\phi_0 = \psi_0\phi_0^H$ with ψ_0 to determined. Plugging into (0) yields

$$\alpha_0 \stackrel{!}{=} q_0(\phi_0^H(t)\psi_0(t))' + \alpha_0(\phi_0^H(t)\psi_0(t)) = q_0e^{-\gamma_0 t}\psi_0'(t).$$

Hence there is some constant C such that $\psi_0(t) = C + e^{\gamma_0 t}$ is satisfied. Since $0 = \phi_0^H(0)\psi_0(0) = C + 1$ holds, we obtain $C = -1$. We have proved $\phi_0(t) = 1 - e^{-\gamma_0 t}$.

Now assume $k > 1$. Again, we have a simplication for (I) for $t \in I_k$, i.e.

$$\begin{aligned}
\alpha_k t &= q_0\phi(t) + \sum_{l=0}^{k} \alpha_l \int_0^t \phi(s)1_{I_l}(t-s)\,ds \\
&= q_0\phi(t) + \sum_{l=0}^{k} \alpha_l \int_{[0,t]\cap(t-(l+1)h,t-lh]} \phi(s)\,ds \\
&= q_0\phi(t) + \alpha_k \int_0^{t-kh} \phi(s)\,ds + \alpha_0 \int_{t-h}^{t} \phi(s)\,ds \\
&\quad + \sum_{l=1}^{k-1} \alpha_l \int_{(t-(l+1)h,t-lh]} \phi(s)\,ds \\
&= q_0\phi_k(t) + \alpha_k \int_0^{t-kh} \phi_0(s)\,ds \\
&\quad + \alpha_0 \int_{t-h}^{kh} \phi_{k-1}(s)\,ds + \alpha_0 \int_{kh}^{t} \phi_k(s)\,ds \\
&\quad + \sum_{l=1}^{k-1} \alpha_l \int_{(t-(l+1)h,(k-l)h]} \phi_{k-l-1}(s)\,ds + \sum_{l=1}^{k-1} \alpha_l \int_{((k-l)h,t-lh]} \phi_{k-l}(s)\,ds \\
(*) &= q_0\phi_k(t) + \alpha_0 \int_{kh}^{t} \phi_k(s)\,ds + \sum_{l=0}^{k-1} \alpha_l \int_{(t-(l+1)h,(k-l)h]} \phi_{k-l-1}(s)\,ds \\
&\quad + \sum_{l=1}^{k} \alpha_l \int_{((k-l)h,t-lh]} \phi_{k-l}(s)\,ds.
\end{aligned}$$

Differentiating with respect to t yields

$$\begin{aligned}\alpha_k &= q_0\phi_k'(t)+\alpha_0\phi_k(t)\\ &\quad +\sum_{l=1}^{k}\alpha_l\phi_{k-l}(t-lh)-\sum_{l=0}^{k-1}\alpha_l\phi_{k-l-1}(t-(l+1)h)\,.\end{aligned}$$

This is again solvable using separation of variables. $\phi_k^H(t)=e^{-\gamma_0(t-kh)}$ solves the homogeneous equation $0=q_0\phi_k^{H\prime}+\alpha_0\phi_k^H$. Consider again $\phi_k=\psi_k\phi_k^H$ with some ψ_k to determined. We have

$$\alpha_k=q_0\phi_k^H(t)\psi_k'(t)+\sum_{l=1}^{k}\alpha_l\phi_{k-l}(t-lh)-\sum_{l=0}^{k-1}\alpha_l\phi_{k-l-1}(t-(l+1)h),$$

hence

$$\begin{aligned}\psi_k(t) &= C+\frac{1}{q_0}\Bigg(\alpha_k\int_{kh}^{t}e^{\gamma_0(s-kh)}\,ds\\ &\quad -\int_{kh}^{t}\sum_{l=1}^{k}\alpha_l\Big\{e^{\gamma_0(s-kh)}+e^{\gamma_0(s-kh)-\gamma_0(s-lh-(k-l)h)}\times\\ &\qquad\qquad \times\sum_{j=0}^{k-l}D_{k-l,j}(t-lh-(k-l)h)^j\Big\}\\ &\quad +\int_{kh}^{t}\sum_{l=0}^{k-1}\alpha_l\Big\{e^{\gamma_0(s-kh)}+e^{\gamma_0(s-kh)-\gamma_0(s-(l+1)h-(k-l-1)h)}\times\\ &\qquad\qquad \times\sum_{j=0}^{k-l-1}D_{k-l-1,j}\big[t-(l+1)h-(k-l-1)h\big]^j\Big\}\Bigg)\\ &= C'+\frac{1}{q_0}\Bigg(\frac{\alpha_k}{\gamma_0}e^{\gamma_0(t-kh)}\\ &\quad -\sum_{l=1}^{k}\alpha_l\Big\{\frac{1}{\gamma_0}e^{\gamma_0(t-kh)}+\sum_{j=0}^{k-l}D_{k-l,j}\frac{1}{j+1}(t-kh)^{j+1}\Big\}\\ &\quad +\sum_{l=0}^{k-1}\alpha_l\Big\{\frac{1}{\gamma_0}e^{\gamma_0(t-kh)}+\sum_{j=0}^{k-l-1}D_{k-l-1,j}\frac{1}{j+1}(t-kh)^{j+1}\Big\}\Bigg).\end{aligned}$$

We therefore obtain

$$
\begin{aligned}
& -\sum_{l=1}^{k}\alpha_l\sum_{j=0}^{k-l}D_{k-l,j}\frac{1}{j+1}(t-kh)^{j+1}+\sum_{l=0}^{k-1}\alpha_l\sum_{j=0}^{k-l-1}D_{k-l-1,j}\frac{1}{j+1}(t-kh)^{j+1}\\
= & -\sum_{j=0}^{k-1}\frac{1}{j+1}(t-kh)^{j+1}\sum_{l=1}^{k-j}\alpha_l D_{k-l,j}+\sum_{j=0}^{k-1}\frac{1}{j+1}(t-kh)^{j+1}\sum_{l=0}^{k-j-1}\alpha_l D_{k-l-1,j}\\
= & -\sum_{j=1}^{k}\frac{1}{j}(t-kh)^{j}\sum_{l=1}^{k-j+1}\alpha_l D_{k-l,j-1}+\sum_{j=1}^{k}\frac{1}{j}(t-kh)^{j}\sum_{l=0}^{k-j}\alpha_l D_{k-l-1,j-1}\\
= & \sum_{j=1}^{k}\frac{1}{j}(t-kh)^{j}\Bigg(-\alpha_{k-j+1}D_{j-1,j-1}\\
& \qquad +\sum_{l=1}^{k-j}\alpha_l(D_{k-l-1,j-1}-D_{k-l,j-1})+\alpha_0 D_{k-1,j-1}\Bigg)
\end{aligned}
$$

Plugging the latter equation into the equation above yields

$$
\begin{aligned}
\psi_k(t) &= C'+\underbrace{\frac{1}{q_0\gamma_0}\underbrace{\left(\alpha_k-\sum_{l=1}^{k}\alpha_l+\sum_{l=0}^{k-1}\alpha_l\right)}_{=\alpha_0}}_{=1}e^{\gamma_0(t-kh)}\\
&\quad +\frac{1}{q_0}\Bigg(\sum_{j=1}^{k}\frac{1}{j}(t-kh)^{j}\Bigg(\alpha_0 D_{k-1,j-1}\\
&\qquad +\sum_{l=1}^{k-j}\alpha_l\big(D_{k-l-1,j-1}-D_{k-l,j-1}\big)-\alpha_{k-j+1}D_{j-1,j-1}\Bigg)\Bigg)\\
&= C'+e^{\gamma_0(t-kh)}+\sum_{j=1}^{k}D_{k,j}(t-kh)^j
\end{aligned}
$$

Hence

$$
\phi_k(t)=1+e^{-\gamma_0(t-kh)}\left(C'+\sum_{j=1}^{k}D_{k,j}(t-kh)^j\right).
$$

The constant $C'=D_{k,0}$ can be calculated plugging the initial value $t=kh$ into $(*)$:

$$
\alpha_k kh=q_0(1+D_{k,0})+\sum_{l=0}^{k-1}\alpha_l\int_{(l-l-1)h}^{(k-l)h}\phi_{k-l-1}(s)\,ds,
$$

hence

$$\begin{aligned} D_{k,0} &= \frac{1}{q_0}\left(\alpha_k kh - \sum_{l=0}^{k-1}\alpha_l \int_{(l-l-1)h}^{(k-l)h} \phi_{k-l-1}(s)\,ds\right) - 1 \\ &= \frac{1}{q_0}\Bigg(\alpha_k kh - \sum_{l=0}^{k-1}\alpha_l \sum_{j=0}^{k-l-1} D_{k-l-1,j} \times \\ &\qquad \int_{(k-l-1)h}^{(k-l)h} e^{-\gamma_0(s-(k-l-1)h)}(s-(k-l-1)h)^j\,ds\Bigg) - 1 \\ &= \frac{1}{q_0}\left(\alpha_k kh - h\sum_{l=0}^{k-1}\alpha_l + \sum_{l=0}^{k-1}\alpha_l \sum_{j=0}^{k-l-1} D_{k-l-1,j}\int_0^{\gamma_0 h} e^{-\gamma_0 s}\left(\frac{s}{\gamma_0}\right)^j \frac{ds}{\gamma_0}\right) - 1\ . \end{aligned}$$

This proves the first representation of $D_{k,0}$. The form stated in the assertion of the theorem is derived under some additional reasoning. Note that $t \longmapsto \int_0^t \phi(s)a(t-s)ds$ is continuous. Considering the integral equation at the jumps, we derive

$$\begin{aligned} &kh(\alpha_k - \alpha_{k-1}) = kh(q(kh) - q(kh-)) = \\ &\qquad q_0(\phi(kh) - \phi(kh-)) = q_0(1 + D_{k,0} - (1 + \exp(-\gamma_0 h)\sum_{j=0}^{k-1} D_{k-1,j}h^j)). \end{aligned}$$

Hence

$$D_{k,0} = \frac{kh(\alpha_k - \alpha_{k-1})}{q_0} + e^{-\gamma_0 h}\sum_{j=0}^{k-1} D_{k-1,j}h^j.$$

□

If the first cell is empty, then we have the following simplification:

Theorem 5.2 *Assume* $\alpha_0 = 0$.
Then

$$p(t) = \frac{1}{\lambda t}\sum_{k\geq 0}\phi_k(t)1_{I_k}(t)$$

with

$$\phi_k(t) = \sum_{l=0}^{k} D_{k,l}(t-kh)^l,\ k \geq 1, \phi_0 \equiv 0.$$

Recursion formulas for the coefficients:

$$
\begin{aligned}
D_{1,0} &= \frac{\alpha_1}{q_0}, \\
D_{1,1} &= \frac{\alpha_1}{q_0}h, \\
D_{k,0} &= \frac{kh(\alpha_k - \alpha_{k-1})}{q_0} + \sum_{l=0}^{k-1} D_{k-1,j}h^j, \\
D_{k,k} &= -\frac{\alpha_1}{q_0 k} D_{k-1,k-1}, \\
D_{k,1} &= \frac{1}{q_0}\left(\alpha_k + \sum_{l=1}^{k-1} \alpha_l (D_{k-l-1,0} - D_{k-l,0})\right), \\
D_{k,j} &= \frac{1}{q_0}\left(\sum_{l=1}^{k-j} \alpha_l (D_{k-l-1,j-1} - D_{k-l,j-1}) - \alpha_{k-j+1} D_{j-1,j-1}\right).
\end{aligned}
$$

Proof: It is easy to calculate that $\phi_0 \equiv 0$ and $\phi_1(t) = \frac{\alpha_1}{q_0}(t-h) + \frac{\alpha_1}{q_0}h$. We consider (∗) in the proof of the latter theorem. For $t \in I_k$ it holds

$$\begin{aligned}
q_0\phi_k(t) &= \alpha_k t - \sum_{l=1}^{k-1}\alpha_l \int_{t-(l+1)h}^{(k-l)h} \sum_{j=0}^{k-l-1} D_{k-l-1,j}\big(s-(k-l-1)\,h\big)^j ds \\
&\quad - \sum_{l=1}^{k}\alpha_l \int_{(k-l)h}^{t-lh} \sum_{j=0}^{k-l} D_{k-l,j}(s-(k-l)h)^j ds \\
&= \alpha_k t - \sum_{l=1}^{k-1}\alpha_l \sum_{j=0}^{k-l-1} D_{k-l-1,j}\frac{1}{j+1}\left(h^{j+1}-(t-kh)^{j+1}\right) \\
&\quad - \sum_{l=1}^{k}\alpha_l \sum_{j=0}^{k-l} D_{k-l,j}\frac{1}{j+1}(t-kh)^{j+1} \\
&= \alpha_k t - \sum_{l=1}^{k-1}\alpha_l \sum_{j=0}^{k-l-1} D_{k-l-1,j}\frac{1}{j+1}h^{j+1} \\
&\quad + \sum_{j=0}^{k-2}\frac{1}{j+1}(t-kh)^{j+1}\sum_{l=1}^{k-j-1}\alpha_l D_{k-l-1,j} \\
&\quad - \sum_{j=0}^{k-1}\frac{1}{j+1}(t-kh)^{j+1}\sum_{l=1}^{k-j}\alpha_l D_{k-l,j} \\
&= \alpha_k t - \sum_{l=1}^{k-1}\alpha_l \sum_{j=0}^{k-l-1} D_{k-l-1,j}\frac{1}{j+1}h^{j+1} \\
&\quad + \sum_{j=1}^{k-1}\frac{1}{j}(t-kh)^{j}\sum_{l=1}^{k-j}\alpha_l D_{k-l-1,j-1} \\
&\quad - \sum_{j=1}^{k}\frac{1}{j}(t-kh)^{j}\sum_{l=1}^{k-j+1}\alpha_l D_{k-l,j-1}.
\end{aligned}$$

Comparing the coefficients yields the theorem. The representation for $D_{k,0}$ follows once again by reasoning about the jumps.□

Assume that (Y_i) is an iid-sequence of random variables with

$$Y_i \sim q_0\delta_0 + q = e^{\lambda(p-\delta_0)}, \text{ for some } p, q \in L^1, \lambda > 0.$$

We estimate q_0 with the relative frequency $q_0^n = \frac{1}{n}\sum_{l=1}^n 1_{Y_l=0}$ and q by a histogram

$$\bar{q}_h^n = \sum_{l=0} \alpha_k^{n,h} 1_{I_l^h}, \qquad I_l^h = [hl, h(l+1)), \quad \alpha_l^{n,h} = \frac{1}{nh}\sum_{k=1}^n 1_{Y_k \in I_l}.$$

We have the following consistency result. Note that h depends on n.

Lemma 5.3 *Assume* $\lim_{n\to\infty} h \to 0$ *and* $\lim_{n\to\infty} nh \to \infty$*, then*

$$q_0^n\delta_0 + q^{nh} \quad \longrightarrow \quad q_0\delta_0 + q$$

in $\delta_0 \oplus L^1$ *a.s..*

Proof: Obviously, $q_0^n \xrightarrow{\text{a.s.}} q_0$.

Consider an iid-sequence of random variables $(Z_i)_i$ with a distribution that has a Lebesgue density g. Define the histograms

$$g^{n,h}(x) = \frac{1}{nh}\sum_{l\in\mathbb{Z}} 1_{[lh,(l+1)h)}(x)\sum_{m=1}^{n} 1_{Z_m\in[lh,(l+1)h)}.$$

It is known that $J_n := \int |g^{nh} - g|$ tends to 0 a.s. iff $h \to 0$ and $nh \to \infty$ (see theorem 2 in [De85] and the concluding remarks for histograms). Now define $Z_i = Y_i 1_{Y_i>0} - U_i 1_{Y_i=0}$ for some iid-sequence of random variables (U_i), uniformly distributed on $(0,1)$ and independent from (Y_i). Then (Z_i) is an iid-sequence of random variables. The distribution of Z_i has the Lebesgue density $g = e^{-\lambda}1_{(-1,0)} + q$. Hence $\lim_{n\to\infty} g^{n,h} = g$ in L^1 a.s. if the conditions given on h and n are satisfied. The assertion of the lemma is proved with the help of the inequality $\int |q^{n,h} - q| \leq \int |g^{n,h} - g|$.

□

This lemma shows that the estimator $q_0^n\delta_0 + q^{nh}$ is strongly consistent in $\delta_0 + L^1$. Once again note that the compounding mapping $\lambda\delta_0 + p \mapsto e^{\lambda(p-\delta_0)}$ has exactly the same properties as in the case ℓ^1. Therefore $q_0^n\delta_0 + q^{nh}$ will be in the codomain of C for n big enough. We can apply the unique real valued logarithm to it. This gives us a strongly consistent estimators for λ and p which can be calculated via $\lambda = -\log q_0^n$ and the formulas given in the theorems above. The empirical Fourier transform is null-homotopic in $\mathbb{C}^*$.

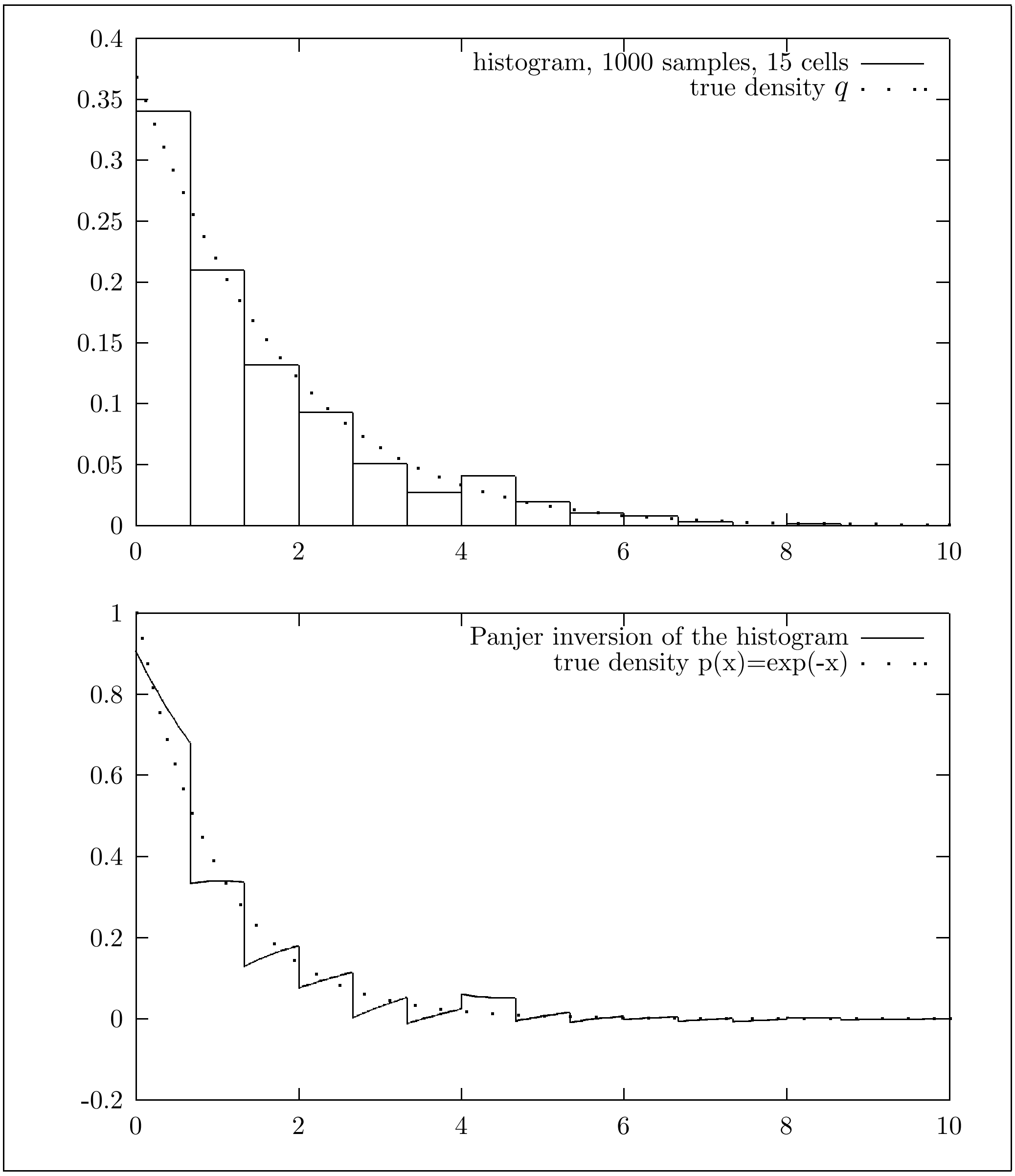

Figure 5.1: At the top q and a histogram with 15 cells is plotted based on 1000 samples from the distribution $q_0\delta_0 + q = \exp(\lambda(p - \delta_0))$ with $\lambda = 1$ and $p(x) = \exp(-x)1_{[0,\infty)}(x)$ is plotted. The figure at the bottom shows the corresponding p and the Panjer inversion of the histogram.

Chapter 6

The General Case

6.1 Introduction

We have investigated the 'discrete decompounding' in Chapter 2. The methods developed there are tailored to discrete data which arise in queueing systems. In the context of insurance risk models it is common to work within a continuous model instead of a discrete one. Suppose that the intensity $\lambda > 0$ is known. Assume that only positive damages occur, i.e. P is some probability measure concentrated on the positive real numbers.

Of course, a reasonable estimator for the total claim distribution Q is the empirical $Q^n = \frac{1}{n}\sum_{l=1}^{n} \delta_{Y_l}$ based on the observations of the $Y's$. The naive method is to consider Q^n in the Banach algebra of signed measures and to take the logarithm. This would be the same method that was quite fertile in the discrete case. This approach fails here for general reasons. On the one hand, the exponential function is locally invertible in the space of signed measures near the true distribution Q. On the other hand, note that Q^n converges to Q typically in norms connected to the sup norm or weighted variants (see for example [Sh86]). However, if Q has an absolute continuous part with density $h(x)$, then $\|Q^n - Q\|_{TV} \geq \int h(x)dx$ for all n. Hence there is no convergence in total variation norm and a simple application of the inverse mapping theorem is not possible.

Since we know that the existence of a logarithm depends on some topological property of the Gelfand transform, the null-homotopy in $\mathbb{C}^*$, we could reduce our claims to the weak convergence. It turns out that this is hopeless too. First the maximal ideal space is not satisfactorily known (for papers discussing the maximal ideal space in the measure algebra see [Ro79], [Ta73]). Furthermore, even the known part does not behave in the right way. To have null-homotopy of the whole Gelfand transform of Q^n for n large, we need to have uniform convergence of the Gelfand transforms to the Gelfand transform of Q. Consider the empirical characteristic function $\theta \longmapsto \frac{1}{n}\sum_{l=1}^{n} e^{i\theta Y_l}$ which represents a part of the Gelfand transform of Q^n. It is well known that the characteristic function of a distribution with an absolute continuous part does not converge uniformly on the whole line (see [Fe77]). At best, it converges uniformly on some compact sets whose end points spread out over the real line like $o((n/\log\log n)^{1/2})$ (see [Cs81]).

Therefore the functional analytical and statistical aspects of our estimation problem do not match. We will try a more direct approach. Let us be a little bit informal. Decompose Q into two measures, i.e. $Q = e^{-\lambda}\delta_0 + NQ$ with $NQ(\cdot) := Q(\cdot \cap (0,\infty))$. We have the equalities

$$\delta_0 + e^{\lambda} NQ = \exp(\lambda)Q = \exp(\lambda\delta_0) * \exp(\lambda(P - \delta_0)) = \exp(\lambda P).$$

Take the logarithm on both sides. The right-hand side simplifies to λP. For the left-hand side this is not clear, hence write down the usual power series expansion about δ_0, the identity. This yields

$$\sum_{k=1}^{\infty} (-1)^{k+1} \frac{\exp(\lambda k)}{k} (NQ)^{*k} = \lambda P.$$

We have found the following representation for P

$$(\dagger) \qquad P = \Lambda(Q) := \frac{1}{\lambda} \sum_{k=1}^{\infty} (-1)^{k+1} \frac{\exp(\lambda k)}{k} (NQ)^{*k}.$$

We can read this identity in terms of distribution functions too. If F^n is the empirical distribution based on a sample $Y_1, \ldots, Y_n$ then $\Lambda(\hat{F}^n)$ defines a plug-in estimator for the distribution function of P.

First note that $\Lambda(\mu)$ is some well defined signed measure for all measures μ with $\|\mu\|_{TV} < e^{-\lambda}$. However, plugging in the probability distribution Q, we need $\|NQ\|_{TV}$ to be smaller than $e^{-\lambda}$ to have convergence in total variation norm. Since $\|NQ\|_{TV} = 1 - e^{-\lambda}$, this is smaller than $e^{-\lambda}$ iff $\lambda < \log 2$. If $\lambda \geq \log 2$ is satisfied then the right hand side of $(*)$ converges neither in total variation norm nor in sup norm (again, identify Q with its distribution function). On the other hand, if we restrict ourselves to uniform convergence on compact sets then we should not be surprised to have convergence. Indeed: Since we are in the one-sided case, increasing the power of convolution of NQ will transport more and more mass to infinity, i.e. there is a loss of mass. We will come back to this phenomenon in the next section.

The statistical behaviour of a convolution series can be investigated by the methods of the functional approach, developed for the analysis of the empirical renewal function (see [Gr93]) and the compounding problem (as mentioned in the introduction, see [Pi94]). On the one hand our problem is easier, because we discuss the one-sided case. This simplifies the proofs of the convolution inequalities a good deal. On the other hand we should be more careful dealing with signed measures because of the alternating sign in $(*)$. However, the main problem turns out to be the exponential growth of the factor $\exp(\lambda k)$, $k \in \mathbb{N}$. We will use the method of exponential tilting.

[He97] has used exponential tilting in the computational compounding problem. The idea is to thin out the tails of a measure by an operation T_τ that commutes with convolutions, then work in whatever better settings and perform a backtilting operation S_τ.

To be more precise, for a signed measure μ on the nonnegative reals define $T_\tau \mu$ to be the measure with μ-density $x \longmapsto e^{-\tau x}$. T_τ is the exponential tilting operator, τ is the tilting parameter. It is easy to see that

$$T_\tau(\mu * \nu) = T_\tau \mu * T_\tau \nu$$

for signed measures μ, ν. Obviously, the larger the tilting parameter the thinner the tails of the tilted measure. Moreover and of some importance for us, if the singleton $\{0\}$ has μ-measure 0, even the total variation norm of $T_\tau \mu$ can be made arbitrarily small by choosing τ large enough. We write $\mu_\tau = T_\tau \mu$.

We define an inverse tilting procedure S_τ. The domain of S_τ is the space of signed measures on the nonnegative real numbers. The codomain is some function space. We further define

$$S_\tau \nu(x) := \int_{(0,x]} e^{\tau y} \, \nu(dy).$$

It is easy to check that $S_\tau(T_\tau \nu)(x) = \nu((0,x])$. Therefore S_τ can be viewed as inverse mapping with respect to T_τ if we restrict T_τ to the measures on the positive real numbers. Note that the function $x \longmapsto S_\tau \nu(x)$ is locally of bounded variation. Furthermore, we have

$$\int_{(0,x]} S_\tau(x-y) dS_\tau(\mu) = S_\tau(\nu * \mu)(x).$$

Hence the inversion of exponential tilting is commuting with convolutions as well.

We will see that the following diagram commutes in some sense:

$$\begin{array}{ccc} F & \xrightarrow{T_\tau} & \nu_\tau \\ \Lambda \Big\downarrow & & \Big\downarrow \Lambda \\ G & \xleftarrow{S_\tau} & \mu_\tau \end{array} .$$

In summary the procedure of proving limit theorems is the following: Choose a tilting parameter τ large enough to make $\|T_\tau N Q^n\|_{TV}$ smaller then $e^{-\lambda}$. Hence $\Lambda(T_\tau Q^n)$ defines a signed measure. Then apply the functional approach to the convolution series. Finally, tilt back.

6.2 Inequalities

This section provides some inequalities. Everything is based on integration by parts. Let us introduce some notation. If F is a distribution function with $F(0-) = 0$ then define $NF(x) := F(x) - F(0)$. NF is the distribution function of some subprobability measure. This definition is consistent with the definition already made for Q in the previous section.

Consider the space D of cadlag functions on the nonnegative real numbers, i.e. the space of bounded right-continuous functions having left-hand limits and being left-continuous at ∞. Equipped with the usual sup norm $\|f\|_\infty := \sup_{x\geq 0} |f(x)|$, this is a Banach space and the natural habitat of distribution functions. We introduce a weighted variant of it. Fix some $\epsilon \in \mathbb{R}$. We define a weighted sup norm for functions $f : [0, \infty) \to \mathbb{R}$,

$$\|f\|_\epsilon := \sup_{x\geq 0} e^{\epsilon x} |f(x)|.$$

Define D_ϵ to be the space of right-continuous functions having left-hand limits on the nonnegative real numbers with $\|f\|_\epsilon < \infty$. $(D_\epsilon, \|\cdot\|_\epsilon)$ is a Banach space.

The first inequality deals with exponential tilting. For a distribution function F we write F_τ for the distribution function of the tilted measure induced by F.

Lemma 6.1 *Consider two distribution functions F^1 and F^2 and some $G \in D, \alpha \geq 0$. Assume $\tau > 0$, $\epsilon + \tau > 0$. Define*

$$\overline{NG}_\tau(x) := -e^{-\tau x} G(x) + \tau \int_{(x,\infty)} G(y) e^{-\tau y} dy.$$

i) Then there is some constant $C_1(\epsilon, \tau)$ not depending on F^1, F^2 with

$$\| \, \|NF^1_\tau\|_\infty - NF^1_\tau - \|NF^2_\tau\|_\infty + NF^2_\tau \|_{\tau+\epsilon} \leq C_1(\epsilon, \tau) \|F^2 - F^1\|_\epsilon.$$

ii) Then with the same constant $C_1(\epsilon, \tau)$ not depending on F^1, F^2, G, α

$$\begin{aligned} &\| \, \alpha \, (\|NF^1_\tau\|_\infty - NF^1_\tau - \|NF^2_\tau\|_\infty + NF^2_\tau \,) - \overline{NG}_\tau \, \|_{\tau+\epsilon} \\ &\qquad \leq \; C_1(\epsilon, \tau) \| \, \alpha \, (\, F^1 - F^2 \,) \, - G \, \|_\epsilon. \end{aligned}$$

Proof: i) The proof is based on integration by parts:

$$\begin{aligned}
&\int_{(x,\infty)} e^{-\tau y}\, NF^1(dy) - \int_{(x,\infty)} e^{-\tau y}\, NF^2(dy) \\
&\quad = \; e^{-\tau y} NF^1(y)|_x^\infty + \tau \int_{(x,\infty)} NF^1(y) e^{-\tau y}\, dy \\
&\qquad - e^{-\tau y} NF^2(y)|_x^\infty - \tau \int_{(x,\infty)} NF^2(y) e^{-\tau y}\, dy \\
&\quad = \; \left(-NF^1(x) + NF^2(x)\right)\, e^{-\tau x} + \tau \int_{(x,\infty)} e^{-\tau y} (NF^1(y) - NF^2(y))\, dy \\
&\quad = \; \left(-F^1(x) + F^2(x)\right)\, e^{-\tau x} + \tau \int_{(x,\infty)} e^{-\tau y} (F^1(y) - F^2(y))\, dy \\
&\qquad + \left(F^1(0) - F^2(0)\right)\, e^{-\tau x} + \tau \int_{(x,\infty)} e^{-\tau y} (-F^1(0) + F^2(0))\, dy \\
&\quad = \; \underbrace{\left(-F^1(x) + F^2(x)\right)\, e^{-\tau x}}_{=:(1)} + \underbrace{\tau \int_{(x,\infty)} e^{-\tau y}\, (F^1(y) - F^2(y))\, dy}_{=:(2)}\,.
\end{aligned}$$

We have the inequalities

$$\begin{aligned}
|(2)| \;&\leq\; \tau \int_{(x,\infty)} e^{-\tau y} |F^1(y) - F^2(y)|\, dy \\
&\leq\; \|F^1 - F^2\|_\epsilon \tau \int_{(x,\infty)} e^{-(\tau+\epsilon)y}\, dy \\
&\leq\; \frac{\tau}{\tau+\epsilon} e^{-(\tau+\epsilon)x} \|F^1 - F^2\|_\epsilon
\end{aligned}$$

and

$$|(1)| \leq e^{-(\tau+\epsilon)x} \|F^1 - F^2\|_\epsilon.$$

Putting all together we get the desired inequality with $C_T(\epsilon,\tau) := 1 + \frac{\tau}{\tau+\epsilon}$.

ii) Again with integration by parts,

$$\alpha(\|NF^1_\tau\|_\infty - NF^1_\tau(x) - \|NF^2_\tau\|_\infty + NF^2_\tau(x)) + e^{-\tau x}G(x) - \tau\int_{(x,\infty)} G(y)e^{-\tau y}dy$$

$$= \alpha\left(\int_{(x,\infty)} e^{-\tau y}NF^1(dy) - \int_{(x,\infty)} e^{-\tau y}NF^2(dy)\right)$$

$$+ e^{-\tau x}G(x) - \tau\int_{(x,\infty)} G(y)e^{-\tau y}dy$$

$$= \alpha\Bigg(e^{-\tau y}NF^1(y)|_x^\infty - e^{-\tau y}NF^2(y)|_x^\infty$$

$$+ \tau\int_{(x,\infty)} e^{-\tau y}NF^1(dy) - \tau\int_{(x,\infty)} e^{-\tau y}NF^2(dy)\Bigg)$$

$$+ e^{-\tau x}\, G(x) - \tau\int_{(x,\infty)} G(y)\, e^{-\tau y}dy.$$

Therefore we obtain

$$\begin{aligned} &\alpha\big(\|NF^1_\tau\|_\infty - NF^1_\tau(x) - \|NF^2_\tau\|_\infty + NF^2_\tau(x)\big) + e^{-\tau}G(x) - \tau\int_{(x,\infty)} G(y)e^{-\tau y}dy \\ = \quad & -e^{-\tau x}\big(\alpha(F^1(x) - F^2(x)) - G(x)\big) + \tau\int_{(x,\infty)} e^{-\tau y}\Big[\alpha\big(F^1(y) - F^2(y)\big) - G(y)\Big]dy. \end{aligned}$$

This yields the inequality

$$\begin{aligned} &\left|\alpha(\|NF^1_\tau\|_\infty - NF^1_\tau(x) - \|NF^2_\tau\|_\infty + NF^2_\tau(x)) + e^{-\tau}G(x) - \tau\int_{(x,\infty)} G(y)e^{-\tau y}dy\right| \\ = \quad & \left|-e^{-\tau x}(\alpha(F^1(x) - F^2(x)) - G(x)) + \tau\int_{(x,\infty)} e^{-\tau y}\Big[\big(\alpha(F^1(y) - F^2(y)\big) - G(y)\Big]\, dy\right| \\ \leq \quad & e^{-(\tau+\epsilon)x}\|\alpha(F^1 - F^2) - G\|_\epsilon\left(1 + \frac{\tau}{\tau+\epsilon}\right). \end{aligned}$$

The desired inequality is now an easy consequence.□

We now consider the back-tilting operation.

Lemma 6.2 *Suppose μ to be a signed measure on the nonnegative real numbers. Suppose $G \in D$. Let $\tau > \tau' > 0$. Define*

$$\tilde{S}_\tau G(x) := -e^{\tau x}G(x) + G(0) + \int_{(0,x]} \tau e^{\tau y}G(y)\; dy.$$

Then there exists a constant $C_S(\tau, \tau')$ not depending on μ and G with

$$\begin{aligned} \|S_\tau\nu(\cdot)\|_{\tau'-\tau} \quad &\leq \quad C_S(\tau,\tau')\|\nu(\cdot,\infty)\|_{\tau'}, \\ \|S_\tau\nu(\cdot) - \tilde{S}_\tau G\|_{\tau'-\tau} \quad &\leq \quad C_S(\tau,\tau')\|\nu(\cdot,\infty) - G\|_{\tau'}. \end{aligned}$$

Proof: The first inequality follows from the second with $G \equiv 0$.

If F with $F(\infty) := \lim_{x\to\infty} F(x)$ is the distribution function of some positive measure μ then an integration by parts yields

$$S_\tau\mu(x) = e^{\tau x}(F(x) - F(\infty)) - e^{\tau 0}(F(0) - F(\infty)) - \int_{(0,x]} \tau e^{\tau y}(F(y) - F(\infty))\, dy.$$

Since an arbitrary signed measure ν can be decomposed into $\nu = \nu^+ - \nu^-$ with nonnegative measures ν^+ and ν^-, linearity yields

$$S_\tau\mu(x) = e^{\tau x}\nu((x,\infty)) - e^{\tau 0}\nu((0,\infty)) - \int_{(0,x]} \tau e^{\tau y}\nu((y,\infty))\, dy.$$

Hence

$$\begin{aligned}
& |S_\tau\nu(x) - \tilde{S}_\tau G| \\
\leq\ & e^{\tau x}e^{-\tau' x}\|\nu((\cdot,\infty)) - G\|_{\tau'} + e^{\tau 0}e^{-\tau' 0}\|\nu((\cdot,\infty)) - G\|_{\tau'} \\
& + \|\nu((\cdot,\infty)) - G\|_{\tau'} \int_{(0,x]} \tau e^{\tau y}e^{-\tau' y}\, dy \\
\leq\ & \underbrace{\left(2 + \frac{\tau}{\tau - \tau'}\right)}_{=:C_S(\tau,\tau')} \|\nu((\cdot,\infty)) - G\|_{\tau'}.
\end{aligned}$$

This proves the assertion. □

Note that S_τ is a continuous operator that maps the signed measures, topologized with the total variation norm, to $(D_{-\tau}, \|\cdot\|_{-\tau})$.

We need a definition for convolutions $h * H$ if H is the distribution function of some positive measure ν on the nonnegative reals and $G \in D_\tau$, $\tau \in \mathbb{R}$. If G is locally bounded and the function $[0,x] \ni y \longmapsto G(x-y)1_{[0,x]}(y)$ is measurable and H-integrable then we define $G*H(x) := G*\nu(x) := \int G(x-y)1_{[0,x]}(y)\, H(dy)$. Note that $x \mapsto G * H(x)$ is cadlag as well, at least on $[0,\infty)$.

The next lemma provides a convolution inequality which bounds the exponential decay of a convolution by that of its factors.

Lemma 6.3 *Let H be the distribution function of a positive finite measure with $H(0) = 0$. Let h be a function such that $h * H(x)$ is well defined for all $x \geq 0$. Assume $\tau_1, \tau_2 > 0$ with $\tau_1 \neq \tau_2$. Then there is a constant $C'(\tau_1,\tau_2)$ not depending on H or h such that for all $\tau \leq \tau_1 \wedge \tau_2$*

$$\|h * H\|_\tau \leq C'(\tau_1,\tau_2)\|h\|_{\tau_1}\left(\|H\|_\infty + \|\,\|H\|_\infty - H\|_{\tau_2}\right).$$

Proof: Without loss of generality we may assume the norms at the right-hand side of the inequalities to be finite. We note

$$\begin{aligned}
(*) \qquad |h(x-y)| &\leq \exp(-\tau_1(x-y))\|h\|_{\tau_1}, \\
(**) \qquad -H(x) &\leq \exp(-\tau_2 x)\|\,\|H\|_\infty - H\|_{\tau_2} - \|H\|_\infty.
\end{aligned}$$

Integration by parts yields

$$
\begin{aligned}
& \left| \int h(x-y) 1_{[0,x]}(y) H(dy) \right| \\
\overset{(*)}{\leq} \quad & \|h\|_{\tau_1} \int \exp(-\tau_1(x-y)) 1_{[0,x]}(y) H(dy) \\
= \quad & \|h\|_{\tau_1} \left(H(x) - \tau_1 \int \exp(-\tau_1(x-y)) H(y) 1_{[0,x]}(y) dy \right) \\
\overset{(**)}{\leq} \quad & \|h\|_{\tau_1} \left(H(x) + \| \, \|H\|_\infty - H\|_{\tau_2} \tau_1 \exp(-\tau_1 x) \int_{[0,x]} \exp((\tau_1 - \tau_2) y) dy \right. \\
& \left. - \|H\|_\infty \tau_1 \exp(-\tau_1 x) \int_{[0,x]} \exp(\tau_1 y) dy \right) \\
= \quad & \|h\|_{\tau_1} \big(H(x) - \|H\|_\infty \exp(-\tau_1 x)(\exp(\tau_1 x) - 1) \\
& + \| \, \|H\|_\infty - H\|_{\tau_2} \tau_1 \exp(-\tau_1 x) \frac{1}{\tau_1 - \tau_2} (\exp((\tau_1 - \tau_2)x) - 1) \big) \\
\leq \quad & \|h\|_{\tau_1} \big(\exp(-\tau_1 x) \|H\|_\infty \\
& + \| \, \|H\|_\infty - H\|_{\tau_2} \frac{\tau_1}{\tau_1 - \tau_2} (\exp(-\tau_2 x) - \exp(-\tau_1 x)) \big)
\end{aligned}
$$

Consider some $\tau \leq \tau' := \tau_1 \wedge \tau_2$. Then

$$
\begin{aligned}
& \exp(\tau x) |h * H|(x) \leq \exp(\tau' x) |g * \nu|(x) \\
& \leq \|g\|_{\tau_1} \left(\|H\|_\infty + \| \, \|H\|_\infty - H\|_{\tau_2} \underbrace{\frac{\tau_1}{\tau_1 - \tau_2} (\exp(-(\tau_2 - \tau')x - \exp(-(\tau_1 - \tau')x))}_{\leq \frac{\tau_1}{|\tau_1 - \tau_2|}} \right).
\end{aligned}
$$

Hence the assertion holds with $C'(\tau_1, \tau_2) = 1 \vee \frac{\tau_1}{|\tau_1 - \tau_2|}$.□

Later on we need another useful bound:

Lemma 6.4 *Let H be the distribution function of a positive measure. Then for every $\tau \geq 0$*

$$
\| \, \|H\|_\infty - H\|_\tau \leq \int_{[0,\infty)} \exp(\tau x) \, dH(x).
$$

Proof:

$$
\begin{aligned}
\sup_{x \geq 0} \exp(\tau x)(\|H\|_\infty - H(x)) \quad &= \quad \sup_{x \geq 0} \exp(\tau x) \int_{(x,\infty)} dH(y) \\
&\leq \quad \sup_{x \geq 0} \int_{(x,\infty)} \exp(\tau y) dH(y) \leq \int_{[0,\infty)} \exp(\tau y) dH(y). \ \square
\end{aligned}
$$

Remark: The operator $\tilde{S}_\tau$ is the inverse operator of $G \mapsto \overline{NG_\tau}$ in some sense. Indeed: $G \in D$ satisfies

$$\begin{aligned}
\tilde{S}_\tau \overline{NG_\tau}(x) &= -\tilde{S}_\tau e^{-\tau\cdot} G(\cdot)(x) + \tilde{S}_\tau \left[\int_{(\cdot,\infty)} \tau e^{-\tau s} G(s)\, ds \right](x) \\
&= -\left(-e^{\tau x} e^{-\tau x} G(x) + G(0) + \int_{(0,x]} \tau e^{\tau y} e^{-\tau y} G(y)\, dy \right) \\
&\quad - e^{\tau x} \int_{(x,\infty)} \tau e^{-\tau s} G(s)\, ds + \int_{(0,\infty)} \tau e^{-\tau s} G(s)\, ds \\
&\quad + \int_{(0,x]} \tau e^{\tau y} \int_{(y,\infty)} \tau e^{-\tau s} G(s)\, ds\, dy \\
&= G(x) - G(0) - \int_{(0,x]} \tau G(s)\, ds - e^{\tau x} \int_{(x,\infty)} \tau e^{-\tau s} G(s)\, ds \\
&\quad + \int_{(0,\infty)} \tau e^{-\tau s} G(s)\, ds + \left[e^{\tau y} \int_{(y,\infty)} \tau e^{-\tau s} G(s)\, ds \right]_{(0,x]} \\
&\quad + \int_{(0,x]} e^{\tau y} \tau e^{-\tau y} G(y)\, dy \\
&= G(x) - G(0) =: NG(x)\ .
\end{aligned}$$

There is also some compatibility with convolutions. It holds that

$$\tilde{S}_\tau \left(G * F\right)(x) = (\tilde{S}_\tau G) * (S_\tau F)(x).$$

If G is some distribution function of a positive measure, this is again an integration by parts. For an arbitrary $G \in D$ this follows using the linearity and a density argument.

6.3 Exponential Mass Loss and Exponential Tilting

For a probability distribution Q concentrated on the nonnegative real numbers we define

$$\tau^*(Q) := \inf\{t \geq 0 : \ \widehat{NQ}(t) < e^{-\lambda}\},$$

with the Laplace transform $\widehat{NQ}(t) = \int e^{-tx} NQ(dx)$. Note that $\tau^*(Q)$ is a well defined nonnegative real number since Lebesgue's theorem implies $\lim_{\tau\to\infty} \widehat{NQ}(\tau) = 0$. Define $\tau^*(F)$ analogously for a distribution function F with $F(0-) = 0$.

The next theorem investigates the connections between the different ingredients.

Theorem 6.5 *Let F be some distribution function.*

i) The series $\Lambda(F)$ is absolutely converging in $\|\cdot\|_{-\tau}$-norm for all $\tau > \tau^*(F)$. Hence $\Lambda(F) \in D_{-\tau}$.
ii) Assume that $Q = e^{\lambda(P-\delta_0)}$ holds for a probability measure P concentrated on the positive real numbers. Then for every $\tau > \tau^*(Q)$ the Laplace transform satisfies

$$\hat{P}(\tau) = \frac{1}{\lambda} \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} e^{\lambda k} \left(\widehat{(NQ)}(\tau)\right)^k .$$

iii) It holds that $S_\tau \circ \Lambda \circ T_\tau(F) = \Lambda(F)$ for all distribution functions F and $\tau > \tau^*(F)$.
iv) If F and G are distribution functions with $F(0-) = 0$ and $G(0) = 0$ then

$$F = \sum_{l=0}^{\infty} e^{-\lambda} \frac{\lambda^l}{l!} G^{*l} \Rightarrow F = \Lambda(G).$$

v) Let $G \in D$. If F is a distribution function then $\Lambda'_F G := \frac{1}{\lambda} \sum_{k=1}^{\infty} e^{\lambda k} (NG) * (NF)^{*(k-1)}$ converges in $D_{-\tau}$ for every $\tau > \tau^*(F)$. Λ'_F defines a bounded linear operator $G \mapsto \Lambda'_F G$ with domain D and codomain $D_{-\tau}$. Furthermore, the decomposition

$$\Lambda'_F G = \tilde{S}_\tau \frac{1}{\lambda} \sum_{k=1}^{\infty} (-1)^{k+1} e^{\lambda k} \overline{NG}_\tau * NF_\tau^{*k-1}$$

is valid with $F^{*0} := 1_{[0,\infty)}$.

Proof: i) For $k \in \mathbb{N}_0$ and $\tau > 0$ we have the inequality

$$\begin{aligned} (NF)^{*k}(x) &= \int 1_{(0,x]}(t) d(NF)^{*k} \\ &\leq \exp(\tau x) \left(\int \exp(-\tau t) NF(dt) \right)^k = \exp(\tau x) \left(\widehat{NF}(\tau) \right)^k . \end{aligned}$$

For $\tau > \tau^*(F)$ we have

$$\|(NF)^{*k}\|_{-\tau} \leq \left(\widehat{NF}(\tau)\right)^k < e^{-k\lambda}.$$

Hence $\Lambda(F)$ is an absolutely convergent series with respect to the $\|\cdot\|_{-\tau}$-norm. Since $D_{-\tau}$ is a Banach space, $\Lambda(F)$ converges to some $G \in D_{-\tau}$.
ii) This is the same calculation as has been carried out for the derivation of (†) above. For $\tau \geq 0$ we obtain

$$e^{-\lambda} + \widehat{(NQ)}(\tau) = \hat{Q}(\tau) = \int e^{-\tau y}\, Q(dy) = \exp\left(\lambda \left(\int e^{-\tau y} P(dy) - 1 \right)\right) = \exp(\lambda(\hat{P}(\tau) - 1)).$$

Multiplying both sides of the equation with $\exp(\lambda)$ yields

$$1 + e^{\lambda}\widehat{NQ}(\tau) = \exp(\lambda\hat{P}(\tau)).$$

For all $\tau > \tau^*(Q)$ we again have $\exp(\lambda)\widehat{NQ}(\tau) < 1$. So we can take the logarithm on both sides. Writing down the power series for the left hand side we see that the assertion follows.

iii) Remember that T_τ and S_τ commute with convolutions. Furthermore, $S_\tau T_\tau NF = NF$. Therefore for the finite sum:

$$S_\tau \frac{1}{\lambda}\sum_{k=1}^{N}\frac{(-1)^{k+1}}{k}e^{k\lambda}(T_\tau NF)^{*k} = \frac{1}{\lambda}\sum_{k=1}^{N}\frac{(-1)^{k+1}}{k}e^{k\lambda}(NF)^{*k}.$$

The right-hand side converges to $\Lambda(F)$ in $D_{-\tau}$. Hence in $D_{-\tau}$

$$\lim_{N\to\infty} S_\tau \frac{1}{\lambda}\sum_{k=1}^{N}\frac{(-1)^{k+1}}{k}e^{k\lambda}(T_\tau NF)^{*k} = \Lambda(F).$$

Obviously,

$$\lim_{n\to\infty}\frac{1}{\lambda}\sum_{k=1}^{N}\frac{(-1)^{k+1}}{k}e^{k\lambda}(T_\tau NF)^{*k} = \frac{1}{\lambda}\sum_{k=1}^{\infty}\frac{(-1)^{k+1}}{k}e^{k\lambda}(T_\tau NF)^{*k}.$$

in total variation norm. Since S_τ is a continuous operator, mapping signed measures to elements of $D_{-\tau}$, we obtain

$$\begin{aligned}\Lambda(F) &= \lim_{N\to\infty} S_\tau \frac{1}{\lambda}\sum_{k=1}^{N}\frac{(-1)^{k+1}}{k}e^{k\lambda}(T_\tau NF)^{*k}\\ &= S_\tau \lim_{N\to\infty}\frac{1}{\lambda}\sum_{k=1}^{N}\frac{(-1)^{k+1}}{k}e^{k\lambda}(T_\tau NF)^{*k} = S_\tau\circ\Lambda\circ T_\tau(F).\end{aligned}$$

iv) Write G, F for the measures associated with G, F. Let $F = \sum_{l=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!}G^{*k}$. Part ii) of the theorem shows that $\widehat{T_\tau G}(\tau') = \widehat{\Lambda(F_\tau)}(\tau')$ for all $\tau' \geq 0$ and $\tau > \tau^*(G)$. Hence the measures $T_\tau G$ and $\Lambda(G_\tau)$ coincide, since their Laplace transforms do. Applying the backtilting operator, we have $S_\tau T_\tau G = S_\tau \Lambda(T_\tau F) = \Lambda(F)$.

v) Just note that $G * NF^{*(k-1)}(x) \leq \|G\|_\infty NF^{*(k-1)}$. Hence proving the convergence of Λ'_F in $D_{-\tau}$ follows with the same reasoning as in i). The rest follows like as iii) using the convolution compatibility of the operators $\tilde{S}_\tau, S_\tau, G \mapsto \overline{NG}_\tau$.□

Remarks: Part i) explains that $\Lambda(F)$ can be understood as a member of the weighted Banach space $D_{-\tau}$. Convergence of $\Lambda(F)$ on compact sets is included here. Part iii) justifies the diagram noted in the first section. Part iv) shows that Λ is the logarithm of a distribution function up to an affine transformation, hence the desired object.

The operator $\Lambda'_F G$ defines the derivative of Λ at F in direction G. The differentiability of the functional Λ is discussed in the next section. The additional remark in v) shows that we also have a commuting diagram for the derivative of Λ, i.e.

$$\begin{array}{ccccc} & G & \xrightarrow{\overline{N\cdot_\tau}} & \overline{NG_\tau} & \\ \Lambda'_F & \Big\downarrow & & \Big\downarrow & \Lambda'_{F_\tau} \\ & \Lambda'_F G & \xleftarrow{\tilde{S}_\tau} & \Lambda'_{F_\tau}\overline{NG_\tau} & . \end{array}$$

The series defining the operator $\Lambda'_{F_\tau} G := \sum_{k=1}^{\infty} e^{\lambda k} G * NF_\tau^{*(k-1)}$ again converges in a stronger norm than its opponent Λ'_F.

6.4 Consistency and Asymptotic Normality

We want to show continuity and differentiability properties for Λ. First note the following approximation result.

Lemma 6.6 *Let $G \in D_\tau$ and $\tau' < \tau$ for some $\tau' > 0$. For every $\epsilon > 0$ there is a linear combination of indicator functions $g = \sum_{l=1}^{N} \alpha_l 1_{[a_l,b_l)}$ with $\alpha_i \in \mathbb{R}$, $0 \leq a_i < b_i < \infty$, $i = 1, \ldots, N$ such that $\|G - g\|_{\tau'} < \epsilon$ holds.*

Proof: Suppose that $D([0,S])$ is the space of cadlag functions on $[0,S]$. Then the set $I_S := \{\sum_{l=1}^{N} \alpha_l 1_{[a_l,b_l)} : \ \alpha_i \in \mathbb{R},\ 0 \leq a_i < b_i < \infty,\ i = 1, \ldots, N\}$ is a dense subset of $D[0,S]$ with respect to the sup norm. Fix some ϵ. For $S > 0$ we have

$$\sup_{x \geq S} e^{\tau' x} |G(x)| \leq \|G\|_\tau \sup_{x \geq S} e^{(\tau'-\tau)x} \leq \|G\|_\tau e^{(\tau'-\tau)S}.$$

Hence if S has been chosen large enough then $\sup_{x \geq S} e^{\tau' x} |G(x)| \leq \epsilon/2$ holds. For such an S there is a linear combination of indicator functions $g \in I_S$ with

$$\sup_{0 \leq x \leq S} |g(x) - G(x)\| \leq e^{-\tau' S} \epsilon/2.$$

Then

$$\|g - G\|_{\tau'} \leq \sup_{0 \leq x \leq S} e^{\tau' x} |g(x) - G(x)| + \sup_{x \geq S} e^{\tau' x} |G(x)| \leq \epsilon/2 + \epsilon/2.$$

This proves our assertion.$\Box$

Lemma 6.7 *Suppose that μ is a signed measure on the nonnegative reals and that $\tau > 0$. Assume that $g = \sum_{l=1}^{N} \alpha_l 1_{[a_l,b_l)}$ with $\alpha_i \in \mathbb{R}$, $0 \le a_i < b_i < \infty$, $i = 1, \dots, N$. Then there exists a constant $C(g,\tau)$ such that*

$$\|g * \mu\|_\tau \le C(g,\tau)\|\mu((\cdot,\infty))\|_\tau \,.$$

Proof: First assume that $g = 1_{[a,b)}$. Then

$$g * \mu(x) = \int_{(0,x]} 1_{[a,b)}(x-y)\, d\mu(y) = \begin{cases} 0, & \text{if} \quad x < a, \\ \mu((0,x-a]), & \text{if} \quad x \le b, \\ \mu((x-b,x-a]), & \text{else.} \end{cases}$$

We obtain

$$\begin{aligned} \sup_{a \le x \le b} e^{\tau x}|\mu((0,x-a])| &\le \sup_{a \le x \le b} e^{\tau x}(|\mu((0,\infty))| + |\mu((a,\infty))|) \\ &\le (e^{\tau b} + e^{\tau a})\|\mu((\cdot,\infty))\|_\tau \end{aligned}$$

and

$$\begin{aligned} \sup_{x \ge b} e^{\tau x}|\mu((x-b,x-a])| &\le \sup_{x \ge b} e^{\tau x}(|\mu((x-b,\cdot))| + |\mu((x-a,\infty))|) \\ &\le (e^{\tau b} + e^{\tau a})\,\|\mu((\cdot,\infty))\|_\tau. \end{aligned}$$

Hence $\|g\|_\tau \le (e^{\tau a} + e^{\tau b})\|\mu((\cdot,\infty))\|_\tau$. The general case is treated using the triangle inequality. $\square$

Theorem 6.8 *Let F^n, F, $n \in \mathbb{N}$, be distribution functions with $F(0-) = F_n(0-) = 0$. Suppose that $G \in D$. Then the following is true:*
i) If $\|F^n - F\|_\infty \to 0$ for $n \to \infty$ and $\epsilon > \tau^*(F)$ then

$$\|\Lambda(F^n) - \Lambda(F)\|_{-\epsilon} \to 0$$

as $n \to \infty$.
ii) If $\|\sqrt{n}(F^n - F) - G\|_\infty \to 0$ for $n \to \infty$ and $\epsilon > \tau^*(F)$ then

$$\|\sqrt{n}(\Lambda(F^n) - \Lambda(F)) - \Lambda'_F G\|_{-\epsilon} \to 0$$

as $n \to \infty$.

Proof: Remember the diagram in the first section.

i) Choose some τ, ϵ with $\tau > \epsilon > \tau^*(F)$. An integration by parts yields the following inequality

$$\left| \int e^{-\sigma y} dNF^n - \int e^{-\sigma y} dNF \right| \le 2\|F^n - F\|_\infty.$$

This shows that $\int e^{-\sigma y} dNF^n < e^{-\lambda}$ for n large enough and $\sigma > \tau^*(F)$. In particular, $\Lambda(F^n) = S_\tau \nu_\tau(F^n)$ for n large enough. Define $\nu_\tau(F^n)$ and $\nu_\tau(F)$ to be the signed measures given by

$$\nu_\tau(F^n) := \frac{1}{\lambda}\sum_{k=1}^{\infty}\frac{e^{k\lambda}}{k}(-1)^k(NF_\tau^n)^{*k}, \qquad \nu_\tau(F) := \frac{1}{\lambda}\sum_{k=1}^{\infty}\frac{e^{k\lambda}}{k}(-1)^k(NF_\tau)^{*k}.$$

Let $\tau' := \tau - \epsilon$. We want to show that

$$(A1) \qquad \|\nu_\tau(F^n)((\cdot,\infty)) - \nu_\tau(F)((\cdot,\infty))\|_{\tau'} \to 0.$$

We define $H_{n,k} := \sum_{l=0}^{k-1}(NF_\tau^n)^{*l} * (NF_\tau)^{*(k-l-1)}$ and use the simple trick

$$(NF_\tau^n)^{*k} - (NF_\tau)^{*k} = (NF_\tau^n - NF_\tau) * H_{n,k}.$$

Now consider the tails. Let μ_n, μ be the measure associated with NF_τ^n, NF_τ. Then we obtain

$$\begin{aligned}
&\|(NF_\tau^n)^{*k}\|_\infty - (NF_\tau)^{*k}(x) - (\,\|(NQ_\tau)^{*k}\|_\infty - (NF_\tau)^{*k}(x))\\
&= \mu_1^{*k}((x,\infty)) - \mu_2^{*k}((x,\infty))\\
&= \int_{[0,x]} \mu_1((x,\infty)-y) - \mu_2((x,\infty)-y))\, H_{n,k}(dy)\\
&= \int_{[0,x]} \|NF_\tau^n\|_\infty - NF_\tau^n(x-y) - \|NF_\tau\|_\infty + NF_\tau(x-y)\, H_{n,k}(dy).
\end{aligned}$$

First apply the convolution inequality, lemma 6.3, and then lemma 6.1:

$$\begin{aligned}
&\|\,\|(NF_\tau^n)^{*k}\|_\infty - (NF_\tau^n)^{*k} - (\,\|(NF_\tau)^{*k}\|_\infty - (NF_\tau)^{*k})\|_{\tau'}\\
&\le C'(\tau',\tau)\, \|\,\|NF_\tau^n\|_\infty - NF_\tau^n - \|NF_\tau\|_\infty + NF_\tau\|_\tau \times\\
&\qquad \times (\,\|H_{n,k}\|_\infty + \|\|H_{n,k}\|_\infty - H_{n,k}\|_{\tau'})\\
&\le C'(\tau',\tau)\, C_1(0,\tau)\, \|F^n - F\|_\infty\, (\,\|H_{n,k}\|_\infty + \|\|H_{n,k}\|_\infty - H_{n,k}\|_{\tau'})\,.
\end{aligned}$$

Note that $\|F^n - F\|_\infty$ converges to zero and that

$$\begin{aligned}
&\|\nu_\tau(F^n)(\cdot,\infty) - \nu_\tau(F)(\cdot,\infty)\|_{\tau'}\\
&\le D\, \|F^n - F\|_\infty\, \frac{1}{\lambda}\sum_{k=1}^{\infty}\frac{\exp(\lambda k)}{k}\left(\|H_{n,k}\|_\infty + \|\|H_{n,k}\|_\infty - H_{n,k}\|_{\tau'}\right)
\end{aligned}$$

with $D := C'(\tau',\tau)C_1(0,\tau)$.

Hence to establish $(A1)$ it is sufficient to show that

$$\sum_{k=1}^{\infty}\frac{\exp(\lambda k)}{k}\left(\|H_{n,k}\|_\infty + \|\|H_{n,k}\|_\infty - H_{n,k}\|_{\tau'}\right)$$

remains bounded as $n \to \infty$. Use lemma 6.4 to obtain

$$\begin{aligned} &\| \|H_{n,k}\|_\infty - H_{n,k}\|_{\tau'} \\ &\leq \int \exp(\tau' y) H_{n,k}(dy) \\ &= \sum_{l=0}^{k-1} \left(\int \exp((\tau' - \tau)y) NF^n(dy) \right)^l \left(\int \exp((\tau' - \tau)y) NF dy \right)^{k-l-1}. \end{aligned}$$

Then with the obvious inequality

$$\|H_{n,k}\|_\infty \leq \sum_{l=0}^{k-1} \|NF_\tau^n\|_\infty^l \|NF_\tau\|_\infty^{k-l-1}.$$

we have the upper bound

$$\begin{aligned} &\|H_{n,k}\|_\infty + \| \|H_{n,k}\|_\infty - H_{n,k}\|_{\tau'} \\ &\leq \sum_{l=0}^{k-1} \|NF_\tau^n\|_\infty^l \|NF_\tau\|_\infty^{k-l-1} \\ &+ \sum_{l=0}^{k-1} \left(\int \exp((\tau' - \tau)y) NF^n(dy) \right)^l \left(\int \exp((\tau' - \tau)y) NF dy \right)^{k-1-l}. \end{aligned}$$

By definition of τ and τ',

$$\begin{aligned} \|NF_\tau^n\|_\infty &= \int e^{-\tau y}\, NF^n(dy) \quad \to \int e^{-\tau y} NF(y) < e^{-\lambda}, \\ \int \exp((\tau' - \tau)y) NF^n(dy) &= \int \exp(-\epsilon y) NF^n(dy) \to \int e^{-\epsilon y} NQ(y) < e^{-\lambda}. \end{aligned}$$

Therefore as an easy consequence of Lebesgue's dominated convergence theorem,

$$\begin{aligned} &\limsup_{n\to\infty} \sum_{k=1}^\infty \frac{\exp \lambda k}{k} \left(\|H_{n,k}\|_\infty + \| \|H_{n,k}\|_\infty - H_{n,k}\|_{\tau'} \right) \\ &\leq \sum_{k=1}^\infty e^{\lambda k} \left(\left(\int e^{-\tau y} NF(y) \right)^k + \left(\int e^{-\epsilon y} NF(y) \right)^k \right). \end{aligned}$$

This proves $(A1)$. The assertion i) is then a consequence of the back-tilting inequality, i.e.

$$\|S_\tau \nu_\tau(F^n) - S_\tau(F)\|_{-\epsilon} \leq C_S(\tau, \tau')\, \|\nu_\tau(F^n) - \nu_\tau(F)\|_{\tau'}.$$

ii) Again choose $\epsilon > \tau^*(F)$ and $\tau > \epsilon$. Define $\tau' := \tau - \epsilon$. We then have $\Lambda(F^n) = S_\tau \nu_\tau(F^n)$ for n large enough and $\Lambda(F) = S_\tau \nu_\tau(F)$. Theorem 6.5 v) provides the factorization $\tilde{S}_\tau \Lambda'_{F_\tau} \overline{NG_\tau} = \Lambda'_F G$.

We show that

$$(A2) \quad \|\sqrt{n}\big(\nu_\tau(F^n)((\cdot,\infty)) - \nu_\tau(F)((\cdot,\infty))\big) - \Lambda'_{F_\tau}\overline{NG_\tau} * (NF)^{*(k-1)}\|_{\tau'} \to 0.$$

We define

$$\begin{aligned}
I_1(N,n) &:= \frac{1}{\lambda}\sum_{k=N+1}^{\infty}\frac{e^{\lambda k}}{k}\sqrt{n}\left\|\int_{(0,\cdot]}\|NF^n_\tau\|_\infty - NF^n_\tau(\cdot - y)\right. \\
&\qquad\qquad \left. - \|NF\|_\infty + NF_\tau(\cdot - y)\; H_{n,k}(dy)\right\|_{\tau'}, \\
I_2(N) &:= \frac{1}{\lambda}\sum_{k=N+1}^{\infty} e^{\lambda k}\,\|\,\overline{NG_\tau} * NF_\tau^{*(k-1)}\,\|_{\tau'}, \\
I_3(N,n) &:= \frac{1}{\lambda}\sum_{k=1}^{N}\frac{e^{\lambda k}}{k}\left\|\sqrt{n}\int_{(0,\cdot]}\|NF^n_\tau\|_\infty - NF^n_\tau(x-y)\right. \\
&\qquad\qquad - \|NF\|_\infty + NF_\tau(x-y)\; H_{n,k}(dy) \\
&\qquad\quad \left. - \overline{NG_\tau} * H_{n,k}\right\|_{\tau'}, \\
I_4(N,n) &:= \frac{1}{\lambda}\sum_{k=1}^{N}\frac{e^{\lambda k}}{k}\left\|\overline{NG_\tau} * H_{n,k} - k\overline{NG_\tau} * (NF_\tau)^{*(k-1)}\right\|_{\tau'}.
\end{aligned}$$

With $H_{n,k}$ as in the proof of i) we use the triangle inequality to obtain

$$\begin{aligned}
&\|\sqrt{n}\,(\,\nu_\tau(F^n)((\cdot,\infty)) - \nu_\tau(F)((\cdot,\infty))) - \sum_{k=1}^{\infty} e^{k\lambda}\overline{NG_\tau} * (NF)^{*(k-1)}\,\|_{\tau'} \\
&\qquad\qquad \leq\; I_1(N,n) + I_2(N) + I_3(N,n) + I_4(N,n)
\end{aligned}$$

$I_1(n)$ once again can be estimated by

$$\begin{aligned}
I_1(N,n) &\leq D\,\sqrt{n}\|F^n - F\|\,\frac{1}{\lambda}\sum_{k=N+1}^{\infty}\frac{e^{\lambda k}}{k}\left(\big\|H_{n,k}\big\|_\infty + \big\|\,\|H_{n,k}\|_\infty - H_{n,k}\big\|_{\tau'}\right) \\
&\leq D\,(\,\|G\|_\infty + 1)\;\frac{1}{\lambda}\sum_{k=N+1}^{\infty} e^{\lambda k}\epsilon_1^k
\end{aligned}$$

for n large enough, choosing some ϵ_1 with $\int e^{-\epsilon y}NF(dy) < \epsilon_1 < e^{-\lambda}$. Hence $I_1(n)$ can be made arbitrary small uniformly in n for $n \to \infty$, choosing N large enough.

Using the same inequalities as in the proof of assertion i) we have

$$\begin{aligned}
I_2(N) &\leq D\|G\|_\infty\frac{1}{\lambda}\sum_{k=N+1}^{\infty} e^{\lambda k}\left(\|NF_\tau^{*(k-1)}\|_\infty + \|\,\|NF_\tau^{*(k-1)}\|_\infty - NF_\tau^{*(k-1)}\|_{\tau'}\right) \\
&\leq D\|G\|_\infty\frac{1}{\lambda}\sum_{k=N+1}^{\infty} e^{\lambda k}\left(\left(\int e^{-\tau y}\,NF_\tau(dy)\right)^{k-1} + \left(\int e^{-\epsilon y}\,NF_\tau(dy)\right)^{k-1}\right).
\end{aligned}$$

$I_2(N)$ again can be made arbitrarily small by choosing N large enough.

We have

$$I_3(N,n) \le D\|\sqrt{n}(F^n - F) - G\|_\infty \frac{1}{\lambda}\sum_{k=1}^{N} \frac{e^{\lambda k}}{k} \left(\|H_{n,k}\|_\infty + \|\|H_{n,k}\|_\infty - H_{n,k}\|_{\tau'} \right).$$

Hence for fixed N we obtain $\lim_{n\to\infty} I_3(N,n) = 0$, since the sum stays bounded.

Now we have a closer look on $I_4(N,n)$. Note that $\overline{NG_\tau} \in D_\tau$.

Fix some $\tau_1 \in (\tau', \tau)$. Then $\overline{NG}_\tau$ can be approximated by a linear combination g of indicator funtions in $\|\cdot\|_{\tau_1}$ for $\tau_1 < \tau$ as proved in lemma 6.6, and we obtain

$$\begin{aligned}
&\left\|\overline{NG_\tau} * H_{n,k} - k\overline{NG_\tau} * NF_\tau^{k-1}\right\|_{\tau'} \\
&\le \left\| \left(\overline{NG_\tau} - g\right) * H_{n,k}\right\|_{\tau'} + \left\| g * H_{n,k} - kg * NF_\tau^{k-1} \right\|_{\tau'} \\
&\quad + \left\| k\left(g * NF_\tau^{k-1} - \overline{NG_\tau} * NF_\tau^{k-1}\right)\right\|_{\tau'} \\
&\le C'(\tau_1,\tau')\, \|g - \overline{NG_\tau}\|_{\tau_1} \left(\|H_{n,k}\|_\infty + \|\|H_{n,k}\|_\infty - H_{n,k}\|_{\tau'}\right) \\
&\quad + \|g * H_{n,k} - kg * NF_\tau^{k-1}\|_{\tau'} \\
&\quad + C'(\tau_1,\tau')\, \|g - \overline{NG_\tau}\|_{\tau_1} \left(\|kNF_\tau^{*(k-1)}\|_\infty + \|\|kNF_\tau^{*(k-1)}\|_\infty - kNF_\tau^{*(k-1)}|_{\tau'}\right).
\end{aligned}$$

The first and the third term can be made arbitrarily small by choosing an appropriate approximation g and by the boundedness of the corresponding second factors as done before. Therefore to establish $(A2)$ it is enough to discuss the term in the middle.

Without loss of generality assume $k > 1$. If $\mu_{n,k}$ is the signed measure defined by $H_{n,k} - kNF_\tau^{*(k-1)}$ then we have by lemma 6.7

$$\| g * H_{n,k} - kg * NF_\tau^{k-1} \|_{\tau'} \le C(g,\tau')\, \| \mu((\cdot,\infty)) \|_{\tau'}.$$

Recall that $\mu_{n,k} = \sum_{l=0}^{k-1} (NF_\tau)^{k-l-1} * (NF_\tau^n)^{*l} - kNF_\tau^{*(k-1)}$. Then with

$$\nu_{k,l} = (NF_\tau^n)^{*l} * NF_\tau^{*(k-l-1)} - NF_\tau^{*(k-1)}$$

we obtain on using the triangle inequality ($\nu_{k,0} \equiv 0$)

$$\| \mu_{n,k}((\cdot,\infty)) \|_{\tau'} \le C(g,\tau') \sum_{l=1}^{k} \| \nu_{k,l}((\cdot,\infty)) \|_{\tau'}.$$

Since

$$\begin{aligned}
(NF_\tau^n)^{*l} * NF_\tau^{*(k-l-1)} - NF_\tau^{*(k-1)} &= ((NF_\tau^n)^{*l} - NF_\tau^{*l}) * NF_\tau^{*(k-l-1)} \\
&= (NF_\tau^n - NF_\tau) * H_{n,l-1} * NF_\tau^{*(k-l-1)}
\end{aligned}$$

we have once again the upper bound

$$\begin{aligned}
\| g * H_{n,k} - kg * NF_\tau^{*(k-1)} \|_{\tau'} \\
\leq \; & C(g,\tau')\; D\; \|F^n - F\|_\infty \times \\
& \times \sum_{l=1}^{k} \Bigg(\| H_{n,l-1} * NF_\tau^{*(k-l-1)} \|_\infty \\
& \qquad + \| \|H_{n,l-1} * NF_\tau^{*(k-l-1)}\|_\infty \\
& \qquad\qquad - H_{n,l-1} * NF_\tau^{*(k-l-1)} \|_{\tau'} \Bigg).
\end{aligned}$$

The sum of the terms of the right-hand side remains bounded with $n \to \infty$. Hence $\|g * H_{n,k} - kg * NF_\tau^{k-1}\|_{\tau'}$ tends to zero.

$(A2)$ is established. Tilting back proves assertion ii). □

Now let (Y_i) be an iid-sequence of random variables with distribution function $Q = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} P^{*k}$ for some distribution function P with $P(0) = 0$.

Let $\tau > \tau^*(Q)$. Suppose F^n to be the empirical distribution function associated with $Y_1, \ldots, Y_n$, i.e.

$$F^n(t) := \frac{1}{n} \sum_{l=1}^{n} 1_{[0,t]}(Y_l), \quad t \geq 0.$$

Fix some $\tau > \tau^*(Q)$. Consider the space $D_{-\tau}$ equipped with the σ-algebra induced by the family of pointwise evaluations, i.e. $\pi_t : D_{-\tau} \longrightarrow \mathbb{R}$, $\pi_t(x) := x(t)$.

Define $T_n := \Lambda(F^n)$, if $\int e^{-\tau x}\; NF^n(dy) < e^{-\lambda}$, and $T_n := 1_{[1,\infty)}$, otherwise. Then it is easy to check that T_n is a mapping taking values in D_τ and is measurable (use the same argument as in [Gr93] (p. 1433)). Moreover, since

$$\int e^{-\tau x}\; NF^n(dy) = \frac{1}{n} \sum_{l=1}^{n} e^{-\tau Y_l} 1_{Y_l > 0} \to \int e^{-\tau y}\; NQ(dy) < e^{-\lambda} \qquad \text{a.s.},$$

we have $T_n = \Lambda(F^n)$ for n large enough a.s.. From theorem 6.8 and theorem 6.5 we have the consistency $\|T_n - P\|_{-\tau} \to 0$ a.s.. Note, that $\|T_n - P\|_{-\tau}$ is measurable, since the supremum can be calculated over a countable dense subset.

The asymptotic normality now follows as in [Pi94]. Consider the empirical process $E_n := \sqrt{n}(F^n - Q)$. It is well known that $E_n \xrightarrow{\mathcal{D}} B_Q$ with B_Q denoting the scaled Brownian bridge (see [Po84], p. 97). To be more specific, B_Q is a centered Gaussian process with covariance kernel $EB_{Q(r)}B_{Q(s)} = Q(r \wedge s) - Q(r)Q(s)$. B_Q takes its values a.s. in $C(Q)$, the set of functions in D having all its jumps at the jumps of F. This is a separable subspace of D.

We should mention that $E_n \xrightarrow{\mathcal{D}} B_F$ is defined to mean $Ef(E_n) \xrightarrow{\mathcal{D}} Ef(B_F)$ for all bounded continuous mappings f from $(D, \|\cdot\|_\infty)$ to $\mathbb{R}$ that are measurable as mappings from $(D, \sigma(\pi_t))$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ (see [Po84], p. 65).

Since the distribution of B_Q is concentrated on the separable subset $C(Q)$ of D, we can apply the Skohorod representation theorem (again [Po84], p. 71) to the processes E_n and B_Q to derive the asymptotic normality

$$\sqrt{n}(\Lambda(F^n) - \Lambda(Q)) = \sqrt{n}(\Lambda(F^n) - P) \xrightarrow{\mathcal{D}} G := \Lambda'_Q B_F.$$

Since Λ'_Q is a linear bounded operator from D to $D_{-\tau}$, $\Lambda'_Q B_Q$ is a centered Gaussian random variable on $D_{-\tau}$. Note that $B_Q(0) = 0$ a.s., hence $NB_Q = B_Q$. If $1_{[0,x]}(NQ) * k$ denotes the positive measure with density $1_{[0,x]}$ with respect to the measure $(NQ)^{*k}$ then we have $\|1_{[0,x]}(NQ) * k\|_{TV} \leq e^{-\lambda}$ for k large enough, hence $\nu_x := \frac{1}{\lambda}\sum_{k=1}^{\infty} e^{\lambda k}(-1)^k 1_{[0,x]}(NQ)^{*(k-1)}$ is well defined as a signed measure on $[0, x]$ for every fixed x. The covariance kernel is then easily calculated as

$$\begin{aligned} EG(r)G(s) &= E(\Lambda'_Q B_Q)(r)(\Lambda'_Q B_Q)(s) \\ &= \int\int F((r - z_1) \wedge (s - z_2)) d\nu_r(z_1) d\nu_s(z_1) - \Lambda'_Q(F)(r)\Lambda'_Q(F)(s) \ . \end{aligned}$$

Appendix A

Quadratic Programming

Suppose that Q is a symmetric and positive definite $n \times n$-matrix and $A \in \mathbb{R}^{m\times n}$ is a matrix with rank m. Let $y \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$. The calculation of the projection $\pi_Q(y|K)$ for some $K = \{x \in \mathbb{R}^d : Ax \leq b\}$ is equivalent to solving the following general quadratic program (see [Lu89], p.427-423).

$$(*) \qquad \begin{array}{ll} \text{minimize} & \frac{1}{2}x^TQx + x^Tc \\ \text{subject to} & Ax \leq b. \end{array}$$

with $c := -Qy$. First we should solve the same problem with equality constraints, i.e.

$$(\text{EQC}) \qquad \begin{array}{ll} \text{minimize} & \frac{1}{2}x^TQx + x^Tc \\ \text{subject to} & Ax = b. \end{array} .$$

The Lagrange necessary conditions for this problem are

$$\begin{pmatrix} Q & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} -c \\ b \end{pmatrix}.$$

The vector λ is the vector of Lagrange mutlipliers.

Under the assumptions on Q and A the matrix on the left-hand side is invertible. The solution x of the system of linear equations is the solution of (EQC).

The quadratic program $(*)$ can be solved by the active set method. For a set $W \subset \{1, \ldots, m\}$ cancel every row of A with $i \notin W$. This matrix is denoted by A_W, i.e.

$$A_W = (A)_{\substack{1\leq i\leq m,\, i\in W \\ 1\leq i\leq n}}$$

The analogue definition is made for b_W, i.e. if $W = \{i_1, \ldots, i_k\}$, then $b_W = (b_{i_1}, \ldots, b_{i_k})^T$.

The active set method works as follows: To start the algorithm choose x_0 with $Ax_0 \leq b$ and a set W_0 with $A_{W_0}x = b_{W_0}$. The set W_0 is called the current active set.

(S): Suppose that x_k has been computed and that the current active set is $W_k = \{i_1^k, \ldots, i_{m_k}^k\}$. Now calculate a d_k such that $z = x_k + d_k$ solves

$$(**) \qquad \text{minimize } \frac{1}{2} z^t Q z + c^T z \quad \text{subject to } A_{W_k} z = b_{W_k}.$$

The solution is given by the quadratic program subject to equality constraints. Hence we have to solve

$$\begin{pmatrix} Q & A_{W_k}^T \\ A_{W_k} & 0 \end{pmatrix} \begin{pmatrix} d_k \\ \lambda_k \end{pmatrix} = \begin{pmatrix} -c - Qx_k \\ 0 \end{pmatrix},$$

with $\lambda = (\lambda_{i_1^k}, \ldots, \lambda_{i_{m_k}^k})$. If $W = \emptyset$, then solve $Qd_k = -c - Qx_k$. There are three possibilities:

1. If $d_k = 0$ and $\lambda_k \geq 0$, then x_k is the solution of $(*)$.
2. If $d_k = 0$, then x_k is the solution of $(**)$. If k is the index of the smallest entry of λ, then put $W_{k+1} := W_k \backslash \{k\}$ and $x_{k+1} := x_k$ (method of steepest descent). Go back to (S).
3. Else calculate
$$\alpha_k = \min_{i \in W_k} \left(1, \frac{b_i - (Ax_k)_i}{(Ad_k)_i} \right).$$
$\alpha_k d_k$ is the greatest feasible vector that can be added to x_k without violating the constraints. If $\alpha_k = 1$, then put $x_{k+1} := x_k + d_k$ and $W_{k+1} := W_k$, otherwise choose an i_l^k with $\alpha_k = \frac{b_i - (Ax_k)_i}{(Ad_k)_i}$ and put $W_{k+1} := W_k \cup \{i_l^k\}$ and $x_k := x_k + \alpha_k d_k$. Go back to (S).

The Maple V/Release 5-procedure given below calculates the projection. It is necessary to include the `linalg` package. For some odd reasons that are due to Maple it works only if the matrix Q has been computed and defined in the work sheet before. The subroutines have self explanatory names. A, b, d, x are treated as matrices, e.g. `b=matrix(m,1,[...])`.

```
minialphaindex:=proc(W,A,b,d,x)
   local alpha,i,k,bminAmalx,Amald,wert;
    alpha:=1;;k:=0;Amald:=evalm(A&*d);bminAmalx:=evalm(b-A&*x);
    for i from 1 to rowdim(A) do
      if (Amald[i,1]>0) then
       if not member(i,W) then
         wert:=bminAmalx[i,1]/Amald[i,1];
         if alpha>wert then
         alpha:=wert:k:=i fi;fi;fi;od;[alpha,k]
 end:
 prolagnull1:=proc(Q,A,c,x,W)
```

```
             local dimQ,R,M,Qzerl,dimN,v;
             if W={} then R:=QRdecomp(Q,Q='Qzerl');
             [evalm(-inverse(R)&*transpose(Qzerl)&*(c+Q&*x)),0]
             else
             dimQ:=rowdim(Q);
             R:=matrix(nops(W),dimQ,[seq(seq(A[i,j],j=1..dimQ),i=W)]);
             dimN:=rowdim(R);
             M:=blockmatrix(2,2,
                [Q,transpose(R),R,matrix(dimN,dimN,(k,l)->0)]);
             R:=QRdecomp(M,Q='Qzerl');
             v:=evalm(inverse(R)&*transpose(Qzerl)&*
             blockmatrix(2,1,[-(c+Q&*x),matrix(dimN,1,(k,l)->0)]));
             [matrix(dimQ,1,[seq(v[i,1],i=1..dimQ)]),
              matrix(dimN,1,[seq(v[i,1],i=dimQ+1..dimQ+dimN)])];fi;
           end:

#Projection of a vector x onto {Ax<=b }.
#Starting vector x0 Starting working set W0
projektion:=proc(Q::matrix,x::matrix,A::matrix,
                b::matrix,x0::matrix,W0::set,eps1::float)
local k,j,m,c,y,d,alpha,Ende,W;
     Ende:=false;c:=evalm(-Q&*x);W:=W0;y:=x0;
     while not Ende do
        d:=prolagnull1(Q,A,c,y,W);
        alpha:=evalm(transpose(d[1])&*Q&*d[1])[1,1];
        if (alpha<eps1)
           then
             k:=0;m:=0;
             for j from 1 to nops(W) do
               if (d[2][j,1]<m) then
                  k:=W[j];m:=d[2][j,1];fi;od;
               if k=0 then Ende:=true else
                  W:=W minus {k} fi;
        else alpha:=minialphaindex(W,A,b,d[1],y);
             if alpha[2]>0 then W:=W union {alpha[2]};y:=evalm(y+alpha[1]*d[1])
             else y:=evalm(y+d[1]); fi;
        fi;
     od; matrix(rowdim(y),1,[seq(y[k,1],k=1..rowdim(y))]),W;
end:
```

Appendix B

Inverse Panjer Inversion for Histograms

The following routines have been used to plot figure 5.1 in Chapter 5. The random generator is a substract-with-borrow generator [Ma91].

The simulation has been done for exponentially distributed claims with the density $p(x) = \exp(-x)$, $x \geq 0$. If $q_0\delta_0 + q = \exp(p - \delta_0)$, then q is given by the formula

$$q(x) = \exp(-1-x)\left(1 + \sum_{n=2}^{\infty} \frac{x^{n-1}}{n(n-1)!^2}\right).$$

This is the C-code for computing q.

```
#include <stdio.h>
#include <math.h>
#include <stdlib.h>


#define lambda 1.0
#define S 10.0
#define mopt 2100
#define unendlich 1000


main(){
int j,i;
double x,sum,prod;


for(i=0;i<=mopt-1;i++)
{x=lambda*i*S/mopt;sum=1;prod=1;
```

```
for(j=1;j<=unendlich;j++){prod=prod*x/(j*j);sum=sum+prod/(j+1);}
printf("%lf %lf\n",i*S/mopt,sum*exp(-(lambda+i*S/mopt)));
}
```

This is the program for the histogram for q.

```
#include <stdio.h>
#include <math.h>
#include <stdlib.h>
#define  NSAMPLE 1000
#define mopt 15 /*optimal  choices  for S=10 and lambda=1: = 158
                                        S=10 and lambda=2: =118
        S=10 lambda=10: =48
        S=10 lambda=10
        NSAMPLE=4294967296: =739*/


#define lambda 1.0
#define S 10.0


double unif(void);
double compois(void);


main()
{unsigned long int i,m,j;
double q0,X;
double hatq[mopt];
q0=0;
for(i=0;i<=mopt-1;i++){hatq[i]=0;}
for(i=1;i<=NSAMPLE;i++)
  {X=compois();
   if(X==0){q0=q0+1;}
   else
   {if (X<=S){j=ceil(X*mopt/S);
              hatq[j-1]=hatq[j-1]+1;}}
     }



/*printf("%lf\n",q0/NSAMPLE);*/
```

```
/*for(i=0;i<=mopt-1;i++){
printf("%lf %lf\n",i*S/mopt,0.0);
printf("%lf %lf\n",i*S/mopt,hatq[i]*mopt/(NSAMPLE*S));
printf("%lf %lf\n",(i+1)*S/mopt,hatq[i]*mopt/(NSAMPLE*S));}*/

for(i=0;i<=mopt-1;i++){
printf("%li %lf\n",i,hatq[i]*mopt/(NSAMPLE*S));}



}/*end of main program*/


double compois(void){
   double sumN,sum;
   sumN=-log(unif())/lambda;sum=0;
   while(sumN<1){sumN=sumN-log(unif())/lambda;sum=sum-log(unif());}
   return(sum);

}
/* returns U(0,1)-variates, Marsaglia-Zaman algorithm */
double unif(void){
static unsigned long x[] =
{1276610355UL, 4193469394UL, 2057566612UL, 1886580328UL, 1694206606UL,
 2633431637UL, 1265626433UL,  885029446UL, 3417643270UL, 3311627661UL,
 2615330922UL, 2585171253UL, 2061319010UL,   76799462UL,  217610450UL,
 1970157156UL, 3650280925UL, 3031778051UL, 3936002891UL, 1455404536UL,
 3581605850UL,  978584193UL, 1392725752UL,  424558724UL,  718634923UL,
 2602380921UL, 1073859225UL, 2260449986UL,  437368889UL,  111202475UL,
  430748330UL,  860297108UL,  469595518UL, 2956147077UL, 2998566928UL,
 3679001976UL, 1174826611UL, 3589929608UL, 2670654217UL,  999890898UL,
 3874011621UL, 3680146780UL, 3569051095UL };
static int r = 0, s = 21, carry = 0;

if (r > 42) r -= 43;
if (x[s] >= x[r] + carry){
x[r] =  x[s] - x[r] - carry;
carry = 0;
}
else{
x[r] =  (4294967291UL - x[r] - carry) + x[s];
carry = 1;
}
```

```
if (++s > 42) s -= 43;
return (((double) x[r++] + 0.5) / 4294967291.0);
```

The next program computes the Panjer inversion.

```
#include <stdio.h>
#include <math.h>
#include <stdlib.h>
#define  NSAMPLE 1000
#define Ngraph 10
#define mopt 15 /*optimal choices for    S=10 and lambda=1: = 158
                                         S=10 and lambda=2: =118
         S=10 lambda=10: =48
         S=10 lambda=10
           NSAMPLE=4294967296: =739*/

#define lambda 1.0
#define S 10.0


double unif(void);
double compois(void);

main()
{unsigned long int N,i,m,j,k,l;
double q0,X,h,prodh,xx,yy,gamma0,sum,summinus,sumplus,fak,prod;
double hatq[mopt];
double alpha[2*mopt];
double D[2*mopt][2*mopt];

q0=0;
for(i=0;i<=mopt-1;i++){hatq[i]=0;}
for(i=1;i<=NSAMPLE;i++)
  {X=compois();
   if(X==0){q0=q0+1;}
   else
   {if (X<=S){j=ceil(X*mopt/S);
              hatq[j-1]=hatq[j-1]+1;}}
     }

q0=q0/NSAMPLE;h=S/mopt;

for(i=0;i<=mopt-1;i++){alpha[i]=hatq[i]/(h*NSAMPLE);}
```

```
for(i=mopt;i<=2*mopt-1;i++){alpha[i]=0;}

gamma0=alpha[0]/q0;


 D[0][0]=-1;

for(k=1;k<=2*mopt-1;k=k++)
  {


      sum=D[k-1][k-1];
      for(l=1;l<=k-1;l++)
        {sum=h*sum+D[k-1][k-1-l];}

      D[k][0]=(alpha[k]-alpha[k-1])*k*h/q0+sum*exp(-gamma0*h);

      for(j=1;j<=k;j=j++)
       {sum=0;
        for(l=1;l<=k-j;l++)
    {sum=sum+alpha[l]*(D[k-l-1][j-1]-D[k-l][j-1]);}

D[k][j]=(alpha[0]*D[k-1][j-1]+sum-alpha[k-j+1]*D[j-1][j-1])/(q0*j);}
}


printf("%lf %lf\n",0.0,gamma0);
for(i=1;i<=Ngraph;i++)
       {xx=(h*i)/Ngraph;yy=1-exp(-gamma0*i*h/Ngraph);
        printf("%lf %lf\n",xx,yy/xx);
       }


for(k=1;k<=2*mopt-1;k=k++)

{
    for(i=0;i<=Ngraph;i++)
       {sum=D[k][k];
        for(j=1;j<=k;j++)
   {sum=sum*i*h/Ngraph+D[k][k-j];}
        xx=(h*i)/Ngraph;xx=xx+h*k;yy=1+exp(-gamma0*i*h/Ngraph)*sum;
        printf("%lf %lf\n",xx,yy/xx);
       }
```

```
}



}/*end of main*/


double compois(void){
   double sumN,sum;
   sumN=-log(unif())/lambda;sum=0;
   while(sumN<1){sumN=sumN-log(unif())/lambda;sum=sum-log(unif());}
   return(sum);

}
/* returns U(0,1)-variates, Marsaglia-Zaman algorithm */
double unif(void){
static unsigned long x[] =
{1276610355UL, 4193469394UL, 2057566612UL, 1886580328UL, 1694206606UL,
 2633431637UL, 1265626433UL,  885029446UL, 3417643270UL, 3311627661UL,
 2615330922UL, 2585171253UL, 2061319010UL,   76799462UL,  217610450UL,
 1970157156UL, 3650280925UL, 3031778051UL, 3936002891UL, 1455404536UL,
 3581605850UL,  978584193UL, 1392725752UL,  424558724UL,  718634923UL,
 2602380921UL, 1073859225UL, 2260449986UL,  437368889UL,  111202475UL,
  430748330UL,  860297108UL,  469595518UL, 2956147077UL, 2998566928UL,
 3679001976UL, 1174826611UL, 3589929608UL, 2670654217UL,  999890898UL,
 3874011621UL, 3680146780UL, 3569051095UL };
static int r = 0, s = 21, carry = 0;

if (r > 42) r -= 43;
if (x[s] >= x[r] + carry){
x[r] =  x[s] - x[r] - carry;
carry = 0;
}
else{
x[r] =  (4294967291UL - x[r] - carry) + x[s];
carry = 1;
}
if (++s > 42) s -= 43;
return (((double) x[r++] + 0.5) / 4294967291.0);
}
```

Appendix C

A Semicone Probability Generator

This program implements the semicone probabilities using the Monte-Carlo method proposed at the end of Chapter **??**. We have used nearly the same notation within the program. A set $I \subset \{1, \ldots, S\}$ is encoded as `Index(x)`, i.e. `Index(I)`$=\sum_{i \in I} 2^{i-1}$. The observed failure rate is encoded in 2^S. The matrix f is equal to N from section 3.5.

```
#include <stdio.h>
#include <math.h>
#include <stdlib.h>
#define S 5
#define zweihochSminus2 32-2
#define  NSAMPLE 10000
```

```
double lambda;
double Q[S+1][S+1],QIminus[S+1][S+1],Lminus[S+1][S+1],
   DI[zweihochSminus2+1][S+1][S+1];
int Imenge[S+1],DIberechnet[zweihochSminus2+1];

void initQLDIber();
void newDecider(int index);
int IndexI();
double unif(void);
```

```
double gauss(void);


main()
{int k,i,j,imin,samplei,algorfails;
double Y[S+1],Z[S+1],halbraum[zweihochSminus2+3];
double Min,summand,sumplus,summinus;

printf("This is a Monte-Carlo semicone calculator\n");
/*printf("I is encoded by Index(I)=sum_{i in I} 2^{i-1}\n");*/
printf("intensity lambda =");scanf("%lf",&lambda);
printf("truncation parameter S=%i\n",S);
printf("sample size S=%i\n",NSAMPLE);

initQLDIber();
for(k=0;k<=zweihochSminus2+2;k=k+1){halbraum[k]=0;}

for(samplei=1;samplei<=NSAMPLE;samplei=samplei+1)
{




   for(i=1;i<=S;i=i+1){Z[i] =gauss();}
   /*calcukation of  Y=L^{-T}Z\sim N(0,Q^{-1})*/
   for(i=1;i<=S;i=i+1)
     {sumplus=0;summinus=0;
      for(j=i;j<=S;j=j+1)
        {summand=Lminus[j][i]*Z[j];
         if(summand>=0){sumplus=sumplus+summand;}
         else{summinus=summinus+summand;}}
      Y[i]=sumplus+summinus;}




   Min=Y[1];imin=1;
   for(i=2;i<=S;i=i+1){if(Y[i]<=Y[imin]){Min=Y[i];imin=i;}}
   if(Min>=0){halbraum[zweihochSminus2+1]=halbraum[zweihochSminus2+1]+1;}
    else
     {imin=1;
      for(i=1;i<=S;i=i+1){Imenge[i]=0;}
      sumplus=0;summinus=0;
      for(j=1;j<=S;j=j+1)
```

```
        {summand=-Q[1][j]*Y[j];
         if(summand>=0){sumplus=sumplus+summand;}
         else{summinus=summinus+summand;}}
      Min=summinus+sumplus;
      for(i=2;i<=S;i=i+1)
         {sumplus=0;summinus=0;
          for(j=1;j<=S;j=j+1)
            {summand=-Q[i][j]*Y[j];
               if(summand>=0){sumplus=sumplus+summand;}
               else{summinus=summinus+summand;}}

         if(sumplus+summinus<=Min){Min=sumplus+summinus;imin=i;}}
         if(Min>=0){halbraum[0]=halbraum[0]+1;}
         else
           {Imenge[imin]=1;algorfails=0;
            do
             {k=IndexI();algorfails=algorfails+1;
      if(DIberechnet[k]==0){newDecider(k);DIberechnet[k]=1;}
      imin=1;
              sumplus=0;summinus=0;
              for(j=1;j<=S;j=j+1)
                {summand=DI[k][1][j]*Y[j];
                 if(summand>=0){sumplus=sumplus+summand;}
                 else{summinus=summinus+summand;}}
              Min=summinus+sumplus;
              for(i=2;i<=S;i=i+1)
                {sumplus=0;summinus=0;
                 for(j=1;j<=S;j=j+1)
                   {summand=DI[k][i][j]*Y[j];
                    if(summand>=0){sumplus=sumplus+summand;}
                    else{summinus=summinus+summand;}}
                 if(sumplus+summinus<=Min){Min=sumplus+summinus;imin=i;}}
                 if(Min<0)
           {if(Imenge[imin]>0){Imenge[imin]=0;}else{Imenge[imin]=1;}}}
            while(Min<0 && algorfails<zweihochSminus2+1);
            if(Min<0){halbraum[zweihochSminus2+2]=halbraum[zweihochSminus2+2]+1;}
            else{halbraum[k]=halbraum[k]+1;}
    }
      }}
 printf("\n DI calculated:  ");
 for(i=1;i<=zweihochSminus2;i=i+1)
    {printf("%i",DIberechnet[i]);}printf("\n\n");
 printf("\n semicone probabilities:\n\n");
```

```
 for(i=0;i<=zweihochSminus2+2;i=i+1)
     {printf("%lf  ",halbraum[i]/NSAMPLE);}
  }/*Ende main*/





void initQLDIber()
{
double Delta[S+1][S+1];
double q[S+1],qminus[S+1],Deltasum[S+1];
double normtildeq,sumplus,summinus,summand;
int i,j,k;


for(k=1;k<=zweihochSminus2;k=k+1){DIberechnet[k]=0;}




q[0]=exp(-lambda);
for(i=1; i<=S; i=i+1){q[i]=q[i-1]*lambda/i;}
qminus[0]=exp(lambda);
for(i=1; i<=S; i=i+1){qminus[i]=qminus[i-1]*(-lambda)/i;}

sumplus=0;summinus=0;
for(k=1;k<=S;k=k+2){summinus=summinus+qminus[k];}
for(k=2;k<=S;k=k+2){sumplus=sumplus+qminus[k];}
normtildeq=sumplus+summinus;





/*calculation of Delta*/


for(i=0;i<=S;i=i+1)
   {for(j=1;j<=i;j=j+1)
      {sumplus=0;summinus=0;
       for(k=0;k<=j;k=k+1)
         {summand=qminus[i-k]*qminus[j-k]*q[k];
```

```
  if(summand>=0)
    {sumplus=sumplus+summand;}
  else
    {summinus=summinus+summand;}}
       Delta[i][j]=sumplus+summinus;
       Delta[j][i]=sumplus+summinus;
    }}

for(i=1;i<=S;i=i+1)
   {sumplus=0;summinus=0;
    for(k=1;k<=S;k=k+1)
         {summand=Delta[i][k];
  if(summand>=0)
    {sumplus=sumplus+summand;}
  else
    {summinus=summinus+summand;}}
    Deltasum[i]=(sumplus+summinus);
 }




/*calculation of Q=f^TA^{-1}(diag(q_0,\dots,q_S)-q\otimes q)A^{-T}f*/


sumplus=0;summinus=0;
for(k=1;k<=S;k=k+1)
   {summand=Deltasum[k];
    if(summand>=0) {sumplus=sumplus+summand;}
    else {summinus=summinus+summand;}}

Q[1][1]=(sumplus+summinus+2*normtildeq+exp(lambda)-1)/(lambda*lambda);


for(i=2;i<=S;i=i+1)
   {Q[1][i]=-(Deltasum[i]+qminus[i])/(lambda*lambda);Q[i][1]=Q[1][i];
 }


for(i=2;i<=S;i=i+1)
  {for(j=i;j<=S;j=j+1)
     {Q[i][j]=Delta[i][j]/(lambda*lambda);Q[j][i]=Q[i][j];}
  }
```

```
/*Cholesky-decomposition of Q=f^TA^{-1}(diag(q_0,\dots,q_S)-q\otimes q)A^{-T}f*/

for(i=1;i<=S;i=i+1)
  {sumplus=0;
   for(k=1;k<=i-1;k=k+1)
     {sumplus=sumplus+Delta[i][k]*Delta[i][k];}
   Delta[i][i]=sqrt(Q[i][i]-sumplus);
   for(j=i+1;j<= S;j=j+1)
     {sumplus=0;summinus=0;
      for(k=1;k<=i-1;k=k+1)
          {summand=Delta[i][k]*Delta[j][k];
           if(summand>=0)
   {sumplus=sumplus+summand;}
           else
  {summinus=summinus+summand;}}
      Delta[j][i]=(Q[i][j]-summinus-sumplus)/Delta[i][i];}
   }


/*calculation of  L^{-1} of the Cholesky decomposition of Q*/

for(i=1;i<=S;i=i+1){Lminus[i][i]=1/Delta[i][i];}
for(i=2;i<=S;i=i+1)
  {for(j=i-1;j>=1;j=j-1)
     {sumplus=0;summinus=0;
      for(k=j+1;k<=i;k=k+1)
          {summand=Lminus[i][k]*Delta[k][j];
           if(summand>=0)
   {sumplus=sumplus+summand;}
           else
  {summinus=summinus+summand;}}
      Lminus[i][j]=-(sumplus+summinus)/Delta[j][j];}}

}/*end of initQL*/




void newDecider(int index)
 {
double sumplus,summinus,summand;
```

```
double LI[S+1][S+1],LIminus[S+1][S+1],QDI[S+1][S+1];
int cardImenge,k,k1,l,i,j;


/*calculation of E_I(E_I^T Q E_I)^{-1} E_I^T*/

/*calculation of Q_I=(E_I^T Q E_I)*/

i=0;cardImenge=0;
for(k=1;k<=S;k=k+1)
 {if(Imenge[k]>0){cardImenge=cardImenge+1;i=i+1;j=i-1;
             for(k1=k;k1<=S;k1=k1+1)
       {if(Imenge[k1]>0){j=j+1;QIminus[i][j]=Q[k][k1];QIminus[j][i]=Q[k][k1];}}}
       }


/*Cholesky decomposition of QIminus*/

for(i=1;i<=cardImenge;i=i+1)
  {sumplus=0;
   for(k=1;k<=i-1;k=k+1)
     {sumplus=sumplus+LI[i][k]*LI[i][k];}
   LI[i][i]=sqrt(QIminus[i][i]-sumplus);
   for(j=i+1;j<= cardImenge;j=j+1)
     {sumplus=0;summinus=0;
      for(k=1;k<=i-1;k=k+1)
          {summand=LI[i][k]*LI[j][k];
           if(summand>=0)
   {sumplus=sumplus+summand;}
           else
  {summinus=summinus+summand;}}
      LI[j][i]=(QIminus[i][j]-summinus-sumplus)/LI[i][i];}
   }
/*inversion of QIminus*/

for(i=1;i<=cardImenge;i=i+1){LIminus[i][i]=1/LI[i][i];}
for(i=2;i<=cardImenge;i=i+1)
  {for(j=i-1;j>=1;j=j-1)
     {sumplus=0;summinus=0;
      for(k=j+1;k<=i;k=k+1)
          {summand=LIminus[i][k]*LI[k][j];
           if(summand>=0)
   {sumplus=sumplus+summand;}
```

```
           else
  {summinus=summinus+summand;}}
      LIminus[i][j]=-(sumplus+summinus)/LI[j][j];}}




i=0;
for(k=1;k<=S;k=k+1)
 {if(Imenge[k]>0){
             i=i+1;j=i-1;
             for(k1=k;k1<=S;k1=k1+1)
       {if(Imenge[k1]>0)
         {j=j+1;
          sumplus=0;summinus=0;
  for(l=j;l<=cardImenge;l=l+1)
  {summand=LIminus[l][i]*LIminus[l][j];
   if(summand>=0)
      {sumplus=sumplus+summand;}
   else {summinus=summinus+summand;}}
   QIminus[k][k1]=sumplus+summinus;
   QIminus[k1][k]=QIminus[k][k1];}
 else{QIminus[k][k1]=0;QIminus[k1][k]=0;}
  }}
   else{for(k1=1;k1<=S;k1=k1+1){QIminus[k][k1]=0;QIminus[k1][k]=0;}}}

/*calculation of the decision matrix
   DI=E_I(E_I^TQE_I)E_I^TQ+E_I^C(E_I^C^TQ-E_I^C^TE_I(E_I^TQE_I)QE_I^TQ)*/


for(i=1;i<=S;i=i+1)
{for(j=1;j<=S;j=j+1)
      {sumplus=0;summinus=0;
       for(k=1;k<=S+1;k=k+1)
  {summand=QIminus[i][k]*Q[k][j];
  if(summand>=0){sumplus=sumplus+summand;}
  else{summinus=summinus+summand;}}
       DI[index][i][j]=sumplus+summinus;}}



for(i=1;i<=S;i=i+1)
{for(j=1;j<=S;j=j+1)
```

```
      {sumplus=0;summinus=0;
       for(k=1;k<=S+1;k=k+1)
  {summand=Q[i][k]*DI[index][k][j];
  if(summand>=0){sumplus=sumplus+summand;}
  else{summinus=summinus+summand;}}
       QDI[i][j]=sumplus+summinus;}}




for(i=1;i<=S;i=i+1)
{if(Imenge[i]==0)
  {for(j=1;j<=S;j=j+1){if(Imenge[j]==0){DI[index][i][j]=-(Q[i][j]-QDI[i][j]);}}}}





}/*end of newDecider*/


int IndexI()
{int ergebnis,k,zweihochk;

zweihochk=1;ergebnis=0;
for(k=1;k<=S;k=k+1)
{ergebnis=ergebnis+zweihochk*Imenge[k];zweihochk=2*zweihochk;}
return(ergebnis);}

double gauss(void){
static int i = 0;   /* 1 if value in stock */
static double x1, x2, y1, y2;

if (i == 0){
x1 = unif();
x2 = unif();
y1 = sqrt(-2 * log(x1));
y2 = 2 * M_PI * x2;
x1 = y1 * sin(y2);
x2 = y1 * cos(y2);
i = 1;
return x1;
}
```

```
else{
i = 0;
return x2;
}
}
/* returns U(0,1)-variates, Marsaglia-Zaman algorithm */
double unif(void){
static unsigned long x[] =
{1276610355UL, 4193469394UL, 2057566612UL, 1886580328UL, 1694206606UL,
 2633431637UL, 1265626433UL,  885029446UL, 3417643270UL, 3311627661UL,
 2615330922UL, 2585171253UL, 2061319010UL,   76799462UL,  217610450UL,
 1970157156UL, 3650280925UL, 3031778051UL, 3936002891UL, 1455404536UL,
 3581605850UL,  978584193UL, 1392725752UL,  424558724UL,  718634923UL,
 2602380921UL, 1073859225UL, 2260449986UL,  437368889UL,  111202475UL,
  430748330UL,  860297108UL,  469595518UL, 2956147077UL, 2998566928UL,
 3679001976UL, 1174826611UL, 3589929608UL, 2670654217UL,  999890898UL,
 3874011621UL, 3680146780UL, 3569051095UL };
static int r = 0, s = 21, carry = 0;

if (r > 42) r -= 43;
if (x[s] >= x[r] + carry){
x[r] =  x[s] - x[r] - carry;
carry = 0;
}
else{
x[r] =  (4294967291UL - x[r] - carry) + x[s];
carry = 1;
}
if (++s > 42) s -= 43;
return (((double) x[r++] + 0.5) / 4294967291.0);
}
```

Bibliography

[Ba92] Bauer, Heinz, Maß- und Integrationstheorie, 2. ed., de Gruyter, 1992.

[Bu70] Bühlmann, Hans, Mathematical Methods in Risk Theory, Springer Verlag, 1970

[Cs81] Csörgő, Sándor, Limit Behaviour of Empirical Characteristic Function, Annals of Probability 1981, Vol. 9, No. 1, p.130-144

[Cs89] Csörgő, Sándor and Mason, David M., Bootstrapping Empirical Functions, The Annals of Statistics, 17, no. 4, 1989, p.1447-1471

[De85] Devroye, Luc and Györfi, László, Nonparametric Density Estimation: The L_1 View, John Wiley & Sons, 1985

[De93] Dembo, A., Zeitouni, O, Large Deviations Techniques and Applications, Boston, Jones and Bartlett Publishers 1993

[Es91] van Es, A. J. Uniform deconvolution: nonparametric maximum likelihood and inverse estimation. Nonparametric functional estimation and related topics (Spetses, 1990), 191–198, NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci., 335, Kluwer Acad. Publ., Dordrecht, 1991.

[Fa91] Fan, Yanqing, Global Behavior of Deconvolution Kernel Estimates, Statistica Sinica, p.541-551, 1991

[Fa97] Fan, Yanqing and Liu Yanjun, A Note on Asymptotic Normality for Deconvolution Kernel Density Estimators, Sankhyã, The Indian Journal of Statistics, 59, series A, p.138-141

[Fe66] Feller, William, An Introduction to Probability Theory and Its Applications, John Wiley, & Sons, 1966

[Fe77] Feuerverger, A., Mureika, R.A., The Empirical Characteristic Function and its Applications, Annals of Statistics 1977, Vol.5, No.1, p.88-97

[Ga69] Gamelin, Theodore W. Uniform Algebras, Prentice Hall, inc. 1969

[Ga87] Galambos, Janos, The Asymptotic Theory of Extreme Order Statistics, sec. ed. 1987, Robert E. Krieger Publishing Co., 1987

[Gr92] Groeneboom, Piet and Wellner, Jon A., Information Bounds and Nonparametric Maximum Likelihood Estimation. DMV Seminar, 19. Birkhäuser Verlag, Basel, 1992

[Gr93] Grübel, Rudolf and Pitts, Susan M., Nonparametric Estimation in Renewal Theory I: The Empirical Renewal Function, The Annals of Statistics, 21, no. 3, p.1431-1451, 1993

[He97] Hermesmeier, Renate, Transformationsmethoden und Diskretisierungseffekte bei stochastischen Modellen in der Versicherungsmathematik, Dissertation, Universität Hannover, 1997

[Hu90] Hürlimann, Werner, On Maximum Likelihood Estimation for Count Data Models, Insurance: Mathematics and Economics 9 (1990), p.39-49, North Holland

[Jo95] Jongbloed, Geurt, Three Statistical Inverse Problems, Thesis Delft University of Technology, 1995

[Kl78] Klingenberg, Wilhelm, A Course in Differential Geometry, Springer, 1978

[La93] Lang, Serge, Real and Functional Analysis, 3. ed, Springer Verlag, 1993

[Li95] Lifshitz, M.A., Gaussian Random Functions, Kluwer Academic Publisher, 1995

[Lu70] Lukacs, Eugene, Characteristic Functions, sec. ed., Griffin, 1970

[Lu89] Luenberger, David G., Linear and Nonlinear Programming, 2. ed, Reading, Massachusetts, Addison-Wesley Publishing Company

[Ma91] Marsaglia, George, Zaman, Arif, A New Class of Random Number Generators, The Annals of Applied Probability, 1,1991, p.462-480

[Ne79] Neyman, Jerzy, $C(\alpha)$-Tests and Their Use, Sankhyã: The Indian Journal of Statistics, 1979, 41, Series A, p.1-21

[Ni00] Ni, Zhongren, Normal Orthant Porbabilities in the Equicorrelated Case, Journal of Mathematical Analysis and Applications, 246, 2000, p.280-295

[No98] Nomakuchi, Kentaro, A Generalization of the Childs-Moran Result for Orthoscheme Probabilities, Austr. & New Zealand J. Statist. 40, no.1, 1998, p.103-109

[Pa81] Panjer, Harry, Recursive Evaluation of a Family of Compound Distributions, Astin Bulletin, 12, p.22-26

[Pa92] Panjer, Harry H., Willmot Gordon E., Insurance Risk Models, Society of Actuaries,

[Pi94] Pitts Susan M., Nonparametric Estimation of Compound Distributions with Applications in Insurance, Ann. Inst. Statist. Math. Vol. 46, No. 3, p.537-555, 1994

[Po84] Pollard, David, Convergence of Stochastic Processes, Springer Series in Statistics, 1984

[Pu85] Puri, Prem S., On an Optimal $C(\alpha)$-Test of Poisson Hypothesis against Compound Poisson Alternatives, Statistical Theory and Data Analysis, p.565-583, 1985

[Qu87] Quine, M.P. , Seneta, E., Bortkiewicz Data and the Law of Small Numbers, International Statistical Review 55, no.2, p.173-181, 1987

[Ro79] Rogozin, B.A., Sgibnev, M.S., Banach Algebras of Measures on the line, English trans. from Sibirian Mathematical Journal, 21, no.2, p.265-273, 1979

[Ro88] Robertson, Tim, Wright, F.T., Dykstra, R.L., Order restricted statistical inference, John Wiley & Sons, 1988

[Ru91] Rudin, Walter, Functional Analysis, Second Edition, New York et al., Mc Craw-Hill, Inc, 1991

[Sh86] Shorack, Galen R. and Wellner, Jon A., Empirical Process with Applications to Statistics, John Wiley & Sons, Inc, 1986

[Si86] Silverman, Bernard W., Density Estimation for Statistics and Data Analysis, Chapman and Hall, 1986

[Si76] Simar, Léopold, Maximum Likelhood Estimation of a Compuond Poisson Process, The Annals of Statistics, 1976, 4, No. 6, p.1200-1209

[Sw92] Swartz, Charles, An Introduction to Functional Analysis, Marcel Dekker, Inc, 1992

[Ta73] Taylor, Joseph L., Measure Algebras, Expository Lectures from the CBMS Regional Conference held at the University of Montana, June 1972, Regional Conference Series in Mathematics, 1973

[Te84] Teugels, Jozef L., Approximations and Estimation of Some Compound Distrinutions, Insurance: Mathematics and Economics 4, p. 143-153, 1985

[To90] Tong, Y. L., The Multivariate Normal Distribution, Springer-Verlag, 1990

[Va87] N.N. Vakhania, V.I. Tarieladze, S.A. Chobanyan, Probability Distributions on Banach Spaces, D. Reidel Publishing Company, Dordrecht, Holland, 1987

Curriculum Vitae

Boris Buchmann was born as son of Klaus Buchmann and Dr.med. Ursula Buchmann, on October 22, 1969 in Göttingen, Germany. He visited the primary school in Edemissen (1976-1980), the Ratsgymnasium in Peine (1982-1983) and the grammar school in Neustadt am Rübenberge, (1983-1989) where he achieved the Allgemeine Hochschulreife in 1989. He spent his mandatory social service at the seniors' home St. Nicolai Stift, Neustadt (1.6.1989-30.9.1990). While preparing to study music (12/1990-06/1991), he extended his employment was again employee of the mentioned seniors' home (07/1991-09/1991). From October 1991 he studied mathematics as major and physics as minor field at the University of Hannover (10/1991-03/1997/ graduation day was March 26, 1997). During his studies he worked as student teaching assistant for the institute of pure mathematics. His diploma thesis at the institute of stochastics is entitled *Große Abweichungen bei halbdeterministischen Sprungprozessen* and deals with large deviations of semideterministic jumping processes. From April 1, 1997 to March 31, 2001 he held a position as teaching and research assistant at the "Institut für Mathematische Stochastik" at the University of Hannover. On April 1, 2001 he started employment at the Center of Mathematical Sciences, Munich University of Technology.