ALEA, Lat. Am. J. Probab. Math. Stat. 20, 561–575 (2023)
DOI: 10.30757/ALEA.v20-20

On the existence of maximum likelihood estimates for the parameters of the Conway-Maxwell-Poisson distribution

Stefan Bedbur, Udo Kamps and Anton Imm

Institute of Statistics, RWTH Aachen University, 52056 Aachen, Germany
E-mail address: bedbur@isw.rwth-aachen.de
URL: https://www.isw.rwth-aachen.de

Institute of Statistics, RWTH Aachen University, 52056 Aachen, Germany
E-mail address: udo.kamps@rwth-aachen.de
URL: https://www.isw.rwth-aachen.de

Institute of Statistics, RWTH Aachen University, 52056 Aachen, Germany
E-mail address: anton.imm@rwth-aachen.de

Abstract. As a well-known and important extension of the common Poisson model with an additional parameter, Conway-Maxwell-Poisson (CMP) distributions allow for describing under- and overdispersion in discrete data. Constituting a two-parameter exponential family, CMP distributions possess useful structural and statistical properties. However, the exponential family is not steep and maximum likelihood estimation may fail even for non-trivial data sets, which is different from the Poisson case, where maximum likelihood estimation only fails if all data outcomes are zero. Conditions are examined for existence and non-existence of maximum likelihood estimates in the full family as well as in subfamilies of CMP distributions, and several figures illustrate the problem.

1. Introduction

As a particular extension of the Poisson model, the Conway-Maxwell-Poisson (CMP) distribution allows for modeling under- and overdispersion in count data via an additional parameter. Although its origin dates back to Conway and Maxwell (1962), the usefulness, simplicity, and statistical elegance of the model has been pointed out later in Shmueli et al. (2005), and there has been a broad interest in theory and applications of CMP distributions since then.

The counting density of the CMP distribution $P_{\boldsymbol{\vartheta}}$ is given by

$$f_{\boldsymbol{\vartheta}}(x) \,=\, C(\boldsymbol{\vartheta})\,\frac{\lambda^x}{x!^\nu}\,, \qquad x \in \mathbb{N}_0 = \{0,1,2,\dots\}\,, \tag{1.1}$$

Received by the editors August 24th, 2022; accepted March 8th, 2023.
2010 *Mathematics Subject Classification.* 62F10, 62E10.
Key words and phrases. Conway-Maxwell-Poisson distribution, exponential family, steepness, maximum likelihood estimation, non-existence of a maximum likelihood estimate.

with

$$C(\boldsymbol{\vartheta}) = \left(\sum_{k=0}^{\infty} \frac{\lambda^k}{k!^\nu} \right)^{-1}$$

for $\boldsymbol{\vartheta} = (\lambda, \nu) \in \Theta = (0,\infty)^2 \cup [(0,1) \times \{0\}]$; see, e.g., Shmueli et al. (2005).

Some basic distributions are contained in the class of CMP distributions. For $\nu = 1$, $P_{\boldsymbol{\vartheta}} = po(\lambda)$ is the Poisson distribution with parameter $\lambda > 0$. By choosing $\nu = 0$, $P_{\boldsymbol{\vartheta}} = geo(1-\lambda)$ is nothing but the geometric distribution with success probability $1 - \lambda \in (0,1)$. Moreover, as ν tends to infinity, $P_{\boldsymbol{\vartheta}}$ converges in distribution to a Bernoulli distribution $ber(\lambda/(1+\lambda))$ with success probability $\lambda/(1+\lambda) \in (0,1)$.

The parameter ν can be considered a dispersion parameter. For $\nu > 1$ ($\nu < 1$), the variance of $P_{\boldsymbol{\vartheta}}$ is smaller (larger) than or equal to its mean; see, e.g., Daly and Gaunt (2016, Section 2.3.1). Of course, for $\nu = 1$, i.e., for the family of Poisson distributions, we have equality. The counting density of $P_{\boldsymbol{\vartheta}}$ is shown in Figure 1.1 for different values of λ and ν.

CMP distributions possess useful structural and statistical properties, and they turn out to be included in several more general families of distributions. For respective reviews along with applications as well as related extensions and generalizations, we refer to Shmueli et al. (2005), Sellers et al. (2012), Daly and Gaunt (2016), Sellers et al. (2017), Li et al. (2020), and Sellers and Premeaux (2021). As stated in Shmueli et al. (2005), CMP distributions constitute a two-parameter exponential family (EF), which yields a variety of helpful properties. In steep EFs, such properties comprise necessary and sufficient conditions for the existence of maximum likelihood (ML) estimates. However, as reported in Krivitsky (2012), the EF of CMP distributions is not steep, such that respective statements cannot be applied, here.

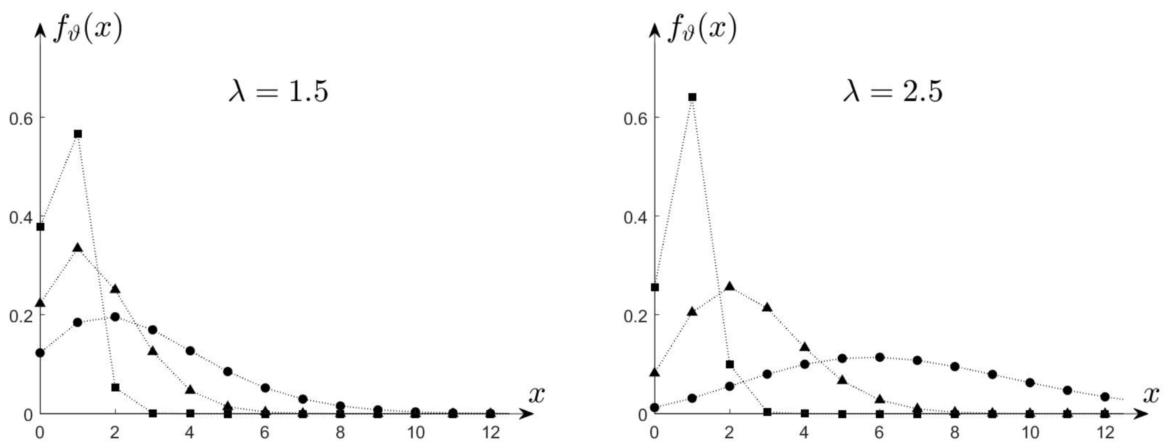
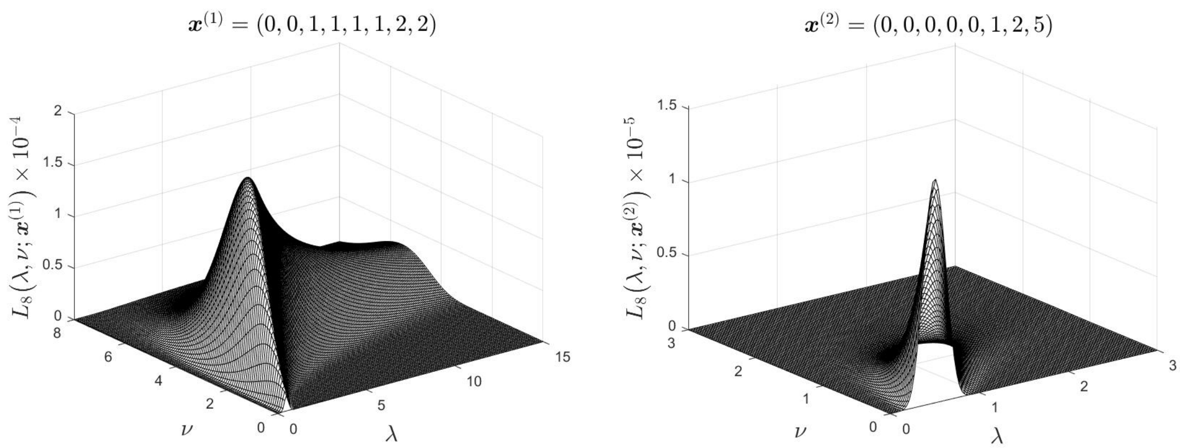


Figure 1.1. Counting density of the CMP distribution $P_{\boldsymbol{\vartheta}}$ for $\lambda \in \{1.5, 2.5\}$ and $\nu = 0.5$ (circles), $\nu = 1$ (triangles), and $\nu = 4$ (squares), where adjacent points are connected by dotted lines for illustration.

In this paper, the focus is on ML estimation of the parameters λ and ν of the CMP distribution based on a sample of independent and identically distributed (iid) random variables. Although the corresponding likelihood function and likelihood equation have a simple form (see, e.g., Shmueli et al., 2005, Section 3.2, and Sellers et al., 2012, Section 2.3), the problem of whether an ML estimate for $\boldsymbol{\vartheta}$ exists at all and, if so, whether it is a solution of the likelihood equation does not seem to have been addressed in the literature so far. In contrast to the common Poisson model, where ML estimation of the parameter λ only fails if all data outcomes are zero and the ML estimate is obtained as the unique solution of the likelihood equation in case of existence, deriving conditions for the (non-)existence of ML estimates for the CMP parameters is analytically problematic. It turns out that existence of ML estimates is indeed an issue (unlike uniqueness, which is guaranteed by the EF structure). We examine conditions for the existence and non-existence of the ML estimate, and it is found that existence may fail even in non-trivial data situations. In Figure 1.2, the likelihood function $L_8(\boldsymbol{\vartheta};\boldsymbol{x}) = \prod_{i=1}^{8} f_{\boldsymbol{\vartheta}}(x_i)$, $\boldsymbol{\vartheta} \in \Theta$, is depicted for three different data sets $\boldsymbol{x} = (x_1, \ldots, x_8) \in \mathbb{N}_0^8$ of size 8. We find that, based on $\boldsymbol{x}^{(1)} = (0,0,1,1,1,1,2,2)$, the ML estimate for $\boldsymbol{\vartheta}$ exists, lies in the interior $int(\Theta)$ of Θ, and is the unique solution of the likelihood equation. Based on $\boldsymbol{x}^{(2)} = (0,0,0,0,0,1,2,5)$, the ML estimate for $\boldsymbol{\vartheta}$ is seen to exist, but it lies on the boundary $\{(\lambda, 0) : \lambda \in (0,1)\} \subset \Theta$ of Θ and can therefore not be obtained as a solution of the likelihood equation. Finally, based on $\boldsymbol{x}^{(3)} = (1,1,1,1,2,2,2,2)$, an ML estimate for $\boldsymbol{\vartheta}$ does not exist at all. These examples motivate to look for conditions, which ensure existence or non-existence of the ML estimate or of a solution of the likelihood equation.



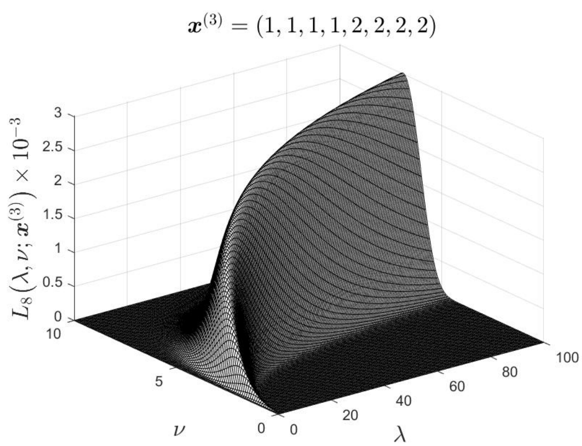


Figure 1.2. Likelihood function $L_8(\lambda,\nu;\boldsymbol{x})$ based on data set $\boldsymbol{x} \in \{\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \boldsymbol{x}^{(3)}\}$ with $\boldsymbol{x}^{(1)} = (0,0,1,1,1,1,2,2)$, $\boldsymbol{x}^{(2)} = (0,0,0,0,0,1,2,5)$, and $\boldsymbol{x}^{(3)} = (1,1,1,1,2,2,2,2)$.

By using the EF structure of the CMP distribution, we show that for data sets with range less than 2, ML estimates for the CMP parameters do not exist. Although the result is obtained by utilizing properties of EFs, the derivation is not standard, since the EF of CMP distributions is not steep (see Krivitsky, 2012), such that the general theory on steep EFs is not applicable, here. ML estimation in EFs is highly connected with the mean value function, which maps inner points of the natural parameter space of the EF to the corresponding expectation of the sufficient statistic. Non-steepness of the EF then goes along with some unpleasant properties of this mean value function, and ML estimation gets more involved. For the full EF of CMP distributions, we state sufficient conditions for the existence and non-existence of the ML estimate. In restricted models with an upper bound for one of the natural parameters of the EF, necessary and sufficient conditions for the existence of the ML estimate are derived. For the subfamily of equi- and underdispersed CMP distributions, we provide a sufficient condition for the non-existence of a solution of the likelihood equation. Several figures indicate whether, based on a given data set, a solution of the likelihood equation exists or does not exist.

The remainder of this article is organized as follows. In Section 2, the CMP distribution is represented and discussed in the context of EFs. Basic properties resulting from the EF structure are derived including non-steepness of the family. The subsequent discussion on ML estimation of the CMP parameters then makes use of these findings and is divided into two parts. In Section 3, ML estimation is studied within the full EF of CMP distributions, and ML estimation in restricted models of CMP distributions is considered in Section 4. Illustrations support and supplement the mathematical findings.

2. The CMP distribution as exponential family

By introducing the mapping $\boldsymbol{Z} = (Z_1, Z_2)$ with

$$Z_1(\boldsymbol{\vartheta}) = \ln(\lambda)\,, \qquad Z_2(\boldsymbol{\vartheta}) = \nu\,, \qquad \boldsymbol{\vartheta} = (\lambda, \nu) \in \Theta\,,$$

and the statistic $\boldsymbol{T} = (T_1, T_2)$ with

$$T_1(x) = x\,, \qquad T_2(x) = -\ln(x!)\,, \qquad x \in \mathbb{N}_0\,, \tag{2.1}$$

density (1.1) can be rewritten as

$$f_{\boldsymbol{\vartheta}}(x) \;=\; C(\boldsymbol{\vartheta})\, e^{\langle \boldsymbol{Z}(\boldsymbol{\vartheta}), \boldsymbol{T}(x)\rangle}\,, \qquad x \in \mathbb{N}_0\,, \tag{2.2}$$

where $\langle\cdot,\cdot\rangle$ denotes the standard scalar product. Hence, $\mathcal{P} = \{P_\vartheta : \vartheta \in \Theta\}$ forms an EF; see also Shmueli et al. (2005). In the following, we address some basic but important properties of $\mathcal{P}$ related to EFs. Firstly, we show that representation (2.2) is minimal meaning that $f_{\boldsymbol{\vartheta}}$ has no representation of type (2.2) with fewer than 2 summands in the exponent; see, e.g., Bedbur and Kamps (2021, Section 2.2) for a precise definition.

Lemma 2.1. *Representation* (2.2) *is minimal.*

Proof: Let $a_0, a_1, a_2 \in \mathbb{R}$ with $a_0 + a_1 \ln(\lambda) + a_2\nu = 0$ for all $\boldsymbol{\vartheta} = (\lambda, \nu) \in \Theta$. For $\nu = 0$ and $\lambda \nearrow 1$, this yields $a_0 = 0$. Then, setting $\lambda = 1/2$ and $\nu = 0$ gives $a_1 = 0$, from which $a_2 = 0$ immediately follows. Hence, Z_1 and Z_2 are affinely independent.

Let $b_0, b_1, b_2 \in \mathbb{R}$ with $b_0 + b_1 x - b_2 \ln(x!) = 0$ for all $x \in \mathbb{N}_0$. For $x = 0$, $b_0 = 0$ is obtained. Then, setting $x = 1$ yields that $b_1 = 0$, from which $b_2 = 0$ directly follows. Hence, T_1 and T_2 are affinely independent.

The assertion now follows from Corollary 8.1 in Barndorff-Nielsen (2014). □

Next, it is suitable for our purposes to switch to the natural parametrization of $\mathcal{P}$; see, e.g., Bedbur and Kamps (2021, Section 2.3) for details. In virtue of the parameter transformation $(\zeta_1, \zeta_2) = (\ln(\lambda), \nu)$, this parametrization is given by $\tilde{P}_{\boldsymbol{\zeta}}$, $\boldsymbol{\zeta} \in \Xi$, with counting densities

$$\tilde{f}_{\boldsymbol{\zeta}}(x) \;=\; e^{\langle \boldsymbol{\zeta}, \boldsymbol{T}(x)\rangle - \kappa(\boldsymbol{\zeta})}\,, \quad x \in \mathbb{N}_0\,, \qquad \boldsymbol{\zeta} \in \Xi\,, \tag{2.3}$$

where the so-called cumulant function $\kappa : \Xi \to \mathbb{R}$ is defined by

$$\kappa(\boldsymbol{\zeta}) = \ln\left(\sum_{k=0}^{\infty} \frac{e^{\zeta_1 k}}{k!^{\zeta_2}}\right), \qquad \boldsymbol{\zeta} = (\zeta_1, \zeta_2) \in \Xi\,, \tag{2.4}$$

and the new parameter space is denoted by

$$\Xi \;=\; [\mathbb{R} \times (0, \infty)] \cup [(-\infty, 0) \times \{0\}]\,. \tag{2.5}$$

The following lemma states that $\mathcal{P}$ is full in the sense that Ξ coincides with the natural parameter space of $\mathcal{P}$ formally defined as the set $\{\boldsymbol{\zeta} \in \mathbb{R}^2 : \sum_{x=0}^{\infty} e^{\langle \boldsymbol{\zeta}, \boldsymbol{T}(x)\rangle} < \infty\}$, which is the maximum set of parameter vectors $\boldsymbol{\zeta} \in \mathbb{R}^2$ such that $\tilde{f}_{\boldsymbol{\zeta}}$ in formula (2.3) is indeed a density. Moreover, since Ξ is not open, $\mathcal{P}$ is not regular. Crucial and problematic for ML estimation, however, is the fact that $\mathcal{P}$ is not even steep, as shown in Krivitsky (2012). Since this finding is important for our issues, we give an alternative proof in what follows. For this, note that a full EF with natural parameter space Ξ is called steep if for all $\boldsymbol{\eta} \in \Xi \setminus int(\Xi)$ and for all $\boldsymbol{\zeta} \in int(\Xi)$ it holds that

$$\lim_{\alpha \nearrow 1} \frac{\partial}{\partial \alpha} \kappa(\boldsymbol{\zeta} + \alpha(\boldsymbol{\eta} - \boldsymbol{\zeta})) = \infty\,,$$

i.e., if its cumulant function has infinite slope when moving towards boundary points in Ξ on straight lines; see, e.g., Bedbur and Kamps (2021, p. 39).

Lemma 2.2. *$\mathcal{P}$ is full but not steep.*

Proof: Let $\zeta_1, \zeta_2 \in \mathbb{R}$. Moreover, let $a_k = e^{\zeta_1 k}/k!^{\zeta_2}$ for $k \in \mathbb{N}_0$. Then,

$$\frac{a_{k+1}}{a_k} = \frac{e^{\zeta_1}}{(k+1)^{\zeta_2}}\,, \qquad k \in \mathbb{N}_0\,.$$

For $k \to \infty$, this ratio converges to 0 if $\zeta_2 > 0$ and to infinity if $\zeta_2 < 0$. Moreover, for $\zeta_2 = 0$, the ratio equals e^{ζ_1}, which is smaller than 1 if $\zeta_1 < 0$ and larger than 1 if $\zeta_1 > 0$. By the ratio test, the series $\sum_{k=0}^{\infty} a_k$ is then convergent if and only if $\boldsymbol{\zeta} = (\zeta_1, \zeta_2) \in \Xi$. Hence, Ξ is the natural parameter space of $\mathcal{P}$, and $\mathcal{P}$ is full.

To show that $\mathcal{P}$ is not steep, consider the boundary point $\boldsymbol{\eta} = (-1, 0) \in \Xi$ of Ξ and the inner point $\boldsymbol{\zeta} = (-1, 1) \in int(\Xi)$. Moreover, let $\nabla\kappa = (\partial\kappa/\partial\zeta_1, \partial\kappa/\partial\zeta_2)$ denote the gradient of κ. Then, we obtain for $\alpha \in (0, 1)$ that

$$\begin{aligned}\frac{\partial}{\partial \alpha}\kappa(\boldsymbol{\zeta} + \alpha(\boldsymbol{\eta} - \boldsymbol{\zeta})) &= \langle \boldsymbol{\eta} - \boldsymbol{\zeta}\,,\, \nabla\kappa(\boldsymbol{\zeta} + \alpha(\boldsymbol{\eta} - \boldsymbol{\zeta}))\rangle = -\frac{\partial}{\partial \zeta_2}\kappa(\boldsymbol{\zeta})\Big|_{\boldsymbol{\zeta}=(-1,1-\alpha)} \\ &= \left(\sum_{k=0}^{\infty} \frac{e^{-k}}{k!^{1-\alpha}}\right)^{-1} \left(\sum_{k=0}^{\infty} \frac{\ln(k!)\, e^{-k}}{k!^{1-\alpha}}\right),\end{aligned}$$

which, for $\alpha \nearrow 1$, converges to

$$\left(\sum_{k=0}^{\infty} e^{-k}\right)^{-1} \left(\sum_{k=0}^{\infty} \ln(k!) e^{-k}\right) = \frac{e-1}{e}\left(\sum_{k=0}^{\infty} b_k\right),$$

where $b_k = \ln(k!)/e^k$, $k \in \mathbb{N}_0$. The latter series converges by the ratio test, since

$$\frac{b_{k+1}}{b_k} = \frac{1}{e}\left(1 + \frac{\ln(k+1)}{\ln(k!)}\right) \le \frac{2}{e}\,, \qquad k \ge 3\,.$$

By definition, $\mathcal{P}$ is then not steep. □

The mean value function plays a key role regarding ML estimation in EFs. It is defined as the mapping

$$\boldsymbol{\pi} = (\pi_1, \pi_2) : \; int(\Xi) \to int(M) : \; \boldsymbol{\zeta} \mapsto E_{\boldsymbol{\zeta}}[\boldsymbol{T}] = \int \boldsymbol{T}\, d\tilde{P}_{\boldsymbol{\zeta}}\,, \tag{2.6}$$

where M denotes the convex support of $\boldsymbol{T}$ defined as the closed convex hull of the support of $\boldsymbol{T}$. According to Brown (1986, Theorem 2.2, Corollary 2.3, and Corollary 2.5), $\boldsymbol{\pi}$ is one-to-one, infinitely

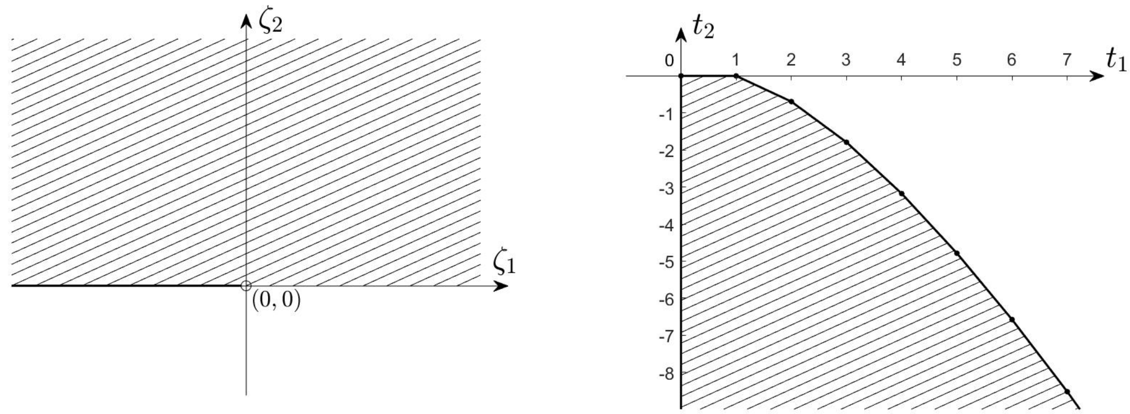


FIGURE 2.1. Natural parameter space Ξ of $\mathcal{P}$ (left, hatched) and convex support M of $\boldsymbol{T}$ (right, hatched). Boundary points belonging to Ξ and M are indicated by bold black lines.

often differentiable, and can be deduced from the identity $\boldsymbol{\pi} = \nabla\kappa$. Using formula (2.4), we find

$$\boldsymbol{\pi}(\boldsymbol{\zeta}) \;=\; \tilde{C}(\boldsymbol{\zeta}) \left(\sum_{k=0}^{\infty} \frac{k e^{\zeta_1 k}}{k!^{\zeta_2}} \,,\, -\sum_{k=0}^{\infty} \frac{\ln(k!)\, e^{\zeta_1 k}}{k!^{\zeta_2}} \right)$$

with

$$\tilde{C}(\boldsymbol{\zeta}) \;=\; \left(\sum_{k=0}^{\infty} \frac{e^{\zeta_1 k}}{k!^{\zeta_2}} \right)^{-1}$$

for $\boldsymbol{\zeta} = (\zeta_1, \zeta_2) \in \mathit{int}(\Xi)$. Here, interchanging differentiation and summation is permitted, for instance, by the differentiation lemma in Bauer (2001, p. 89).

We end this section of preliminary results by stating an explicit representation for the convex support of $\boldsymbol{T}$, the interior of which forms the co-domain of $\boldsymbol{\pi}$. An illustration of Ξ and M is presented in Figure 2.1.

Lemma 2.3. *The convex support of* $\boldsymbol{T}$ *is given by*

$$M \;=\; \{(y+\alpha, z) \in \mathbb{R}^2 : \, y \in \mathbb{N}_0 \,,\, \alpha \in [0,1) \,, z \leq \tilde{T}_2(y,\alpha)\} \,,$$

where

$$\tilde{T}_2(y,\alpha) \;=\; \alpha T_2(y+1) + (1-\alpha) T_2(y) \,, \qquad y \in \mathbb{N}_0 \,, \qquad \alpha \in [0,1) \,, \tag{2.7}$$

with $T_2(y) = -\ln(y!)$, $y \in \mathbb{N}_0$.

Proof: According to formula (2.1), the support of $\boldsymbol{T}$ is given by the graph of $T_2(x) = -\ln(x!)$, $x \in \mathbb{N}_0$. By using the identity $T_2(x) = -\ln\Gamma(x+1)$ for $x \in \mathbb{N}_0$ and that the mapping $t \mapsto -\ln\Gamma(t+1)$ is strictly concave on $[0,\infty)$, it follows that

$$M \;\subset\; \{(y+\alpha, z) \in \mathbb{R}^2 : \, y \in \mathbb{N}_0 \,,\, \alpha \in [0,1) \,, z \leq \tilde{T}_2(y,\alpha)\} \,. \tag{2.8}$$

Now, let $x = y + \alpha > 0$ for some $y \in \mathbb{N}_0$ and $\alpha \in [0,1)$, and let $z \leq \tilde{T}_2(y,\alpha)$. To show that $(x,z) \in M$, note that the line segment $\tau_k(t) = -\ln(k!)t/k$, $t \in [0,k]$, connecting $(0,0) \in M$ and $(k, T_2(k)) \in M$ lies in M for every $k \in \mathbb{N}_0$. By l'Hospital, we find

$$\lim_{k\to\infty} \frac{-\ln(k!)}{k} \;=\; \lim_{t\to\infty} \frac{-\ln\Gamma(t+1)}{t} \;=\; \lim_{t\to\infty} -\psi(t+1) \;=\; -\infty \,,$$

where ψ denotes the digamma function. Hence, there exists some $k_0 \in \mathbb{N}$, $k_0 \geq x$, such that $\tau_{k_0}(x) < z$, and, since (x,z) lies on the vertical line segment connecting $(x, \tilde{T}_2(y,\alpha)) \in M$ and

$(x, \tau_{k_0}(x)) \in M$, it follows that $(x, z) \in M$. Since M is closed by definition, this finding moreover implies that $(0, z) \in M$ for $z \leq 0$. Hence, we have equality in formula (2.8). $\square$

3. ML estimation in the full exponential family of CMP distributions

We now turn to ML estimation of the natural parameter of the CMP distribution based on an iid sample. In this section, ML estimation is studied within the full EF of CMP distributions. For subfamilies of CMP distributions, ML estimation is considered in Section 4.

Let $n \in \mathbb{N}$ and $X_1, \ldots, X_n$ be iid random variables with distribution $\tilde{P}_{\boldsymbol{\zeta}}$ for some $\boldsymbol{\zeta} \in \Xi$. Moreover, let $x_1, \ldots, x_n \in \mathbb{N}_0$ be realizations of $X_1, \ldots, X_n$, and let $\boldsymbol{t}_i = \boldsymbol{T}(x_i)$, $1 \leq i \leq n$. Furthermore, we use the notation $\overline{x}_n = \sum_{i=1}^n x_i/n$ and $\overline{\boldsymbol{t}}_n = \sum_{i=1}^n \boldsymbol{t}_i/n$ for the arithmetic mean of $x_1, \ldots, x_n$ and $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n$, respectively. Note that then $\overline{\boldsymbol{t}}_n = (\overline{x}_n, \sum_{i=1}^n T_2(x_i)/n)$.

We are aiming for ML estimation of the parameter $\boldsymbol{\zeta}$ based on $x_1, \ldots, x_n$. By using formula (2.3), the likelihood function based on $\boldsymbol{x} = (x_1, \ldots, x_n)$ is obtained as

$$\tilde{L}_n(\boldsymbol{\zeta}; \boldsymbol{x}) = \prod_{i=1}^n \tilde{f}_{\boldsymbol{\zeta}}(x_i) = e^{n\langle \boldsymbol{\zeta}, \overline{\boldsymbol{t}}_n\rangle - n\kappa(\boldsymbol{\zeta})}, \qquad \boldsymbol{\zeta} \in \Xi.$$

The corresponding log-likelihood function is then given by

$$\tilde{\ell}_n(\boldsymbol{\zeta}; \boldsymbol{x}) = n[\langle \boldsymbol{\zeta}, \overline{\boldsymbol{t}}_n\rangle - \kappa(\boldsymbol{\zeta})] = n\left[\zeta_1 \overline{x}_n - \frac{\zeta_2}{n}\sum_{i=1}^n \ln(x_i!) - \ln\left(\sum_{k=0}^\infty \frac{e^{\zeta_1 k}}{k!^{\zeta_2}}\right)\right], \qquad \boldsymbol{\zeta} \in \Xi, \tag{3.1}$$

which, by differentiation and by using that $\nabla\kappa = \boldsymbol{\pi}$, leads to the likelihood equation $\boldsymbol{\pi}(\boldsymbol{\zeta}) = \overline{\boldsymbol{t}}_n$ for $\boldsymbol{\zeta} \in int(\Xi)$.

ML estimation for the natural parameter $\boldsymbol{\zeta}$ of the general (minimal) EF has extensively been discussed in the literature, and there exists a simple rule for the existence and uniqueness of an ML estimate for $\boldsymbol{\zeta}$ if the EF is steep; see Barndorff-Nielsen (2014) and Brown (1986). It states that an ML estimate of $\boldsymbol{\zeta}$ based on $x_1, \ldots, x_n$ exists if and only if $\overline{\boldsymbol{t}}_n \in int(M)$, and that, in case of existence, the ML estimate is uniquely determined and given by $\boldsymbol{\pi}^{-1}(\overline{\boldsymbol{t}}_n)$, where $\boldsymbol{\pi}^{-1}$ denotes the inverse function of $\boldsymbol{\pi}$. This simple statement results from the steepness property of the EF in the following way. Firstly, steepness of the EF ensures that $\boldsymbol{\pi}$ is onto (and thus bijective), i.e., $\boldsymbol{\pi}(int(\Xi)) = int(M)$, such that there exists an inverse function $\boldsymbol{\pi}^{-1} : int(M) \to int(\Xi)$ of $\boldsymbol{\pi}$. If $\overline{\boldsymbol{t}}_n \in int(M)$, the likelihood equation then has the unique solution $\boldsymbol{\pi}^{-1}(\overline{\boldsymbol{t}}_n) \in int(\Xi)$, which is also the unique global maximum of $\tilde{\ell}_n$, since $\tilde{\ell}_n$ is strictly concave on Ξ; see, e.g., Bedbur and Kamps (2021, Lemma 4.2(a)). Secondly, steepness of $\mathcal{P}$ guarantees that maxima of the (log-)likelihood function cannot lie on the boundary of Ξ. Putting both findings together then yields the above rule.

The EF of CMP distributions, however, is not steep, such that ML estimation gets more involved. We have the following two major drawbacks. The first one is that $\overline{\boldsymbol{t}}_n \in int(M)$ does not imply existence of a solution of the likelihood equation, since it may happen that $\overline{\boldsymbol{t}}_n \in int(M) \setminus \boldsymbol{\pi}(int(\Xi))$. Hence, ML estimation of $\boldsymbol{\zeta}$ requires to study the image of $\boldsymbol{\pi}$. Since the properties hold true in the general EF, $\boldsymbol{\pi}(int(\Xi))$ is open, connected, and without holes, but it need not to be convex; see, e.g., Kotz et al. (2000, p. 668). An illustration of $\boldsymbol{\pi}(int(\Xi))$ is provided in Figure 3.1. The second drawback is that the boundary of Ξ has to be taken into account when maximizing $\tilde{\ell}_n$.

Our aim is now to find conditions for the existence or non-existence of an ML estimate of $\boldsymbol{\zeta}$ based on $x_1, \ldots, x_n$. Note that in case of existence, regardless whether it is the (unique) solution of the likelihood equation or lies on the boundary of Ξ, the ML estimate is uniquely determined, since $\tilde{\ell}_n$ is strictly concave on Ξ.

Firstly, the following statement is obvious.

Lemma 3.1. *Let $x_1, \ldots, x_n \in \mathbb{N}_0$. If $\overline{\boldsymbol{t}}_n \in \boldsymbol{\pi}(int(\Xi))$, then $\pi^{-1}(\overline{\boldsymbol{t}}_n)$ is the solution of the likelihood equation and the ML estimate of $\boldsymbol{\zeta}$ based on $x_1, \ldots, x_n$.*

Unfortunately, the region $\boldsymbol{\pi}(int(\Xi))$, depicted in Figure 3.1, cannot be described in a useful and convenient form, such that the benefit of Lemma 3.1 is rather low. In contrast, a simple sufficient condition for the non-existence of the ML estimate can be provided, which is developed in the following.

By definition of M, $\overline{\boldsymbol{t}}_n \in M$ is always true (with probability 1). Since $\overline{\boldsymbol{t}}_n \in \partial M = M \setminus int(M)$ implies that the likelihood equation has no solution in $int(\Xi)$, we give a respective characterization result. It turns out that $\overline{\boldsymbol{t}}_n \in int(M)$ is equivalent to the range of observations being at least 2. In what follows, let $x_{(1)} \leq \cdots \leq x_{(n)}$ denote the ascendingly ordered observations.

Lemma 3.2. *Let $x_1, \ldots, x_n \in \mathbb{N}_0$. Then, it holds:*

(a) *For $n = 1$, we have $\overline{\boldsymbol{t}}_1 = (x_1, T_2(x_1)) \notin int(M)$.*

(b) *For $n \geq 2$, we have*

$$\overline{\boldsymbol{t}}_n \in int(M) \qquad \Leftrightarrow \qquad x_{(n)} - x_{(1)} \geq 2\,.$$

Proof: Let $x_1, \ldots, x_n \in \mathbb{N}_0$. Assertion (a) is obvious by Lemma 2.3. Hence, let $n \geq 2$. Moreover, let

$$\overline{x}_n = y + \alpha\,,$$

where $y = \lfloor \overline{x}_n \rfloor \in \mathbb{N}_0$ denotes the integer part of $\overline{x}_n$, and $\alpha = \overline{x}_n - \lfloor \overline{x}_n \rfloor \in [0, 1)$.

Then, we have to show that

$$x_{(n)} - x_{(1)} \geq 2 \qquad \Leftrightarrow \qquad \frac{1}{n}\sum_{i=1}^{n} T_2(x_i) < \tilde{T}_2(y, \alpha)$$

with $\tilde{T}_2(y, \alpha) = \alpha T_2(y+1) + (1-\alpha) T_2(y)$ or, equivalently,

$$x_{(n)} - x_{(1)} \geq 2 \qquad \Leftrightarrow \qquad \left(\prod_{i=1}^{n} x_i!\right)^{\frac{1}{n}} > y!\,(y+1)^{\alpha}\,. \tag{3.2}$$

Firstly, let $x_{(n)} - x_{(1)} \leq 1$. Then, $y = x_{(1)}$. Let $K = \{i \in \{1, \ldots, n\} : x_i = y\}$ and $k = |K|$. Then, we have

$$\alpha = \overline{x}_n - y = \frac{ky + (n-k)(y+1)}{n} - y = 1 - \frac{k}{n}$$

and

$$\left(\prod_{i=1}^{n} x_i!\right)^{\frac{1}{n}} = \left(\prod_{i \in K} x_i!\right)^{\frac{1}{n}} \left(\prod_{i \notin K} x_i!\right)^{\frac{1}{n}} = y!^{\frac{k}{n}}(y+1)!^{1-\frac{k}{n}} = y!(y+1)^{\alpha} \tag{3.3}$$

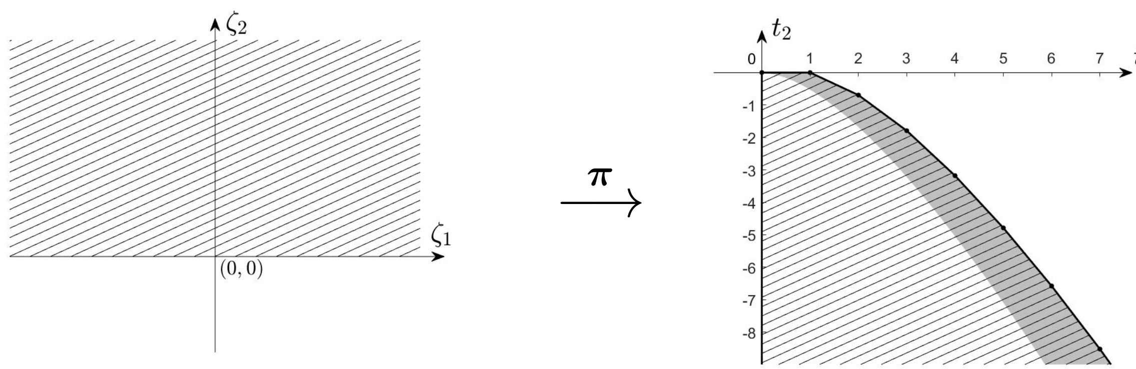


Figure 3.1. Interior $int(\Xi)$ of Ξ (left, hatched) and its image $\boldsymbol{\pi}(int(\Xi))$ (right, grey area) being a proper subset of $int(M)$ (right, hatched).

Next, let $x_{(n)} - x_{(1)} \geq 2$. Then, we apply the following algorithm:

(1) Choose x_i minimally with $x_i + 1 < x_j$ for some $j \in \{1, \ldots, n\}$.

(2) Replace x_i by $x_i + 1$ and x_j by $x_j - 1$, which preserves the sum $\sum_{m=1}^n x_m$ (and thus y and α) but decreases the product $\prod_{m=1}^n x_m!$, since

$$(x_i + 1)!(x_j - 1)! \, < \, x_i! x_j! \quad \Leftrightarrow \quad x_i + 1 < x_j \, .$$

(3) If $x_{(n)} - x_{(1)} \geq 2$, return to step (1). Otherwise, the algorithm is completed and ends up with $x_1, \ldots, x_n \in \{y, y+1\}$ for which equation (3.3) holds true.

Convergence of the algorithm is ensured, since the range of $x_1, \ldots, x_n$ is non-increasing over all iterations and decreases by at least 1 in n iterations.

The assumption that $x_{(n)} - x_{(1)} \geq 2$ now guarantees that we have to go through at least one iteration and, hence, that the right-hand side of equivalence (3.2) is true. □

In Theorem 3.3, a sufficient condition for the non-existence of the ML estimate for $\boldsymbol{\zeta}$ is given, namely, that all observations take only two different values m and $m+1$, say. This is different from the common Poisson model forming a regular EF, where ML estimation of the parameter only fails in the trivial case $x_1 = \cdots = x_n = 0$.

Theorem 3.3. *Let $x_1, \ldots, x_n \in \mathbb{N}_0$. If $x_{(n)} - x_{(1)} \leq 1$, then the ML estimate for $\boldsymbol{\zeta}$ based on $x_1, \ldots, x_n$ does not exist.*

Proof: Let $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{N}_0^n$ with $x_{(n)} - x_{(1)} \leq 1$. For $\overline{x}_n = 0$, the assertion is obvious from $\tilde{L}_n(\boldsymbol{\zeta}; \boldsymbol{x}) = \tilde{C}^n(\boldsymbol{\zeta}) < 1$, $\boldsymbol{\zeta} \in \Xi$, which approximates 1, e.g., for $\zeta_1 \to -\infty$ and $\zeta_2 = 1$. Hence, let $\overline{x}_n > 0$. According to Lemma 3.2, we have that $\overline{\boldsymbol{t}}_n \notin int(M)$. Since

$$\boldsymbol{\pi}(int(\Xi)) \, \subset \, int(M) \, ,$$

the likelihood equation $\boldsymbol{\pi}(\boldsymbol{\zeta}) = \overline{\boldsymbol{t}}_n$ therefore has no solution in $int(\Xi)$. What is left to show is that the maximum of the likelihood function $\tilde{L}_n(\cdot; \boldsymbol{x})$ is not attained at the boundary $\{(\zeta_1, 0) : \zeta_1 < 0\}$ of Ξ. For this, we show that

$$\tilde{L}_n(\ln(\overline{x}_n), 1; \boldsymbol{x}) \, \geq \, \tilde{L}_n\left(\ln\left(\frac{\overline{x}_n}{1 + \overline{x}_n}\right), 0; \boldsymbol{x} \right) , \tag{3.4}$$

where $\ln(\overline{x}_n)$ is the ML estimate of $\zeta_1 = \ln(\lambda)$ in the submodel of Poisson distributions, and $\ln(\overline{x}_n / (1 + \overline{x}_n))$ is the ML estimate of ζ_1 in the submodel of geometric distributions maximizing $\tilde{L}_n$ on the boundary of Ξ. We have

$$\tilde{L}_n(\ln(\overline{x}_n), 1; \boldsymbol{x}) \; = \; e^{-n\overline{x}_n} \frac{\overline{x}_n^{n\overline{x}_n}}{\prod_{i=1}^n x_i!} \qquad \text{and} \qquad \tilde{L}_n\left(\ln\left(\frac{\overline{x}_n}{1 + \overline{x}_n}\right), 0; \boldsymbol{x} \right) \; = \; \frac{\overline{x}_n^{n\overline{x}_n}}{(1 + \overline{x}_n)^{n\overline{x}_n + n}}$$

such that inequality (3.4) is equivalent to

$$e^{-\overline{x}_n} (1 + \overline{x}_n)^{1 + \overline{x}_n} \; \geq \; \left(\prod_{i=1}^n x_i! \right)^{\frac{1}{n}} . \tag{3.5}$$

By inspecting the proof of Lemma 3.2 and by using formula (3.3), the assumption $x_{(n)} - x_{(1)} \leq 1$ yields that

$$\left(\prod_{i=1}^n x_i! \right)^{\frac{1}{n}} \; = \; \lfloor \overline{x}_n \rfloor ! (\lfloor \overline{x}_n \rfloor + 1)^{\overline{x}_n - \lfloor \overline{x}_n \rfloor} \, ,$$

such that inequality (3.5) is equivalent to $g(\overline{x}_n) \geq 0$ with mapping

$$g(a) \; = \; (1 + a) \ln(1 + a) - a - \ln(\lfloor a \rfloor !) - (a - \lfloor a \rfloor) \ln(\lfloor a \rfloor + 1) \, , \qquad a \geq 0 \, .$$

Note that g is continuous. Moreover, g is differentiable at $a \notin \mathbb{N}_0$ with derivative

$$g'(a) \;=\; \ln(1+a) - \ln(1+\lfloor a \rfloor) \;>0\,,$$

which implies that g is non-decreasing on $[0,\infty)$. It follows that $g(\overline{x}_n) \geq g(0) = 0$ and, hence, that inequalities (3.5) and (3.4) are true. □

Remark 3.4. Based on an iid sample $X_1, \ldots, X_n$ from $P_{\boldsymbol{\vartheta}}$ for some $\boldsymbol{\vartheta} = (\lambda, \nu) \in \Theta$, we consider the probability $\mathbb{P}_{\boldsymbol{\vartheta}}(X_{(n)} - X_{(1)} \leq 1)$ of observing a range of at most 1.

On the one hand, since $P_{\boldsymbol{\vartheta}}$ converges in distribution to a Bernoulli distribution (with support $\{0,1\}$) as ν increases, we have

$$\begin{aligned}\mathbb{P}_{\boldsymbol{\vartheta}}(X_{(n)} - X_{(1)} \leq 1) \;&\geq \mathbb{P}_{\boldsymbol{\vartheta}}\left(\bigcap_{i=1}^{n}\{X_i \in \{0,1\}\}\right)\\ &= \prod_{i=1}^{n}\mathbb{P}_{\boldsymbol{\vartheta}}(X_i \in \{0,1\}) \;\longrightarrow\; \prod_{i=1}^{n} ber\Big(\frac{\lambda}{1+\lambda}\Big)(\{0,1\}) \;=\; 1\end{aligned}$$

for $\nu \to \infty$. Hence, for large values of ν, the probability of $X_{(n)} - X_{(1)} \leq 1$ is approximately 1.

On the other hand, if we set $\nu = 0$ (and thus have $\lambda \in (0,1)$), we find

$$\begin{aligned}\mathbb{P}_{\boldsymbol{\vartheta}}(X_{(n)} - X_{(1)} \leq 1) \;&=\; \sum_{y=0}^{\infty}\sum_{k=1}^{n}\binom{n}{k} f_{\boldsymbol{\vartheta}}^{k}(y)\, f_{\boldsymbol{\vartheta}}^{n-k}(y+1) \;=\; (1-\lambda)^n\lambda^n \sum_{y=0}^{\infty}\lambda^{ny}\sum_{k=1}^{n}\binom{n}{k}\lambda^{-k}\\ &=\; (1-\lambda)^n\lambda^n\sum_{y=0}^{\infty}\lambda^{ny}\left[\Big(1+\frac{1}{\lambda}\Big)^n - 1\right] \;=\; \frac{[1-\lambda^2]^n - [\lambda(1-\lambda)]^n}{1-\lambda^n}\,,\end{aligned}$$

which converges to 1 as λ tends to 0.

These examples demonstrate that there are under- and overdispersed CMP distributions, for which the event $X_{(n)} - X_{(1)} \leq 1$ (and thus that ML estimation fails) is likely to occur or at least has a probability that may not be neglected.

We can rephrase the result of Theorem 3.3 as follows.

Corollary 3.5. *$x_{(n)} - x_{(1)} \geq 2$ is a necessary condition for the existence of the ML estimate of $\boldsymbol{\zeta}$ based on $x_1, \ldots, x_n$.*

So far, we have established that an ML estimate of $\boldsymbol{\zeta}$ does not exist if $\overline{\boldsymbol{t}}_n \in \partial M$ and that it exists and is the only solution of the likelihood equation if $\overline{\boldsymbol{t}}_n \in \boldsymbol{\pi}(int(\Xi))$. What is left to discuss is the case $\overline{\boldsymbol{t}}_n \in int(M) \setminus \boldsymbol{\pi}(int(\Xi))$, i.e., when $\overline{\boldsymbol{t}}_n$ lies below the image of $\boldsymbol{\pi}$ depicted in Figure 3.1. Numerical studies indicate that in this case the ML estimate of $\boldsymbol{\zeta}$ exists and is given by the boundary point $\hat{\boldsymbol{\zeta}} = (\ln(\overline{x}_n/(1+\overline{x}_n)), 0)$ of Ξ corresponding to a geometric distribution; see Figure 1.2, case $\boldsymbol{x} = \boldsymbol{x}^{(2)}$, for an example. In this context, it is worth mentioning that the lower boundary of $\boldsymbol{\pi}(int(\Xi))$ seems to be the image of $\{\boldsymbol{\zeta} \in \Xi : \zeta_2 = 0\}$ under $\boldsymbol{\pi}$ when extending the domain of $\boldsymbol{\pi}$ in formula (2.6) to Ξ. By doing so we have

$$\pi_1(\hat{\boldsymbol{\zeta}}) \;=\; E_{\hat{\boldsymbol{\zeta}}}[T_1] \;=\; \frac{e^{\ln(\overline{x}_n/(1+\overline{x}_n))}}{1 - e^{\ln(\overline{x}_n/(1+\overline{x}_n))}} \;=\; \overline{x}_n$$

by using the formula for the mean of the geometric distribution, and $\boldsymbol{\pi}$ therefore maps the ML estimate to the point $(\overline{x}_n, \pi_2(\hat{\boldsymbol{\zeta}}))$ on the lower boundary of $\boldsymbol{\pi}(int(\Xi))$ lying above $\overline{\boldsymbol{t}}_n$; see Figure 3.2 for an illustration. However, analytic proofs of these findings cannot be provided yet. When restricting the model to equi- and underdispersed CMP distributions, some related results can be derived and will be presented later.

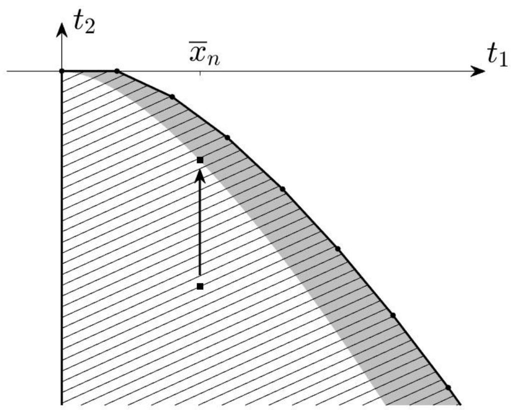


FIGURE 3.2. Case $\overline{\boldsymbol{t}}_n \in int(M) \setminus \pi(int(\Xi))$: the observation $\overline{\boldsymbol{t}}_n$ (lower square) leads to the ML estimate $\hat{\boldsymbol{\zeta}} = (\ln(\overline{x}_n/(1+\overline{x}_n)), 0)$ whose image $\boldsymbol{\pi}(\hat{\boldsymbol{\zeta}})$ (upper square) lies on the lower boundary of $\pi(int(\Xi))$ (grey area) and has coordinate $\overline{x}_n$ on the abscissa such as $\overline{\boldsymbol{t}}_n$.

4. ML estimation in subfamilies of CMP distributions

Having discussed ML estimation in the full EF of CMP distributions in Section 3, we now focus on conditions for the existence and non-existence of ML estimates in restricted models. For this, we first introduce the subsets

$$\Xi_{a,\bullet} \;=\; \{\boldsymbol{\zeta} = (\zeta_1, \zeta_2) \in \Xi : \zeta_1 \leq a\}$$

$$\text{and} \qquad \Xi_{\bullet,b} \;=\; \{\boldsymbol{\zeta} = (\zeta_1, \zeta_2) \in \Xi : \zeta_2 \leq b\}$$

of the natural parameter space Ξ given by formula (2.5), where $a \in \mathbb{R}$ and $b \geq 0$. According to the relations $\zeta_1 = \ln(\lambda)$ and $\zeta_2 = \nu$, $\Xi_{a,\bullet}$ and $\Xi_{\bullet,b}$ are associated with families of CMP distributions with an upper bound for either λ or ν. In particular, $\Xi_{\bullet,1}$ corresponds to the family of equi- and overdispersed CMP distributions.

Necessary and sufficient conditions for the existence of the ML estimates in $\Xi_{a,\bullet}$ and $\Xi_{\bullet,b}$ are stated in the following theorem. Note that, in case of existence, the ML estimates are uniquely determined, since $\Xi_{a,\bullet}$ and $\Xi_{\bullet,b}$ are convex sets and $\tilde{\ell}_n$ is strictly concave on Ξ.

Theorem 4.1. *Let $x_1, \ldots, x_n \in \mathbb{N}_0$. Moreover, let $a \in \mathbb{R}$ and $b \geq 0$ be arbitrary.*

(a) *The ML estimate for $\boldsymbol{\zeta}$ in $\Xi_{a,\bullet}$ exists based on $x_1, \ldots, x_n$ if and only if $x_i \geq 2$ for some $i \in \{1, \ldots, n\}$.*
(b) *The ML estimate for $\boldsymbol{\zeta}$ in $\Xi_{\bullet,b}$ exists based on $x_1, \ldots, x_n$ if and only if $\overline{x}_n > 0$.*

Proof: (a) Let $a \in \mathbb{R}$ be arbitrary. Firstly, let $x_i \in \{0, 1\}$ for all $i \in \{1, \ldots, n\}$. According to Lemma 3.2, the likelihood equation $\boldsymbol{\pi}(\boldsymbol{\zeta}) = \overline{\boldsymbol{t}}_n$ $(= (\overline{x}_n, 0))$ has no solution in $int(\Xi)$ and, thus, no solution in $int(\Xi_{a,\bullet})$. Hence, if $\tilde{\ell}_n$ would have a global maximum over $\Xi_{a,\bullet}$, it must be attained at the boundary of $\Xi_{a,\bullet}$. Since $poi(e^{\zeta_1})(\{x\}) > geo(1 - e^{\zeta_1})(\{x\})$ for $x \in \{0, 1\}$ and all $\zeta_1 < 0$, a global maximum of $\tilde{\ell}_n$ cannot lie on the lower boundary $(-\infty, \min\{a, 0\}) \times \{0\}$. Moreover, regarding the right boundary of $\Xi_{a,\bullet}$, formula (3.1) yields that

$$\tilde{\ell}_n(a, \zeta_2; \boldsymbol{x}) \;=\; n \left[a\overline{x}_n - \ln\left(\sum_{k=0}^{\infty} \frac{e^{ak}}{k!^{\zeta_2}} \right) \right]$$

is strictly increasing in ζ_2. Together, these findings imply that the ML estimate for $\boldsymbol{\zeta}$ in $\Xi_{a,\bullet}$ does not exist.

Now, let $x_i \geq 2$ for some $i \in \{1, \ldots, n\}$. From formula (3.1), we obtain for $\boldsymbol{\zeta} \in \Xi_{a,\bullet}$ that

$$\tilde{\ell}_n(\boldsymbol{\zeta}; \boldsymbol{x}) \;\leq\; n\zeta_1\overline{x}_n \overset{\zeta_1 \to -\infty}{\longrightarrow} -\infty$$

uniformly in ζ_2, and

$$\tilde{\ell}_n(\boldsymbol{\zeta};\boldsymbol{x}) \;\leq\; n\left[a\overline{x}_n - \frac{\zeta_2}{n}\sum_{i=1}^{n}\ln(x_i!)\right] \stackrel{\zeta_2\to\infty}{\longrightarrow} -\infty$$

uniformly in ζ_1. Hence, for every $M > 0$, there exists some $\eta > \max\{0, -a\}$, such that $\tilde{\ell}_n(\boldsymbol{\zeta};\boldsymbol{x}) < -M$ for $\zeta_1 < -\eta$ or $\zeta_2 > \eta$. With the common convention $\ln(\infty) = \infty$, $\tilde{\ell}_n$ forms an upper semi-continuous function on $\mathbb{R}^2$ (see, e.g., Bedbur and Kamps, 2021, Lemma 4.2), which has a global maximum on the compact set $[-\eta, a] \times [0, \eta]$. Since $\tilde{\ell}_n(\zeta_1, 0; \boldsymbol{x}) = -\infty$ for $\zeta_1 \geq 0$, this maximum is attained at some $\boldsymbol{\zeta} \in \Xi_{a,\bullet}$, and existence of the ML estimate for $\boldsymbol{\zeta}$ in $\Xi_{a,\bullet}$ is guaranteed.

(b) Let $b \geq 0$ be arbitrary. For $\overline{x}_n = 0$, the ML estimate for $\boldsymbol{\zeta}$ in $\Xi_{\bullet,b}$ does not exist, since then $\tilde{L}_n(\boldsymbol{\zeta};\boldsymbol{x}) = (1 - e^{\zeta_1})^n$ for $\boldsymbol{\zeta} = (\zeta_1, 0) \in \Xi_{\bullet,b}$, which tends to 1 for $\zeta_1 \to -\infty$. Now, let $\overline{x}_n > 0$. By formula (3.1), we have that

$$\tilde{\ell}_n(\boldsymbol{\zeta};\boldsymbol{x}) \;\leq\; n\left[\zeta_1\overline{x}_n - \ln\left(\sum_{k=0}^{\infty}\frac{e^{\zeta_1 k}}{k!^{\zeta_2}}\right)\right] \;\leq\; n\left[\zeta_1(\overline{x}_n - k_0) + b\ln(k_0!)\right]$$

for $\boldsymbol{\zeta} \in \Xi_{\bullet,b}$ and every $k_0 \in \mathbb{N}_0$. Choosing $k_0 > \overline{x}_n$ then yields that $\tilde{\ell}_n(\boldsymbol{\zeta};\boldsymbol{x}) \to -\infty$ for $\zeta_1 \to \infty$ uniformly in ζ_2. Moreover, choosing $k_0 < \overline{x}_n$ ensures that $\tilde{\ell}_n(\boldsymbol{\zeta};\boldsymbol{x}) \to -\infty$ for $\zeta_1 \to -\infty$ uniformly in ζ_2. Hence, for every $M > 0$, there exists some $\eta > 0$, such that $\tilde{\ell}_n(\boldsymbol{\zeta};\boldsymbol{x}) < -M$ for $|\zeta_1| > \eta$. Again, upper semi-continuity of $\tilde{\ell}_n$ then ensures that $\tilde{\ell}_n$ has a global maximum on $[-\eta, \eta] \times [0, b]$, which must be attained at some $\boldsymbol{\zeta} \in \Xi_{\bullet,b}$. Hence, existence of the ML estimate for $\boldsymbol{\zeta}$ in $\Xi_{\bullet,b}$ is guaranteed. □

As seen in Figure 3.1, the set $\boldsymbol{\pi}(int(\Xi))$ is bounded from above by the graph of $\tilde{T}_2$, which results from the graph of T_2 by connecting adjacent points with lines; see formula (2.7). Moreover, as mentioned in Section 3, simulations suggest that it is bounded from below by the image of the boundary $\{\boldsymbol{\zeta} \in \Xi : \zeta_2 = 0\}$ of Ξ under $\boldsymbol{\pi}$, a proof of which is missing. However, in the restricted model with parameter space

$$\Delta \;=\; \{\boldsymbol{\zeta} = (\zeta_1, \zeta_2) \in \Xi : \zeta_2 \geq 1\}$$

consisting of equi- and underdispersed CMP distributions, a lower bound for $\boldsymbol{\pi}(\Delta)$ can be derived analytically, which may then be used to ensure that, in Δ, a solution of the likelihood equation does not exist.

Lemma 4.2. *For $\boldsymbol{\zeta} \in \Delta$, we have*

$$\pi_2(\boldsymbol{\zeta}) \;\geq\; \max\left\{-\,\frac{\pi_1^2(\boldsymbol{\zeta})}{2}\,,\, -c\left(\frac{\pi_1^2(\boldsymbol{\zeta})}{2} + \pi_1(\boldsymbol{\zeta})\right)\right\}$$

$$= \begin{cases} -\,\frac{\pi_1^2(\boldsymbol{\zeta})}{2}\,, & \pi_1(\boldsymbol{\zeta}) \;\leq\; \frac{2c}{1-c}\,, \\ -\,c\left(\frac{\pi_1^2(\boldsymbol{\zeta})}{2} + \pi_1(\boldsymbol{\zeta})\right), & \pi_1(\boldsymbol{\zeta}) \;>\; \frac{2c}{1-c}\,, \end{cases}$$

with constant $c = e^{1/e} - 1$.

Proof: Let $\boldsymbol{\zeta} = (\zeta_1, \zeta_2) \in \Delta$ and $X \sim \tilde{P}_{\boldsymbol{\zeta}}$. By using that $\ln(y) \leq y - 1$ for $y > 0$, we find

$$\pi_2(\boldsymbol{\zeta}) \;=\; -E_{\boldsymbol{\zeta}}[\ln(X!)] \;=\; -E_{\boldsymbol{\zeta}}\left[\sum_{k=1}^{X}\ln(k)\right] \;\geq\; -E_{\boldsymbol{\zeta}}\left[\sum_{k=1}^{X}(k-1)\right]$$

$$= -E_{\boldsymbol{\zeta}}\left[\frac{X(X+1)}{2} - X\right] = -\,\frac{E_{\boldsymbol{\zeta}}[X^2 - X]}{2} \;=\; -\,\frac{\mathrm{Var}_{\boldsymbol{\zeta}}(X) + E_{\boldsymbol{\zeta}}^2[X] - E_{\boldsymbol{\zeta}}[X]}{2}\,.$$

According to, e.g., Daly and Gaunt (2016, Prop. 2.12(ii)), it holds that $\mathrm{Var}_{\boldsymbol{\zeta}}(X) \le E_{\boldsymbol{\zeta}}[X]$, such that

$$\pi_2(\boldsymbol{\zeta}) \ge -\frac{E_{\boldsymbol{\zeta}}^2[X]}{2} = -\frac{\pi_1^2(\boldsymbol{\zeta})}{2}$$

is shown.

The second bound is obtained by using the inequality $\ln(k) \le m(k^{1/m} - 1)$, $m, k > 0$. Choosing $m = k$, we find

$$\begin{aligned}\pi_2(\boldsymbol{\zeta}) &= -E_{\boldsymbol{\zeta}}\left[\sum_{k=1}^{X} \ln(k)\right] \ge -E_{\boldsymbol{\zeta}}\left[\sum_{k=1}^{X} k(k^{\frac{1}{k}} - 1)\right] \ge -c\,E_{\boldsymbol{\zeta}}\left[\sum_{k=1}^{X} k\right] \\ &= -cE_{\boldsymbol{\zeta}}\left[\frac{X(X+1)}{2}\right] = -c\left(\frac{\mathrm{Var}_{\boldsymbol{\zeta}}(X) + E_{\boldsymbol{\zeta}}^2[X] + E_{\boldsymbol{\zeta}}[X]}{2}\right) \ge -c\Big(\frac{\pi_1^2(\boldsymbol{\zeta})}{2} + \pi_1(\boldsymbol{\zeta})\Big),\end{aligned}$$

by using that $k^{1/k} \le e^{1/e}$, $k > 0$, and $\mathrm{Var}_{\boldsymbol{\zeta}}(X) \le E_{\boldsymbol{\zeta}}[X]$, again. □

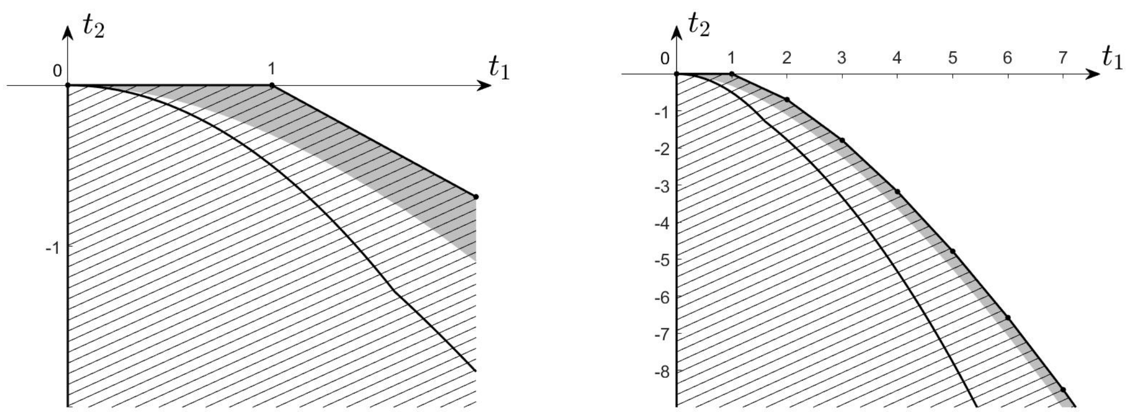


FIGURE 4.1. Lower bound in Lemma 4.2 (dark grey curve) and image $\boldsymbol{\pi}(\Delta)$ (grey area) for $t_1 = \pi_1(\boldsymbol{\zeta}) \in [0, 2]$ and $t_1 = \pi_1(\boldsymbol{\zeta}) \in [0, 7]$.

Figure 4.1 illustrates the lower bound for $\boldsymbol{\pi}(\Delta)$ stated in Lemma 4.2. Additionally, the bound is depicted in Figure 4.2 together with the superset $\boldsymbol{\pi}(int(\Xi))$ of $\boldsymbol{\pi}(\Delta)$. It is found that the lower bound established for $\boldsymbol{\pi}(\Delta)$ also yields an approximate lower bound for $\boldsymbol{\pi}(int(\Xi))$, which is violated only for $\pi_1 \le 1$ and there only slightly.

Lemma 4.2 can now be used to formulate the following result.

Theorem 4.3. *Let* $x_1, \ldots, x_n \in \mathbb{N}_0$. *If*

$$\frac{1}{n}\sum_{i=1}^{n} \ln(x_i!) > \min\left\{\frac{\overline{x}_n^2}{2}, c\left(\frac{\overline{x}_n^2}{2} + \overline{x}_n\right)\right\}, \tag{4.1}$$

where $c = e^{1/e} - 1$, *then a solution of the likelihood equation* $\boldsymbol{\pi}(\boldsymbol{\zeta}) = \overline{\boldsymbol{t}}_n$, $\boldsymbol{\zeta} \in \Delta$, *does not exist.*

Proof: If condition (4.1) is true, we have

$$\frac{1}{n}\sum_{i=1}^{n} T_2(x_i) < \max\left\{-\frac{\overline{x}_n^2}{2}, -c\left(\frac{\overline{x}_n^2}{2} + \overline{x}_n\right)\right\},$$

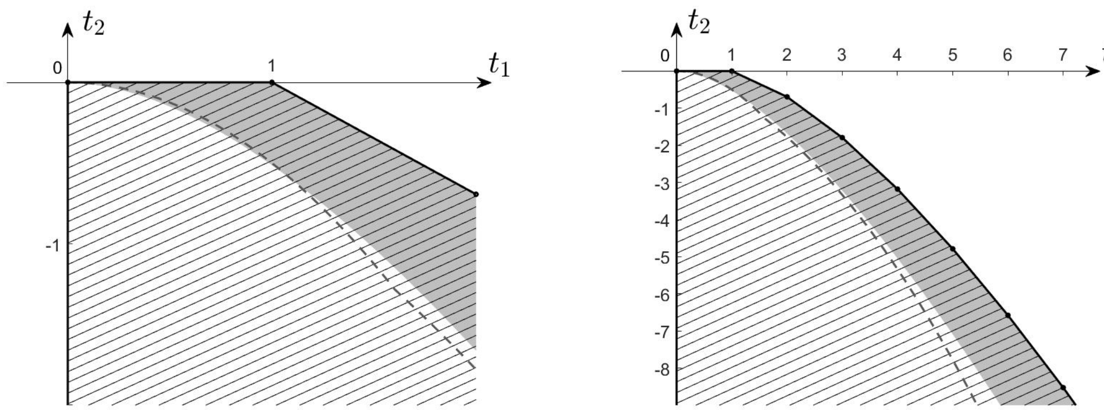


FIGURE 4.2. Lower bound in Lemma 4.2 (dark grey curve) and image $\boldsymbol{\pi}(int(\Xi))$ (grey area) for $t_1 = \pi_1(\boldsymbol{\zeta}) \in [0,2]$ and $t_1 = \pi_1(\boldsymbol{\zeta}) \in [0,7]$. The bound is dashed to indicate that it is not exact.

Lemma 4.2 then yields that the likelihood equation

$$\boldsymbol{\pi}(\boldsymbol{\zeta}) \;=\; \left(\overline{x}_n\,,\; \frac{1}{n}\sum_{i=1}^{n} T_2(x_i)\right)$$

has no solution in Δ. □

Note that Theorem 4.3 remains valid if Δ is replaced by the subset $\{(\zeta_1,\zeta_2) \in \Xi : \zeta_2 \geq b\}$ for some arbitrary $b \geq 1$. Here, the condition $b \geq 1$ ensures that every CMP distribution in the associated family is not overdispersed and, hence, has variance not greater than its mean, which is used in the proof of Lemma 4.2.

Finally, we give some data examples in which condition (4.1) is or is not met.

Example 4.4. (a) Condition (4.1) is fulfilled for $x_1,\dots,x_n \in \{0,1\dots,10\}$ only for the following data vectors (and their permutations):
- $n = 2$: $-$,
- $n = 3$: $(2,0,0),\dots,(10,0,0)$, $(7,1,0),\dots,(10,1,0)$,
- $n = 4$: $(2,0,0,0),\dots,(10,0,0,0)$, $(4,1,0,0),\dots,(10,1,0,0)$, $(6,2,0,0),\dots,(10,2,0,0)$, $(7,3,0,0),\dots,(10,3,0,0)$, $(7,4,0,0),\dots,(10,4,0,0)$, $(8,5,0,0),\dots,(10,5,0,0)$, $(8,6,0,0),\dots,(10,6,0,0)$, $(10,7,0,0)$, $(7,1,1,0),\dots,(10,1,1,0)$, $(9,2,1,0)$, $(10,2,1,0)$.

(b) Condition (4.1) is not fulfilled for $x_1,\dots,x_n \in \mathbb{N}_0$ with $x_{(n)} - x_{(1)} \leq 1$: In that case, formula (3.3) yields

$$\frac{1}{n}\sum_{i=1}^{n} \ln(x_i!) \;=\; \sum_{k=1}^{y} \ln(k) + \alpha\ln(y+1)$$

with $y = \lfloor \overline{x}_n \rfloor \in \mathbb{N}_0$ and $\alpha = \overline{x}_n - \lfloor \overline{x}_n \rfloor \in [0,1)$, where

$$\sum_{k=1}^{y} \ln(k) + \alpha\ln(y+1) \;\leq\; \sum_{k=1}^{y}(k-1) + \alpha y \;=\; \frac{y^2-y}{2} + \alpha y \;\leq\; \frac{(y+\alpha)^2}{2} \;=\; \frac{\overline{x}_n^2}{2}$$

and also

$$
\begin{aligned}
\sum_{k=1}^{y} \ln(k) + \alpha \ln(y+1) &\leq \sum_{k=1}^{y} k(k^{\frac{1}{k}} - 1) + \alpha(y+1)[(y+1)^{\frac{1}{y+1}} - 1] \\
&\leq c\left[\frac{y^2+y}{2} + \alpha(y+1)\right] \leq c\left[\frac{(y+\alpha)^2}{2} + y + \alpha\right] \\
&= c\left(\frac{\overline{x}_n^2}{2} + \overline{x}_n\right).
\end{aligned}
$$

Nevertheless, the likelihood equation has no solution in Δ in case $x_{(n)} - x_{(1)} \leq 1$, since it has no solution in the superset $int(\Xi)$ by Theorem 3.3.

References

Barndorff-Nielsen, O. *Information and exponential families in statistical theory*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester (2014). ISBN 978-1-118-85750-2. MR3221776.

Bauer, H. *Measure and integration theory*, volume 26 of *De Gruyter Studies in Mathematics*. Walter de Gruyter & Co., Berlin (2001). ISBN 3-11-016719-0. MR1897176.

Bedbur, S. and Kamps, U. *Multivariate exponential families: a concise guide to statistical inference*. SpringerBriefs in Statistics. Springer, Cham (2021). ISBN 978-3-030-81899-9; 978-3-030-81900-2. MR4385530.

Brown, L. D. *Fundamentals of statistical exponential families with applications in statistical decision theory*, volume 9 of *Institute of Mathematical Statistics Lecture Notes—Monograph Series*. Institute of Mathematical Statistics, Hayward, CA (1986). ISBN 0-940600-10-2. MR882001.

Conway, R. and Maxwell, W. A queuing model with state dependent service rates. *J. Ind. Eng.*, **12**, 132–136 (1962).

Daly, F. and Gaunt, R. E. The Conway-Maxwell-Poisson distribution: distributional theory and approximation. *ALEA Lat. Am. J. Probab. Math. Stat.*, **13** (2), 635–658 (2016). MR3522822.

Kotz, S., Balakrishnan, N., and Johnson, N. L. *Continuous Multivariate Distributions: Models and Applications*, volume 1 of *Wiley Series in Probability and Statistics*. Wiley, New York, 2nd edition (2000). DOI: 10.1002/0471722065.

Krivitsky, P. N. Exponential-family random graph models for valued networks. *Electron. J. Stat.*, **6**, 1100–1128 (2012). MR2988440.

Li, B., Zhang, H., and He, J. Some characterizations and properties of COM-Poisson random variables. *Comm. Statist. Theory Methods*, **49** (6), 1311–1329 (2020). MR4055520.

Sellers, K. F., Borle, S., and Shmueli, G. The COM-Poisson model for count data: a survey of methods and applications. *Appl. Stoch. Models Bus. Ind.*, **28** (2), 104–116 (2012). MR2911711.

Sellers, K. F. and Premeaux, B. Conway-Maxwell-Poisson regression models for dispersed count data. *Wiley Interdiscip. Rev. Comput. Stat.*, **13** (6), Paper No. e1533, 13 (2021). MR4336077.

Sellers, K. F., Swift, A. W., and Weems, K. S. A flexible distribution class for count data. *J. Stat. Distrib. App.*, **4** (22) (2017). DOI: 10.1186/s40488-017-0077-0.

Shmueli, G., Minka, T. P., Kadane, J. B., Borle, S., and Boatwright, P. A useful distribution for fitting discrete data: revival of the Conway-Maxwell-Poisson distribution. *J. Roy. Statist. Soc. Ser. C*, **54** (1), 127–142 (2005). MR2134602.