Research Article

A Characterization of the Compound Multiparameter Hermite Gamma Distribution via Gauss’s Principle

Werner Hürlimann

Wolters Kluwer Financial Services, Seefeldstrasse 69, 8008 Zürich, Switzerland

Correspondence should be addressed to Werner Hürlimann; whurlimann@bluewin.ch

Received 13 August 2013; Accepted 22 September 2013

Academic Editors: N. Marwan, J. Pacheco, and S. Umarov

Copyright © 2013 Werner Hürlimann. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We consider the class of those distributions that satisfy Gauss’s principle (the maximum likelihood estimator of the mean is the sample mean) and have a parameter orthogonal to the mean. It is shown that this so-called “mean orthogonal class” is closed under convolution. A previous characterization of the compound gamma characterization of random sums is revisited and clarified. A new characterization of the compound distribution with multiparameter Hermite count distribution and gamma severity distribution is obtained.

1. Introduction

The topic of maximum likelihood characterizations of distributions has a long history and is an active field of contemporary mathematical sciences. It concerns the characterization of a (class of) probability distribution(s) through the structure of the maximum likelihood estimator (MLE) of one or more parameters of interest (e.g., location, scale, etc.). Starting point is a famous result by Gauss [1] on the foundation of least squares theory (see, e.g., [2], [3, afterword, pages 208 and 215]). Given is a location family with continuous derivative. If the maximum likelihood estimator of the location parameter is the sample mean, then the distribution is normal. This important result has been discussed by Poincaré [4, Chapter 10], Teicher [5], Ferguson [6, 7], Marshall and Olkin [8], Bondesson [9], and Azzalini and Genton [10], by many other authors. The property that the MLE of the mean is the sample mean has been called Gauss’s principle by Campbell [11] (see also [12, 13] and references therein). A brief account of the present contribution follows.

Within the framework of multiparameter distributions, consider the “mean orthogonal class” of those distributions that besides Gauss’s principle have a parameterization such that the mean is orthogonal to some parameter vector, a property which can always be satisfied by Amari [14], Section 8. This class has been considered by Sprott [15]. A characterization of the mean orthogonal class through the cumulant generating function (cgf) has been formulated by Hürlimann [16]. Extending a result by Hudson [17], it is shown in Theorem 3 that this class is closed under convolution. Section 3 is devoted to a characterization of random sums through the mean orthogonal class. Hürlimann [18] has established that the mean scaled severity in a compound model is necessarily gamma distributed provided that the count distribution and the distribution of the random sum belong to the mean orthogonal class, and some additional partial differential parameter equation can be solved. A followup to this construction for the individual model of risk theory is Hürlimann [19]. We clarify and simplify the original proof to obtain a characterization that is further used in Section 4. Based on a result by Puig and Valero [20], we derive in Theorem 17 a most stringent characterization, which allows compounding of the gamma distribution under a single count data family, namely, the multiparameter Hermite distribution. This one requires that the count distribution is closed under convolution and binomial subsampling.

2. Distributions with the Mean Orthogonal Property

Let $X$ be a random variable whose distribution depends upon a vector $(\mu, \theta) = (\mu, \theta_1, \ldots, \theta_m)$ of $m + 1$ parameters, where the mean $\mu$ is functionally independent of $\theta$, that is,
\(\frac{\partial \mu}{\partial \theta_k} = 0, \ k = 1, \ldots, m.\) The log likelihood of \(X\) is denoted by \(\ell(x; \mu, \theta).\) We assume throughout that the cumulant generating function (cgf) \(C(t; \mu, \theta) = \ln[E(\exp(tX))]\) exists and denotes the variance by \(\sigma^2 = \sigma^2(\mu, \theta).\) The standard regularity conditions for maximum likelihood estimation are supposed to hold. The vector \(X = (X_1, \ldots, X_n)\) denotes a random sample of size \(n\), which realizes the random variable \(X\), and \(\bar{X}\) denotes the sample mean. We are interested in the class \(\mathcal{C}\) of distributions that satisfy Gauss’s principle (the maximum likelihood estimator of the mean is the sample mean), that is, such that \(\hat{\mu} = \bar{X}.\) A distribution belongs to this class if and only if there are functions \(g_k = g_k(\mu, \theta), k = 1, \ldots, m, h = h(\mu, \theta),\) such that the following equivalent partial differential equations hold (e.g., [15–17]):

\[
\frac{\partial \ell(x; \mu, \theta)}{\partial \mu} + \sum_{k=1}^{m} g_k \cdot \frac{\partial \ell(x; \mu, \theta)}{\partial \theta_k} + h \cdot (x - \mu) = 0,
\]

\[
\frac{\partial C}{\partial \mu} + \sum_{k=1}^{m} g_k \frac{\partial C}{\partial \theta_k} + h \cdot \left( \frac{\partial C}{\partial t} - \mu \right) = 0.
\]

**Definition 1.** The mean \(\mu\) is called orthogonal to the parameter vector \(\theta\), denoted by \(\mu \perp \theta\), if on has \(E[(\partial^2 \ell/\partial \mu \partial \theta)] = E[(\partial C/\partial \mu)(\partial \ell/\partial \theta)] = 0, \ k = 1, \ldots, m.\)

The original motivation for parameter orthogonality is improvement of maximum likelihood estimation by reparameterization. In the class \(\mathcal{C}\) the number of maximum likelihood equations is reduced by one and parameter orthogonality decreases the often high correlation between the MLEs of the parameters since the MLEs of orthogonal parameters are asymptotically uncorrelated. Indeed, the expectations in (2) are elements of the (expected) Fisher information matrix, which determines the asymptotic covariance matrix of \((\mu, \theta).\)

In this respect, one is interested in the subclass \(\mathcal{C}^\perp\) of \(\mathcal{C}\) of all distributions satisfying besides \(\hat{\mu} = \bar{X}\) the mean orthogonal property \(\mu \perp \theta.\) This so-called mean orthogonal class is characterized as follows.

**Theorem 2** (Characterization of the mean orthogonal class). Let \(X\) be a random variable with cgf \(C(t; \mu, \theta)\) satisfying the above assumptions. Then, one has \(X \in \mathcal{C}^\perp\) if and only if the following quasi-linear partial differential equation is satisfied:

\[
\sigma^2 \cdot \frac{\partial C}{\partial \mu} - \left( \frac{\partial C}{\partial t} - \mu \right) = 0.
\]

**Proof.** This is shown in Hürlimann [16].

Hudson [17, Theorem 1] has shown that the class \(\mathcal{C}\) is closed under convolution. In fact, convolution invariance holds under the more stringent mean orthogonal property.

**Theorem 3** (Convolution invariance of the mean orthogonal class). If \(X_1, X_2 \in \mathcal{C}^\perp\) are independent, then \(X = X_1 + X_2 \in \mathcal{C}^\perp.\) More precisely, if \(X_i \in \mathcal{C}^\perp\) has cgf \(C_i(t; \mu_i, \theta_i),\) with \(\mu_i \perp \theta_i, i = 1, 2,\) then the cgf \(C(t; \mu, \delta, \theta)\) of \(X = X_1 + X_2,\) with

\[
\mu = \mu_1 + \mu_2, \ \delta = |\mu_1\sigma_2^2 - \mu_2\sigma_1^2|, \ \text{and} \ \theta = (\theta^{(1)}, \theta^{(2)}),
\]

satisfies (2) and one has \(\hat{\mu} = \bar{X}, \ \mu \perp (\delta, \theta).\)

**Proof.** Without loss of generality we assume that \(\delta = \mu_1\sigma_2^2 - \mu_2\sigma_1^2 > 0.\) Since \(\mu = \mu_1 + \mu_2\) and \(\sigma^2 = \sigma^2_1 + \sigma^2_2,\) one can express \((\mu_1, \mu_2)\) as a function of \((\mu, \delta)\) through the parameter transformation \(\mu_1 = (\mu\sigma_2^2 + \delta)/\sigma^2_2,\mu_2 = (\mu\sigma_2^2 - \delta)/\sigma^2_2.\) Since \(X_1, X_2 \in \mathcal{C}^\perp\) and \(C(t; \mu, \delta, \theta) = C_1(t; \mu_1, \theta^{(1)}) + C_2(t; \mu_2, \theta^{(2)}),\) one obtains

\[
\sigma^2 \cdot \frac{\partial C}{\partial \mu} = \sigma^2_1 \cdot \frac{\partial C_1}{\partial \mu_1} + \sigma^2_2 \cdot \frac{\partial C_2}{\partial \mu_2},
\]

which implies the result by (2) of Theorem 2.

**Example 4.** Binomial random variables and their convolutions belong to the class \(\mathcal{C}^\perp\). For two binomials this is shown in Hürlimann [16, Example 2] (see also [21]). For an arbitrary number of binomials this is derived in the Appendix of Hudson [17].

### 3. Mean Orthogonal Characterization of the Compound Gamma Distribution

Consider random sums of the type

\[
X = \sum_{i=1}^{N} Y_i,
\]

where the \(Y_i\)'s are independent and identically distributed nonnegative random variables, and \(N\) is a counting random variable defined on the nonnegative integers, which is independent of the \(Y_i\)'s. The mean and variance of \(X, N,\) and \(Y \sim Y_i\) are denoted, respectively, by \(\mu, \sigma^2, \mu_N, \sigma^2_N,\) and \(\mu_t, \sigma^2_t.\) The coefficient of variation of \(X\) is denoted by \(\kappa = \sigma/\mu.\) In some applications, it is convenient to scale the severity \(Y\) by the mean \(\mu\) such that the mean scaled severity \(Z = Y/\mu \sim Z_i = Y_i/\mu\) has mean 1/\(\mu_N.\) The resulting sum

\[
X = \mu \cdot \sum_{i=1}^{N} Z_i
\]

is called mean scaled compound random sum. The mean scaled compound model has important insurance risk applications. It has been studied in Hürlimann [18], which establishes that the mean scaled severity is necessarily gamma.
distributed provided that the random variables $N$ and $X$ belong to the mean orthogonal class and some additional partial differential parameter equation can be solved. A followup to this construction for the individual model of risk theory is H"urlimann [19]. We clarify and simplify the original proof to obtain a characterization of (4), which will be used in Section 4. In particular, (3.20) in H"urlimann [18] is not a consequence but an assumption. Since this equation is satisfied in the provided examples, this error does not harm the obtained result but must be rectified from a mathematical logical point of view. Also, the proof of Lemma 7 there will be simplified (proof of Lemma 8 below).

**Theorem 5** (Compound gamma characterization). Let $N \in \mathbb{C}^+$ be a counting random variable with cgf $C_N(t; p, \theta_N)$. Suppose there exists a one-to-one coordinate transformation mapping $(p, \theta_N)$ to $(\mu_N, \theta_N)$ such that $\mu_N = \mu_N(p, \theta_N)$ and set $\lambda_N = -\ln \Pr(N = 0)$. Then $N \in \mathbb{C}^+$ with $\mu_N = \mu_N(p, \theta_N) \perp \theta_N$ is equivalent to the following conditions:

$$
\sigma_N^2 \cdot \frac{\partial C_N(t)}{\partial \mu_N} = \frac{\partial C_N(t)}{\partial t} - \mu_N, \quad (9)
$$

$$
\mu_N \cdot \frac{\partial G(s)}{\partial p} = \frac{\partial \lambda_N}{\partial p} \cdot s \cdot G'(s), \quad (10)
$$

$$
\mu_N \cdot \frac{\partial g(k)}{\partial p} = \frac{\partial \lambda_N}{\partial p} \cdot k g(k), \quad k = 1, 2, \ldots \quad (11)
$$

**Proof.** The condition (9) is a restatement of Theorem 2. Applying the chain rule of differential calculus, this condition transforms to

$$
\sigma_N^2 \cdot \frac{\partial C_N(t)}{\partial p} = \frac{\partial \mu_N}{\partial p} \cdot \left( \frac{\partial C_N(t)}{\partial t} - \mu_N \right). \quad (12)
$$

Now, by Lemma 8 below and the chain rule, one has

$$
\sigma_N^2 \cdot \frac{\partial \lambda_N}{\partial p} = \frac{\partial \mu_N}{\partial p} \cdot \mu_N. \quad (13)
$$

Inserting into (12) shows that

$$
\mu_N \cdot \frac{\partial C_N(t)}{\partial p} = \frac{\partial \lambda_N}{\partial p} \cdot \left( \frac{\partial C_N(t)}{\partial t} - \mu_N \right). \quad (14)
$$

The statements (10) and (11) follow by using the representation (7).

**Lemma 7.** Let $N$ be a counting random variable with cgf $C_N(t; p, \theta_N)$ of the form (7). Suppose there exists a one-to-one coordinate transformation mapping $(p, \theta_N)$ to $(\mu_N, \theta_N)$, and set $\lambda_N = -\ln \Pr(N = 0)$. Then $N \in \mathbb{C}^+$ with $\mu_N = \mu_N(p, \theta_N) \perp \theta_N$ is equivalent to the following conditions:

$$
\sigma_N^2 \cdot \frac{\partial C_N(t)}{\partial \mu_N} = \frac{\partial C_N(t)}{\partial t} - \mu_N, \quad (9)
$$

$$
\mu_N \cdot \frac{\partial G(s)}{\partial p} = \frac{\partial \lambda_N}{\partial p} \cdot s \cdot G'(s), \quad (10)
$$

$$
\mu_N \cdot \frac{\partial g(k)}{\partial p} = \frac{\partial \lambda_N}{\partial p} \cdot k g(k), \quad k = 1, 2, \ldots \quad (11)
$$

**Proof.** The representation (7) implies that $\lambda_N = G(1) = \sum_{k=1}^{\infty} g(k), C_N(0) = \sum_{k=1}^{\infty} kg(k) = \mu_N$. Now, using (7) one sees that (9) is equivalent to $\sigma_N^2 \cdot (\partial g(k)/\partial \mu_N) = kg(k), \ k = 1, 2, \ldots$. It follows that $\sigma_N^2 \cdot (\partial \lambda_N/\partial \mu_N) = \sum_{k=1}^{\infty} \sigma_N^2 \cdot (\partial g(k)/\partial \mu_N) = \sum_{k=1}^{\infty} kg(k) = \mu_N$.

**Proof of Theorem 5.** Let $M_Y(t) = \exp[C_Y(t)]$ be the moment generating function of $Y$. Expressed in terms of the mean scaled severity $Z = Y/\mu$ one has $M_Y(t) = M_Z(\mu t)$. The relationship (7) for the cgf $C_N(t)$ yields the series expansion:

$$
C(t) = C_N(\ln M_Z(\mu t)) = \sum_{k=1}^{\infty} g(k) \cdot M_Z^k(\mu t) - \lambda_N. \quad (15)
$$

By Theorem 2 one has $X \in \mathbb{C}^+$ if and only if the equation $\sigma_N^2 \cdot (\partial C/\partial \mu) - (\partial C/\partial t) = 0$ is satisfied. With the series representation for $C(t)$ and the assumption $\sigma_N^2 \cdot (\partial \lambda_N/\partial \mu) = \mu$, this equation is equivalent to the following condition (use that $M_Z(t)$ does not depend on $\mu$):

$$
\left( \mu - \sigma_N^2 t \right) \cdot M_Z'(\mu t) \cdot \sum_{k=1}^{\infty} kg(k) \cdot M_Z^{k-1}(\mu t)
$$

$$
= M_Z(\mu t) \cdot \sigma_N^2 \cdot \sum_{k=1}^{\infty} \frac{\partial g(k)}{\partial \mu} \cdot M_Z^{k-1}(\mu t). \quad (16)
$$
Now, by Lemma 7 and (10), one has the identity (use the differential chain rule)

$$
\frac{\partial g(k)}{\partial \mu} = \frac{\partial \lambda_N}{\partial \mu} \cdot kg(k),
$$

which, together with \(\sigma^2 \cdot (\partial \lambda_N/\partial \mu) = \mu\), implies that

$$
\sigma^2 \cdot \frac{\partial g(k)}{\partial \mu} = \sigma^2 \cdot \frac{\partial \lambda_N}{\partial \mu} \cdot kg(k) = \mu \cdot kg(k).
$$

Inserted into the above expression one obtains the ordinary differential equation:

$$
\mu_N \cdot (1 - \kappa^2 t) \cdot M'_Z(t) = M_Z(t),
$$

whose unique solution is \(M_Z(t) = (1 - \kappa^2 t)^{-\alpha}\). Since \(\kappa^2 \mu = \mu_N/\alpha\), one sees that

$$
C_Y(t) = \ln M_Z(\mu) = \alpha \cdot \ln \left( \frac{\alpha}{(\alpha - \mu_N t)} \right)
$$

is the cdf of a gamma-distributed random variable. The proof is complete.

The proof uses the so-called natural parameterization \((p, \theta_N, \mu_Y, \alpha)\) of the compound gamma distribution. It is interesting to obtain explicit parameters orthogonal to the means of \(N, Y\), and \(X\). By the assumption \(N \in C^\perp\) one has \(\mu_N = \mu_N(p, \theta_N) \perp \theta_N\), and since \(Y\) is gamma distributed, one has \(Y \in C^\perp\) with \(\mu_Y \perp \alpha\). It remains to construct a parameter vector orthogonal to the mean of \(X\) such that

$$
\mu = \mu_N(p, \theta_N) \cdot \mu_Y \perp (\theta_N, \alpha, \theta),
$$

where \(\theta = \theta(p, \theta_N, \mu_Y, \alpha)\) must be determined. This task can be solved in a unified way for a lot of counting distributions (see [18, Section 4]). To illustrate the method, it suffices to consider here a single example.

Example 9 (compound negative binomial gamma distribution). Let \(N \sim NB(\theta_N, p), \theta_N \geq 0, p \in (0, 1)\), be a negative binomial random variable. Its cumulant pgf (6) reads \(G(s) = -\theta_N \cdot \ln(1 - p \cdot s), \lambda_N = -\theta_N \cdot \ln(1 - p)\). One has the following identity (see [18, equation (4.7)]):

$$
p \frac{\partial G(s)}{\partial p} = s \cdot G'(s),
$$

which implies for \(s = 1\) that \(p(\partial \lambda_N/\partial p) = \mu_N\). Together, this shows that (10) is satisfied. Therefore, one has \(N \in C^\perp\) and \(\mu_N = \theta_N \cdot p/(1 - p) \perp \theta_N\). Now, by Theorem 5 the compound negative binomial will be a compound negative binomial gamma if \(X \in C^\perp\) and \(\sigma^2 \cdot (\partial \lambda_N/\partial \mu) = \mu\) is satisfied. Written in terms of the parameter \(\alpha = (\mu_N \cdot \kappa^2)^{-1}, \kappa = \sigma/\mu\), the latter equation is equivalent to the condition \((1/\mu_N) \cdot (\partial \lambda_N/\partial \mu) = \alpha/\mu\). With \(\lambda_N/\partial \mu = (\partial \lambda_N/\partial p) \cdot (\partial p/\partial \mu) = (\mu_Y/\alpha \cdot (\partial p/\partial \mu)\) one obtains the differential equation \((1/p) \cdot (\partial p/\partial \mu) = \alpha/\mu\), which has the solution \(p = \theta^\mu/\alpha\) for some \(\theta\). In the coordinates \((p, \theta_N, \mu_Y, \alpha)\), this constant is equal to

$$
\theta = \theta(p, \theta_N, \mu_Y, \alpha) = p \cdot \mu^{\alpha} = \left(\frac{\mu_Y \theta_N p}{1 - p}\right)^\alpha.
$$

Since \(X \in C^\perp\) one must have \(\mu = \mu_N(p, \theta_N) \cdot \mu_Y \perp (\theta_N, \alpha, \theta)\).

4. Mean Orthogonal Characterization of the Compound Multiparameter Hermite Gamma

The mean orthogonal characterization of the compound gamma distribution allows for a wide variety of count data distributions in the mean orthogonal class. In order to reduce further the possible set of count distributions that can be used, one can ask for characterizations in terms of additional assumptions. For example, Puig [25] and Puig and Valero [26] characterize count data distributions satisfying Gauss’s principle and several notions of additivity, which via Theorem 5 can be translated to characterizations of compound gamma distributions. Based on a result by Puig and Valero [20], we derive a most stringent characterization, which allows compounding of the gamma distribution under a single count data family, namely, the multiparameter Hermite distribution. To show this, some additional preliminaries are required.

Definition 10. Let \(N\) be a counting random variable, let \(B_1(p), B_2(p), \ldots\), be independent and identically distributed Bernoulli random variables with probability of success \(p \in (0, 1)\), and let \(N\) be independent of \(B_i(p), i = 1, 2, \ldots\) Then \(N(p) = \sum_{i=1}^{\infty} B_i(p) (X(p) = 0 \text{ if } N = 0)\) is called an independent \(p\)-thinning of \(N\).

Definition 11. Let \(F\) be a family of count distributions. It is called closed under binomial subsampling if, for any random variable \(N\) with distribution in \(F\), all its independent \(p\)-thinnings, for all \(p \in (0, 1)\), have distributions in \(F\).

Definition 12. Let \(F\) be a family of distributions. It is called closed under convolution if, for any two independent random variables \(X, Y\) with distributions in \(F\), the distribution of the sum \(X + Y\) also belongs to \(F\).

Definition 13. Let \(N\) be an integer random variable with pgf \(P(s)\) and factorial cumulant generating function (fcgf) \(C^{(f)}(s) = \ln P(s + 1)\). For any integer \(n \geq 1\) the \(n\)th factorial cumulant of \(N\) is defined and denoted by \(\kappa_{(n)} = d^n C^{(f)}(s)/ds^n|_{s=0}\).

There is only one count distribution family closed under convolution and binomial subsampling.

Theorem 14 (Characterization of the multiparameter Hermite distribution). Let \(F\) be a family of count distributions parameterized by its \(r\) first factorial cumulants \(\theta = (\kappa_{(1)}, \ldots, \kappa_{(r)})\) and assume that its pgf \(P(s)\) is continuous in \(\theta\)
over its parameter space. Then \( F \) is closed under convolution and binomial subsampling if and only if the pgf is of the form

\[
P(s) = \exp \left\{ \sum_{k=1}^{r} \frac{\kappa(k)}{k!} (s-1)^k \right\}.
\]

(24)

Proof. See Puig and Valero [20], proof of Theorem 1. \( \square \)

Some comments are in order. The case \( r = 1 \) corresponds to the Poisson distribution, \( r = 2 \) is the Hermite distribution (e.g., [27]). For arbitrary \( r \), this distribution is called the \textit{multiparameter Hermite} distribution of order \( r \) by Milne and Westcott [28]. In terms of the cumulant pgf (6), the representation (24) can be rewritten as

\[
P(s) = \exp \left\{ \sum_{k=1}^{r} g(k) \cdot (s^k - 1) \right\}.
\]

(25)

where \( g(k), k = 1, 2, \ldots, r \), solves the system in (8), that is,

\[
n \cdot p(n) = \sum_{k=1}^{\min(n,k)} k \cdot g(k) \cdot p(n-k), \quad n = 1, 2, \ldots,
\]

(26)

The case \( r = 2 \) (26) is already in A.W. Kemp and C.D. Kemp [29], and for arbitrary \( r \) this assertion is equivalent to Lemma 2 in Puig and Valero [20]. The special case \( g(1) > 0 \), \( g(r) \geq 0 \), \( g(k) = 0, k = 2, \ldots, r-1 \) is the generalized Hermite by Gupta and Jain [30]. The multiparameter Hermite belongs also to the Kumar [31] family of distributions. In general, the conditions on the sequence \( g(k), k = 1, 2, \ldots, r \), under which (25) defines a true probability distribution are identified in Lévy [32]. According to Lukacs [33, page 252] and Johnson et al. [34, page 356], this is the case provided that a negative value \( g(k) < 0 \) is preceded by a positive value and followed by at least two positive values. In particular, if at least \( g(1), g(r-1), g(r) \) are nonzero, then \( g(1) > 0, g(r-1) > 0, g(r) > 0 \), are necessary conditions for (25) to be a pgf [28, Remark 1]. If \( g(k) \geq 0 \) for \( k = 1, 2, \ldots, r \), then the multiparameter Hermite is compound Poisson with parameter \( \Lambda_N = \sum_{k=1}^{r} g(k) \) and severity \( h(k) = g(k)/\lambda_N \), thus infinitely divisible by Feller [35, Section XII.2]. Due to the next result, the multiparameter Hermite is of interest in the context of Gauss’s principle, orthogonal parameters to the mean, and the related compound gamma characterization of random sums.

**Lemma 15.** Let \( \xi_i(\theta), k = 1, 2, \ldots, r \), be continuous real functions in the parameter vector \( \theta \) over some parameter space, and set \( g(k) = \xi_i(\theta) \cdot p^k \), \( k = 1, 2, \ldots, r \), for a parameter \( p > 0 \). Assume that the cumulant pgf \( G(s) = \sum_{k=1}^{r} \xi_i(\theta) \cdot (ps)^k \) defines a feasible multiparameter Hermite random variable \( N \) of order \( r \) over the parameter space. Then \( \Lambda_N \in \mathbb{C}^+ \) and \( \mu_N = \mu_N(p, \theta) = \sum_{k=1}^{r} k \cdot \xi_i(\theta) \cdot p^k \perp \theta \).

Proof. Set \( \lambda_N = G(1) = \sum_{k=1}^{r} \xi_i(\theta) \cdot p^k \). Then one has

\[
p \cdot \frac{\partial G(s)}{\partial p} = s \cdot G'(s), \quad p \cdot \frac{\partial \lambda_N}{\partial p} = \mu_N.
\]

(27)

Together, this shows that (10) is satisfied. The result follows by Lemma 7. \( \square \)

**Example 16** (Hermite distribution \( r = 2 \)). Suppose the Hermite distribution is parameterized by its first two factorial cumulants \( \kappa(1), \kappa(2) \). Since \( \kappa(1) = \mu_N, \kappa(2) = \sigma_N^2 - \mu_N \), it can equivalently be parameterized by its mean \( \mu_N \) and variance \( \sigma_N^2 \). Consider a parameterization \( p > 0, \theta_N > 0 \) such that \( g(k) = \theta_N \cdot p^k, k = 1, 2 \). There exists a one-to-one mapping between \( (p, \theta_N) \) and \( (\mu_N, \sigma_N^2) \). Since \( \mu_N = g(1) + 2g(2), \sigma_N^2 = g(1) + 4g(2) \), it is determined by the coordinate transformation:

\[
p = \frac{1}{2} \cdot \frac{\sigma_N^2 - \mu_N}{\sigma_N^2 + 2\mu_N}, \quad \theta_N = 2 \cdot \left( \frac{\sigma_N^2 + 2\mu_N}{\sigma_N^2 - \mu_N} \right)^2.
\]

(28)

Therefore, the cumulant pgf \( G(s) = \theta \cdot (ps + p^2s^2) \) defines a feasible two-parameter Hermite distribution such that the corresponding random variable belongs to \( \mathbb{C}^+ \) and \( \mu_N = \mu_N(p, \theta_N) = \theta_N \cdot (p + 1) - \theta_N \cdot p \cdot (1 + 2p) \perp \theta_N \). Since \( \sigma_N^2 > \mu_N \), one notes that the Hermite distribution is necessarily overdispersed. As noted by Puig and Valero [20], overdispersion holds for all infinitely divisible multiparameter Hermite distributions of arbitrary order \( r \geq 2 \). Therefore, it should be useful to analyze data with this property (e.g., claim number data in automobile insurance, up-to-date H¨urlimann [36, page 802], and multiparameter Hermite for \( r \geq 3 \)).

We are ready for the following new characterization result.

**Theorem 17** (Compound multiparameter Hermite gamma characterization). Let \( N \) be a counting random variable parameterized by its \( r \) first factorial cumulants \( \kappa(1), \ldots, \kappa(r) \) and assume that its cgf \( C_N(t, \xi) \) is continuous in \( \xi \) over its parameter space and set \( \mu_N = \kappa(1), \lambda_N = -\ln Pr(N = 0) \). Suppose the cgf \( C_N(t) \) of the severity \( Y \) exists, and let \( C(t) = C_N(C_N(t)) \) be the cgf of the random sum \( X = \sum_{i=1}^{N} Y_i \). Assume the cgf of the mean scaled severity \( C_N(t) = C_N(t/\mu_N) \) is functionally independent of \( \mu_N \) and set \( \alpha = (\mu_N\kappa^2)^{-1}, \kappa = \sigma/\mu \). Assume \( N \) is closed under convolution and binomial subsampling, \( X \in \mathbb{C}^+ \) and \( \alpha^2 \cdot (\partial \lambda_N/\partial \mu) = \mu \). Then \( N \) is a multiparameter Hermite distribution of order \( r \), and \( Y \) is gamma distributed with cgf \( C_N(t) = \alpha \ln(\alpha/(\alpha - \mu_N t)) \). Furthermore, there exists a parameterization \( (p, \theta_N) \) of \( N \) such that its cumulant pgf reads \( G(s) = \sum_{k=1}^{r} \xi_i(\theta_N) \cdot (ps)^k \). One has \( N,Y,X \in \mathbb{C}^+ \) with \( \mu_N = \mu_N(p, \theta_N) = \sum_{k=1}^{r} k \cdot \xi_i(\theta_N) \cdot p^k \perp \theta_N \).
\[ \mu_y \perp \alpha, \mu = \mu_N(p, \theta_N); \mu_y \perp (\theta_N, \alpha, \theta), \text{ and, in the coordinates} \ (p, \theta_N, \mu_y, \alpha), \text{ the constant } \theta \text{ is equal to} \]

\[ \theta = \theta (p, \theta_N, \mu_y, \alpha) = p \cdot \mu^{-\alpha} \]

\[ = p \cdot \left( \mu_y \cdot \sum_{k=1}^{r} k \cdot c_k (\theta_N) \cdot p^k \right)^{-\alpha} \quad (29) \]

**Proof.** The result follows by combining Theorems 5 and 14 making the observation that a multiparameter Hermite distribution can always be put in the form of Lemma 15 (generalization of Example 16). The assertion about the orthogonal parameters to the means \( \mu_N, \mu_y, \mu \) follows along the same arguments as in Example 9 using (27).

**References**

[1] C. F. Gauss, *Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientum*, Perthes und Besser, Hamburg, Germany, 1809, English Translation by C. H. Davis, *Theory of the Motions of the Heavenly Bodies Moving about the Sun in Conic Sections*, Dover, New York, NY, USA, 1857, Reprint 1963.

[2] S. M. Stigler, “Gauss and the invention of least squares,” *The Annals of Statistics*, vol. 9, no. 3, pp. 465–474, 1981.

[3] G. W. Stewart, *Theory of the Combination of Observations Least Subject to Errors: Part One, Part Two, Supplement*, vol. 11 of Classics in Applied Mathematics, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1995, English Translation of C. F. Gauss, *Theoria Combinationis Observationum Erroribus Minimis Obnoxiae*, 1821–1826.

[4] H. Poincaré, *Calcul des Probabilities*, Gauthier-Villars, Paris, France, 2nd edition, 1912.

[5] H. Teicher, “Maximum likelihood characterizations of distributions,” *The Annals of Mathematical Statistics*, vol. 32, no. 4, pp. 1214–1222, 1961.

[6] T. S. Ferguson, “Location and scale parameters in exponential families of distributions,” *The Annals of Mathematical Statistics*, vol. 33, no. 3, pp. 986–1001, 1962.

[7] T. S. Ferguson, “Correction to ‘location and scale parameters in exponential families of distributions’,” *The Annals of Mathematical Statistics*, vol. 34, no. 4, p. 1603, 1963.

[8] A. W. Marshall and I. Olkin, “Maximum likelihood characterizations of distributions,” *Statistica Sinica*, vol. 3, pp. 157–171, 1993.

[9] L. Bondesson, “A generalization of Poincaré’s characterization of exponential families,” *Journal of Statistical Planning and Inference*, vol. 63, no. 2, pp. 147–155, 1997.

[10] A. Azzalini and M. G. Gentoon, “On Gauss’s characterization of the normal distribution,” *Bernoulli*, vol. 13, no. 1, pp. 169–174, 2007.

[11] L. L. Campbell, “Equivalence of Gauss’s principle and minimum discrimination information estimation of probabilities,” *The Annals of Mathematical Statistics*, vol. 41, no. 3, pp. 1011–1015, 1970.

[12] P. Puig, “A note on the harmonic law: a two-parameter family of distributions for ratios,” *Statistics and Probability Letters*, vol. 78, no. 3, pp. 320–326, 2008.

[13] M. Duerinckx, C. Ley, and Y. Swan, “Maximum likelihood characterization of distributions,” preprint, http://arxiv-web3.library.cornell.edu/abs/1207.2805.

[14] S. I. Amari, *Differential-Geometrical Methods in Statistics*, vol. 28 of Lecture Notes in Statistics, Springer, Heidelberg, Germany, 1985.

[15] D. A. Sprott, “Estimating the parameters of a convolution by maximum likelihood estimation,” *Journal of the American Statistical Association*, vol. 78, no. 382, pp. 457–460, 1983.

[16] W. H"urlimann, “On parameter orthogonality to the mean,” *Statistical Papers*, vol. 33, no. 1, pp. 69–74, 1992.

[17] S. Hudson, “New results for a class of univariate distributions,” *Communications in Statistics*, vol. 31, no. 2, pp. 239–247, 2002.

[18] W. H"urlimann, “On mean scaled individual risk models,” *Blätter der DGVFM*, vol. 23, no. 3, pp. 251–265, 1998.

[19] W. H"urlimann, “Measuring operational risk using a mean scaled individual risk model,” *Applied Mathematics and Computation*, vol. 152, no. 2, pp. 425–447, 2004.

[20] P. Puig and J. Valero, “Characterization of count data distributions involving additivity and binomial subsampling,” *Bernoulli*, vol. 13, no. 2, pp. 544–555, 2007.

[21] G. E. Willmot and D. A. Sprott, “A note on parameters orthogonal to the mean,” *Biometrika*, vol. 81, no. 2, pp. 409–412, 1994.

[22] W. H"urlimann, “On maximum likelihood estimation for count data models,” *Insurance Mathematics and Economics*, vol. 9, pp. 39–49, 1990.

[23] W. H"urlimann, “Correction note: on maximum likelihood estimation for count data models,” *Insurance Mathematics and Economics*, vol. 10, no. 1, pp. 81, 1991.

[24] W. H"urlimann, “Pseudo compound Poisson distributions in risk theory,” *ASTIN Bulletin*, vol. 20, pp. 57–79, 1990.

[25] P. Puig, “Characterizing additively closed discrete models by a property of their maximum likelihood estimators, with an application to generalized Hermite distributions,” *Journal of the American Statistical Association*, vol. 98, no. 463, pp. 687–692, 2003.

[26] P. Puig and J. Valero, “Count data distributions: some characterizations with applications,” *Journal of the American Statistical Association*, vol. 101, no. 473, pp. 332–340, 2006.

[27] C. D. Kemp and A. W. Kemp, “Some properties of the Hermite distribution,” *Biometrika*, vol. 52, no. 3, pp. 381–394, 1965.

[28] R. K. Milne and M. Westcott, “Generalized multivariate Hermite distributions and related point processes,” *Annals of the Institute of Statistical Mathematics*, vol. 45, no. 2, pp. 367–381, 1993.

[29] A. W. Kemp and C. D. Kemp, “An alternative derivation of the Hermite distribution,” *Biometrika*, vol. 53, pp. 627–628, 1966.

[30] R. P. Gupta and G. C. Jain, “A generalized Hermite distribution and its properties,” *SIAM Journal on Applied Mathematics*, vol. 27, no. 2, pp. 359–363, 1974.

[31] C. S. Kumar, “Extended generalized hypergeometric probability distributions,” *Statistics and Probability Letters*, vol. 59, no. 2, pp. 1–7, 2002.

[32] P. Lévy, “Sur les exponentielles de polynômes et sur l’arithmétique des produits de lois de Poisson,” *Annales Scientifiques de l’Ecole Normale Supérieure*, vol. 54, pp. 231–292, 1937.

[33] E. Lukacs, *Characteristic Functions*, Griffin, London, UK, 2nd edition, 1970.

[34] L. N. Johnson, S. Kotz, and A. W. Kemp, *Univariate Discrete Distributions*, John Wiley & Sons, New York, NY, USA, 2nd edition, 1992.
[35] W. Feller, *An Introduction to Probability Theory and Its Applications*, vol. 1, John Wiley & Sons, 3rd edition, 1968.

[36] W. Hërlimann, “Robust confidence bounds for the mean of some count data models,” *Blätter der DGVFM*, vol. 25, no. 4, pp. 795–811, 2002.