Stein’s Lemma for elliptical random vectors
Zinoviy Landsman\textsuperscript{a}, Johanna Nešlehová\textsuperscript{b,*}
\textsuperscript{a}Department of Statistics, University of Haifa, Mount Carmel, 31905 Haifa, Israel
\textsuperscript{b}Department of Mathematics, ETH Zurich, CH-8092 Zurich, Switzerland
Received 31 July 2006
Available online 31 May 2007

Abstract
For the family of multivariate normal distribution functions, Stein’s Lemma presents a useful tool for calculating covariances between functions of the component random variables. Motivated by applications to corporate finance, we prove a generalization of Stein’s Lemma to the family of elliptical distributions.
© 2007 Elsevier Inc. All rights reserved.

AMS 2000 subject classification: 62E10; 62H05

Keywords: Density generator; Elliptical distribution; Generalized hyperbolic distribution; Normal distribution; Normal mixtures; Pearson Type VII distribution; Stein’s Lemma; Spherical distribution; Student \( t \) distribution

1. Introduction
One of the important tools in the capital asset pricing model (CAPM) is Stein’s Lemma which in the bivariate normal case allows to calculate the covariance between some function of the first component and the second component, and this in terms of the covariance between the two components. More precisely, Stein’s Lemma states that (see [18,3]):

\textbf{Theorem 1.} Suppose that the random vector \((X_1, X_2)’\) has a bivariate normal distribution and \(h\) is a differentiable function fulfilling \(E\left(|h’(X_1)|\right) < \infty\), then

\[
\text{Cov}(h(X_1), X_2) = \text{Cov}(X_1, X_2) E\left(h’(X_1)\right). 
\]

The link of Stein’s Lemma to the CAPM can for instance be found in Cochrane [3, p. 164] and Panjer et al. [17, Section 4.5].

\* Corresponding author.
E-mail address: johanna@math.ethz.ch (J. Nešlehová).
In the context of a corporate finance model for insurance, Froot [8] uses a three-dimensional version of Stein’s Lemma, still based on the assumption of multivariate normality. It is a special case of the extension of Stein’s Lemma to multivariate normal random vectors introduced in Stein [18] and states that if the random vector \( X := (X_1, X_2, X_3)' \) follows a multivariate normal distribution and \( h: \mathbb{R}^2 \to \mathbb{R} \) is a function satisfying certain regularity conditions, which are further specified in Section 3, then

\[
\text{Cov}(h(X_1, X_2), X_3) = \text{Cov}(X_1, X_3) \mathbb{E}(\nabla_1 h(X_1, X_2)) + \text{Cov}(X_2, X_3) \mathbb{E}(\nabla_2 h(X_1, X_2)).
\]

(2)

The function \( \nabla_i h \) is essentially the first partial derivative of \( h \) with respect to the \( i \)th coordinate; see Definition 5.

In the context discussed in Froot [8], the assumption of normality restricts the range of possible applications. Especially in insurance, the liability components are typically heavier-tailed than the normal (see for instance [4,15]). The CAPM model can be derived under various assumptions, one of them being the assumption of multivariate normality. We want to stress that the CAPM can also be arrived at without such a distributional assumption if “one is willing to swallow quadratic utility instead” as stated in [3, p. 152]. Our generalization of Stein’s Lemma below is, however, motivated by the distributional approach.

Asset pricing models, such as the CAPM, have been studied for more general asset return distributions. For instance, the class of elliptical distributions has been shown to be consistent with a form of the CAPM; see Ingersoll [10], Owen and Rabinovitch [16] and Hamada and Valdez [9]. In Landsman [13], a generalization of (1) for bivariate elliptical distributions is discussed. In the present paper, we derive a result which extends both the classical Stein’s Lemma and Landsman’s generalization to the class of multivariate elliptical distributions.

The paper is organized as follows. Section 2 provides necessary background in that it reviews several results on elliptical distributions and introduces the notions of cumulative density generator and associate measure. The generalization of Stein’s Lemma is given in Section 3. Section 4 illustrates Stein’s Lemma for selected families of elliptical distributions. Section 5 concludes.

2. Cumulative density generator and associate measure

We begin by recalling several results on elliptical distributions which will be needed later on. For detailed information on this extensively studied class of distributions, we refer in particular to Fang et al. [6], Fang and Zhang [7], Cambanis et al. [2], Kelker [12], Embrechts et al. [5], Bingham and Kiesel [1] and McNeil et al. [15, Chapter 3], which also includes a discussion and further references on relevant statistical inference.

Throughout, \( x \in \mathbb{R}^n \) will denote an \( n \)-dimensional vector and \( x' = (x_1, \ldots, x_n) \) its transpose. For an \( n \times n \) matrix \( \Sigma \in \mathbb{R}^{n \times n} \), \( |\Sigma| \) is the determinant of \( \Sigma \). Note that if \( \Sigma \) is positive definite, the Cholesky decomposition is unique. In this case, \( \sqrt{\Sigma} \) stands for the corresponding Cholesky factor, i.e. \( \sqrt{\Sigma}' \sqrt{\Sigma} = \Sigma \). Furthermore, we will use the following result, see e.g. Fang et al. [6, Lemma 1.4]:

**Lemma 1.** Every non-negative measurable function \( f: \mathbb{R} \to \mathbb{R} \) satisfies

\[
\int_{\mathbb{R}^n} f(x'x/2) \, dx = \frac{(2\pi)^{n/2}}{\Gamma(n/2)} \int_0^\infty u^{n/2-1} f(u) \, du.
\]

(3)
As is well known, elliptical distributions emerge as multivariate affine transformations of spherical distributions and can be distinguished by a specific form of its characteristic function. In the present paper, however, we will only consider absolutely continuous elliptical distributions with positive definite dispersion matrices.

**Definition 1.** An absolutely continuous $n$-dimensional random vector $X$ follows an elliptical distribution if its density is of the form

$$f_X(x) = |\Sigma|^{-1/2} g \left( \frac{(x - \mu)' \Sigma^{-1} (x - \mu)}{2} \right)$$

for all $x \in \mathbb{R}^n$ where $\mu$ is an $n \times 1$ vector, $\Sigma = (\sigma_{ij})$ a positive definite $n \times n$ dispersion matrix and $g : [0, \infty) \to [0, \infty)$ a measurable function referred to as the density generator of $X$. We write $X \sim \mathcal{E}_n(\mu, \Sigma, g)$. In particular, $\mathcal{E}_n(0, I_n, g)$ is called a spherical distribution, $S_n(g)$.

Note that for $X \sim \mathcal{E}_n(\mu, \Sigma, g)$, $\Sigma$ is not necessarily the covariance matrix of $X$. The latter may not even exist.

Not every measurable non-negative function $g$ can act as a density generator of an elliptical distribution unless additional integrability properties are ensured. As discussed e.g. by Kelker [12] or Fang and Zhang [7], a necessary and sufficient condition can be easily derived from (3): $g$ is a density generator (up to a normalization) of an $n$-dimensional elliptical distribution if and only if

$$\int_0^\infty u^{n/2-1} g(u) \, du < \infty. \tag{5}$$

The next proposition recalls the well-known stochastic representation of elliptical distributions.

**Proposition 1.** Let $X$ be an $n$-dimensional absolutely continuous random vector. Then $X \sim \mathcal{E}_n(\mu, \Sigma, g)$ if and only if

$$X \overset{d}{=} \mu + R \sqrt{\Sigma} U_n,$$

where $U_n$ is a random vector uniformly distributed on the unit sphere in $\mathbb{R}^n$ and $R$ is a non-negative random variable independent of $U_n$ with density

$$f_R(r) = \frac{2\pi^{n/2}}{\Gamma(n/2)} r^{n-1} g(r^2/2), \quad r \in [0, \infty). \tag{6}$$

Furthermore, $RU_n \sim S_n(g)$. The random variable $R$ is called the radial part of $X$.

**Proof.** See Cambanis et al. [2, Section 4]. \quad \square

It may be noted that neither the dispersion matrix nor the density generator of an elliptical distribution is unique. There, however, always exists a positive constant $c$ such that, if $X \sim \mathcal{E}_n(\mu, \Sigma, \phi)$ and at the same time $X \sim \mathcal{E}_n(\tilde{\mu}, \tilde{\Sigma}, \tilde{g})$, $\tilde{\mu} = \mu$, $\tilde{\Sigma} = c \Sigma$ and $\tilde{g}(x) = c^{n/2} g(cx)$ for all $x \in [0, \infty)$; see Cambanis et al. [2, Theorem 3].

A further useful fact is that if $X \sim \mathcal{E}_n(\mu, \Sigma, g)$, all lower dimensional margins have densities whose generators can be explicitly stated in terms of $g$. 

**Proposition 2.** Suppose \( X \sim \mathcal{E}_n(\mu, \Sigma, g) \) and \( m \) an integer with \( 1 \leq m < n \). Then all \( m \)-dimensional margins are elliptical with the same generator \( g_m \) given by
\[
g_m(u) = \frac{(2\pi)^{(n-m)/2}}{\Gamma((n-m)/2)} \int_{u}^{\infty} (y - u)^{(n-m)/2 - 1} g(y) dy \tag{7}
\]
for almost all \( u \in [0, \infty) \).

**Proof.** The assertion follows directly from (2.44) of Fang et al. [6] and (6). \( \square \)

Finally, note that the existence of moments of an elliptical random vector is related to the existence of moments of its radial part.

**Proposition 3.** Let \( X \sim \mathcal{E}_n(\mu, \Sigma, g) \) with radial part \( R \). Then
\begin{enumerate}
\item \( \mathbb{E}X < \infty \) if and only if \( \mathbb{E}R < \infty \). In that case, \( \mathbb{E}X = \mu \).
\item \( \text{Cov}(X) \) exists if and only if \( \mathbb{E}R^2 < \infty \). In that case,
\[
\text{Cov}(X) = \frac{\mathbb{E}(R^2)}{n} \Sigma.
\]
\end{enumerate}

**Proof.** See Theorem 2.6.4 of Fang and Zhang [7]. \( \square \)

Generalizing Landsman and Valdez [14], we now introduce cumulative generators and associative measures which will be of key importance in the next section.

**Definition 2.** Let \( X \sim \mathcal{E}_n(\mu, \Sigma, g) \) and \( g_m, 1 \leq m < n \), denote the density generators of the \( m \)-dimensional margins of \( X \) as specified by Proposition 2. The \( m \)-dimensional cumulative generator \( \overline{G}_m \) of \( X \) is a function on \( [0, \infty) \) given by
\[
\overline{G}_m(x) = \int_{x}^{\infty} g_m(u) du, \quad 1 \leq m < n.
\]

First note that, with (7),
\[
\overline{G}_m(x) = \int_{x}^{\infty} \frac{(2\pi)^{(n-m)/2}}{\Gamma((n-m)/2)} \int_{u}^{\infty} (y - u)^{(n-m)/2 - 1} g(y) dy du
\begin{align*}
&= \frac{(2\pi)^{(n-m)/2}}{\Gamma((n-m)/2)} \int_{x}^{\infty} \int_{x}^{y} (y - u)^{(n-m)/2 - 1} g(y) du dy \\
&= \frac{(2\pi)^{(n-m)/2}}{\Gamma((n-m)/2 + 1)} \int_{x}^{\infty} (y - x)^{(n-m)/2} g(y) dy.
\end{align*}
\]

In particular, therefore,
\[
\overline{G}_m(x) = \frac{(2\pi)^{(n-m+2)/2}}{2\pi \Gamma((n - m + 2)/2)} \int_{x}^{\infty} (y - x)^{(n-m+2)/2 - 1} g(y) dy = \frac{1}{2\pi} g_{m-2}(x), \tag{8}
\]
when \( m \geq 3 \). Hence \( \overline{G}_m(x) \) is finite for almost all \( x \in [0, \infty) \) whenever \( 3 \leq m < n \). For \( m = 1, 2 \) we have that
\[
\overline{G}_m(x) = \frac{(2\pi)^{(n-m)/2}}{\Gamma((n-m)/2 + 1)} \int_{x}^{\infty} y^{(n-m)/2} (1 - x/y)^{(n-m)/2} g(y) dy
\]
\[ \frac{(2\pi)^{(n-m)/2}}{\Gamma((n-m)/2 + 1)} \int_0^\infty y^{(n-m)/2} g(y) \, dy \]

(6)

\[ \frac{2\pi^{m/2}}{\Gamma((n-m)/2 + 1)} \int_0^\infty u^{2-m} f_R(u) \, du. \]

This immediately implies that \( \overline{G}_2(x) \) is finite for every \( x \in [0, \infty) \) as \( \int_0^\infty f_R(u) \, du = 1 \). In addition, \( \overline{G}_1(x) \) finite on \([0, \infty)\) if \( ER < \infty \). Recall from Proposition 3(i) that the latter condition is equivalent to \( \text{EX} < \infty \).

Put together, if \( X \) has a finite mean, \( \overline{G}_m(x) \), \( 1 \leq m < n \), is finite for almost all \( x \in [0, \infty) \). In particular, \( \overline{G}_m \) can then be used to define a measure \( v_m \) as follows.

**Definition 3.** Let \( X \sim \mathcal{E}_n(\mu, \Sigma, g) \) and \( g_m, 1 \leq m < n \), denote the density generators of the \( m \)-dimensional margins of \( X \) as specified by Proposition 2. Furthermore, assume that \( \text{EX} < \infty \). The measure \( v_m \) on the Borel \( \sigma \)-algebra \( \mathcal{B}(\mathbb{R}^m) \) given by

\[ v_m(B) = \int_B \overline{G}_m(x'x/2) \, dx \]

will be referred to as the \( m \)-dimensional associate measure of \( X \).

Note that the associate measure \( v_m \) is \( \sigma \)-finite. Moreover, when the covariance of \( X \) exists, \( v_m \) is finite and can be normalized to be a probability measure; the normalization is given by the following lemma.

**Lemma 2.** Let \( X \sim \mathcal{E}_n(\mu, \Sigma, g) \) be an elliptical random vector with radial part \( R \). If the covariance of \( X \) exists, then

\[ v_m(\mathbb{R}^m) = \frac{ER^2}{n}, \quad 1 \leq m < n. \]

**Proof.** Recall first that \( ER^2 < \infty \) by Proposition 3. Since \( \overline{G}_m \) is non-negative and measurable, we obtain the following identities by (3) and Fubini’s Theorem:

\[
\int_{\mathbb{R}^m} \overline{G}_m \left( \frac{x'x}{2} \right) \, dx \\
= \frac{(2\pi)^{m/2}}{\Gamma(m/2)} \int_0^\infty u^{m/2-1} \overline{G}_m(u) \, du \\
= \frac{(2\pi)^{m/2}}{\Gamma(m/2)} \int_0^\infty \int_u^\infty t^{m/2-1} g_m(t) \, dt \, du \\
= \frac{(2\pi)^{m/2}}{\Gamma(m/2 + 1)} \int_0^\infty t^{m/2} g_m(t) \, dt \\
(7) = \frac{(2\pi)^{n/2}}{\Gamma(m/2 + 1)\Gamma((n-m)/2)} \int_0^\infty \int_t^\infty t^{m/2} (u-t)^{(n-m)/2-1} g(u) \, du \, dt
\]
With (6), the last expression equals \( \frac{1}{n} \int_0^\infty v^2 f_R(v) \, dv \), where \( f_R \) is the density of the radial part corresponding to \( X \). □

In other words, if the covariance of \( X \) exists, the measures \( \frac{n}{E R^2} v_m, 1 \leq m < n \) are spherical probability distributions with densities \( \frac{n}{E R^2} \overline{G}_m \left( \frac{x^T x}{2} \right) \).

3. Main result

Throughout this section, we will only consider elliptical distributions with finite mean; the latter condition is needed for the definition of the associate measure \( v_m \) as discussed in the preceding paragraphs.

Furthermore, we will fix the following notation. Consider a partition of a random vector \( X \sim \mathcal{E}_n(\mu, \Sigma, g) \) according to \( X' = (X'_1, X'_2) \) where \( X'_1 = (X_1, \ldots, X_m) \) and \( X'_2 = (X_{m+1}, \ldots, X_n) \). Correspondingly, we write \( \mu' = (\mu'_1, \mu'_2) \) with \( \mu_i = EX_i, i = 1, 2 \) and partition the dispersion matrix \( \Sigma \) and its inverse \( Q = \Sigma^{-1} \) as

\[
\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \quad \text{and} \quad Q = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix},
\]

where \( \Sigma_{11} \in \mathbb{R}^{m \times m} \) and \( \Sigma_{22} \in \mathbb{R}^{(n-m) \times (n-m)} \). Note that, in particular, \( X_1 \sim \mathcal{E}_m(\mu_1, \Sigma_{11}, g_m) \).

With \( Q_{11.2} = Q_{11} - Q_{12}Q_{22}^{-1}Q_{21} \) it is well known that \( \Sigma_{11}^{-1} = Q_{11.2} \) as well as that \( Q_{11.2} \) is positive definite as soon as \( Q \) is. There hence exists a unique upper triangular and positive definite matrix \( C_{11} \) such that \( C_{11}^{-1} = \Sigma_{11} \). Note that \( C_{11}^{-1} \) is then the Cholesky factor of \( Q_{11.2} \), i.e. \( Q_{11.2} = (C_{11}^{-1})'C_{11}^{-1} \). Furthermore, recall the following well-known identities

\[
Q_{11} = \Sigma_{11}^{-1.2}, \quad Q_{21} = -\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11.2}^{-1} \quad \text{and} \quad Q_{22} = \Sigma_{22}^{-1},
\]

where \( \Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \).

With the above notation in mind, we now introduce the following measure which will be an important ingredient in our generalization of Stein’s Lemma.

**Definition 4.** Consider a random vector \( X \sim \mathcal{E}_n(\mu, \Sigma, g) \) with \( EX < \infty \). Furthermore, let \( C_{11} \) and \( \mu_1 \) be as above and \( v_m \) denote the \( m \)-dimensional associate measure of \( X \). Then \( v_m^* \) denotes the image measure of \( v_m \) with respect to the affine transformation \( x_1 \mapsto \mu_1 + C_{11}x_1 \).

If the covariance of \( X \) exists, \( v_m^* \) has a specific interpretation. To see this, let \( Z_1^* \) denote the spherical random vector generated by \( \overline{G}_m \), i.e. \( Z_1^* \sim S_m(n\overline{G}_m/ER^2) \). The random vector

\[
X_1^* = \mu_1 + C_{11}Z_1^*
\]

then follows an elliptical distribution, \( X_1^* \sim \mathcal{E}_m(\mu_1, \Sigma_{11}, n\overline{G}_m/ER^2) \), and \( v_m^* \) coincides with the distribution of \( X_1^* \) up to a normalizing constant. Recall that for \( 3 \leq m < n \) we also have that \( X_1^* \sim \mathcal{E}_m(\mu_1, \Sigma_{11}, ng_{m-2}/(2\pi R^2)) \).
As in the original version of Stein’s Lemma, our generalization concerns functions $h : \mathbb{R}^n \to \mathbb{R}$ with certain differentiability properties. More specifically, we define as in Stein [18]:

**Definition 5.** A function $h : \mathbb{R}^n \to \mathbb{R}$ is called almost differentiable if there exists a vector-function $\nabla h : \mathbb{R}^n \to \mathbb{R}^n$ such that, for all $z \in \mathbb{R}^n$,

$$h(x + z) - h(x) = \int_0^1 z' \nabla h(x + tz) \, dt$$

for almost all $x \in \mathbb{R}^n$.

Note that a function is almost differentiable if and only if all its coordinate functions are; the derivative of the $i$th coordinate function is a.e. the $i$th coordinate function of $\nabla h$ which will be denoted by $\nabla_i h$.

The following theorem now gives the desired generalization of Stein’s Lemma.

**Theorem 2.** Let $X \sim \mathcal{E}_n(\mu, \Sigma, g)$ be an $n$-dimensional elliptical vector with density generator $g$ and finite expectation $\mu$ and let $v_m^*$ be as in Definition 4. Furthermore, let $h : \mathbb{R}^m \to \mathbb{R}$, $1 < m < n$ be an almost differentiable function satisfying

$$\int_{\mathbb{R}^m} \| \nabla h(x_1) \| \, dv_m^*(x_1) < \infty,$$

where $\| \cdot \|$ denotes the Euclidean norm on $\mathbb{R}^m$. Then

$$E\left(h(X_1)(X_2 - \mu_2)\right) = \Sigma_{21} \int_{\mathbb{R}^m} \nabla h(x_1) \, dv_m^*(x_1).$$

(12)

If, in addition, the covariance matrix of $X$ is finite, then, with $X_1^*$ given by (11),

$$E\left(h(X_1)(X_2 - \mu_2)\right) = \text{Cov}(X_2, X_1)E\nabla h(X_1^*).$$

(13)

**Remark 1.** According to the discussion preceding Definition 3 it follows that the condition $E(X) < \infty$ is only needed when $m = 1$. Furthermore, observe that Theorem 2 does not require the existence of $\text{Cov}(X)$.

The proof of Theorem 2 uses a result which we formulate first as a separate lemma.

**Lemma 3.** Let $X$ be an elliptical vector and $f$ an almost differentiable function, which meet the assumptions of Theorem 2. Then

$$\int_{\mathbb{R}^m} \nabla f(z) \, dv_m(z) = \int_{\mathbb{R}^m} z f(z) g_m\left(\frac{z'z}{2}\right) \, dz.$$  

(14)

**Proof.** The lemma is essentially an application of Fubini’s Theorem. Fix some $1 < i < m$, without loss of generality $i = 1$, and partition $z$ according to $z' = (z_1, z_{1-1}')$ where $z_{1-1}' = (z_2, \ldots, z_m)$. Then

$$\int \nabla_1 f(z) \, dv_m(z) = \int \nabla_1 f(z) G_m\left(\frac{1}{2}(z'z)\right) \, dz = \int \nabla_1 f(z) \int_{(z'z)/2}^{\infty} g_m(u) \, du \, dz.$$
Substituting \( u = s^2/2 + (z'_{-1}z_{-1})/2 \) gives
\[
\int \nabla f(z) \int_{z'/2}^{\infty} g_m(u) \, du \, dz
\]
\[
= \int_{\mathbb{R}^m-1} \int_0^\infty \nabla f(z_1, z_{-1}) \int_{z_1}^\infty s g_m(s^2/2 + (z'_{-1}z_{-1})/2) \, ds \, dz_1 \, dz_{-1}
\]
\[
- \int_{\mathbb{R}^m-1} \int_{-\infty}^0 \nabla f(z_1, z_{-1}) \int_{-\infty}^{z_1} s g_m(s^2/2 + (z'_{-1}z_{-1})/2) \, ds \, dz_1 \, dz_{-1}
\]
\[
= \int_{\mathbb{R}^m-1} \int_0^\infty s g_m(s^2/2 + (z'_{-1}z_{-1})/2) \int_0^s \nabla f(z_1, z_{-1}) \, dz_1 \, ds \, dz_{-1}
\]
\[
- \int_{\mathbb{R}^m-1} \int_{-\infty}^0 s g_m(s^2/2 + (z'_{-1}z_{-1})/2) \int_0^0 \nabla f(z_1, z_{-1}) \, dz_1 \, ds \, dz_{-1}
\]
\[
= \int_{\mathbb{R}^m-1} \int_{-\infty}^\infty s g_m(s^2/2 + (z'_{-1}z_{-1})/2)(f(s, z_{-1}) - f(0, z_{-1})) \, ds \, dz_{-1}
\]
where the step (*) follows from Fubini’s Theorem. In addition,
\[
\int_{-\infty}^{\infty} z_1 g_m(z_1^2/2 + (z'_{-1}z_{-1})/2)f(0, z_{-1}) \, dz_1 = 0,
\]
which finally gives (14). □

**Proof of Theorem 2.** Partitioning \( x \) in the same way as \( X \), i.e. writing \( x' = (x'_1, x'_2) \) and using the notation (9) we have that
\[
(x - \mu)'Q(x - \mu) = (x_1 - \mu_1)'Q_{11}(x_1 - \mu_1) + 2(x_1 - \mu_1)'Q_{12}(x_2 - \mu_2)
\]
\[
+ (x_2 - \mu_2)'Q_{22}(x_2 - \mu_2),
\]
which can be further rewritten as
\[
(x - \mu)'Q(x - \mu) = (x_1 - \mu_1)'Q_{11,2}(x_1 - \mu_1)
\]
\[
+ \left( \sqrt{Q_{22}^{-1}}Q_{21}(x_1 - \mu_1) + \sqrt{Q_{22}^{-1}}(x_2 - \mu_2) \right)
\]
\[
\times \left( \sqrt{Q_{22}^{-1}}Q_{21}(x_1 - \mu_1) + \sqrt{Q_{22}^{-1}}(x_2 - \mu_2) \right).
\]
Setting
\[
z_1 = C_{11}^{-1}(x_1 - \mu_1) \quad \text{and} \quad z_2 = \sqrt{Q_{22}^{-1}}Q_{21}(x_1 - \mu_1) + \sqrt{Q_{22}^{-1}}(x_2 - \mu_2)
\]
yields that \((x - \mu)'Q(x - \mu) = z'z\) where \(z' = (z_1', z_2')\). This can be used in the evaluation of the left-hand side of (12) as follows:

\[
E\left( h(X_1)(X_2 - \mu_2) \right) = \frac{1}{|\Sigma|^{1/2}} \int h(x_1)(x_2 - \mu_2)g_n \left( \frac{1}{2}(x - \mu)'Q(x - \mu) \right) dx
\]

\[
= \frac{|C|}{|\Sigma|^{1/2}} \int h(\mu_1 + C_{11}z_1) \left( -Q_{22}^{-1}Q_{21}C_{11}z_1 + \sqrt{Q_{22}^{-1}}z_2 \right) g_n \left( \frac{1}{2}(z'z) \right) dz.
\]

where \(|C|\) denotes the corresponding Jacobian. As

\[
C = \begin{pmatrix}
C_{11} & 0 \\
-Q_{22}^{-1}Q_{21}C_{11} & \sqrt{Q_{22}^{-1}}
\end{pmatrix},
\]

it further follows that \(|C| = |C_{11}| \left| \sqrt{Q_{22}^{-1}} \right| = \sqrt{|Q_{11,2}^{-1}|Q_{22}^{-1}} = \sqrt{|\Sigma|}\) which leads to

\[
E\left( h(X_1)(X_2 - \mu_2) \right) = \int h(\mu_1 + C_{11}z_1) \left( -Q_{22}^{-1}Q_{21}C_{11}z_1 + \sqrt{Q_{22}^{-1}}z_2 \right) g_n \left( \frac{1}{2}(z'z) \right) dz
\]

\[
= \sqrt{Q_{22}^{-1}} \int z_2 h(\mu_1 + C_{11}z_1)g_n \left( \frac{1}{2}(z'z) \right) dz
\]

\[
- Q_{22}^{-1}Q_{21}C_{11} \int z_1 h(\mu_1 + C_{11}z_1)g_n \left( \frac{1}{2}(z'z) \right) dz.
\]

Spherical symmetry implies that

\[
\int z_2 g_n \left( \frac{1}{2}(z_1'z_1 + z_2'z_2) \right) dz_2 = 0
\]

and we are left with

\[
E\left( h(X_1)(X_2 - \mu_2) \right) = -Q_{22}^{-1}Q_{21}C_{11} \int z_1 h(\mu_1 + C_{11}z_1)g_n \left( \frac{1}{2}(z'z) \right) dz
\]

\[
= -Q_{22}^{-1}Q_{21}C_{11} \int z_1 h(\mu_1 + C_{11}z_1)g_n \left( \frac{1}{2}(z'_1z_1) \right) dz_1.
\]

With (10) we first obtain that

\[
- Q_{22}^{-1}Q_{21}C_{11} = Q_{22}^{-1}Q_{21}Q_{11,2} = -\Sigma_{22,1}(-\Sigma_{22,1}^{-1}\Sigma_{21}\Sigma_{11}^{-1})\Sigma_{11} = \Sigma_{21}.
\]

As can be readily verified, for \(f(z_1) = h(\mu_1 + C_{11}z_1)\) we further have that \(\nabla f(z_1) = C_{11} \nabla h(\mu_1 + C_{11}z_1)\). Lemma 3 then yields

\[
E\left( h(X_1)(X_2 - \mu_2) \right) = \Sigma_{21} \int \nabla h(\mu_1 + C_{11}z_1) d\nu_m(z_1)
\]

\[
= \Sigma_{21} \int \nabla h(z_1) d\nu_m^*(z_1).
\]
In case \( \text{Cov}(X) \) exists, \( n^{m_1}/ER^2 \) is the distribution of \( X_1^* \) and we have that
\[
E \left( h(X_1)(X_2 - \mu_2) \right) = \frac{ER^2}{n} \Sigma_{21} E \nabla h(X_1^*).
\]
Proposition 3 yields \( \frac{ER^2}{n} \Sigma_{21} = \text{Cov}(X_2, X_1) \), which completes the proof. \( \square \)

Remark 2. Note that the only difference between the classical result of Stein for multivariate normal distributions as given in Stein [18], is that in Eq. (13), \( E \nabla h(X_1^*) \) is replaced by \( E \nabla h(X_1) \).

According to Fang et al. [6, Theorem 2.7.5], we have that an elliptical distribution is a normal distribution if and only if two marginal densities of different dimensions exist and have functional forms which agree up to a positive constant. As \( G_m = g_m - 2/m \) for \( 3 \leq m < n \) by (8), it follows that \( X_1^* \overset{d}{=} X_1 \) if and only if \( X \) follows a multivariate normal distribution. If \( n < 3 \), one can for instance use the fact that \( G_m \) is proportional to the \( m \)-dimensional marginal density of \( S_{m+2}(g_m) \).

4. Examples

In this section, we illustrate the generalization of Stein’s Lemma given by Theorem 2. More specifically, we calculate the distribution of \( X_1^* \) featuring in (13) for several families of elliptical distributions which may be useful for applications. First, recall that if the covariance of \( X \sim E_n(\mu, \Sigma, g) \) exists, (11) yields that \( X_1^* \sim E_m(\mu_1, \Sigma_{11}, g_m^*) \) where \( g_m^* = nG_m/ER^2 \).

We begin with the simplest example of the multivariate normal distribution.

Example 1. Consider \( X \sim N_n(\mu, \Sigma) \). As is well known, the marginal density generators are given by
\[
g_m(x) = \frac{1}{(2\pi)^{m/2}} \exp(-x), \quad 1 \leq m < n.
\]
Furthermore, it is straightforward to verify that \( \overline{G}_m(x) = g_m(x), \quad x \in [0, \infty) \) and \( 1 \leq m < n \).

Therefore, \( X_1^* \overset{d}{=} X_1 \sim N_m(\mu_1, \Sigma_{11}) \) and formula (13) reduces to
\[
E(h(X_1)(X_2 - \mu_2)) = \text{Cov}(X_2, X_1)E \nabla h(X_1),
\]
which is the classical version of Stein’s Lemma.

As already mentioned in Remark 2, the multivariate normal distribution is the only example where the general version of Stein’s Lemma reduces to the classical one. The identity (8), however, suggests that the form of the cumulative generator simplifies for families of distributions with generators exhibiting certain stability, referred to as consistency. To make this explicit, let us first emphasize the dependence of the density generator on the dimension \( n \) by writing \( g(x, n) \) instead of \( g(x) \). A family of density generators \( \{g(x, n) \in \mathbb{N}\} \) is called consistent if
\[
\int_{-\infty}^{\infty} g \left( \frac{1}{2} \sum_{i=1}^{n+1} x_i^2, n + 1 \right) dx_{n+1} = g \left( \frac{1}{2} \sum_{i=1}^{n} x_i^2, n \right).
\]
In other words, the \( m \)-dimensional marginal density generator corresponding to \( g(x, n) \) is precisely \( g(x, m), 1 \leq m \leq n \). However, as shown in Kano [11], elliptical families with consistent generators
are precisely normal variance mixtures with a mixing variable which is independent of \( n \). We discuss these in the next example.

**Example 2.** For the sake of simplicity, consider random vectors \( X \) with the stochastic representation

\[
X \overset{d}{=} \sqrt{W} Y,
\]

where \( Y \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I}_n) \) and \( W \) is a non-negative scalar-valued random variable with distribution function \( F_W \) and independent of \( Y \). If \( W \) has no point mass at zero, \( X \sim \mathcal{E}_n(\mathbf{0}, \mathbf{I}_n, g) \) where

\[
g(x) = \frac{1}{(2\pi)^{n/2}} \int_0^\infty \frac{1}{w^{n/2}} \exp(-x/w) \, dF_W(w),
\]

see McNeil et al. [15, p. 74]. With (15) it is immediate that the \( m \)-dimensional margin \( X_1 \) of \( X \) is again a normal mixture with the same mixing variable \( W \) for \( 1 \leq m < n \), i.e. \( X_1 \overset{d}{=} \sqrt{W} Y_1 \) where \( Y_1 \) denotes the \( m \)-dimensional margin of \( Y \). The density generator of \( X_1 \) is hence given by

\[
g_m(x) = \frac{1}{(2\pi)^{m/2}} \int_0^\infty \frac{1}{w^{m/2}} \exp(-x/w) \, dF_W(w).
\]

In particular, therefore, \( g_m \) is of the same form as \( g \) which confirms the consistency of the family of density generators given by (16).

If \( \mathbb{E}(\sqrt{W}) < \infty \) then \( \mathbb{E}X \) is finite (and equal to \( \mathbf{0} \)). Under this condition, the cumulative density generator is a.s. finite and we obtain that

\[
\overline{G}_m(x) = \int_x^\infty g_m(u) \, du = \frac{1}{(2\pi)^{m/2}} \int_0^\infty \frac{1}{w^{m/2}} \int_x^\infty \exp(-u/w) \, du \, dF_W(w)
\]

\[
= \frac{1}{(2\pi)^{m/2}} \int_0^\infty \frac{1}{w^{m/2}} \exp(-x/w/w) \, dF_W(w).
\]

Provided \( W \) has finite expectation, \( \text{Cov}(X) \) exists and \( X_1^* \sim \mathcal{E}_m(\mathbf{0}, \mathbf{I}_m, n\overline{G}_m/\mathbb{E}R^2) \) where \( R \) denotes the radial part of \( X \). Furthermore, as shown e.g. in McNeil et al. [15, p. 74], \( \text{Cov}(X) = \mathbb{E}W\mathbb{I}_n \). Proposition 3(ii) then implies \( \mathbb{E}R^2 = n\mathbb{E}W \). Hence,

\[
\frac{n\overline{G}_m(x)}{\mathbb{E}R^2} = \frac{\overline{G}_m(x)}{\mathbb{E}W} = \frac{1}{(2\pi)^{m/2}} \int_0^\infty \frac{1}{w^{m/2}} \exp(-x/w) \, w \, dF_W(w).
\]

Comparing this expression with (16) yields \( X_1^* \overset{d}{=} \sqrt{W^*} Y_1 \) where \( W^* \) is a non-negative random variable, independent of \( Y_1 \). Furthermore, the distribution of \( W^* \) is absolutely continuous with respect to \( F_W \) with Radon–Nikodym derivative \( \frac{dF_{W^*}}{dF_W}(x) = \frac{x}{\mathbb{E}W} \) a.e. Consequently, if \( F_W \) is absolutely continuous then so is \( W^* \) and

\[
f_{W^*}(x) = \frac{x}{\mathbb{E}W} f_W(x) \quad \text{a.e.} \tag{17}
\]

holds for the corresponding densities.

Examples of normal variance mixture distributions are for instance the Student \( t \) and the symmetric generalized hyperbolic distributions which are useful in many applications, in particular in quantitative risk management and mathematical finance; see e.g. McNeil et al. [15]. The next two examples therefore discuss these distributions in greater detail.
Example 3. The Student $t$ distribution with $k$ degrees of freedom can be viewed as a normal variance mixture where $W$ follows an inverse gamma distribution, $W \sim \text{Ig}(k/2, k/2)$; see e.g. McNeil et al. [15]. Following the notation therein, we first consider a random vector $X \sim t_n(0, I_n, k)$, i.e. $X \overset{d}{=} \sqrt{W}Y$ where $Y \sim N_n(0, I_n)$. As is well known, $EW$ exists if $k > 2$ and equals $k/(k - 2)$. By (17), the density of $W^*$ is

$$f_{W^*}(x) = \frac{(k - 2)(k/2)^{k/2}}{k\Gamma(k/2)} x^{-(k/2)} \exp(-k/(2x))$$

and hence $W^* \sim \text{Ig}((k - 2)/2, k/2)$. In other words, $X_1^*$ is still a normal variance mixture with inverse gamma mixing variable, but no longer $t$-distributed. However, we can remedy this inconvenience by scaling. Observe that $(k - 2)W^*/k \sim \text{Ig}((k - 2)/2, (k - 2)/2)$ and hence $\sqrt{(k - 2)/k}X_1^* \sim t_m(0, I_m, k - 2)$. This in particular shows that $X_1^*$ has heavier tails than $X_1$.

The case $X \sim t_n(\mu, \Sigma, k)$ can be treated similarly by noting that $X \overset{d}{=} \mu + \sqrt{\Sigma}Y$ where $Y \sim t_n(0, I_n, k)$. Then, $\sqrt{(k - 2)/k}X_1^* \sim t_m(\sqrt{(k - 2)/k} \mu_1, \Sigma_{11}, k - 2)$.

Example 4. In the case of the symmetric hyperbolic distribution the mixing variable $W$ is distributed according to a generalized inverse Gaussian distribution, $W \sim N^- (\lambda, \chi, \psi)$. The latter has a density of the form

$$f_W(x) = \frac{\chi^{\lambda/2} (\sqrt{\chi} \psi)^{\lambda/2}}{2 K_{\lambda}(\sqrt{\chi} \psi)} x^{\lambda-1} \exp\left(-\frac{1}{2}(\chi x^{-1} + \psi x)\right), \quad x > 0,$$

where $K_{\lambda}$ denotes the modified Bessel function of the third kind with index $\lambda$ and the parameters satisfy $\chi > 0, \psi \geq 0$ if $\lambda < 0, \chi > 0, \psi > 0$ if $\lambda = 0$ and $\chi \geq 0, \psi > 0$ if $\lambda > 0$. For $\chi > 0$ and $\psi > 0$, the expectation of $W$ is given by

$$EW = \frac{\chi^{\lambda/2} K_{\lambda+1}(\sqrt{\chi} \psi)}{\psi^{\lambda/2} K_{\lambda}(\sqrt{\chi} \psi)},$$

see for instance McNeil et al. [15]. With (17) the density of $W^*$ then becomes

$$f_{W^*}(x) = \frac{\chi^{-(\lambda+1)} (\sqrt{\chi} \psi)^{\lambda+1}}{2 K_{\lambda+1}(\sqrt{\chi} \psi)} x^{\lambda} \exp\left(-\frac{1}{2}(\chi x^{-1} + \psi x)\right).$$

In other words, $W^*$ follows again a generalized inverse Gaussian distribution, $W^* \sim N^- (\lambda + 1, \chi, \psi)$. Consequently, $X_1^*$ is symmetric generalized hyperbolic. Moreover, the parameters are the same as for $X_1$ with the sole exception of $\lambda$, which in the case of $X_1^*$ is replaced by $\lambda + 1$.

There exist other interesting families of elliptical distributions which are not consistent. These include for instance the Logistic, Pearson Type II and Pearson Type VII, symmetric Kotz Type or multivariate Bessel distributions; see Kano [11] for the proof that the above mentioned families are indeed inconsistent. In such cases, the calculation of the marginal density generators and consequently of the cumulative density generators as well is usually more complicated and seldom leads to closed expressions. An exception are families of normal variance mixtures where the mixing variable depends on $n$. According to Kano [11], such families are no longer consistent.
However, several results from Example 2 still do apply leading to tractable expressions for the marginal as well as cumulative density generators. We illustrate this for the example of Pearson Type VII distributions.

**Example 5.** A random vector $X \sim \mathcal{E}_n(\mu, \Sigma, g)$ follows a multivariate Pearson Type VII distribution if the density generator is of the form

$$g(x) = \frac{\Gamma(p)}{\Gamma(p - n/2)(\pi k)^{n/2}} \left(1 + \frac{2x}{k}\right)^{-p}, \quad p > n/2, \quad k > 0. \quad (18)$$

We will use the notation $X \sim \text{MPVII}_n(\mu, \Sigma, p, k)$. Observe that $X$ allows a stochastic representation $X \stackrel{d}{=} \mu + \sqrt{W_n} \sqrt{\Sigma} Y$ where $Y \sim \mathcal{N}_n(0, I_n)$ and $W_n$ is inverse gamma, i.e. $W_n \sim \text{Ig}(p - n/2, k/2)$. The specific choices $p = n/2 + k/2$ where $k$ is an integer and $p = n/2 + 1/2$ yield the Student $t$ distribution and the multivariate Cauchy distribution, respectively.

The $m$-dimensional marginals of $X$ have density generators of the form

$$g_m(x) = \frac{\Gamma(p - (n - m)/2)}{\Gamma(p - n/2)(\pi k)^{m/2}} \left(1 + \frac{2x}{k}\right)^{-p+(n-m)/2}.$$ 

Setting $\tilde{p} = p - n/2 + m/2$ yields that $X_1 \sim \text{MPVII}_m(\mu_1, \Sigma_{11}, \tilde{p}, k)$.

The expectation of $X$ exists if and only if $\text{E}(\sqrt{W_n}) < \infty$ which is the case if and only if $p > n/2 + 1/2$. The $m$-dimensional cumulative density generator of $X$ is then

$$G_m(x) = \frac{k\Gamma(p - (n - m)/2 - 1)}{2\Gamma(p - n/2)(\pi k)^{m/2}} \left(1 + \frac{2x}{k}\right)^{-p+(n-m)/2+1}.$$ 

Note that for the Cauchy distribution, the mean is not finite. Consequently, Theorem 2 is not applicable in this case (recall, however, Remark 1).

If, in addition, $p > n/2 + 1$ then $\text{Cov}(X)$ exists and the radial part of $X$ satisfies $\text{E}R^2 = (nk)/(2p - n - 2)$. Setting $p^* = p - (n - m)/2 - 1$ gives

$$\frac{n}{\text{E}R^2} G_m(x) = \frac{\Gamma(p^*)}{\Gamma(p^* - m/2)(\pi k)^{m/2}} \left(1 + \frac{2x}{k}\right)^{p^*}.$$ 

A comparison with (18) yields that $X_1^* \sim \mathcal{E}_m(\mu_1, \Sigma_{11}, nG_m/\text{E}R^2)$ satisfies

$$X_1^* \sim \text{MPVII}_m(\mu_1, \Sigma_{11}, p^*, k).$$

Hence, the sole difference between the distributions of $X_1$ and $X_1^*$ are the parameters $\tilde{p}$ and $p^*$, respectively. Finally, note that $\text{Cov}(X_1^*)$ does not exist while at the same time $\text{Cov}(X_1)$ is finite whenever $n/2 + 1 < p \leq n/2 + 2$. For $p > n/2 + 2$, the covariance of both $X_1$ and $X_1^*$ is finite.

**5. Conclusions and directions for further research**

This paper introduces a generalization of Stein’s Lemma for multivariate elliptical distributions. Under some regularity conditions, this result allows for the calculation of the moments $\text{E}(h(X_1)(X_2 - \mu_2))$ in terms of the dispersion matrix of $X = (X_1, X_2)$. If the covariance of $X$
exists, (13) in particular states that

\[ \mathbb{E}(h(X_1)(X_2 - \mu_2)) = \text{Cov}(X_2, X_1)\nabla h(X_1^*). \]

The random vector \( X_1^* \) is again elliptical, its density generator being equal to the \( m \)-dimensional cumulative density generator of \( X \) up to a normalizing constant. For \( X \) multivariate normal, the distributions of \( X_1^* \) and \( X_1 \) are identical.

Apart from theory and applications in asset pricing, Stein’s Lemma, in particular the identity (13), may be useful for estimation. An example is the following estimation procedure for \( \text{Cov}(h(X_1), X_2) \):

1. for a given data set, select a best-fitting elliptical model and estimate the mean, dispersion matrix and additional parameters if present;
2. evaluate the right-hand side of (13) for the best-fitting elliptical model and substitute estimates obtained in step (1) for unknown parameters.

The question under which conditions this estimation procedure outperforms the standard sample covariance estimator of \( \text{Cov}(h(X_1), X_2) \) is too complex to be treated here in detail; we merely include an illustration.

**Example 6.** Consider a bivariate random vector \((X_1, X_2) \sim t_2(\mu, \Sigma, k)\) and set \( h(x) = x^3 \). As shown in Example 3, \( \sqrt{(k-2)/k}X_1^* \sim t_1(\sqrt{(k-2)/k}\mu_1, \Sigma_{11}, k-2) \), provided \( k > 2 \). Therefore, \( \mathbb{E}(h(X_1^*)) = 3\mathbb{E}(X_1^*)^2 \) is finite if and only if \( k > 4 \). Under this assumption, (13) and elementary calculations yield

\[ \text{Cov}(h(X_1), X_2) = 3\Sigma_{12}\frac{k^2}{(k-2)^2} \left( \Sigma_{11}\frac{k-2}{k-4} + \mu_1^2\frac{k-2}{k} \right). \]  

(19)

For an iid sample from \( t_2(\mu, \Sigma, k) \), \( \mu, \Sigma \) and \( k \) can be estimated by maximum likelihood using EM algorithm, see McNeil et al. [15, Section 3.2] for details and further references. Substituting for the unknown parameters on the right-hand side of (19) then yields an estimator of \( \text{Cov}(h(X_1), X_2) \) which will be denoted by \( \hat{r}_{\text{MLE}} \). The special case when \( k \) is known leads to another estimator referred to as \( \hat{r}_{\text{CMLE}} \).

To study the performance of \( \hat{r}_{\text{MLE}} \) and \( \hat{r}_{\text{CMLE}} \) in comparison with the standard sample covariance estimator \( \hat{r}_{SC} \) of \( \text{Cov}(h(X_1), X_2) \), we have simulated 10,000 iid samples of size 100 from a bivariate \( t \) distribution with \( \mu = 0, k = 6 \) and \( \Sigma = \begin{pmatrix} 1 & 0.7 \\ 0.7 & 1 \end{pmatrix} \). In that case, it may be easily calculated that \( \text{Cov}(h(X_1), X_2) = 9.45 \). Fig. 1 displays the outcome of this simulation experiment.

It is first clear that \( \hat{r}_{\text{CMLE}} \) offers a significant improvement. However, the knowledge of the degrees of freedom may be a too optimistic assumption in practice. It is therefore unfortunate to note that \( \hat{r}_{\text{MLE}} \) performs poorly, even worse than \( \hat{r}_{SC} \): the ranges of \( \hat{r}_{\text{MLE}} \) and \( \hat{r}_{SC} \) were \((1.22, 36048.32) \) and \((-20.48, 3086.41) \), respectively. This is most likely due to the fact that \( \hat{r}_{\text{MLE}} \) is very sensitive to \( \hat{k} \) when \( k \) is close to 4 and \( \hat{k} \) may, for small samples, still be too far from the true value \( k = 6 \). In fact, ca. 9% of the samples even lead to \( \hat{k} < 4 \) and hence to an undefined value of \( \hat{r}_{\text{MLE}} \). The situation improves with increasing sample size; for \( N = 1000 \) (not shown here) \( \hat{r}_{\text{MLE}} \) and \( \hat{r}_{\text{SC}} \) perform similarly.

To conclude, we can say that our generalization of Stein’s Lemma may, in some situations, lead to efficient and robust estimators of \( \text{Cov}(h(X_1), X_2) \). This intuition is further supported by the fact that many attractive estimators of \( \text{Cov}(X_1, X_2) \) exist in the literature, see e.g. McNeil...
et al. [15] for details. However, the reader may recall that the distribution of $X_1^*$ depends on the underlying elliptical model and the above proposed estimation procedure may hence be sensitive to model misspecification. Furthermore, the conditions under which the new estimation procedure performs well are yet to be determined.

Acknowledgments

The first author acknowledges financial support by the FIM, ETH Zurich. The second author was supported by RiskLab, ETH Zurich. The authors thank Paul Embrechts for bringing Froot [8] to their attention and for fruitful discussions. They are grateful to the referees for their thorough comments and suggestions.
References