ims **Monographs**

The Skew-Normal and Related Families

Adelchi Azzalini
with the collaboration of Antonella Capitanio

CAMBRIDGE

The Skew-Normal and Related Families

ADELCHI AZZALINI
Università degli Studi di Padova

with the collaboration of

ANTONELLA CAPITANIO
Università di Bologna

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

Published in the United States of America by Cambridge University Press, New York

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781107029279

First published 2014

Printed in the United Kingdom by TJ International Ltd. Padstow Cornwall

A catalogue record for this publication is available from the British Library

Library of Congress Cataloguing in Publication data

Azzalini, Adelchi, author.
The skew-normal and related families / Adelchi Azzalini, Università degli Studi di Padova
with the collaboration of Antonella Capitanio, Università di Bologna.
pages cm
Includes bibliographical references and index.
ISBN 978-1-107-02927-9 (Hardback)
1. Distribution (Probability theory) I. Capitanio, Antonella, 1964– author. II. Title.
QA273.6.A98 2014
519.2′4–dc23 2013030070

ISBN 978-1-107-02927-9 Hardback

The Skew-Normal and Related Families

Interest in the skew-normal and related families of distributions has grown enormously over recent years, as theory has advanced, challenges of data have grown and computational tools have become more readily available. This comprehensive treatment, blending theory and practice, will be the standard resource for statisticians and applied researchers. Assuming only basic knowledge of (non-measure-theoretic) probability and statistical inference, the book is accessible to the wide range of researchers who use statistical modelling techniques.

Guiding readers through the main concepts and results, the book covers both the probability and the statistics sides of the subject, in the univariate and multivariate settings. The theoretical development is complemented by numerous illustrations and applications to a range of fields including quantitative finance, medical statistics, environmental risk studies and industrial and business efficiency. The authors' freely available R package `sn`, available from CRAN, equips readers to put the methods into action with their own data.

ADELCHI AZZALINI was Professor of Statistics in the Department of Statistical Sciences at the University of Padua until his retirement in 2013. Over the last 15 years or so, much of his work has been dedicated to the research area of this book. He is regarded as the pioneer of this subject due to his 1985 paper on the skew-normal distribution; in addition, several of his subsequent papers, some of which have been written jointly with Antonella Capitanio, are considered to represent fundamental steps. He is the author or co-author of three books, over 70 research papers and four packages written in the R language.

ANTONELLA CAPITANIO is Associate Professor of Statistics in the Department of Statistical Sciences at the University of Bologna. She began working on the skew-normal distribution about 15 years ago, co-authoring with Adelchi Azzalini a series of papers, related to the skew-normal and skew-elliptical distributions, which have provided key results in this area.

INSTITUTE OF MATHEMATICAL STATISTICS
MONOGRAPHS

IMS Monographs are concise research monographs of high quality on any branch of statistics or probability of sufficient interest to warrant publication as books. Some concern relatively traditional topics in need of up-to-date assessment. Others are on emerging themes. In all cases the objective is to provide a balanced view of the field.

Preface

Since about the turn of the millennium, the study of parametric families of probability distributions has received new, intense interest. The present work is an account of one approach which has generated a great deal of activity.

The distinctive feature of the construction to be discussed is to start from a symmetric density function and, by suitable modification of this, generate a set of non-symmetric distributions. The simplest effect of this process is represented by skewness in the distribution so obtained, and this explains why the prefix 'skew' recurs so often in this context. The focus of this construction is not, however, skewness as such, and we shall not discuss the quintessential nature of skewness and how to measure it. The target is instead to study flexible parametric families of continuous distributions for use in statistical work. A great deal of those in standard use are symmetric, when the sample space is unbounded. The aim here is to allow for possible departure from symmetry to produce more flexible and more realistic families of distributions.

The concentrated development of research in this area has attracted the interest of both scientists and practitioners, but often the variety of proposals and the existence of related but different formulations bewilders them, as we have been told by a number of colleagues in recent years. The main aim of this work is to provide a key to enter this theme. Besides its role as an introductory text for the newcomer, we hope that the present book will also serve as a reference work for the specialist.

This is not the first book covering this area: there exists a volume, edited by Marc Genton in 2004, which has been very beneficial to the dissemination of these ideas, but since its publication many important results have appeared and the state of the art is now quite different. Even today a definitive stage of development of this field has not been reached, if one assumes for a moment that such a state can ever be achieved, but we feel that the material is now sufficiently mature to also be fruitfully used for routine work of non-specialists.

The general framework and the key concepts of our development are formulated in Chapter 1. Subsequent chapters develop specific directions, in the univariate and in the multivariate case, and discuss why other directions are given lesser importance or even neglected. Some people may find it surprising that quite ample space is given to univariate distributions, considering that the context of multivariate distributions is where the new proposals appear more significant. However, besides its interest *per se*, the univariate case facilitates the exposition of many concepts, even when their main relevance is in the multivariate context.

There is a noticeable difference in the more articulate expository style of Chapters 1 to 6 compared with the briefer – even meagre one might say – summaries employed in Chapters 7 and 8, which deal with more specific themes. One reason for this choice is the greater importance given to the exposition of the basic concepts, recalling our main target in writing the book, and certain applied topics do not require a detailed discussion after the foundations of the construction are in place. Moreover, some of the more specialized or advanced topics are still in an evolutionary state, and any attempt to arrange them in an organized system is likely to become obsolete quite rapidly.

Chapters 1 to 6 are organized with a set of complements each, dealing with some more specialized topics. At first reading or if a reader is interested in getting a grasp of the key concepts only, these complements can be skipped without hindrance to understanding the core parts. At the end of these chapters there are sets of problems of varied levels of difficulty. As a rule of thumb, the harder ones are those with a reference at the end.

The development of this work has greatly benefited from the generous help of Giuliana Regoli, who has dedicated countless hours to examining and discussing with us many mathematical aspects. Obviously, any remaining errors are our own responsibility. We are also grateful to Elvezio Ronchetti, Marco Minozzo and Chris Adcock for comments on aspects of robustness, time series and quantitative finance, respectively, and to Marc Genton for several remarks on the nearly final draft. Even if in a less tangible form, our views on this research area have benefited from interactions with people of the 'skew community', with whom we have shared our enthusiasm during these years. It has been a stimulating and rewarding enterprise.

Adelchi Azzalini and Antonella Capitanio
February 2013

Contents

1

Modulation of symmetric densities

1.1 Motivation

This book deals with a formulation for the construction of continuous probability distributions and connected statistical aspects. Before we begin, a natural question arises: with so many families of probability distributions currently available, do we need any more?

There are three motivations for the development ahead. The first motivation lies in the essence of the mechanism itself, which starts with a continuous symmetric density function that is then modified to generate a variety of alternative forms. The set of densities so constructed includes the original symmetric one as an 'interior point'. Let us focus for a moment on the normal family, obviously a case of prominent importance. It is well known that the normal distribution is the limiting form of many non-normal parametric families, while in the construction to follow the normal distribution is the 'central' form of a set of alternatives; in the univariate case, these alternatives may slant equally towards the negative and the positive side. This situation is more in line with the common perception of the normal distribution as 'central' with respect to others, which represent 'departures from normality' rather than 'incomplete convergence to normality'.

The second motivation derives from the applicability of the mechanism to the multivariate context, where the range of tractable distributions is much reduced compared to the univariate case. Specifically, multivariate statistics for data in Euclidean space is still largely based on the normal distribution. Some alternatives exist, usually in the form of a superset, of which the most notable example is represented by the class of elliptical distributions. However, these retain a form of symmetry and this requirement may sometimes be too restrictive, especially when considering that symmetry must hold for all components.

The third motivation derives from the mathematical elegance and tractability of the construction, in two respects. First, the simplicity and generality of the construction is capable of encompassing a variety of interesting subcases without requiring particularly complex formulations. Second, the mathematical tractability of the newly generated distributions is, at least in some noteworthy cases, not much reduced compared to the original symmetric densities we started with. A related but separate aspect is that these modified families retain some properties of the parent symmetric distributions.

1.2 Modulation of symmetry

The rest of this chapter builds the general framework within which we shall develop specific directions in subsequent chapters. Consequently, the following pages adopt a somewhat more mathematical style than elsewhere in the book. Readers less interested in the mathematical aspects may wish to move on directly to Chapter 2. While this is feasible, it would be best to read at least to the end of the current section, as this provides the core concepts that will recur in subsequent chapters.

1.2.1 A fairly general construction

Many of the probability distributions to be examined in this book can be obtained as special instances of the scheme to be introduced below, which allows us to generate a whole set of distributions as a perturbed, or modulated, version of a symmetric probability density function f_0, which we shall call the *base density*. This base is *modulated*, or *perturbed*, by a factor which can be chosen quite freely because it must satisfy very simple conditions.

Since the notion of symmetric density plays an important role in our development, it is worth recalling that this idea has a simple and commonly accepted definition only in the univariate case: we say that the density f_0 is symmetric about a given point x_0 if $f_0(x - x_0) = f_0(x_0 - x)$ for all x, except possibly a negligible set; for theoretical work, we can take $x_0 = 0$ without loss of generality. In the d-dimensional case, the notion of symmetric density can instead be formulated in a variety of ways. In this book, we shall work with the condition of central symmetry: according to Serfling (2006), a random variable X is centrally symmetric about 0 if it is distributed as $-X$. In case X is a continuous variable with density function denoted $f_0(x)$, then central symmetry requires that $f_0(x) = f_0(-x)$ for all $x \in \mathbb{R}^d$, up to a negligible set.

Proposition 1.1 *Denote by f_0 a probability density function on $\mathbb{R}^d$, by $G_0(\cdot)$ a continuous distribution function on the real line, and by $w(\cdot)$ a real-valued function on $\mathbb{R}^d$, such that*

$$f_0(-x) = f_0(x), \quad w(-x) = -w(x), \quad G_0(-y) = 1 - G_0(y) \tag{1.1}$$

for all $x \in \mathbb{R}^d$, $y \in \mathbb{R}$. Then

$$f(x) = 2\, f_0(x)\, G_0\{w(x)\} \tag{1.2}$$

is a density function on $\mathbb{R}^d$.

Technical proof Note that $g(x) = 2\,[G_0\{w(x)\} - \frac{1}{2}]\, f_0(x)$ is an odd function and it is integrable because $|g(x)| \le f_0(x)$. Then

$$0 = \int_{\mathbb{R}^d} g(x)\,\mathrm{d}x = \int_{\mathbb{R}^d} 2\, f_0(x)\, G_0\{w(x)\}\,\mathrm{d}x - 1\,. \qquad \text{QED}$$

Although this proof is adequate, it does not explain the role of the various elements from a probability viewpoint. The next proof of the same statement is more instructive. In the proof below and later on, we denote by $-A$ the set formed by reversing the sign of all elements of A, if A denotes a subset of a Euclidean space. If $A = -A$, we say that A is a symmetric set.

Instructive proof Let Z_0 denote a random variable with density f_0 and T a variable with distribution G_0, independent of Z_0. To show that $W = w(Z_0)$ has distribution symmetric about 0, consider a Borel set A of the real line and write

$$\mathbb{P}\{W \in -A\} = \mathbb{P}\{-W \in A\} = \mathbb{P}\{w(-Z_0) \in A\} = \mathbb{P}\{w(Z_0) \in A\}\,,$$

taking into account that Z_0 and $-Z_0$ have the same distribution. Since T is symmetric about 0, then so is $T - W$ and we conclude that

$$\tfrac{1}{2} = \mathbb{P}\{T \le W\} = \mathbb{E}_{Z_0}\{\mathbb{P}\{T \le w(Z_0)|Z_0 = x\}\} = \int_{\mathbb{R}^d} G_0\{w(x)\}\, f_0(x)\,\mathrm{d}x\,.$$

QED

On setting $G(x) = G_0\{w(x)\}$ in (1.2), we can rewrite (1.2) as

$$f(x) = 2\, f_0(x)\, G(x) \tag{1.3}$$

where

$$G(x) \ge 0, \quad G(x) + G(-x) = 1\,. \tag{1.4}$$

Vice versa, any function G satisfying (1.4) can be written in the form $G_0\{w(x)\}$. For instance, we can set

$$\begin{aligned} G_0(y) &= \left(y + \tfrac{1}{2}\right) I_{(-1,1)}(2\,y) + I_{[1,+\infty)}(2\,y) \quad (y \in \mathbb{R}),\\ w(x) &= G(x) - \tfrac{1}{2} \quad (x \in \mathbb{R}^d), \end{aligned} \tag{1.5}$$

where $I_A(\cdot)$ denotes the indicator function of set A; more simply, this G_0 is the distribution function of a $\mathrm{U}(-\frac{1}{2}, \frac{1}{2})$ variate. We have therefore obtained the following conclusion.

Proposition 1.2 *For any given density f_0 in $\mathbb{R}^d$, such that $f_0(x) = f_0(-x)$, the set of densities of type (1.1)–(1.2) and those of type (1.3)–(1.4) coincide.*

Which of the two forms, (1.2) or (1.3), will be used depends on the context, and is partly a matter of taste. Representation of $G(x)$ in the form $G_0\{w(x)\}$ is not unique since, given any such representation,

$$G(x) = G_*\{w_*(x)\}, \quad w_*(x) = G_*^{-1}[G_0\{w(x)\}]$$

is another one, for any monotonically increasing distribution function G_* on the real line satisfying $G_*(-y) = 1 - G_*(y)$. Therefore, for mathematical work, the form (1.3)–(1.4) is usually preferable. In contrast, $G_0\{w(x)\}$ is more convenient from a constructive viewpoint, since it immediately ensures that conditions (1.4) are satisfied, and this is how a function G of this type is usually constructed. Therefore, we shall use either form, $G(x)$ or $G_0\{w(x)\}$, depending on convenience.

Since $w(x) = 0$ or equivalently $G(x) = \frac{1}{2}$ are admissible functions in (1.1) and (1.4), respectively, the set of modulated functions generated by f_0 includes f_0 itself. Another immediate fact is the following *reflection property*: if Z has distribution (1.2), $-Z$ has distribution of the same type with $w(x)$ replaced by $-w(x)$, or equivalently with $G(x)$ replaced by $G(-x)$ in (1.3).

The modulation factor $G_0\{w(x)\}$ in (1.2) can modify radically and in very diverse forms the base density. This fact is illustrated graphically by Figure 1.1, which displays the effect on the contour level curves of the base density f_0 taken equal to the $\mathrm{N}_2(0, I_2)$ density when the perturbation factor is given by $G_0(y) = e^y/(1 + e^y)$, the standard logistic distribution function, evaluated at

$$w(x) = \frac{\sin(p_1\, x_1 + p_2\, x_2)}{1 + \cos(q_1\, x_1 + q_2\, x_2)}, \qquad x = (x_1, x_2) \in \mathbb{R}^2, \tag{1.6}$$

for some choices of the real parameters p_1, p_2, q_1, q_2.

Densities of type (1.2) or (1.3) are often called *skew-symmetric*, a term which may be surprising when one looks for instance at Figure 1.1, where

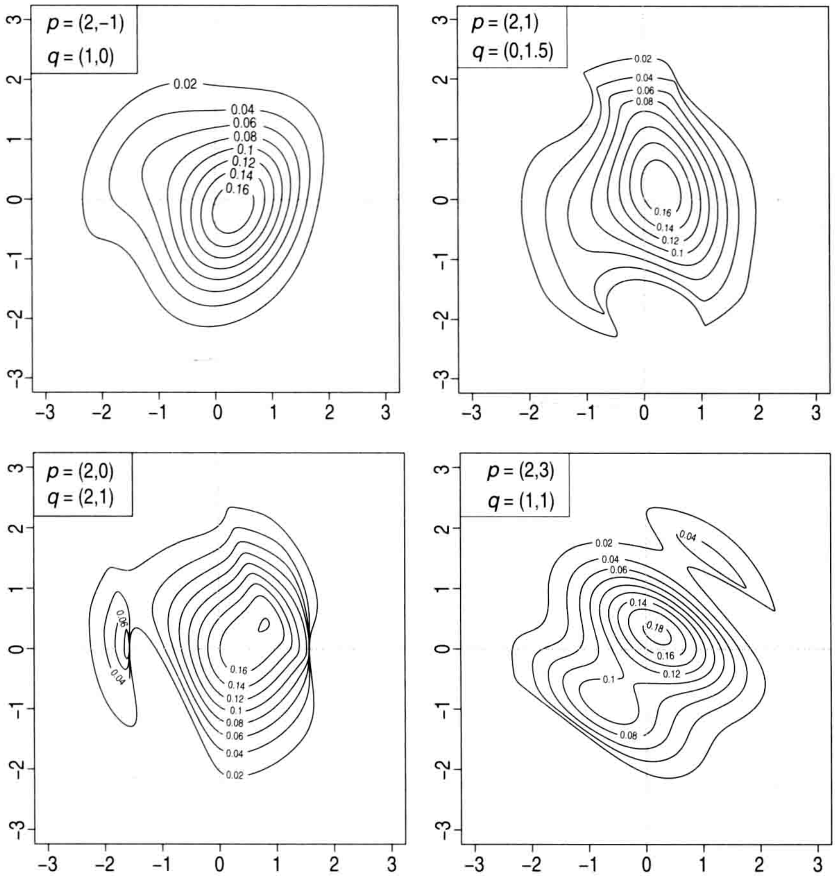

Figure 1.1 Density function of a bivariate standard normal variate with independent components modulated by a logistic distribution factor with argument regulated by (1.6) using parameters indicated in the top-left corner of each panel.

skewness is not the most distinctive feature of these non-normal distributions, apart from possibly the top-left plot. The motivation for the term 'skew-symmetric' originates from simpler forms of the function $w(x)$, which actually lead to densities where the most prominent feature is asymmetry. A setting where this happens is the one-dimensional case with linear form $w(x) = \alpha x$, for some constant α, a case which was examined extensively in the earlier stages of development of this theme, so that the prefix 'skew' came into use, and was later used also where skewness is not really the most distinctive feature. Some instances of the linear type will be discussed in detail later in this book, especially but not only in Chapter 2. However, in the more general context discussed in this chapter, the prefix 'skew' may be slightly misleading, and we prefer to use the term modulated or perturbed symmetry.

The aim of the rest of this chapter is to examine the general properties of the above-defined set of distributions and of some extensions which we shall describe later on. In subsequent chapters we shall focus on certain subclasses, obtained by adopting a specific formulation of the components f_0, G_0 and w of (1.2). We shall usually proceed by selecting a certain parametric set of functions for these three terms. We make this fact more explicit with notation of the form

$$f(x) = 2\, f_0(x)\, G_0\{w(x;\alpha)\}, \qquad x \in \mathbb{R}^d, \tag{1.7}$$

where $w(x;\alpha)$ is an odd function of x, for any fixed value of the parameter α. For instance, in (1.6) α is represented by (p_1, p_2, q_1, q_2). However, later on we shall work mostly with functions w which have a more regular behaviour, and correspondingly the densities in use will usually fluctuate less than those in Figure 1.1. In the subsequent chapters, we shall also introduce location and scale parameters, not required for the aims of the present chapter.

A word of caution on this programme of action is appropriate, even before we start to expand it. The densities displayed in Figure 1.1 provide a direct perception of the high flexibility that can be achieved with these constructions. And it would be very easy to proceed further, for instance by adding cubic terms in the arguments of $\sin(\cdot)$ and $\cos(\cdot)$ in (1.6). Clearly, this remark applies more generally to parametric families of type (1.7). However, when we use these distributions in statistical work, one must match flexibility with feasibility of the inferential process, in light of the problem at hand and of the available data. The results to be discussed make available powerful tools for constructing very general families of probability distributions, but power must be exerted with wisdom, as in other human activities.

1.2.2 Main properties

Proposition 1.3 (Stochastic representation) *Under the setting of Propositions 1.1 and 1.2, consider a d-dimensional variable Z_0 with density function $f_0(x)$ and, conditionally on Z_0, let*

$$S_{Z_0} = \begin{cases} +1 & \text{with probability } G(Z_0), \\ -1 & \text{with probability } G(-Z_0). \end{cases} \tag{1.8}$$

Then both variables

$$Z' = (Z_0 | S_{Z_0} = 1), \tag{1.9}$$

$$Z = S_{Z_0} Z_0 \tag{1.10}$$

have probability density function (1.2). The variable S_{Z_0} *can be represented in either of the forms*

$$S_{Z_0} = \begin{cases} +1 & \text{if } T < w(Z_0), \\ -1 & \text{otherwise,} \end{cases} \qquad S_{Z_0} = \begin{cases} +1 & \text{if } U < G(Z_0), \\ -1 & \text{otherwise,} \end{cases} \tag{1.11}$$

where $T \sim G_0$ *and* $U \sim U(0,1)$ *are independent of* Z_0.

Proof First note that marginally $\mathbb{P}\{S = 1\} = \int_{\mathbb{R}^d} G(x)\, f_0(x)\, \mathrm{d}x = \frac{1}{2}$, and then apply Bayes' rule to compute the density of Z' as the conditional density of $(Z_0|S = 1)$, that is

$$f_{Z'}(x) = \frac{\mathbb{P}\{S = 1 | Z_0 = x\}\; f_0(x)}{\mathbb{P}\{S = 1\}} = 2\,G(x)\, f_0(x)\,.$$

Similarly, the variable $Z'' = (Z_0|S_{Z_0} = -1)$ has density $2\,G(-x)\, f_0(x)$. The density of Z is an equal-weight mixture of Z' and $-Z''$, namely

$$\tfrac{1}{2}\{2\, f_0(x)\, G(x)\} + \tfrac{1}{2}\{2\, f_0(-x)\, G(x)\} = 2\, f_0(x)\, G(x)\,.$$

Representations (1.11) are obvious. QED

An immediate corollary of representation (1.10) is the following property, which plays a key role in our construction.

Proposition 1.4 (Modulation invariance) *If the random variable* Z_0 *has density* f_0 *and* Z *has density* f, *where* f_0 *and* f *are as in Proposition 1.1, then the equality in distribution*

$$t(Z) \overset{\mathrm{d}}{=} t(Z_0) \tag{1.12}$$

holds for any q-valued function $t(x)$ *such that* $t(x) = t(-x) \in \mathbb{R}^q$, $q \geq 1$.

We shall refer to this property also as *perturbation invariance*. An example of the result is as follows: if the density function of the two-dimensional variable (Z_1, Z_2) is one of those depicted in Figure 1.1, we can say that $Z_1^2 + Z_2^2 \sim \chi_2^2$, since this fact is known to hold for their base density f_0, that is when $(Z_1, Z_2) \sim \mathrm{N}_2(0, I_2)$ and $t(x) = x_1^2 + x_2^2$ is an even function of $x = (x_1, x_2)$.

An implication of Proposition 1.4 which we shall use repeatedly is that

$$|Z_r| \overset{\mathrm{d}}{=} |Z_{0,r}| \tag{1.13}$$

for the rth component of Z and Z_0, respectively, on taking $t(x) = |x_r|$. This fact in turn implies invariance of even-order moments, so that

$$\mathbb{E}\{Z_r^m\} = \mathbb{E}\left\{Z_{0,r}^m\right\}, \qquad m = 0, 2, 4, \ldots, \tag{1.14}$$

when they exist. Clearly, equality of even-order moments holds also for more general forms such as

$$\mathbb{E}\left\{Z_r^k\, Z_s^{m-k}\right\} = \mathbb{E}\left\{Z_{0,r}^k\, Z_{0,s}^{m-k}\right\}, \quad m = 0, 2, 4, \ldots; \quad k = 0, 1, \ldots, m.$$

It is intuitive that the set of densities of type (1.2)–(1.3) is quite wide, given the weak requirements involved. This impression is also supported by the visual message of Figure 1.1. The next result confirms this perception in its extreme form: all densities belong to this class.

Proposition 1.5 *Let f be a density function with support $S \subseteq \mathbb{R}^d$. Then a representation of type (1.3) holds, with*

$$f_0(x) = \tfrac{1}{2}\{f(x) + f(-x)\},$$
$$G(x) = \begin{cases} \dfrac{f(x)}{2 f_0(x)} & \text{if } x \in S_0, \\ \text{arbitrary} & \text{otherwise}, \end{cases} \tag{1.15}$$

where $S_0 = S \cup (-S)$ is the support of $f_0(x)$ and the arbitrary branch of G satisfies (1.4). Density f_0 is unique, and G is uniquely defined over S_0.

The meaning of the notation $-S$ is explained shortly after Proposition 1.1.

Proof For any $x \in S_0$, the identity

$$f(x) = 2\, \frac{f(x) + f(-x)}{2}\, \frac{f(x)}{f(x) + f(-x)}$$

holds, and its non-constant factors coincide with those stated in (1.15). To prove uniqueness of this factorization on S_0, assume that there exist f_0 and G such that $f(x) = 2\, f_0(x)\, G(x)$ and they satisfy $f_0(x) = f_0(-x)$ and (1.4). From

$$f(x) + f(-x) = 2\, f_0(x)\{G(x) + G(-x)\} = 2\, f_0(x),$$

it follows that f_0 must satisfy the first equality in (1.15). Since $f_0 > 0$ and it is uniquely determined over S_0, then so is $G(x)$. QED

Rewriting the first expression in (1.15) as $f(-x) = 2\, f_0(x) - f(x)$, followed by integration on $(-\infty, x_1] \times \cdots \times (-\infty, x_d]$, leads to

$$\overline{F}(-x) = 2\, F_0(x) - F(x), \qquad x = (x_1, \ldots, x_d) \in \mathbb{R}^d, \tag{1.16}$$