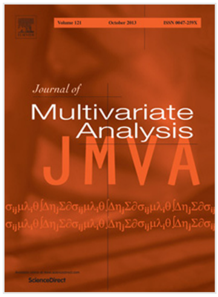
Accepted Manuscript



A general setting for symmetric distributions and their relationship to general distributions

P.E. Jupp, G. Regoli, A. Azzalini

PII: S0047-259X(16)00033-6
DOI: http://dx.doi.org/10.1016/j.jmva.2016.02.011
Reference: YJMVA 4090

To appear in: *Journal of Multivariate Analysis*

Received date: 30 May 2015

Please cite this article as: P.E. Jupp, G. Regoli, A. Azzalini, A general setting for symmetric distributions and their relationship to general distributions, *Journal of Multivariate Analysis* (2016), http://dx.doi.org/10.1016/j.jmva.2016.02.011

A general setting for symmetric distributions and their relationship to general distributions

P.E. Jupp[a,*], G. Regoli[b], A. Azzalini[c]

[a]*School of Mathematics and Statistics, University of St Andrews, St Andrews KY16 9SS, UK*
[b]*Dipartimento di Matematica e Informatica, Università di Perugia, 06123 Perugia, Italy*
[c]*Dipartimento di Scienze Statistiche, Università di Padova, 35121 Padova, Italy*

Abstract

A standard method of obtaining non-symmetrical distributions is that of modulating symmetrical distributions by multiplying the densities by a perturbation factor. This has been considered mainly for central symmetry of a Euclidean space in the origin. This paper enlarges the concept of modulation to the general setting of symmetry under the action of a compact topological group on the sample space. The main structural result relates the density of an arbitrary distribution to the density of the corresponding symmetrised distribution. Some general methods for constructing modulating functions are considered. The effect that transformations of the sample space have on symmetry of distributions is investigated. The results are illustrated by general examples, many of them in the setting of directional statistics.

Keywords: Directional statistics, Skew-symmetric distribution, Symmetry-modulated distribution, Transformation
2010 MSC: 62E10, 62H05

1. Introduction

Because many appealing distributions are symmetrical but many data sets are not, intense work has been dedicated to the study of families of tractable distributions obtained by modifying standard symmetrical distributions such

*Corresponding author
Email addresses: pej@st-andrews.ac.uk (P.E. Jupp), regoli@dipmat.unipg.it (G. Regoli), azzalini@stat.unipd.it (A. Azzalini)

as normal distributions. A major impetus in this development was [4], which introduced families of univariate distributions of the form

$$f(x;\lambda) = 2\,G_0(\lambda x)\,f_0(x), \qquad x \in \mathbb{R}, \tag{1}$$

where f_0 is a given density that is symmetrical about 0, G_0 is a distribution function having density that is symmetrical about 0 and λ is a real parameter. Hence a 'baseline' symmetric density f_0 is *modulated*, i.e., multiplied by a perturbation factor $G_0(\lambda x)$, to give an asymmetric distribution. If $\lambda = 0$ then $f = f_0$; otherwise f is asymmetric to the left or to the right, depending on the sign of λ. The most popular example of this construction is the skew-normal distribution, obtained when f_0 and G_0 are the standard normal density and its distribution function, respectively. This initial construction has subsequently been extended considerably, leading to an extensive literature. A comprehensive discussion is provided by [7].

A substantially more general version of (1), proposed independently in [25] and in [8] in slightly different forms, starts from a multivariate density f_0, satisfying the condition of central symmetry $f_0(x) = f_0(-x)$ for all $x \in \mathbb{R}^d$. Modulation of this baseline density to

$$f(x) = 2\,G(x)\,f_0(x), \qquad x \in \mathbb{R}^d, \tag{2}$$

is achieved via the perturbation factor $G(x)$ which satisfies

$$G(x) \geq 0, \qquad G(x) + G(-x) = 1. \tag{3}$$

A convenient mechanism for building a suitable such function G is to set $G(x) = G_0\{w(x)\}$, where w is an odd real-valued function, i.e., $w(-x) = -w(x)$. For any G of type (3), there are infinitely many G_0 and w such that $G(x) = G_0\{w(x)\}$. If $G(x) = 1/2$ or, equivalently $w(x) = 0$, then $f = f_0$. Densities (2) can take on shapes very different from that of the baseline f_0, for instance allowing for multimodality in cases where f_0 is unimodal.

An important property of density function (2) is that a random variable Z having density (2) can be represented as

$$Z = \begin{cases} Z_0 & \text{with probability } G(Z_0), \\ -Z_0 & \text{with probability } G(-Z_0), \end{cases} \tag{4}$$

where Z_0 has density f_0. One route to the constructive use of formula (4) is to generate a random variable $U \sim \mathcal{U}(0,1)$ independent of Z_0 and to examine whether or not $U \leq G(Z_0)$; an equivalent route is to generate $T \sim G_0$ and examine whether or not $T \leq w(Z_0)$. A corollary of construction (4) is the *modulation-invariance property*, which states that $t(Z_0)$ and $t(Z)$ have the same distribution for any $\mathbb{R}^q$-valued function t such that $t(x) = t(-x)$ for all x in $\mathbb{R}^d$.

Even more general formulations of (2) exist; see [3]. However, almost all formulations assume central symmetry of f_0, i.e., $f_0(x) = f_0(-x)$. One paper which moves away from this assumption is [6], where central symmetry is replaced by 'generalised symmetry', meaning that a random variable X on $\mathbb{R}^d$ has the same distribution as $R^{-1}(X)$ for some invertible mapping R from $\mathbb{R}^d$ to $\mathbb{R}^d$, a condition ensured by the requirement that if f_0 is the density of X then $f_0(x) = f_0\{R(x)\}$ and $|\det R'(x)| = 1$ for some invertible mapping R from $\mathbb{R}^d$ to $\mathbb{R}^d$. If $R(x) = -x$ then we return to the earlier construction. With an additional condition on R, one can establish an analogue of (4) and hence of the modulation-invariance property. A similar treatment for discrete variates is provided in [11].

The phrase 'skew-symmetric' is often used to refer to distributions of type (2). This arose because the construction evolved from that of the more restricted 'skew-elliptical' class. Since the phrase 'skew-symmetric' may convey the misleading message that skewness is the essential feature of distributions of this type, we prefer to call them 'symmetry-modulated' distributions.

Since all symmetry is symmetry under the action of some group, the aim of this paper is to extend known results on symmetrical and asymmetrical distributions to a very general setting. This programme is developed in the subsequent sections, where we consider measure-preserving actions of compact topological groups (such as finite groups or compact Lie groups) on sample spaces that are measure spaces.

2. A general setting for symmetry

Let K be a group and $K \times \mathcal{X} \to \mathcal{X}$ be an action of K on some measurable space $\mathcal{X}$. We shall write the action as $(k, x) \mapsto k.x$. By the definition of an action,

$$\begin{aligned} e.x &= x \\ k_2.(k_1.x) &= (k_2 k_1).x, \qquad k_1, k_2 \in K, \end{aligned}$$

where e denotes the identity element of K. A function f on $\mathcal{X}$ is K-*symmetric* (or K-*invariant*) if

$$f(k.x) = f(x), \qquad k \in K, x \in \mathcal{X}. \tag{5}$$

Similarly, a measure μ on $\mathcal{X}$ is K-*invariant* if

$$\mu(k.A) = \mu(A)$$

for all k in K and all measurable subsets A of $\mathcal{X}$. If μ is a K-invariant measure and X is a random variable with K-symmetric density then $k.X$ has the same distribution as X, for all k in K.

If K is a compact topological group then there is a (unique) left K-invariant and right K-invariant probability measure (Haar measure) ν on K. We say that a real-valued function f on $\mathcal{X}$ is K-*odd* if

$$\int_K f(k.x)\, d\nu(k) = 0, \qquad x \in \mathcal{X}.$$

If $K = C_2$, the group with 2 elements, acts by reflection in the origin then f is K-odd if and only if $f\{(-1).x\} = -f(x)$, where -1 denotes the non-identity element of C_2.

Define the quotient map $\pi : \mathcal{X} \to \mathcal{X}/K$ as the map that sends x in $\mathcal{X}$ to the corresponding equivalence class $[x] = \{kx : k \in K\}$. Then π sends any measure μ on $\mathcal{X}$ into a corresponding measure $\pi_*\mu$ on $\mathcal{X}/K$, given by

$$\pi_*\mu(B) = \mu(\pi^{-1}B) \tag{6}$$

for all measurable subsets B of $\mathcal{X}/K$.

If μ is a measure on $\mathcal{X}$ that is invariant under the action of K then there is a bijection $f_0 \mapsto \tilde{f}_0$ between densities f_0 (with respect to μ) on $\mathcal{X}$ that are K-symmetric (in the sense of (5)) and densities $\tilde{f}_0$ on $\mathcal{X}/K$ (with respect to $\pi_*\mu$), where $\tilde{f}_0([x]) = f_0(x)$.

Example 1. *Let $\mathcal{X}$ be the unit sphere S^{p-1} and $K = C_2$ act by $(-1).\mathbf{x} = -\mathbf{x}$. Then $\mathcal{X}/K$ is the projective space, $\mathbb{R}P^{p-1}$, and distributions on $\mathbb{R}P^{p-1}$ are identified with antipodally symmetric distributions on S^{p-1}.*

The main structural result relating arbitrary distributions to symmetric distributions is the following proposition. It states that the density of a random variable X on $\mathcal{X}$ is equivalent to the pair consisting of the marginal density of $[X]$ on $\mathcal{X}/K$ and the conditional density of X given $[X]$. This equivalence is a generalisation of that in (2) for the case of $\mathcal{X} = \mathbb{R}^d$ and $K = C_2$ acting by reflection in the origin. In this case, $|X|$ can be identified with $[X]$, $f_0(|x|)$ is the marginal density of $[X]$ at $[x]$, and $G(x)$ (given $|x|$) is regarded as the conditional density of X given $|X|$. This case is considered in greater detail after the proof of the proposition. For ease of exposition, Proposition 1 is stated as if densities and conditional densities were well-defined functions. There is a more careful version in which it is acknowledged that densities are equivalence classes of functions that differ on sets of measure zero.

Proposition 1. *Let the compact group K act on a measure space $\mathcal{X}$, $\pi : \mathcal{X} \to \mathcal{X}/K$ be the quotient map, and μ be a measure on $\mathcal{X}$ that is invariant under the action of K. Suppose that either K is finite or μ is σ-finite. Then the equations*

$$f(x) = \tilde{f}_0([x])\tilde{\Gamma}(x|[x]) \tag{7}$$

$$= f_0(x)\Gamma(x) \tag{8}$$

give bijections between

(a) densities f on $\mathcal{X}$ with respect to μ,
(b) pairs $(\tilde{f}_0, \tilde{\Gamma})$ with $\tilde{f}_0$ a density on $\mathcal{X}/K$ with respect to $\pi_*\mu$ and $\tilde{\Gamma}$ a family of conditional densities $\tilde{\Gamma}(\cdot|[x])$ on $\pi^{-1}([x])$ given $[x]$,
(c) pairs (f_0, Γ) with f_0 a K-symmetric density on $\mathcal{X}$ (with respect to μ) and Γ a non-negative function on $\{x \in \mathcal{X} : f_0(x) > 0\}$ such that

$$\int_K \Gamma(k.x)\, d\nu(k) = 1. \tag{9}$$

Proof. For x in $\mathcal{X}$, define $i_x : K \to \pi^{-1}([x])$ by $i_x(k) = k.x$ and define the measure $\nu_{[x]}$ on $\pi^{-1}([x])$ by $\nu_{[x]} = (i_x)_*\nu$, i.e.,

$$\int_{\pi^{-1}[x]} h(y)\, d\nu_{[x]}(y) = \int_K h(k.x)\, d\nu(k) \tag{10}$$

for every function h on $\pi^{-1}([x])$ for which the right hand side exists. It follows from K-invariance of ν that $\nu_{[x]}$ is well-defined.

Let $g : \mathcal{X} \to \mathbb{R}$ be μ-measurable. Because μ is K-invariant,

$$\int_{\mathcal{X}} g(k.x)\, d\mu(x) = \int_{\mathcal{X}} g(x)\, d\mu(x)$$

for all k in K and all integrable $g : \mathcal{X} \to \mathbb{R}$. Integrating this over K gives

$$\int_K \int_{\mathcal{X}} g(k.x) d\mu(x)\, d\nu(k) = \int_{\mathcal{X}} g(x)\, d\mu(x).$$

Since $(k, x) \mapsto g(k.x)$ is integrable and either K is finite or μ is σ-finite, Fubini's theorem gives

$$\begin{aligned} \int_{\mathcal{X}} g(x)\, d\mu(x) &= \int_K \int_{\mathcal{X}} g(k.x)\, d\mu(x)\, d\nu(k) \\ &= \int_{\mathcal{X}} \int_K g(k.x)\, d\nu(k)\, d\mu(x) \\ &= \int_{\mathcal{X}/K} \left\{ \int_{\pi^{-1}([x])} g(y)\, d\nu_{[x]}(y) \right\} d\pi_*\mu([x]), \end{aligned} \tag{11}$$

where equality of the inner integrals in the last two lines follows from (10).

Thus the measure μ is decomposed (disintegrated) in the sense of [12, Section 5] or [14] into the marginal measure $\pi_*\mu$ on $\mathcal{X}/K$ and the conditional measures $\nu_{[x]}$ on the fibres $\pi^{-1}([x])$.

Given a density f on $\mathcal{X}$, define $\tilde{f}_0$ and $\tilde{\Gamma}$ by

$$\begin{aligned} \tilde{f}_0([x]) &= \int_K f(k.x)\, d\nu(k), \\ \tilde{\Gamma}(y|[x]) &= f(y)/\tilde{f}_0([x]), \qquad y \in \pi^{-1}([x]), f_0(x) > 0. \end{aligned} \tag{12}$$

Then (7) holds. Replacing g in (11) by gf gives

$$\begin{aligned} &\int_{\mathcal{X}} g(x) f(x)\, d\mu(x) \\ &= \int_{\mathcal{X}/K} \left\{ \int_{\pi^{-1}([x])} g(y)\tilde{\Gamma}(y|[x])\, d\nu_{[x]}(y) \right\} \tilde{f}_0([x])\, d\pi_*\mu([x]). \end{aligned}$$

Thus the marginal and conditional densities are $\tilde{f}_0$ and $\tilde{\Gamma}(\cdot|[x])$, respectively. Conversely, such a $\tilde{f}_0$ and $\tilde{\Gamma}$ can be combined by (7) to give a density f on $\mathcal{X}$. Equivalence between (7) and (8) comes from putting $f_0(x) = \tilde{f}_0([x])$ and $\Gamma(x) = \tilde{\Gamma}(x|[x])$. Verification that (7) and (8) are bijections is straightforward. In particular, uniqueness of the decomposition in (7) follows from $f_0(x) = \tilde{f}_0([x])$ and (12). $\square$

In the case of $\mathcal{X} = \mathbb{R}^d$ and $K = C_2$ acting by reflection in the origin, i.e., $e.x = x$ and $(-1).x = -x$, (8) is usually written as

$$f(x) = 2f_0(x)G(x),$$

as in (2) (cf. ([4])), where $G(x) = \Gamma(x)/2$, i.e., G expresses the conditional density with respect to counting measure on C_2, whereas Γ expresses the conditional density with respect to the Haar *probability* measure ν. See Example 2. In this case, Proposition 1 becomes Propositions 1 and 3 of [25].

Remark 1. *It follows from (8) that if $f(x) > 0$ then $f_0(x) > 0$. If K is finite then it follows from (9) that $\{x \in \mathcal{X} : f_0(x) > 0\} = \bigcup_{k\in K} k.\{x \in \mathcal{X} : f(x) > 0\}$.* □

If K is a finite group of order m then (9) becomes

$$\frac{1}{m}\sum_{k\in K}\Gamma(k.x) = 1. \tag{13}$$

If $m = 2$ and the action of the non-identity element, -1, of C_2 on $\mathcal{X}$ is written as $(-1).x = R(x)$, as in [6], then (13) is equivalent to $G(x) + G\{R(x)\} = 1$, where $G(x) = \Gamma(x)/2$.

One way of looking at Proposition 1 is to say that modulation is not just one method of creating non-symmetrical distributions but the only method. The job of Γ is to break the symmetry of f_0 by redistributing probability within each $\pi^{-1}([x])$. The interesting non-symmetrical distributions are those in which the Γ have pleasant properties or are plausible ways of modelling the asymmetry of the data.

Remark 2. *Denote by $L^1(\mathcal{X})$ the space of real functions on $\mathcal{X}$ that are integrable with respect to μ. The subspaces $L^1_K(\mathcal{X})$ and $L^1_0(\mathcal{X})$ of K-symmetric and of K-odd functions in $L^1(\mathcal{X})$ are*

$$\begin{aligned} L^1_K(\mathcal{X}) &= \left\{f \in L^1(\mathcal{X}) : f(k.x) = f(x)\right\}, \\ L^1_0(\mathcal{X}) &= \left\{f \in L^1(\mathcal{X}) : \int_K f(k.x)\, d\nu(k) = 0\right\}, \end{aligned}$$

respectively. If $f \in L^1_K(\mathcal{X}) \cap L^1_0(\mathcal{X})$ then for $x \in \mathcal{X}$, $f(x) = \int_K f(k.x)\, d\nu(k) = 0$. Thus $L^1_K(\mathcal{X}) \cap L^1_0(\mathcal{X}) = \{0\}$. On the other hand, for f in $L^1(\mathcal{X})$, define $\bar{f}$ by

$$\bar{f}(x) = \int_K f(k.x)\, d\nu(k). \tag{14}$$

Then $f = \bar{f} + (f - \bar{f})$ *and* $\bar{f} \in L^1_K(\mathcal{X})$, $f - \bar{f} \in L^1_0(\mathcal{X})$. *Thus there is a direct sum decomposition*

$$L^1(\mathcal{X}) = L^1_K(\mathcal{X}) \oplus L^1_0(\mathcal{X}). \tag{15}$$

(The L^2 *analogue of (15) is considered in [18, Section 2].) For a density* f, *the functions* f_0 *and* Γ *in (8) satisfy* $f_0 \in L^1_K(\mathcal{X})$ *and* $\Gamma - 1 \in L^1_0(\mathcal{X})$. *The decomposition of* f *given by (15) is*

$$f = f_0 + f_0(\Gamma - 1). \tag{16}$$

This is an additive version of the multiplicative decomposition (8) of f. *Averaging (16) over* K *shows that*

$$f_0 = \bar{f},$$

where $\bar{f}$ *is the symmetrisation of* f *given by (14).* □

Remark 3. *Let the group* K *act both on* $\mathcal{X}$ *and on a space* Θ. *Among the parametric statistical models on* $\mathcal{X}$ *that are parameterised by* Θ, *a particularly nice class consists of those having densities (with respect to a* K*-invariant measure)* $f(\cdot;\theta)$ *that satisfy*

$$f(k.x; k.\theta) = f(x;\theta), \qquad k \in K. \tag{17}$$

Models satisfying (17) are composite transformation models in the sense of [13, Section 2.8]. Some important examples are those with $\mathcal{X} = \Theta = \mathbb{R}^d$, $K = C_2$ *acting by*

$$e.x = x, e.\theta = \theta, (-1).x = -x, (-1).\theta = -\theta,$$

where (-1) *denotes the non-identity element of* C_2, *and densities (with respect to Lebesgue measure)*

$$f(x;\theta) = 2\, G_0(\theta^\top x)\, f_0(x)$$

with f_0 *a centrally-symmetric density on* $\mathbb{R}^d$. *Then a simple calculation shows that these models satisfy (17).* □

Proposition 2. *Let* Z_0 *and* Z *be random variables on* $\mathcal{X}$ *with densities* f_0 *and* f, *respectively, where* f_0 *is* K*-symmetric and*

$$f(x) = f_0(x)\tilde{\Gamma}(x|[x]), \tag{18}$$

$\tilde{\Gamma}(\cdot|[x])$ *being the conditional density of* x *given* $[x]$. *If* $t : \mathcal{X} \to \mathcal{Y}$ *is invariant, i.e.,* $t(k.x) = t(x)$ *for all* k *in* K *then* $t(Z_0)$ *and* $t(Z)$ *have the same distribution.*

PROOF. The function $x \mapsto [x]$ is a maximal invariant, i.e., any invariant $t : \mathcal{X} \to \mathcal{Y}$ must have the form $t(x) = u([x])$ for some function $u : \mathcal{X}/K \to \mathcal{Y}$. By (18), $[Z_0]$ and $[Z]$ have the same distribution and so, therefore, do $u([Z_0])$ and $u([Z])$. □

In the case of $\mathcal{X} = \mathbb{R}^d$ and $K = C_2$ acting by reflection in the origin, Proposition 2 gives the modulation-invariance property described after (4).

Example 2. *(central symmetry on $\mathbb{R}^d$) Here $\mathcal{X} = \mathbb{R}^d$ and $K = C_2$ acts by reflection in the origin, i.e., $e.\mathbf{x} = \mathbf{x}$ and $(-1).\mathbf{x} = -\mathbf{x}$. In this case, a useful representation of functions Γ in (8) is*

$$\Gamma(\mathbf{x}) = 2\,G_0\{w(\mathbf{x})\}, \tag{19}$$

where G_0 is the cumulative distribution function of a univariate random variable symmetric about 0 and w is a real-valued function on $\mathbb{R}^d$ satisfying $w(-\mathbf{x}) = -w(\mathbf{x})$. Expression (19) is the traditional form of the perturbation factor recalled in the passage following (3). The traditional factor of 2 does not appear in (8), because the conditional density Γ is taken with respect to the probability *measure ν rather than counting measure on C_2.*

Example 3. *Here K is the additive group of the integers, acting by $(k, x) \mapsto R^k x$, where R is an invertible μ-preserving transformation of $\mathcal{X}$. The case of $\mathcal{X} = \mathbb{R}^d$ and functions Γ of the form (19) with $w\{R(x)\} = -w(x)$ is considered in [6, Sect. 2].*

Example 4. *Let $\mathcal{X} = \mathbb{R}^2$ and $K = C_s$, the cyclic group of order s. Consider the case in which K acts on $\mathcal{X}$ by rotations through multiples of $2\pi/s$, so that $k.\mathbf{x} = \mathbf{M}^k\mathbf{x}$ for $k = 0, \ldots, s-1$, where*

$$\mathbf{M} = \begin{pmatrix} \cos 2\pi/s & -\sin 2\pi/s \\ \sin 2\pi/s & \cos 2\pi/s \end{pmatrix}.$$

Let f_0 be a density on $\mathbb{R}^2$ with circular symmetry about the origin, e.g., the standard bivariate normal density. Consider a function $G : \mathbb{R}^2 \to [0,1]$ such that

$$\sum_{k=0}^{s-1} G(\mathbf{M}^k\mathbf{x}) = 1 \tag{20}$$

and define Γ *by* $\Gamma(\mathbf{x}) = s\,G(\mathbf{x})$. *Then* f, *defined by*

$$f(\mathbf{x}) = sf_0(\mathbf{x})G(\mathbf{x}), \tag{21}$$

is a probability density on $\mathbb{R}^2$.

The case $s = 2$ *is included in (19). The next step is to consider* $s = 3$. *One specific function* G *satisfying (20) is given by*

$$G(\mathbf{x}) = \begin{cases} 1/2 & \text{if } 1/2 < \|\mathbf{x}\|^{-1}\mathbf{x}^\top \mathbf{u}_0 \leq 1, \\ 1/4 & \text{otherwise} \end{cases} \tag{22}$$

for some fixed unit vector $\mathbf{u}_0$. *If* $f_0 : \mathbb{R}^2 \to (0, \infty)$ *is a probability density function which is constant on every circle with centre at the origin (such as an isotropic bivariate normal distribution with mean zero) then* f, *defined by*

$$f(\mathbf{x}) = 3\, f_0(\mathbf{x})\, G(\mathbf{x}), \tag{23}$$

is also a probability density function on $\mathbb{R}^2$. *An alternative, continuous, choice of* G *is the following. Define* $G_0 : [0, 2\pi/3) \to [0, 1/2)$ *and* $G_2 : [4\pi/3, 2\pi) \to (0, 1/2]$ *by*

$$G_0(\theta) = \frac{3}{4\pi}\sqrt{\left(\frac{2}{3}\pi\right)^2 - \left(\theta - \frac{2}{3}\pi\right)^2}, \quad G_2(\theta) = \frac{3}{4\pi}\sqrt{\left(\frac{2}{3}\pi\right)^2 - \left(\theta - \frac{4}{3}\pi\right)^2}$$

and define $G : \mathcal{X} \to \mathbb{R}$ *by*

$$G(\mathbf{x}) = \begin{cases} G_0(\theta) & \text{if } 0 \leq \theta < 2\pi/3, \\ G_2(\theta) & \text{if } 4\pi/3 \leq \theta < 2\pi, \\ 1 - G_0(\theta - 2\pi/3) - G_2(\theta + 2\pi/3) & \text{otherwise,} \end{cases} \tag{24}$$

where $\mathbf{x} = (r\cos\theta, r\sin\theta)^\top$. *With this choice of* G, $0 \leq G(\mathbf{x}) \leq 1/2$ *and (20) holds with* $s = 3$. *Then (23) is a density on* $\mathbb{R}^2$.

Now consider the case $s = 4$. *A direct adaptation of (22) is*

$$G(\mathbf{x}) = \begin{cases} 1/2 & \text{if } (\sqrt{2}/2) < \|\mathbf{x}\|^{-1}\mathbf{x}^\top r\mathbf{u}_0 \leq 1, \\ 1/6 & \text{otherwise.} \end{cases}$$

An alternative choice of G *is* $G(\mathbf{x}) = 2^{-1}\Phi(\alpha x_1 x_2)$, *where* Φ *is the standard normal distribution function,* α *is a real parameter, and* $\mathbf{x} = (x_1, x_2)^\top$. *Taking* f_0 *to be the standard bivariate normal density, (21) gives*

$$f(x,\, y) = 4f_0(x, y)\{2^{-1}\Phi(\alpha xy)\} = 2\, f_0(x, y)\, \Phi(\alpha xy),$$

which is the density of the distribution studied in [1, Section 4.1] and appearing in [6, (15)].

Example 5. *The action of C_4 on $\mathbb{R}^2$ described in Example 4 extends to an action of the dihedral group D_8 of rotations and reflections of the square. This appears in (16) of [6].*

Example 6. *(rotations of $\mathbb{R}^2$) The group, $SO(2)$, of rotations of the plane can be identified with the unit circle, S^1, regarded as the interval $[-\pi, \pi]$ with its ends identified. Use of polar coordinates (r, θ) on $\mathbb{R}^2$ enables the standard action of $SO(2)$ on $\mathbb{R}^2$ to be written as $\phi.(r, \theta) = (r, \theta + \phi)$ for $\phi, \theta \in S^1$. For $\mathcal{X} = \mathbb{R}^2$ and $K = SO(2)$, the above action gives $\mathcal{X}/K = [0, \infty)$. Let $\mathbf{x}$, with polar coordinates (r, θ), be a random point in the plane, having density f with respect to $dr\,(2\pi)^{-1}d\theta$. (The factor $(2\pi)^{-1}$ is used to get a* probability *measure on S^1.) Then (7) can be interpreted as the decomposition of the joint density of (r, θ) on the product space $[0, \infty) \times S^1$ into the marginal density of r and the conditional density of θ given r. If $\mathbf{x}$ has the standard bivariate normal distribution then the marginal density of r is*

$$\tilde{f}_0(r) = r \exp(-r^2/2).$$

The standard bivariate normal distribution can be modulated by using as perturbation factor the conditional densities

$$\tilde{\Gamma}(\theta|r) = r\,|\pi^{-1}\theta|^{r-1}/2, \quad \theta \in [-\pi, \pi).$$

By Proposition 1, the function given in polar coordinates by

$$f(r, \theta) = \tilde{f}_0(r)\tilde{\Gamma}\{(r, \theta)|r\} \tag{25}$$

is a density on the plane, with respect to $dr\,(2\pi)^{-1}d\theta$. In contrast to densities of type (2), where necessarily $0 \leq f(\mathbf{x})/f_0(\mathbf{x}) \leq 2$, densities produced by this construction have the property that the ratio $f(\mathbf{x})/f_0(\mathbf{x})$ is unbounded.

The plots in Figure 1 below display (a) some contours of the bivariate density (25), transformed to rectangular co-ordinates, (b) the marginal density of the first component x_1 of $\mathbf{x}$, obtained by numerical integration of the bivariate density, with the $\mathcal{N}(0, 1)$ density superimposed as a dashed curve, (c) the ratio of this marginal density to the $\mathcal{N}(0, 1)$ density. Plot (c) highlights the fact that the ratio of the densities in plot (b) diverges to $+\infty$ as x_1 moves towards $-\infty$. In other words, the marginal density of the new distribution on $\mathbb{R}^2$ appears to have a heavier left tail than the $\mathcal{N}(0, 1)$ density, a feature which cannot be achieved by perturbation of the normal density via (2).

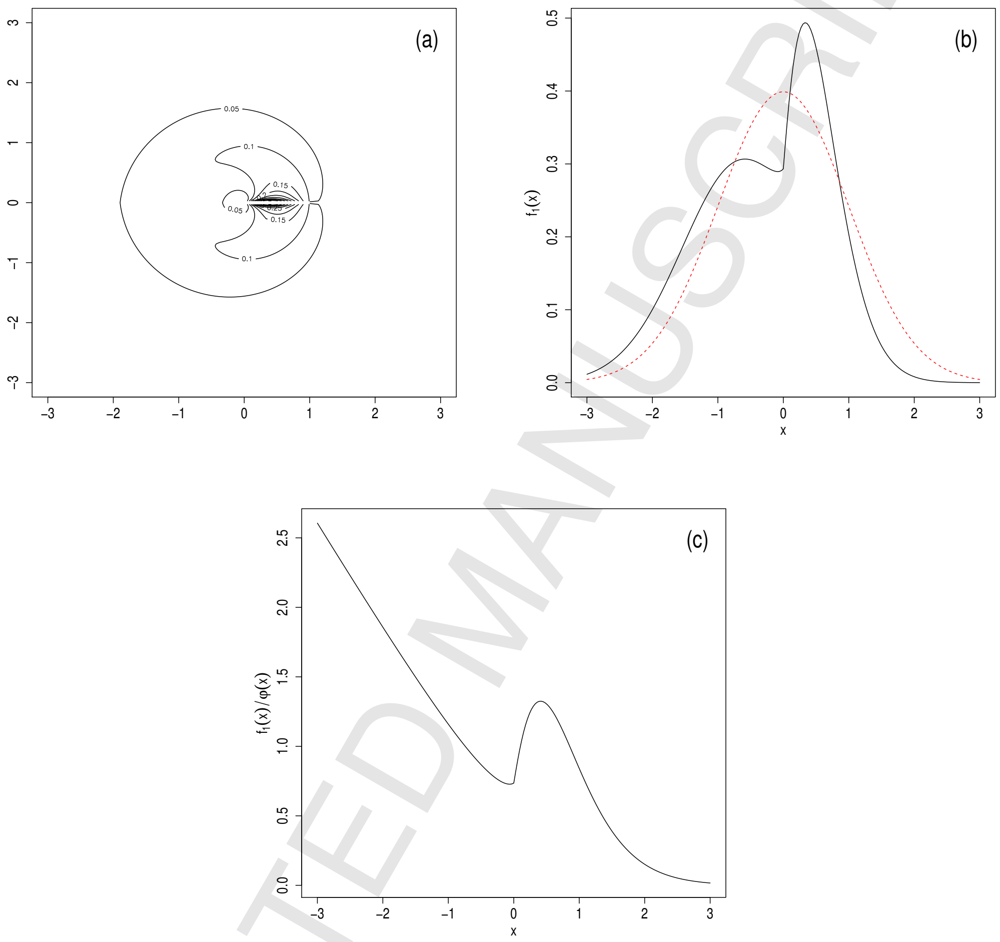


Figure 1: (a) contours of density (25), (b) marginal density of first component, with standard normal density (dashed), (c) ratio of marginal density to standard normal density.

Example 7. *(exchangeability) If $\mathcal{X} = \mathcal{Y}^n$ then the permutation group Σ_n acts on $\mathcal{X}$ by permuting the copies of $\mathcal{Y}$. A distribution is Σ_n-invariant iff it is exchangeable.*

Taking $\mathcal{Y} = \mathbb{R}$ and $n = 3$ gives $\mathcal{X} = \mathbb{R}^3$, on which the permutation group Σ_3 acts by permuting the coordinates. Each such permutation can be written *uniquely as the rotation* $\mathbf{R}^r \tilde{\mathbf{R}}^s$ *of* $\mathbb{R}^3$ *for* $r = 0, 1, 2$ *and* $s = 0, 1$*, where*

$$\mathbf{R} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \qquad \text{and} \qquad \tilde{\mathbf{R}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}.$$

Given any probability density g *on* $\mathbb{R}$*, define* $f_0 : \mathbb{R}^3 \to \mathbb{R}$ *by* $f_0(\mathbf{x}) = \prod_{j=1}^{3} g(x_j)$ *for* $\mathbf{x} = (x_1, x_2, x_3)^\top$*. Then* f_0 *is a* Σ_3*-invariant density on* $\mathbb{R}^3$*. Define the subsets* $D_1, \ldots, D_6$ *of* $\mathcal{X}$ *by*

$$\begin{aligned} D_1 &= \{\mathbf{x} : x_1 \le x_2 \le x_3\}, & D_2 &= \{\mathbf{x} : x_1 \le x_3 \le x_2\}, \\ D_3 &= \{\mathbf{x} : x_2 \le x_3 \le x_1\}, & D_4 &= \{\mathbf{x} : x_2 \le x_1 \le x_3\}, \\ D_5 &= \{\mathbf{x} : x_3 \le x_1 \le x_2\}, & D_6 &= \{\mathbf{x} : x_3 \le x_2 \le x_1\}. \end{aligned}$$

The group Σ_3 *permutes* $D_1, \ldots, D_6$ *and each* D_j *can be obtained from* D_1 *by one of these permutations. If sets of measure 0 are ignored then* $\{D_1, \ldots, D_6\}$ *is a partition of* $\mathcal{X}$*. Thus, by symmetry under* Σ_3*, each* D_j *has probability* $1/6$ *under the distribution with density* f_0*. Given any non-negative numbers* $p_1, \ldots, p_6$ *such that* $\sum_{j=1}^{6} p_j = 1$*, define* G *by*

$$G(\mathbf{x}) = p_j \qquad \text{if } \mathbf{x} \in D_j.$$

Then

$$\sum_{r=0}^{2} \sum_{s=0}^{1} G(\mathbf{R}^r \tilde{\mathbf{R}}^s \mathbf{x}) = 1,$$

and so f*, defined by*

$$f(\mathbf{x}) = 6\, f_0(\mathbf{x})\, G(\mathbf{x}), \qquad \mathbf{x} \in \mathcal{X},$$

is a probability density.

We now introduce a variant construction in which G *is smoother than the above step function. To simplify the problem, we merge the sets* D_j *in pairs by defining*

$$E_j = D_{2j-1} \cup D_{2j}, \qquad j = 1, 2, 3.$$

Then $E_2 = \mathbf{R}E_1$ *and* $E_3 = \mathbf{R}^2 E_1$*. The function* G*, defined by*

$$G(\mathbf{x}) = \begin{cases} x_1/(x_1 + x_2 + x_3) & \text{if } \mathbf{x} \in E_1, \\ x_2/(x_1 + x_2 + x_3) & \text{if } \mathbf{x} \in E_2, \\ 1 - 2\, x_3/(x_1 + x_2 + x_3) & \text{if } \mathbf{x} \in E_3, \end{cases}$$

is continuous on $\mathbb{R}^3 \setminus \{\mathbf{0}\}$ *and the function* f *with the same formal expression as (23) is a density on* $\mathcal{X}$*.*

Some modulations of products of independent identically distributed Poisson random variables on $\mathbb{N}^2$ *are used in [11] to model scores in sporting tournaments.*

Example 8. *(reflection on the circle) The choice of a reference direction,* $\theta = 0$*, on the circle,* S^1*, makes it possible to define an action of* C_2 *on* S^1 *that maps the angle* θ *to* $-\theta$*. This is equivalent to reflection of* S^1 *in the line through* $\theta = 0$ *and* $\theta = \pi$*.*

One standard distribution on S^1 *that is symmetric about* 0 *is the wrapped Cauchy distribution with mean direction* 0 *and mean resultant length* ρ*. Its density (with respect to* $d\theta$*),* f_0*, is given by*

$$f_0(\theta) = \frac{1}{2\pi} \frac{1-\rho^2}{1-2\rho\cos\theta+\rho^2}.$$

A skewed version of f_0*, considered in [23, Section 2], is* f*, given by*

$$f(\theta) = f_0(\theta)\,(1+\sin\theta). \tag{26}$$

The generalisation of (26) introduced in [23] takes f_0 *to be any probability density on* S^1 *that is symmetric about* 0 *and replaces the factor* $1+\sin\theta$ *by* $\Gamma(\theta)$*, where*

$$\Gamma(x) = 2\,G_0\{w(x)\} \tag{27}$$

with $w : S^1 \to S^1$ C_2*-equivariant, i.e.,* $w(-x) = -w(x)$*, and*

$$G_0(x) = \int_{-\pi}^{x} g(u)\,du$$

for some density g *on* S^1 *which is symmetric about 0. If* f *is defined by* $f(\theta) = f_0(\theta)\Gamma(\theta)$*, as in (8), then* f *is a probability density on* S^1*.*

The bijection in Proposition 1 relating densities f on $\mathcal{X}$ to pairs of the form (f_0, Γ) can be expressed in terms of random variables as follows.

Proposition 3. *Let* f_0 *be a* K*-symmetric density on* $\mathcal{X}$ *and* Γ *be a non-negative function on* $\{x \in \mathcal{X} : f_0(x) > 0\}$ *satisfying (9). Let* Z_0 *and* S *be random variables on* $\mathcal{X}$ *and* K*, respectively, such that* Z_0 *has density* f_0 *and the conditional density of* S *given* Z_0 *is* $h(s|x) = \Gamma(s.x)$*. Put* $Z = S.Z_0$*. Then the density of* Z *is*

$$f(z) = f_0(z)\Gamma(z).$$

PROOF. The density of Z at z is

$$\begin{aligned} f(z) &= \int_K f_0(k.z)h(k|k^{-1}.z)\,d\nu(k) \\ &= \int_K f_0(z)\Gamma\{k.(k^{-1}.z)\}\,d\nu(k) \\ &= f_0(z)\Gamma(z). \end{aligned}$$

□

This result is an extension of [5, Proposition 2] for distributions of type (1) and of (4) for distributions of type (2). There is a similar result based on $[X]$ instead of X, paralleling [5, Proposition 3] and its extension in [7, Proposition 1.6]. In the case $K = C_2$, the representation $Z = S.Z_0$ is known as the 'stochastic representation' of Z.

Remark 4. *In contrast to earlier developments, this paper does not derive modulation invariance as a corollary of the stochastic representation. However, this connection still exists and it is possible to derive Proposition 2 from Proposition 3.* □

If the group K is finite then the stochastic representation given in Proposition 3 can be considered as a selection mechanism which is applied to values of Z_0 in order to obtain values of Z. The construction is described in Corollary 1 below and is a direct extension of that for $s = 2$ appearing at the end of Section 2.1 in [11], which in turn is related to expression (9) of [10]; all these formulations are generalisations of representation (4) for the case $K = C_2$.

Corollary 1. *Suppose that the group K is finite and choose an enumeration $k_1, \ldots, k_s$ of its elements. Let f_0 be a K-symmetric density on $\mathcal{X}$ and G be a non-negative function on $\mathcal{X}$ such that $\sum_{i=1}^s G(k_i.x) = 1$. Define $w_0(x), \ldots, w_s(x)$ by $w_0(x) = 0$ and $w_i(x) = \sum_{j=1}^i G(k_j.x)$ for $i = 1, \ldots, s$. Let Z_0 and Y be independent random variables on $\mathcal{X}$ and $[0, 1]$, respectively, such that Z_0 has density f_0 and Y is uniformly distributed. Define the random variable Z on $\mathcal{X}$ by*

$$Z = k_i.Z_0 \quad \textit{if} \;\; w_{i-1}(Z_0) \le Y < w_i(Z_0).$$

Then the density of Z at z is $f(z) = sf_0(z)G(z)$.

PROOF. Define Γ by $\Gamma(x) = sG(x)$ and define the random variable S on K by

$$S = k_i \quad \text{if} \quad w_{i-1}(Z_0) \leq Y < w_i(Z_0).$$

Then Γ satisfes (9) and the result follows from Proposition 3. $\square$

3. Parametric models for the conditional density Γ

A convenient class of conditional densities Γ in (8) consists of those of the form

$$\Gamma(x; \boldsymbol{\alpha}) = c([x], \boldsymbol{\alpha})^{-1} q\{\boldsymbol{\alpha}^\top \mathbf{h}(x)\}, \tag{28}$$

where $\mathbf{h} : \mathcal{X} \to \mathbb{R}^d$, $\boldsymbol{\alpha}$ runs through a subset of $\mathbb{R}^d$, $q : \mathbb{R} \to \mathbb{R}$, and

$$c([x], \boldsymbol{\alpha}) = \int_K q\{\boldsymbol{\alpha}^\top \mathbf{h}(k.x)\}\, d\nu(k).$$

Special cases of interest include the following.

(a) If $\mathcal{X} = \mathbb{R}^d$, $K = C_2$ acts by central symmetry, $\mathbf{h}(\mathbf{x}) = \mathbf{x}$ and $q(t) = \Phi(t)$ then $c([\mathbf{x}], \boldsymbol{\alpha}) = 1/2$ and the functions $\Gamma(\cdot; \boldsymbol{\alpha})$ are those arising in the multivariate skew-normal distributions [9].

(b) If $q(t) = \exp(t)$ then, for each $[x]$, the family of conditional densities $\tilde{\Gamma}(\cdot|[x]; \boldsymbol{\alpha})$ is an exponential family with canonical statistic $\mathbf{h}$. If also $\mathcal{X} = \mathbb{R}^d$, $K = C_2$ acts by central symmetry, and $\mathbf{h}(\mathbf{x}) = \mathbf{x}$ then $c([\mathbf{x}], \boldsymbol{\alpha}) = \cosh(\boldsymbol{\alpha}^\top \mathbf{x})$ and so

$$f(\mathbf{x}) = 2 f_0(\mathbf{x}) L_0(2\boldsymbol{\alpha}^\top \mathbf{x}),$$

where L_0 is the standard logistic distribution function. In the case $d = 1$ this is of the form (1) with $\lambda = 2\alpha$ and $G_0 = L_0$.

(c) If $q(t) = 1 + t$ and $\mathbf{h}$ is bounded and satisfies

$$\int_K \mathbf{h}(k.x)\, d\nu(k) = 0$$

then $c([x], \boldsymbol{\alpha}) = 1$. If $\boldsymbol{\alpha}$ is near enough to $\mathbf{0}$ then $\Gamma(x; \boldsymbol{\alpha})$ is non-negative. For $\mathcal{X} = S^1$, $K = C_2$ acting by reflection in the diameter through 0 and π, $d = 1$ and $h(x) = \sin x$, the functions $\Gamma(\cdot; \alpha)$ are those arising in the asymmetrical circular distributions of [23].

Suppose given a parametric model $f_0(\cdot;\theta)$ of symmetric densities, where θ runs through Θ. Modulating $f_0(\cdot;\theta)$ by the conditional densities (28) produces the parametric model with densities

$$f(x;\theta,\boldsymbol{\alpha}) = f_0(x;\theta)c([x],\boldsymbol{\alpha})^{-1}q\{\boldsymbol{\alpha}^\top \mathbf{h}(x)\}. \tag{29}$$

In (29), θ parameterises the symmetric part of the distribution, whereas $\boldsymbol{\alpha}$ parameterises the departure from symmetry. In the language of cuts [13, p. 38], $[x]$ is a cut, being S-sufficient for θ and S-ancillary for $\boldsymbol{\alpha}$. Thus inference can be carried out separately on the parameters θ and $\boldsymbol{\alpha}$. In particular, θ and $\boldsymbol{\alpha}$ are orthogonal. Examples of densities (29) with $\mathcal{X}$ a shape space and h and q satisfying (b) or (c) above are given in [16, Sections 10.3.1, 10.3.2].

4. Transformations and symmetry

One of the main techniques for generating new families of distributions from old ones is transformation of the sample space. This section investigates the effect that such transformations have on symmetry of distributions and the decomposition (8) of general probability density functions into their symmetric and modulating parts.

Given a probability measure P on $\mathcal{X}$ and a measurable function $t : \mathcal{X} \to \mathcal{Y}$, we define t_*P by

$$t_*P(A) = P\{t^{-1}(A)\} \tag{30}$$

for all measurable subsets A of $\mathcal{Y}$. (The definition (6) of $\pi_*\mu$ is a special case of (30).) Let K act on $\mathcal{X}$ and $\mathcal{Y}$. The transformation t is *K-equivariant* if

$$t(k.x) = k.t(x), \qquad k \in K. \tag{31}$$

If $\mathcal{X} = \mathcal{Y} = \mathbb{R}^d$ and C_2 acts by central symmetry then t is C_2-equivariant if and only if it is an odd function.

Proposition 4. *Let $t : \mathcal{X} \to \mathcal{Y}$ be a measurable function and suppose that K acts on $\mathcal{X}$ and $\mathcal{Y}$.*

(i) If t is K-equivariant then it induces a transformation $\tilde{t} : \mathcal{X}/K \to \mathcal{Y}/K$, given by $\tilde{t}([x]) = [t(x)]$;
(ii) if t is K-equivariant and P is K-invariant then t_*P is K-invariant;
(iii) if Y has probability distribution t_*P then $k.Y$ has probability distribution $(k * t)_*P$, where $k * t : \mathcal{X} \to \mathcal{Y}$ is defined by $(k * t)(x) = k.t(x)$;

(iv) *let μ and λ be K-invariant measures on $\mathcal{X}$ and $\mathcal{Y}$, respectively, such that $t_*\mu$ is absolutely continuous with respect to λ. Let X be a random variable on $\mathcal{X}$ having density (with respect to μ) f expressed as*

$$f(x) = f_0(x)\Gamma(x), \tag{32}$$

as in (8). Suppose that t is equivariant and one-to-one. Then the density, t_*f, of $t(X)$ with respect to λ can be expressed analogously to (32) as

$$(t_*f)(y) = f_0\{t^{-1}(y)\}\,\frac{dt_*\mu}{d\lambda}(y)\,\Gamma\{t^{-1}(y)\} \tag{33}$$

with $f_0\{t^{-1}(y)\}\,(dt_*\mu/d\lambda)(y)$ as the K-symmetric part and $\Gamma\{t^{-1}(y)\}$ as the modulating factor. If $\mathcal{X} = \mathcal{Y} = \mathbb{R}^d$ and $\mu = \lambda$ is Lebesgue measure then $dt_*\mu/d\mu(y) = |t'\{t^{-1}(y)\}|^{-1}$ (the inverse Jacobian determinant), whereas if $\mathcal{X}$ is discrete and $\mu = \lambda$ is counting measure then $dt_*\mu/d\lambda(y) = 1$.

Proof. These follow from straightforward calculations. □

Note:

In the case in which $\mathcal{X} = \mathcal{Y} = \mathbb{R}^d$, $K = C_2$ acts by central symmetry, and μ and λ are Lebesgue measure, Proposition 4 yields Theorem 2.2 of [19].

Example 9. *(inversion in the unit circle)*

Let $\mathcal{X}$ be the punctured plane, $\mathbb{R}^2 \setminus \{\mathbf{0}\}$. The restriction to $\mathcal{X}$ of the action of the cyclic group C_3 by rotation of $\mathbb{R}^2$ (as in Example 4 with $s = 3$) gives an action of C_3 on $\mathcal{X}$. The geometrical operation of inversion of the punctured plane in the unit circle can be expressed algebraically as the transformation $t : \mathcal{X} \to \mathcal{X}$ given by $\mathbf{x} \mapsto \|\mathbf{x}\|^{-2}\mathbf{x}$. Let $\mathbf{X}$ be a random variable on $\mathcal{X}$ having density (with respect to Lebesgue measure) f expressed as

$$f(\mathbf{x}) = f_0(\mathbf{x})\Gamma(\mathbf{x}),$$

where f_0 is the standard bivariate normal density (restricted to $\mathcal{X}$) and $\Gamma(\mathbf{x}) = 3G(\mathbf{x})$ with given G given by restriction of (22). Put $\mathbf{Y} = t(\mathbf{X})$. It follows from (33) (and the fact that here $\Gamma(\|\mathbf{y}\|^{-2}\mathbf{y}) = \Gamma(\mathbf{y})$) that the density, $\check{f}$, of $\mathbf{Y}$ (with respect to Lebesgue measure) is given by

$$\check{f}(\mathbf{y}) = f_0(\|\mathbf{y}\|^{-2}\mathbf{y})\|\mathbf{y}\|^{-2}\Gamma(\mathbf{y}),$$

the symmetric part being $f_0(\|\mathbf{y}\|^{-2}\mathbf{y}))\|\mathbf{y}\|^{-2}$ and the modulating factor being $\Gamma(\mathbf{y})$.

Example 10. *(Multivariate skew Birnbaum–Saunders distributions)*

Let $\mathcal{X} = \mathbb{R}^d$, $\mathcal{Y} = (\mathbb{R}^+)^d$ *and* $K = C_2$ *Then* K *acts on* $\mathcal{X}$ *by central symmetry, i.e.,* $(-1).\mathbf{x} = -\mathbf{x}$ *for* $\mathbf{x}$ *in* $\mathcal{X}$, *as in Example 2, and on* $\mathcal{Y}$ *by* $(-1).(y_1, \dots, y_d) = (1/y_1, \dots, 1/y_d)$ *for* $(y_1, \dots, y_d)$ *in* $\mathcal{Y}$. *For* $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_d)$ *with* $\alpha_i, \beta > 0$, *the multivariate Birnbaum–Saunders transformation* $t_{\boldsymbol{\alpha},\beta} : \mathbb{R}^d \to (\mathbb{R}^+)^d$ *is defined by* $t_{\boldsymbol{\alpha},\beta}(x_1, \dots, x_d) = (t_1(x_1), \dots, t_d(x_d))$, *where*

$$t_i(x_i) = \beta \left\{ \frac{\alpha_i}{2} x_i + \sqrt{\left(\frac{\alpha_i}{2} x_i\right)^2 + 1} \right\}^2 .$$

The transformation $t_{\boldsymbol{\alpha},1}$ *is* K*-equivariant for the above actions of* K *on* $\mathcal{X}$ *and* $\mathcal{Y}$.

For $\boldsymbol{\lambda}$ *in* $\mathbb{R}^d$, *the corresponding standard skew d-variate normal distribution [9] is obtained from the standard d-variate normal distribution by modulating it by the function* $\mathbf{x} \mapsto 2\Phi(\boldsymbol{\lambda}^\top \mathbf{x})$, *where* Φ *denotes the standard normal cumulative distribution function. The resulting density on* $\mathbb{R}^d$ *is* $2\phi_d(\mathbf{x})\, \Phi(\boldsymbol{\lambda}^\top \mathbf{x})$, *where* ϕ_d *denotes the density of the standard d-variate normal distribution. Transformation by* $t_{\boldsymbol{\alpha},\beta}$ *sends the skew d-variate normal distribution to the generalised Birnbaum–Saunders distribution of [17], which has density*

$$\phi_d(\mathbf{x}) \left\{ \prod_1^d \frac{y_i^{-3/2}(y_i + 1)}{2\alpha_i} \right\} 2\Phi(\boldsymbol{\lambda}^\top \mathbf{x})$$

at $\mathbf{y}$, *where* $\mathbf{x} = (x_1, \dots, x_d)$ *with* $x_i = [\sqrt{y_i} - 1/\sqrt{y_i}]/\alpha_i$.

More generally, multivariate Birnbaum–Saunders transformations can be used to generate a new class of multivariate distributions as follows. Let f_0 *be any centrally symmetric density with respect to Lebesgue measure on* $\mathbb{R}^d$ *and let* Γ *be any non-negative function on* $\mathbb{R}^d$ *satisfying* $\Gamma(\mathbf{x}) + \Gamma(-\mathbf{x}) = 2$. *Let* $\mathbf{Z}$ *be a random variable on* $\mathbb{R}^d$ *having density* $f(\mathbf{x}) = f_0(\mathbf{x})\Gamma(\mathbf{x})$. *Then the density of* $t_{\boldsymbol{\alpha},1}(\mathbf{Z})$ *at* $\mathbf{y}$ *is*

$$f_0(\mathbf{x}) \left\{ \prod_1^d \frac{y_i^{-3/2}(y_i + 1)}{2\alpha_i} \right\} \Gamma(\mathbf{x}),$$

where $\mathbf{x} = (x_1, \dots, x_d)$ *with* $x_i = [\sqrt{y_i} - 1/\sqrt{y_i}]/\alpha_i$. *It is worth underlining that in this case, although all other conclusions of Proposition 4 hold, the factor*

$$f_0\left(\mathbf{x}\right) \left\{ \prod_{i=1}^d \frac{y_i^{-3/2}(y_i + 1)}{2\alpha_i} \right\}$$

fails to be K-symmetric, because Lebesgue measure is not K-invariant under the action of K on $\mathcal{Y}$.

5. Symmetry in directional statistics

This section illustrates the above general theory with some examples in the areas of directional statistics and shape statistics.

Example 11. *(central symmetry on spheres) For a point $\mathbf{n}$ in the $(p-1)$-sphere S^{p-1}, central symmetry about $\mathbf{n}$ is invariance under the C_2-action given by restriction of the linear transformation $2\mathbf{n}\mathbf{n}^\top - \mathbf{I}_p$ of $\mathbb{R}^p$. This transformation reverses each great circle through $\mathbf{n}$. (Technical note: this can be generalised to symmetric spaces with maximal tori of dimension 1.) Here $\mathcal{X} = S^{p-1}$ and $K = C_2$. Distributions on S^{p-1} with central symmetry about $\mathbf{n}$ include the Watson distributions [22, Section 9.4.2] with log density proportional to $(\mathbf{x}^\top\mathbf{n})^2$. Suitable densities Γ of conditional distributions for use in (8) can be obtained by applying the functions (27) of [23] to each great circle through $\mathbf{n}$. To do this, we exploit the tangent-normal decomposition*

$$\mathbf{x} = (t, (1-t^2)^{1/2}\mathbf{z}), \qquad t \in (-1, 1), \mathbf{z} \in S^{p-2} \tag{34}$$

(see [22, (9.1.2)]) and define the colatitude θ by $\cos\theta = t$. Then

$$\Gamma(\mathbf{x}) = \Gamma_{\pm\mathbf{z}}\{w_{\pm\mathbf{z}}(\theta)\},$$

where, for each $\{-\mathbf{z}, \mathbf{z}\}$, $w_{\pm\mathbf{z}} : S^1 \to S^1$ is $\mathbb{Z}_2$-equivariant and

$$\Gamma_{\pm\mathbf{z}}(x) = \int_{-\pi}^{x} g_{\pm\mathbf{z}}(u)\, du$$

for some density $g_{\pm\mathbf{z}}$ on S^1 which is symmetric about 0.

Example 12. *(rotational symmetry about an axis on S^{p-1}) For an axis $\pm\mathbf{n}$ of S^{p-1}, consider the $SO(p-1)$-action of rotations about $\pm\mathbf{n}$. Here $\mathcal{X} = S^{p-1}$, $K = SO(p-1)$ and t in the tangent-normal decomposition (34) is invariant. Distributions on S^{p-1} with rotational symmetry about $\pm\mathbf{n}$ include the von Mises–Fisher distributions [22, Section 9.3.2] with log density proportional to $\mathbf{x}^\top\mathbf{n}$.*

Choose (a) a family of density functions $g(\cdot;\lambda)$ on S^{p-2} indexed by a parameter λ such that $g(\cdot; 0)$ is constant, (b) a function $t \mapsto \mathbf{U}_t$ from $[-1, 1]$

to the rotation group $SO(p-1)$ with $\mathbf{U}_{-1} = \mathbf{U}_1 = \mathbf{I}_{p-1}$. Then functions Γ of the form

$$\Gamma(\mathbf{x}) = g(\mathbf{U}_t\mathbf{z}; \lambda),$$

where $\mathbf{x} = t\mathbf{n} + (1-t^2)^{1/2}\mathbf{z}$ is the tangent-normal decomposition of $\mathbf{z}$, are non-negative and satisfy (9), and so can be used as conditional densities in (7). Some other functions Γ that are suitable in this context are those introduced by Ley & Verdebout [20]. These have the form

$$\Gamma(\mathbf{x}) = 2\Pi\{m(t)\boldsymbol{\delta}^\top(\mathbf{I}_p - \mathbf{n}\mathbf{n}^\top)\mathbf{x}\},$$

where $m : [0,1] \to \mathbb{R}$ is continuous a.e., $\Pi : \mathbb{R} \to [0,1]$ satisfies $\Pi(y) + \Pi(1-y) = 1$ for all y in $\mathbb{R}$ and $\boldsymbol{\delta} \in \mathbb{R}^p$ with $\boldsymbol{\delta}^\top\mathbf{n} = 0$.

Example 13. *(Stiefel manifolds and Grassmann manifolds) The* Stiefel manifold *$V_r(\mathbb{R}^p)$ of orthonormal r-frames in $\mathbb{R}^p$ can be written as*

$$V_r(\mathbb{R}^p) = \left\{(\mathbf{u}_1, \ldots, \mathbf{u}_r) : \mathbf{u}_i^\top\mathbf{u}_j = \delta_{ij}\right\},$$

where δ_{ij} denotes the Kronecker δ. The Grassmann manifold *$G_r(\mathbb{R}^p)$ is the space of r-dimensional subspaces of $\mathbb{R}^p$. The rotation group $SO(r)$ acts on $V_r(\mathbb{R}^p)$ by $(\mathbf{V}, (\mathbf{u}_1, \ldots, \mathbf{u}_r)) \mapsto (\mathbf{V}\mathbf{u}_1, \ldots, \mathbf{V}\mathbf{u}_r)$. Then $\mathcal{X} = V_r(\mathbb{R}^p)$, $K = SO(p-2)$ and $\mathcal{X}/K = G_r(\mathbb{R}^p)$. Distributions on $V_r(\mathbb{R}^p)$ with $SO(r)$-symmetry include the matrix Bingham distributions [22, Section 13.3.3] with log density proportional to* $\mathrm{trace}(\mathbf{X}^\top\mathbf{B}\mathbf{X})$, *where $\mathbf{X} = (\mathbf{u}_1, \ldots, \mathbf{u}_r)$ and $\mathbf{B}$ is a symmetric $p \times p$ matrix. Functions $\Gamma : V_r(\mathbb{R}^p) \to \mathbb{R}$ of the form*

$$\Gamma(\mathbf{U}) = 1 + \mathrm{trace}(\mathbf{A}^\top\mathbf{U}),$$

where $\mathrm{trace}\{(\mathbf{A}^\top\mathbf{A})^{1/2}\} < 1$, are suitable for use as conditional densities in (7). These functions are of the form (28) with q as in special case (c) in Section 3.

Example 14. *(orthogonal axial frames) The group ${C_2}^r = \{(\varepsilon_1, \ldots, \varepsilon_r) | \varepsilon_j = \pm 1\}$ acts on $V_r(\mathbb{R}^p)$ by changing the signs of the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_r$, i.e., $(\varepsilon_1, \ldots, \varepsilon_r)$ acts by $(\mathbf{u}_1, \ldots, \mathbf{u}_r) \mapsto (\varepsilon_1\mathbf{u}_1, \ldots, \varepsilon_r\mathbf{u}_r)$. Then $\mathcal{X} = V_r(\mathbb{R}^p)$, $K = {C_2}^r$ and $\mathcal{X}/K = V_r(\mathbb{R}^p)/{C_2}^r$ is the space of orthogonal axial frames considered in [2]. One of the simplest families of distributions on $V_r(\mathbb{R}^p)$ consists of the matrix Fisher distributions. These have densities*

$$f(\mathbf{U}; \mathbf{A}) = \left\{{}_0F_1\left(p/2; 1/4\,\mathbf{A}^\top\mathbf{A}\right)\right\}^{-1}\exp\{\mathrm{trace}(\mathbf{A}^\top\mathbf{U})\},$$

where $\mathbf{U} \in V_r(\mathbb{R}^p)$ *and* $\mathbf{A}$ *is a* $p \times r$ *matrix. (See, e.g., [22, Section 13.2.3].) Then (8) becomes*

$$f(\mathbf{U}; \mathbf{A}) = f_0([\mathbf{U}]; [\mathbf{A}])\Gamma(\mathbf{U}),$$

where

$$\begin{aligned} f_0([\mathbf{U}]; [\mathbf{A}]) &= \left\{ {}_0F_1\left(p/2; 1/4\, \mathbf{A}^\top \mathbf{A}\right) \right\}^{-1} \prod_{j=1}^{r} \cosh(\mathbf{a}_j^\top \mathbf{u}_j) \qquad (35) \\ \Gamma(\mathbf{U}) &= \prod_{j=1}^{r} \frac{\exp(\mathbf{a}_j^\top \mathbf{u}_j)}{\cosh(\mathbf{a}_j^\top \mathbf{u}_j)}, \end{aligned}$$

with $\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_r)$ *and* $\mathbf{A} = (\mathbf{a}_1, \ldots, \mathbf{a}_r)$. *The densities (35) are those of the symmetrised matrix Fisher distributions, which were introduced in the case* $r = p = 3$ *in [24].*

Example 15. *(shape versus reflection shape) Two configurations of* k *non-identical (labelled) landmarks in* $\mathbb{R}^m$ *have the same (similarity) shape if one can be transformed into the other by translation, rotation and change of scale; they have the same reflection shape if the transformation may include reflections also. The sets of (similarity) shapes and of reflection shapes are the* shape space Σ_m^k *and the* reflection shape space $R\Sigma_m^k$, *respectively. The group* C_2 *acts on* Σ_m^k *by reflections in* $\mathbb{R}^m$ *and* $\Sigma_m^k/C_2 = R\Sigma_m^k$. *The space* Σ_m^k *is the quotient space* $\mathcal{S}_m^k/SO(m)$, *where* $\mathcal{S}_m^k$ *is the space of* $m \times (k-1)$ *real matrices* $\mathbf{Z}$ *with* $\mathrm{trace}(\mathbf{Z}\mathbf{Z}^\top) = 1$ *and the rotation group* $SO(m)$ *of* R^m *acts on* $\mathcal{S}_m^k$ *by left multiplication. For* $\mathbf{Z}$ *in* $\mathcal{S}_m^k$, $[\mathbf{Z}]$ *denotes the corresponding shape. Various distributions on* Σ_m^k *are considered in [16], including those with densities of the form*

$$f([\mathbf{Z}]; \mathbf{A}, \kappa, [\mathbf{M}]) = c(\mathbf{A}) \exp\{\mathrm{trace}(\mathbf{A}\mathbf{Z}^\top \mathbf{Z})\} \left\{1 + \kappa |\mathbf{M}\mathbf{Z}^\top|\right\}, \qquad (36)$$

where $\mathbf{Z}$ *and* $\mathbf{M}$ *are in* $\mathcal{S}_m^k$, $\mathbf{A}$ *is a symmetric real* $(k-1) \times (k-1)$ *matrix,* $c(\mathbf{A}) = {}_1F_1\left\{1/2; m(k-1)/2; \mathbf{A} \otimes \mathbf{I}_m\right\}^{-1}$, $0 \le \kappa m^{-m/2} \left|\mathbf{M}\mathbf{M}^\top\right|^{1/2} \le 1$ *and* $|\cdot|$ *denotes a determinant. The density (36) can be written in the form of (8) as*

$$f([\mathbf{Z}]; \mathbf{A}, \kappa, [\mathbf{M}]) = f_0([\mathbf{Z}]; \mathbf{A})\Gamma([\mathbf{Z}]; \kappa, [\mathbf{M}]),$$

where

$$\begin{aligned} f_0([\mathbf{Z}]; \mathbf{A}) &= c(\mathbf{A}) \exp\{\mathrm{trace}(\mathbf{A}\mathbf{Z}^\top \mathbf{Z})\} \qquad (37) \\ \Gamma([\mathbf{Z}]; \kappa, [\mathbf{M}]) &= 1 + \kappa |\mathbf{M}\mathbf{Z}^\top|. \end{aligned}$$

The densities (37) are those of the shape Bingham distributions introduced in [15]. The functions Γ *are of the form (28) with* q *as in special case (c) in Section 3.*

6. Discussion

The broad aim of the present contribution is to embed the 'standard' modulation of symmetry, represented by density (2) with corresponding stochastic representation (4), into a wider framework in which the mechanism of sign reversal, together with the identity, underlying the standard formulation is replaced by a more general group of transformations. One benefit of this construction is the inclusion of distributions not generated in the standard formulation recalled above, hence providing a more comprehensive view of a broad set of constructions. Another benefit is the generation of new distributions. For instance, some cases discussed in Examples 4 and 7 are of this form, as is evident from the normalising factor which differs from the ubiquitous value 2 in the density of the standard formulation. Another case in point is Example 6, the aim of which is to show that there exist selection mechanisms capable of producing distributions with tails thicker than those of the original distribution prior to selection. Examples 9 and 10 also appear to be new, at least at this level of generality.

As already stated, the aim of the above-mentioned examples is to illustrate some novel features of the present general formulation, not to put forward models for practical work. Progress in this sense would, of course, be valuable. Examples 7 and 10 take a step in this direction but we hope that future work can lead to further practical constructions.

References

[1] B.C. Arnold, E. Castillo, J.M. Sarabia, Conditionally specified multivariate skewed distributions. Sankhyā A, 64 (2002) 206–266.

[2] R. Arnold, P.E. Jupp, Statistics of orthogonal axial frames, Biometrika 100 (2013) 571–586.

[3] R.B. Arellano-Valle, M.D. Branco, M.G. Genton, A unified view on skewed distributions arising from selections, Canad. J. Statist. 34 (2006) 581-601.

[4] A. Azzalini, A class of distributions which includes the normal ones, Scand. J. Statist. 12 (1985) 171–178.

[5] A. Azzalini, Further results on a class of distributions which includes the normal ones, Statistica 46 (1986) 199-208.

[6] A. Azzalini, Selection models under generalized symmetry settings, Ann. Inst. Stat. Math. 64 (2012) 737–750.

[7] A. Azzalini with the collaboration of A. Capitanio, The Skew-Normal and Related Families, Cambridge University Press, Cambridge, 2014.

[8] A. Azzalini, A. Capitanio, Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution, J. R. Statist. Soc. B 65 (2003) 367-389. Full version of the paper at `arXiv.org:0911.2342`.

[9] A. Azzalini, A. Dalla Valle, The multivariate skew-normal distribution, Biometrika 83 (1996) 571–586.

[10] A. Azzalini, G. Regoli, Some properties of skew-symmetric distributions, Ann. Inst. Stat. Math. 64 (2012) 857-879.

[11] A. Azzalini, G. Regoli, Modulation of symmetry for discrete variables and some extensions, Stat 3 (2014) 56–67.

[12] O.E. Barndorff-Nielsen, P. Blaesild, P.S. Eriksen, Decomposition and Invariance of Measures, and Statistical Transformation Models, Lecture Notes in Statistics 58, Springer-Verlag, New York, 1989.

[13] O.E. Barndorff-Nielsen, D.R. Cox, Inference and Asymptotics, Chapman & Hall, London, 1994.

[14] J.T. Chang, D. Pollard, Conditioning as disintegration, Statistica Neerlandica 51 (1997) 287–317.

[15] Y. Chikuse, P.E. Jupp, A test of uniformity on shape spaces, J. Mult. Anal. 88 (2004) 163–176.

[16] Y. Chikuse, P.E. Jupp, Some families of distributions on higher shape spaces, in *Geometry Driven Statistics: a Festschrift for Kanti Mardia* (I.L. Dryden and J.T. Kent, eds.) 206–217. Wiley, Chichester, 2015.

[17] A. Jamalizadeh, D. Kundu, Multivariate Birnbaum-Saunders distribution based on multivariate skew normal distribution, J. Japan Statist. Soc. 45 (2015) 1–20.

[18] P.E. Jupp, B.D. Spurr, Sobolev tests for symmetry of directional data, Ann. Statist. 11 (1983) 1225–1231.

[19] C. Ley, D. Paindaveine, Multivariate skewing mechanisms: A unified perspective based on the transformation approach, Statist. Probab. Lett. 80 (2010) 1658-1694.

[20] C. Ley, T. Verdebout, Skew-rotsymmetric distributions on unit spheres and related efficient inferential procedures, ECARES working paper 2014-46, ECARES, Université Libre de Bruxelles, 2014.

[21] Y. Ma, M.G. Genton, Flexible class of skew-symmetric distributions, Scand. J. Statist. 31 (2004) 459–468.

[22] K.V. Mardia, P.E. Jupp, Directional Statistics, Wiley, Chichester, 2000.

[23] D. Umbach, S.R. Jammalamadaka, Building asymmetry into circular distributions, Statist. Probab. Lett. 79 (2009) 659–663.

[24] D. Walsh, R. Arnold, J. Townend, Bayesian approach to determining and parameterising earthquake focal mechanisms, Geophys. J. Int. 176 (2009) 235–255.

[25] J. Wang, J. Boyer, M.G. Genton, A skew-symmetric representation of multivariate distributions, Statistica Sinica 14 (2004) 1259–1270.