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

https://doi.org/10.1017/CBO9781139248891.002 Published online by Cambridge University Press

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

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

https://doi.org/10.1017/CBO9781139248891.002 Published online by Cambridge University Press

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

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

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,

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

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

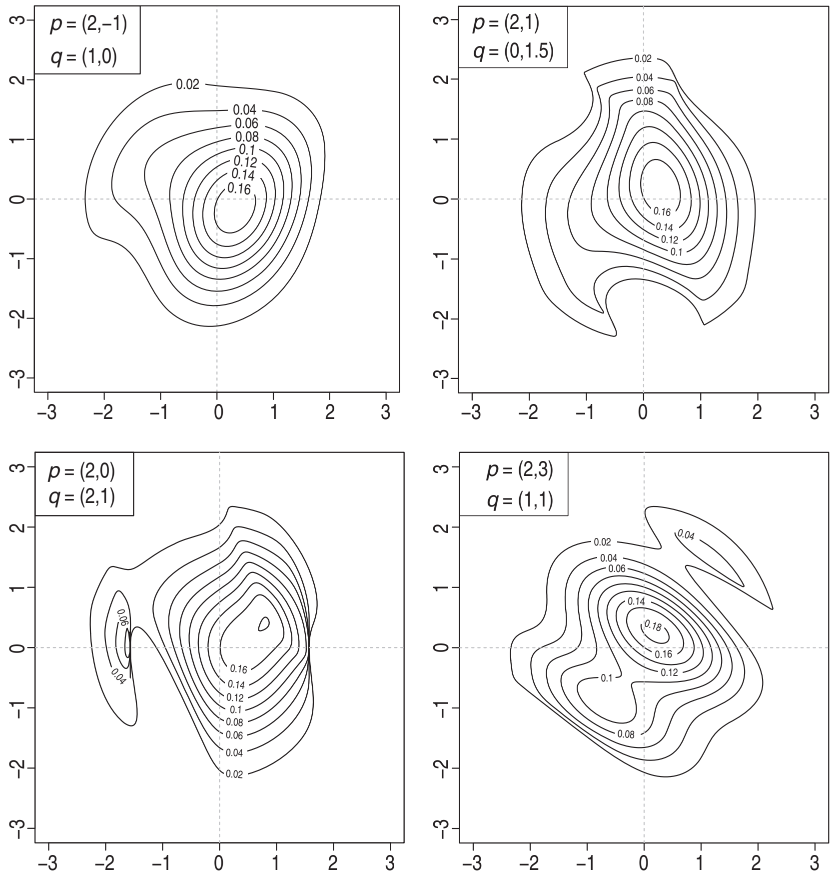

Figure 1.1 Density function of a bivariate standard normal variate with independent components modulated by a logistic distribution factor with argument regulated by (1.6) using parameters indicated in the top-left corner of each panel.

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*

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

https://doi.org/10.1017/CBO9781139248891.002 Published online by Cambridge University Press

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 $\mathbb{P}\{S = 1\} = \int_{\mathbb{R}^d} G(x) f_0(x) \, dx = \frac{1}{2}$, 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*

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

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

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

https://doi.org/10.1017/CBO9781139248891.002 Published online by Cambridge University Press

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

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

https://doi.org/10.1017/CBO9781139248891.002 Published online by Cambridge University Press

if F_0 denotes the cumulative distribution function of f_0 and $\overline{F}$ denotes the survival function, which is defined for a variable $Z = (Z_1, \ldots, Z_d)$ as

$$\overline{F}(x) = \mathbb{P}\{Z_1 \geq x_1, \ldots, Z_d \geq x_d\} . \tag{1.17}$$

1.2.3 The univariate case

Additional results can be obtained for the case $d = 1$. An immediate consequence of (1.16) is

$$1 - F(-x) = 2\, F_0(x) - F(x), \qquad x \in \mathbb{R}, \tag{1.18}$$

which will be useful shortly.

The following representation can be obtained with an argument similar to Proposition 1.3. Note that $V = |Z|$ has distribution $2\, f_0(\cdot)$ on $[0, \infty)$, irrespective of the modulation factor, and is of type (1.2). See Problem 1.2.

Proposition 1.6 *If Z_0 is a univariate variable having density f_0 symmetric about* 0*, $V = |Z_0|$ and G satisfies (1.4), then*

$$Z = S_V\, V, \qquad S_V = \begin{cases} +1 & \text{with probability } G(V), \\ -1 & \text{with probability } G(-V) \end{cases} \tag{1.19}$$

has density function (1.3).

We know that $\mathbb{E}\{Z^m\} = \mathbb{E}\{Z_0^m\} = \mathbb{E}\{V^m\}$ for $m = 0, 2, 4 \ldots$ The odd moments of Z can be expressed with the aid of (1.19) as

$$\begin{aligned} \mathbb{E}\{Z^m\} &= \mathbb{E}\{S_V\, V^m\} \\ &= \mathbb{E}_V\{\mathbb{E}\{S_V | V\}\, V^m\} \\ &= \mathbb{E}\{[G(V) - G(-V)] V^m\} \\ &= \mathbb{E}\{[2\, G(V) - 1] V^m\} \\ &= 2\, \mathbb{E}\{V^m\, G(V)\} - \mathbb{E}\{V^m\} , \quad m = 1, 3, \ldots \end{aligned} \tag{1.20}$$

Consider now a fixed base density f_0 and a set of modulating functions G_k, all satisfying (1.4). What can be said about the resulting perturbed versions of f_0? This broad question can be expanded in many directions. An especially interesting one, tackled by the next proposition, is to find which conditions on the G_k ensure that there exists an ordering on the distribution functions

$$F_k(x) = \int_{-\infty}^{x} 2\, f_0(u)\, G_k(u)\, \mathrm{d}u\,, \tag{1.21}$$

since this fact implies a similar ordering of moments and quantiles. If the

https://doi.org/10.1017/CBO9781139248891.002 Published online by Cambridge University Press

variables X_1 and X_2 have distribution functions F_1 and F_2, respectively, recall that X_2 is said to be stochastically larger than X_1, written $X_2 \geq_{\rm st} X_1$, if $\mathbb{P}\{X_2 > x\} \geq \mathbb{P}\{X_1 > x\}$ for all x, or equivalently $F_1(x) \geq F_2(x)$. In this case we shall also say that X_1 is stochastically smaller than X_2, written $X_1 \leq_{\rm st} X_2$. An introductory account of stochastic ordering is provided by Whitt (2006).

Proposition 1.7 *Consider functions G_1 and G_2 on $\mathbb{R}$ which satisfy condition (1.4) and additionally $G_2(x) \geq G_1(x)$ for all $x > 0$. Then distribution functions (1.21) satisfy*

$$F_1(x) \geq F_2(x)\,, \qquad x \in \mathbb{R}. \tag{1.22}$$

If $G_1(x) > G_2(x)$ for all x in some interval, (1.22) holds strictly for some x.

Proof Consider first $s \leq 0$ and notice that $G_1(x) \geq G_2(x)$ for all $x < s$. This clearly implies $F_1(s) \geq F_2(s)$. If $s > 0$, the same conclusion holds using (1.18) with $x = -s$. QED

To illustrate, consider variables Z_0, Z and $|Z_0|$ whose respective densities are: (i) $f_0(x)$, (ii) $2\,f_0(x)\,G(x)$ with G continuous and $\frac{1}{2} < G(x) < 1$ for $x > 0$, and (iii) $2\,f_0(x)\,I_{[0,\infty)}(x)$. They can all be viewed as instances of (1.3), recalling that the first distribution is associated with $G(x) \equiv \frac{1}{2}$ and the third one with $G(x) = I_{[0,\infty)}(x)$, both fulfilling (1.4). From Proposition 1.7 it follows that

$$Z_0 \;\leq_{\rm st}\; Z \;\leq_{\rm st}\; |Z_0| \tag{1.23}$$

and correspondingly, for any increasing function $t(\cdot)$, we can write

$$\mathbb{E}\{t(Z_0)\} \;< \mathbb{E}\{t(Z)\} \;<\; \mathbb{E}\{t(|Z_0|)\}\,, \tag{1.24}$$

provided these expectations exist. Here strict inequalities hold because of analogous inequalities for the corresponding G functions, which implies strict inequality for some x in (1.22). A case of special interest is when $t(x) = x^{2k-1}$, for $k = 1, 2, \ldots$, leading to ordering of odd moments. Another implication of stochastic ordering is that p-level quantiles of the three distributions are ordered similarly to expectations in (1.24), for any $0 < p < 1$.

We often adopt the form of (1.2), with pertaining conditions, and it is convenient to formulate a version of Proposition 1.7 for this case.

Corollary 1.8 *Consider $G_1(x) = G_0\{w_1(x)\}$ and $G_2(x) = G_0\{w_2(x)\}$, where G_0, w_1 and w_2 satisfy (1.1) and additionally G_0 is monotonically increasing. If $w_2(x) \geq w_1(x)$ for all $x > 0$, then (1.22) holds. If $w_1(x) > w_2(x)$*

for all x in some interval of the positive half-line, (1.22) holds strictly for some x.

A further specialization occurs when $w_j(\cdot)$ represents an instance of the linear form $w(x) = \alpha x$, where α is an arbitrary constant, leading to the form (quite popular in this stream of literature)

$$f(x;\alpha) = 2\, f_0(x)\, G_0(\alpha x)\,, \qquad x \in \mathbb{R}, \tag{1.25}$$

where of course f_0 and G_0 are as in Proposition 1.1.

Corollary 1.9 *If f_0 and G_0 are as in Proposition 1.1 with $d = 1$, the set of densities (1.25) indexed by the real parameter α have distribution functions stochastically ordered with α.*

1.2.4 Bibliographic notes

A simplified version of Proposition 1.1 for the linear case of type $w(x) = \alpha x$ when $d = 1$ has been presented by Azzalini (1985); the rest of that paper focuses on the skew-normal distribution, which is the theme of the next two chapters. A follow-up paper (Azzalini, 1986) included, in the restricted setting indicated, stochastic representations analogous to those presented in § 1.2.2 and § 1.2.3, and a statement (his Proposition 1) equivalent to modulation invariance. Azzalini and Capitanio (1999, Section 7) introduced a substantially more general result, which will be examined later in this chapter.

The present version of Proposition 1.1 is as given by Azzalini and Capitanio (2003); the matching formulation (1.3)–(1.4) was developed independently by Wang *et al*. (2004), who showed the essential equivalence of the two constructions. Both papers included the corresponding general forms of stochastic representation and perturbation invariance. Wang *et al*. (2004) included also Proposition 1.5, up to an inessential modification. An intermediate formulation of similar type, where f_0 is a density of elliptical type, has been presented by Genton and Loperfido (2005).

The content of § 1.2.3 is largely based on § 3.1 of Azzalini and Regoli (2012a), with some exceptions: Proposition 1.6 and (1.20) have been given by Azzalini (1986), the latter up to a simple extension; inequalities similar to (1.24) have been obtained by Umbach (2006) for the case of an odd function $t(\cdot)$ such that $t(x) > 0$ for $x > 0$.

1.3 Some broader formulations

1.3.1 Other conditioning mechanisms

We want to examine more general constructions than that of Proposition 1.1, by relaxing the conditions involved. At first sight this programme seems pointless, recalling that, by Proposition 1.5, the set of distributions already encompassed is the widest possible. Such explorations make sense when we fix in advance some of the components; quite commonly, we want to pre-select the base density f_0. With these restrictions, the statement of Proposition 1.5 is affected.

As a first extension to the setting of Proposition 1.1, we replace the component $G_0\{w(x)\}$ by $G_0\{\alpha_0 + w(x)\}$, where α_0 is some fixed but arbitrary real number. This variant is especially natural if one thinks of the linear case $\alpha_0 + \alpha x$, which has been examined by various authors. With the same notation and type of argument adopted in the proof of Proposition 1.1, it follows that

$$f(x) = f_0(x)\,\frac{G_0\{\alpha_0 + w(x)\}}{\mathbb{P}\{T < \alpha_0 + w(Z_0)\}} \tag{1.26}$$

is a density function on $\mathbb{R}^d$. We shall commonly refer to this distribution as an *extended* version of the similar one without α_0.

Such a simple modification of the formulation has an important impact on the whole construction, unless of course $\alpha_0 = 0$. One effect is that the denominator of (1.26) must be computed afresh for any choice of components. This computation is feasible in closed form only in favourable cases, while an appealing aspect of (1.2) is to have a fixed $\frac{1}{2}$ here.

In addition, the associated stochastic representation is affected. If we now set

$$S_{Z_0} = \begin{cases} +1 & \text{if } T < \alpha_0 + w(Z_0), \\ -1 & \text{otherwise,} \end{cases} \tag{1.27}$$

then the distribution of $Z = (Z_0 | S_{Z_0} = 1)$ turns out to be (1.26), arguing as in Proposition 1.3. However, a representation similar to (1.10) does not hold because now $G(x) = G_0\{\alpha_0 + w(x)\}$ does not satisfy (1.4). In turn, this removes the modulation invariance property (1.12).

In spite of the above limitations, there are good reasons to explore this direction further. Although an explicit computation of the denominator in (1.26) cannot be worked out in general, still it can be pursued in a set of practically important cases. In addition, strong motivations arise from applications to consider this construction, and even more elaborate ones. In this section we only sketch a few general aspects, since a fuller treatment is feasible only in some specific cases, partly for the reasons explained; these developments will take place in later chapters.

It is convenient to reframe the probability context in a slightly different, but eventually equivalent, manner. Consider a $(d+m)$-dimensional variable (Z_0, Z_1) with joint density $f_*(x_0, x_1)$ such that Z_0 has marginal density f_0 on $\mathbb{R}^d$ and Z_1 has marginal density f_1 on $\mathbb{R}^m$. For a fixed Borel set $C \in \mathbb{R}^m$ having positive probability, consider the distribution of $(Z_0|Z_1 \in C)$, that is

$$f(x) = \frac{\int_C f_*(x, z)\,\mathrm{d}z}{\int_C f_1(z)\,\mathrm{d}z} = f_0(x)\,\frac{\mathbb{P}\{Z_1 \in C|Z_0 = x\}}{\mathbb{P}\{Z_1 \in C\}} \tag{1.28}$$

for $x \in \mathbb{R}^d$; from the first equality we see that $f(x)$ integrates to 1. In the special case when Z_0 and Z_1 are independent, the final fraction in (1.28) reduces to 1, and $f = f_0$.

The appeal of (1.28) comes from its meaningful interpretation from the viewpoint of applied work: $f(x)$ represents the joint distribution of a set of quantities of interest, Z_0, which are observed only for cases fulfilling a certain condition, that is $Z_1 \in C$, determined by another set of variables. As a simple illustration, think of Z_0 as the set of scores obtained by a student in certain university exams, and of Z_1 as the score(s) obtained by the same student in university admission test(s); we can observe Z_0 only for students whose Z_1 belongs to the admission set C. Situations of this type usually go under the heading 'selective sampling' or similar terms; it is then quite natural to denote (1.28) a *selection distribution*.

Expression (1.2) can be obtained as a special case of (1.28) when $m = 1$, $C = (-\infty, 0]$ and $Z_1 = T - w(Z_0)$, where T is a variable with distribution function G_0, independent of Z_0, and conditions (1.1) hold. Clearly (1.28) encompasses much more general situations, of which (1.26) is a subset. The next example is provided by two-sided constraints of the form $a < Z_1 < b$, again when $m = 1$. A much wider scenario is opened up by consideration of multiple constraints when $m > 1$.

Some general conclusions can be drawn about distributions of type (1.28). One of these is that, if Z_0 is transformed to $t(Z_0)$, the conditional distribution of $(t(Z_0)|Z_1 \in C)$ is still computed using (1.28), replacing the distribution of Z_0 with that of $t(Z_0)$. One implication is that, if f_0 belongs to a parametric family closed under a set of invertible transformations $t(\cdot)$, such as the set of affine transformations, then the same closure property holds for (1.28). See also Problem 1.8.

Because of its ample generality, it is difficult to develop more general conclusions for (1.28). As already indicated, in later chapters we shall examine important subcases, in particular those which allow a manageable computation of the two integrals involved, in connection with a symmetric density f_0, usually of elliptical type. The case of interest here is $m > 1$ since the case with $m = 1$ falls under the umbrella of the modulation invariance property.

Bibliographic notes

Emphasis has been placed on distributions of type (1.26), especially when $w(\cdot)$ is linear, by Barry Arnold and co-workers in a series of papers, many of which are summarized in Arnold and Beaver (2002); some will be described specifically later on. An initial formulation of (1.28) has been presented by Arellano-Valle *et al*. (2002), referring to the case where C is an orthant of $\mathbb{R}^m$, extended first by Arellano-Valle and del Pino (2004) and subsequently by Arellano-Valle and Genton (2005) and Arellano-Valle *et al*. (2006). The last paper shows how (1.28) formally encompasses a range of specific families of distributions examined in the literature. The focus on their development lies in situations where f_0 in (1.28) is a symmetric density; this case gives rise to what they denote *fundamental skew-symmetric* (FUSS) distributions. As already remarked, a unified theory does not appear to be feasible much beyond this point and specific, although very wide, subclasses must be examined. Some general results, however, have been provided by Arellano-Valle and Genton (2010a) with special emphasis on the distribution of quadratic forms when the parent population before selection has a normal or an elliptically contoured distribution.

1.3.2 Working with generalized symmetry

Proposition 1.10 *Denote by T a continuous real-valued random variable with distribution function G_0 symmetric about* 0 *and by Z_0 a d-dimensional variable with density function f_0, independent of T, such that the real-valued variable $W = w(Z_0)$ is symmetric about* 0*. Then*

$$f(x) = 2\, f_0(x)\, G_0\{w(x)\}\,, \qquad x \in \mathbb{R}^d, \tag{1.29}$$

is a density function.

Proof See the final line of the ‘instructive proof’ of Proposition 1.1. QED

https://doi.org/10.1017/CBO9781139248891.002 Published online by Cambridge University Press

Proposition 1.1 can be seen as a restricted version of this result, since conditions (1.1) are sufficient to ensure that $w(Z_0)$ has a symmetric distribution about 0. From an operational viewpoint the formulation in Proposition 1.1 is more convenient because checking conditions (1.1) is immediate, but does not embrace all possible settings falling within Proposition 1.10. Notice that Proposition 1.10 does not require that f_0 is symmetric about 0.

For a simple illustration, consider the density function on $\mathbb{R}^2$ obtained by modulating the bivariate normal with standardized marginals and correlations ρ, denoted $\varphi_B(x_1, x_2; \rho)$, as follows:

$$f(x) = 2\,\varphi_B(x_1, x_2; \rho)\,\Phi\{\alpha(x_1^2 - x_2^2)\}, \qquad x = (x_1, x_2) \in \mathbb{R}^2, \tag{1.30}$$

where α is a real parameter and Φ is the standard normal distribution function. In this case the perturbation factor modifies the base density, preserving central symmetry. Figure 1.2 shows two instances of this density.

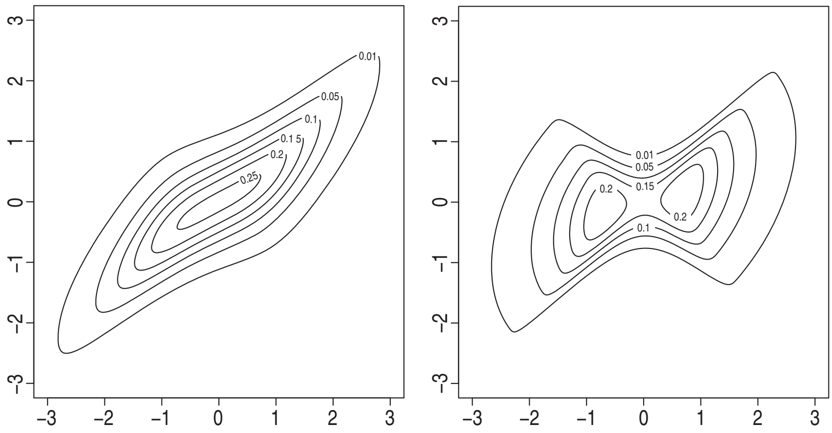

Figure 1.2 Density functions of type (1.30), displayed as contour level plots: in the left panel $\alpha = 1, \rho = 0.8$; in the right panel $\alpha = 3, \rho = 0.4$.

The fact that $f(x)$ integrates to 1 does not follow from Proposition 1.1 which requires an odd function $w(x)$, while $w(x) = \alpha(x_1^2 - x_2^2)$ is even; equivalently, $G(x) = \Phi\{\alpha(x_1^2 - x_2^2)\}$ does not satisfy (1.4). However, if $Z_0 = (Z_{01}, Z_{02})^\top \sim N_2(0, \Omega)$ where Ω is the 2×2 correlation matrix with off-diagonal entries ρ, it is true that $w(Z_0) = \alpha(Z_{01}^2 - Z_{02}^2)$ has a symmetric distribution about 0, and so Proposition 1.10 can be applied to conclude that (1.30) integrates to 1. In this respect, it would be irrelevant to replace Φ in (1.30) by some other symmetric distribution function G_0.

https://doi.org/10.1017/CBO9781139248891.002 Published online by Cambridge University Press

From the argument of the proof, it is immediate that a random variable with distribution (1.29) admits a representation of type (1.9). For the reasons already discussed in connection with (1.26), it is desirable that a representation similar to (1.10) also exists. The next result provides a set of sufficient conditions to this end.

Proposition 1.11 *Let T and Z_0 be as in Proposition 1.10, and suppose that there exists an invertible transformation $R(\cdot)$ such that, for all $x \in \mathbb{R}^d$,*

$$f_0(x) = f_0[R(x)], \quad |\det R'(x)| = 1, \quad w[R(x)] = -w(x)\,, \tag{1.31}$$

where $R'(x)$ denotes the Jacobian matrix of the partial derivatives, then

$$Z = \begin{cases} Z_0 & \text{if } T \le w(Z_0), \\ R^{-1}(Z_0) & \text{otherwise} \end{cases} \tag{1.32}$$

has distribution (1.29).

Proof The density function of Z at x is

$$\begin{aligned} f(x) &= f_0(x)\, G_0\{w(x)\} + f_0(R(x))\, |\det R'(x)|\, [1 - G_0\{w(R(x))\}] \\ &= f_0(x)\, G_0\{w(x)\} + f_0(x)\, [1 - G_0\{-w(x)\}] \\ &= 2\, f_0(x)\, G_0\{w(x)\} \end{aligned}$$

using (1.31) and $G_0(-x) = 1 - G_0(x)$. QED

In this formulation the condition of (central) symmetry $f_0(x) = f_0(-x)$ has been replaced by the first requirement in (1.31), $f_0(x) = f_0[R(x)]$, which represents a form of *generalized symmetry*. Usual symmetry is recovered when $R(x) = -x$. The requirement of an odd function w is replaced here by the similarly generalized condition given by the last expression in (1.31).

For the corresponding extension of the modulation invariance property (1.12), consider a transformation from $\mathbb{R}^d$ to $\mathbb{R}^q$ which is even in the generalized sense adopted here, that is

$$t(x) = t(R^{-1}(x)), \qquad x \in \mathbb{R}^d\,.$$

It is immediate from representation (1.32) that (1.12) then holds.

For distribution (1.30), conditions (1.31) are fulfilled by the transformation

$$R(x) = R_0\, x\,, \qquad R_0 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = R_0^{-1}\,,$$

which swaps the two coordinates, and $w(x) = \alpha(x_1^2 - x_2^2)$. Therefore, if $Z = (Z_1, Z_2)$ has density (1.30), perturbation invariance holds for any transformation $t(Z)$ such that $t((Z_1, Z_2)) = t((Z_2, Z_1))$. One implication is that

https://doi.org/10.1017/CBO9781139248891.002 Published online by Cambridge University Press

$Z^\top\Omega^{-1}Z \sim \chi^2_2$. Another consequence is that, since $t(x) = x_1x_2 = x_2\,x_1 = t(R_0x)$, then $\mathbb{E}\{Z_1\,Z_2\} = \rho$. Since central symmetry holds for $f(x)$, then $\mathbb{E}\{Z_1\} = \mathbb{E}\{Z_2\} = 0$ and so $\text{cov}\{Z_1, Z_2\} = \rho$.

Using Proposition 1.10, one can construct distributions also with non-symmetric base density; see Problem 5.17 for an illustration.

Finally, note that the statement of Proposition 1.10 is still valid under somewhat weaker assumptions, as follows. We can relax the assumption about absolute continuity of all distributions involved, and allow G or the distribution of $w(Z_0)$ to be of discrete or mixed type, provided the condition $\mathbb{P}\{T - W(Z_0) \le 0\} = \frac{1}{2}$ in (1.29) still holds. A sufficient condition to meet this requirement is that at least one of T and $W(Z_0)$ is continuous.

Bibliographic notes

Proposition 1.10 has been presented by Azzalini and Capitanio (1999, Section 7). Although it was followed by a remark that the base density does not need to be symmetric, the ensuing development focused on elliptical distributions, and this route was followed in a number of subsequent papers, including extensions to the weaker condition of central symmetry; these have been quoted in earlier sections. The broader meaning of Proposition 1.10 has been reconsidered by Azzalini (2012), on which this section is based. Since exploration of this direction started only recently, no further discussion along this line will take place in the following chapters.

1.4 Complements

Complement 1.1 (Random number generation) For sampling from distribution (1.2), both (1.9) and (1.10) provide a suitable technique for random number generation. However, in practice the first one is not convenient, since it involves rejection of half of the sampled Z_0's, on average.

To generate S_{Z_0}, both forms (1.11) are suitable. Which of the two variants is computationally more convenient depends on the specific instance under consideration. The second form involves computation of $G(x)$, which in practice is expressed as $G_0\{w(x)\}$. Since evaluation of $w(\cdot)$ is required in both cases, the comparison is then between computation of G_0 and generation of U versus generation of T. A general statement on which route is preferable is not possible, because the comparison depends on a number of factors, including the computing environment in use.

Further stochastic representations may exist for specific subclasses of (1.2), to be discussed in subsequent chapters. In these cases, they provide additional generation algorithms for random number generation.

https://doi.org/10.1017/CBO9781139248891.002 Published online by Cambridge University Press

Sampling from a distribution of type (1.26) is a somewhat different problem compared with (1.2), because only representation following (1.27) holds in general here. Its use implies rejection of a fraction of the sampled Z_0's, and the acceptance fraction can be as low as 0 if α_0 approaches $-\infty$. The more general set of distributions (1.28) can be handled in a similar manner: sample values (Z_0, Z_1) are drawn from f_*, and we accept only those Z_0's such that $Z_1 \in C$. For both situations, the problem of non-constant, and possibly very low, acceptance rate can be circumvented for specific subclasses of (1.26) or of (1.28) which allow additional stochastic representations that do not involve an acceptance–rejection technique; again, these will be discussed in subsequent chapters.

Complement 1.2 (A characterization) The property of modulation invariance (1.12) leads to a number of corollaries for distributions of type (1.3) which share the same base density f_0; some of these corollaries appear in the next proposition. However, the interesting fact is not their isolated validity, but instead the fact that they are equivalent to each other and to representation (1.3), hence providing a characterization result.

More explicitly, if modulation invariance holds for all even $t(\cdot)$, this implies that the underlying distributions allow a representation of type (1.3) with common base f_0.

Proposition 1.12 *Consider variables* $Z = (Z_1, ..., Z_d)^\top$ *and* $Y = (Y_1, ..., Y_d)^\top$ *with distribution functions* F *and* H*, and density functions* f *and* h*, respectively; denote by* $\overline{F}$ *and* $\overline{H}$ *the survival functions of* Z *and* Y*, respectively, defined as in (1.17). The following conditions are then equivalent:*

(a) densities $f(x)$ *and* $h(x)$ *admit a representation of type (1.3) with the same symmetric base density* $f_0(x)$*;*

(b) $t(X) \stackrel{d}{=} t(Y)$*, for any even q-dimensional function t on* $\mathbb{R}^d$*;*

(c) $\mathbb{P}\{Z \in A\} = \mathbb{P}\{Y \in A\}$*, for any symmetric set* $A \subset \mathbb{R}^d$*;*

(d) $F(x) + \overline{F}(-x) = H(x) + \overline{H}(-x)$*;*

(e) $f(x) + f(-x) = h(x) + h(-x) \quad$ *(a.e.).*

Proof

(a)⇒(b) This follows from the perturbation invariance property of Proposition 1.4.

(b)⇒(c) Simply note that the indicator function of a symmetric set A is an even function.

(c)⇒(d) On setting

$$A_+ = \{s = (s_1, \dots, s_d) \in \mathbb{R}^d : s_j \le x_j, \forall j\},$$
$$A_- = \{s = (s_1, \dots, s_d) \in \mathbb{R}^d : -s_j \le x_j, \forall j\} = -A_+,$$
$$A_\cup = A_+ \cup A_-\,,$$
$$A_\cap = A_+ \cap A_-\,,$$

both $A_\cup$ and $A_\cap$ are symmetric sets. Hence we obtain:

$$\begin{aligned} F(x) + \overline{F}(-x) &= \mathbb{P}\{Z \in A_+\} + \mathbb{P}\{Z \in A_-\} \\ &= \mathbb{P}\{Z \in A_\cup\} + \mathbb{P}\{Z \in A_\cap\} \\ &= \mathbb{P}\{Y \in A_\cup\} + \mathbb{P}\{Y \in A_\cap\} \\ &= H(x) + \overline{H}(-x)\,. \end{aligned}$$

(d)⇒(e) Taking the dth mixed derivative of the final relationship in (d), relationship (e) follows.

(e)⇒(a) This follows from the representation given in Proposition 1.5.

QED

This proof is taken from Azzalini and Regoli (2012a). For the case $d = 1$, an essentially equivalent result has been given by Huang and Chen (2007, Theorem 1).

Complement 1.3 (On uniqueness of the mode) Another interesting theme concerns the range of possible shapes of the modulated density f, for a given base f_0. This is a very broad issue, only partly explored so far. A specific but important question is as follows: if f_0 has a unique mode, when does f also have a unique mode?

In the case $d = 1$, it is tempting to conjecture that a monotonic G preserves uniqueness of the mode of f_0, but this is dismissed by the example having $f_0 = \varphi$, N(0, 1) density and $G(x) = \Phi(x^3)$, where Φ is the N(0, 1) distribution function. Figure 1.3 illustrates graphically this case; the left panel displays G, the right panel shows f.

Sufficient conditions for uniqueness of the mode of f are given by the next statement, which we reproduce without proof from Azzalini and Regoli (2012a). Recall that *log-concavity* of a density means that the logarithm of the density is a concave function; in the univariate case, this property is equivalent to *strong unimodality* of the density (Dharmadhikari and Joag-dev, 1988, Theorem 1.10).

Proposition 1.13 *In case $d = 1$, if $G(x)$ in (1.3) is an increasing function*

https://doi.org/10.1017/CBO9781139248891.002 Published online by Cambridge University Press

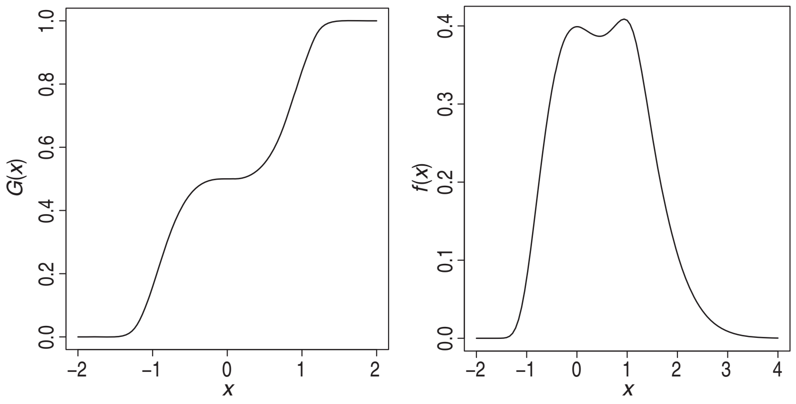

Figure 1.3 Example of a bimodal density produced with f_0 equal to the N(0, 1) density and $G(x) = \Phi(x^3)$; the left panel displays $G(x)$, the right panel the modulated density.

and $f_0(x)$ is unimodal at 0, then no negative mode exists. If we assume that f_0 and G have continuous derivatives everywhere on the support of f_0, $G(x)$ is concave for $x > 0$ and $f_0(x)$ is log-concave, where at least one of these properties holds in a strict sense, then there is a unique positive mode of $f(x)$. If $G(x)$ is decreasing, similar statements hold, with reversed sign of the mode; the uniqueness of the negative mode requires that $G(x)$ is convex for $x < 0$.

A popular situation where the conditions of this proposition are readily checked is (1.25) with linear w.

Corollary 1.14 *In case $d = 1$, if f_0 in Proposition 1.1 is log-concave and G_0' is continuous everywhere and unimodal at 0, then density (1.25) is unimodal for all α, and the mode has the same sign as α.*

A related issue, which includes uniqueness of the mode as a byproduct, will be discussed in Chapter 6, for general d.

Complement 1.4 (Transformation of scale) Jones (2013) has put forward an interesting proposal for the construction of flexible families of distributions which has a direct link with our main theme. We digress briefly in that direction for the aspects which illustrate this connection, without attempting a full summary of his formulation.

On the real line, consider a density f_0, symmetric about 0, having support

S_0. For a transformation t from the set S to $D \supseteq S_0$, it may happen that

$$f(x) = 2\, f_0\{t(x)\}, \qquad x \in S \tag{1.33}$$

is a density function; in this case, the mechanism leading from f_0 to f is called *transformation of scale*, as opposed to the familiar transformation of variable. The next statement provides conditions to ensure that (1.33) is indeed a proper density.

Proposition 1.15 *Let $\bar{G} : D \to S$ denote a piecewise differentiable monotonically increasing function with inverse t, where $D \supseteq S_0 \ni 0$. If*

$$\bar{G}(z) - \bar{G}(-z) = z, \qquad \textit{for all } z \in D \tag{1.34}$$

and f_0 is density symmetric about 0 *with support S_0, then (1.33) is a density on S.*

Proof Non-negativity of f follows from that of f_0, so we only need to prove that it integrates to 1. We consider the case where $\bar{G}$ is differentiable everywhere, with obvious extension to the case of piecewise differentiability. Making the substitution $z = t(x)$ and writing $G(z) = \bar{G}'(z)$, which is positive for all $z \in D$, write

$$\int_S 2\; f_0\{t(x)\}\, \mathrm{d}x = 2 \int_D f_0(z)\, G(z)\, \mathrm{d}z = 2 \int_{S_0} f_0(z)\, G(z)\, \mathrm{d}z\,.$$

Since function G is positive and, on differentiating (1.34), fulfils conditions (1.4), then the above integral equals 1. QED

The argument of the proof shows that a variable X with distribution (1.33) can be obtained as $X = \bar{G}(Z)$, where Z has distribution of type (1.3) with $G = \bar{G}'$.

However, not all transformations of Z achieve the form (1.33), since $\bar{G}$ must satisfy (1.34). It can be shown that $t = \bar{G}^{-1}$ must be of the type $t(x) = x - s(x)$ where $s : \mathbb{R}^+ \to \mathbb{R}^+$ is an onto monotone decreasing function that is a self-inverse, i.e. $s^{-1}(x) = s(x)$. The proof of this fact is given by Jones (2013), together with various additional results. See also the related work of Jones (2012).

Complement 1.5 (Fechner-type distributions) A number of authors have considered asymmetric distributions on the real line obtained by applying different scale factors to the half-line $x > x_0$ and to the half-line $x < x_0$ of a density symmetric about x_0, which we can take equal to 0. This idea goes back to Fechner (1897, Chapter XIX) who applied it to the normal density, and it has re-emerged several times since then, in various forms of

https://doi.org/10.1017/CBO9781139248891.002 Published online by Cambridge University Press

parameterization. See Mudholkar and Hutson (2000) for a variant form and an overview of others. Hansen (1994) employed the same device to build an asymmetric form of Student's distribution. A similar type of construction has been developed by Hinkley and Revankar (1977), by an independent argument, leading to a form of asymmetric Laplace distribution.

With similar logic, Arellano-Valle *et al.* (2005b) consider the class of densities

$$\frac{2}{a(\alpha)+b(\alpha)}\left[f_0\left(\frac{x}{a(\alpha)}\right)I_{[0,\infty)}(x)+f_0\left(\frac{x}{b(\alpha)}\right)I_{(-\infty,0)}(x)\right], \tag{1.35}$$

where f_0 is a density symmetric about 0 and α is a parameter which regulates asymmetry via the positive-valued functions $a(\cdot)$ and $b(\cdot)$. On setting $a(\alpha) = \alpha$ and $b(\alpha) = 1/\alpha$ where $\alpha > 0$, (1.35) leads to the class of Fernández and Steel (1998).

If X is a random variable with density (1.35), a stochastic representation is $X = W_\alpha\,|X_0|$ where X_0 has density $f_0(x)$ and W_α is an independent discrete variate such that

$$\mathbb{P}\{W_\alpha = a(\alpha)\} = \frac{a(\alpha)}{a(\alpha)+b(\alpha)}, \quad \mathbb{P}\{W_\alpha = -b(\alpha)\} = \frac{b(\alpha)}{a(\alpha)+b(\alpha)}\,.$$

Arellano-Valle *et al.* (2006) noted that this stochastic representation allows us to view (1.35) as an instance of the selection distributions (1.28). First note that $|X_0| \stackrel{d}{=} (X_0|X_0 > 0)$; hence set $X \stackrel{d}{=} (Z_0|Z_1 \in C)$ where $Z_0 = W_\alpha X_0$, $Z_1 = X_0$, $C = (0,\infty)$. Combining these settings, rewrite $X = W_\alpha\,|X_0|$ as $X = (W_\alpha X_0|X_0 > 0)$, which coincides with $X = (Z_0|Z_1 > 0)$.

Problems

1.1 Consider two independent real-valued continuous random variables, U and V, with common density f_0, symmetric about 0. Show that $Z_1 = \min\{U, V\}$ and $Z_2 = \max\{U, V\}$ have density of type (1.2) with base f_0.

1.2 Confirm that $V = |Z|$ introduced right before Proposition 1.6 has density $2\,f_0(\cdot)$ on $[0,\infty)$ and find the expression of $G(x)$ to represent this distribution in the form (1.3).

1.3 Prove Proposition 1.6.

1.4 Assume that Z, conditionally on α, is a random variable with density function (1.25) and that α is a random variable with density symmetric about 0. Show that the unconditional density of Z is f_0. Extend this result to the general case (1.7) provided w is both an odd function of x for any fixed α and an odd function of α for any fixed x.

1.5 The product of two symmetric Beta densities rescaled to the interval $(-1, 1)$ takes the form

$$f_0(x, y) = \frac{(1-x^2)^{a-1}\,(1-y^2)^{b-1}}{4^{a+b-1}\,B(a,a)\,B(b,b)}, \qquad (x, y) \in (-1, 1)^2,$$

for some positive a and b. Define $f(x, y) = 2\, f_0(x, y)\, L[w(x, y)]$, where $L(t) = (1 + \exp(-t))^{-1}$ is the standard logistic distribution function and

$$w(x, y) = \frac{\sin(p_1 x + p_2 y)}{1 + \cos(q_1 x + q_2 y)}\,.$$

Check that $f(x, y)$ is a properly normalized density on $(-1, 1)^2$. Choose constants $(a, b, p_1, p_2, q_1, q_2)$ as you like and plot the density using your favourite computing environment; repeat this step 11 more times.

1.6 For the variables in (1.23), show that $\mathrm{var}\{Z_0\} > \mathrm{var}\{Z\} > \mathrm{var}\{|Z_0|\}$, provided $\mathrm{var}\{Z_0\}$ exists.

1.7 Confirm that (1.26) is a density function.

1.8 Prove that, if a variable Z having selection distribution (1.28) with f_0 centrally symmetric is partitioned as $Z = (Z', Z'')$, then both the marginal distribution of Z' and that of Z' conditional on the value taken on by Z'' are still of the same type (Arellano-Valle and Genton, 2005).

1.9 Show that in (1.30) we can replace $w(x) = \alpha(x_1^2 - x_2^2)$ by

$$w(x) = \alpha_1(x_1 - x_2) + \cdots + \alpha_m(x_1^m - x_2^m)$$

for any natural number m and any choice of the coefficients $\alpha_1, \ldots, \alpha_m$, and still obtain a proper density function. Discuss the implication of selecting coefficients α_j where (i) only odd-order terms are non-zero, (ii) only even-order terms are non-zero (Azzalini, 2012).

1.10 If $\varphi_B(x_1, x_2; \rho)$ denotes the bivariate normal density with standardized marginals and correlation ρ, show that

$$2\,\varphi_B(x_1, x_2; \rho)\,\Phi\{\alpha x_1(x_2 - \rho x_1)\}, \quad 2\,\varphi_B(x_1, x_2; \rho)\,\Phi\{\alpha x_2(x_1 - \rho x_2)\},$$

for $(x_1, x_2) \in \mathbb{R}^2$, are density functions. Establish whether a representation of type (1.32) holds (Azzalini, 2012). *Note:* when $\rho = 0$, both forms reduce to a distribution examined by Arnold *et al.* (2002), which enjoys various interesting properties – its marginals are standardized normal densities and the conditional distribution of one component given the other is of skew-normal type, to be discussed in Chapter 2.

1.11 Consider the family of d-dimensional densities of type (1.2) where the base density is multivariate normal, $\varphi_d(x; \Sigma)$. Show that this family is closed under h-dimensional marginalization, for $1 \le h < d$ (Lysenko *et al.*, 2009).

https://doi.org/10.1017/CBO9781139248891.002 Published online by Cambridge University Press