CAMBRIDGE

Cambridge University Press
978-1-107-02927-9 - The Skew-Normal and Related Families
Adelchi Azzalini and Antonella Capitanio
Excerpt
More information

# 1

# Modulation of symmetric densities

## 1.1 Motivation

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

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

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

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

## 1.2 Modulation of symmetry

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

### *1.2.1 A fairly general construction*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

QED

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

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

where

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

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

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

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

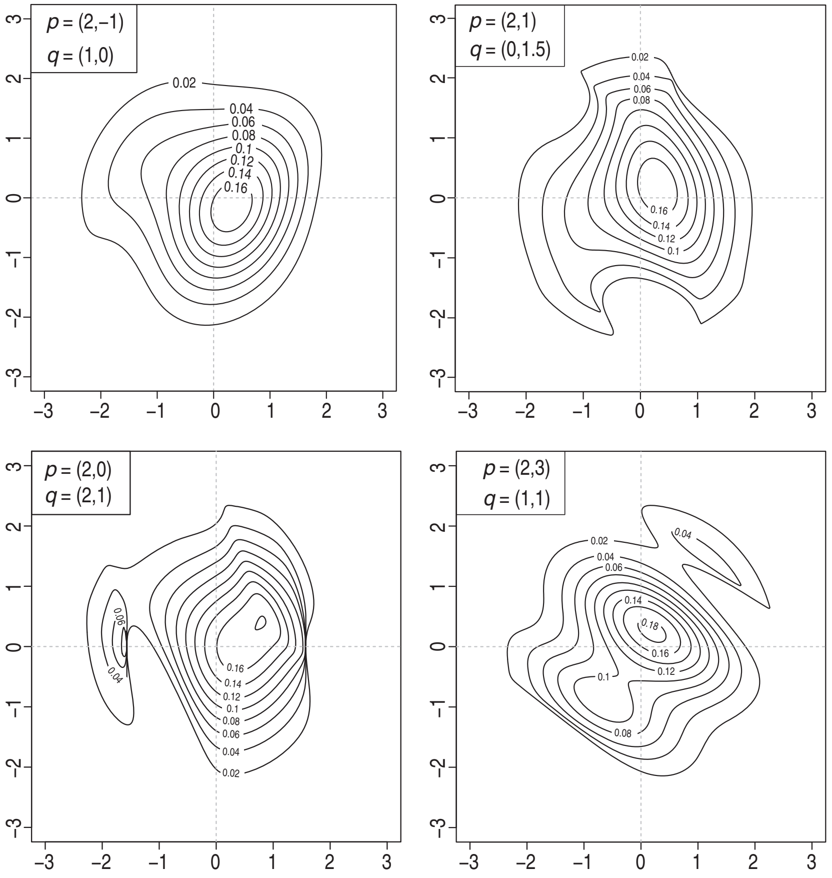


**Figure 1.1** Density function of a bivariate standard normal variate with independent components modulated by a logistic distribution factor with argument regulated by (1.6) using parameters indicated in the top-left corner of each panel.

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

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

### *1.2.2 Main properties*

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

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

*Then both variables*

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

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

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

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

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

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

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

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

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

Representations (1.11) are obvious. QED

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

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

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

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

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

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

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

for the $r$th 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

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

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 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 $w_1(x) > w_2(x)$*