Stochastic dominance with parametric distributions

Suman SETH∗ Gaston YALONETZKY†

September 16, 2014

Abstract

We propose a method to derive first- and second-order dominance conditions for parametric distributions whenever their respective cumulative distribution functions are either not readily available or not easily tractable. Our necessary conditions rely mostly on the distributions’ moment-generating functions, whereas our sufficient conditions are mostly derived directly from manipulations of the distributions’ density functions. We provide results for the Log-normal, Gamma, Fisk, and Weibull distributions.

Keywords: Stochastic dominance, parametric distributions.
JEL Classification: C46, I31.

∗Oxford Poverty and Human Development Initiative. E-mail: suman.seth@qeh.ox.ac.uk.
†University of Leeds; Maurice Keyworth Building LS2 9JT, UK. Phone: (+44) 113 343 0199. E-mail: G.Yalonetzky@leeds.ac.uk.
1 Introduction

Stochastic dominance provides us with an incomplete ordering of distributions, whose meaning depends on the field of application. For instance, in the realm of poverty measurement, the fulfillment of some dominance condition may ensure that a monetary poverty comparison is robust to any choice of poverty line and/or of individual poverty function. The theoretical advantages and disadvantages of stochastic dominance analysis are well known. Meanwhile, in practice, stochastic dominance testing applied to continuous distributions requires comparing them over several points along their common domain (e.g. Davidson and Duclos, 2000, Barret and Donald, 2003). However, if our continuous distributions were parameterised, would it be possible to produce dominance orderings straightforwardly by just comparing the respective parameters? The aim of this note is precisely to derive parametric conditions that ensure first- and second-order stochastic dominance among parametric distributions that can be useful in different applications, e.g. for model simulation purposes. Resorting to these conditions, practitioners would be spared the more cumbersome (but more generally applicable) non-parametric tests.

There is already a substantial literature on the parametric conditions for Lorenz orderings (Kleiber and Kotz, 2003), except for generalised beta distribution of first type (GB1). By contrast, there is not enough work on the parametric stochastic dominance conditions for different distributions. Levy worked on Truncated Normal (Levy, 1982) and Log Normal Distributions (Levy, 1973). Klonner derived the first-order dominance conditions for Singh-Maddala distributions (Klonner, 2000), but not the second-order conditions. Klenke and Mattner (2010) developed dominance conditions for classical discrete distributions. As a tractable cumulative distribution function is not readily available for many parametric distributions, it is difficult to derive their dominance conditions in the usual way, i.e. by comparing cumulative distribution functions, or their sums (e.g. in the case of higher orders).

In this paper we propose a novel approach resorting to the distributions’ density functions, which are tractable and readily available. The approach yields three sets conditions based solely on the distributions’ parameters, which we derive for Log Normal, Gamma, Fisk, and Weibull distributions; all with support on the positive segment of the real line. Their fulfillment guarantees, respectively, first-order dominance, second-order dominance, and Lorenz consistency. Hence, as soon as we know the parameters of two distributions of the same type, we can ascertain the existence of a dominance relationship between them.

The rest of the paper proceeds as follows: The next section introduces required notation and stochastic dominance definitions. Then the methodology section explains our method for the derivation of dominance conditions based on density functions of parametric distributions. Finally, the results section shows the first-order, second-order dominance conditions, and the Lorenz consistency conditions, for the distributions mentioned above. The paper ends with some final remarks.
2 Preliminaries

Let $\Omega$ be the domain of a continuous random variable $x$ with distribution $A$. Also $f_A(x) : \Omega \to \mathbb{R}_+ | \int_{\Omega} f_A(x) \, dx = 1$ is its density function, $F_A(x) \equiv \int_x^\infty f(t) \, dt$ is its cumulative distribution function (with $x$ standing for the lowest boundary of the domain), and $E_A[x]$ is its expected value. By stating $A \succ^1 B$ we mean that distribution $A$ first-order dominates $B$. Likewise $A \succ^2 B$ means that $A$ second-order dominates $B$, and $A \succ^L B$ means that $A$ Lorenz-dominates $B$. Now we define the three relationships in terms of their welfare-utility interpretations:

Definition 1. First-order Stochastic Dominance ($\succ^1$): $E_A[u(x)] > E_B[u(x)]$ for all strictly increasing, continuous utility functions $u(x) : \Omega \to \mathbb{R}$ if and only if $F_A(x) \preceq F_B(x) \quad \forall x \in \Omega \land \exists x | F_A(x) < F_B(x)$.

Definition 2. Second-order Stochastic Dominance ($\succ^2$): $E_A[u(x)] > E_B[u(x)]$ for all strictly increasing, strictly concave, continuous utility functions $u(x) : \Omega \to \mathbb{R}$ if and only if $\int_x^\infty F_A(t) \, dt \leq \int_x^\infty F_B(t) \, dt \quad \forall x \in \Omega \land \exists x | \int_x^\infty F_A(t) \, dt < \int_x^\infty F_B(t) \, dt$.

Definition 3. Lorenz dominance ($\succ^L$): $I_A(x) > I_B(x)$ for all inequality indices $I : A \to \mathbb{R}_+$ satisfying the properties of anonymity, population replication invariance, scale invariance and principle of transfers, if and only if $\frac{1}{E_A[x]} \int_x^y x \, dF_A(x) \leq \frac{1}{E_B[x]} \int_x^y x \, dF_B(x) \quad \forall y \in \Omega \land \exists y | \frac{1}{E_A[x]} \int_x^y x \, dF_A(x) < \frac{1}{E_B[x]} \int_x^y x \, dF_B(x)$.

3 Methodology

Our approach relies on density functions which are readily available for parametric distributions, as opposed to cumulative distribution functions. We work with the following set of assumptions:

Assumption 1. Density functions are absolutely continuous.

Assumption 2. For any absolutely continuous density function $f(x)$, the sign of $f'(x)$ changes at most once over the domain of $x$.

The next step is the derivation of two theorems, whose respective corollaries are used to derive sufficient dominance conditions when applied to the distributions’ density functions. Then, the necessary counterparts are derived using the utility functions introduced by Klonner (2000), again in combination with the parametric features of the distributions’ density functions.

3.1 Tools for the derivation of sufficient conditions

The first theorem is the following:

Theorem 1. If $f_A(x) - f_B(x)$ starts negative and changes sign at most once, then $A \succ^1 B$. 

The following corollary provides the operationalization of theorem 1 for our purpose of deriving first-order dominance conditions for parametric distributions:

**Corollary 1.** Let \( z(x) \equiv \frac{f_A(x)}{f_B(x)} \). Then, for first-order dominance favouring \( A \), given the above assumptions, we require: (1) \( \lim_{x \to -\infty} z(x) < 1 \); (2) \( \lim_{x \to \infty} z(x) > 1 \); and (3) \( z(x) = 1 \) is unique in the domain of \( x \).

For the second-order stochastic dominance we rely on the following theorem:

**Theorem 2.** If (1) \( f_A(x) - f_B(x) \) starts negative, (2) changes sign at most twice, and (3) \( E_A(x) \geq E_B(x) \): then \( A \succcurlyeq B \).

**Proof.** From theorem 1 we know that if \( f_A(x) - f_B(x) \) starts negative and changes sign at most once, then \( A \succcurlyeq B \), which in turn implies \( A \succcurlyeq B \). So we need to prove that, if (1) \( f_A(x) - f_B(x) \) starts negative, (2) changes sign twice, and (3) \( E_A(x) \geq E_B(x) \): then \( A \succcurlyeq B \).

The first step is to realize that if \( f_A(x) - f_B(x) \) starts negative and changes sign twice, then, using an argument analogous to the one put forward in the proof of theorem 1, it is easy to show that \( F_A(x) - F_B(x) \) starts negative and changes sign only once, i.e. the cumulative distribution functions cross only once.

If the cumulative distribution functions cross at point \( y \) (in the domain of \( x \)) then \( \int_y^\infty [F_A(x) - F_A(B)] dx < 0 \). However for second-order dominance we need that relationship to hold also for the interval \([y, \pi]\) (where \( \pi \) is the relevant upper boundary value of the domain). The single crossing of the cumulative distribution functions does not guarantee this required result. However, we have imposed that \( E_A(x) \geq E_B(x) \). Since, it is easy to show that: \( E_A(x) - E_B(x) = \int_y^\infty [F_B(x) - F_A(x)] dx \), and we are imposing \( E_A(x) \geq E_B(x) \), then it must be the case that \( \int_y^\infty [F_A(x) - F_A(B)] dx < 0 \) for the whole relevant domain, which implies: \( A \succcurlyeq B \).

As before, we derive a corollary enabling us to apply theorem 2 in practice:

**Corollary 2.** Let \( z(x) \equiv \frac{f_A(x)}{f_B(x)} \). Then, for second-order dominance favouring \( A \), given the above assumptions, we require: (1) \( \lim_{x \to -\infty} z(x) < 1 \); (2) \( \lim_{x \to \infty} z(x) < 1 \); (3) \( z(x) \) has a global maximum; (4) \( E_A(x) \geq E_B(x) \).
3.2 Tools for the derivation of necessary conditions

We derive the necessary conditions relying on the definitions of stochastic dominance used by Klonner (2000). For first-order dominance we use the utility function: \( u(x) = \frac{1}{t} x^t \forall t \neq 0 \); whereas for second-order we use the utility function: \( u(x) = \frac{1}{t} x^t \forall t < 0 \). In both cases, we derived the expected values of \( u(x) \) for each distribution using its corresponding moment-generating function. Then taking certain limits of \( t \) (e.g. toward 0 or \( \infty \)) we can deduce the necessary dominance conditions. A full-fledged illustration of how this method works, in tandem with the method for sufficient conditions, is shown in the Appendix.

4 Results

Using the above tools we report the following results for several two-parameter distributions: Log Normal, Log Logistic, Gamma, Weibull and Uniform distribution:

4.1 The Log Normal distribution

The lognormal distribution can be defined as:

\[
 f_A(x) = \frac{1}{x\sigma_A\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma_A^2}(\ln x - \mu_A)^2\right); \quad x, \mu, \sigma > 0
\]  

The conditions ensuring dominance of \( A \) over \( B \) are as follows:

- First-order dominance - necessary and sufficient: \( \mu_A > \mu_B \) and \( \sigma_A = \sigma_B \).
- Second-order dominance - necessary and sufficient: \( \mu_A > \mu_B \), \( \sigma_A \leq \sigma_B \), and \( \mu_A + \frac{\sigma_A^2}{2} > \mu_B + \frac{\sigma_B^2}{2} \).
- Lorenz dominance - necessary and sufficient: \( \sigma_A < \sigma_B \)

Of course, as it is to be expected generally, first-order dominance implies second-order dominance, which in turn implies Lorenz dominance (but the reverses are not true).

Figure 1 shows the dominance regions of the log-normal distribution, defined by combinations of its two parameters. For an hypothetical distribution \( g \) we show its parameters \( \mu_g \) and \( \sigma_g \), together with an iso-mean function denoted by \( E_g(x) \) (i.e. \( \exp(\mu_g + \frac{\sigma_g^2}{2}) = \exp(\mu + \frac{\sigma^2}{2}) \)), and an iso-variance function denoted by \( Var_g(x) \) (i.e. \( (\exp(\sigma_g^2) - 1)\exp(2\mu_g + \sigma_g^2) = (\exp(\sigma^2) - 1)\exp(2\mu + \sigma^2) \). The upper contour sets of both curves represent higher mean and variance, respectively.

Therefore, the area to the right of the blue (solid and dotted) line denoting \( \sigma_g \) comprehends all the log-normal distributions Lorenz dominated by \( g \), whereas the area to the left (light blue arrows) represents all the distributions that Lorenz dominate \( g \). Likewise, the solid blue line hosts all the distributions first-order dominating \( g \), while the dotted blue line collects all the distributions first-order dominated by \( g \). (The rest of the quadrant cannot be ordered vis-a-vis \( g \) in terms of first-order dominance). Finally, the area delimited by the iso-mean function, the solid blue line and the vertical axis, shows the distributions
that second-order dominate $g$, whereas the area bounded by the iso-mean function, the
dotted blue line and the horizontal axis hosts the distributions second-order dominated by
$g$.

Figure 1: Dominance regions in the Log-normal distribution

4.2 The Gamma distribution

The Gamma distribution is defined as:

$$f_A(x) = \frac{1}{b_A^p \Gamma(p_A)} x^{p_A-1} \exp(-\frac{x}{b_A}),$$

(2)

where $x, b, p > 0$, and $\Gamma(p) = \int_0^\infty x^{p-1} \exp(-x) dx$ is the Gamma function. The dominance
conditions are the following:

- First-order dominance - sufficient: $p_A > p_B$ and $b_A > b_B$; necessary: $p_A > p_B$ and
  $p_A b_A > p_B b_B$.
- Second-order dominance - necessary and sufficient: $p_A > p_B$ and $p_A b_A > p_B b_B$.
- Lorenz dominance - necessary and sufficient: $p_A > p_B$. 

Figure 2: Dominance regions in the Gamma distribution

Figure 2 shows now the dominance regions for the Gamma distribution. The light blue arrows point to the part of the quadrant where distributions Lorenz dominate distribution \( g \), i.e. toward the right of \( p_g \). To the left, distributions are Lorenz dominated by \( g \). The region between the two solid blue lines (with origin in the point \( (p_g, b_g) \)) contains the parametric combinations that yield distributions first-order dominating \( g \); whereas the region bounded by the dotted blue lines and the two axes hosts all the distributions first-order dominated by \( g \). Finally, the region made by the upper contour set of the intersection between the vertical solid blue line and the iso-mean curve crossing through \( (p_g, b_g) \), provides the distributions that second-order dominate \( g \). By contrast, the lower contour set of the intersection between the vertical dotted blue line and the same iso-mean curve collect the distributions that are second-order dominated by \( g \).

4.3 The Log Logistic or Fisk distribution

The Log Logistic distribution is defined as:

\[
 f_A(x) = \frac{a_A x^{a_A-1}}{b_A^{a_A} \left[1 + \left(\frac{x}{b_A}\right)^{a_A}\right]^2}; \quad x, a, b > 0
\]  

As the first moment of this distribution is not defined unless \( a_A > 1 \), we impose \( a > 1 \). The dominance conditions are the following:

- First-order dominance - necessary and sufficient: \( b_A > b_B \) and \( a_A = a_B \).
- Second-order dominance - necessary and sufficient: \( a_A \geq a_B \) and \( \frac{\pi b_B}{a_A \sin\left(\frac{\pi}{a_A}\right)} \geq \frac{\pi b_B}{a_B \sin\left(\frac{\pi}{a_B}\right)} \).
- Lorenz dominance - necessary and sufficient: \( a_A \geq a_B \).
Figure 3 shows the dominance regions for the log-logistic or Fisk distribution. As before, the light blue arrows point toward the region where the distributions Lorenz-dominate \( g \) (right of the vertical blue line passing through \( a_g \)). The rest of the quadrant is made of distributions which are Lorenz-dominated by \( g \). The solid blue line represents all the distributions that first-order dominate \( g \), whereas the dotted blue line represents the distributions first-order dominated by \( g \). Finally, the solid black curve represents the iso-mean curve passing through \((a_g, \beta_g)\). The region made by the upper contour set of the intersection between the solid blue line and the iso-mean curve collects all the distributions that second-order dominate \( g \). By contrast, the region delimited between the iso-mean curve, the dotted blue line and the horizontal axis represents the distributions that are second-order dominated by \( g \).

4.4 The Weibull distribution

The Weibull distribution is defined as:

\[
f_A(x) = \frac{a_A}{\beta_A} x^{a_A-1} \exp\left[-\left(\frac{x}{\beta_A}\right)^{a_A}\right]; \quad x, a, \beta > 0
\]  

(4)

The dominance conditions are the following:

- First-order dominance - necessary and sufficient: \( \beta_A > \beta_B \) and \( a_A = a_B \).
- Second-order dominance - necessary and sufficient: \( a_A \geq a_B \) and \( \beta_A \Gamma(1 + \frac{1}{a_A}) \geq \beta_B \Gamma(1 + \frac{1}{a_B}) \).
- Lorenz dominance - necessary and sufficient: \( a_A \geq a_B \).
5 Conclusions

In this paper we have characterized the parametric conditions for stochastic dominance pertaining to the Log-normal, Gamma, Log-logistic, and Weibull distributions. Our method avoids the use of cumulative distribution functions (and its respective accumulations), therefore being useful whenever these are neither readily available nor tractable.

Regarding symmetric distributions with the whole real line as support, it is worth noting that, if we look at the Theorem 1, it is evident that we require one density to intersect the other only once and from below, in order to dominate the later by first-order. But, if the variance of two symmetric distributions of the same family are different then the densities are going to intersect each other more than once and the corresponding cumulative distribution functions will cross, preventing first-order dominance. Therefore, in this context, first-order dominance requires higher mean with equal variance. It can also be shown that second-order dominance requires the mean of the dominance distribution to be at least as high as the other distribution’s, and its variance to be at least as low as that of the dominated distribution.

Characterizing the whole set of parametric conditions for a much broader range of distributions could intensify the application of stochastic dominance technique. Hence our

Figure 4: Dominance regions in the Weibull distribution

Figure 4 shows the dominance regions for the Weibull distribution. As in many cases before, the distributions Lorenz dominating \( g \) are found on the right of the vertical blue line intersecting the horizontal axis at \( a_g \). The rest of the quadrant corresponds to distributions Lorenz dominated by \( g \). First-order dominance against \( g \) occurs along the solid blue line, whereas first-order dominance favouring \( g \) takes place along the dotted blue line. Finally, the region formed by the upper contour set of the intersection between the solid blue line and the black iso-mean curve provides the distributions that second-order dominate \( g \), whereas the region formed by the iso-mean curve, the dotted blue line, and the horizontal axis, contains all the distribution second-order dominated by \( g \).
next attempt is to find out all the necessary and sufficient conditions for all the distributions defined by McDonald (1984). Further extensions may include conditions for distributions with support throughout the whole real line, bivariate dominance conditions, comparisons between two different distributions, and/or mixtures.

Finally, we have not tackled the derivation of third-order dominance conditions. While this task is possible, we are less inclined to pursue it, as long as little attention is paid to the concept in the applied literature.

References


6 Appendix: Proofs of the dominance conditions for the log-normal distribution

We provide the following proof of the dominance conditions for the log-normal distribution as an illustration of how the methods for necessary and sufficient conditions work. Proofs for the other distributions are available upon request.
6.1 First order dominance

6.1.1 Necessary conditions

As we know, first-order dominance is equivalent to \( E_A[u(x)] \geq E_B[u(x)] \) for all strictly increasing, continuous utility functions \( u \). Then this has to be the case for \( u_t \equiv \frac{1}{t}x^t \) \( \forall t \) \( \notin \emptyset \), with \( x \in \mathbb{R}_{++} \).

We can compute the expected value of \( u_t \) when it follows a log-normal distribution, using the moment generating function:

\[
E_A[u_t(x)] = \frac{1}{t} \int x^t f_A(x) dx = \frac{1}{t} \exp(t\mu + \frac{1}{2}t^2\sigma^2).
\]

(5)

If \( A >_1 B \) then it must be the case that: \( E_A[u_t(x)] > E_B[u_t(x)] \). If \( t > 0 \) then the latter implies:

\[
\exp(t\mu_A + \frac{1}{2}t^2\sigma_A^2) > \exp(t\mu_B + \frac{1}{2}t^2\sigma_B^2)
\]

(6)

\[
\mu_A + \frac{1}{2}t\sigma_A^2 > \mu_B + \frac{1}{2}t\sigma_B^2
\]

(7)

\[
\frac{\mu_A}{t} + \frac{1}{2}t\sigma_A^2 > \frac{\mu_B}{t} + \frac{1}{2}t\sigma_B^2
\]

(8)

Now let \( \mu_B > \mu_A \), and consider \( t \) to be very small such that \( t \to 0 \Rightarrow t\sigma_i^2 \to 0 \) then \( \mu_i \) is the deciding factor and \( E_B[u_t(x)] > E_A[u_t(x)] \) by [7]. But this is a contradiction, since we posited that \( E_A[u_t(x)] > E_B[u_t(x)] \). Hence a necessary condition for \( A >_1 B \) is that: \( \mu_A > \mu_B \).

Similarly, consider \( \sigma_A < \sigma_B \) and \( t \to \infty \), then \( E_B[u_t(x)] > E_A[u_t] \) by [8]. Again, this contradicts the premise of \( A \) dominating \( B \). Hence, we need to have \( \sigma_A \geq \sigma_B \).

Next consider, \( t < 0 \). Then \( E_A[u_t(x)] > E_B[u_t] \) implies:

\[
\exp(t\mu_A + \frac{1}{2}t^2\sigma_A^2) < \exp(t\mu_B + \frac{1}{2}t^2\sigma_B^2)
\]

(9)

\[
\mu_A + \frac{1}{2}t\sigma_A^2 > \mu_B + \frac{1}{2}t\sigma_B^2
\]

(10)

\[
\frac{\mu_A}{t} + \frac{1}{2}t\sigma_A^2 < \frac{\mu_B}{t} + \frac{1}{2}t\sigma_B^2
\]

(11)

By a similar argument as above we require \( \mu_A > \mu_B \) by [10]. Now suppose \( \sigma_A > \sigma_B \). Then, if \( t \to -\infty \), \( E_A[u_t(x)] < E_B[u_t(x)] \) by [11]. This is a contradiction. Hence, we require: \( \sigma_A = \sigma_B \). To summarize, the necessary conditions for first-order dominance with log-normal distributions involve both \( \mu_A > \mu_B \) and \( \sigma_A = \sigma_B \).

6.1.2 Sufficient conditions

We construct the ratio \( z(x) = \frac{f_A(x)}{f_B(x)} \) for the log-normal case:
Second-order dominance is equivalent to $E_A[u(x)] \geq E_B[u(x)]$ for all strictly increasing, continuous, and concave utility functions $u$. Then this has to be the case for $u_t \equiv \frac{1}{2}x^t \forall t < 0$, with $x \in \mathbb{R}_{++}$.

Using [9][10] and [11], and proceeding by taking the limits of $t$ toward 0 or $-\infty$ as before, it is easy to deduce that $\mu_A(x) > \mu_B(x)$ and $\sigma_A(x) \leq \sigma_B(x)$ are both necessary conditions. However, since $A >^2 B$ implies $E_A(x) \geq E_B(x)$, then a condition ensuring the latter is necessary (but not sufficient) for the former. Now, note that $\mu_A(x) > \mu_B(x)$ and $\sigma_A(x) \leq \sigma_B(x)$ do not imply $E_A(x) \geq E_B(x)$. Therefore we need a third necessary condition based on the means of the log-normal distributions, namely: $\mu_A + \frac{\sigma_A^2}{2} \geq \mu_B + \frac{\sigma_B^2}{2}$.

### 6.2.2 Sufficient conditions

We consider again $z(x)$, but without the need to impose $\sigma_A = \sigma_B$:

$$z(x) = \frac{\sigma_B}{\sigma_A} \exp\left\{ \frac{1}{2} \left( \ln x - \mu_A \right)^2 \right\} \left( \frac{1}{\sigma_B^2} - \frac{1}{\sigma_A^2} \right) + \ln(x) \left( \frac{1}{\sigma_A^2} \mu_A - \frac{1}{\sigma_B^2} \mu_B \right) + \frac{1}{2} \left( \frac{1}{\sigma_B^2} \mu_B^2 - \frac{1}{\sigma_A^2} \mu_A^2 \right) \right\} \tag{15}$$

It is straightforward to show that, if $\sigma_A \leq \sigma_B$ and $\mu_A > \mu_B$ then $\lim_{x \to 0} z(x) = \lim_{x \to -\infty} z(x) = 0$. Moreover, deriving and inspecting $\frac{\partial z(x)}{\partial x}$, and $\frac{\partial^2 z(x)}{\partial x^2}$ it is also easy to realize that $z(x)$ is strictly concave and has a global maximum. Hence $\sigma_A \leq \sigma_B$ and $\mu_A > \mu_B$ are sufficient conditions for second-order dominance, once we also impose the condition on the mean alongside them: $\mu_A + \frac{\sigma_A^2}{2} \geq \mu_B + \frac{\sigma_B^2}{2}$ (which is also necessary as we mentioned above).

In summary, the three necessary conditions are also sufficient for $A >^2 B$. 