What If We Only Have Approximate Stochastic Dominance?

Vladik Kreinovich[1], Hung T. Nguyen[2,3], and Songsak Sriboonchitta[3]

[1] Department of Computer Science, University of Texas at El Paso
500 W. University, El Paso, TX 79968, USA, vladik@utep.edu
[2] Department of Mathematical Sciences, New Mexico State University
Las Cruces, New Mexico 88003, USA, hunguyen@nmsu.edu
[3] Department of Economics, Chiang Mai University
Chiang Mai, Thailand, songsak@econ.chiangmai.ac.th

Abstract. In many practical situations, we need to select one of the two alternatives, and we do not know the exact form of the user's utility function – e.g., we only know that it is increasing. In this case, stochastic dominance result says that if the cumulative distribution function (cdf) corresponding to the first alternative is always smaller than or equal than the cdf corresponding to the second alternative, then the first alternative is better. This criterion works well in many practical situations, but often, we have situations when for most points, the first cdf is smaller but at some points, the first cdf is larger. In this paper, we show that in such situations of approximate stochastic dominance, we can also conclude that the first alternative is better – provided that the set of points x at which the first cdf is larger is sufficiently small.

1 Stochastic Dominance: Reminder and Formulation of the Problem

In finance, we need to make decisions under uncertainty. In financial decision making, we need to select one of the possible decisions: e.g., whether we sell or buy a given financial instrument (share, option, etc.). Ideally, we should select a decision which leaves us with the largest monetary value x. However, in practice, we cannot predict exactly the monetary consequences of each action: because of the changing external circumstances, in similar situations the same decision can lead to gains and to losses. Thus, we need to make a decision in a situation when we do not know the exact consequences of each action.

In finance, we usually have probabilistic uncertainty. Numerous financial transactions are occurring every moment. For the past transactions, we know the monetary consequences of different decisions. By analyzing these past transactions, we can estimate, for each decision, the frequencies with which this decision leads to different monetary outcomes x. When the sample size is large – and for financial transactions it is large – the corresponding frequencies become very close to the actual probabilities. Thus, in fact, we can estimate the probabilities of different values x.

Comment. Strictly speaking, this is not always true: we may have new circumstances, we can have a new financial instrument for which we do not have many records of its use – but in most situations, knowledge of the probabilities is a reasonable assumption.

How to describe the corresponding probabilities. As usual, the corresponding probabilities can be described either by the probability density function $f(x)$ or by the cumulative distribution function $F(t) \stackrel{\text{def}}{=} \text{Prob}(x \le t)$.

If we know the probability density function $f(x)$, then we can reconstruct the cumulative distribution function as $F(t) = \int_{-\infty}^{t} f(x)\,dx$. Vice versa, if we know the cumulative distribution function $F(t)$, we can reconstruct the probability density function as its derivative $f(x) = F'(x)$.

How to make decisions under probabilistic uncertainty: a theoretical recommendation. Let us assume that we have several possible decisions whose outcomes are characterized by the probability density functions $f_1(x)$, $f_2(x)$, ... According to the traditional decision making theory (see, e.g., [3, 5–7]), the decisions of a rational person can be characterized by a function $u(x)$ called *utility function* such that this person always selects a decision with the largest value of expected utility $\int f_i(x) \cdot u(x)\,dx$.

A decision corresponding to the probability distribution function $f_1(x)$ is preferable to the decision corresponding to the probability distribution function $f_2(x)$ if

$$\int f_1(x) \cdot u(x)\,dx > \int f_2(x) \cdot u(x)\,dx,$$

i.e., equivalently, if

$$\int \Delta f(x) \cdot u(x)\,dx > 0,$$

where we denoted $\Delta f(x) \stackrel{\text{def}}{=} f_1(x) - f_2(x)$.

Comment. It is usually assumed that small changes in x lead to small changes in utility, i.e., in formal terms, that the function $u(x)$ is differentiable.

From a theoretical recommendation to practical decision. Theoretically, we can determine the utility function of the decision maker. However, since such a determination is very time-consuming, it is rarely done in real financial situations. As a result, in practice, we only have a partial information about the utility function.

One thing we know for sure if that the larger the monetary gain x, the better the resulting situation; in other words, we know that the utility $u(x)$ grows with x, i.e., the utility function $u(x)$ is increasing.

Often, this is the only information that we have about the utility function. How can we make a decision in such a situation?

How to make decisions when we only know that utility function is increasing: analysis of the problem. When is the integral $\int \Delta f(x) \cdot u(x)\,dx$ positive?

To answer this question, let us first note that while theoretically, we have gains and losses which can be arbitrarily large, in reality, both gains and losses are bounded by some value T. In other words, $f_i(x) = 0$ for $x \le -T$ and for $x \ge T$ and thus,

$$F_i(-T) = \text{Prob}_i(x \le -T) = 0$$

and

$$F_i(T) = \text{Prob}_i(x \le T) = 1.$$

In this case,

$$\int \Delta f(x) \cdot u(x)\, dx = \int_{-T}^{T} \Delta f(x) \cdot u(x)\, dx.$$

Let us now take into account that since $\Delta f(x) = f_1(x) - f_2(x)$, $f_1(x) = F_1'(x)$, and $f_2(x) = F_2'(x)$, we can conclude that $\Delta f(x) = \Delta F'(x)$, where

$$\Delta F(x) \stackrel{\text{def}}{=} F_1(x) - F_2(x).$$

We can therefore apply integration by parts

$$\int_{\ell}^{u} a'(x) \cdot b(x)\, dx = a(x) \cdot b(x)|_{\ell}^{u} - \int_{\ell}^{u} a(x) \cdot b'(x)\, dx,$$

with $a(x) = \Delta f(x)$ and $b(x) = u(x)$, to the above integral. As a result, we get the formula

$$\int_{-T}^{T} \Delta f(x) \cdot u(x)\, dx = \Delta F(x) \cdot u(x)|_{-T}^{T} - \int \Delta F(x) \cdot u'(x)\, dx.$$

Since $F_1(-T) = F_2(-T) = 0$, we have

$$\Delta F(-T) = F_1(-T) - F_2(-T) = 0.$$

Similarly, from $F_1(T) = F_2(T) = 1$, we conclude that

$$\Delta F(T) = F_1(T) - F_2(T) = 0.$$

Thus, the first term in the above expression for integration by parts is equal to 0, and we have

$$\int_{-T}^{T} \Delta f(x) \cdot u(x)\, dx = -\int \Delta F(x) \cdot u'(x)\, dx.$$

We know that the utility function is increasing, so $u'(x) \ge 0$ for all x. Thus, if $\Delta F(x) \le 0$ for all x – i.e., if $F_1(x) \le F_2(x)$ for all x – then the difference $\int \Delta f(x) \cdot u(x)\, dx$ is always non-negative and thus, the decision corresponding to the probability distribution function $f_1(x)$ is preferable to the decision corresponding to the probability distribution function $f_2(x)$.

This is the main idea behind *stochastic dominance* (see, e.g., [4, 8]):

Stochastic dominance: summary. If $F_1(x) \le F_2(x)$ for all x and the utility function $u(x)$ is increasing, then the decision corresponding to the probability distribution function $f_1(x)$ is preferable to the decision corresponding to the probability distribution function $f_2(x)$.

Comments.

- The condition $F_1(x) \le F_2(x)$ for all x is not only sufficient to conclude that the first alternative is better, it is also necessary. Indeed, if $F_1(x_0) > F_2(x_0)$ for some x_0, then, since both cumulative distribution functions $F_i(x)$ are differentiable and thus, continuous, there exists an $\varepsilon > 0$ such that $F_1(x) > F_2(x)$ for all $x \in (x_0 - \varepsilon, x_0 + \varepsilon)$.
  We can then take a utility function which:
  - is equal to 0 for $x \le x_0 - \varepsilon$,
  - is equal to 1 for $x \ge x_0 + \varepsilon$, and
  - is, e.g., linear for x between $x_0 - \varepsilon$ and $x_0 + \varepsilon$.

  For this utility function, we have
  $$\int F_1(x) \cdot u'(x)\, dx > \int F_2(x) \cdot u'(x)\, dx,$$
  and thus,
  $$\int f_1(x) \cdot u(x), dx = -\int F_1(x) \cdot u'(x)\, dx <$$
  $$-\int F_2(x) \cdot u'(x)\, dx = \int f_1(x) \cdot u(x), dx,$$
  so the first alternative is worse.
- Sometimes, we have additional information about the utility function. For example, the same amount of additional money h is more valuable for a poor person than for the rich person. This can be interpreted as saying that for every value $x < y$ and, the increase in utility $u(x+h) - u(x)$ is larger than (or equal to) the increase $u(y+h) - u(y)$. If we take the resulting inequality
  $$u(x+h) - u(x) \ge u(y+h) - u(y),$$
  divide both sides by h, and tends h to 0, we conclude that $u'(x) \ge u'(y)$ when $x < y$. In other words, it is reasonable to conclude that the derivative $u'(x)$ of the utility function is decreasing with x – and thus, that its second derivative is negative.
  If this property is satisfied, then we can perform one more integration by parts and get a more powerful criterion for decision making – for situations when we do not know the exact utility function.

What if the stochastic dominance condition is satisfied "almost always": formulation of the problem. Let us return to the simple situation when we only know that utility is increasing, i.e., that $u'(x) \ge 0$. In this case, as we have mentioned, if we know that $F_1(x) \le F_2(x)$ for *all* x, then the first alternative is better. In many cases, we can use this criterion and make a decision.

However, often, in practice, the inequality $F_1(x) \le F_2(x)$ holds for "almost all" values x – i.e., it is satisfied for most values x except for the values x from some small interval. Unfortunately, in this case, as we have shown, the traditional stochastic dominance approach does now allow to make any conclusion – even when the interval is really small. It would be nice to be able to make decisions even if we have *approximate* stochastic dominance.

What we plan to do in this paper. In this paper, we show that, under reasonable assumptions, we can make definite decisions even under approximate stochastic dominance – provided, of course, that the deviations from stochastic dominance are sufficiently small.

Comment. A similar – but somewhat different – problem is analyzed in [1], where it is shown that under certain assumptions, approximate stochastic dominance implies that the first alternative is not much worse than the second one – i.e., if we select the first alternative instead of the second one, we may experience losses, but these losses are bounded, and the smaller the size of the area where $F_1(x)$ is larger than $F_2(x)$, the smaller this bound.

2 How To Make Decisions Under Approximate Stochastic Dominance: Analysis of the Problem

Additional reasonable assumptions about the utility function $u(x)$. In the previous text, we used the fact that the utility function $u(x)$ increases with x, i.e., that its derivative $u'(x)$ is non-negative. Theoretically, we are thus allowing situations when this derivative is extremely small – e.g., equal to 10^{-40} – or, vice versa, extremely large – e.g., equal to 10^{40}.

From the economical viewpoint, however, such too small or too large numbers make no sense. If the derivative is too small, this means that for all practical purposes, the person does not care whether he or she gets more money – which may be true for a monk leading a spiritual life, but not for agents who look for profit. Similarly, if the derivative $u'(x)$ is, for some x, too large, this means that, in effect, the utility function is discontinuous at this x, i.e., that adding a very small amount of money leads to a drastic increase in utility – and this is usually not the case.

These examples show that not only the derivative $u'(x)$ should be non-negative, it cannot be too small and it cannot be too large. In other words, there should be some values $0 < s < L$ for which

$$s \le u'(x) \le L$$

for all x.

This additional assumption helps us deal with situation of approximate stochastic dominance. Let us show that the above additional assumption $0 < s \le u'(x) \le L$ enables us to deal with approximate stochastic dominance. Indeed, we want to make sure that

$$\int \Delta F(x) \cdot u'(x)\,dx \le 0.$$

In the case of stochastic dominance, we have $\Delta F(x) \le 0$ for all x, but we consider the case of *approximate* stochastic dominance, when $\Delta F(x) > 0$ for some values x. To deal with this situation, let us represent the desired integral as the sum of the two component integrals:

- an integral over all the values x for which $\Delta F(x) \le 0$, and
- an integral over all the values x for which $\Delta F(x) > 0$:

$$\int \Delta F(x) \cdot u'(x)\,dx = \int_{x:\Delta F(x)\le 0} \Delta F(x) \cdot u'(x)\,dx + \int_{x:\Delta x>0} \Delta F(x) \cdot u'(x)\,dx.$$

We want to prove that the sum of these two component integrals is bounded, from above, by 0. To prove this, let us find the upper bound for both integrals.

For the values x for which $\Delta F(x) \le 0$, the largest possible value of the product $\Delta F(x) \cdot u'(x)$ is attained when the derivative $u'(x)$ is the smallest possible – i.e., when this derivative is equal to s. Thus, we conclude that

$$\Delta F(x) \cdot u'(x) \le s \cdot \Delta F(x).$$

Therefore,

$$\int_{x:\Delta F(x)\le 0} \Delta F(x) \cdot u'(x)\,dx \le s \cdot \int_{x:\Delta F(x)\le 0} \Delta F(x)\,dx.$$

Since $\Delta F(x) \le 0$, we have $\Delta F(x) = -|\Delta F(x)|$ and thus,

$$\int_{x:\Delta F(x)\le 0} \Delta F(x) \cdot u'(x)\,dx \le -s \cdot \int_{x:\Delta F(x)\le 0} |\Delta F(x)|\,dx.$$

For the values x for which $\Delta F(x) > 0$, the largest possible value of the product $\Delta F(x) \cdot u'(x)$ is attained when the derivative $u'(x)$ is the largest possible – i.e., when this derivative is equal to L. Thus, we conclude that

$$\Delta F(x) \cdot u'(x) \le L \cdot \Delta F(x).$$

Therefore,

$$\int_{x:\Delta F(x)>0} \Delta F(x) \cdot u'(x)\,dx \le L \cdot \int_{x:\Delta F(x)>0} \Delta F(x)\,dx.$$

By combining the bounds on the two component integrals, we conclude that

$$\int \Delta F(x) \cdot u'(x)\,dx \le -s \cdot \int_{x:\Delta F(x)\le 0} |\Delta F(x)|\,dx + L \cdot \int_{x:\Delta F(x)>0} \Delta F(x)\,dx.$$

The integral $\int \Delta F(x) \cdot u'(x)\,dx$ is non-negative if the right-hand side bound is non-negative, i.e., if

$$-s \cdot \int_{x:\Delta F(x)\le 0} |\Delta F(x)|\,dx + L \cdot \int_{x:\Delta F(x)>0} \Delta F(x)\,dx \le 0,$$

i.e., equivalently, if

$$L \cdot \int_{x:\Delta F(x)>0} \Delta F(x)\,dx \le s \cdot \int_{x:\Delta F(x)\le 0} |\Delta F(x)|\,dx,$$

or

$$\int_{x:\Delta F(x)>0} \Delta F(x)\,dx \le \frac{s}{L} \cdot \int_{x:\Delta F(x)\le 0} |\Delta F(x)|\,dx.$$

This condition is satisfied when the set of all the values x for which $\Delta F(x) > 0$ is small – in this case the integral over this set is also small and thus, smaller than the right-hand side.

Let us describe the resulting criterion in precise terms.

3 How To Make Decisions Under Approximate Stochastic Dominance: Main Result

Formulation of the problem. We have two alternatives, characterized by the cumulative distribution functions $F_1(x)$ and $F_2(x)$. We need to decide which of these two alternatives is better.

What we know about the utility function $u(x)$. We know that the utility function $u(x)$ describing the agent's attitude to different monetary values x is non-decreasing: $u'(x) \ge 0$. Moreover, we assume that we know two positive numbers $s < L$ such that for every x, we have

$$s \le u'(x) \le L.$$

Stochastic dominance: reminder. If $F_1(x) \le F_2(x)$ for all x, i.e., if $\Delta F(x) \le 0$ for all x (where we denoted $\Delta F(x) = F_1(x) - F_2(x)$), then the first alternative is better.

New criterion for the case of approximate stochastic dominance. If $\Delta F(x) > 0$ for some values x, but the set of all such x is small, in the sense that

$$\int_{x:\Delta F(x)>0} \Delta F(x)\,dx \le \frac{s}{L} \cdot \int_{x:\Delta F(x)\le 0} |\Delta F(x)|\,dx,$$

then the first alternative is still better.

Comments.

- It is interesting that a similar expression appears in another context: namely, in the study of different notions of transitivity of stochastic relations; see, e.g., [2]. Indeed, adding $\int_{x:\Delta F(x)\le 0} |\Delta F(x)|\,dx$ to both sides of the above inequality, and taking into account that the resulting integral in the left-hand side is simply an integral of $|\Delta F(x)| = |F_2(x) - F_1(x)|$ over all possible x, we conclude that

$$\int |F_2(x) - F_1(x)|\,dx \le \left(1 + \frac{s}{L}\right) \cdot \int_{x:F_2(x)-F_1(x)>0} (F_2(x) - F_1(x))\,dx.$$

  The right-hand side of the new inequality can be described as the interval, over all possible x, of the function $(F_2(x) - F_1(x))_+$, where, as usual, for any function $f(x)$, its positive part $f_+(x)$ is defined as $f_+(x) \stackrel{\text{def}}{=} \max(f(x), 0)$. Thus, this inequality can be represented as

$$\int |F_2(x) - F_1(x)|\,dx \le \left(1 + \frac{s}{L}\right) \cdot \int (F_2(x) - F_1(x))_+\,dx,$$

  or, equivalently, as

$$\frac{\int (F_2(x) - F_1(x))_+\,dx}{\int |F_2(x) - F_1(x)|\,dx} \ge \frac{1}{1 + \dfrac{s}{L}}.$$

  The left-hand side of this inequality is known as the *Proportional Expected Difference*, it is used in several results about transitivity [2].
- The same idea can extend the stochastic dominance criterion corresponding to $u''(x) \le 0$ to the case when this criterion is satisfied for "almost all" values x.

Acknowledgments. We acknowledge the partial support of the Center of Excellence in Econometrics, Faculty of Economics, Chiang Mai University, Thailand.

This work was also supported in part by the National Science Foundation grants HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and DUE-0926721.

The authors are thankful to Bernard de Baets for valuable discussions.

References

1. R. Bruni, F, Cesarone, A. Scozzari, and F. Tardella, "A new stochastic dominance approach to enhanced index tracking problems", *Economics Bulletin*, 2012, Vol. 32, No. 4, pp. 3460–3470.
2. B. de Baets, "State-of-the-art in reciprocal relations", *Proceedings of the International Workshop "Information, Uncertainty, and Imprecision" WIUI'2014*, Olomouc, Czech Republic, April 7–11, 2014.

3. P. C. Fishburn, *Nonlinear Preference and Utility Theory*, John Hopkins Press, Baltimore, Maryland, 1988.
4. H. Levy, *Stochastic Dominance: Investment Decision Making under Uncertainty*, Springer Verlag, New York, 2006.
5. R. D. Luce and R. Raiffa, *Games and Decisions: Introduction and Critical Survey*, Dover, New York, 1989.
6. H. T. Nguyen, V. Kreinovich, B. Wu, and G. Xiang, *Computing Statistics under Interval and Fuzzy Uncertainty*, Springer Verlag, Berlin, Heidelberg, 2012.
7. H. Raiffa, *Decision Analysis*, McGraw-Hill, Columbus, Ohio, 1997.
8. S. Sriboonchitta, W.-K. Wong, S. Dhompongsa, and H. T. Nguyen, *Stochastic Dominance and Applications to Finance, Risk and Economics Hardcover*, CRC Press, Boca Raton, Florida, 2009.