Stochastic dominance

Motivation

Recall that to apply the expected utility model

$$a^* = \underset{a \in \{a_1, a_2, \dots\}}{\operatorname{argmax}} \; \mathbb{E}[U(a)],$$

we need both the probability distribution $P(a_i)$ for each a_i and the risk preferences $U(\cdot)$. In reality, we do not necessarily have access to all the required information. Stochastic dominance addresses cases where we know $P(a_i)$ for all a_i but do not know much about $U(\cdot)$ so that we can still make a choice, a^*, in a consistent manner.

Suppose that the domain of the utility function is a closed interval $[L, M]$,

$$U : [L, M] \to \mathbb{R}.$$

Now, consider two risky choices a_f and a_g, whose distributions are characterized by PDFs $f(x)$ and $g(x)$ respectively. We assume

$$f(x) = g(x) = 0 \text{ for } x \notin [L, M].$$

Then,

$$a_f \succsim a_g \tag{1}$$

$$\Leftrightarrow \quad \mathbb{E}[U(a_f)] \geq \mathbb{E}[U(a_g)] \tag{2}$$

$$\Leftrightarrow \quad \int_L^M U(x)f(x)dx \geq \int_L^M U(x)g(x)dx \tag{3}$$

$$\Leftrightarrow \quad \int_L^M U(x)[f(x) - g(x)]dx \geq 0 \tag{4}$$

Using stochastic dominance, therefore, we try to deduce either $a_f \succsim a_g$ or $a_f \precsim a_g$ from the sign of the left-hand side of (4) despite the limited information about U.

Key notations & equations

$\boldsymbol{U_1}$: a set of utility functions such that $U \in \boldsymbol{U_1}$ implies $U'(x) > 0$
$\boldsymbol{U_2}$: $\boldsymbol{U_2} \subseteq \boldsymbol{U_1}$ such that $U \in \boldsymbol{U_2}$ implies $U''(x) < 0$
$\boldsymbol{U_3}$: $\boldsymbol{U_3} \subseteq \boldsymbol{U_2}$ such that $U \in \boldsymbol{U_3}$ implies $U'''(x) > 0$

For $x \in [L, M]$,

$$f(x), \ g(x) \tag{PDF}$$

$$F(x) = \int_L^x f(t)dt, \ G(x) = \int_L^x g(t)dt \tag{CDF}$$

$$D_1(x) = G(x) - F(x)$$

$$D_2(x) = \int_L^x D_1(t)dt$$

$$D_3(x) = \int_L^x D_2(t)dt$$

We try to determine the sign of (5) by using one of (7), (8) or (9).

$$\mathbb{E}[U(a_f)] - \mathbb{E}[U(a_g)] \tag{5}$$

$$= \int_L^M U(x)[f(x) - g(x)]dx \tag{6}$$

$$= \int_L^M U'(x)D_1(x)dx \tag{7}$$

$$= U'(M)D_2(M) - \int_L^M U''(x)D_2(x)dx \tag{8}$$

$$= U'(M)[\mathbb{E}[a_f] - \mathbb{E}[a_g]] - U''(M)D_3(M) + \int_L^M U'''(x)D_3(x)dx \tag{9}$$

Please refer to Chavas (2004) for derivation, which is basically a patient application of integration by parts.

As you can see in (7), (8) and (9), when applying stochastic dominance, there is a tradeoff between the required conditions on CDFs (hidden in D_i) and the amount of information assumed on U. Depending on how we balance these competing assumptions, we make different statements.

First-order stochastic dominance

Using (7), we get the following.

$$\text{With } U \in \boldsymbol{U_1} \text{ assumed, } a_f \succsim a_g \quad \Leftrightarrow \quad D_1(x) \geq 0, \forall x.$$

This should be obvious and intuitive. $D_1(x) \geq 0$ means that a_f always has a higher chance for a better outcome than a_g. With $U \in \boldsymbol{U_1}$, then, we do not need math to choose a_f.

Second-order stochastic dominance

Here, the mathematical model becomes useful in more complicated situations, i.e. relaxing $D_1(x) \geq 0$ assumption. Using (8),

$$\text{With } U \in \boldsymbol{U_2} \text{ assumed, } a_f \succsim a_g \quad \Leftrightarrow \quad D_2(x) \geq 0, \forall x.$$

It is easy to see it because $D_2(x) \geq 0$ makes both first term and integrand in (8) positive. Although $D_2(x)$ is already not-so-intuitive a quantity, we can certainly compute it for any pair of random variables. Since $U \in \boldsymbol{U_2}$ is a relatively mild assumption, it is a powerful statement.

Third-order stochastic dominance

In different situations, we may find the following statement more applicable.

$$\text{With } U \in \boldsymbol{U_3} \text{ assumed, } a_f \succsim a_g \quad \Leftrightarrow \quad \mathbb{E}[a_f] \geq \mathbb{E}[a_g] \text{ and } D_3(x) \geq 0, \forall x.$$

Notes

- Above, we motivated stochastic dominance (SD) by using continuous random variables. But, SD holds for discrete random variables as well.
- 1st order SD $\Rightarrow$ 2nd order SD $\Rightarrow$ 3rd order SD

  [Proof]
  By definition,
  $$\boldsymbol{U_3} \subseteq \boldsymbol{U_2} \subseteq \boldsymbol{U_1}.$$
  Also, remember a property of integral:
  $$f \geq 0 \Rightarrow \int f \geq 0.$$
  So, by each difinition of D_1, D_2 and D_3,
  $$D_1(x) \geq 0 \Rightarrow D_2(x) \geq 0 \Rightarrow D_3(x) \geq 0.$$
  Finally,
  $$D_2(x) \geq 0 \Rightarrow D_2(M) \geq 0 \Leftrightarrow \mathbb{E}[a_f] - \mathbb{E}[a_g] \geq 0.$$

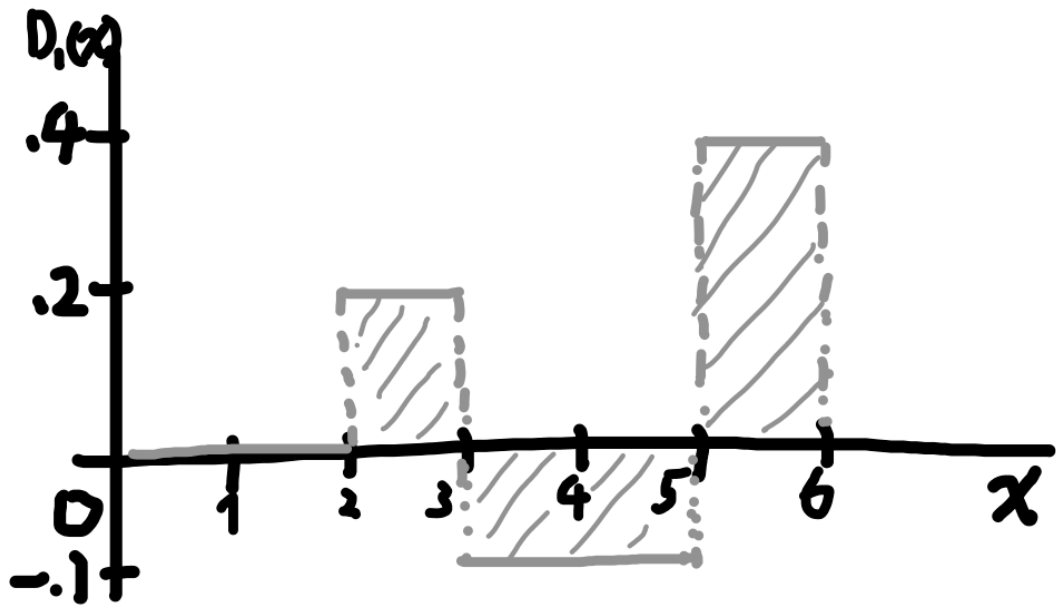
Because,

$$
\begin{aligned}
\mathbb{E}[a_f] - \mathbb{E}[a_g] &= \int_L^M x[f(x) - g(x)]dx \\
&= x[F(x) - G(x)]\Big|_L^M - \int_L^M [F(x) - G(x)]dx \\
&= 0 + \int_L^M [G(x) - F(x)]dx \\
&= \int_L^M D_1(x)dx \\
&= D_2(M)
\end{aligned}
$$

□

- An example of D_1, D_2 and D_3 for discrete random variables. Suppose that we have the following information about two risky choices, characterized by CDFs, $F(x)$ and $G(x)$.

x	1	2	3	4	5	6
$F(x)$	0	0.2	0.5	0.5	1	1
$G(x)$	0	0	0.6	0.6	0.6	1
$D_1(x)$	0	0.2	-0.1	-0.1	0.4	0



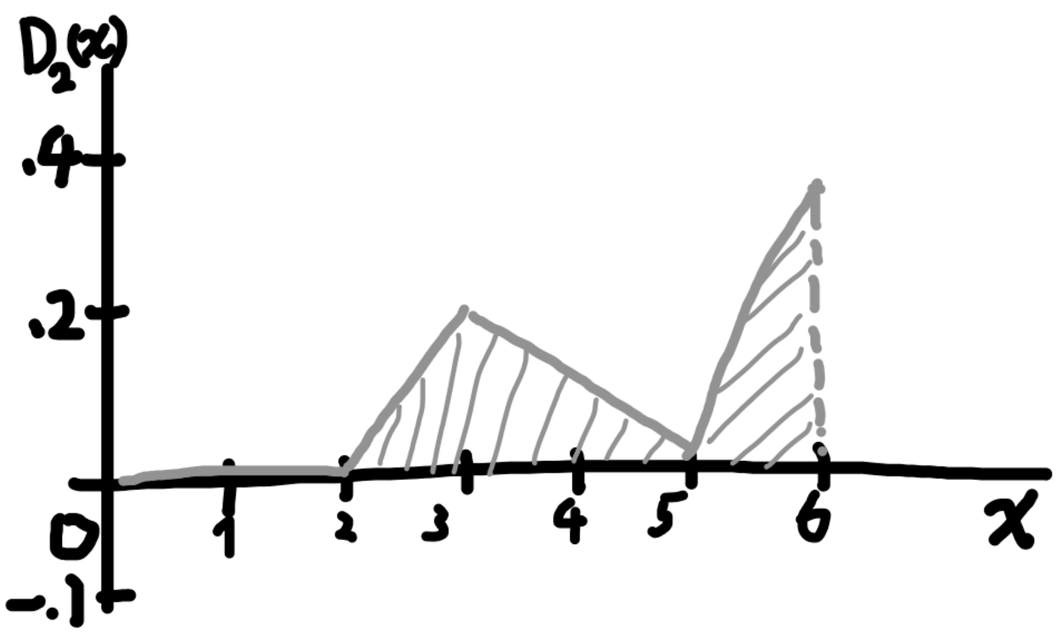
Remember that an integral is the area under a curve. So, the total area is

$$\begin{aligned} D_2(6) &= \int_0^6 D_1(t)dt \\ &= \int_2^3 D_1(t)dt + \int_3^5 D_1(t)dt + \int_5^6 D_1(t)dt \end{aligned}$$

Instead of calculating the total area, how about calculating up to some point $x \leq 6$? We can do this intuitively by combining rectangles.

$$\begin{aligned} D_2(x) &= \begin{cases} 0 & 0 \leq x \leq 2 \\ 0.2(x-2) & 2 \leq x \leq 3 \\ 0.2 - 0.1(x-3) & 3 \leq x \leq 5 \\ 0.2 + (-0.2) + 0.4(x-5) & 5 \leq x \leq 6 \end{cases} \\ &= \begin{cases} 0 & 0 \leq x \leq 2 \\ 0.2x - 0.4 & 2 \leq x \leq 3 \\ -0.1x + 0.5 & 3 \leq x \leq 5 \\ 0.4x - 2 & 5 \leq x \leq 6 \end{cases} \end{aligned}$$

This is a piece-wise linear function.



Do the same for D_3 by combining triangles and trapezoids.

$$\begin{aligned} D_3(6) &= \int_0^6 D_2(t)dt \\ &= \int_2^3 D_2(t)dt + \int_3^5 D_2(t)dt + \int_5^6 D_2(t)dt \\ D_3(x) &= \begin{cases} 0 & 0 \le x \le 2 \\ 0.5(x-2) \times D_2(x) & 2 \le x \le 3 \\ 0.1 + 0.5(x-3) \times [0.2 + D_2(x)] & 3 \le x \le 5 \\ 0.3 + 0.5(x-5) \times D_2(x) & 5 \le x \le 6 \end{cases} \\ &= \begin{cases} 0 & 0 \le x \le 2 \\ 0.5(x-2) \times [0.2x - 0.4] & 2 \le x \le 3 \\ 0.1 + 0.5(x-3) \times [0.2 - 0.1x + 0.5] & 3 \le x \le 5 \\ 0.3 + 0.5(x-5) \times [0.4x - 2] & 5 \le x \le 6 \end{cases} \end{aligned}$$

Rearranging, you see it as a piece-wise quadratic function.