Stochastic order and affiliation

Stefano Lovo*

February 7, 2021

1 Some probability theory preliminaries

1.1 Stochastic Order

Consider two lotteries A and B. Lottery A pays a random amount $\tilde{x}$ and lottery B pays a random amount $\tilde{y}$. If the price for one lottery ticket is the same for lotteries A and B, which lottery would you prefer? Because today the amount paid by each lottery ticket is unknown, it is difficult to say whether A is better than B, or viceversa. This can be known with certainty only after the outcome of both lotteries has realized. However, in some cases, lotteries can be compared from an ex-ant perspective.

1.1.1 First order stochastic dominance

Let denote with F and G the c.d.f of random variables $\tilde{x}$ and $\tilde{y}$, respectively. Suppose that for any real number x the probability that lottery B pays less then x is weakly larger than the probability that lottery A pays less than x. Then, intuitively, from an ex-ante perspective, one would prefer A to B, as for any fixed amount x, it is more likely to gain less then x with lottery B than with lottery A.

In situations like this, we say that c.d.f F **first order stochastically dominates** c.d.f. G, that is

$$\forall x \in \mathbf{R}, F(x) \leq G(x) \tag{1}$$

*HEC,Paris and GREGHEC, 78351 Jouy-en-Josas, France. lovo@hec.fr.

Now, take any increasing and differentiable function γ mapping real number into another real number. Let $\gamma(\tilde{x})$ associate to any realization x of the random variable $\tilde{x}$ the number $\gamma(x)$. Then if $\tilde{x}$ first order stochastically dominates $\tilde{y}$ the expected value of $\gamma(\tilde{x})$ is larger than the expected value of $\gamma(\tilde{y})$. In facts for $\tilde{x}$ and $\tilde{y}$ continuously distributed with density f and g, respectively, one has:

$$\begin{aligned} E[\gamma(\tilde{x})] - E[\gamma(\tilde{y})] &= \int_{-\infty}^{+\infty} \gamma(z)(f(z) - g(z))dz \\ &= -\int_{-\infty}^{+\infty} \gamma'(z)(F(z) - G(z))dz \geq 0 \end{aligned}$$

where the second equality is obtained by integrating by parts and the inequality follows from $\gamma' \geq$ because γ is non-decreasing and $F(z) - G(z) \geq 0$ because of (1).

1.1.2 Hazard rate dominance

Consider a continuously distribute random variable with c.d.f F. Let define the **hazard rate** as the function

$$\lambda_F(x) = \frac{f(x)}{1 - F(x)}.$$

We say that F dominates G in terms of the hazard rate if for any real number x one has

$$\lambda_F(x) \leq \lambda_G(x) \tag{2}$$

Proposition 1 *If F hazard rate dominates G, then F first order stochastically dominates G.*

Proof:

$$F(x) = 1 - \exp\left(-\int_{-\infty}^{x} \lambda_F(z)dz\right) \leq 1 - \exp\left(-\int_{-\infty}^{x} \lambda_G(z)dz\right) = G(x).$$

Q.E.D.

1.1.3 Reverse hazard rate dominance

Consider a continuously distribute random variable with c.d.f F. Let define the **reverse hazard rate** as the function

$$\sigma_F(x) = \frac{f(x)}{F(x)}.$$

We say that F dominates G in terms of the reverse hazard rate if for any real number x one has

$$\sigma_F(x) \geq \sigma_G(x) \tag{3}$$

Proposition 2 *If F reverse hazard rate dominates G, then F first order stochastically dominates G.*

Proof:

$$F(x) = \exp\left(-\int_x^\infty \sigma_F(t)dt\right) \leq \exp\left(-\int_x^\infty \sigma_G(t)dt\right) = G(x)$$

Q.E.D.

1.1.4 Likelihood ratio dominance

Take two random variables $\tilde{x}$ and $\tilde{y}$. Let denote with F and G their c.d.f, respectively, and with f and g their densities, respectively.

The c.d.f. F is said to dominates c.d.f. G **in terms of the likelihood ratio** if for any $x < y$ one has

$$\frac{f(x)}{g(x)} \leq \frac{f(y)}{g(y)} \tag{4}$$

or equivalently $f(x)/g(x)$ is non-decreasing in x.

Observe that likelihood ratio dominance implies hazard rate dominance. In facts, condition (4) is equivalent to

$$\begin{aligned}
\forall x < y \Rightarrow \frac{f(y)}{f(x)} &\geq \frac{g(y)}{g(x)} \\
&\Rightarrow \\
\int_x^\infty \frac{f(y)}{f(x)}dy &\geq \int_x^\infty \frac{g(y)}{g(x)}dy \\
&\Rightarrow \\
\frac{1-F(x)}{f(x)} &\geq \frac{1-G(x)}{g(x)} \\
&\Rightarrow \\
\lambda_F(x) &\leq \lambda_G(x)
\end{aligned}$$

It can also be shown that likelihood ratio dominance implies reverse hazard rate dominance. In facts,

$$\begin{aligned}
\forall x < y \Rightarrow \frac{f(x)}{f(y)} &\leq \frac{g(x)}{g(y)} \\
&\Rightarrow \\
\int_{-\infty}^y \frac{f(x)}{f(y)}dx &\leq \int_{-\infty}^y \frac{g(x)}{g(y)}dx \\
&\Rightarrow \\
\frac{F(y)}{f(y)} &\leq \frac{G(y)}{g(Y)} \\
&\Rightarrow \\
\sigma_F(x) &\geq \sigma_G(x)
\end{aligned}$$

We can conclude that likelihood dominance implies hazard rate dominance and reverse hazard rate dominant, that in turn imply first order stochastic dominance.

1.2 Affiliated random variables

Consider the random variables $\tilde{x}_1\tilde{x}_2\ldots,\tilde{x}_N$. Let D_i be the domain of random variable $\tilde{x}_i$, let $D = \times_i D_i$, that is the product of the domains, and let $f : D \rightarrow \mathbf{R}^+$ be the joint density function. The variables $\tilde{x}_1,\tilde{x}_2\ldots,\tilde{x}_N$ are said to be **affiliated** if for all $x, y \in D$

$$f(x \vee y)f(x \wedge y) \geq f(x)f(y) \tag{5}$$

Where

$$x \vee y = \{\max(x_1, y_1), \max(x_2, y_2), \ldots, \max(x_N, y_N)\}$$

is the component-wise maximum of x and y, and

$$x \wedge y = \{\min(x_1, y_1), \min(x_2, y_2), \ldots, \min(x_N, y_N)\}$$

is the component-wise minimum of x and y.

If the density function f is strictly positive twice continuously differentiable in D, then affiliation is equivalent to having for any $i \neq j$,

$$\frac{\partial^2 \ln f}{\partial x_i, \partial x_j} \geq 0.$$

That is, the derivative of the log of f with respect to a variable x_i is increasing in any other variable x_j.

Affiliation has an important implication when computing conditional distributions. Suppose two random variable $\tilde{x}$ and $\tilde{y}$ are affiliated. Suppose that you observe the realization of $\tilde{x}$ but not the realization of $\tilde{y}$. How does the realization of $\tilde{y}$ affect the distribution of $\tilde{x}$?

Namely what can we say about how the $F(y|x)$, that is the c.d.f. of $\tilde{y}$ conditionally on observing $\tilde{x} = x$, is affected by the value of y?

Proposition 3 *If $\tilde{x}$ and $\tilde{y}$ are affiliated then for any $x' \geq x$, one has that $F(y|x')$ dominates in term of the likelihood ratio $F(y|x)$.*

Proof: Because $\tilde{x}$ and $\tilde{y}$ are affiliated for any $x' > x$ and $y' > y$ it must be that

$$f(x, y')f(x', y) \leq f(x, y)f(x', y')$$

or equivalently

$$\frac{f(x, y')}{f(x, y)} \leq \frac{f(x', y')}{f(x', y)} \tag{6}$$

Now recall that from Baye's rule we have that that

$$f(y|x) = \frac{f(y, x)}{f_{\tilde{x}}(x)}$$

where $f_{\tilde{x}}(x) = \int_{D_y} f(x,y)dy$ is the the unconditional marginal density of x. Or equivalently,

$$f(x,y) = f(y|x)f_{\tilde{x}}(x)$$

Replacing this expression for $x, y, x'y'$ in (6), we get

$$\frac{f(y'|x)f_{\tilde{x}}(x)}{f(y|x)f_{\tilde{x}}(x)} \leq \frac{f(y'|x')f_{\tilde{x}}(x')}{f(y|x')f_{\tilde{x}}(x')}$$

simplifying and rearranging

$$\frac{f(y|x')}{f(y|x)} \leq \frac{f(y'|x')}{f(y'|x)}$$

Because this is true for all $y' \geq y$ and $x' > x$, we can conclude that the likelihood ratio

$$\frac{f(\cdot|x')}{f(\cdot|x)}$$

is increasing and hence $F(\cdot|x')$ dominates $F(\cdot|x)$ in terms of the likelihood ratio. As we have seen in the section on stochastic order, this dominance implies the other forms of stochastic dominance. Q.E.D.

Affiliated random variable also satisfy a number of other properties that we will not prove here. These properties are:

Proposition 4 *Let $\tilde{x} = \{\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_N\}$ be affiliated random variables, then*

1. *If $\tilde{x}$ and $\tilde{y}$ are affiliated then for any $x' \geq x$, on has that $F(y|x')$ dominates $F(y|x)$ both in terms of the hazard rate rate and in terms of the reverse hazard rate.*

2. *$E[\tilde{x}_i|\tilde{x}_j = x_j]$ is an increasing function of x_j.*

3. *If γ is an increasing function from D to $\mathbf{R}$, then*

$$E[\gamma(\tilde{x})|\tilde{x}_1 \leq x_1, \tilde{x}_2 \leq x_2 \ldots, \tilde{x}_N \leq x_n]$$

Is an increasing function of $x_1, x_2, \ldots, x_N$

4. *Let* $b_1(\cdot), b_2(\cdot), \ldots, b_N(\cdot)$ *strictly increasing function. Then* $b_1(\tilde{x}_1), b_2(\tilde{x}_2), \ldots, b_N(\tilde{x}_N)$ *are affiliated random variables.*

5. *Fix* $\tilde{x}_1$ *and let* $\tilde{y}_1, \tilde{y}_2, \tilde{y}_{N-1}$ *denote the highest, second highest and so on up to the* $(N-1)$*-th highest realization of* $\tilde{x}_2, \tilde{x}_2, \ldots, \tilde{x}_N$*. Then* $\tilde{x}_1, \tilde{y}_1, \tilde{y}_2, \tilde{y}_{N-1}$ *are affiliated random variables.*