The Cumulative Distribution and Stochastic Dominance

LaTeX file: *StochasticDominance* — Daniel A. Graham, September 1, 2011

A decision problem under uncertainty is frequently cast as the problem of choosing among a set of stochastic variables. Not much is lost if these stochastic variables are interpreted as monetary gains with (cumulative) distribution functions $F_i(x)$, $i = 1, 2, \cdots$. Such distributions are commonly called *lotteries* or *gambles*. Lotteries are to the analysis of uncertainty what commodity bundles are to consumption theory.

Consider, for example, the decision of a person with an income of \$100 of whether or not to accept a \$10 bet on the toss of a fair coin. This decision corresponds to the choice between the following two lotteries:

$$\text{Reject Bet: } F_1(x) \equiv \begin{cases} 0 & \text{if } x < 100 \\ 1 & \text{if } 100 \le x \end{cases}$$

$$\text{Accept Bet: } F_2(x) \equiv \begin{cases} 0 & \text{if } x < 90 \\ 1/2 & \text{if } 90 \le x < 110 \\ 1 & \text{if } 110 \le x \end{cases}$$

Our initial interest focuses upon two questions. (i) When is one lottery unambiguously better than another? (ii) When is one lottery unambiguously less risky than another? As will be shown, the answers involve, respectively, the statistical concepts of first and second order *stochastic dominance*. The cumulative distribution is the key to understanding both concepts.

The Cumulative Distribution

The best way to visualize a lottery is by considering the graph of the corresponding *cumulative distribution*. Continuing the coin-toss example, the graphs of the cumulative distribution functions are as follows:
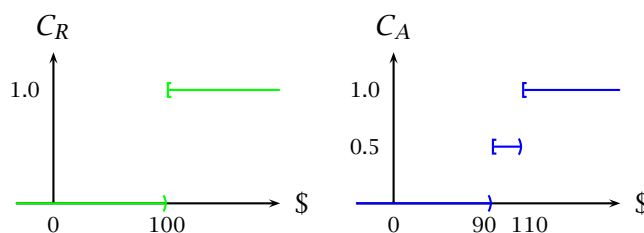
The Mean or Expected Value

The *mean* or *expected value* of a lottery is just a weighted average of the prizes using the probabilities of the prizes as weights. For our coin-toss example, the expected values are as follows:

$$E(L_R) \equiv 1 \times \$100 = \$100$$
$$E(L_A) \equiv 1/2 \times \$90 + 1/2 \times \$110 = \$100$$

Since the expected values are the same, the gamble is called a *fair bet*.

The expected values can also be determined from the graph of the cumulative distribution as the area to the left of the cumulative.[1]
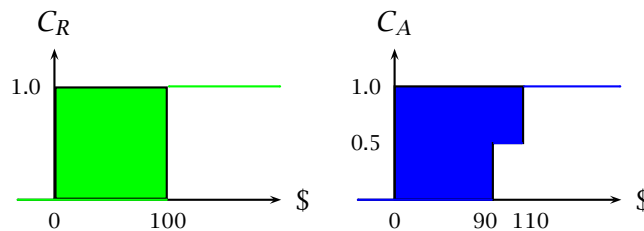
The fact that the lotteries have the same expected value is reflected by the fact that the shaded areas are equal.

Better Lotteries

When is one lottery necessarily better than another lottery? Consider the two lotteries whose cumulatives are illustrated below.

---
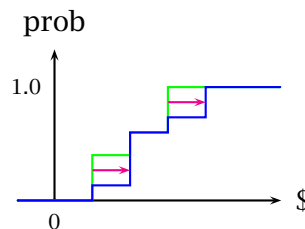
[1]This also holds for continuous distributions. Let $F(x)$ denote the cumulative, $f(x) = F'(x)$ the associated density and suppose that $f(x) > 0$ for $x \in (a, b)$ where $[a, b]$ is the support of the distribution. Now define $y \equiv F(x)$. Then $x = F^{-1}(y)$ since $f(x) = F'(x) > 0 \in (a, b)$ means that the inverse exists. Finally note that since $dy = f(x)dx$ we can write $F^{-1}(y)dy = xf(x)dx$ and thus the expected value of x is

$$E_F(x) = \int_a^b xf(x)\,dx = \int_0^1 F^{-1}(y)\,dy$$

If $b > a \geq 0$, the latter integral is just the area between the (vertical) y-axis and the cumulative. In general, when the support includes both positive and negative parts, the expected value is: (positive part) plus the area to the right of the y-axis and to the left of the cumulative; (negative part) minus the area to the left of the y-axis and to the right of the cumulative.

Note that the blue cumulative lies on or to the right of the green cumulative *everywhere* and strictly to the right over at least one interval. When this is true the blue cumulative is said to *first-order stochastically dominate* the green cumulative. When this is true, *anyone* who prefers larger money prizes to smaller ones, will prefer the blue cumulative. Why? Look at the magenta arrows and note that the blue cumulative entails taking probability that the green cumulative attached to a smaller prize (the vertical green segment to the left of an arrow) and reassigning this probability to a larger prize (the vertical blue segment to the right of the arrow).

Definition 1 (First-Order Stochastic Dominance). A cumulative distribution F first-order stochastically dominates another distribution G iff

$$F(x) \leq G(x)$$

for all x with a strict inequality over some interval.

Note also that if F first-order stochastically dominates G then F necessarily has a strictly larger expected value, $E_F(x) > E_G(x)$. The implication doesn't go the other way though. Just because one distribution has a larger expected value than another doesn't mean that the first stochastically dominates the second — see if you can construct a counter example — and it doesn't mean that anyone would necessarily prefer the distribution with the larger expected value. More on this later.

Less Risky Lotteries

When is one lottery less risky than another? The coin-toss example is again illustrated below, this time with both the green (reject) and blue (accept) cumulatives drawn in the same graph.
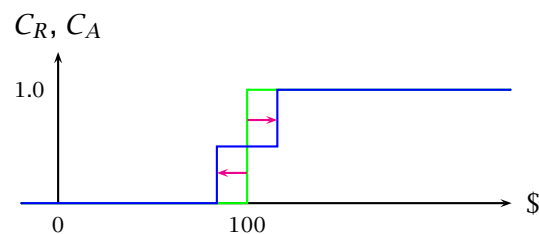
The magenta arrows emphasize the relationship between these cumulatives — moving from green to blue shifts probability of 1/2 from getting $100 to getting $90 and shifts probability of 1/2 from getting $100 to getting $110. This shift of probability away from the center of the distribution and into the tails while simultaneously preserving the mean (leaving the expected value unchanged) clearly increases the risk — accepting the bet clearly entails a riskier lottery than rejecting the bet.

A lottery with cumulative distribution F is called a *mean preserving spread* of a lottery with cumulative distribution G if F and G (1) have the same mean (expected value), (2) the graph of F crosses the graph of G exactly once, (3) F lies on or above G to the left of the crossing

point and F on or below G to the right of the crossing point and (4) F and G are not the same distribution, i.e., they differ on some non-null interval.

Note that accepting the bet is a *mean preserving spread* of rejecting the bet.

Now let's extend the coin toss example to include the possibility of accepting two bets:

$$\text{Accept Two Bets: } C_{AA}(x) \equiv \begin{cases} 0 & x < 80 \\ 1/4 & 80 \le x < 100 \\ 3/4 & 100 \le x < 120 \\ 1 & 120 \le x \end{cases}$$

The cumulative for accepting two bets (orange), is added to the reject (green) and accept one bet (blue) cumulatives below.
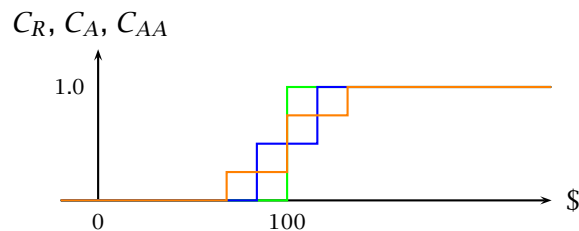
Note that moving from rejecting (green) to accepting two bets (orange) shifts probability of 1/2 away form getting \$100, 1/4 to getting \$80 (lost both bets) and 1/4 to \$120 (won both bets). Thus accepting two bets is a mean preserving spread of rejecting.

Moving from accepting one bet (blue) to accepting two bets (orange) involves a similar sort of spreading but is not a mean preserving spread since the cumulatives cross twice. Suppose, though, that we break the move from accepting one bet to accepting two bets into two steps.

1. Accept the second bet if and only if the first bet entails a loss. The cumulative for the resulting distribution would coincide with the two bet (orange) cumulative to the left of 100 and would coincide with the one bet (blue) cumulative to the right. Note that this is a mean preserving spread of accepting only one bet.

2. Accept the second bet whether or not the first bet entails a loss. The cumulative for this distribution coincides with the cumulative for "accept second bet if loss on first" to the left of 100 and coincides with accept two bets to the right. Note that this is a mean preserving spread of 'accept second bet if loss on first'.

Thus while accepting two bets is not a mean preserving spread of accepting one bet, there is a sequence of mean preserving spreads leading from the latter to the former.

The generalization of this idea is *second-order stochastic domination*.

Definition 2 (Second-Order Stochastic Domination). A cumulative distribution F second-order stochastically dominates another distribution G iff

$$\int_{-\infty}^{c} F(x)\, dx \le \int_{-\infty}^{c} G(x)\, dx$$

for all c with a strict inequality over some interval.

If F second-order stochastically dominates G then $E_F(x) \ge E_G(x)$. They would have the same expected value if, in the definition above, the equality held for all $c \ge c^*$ for some sufficiently large c^*.

Check this out for reject (green) versus accept one bet (blue). For $c < 90$ and for $c \ge 110$ the areas (integrals) are the same but for $90 \le c < 100$, the area under green is strictly less than the area under blue. Thus reject second-order stochastically dominates one bet.

What about accepting one bet (blue) versus accepting two bets (orange)? For $c < 80$, for $c = 100$ and for $c \ge 120$, the areas under blue and orange are the same. For $80 \le c < 100$ and for $100 \le c < 120$, the area under blue is strictly less than the area under orange. Thus accepting one bet second-order stochastically dominates accepting two bets.

The Ranking Theorems

There are two important connections between qualitative properties of the utility function and stochastic dominance.

Proposition 1 (First-Order Stochastic Ranking). If $u(x)$ is strictly increasing and piecewise differentiable, and cumulative F *first-order* stochastically dominates cumulative G, then

$$E_F[u(x)] > E_G[u(x)]$$

Additionally, if neither F nor G first-order stochastically dominates the other, then there exist strictly increasing and piecewise differentiable functions, $u(x)$ and $v(x)$, such that

$$E_F[u(x)] > E_G[u(x)]$$
$$E_F[v(x)] < E_G[v(x)]$$

i.e, *all* "normal" people will strictly prefer F to G iff F first-order stochastically dominates G.

Proposition 2 (Second-Order Stochastic Ranking). If $u(x)$ is strictly increasing, twice piecewise differentiable and concave with the concavity being somewhere strict, and cumulative F *second-order* stochastically dominates cumulative G, then

$$E_F[u(x)] > E_G[u(x)]$$

Additionally, if neither F nor G second-order stochastically dominates the other, then there exist strictly increasing, twice piecewise differentiable and concave functions with the concavity somewhere strict, $u(x)$ and $v(x)$, such that

$$E_F[u(x)] > E_G[u(x)]$$
$$E_F[v(x)] < E_G[v(x)]$$

i.e., *all* "risk-averse" people will strictly prefer F to G iff F second-order stochastically dominates G.

The Equivalence Theorem

Consider the following, apparently distinct, definitions of what it means for a distribution G to be at least as *risky* as a distribution F with the same mean.

1. Risk is what risk averse people dislike, i.e., it is not possible to find a risk averse person who prefers G to F. More precisely, for *any* concave utility function U, the expected utility of G is not greater than expected utility of F.

$$\int U(x)dG(x) \not> \int U(x)dF(x)$$

2. Risk corresponds to adding noise, i.e., F and G are cumulative distributions corresponding to the random variables x and $x + \epsilon$ where $E[\epsilon|x] = 0$ for all x.

3. G can be obtained from F by a finite sequence or, as the limit of an infinite sequence of mean preserving spreads.

4. F second-order stochastically dominates G.

Rothschild and Stigliz[2] have shown that these definitions are *equivalent*.

---

[2]Rothschild, M. and Stiglitz, J. "Increasing Risk: A Definition", *Journal of Economic Theory*, 1970, 2, 225-43.