

General structure

Here we suppose that the consequences are wealth amounts denoted by W , which can take on any value between a and b . Thus $[a, b]$ is the maximal support of all the probability distributions we will consider. (In practice a may be zero if there is limited liability, but could become negative; the upside potential could be unbounded. Thus we may want to allow $a = -\infty$ and $b = \infty$. This raises some technical complications which we ignore.)

Thus we have a new feature: W can be a continuous random variable, and a lottery is then a probability distribution over $[a, b]$. Consider an expected-utility maximizer with a utility-of-consequences function $u(W)$, evaluating particular lottery with a cumulative distribution function $F(W)$ and a density function $f(W)$. The expected utility is

$$u(L) = \int_a^b u(W) f(W) dW.$$

(If you find integrals difficult, try writing out the sums for a discrete distribution taking on values $W_1 < W_2 < \dots < W_n$, probabilities f_i , and cumulative probabilities

$$F_i = \sum_{j=1}^i f_j.$$

But actually that is algebraically more messy.)

Once the u -function is known, the expected utility can be evaluated for any lottery, and any pair of lotteries can be compared. But we are often interested in making broader comparisons that are valid for very general classes of utility functions, that is, in statements such as “any person whose u -function satisfies such and such a property will prefer lottery L^1 to lottery L^2 .” There are two especially important questions of this kind:

[1] What kind of relationship must exist between lotteries L^1 and L^2 , or equivalently, between their CDFs F_1 and F_2 , to ensure that everyone, regardless of his attitude to risk, will prefer L^1 to L^2 (so long as he likes more wealth to less). In other words, to ensure that for *all increasing* utility functions $u(W)$,

$$u(L^1) \equiv \int_a^b u(W) f_1(W) dW > \int_a^b u(W) f_2(W) dW \equiv u(L^2).$$

The relation of F_1 to F_2 that ensures this property is called *first-order stochastic dominance* (FOSD).

[2] What kind of relationship must exist between lotteries L^1 and L^2 , or equivalently, between their CDFs F_1 and F_2 , to ensure that every risk-averse person will prefer L^1 to L^2 . In other words, to ensure that for *all concave* utility functions $u(W)$,

$$u(L^1) \equiv \int_a^b u(W) f_1(W) dW > \int_a^b u(W) f_2(W) dW \equiv u(L^2).$$

The relation of F_1 to F_2 that ensures this property is called *second-order stochastic dominance* (SOSD).

Comments: (1) For arbitrary L^1 and L^2 , neither may dominate the other in either of these senses. In other words, both of these dominance concepts are *partial* orderings of lotteries, not *complete* orderings. (2) Your first thought may be that first-order stochastic dominance should correspond to having a higher mean, and second-order stochastic dominance should correspond to having a lower variance, but neither of these intuitions is correct. You should be able to see after a moment's thought that a higher mean is not going to suffice for FOSD: F_1 may have a higher mean but so much higher variance than F_2 that someone with an increasing but concave utility function will like L^1 less than L^2 . But the point about variance and SOSD is more subtle. We will see counterexamples soon.

First-Order Stochastic Dominance

To find the mathematical concept of FOSD, begin by observing that for any distribution function F and its density function f , and for any utility function u

$$\begin{aligned} \int_a^b u(W) f(W) dW &= [u(W) F(W)]_a^b - \int_a^b u'(W) F(W) dW && \text{integrating by parts} \\ &= u(b) * 1 - u(a) * 0 - \int_a^b u'(W) F(W) dW \\ &= u(b) - \int_a^b u'(W) F(W) dW \end{aligned}$$

Therefore comparing expected utility under two different distributions (if they have different “supports” $[a, b]$, take the biggest and define the other density outside its support to be 0):

$$\begin{aligned} \int_a^b u(W) f_1(W) dW - \int_a^b u(W) f_2(W) dW &= - \int_a^b u'(W) F_1(W) dW + \int_a^b u'(W) F_2(W) dW \\ &= \int_a^b u'(W) [F_2(W) - F_1(W)] dW. \end{aligned}$$

We want this to be positive for all increasing functions u , that is, for all functions with $u'(W) > 0$ for all $W \in [a, b]$. That will obviously be true if $F_2(W) > F_1(W)$ for all $W \in (a, b)$. We cannot avoid equality at the endpoints a and b because $F_1(a) = F_2(a) = 0$ and $F_1(b) = F_2(b) = 1$. But that does not hurt the result. We can also allow equalities at isolated points within the interval. Again I leave more rigorous proofs to more advanced (graduate) courses.

But the converse is also true: if $F_2(W) < F_1(W)$ for any ranges of $W \in (a, b)$, then we will be able to construct a function u for which $u'(W)$ is very positive and significant, say equal to 1, in these ranges, and still positive but very small, say some ϵ , everywhere else. By choosing ϵ small enough, we can then make the integral on the right hand side negative. This counterexample shows that $F_1(W) < F_2(W)$ for all $W \in (a, b)$ is also necessary for $u(L^1) > u(L^2)$ to be true for *all* increasing u functions.

So we have the desired result if, and only if, $F_2(W) > F_1(W)$ for all $W \in (a, b)$. We might make this our definition of FOSD, and state the result proved in the above argument as a theorem:

Definition 1: The distribution F_1 is first-order stochastic dominant over F_2 if and only if $F_1(W) < F_2(W)$ for all $W \in (a, b)$.

Theorem 1: Every expected-utility maximizer with an increasing utility function of wealth prefers the lottery L^1 with distribution F_1 to L^2 with distribution F_2 if and only if F_1 is FOSD over F_2 .

Figure 1 shows two such cumulative distribution functions. I have taken $a = 0$ and $b = 1$ for convenience. Since both distributions start out at 0 when $W = a = 0$, and then $F_1(W)$ becomes less than $F_2(W)$, it must be the case that $F_1(W)$ is flatter than $F_2(W)$ for small W . Since the slope of the cumulative distribution function is the probability density function, the density $f_1(W)$ must be less than the density $f_2(W)$ for small W . Conversely, both distributions climb to 1 at $W = b = 1$, but $F_1(W)$ climbs from smaller values than $F_2(W)$, it must be the case that $F_1(W)$ is steeper than $F_2(W)$ for W close to 1. That is, the density $f_1(W)$ is higher than the density $f_2(W)$ for W close to 1. Figure 2 shows the density functions.

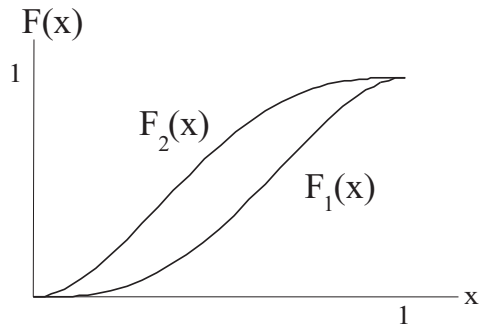


Figure 1: FOSD: CDF comparison

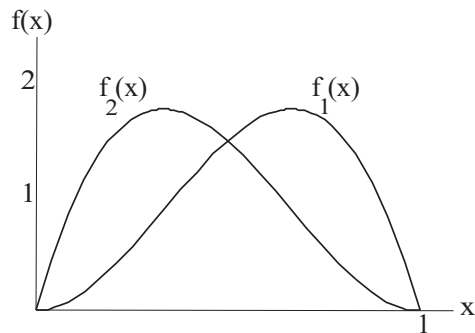


Figure 2: FOSD: Density comparison

The density comparison can be restated as saying: $f_1(W)$ can be obtained from $f_2(W)$

by shifting some probability weight from lower to higher values of W . That is an intuition for why anyone who prefers more wealth to less prefers the lottery L^1 to L^2 .

For your information, the functions are

$$\begin{aligned} f_1(W) &= 12 W^2 (1 - W), & f_2(W) &= 12 W (1 - W)^2, \\ F_1(W) &= 4 W^3 - 3 W^4, & F_2(W) &= 6 W^2 - 8 W^3 + 3 W^4. \end{aligned}$$

Second-Order Stochastic Dominance

The procedure is similar except that we have to integrate by parts twice. Write

$$S(W) = \int_a^W F(w) dw.$$

This is the integral of the cumulative distribution function, so its value at any point W is the area under the $F(w)$ curve for w going from a to W . And by the fundamental theorem of the calculus, the cumulative distribution function is its derivative: $F(W) = S'(W)$ for all W . Call the S function the “super-cumulative” distribution function for short. Note that $S(a) = 0$.

Now carry out the process of integration by parts twice:

$$\begin{aligned} \int_a^b u(W) f(W) dW &= [u(W) F(W)]_a^b - \int_a^b u'(W) F(W) dW && \text{integrating by parts} \\ &= u(b) * 1 - u(a) * 0 - \int_a^b u'(W) F(W) dW \\ &= u(b) - \int_a^b u'(W) F(W) dW \\ &= u(b) - [u'(W) S(W)]_a^b + \int_a^b u''(W) S(W) dW && \text{int. by parts again} \\ &= u(b) - u'(b) S(b) + \int_a^b u''(W) S(W) dW \end{aligned}$$

Therefore

$$\begin{aligned} &\int_a^b u(W) f_1(W) dW - \int_a^b u(W) f_2(W) dW \\ &= u'(b) [S_2(b) - S_1(b)] + \int_a^b u''(W) [S_1(W) - S_2(W)] dW \end{aligned}$$

We want our dominance definition to be such that every risk-averse person prefers L^1 and every risk-loving person prefers L^2 . Therefore we want the expression on the right hand side of the last line to be positive for all concave functions u , that is, for all functions with $u''(W) < 0$ for all $W \in [a, b]$; we also want the same expression to be negative for all convex functions, that is, for all functions with $u''(W) > 0$.

And therefore, on the borderline, we want a risk-neutral person to be indifferent between L^1 and L^2 . That is, the whole expression on the right hand side of the last line should be

zero if utility function satisfies $u''(W) = 0$ for all $W \in (a, b)$. But this condition reduces the integral on the right hand side to zero. So the first term $u'(b) [S_2(b) - S_1(b)]$ must be zero as well. Since $u'(b) \neq 0$, it must be that $S_2(b) = S_1(b)$. What does this condition signify?

For any distribution F , we have

$$\begin{aligned} S(b) &= \int_a^b F(W) dW = \int_a^b F(W) * 1 dW \\ &= [F(W) W]_a^b - \int_a^b f(W) W dW \quad \text{int. by parts} \\ &= F(b) b - F(a) a - \int_a^b f(W) W dW \\ &= b - E[W] \end{aligned}$$

Therefore, if the two distributions have $S_1(b) = S_2(b)$, then the expected values of W under the two distributions should also be equal, which we can write as $E_1(W) = E_2(W)$. In other words, the mean wealth should be the same under the two lotteries. This is a very natural condition to require when we want preference between them to depend only on the attitudes toward risk.

So we are left with

$$\int_a^b u(W) f_1(W) dW - \int_a^b u(W) f_2(W) dW = \int_a^b u''(W) [S_1(W) - S_2(W)] dW.$$

Proceeding exactly as the case of first-order stochastic dominance, we see that the right hand side will be positive for all utility functions satisfying $u'' < 0$ if, and only if, $S_1(W) < S_2(W)$ for all $W \in (a, b)$, and $S_1(b) = S_2(b)$.

So we have our definition of second-order stochastic dominance (SOSD), and a theorem:

Definition 2: The distribution F_1 is second-order stochastic dominant over F_2 if and only if $S_1(W) < S_2(W)$ for all $W \in (a, b)$ and $S_1(b) = S_2(b)$, where

$$S(W) = \int_a^W F(w) dw.$$

Recalling that the end-point equality of the S -functions imposes equal means, and F_2 is going to be riskier than F_1 , we also say that F_2 is a *mean-preserving spread* of F_1 .

Theorem 2: Every expected-utility maximizer with a concave utility function of wealth prefers the lottery L^1 with distribution F_1 to L^2 with distribution F_2 if and only if F_1 is SOSD over F_2 .

Actually the result is even more general. In the whole argument we nowhere used the sign of u' . So all we needed was a concave function, whether it be increasing or decreasing. So we have actually proved a more general

Theorem 2': The inequality

$$\int_a^b u(W) f_1(W) dW > \int_a^b u(W) f_2(W) dW$$

holds for *every* concave function u if, and only if, F_1 is SOSD over F_2 .

Figure 3 shows such a comparison of two S functions. But S functions are not immediately amenable to intuition, so we want to see how the underlying F or f functions compare. Figures 4 and 5 show them.

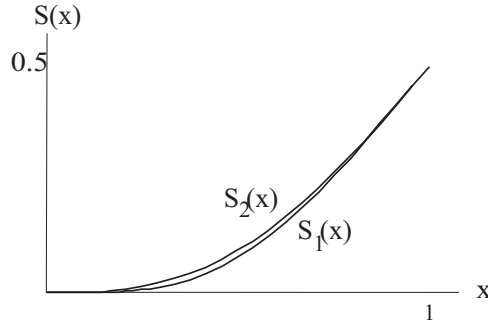


Figure 3: SOSD: S -function comparison

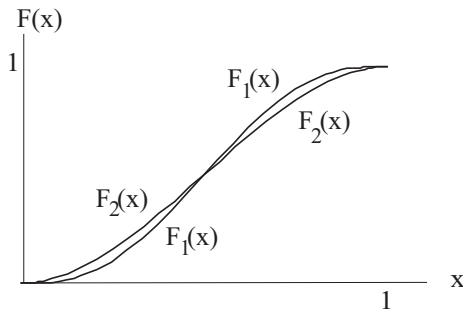


Figure 4: SOSD: CDF comparison

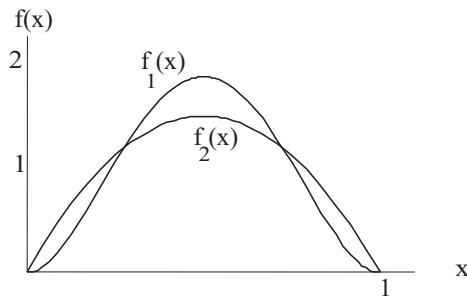


Figure 5: SOSD: density comparison

For your information, the functions are

$$\begin{aligned}
 f_1(W) &= 30 W^2 (1 - W)^2, & f_2(W) &= 6 W (1 - W), \\
 F_1(W) &= 10 W^3 - 15 W^4 + 6 W^5, & F_2(W) &= 3 W^2 - 2 W^3, \\
 S_1(W) &= 2.5 W^4 - 3 W^5 + W^6, & S_2(W) &= W^3 - 0.5 W^4.
 \end{aligned}$$

We can now interpret the differences between the two distributions. It helps to compare the simpler but somewhat similar analysis in the case of first-order stochastic dominance. There the cumulative $F_1(W)$ was entirely below $F_2(W)$, and the densities $f_i(W)$ were the derivatives (slopes) of the cumulatives $F_i(W)$; therefore $f_1(W)$ was a rightward shift of $f_2(W)$. (See Figures 1 and 2.) Here, the super-cumulative $S_1(W)$ is entirely below $S_2(W)$, and the cumulatives (CDFs) $F_i(W)$ are the derivatives (slopes) of the super-cumulatives $S_i(W)$; therefore $F_1(W)$ is a rightward shift of $F_2(W)$. But the $F_i(W)$ are increasing functions, unlike the densities $f_i(W)$; contrast Figures 4 and 2. Thus $F_1(W)$ starts out flatter than $F_2(W)$ (is smaller for smaller W values), then crosses it (is larger for larger W values), but must flatten out again as the two approach the common value 1 for $W = b$. In other words, $F_1(W)$ is flatter for small and large values of W , and is steeper in the middle. The density functions f_i are the slopes of the cumulatives F_i ; therefore the comparison of densities is that $f_1(W)$ is smaller than $f_2(W)$ for small and large values of W , and is larger in the middle. This is shown in Figure 5. In other words, the density $f_1(W)$ has a higher peak, and smaller tails, than the density $f_2(W)$. The distribution of the outcomes of lottery L^2 is more spread out than that of the outcomes of L^1 . That also explains why the switch from $f_1(W)$ to $f_2(W)$ is called a *mean-preserving* (because $S_1(b) = S_2(b)$) *spread*. And it is quite a natural reflection of the idea we want to capture, namely that *all* risk-averse persons prefer L^1 to L^2 , or that L^2 is in some uniform sense riskier than L^1 .

Being a mean-preserving spread implies having a higher variance, but not conversely. To prove the implication, just consider the function

$$\phi(W) = (W - \bar{W})^2$$

where \bar{W} is the expectation of W , the same under the two distributions (remember mean-preserving). The function ϕ is convex. If F_1 is SOSD over F_2 , our general Theorem 2' gives

$$\int_a^b \phi(W) f_1(W) dW < \int_a^b \phi(W) f_2(W) dW.$$

But each side is just the variance of W under the respective distribution.

Actually these pictures show only the simplest way in which the condition in the theorem can be fulfilled. Figure 3 remains the true story, but the Figures 4 and 5 derived from it can have more complicated shapes. In a more general Figure 4, the graphs of $F_1(W)$ and $F_2(W)$ can intersect multiple times. $F_1(W)$ must still start out flatter and end up flatter, so the number of intersections must be odd. And because $F_1(W)$ starts out flatter, the area under it (its integral $S_1(W)$) starts out smaller than the corresponding area for $F_2(W)$. In subsequent intervals where $F_1(W) > F_2(W)$, the gap between $S_2(W)$ and $S_2(W)$ starts to close. (And in other intervals where $F_1(W) < F_2(W)$ again, the gap widens again.) But the requirement of SOSD is that the gap should never close entirely until we reach the right hand end-point b . These figures are difficult to draw; see Fig. 3.7 on p. 111 of the book by Hirshleifer and Riley that is cited as occasional reading in the syllabus.

The ideas generated by Figure 5 can be formulated into alternative (and equivalent) definitions of SOSD:

Definition 2(a): Let W_i denote the random variable following distribution F_i , for $i = 1, 2$, and $E_1(W_1) = E_2(W_2)$. Then F_1 is said to be SOSD over F_2 if there exists a random variable z with zero expectation conditional on any given value of W_1 , such that w_2 has the same distribution as $w_1 + z$, or in other words, w_2 equals W_1 plus some added pure uncertainty or “noise.”

Definition 2(b): Of two distributions yielding equal expected values, F_1 is said to be SOSD over F_2 if it is possible to get from F_1 to F_2 by a sequence of operations which shift pairs of probability weights on either side of the mean farther away, while leaving the mean unchanged.

The “adding noise” idea is simple and we will use it occasionally. We won’t need the details of how to construct basic mean-preserving spreads so I omit them; if you want to, you can find them in the Rothschild-Stiglitz article pp. 233–237, reprinted in the Diamond-Rothschild book pp. 103–107.

But it is very important to understand how and why higher variance does not ensure SOSD. Since we want to show that it is possible to have an instance where L^2 has higher variance than L^1 but a risk-averse person prefers L^2 to L^1 , so F^1 is not SOSD over F^2 , we need only construct an example. There are various examples of this kind in the literature. Here is a relatively simple and comprehensible one:

The two lotteries, defined by their vectors of wealth consequences and probabilities, are:

$$L^1 = (0, 2, 4, 6; 1/4, 1/4, 1/4, 1/4)$$

$$L^2 = (0.1, 3, 5.9; 1/3, 1/3, 1/3)$$

It is easy to calculate that the means and variances are

$$L^1: \text{Mean} = 3, \text{Variance} = 5.000$$

$$L^2: \text{Mean} = 3, \text{Variance} = 5.607$$

Take the utility function

$$u(W) = \begin{cases} 2W & \text{if } W \leq 3 \\ 3 + W & \text{if } W > 3 \end{cases}$$

Figure 6 shows the two lotteries and Figure 7 shows the utility function:

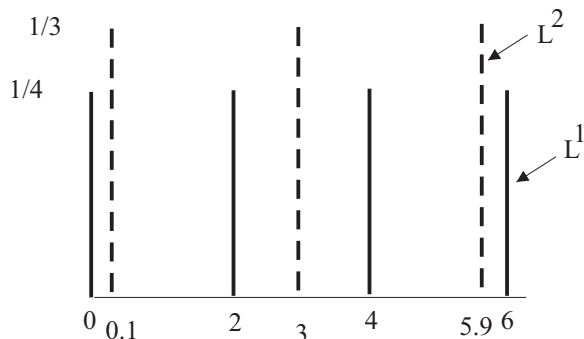


Figure 6: Lotteries in example for SOSD \neq lower variance

This gives expected utilities:

$$EU(L^1) = 5.000, \quad EU(L^2) = 5.033$$

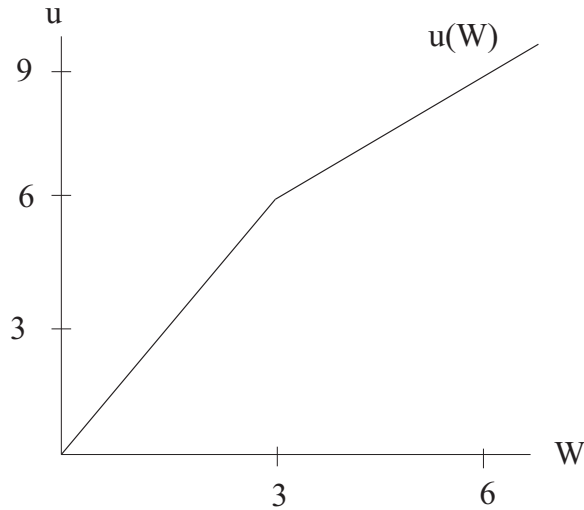


Figure 7: Utility in example for SOSD *neq* low variance

so the person prefers L^2 despite the fact that it has a higher variance than L^1 and the same mean.

Note that the utility function is a straight line in the regions below and above $W = 3$, but has a kink at 3. In other words, the concavity (risk-aversion) arises solely because the person dislikes downside risk more than he values upside potential. This is not unusual; in fact it is one aspect of actual behavior emphasized by many critics of the usual smooth expected utility, and we will pick up on this idea later. But you can get another example by making a very slight change to the u function to make it smooth and strictly concave everywhere; you just need a rapid change of slope in a small interval around $W = 3$.

Now consider the expected *absolute* deviations, that is, $E(|W - \bar{W}|)$ for the two distributions: for L^1 it is 2, and for L^2 it is 1.933. Therefore L^1 has a larger absolute deviation, even though it has a smaller variance. And our utility function, which has a constant slope below the mean and another smaller constant slope above the mean, can be expressed as if the person dislikes absolute deviation. We can write

$$u(W) = \begin{cases} 1.5 + 1.5 W - 0.5 (3 - W) & \text{if } W \leq 3, \\ 1.5 + 1.5 W - 0.5 (W - 3) & \text{if } W > 3, \end{cases}$$

or

$$u(W) = 1.5 + 1.5 W - 0.5 |W - 3|.$$

Therefore

$$EU(W) = 1.5 + 1.5 E(W) - 0.5 AbsDev(W)$$

where $AbsDev(W)$ denotes the absolute deviation of W . So it is natural that the lotteries are ranked by absolute deviation rather than by variance. And once you understand this idea, you can get similar counterexamples by constructing utility functions that involve some other measure of central tendency than the variance.

And how does this relate to the definition of SOSD? For that, we need to compare the super-cumulative functions for the two distributions. A little work shows yields Figure 6. I have shown S_1 thicker, and have shown the points 0.1 and 5.9 out of scale for clarity of appearance. We see that the two super-cumulatives S_1 and S_2 cross, so F_1 is not SOSD over F_2 .

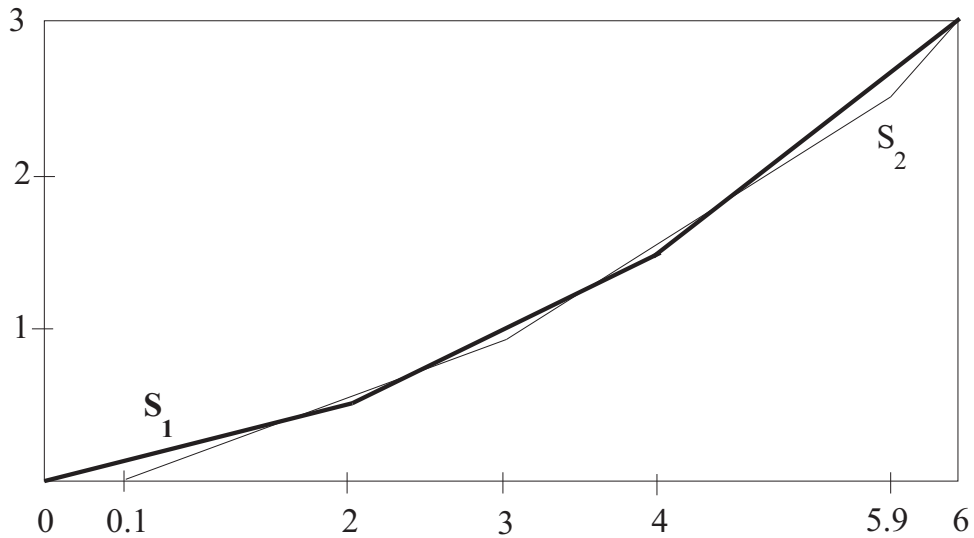


Figure 8: SOSD vs. variance comparison

Another way to look at this is that SOSD is a kind of comprehensive requirement that requires every conceivable measure of central tendency under one distribution to be smaller than that under the other. The variance is only one such measure; therefore it being smaller does not ensure that the person prefers that distribution if he actually cares about some other measure.