The relation of $F_1$ to $F_2$ that ensures this property is called \textit{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 \textit{partial} orderings of lotteries, not \textit{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.

\section*{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 $u$

\[
\int_a^b u(W) f(W) dW = \left[ u(W) F(W) \right]_a^b - \int_a^b u'(W) F(W) dW \quad \text{integrating by parts}
\]

\[
= u(b) * 1 - u(a) * 0 - \int_a^b u'(W) F(W) dW
\]

\[
= u(b) - \int_a^b u'(W) F(W) dW
\]

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):

\[
\int_a^b u(W) f_1(W) dW - \int_a^b u(W) f_2(W) dW = - \int_a^b u'(W) F_1(W) dW + \int_a^b u'(W) F_2(W) dW
\]

\[
= \int_a^b u'(W) \left[ F_2(W) - F_1(W) \right] dW.
\]

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 $\epsilon$, everywhere else. By choosing $\epsilon$ 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 \textit{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](image1)

![Figure 2: FOSD: Density comparison](image2)

The density comparison can be restated as saying: \( f_1(W) \) can be obtained from \( f_2(W) \).
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](image1.png)

![Figure 2: FOSD: Density comparison](image2.png)

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

$$f_1(W) = 12W^2 (1 - W), \quad f_2(W) = 12W (1 - W)^2,$$

$$F_1(W) = 4W^3 - 3W^4, \quad F_2(W) = 6W^2 - 8W^3 + 3W^4.$$

**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:

$$\int_a^b u(W) f(W) \, dW = \left[ u(W) F(W) \right]_a^b - \int_a^b u'(W) F(W) \, dW$$

integrating by parts

$$= u(b) \ast 1 - u(a) \ast 0 - \int_a^b u'(W) F(W) \, dW$$

$$= u(b) - \int_a^b u'(W) F(W) \, dW$$

$$= u(b) - \left[ u'(W) S(W) \right]_a^b + \int_a^b u''(W) S(W) \, dW$$

int. by parts again

$$= u(b) - u'(b) S(b) + \int_a^b u''(W) S(W) \, dW$$

Therefore

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

$$= u'(b) [S_2(b) - S_1(b)] + \int_a^b u''(W) [S_1(W) - S_2(W)] \, dW$$

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

\[
S(b) = \int_a^b F(W) \, dW = \int_a^b F(W) \times 1 \, dW \\
= \left[ F(W) W \right]^b_a - \int_a^b f(W) W \, dW \quad \text{int. by parts}
\]

\[
= F(b) b - F(a) a - \int_a^b f(W) W \, dW \\
= b - E[W]
\]

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](image)

![Figure 4: SOSD: CDF comparison](image)

![Figure 5: SOSD: density comparison](image)

For your information, the functions are

$$f_1(W) = 30 W^2 (1 - W)^2, \quad f_2(W) = 6 W (1 - W),$$

$$F_1(W) = 10 W^3 - 15 W^4 + 6 W^5, \quad F_2(W) = 3 W^2 - 2 W^3,$$

$$S_1(W) = 2.5 W^4 - 3 W^5 + W^6, \quad S_2(W) = W^3 - 0.5 W^4.$$
**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

$L^1$: Mean = 3, Variance = 5.000
$L^2$: Mean = 3, 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 ≠ lower variance](image)

This gives expected utilities:

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

$L^1$: Mean = 3, Variance = 5.000
$L^2$: Mean = 3, 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 ≠ lower variance](image)

This gives expected utilities:

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

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.