Abstract

It has been shown that if two probability distributions satisfy the monotone likelihood ratio property (MLRP), and are independently updated using common public information and traditional Bayesian updating, then the resulting two posterior distributions will also satisfy the MLRP. I discuss this result and extend it by characterizing the full set of updating operations which preserve the MLRP in this manner, of which Bayesian updating is an element. I also find the set of updating operations which preserve first order stochastic dominance (FOSD). The only operator which preserves both the MLRP and FOSD is that which ignores all new information.

JEL Classifications: D81, D83.

When do ordered prior distributions induce ordered posterior distributions?

Ben Klemens*

February 24, 2005

1 Introduction

First-order stochastic dominance (FOSD) is a useful property for a set of probability distributions to have. A lottery whose distribution FOSDs that of another lottery can be said to be 'better' or 'more optimistic',[1] in the sense that the expected utility given the the dominant lottery is higher than that of the dominated lottery. This statement assumes that a higher payoff is better, but assumes nothing about risk preferences, returns to scale, or anything else about the form of utility functions. See for example Milgrom [1981].

This result is directly used in Milgrom and Weber [1982] and a broad swatch of the auction theory literature based thereon. In the realm of social choice theory, Feddersen and Pesendorfer [1997] rely on the MLRP to ensure that beliefs are ordered in their work on information revelation through voting.

For some applications, FOSD can be used as a sanity check: the distribution of a stock's expected ask price should always FOSD the distribution of the expected bid, and as new information comes in, the updated ask distribution should still FOSD the updated bid's distribution. In other applications, it is a useful assumption: one bidder may have a higher expected value for a good than another, which we can model by saying that the first bidder's distribution of the good's expected value FOSDs that of the second.

What sort of information from the auctioneer would or would not affect this ordering?

If the prior distributions satisfy the monotone likelihood ratio property (the MLRP, defined below), then the posterior distributions after Bayesian updating with common information will also satisfy the MLRP, regardless of what common information is used to update. Since any pair of distributions which satisfy

*Thanks to Kim Border, Peter Bossaerts, Matt Jackson, Richard McKelvey, and Bob Sherman.

[1]The phrase 'f FOSDs g' may be read as 'f first-order stochastically dominates g'. The clarity of the abbreviation more than makes up for any æsthetic shortcomings it may have. Similarly, the reader may take 'f MLRPs g' to mean 'f satisfies the monotone likelihood ratio property with respect to g'.

the MLRP also satisfy FOSD, MLRP priors guarantee Bayesian posteriors which are ordered by FOSD.

This is the state of the literature. This result appears in a discrete form in Whitt [1979], in Milgrom [1981], and in some of the papers that they cite. Bikhchandani et al. [1992] do a similar thing with a simplified setup.

The debate over the use of Bayes's rule is a long-running one, though some authors brush it aside; to select a representative quote, Calvert [1985] claims that "[...] there is nothing magical about Bayes's rule that should cause us to believe, in advance, that a different rule would qualitatively change our conclusions about the rational use of [...] information." Section 4 shows that there *is* something magical about Bayes's rule, and that other reasonable methods of updating do not have the consistency property of preserving the MLRP for which Bayes's rule is frequently cited in the literature. If we believe that the MLRP should be preserved, then we believe that Bayesian updating is the correct method of aggregating information.

It is easy to give counterexamples to show that Bayesian updating does not guarantee posteriors ordered by FOSD given arbitrary priors ordered by FOSD—the priors need the more restrictive monotone likelihood ratio property. Theorem 8 describes the full set of updating operations which will guarantee FOSD posteriors given FOSD priors. This set of operators consists of those which take linear weighted averages of their inputs, and conspicuously excludes Bayesian updating from the set.

2 Preliminary notes

Say that there are two competing researchers, who both know the conditional distribution $f(x|t)$ (with $x, t \in \mathbb{R}$), but have private probability density functions for t, $g_1(t)$ and $g_2(t)$.[2] Assume that all functions have strictly positive values for all $x \in \mathbb{R}$ and $t \in \mathbb{R}$ (that is, they have full support).

Given a probability density function (PDF) $g_i(t)$ describing the parameter t, Bayes's rule tells us how to use $f(x|t)$ to update to posterior distributions:[3]

$$post_i(t|x) = \frac{f(x|t)g_i(t)}{\int_{\mathbb{R}} f(x|\tau)g_i(\tau)d\tau}. \tag{1}$$

Typically, t is a parameter such as a mean, and x is new data such as experimental observation. Let $\tilde{x}$ be the observed values of x; after $\tilde{x}$ is observed, it has a fixed value, so $post_i(t|\tilde{x})$ is a function of only t.

[2]Compare with Milgrom's 'Good news and bad news' setup, in which one person could receive two different types of signals. The symmetry of Bayesian updating indicates that both types of setup are identical save for the storyline.

[3]'Bayesian updating' means different things to different people. For example, Majumdar [2002] and Whitt [1979] prefer a form which is applicable to a finite number of states of the world. New information eliminates states of the world from consideration, and thus changes the probabilities placed on the other states. Here, I use a continuous form, and new information comes in the form of a likelihood function with full support.

That said, questions of whether the various functions $g_i(t)$ and $post_i(t|x)$ are ordered of course require some methods of ordering distributions.

2.1 Ways to order a family of distributions

Since this section gives general definitions, take $p(t)$ to be any PDF, mapping from values of $t \in \mathbb{R}$ to $\mathbb{R}^+$.

Often, we have a family of related PDFs. For example, each researcher may have beliefs which differ slightly from the beliefs of other researchers. Let $i \in \mathbb{R}$ be an index, and denote a member of the family of distributions as either $p_i(t)$ or, equivalently, $p(t|i)$. A family of PDFs is ordered when any two members $p_i(t)$ and $p_j(t)$ are ordered.

MLRP, the monotone likelihood ratio property: if $i > j$,

$$\frac{p(t|i)}{p(t|j)}$$

is an increasing function of t; if $i < j$ it is decreasing, and is constant (one) if $i = j$. Notice that these are strict inequalities.

Having two functions $p_1(x)$ and $p_2(x)$ is equivalent to having one function $p'(x|i)$ where $p'(x|1) = p_1(x)$ and $p'(x|2) = p_2(x)$. Therefore, to say '$g_1(t)$ MLRPs $g_2(t)$' means that $g_1(t)/g_2(t)$ is an increasing function of t.

FOSD, first-order stochastic dominance: $p_1(t)$ FOSDs $p_2(t)$ iff for any constant k:

$$\int_k^\infty p_1(t)dt > \int_k^\infty p_2(t)dt. \tag{2}$$

Alternatively, a family $p(\cdot|t)$ satisfies FOSD iff $p(\cdot|t_1)$ FOSDs $p(\cdot|t_2)$ for all $t_1 > t_2$.

single-crossing: $p_1(t) > p_2(t)$ for all t less than some point K, and $p_1(t) < p_2(t)$ for all $t > K$.

These terms allow the following definitions:

Definition 1 *Consider PDFs f, g_1, and g_2, all of which have full support over the reals. Then an updating operation* preserves FOSD *iff $g_1(t)$ FOSDs $g_2(t)$ implies that $post_1(t|x)$ FOSDs $post_2(t|x)$ for all likelihood functions $f(x|t)$ and any fixed x.*

Definition 2 Preserving the MLRP *is similarly defined.*

3 Bayesian updating preserves the MLRP

As discussed above, FOSD provides an ideal method of ordering distributions. But unfortunately, it is easy to construct examples which show that Bayesian updating does *not* preserve FOSD. That is, for any two PDFs such that $g_1(t)$ FOSDs $g_2(t)$, there exists a likelihood function $f(x|t)$ such that $post_1(t|x)$ does not FOSD $post_2(t|x)$—in fact, it is easy to construct examples where $post_2(t|x)$ FOSDs $post_1(t|x)$. This section shows how the problem can be surmounted using the MLRP.

Diagrammatically, Figure 1 shows the trick used in the literature to ensure that posteriors satisfy FOSD: MLRP priors imply MLRP posteriors, which in turn imply FOSD posteriors.

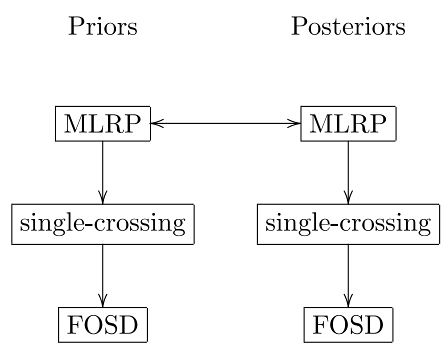

Figure 1: Relations to be proven in the sequel

3.1 Ordering posteriors with the MLRP

The validity of the arrows in Figure 1 are proven in the appendix, Section 6, via Lemma 11 through Lemma 13. But the main result is that if priors $g_1(t)$ and $g_2(t)$ are ordered via the MLRP, then the posteriors they induce will be ordered via FOSD. Formally,

Theorem 1 *MLRP priors (on t) $\Rightarrow$ FOSD posteriors on t for any fixed x, within the support of $f(x|t)$. That is, $g_1(t)$ MLRPs $g_2(t)$ implies $post_1(t|x)$ FOSDs $post_2(t|x)$ for any x.*

[This and other proofs not given in the body of the text are in the appendix.]

Theorem 1 gives a detailed description of an oft-used technique which is effectively the state of the art in the literature on auctions and mechanism design. The key point of interest from an economic perspective is FOSD, since it can be shown that if the probability distribution $g_1(x)$ FOSDS $g_2(x)$, and $U(x)$ is any monotonically increasing utility function, then, without any further assumptions about utility, the expected utility given the first PDF is larger than the expected utility given the second: $\int U(x)g_1(x)dx > \int U(x)g_2(x)dx$. But

FOSD is not preserved under Bayesian updating, so the only way to preserve the ordering of utility functions is via the MLRP. The following sections show that if we assume Bayesian updating, then the MLRP is the only type of ordering which is preserved; while if we are only interested in preserving FOSD, then we should instead prefer an averaging-type operator.

4 General updating operations

Say that a decision maker would like to know the likelihood of $x \in \mathbb{R}$, and toward that end two different authorities each present her with a different PDF for x, $f(x)$ and $g(x)$. The decision maker herself can only have one set of beliefs, so she must amalgamate $f(x)$ and $g(x)$ into one posterior PDF. For example, it may be reasonable for the decision maker to simply take the average of the two densities:

$$post(x) = \frac{f(x) + g(x)}{2}. \tag{3}$$

Now say that the second authority has two distributions to choose from, $g_1(t)$ and $g_2(t)$. The same question as above can be asked: if $g_1(t)$ MLRPs $g_2(t)$, then what sort of updating operations will preserve the MLRP? This section shows that the result that the Bayesian updating operation preserves the MLRP can be expanded to a class of updating operations which I dub 'weighted Bayesian rules'—but this class does not include many reasonable updating operations such as the averaging operator above. We are guaranteed that if an updating operation that is not in this class is used, then there is some information $f(\cdot)$ which will destroy the order of the posteriors, using the MLRP ordering. Meanwhile, the averaging operator of Equation 3 is a member of the class of functions which preserve FOSD, but which may destroy the MLRP ordering.

4.1 Definition of an updating operator

Let $\tau(x)$ be the true, unobservable PDF of x. Say that one source claims that for $x = 1$, $\tau(1) = a$, and another source claims $\tau(1) = b$. The decision maker must amalgamate these two data points into one belief, so let $op(a, b)$ be the updating operation, mapping $\mathbb{R}^+ \times \mathbb{R}^+$ to $\mathbb{R}^+$.

The same may be done for any value of x: let $a(x)$ be one source's claims about the true distribution of x, and let $b(x)$ be the other's. Then $op(a(x), b(x))$ defines an implicit function mapping x to $\mathbb{R}^+$. A few more caveats ensure that this will lead to a valid output given valid PDFs as inputs.

Definition 3 *An updating operator is any two-variable function $op(a, b) : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$, where $op(a, b) \geq 0$ for all $a \geq 0$ and $b \geq 0$; and if $\int_{-\infty}^{\infty} a(x)dx = 1$ and $\int_{-\infty}^{\infty} b(x)dx = 1$, then $\int_{-\infty}^{\infty} op(a(x), b(x))dx$ is finite.*

I will restrict attention to operators which are continuous in both variables. It will also be occasionally useful to restrict attention to those updating operators which are increasing in both arguments. That is, if $\alpha > \beta$, then

$op(\alpha, \gamma) > op(\beta, \gamma)$ and $op(\gamma, \alpha) > op(\gamma, \beta)$ for all γ. This is sensible because if one source modifies its report from $a(x_1) = \beta$ to $a(x_1) = \alpha$, thus putting more weight on x_1, a listener should take that into account when aggregating the information and put a little more weight on x_1 as well.

Although the discussion to this point has been oriented toward symmetric applications, the first and second inputs into the updating operation will have different interpretations for many of the applications below. Further, symmetric operators will be shown to have some additional implications over the general specification of Definition 3, where $op(a, b)$ may or may not equal $op(b, a)$.

There is no reason why $\int_{-\infty}^{\infty} op(f(x), g(x))dx$ should be one, meaning that the actual posterior requires normalization.

Definition 4 *The posterior distribution given updating operator $op(\cdot, \cdot)$ and inputs $f(\cdot)$ and $g(\cdot)$ is*

$$post(f(\cdot), g(\cdot), op(\cdot, \cdot), x) = \frac{op(f(x), g(x))}{\int_{-\infty}^{\infty} op(f(y), g(y))dy},$$

which is guaranteed to integrate to one.

In this notation, Bayesian updating is equivalent to $op(f(x|\cdot), g(\cdot)) = f(x|\cdot)g(\cdot)$, for any fixed x, giving a posterior exactly as described in Equation 1. For example, let x be an observable variable and t be a parameter of x such as the mean. A Bayesian researcher would begin with prior beliefs that t has a PDF $g(t)$, that the data has distribution $f(x|t)$. The researcher then runs an experiment which provides a value of x, χ. At this point, the researcher applies $op(f(\chi|t), g(t)) = f(\chi|t)g(t)$ to arrive at a posterior $post(t|\chi)$, which is a function of t but not x.

Shafer [1976] describes how this is equivalent to Dempster's rule within the probability framework here.

For a posterior which averages the priors, $op(f(\cdot), g(\cdot)) = f(\cdot)+g(\cdot)$, and the posterior is

$$post(f(\cdot), g(\cdot), op(\cdot, \cdot), x) = \frac{f(x) + g(x)}{\int_{-\infty}^{\infty} f(y) + g(y)dy} = \frac{f(x) + g(x)}{2}.$$

Finally, notice that if k is any positive constant,

$$\frac{k \cdot op(f(x), g(x))}{\int_{-\infty}^{\infty} k \cdot op(f(y), g(y))dy} = \frac{op(f(x), g(x))}{\int_{-\infty}^{\infty} op(f(y), g(y))dy}.$$

In other words, any statement below about the properties of $op(f(\cdot), g(\cdot), x)$ also applies to $op'(f(\cdot), g(\cdot), x) = k \cdot op(f(\cdot), g(\cdot), x)$. Having made this note here, I will state all results below ignoring the fact that the operators can trivially be multiplied by any positive constant.

4.2 Operators which preserve the MLRP

Recall from Definition 4 that the updating operator must be rescaled to be a true PDF. However, the following statement tells us that we can simplify the problem by ignoring the rescaling which converted the updating operator to a posterior.

Proposition 2 *The posterior distribution* $post(f, g_1, op, x)$ *MLRPs* $post(f, g_2, op, x)$ *iff the function* $op(f(x), g_1(x))$ *MLRPs the function* $op(f(x), g_2(x))$*.*

That said, we can consider arbitrary positive functions without regard to whether they integrate to one. Then:

Lemma 3 *A continuous operator preserves the MLRP iff it is of the form*

$$op(a, b) = t(a) \cdot b^p,$$

with $p > 0$*, and* $t(\cdot)$ *any transformation.*

The fact that the first argument may undergo any transformation is not surprising: an operator preserves the MLRP if $op(f(x), g_1(x))$ MLRPs $op(f(x), g_2(x))$ for any pair of gs which satisfy the MLRP and *any* density $f(x)$. So naturally, if $op(\cdot, \cdot)$ preserves the MLRP given $f(x)$, it also preserves the MLRP given $t(f(x))$.

More interesting would be to restrict ourselves to operators which preserve the MLRP symmetrically: $op(f, g_1)$ MLRPs $op(f, g_2)$, and also $op(g_1, f)$ MLRPs $op(g_2, f)$.

Theorem 4 *A continuous operator preserves the MLRP symmetrically iff it is of the form*

$$op(a, b) = a^p \cdot b^q,$$

with $p, q > 0$*.*

Proof: In this case, we have two conditions: $op(a, b) = t(a) \cdot b^q$ and $op(a, b) = a^p \cdot t(b)$. Any operator which satisfies both of these conditions must take the form $op(a, b) = a^p \cdot b^q$. $\diamondsuit$

That is, the MLRP is symmetrically preserved in all cases only when decision makers update based on a monomial operator. The Bayesian updating operator, $op(a, b) = a \cdot b$, is the special case where $p = q = 1$.

The exponents p and q allow the decision maker to place more or less weight on the distributions $f(\cdot)$ and $g(\cdot)$. Say that $g(x)$ is a single-peaked distribution; the convex transformation of squaring would exacerbate the peak, reducing the variance of the distribution. In the context of receiving information from an advisor, this means that $op(f, g, x) = f^1(x)g^2(x)$ places more weight on the second advisor's claim and puts more of the posterior density around the center of that advisor's distribution. In a similar manner, $op(f, g, x) = f^{1/2}(x)g^1(x)$ discounts the first advisor's advice. At the extreme, $op(f, g, x) = f^0(x)g^1(x)$ is a long way of writing $op(f, g, x) = g(x)$, ignoring the input of the function $f(\cdot)$ entirely. This justifies the use of the term 'weighted Bayes's rule' to describe this class of updating operators, where the weights are the exponents p and q.

But the set of weighted Bayes's rules is a small subset within the class of possible methods of updating, excluding such intuitive methods as averaging and convolution. If the reader does not believe that preserving the MLRP is essential, then he or she should take the discussion of Sections 2 and 3 with a grain of salt, since they describe only characteristics of Bayesian updating and slight variants thereof. But if the reader believes that the MLRP *should* be preserved by new information, then the above shows that the reader must also believe that people use Bayesian updating (possibly weighted) to assimilate new information.

4.3 Operators which preserve single-crossing

Here is a characterization of the set of updating operators which lead to posteriors ordered by FOSD:

Proposition 5 *Given any pair of priors g_1 and g_2 which satisfy single-crossing, and any function f, and an updating operator $op(f, g)$ which is monotonically increasing in f and g, the operator $op(\cdot, \cdot)$ provides posteriors ordered by FOSD iff the function is of the form*

$$op(f, g) = t(f) + qg,$$

where $t(\cdot)$ is a transformation function of any form, and q is any positive constant.

This also describes the set of updating operators which preserve single-crossing:

Lemma 6 *An updating operator $op(f, g)$ which is monotonically increasing in f and g preserves single-crossing iff the function is of the form*

$$op(f, g) = t(f) + qg,$$

where $t(\cdot)$ is a transformation function of any form, and q is any positive constant.

Notice that if $g_1(x) = g_2(x)$, then $t(f(x)) + qg_1(x) = t(f(x)) + qg_2(x)$, meaning that the point at which the priors cross is also the point at which the posteriors cross. That is, if decision makers use an updating operator which preserves single-crossing, no new news can move the point of crossing.

4.4 Operators which preserve FOSD

The set of operators which preserve FOSD matches the set of operators which preserve single-crossing.

Theorem 7 *Within the class of operators $op(f, g)$ which are monotonically increasing in both arguments, an operator preserves FOSD iff it is of the form $op(f, g) = t(f) + qg$, where $t(\cdot)$ is a transformation function of any form and q is a positive constant.*

Thus, the only updating method which preserved first-order stochastic dominance in all cases is that of taking the weighted mean of the two sets, perhaps transforming the first before knowing the second. Bayesian updating is not in this set.

As we did with the MLRP, it is worth considering the problem of symmetrically preserving FOSD as well.

Theorem 8 *Within the class of operators $op(f, g)$ which are monotonically increasing in both arguments, an operator symmetrically preserves FOSD iff it is of the form $op(f, g) = pf + qg$, where p and q are arbitrary positive weights.*

Proof: We require both $op(f, g) = t(f) + qg$ and $op(f, g) = pf + t'(g)$. The only operators which satisfy both of these conditions are those listed in the theorem. $\diamondsuit$

4.5 Preserving both the MLRP and FOSD

Both the MLRP and FOSD are preserved in the case of only one updating operator—the trivial operator where new information is ignored entirely.

Lemma 9 *The only updating operator which preserves both the MLRP and FOSD is:*

$$op(f(\cdot), g(\cdot), x) = g(x).$$

Proof: This is the only operator which satisfies both Lemmas 3 and 7.$\diamondsuit$

There is no function which symmetrically preserves both the MLRP and FOSD. The definitions given here are strict, but can be weakened: the ratio in the definition of MLRP must be increasing, which can be weakened to nondecreasing; and the comparison in the definition of FOSD may be weakened from 'less than' to 'less than or equal to'. Then the class of updating operators which symmetrically preserve both the weak MLRP and weak FOSD becomes a large one:

$$op(f(\cdot), g(\cdot), x) = h(x),$$

where $h(x)$ is any arbitrary PDF which depends on neither $f(x)$ nor $g(x)$.

5 Conclusion

This paper discussed the result that Bayesian updating preserves the monotone likelihood ratio property. This means that if a pair of priors satisfy the MLRP, and both are updated with the same likelihood function and the Bayesian updating operation, then the two posteriors will also satisfy the MLRP.

But (weighted) Bayesian updating is necessary and sufficient for all of the above results. If a modeler believes that the MLRP is always preserved when people update their priors, or that posterior medians should be ordered as discussed above, then the modeler must describe people as using Bayesian updating. If a modeler believes that people do not use Bayesian updating, then he must also believe that the ordering among people's beliefs can be destroyed by certain pieces of new information.

Similarly, the class of averaging operators preserves first-order stochastic dominance. A researcher who desires that FOSD be preserved must assume an averaging operator, and a researcher who does not assume an averaging operator must accept that the FOSD ordering will not always be preserved.

6 Appendix

This section gives proofs for those statements and theorems not proven above. It also states and proves Lemma 11 through Lemma 13, represented by the arrows in Figure 1, and presents Proposition 14 for use in later proofs.

6.1 Proof of Figure 1

Lemma 10 *Let $t \in \mathbb{R}$, and $p_1(t)$ and $p_2(t)$ be continuous PDFs. Define K to be the set of ts such that $p_1(t)/p_2(t) = 1$. If $p_1(t)$ MLRPs $p_2(t)$, then K is a single point. Also, $p_1(t)/p_2(t) < 1$ for all $t < K$ and $p_1(t)/p_2(t) > 1$ for all $t > K$. In other words, MLRP $\Rightarrow$single-crossing.*

Proof: If $K = \{\emptyset\}$, then it must be that either $p_1(t) > p_2(t)$ for all t or $p_1(t) < p_2(t)$ for all t. The first case implies that

$$\int_{-\infty}^{\infty} p_1(t)dt > \int_{-\infty}^{\infty} p_2(t)dt,$$

but since both integrals must equal one, this is a contradiction. Similarly for the second ($<$) case. Therefore $K \neq \{\emptyset\}$. [Note that this is true for all continuous PDFs, without regard to the MLRP.]

If there are two points $t_1, t_2 \in K$, meaning $p_1(t_1)/p_2(t_1) = p_1(t_2)/p_2(t_2)$, then the MLRP is violated.

By the MLRP, if $t < K$, then $p_1(t)/p_2(t) < p_1(K)/p_2(K)$; since $p_1(K)/p_2(K) = 1$, this means that $p_1(t)/p_2(t) < 1$. Similarly for $t > K$. Thus, $p_1(t)$ and $p_2(t)$ satisfy single-crossing.$\diamondsuit$

Lemma 11 *If $p_i(t)$ is a family of single-crossing distributions then the family satisfies FOSD.*

Proof: Take K defined as in Lemma 10, and consider two members of the family, $p_1(t)$ and $p_2(t)$. Then for any point $k > K$, $p_1(t) > p_2(t)$ for all $t > k$, so

$$\int_k^{\infty} p_1(t)dt > \int_k^{\infty} p_2(t)dt$$

and FOSD is demonstrated. Now considering $k \leq K$,

$$\int_{-\infty}^{k} p_1(t)dt < \int_{-\infty}^{k} p_2(t)dt \tag{4}$$

Since these are PDFs, they integrate to one, and so Inequality 4 is equivalent to

$$1 - \int_k^{\infty} p_1(t)dt < 1 - \int_k^{\infty} p_2(t)dt$$

Subtracting one and negating both sides demonstrates FOSD again.$\diamondsuit$

Lemma 12 *Given $x \in \mathbb{R}$ and $t \in \mathbb{R}$, and a functions $f(x,t)$ with full support over the ranges of x and t. Then if $f(x,t)$ satisfies the MLRP with respect to t [that is, $f(x,2)$ MLRPs $f(x,1)$], then it also satisfies the MLRP with respect to x [that is, $f(2,t)$ MLRPs $f(1,t)$].*

Proof: Let $t_1 > t_2$ and $x_1 > x_2$. We assume that

$$\frac{f(x_1,t_1)}{f(x_2,t_1)} > \frac{f(x_1,t_2)}{f(x_2,t_2)}.$$

Cross multiplying gives

$$\frac{f(x_1,t_1)}{f(x_1,t_2)} > \frac{f(x_2,t_1)}{f(x_2,t_2)},$$

proving our result.$\diamondsuit$

Lemma 13 *Bayesian updating is done using a draw from x and the single distribution $f(x|t)$. Then a distribution family $g_i(t)$ satisfies the MLRP $\Leftrightarrow$ within the support of $f(x|t)$, the family of posterior distributions satisfies the MLRP with respect to t.*

Proof: This is proven in a more general setting by Proposition 2 and Lemma 3. $\diamondsuit$

Note also that by Lemma 12, statisfying the MLRP with respect to t is equivalent to satisfying the MLRP with respect to x.

Theorem 1 *MLRP priors (on t) $\Rightarrow$ FOSD posteriors on t for any fixed x, within the support of $f(x|t)$. That is, $g_1(t)$ MLRPs $g_2(t)$ implies $post_1(t|x)$ FOSDs $post_2(t|x)$ for any x.*

Proof: MLRP priors on $t \Rightarrow$ MLRP posteriors on t (by Lemma 13); MLRP posteriors $\Rightarrow$ single-crossing posteriors (by Lemma 10); single-crossing posteriors $\Rightarrow$ FOSD posteriors on t (by Lemma 11). $\diamondsuit$

6.2 A result used in later proofs

Proposition 14 *Let $op(f,g)$ be monotonically increasing in g, let $f(x)$ be any PDF, and let $g_1(x)$ and $g_2(x)$ be any pair of PDFs which satisfy single-crossing. Then $post(op,f,g_1,x)$ FOSDs $post(op,f,g_2,x)$ for any such functions iff*

$$\int_{-\infty}^{\infty} op(f,g_1,x)dx = \int_{-\infty}^{\infty} op(f,g_2,x)dx.$$

First, I will show that this equality condition is sufficient for the posteriors to satisfy FOSD. Since $op(f,g)$ is monotonically increasing in g, it preserves single-crossing, meaning that $op(f,g_1)$ single-crosses $op(f,g_2)$.

Since $\int_{-\infty}^{\infty} op(f(x),g_1(x))dx = \int_{-\infty}^{\infty} op(f(x),g_2(x))dx$, this means that the posteriors

$$\frac{op(f,g_1)}{\int_{-\infty}^{\infty} op(f(x),g_1(x))dx} \text{ and } \frac{op(f,g_2)}{\int_{-\infty}^{\infty} op(f(x),g_2(x))dx}$$

also satisfy single-crossing, and therefore FOSD is satisfied (by Lemma 11).

Say that, contrary to the above premise, there exist three functions $g_1(\cdot)$, $g_2(\cdot)$, and $f(\cdot)$ such that

$$\int_{-\infty}^{\infty} op(f(x),g_1(x))dx > \int_{-\infty}^{\infty} op(f(x),g_2(x))dx. \tag{5}$$

FOSD is satisfied when

$$\frac{\int_{k}^{\infty} op(f(x),g_1(x))dx}{\int_{-\infty}^{\infty} op(f(x),g_1(x))dx} > \frac{\int_{k}^{\infty} op(f(x),g_2(x))dx}{\int_{-\infty}^{\infty} op(f(x),g_2(x))dx},$$

which we can rewrite as

$$\frac{\int_{k}^{\infty} op(f(x),g_1(x))dx}{\int_{k}^{\infty} op(f(x),g_2(x))dx} > \frac{\int_{-\infty}^{\infty} op(f(x),g_1(x))dx}{\int_{-\infty}^{\infty} op(f(x),g_2(x))dx} \equiv 1+\delta. \tag{6}$$

If the premise of the theorem held, then δ would always equal zero. Assuming Inequality 5 means that $\delta > 0$, but is constant with respect to k, so if the ratio of $\int_{k}^{\infty} op(f(x),g_1(x))$ to $\int_{k}^{\infty} op(f(x),g_2(x))$ approaches one for some sequence of ks, then FOSD will not be not satisfied.

If it is not the case that this ratio approaches one, then we may easily construct a set of distributions where it does. Recall that one of the conditions on $op(f(x),g_1(x))$ was that its integral over all $x \in \mathbb{R}$ be finite. For this to be true, it must be that $\int_{c}^{\infty} op(f(x),g_1(x))$ approaches zero as c approaches ∞. Therefore, there is some c such that

$$\frac{\int_{c}^{\infty} op(f(x),g_1(x))dx}{\int_{-\infty}^{\infty} op(f(x),g_2(x))dx} < \frac{\delta}{2}.$$

Now define

$$g_1'(x) = \begin{cases} g_1(x) & x < c \\ \text{any } y \text{ such that } op(f(x), y) = \left(1 + \frac{\delta}{4}\right) op(f(x), g_2(x)) & x \geq c \end{cases}.$$

That is, $g_1'(x)$ is the same as $g_1(x)$ for all x up to c, and then takes on any value such that we are assured that $op(f(x), g_1'(x))$ is slightly larger than $op(f(x), g_2(x))$ for all $x \geq c$. Single-crossing is still satisfied, and any offending discontinuity may be easily smoothed out.

Notice that

$$\begin{aligned} \frac{\int_{-\infty}^{\infty} op(f(x), g_1(x))dx}{\int_{-\infty}^{\infty} op(f(x), g_2(x))dx} &> \frac{\int_{-\infty}^{c} op(f(x), g_1(x))dx}{\int_{-\infty}^{c} op(f(x), g_2(x))dx} \\ &> 1 + \frac{\delta}{4}, \end{aligned} \tag{7}$$

since $g_1'(x)$ is at its smallest if it were zero above c, and c was defined so that Inequality 7 is true.

We now have

$$\frac{\int_{c}^{\infty} op(f(x), g_1'(x))dx}{\int_{c}^{\infty} op(f(x), g_2(x))dx} < 1 + \frac{\delta}{4},$$

while

$$\frac{\int_{-\infty}^{\infty} op(f(x), g_1'(x))dx}{\int_{-\infty}^{\infty} op(f(x), g_2(x))dx} = 1 + \frac{\delta}{2}.$$

Inequality 6 is not satisfied for $g_1'(x)$, $g_2(x)$, and $k = c$, meaning that FOSD is not satisfied for these functions.

If the reverse of Inequality 5 is true, then there are two ways we can construct single-crossing functions which do not lead to FOSD posteriors. One is to repeat the above procedure, instead modifying the left-hand tail of $g_2(x)$. A neater way is to let $f^m(x) = f(-x)$, $g_1^m(x) = g_2(-x)$, and $g_2^m(x) = g_1(-x)$. By taking the mirror image in this way, we still have $g_1^m(x)$ single-crossing $g_2^m(x)$, and

$$\begin{aligned} \int_{-\infty}^{\infty} op(f^m(x), g_1^m(x))dx &= \int_{-\infty}^{\infty} op(f(x), g_2(x))dx \\ &\text{and} \\ \int_{-\infty}^{\infty} op(f^m(x), g_2^m(x))dx &= \int_{-\infty}^{\infty} op(f(x), g_1(x))dx. \end{aligned}$$

meaning that either Inequality 5 is true or if it is false, then

$$\int_{-\infty}^{\infty} op(f^m(x), g_1^m(x))dx > \int_{-\infty}^{\infty} op(f^m(x), g_2^m(x))dx.$$

The conditions up to Inequality 5 are now satisfied, and we may apply the procedure which followed that inequality to show that there exist functions $f^m(x)$, $g_1^{m\prime}(x)$, and $g_2(x)$ which do not satisfy FOSD.

In conclusion, if there exist three distributions such that

$$\int_{-\infty}^{\infty} op(f(x), g_1(x))dx \neq \int_{-\infty}^{\infty} op(f(x), g_2(x))dx,$$

then either FOSD is not satisfied for the distributions as given, or new PDFs can be constructed for which FOSD does not hold.

6.3 Proofs of results from the text

Proposition 2 *The posterior distribution $post(f, g_1, op, x)$ MLRPs $post(f, g_2, op, x)$ iff the function $op(f(x), g_1(x))$ MLRPs the function $op(f(x), g_2(x))$.*

Proof: MLRP means that the ratio

$$\frac{op(f(x), g_1(x))}{op(f(x), g_2(x))}$$

is increasing in x. Of course, multiplying the ratio by a constant (in terms of x) won't change this, so

$$\frac{op(f(x), g_1(x))}{op(f(x), g_2(x))} \cdot \frac{\int_{-\infty}^{\infty} op(f(y), g_2(y))dy}{\int_{-\infty}^{\infty} op(f(y), g_1(y))dy}$$

is also increasing in x. But this is the ratio of the posteriors, so the posteriors satisfy the MLRP.

The steps reverse to show the 'only if' part of the statement. $\diamondsuit$

Lemma 3 *A continuous operator preserves the MLRP iff it is of the form*

$$op(a, b) = t(a) \cdot b^p,$$

with $p > 0$, and $t(\cdot)$ any transformation.

Proof: First, consider four positive real numbers, a, b, c, and d, and a continuous function $\phi : \mathbb{R} \to \mathbb{R}$ such that

$$\frac{a}{b} > \frac{c}{d} \Rightarrow \frac{\phi(a)}{\phi(b)} > \frac{\phi(c)}{\phi(d)} \tag{8}$$

for all positive a, b, c, and d. What can be said about the function $\phi(\cdot)$?

Setting $b \equiv (ad/c)+\epsilon$, $ad > cb$, so $\phi(a)\phi(d) > \phi(c)\phi(b)$; setting $b \equiv (ad/c)-\epsilon$, $ad < cb$, so $\phi(a)\phi(d) < \phi(c)\phi(b)$; so by continuity, it must be the case that $\phi(a)\phi(d) = \phi(c)\phi(b)$ whenever $ad = cb$. This means that $\phi(a)\phi(d)$ is a function only of the product ad, and can be expressed as a new one-argument function $\pi(ad) \equiv \phi(a)\phi(d)$.

Taking the derivative of this identity gives two new equations:

$$\phi'(a)\phi(d) = \pi'(ad)\cdot d$$

and

$$\phi(a)\phi'(d) = \pi'(ad)\cdot a$$

Rearranging:

$$\frac{\phi'(a)}{\phi(a)}\cdot a = \frac{\phi'(d)}{\phi(d)}\cdot d.$$

Rewrite $\phi'(a)/\phi(a)$ as $d\ln(\phi(a))/da$, and notice that this equation holds for any a and d. So for some constant p and for any $x \in \mathbb{R}$,

$$\frac{d\ln(\phi(x))}{dx}\cdot x = p.$$

Integrating p/x and exponentiating gives:

$$\phi(x) = Cx^p,$$

where $\ln C$ is the constant of integration.

Let $\alpha > \beta$ and take $a = g_1(\alpha)$, $b = g_2(\alpha)$, $c = g_1(\beta)$, $d = g_2(\beta)$, and $\phi(x) = op(f, g, x)$. Then making these substitutions into Implication 8 gives:

$$\frac{g_1(\alpha)}{g_2(\alpha)} > \frac{g_1(\beta)}{g_2(\beta)} \Rightarrow \frac{op(f, g_1, \alpha)}{op(f, g_2, \alpha)} > \frac{op(f, g_1, \beta)}{op(f, g_2, \beta)}. \tag{9}$$

That is: if $g_1(\cdot)$ MLRPs $g_2(\cdot)$, then $op(f(\cdot), g_1(\cdot), \cdot)$ MLRPs $op(f(\cdot), g_2(\cdot), \cdot)$. The proof shows that this can only be the case when $op(f, g)$ is of the form Cg^p. The constant C may be a function of f but not of g.

This proves that any function which preserves the MLRP must be a monomial.

To show that any monomial updating operator preserves the MLRP, let $g_1(x)/g_2(x)$ be an increasing function of x. Then

$$\frac{Cg_1^p(x)}{Cg_2^p(x)} = \left(\frac{g_1(x)}{g_2(x)}\right)^p$$

must also be increasing, so long as $p > 0$. $\diamondsuit$

Proposition 5 *Given any pair of priors g_1 and g_2 which satisfy single-crossing, and any function f, and an updating operator $op(f, g)$ which is monotonically increasing in f and g, the operator $op(\cdot, \cdot)$ provides posteriors ordered by FOSD iff the function is of the form*

$$op(f, g) = t(f) + qg,$$

where $t(\cdot)$ is a transformation function of any form, and q is any positive constant.

Proof: We need only prove that

$$\int_{-\infty}^{\infty} op(f, g_1, x)dx = \int_{-\infty}^{\infty} op(f, g_2, x)dx \tag{10}$$

if and only if $op(f, g, x)$ can be written in the form $t(f) + qg$; Proposition 14 (page 13) proves the rest.

To show that the integral of $t(f(x)) + qg(x)$ is constant in changes in $g(\cdot)$, we need only break down the integral:

$$\begin{aligned} \int_{-\infty}^{\infty} t(f(x)) + qg(x)dx &= \int_{-\infty}^{\infty} t(f(x))dx + \int_{-\infty}^{\infty} qg(x)dx \\ &= \int_{-\infty}^{\infty} t(f(x))dx + q. \end{aligned}$$

This is constant with respect to $g(\cdot)$.

To prove the other direction, I use the following three sample functions to show that if the updating operator is not of the form given here, then Equation 10 does not hold (and so posteriors will not satisfy FOSD, by Proposition 14).

$$\begin{aligned} f(x) &= \begin{cases} \frac{1}{\epsilon} & x \in [0, \epsilon] \\ 0 & x \notin [0, \epsilon] \end{cases} \\ g_1(x) &= \begin{cases} \frac{1}{\epsilon} & x \in [0, \epsilon] \\ 0 & x \notin [0, \epsilon] \end{cases} \\ g_2(x) &= \begin{cases} \frac{1}{\epsilon} & x \in [1, 1+\epsilon] \\ 0 & x \notin [1, 1+\epsilon] \end{cases} \end{aligned}$$

For simplicity, $g_1(x)$ and $g_2(x)$ are weakly single-crossing, but they may be smoothed out to functions which are strictly single-crossing.

Here are some useful integrals, which hold for all $n, m > 0$:

$$\begin{aligned} \int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{\infty} g_1(x)dx = \int_{-\infty}^{\infty} g_2(x)dx &= 1 \\ \int_{-\infty}^{\infty} f^n(x)g_1^m(x)dx &= \left(\frac{1}{\epsilon}\right)^{(n+m-1)} \\ \int_{-\infty}^{\infty} f^n(x)g_2^m(x)dx &= 0 \end{aligned}$$

Now consider any function $op(a, b)$. Weirstrass's polynomial approximation theorem tells us that any continuous function can be approximated arbitrarily well by a polynomial of the form

$$op(a, b) = \sum_{\substack{i=0,1,2,\ldots \\ j=0,1,2,\ldots}} k_{ij} a^i b^j$$

When will $\int_{-\infty}^{\infty} op(f(x), g_1(x))dx = \int_{-\infty}^{\infty} op(f(x), g_2(x))dx$? If the expansion includes terms of the form $k_{i0} f^i(x)$, where $g_1(x)$ or $g_2(x)$ do not appear, these will be equal between both integrals. A term of the form $k_{01} g(x)$ will integrate to one in both cases. But any other term will differ from one integral to the other. Letting ϵ get arbitrarily large will make this difference as large as wished, outstripping the error term in the polynomial approximation. So if $op(a, b)$ is not of the form given in the theorem, there exist functions $f(x)$, $g_1(x)$, and $g_2(x)$ such that $g_1(x)$ and $g_2(x)$ are single-crossing, but Equation 10 does not hold, and so the posteriors do not satisfy FOSD. $\diamondsuit$

Lemma 6 *An updating operator $op(f, g)$ which is monotonically increasing in f and g preserves single-crossing iff the function is of the form*

$$op(f, g) = t(f) + qg,$$

where $t(\cdot)$ is a transformation function of any form, and q is any positive constant.

Proof: Any pair of single-crossing priors satisfies FOSD (Lemma 11). Therefore, if $op(a, b_1)$ and $op(a, b_2)$ take any pair of functions which satisfies FOSD and return a pair of functions which satisfy FOSD, then this updating operator takes any pair of functions which satisfies single-crossing, and returns a pair of functions which satisfy FOSD. By Proposition 5, it must be the case that this operator is of the linear form given in that proposition.

If $g_1(x) > g_2(x)$, then since

$$\begin{aligned} \int_{-\infty}^{\infty} t(f(x)) + g_1(x)dx &= \int_{-\infty}^{\infty} t(f(x)) + g_2(x)dx, \\ \frac{t(f(x)) + g_1(x)}{\int_{-\infty}^{\infty} t(f(x)) + g_1(x)dx} &> \frac{t(f(x)) + g_2(x)}{\int_{-\infty}^{\infty} t(f(x)) + g_2(x)dx}. \end{aligned}$$

So if $g_1(x) > g_2(x)$ for all $x < K$, where K is the crossing point of the priors, then $post_1(x) > post_2(x)$ for all $x < K$; conversely for all $x \geq K$, meaning that if the priors satisfy single-crossing, so do the posteriors. $\diamondsuit$

Theorem 7 *Within the class of operators $op(f, g)$ which are monotonically increasing in both arguments, an operator preserves FOSD iff it is of the form $op(f, g) = t(f) + qg$, where $t(\cdot)$ is a transformation function of any form and q is a positive constant.*

Proof: Necessity: Single-crossing priors are a subset of FOSD priors, so it must be the case that this operator is in the class of operators described by Proposition 5 (or equivalently, Proposition 14).

Sufficiency is easily checked: if $g_1(x)$ FOSDs $g_2(x)$, then Inequality 11 holds for any k; the rest follows from algebra (and the fact that the integral in the denominator of Inequality 12 is the same for both g_1 and g_2).

$$\int_k^{\infty} g_1(x)dx > \int_k^{\infty} g_2(x)dx \tag{11}$$

$$
\begin{aligned}
\int_k^\infty t(f(x)) + q\int_k^\infty g_1(x)dx &> \int_k^\infty t(f(x))dx + q\int_k^\infty g_2(x)dx \\
\int_k^\infty t(f(x)) + qg_1(x)dx &> \int_k^\infty t(f(x)) + qg_2(x)dx \\
\frac{\int_k^\infty t(f(x)) + qg_1(x)dx}{\int_{-\infty}^\infty (f(x)) + qg_1(x)dx} &> \frac{\int_k^\infty t(f(x)) + qg_2(x)dx}{\int_{-\infty}^\infty t(f(x)) + qg_2(x)dx} \qquad (12)
\end{aligned}
$$

$\diamondsuit$

References

Sushil Bikhchandani, Uzi Segal, and Sunil Sharma. Stochastic dominance under Bayesian learning. *Journal of Economic Theory*, 56:352–377, 1992.

Randall L Calvert. The value of biased information: A rational choice model of political advice. *Journal of Politics*, 47:530–555, June 1985.

Timothy J Feddersen and Wolfgang Pesendorfer. Voting behavior and information aggregation in elections with private information. *Econometrica*, 65(5):1029–1058, September 1997.

Dipjyoti Majumdar. *Essays in Social Choice Theory*. PhD thesis, Indian Statistical Institute, October 2002.

Paul R Milgrom. Good news and bad news: Representation theorems and applications. *The Bell Journal of Economics*, 12(2):380–391, Autumn 1981.

Paul R Milgrom and Robert J Weber. A theory of auctions and competitive bidding. *Econometrica*, 50(5):1089–1122, September 1982.

Glenn Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

Ward Whitt. A note on the influence of the sample on the posterior distribution. *Journal of the American Statistical Association*, 74(366): 424–426, June 1979.