Journal of Statistical Planning and Inference 40 (1994) 345–356
North-Holland

The dilation phenomenon in robust Bayesian inference*

Larry Wasserman** and Teddy Seidenfeld***

Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213, USA

Received 1 June 1992; revised manuscript received 22 April 1993

Abstract

Upper and lower probabilities may become uniformly less precise after conditioning. We call this dilation. We review some results about dilation, present some examples and explore the effect of Bayesian updating. Also, we show a connection between dilation and nonconglomerability. Finally, we consider the implications of this phenomenon.

AMS Subject Classifications: Primary 62F15; Secondary 62F35.

Key words: Conditional probability; density ratio neighborhoods; ε-contaminated neighborhoods; robust Bayesian inference; upper and lower probabilities.

1. Introduction

Let $(\Theta, \mathscr{A})$ be a measurable space and let M be a nonempty, closed, convex set of probability measures. The lower expectation of an $\mathscr{A}$-measurable function X is defined by $\underline{P}X=\inf_{P\in\mathscr{M}} PX$, where $PX=\int X(\theta)P(\mathrm{d}\theta)$. Upper expectations and upper and lower probabilities are defined in the obvious way. If $\underline{P}(B)>0$ then the lower conditional expectation of X given B is defined by $\underline{P}(X\mid B)=\inf_{P\in M} P(XB)/P(B)$. Let $\mathscr{B}$ be a measurable partition. If $\underline{P}(B)=0$ for all $B\in\mathscr{B}$ then the above definition is replaced in the usual way by conditional probability given a σ field (Ash, 1972, p. 252). We say that *$\mathscr{B}$ dilates X* if $\underline{P}(X\mid B)\leqslant\underline{P}X\leqslant\bar{P}X\leqslant\bar{P}(X\mid B)$ for all $B\in\mathscr{B}$ with the lower or upper inequality being strict for all $B\in\mathscr{B}$. If the outer inequalities are strict, we say that strict dilation has occurred.

When dilation occurs, there is a uniform loss of precision by conditioning. This conflicts with a familiar Bayesian result (Ramsey, 1990; Good, 1967) that cost-free,

Correspondence to: Dr. Larry Wasserman, Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213, USA.

*Presented at the Milano Workshop on Bayesian Robustness, May 1992.
**Research supported by NSF grant DMS-9005858.
***Research supported by NSF grant SES-92089428.

SSDI 0378-3758(93)E0069-S

new information is worth acquiring before making a terminal decision. In Seidenfeld and Wasserman (1993), we explore the phenomenon and characterize dilation prone sets of probabilities. Here we review these results and investigate some cases in greater detail. We also make a connection with nonconglomerability. Further, we note that if we demand dilation immunity and admit the existence of fair coins, nontrivial upper and lower probabilities are ruled out.

Before proceeding we give a simple example. A coin is flipped twice. Let H_i mean heads on flip i and let T_i mean tails on flip i. Suppose $P(H_1)=P(H_2)=P(T_1)=P(T_2)=1/2$, and let M consist of *all* probabilities with this constraint. Then $\underline{P}(H_2)=\bar{P}(H_2)=1/2$ but $\underline{P}(H_2\,|\,x)=0<1/2<1=\bar{P}(H_2\,|\,x)$ regardless of whether $x=H_1$ or $x=T_1$. Thus, dilation occurs; see Walley (1991 pp. 298–299).

Levi and Seidenfeld pointed out the dilation phenomenon to Good in relation to his argument (Good, 1967) about the value of new information. Good's reply is given in Good (1974). Other discussions related to dilation and conditionalization are in Kyburg (1961, 1977), Seidenfeld (1981), and Walley (1991).

2. Review of results

Here, we briefly review the results from Seidenfeld and Wasserman (1993). Given M, define $M_*(A)=\{P\in M;\ P(A)=\underline{P}(A)\}$, $M^*(A)=\{P\in M;\ P(A)=\bar{P}(A)\}$, $M_*(A\,|\,B)=\{P\in M; P(A\,|\,B)=\underline{P}(A\,|\,B)\}$ and $M^*(A\,|\,B)=\{P\in M; P(A\,|\,B)=\bar{P}(A\,|\,B)\}$. For $P\in\mathscr{P}$, the set of all probabilities, define $S_P(A,B)=P(A\cap B)/(P(A)P(B))$, $d_P(A,B)=P(A\cap B)-P(A)P(B)$, $\Sigma^+(A,B)=\{P\in\mathscr{P};\ d_P(A,B)>0\}$ and $\Sigma^-(A,B)=\{P\in\mathscr{P};\ d_P(A,B)<0\}$. The surface of independence for events A and B is defined by $\mathscr{I}(A,B)=\{P\in\mathscr{P};\ d_P(A,B)=0\}$. Finally, define

$$S_0=\inf_{P\in M_*(A\,|\,B)} S_P(A,B),\qquad S^0=\sup_{P\in M^*(A\,|\,B)} S_P(A,B),$$

$$M_0=\{P\in M_*(A\,|\,B): S_P=S_0\},\qquad M^0=\{P\in M^*(A\,|\,B); S_P=S_0\},$$

and

$$P_0=\inf_{P\in M_0} P(A),\qquad P^0=\sup_{P\in M^0} P(A).$$

Result 1. Suppose that $\mathscr{B}=\{B,B^c\}$. A necessary and sufficient condition for strict dilation is that

$$S_0<\frac{\underline{P}(A)}{P_0(A)}\leqslant 1\leqslant\frac{\bar{P}(A)}{P^0(A)}<S^0.$$

Result 2. Let $\mathscr{B}=\{B,B^c\}$. If $\mathscr{B}$ dilates A, then $M\cap\mathscr{I}(A,B)\neq\emptyset$.

Result 3. If $\mathscr{B}$ strictly dilates A, then for every $B\in\mathscr{B}$, $M_*(A\,|\,B)\subset\Sigma^-(A, B)$ and $M^*(A\,|\,B)\subset\Sigma^+(A, B)$.

Result 4. If for every $B\in\mathscr{B}$,

$$M_*(A)\cap\Sigma^-(A, B)\neq\emptyset \quad\text{and}\quad M^*(A)\cap\Sigma^+(A, B)\neq\emptyset,$$

then $\mathscr{B}$ strictly dilates A.

The next result says that if we include independent coin flips, then dilation always occurs.

Result 5. Suppose that $0<\underline{P}(E)\leqslant\bar{P}(E)<1$ for some $E\in\mathscr{B}$. Extend the algebra $\mathscr{B}$ to a larger algebra $\mathscr{B}^*$ to include events A and B and extend M to M^* so that (i) A, B and E are independent under every $P\in M^*$, (ii) $P(A)=1/2$ for every $P\in M^*$ and (iii) $P(B)=\lambda$ for every $P\in M^*$. Then there exists $\lambda\in(0,1)$ and an event F in the enlarged algebra such that $\{A, A^c\}$ dilates F.

Now let M be an ε-contaminated class (Huber, 1973, 1981; Berger, 1984, 1985, 1990), i.e. $M=\{(1-\varepsilon)P+\varepsilon Q;\ Q\in\mathscr{P}\}$, where P is a fixed probability measure and ε is a fixed number in $[0, 1]$. To avoid triviality, assume $\varepsilon>0$ and that P is an internal point in the set of all probability measures.

Result 6. Dilation occurs for this class if and only if $\varepsilon>c$, where

$$c=\max\left\{\frac{d_P}{P(A)P(B^c)},\frac{d_P}{P(A^c)P(B)},-\frac{d_P}{P(A^c)P(B^c)},-\frac{d_P}{P(A)P(B)}\right\}.$$

We see that dilation occurs if A and B are 'sufficiently independent' under the measure P. If P is a nonatomic measure on the real line, then there always exist A and B with positive probability that are independent under P, hence, $d_P=0$, and dilation occurs for every $\varepsilon>0$.

Define the total variation metric by $d(P,Q)=\sup_A|P(A)-Q(A)|$. Fix P and ε and assume that P is internal. Let $M=\{Q;\ d(P,Q)\leqslant\varepsilon\}$. Then $\underline{P}(A)=\max\{P(A)-\varepsilon,0\}$ and $\bar{P}(A)=\min\{P(A)+\varepsilon,1\}$. Also,

$$\underline{P}(A\,|\,B)=\frac{\max\{P(AB)-\varepsilon,0\}}{\max\{P(AB)-\varepsilon,0\}+\min\{P(A^cB)+\varepsilon,1\}}.$$

Let $d=d_P$. There are four cases:

Case 1: $P(AB)$, $P(AB^c)\leqslant\varepsilon$. Dilation occurs if and only if

$$\varepsilon>\max\{-d/P(B^c),\ d/P(B)\}.$$

Case 2: $P(AB) \leqslant \varepsilon < P(AB^c)$. Dilation occurs if and only if

$$\varepsilon > \max\{-d/P(B),\ -d/P(B^c),\ d/P(B)\}.$$

Case 3: $P(AB^c) \leqslant \varepsilon < P(AB)$. Dilation occurs if and only if

$$\varepsilon > \max\{d/P(B^c),\ -d/P(B^c),\ d/P(B)\}.$$

Case 4: $\varepsilon < P(AB),\ P(AB^c)$. Dilation occurs if and only if

$$\varepsilon > \max\{-d/P(B),\ d/P(B^c),\ -d/P(B^c),\ d/P(B)\}.$$

A question that arises is whether dilation with respect to a binary partition $\{B, B^c\}$ is implied by dilation on a more general partition. A partial answer is given by the following result.

Result 7. If M is an ε-contamination model, then dilation on a partition $\{B_1, \ldots, B_n\}$ implies dilation on a binary subpartition.

Further progress may be made by focusing on neighborhoods of the uniform measure on the unit interval. Let $\Omega = [0, 1]$, let $\mathscr{C}(\Omega)$ be the Borel sets and let μ be Lebesgue measure. Given two measurable functions f and g, say that f and g are *equimeasurable* and write $f \sim g$ if $\mu(\{\omega; f(\omega) > t\}) = \mu(\{\omega; g(\omega) > t\})$ for all t. Given f, there is a unique, nonincreasing, right continuous function f^* such that $f^* \sim f$. The function f^* is called the *decreasing rearrangement* of f. We say that f is *majorized* by g and we write $f \prec g$ if $\int_0^1 f = \int_0^1 g$ and $\int_0^s f^* \leqslant \int_0^s g^*$ for all s. Here, $\int f$ means $\int f(\omega)\, \mu(\mathrm{d}\omega)$. Let $\Lambda(f)$ be the convex closure of $\{g;\ g \sim f\}$. Ryff (1965) shows that $\Lambda(f) = \{g;\ g \prec f\}$. We define the *increasing rearrangement* of f to be unique, nondecreasing, right continuous function f_* such that $f_* \sim f$.

Let $u(\omega) = 1$ for all $\omega \in \Omega$. Let m be a weakly closed, convex set of bounded density functions with respect to Lebesgue measure on Ω, let M be the corresponding set of probability measures and let $\bar{P}$ and $\underline{P}$ be the upper and lower probability generated by M. We call m a neighborhood of u if $f \in m$ implies that $g \in m$ whenever $g \sim f$. This condition is like requiring permutation invariance for neighborhoods of the uniform measure on finite sets. All common neighborhoods satisfy this regularity condition. From Ryff's theorem, it follows that if $f \in m$ and $g \prec f$, then $g \in m$. The properties of such sets are studied in Wasserman and Kadane (1992). If m is a neighborhood of u, we shall say that M is a neighborhood of μ. To avoid triviality, we assume that $M \neq \{\mu\}$. For every f define

$$\rho(f) = \frac{\operatorname{ess\,sup} f}{\operatorname{ess\,inf} f},$$

where $\operatorname{ess\,sup} f = \inf\{t;\ \mu(\{\omega; f(\omega) > t\}) = 0\}$ and $\operatorname{ess\,inf} f = \sup\{t;\ \mu(\{\omega; f(\omega) < t\}) = 0\}$. For $k \geqslant 1$, define $\gamma_k = \{f;\ \rho(f) \leqslant k\}$. This is the *density ratio neighborhood of* μ (DeRobertis and Hartigan, 1981).

Result 8. Suppose that M is a neighborhood of μ. Then M is dilation immune if and only if $m=\gamma_k$ for some k.

We may summarize all this by saying that A is dilated by B when they are sufficiently independent. Result 8 says that it is the shape, not the size of M that determines whether dilation occurs. In particular, density ratio classes have just the right shape to avoid dilation. Wasserman (1992) shows that they have other wonderful properties.

3. A closer look at a simple case

Here we examine the case $M=\text{hull}(\{P, Q\})$ more closely. The results so far show that the independence surface must cut through M to induce dilation. Let $P\in\Sigma^-(A,B)$ and $Q\in\Sigma^+(A,B)$, otherwise there is no dilation. Without loss of generality, suppose that $P(A)\leqslant Q(A)$. Let $\gamma_P=|P(AB)-P(A)P(B)|$. Fix γ for both P and Q to be equal so that $\gamma_P=P(A)P(B)-P(AB)=Q(AB)-Q(A)Q(B)=\gamma$, say. Define $\theta=|P(A)-Q(A)|$ and define the standardized distance $i(M)$ from the independence surface by

$$i(M)=\frac{\gamma}{\theta}\{\max\{P(B^c), Q(B^c)\}\}^{-1}.$$

After some algebra, we see that there is dilation if and only if $i(M)>1$.

To summarize, if P and Q give dependence to A and B in the same direction so that there is no uncertainty about the direction of dependence, there is no dilation. But if P and Q are separated by the independence surface, then dilation occurs if the extreme points are sufficiently far from independence, i.e. if the uncertainty about the dependence between A and B is large relative to the uncertainty about $P(A)$. Perhaps, this may be interpreted as a statement about the distance between the projections of P and Q onto the independence surface.

4. Bayesian updating

So far, we have considered the effect of conditioning, but we have not explored more general Bayesian updating. Now consider a parameter space Θ, which we take to be the real line. Let $H=\{h_1,\ldots,h_k\}$ be a partition which is assumed to be a set of contiguous intervals. Let $p=(p_1,\ldots,p_k)$ where $p_i>0$ and $\sum p_i=1$. A useful class of priors is $M=\{P;\, P(h_i)=p_i,\, i=1,\ldots,k\}$. Let $A=h_i$ be such that $p_i<1$ and let B be such that $A\cap B$ and $A-B$ are nonempty. Then, $\underline{P}(A\mid B)=0<\underline{P}(A)=p_i=\bar{P}(A)<1=\bar{P}(A\mid B)$ and $\underline{P}(A\mid B^c)=0<\underline{P}(A)=p_i=\bar{P}(A)\leqslant\bar{P}(A\mid B^c)$. Thus, we have dilation. But now let $L^y(\theta)$ be a likelihood function for data $Y=y$. Assume that L^y is bounded and continuous. Dilation in this context means $\underline{P}(A\mid y)\leqslant\underline{P}(A)\leqslant\bar{P}(A)\leqslant\bar{P}(A\mid y)$ for all y in the sample space $\mathscr{Y}$.

Result 9. Dilation occurs if and only if, for all $y \in \mathcal{Y}$,

$$\underline{L}_i^y \leqslant \sum_{j \neq i} w_j \bar{L}_j^y \quad \text{and} \quad \bar{L}_i^y \geqslant \sum_{j \neq i} w_j \underline{L}_j^y,$$

where $\bar{L}_i^y = \sup_{\theta \in h_i} L^y(\theta)$, $\underline{L}_i^y = \inf_{\theta \in h_i} L^y(\theta)$ and $w_j = p_j / \sum_{r \neq i} p_r$. If there exists y such that $L^y(\theta) \geqslant L^y(\theta')$ for all $\theta \in h_i$, $\theta' \in h_i^c$, then dilation does not occur.

Proof. Note that

$$\underline{P}(A \mid y) = \frac{p_i \underline{L}_i^y}{p_i \underline{L}_i^y + \sum_{j \neq i} p_j \bar{L}_j^y}.$$

Set this less than p_i to obtain the first inequality. Similarly, for the second. The final claim is obvious. □

Thus, it seems that likelihoods mitigate dilation. Dilation might still occur for y such that the mode is not in A. This might be a high probability event. Further investigation is needed to determine if this is so.

Consider an example. Let Y be Bernoulli(p) and let $h_1 = [0, 1/k), \ldots, h_k = [(k-1)/k, 1]$ and $p_1 = \ldots = p_k = 1/k$. Then dilation occurs for h_i if and only if

$$h_i \subset h^* = \left\{ p; \frac{k^2 - k + 2}{2k^2} \leqslant p \leqslant \frac{k^2 + 3k - 2}{2k^2} \right\}.$$

Let μ be Lebesgue measure. Then $\mu(h^*) = O(1/k)$ so that dilation disappears with increasing precision in the class of priors.

5. Nonconglomerability and dilation

In this section, explicitly, we do not assume that probabilities are countably additive. A bounded random variable X can be thought of as a deFinetti-act, a function from states to cardinal utilities, also called a gamble. Let $\mathcal{G} = \{X; \underline{P}X > 0\}$ be the set of acceptable gambles. (We remind the reader that we assume the set M is closed. We use the topology of pointwise convergence. This assumption allows us to avoid the distracting complication that, with open sets, $PX > 0$ for every $P \in M$ does not entail $X \in \mathcal{G}$). Similarly, let $\mathcal{G}_B = \{X; \underline{P}(X \mid B) > 0\}$ be the set of conditionally acceptable gambles. (Recall that X is a B-called-off gamble provided $X(\theta) = 0$ for θ not in B. Where $P(B) = 0$ for some $P \in M$, $\mathcal{G}_B$ may be larger than the set of acceptable B-called-off gambles, i.e. larger than the set of B-called-off gambles that have positive expected utility for each $P \in M$.)

Now assume that $\mathcal{B} = \{B_i\}$ is countable. M forms B-conglomerable preferences if:

(C) $X \in \mathcal{G}$ whenever there exists $\varepsilon > 0$ such that, for every $B_i \in \mathcal{B}$, $X - \varepsilon \in \mathcal{G}_{B_i}$.

Say that M forms conglomerable preferences if it forms B-conglomerable preferences for every countable partition $\mathscr{B}$.

Claim. A singleton $M=\{P\}$ forms conglomerable preferences just when P is countably additive. (This follows from Theorems 2.3, 3.1 and 3.3 in Schervish et al. (1984).)

The conglomerability condition suggests to us the following dual principle:

(D) If there exists $\varepsilon>0$ such that for each $B_i \in \mathscr{B}$, $X+\varepsilon \notin \mathscr{G}_{B_i}$, then $X \notin \mathscr{G}$.

In the case of a singleton $M=\{P\}$, it is easy to see that conditions (C) and (D) are equivalent; hence, each is equivalent to countable additivity of P. The situation is more interesting when M is not a singleton. Then, in light of dilation, even with a finite partition, a set of countably additive probabilities can fail (D). More precisely,

Result 10. If there is dilation then condition (D) is violated.

Proof. Let $X=I_A-\underline{P}(A)$. For every $P\in M$, $PX=P(A)-\underline{P}(A)\geqslant 0$ so $\underline{P}X\geqslant 0$ and $X\in\mathscr{G}$. But, $\underline{P}(X\mid B)=\inf P(X\mid B)=\inf P(A\mid B)-\underline{P}(A)=\underline{P}(A\mid B)-\underline{P}(A)<0$ for every B. Hence, $X\notin\mathscr{G}_B$ for every $B\in\mathscr{B}$. □

Curiously, Walley (1991) declares violation of (C) 'incoherent', but failures of (D) carry no parallel sanction.

6. Interpreting dilation: imprecision versus indeterminacy

There are at least two ways to approach upper and lower probabilities and sets of probabilities more generally. Following Levi (1985), we need to distinguish *imprecise* from *indeterminate* probabilities. One position is to regard traditional 'Bayesian' theory, with its single precise probability, as the ideal for a rational expression of uncertainty. Then sets of probabilities arise from an incomplete specification of that ideal, as might occur in partial elicitation. An alternative point of view (Smith, 1961; Walley, 1991) is to change the norm and to allow that sets of probabilities provide a complete description of an agent's degrees of beliefs — only they are indeterminate degrees of belief. Of course, both factors may operate at the same time. There can be imprecise evaluation of indeterminate beliefs.

The difference is nontrivial. For one, imprecise probabilities can, in principle, be sharpened and dilation can thus be made to disappear or, at least, fade away as the Bayesian ideal is approached through more accurate elicitation. But with an indeterminate set of probabilities, there is nothing to be sharpened through further elicitation and we face dilation; hence, indeterminacy in probability leads to violation of condition (D) through dilation.

Failures of conditions (C) and (D) are unacceptable to many, including one of us (W). Can dilation be avoided? In the light of Result 5, upper and lower indeterminate probabilities lead to dilation once the apparatus of coin flips is available. And given Result 8, even with imprecise upper and lower probabilities, many robust Bayesian models have a transient problem with dilation — 'transient' until the extra work of eliminating the imprecision is expended and the robustness question is resolved.

We conclude with some pragmatic questions. When does dilation arise in practice? How can we indicate the intensity and frequency of its appearances? We point out that real examples of dilation exist. Weichselberger and Pöhlmann (1990) noted the dilation phenomenon in an expert system scenario and suggested that the information leading to dilation not be gathered. Lavine (1987) also noted that dilation occurs in predictive inference for an exponential distribution when standard neighborhoods are used for robustifying the inferences. At present, however, we have no measure of how often dilation occurs with respect to the events of interest. In a dilation prone set M, there will be some events $\mathscr{D}$ that are dilated. But the event of interest may not be in $\mathscr{D}$. Further work is needed to measure the size of $\mathscr{D}$. At present, robust Bayesians should not worry. But neither should they dismiss the phenomenon as unimportant.

References

Ash, R. (1972). *Real Analysis and Probability.* Academic Press, New York.

Berger, J. (1984). The robust Bayesian viewpoint (with discussion). In: J. Kadane, Ed., *Robustness in Bayesian Statistics.* North-Holland, Amsterdam.

Berger, J. (1985). *Statistical Decision Theory*, 2nd ed. Springer, New York.

Berger, J. (1990). Robust Bayesian analysis: sensitivity to the prior. *J. Statist. Plann. Inference* **25**, 303–328.

Berger, J. and M. Berliner (1986). Robust Bayes and empirical Bayes analysis with ε-contaminated priors. *Ann. Statist.* **14**, 461–486.

DeGroot, M.H. (1970). *Optimal Statistical Decisions.* McGraw-Hill, New York.

de Finetti, B. (1964). Foresight: its logical laws, its subjective sources. In: H.E. Kyburg and H.E. Smokler, Eds., *Studies in Subjective Probability.* Wiley, New York.

DeRobertis, L. and J.A. Hartigan (1981). Bayesian inference using intervals of measures. *Ann. Statist.* **9**, 235–244.

Fine, T.L. (1988). Lower probability models for uncertainty and nondeterministic process. *J. Statist. Plann. Inference* **20**, 389–411.

Good, I.J. (1952). Rational decisions. *J. Roy. Statist. Soc. B.* **14**, 107–114.

Good, I.J. (1967). On the principle of total evidence. *British J. Phil. of Sci.* **17**, 319–321.

Good, I.J. (1974). A little learning can be dangerous. *British J. Phil. of Sci.* **25**, 340–342.

Hardy, G.H., J.E. Littlewood and G. Pólya (1952). *Inequalities*, 2nd ed. Cambridge Univ. Press, New York.

Huber, P.J. (1981). *Robust Statistics.* Wiley, New York.

Huber, P.J. (1973). The use of Choquet capacities in statistics. *Bull. Inst. Internat. Statist.* **45**, 181–191.

Huber, P.J. and V. Strassen (1973). Minimax tests and the Neyman–Pearson lemma for capacities. *Ann. Statist.* **1**, 251–263.

Koopman, B.O. (1940). The bases of probability. *Bull. Amer. Math. Soc.* **46**, 763–774.

Kyburg, H. (1961). *Probability and the Logic of Rational Belief.* Wesleyan Univ. Press, Middleton, CT.

Kyburg, H. (1977). Randomness and the right reference class. *J. Phil.* **74**, 501–520.
Lavine, M. (1987). Ph.D. dissertation, School of Statistics, Univ. of Minnesota.
Lavine, M. (1991). Sensitivity in Bayesian statistics: the prior and the likelihood. *J. Amer. Statist. Assoc.* **86**, 396–399.
Levi, I. (1974). On indeterminate probabilities. *J. Phil.* **71**, 391–418.
Levi, I. (1977). Direct inference. *J. Phil.* **74**, 5–29.
Levi, I. (1982). Conflict and social agency. *J. Phil.* **79**, 231–247.
Levi, I. (1985). Imprecision and indeterminacy in probability judgment. *Phil. of Sci.* **52**, 390–409.
Marshall, A.W. and I. Olkin (1979). *Inequalities: Theory of Majorization and its applications.* Academic Press, New York.
Ramsey, F.P. (1990). Weight or the value of knowledge. *British J. Phil. Sci.* **41**, 1–3.
Ryff, J. (1965). Orbits of L^1 functions under doubly stochastic transformations. *Trans. Amer. Math. Soc.* **117**, 92–100.
Savage, L.J. (1972). *The Foundations of Statistics*, 2nd ed. Dover, New York.
Schervish, M.J., T. Seidenfeld and J. Kadane (1984). The extent of non-conglomerability of finitely additive probabilities. *Z. Wahr. v. Gebiete* **66**, 205–226.
Seidenfeld, T. (1981). Levi on the dogma of randomization. In: R. Bogdan, Ed., *Henry E. Kyburg and Isaac Levi.* Dordrecht, D. Reidel, 263-291.
Seidenfeld, T. (1991). When normal and extensive form differ. *Proc. of the 9th International Congress of Logic, Methodology and the Philosophy of Science*, Uppsala, Sweden (to appear).
Seidenfeld, T., M. Schervish and J. Kadane (1989). On the shared preferences of two Bayesian decision makers. *J. Phil.* **86**, 225–244.
Seidenfeld, T., M. Schervish and J. Kadane (1990). Decisions without ordering. In: W. Sieg, Ed., *Acting and reflecting.* Kluwer, 143–170.
Seidenfeld, T. and L. Wasserman (1993). Dilation for sets of probabilities. *Ann. Statist.* (to appear).
Smith, C.A.B. (1961). Consistency in statistical inference and decision. *J. Roy. Statist. Soc. B.* **23**, 1–25.
Walley, P. (1991). *Statistical Reasoning With Imprecise Probabilities.* Monographs on Statistics and Applied Probability, Chapman and Hall, London.
Walley, P. and T.L. Fine (1982). Towards a frequentist theory of upper and lower probabilities. *Ann. Statist.* **10**, 741–761.
Wasserman, L. (1992). Invariance properties of density ratio classes. *Ann. Statist.* **20**, 2177–2182.
Wasserman, L.A. and J. Kadane (1990). Bayes' theorem for Choquet capacities. *Ann. Statist.* **18**, 1328–1339.
Wasserman, L.A. and J. Kadane (1992). Symmetric upper probabilities. *Ann. Statist.* **20**, 1720–1736.
Weichselberger, K. and S. Pöhlmann (1989). *A Methodology for Uncertainty in Knowledge-Based Systems.* Lecture Notes in Artificial Intelligence, Vol. 419. Springer, New York.

Discussion on 'The dilation phenomenon in robust Bayesian inference' by Larry Wasserman and Teddy Seidenfeld

Constantinos Goutis

University College, London

First of all, I would like to thank Larry Wasserman for presenting us an interesting paper describing the dilation phenomenon which I was certainly not aware of.

There are two general questions that could be asked here. What is the intuitive meaning of dilation and why does it occur when it does occur? A nonrigorous way of describing dilation is to say that we know less about $P(A|\mathscr{B})$ than we knew about $P(A)$. That seems counterintuitive at the first sight. I believe that the main poblem is that we knew little about $P(A\cap\mathscr{B})$ to start with. This may or may not be evident when we consider only the upper and lower probabilities of the event A itself. We can think of the marginal $P(A)$ as the projection of $P(A\cap\mathscr{B})$ to some 'subspace'. The projection might be a point, hence, precise, but this does not guarantee that $P(A\cap\mathscr{B})$ is also precise. The imprecision of $P(A\cap\mathscr{B})$ is hidden if we look only at $P(A)$ and surfaces when $B\in\mathscr{B}$ becomes known. In that sense, dilation does not introduce any new uncertainty, but it helps revealing it.

In order to do a proper updating of $P(A)$, i.e. some updating that would not show more uncertainty than we had, we need the relation of the event A with any B in any $\mathscr{B}$ which may occur. In other words, we must be able to say in which sense A and B depend on each other. This is exactly the reason why quantities like $P(A\cap B)-P(A)P(B)$ are so crucial as to when dilation occurs. Thinking of all possible $\mathscr{B}$ and the relations with A might be too large a task, but failing to do so leads to dilation, as Result 5 shows.

A point I would like to discuss further, relevant to Bayesian robustness and dilation, is the meaning of the word uncertainty itself. If we take probability to be a measure of our own subjective unertainty, being uncertain about it appears to be a pleonasm. If probability is a property of the system then, of course, uncertainty is perfectly justifiable. It is interesting to contrast the two meanings with the statistical physicists' perception of uncertainty, where probability seems to be a little bit of both: some kind of subjective uncertainty which is a property of nature itself.

Following Walley (1991) and his slightly different terminology, there may be indeterminate or incomplete probabilities. The sources of indeterminacy are, among others, lack of or conflicting information, a physical indeterminacy in the system, etc. Essentially, the probability assessor will resist any attempt to be more precise. On the other hand, incompleteness while making probability statements implies a 'true probability', but the assessor is too busy to think about it. In the first case, a probability does not really exist whereas in the latter it certainly does.

A question we might ask ourselves is which approach do robust Bayesians take. It seems that it is a little bit of both, though in most cases there is an implicit assumption that using robust priors is a safeguard against the prior that does not really express the beliefs (hence, it is 'wrong') rather than the acknowledgement that there is no such prior. In both cases, the result is that the process of obtaining the prior information is not through. Whether the situation is transient or not, I do not know. Although specifying a probability of a single event might not be a major problem, practitioners are busy people and are unlikely to have the time or the energy to think about uncountably many probabilities to construct an appropriate prior density. The result in either case is that they end up with imprecise probabilities and they somehow have to live with them. Dilation shows that it is not always a satisfactory situation. I fully agree that dilation cannot be easily dismissed.

The main problem with dilation is not that, when it occurs, people do not obey some rationality criteria, arising from gambling scenarios. Thinking about gambling might be useful operationally since it guarantees that you can make some kind of probability statement. It should not be forgotten that probabilities arose as answers to questions about fair games by the great grandparents of modern statisticians. However, not too much weight should be attached to the fact that people might not obey certain axioms, especially if it is not immediately obvious where a violation of the axioms would lead. After all, people are not rational in so many other aspects of life, why would somebody expect them to be so when stating probabilities? Of course, I do not dismiss the usefulness of the gambling approach, it should just not be taken too seriously.

Some more comments on details of the paper. Result 9 tells us in rigorous terms that dilation occurs for sets with low likelihood. In a somewhat different context Pericchi and Walley (1991) observed that posterior upper and lower probability of intervals tend be far apart for sets where the likelihood is low relative to the prior information. I was wondering if this is simply a coincidence or if there is some deeper relation. Do both phenomena represent the difficulty that upper and lower probabilities have when processing too 'little information', where information is translated here as (relative) height of likelihood?

Another general, though somewhat unformulated, question which might be of interest is whether one can demonstrate phenomena analogous to dilation for other measures which are not defined as simple expectations. Is there any satisfactory way, for our purposes, of defining some global uncertainty for distributions? If yes, how is the prior uncertainty related to the posterior uncertainty? Could one exhibit similar behaviour?

Additional reference

Pericchi, L.R. and P. Walley (1991). Robust Bayesian credible intervals and prior ignorance. *Internat. Statist. Rev.* **59**, 1–23.

Rejoinder

Larry Wasserman and Teddy Seidenfeld

We thank Dr. Goutis for his stimulating comments. Dr. Goutis makes a good point: conditioning does not create indeterminacy, rather, it reveals it. The indeterminacy is already lurking in $A \cap \mathscr{B}$. Conditioning merely lets dilation out of the bag. Seen this way, dilation is a diagnostic for revealing indeterminacy hidden in the initial joint space.

We agree, also, that many robust Bayesians use sets of probabilities because of imprecision, not because of indeterminacy. In practice, imprecision is common, hence, dilation is an issue for all kinds of robust Bayesians.

Regarding popular claims about systematic descriptive failings of Bayesian theory and the fault of particular axioms, the situation is complex. Just because people have predictable difficulties adding two large numbers is no reason to reprogram our computer with a new addition rule. In that sense, subjective probability is normative. Nonetheless, we think there are good normative reasons for liberalizing the theory of personal probability to allow for indeterminacy. For us, the key is to relax the so-called 'ordering' postulate. (The axiom requires that each pair of events be comparable by a qualitative probability, '... is more probably than ...'.) The verdict, of course, depends on what kind of theory is possible without the postulate — see Seidenfeld et al. (1990).

Dr. Goutis writes: 'Pericchi and Walley (1991) observed that [the] posterior upper and lower probability of intervals tend [to] be far apart for sets where the likelihood is low relatively to the prior information' and asks whether this is related to dilation. Perhaps, he is referring to the result on page 15 of Pericchi and Walley (1991). They consider normal i.i.d. data with a prior ε-contamination class around a conjugate prior. When the sample average is far from the prior mean, the upper and lower posterior probability of the standard likelihood interval diverge. (This happens also, as sample size increases or as prior variance increases.) This result is different from out Result 9. We show that, in a prior class defined by fixed probabilities on a partition, dilation is limited to sets that do not have extreme likelihood.

Finally, Dr. Goutis asks the following: For which indices of uncertainty $U(\cdot)$ on a set of probabilities is it the case that $U(\cdot \mid y) > U(\cdot)$ for all outcomes $y \in \mathscr{Y}$? This is an excellent question — currently we do not have an answer.

Reference

Seidenfeld, T., M. Schervish and J. Kadane (1990). Decisions without ordering. In: W. Sieg, Ed., *Acting and Reflecting*. Kluwer, Dordrecht, 143–170.