The Extent of Dilation of Sets of Probabilities and the Asymptotics of Robust Bayesian Inference

Timothy Herron; Teddy Seidenfeld; Larry Wasserman

The Extent of Dilation of Sets of Probabilities and the Asymptotics of Robust Bayesian Inference[1]

Timothy Herron, Teddy Seidenfeld, and Larry Wasserman

Carnegie-Mellon University

1. Overview

We discuss two general issues concerning diverging sets of Bayesian (conditional) probabilities—divergence of "posteriors"— that can result with increasing evidence. Consider a set $\mathcal{P}$ of probabilities typically, but not always, based on a set of Bayesian "priors." Incorporating sets of probabilities, rather than relying on a single probability, is a useful way to provide a rigorous mathematical framework for studying sensitivity and robustness in Classical and Bayesian inference. See: Berger (1984, 1985, 1990); Lavine (1991); Huber and Strassen (1973); Walley (1991); and Wasserman and Kadane (1990). Also, sets of probabilities arise in group decision problems. See: Levi (1982); and Seidenfeld, Kadane, and Schervish (1989). Third, sets of probabilities are one consequence of weakening traditional axioms for uncertainty. See: Good (1952); Smith (1961); Kyburg (1961); Levi (1974); Fishburn (1986); Seidenfeld, Schervish, and Kadane (1990); and Walley (1991).

Fix E, an event of interest, and X, a random variable to be observed. With respect to a set $\mathcal{P}$, when the set of conditional probabilities for E, given X, strictly contains the set of unconditional probabilities for E, for each possible outcome $X = x$, call this phenomenon *dilation* of the set of probabilities (Seidenfeld and Wasserman 1993). Thus, *dilation* contrasts with the asymptotic merging of posterior probabilities reported by Savage (1954) and by Blackwell and Dubins (1962), which we discuss briefly in section §2.

For a wide variety of models for Robust Bayesian inference the extent to which X dilates E is related to a model specific index of how far key elements of $\mathcal{P}$ are from a distribution that makes X and E independent. Some sets $\mathcal{P}$ use a class of priors generated by a "neighborhood" of a focal distribution P. These include: the ε-contamination class of priors, the Total Variation class of priors, and symmetric neighborhoods of a prior. The extent to which X dilates E in these sets is related to a model specific index of how far P is from a distribution that makes X and E independent. In other sets $\mathcal{P}$, e.g., in the Frechet class, and models given by lower and upper probabilities for atoms, the extent of dilation may be indexed by departures from independence of the probabilities that are extreme points of (the convex closure of) $\mathcal{P}$ rather than by re-

ferring to some obvious focal member of $\mathcal{P}$. In section §3, we discuss this connection between independence and indices for the extent of dilation.

In section §4, we consider phenomena related to asymptotic dilation. At a fixed confidence level, $(1-\alpha)$, Classical interval estimates (based on an m.l.e., $\hat{\theta}$), $A_n = [\hat{\theta} - a_n, \hat{\theta} + a_n]$, have length $O(n^{-1/2})$ (for a sample of size n). Of course, the confidence level correctly reports the (prior) probability that $\theta \in A_n$ for each $P \in \mathcal{P}$, $P(A_n) = 1-\alpha$, independent of the prior for θ. However, as shown by Pericchi and Walley (1991), if an ε-contamination class is used for the prior on the parameter, there is asymptotic (posterior) dilation for the A_n, given the data $(x_1, ..., x_n)$. That is, the asymptotic *lower* posterior probability $P_* (= \inf_{P \in \mathcal{P}})$ for A_n is 0,

$$\lim_{n \to \infty} P_*(A_n \mid x_1, ..., x_n) = 0. \qquad \text{(a.s.)}$$

By contrast, if the intervals A'_n are chosen with length $O(\sqrt{\log (n)/n})$, then there is no asymptotic dilation. This is explained by using H.Jeffreys' (1967, §5) theory of Bayesian hypothesis testing. In section §4, we discuss how the class of priors and the asymptotic rate of dilation for Bayesian (posterior) and Classical interval estimates are related. First, however,we summarize two familiar results about the merging of conditional probabilities since these are in sharp contrast with the effects of dilation.

2. The merging of conditional probabilities with increasing shared evidence

As a backdrop to the discussion of dilation, we begin by pointing to two well known results about the asymptotic merging of Bayesian posterior probabilities.

2.1 Savage (1954, §3.6) provides an (almost everywhere) approach simultaneously to consensus and to certainty among a few Bayesian investigators, provided:

(1) they investigate finitely many statistical hypotheses $\Theta = \{\theta_1, ..., \theta_k\}$
(2) they use Bayes' rule to update probabilities about Θ given a growing sequence shared data $\{x_1, ...\}$. These data are identically, independently distributed (i.i.d.) given θ (where the Bayesians agree on the statistical model parametrized by Θ).
(3) they have prior agreement about *null* events. Specifically (given condition 2), there is agreement about which parameter values have positive prior probability.

By a simple application of the strong law of large numbers, Savage concludes that, almost surely, the agents' posterior probabilities will converge to 1 for the true value of Θ. Asymptotically, with probability 1, they achieve consensus and certainty about Θ.

2.2 Blackwell and Dubins (1962) give an impressive generalization about consensus without using either "i" of Savage's i.i.d. condition (2). Theirs is a standard martingale convergence result which we summarize next.

Consider a denumerable sequence of sets X_i $(i = 1,...)$ with associated σ-fields $\mathcal{B}_i$. Form the infinite Cartesian product $X = X_1 \otimes ...$ of sequences $(x_1, x_2, ...) = x \in X$, where $x_i \in X_i$. That is, each x_i is an atom of its algebra $\mathcal{B}_i$. Let the measurable sets in X (the events) be the elements of of the σ-algebra $\mathcal{B}$ generated by the set of measurable rectangles. Define the spaces of histories $(H_n, \mathcal{H}_n)$ and futures $(F_n, \mathcal{F}_n)$ where $H_n = X_1 \otimes ... \otimes X_n$, $\mathcal{H}_n = \mathcal{B}_1 \otimes ... \otimes \mathcal{B}_n$, and where $F_n = X_{n+1} \otimes ...$ and $\mathcal{F}_n = \mathcal{B}_{n+1} \otimes ...$.

Blackwell and Dubins' argument requires that P is a *predictive*, σ-additive probability on the measure space $(X, \mathcal{B})$. (That P is *predictive* means that there exist conditional probability *distributions* of events given past events, $P^n(\bullet \mid \mathcal{H}_n)$.) Consider a probability Q which is in agreement with P about events of measure 0 in $\mathcal{B}$: $\forall E \in \mathcal{B}$, $P(E) = 0$ iff $Q(E) = 0$. That is, P and Q are mutually absolutely continuous [**m.a.c.**]. Then Q, too, is σ-additive and predictive if P is, with conditional distributions $Q^n(\mathcal{F}_n \mid \mathcal{H}_n)$.

Blackwell & Dubins (1962) prove there is almost certain asymptotic consensus between the conditional probabilities P^n and Q^n.

Theorem 1. For each P^n there is (a version of) Q^n so that, almost surely, the distance between them vanishes with increasing histories:$\lim_{n \to \infty} \rho(P^n, Q^n) = 0$ [a.e. P or Q], where ρ is the uniform distance (total variation) metric between distributions. (That is, with μ and ν defined on the same measure space $(M, \mathcal{M})$, $\rho(\mu, \nu)$ is the least upper bound over events $E \in \mathcal{M}$ of $\mid \mu(E) - \nu(E) \mid$.)

Thus, the powerful assumption that P and Q are mutually absolutely continuous (Savage's condition 3) is what drives the merging of the two families of conditional probabilities P^n and Q^n.

3. Dilation and short run divergence of posterior probabilities.

Throughout this section, let $\mathcal{P}$ be a (convex) set of probabilities on a (finite) algebra $\mathcal{A}$. For a useful contrast with Savage-styled, or Blackwell-Dubins-styled asymptotic consensus, the following discussion focuses on the short run dynamics of upper and lower conditional probabilities in Robust Bayesian models.

For an event E, denote by $P_*(E)$ the "lower" probability of E: $\inf_{\mathcal{P}} \{P(E)\}$ and denote by $P^*(E)$ the "upper" probability of E: $\sup_{\mathcal{P}} \{P(E)\}$. Let $(b_1, ..., b_n)$ be a (finite) partition generated by an observable B.

Definition. The set of conditional probabilities $\{P(E \mid b_i)\}$ **dilate** if
$$P_*(E \mid b_i) < P_*(E) \leq P^*(E) < P^*(E \mid b_i) \qquad (i = 1, ..., n).$$

That is, dilation occurs provided that, for each event b_i in a partition B, the conditional probabilities for an event E, given b_i, properly include the unconditional probabilities for E.

Here is an illustration of dilation.

Heuristic Example. Suppose A is a highly "uncertain" event with respect to the set $\mathcal{P}$. That is, $P^*(A) - P_*(A) \approx 1$. Let $\{H,T\}$ indicate the flip of a fair coin whose outcomes are independent of A. That is, $P(A,H) = P(A)/2$ for each $P \in \mathcal{P}$. Define event E by, $E = \{(A,H), (A^c,T)\}$.

It follows, simply, that $P(E) = .5$ for each $P \in \mathcal{P}$. (E is *pivotal* for A.) But then,
$$0 \approx P_*(E \mid H) < P_*(E) = P^*(E) < P^*(E \mid H) \approx 1$$
and $\qquad 0 \approx P_*(E \mid T) < P_*(E) = P^*(E) < P^*(E \mid T) \approx 1.$

Thus, regardless of how the coin lands, conditional probability for event E dilates to a large interval, from a determinate unconditional probability of .5. Also, this example mimics Ellsberg's (1961) "paradox," where the mixture of two uncertain events has a determinate probability.

3.1 Dilation and Independence.

The next two theorems on existence of dilation serve to motivate using indices of departures from independence to gauge the extent of dilation. They appear in (Seidenfeld and Wasserman 1993).
Independence is sufficient for dilation.

Let Q be a convex set of probabilities on algebra $\mathcal{A}$ and suppose we have access to a "fair" coin which may be flipped repeatedly: algebra C. Assume the coin flips are independent and, with respect to Q, also independent of events in $\mathcal{A}$. Let P be the resulting convex set of probabilities on $\mathcal{A} \times C$. (This condition is similar to, e.g., DeGroot's assumption of an extraneous continuous random variable, and is similar to the "fineness" assumptions in the theories of Savage, Ramsey, Jeffrey, etc.)

Theorem 2: If Q is not a singleton, there is a 2×2 table of the form $(E, E^c) \times (H, T)$ where both:
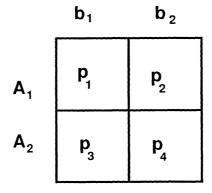$$P_*(E \mid H) < P_*(E) = .5 = P^*(E) < P^*(E \mid H)$$
$$P_*(E \mid T) < P_*(E) = .5 = P^*(E) < P^*(E \mid T).$$

That is, then dilation occurs.

Independence is necessary for dilation.

Let P be a convex set of probabilities on algebra $\mathcal{A}$. The next result is formulated for subalgebras of 4 atoms: (p_1, p_2, p_3, p_4)

The case of 2×2 tables.

	b_1	b_2
A_1	p_1	p_2
A_2	p_3	p_4

Define the quantity $S_P(A_1, b_1) = P(A_1, b_1) / P(A_1)P(b_1) = p_1/(p_1+p_2)(p_1+p_3)$, and we stipulate that $S_P(A_1, b_1) = 1$ if $P(A_1)P(b_1) = 0$. Thus, $S_P(A_1, b_1) = 1$ iff A and B are independent under P and "S_P" is an index of dependence between events.

Lemma 1: If P displays dilation in this sub-algebra, then

$$\inf_P \{S_P(A_1, b_1)\} < 1 < \sup_P \{S_P(A_1, b_1)\}.$$

Theorem 3: If P displays dilation in this sub-algebra, then there exists $P^\# \in P$ such

$S_P\#(A_1,b_1) = 1$.

Thus, independence is also necessary for dilation.

3.2 The extent of dilation

We begin by reviewing some results that obtain for the ε-contaminated model (Seidenfeld and Wasserman 1993). Given probability P and $1 > \varepsilon > 0$, define the convex set $\mathcal{P}_\varepsilon(P) = \{(1-\varepsilon)P + \varepsilon Q: Q$ an arbitrary probability}. This model is popular in studies of Bayesian Robustness. (See Huber 1973 and 1981; Berger 1984.)

Lemma 2. In the ε-contaminated model, dilation occurs in algebra $\mathcal{A}$ iff it occurs in some 2x2 subalgebra of $\mathcal{A}$.

So without loss of generality, the next result is formulated for 2x2 tables using the notation of Lemma 1.

Theorem 4: $\mathcal{P}_\varepsilon(P)$ experiences dilation if and only if
case 1: $S_P(A_1,b_1) > 1$
$\varepsilon > [S_P(A_1,b_1) - 1] \bullet \max\{ P(A_1)/P(A_2) ; P(b_1)/P(b_2) \}$
case 2: $S_P(A_1,b_1) < 1$
$\varepsilon > [1 - S_P(A_1,b_1)] \bullet \max\{ 1 ; P(A_1)P(b_1)/P(A_2)P(b_2) \}$
and case 3: $S_P(A_1,b_1) = 1$
P is internal to the simplex of all distributions.

Thus, dilation occurs in the ε-contaminated model if and only if the focal distribution, P, is close enough (in the tetrahedron of distributions on four atoms) to the saddle-shaped surface of distributions which make A and B independent. Here, S_P provides one relevant index of the proximity of the focal distribution P to the surface of independence.

Definition: For $B \subset \mathcal{B}$ (*B* not necessarily a binary outcome) define the *extent of dilation* by $\Delta(A, B) = \min_{b \in B} [P^*(A|b) - P^*(A) + (P_*(A) - P_*(A|b))]$.

For the ε-contamination model we have

Theorem 5: $\Delta(A, B) = \min_{b \in B} [\varepsilon(1-\varepsilon)P(b^c) \div \varepsilon+(1-\varepsilon)P(b)]$

In this model, $\Delta(A, B)$ does not depend upon the event A. Moreover, the extent of dilation is maximized when $\varepsilon = \sqrt{P(b_\Delta)} \div (1+\sqrt{P(b_\Delta)})$, where $b_\Delta \in B$ achieves the minimum for $\Delta(A, B)$.

Similar findings obtain for *total variation neighborhoods*. Given a probability P and $1 > \varepsilon > 0$, define the convex set $\mathcal{U}_\varepsilon(P) = \{Q: \rho(P, Q) \le \varepsilon\}$. Thus $\mathcal{U}_\varepsilon(P)$ is the uniform distance (total variation) neighborhood of P, corresponding to the metric of Blackwell-Dubins' consensus. As before, consider dilation in 2x2 tables. Define a second index of association: $d_P(A,B) = P(AB) - P(A)P(B)$

(Informal version of) **Theorem 6**: $\mathcal{U}_\varepsilon(P)$ experiences dilation if and only if P is sufficiently close to the surface of independence, as indexed by d_P.

The extent of dilation for the total variation model also may be expressed in terms of the d_P- index, though there are annoying cases depending upon whether the set $\mathcal{U}_\varepsilon(P)$ is truncated by the simplex of all distributions.

Whereas, in the previous two models, each of the sets $\mathcal{P}_\varepsilon(P)$ and $\mathcal{U}_\varepsilon(P)$ has a single distribution that serves as its natural focal point, some sets of probabilities are created through constraints on extreme points directly. For example, consider a model where $\mathcal{P}$ is defined by the lower and upper probabilities on the *atoms* of the algebra $\mathcal{A}$. In section §2 of "Divisive Conditioning" (Herron, Seidenfeld, and Wasserman 1993 - hereafter referred to as **DC**), these sets are called **ALUP** models. For convenience, take the algebra to be finite with atoms $a_{i,j}$ ($i = 1,2;\ j = 1, ..., n$) and where $A = \cup_j a_{1,j}$ and $b_j = \{a_{1,j}, a_{2,j}\}$. For each atom $a_{i,j}$, denote the lower and upper probability bounds achieved within the (closed set) $\mathcal{P}$ by $\beta_{i,j}$ and $\gamma_{i,j}$, respectively. Likewise, for an event E let β_E and γ_E, denote the values $P_*(E)$ and $P^*(E)$. We discuss dilation of the event A given the outcome of random quantity $B = \{b_1, ..., b_n\}$.

Dilation conditions for an ALUP model are easy to express in terms of extreme values within $\mathcal{P}$.

Theorem 7. (i) $P^*(A) < P^*(A|b_j)$ *iff* $\gamma_{1,j} \beta_A c - \gamma_A \beta_{2,j} > 0$

and (ii) $P_*(A) > P_*(A|b_j)$ *iff* $\gamma_{2,j} \beta_A - \gamma_A c \beta_{1,j} > 0$.

Next, given events E and F and a probability P, define the (covariance-) index

$$\delta_P(E,F) = P(EF)P(E^c F^c) - P(E^c F)P(EF^c).$$

Within ALUP models, the extent of dilation for A given $B = b_j$ is provided by the $\delta_P(A,b_j)$ (covariance-) index. Given an event E, use the notation $\{P_*(E)\}$ and $\{P^*(E)\}$ for denoting, respectively, the set of probabilities within $\mathcal{P}$ that achieve the lower and upper probability bounds for event E. Specifically: let $P_{1,j}$ be a probability such that $P_{1,j} \in \{P^*(A)\} \cap \{P^*(a_{1,j})\} \cap \{P_*(a_{2,j})\}$, and let $P_{2,j}$ be a probability such that, $P_{2,j} \in \{P_*(A)\} \cap \{P_*(a_{1,j})\} \cap \{P^*(a_{2,j})\}$. (Existence of $P_{1,j}$ and $P_{2,j}$ are demonstrated in §2 of **DC**.) Then a simple calculation shows:

Theorem 8. $\Delta(A, B) = \min_j [\ \delta_{P1,j}(A,b_j)P_{2,j}(b_j) - \delta_{P2,j}(A,b_j)P_{1,j}(b_j)\ ]$.

Thus, as with the ε-contamination and total variation models, the extent of dilation in ALUP models also is a function of an index of probabilistic independence between the events in question.

Observe that the ε-contamination models are a special case of the ALUP models: they correspond to ALUP models obtained by specifying the lower probabilities for the atoms and letting the upper probabilities be as large as possible consistent with these constraints on lower probabilities. Then the extent of dilation for a set $\mathcal{P}_\varepsilon(P)$ of probabilities may be reported either by attending to the S_p-index for the focal distribution of the set (as in Theorem 5), or by attending to the δ_P-index for the extreme points of the set (as in Theorem 8).

4. Asymptotic Dilation for Classical and Bayesian interval estimates

In an interesting essay, L.Pericchi and P.Walley (1991, pp. 14-16), calculate the upper and lower probabilities of familiar Normal confidence interval estimates under an ε-contaminated model for the "prior" of the unknown Normal mean. Specifically, they consider data $x = (x_1, ..., x_n)$ which are i.i.d. $N(\theta, \sigma^2)$ for an unknown mean θ and known variance σ^2. The "prior" class $\mathcal{P}_\varepsilon(P_0)$ is an ε-contaminated set $\{(1-\varepsilon)P_0 + \varepsilon Q\}$, where P_0 is a conjugate Normal distribution, $N(\mu, v^2)$, and Q is arbitrary. Note

that pairs of elements of $\mathcal{P}_\varepsilon(P_0)$ are *not* all mutually absolutely continuous since Q ranges over one-point distributions that concentrate mass at different values of θ. Hence, Theorem 1 does not apply.

For $\varepsilon = 0$, $\mathcal{P}_\varepsilon(P_0)$ is the singleton Bayes' (conjugate) prior, P_0. Then the Bayes' posterior for θ, $P_0(\theta|x)$, is a Normal $N(\mu',\tau^2)$; where $\tau^2 = (v^{-2} + n\sigma^{-2})^{-1}$, $\mu' = \tau^2[(\mu/v^2) + (n\hat{x}/\sigma^2)]$, and where $\hat{x}$ is the sample average (of x). The standard 95% confidence interval for θ is $A_n = [\hat{x} \pm 1.96\sigma/n^{.5}]$. Under the Bayes' prior P_0 (for $\varepsilon = 0$), the Bayes' posterior of A_n, $P_0(A_n \mid x)$, depends upon the data, x. When n is large enough that τ^2 is approximately equal to σ^2/n, i.e., when $\sigma/vn^{.5}$ is sufficiently small, then $P_0(A_n \mid x)$ is close to .95. Otherwise, $P_0(A_n \mid x)$ may fall to very low values. Thus, asymptotically, the Bayes' posterior for A_n approximates the usual confidence level. However, under the ε-contaminated model $\mathcal{P}_\varepsilon(P_0)$ (for $\varepsilon > 0$), Pericchi and Walley show that, with increasing sample size n, $P^n_*(A_n) \to 0$ while $P^{n*}(A_n) \to 1$. That is, in terms of dilation, the sequence of standard confidence intervals estimates (each at the same fixed confidence level) dilate their unconditional probability or coverage level.

What sequence of confidence levels avoids dilation? That is, if it is required that $P^n_*(A'_n) \geq .95$, how should the intervals, A'_n, grow as a function of n? Pericchi and Walley (1991, 16) report that the sequence of intervals $A'_n = [\hat{x} \pm \zeta_n\sigma/n^{.5}]$ has a posterior probability which is bounded below, e.g., $P^n_*(A'_n) \geq .95$, provided that ζ_n increases at the rate $(\log n)^{.5}$. They call intervals whose lower posterior probability is bounded above some constant, "credible" intervals.

A connection exists between this the rate of growth for ζ_n that makes A'_n credible, due to Walley and Pericchi, and an old but important result due to Sir Harold Jeffreys (1967, 248). The connection to Jeffreys' theory offers another interpretation for the lower posterior probabilities $P^n_*(A'_n)$ arising from the ε-contaminated class.

Adapt Jeffreys' Bayesian hypothesis testing, as follows. Consider a (simple) "null" hypothesis, H_0: $\theta = \theta_0$, against the (composite) alternative $H_0{}^c$: $\theta \neq \theta_0$. Let the prior ratio $P(H_0) / P(H_0{}^c)$ be specified as $\gamma{:}(1-\gamma)$. (Jeffreys uses $\gamma = .5$.) Given H_0, the x_i are i.i.d. $N(\theta_0,\sigma^2)$. Given $H_0{}^c$, let the parameter θ be distributed as $N(\mu,v^2)$. Then, when the data make $|\hat{x}-\theta_0|$ large relative to $\sigma/n^{.5}$ the posterior ratio $P(H_0|x) / P(H_0{}^c|x)$ is smaller than the prior ratio, and when $|\hat{x}-\theta_0|$ is small relative to $\sigma/n^{.5}$ the posterior odds favor the null hypothesis. But to maintain a constant posterior odds ratio with increasing sample size *rather than being constant —as a fixed significance level would entail*— the quantity $|\hat{x}-\theta_0| / (\sigma/n^{.5})$ has *to grow* at the rate $(\log n)^{.5}$ though, of course, the difference $|\hat{x}-\theta_0|$ approaches 0.

In other words, Jeffreys' analysis reveals that, from a Bayesian point of view, the posterior odds for the usual two-sided hypothesis test of H_0 versus the alternative $H_0{}^c$ depends upon *both* the observed type$_1$ error (or significance level), α and the sample size, n. At a fixed significance level, e.g. at observed significance $\alpha = .05$, larger samples yield ever higher (in fact, unbounded) posterior odds in favor of H_0. To keep posterior odds constant as sample size grows, the observed significance level must decrease towards 0.

It is well known that Classical confidence intervals can be obtained by inverting a family of hypothesis tests, generated by varying the "null" hypothesis. That is, the interval $A_n = [\hat{x} \pm 1.96\sigma/n^{.5}]$, with confidence 95%, corresponds to the family of unrejected null hypotheses: each value θ belonging to the interval is a null hypothesis that is not rejected on a standard two-sided test at significance level $\alpha = .05$.

rejected null hypotheses: each value θ belonging to the interval is a null hypothesis that is not rejected on a standard two-sided test at significance level $\alpha = .05$. Consider a family of Jeffreys' hypothesis tests obtained by varying the "null" through the parameter space and, corresponding to each null hypothesis, varying the prior probability which puts mass γ on H_0. Say that a value of θ, $\theta = \theta_0$, is rejected when its posterior probability falls below a threshold, e.g., when $P(H_0|x) < .05$ for the Jeffreys' prior $P(\theta = \theta_0) = \gamma$. The class of probabilities obtained by varying the null hypothesis forms an ε-contaminated model: $\{(1-\gamma)P(\theta|H_0{}^c) + \gamma Q\}$, with extreme points (for Q) corresponding to all the one-point "null" hypotheses.

Define the interval B_n of null hypotheses, with sample size n, where each survives rejection under Jeffreys' tests. The B_n are the intervals $A'_n = [\overset{.}{x} \pm \zeta_n \sigma/n^{.5}]$ of Pericchi and Walley's analysis, reported above. What Pericchi and Walley observe, expressed in terms of the required rate of growth of ζ_n for credible intervals (intervals that have a fixed lower posterior probability with respect to the class $\mathbf{P}_\varepsilon(P_0)$) is exactly the result Jeffreys reports about the shrinking α-levels in hypothesis tests in order that posterior probabilities for the "null" be constant, regardless of sample size. In short, credible intervals from the ε-contaminated model $\mathbf{P}_\varepsilon(P_0)$ are the result of inverting on a family of Jeffreys' hypothesis tests that use a fixed lower bound on posterior odds to form the rejection region of the test.

We conclude our discussion of asymptotic dilation by relating the length of an interval estimate of a parameter θ to the shape of a (symmetic) class of priors for θ. Consider interval estimation of a normal mean, $0 < \theta < 1$. (This restriction to $\theta = (0,1)$ is for mathematical convenience.) We use a prior (symmetric) family $\mathbf{S}_\alpha$ of rearrangements of the density $p_\alpha(\theta) = (1-\alpha)\theta^{-\alpha}$, for $0 < \alpha < 1$. Let the interval estimate of θ be $A_n = [\hat{\theta} - a_n, \hat{\theta} + a_n]$. For constants $C > 0$ and d, write $a_n = \{n^{-1}(C+d log n)\}^{1/2}$.

Theorem 9. For the $\mathbf{S}_\alpha$ model, there is asymptotic dilation of A_n if and only if $d < \alpha$.

(The proof is given in §7 of **DC**.)

5. Summary

In contrast with Savage's, and Blackwell and Dubins' well known results about the merging of Bayesian posterior probabilities given sufficient shared evidence, in this paper we reported two aspects of the contrary case, which we call dilation of sets of probabilities. Let $\mathbf{P}$ be a set of probabilities. The quantity X **dilates** the probabilities for an event E provided the set of conditional probabilities for E, given $X = x$, properly contains the set of unconditional probabilities for E, for each possible outcome of X. Thus, when X dilates E and probabilities are updated by Bayes' rule, the revised opinions about E given X diverge for certain.

In section §3 we indicated how, for several classes of probabilities used in Robust Bayesian inference, the extent of dilation may be gauged by an index of how far away key elements of $\mathbf{P}$ are from distributions that make X and E independent. In section §4 we discussed how ordinary Classical confidence intervals (at a fixed confidence level) experience asymptotic dilation as a function of sample size. Relative to the choice of the class of priors, we explain how to adjust the length of the intervals to avoid asymptotic dilation.

This inquiry, we think, points out two new ways in which Classical and Robust Bayesian statistical inference may be related to each other. We hope to continue our

258

Note

References

Berger, J.O. (1984), "The Robust Bayesian Viewpoint", with discussion, in J.B. Kadane, (ed.), *Robustness of Bayesian Analysis* Amsterdam: North-Holland, 63-144.

_ _ _ _ _ _ . (1985), *Statistical Decision Theory*, (2nd edition). N.Y.: Springer-Verlag.

_ _ _ _ _ _ . (1990), "Robust Bayesian analysis: sensitivity to the prior", *J. Stat. Planning & Inference* 25: 303-328.

Blackwell, D. & Dubins, L. (1962), "Merging of opinions with increasing information", *Ann. Math. Stat.* 33: 882-887.

DeRobertis, L. and Hartigan, J. A. (1981), "Bayesian inference using intervals of measures", *Annals of Statistics* 9: 235-244.

Ellsberg, D. (1961), "Risk, Ambiguity, and the Savage Axioms", *Quart. J. Econ.* 75: 643-669.

Fishburn, P.C. (1986), "The Axioms of Subjective Probability", *Statistical Science* 1: 335-358.

Good, I.J. (1952), "Rational Decisions", *J. Royal Stat. Soc.* B, 14: 107-114.

Herron, T., Seidenfeld, T., and Wasserman, L. (1993), "Divisive Conditioning", Tech.Report #585, Dept. of Statistics, Carnegie Mellon University. Pgh., PA 15213.

Huber, P.J. (1973), "The use of Choquet capacities in statistics", *Bull. Inst. Int. Stat.* 45:181-191.

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

Huber, P.J. and Strassen, V. (1973), "Minimax tests and the Neyman-Pearson lemma for capacities", *Annals of Statistics* 1: 241-263.

Jeffreys, H. (1967), *Theory of Probability* (3rd ed.), Oxford: Oxford University Press.

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

Lavine, M. (1991), "Sensitivity in Bayesian statistics: the prior and the likelihood", *Journal of American Statistics Association* 86: 396-399.

Levi, I. (1974), "On indeterminate probabilities", *Journal of Philosophy* 71: 391-418.

_ _ _ _ . (1982), "Conflict and Social Agency", *Journal of Philosophy* 79: 231-247.

Pericchi, L.R. and Walley, P. (1991), "Robust Bayesian Credible Intervals and Prior Ignorance", *Int. Stat. Review* 58: 1-23.

Savage, L.J. (1954), *The Foundations of Statistics*, New York: Wiley.

Seidenfeld, T., Kadane, J.B., and Schervish, M.J. (1989), "On the shared preferences of two Bayesian decision makers", *Journal of Philosophy* 86: 225-244.

Seidenfeld, T., Schervish, M.J., and Kadane, J.B. (1990), "Decisions without Ordering", in W.Sieg (ed.), *Acting and Reflecting*, Dordrecht: Kluwer Academic, pp. 143-170.

Seidenfeld, T. and Wasserman, L. (1993), "Dilation for sets of probabilities", *Annals of Statistics*. 21: 1139-1154.

Smith, C.A.B. (1961), "Consistency in statistical inference and decisions", *J. Royal Stat. Soc.* B. 23: 1-25.

Walley, P. (1991), *Statistical Reasoning with Imprecise Probabilities*, London: Chapman Hall.

Wasserman, L. and Kadane, J.B. (1990), "Bayes' theorem for Choquet capacities", *Annals of Statistics* 18: 1328-1339.