The Annals of Statistics
1982, Vol. 10, No. 1. 174–183

CONTAMINATION DISTRIBUTIONS

BY MICHAEL GOLDSTEIN

University of Hull

A simple class of models for possibly contaminated data is considered, for which the effect of large observations on our beliefs and on our procedures is small. Various properties are derived, and the effects of differing prior opinions are considered.

1. **Introduction.** Problems which involve samples with possibly contaminated data tend by their very nature to be vaguely formulated, since we seldom have easily quantifiable information on the nature of the contamination process. Because Bayesian modelling of such problems typically requires detailed specification of the contamination procedure, usually we are forced to employ highly simplified models for the spurious observations. Thus, we must consider carefully how well the consequences of the model we have employed correspond to our intuitive understanding of the problem before we may rely on the Bayesian analysis to produce sensible answers.

However, once we admit the possibility that some of the observations may be drawn from a spurious alternative distribution to the distribution of interest, then coherence implies a posterior distribution of sufficient complexity that the properties of the posterior distribution and the nature of the prior-data interaction are very hard to determine. In particular, the analysis will depend crucially on the tail behaviour of the distribution of interest, which is very difficult to specify meaningfully.

Barnett and Lewis (1978) in their survey on the Bayesian approach to outliers conclude that "very much remains to be done to achieve a convincing advance in the Bayesian study of outliers". Their survey illustrates the general point that the formal treatment of the Bayesian model for outliers is straightforward, but that there are major difficulties in deciding upon a reasonable model, and assessing the consequences of its application. Thus in contrast to the material surveyed by Barnett and Lewis, which provides detailed analysis of specific applications, we shall consider qualitative aspects of the behaviour of the simplest general model for contamination, namely that for which the sample value is either drawn from a member of a parametric family, the parameter of which we wish to determine, or the sample value is drawn from a spurious alternative distribution, in which case it carries no information about the parameter of interest.

We formulate a general criterion for the model, corresponding to the requirement that sufficiently discrepant observations have little effect on our beliefs and on our estimation procedures, and the remainder of the paper is a consideration of the consequences of this modelling criterion. Of course, we are not claiming that a model which satisfies our criterion is necessarily a good model for treating contaminated data, only that if we discover that the model does not satisfy the criterion, we might be hesitant to employ the model, and would at the very least find it useful to have the information that the criterion had not been met.

2. **Notation and definitions.** A random sample $x = (x_1, \cdots, x_n)$, of size n, is taken in the following manner: each observation is independently drawn, with specified probability p from a specified distribution $G(x)$, and with probability $(1-p)$ from a distribution $F(x \mid \theta)$, indexed by an unknown parameter θ.

Note that this model applies specifically to those problems for which outliers contain

Received May 6, 1981; revised July, 1978.
AMS 1980 *Subject Classification.* Primary 62F15, Secondary 62G35.
Keywords. Contamination, outliers, robust estimation, Bayesian estimation

no information whatsoever about the unknown parameter θ. This will often be the case, as it will be unusual for us to be able to specify meaningful models linking θ with values which are known to be contaminated. However this does not mean that we must refrain from allowing our prior beliefs about θ to influence our prior choice of $G(\cdot)$.

We shall require that all the distributions are absolutely continuous with respect to some common dominating measure K, which will be either Lebesgue measure or a discrete counting measure on a specified countable subset. Thus, the notations $f(x \mid \theta)$, $g(x)$ will denote either the p.d.f. of F, G with respect to Lebesgue measure, or the p.f. of each with respect to some specified countable set of x values.

We make the following definition. C is a class of real-valued functions of θ, each of which has finite expectation with respect to a specified prior measure $P(\cdot)$ for θ. Let $\underset{\sim}{x}_{(i)}$ be the sample omitting the value x_i. We define $G(x)$ to be a *contamination distribution* for the class C with respect to the prior measure P if, for any value $p \in (0, 1)$, for any sample size n, and for any fixed $\underset{\sim}{x}_{(i)}$, we have, for every $h \in C$,

$$E\{h(\theta) \mid \underset{\sim}{x}\} \to E\{h(\theta) \mid \underset{\sim}{x}_{(i)}\}, \qquad |x_i| \to \infty.$$

In certain contexts, we may only require the above property for $x_i \to \infty$.

As a very important special case, we define the class C_0 to be the class of indicator functions for all sets of values of θ which are measurable with respect to P. G is a contamination distribution for C_0 if for each measurable set A,

$$P(\theta \in A \mid \underset{\sim}{x}) \to P(\theta \in A \mid \underset{\sim}{x}_{(i)}), \; |x_i| \to \infty.$$

We shall refer to such a distribution G as a *weak contamination* distribution. The intuitive interpretation of the weak contamination criterion is straightforward. Essentially, we are aware that the observation may be drawn from a source other than the distribution of interest, and we are requiring of our contamination model that sufficiently discrepant observations do not affect our beliefs concerning the parameter of this distribution. Typically, the one feature common to most problems involving outliers is that if a single observation is sufficiently separated from the remaining observations, then we would expect to be able to identify this value as an outlier. Precisely how separate this value would need to be to make this identification with confidence will depend strongly on the problem at hand. However, we employ outlier models precisely because we believe that it is possible to detect outliers, so that if our model implies that no possible sample values can be so detected, then we must either change our model or at the least, consider carefully the implications of using this model.

Clearly, however, there will be richer classes than C_0 which will be of practical interest. Indeed, it is frequently stressed that outliers will cause us to make bad decisions. Thus, if we wish to estimate a variety of utility-based functions of θ, on the basis of our sample, then by appropriate construction of the class C we may be assured that highly discrepant sample values have little effect on estimates based on posterior mean values, for a wide variety of functions of θ. Rarely would we make a virtually certain identification of an observation as an outlier and yet still wish to allow it considerable weight in determining our estimate, and even in such cases we would presumably find it useful to be aware that our model implied such behaviour.

This raises an immediate problem. Clearly, under the specifications we have made, convergence of the first few posterior moments does not ensure convergence of the higher moments, as the following example shows.

EXAMPLE. Let $F(x \mid \theta)$ be the uniform distribution on $(0, \theta)$, $dP(\theta) \propto (d\theta/\theta^r)[a, \theta \in \infty)$ for some $a > 0$, and $dG(x) \propto (dx/x^{r-k})$, $(r > k+1)$. Then $G(x)$ is a weak contamination distribution and for a single observation x, as $|x| \to \infty$, $E(|\theta|^i \mid x) \to E(|\theta|^i)$, $i < k$, $E(|\theta|^k \mid x) \to c < \infty$, where $c \neq E(|\theta|^k)$, and $E(|\theta|^i \mid x) \to \infty$, $i > k$.

Obvious questions that must be settled are (i) is our definition of a contamination distribution, for example for the class C_1 of polynomials in θ of finite prior expectation, too restrictive to be always satisfied; and (ii) how may we constructively define a contamination distribution for a specified problem?

To approach these questions, it is easier to deal with the apparently weaker requirements that the stated properties must hold only for a sample of size one, and we define $G(x)$ to be a single observation contamination distribution if the contamination property is only required for $n = 1$.

Thus, in Section 3, we consider the above questions for single observation contamination distributions. Then in Section 4, we demonstrate that, under fairly general conditions, single observation and general contamination distributions are equivalent, so that the results of Section 3 are immediately applicable in the more general case. This is accomplished by considering a rather more general problem, namely given two individuals who have different prior distributions for the parameter θ but who both agree on the contamination distribution $G(x)$, then to what extent may their identification of contamination differ? Finally, in Section 5, we consider the reverse of this question, namely under what circumstances may we find possible choices of distribution $G(x)$ which will be contamination distributions for any individual, irrespective of his prior beliefs concerning θ?

3. Single observation contamination models. We begin by establishing some useful simple properties of contamination models for single observations. The subscripted notations $P_0(\theta \mid x)$ and $E_0\{h(\theta) \mid x\}$ refer to posterior calculations if we assume no possibility of contamination, i.e. $p = 0$. Also $P_G(x) =$ prob(observation from $G \mid x$). Thus, given a common dominating measure for F_θ and G, $P_G(x)$ is given by

$$P_G(x) = \frac{pg(x)}{pg(x) + (1-p)f(x)},$$

where $f(x) = \int f(x \mid \theta)\, dP(\theta)$.

Some simple properties of contamination distributions are given in the following theorem.

THEOREM 1. *Let C be a class of functions of θ with the property that, for at least one element $h_1 \in C$,*

$$\liminf_{|x| \to \infty} | E_0\{h_1(\theta) \mid x\} - E\{h_1(\theta)\} | > 0. \tag{3.1}$$

A necessary condition for $g(x)$ to be a single observation contamination distribution for the class C with respect to $P(\theta)$ is that, for each $h \in C$, the function

$$h^*(x) = \int h(\theta) f(x \mid \theta)\, dP(\theta)/g(x) \tag{3.2}$$

tends to zero as $|x| \to \infty$. This condition is also sufficient, provided that condition (3.1) above is replaced by the condition that for at least one element $h_2 \in C$,

$$\liminf_{|x| \to \infty} | E_0\{h_2(\theta) \mid x\} | > 0. \tag{3.3}$$

PROOF. Suppose that $g(x)$ is a single observation contamination distribution. For any $h \in C$ we have

$$E\{h(\theta) \mid x\} = P_G(x) Eh(\theta) + \{1 - P_G(x)\} E_0\{h(\theta) \mid x\}. \tag{3.4}$$

Applying condition (3.1) in this expression with $h = h_1$, we have $P_G(x) \to 1$ as $|x| \to \infty$. For any $h \in C$, and any x for which $g(x) \neq 0$, we have

$$h^*(x) = \frac{p}{1-p} \cdot \frac{(1 - P_G(x)) E_0\{h(\theta) \mid x\}}{P_G(x)}. \tag{3.5}$$

Thus, from (3.4), it is necessary that $h^*(x) \to 0$ as $|x| \to \infty$. Conversely, suppose that condition (3.3) holds. If $h_2^*(x) \to 0$ as $|x| \to \infty$, then from (3.5), $P_G(x) \to 1$ as $|x| \to \infty$.

Thus, from (3.4) and (3.5), if $h^*(x) \to 0$ as $|x| \to \infty$ for all $h \in C$, then $g(x)$ is a single observation contamination distribution.

We have the following corollary.

COROLLARY 1. (i) *A sufficient condition for g to be a weak single observation contamination distribution is that the Bayes factor* $(f(x)/g(x)) \to 0$ *as* $|x| \to \infty$. *This condition is also a necessary condition provided there exists a set A for which*

$$\liminf_{|x|\to\infty} |P_0(\theta \in A \mid x) - P(\theta \in A)| > 0. \tag{3.6}$$

(ii) *A single observation contamination distribution for any class C satisfying condition* (3.1) *above is also a weak single observation contamination distribution.*

Essentially, the conditions of Theorem 1 ensure that the observation does not automatically "reject" itself over some subset of large values of X. If these conditions were not satisfied, we would not need the mechanism of contamination models, in the sense that discrepant observations would automatically tend to discredit themselves.

In certain contexts, it might be reasonable to incorporate the condition (3.1) into the definition of the contamination distribution, in the sense that G is redundant if the posterior measure reaches the appropriate limit even without the contamination model. The main reason that we have not done so is that many of the results we shall derive will be of the form "if G is a contamination distribution under certain circumstances, it will also be a contamination distribution under certain other circumstances". Such results would be rather awkward to state under the modified definition, whereas the practical implication would be the same under each definition. A detailed discussion of the type of situation where conditions such as (3.1) are not met is given in Dawid (1973) and O'Hagan (1979). The essential difference between the model that they discuss and the model we are using is that they consider those situations in which the so-called outliers are not arising from some contamination process but are a natural feature of the data generating process, i.e. models in which outliers are automatically accommodated, whereas we are treating situations in which it is meaningful and indeed necessary to consider the explicit contamination mechanism.

As a simple example, suppose that we are sampling from a normal distribution with unknown mean θ and unit variance, where the prior distribution for θ is normal with mean 0 and precision r_0. The marginal density of x, $f(x)$, is normal with mean 0 and precision $r_0/(1 + r_0)$. Using Theorem 1, we may verify that $g(x) = cf(cx)$, for any $c \in (0, 1)$, is the p.d.f. of a single observation weak contamination distribution. Further, it is a contamination distribution for the class of polynomials in θ. Replacing $P(\theta)$ by $P(\theta \mid \underset{\sim}{x})$, for a general sample $\underset{\sim}{x}$, we may also use Theorem 1 to verify that $g(x)$ has the above properties as a contamination distribution for general sample sizes.

We may use the above theorem to provide a partial answer to the questions posed in Section 2, namely, does a contamination distribution always exist and how can we construct it? We answer both questions in the single observation case by providing, for a general parametric family and prior distribution, a method for constructing a contamination distribution for certain classes of functions C, assuming a common dominating measure for $F(x \mid \theta)$.

THEOREM 2. *Suppose that a class C of real functions of* θ *has the property that there exist a countable sequence of non-negative functions* $\{r_1(\theta), r_2(\theta), \cdots\}$ *with the following properties*:

i) *for each* θ, $r_i(\theta) \le r_j(\theta)$ *for* $i \le j$;
ii) *for each value i, there is a function* $h_i(\theta) \in C$ *with* $\sup_\theta |r_i(\theta) - h_i(\theta)| < \infty$;
iii) *for any function* $h(\theta) \in C$, *there is a value of i for which* $h(\theta) \le r_i(\theta)$ *for all* θ.

Then, for any family $\{f(x \mid \theta)\}$, *we may construct a contamination distribution for the class C with respect to any prior distribution P for which the expected value of each element of C is finite.*

PROOF. For each i, define $f_i(x) = \int r_i(\theta) f(x \mid \theta)\, dP(\theta)$. From property (ii) and the condition imposed on P, $\int f_i(x)\, dx < \infty$, so that we may define a series of increasing finite values C_i such that for each i,

$$\int_{|x|>C_i} f_i(x)\, dx \le \frac{1}{i2^i}.$$

Define $\hat{g}(x)$ by

$$\hat{g}(x) = i f_i(x),\ C_i \le |x| < C_{i+1}.$$

From the construction, the integral of $\hat{g}(x)$ is finite, so that we may define

$$g(x) = \hat{g}(x) \Big/ \int \hat{g}(x)\, dx.$$

From properties i) and (iii), for each $h \in C$, $h^*(x) \to 0$ as $|x| \to \infty$, where h^* is as defined in (3.2). Thus, from Theorem 1, $g(x)$ is a single observation contamination distribution with respect to P, for the class C.

As an illustration of the above theorem, consider the class C of polynomials in θ with finite prior expectation. Specifically, denote $k^* = \sup\{k: E(|\theta|^k) < \infty\}$, and define C_1 to be the class of functions $f_k(\theta) = |\theta|^k$, $k \in [0, k^*)$. We have the following corollary.

COROLLARY 2. *With notation as in Section* 2, *for any family* $\{f(x \mid \theta)\}$, *and any prior distribution* $P(\theta)$, *we may construct a single observation contamination distribution for* C_1.

PROOF. If the quantity k^*, as defined above, is infinite, then define the sequence of functions $r_i(\theta)$ to be

$$\begin{aligned} r_i(\theta) &= 1, & |\theta| &\le 1 \\ &= |\theta|^i, & |\theta| &> 1. \end{aligned}$$

If k^* is finite, then define $r_i(\theta)$ by

$$\begin{aligned} r_i(\theta) &= 1, & |\theta| &\le 1 \\ &= |\theta|^{k^*(1-(i)^{-1})}, & |\theta| &> 1 \end{aligned}$$

unless $E(|\theta|^{k^*}) < \infty$, in which case it is sufficient to set $r_i(\theta) = \sup(1, |\theta|^{k^*})$ for all i. The result now follows from Theorem 2.

4. Contamination with respect to differing prior distributions and the relation between single observation and general contamination distributions. In this section we establish the link between single observation and general contamination models by investigating the following question, which is of interest in its own right: Suppose that two individuals, A and B, have prior distributions $P_A(\theta)$ and $P_B(\theta)$ for θ. Suppose that they both agree on the contamination mechanism, i.e. they have a common distribution $G(x)$. Under what conditions may they draw very different conclusions from discrepant observations? Specifically, define the sets G_A, G_B to be the sets of distributions which are weak single observation contamination distributions with respect to P_A, P_B respectively. What is the relation between the sets G_A, G_B? We have the following theorem, assuming a common dominating measure for F_θ and G.

THEOREM 3. *Suppose that* P_A, P_B *each satisfy property* (3.6) *above. Then*

$$G_A \subseteq G_B \quad \text{if and only if} \quad \limsup_{|x|\to\infty} \frac{\int f(x \mid \theta)\, dP_B(\theta)}{\int f(x \mid \theta)\, dP_A(\theta)} < \infty.$$

PROOF. If the above condition holds, then since

$$\frac{\int f(x \mid \theta)\, dP_B(\theta)}{g(x)} = \frac{\int f(x \mid \theta)\, dP_A(\theta)}{g(x)} \cdot \frac{\int f(x \mid \theta)\, dP_B(\theta)}{\int f(x \mid \theta)\, dP_A(\theta)}$$

we have $G_A \subseteq G_B$. Conversely, suppose that the above condition does not hold and construct a sequence $\{x_i\}$ with $|x_i| \to \infty$, for which

$$\int f(x_i \mid \theta)\, dP_B(\theta) > i^2 \int f(x_i \mid \theta)\, dP_A(\theta)$$

and

$$\int_{|x|>|x_i|} \int f(x \mid \theta)\, dP_A(\theta)\, dx \leq \frac{1}{i2^i}.$$

Define $g(x)$ by

$$g(x) = Ci \int f(x \mid \theta)\, dP_A(\theta), \qquad |x| \in [|x_i|, |x_{i+1}|)$$

where C is the appropriate normalizing constant. By Corollary 1, $g(x) \in G_A$ but $g(x) \notin G_B$, and the result follows.

From the above proof, we may immediately state the following corollary.

COROLLARY 3. *Suppose $h(\theta)$ is a function satisfying properties* (3.1) *and* (3.3) *with respect to P_A and P_B. Define $G_A(h)$, $G_B(h)$ to be the sets of distributions which are single observation contamination distributions for the class consisting of the single function h, with respect to P_A, P_B respectively. Then $G_A(h) \subseteq G_B(h)$ if and only if*

$$\limsup_{|x| \to \infty} \left| \frac{\int h(\theta) f(x \mid \theta)\, dP_B(\theta)}{\int h(\theta) f(x \mid \theta)\, dP_A(\theta)} \right| < \infty.$$

A useful simple corollary is as follows, where we assume a common dominating measure for P_A and P_B, so that we may denote the p.d.f.'s or p.f.'s $P_A(\ )$, $P_B(\ )$ respectively.

COROLLARY 4. *With notation as above, let C denote any class of functions with finite expectations with respect to both P_A and P_B. Let $G_A(C)$, $G_B(C)$ be the set of contamination distributions for the class C with respect to P_A, P_B respectively. If property* (3.1) *holds for individual A and*

$$\sup_\theta \frac{P_B(\theta)}{P_A(\theta)} < \infty,$$

then $G_A(C) \subseteq G_B(C)$.

As an important application of the above corollary, we now provide the link between the single observation and general contamination distributions. Essentially, for many families, these two classes are equivalent, so that we may use the simple structure of the single observation contamination to provide general results for the more complicated situation. In particular the construction of Section 3 is a method of constructing a general contamination distribution.

THEOREM 4. *If $g(x)$ is a single observation contamination distribution for a class C for sampling from $\{f(x \mid \theta)\}$ with respect to $P(\theta)$, and if $f(x \mid \theta)$ is bounded for each x as a function of θ, then $g(x)$ is a contamination distribution for the class C.*

PROOF. Under the above assumptions, $\frac{p(\theta \mid \underset{\sim}{x})}{p(\theta)}$ is bounded as a function of θ for any sample size and sample values. Thus the result follows from the above corollary, with $p_A(\theta) = p(\theta)$ and $p_B(\theta) = p(\theta \mid \underset{\sim}{x}_{(i)})$.

Thus, under certain circumstances, detection of a single contaminant implies the detection of one contaminant out of n for general n. We now state the converse, namely that if $g(x)$ is not a single sample contamination distribution then only in trivial circumstances will it be a more general contamination distribution.

THEOREM 5. *Suppose that $g(x)$ is not a single observation contamination distribution for a class C of nonnegative functions of θ for sampling from $\{f(x \mid \theta)\}$ with respect to $P(\theta)$, and that properties* (3.1) *and* (3.3) *of Theorem* 1 *are satisfied. Then $g(x)$ will not be a single observation contamination distribution with respect to $P(\theta \mid \underset{\sim}{x})$ for any values $\underset{\sim}{x} = (x_1, x_2, \cdots, x_n)$ with condition* (3.1) *holding for $P(\theta \mid \underset{\sim}{x})$ and $g(x_i) \geq 0$, $i = 1, \cdots, n$.*

PROOF. As $g(x)$ is not a single observation contamination distribution, from Theorem 1 there is an element $h \in C$ for which

$$\frac{\int h(\theta) f(x \mid \theta)\, dP(\theta)}{g(x)} \not\to 0, \quad |x| \to \infty.$$

But

$$\int h(\theta) f(x \mid \theta)\, dP(\theta \mid \underset{\sim}{x}) \geq P_G(\underset{\sim}{x}) \int h(\theta) f(x \mid \theta)\, dP(\theta),$$

where $P_G(\underset{\sim}{x})$ is the posterior probability, given $\underset{\sim}{x}$, that all the members of the sample $\underset{\sim}{x}$ were drawn from G. As $g(x_i) > 0$, $i = 1, \cdots, n$, $P_G(\underset{\sim}{x}) > 0$ so that

$$\frac{\int h(\theta) f(x \mid \theta)\, dP(\theta \mid \underset{\sim}{x})}{g(x)} \not\to 0,$$

and by Theorem 1, the result follows.

5. Sufficient classes of contamination distributions. In this section, we consider a complementary question to that posed in Section 4, namely how wide a class of distributions $G(\cdot)$ will we need to consider to ensure that any individual, whatever his prior beliefs concerning θ, may find at least one contamination distribution in this class? Thus, we define a sufficient class of contamination distributions for sampling from the family $F(x \mid \theta)$ to be a class G of distributions such that for any prior distribution $P(\theta)$ there will be at least one member of the class G which is a contamination distribution against P.

If we may find a sufficient class with a small number of elements, then we have a useful reference collection of contamination models, at least one of which must apply for any prior measure. Clearly the simplest such class would be a class consisting of a single member, and we call such a class a unit sufficient class.

Notice that if there is a countable sufficient class $\{g_1, g_2, \cdots,\}$, then $g(x) = \Sigma(\frac{1}{2})^i g_i(x)$ is a contamination distribution for all $P(\theta)$. Thus, either there exists a unit sufficient class for $\{F(x \mid \theta)\}$ or every sufficient class is uncountable.

We now give the necessary and sufficient condition for the existence of the unit sufficient class, this condition being essentially that very large values of x should be unlikely under every member of the parametric family.

THEOREM 6. *There exists a unit sufficient class of weak contamination distributions if and only if, for some k,*

$$\int_{|x|>k} f^*(x)\,dx < \infty, \tag{5.1}$$

where $f^*(x) = \sup_\theta f(x \mid \theta)$.

PROOF. First, suppose that (5.1) holds. We will derive a unit sufficient class for a single observation. Then by Theorem 4 this will also be a unit sufficient class generally because the boundedness in θ of $f(x \mid \theta)$ is ensured by the existence of $f^*(x)$.

Because of (5.1) we may define an increasing unbounded sequence $\{C_i\}$ satisfying the relation

$$\int_{|x|>C_i} f^*(x)\,dx \le 1/(i2^i), \qquad i = 1, 2, \cdots. \tag{5.2}$$

Let

$$\hat{g}(x) = \begin{cases} 0 & |x| < C_1, \\ if^*(x) & C_i \le |x| < C_{i+1}, \qquad i = 1, 2, \cdots \end{cases}$$

From (5.2), $\hat{g}(x)$ is integrable and we can define the density function

$$g(x) = \hat{g}(x) \Big/ \int g(x)\,dx.$$

We now employ Corollary 1. For any prior distribution $P(\theta)$, $g(x)$ is a weak contamination distribution because

$$\frac{f(x)}{g(x)} = \frac{\int f(x \mid \theta)\,dP(\theta)}{g(x)} \le \frac{\int f^*(x)\,dP(\theta)}{g(x)} = \frac{f^*(x)}{g(x)},$$

which tends to zero as $|x| \to \infty$. Therefore $g(x)$ is a unit sufficient class of weak contamination distributions.

Conversely, if (5.1) does not hold, then for any p.d.f. $g(x)$ we will construct a prior distribution $P(\theta)$ for which $g(x)$ is not a weak contamination distribution. We deal first with the case of a single observation.

The failure of (5.1), and the fact that $g(x)$ is a p.d.f., means that the ratio $f^*(x)/g(x)$ is unbounded. More precisely, given any $k > 0$ and $c > 0$ we can find $x_1 > k$ such that $f^*(x_1) \ge cg(x_1)$. Furthermore, having found x_1 in this way, we can also find θ_1 such that $f(x_1 \mid \theta_1) \ge cg(x_1)$. By this procedure we recursively define an increasing unbounded sequence $\{x_i\}$ and a corresponding sequence $\{\theta_i\}$ such that

$$f(x_1 \mid \theta_1) \ge 4g(x_1),$$
$$f(x_i \mid \theta_i) \ge 4^i\{g(x_i) + \textstyle\sum_{j=1}^{i-1} f(x_i \mid \theta_j)\}, \qquad i = 2, 3, \cdots.$$

Notice that the definitions of x_i and θ_i for $i > 1$ are valid because

$$i^{-1}\{g(x) + \textstyle\sum_{j=1}^{i-1} f(x \mid \theta_j)\}$$

is a p.d.f. We now define the prior distribution $P(\theta)$ to be the discrete distribution

$$P(\theta = \theta_i) = 2^{-i}, \qquad i = 1, 2, \cdots$$

This prior distribution satisfies condition (3.6) because for $j < i$

$$\begin{aligned} P_0(\theta_j \mid x_i) &= P(\theta_j) f(x_i \mid \theta_j) / \textstyle\sum_{k=1}^{\infty} P(\theta_k) f(x_i \mid \theta_k) \\ &\le P(\theta_j) f(x_i \mid \theta_j)/P(\theta_i) f(x_i \mid \theta_i) = 2^{i-j} f(x_i \mid \theta_j)/f(x_i \mid \theta_i) \\ &\le 2^{-(i+j)} f(x_i \mid \theta_j)/\{g(x_i) + \textstyle\sum_{k=1}^{i-1} f(x_i \mid \theta_k)\} \le 2^{-(i+j)}, \end{aligned}$$

and therefore as $x \to \infty$, $P(\theta_j) - P_0(\theta_j \mid x)$ does not tend to zero for any j. Therefore, using

Corollary 1, the fact that

$$f(x_i)/g(x_i) = \sum_{j=1}^{\infty} 2^{-j} f(x_i \mid \theta_j)/g(x_i) \geq 2^{-i} f(x_i \mid \theta_i)/g(x_i)$$
$$\geq 2^{i}\{1 + \sum_{j=1}^{i-1} f(x_i \mid \theta_j)/g(x_i)\}$$

implies that $f(x)/g(x) \not\to 0$ as $x \to \infty$, and hence that $g(x)$ is not a weak contamination distribution for this $P(\theta)$ in the case of a single observation. Finally, we use Theorem 5, with C the class C_0 of indicator functions so that (3.1) is automatic and (3.3) reduces to (3.6). It follows that for any $g(x)$ there is a $P(\theta)$ for which $g(x)$ is not a weak contamination distribution, and therefore no unit sufficient class exists.

Surprisingly, the condition for a general contamination unit class is the same as that for the weak unit sufficient class, as shown in the following theorem.

Note that when we refer to a sufficient class of contamination distributions for a class C, it is implicit in the notation that the distribution should be a contamination distribution with respect to each prior measure which gives finite prior expectation to each element of C. We give the proof for a single observation. The result will hold for the general case if the conditions of Theorem 4 are satisfied, i.e. if $f(x \mid \theta)$ is bounded, for each x, as a function of θ.

THEOREM 7. *A unit sufficient class of weak single observation contamination distributions for sampling from $f(x \mid \theta)$ is also a unit sufficient class of single observation contamination distributions with respect to a general class C of functions of θ.*

PROOF. If there is a unit sufficient class of weak contamination distributions, then condition (5.1) must hold. Thus, we may construct a unit sufficient class of weak contamination distributions, $g(x)$, as in (5.2) and (5.3). The Bayes factor $(f(x)/g(x)) \to 0$ as $|x| \to \infty$, for each prior distribution for θ. Further, for any function $h(\theta)$ and any prior distribution $P(\theta)$ with $\int |h(\theta)| \, dP(\theta) < \infty$, we have

$$|h^*(x)| = \frac{\left| \int h(\theta) f(x \mid \theta) \, dP(\theta) \right|}{g(x)} \leq \frac{f^*(x) \int |h(\theta)| \, dP(\theta)}{g(x)}$$

which tends to zero as $|x| \to \infty$, so that $g(x)$ is a unit sufficient class with respect to a general class C.

6. Discussion. As stressed above, models which meet the contamination criteria we have proposed are not necessarily good models for outlier analysis. However, models which fail to meet these criteria should be treated with care, as it is implicit in such models that however discrepant the observation, we will always include it in the analysis. Under certain circumstances, this may be desirable, but at the very least we should be aware that our model implies this. A particular advantage of the specific qualitative features identified by the contamination criteria is that we may assess them for the general sampling problem by considering the more straightforward single observation problem.

As to the assignment of $G(x)$, in certain cases it will be straightforward, in the sense that there is a clear, perhaps subjectively assessed, contaminating mechanism, while in others we may be unwilling to specify precisely our beliefs in the nature of contamination. In such cases, we might use the criteria of contamination with respect to various choices of class C, to specify general classes of possible contamination models, and investigate the effect upon our subsequent procedures of various choices within our class.

More generally, we may consider using the contamination models whenever we are modelling real data by distributions with infinite tails. Typically, we will have information that will place bounds on the range of possible sample values, and the contamination model is a compromise for avoiding the need to truncate the parametric family. As an extension of this argument, most of the common parametric families are employed essentially for reasons of simplicity, and in particular are unlikely to model accurately the tail behaviour of the sampling distribution. Thus, tail values will not, in practice, provide reliable information as to the value of the central "unknown" parameter, and we may employ contamination distributions as a formal device to partially discount such values.

Acknowledgments. I would like to thank the referee and associate editor for their many helpful comments and clarifications.

REFERENCES

BARNETT, V. and LEWIS, T. (1978). Outliers in statistical data. Wiley, New York.
DAWID, A. P. (1973). Posterior expectations for large observations. *Biometrika* **60** 664–667.
O'HAGAN, A. (1979). On outlier rejection phenomena in Bayes inference. *J. Roy. Statist. Soc. Ser. B* **41** 358–367.

DEPARTMENT OF MATHEMATICAL STATISTICS
THE UNIVERSITY OF HULL
COTTINGHAM ROAD
HULL HU6 7RX
ENGLAND