Outliers and Credence for Location Parameter Inference
Author(s): A. O'Hagan
Source: *Journal of the American Statistical Association*, Mar., 1990, Vol. 85, No. 409 (Mar., 1990), pp. 172-176
Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association
Stable URL: https://www.jstor.org/stable/2289540

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at https://about.jstor.org/terms

Taylor & Francis, Ltd. and *American Statistical Association* are collaborating with JSTOR to digitize, preserve and extend access to *Journal of the American Statistical Association*

This content downloaded from
24.240.36.145 on Sun, 16 Jun 2024 04:09:34 +00:00
All use subject to https://about.jstor.org/terms

Outliers and Credence for Location Parameter Inference

A. O'HAGAN*

Heavy-tailed distributions are important for modeling problems in which there may be outlying observations or parameters. Examples of their use are given in O'Hagan (1988). This article develops some general theory, based on the notion of credence, for inference about unknown location parameters in the case of known variances. A density on the real line is defined to have credence c if it is bounded above and below by positive multiples of $(1 + x^2)^{-c/2}$. For instance, a t distribution with d degrees of freedom has credence $1 + d$. I prove that the credence of a sum of independent random variables is the minimum of their individual credences, and that the credence of a posterior density of a location parameter is the sum of the credences of the prior and the observations. More generally, when independent information sources are combined, their credences add. When groups of information sources conflict, outlier rejection occurs, with the group having the greatest total credence dominating all others. Propagation of credence and outlier rejection are considered briefly in the more complex case of a hierarchical model.

KEY WORDS: Bayesian inference; Bimodality; Convolution; Heavy-tailed modeling; Student-t distribution.

1. BACKGROUND

A great deal of statistical theory relies on normal distributions; however, replacing the normal distribution by one with heavy tails, such as a Student-t distribution, can be very useful in problems where there may be outlying observations or parameters. Outlier rejection in Bayesian analysis was first described by De Finetti (1961). Theoretical results were given by Dawid (1973) and Hill (1974). In the simplest case, we can consider a single observation X with mean θ. This is one source of information about Θ; the other source is prior information, which we can suppose to have mean m. The two sources conflict if m is far from the observed x (relative to both standard deviations). If both the likelihood and the prior distribution are normal, then the posterior distribution is normal, with mean between x and m. In the case of conflict, this compromise is equally unsupported by either information source. Dawid (1973) showed that if we give X a t distribution then the conflict is resolved in a very different way: As x moves further from m the posterior distribution converges to the prior distribution. The outlying observation is rejected. Equally, if the likelihood is normal and the prior distribution is t then when conflict arises the prior is rejected and the posterior distribution converges to the normalized likelihood.

More generally, Dawid (1973) gave conditions for one of the two information sources to be rejected in the limit as $|x - m| \to \infty$. In essence, the conditions are that the rejected information source should be represented by a heavy-tailed distribution (likelihood or prior) and that the other source should be represented by a distribution with suitably lighter tails. Hill (1974) complemented Dawid's results by showing that (under simple conditions) if the posterior distribution converges to any limit as $|x| \to \infty$, that limit is the prior distribution. Hill's result also applies to multivariate X and Θ.

Following Dawid's approach, O'Hagan (1979) considered outlier rejection in a sample, simplifying Dawid's conditions somewhat and showing how one or more outlying observations in a sample could be rejected. As the outliers become more extreme, the posterior distribution of Θ converges to the posterior that would be obtained from the reduced sample, excluding the outliers. In particular, this happens for observations having t distributions.

In Section 2 I define a distribution to have credence c if its density is bounded above and below by positive multiples of $(1 + x^2)^{-c/2}$. The definition is motivated by the use of t distributions in O'Hagan (1988), where a t distribution with d degrees of freedom was defined to have "credibility" $1 + d$. The term *credence* replaces *credibility* to avoid confusion with uses of credibility in other contexts. The definition of credence encompasses t distributions but is more general. In particular, a density with given credence may not satisfy any of the regularity conditions, such as continuity and monotonicity in the tails, imposed in the earlier work of Dawid and O'Hagan. Furthermore, the convolution of two t distributions is not a t distribution in general, and cannot even be expressed in a simple, closed form. Nevertheless, in Section 2 we find that the credence of such a convolution is the minimum of the credences of its component distributions. The remaining sections consider the posterior distribution when some of the information sources (data and prior) conflict, assuming only that each information source is represented by a distribution with a well-defined credence.

It should be stressed that many of our results are well-known in the sense that it is generally accepted that heavy-tailed distributions behave in this way. Possibly only Theorem 8 is not well-known in this sense. Certainly, O'Hagan (1988) and some of the work of B.M. Hill and other authors implicitly assumes that heavy-tailed distributions behave in this way, and the primary concern of this article is to provide formal justifications for their intuition.

* A. O'Hagan is Senior Lecturer, Department of Statistics, University of Warwick, Coventry CV4, 7AL, United Kingdom. The author thanks the referees for several constructive comments.

© 1990 American Statistical Association
Journal of the American Statistical Association
March 1990, Vol. 85, No. 409, Theory and Methods

This content downloaded from
24.240.36.145 on Sun, 16 Jun 2024 04:09:34 +00:00
All use subject to https://about.jstor.org/terms

2. BASIC PROPERTIES

Definition. A density f on $\mathbf{R}$ has credence c, denoted by $\text{cred}(f) = c$, if there exist $K \ge k > 0$ such that for all $x \in \mathbf{R}$

$$k \le (1 + x^2)^{c/2} f(x) \le K. \qquad (2.1)$$

I will also write $\text{cred}(X) = c$ if the distribution of X has a density with credence c.

Essentially, (2.1) ensures that for large $|x|$, $f(x)$ is of order $|x|^{-c}$. It is not a completely general measure of tail weight, for various reasons, and $\text{cred}(\cdot)$ is undefined for many densities. It is useful in a wide class of problems, however, where we might wish to consider distributions whose tails are like those of t distributions. If $\text{cred}(f)$ is defined it is unique.

Lemma. For any $a \ne 0$ and b, there exist $L \ge l > 0$ such that for all $x \in \mathbf{R}$

$$l \le \{1 + (ax + b)^2\}/(1 + x^2) \le L. \qquad (2.2)$$

Theorem 1 (Invariance). If $\text{cred}(X) = c$, then for any $a \ne 0$ and b

$$\text{cred}(aX + b) = c. \qquad (2.3)$$

Proof. See the Appendix.

The lemma establishes directly that the standard Student-t distribution with d degrees of freedom has credence $1 + d$. Moreover, through the invariance theorem, this applies to the general t family having location and scale parameters. In general, credence is independent of location and scale.

Theorem 2 (Convolution). Let X be independent of Y, $\text{cred}(X) = c$, and $\text{cred}(Y) = c'$; then

$$\text{cred}(X + Y) = \min(c, c'). \qquad (2.4)$$

Proof. See the Appendix.

The proof of Theorem 2 works whenever $c, c' \ge 0$. Of course, if f is to be a proper density function its credence must be greater than unity. In Bayesian analysis, however, it is often useful to consider improper densities. Our definition says, for instance, that the (improper) uniform density on $\mathbf{R}$ has zero credence. Formal operations with improper densities are often proper, and all of the results in this article are valid and have useful interpretations for credences in [0, 1], as well as for proper distributions.

At the other extreme, the normal distribution can be accommodated as a case of infinite credence. Formally, I define a density to have infinite credence if for any given c and $k > 0$ there exists an A such that for all $|x| > A$, $k > (1 + x^2)^{c/2} f(x)$. I will also require f to be positive and bounded above so that for any c there will exist a K such that that f satisfies the upper bound in (2.1). The Appendix shows that Theorem 2 applies also to infinite credence. That is, if either X or Y (or both) has infinite credence then so does $X + Y$. The Box–Tiao family with densities proportional to $\exp(-|x|^b)$ all have infinite credence.

3. POSTERIOR PROPERTIES

We now consider the context in which heavy-tailed distributions first became of interest in Bayesian inference. A parameter Θ has prior density f. An observation Y is made whose density, given $\Theta = \theta$, is $p(y \mid \theta) = g(y - \theta)$. Thus Θ is a location parameter. We are interested in properties of the posterior distribution of Θ when either f or g (or both) is heavy-tailed. Our results are not confined to Bayesian inference, however. Moreoever, in more complex Bayesian models we may wish to apply them where both Y and Θ are parameters (or "hyperparameters") or when both are observables. To emphasize this generality, we replace Θ with the more neutral symbol X. X and Y are any two random variables.

Throughout this section suppose the following: (a) X has density f, and $\text{cred}(f) = c$; (b) the conditional density of Y given $X = x$ is $p(y \mid x) = g(y - x)$, and $\text{cred}(g) = c'$; (c) the marginal density of Y is denoted by h, and the conditional density of X, given $Y = y$, is denoted by q_y. Bayes's theorem gives

$$q_y(x) = f(x)g(y - x)/h(y), \qquad (3.1)$$

where Theorem 2 tells us that $\text{cred}(h) = \min(c, c') = c^*$. Thus there exist $K \ge k > 0$ such that for all x

$$k \le (1 + x^2)^{c/2} f(x) \le K, \qquad (3.2)$$

there exist $K' \ge k' > 0$ such that for all x, y

$$k' \le \{1 + (y - x)^2\}^{c'/2} g(y - x) \le K', \qquad (3.3)$$

and there exist $K^* \ge k^* > 0$ such that for all y

$$k^* \le (1 + y^2)^{c^*/2} h(y) \le K^*. \qquad (3.4)$$

Theorem 3 (Posterior Credence). For all y, $\text{cred}(q_y) = c + c'$.

Proof. For any given y, $h(y)$ is a constant. Using (3.1), (3.2), and (3.3), for all x, y

$$\{h(y)\}^{-1} kk' \le (1 + x^2)^{c/2} \{1 + (y - x)^2\}^{c'/2}\, q_y(x)$$
$$\le \{h(y)\}^{-1} KK'.$$

I now apply the lemma to give bounds on $\{1 + (y - x)^2\}/(1 + x^2)$ for any given y. It is equally easy to prove that if either c or c' is infinite then so is $\text{cred}(q_y)$.

Theorem 4 (Bimodality). There exists an A such that for all $|y| > A$, q_y is bimodal.

Proof. See the Appendix.

Bimodality arises when there is conflict between the two sources of information—the marginal (prior) distribution of X, suggesting the X is somewhere near 0, and the conditional distribution of Y given x (or likelihood), suggesting that X is somewhere near y. The posterior density q_y fails to reconcile the conflict into a single mode. If either component distribution has infinite credence, we do not in general obtain bimodality.

4. OUTLIERS

This section looks further at the posterior distribution (3.1) of Section 3, as $|y| \to \infty$. The properties that I shall

This content downloaded from
24.240.36.145 on Sun, 16 Jun 2024 04:09:34 +00:00
All use subject to https://about.jstor.org/terms

prove can be thought of as outlier rejection. In interpreting the theorems it is helpful to compare with the case when both f and g are zero-mean normal densities. Then the posterior density q_y is also normal. Its variance is independent of y, and its mean is ay, where $0 < a < 1$. As $|y|$ increases, first note that the posterior probability in any neighborhood of 0 or y tends to 0, despite the fact that these two neighborhoods are precisely where the two information sources—prior and likelihood—suggest that X should lie. Instead, all of the posterior probability lies in a neighborhood of ay. Second, note that all moments of the posterior distribution tend to ∞. In contrast, the outlier rejection phenomenon is that, as $|y| \to \infty$, one information source dominates the other. Posterior probability concentrates in the neighborhood of the dominant information source and tends to 0 everywhere else.

We shall see that the information source with the greater credence will dominate. Assume throughout this section that

$$c > c'; \tag{4.1}$$

therefore, $c^* = c'$ in (3.4). In this case the prior information dominates, and the posterior probablity settles around the origin. Results for $c' > c$ follow immediately by looking at the posterior distribution of $X - y$ so that the roles of f and g are reversed.

Theorem 5 (Rejection). If $c > c'$, then (a) for any given $d > 0$ and $\varepsilon > 0$, there exists an A such that for all $|y| > A$ and for all $|x| \leq d$,

$$\frac{k'}{K^*}(1 - \varepsilon) \leq \{f(x)\}^{-1} q_y(x) \leq \frac{K'}{k^*}(1 + \varepsilon), \tag{4.2}$$

and (b) for all t: $\mathbf{R}^+ \to \mathbf{R}^+$ such that $t(y) \to \infty$ as $y \to \infty$,

$$\lim_{|y|\to\infty} \Pr(|X| > t(|y|) \mid y) = 0. \tag{4.3}$$

Proof. See the Appendix.

Result (a) of Theorem 5 says that posterior probabilities for any *fixed* x are bounded by positive multiples of the prior density. Result (b) says that posterior probability everywhere else tends to 0. In particular, the likelihood suggesting that X is in the neighborhood of y is ignored in the limit. Furthermore, there is ultimately no posterior probability in a neighborhood of ay for any $a > 0$, which contrasts with the case of normal distributions.

The outlier rejection limits of Dawid (1973) and Hill (1974) are stronger than Theorem 5 because in place of (4.2) they assert that $q_y \to f$. My result is weaker because smoothness has not been imposed on the tails of g. If they oscillate then q_y may not even have a limit.

It is simple to show that Theorem 5 holds if c is infinite, but we know from the case of two normal distributions already discussed that outlier rejection in this form will not necessarily occur if both c and c' are infinite. To illustrate the possibilities, O'Hagan (1979) considers outlier rejection in a sample from a Box–Tiao density with index b. If $b \geq 1$ the Box–Tiao distribution is outlier-resistant, meaning that outlier rejection will not occur. On the other hand, if $b < 1$ the distribution is outlier-prone: Outlier rejection does occur, yet these distributions also have infinite credence. Further study is needed into distributions with tails of this type.

Although (4.3) says that posterior probabilities for infinite x go to 0, they may not tend to 0 quickly enough to prevent posterior moments going to infinity.

Theorem 6 (Moment Limits). If $c - c' > 1 + p$, then $\mathbf{E}(|X|^p \mid y)$ is bounded in y.
Proof. See the Appendix.

5. MANY INFORMATION SOURCES

I now extend the theory to the case of many sources of information about a single parameter Θ. Suppose that X_1, $X_2, \ldots, X_n$ are independently distributed given that $\Theta = \theta$, that the density of X_i is $g_i(x_i - \theta)$, and that $\text{cred}(g_i) = c_i$. The prior density of Θ is written as $g_0(x_0 - \theta)$, and we suppose that $\text{cred}(g_0) = c_0$. I introduce x_0 as the location for prior information so that no special significance attaches to the origin. I have also written the prior density as $g_0(x_0 - \theta)$ rather than $g_0(\theta - x_0)$ to emphasize that all $n + 1$ information sources are combined symmetrically in the posterior density

$$q_{\mathbf{x}}(\theta) \propto \prod_{i=0}^{n} g_i(x_i - \theta). \tag{5.1}$$

Repeated use of Theorem 3 shows that $q_{\mathbf{x}}$ has credence $\sum_{i=0}^n c_i$ for all $\mathbf{x}$. When independent information sources are combined, their credences add.

Theorem 7 (Grouping). For $i = 1, 2, \ldots, m$ let $\text{cred}(g_i) = c_i$. Let

$$f(t) = \prod_{i=1}^{m} g_i(y_i + t); \tag{5.2}$$

then $\text{cred}(f) = \sum_{i=1}^m c_i$.
Proof. Apply Definition (2.1) and the lemma to each $g_i(y_i + t)$.

Suppose that the $n + 1$ information sources in (5.1) form two widely separated groups. Then we would expect to find that one group dominates the other, in the sense of Theorem 5, as their separation increases. For this, we need to consider a group moving as a whole so that it has a center that moves, but differences between the group members stay constant. So we consider the group $x_1, x_2, \ldots, x_m$ and write $x_i = y_i + x$, where x is an arbitrary center. Regard the y_i's as fixed. The combination of the information in the group is

$$\prod_{i=1}^{m} g_i(x_i - \theta) = \prod_{i=1}^{m} g_i(y_i + t), \tag{5.3}$$

where $t = x - \theta$. Comparing (5.3) and (5.2), we see that the group acts as a single information source, with density $f(t) = f(x - \theta)$. This source is centered at x, as required, and its credence is the sum of the credences of its m members. This is the significance of Theorem 7.

We can now apply Theorems 5 and 6 to the case of two conflicting groups. As might be expected, as their sepa-

This content downloaded from
24.240.36.145 on Sun, 16 Jun 2024 04:09:34 +00:00
All use subject to https://about.jstor.org/terms

ration increases, the group with the larger (combined) credence dominates. All of the information in the other group is ultimately rejected.

Now suppose that the information sources divide into three or more well-separated groups. We cannot apply Theorem 5 repeatedly (and this is also why Theorem 7 was needed to handle two groups). We might expect that if any group has credence higher than the sum of all of the others, then it will dominate. The following theorem is much stronger. The group with the highest credence dominates all of the others. Without loss of generality, suppose that this is the group indexed by 0 in (5.1). There are $n + 1$ groups centered at $x_0, x_1, \ldots, x_n$, and we let the centers separate further and further by writing $x_i = zy_i$ and letting $z \to \infty$.

Theorem 8 (*Multiple Rejection*). Let $q_{\mathbf{x}}(\theta)$ be as in (5.1) with $\text{cred}(g_i) = c_i$ for $i = 0, 1, \ldots, n$. Let $c_0 > c_i$ for all $i > 0$. Let $x_i = zy_i$ and let $z \to \infty$. Then $q_{\mathbf{x}}(\theta) \to 0$ for all θ in any neighborhood of $x_1, x_2, \ldots$ or x_n. In any neighborhood of x_0, $q_{\mathbf{x}}(\theta)$ is bounded by positive multiples of $g_0(x_0 - \theta)$.

Proof. See the Appendix.

The formulation of Theorem 8 is different from that of Theorem 5 because there is no equivalent of Theorem 2. It is clear that the analog of Theorem 6 would entail $\min(c_0 - c_i) > 1 + p$.

No single source or group dominates if the c_i's do not have a single maximum. Then the posterior will continue to represent two or more sources as $z \to \infty$. It will be multimodal (unlike the case with normal distributions) with modes close to the centers of those sources, and all posterior moments will tend to ∞.

Applications. The theory developed in this article may be applied to quite complex models. For instance, the one-way analysis-of-variance example in section 2 of O'Hagan (1988) may now be analyzed fully. An important restriction of this theory is that all variances are assumed known. To extend it to cover unknown scale parameters is not trivial.

APPENDIX: PROOFS

Proof of Theorem 1. Let X have density f, and (2.1) applies. The density of $aX + b$ is g, where $g(x) = a^{-1}f(a^{-1}(x - b))$. Consider

$$\begin{aligned}(1 + x^2)^{c/2}g(x) &= a^{-1}(1 + x^2)^{c/2}f(a^{-1}(x - b)) \\ &= a^{-1}\{1 + (ay + b)^2\}^{c/2}f(y),\end{aligned}$$

where $y = a^{-1}(x - b)$. Varying x over $\mathbf{R}$ is equivalent to varying y over $\mathbf{R}$, and applying (2.1) and (2.2) establishes (2.3).

Proof of Theorem 2. Let X have density f and Y have density g. Let $\min(c, c') = c^*$. The density of $X + Y$ is the convolution $h(y) = \int_{-\infty}^{\infty} f(x)g(y - x)\, dx$.

Lower Bound: Take any $\varepsilon > 0$. Then from the credences of X and Y, there exist $k, k' > 0$ such that for all $x \in [0, \varepsilon]$ and for all y, $f(x)g(y - x) \geq kk'(1 + \varepsilon^2)^{-c/2}(1 + y^2)^{-c'/2}$. Therefore $h(y) \geq l(1 + y^2)^{-c'/2}$, where $l = kk'\varepsilon(1 + \varepsilon^2)^{-c/2} > 0$. By looking at $x \in [y - \varepsilon, y]$ I prove similarly that $h(y) \geq l'(1 + y^2)^{-c/2}$, where $l' = kk'\varepsilon(1 + \varepsilon^2)^{-c'/2} > 0$. Define $k^* = \min(l, l')$, and we have $h(y) \geq k^*(1 + y^2)^{-c^*/2}$.

Upper Bound: There exist K, K' such that for all $y \geq 0$,

$$\begin{aligned}h(y) &\leq K' \int_{-\infty}^{y/2} f(x)\{1 + (y - x)^2\}^{-c'/2}\, dx \\ &\quad + K \int_{y/2}^{\infty} (1 + x^2)^{-c/2}g(y - x)\, dx \\ &< K'(1 + y^2/4)^{-c'/2} \int_{-in}^{y/2} f(x)\, dx \\ &\quad + K(1 + y^2/4)^{-c/2} \int_{y/2}^{\infty} g(y - x)\, dx \\ &< (K' + K)(1 + y^2/4)^{-c^*/2} = K^*(4 + y^2)^{-c^*/2} \\ &\leq K^*(1 + y^2)^{-c^*/2},\end{aligned}$$

where $K^* = (K' + K)2^{c^*}$. For ε sufficiently small, $k^* \leq K^*$.

Theorem 2 With Infinite Credences: First suppose that only the credence of Y is infinite. The upper-bound part of this proof will go through trivially for any $c' > c$. For the lower bound we only need consider $x \in [y - \varepsilon, y]$. Define $g_0 = \inf_{[0,\varepsilon]} g$. Then $f(x)g(y - x) \geq k\, g_0(1 + y^2)^{-c/2}$; therefore, $h(y) \geq \varepsilon k\, g_0(1 + y^2)^{-c/2}$. Next suppose that both $X$and Y have infinite credences. Consider the case $y > 0$:

$$\begin{aligned}(1 + y^2)^{c/2}h(y) &= (1 + y^2)^{c/2} \int_{-\infty}^{\infty} f(x)g(y - x)\, dx \\ &= \int_{-\infty}^{y/2} f(x)\{(1 + y^2)^{c/2}g(y - x)\}\, dx \\ &\quad + \int_{y/2}^{\infty} \{(1 + y^2)^{c/2}f(x)\}g(y - x)\, dx.\end{aligned}$$

Using the lemma, the fact that $x \leq y/2$, and the finite credence of g, there exists a k^* such that for all sufficiently large y the first integral is less than $\int_{-\infty}^{y/2} f(x)(k^*/2)\, dx < k^*/2$. We can similarly make the second integral less than $k^*/2$ for all sufficiently large y, using the infinite credence of f. For $y < 0$ we attach the term $(1 + y^2)^{c/2}$ to $f(x)$ in the first integral and to $g(y - x)$ in the second.

Proof of Theorem 4. From (3.1), (3.2), and (3.3),

$$\frac{q_y(0)}{q_y(y/2)} \geq \frac{kk'}{KK'}\left(\frac{1 + y^2/4}{1 + y^2}\right)^{c'/2}(1 + y^2/4)^{c/2} \geq m(1 + y^2/4)^{c/2},$$

where $m = kk'l^{c'/2}/(KK')$ and l is given by the lemma. Therefore, $q_y(0) > q_y(y/2)$ when $|y| > 2(m^{-2/c} - 1)^{1/2}$. Similarly, $q_y(y) > q_y(y/2)$ when $|y| > 2(n^{-2/c'} - 1)^{1/2}$, where $n = kk'l^{c/2}/(KK')$. Define A as the maximum of these two bounds.

Proof of Theorem 5. (a) Using (3.1), (3.3), and (3.4),

$$\begin{aligned}q_y(x) &\leq (K'/k^*)f(x)(1 + y^2)^{c'/2}\{1 + (y - x)^2\}^{-c'/2} \\ &\leq (K'/k^*)f(x)(1 + y^2)^{c'/2}\{1 + (|y| - d)^2\}^{-c'/2}\end{aligned}$$

for all $|x| \leq d$. Now for any d there exists an A such that for all $|y| > A$, the upper bound of (4.2) holds. The lower bound is established similarly. (b) By symmetry we need only consider the case $y > 0$. Then using (3.1), (3.2), and (3.3),

$$\begin{aligned}&\Pr(X < -t(y) \mid y) \\ &= \{h(y)\}^{-1} \int_{-\infty}^{-t(y)} f(x)g(y - x)\, dx \\ &\leq \frac{K'}{k^*}(1 + y^2)^{c'/2}[1 + \{y + t(y)\}^2]^{-c'/2} \int_{-\infty}^{-t(y)} f(x)\, dx \\ &\leq \frac{K'}{k^*} \int_{-\infty}^{-t(y)} f(x)\, dx \to 0 \quad \text{as } t(y) \to \infty.\end{aligned}$$

This content downloaded from
24.240.36.145 on Sun, 16 Jun 2024 04:09:34 +00:00
All use subject to https://about.jstor.org/terms

Next consider $\Pr(X > t(y) \mid y)$. We first prove that for all $a > 0$, $\Pr(X > ay \mid y) \to 0$ and

$$\begin{aligned}
&\Pr(X > ay \mid y) \\
&= \{h(y)\}^{-1} \int_{ay}^{\infty} f(x)g(y - x)\,dx \\
&= \{h(y)\}^{-1} \int_{0}^{\infty} f(ay + z)g(y - ay - z)\,dz \\
&\le \frac{K}{k^*}(1 + y^2)^{c'/2} \int_0^{\infty} \{1 + (ay + z)^2\}^{-c/2} g(y - ay - z)\,dz \\
&< \frac{K}{k^*}(1 + y^2)^{c'/2}(1 + a^2y^2)^{-c/2} \int_0^{\infty} g(y - ay - z)\,dz \\
&< \frac{K}{k^*}(1 + y^2)^{c'/2}(1 + a^2y^2)^{-c/2} \to 0 \quad \text{as } y \to \infty. \qquad \text{(A.1)}
\end{aligned}$$

This result is sufficient to complete the proof of (4.3) if $t(y)$ tends to ∞ faster than ay for some $a > 0$. Otherwise, choose any fixed $a > 0$ and supplement (A.1) by

$$\begin{aligned}
\Pr(t(y) < X \le ay \mid y) &= \{h(y)\}^{-1} \int_{t(y)}^{ay} f(x)g(y - x)\,dx \\
&\le \frac{K'}{k^*}(1 + y^2)^{c'/2}\{1 + (1 - a)^2y^2\}^{-c'/2} \int_{t(y)}^{ay} f(x)\,dx.
\end{aligned}$$

The expression before the integral is bounded in y and tends to $(1 - a)^{-c'}$ as $y \to \infty$. The integral tends to 0 as $t(y) \to \infty$.

Proof of Theorem 6. From (3.1) to (3.4),

$$q_y(x) \le \frac{KK'}{k^*}(1 + y^2)^{c'/2}\{1 + (y - x)^2\}^{-c'/2}(1 + x^2)^{-c/2}.$$

Now the following identity is easy to prove: $1 + y^2 < 2(1 + x^2)\{1 + (y - x)^2\}$. Therefore $q_y(x) < M(1 + x^2)^{-(c-c')/2}$, where $M = 2^{c'/2}KK'/k^*$. Therefore, since $c - c' > 1 + p$,

$$\mathbf{E}(|X|^p \mid y) < M \int_{-\infty}^{\infty} |x|^p(1 + x^2)^{-(c-c')/2}\,dx < \infty.$$

Proof of Theorem 8. Consider

$$\frac{q_x(\theta)}{q_x(x_0)} = \prod_{i=0}^{n} \frac{g_i(x_i - \theta)}{g_i(x_i - x_0)}. \qquad \text{(A.2)}$$

From the credences of the g_i's, (A.2) is bounded by positive multiples of

$$\prod_{i=0}^{n} \left\{\frac{1 + (x_i - \theta)^2}{1 + (x_i - x_0)^2}\right\}^{-c_i/2} \qquad \text{(A.3)}$$

Now let $\theta = x_i + d$ for $i > 0$. Then for $j \ne i$ and $j \ne 0$

$$\frac{1 + (x_j - \theta)^2}{1 + (x_j - x_0)^2} = \frac{1 + \{z(y_j - y_i) - d\}^2}{1 + z^2(y_j - y_0)^2} \to \frac{(y_j - y_i)^2}{(y_j - y_0)^2}. \qquad \text{(A.4)}$$

Therefore, all but two terms in (A.3) are bounded. The other two terms yield

$$(1 + d^2)^{-c_i/2} \frac{[1 + z^2(y_i - y_0)^2]^{c_i/2}}{[1 + \{z(y_0 - y_i) - d\}^2]^{c_0/2}} \to 0.$$

Now let $\theta = x_0 + d$. Then (A.4) covers every term in (A.2) except $i = 0$. Therefore, (A.3) is bounded by multiples of $g_0(x_0 - \theta)$. Since in the limit the integral of (A.2) is bounded by those same multiples, $q_{\mathbf{x}}(x_0)$ is neither 0 nor infinite. The theorem now follows.

[*Received October 1988. Revised July 1989.*]

REFERENCES

Dawid, A. P. (1973), "Posterior Expectations for Large Observations," *Biometrika*, 60, 664–667.

De Finetti, B. (1961), "The Bayesian Approach to the Rejection of Outliers," in *Proceedings of the Fourth Berkeley Symposium on Probability and Statistics* (Vol. 1), Berkeley: University of California Press, pp. 199–210.

Hill, B. M. (1974), "On Coherence, Inadmissibility and Inference About Many Parameters in the Theory of Least Squares," in *Studies in Bayesian Econometrics and Statistics,* eds. S. E. Fienberg and A. Zellner, Amsterdam: North-Holland, pp. 555–584.

O'Hagan, A. (1979), "On Outlier Rejection Phenomena in Bayes Inference," *Journal of the Royal Statistical Society,* Ser. B, 41, 358–367.

——— (1988), "Modelling With Heavy Tails," in *Bayesian Statistics 3,* eds. J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith, Oxford, U.K.: Clarendon Press, pp. 345–359.

This content downloaded from
24.240.36.145 on Sun, 16 Jun 2024 04:09:34 +00:00
All use subject to https://about.jstor.org/terms