March 7, 2012

Characterization of Proper and Strictly Proper Scoring Rules for Quantiles

Mark J. Schervish, Joseph B. Kadane, and Teddy Seidenfeld
Carnegie Mellon University

Abstract

We give necessary and sufficient conditions for a scoring rule to be proper (or strictly proper) for a quantile if utility is linear, and the distribution is unrestricted. We also give results when the set of distributions is limited, for example, to distributions that have first moments.

Keywords: Elicitation.

1. Introduction. In subjective Bayesian practice, elicitation is a critical matter (Savage, 1971; O'Hagan, 1998; O'Hagan et al., 2006; Garthwaite, Kadane, and O'Hagan, 2005). Many papers (Kadane et al., 1980 and Garthwaite and Dickey, 1985, 1988) rely on elicitation of quantiles even if the parametric forms they target are expressed in terms of moments. There are two good reasons for this. First, moments are not easy to understand. Second, moments are sensitive to extreme outliers, and may even not exist. Quantiles, on the other hand, always exist and are more intuitive.

There are two principal methods of eliciting quantiles, the direct method and by use of scoring rules. In the direct method, the expert or other person being elicited is asked, for example, "what value would leave you indifferent between betting that the outcome will be greater than the value you name and less than the value you name?", to obtain a median (see Dey and Liu, 2007). Other values are then obtained by bisection.

Scoring rules, by contrast, give the expert an explicit penalty that is a function of the elicited quantile and the outcome. (Some authors, such as Gneiting and Raftery, 2007, define scoring rules to be gains to the expert rather than penalties, in which case a minus sign will convert results from one definition to the other.) The implicit assumption is that the expert's utility is linear in the score. For example, absolute error is often thought of as a scoring rule for the median. A scoring rule is called *proper* if it is minimized at the desired quantile. It is *strictly proper* if it is minimized only at the desired quantile. Gneiting and Raftery (2007), following Cervera and Munoz (1996), propose a specific class of scoring rules that they prove to be proper for a certain restricted class of distributions. Additionally they write "We do not know whether [this class] provides the general form of proper scoring rule for quantiles."

To address this issue, there are three problems to overcome. The first is that the condition that a scoring rule be proper is too weak to be useful. For example, a constant function is a proper scoring rule (because the desired quantile minimizes it), but so does every other

possible elicited value. Hence the main focus has to be on strictly proper scoring rules. Second, quantiles are not necessarily unique. For example, the pdf

$$f(x) = \begin{cases} 1/2 & 0 < x < 1 \\ 1/2 & 2 < x < 3 \\ 0 & \text{otherwise} \end{cases}$$

has the whole interval [1,2] as medians. Finally, to be valid no matter what the underlying distribution is, one has to be careful about infinite expected scores. For example, absolute error under a Cauchy distribution has infinite expected score for all choices of median, making it not strictly proper.

In response to these challenges, we focus on strictly proper scoring rules (although we find a characterization of proper scoring rules as well). We define quantiles in a way that yield a closed interval in general (and specializes to a point in the case of an absolutely continuous distribution with positive density at the quantile). We extend the notion of proper scoring rule to mean that every point in this closed interval minimizes expected score, and the notion of strictly proper scoring rule to mean that only these points do so. Finally, our main result is valid for the class of all distributions, although we also give results for classes of distributions having certain finite moments.

2. Characterization of proper scoring rules. In this section, we give a characterization of (strictly) proper scoring rules for a single quantile in Theorem 1. Throughout this discussion, $0 < \alpha < 1$, P is a probability distribution over a set X of real numbers, and X is a random variable whose distribution is P. Let $F(x) = P(X \le x)$ stand for the c.d.f. of X, and let $G(x) = P(X \ge x)$ be the symmetrically defined function for the upper tail. It is useful to note that, for all $a < b$,

(1) $$P(a < X < b) = 1 - G(b) - F(a):$$

We use the functions F and G to define quantiles.

D

LEMMA 1. *Let X be a random variable with distribution P. The set of quantiles of X is a nonempty closed interval $[q\,;q\,]$. Also,*

If $q\,>q\,$, then $G(q) = 1$ for all $q \in (q\,;q\,]$ and $F(q) =$ for all $q \in [q\,;q\,)$.

If $F(q) =$ and $t > q\,$, then $P[q < X < t] > 0$.

If $G(q) = 1$ and $t < q\,$, then $P[t < X < q] > 0$.

PROOF. Define $A = \{x : F(x) \quad g$ and $B = \{x : G(x) \quad 1 \quad g$. Define $q = \inf A$ and $q = \sup B$. Since $F(x) \quad F(q\,)$, for all $x > q\,$, it follows that $F(q\,)$. Similarly, $G(q\,) \quad 1$. Note that both A and B are semi-infinite intervals, and both q and q are finite. Every $x < q\,$ is in B and every $x > q\,$ is in A, so that $q \quad q\,$. The set of quantiles is $A \setminus B$, which is a closed and bounded interval.

For the first bullet, if $F(q\,) >$, no number greater than $q\,$ is in B, hence a necessary condition for $q\,>q\,$ is $F(q\,) =$. Similarly, a necessary condition for $q\,>q\,$ is $G(q\,) = 1$. For every $t < u$ in the open interval $(q\,;q\,)$, $F(t) = F(q\,)$ and $G(u) = G(q\,)$, hence the first bullet holds.

For the second bullet, it is clear that $G(t) < 1$ for all $t > q\,$. Then (1) implies

$$P(q < X < t) = 1 \quad G(t) \quad F(q) > 1 \quad (1 \quad) \quad = 0:$$

The third bullet is proven in similar fashion, since $F(t) <$ for all $t < q\,$.

Let $g\, : X \quad R \,! \, \mathbb{R}$ be a function, where R is the set of allowed values for the quantile of interest, e.g., the convex hull of X, or even its closure. Suppose that we want to use $g\,(x;q)$ as the penalty to an elicitee for giving q as the quantile of P when $X = x$ is observed238(63 0 Td [(>)]Tv)25(al.)]TJ -24.n.5521/F24/F25 7.96n.55210(in)521/40(oin)52125(vinin)521494

there is a number $d(a; b; \tau)$ such that

(3) $$g_\tau(b; q) = d(a; b; \tau) - \frac{}{1-} g_\tau(a; q);$$

for all $q \in [a; b]$, and

$$g_\tau(a; a) \quad g_\tau(a; b).$$

The scoring rule is strictly proper for the quantiles of P_0 if and only if, in addition to the above conditions, $g_\tau(a; a) < g_\tau(a; b)$.

PROOF. Let P be the distribution such that $P(X = a) = $. Every number in the interval $[a; b]$ is an quantile of X. A necessary condition for g_τ to be proper is that $P[g_\tau(X; q)]$ be constant as a function of q. Trivially,

(4) $$P[g_\tau(X; q)] = g_\tau(a; q) + (1 -)g_\tau(b; q):$$

This is constant in q if and only if there is a number $d(a; b; \tau)$ such that (3) holds for all $q \in [a; b]$. Another way to understand (3) is that, aside from a shift of level, $g_\tau(a;)$ and $(1 -)g_\tau(b;)$ are curves with slopes that are negatives of each other for all $q \in [a; b]$. Another necessary condition for g_τ to be proper comes from consideration of P_p, the distribution such that $P_p(X = a) = p$.

$$P_p[g_\tau(X; q)] = pg_\tau(a; q) + (1 - p)g_\tau(b; q)$$

$$= pg_\tau(a; q) + (1 - p)\left[d(a; b; \tau) - \frac{}{1-}g_\tau(a; q)\right]$$

$$= g_\tau(a; q)\left[p - \frac{(1 - p)}{1-}\right] + (1 - p)d(a; b; \tau):$$

The minimum of this expression, as a function of q, occurs at the minimum or maximum for $g_\tau(a;)$

THEOREM 1. *Let P_0 be a collection of probability distributions on a subset X of $\mathbb{R}$ such that, for every $a, b \in X$, P_0 contains every distribution concentrated on $\{a, b\}$. Let g be a real-valued function defined on $X \times R$, where R is the convex hull of X. Then g is a (strictly) proper scoring rule for the α quantiles of P_0 if and only if for each $P \in P_0$ $P[g^*(X; q)]$ exists (is finite), and there exists a (strictly) increasing function s such that*

(5)
$$g^*(x; q) \equiv g(x; q) - g(x; x) = \begin{cases} \alpha[s(x) - s(q)] & \text{if } x > q, \\ (1 - \alpha)[s(q) - s(x)] & \text{if } x < q. \end{cases}$$

PROOF. First, we prove necessity by assuming that g is (strictly) proper for the set of all two-point distributions. This allows us to assume that g satisfies the two bullets in Lemma 2. Define $g^*(x; q) = g(x; q) - g(x; x)$. The second bullet of Lemma 2, namely that $g^*(a; a) \le g^*(a; b)$, when applied to all $a < b$ implies, among other things, that $g^*(a; q)$ and $g^*(a; q)$ are monotone increasing in q for $q > a$ and monotone decreasing in q for $q < a$ (with strict monotonicity in the strictly proper case). From this it follows that $g^*(x; q) \ge 0$ for all x, q.

Rewrite (3) as

(6)
$$d(a; b; \alpha) = \alpha g^*(b; q) + \frac{\alpha}{1 - \alpha} g^*(a; q);$$

for all $a \le q \le b$. Substitute $q = a$ and $q = b$ on the right side of (6) and set the two results equal to each other to obtain

(7)
$$\alpha g^*(b; a) + \frac{\alpha}{1 - \alpha} g^*(a; a) = \alpha g^*(b; b) + \frac{\alpha}{1 - \alpha} g^*(a; b);$$

which can be rewritten as

(8)
$$(1 - \alpha) g^*(b; a) = \alpha g^*(a; b).$$

In other words, specifying $g^*(x; q)$ for $x > q$ determines its values for $x < q$ by (8).

Let $r(q) = g^*(0; q)$, which we have already shown is (strictly) increasing for $q > 0$ and (strictly) decreasing for $q < 0$. It is more convenient to work with a monotone function such as

(9)
$$s(q) = \begin{cases} \dfrac{r(q)}{1 - \alpha} & \text{if } q \ge 0, \\[2mm] -\dfrac{r(q)}{\alpha} & \text{if } q < 0, \end{cases}$$

which is (strictly) increasing. Next, set the right sides of (6) and (7) equal to each other:

$$\alpha g^*(b; q) + \frac{\alpha}{1 - \alpha} g^*(a; q) = \alpha g^*(b; b) + \frac{\alpha}{1 - \alpha} g^*(a; b);$$

which rearranges to become

(10)
$$g^*(b; q) = \frac{1}{1 - \alpha}[g^*(a; b) - g^*(a; q)];$$

for $a \le q \le b$. Hence, for $0 \le q \le x$, we have from (10) (with $a = 0$ and $b = x$),

$$g^*(x; q) = \frac{1}{1 - \alpha}[r(x) - r(q)] = \alpha[s(x) - s(q)].$$

5

For $q \quad 0 \quad x$, we have from (10),

$$g \ ($$

2. If q is a quantile and t is not a quantile, then $P[h_t(X; q)] \geq 0$, with strict inequality for unique minimization.

Let q be an α quantile of X, and let $t < q$. Then, we can write

$$P[h_t(X; q)] = [s(q) \quad s(t)][(1 \quad)F(t) \quad G(q)] + [(1 \quad)s(q) + \quad s(t)]P[t < X < q]$$
$$\int_{(t;q)} s(x)\,dP(x)$$

$$(14) \qquad = s(t)[\quad F(t)] + s(q)[1 \quad G(q)] \quad \int_{(t;q)} s(x)\,dP(x):$$

First, assume that t is also an α quantile of P. Because $t < q$, $q > q$ and $t < q$ so that $F(t) = \quad$ and $G(q) = 1 \quad$ according to Lemma 1, and $P[t < X < q] = 0$. This establishes condition 1.

Next, let q be an α quantile of X, and let $t < q$ not be an α quantile. In this case, $G(q) \quad 1 \quad$, and

$$(15) \qquad \int_{(t;q)} s(x)\,dP(x) \quad s(t)[1 \quad F(t) \quad G(q)]:$$

There are two cases. (i) If $G(q) = 1 \quad$, then Lemma 1 says that $P(t < X < q) > 0$, which implies that the inequality in (15) is strict. It follows that (14) is strictly less than $[1 \quad G(q)][s(q) \quad s(t)] = 0$. (ii) If $G(q) > 1 \quad$, then (14) is at most $[1 \quad G(q)][s(q) \quad s(t)] \quad 0$ with strict inequality if s is strictly increasing. This establishes condition 2.

Next, let q be an α quantile of X, and let $t > q$. Then, we can write

$$P[h_t(X; q)] = [s(q) \quad s(t)][(1 \quad)F(q) \quad G(t)] \quad [(1 \quad)s(t) + \quad s(q)]P[q < X < t]$$
$$+ \int_{(q;t)} s(x)\,dP(x)$$

$$(16) \qquad = s(q)[\quad F(q)] \quad s(t)[1 \quad G(t)] + \int_{(q;t)} s(x)\,dP(x):$$

First, assume that t is also an α quantile of P. Because $t > q$, $t > q$ and $q < q$ so that $F(q) = \quad$ and $G(t) = 1 \quad$ according to Lemma 1, and $P[q < X < t] = 0$. This establishes condition 1.

Finally, let q be an α quantile of X, and let $t > q$ not be an α quantile. In this case, $F(q)$

EXAMPLE 2. The scoring rule $g_{1=2}(x; q) = |x - q|$ in Example 1 corresponds to $s(x) = 2x$ in Theorem 1 with $g_{1=2}(x; x) \equiv 0$. If desired, one could extend this example to all distributions by setting $g_{1=2}(x; x) = -|x|$. The scoring rule would no longer be $|x - q|$, but $g_{1=2}(x; q) - g_{1=2}(x; t)$ would be the same as it would be in Example 1 for all

where c is chosen to make $\sum_{n=1}^{\infty} P(\{x_n\}) = 1$. Then

$$P[|f(X)|] = \sum_{n=1}^{\infty} \frac{c|f(x_n)|}{[1 + |f(x_n)|]n^2} < 1 \; ;$$

so that $P \in P_0$. Also,

$$P[|h(X)|] = \sum_{n=1}^{\infty} \frac{c|h(x_n)|}{[1 + |f(x_n)|]n^2} \le \sum_{n=1}^{\infty} \frac{cn[1 + |f(x_n)|]}{[1 + |f(x_n)|]n^2} = 1 \; .$$

Lemma 3 lets us identify all strictly proper scoring rules for classes of distributions defined by certain means being finite.

THEOREM 2. *Let F be a collection of real-valued functions defined on $\mathbb{R}$. Let P_0 be the set of all probability distributions on $\mathbb{R}$ that give finite mean to $f(X)$ for every $f \in F$. Define H to be the set of all functions h that satisfy (18) for all $f \in F$. Then g is a strictly proper scoring rule for the τ quantiles of P_0 if and only if there exists $h \in H$ and a strictly increasing function s such that, for all real t, $g(x; q) = h(x) + h_t(x; q)$, where h_t is defined in (13).*

PROOF. First note that all distributions supported on two points are in P_0, no matter what F is. For the "if" direction, we already showed in the proof of Theorem 1 that every h_t of the form (13) is strictly proper for the class of all probability measures. Adding a function that has finite mean for all $P \in P_0$ produces another strictly proper scoring rule for P_0. For the "only if" direction, assume that g is strictly proper. It follows that $P[g(X; t)]$ is finite for every $P \in P_0$ and every real t, hence $h(x) = g(x; t) \in H$. In the proof of Theorem 1, we showed that there is a strictly increasing s such that, for every real t, $g(x; q) - g(x; t) = h_t(x; q)$ from (13). Hence $g(x; q) = h(x) + h_t(x; q)$.

4. Conclusion.   This paper characterizes proper and strictly proper scoring rules for a quantile if utility is linear in score. Kiefer (2010) following Karni (2009) finds a proper scoring rule under more general risk-averse utility. Characterizing proper and strictly proper scoring rules under those conditions remains open.

It is obvious that the sum of (strictly) proper scoring rules for several quantities is (strictly) proper. But are there others? The characterization of such rules is also currently unsolved.

REFERENCES

Cervera, J. and Munoz, J. (1996), "Proper scoring rules for fractiles," in *Bayesian Statistics 5*, eds. J. Bernardo, J. Berger, A. Dawid, and A. Smith, U.K.: Oxford University Press, pp. 513–519.

Dey, D. and Liu, J. (2007), "A quantitative study of quantile based direct prior elicitation from expert opinion," *Bayesian Analysis*, 2, 137–166.

Garthwaite, P. and Dickey, J. (1985), "Double- and single-bisection methods for subjective probability assessment in a location-scale family," *Journal of Econometrics*, 29, 149–163.

——— (1988), "Quantifying expert opinion in linear regression problems," *Journal of the Royal Statistical Society*, 50(Series B), 462–474.

Garthwaite, P., Kadane, J., and O'Hagan, A. (2005), "Statistical Methods for eliciting probability distributions," *Journal of the American Statistical Association*, 100, 680–701.

Gneiting, T. and Raftery, A. (2007), "Strictly Proper Scoring Rules, Prediction and Estimation," *Journal of the American Statistical Association*, 102, 359–378.

Kadane, J., Dickey, J., Winkler, R., Smith, W., and Peters, S. (1980), "Interactive elicitation fro the normal linear model," *Journal of the American Statistical Association*, 75, 845–854.

Karni, E. (2009), "A mechanism for eliciting probabilities," *Econometrica*, 77, 603–606.

Kiefer, N. (2010), "Incentive-compatible elicitation of quantiles," unpublished manuscript.

O'Hagan, A. (1998), "Eliciting expert beliefs in substantial practical applications," *The Statistician*, 47, 21–35.

O'Hagan, A., Buck, C., Daneshkhah, A., Eiser, J., Garthwaite, P., Jenkinson, D., Oakley, J., and Rakow, T. (2006), *Uncertain Judgement: Eliciting Expert Probabilities*, Chichester: J. Wiley and Sons.

Savage, L. (1971), "Elicitation of Personal Probabilities and Expectations," *Journal of the American Statistical Association*, 66, 783–801.