IMS Collections
From Probability to Statistics and Back: High-Dimensional Models and Processes
Vol. 9 (2013) 163–173

DOI: 10.1214/12-IMSCOLL912

Consistent scoring functions for quantiles

Kyrill Grant[1] **and Tilmann Gneiting**[1,*]

Universität Heidelberg

Abstract: A scoring function is consistent for the α-quantile functional if, and only if, it is generalized piecewise linear (GPL) of order α, up to equivalence. Expressed differently, loss functions that yield quantiles as Bayes rules are GPL functions. We review and discuss this basic decision-theoretic result with focus on Thomson's pioneering characterization.

1. Introduction

As is well known, if F is a probability measure on the real line, $\mathbb{R}$, with finite first moment, the function

$$h : \mathbb{R} \to \mathbb{R}, \quad x \mapsto h(x) = \mathbb{E}_F|x - Y|,$$

where the random variable Y has distribution F, attains a global minimum in any median of F. Equivalently, the median is the Bayes rule or Bayes predictor under the linear loss function. Our preferred way of stating this basic decision-theoretic fact is that *the scoring function* $\mathrm{S}(x, y) = |x-y|$ *is strictly consistent for the median functional relative to the class of probability measures on* $\mathbb{R}$ *with finite first moment.*

In this paper, we study scoring functions that are consistent for quantiles. Given an interval $\mathrm{I} \subseteq \mathbb{R}$, we refer to the corresponding Borel σ-algebra as $\mathcal{B}(\mathrm{I})$. A *scoring function* is any map $\mathrm{S} : \mathrm{I} \times \mathrm{I} \to \mathbb{R}$ that is measurable from $\mathcal{B}(\mathrm{I})$ to $\mathcal{B}(\mathbb{R})$ in the second argument, for each fixed value of the first argument. Given a family $\mathcal{F}$ of probability measures on $(\mathrm{I}, \mathcal{B}(\mathrm{I}))$, a *statistical functional*, or simply a *functional*, is a map $\mathrm{T} : \mathcal{F} \to \mathrm{P}(\mathrm{I})$, where $\mathrm{P}(\mathrm{I})$ denotes the power set of the interval I. Statistical functionals feature prominently in Jon Wellner's widely circulated lecture notes [13], where a wealth of examples can be found.

We distinguish four types of consistency, which can be traced to the fundamental work of Savage [11], Thomson [12] and Osband [9].

Definition 1.1. Let $\mathrm{I} \subseteq \mathbb{R}$ be an interval, $\mathcal{F}$ a family of probability measures on $(\mathrm{I}, \mathcal{B}(\mathrm{I}))$, $\mathrm{T} : \mathcal{F} \to \mathrm{P}(\mathrm{I})$ a statistical functional, and $\mathrm{S} : \mathrm{I} \times \mathrm{I} \to \mathbb{R}$ a scoring function.

(a) The scoring function S is *consistent* for the functional T relative to the class $\mathcal{F}$, if $\mathbb{E}_F\, \mathrm{S}(x, Y)$ exists and is finite for all $F \in \mathcal{F}$ and $x \in \mathrm{I}$, and if

$$\mathbb{E}_F\, \mathrm{S}(t, Y) \le \mathbb{E}_F\, \mathrm{S}(x, Y)$$

for all $F \in \mathcal{F}$, all $t \in \mathrm{T}(F)$, and all $x \in \mathrm{I}$.

*Tilmann Gneiting thanks Werner Ehm for helpful discussions and acknowledges support by the Alfried Krupp von und zu Behlen Foundation.

[1] Universität Heidelberg, Institut für Angewandte Mathematik, Im Neuenheimer Feld 294, 69120 Heidelberg, Germany, e-mail: kirill2@gmx.de; t.gneiting@uni-heidelberg.de

AMS 2000 subject classifications: Primary 62C05; secondary 91B06

Keywords and phrases: Bayes rule, consistent scoring function, fractile, optimal point forecast, proper scoring rule, quantile

(b) The scoring function S is *strictly consistent* for the functional T relative to the class $\mathcal{F}$, if S is consistent for T relative to $\mathcal{F}$, and if $\mathbb{E}_F \mathrm{S}(t, Y) = \mathbb{E}_F \mathrm{S}(x, Y)$, where $t \in \mathrm{T}(F)$, implies that $x \in \mathrm{T}(F)$.
(c) The scoring function S is *weakly consistent* for the functional T relative to the class $\mathcal{F}$, if $\mathbb{E}_F[\mathrm{S}(x_1, Y) - \mathrm{S}(x_2, Y)]$ exists and is finite for all $F \in \mathcal{F}$ and $x_1, x_2 \in \mathrm{I}$, and if

$$\mathbb{E}_F\big[\mathrm{S}(t, Y) - \mathrm{S}(x, Y)\big] \le 0$$

for all $F \in \mathcal{F}$, all $t \in \mathrm{T}(F)$, and all $x \in \mathrm{I}$.
(d) The scoring function S is *strictly weakly consistent* for the functional T relative to the class $\mathcal{F}$, if S is weakly consistent for T relative to $\mathcal{F}$, and if $\mathbb{E}_F[\mathrm{S}(t, Y) - \mathrm{S}(x, Y)] = 0$, where $t \in \mathrm{T}(F)$, implies that $x \in \mathrm{T}(F)$.

We consider the following classes $\mathcal{F}$.

Definition 1.2. Let $\mathrm{I} \subseteq \mathbb{R}$ be an interval, and let $\mathcal{F}_\mathrm{I}$ denote the class of the probability measures on $(\mathrm{I}, \mathcal{B}(\mathrm{I}))$.

(a) Given a measurable function $g : \mathrm{I} \to \mathbb{R}$, the class $\mathcal{F}_g$ consists of the probability measures $F \in \mathcal{F}_\mathrm{I}$ for which $\mathbb{E}_F\, g(Y)$ exists and is finite.
(b) The class $\mathcal{D}_n$ consists of the discrete n-point measures within $\mathcal{F}_\mathrm{I}$.
(c) The class $\mathcal{L}$ consists of the probability measures in $\mathcal{F}_\mathrm{I}$ that are absolutely continuous with respect to the Lebesgue measure. The subclass $\mathcal{L}_+$ comprises the members that admit strictly positive densities.

As noted, we study scoring functions that are consistent for the α-quantile functional. Equivalently, we are concerned with loss functions that lead to quantiles as Bayes rules or optimal point forecasts. The characterization of these functions has an interesting and varied history, for which we refer to [2].[1]

Concise results and proofs become available if one puts smoothness conditions on the scoring function and uses the notion of a *generalized piecewise linear* (GPL) function of order α, that is, a function $\mathrm{S}_g : \mathrm{I} \times \mathrm{I} \to [0, \infty)$ of the form

$$\mathrm{S}_g(x, y) = \begin{cases} (1-\alpha)\big(g(x) - g(y)\big), & y \le x, \\ \alpha\,\big(g(y) - g(x)\big), & y > x, \end{cases}$$

where $g : \mathrm{I} \to \mathbb{R}$ is nondecreasing. When g is linear, we recover the asymmetric piecewise linear loss function that lies at the heart of quantile regression [7].

We then have the following characterization. The sufficiency part (Theorem 1.3) is well known and does not depend on smoothness conditions, as opposed to the necessity part (Theorem 1.4), which we state in the form given in [2].

Theorem 1.3. *Let* I *be an interval, and let* S *be a* GPL *function of order* α. *Then* S *is a scoring function and the following holds:*

(a) The scoring function S *is consistent for the* α*-quantile relative to the class* $\mathcal{F}_g$.
(b) If g *is strictly increasing, then* S *is strictly consistent for the* α*-quantile relative to the class* $\mathcal{F}_g$.

Theorem 1.4. *Suppose that the scoring function* $\mathrm{S} : \mathrm{I} \times \mathrm{I} \to \mathbb{R}$ *is consistent for the* α*-quantile relative to the class* $\mathcal{D}_2$*, and such that*

[1] Being unaware of the work of Thomson [12] and Saerens [10], various authors described the characterization as an open problem, including Cervera and Muñoz [1], Gneiting and Raftery [4] and Jose and Winkler [6].

(S0) $\mathrm{S}(x,y) \geq 0$ *for all* $x, y \in \mathrm{I}$ *with equality if* $x = y$,
(S1) $\mathrm{S}(x,y)$ *is continuous in* x,
(S2) *the partial derivative* $\frac{\partial \mathrm{S}}{\partial x}(x,y)$ *exists and is continuous in* x *if* $x \neq y$.

Then S *is a* GPL *function of order* α.

There is an asymmetry in the characterization, in that the necessity part depends on stringent regularity conditions. In a pioneering yet little known paper, Thomson [12] established a characterization under much weaker conditions. However, his results are complicated and do not readily reveal the GPL structure. In the balance of the paper, we review his proof and characterization and provide analogues of Theorems 1.3 and 1.4 that depend on minimal conditions only.

2. Thomson's characterization

In what ways can the regularity conditions (S0), (S1) and (S2) of Theorem 1.4 be relaxed? The nonnegativity condition (S0) is standard and not restrictive. Indeed, if S_0 is such that $\mathrm{S}_0(x,y) \geq \mathrm{S}_0(y,y)$ for all $x, y \in \mathrm{I}$, which is a natural assumption on a loss or scoring function, then $\mathrm{S}(x,y) = \mathrm{S}_0(x,y) - \mathrm{S}_0(y,y)$ satisfies (S0) and leads to the same Bayes rule, subject to integrability conditions. In contrast, assumptions (S1) and (S2) are restrictive.

If we abandon any assumption of continuity on the scoring function, it is natural to let $\mathrm{I} = (a,b)$ be an open interval, where possibly $a = -\infty$ and/or $b = \infty$, and to consider consistency relative to the class $\mathcal{L}$ of the probability measures with Lebesgue densities. The following result then is nearly immediate.

Theorem 2.1. *Let* I *be an open interval. Suppose that* $\mathrm{S}_g : \mathrm{I} \times \mathrm{I} \to [0,\infty)$ *is a* GPL *function of order* α, *and let* $h : \mathrm{I} \to \mathbb{R}$ *be measurable. Suppose that the scoring function* $\mathrm{S} : \mathrm{I} \times \mathrm{I} \to \mathbb{R}$ *is such that for all* $x \in \mathrm{I}$,

$$\mathrm{S}(x,y) = \mathrm{S}_g(x,y) + h(y)$$

for all $y \in \mathrm{I} \setminus \mathrm{N}_x$, *where* N_x *is a Lebesgue null set. Then the following holds:*

(a) The scoring function S *is consistent for the* α*-quantile relative to the class* $\mathcal{L} \cap \mathcal{F}_g \cap \mathcal{F}_h$.
(b) If g *is strictly increasing, then* S *is strictly consistent for the* α*-quantile relative to the class* $\mathcal{L} \cap \mathcal{F}_g \cap \mathcal{F}_h$.
(c) The scoring function S *is weakly consistent for the* α*-quantile relative to the class* $\mathcal{L}$.
(d) If g *is strictly increasing, then* S *is strictly weakly consistent for the* α*-quantile relative to the class* $\mathcal{L}$.

As regards sufficiency, the following theorem assumes consistency relative to the subclass $\mathcal{L}_+$. While the result and the proof are due to Thomson [12], there are distinctions in detail, as summarized in [5, p. 36].

Theorem 2.2 (Thomson 1979)**.** *Let* I *be an open interval. Suppose that the scoring function* $\mathrm{S} : \mathrm{I} \times \mathrm{I} \to \mathbb{R}$ *is weakly consistent for the* α*-quantile relative to the class* $\mathcal{L}_+$. *Then there exist a GPL function* $\mathrm{S}_g : \mathrm{I} \times \mathrm{I} \to \mathbb{R}$ *and a measurable function* $h : \mathrm{I} \to \mathbb{R}$ *such that, for any* $x \in \mathrm{I}$,

$$\mathrm{S}(x,y) = \mathrm{S}_g(x,y) + h(y)$$

for all $y \in \mathrm{I} \setminus \mathrm{N}_x$, *where* N_x *is a Lebesgue null set.*

Any such result depends on a delicate trade-off, in that any assumptions on the scoring function S, the choice of the class $\mathcal{F}$, and the notion of consistency employed need to be balanced. In this regard, Thomson's choices appear to be usefully general and achieving a reasonable balance. That said, a careful inspection of the proof in Section 4 shows that Theorem 2.2 remains valid relative to smaller classes $\mathcal{F}$, with the class of piecewise Gaussian probability measures being one such example.

3. Proof of Theorem 2.1

For completeness, we prove parts (c) and (d) of this well known result. Let $F \in \mathcal{L}$. If $x_1, x_2 \in \mathrm{I}$ with $x_1 < x_2$ then

$$\begin{aligned}\mathbb{E}_F\big[\mathrm{S}(x_1,Y) - \mathrm{S}(x_2,Y)\big] = {} & (1-\alpha)\big(g(x_1) - g(x_2)\big)F((a,x_1]) \\ & + \int_{x_1}^{x_2} \big[g(y) - \alpha g(x_1) - (1-\alpha)g(x_2)\big]\, dF(y) \\ & + \alpha\big(g(x_2) - g(x_1)\big)F\big((x_2,b)\big)\end{aligned}$$

exists and is finite. In particular, if $x_1 = q_\alpha$ is an α-quantile of F and $x_2 = x > q_\alpha$, then $F((x,b)) = F((q_\alpha,b)) - F((q_\alpha,x])) = (1-\alpha) - F((q_\alpha,x])$ so that the above equality simplifies to

$$\mathbb{E}_F\big[\mathrm{S}(q_\alpha,Y) - \mathrm{S}(x,Y)\big] = \int_{q_\alpha}^{x} \big[g(y) - g(x)\big]\, dF(y),$$

where the right-hand side is nonnegative, and strictly positive if g is strictly increasing and x is not an α-quantile of F. An analogous argument applies when $x < q_\alpha$.

4. Proof of Theorem 2.2

To recall the setting, we let $\mathrm{I} = (a,b)$ an open interval, where possibly $a = -\infty$ and/or $b = \infty$. Given $\alpha \in (0,1)$, we suppose that the scoring function $\mathrm{S} : \mathrm{I} \times \mathrm{I} \to \mathbb{R}$ is weakly consistent for the α-quantile relative to the class $\mathcal{L}_+$ of the probability measures on $(\mathrm{I}, \mathcal{B}(\mathrm{I}))$ that admit a strictly positive density with respect to the Lebesgue measure, λ. *All results stated in this section depend on these assumptions.* In what follows, densities are Lebesgue densities, null sets are Lebesgue null sets, and the term almost everywhere is used with respect to the Lebesgue measure. As noted, we review the ingenious proof of Thomson [12], tying up loose ends and closing a few minor gaps, with our contributions being summarized in [5, p. 36].

For $x_1, x_2 \in \mathrm{I}$ with $x_1 < x_2$, and $m_l, M_l, m_r, M_r \in \mathbb{R}$ with $m_l < M_l$ and $m_r < M_r$, we consider the function

$$\Delta(x_1,x_2,y) = \mathrm{S}(x_1,y) - \mathrm{S}(x_2,y)$$

and the Borel sets

$$\begin{aligned}L_{\le m_l}(x_1,x_2) &= \big\{y \in \mathrm{I} : y \le x_1, \Delta(x_1,x_2,y) \le m_l\big\}, \\ L_{\ge M_l}(x_1,x_2) &= \big\{y \in \mathrm{I} : y \le x_1, \Delta(x_1,x_2,y) \ge M_l\big\}, \\ R_{\le m_r}(x_1,x_2) &= \big\{y \in \mathrm{I} : y \ge x_2, \Delta(x_1,x_2,y) \le m_r\big\}, \\ R_{\ge M_r}(x_1,x_2) &= \big\{y \in \mathrm{I} : y \ge x_2, \Delta(x_1,x_2,y) \ge M_r\big\}.\end{aligned}$$

For notational convenience, we frequently suppress the dependence on x_1 and x_2, and write $\Delta(y)$, and $L_{\le m_l}$, $L_{\ge M_l}$, $R_{\le m_r}$ and $R_{\ge M_r}$, respectively.

Lemma 4.1. *There are no $x_1, x_2 \in \mathrm{I}$ with $x_1 < x_2$, and $m_l, M_l, m_r, M_r \in \mathbb{R}$ with $m_l < M_l$ and $m_r < M_r$, such that*

$$\begin{aligned} \lambda\big(L_{\le m_l}(x_1,x_2)\big) &> 0, \qquad \lambda\big(R_{\le m_r}(x_1,x_2)\big) > 0, \\ \lambda\big(L_{\ge M_l}(x_1,x_2)\big) &> 0, \qquad \lambda\big(R_{\ge M_r}(x_1,x_2)\big) > 0. \end{aligned}$$

Proof. For a contradiction, suppose that the above inequalities hold. If we let

$$c > \max\left(\frac{\alpha}{1-\alpha}, \frac{1-\alpha}{\alpha}\right) \tag{1}$$

be a fixed constant, there exists a sequence $(F_n)_{n=1,2,\dots}$ of probability measures in $\mathcal{L}_+$ with densities $(f_n)_{n=1,2,\dots}$ such that

$$F_n(L_{\ge M_l}) = \alpha - \frac{1}{n+c}, \quad F_n((a,x_1]) = \alpha, \quad F_n(R_{\ge M_r}) = 1-\alpha-\frac{1}{n+c},$$

and

$$f_n(y) = \frac{1+c}{n+c}\, f_1(y) \quad \text{for } y \in \mathrm{I} \setminus (L_{\ge M_l} \cup R_{\ge M_r}).$$

As x_1 is an α-quantile of F and S is weakly consistent for the α-quantile relative to the class $\mathcal{L}_+$, we have

$$\begin{aligned} 0 &\ge \mathbb{E}_{F_n}\left[\Delta(x_1,x_2,Y)\right] \\ &= \int_{L_{\ge M_l}} \Delta(y)\,dF_n(y) + \int_{R_{\ge M_r}} \Delta(y)\,dF_n(y) + \int_{\mathrm{I}\setminus(L_{\ge M_l}\cup R_{\ge M_r})} \Delta(y)\,dF_n(y) \\ &\ge M_l\left(\alpha - \frac{1}{n+c}\right) + M_r\left(1-\alpha-\frac{1}{n+c}\right) + \frac{1+c}{n+c}\int_{\mathrm{I}\setminus(L_{\ge M_l}\cup R_{\ge M_r})} \Delta(y)\,dF_1(y) \end{aligned}$$

for $n = 1, 2, \dots$ In the limit as $n \to \infty$ we get $M_l\alpha + M_r(1-\alpha) \le 0$, and an analogous argument leads to $m_l\alpha + m_r(1-\alpha) \ge 0$. Combining the two inequalities, we obtain

$$(m_l - M_l)\alpha + (m_r - M_r)(1-\alpha) \ge 0,$$

contrary to the assumption that $m_l < M_l$ and $m_r < M_r$. □

Lemma 4.2. *There are no $x_1, x_2 \in \mathrm{I}$ with $x_1 < x_2$ and $m_l, M_l, m_r, M_r \in \mathbb{R}$ with $m_l < M_l$ and $m_r < M_r$, such that*

$$\begin{aligned} \lambda\big(L_{\le m_l}(x_1,x_2)\big) &> 0, \qquad \lambda\big(R_{\le m_r}(x_1,x_2)\big) > 0, \\ \lambda\big(L_{\ge M_l}(x_1,x_2)\big) &> 0, \qquad \lambda\big(R_{\ge M_r}(x_1,x_2)\big) = 0. \end{aligned}$$

Proof. For a contradiction, suppose that the stated relations hold. If there is no constant $d \in \mathbb{R}$ such that

$$\lambda\big(\{y \ge x_2 \mid \Delta(y) \ne d\}\big) = 0, \tag{2}$$

we can find constants $m_r', M_r' \in \mathbb{R}$ with $m_r' < M_r'$ such that both $\lambda(R_{\le m_r'})$ and $\lambda(R_{\ge M_r'})$ are strictly positive, contrary to Lemma 4.1. Thus, we may assume that there is some $d \in \mathbb{R}$ for which (2) holds.

Then there exist a constant c satisfying (1) and a sequence $(F_n)_{n=1,2,\dots}$ of probability measures in $\mathcal{L}_+$ with densities $(f_n)_{n=1,2,\dots}$, such that

$$F_n(L_{\ge M_l}) = \alpha - \frac{1}{n+c}, \quad F_n\big((a,x_1]\big) = \alpha, \quad F_n\big([x_2,b)\big) = 1-\alpha-\frac{1}{n+c},$$

and

$$f_n(y) = \frac{1+c}{n+c} f_1(y) \quad \text{for} \quad y \in \mathrm{I} \setminus \big(L_{\geq M_l} \cup [x_2, b)\big).$$

As x_1 is an α-quantile of F and S is weakly consistent for the α-quantile relative to the class $\mathcal{L}_+$, we have

$$\begin{aligned}
0 &\geq \mathbb{E}_{F_n}\big[\Delta(x_1, x_2, Y)\big] \\
&= \int_{L_{\geq M_l}} \Delta(y)\, dF_n(y) + \int_{x_2}^{b} \Delta(y)\, dF_n(y) + \int_{\mathrm{I}\setminus(L_{\geq M_l} \cup (x_2, b])} \Delta(y)\, dF_n(y) \\
&\geq M_l\left(\alpha - \frac{1}{n+c}\right) + d\left(1 - \alpha - \frac{1}{n+c}\right) + \frac{1+c}{n+c} \int_{\mathrm{I}\setminus(L_{\geq M_l} \cup R_{\geq M_r})} \Delta(y)\, dF_1(y)
\end{aligned}$$

for $n = 1, 2, \ldots$ In the limit as $n \to \infty$ we get $M_l \alpha + d(1-\alpha) \leq 0$, and an analogous argument leads to $m_l \alpha + d(1-\alpha) \geq 0$. Combining the two inequalities, we obtain $(m_l - M_l)\alpha \geq 0$, contrary to the assumption that $m_l < M_l$. □

Lemma 4.3. *There are no $x_1, x_2 \in \mathrm{I}$ with $x_1 < x_2$, and $m_l, M_l, m_r, M_r \in \mathbb{R}$ with $m_l < M_l$ and $m_r < M_r$, such that*

$$\begin{aligned}
&\lambda\big(L_{\leq m_l}(x_1, x_2)\big) > 0, \qquad &&\lambda\big(R_{\leq m_r}(x_1, x_2)\big) = 0, \\
&\lambda\big(L_{\geq M_l}(x_1, x_2)\big) > 0, \qquad &&\lambda\big(R_{\geq M_r}(x_1, x_2)\big) > 0.
\end{aligned}$$

Proof. For a contradiction, suppose that the stated relations hold. Then there is an $m_r' \in \mathbb{R}$ such that $\lambda(R_{\leq m_r'}) > 0$. Pick any $M_r' > m_r'$. If $\lambda(R_{\geq M_r'}) > 0$, we have a contradiction to Lemma 4.1; if $\lambda(R_{\geq M_r'}) = 0$, we have a contradiction to Lemma 4.2. □

Lemma 4.4. *There are no $x_1, x_2 \in \mathrm{I}$ with $x_1 < x_2$, and $m_l, M_l, m_r, M_r \in \mathbb{R}$ with $m_l < M_l$ and $m_r < M_r$, such that*

$$\begin{aligned}
&\lambda\big(L_{\leq m_l}(x_1, x_2)\big) > 0, \qquad &&\lambda\big(R_{\leq m_r}(x_1, x_2)\big) = 0, \\
&\lambda\big(L_{\geq M_l}(x_1, x_2)\big) > 0, \qquad &&\lambda\big(R_{\geq M_r}(x_1, x_2)\big) = 0.
\end{aligned}$$

Proof. For a contradiction, suppose that the stated relations hold. Then $\lambda(R_{\geq M_r}) = 0$ yields $\lambda(R_{\leq M_r}) > 0$, and any $M_r' > M_r$ incurs $\lambda(R_{\geq M_r'}) = 0$, contrary to Lemma 4.2. □

Summarizing the above, we obtain part (a) of the following result, with the proof of part (b) being analogous.

Proposition 4.5.

(a) There are no $x_1, x_2 \in \mathrm{I}$ with $x_1 < x_2$ and $m_l, M_l \in \mathbb{R}$ with $m_l < M_l$ such that $\lambda(L_{\leq m_l}(x_1, x_2)) > 0$ and $\lambda(L_{\geq M_l}(x_1, x_2)) > 0$.

(b) There are no $x_1, x_2 \in \mathrm{I}$ with $x_1 < x_2$ and $m_r, M_r \in \mathbb{R}$ with $m_r < M_r$ such that $\lambda(R_{\leq m_r}(x_1, x_2)) > 0$ and $\lambda(R_{\geq M_r}(x_1, x_2)) > 0$.

The next result shows that for all $x_1, x_2 \in \mathrm{I}$ with $x_1 < x_2$ the function $\Delta(x_1, x_2, y)$ is almost everywhere constant outside $[x_1, x_2]$.

Proposition 4.6. *Suppose that $x_1, x_2 \in \mathrm{I}$ with $x_1 < x_2$. Then there exist a constant $G_l(x_1, x_2)$ such that*

$$\Delta(x_1, x_2, y) = G_l(x_1, x_2) \quad \textit{for almost all } y \in (a, x_1],$$

and a constant $G_r(x_1, x_2)$ *such that*

$$\Delta(x_1, x_2, y) = G_r(x_1, x_2) \quad \textit{for almost all } y \in [x_2, b).$$

Proof. There exists a real number M_l such that both $\lambda(L_{\le M_l}(x_1, x_2)) > 0$ and $\lambda(L_{\ge M_l}(x_1, x_2)) > 0$. This choice is unique, because by part (a) of Proposition 4.5 there is no $m_l \ne M_l$ such that $\lambda(L_{\le m_l}(x_1, x_2)) > 0$ and $\lambda(L_{\ge m_l}(x_1, x_2)) > 0$. Therefore, $\Delta(x_1, x_2, y) = M_l = G_l(x_1, x_2)$ for almost all $y \in (a, x_1]$. The proof of the second statement invokes part (b) of Proposition 4.5 and otherwise is analogous. □

Proposition 4.7. *There exist functions* $\mathrm{A}_1 : \mathrm{I} \to \mathbb{R}$, $\mathrm{A}_2 : \mathrm{I} \to \mathbb{R}$ *and measurable functions* $\mathrm{B}_1 : \mathrm{I} \to \mathbb{R}$, $\mathrm{B}_2 : \mathrm{I} \to \mathbb{R}$ *such that, for each* $x \in \mathrm{I}$,

$$\mathrm{S}(x, y) = \begin{cases} \mathrm{A}_1(x) + \mathrm{B}_1(y), & y \le x, \\ \mathrm{A}_2(x) + \mathrm{B}_2(y), & y > x, \end{cases}$$

for all $y \in \mathrm{I} \setminus \mathrm{N}_x$*, where* N_x *is a null set.*

Proof. Suppose that $a_n \downarrow a$ and $b_n \uparrow b$ are strictly monotone sequences in I such that $a_1 < b_1$. By Proposition 4.6, it is true for $n = 1, 2, \ldots$ and all $x \in (a_n, b_n)$ that

$$\mathrm{S}(x, y) = \begin{cases} G_l(x, b_n) + \mathrm{S}(b_n, y), & y \le x, \\ -\,G_r(a_n, x) + \mathrm{S}(a_n, y), & y > x, \end{cases}$$

for all $y \in \mathrm{I} \setminus \mathrm{N}_x$, where N_x is a null set. Thus, we have for $n = 1, 2, \ldots$

$$G_l(x, b_n) - G_l(x, b_{n+1}) = \mathrm{S}(b_{n+1}, y) - \mathrm{S}(b_n, y),$$

for all $x \in (a_n, b_n)$ and all $y \in \mathrm{I} \setminus \mathrm{N}$, where N is a null set. Thus, $G_l(x, b_n) - G_l(x, b_{n+1})$ does not depend on x, and $\mathrm{S}(b_{n+1}, y) - \mathrm{S}(b_n, y)$ does not depend on y, and both equal a constant, say c_n. Put $c_0 = 0$. Let $\mathrm{I}_1 = (a_1, b_1)$ and for $n \ge 2$ let $\mathrm{I}_n = (a_n, b_n) \setminus (a_{n-1}, b_{n-1})$, so that the family $\{\mathrm{I}_n : n = 1, 2, \ldots\}$ forms a partition of I. If we define

$$A_1(x) = \sum_{n=1}^{\infty} \left(G_l(x, b_n) + \sum_{k=0}^{n-1} c_k \right) \mathbf{1}(x \in \mathrm{I}_n)$$

and let

$$B_1(y) = \sum_{n=1}^{\infty} \left(\mathrm{S}(b_n, y) - \sum_{k=0}^{n-1} c_k \right) \mathbf{1}(y \in \mathrm{I}_n),$$

the desired representation holds. Analogously, we find the desired functions $A_2 : \mathrm{I} \to \mathbb{R}$ and $B_2 : \mathrm{I} \to \mathbb{R}$. □

Proposition 4.8. *For all* $x_1, x_2 \in \mathrm{I}$,

$$\alpha \left(\mathrm{A}_1(x_1) - \mathrm{A}_1(x_2) \right) = (1 - \alpha) \big(\mathrm{A}_2(x_2) - \mathrm{A}_2(x_1) \big).$$

Proof. Without loss of generality, we may consider $x_1, x_2 \in \mathrm{I}$ with $x_1 < x_2$. There exist a constant c satisfying (1), and a sequence $(F_n)_{n=1,2,\ldots}$ of probability measures in $\mathcal{L}_+$ with densities $(f_n)_{n=1,2,\ldots}$, such that

$$F_n((a, x_1]) = \alpha, \qquad F_n((x_2, b]) = 1 - \alpha - \frac{1}{n + c},$$

and

$$f_n(y) = \frac{1+c}{n+c}\, f_1(y) \quad \text{for } y \in (x_1, x_2].$$

As x_1 is an α-quantile of F and S is weakly consistent for the α-quantile relative to $\mathcal{L}_+$, we have

$$\begin{aligned} 0 &\geq \mathbb{E}_{F_n}\big[\Delta(x_1, x_2, Y)\big] \\ &= \int_a^{x_1} \Delta(y)\, dF_n(y) + \int_{x_1}^{x_2} \Delta(y)\, dF_n(y) + \int_{x_2}^{b} \Delta(y)\, dF_n(y) \\ &= \alpha\big(\mathrm{A}_1(x_1) - \mathrm{A}_1(x_2)\big) + \frac{1+c}{n+c} \int_{x_1}^{x_2} \Delta(y)\, dF_1(y) \\ &\quad + \left(1 - \alpha - \frac{1}{n+c}\right)\big(\mathrm{A}_2(x_1) - \mathrm{A}_2(x_2)\big) \end{aligned}$$

for $n = 1, 2, \ldots$ In the limit as $n \to \infty$ we obtain

$$\alpha\left(\mathrm{A}_1(x_1) - \mathrm{A}_1(x_2)\right) \leq (1 - \alpha)\big(\mathrm{A}_2(x_2) - \mathrm{A}_2(x_1)\big).$$

An analogous argument with $F_n((a, x_1]) = \alpha - \frac{1}{n+c}$ and $F_n((a, x_2]) = \alpha$ for $n = 1, 2, \ldots$ yields the opposite inequality, thereby completing the proof. □

For subsequent use we define, for $c \in \mathbb{R}$, the Borel set

$$M_c(x_1, x_2) = \big\{y \in [x_1, x_2] : \mathrm{B}_2(y) - \mathrm{B}_1(y) \leq c\big\},$$

where we frequently abbreviate $M_c = M_c(x_1, x_2)$, with the arguments understood to be fixed.

Proposition 4.9. *For all $x_1, x_2 \in \mathrm{I}$ with $x_1 < x_2$,*

$$\operatorname{ess\,inf}_{[x_1,\, x_2]}\big(\mathrm{B}_2(y) - \mathrm{B}_1(y)\big) \geq \mathrm{A}_1(x_1) - \mathrm{A}_2(x_1)$$

and

$$\operatorname{ess\,sup}_{[x_1,\, x_2]}\big(\mathrm{B}_2(y) - \mathrm{B}_1(y)\big) \leq \mathrm{A}_1(x_2) - \mathrm{A}_2(x_2).$$

Proof. We first show that $\operatorname{ess\,inf}_{[x_1,\, x_2]}(\mathrm{B}_2(y) - \mathrm{B}_1(y))$ is finite. For a contradiction, suppose there are $x_1, x_2 \in \mathrm{I}$ with $x_1 < x_2$ with the property that

$$\operatorname{ess\,inf}_{[x_1,\, x_2]}(\mathrm{B}_2(y) - \mathrm{B}_1(y)) = -\infty.$$

Let $\beta \in (0, \alpha)$. Then there exists a sequence $c_n \downarrow -\infty$ such that $c_1 < -(x_2 - x_1)/(\alpha - \beta)$, $\lambda(M_{c_1}) < x_2 - x_1$ and $\lambda(M_{c_n}) > 0$ for $n = 1, 2, \ldots$, whence

$$\frac{1}{|c_n|}\, \lambda\big([x_1, x_2] \setminus M_{c_n}\big) < \frac{1}{|c_1|}\, (x_2 - x_1) < \alpha - \beta \quad \text{for } n = 1, 2, \ldots$$

Thus, there is a sequence $(F_n)_{n=1,2,\ldots}$ of probability measures in $\mathcal{L}_+$ with densities $(f_n)_{n=1,2,\ldots}$, such that $F_n((a, x_2]) = \alpha$, $F_n(M_{c_n}) = \beta$, and $f_n(y) \leq 1/|c_n|$ for $y \in [x_1, x_2] \setminus M_{c_n}$. As x_2 is an α-quantile of F_n and S is weakly consistent for the α-quantile relative to the class $\mathcal{L}_+$, we have

$$\begin{aligned} 0 &\leq \mathbb{E}_{F_n}\big[\Delta(x_1, x_2, Y)\big] \\ &= \big(\mathrm{A}_1(x_1) - \mathrm{A}_1(x_2)\big) \int_a^{x_1} dF_n(y) + \big(\mathrm{A}_2(x_1) - \mathrm{A}_2(x_2)\big) \int_{x_2}^{b} dF_n(y) \\ &\quad + \int_{x_1}^{x_2} \big[\big(\mathrm{A}_2(x_1) + \mathrm{B}_2(y)\big) - \big(\mathrm{A}_1(x_2) + \mathrm{B}_1(y)\big)\big]\, dF_n(y) \end{aligned}$$

$$
\begin{aligned}
&= \big(\mathrm{A}_1(x_1) - \mathrm{A}_1(x_2)\big) \int_a^{x_2} dF_n(y) + \big(\mathrm{A}_2(x_1) - \mathrm{A}_2(x_2)\big) \int_{x_2}^{b} dF_n(y) \\
&\quad + \int_{x_1}^{x_2} \big[\big(\mathrm{A}_2(x_1) + \mathrm{B}_2(y)\big) - \big(\mathrm{A}_1(x_1) + \mathrm{B}_1(y)\big)\big]\, dF_n(y) \\
&= \int_{x_1}^{x_2} \big[\big(\mathrm{B}_2(y) - \mathrm{B}_1(y)\big) + \big(\mathrm{A}_2(x_1) - \mathrm{A}_1(x_1)\big)\big]\, dF_n(y) \\
&= \int_{M_{c_n}} \big(\mathrm{B}_2(y) - \mathrm{B}_1(y)\big)\, dF_n(y) + \int_{[x_1,x_2]\setminus M_{c_n}} \big(\mathrm{B}_2(y) - \mathrm{B}_1(y)\big)\, dF_n(y) \\
&\quad + \big(\mathrm{A}_2(x_1) - \mathrm{A}_1(x_1)\big)\big[F_n(M_{c_n}) + F_n\big([x_1,x_2] \setminus M_{c_n}\big)\big] \\
&\le c_n\beta + \frac{1}{|c_n|} \int_{x_1}^{x_2} |\mathrm{B}_2(y) - \mathrm{B}_1(y)|\, d\lambda(y) + |\mathrm{A}_2(x_1) - \mathrm{A}_1(x_1)| \left(\beta + \frac{x_2 - x_1}{|c_n|}\right)
\end{aligned}
$$

for $n = 1, 2, \ldots$, where the equalities stem from Propositions 4.7 and 4.8, and the inequality contradicts our assumption that $c_n \downarrow -\infty$.

We now prove the lower estimate. Suppose first that $B_2(y) - B_1(y)$ is almost everywhere constant on $[x_1, x_2]$. Pick $\beta \in (0, \alpha)$ and a probability measure $F \in \mathcal{L}_+$ such that $F((x_1, x_2]) = \beta$ and $F((a, x_2]) = \alpha$, whence

$$
\begin{aligned}
0 &\le \mathbb{E}_F\big[\Delta(x_1, x_2, Y)\big] \\
&= \int_{x_1}^{x_2} \big(B_2(y) - B_1(y)\big)\, dF(y) + \big(A_2(x_1) - A_1(x_1)\big)\, F((x_1, x_2]) \\
&= \beta\big(\operatorname{ess\,inf}_{[x_1,x_2]}\big(B_2(y) - B_1(y)\big) + \big[A_2(x_1) - A_1(x_1)\big]\big).
\end{aligned}
$$

If $B_2(y) - B_1(y)$ is not almost everywhere constant on $[x_1, x_2]$, let

$$
\gamma = \operatorname{ess\,inf}_{[x_1,\, x_2]}\big(\mathrm{B}_2(y) - \mathrm{B}_1(y)\big) = \inf\big\{c \in \mathbb{R} : \lambda(M_c) > 0\big\},
$$

which is finite by the above. Let $\beta \in (0, \alpha)$ and pick any sequence $\epsilon_n \downarrow 0$ where $\epsilon_1 < (\alpha - \beta)/(x_2 - x_1)$ is sufficiently small to ensure that $\lambda(M_{\gamma+\epsilon_1}) < x_2 - x_1$. Since

$$
\epsilon_n\, \lambda\big([x_1, x_2] \setminus M_{\gamma+\epsilon_n}\big) < \epsilon_1 (x_2 - x_1) < \alpha - \beta \quad \text{for} \quad n = 1, 2, \ldots
$$

there exists a sequence $(F_n)_{n=1,2,\ldots}$ of probability measures in $\mathcal{L}_+$ with densities $(f_n)_{n=1,2,\ldots}$, such that $F_n((a, x_2]) = \alpha$, $F_n(M_{\gamma+\epsilon_n}) = \beta$, and $f_n(y) \le \epsilon_n$ for all $y \in [x_1, x_2] \setminus M_{\gamma+\epsilon_n}$. Familiar arguments imply that

$$
\begin{aligned}
0 &\le \mathbb{E}_{F_n}\big[\Delta(x_1, x_2, Y)\big] \\
&= \int_{M_{\gamma+\epsilon_n}} \big(\mathrm{B}_2(y) - \mathrm{B}_1(y)\big)\, dF_n(y) + \int_{[x_1,x_2]\setminus M_{\gamma+\epsilon_n}} \big(\mathrm{B}_2(y) - \mathrm{B}_1(y)\big)\, dF_n(y) \\
&\quad + \big(\mathrm{A}_2(x_1) - \mathrm{A}_1(x_1)\big)\big[F_n(M_{\gamma+\epsilon_n}) + F_n\big([x_1, x_2] \setminus M_{\gamma+\epsilon_n}\big)\big] \\
&\le (\gamma + \epsilon_n)\beta + \epsilon_n \int_{x_1}^{x_2} |\mathrm{B}_2(y) - \mathrm{B}_1(y)|\, d\lambda(y) \\
&\quad + \big(\mathrm{A}_2(x_1) - \mathrm{A}_1(x_1)\big)\beta + |\mathrm{A}_2(x_1) - \mathrm{A}_1(x_1)|\epsilon_n (x_2 - x_1)
\end{aligned}
$$

for $n = 1, 2, \ldots$, which yields the claim in the limit as $n \to \infty$. Analogous arguments imply the upper estimate. □

Proposition 4.10. *It is true that* $\mathrm{A}_1(\cdot) - \mathrm{A}_2(\cdot)$ *is nondecreasing everywhere, and that*

$$
\mathrm{B}_2(\cdot) - \mathrm{B}_1(\cdot) = \mathrm{A}_1(\cdot) - \mathrm{A}_2(\cdot)
$$

almost everywhere.

Proof. The first claim is immediate from Proposition 4.9. Further, let $x_1, x, x_2 \in \mathrm{I}$ with $x_1 < x < x_2$. Again by Proposition 4.9,

$$\operatorname{ess\,sup}_{[x_1,\,x]}\big(\mathrm{B}_2(y) - \mathrm{B}_1(y)\big) \le \mathrm{A}_1(x) - \mathrm{A}_2(x) \le \operatorname{ess\,inf}_{[x,\,x_2]}\big(\mathrm{B}_2(y) - \mathrm{B}_1(y)\big),$$

and it follows readily that

$$\mathrm{B}_2(x) - \mathrm{B}_1(x) \le \mathrm{A}_1(x) - \mathrm{A}_2(x) \le \mathrm{B}_2(x) - \mathrm{B}_1(x)$$

for all $x \in \mathrm{I}$ up to a null set, thereby proving the second claim. A more detailed proof is given in [5]. □

We are now ready to complete the proof. By Proposition 4.8, there exists a constant C such that $\alpha \mathrm{A}_1(x) + (1-\alpha)\mathrm{A}_2(x) = C$ everywhere, whence

$$A_1(x) = \frac{C}{\alpha} - \frac{1-\alpha}{\alpha}\mathrm{A}_2(x) \quad \text{for all } x \in \mathrm{I}. \tag{3}$$

Invoking Proposition 4.10, we see that

$$\mathrm{A}_2(x) - \mathrm{A}_1(x) = \mathrm{A}_2(x) - \left(\frac{C}{\alpha} - \frac{1-\alpha}{\alpha}\mathrm{A}_2(x)\right) = \frac{1}{\alpha}\mathrm{A}_2(x) - \frac{C}{\alpha}$$

is nonincreasing. By the same result,

$$\mathrm{B}_1(x) - \mathrm{B}_2(x) = \mathrm{A}_2(x) - \mathrm{A}_1(x) = \frac{1}{\alpha}\mathrm{A}_2(x) - \frac{C}{\alpha}$$

almost everywhere, which gives

$$\mathrm{B}_1(y) = \mathrm{B}_2(y) + \frac{1}{\alpha}\mathrm{A}_2(y) - \frac{C}{\alpha} \quad \text{for almost all } y \in \mathrm{I}. \tag{4}$$

Combining (3) and (4) with Proposition 4.7 we find that, for all $x \in \mathrm{I}$,

$$\mathrm{S}(x,y) = \begin{cases} -\dfrac{1-\alpha}{\alpha}\mathrm{A}_2(x) + \dfrac{1}{\alpha}\mathrm{A}_2(y) + \mathrm{B}_2(y), & y \le x, \\ \mathrm{A}_2(x) + \mathrm{B}_2(y), & y > x, \end{cases}$$

for all $y \in \mathrm{I} \setminus \mathrm{N}_x$, where N_x is a null set. Define

$$g(x) = -\frac{1}{\alpha}\mathrm{A}_2(x) \quad \text{and} \quad h(y) = B_2(y) + A_2(y)$$

and note that g is nondecreasing since A_2 is nonincreasing, thereby completing the proof of Theorem 2.2.

5. Discussion

We have reviewed Thomson's [12] pioneering characterization of the scoring functions that are consistent for quantiles. This fundamental decision-theoretic result deserves broad attention, and our goal here was to make it accessible to an audience of probabilists and statisticians.

Closely related questions arise in the case of the mean or expectation functional. Subject to conditions (S0), (S1) and (S2) of Theorem 1.4, a scoring function is consistent for the mean functional relative to the class $\mathcal{D}_2$ if and only if it is a *Bregman function* of the form

$$\mathrm{S}_\phi(x, y) = \phi(y) - \phi(x) - \phi'(x)(y - x),$$

where ϕ is convex with subgradient ϕ' [3, 11]. Savage [11] suggested that the characterization continues to hold if conditions (S1) and (S2) are dropped, but his proof lacks detail and we have not been able to verify the claim. For a detailed discussion, see Section 3 in [5]. Still greater technical challenges arise in the case in which $\mathcal{F}$ is a class of multivariate distributions, for which we refer to [8] and [3] along with references therein.

References

[1] Cervera, J. L. and Muñoz, J. (1996). Proper scoring rules for fractiles. In *Bayesian Statistics* **5** (Bernardo, J. M., Berger, J. O., Dawid, A. P. and Smith, A. F. M., eds.) 513–519. Oxford Univ. Press.

[2] Gneiting, T. (2011). Quantiles as optimal point forecasts. *Int. J. Forecasting* **27** 197–207.

[3] Gneiting, T. (2011). Making and evaluating point forecasts. *J. Amer. Statist. Assoc.* **106** 746–762.

[4] Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *J. Amer. Statist. Assoc.* **102** 359–378.

[5] Grant, K. (2011). Scoring functions for quantiles. Diploma thesis, Univ. of Heidelberg.

[6] Jose, V. R. R. and Winkler, R. L. (2009). Evaluating quantile assessments. *Operations Res.* **57** 1287–1297.

[7] Koenker, R. (2005). *Quantile Regression*. Cambridge Univ. Press.

[8] Koltchinskii, V. I. (1997). *M*-estimation, convexity and quantiles. *Ann. Statist.* **25** 435–477.

[9] Osband, K. H. (1985). Providing incentives for better cost forecasting. Ph.D. thesis, Univ. of California, Berkeley.

[10] Saerens, M. (2000). Building cost functions minimizing to some summary statistics. *IEEE Trans. Neur. Netw.* **11** 1263–1271.

[11] Savage, L. J. (1971). Elicitation of personal probabilities and expectations. *J. Amer. Statist. Assoc.* **66** 783–801.

[12] Thomson, W. (1979). Eliciting production possibilities from a well-informed manager. *J. Econ. Theo.* **20** 360–380.

[13] Wellner, J. A. (2009). Statistical functionals and the delta method. Lecture notes, available online at http://www.stat.washington.edu/people/jaw/COURSES/580s/581/LECTNOTES/ch7.pdf.