THE HALFSPACE DEPTH CHARACTERIZATION PROBLEM: ADDITIONAL DETAILS

STANISLAV NAGY[1]

ABSTRACT. This note contains some technical details and derivations that complement the analysis from Section 2 of the paper [1]: *The halfspace depth characterization problem.*

HALFSPACE DEPTH FOR DISTRIBUTIONS SUPPORTED ON COORDINATE AXES

Let $P \in \mathcal{P}(\mathbb{R}^2)$ be a probability distribution supported on the coordinate axes $A_x = \left\{(x,y)^{\mathsf{T}} : y = 0\right\}$ and $A_y = \left\{(x,y)^{\mathsf{T}} : x = 0\right\}$. The density of P with respect to the sum of one-dimensional Lebesgue measures concentrated on A_x and A_y is given by

$$f(x,y) = \begin{cases} f_x(x)/2 & \text{for } (x,y)^{\mathsf{T}} \in A_x, \\ f_y(y)/2 & \text{for } (x,y)^{\mathsf{T}} \in A_y, \end{cases} \tag{1}$$

where f_x and f_y are symmetric univariate density functions that are positive and bounded on $\mathbb{R}$. Denote by F_x and F_y the distribution functions that correspond to densities f_x and f_y, respectively, i.e. for F_x we have

$$F_x(x) = \int_{-\infty}^{x} f_x(t)\,\mathrm{d}\,t,$$

and F_y is given analogously.

We now compute the halfspace depth of a point $\boldsymbol{x} = (x_0, y_0)^{\mathsf{T}} \in (0,\infty) \times [0,\infty)$ with respect to a random vector $X \sim P$. For that, define the halfspace function

$$\varphi(\theta) = \mathsf{P}\left(\langle X, u_\theta\rangle \geq \langle \boldsymbol{x}, u_\theta\rangle\right),$$

where $u_\theta = (\cos(\theta), \sin(\theta))^{\mathsf{T}}$. Function $\varphi(\theta)$ provides the probability mass of the halfplane whose boundary passes through $\boldsymbol{x}$, with inner normal u_θ. Of course, φ depends on the choice of $\boldsymbol{x}$. The depth of $\boldsymbol{x}$ is given by

$$D(\boldsymbol{x}; P) = \inf_{\theta \in (-\pi,\pi]} \varphi(\theta). \tag{2}$$

Since the densities f_x and f_y are symmetric and have equal weights in (1), and because we restrict $\boldsymbol{x}$ to lie in first quadrant of $\mathbb{R}^2$, it is sufficient to search for the minimum in (2) only in the interval $\theta \in [0, \pi/2]$.

If $y_0 = 0$, it is easy to see that for the halfspace depth we have

$$D(\boldsymbol{x}; P) = \varphi(0) = \left(1 - F_x(x_0)\right)/2. \tag{3}$$

Suppose then that both x_0 and y_0 are positive. For $\theta \in (-\pi/2, \pi/2) \setminus \{0\}$, the boundary line of the halfspace $H_\theta = \left\{\boldsymbol{y} \in \mathbb{R}^2 \colon \langle \boldsymbol{y}, u_\theta\rangle \geq \langle \boldsymbol{x}, u_\theta\rangle\right\}$ intersects the axis A_x in

[1]CHARLES UNIVERSITY, FACULTY OF MATHEMATICS AND PHYSICS, PRAGUE, CZECH REPUBLIC
E-mail address: nagy@karlin.mff.cuni.cz.
Date: August 4, 2019.

$(x_0 + y_0 \tan(\theta), 0)^\mathsf{T}$, and A_y in $(0, x_0/\tan(\theta) + y_0)^\mathsf{T}$. Likewise, for $\theta = 0$ we have $\partial H_\theta \cap A_x = (x_0, 0)^\mathsf{T}$, $\partial H_\theta \cap A_y = \emptyset$, and for $\theta = \pi/2$ we can write $\partial H_\theta \cap A_x = \emptyset$, $\partial H_\theta \cap A_y = (0, y_0)^\mathsf{T}$. That gives us

$$2\varphi(\theta) = \begin{cases} 1 - F_x(x_0 + y_0 \tan(\theta)) + F_y(x_0/\tan(\theta) + y_0) & \text{for } \theta \in (-\pi/2, 0), \\ 1 - F_x(x_0) & \text{for } \theta = 0, \\ 1 - F_x(x_0 + y_0 \tan(\theta)) + 1 - F_y(x_0/\tan(\theta) + y_0) & \text{for } \theta \in (0, \pi/2), \\ 1 - F_y(y_0) & \text{for } \theta = \pi/2. \end{cases} \tag{4}$$

Because θ enters function $\varphi(\theta)$ only as $\tan(\theta)$, it will be convenient to minimize function $\psi(t) = \varphi(\tan(\theta))$ instead of $\varphi(\theta)$, with $t \in \mathbb{R} \cup \{+\infty\}$. That way, we consider

$$2\psi(t) = \begin{cases} 1 - F_x(x_0 + y_0 t) + F_y(x_0/t + y_0) & \text{for } t \in (-\infty, 0), \\ 1 - F_x(x_0) & \text{for } t = 0, \\ 1 - F_x(x_0 + y_0 t) + 1 - F_y(x_0/t + y_0) & \text{for } t \in (0, \infty), \\ 1 - F_y(y_0) & \text{for } \theta = +\infty. \end{cases}$$

Let us focus on the derivative of ψ around $t = 0$; the situation with function φ around $\theta = \pi/2$ is analogous. Direct computation yields

$$2\psi'(t) = \begin{cases} -f_x(x_0 + y_0 t)\, y_0 - f_y(x_0/t + y_0)\, x_0/t^2 & \text{for } t \in (-\infty, 0), \\ -f_x(x_0 + y_0 t)\, y_0 + f_y(x_0/t + y_0)\, x_0/t^2 & \text{for } t \in (0, \infty), \end{cases} \tag{5}$$

and, provided that all the limits on the right hand sides of the following expressions exist,

$$\begin{aligned} \lim_{t \to 0-} 2\psi'(t) &= -\lim_{t \to 0-} \left(f_x(x_0 + y_0 t)\, y_0 + f_y(x_0/t + y_0)\, x_0/t^2 \right) \\ &= -f_x(x_0)\, y_0 - \lim_{t \to 0-} f_y(x_0/t + y_0)\, x_0/t^2 \\ &= -f_x(x_0)\, y_0 - x_0 \lim_{s \to -\infty} f_y(s) \left(\frac{s - y_0}{x_0} \right)^2 \\ &= -f_x(x_0)\, y_0 - \frac{1}{x_0} \left(\lim_{s \to -\infty} f_y(s)\, s^2 - 2y_0 \lim_{s \to -\infty} f_y(s) s + y_0^2 \lim_{s \to -\infty} f_y(s) \right), \\ \lim_{t \to 0+} 2\psi'(t) &= -\lim_{t \to 0+} \left(f_x(x_0 + y_0 t)\, y_0 - f_y(x_0/t + y_0)\, x_0/t^2 \right) \\ &= -f_x(x_0)\, y_0 + \lim_{t \to 0+} f_y(x_0/t + y_0)\, x_0/t^2 \\ &= -f_x(x_0)\, y_0 + x_0 \lim_{s \to +\infty} f_y(s) \left(\frac{s - y_0}{x_0} \right)^2 \\ &= -f_x(x_0)\, y_0 + \frac{1}{x_0} \left(\lim_{s \to +\infty} f_y(s)\, s^2 - 2y_0 \lim_{s \to +\infty} f_y(s) s + y_0^2 \lim_{s \to +\infty} f_y(s) \right). \end{aligned}$$

Because f_y is a density, if its limit at infinity exists, then necessarily $\lim_{s \to -\infty} f_y(s) = \lim_{s \to +\infty} f_y(s) = 0$. Since f_y was assumed to be a symmetric function, $\lim_{s \to -\infty} f_y(s) s = -\lim_{s \to +\infty} f_y(s) s$, and $\lim_{s \to -\infty} f_y(s) s^2 = \lim_{s \to +\infty} f_y(s) s^2$. Thus, $\psi'(0)$ exists if and only if $\lim_{s \to +\infty} f_y(s) s^2 = 0$, and in that case

$$\psi'(0) = -f_x(x_0)\, y_0/2 < 0.$$

This result is rather intuitive — for $\boldsymbol{x}$ in the positive quadrant, function $\varphi(\theta)$ is decreasing at $\theta = 0$; i.e. tilting the halfplane H_θ at $\theta = 0$ in the counter-clockwise sense results in a smaller probability mass.

If

$$\lim_{s\to+\infty} f_y(s)s^2 > 0, \tag{6}$$

the derivative of $\varphi(\theta)$ does not exist at $\theta = 0$. If $\lim_{s\to+\infty} f_y(s)s^2 = +\infty$, we have

$$\lim_{t\to 0-} 2\psi'(t) = -\infty,$$
$$\lim_{t\to 0+} 2\psi'(t) = +\infty,$$

and $t = 0$ is a local minimum of ψ. In the case $\lim_{s\to+\infty} f_y(s)s^2 = S < \infty$, necessarily $\lim_{s\to+\infty} f_y(s)s = 0$, and

$$\lim_{t\to 0-} 2\psi'(t) = -f_x(x_0)\,y_0 - \frac{S}{x_0},$$
$$\lim_{t\to 0+} 2\psi'(t) = -f_x(x_0)\,y_0 + \frac{S}{x_0}.$$

Point $t = 0$ is a local minimum of ψ if and only if

$$S \geq f_x(x_0)y_0x_0, \tag{7}$$

which holds true at least if $\boldsymbol{x}$ is close to the origin. Note that (6) implies that the expectation of $X \sim P$ cannot exist.

Example: Cauchy distribution. In what follows, let both f_x and f_y be the densities of the standard univariate Cauchy random variable, i.e.

$$f_x(s) = f_y(s) = \frac{1}{\pi\,(1+s^2)}, \quad \text{for } s \in \mathbb{R}.$$

We have that $S = \lim_{s\to+\infty} f_y(s)s^2 = \pi^{-1}$, and $\theta = 0$ is a local minimum of φ if and only if

$$\frac{x_0y_0}{(1+x_0^2)} \leq 1 \tag{8}$$

due to (7). By symmetry considerations, $\theta = \pi/2$ is a local minimum of φ for any $\boldsymbol{x}$ such that

$$\frac{x_0y_0}{(1+y_0^2)} \leq 1. \tag{9}$$

If $y_0 = 0$, the depth of $\boldsymbol{x}$ is given by (3). Thus, we may assume that $y_0 > 0$. Now we establish that for any $\boldsymbol{x} = (x_0, y_0)^{\mathsf{T}} \in (0,\infty) \times (0,\infty)$ function φ is quasi-concave* on $(0, \pi/2)$. From that it follows that

$$D(\boldsymbol{x}; P) = \inf_{\theta\in(0,\pi/2)} \varphi(\theta) = \min\{\varphi(0), \varphi(\pi/2)\}. \tag{10}$$

To show the quasi-concavity of φ, note first that for $\theta \in (0, \pi/2)$ function $\varphi(\theta)$ is continuously differentiable, with finite one-sided derivatives at $\theta \in \{0, \pi/2\}$. For such a function, it is sufficient to show that either φ is monotone on $(0, \pi/2)$, or that it is increasing in a neighbourhood of $\theta = 0$, with at most one critical point. We proceed to find the critical

*By a quasi-concave function φ on the interval $(0, \pi/2)$ we mean that for all $\theta_1, \theta_2 \in (0, \pi/2)$ and $\lambda \in [0, 1]$ we can write $\varphi(\lambda\theta_1 + (1-\lambda)\theta_2) \geq \min\{\varphi(\theta_1), \varphi(\theta_2)\}$.

points of φ in $(0, \pi/2)$, which, by the considerations above, coincide with the critical points of ψ in $(0, \infty)$. For our choice of f_x and f_y, by (5), the derivative of ψ can be written for $t \in (0, \infty)$ as

$$2\pi\psi'(t) = -\frac{y_0}{1 + (x_0 + y_0 t)^2} + \frac{x_0}{t^2\left(1 + (x_0/t + y_0)^2\right)}.$$

Recall that we assume $y_0 > 0$, and re-parametrise the derivative above using $t = \alpha x_0/y_0$ for $\alpha \in (0, \infty)$. We obtain

$$2\pi\psi'\left(\frac{\alpha x_0}{y_0}\right) = -\frac{y_0}{1 + (x_0(1 + \alpha))^2} + \frac{y_0^2}{\alpha^2 x_0\left(1 + (y_0(1 + 1/\alpha))^2\right)}.$$

Set the right hand side of this formula to be zero, and solve for $\alpha \in (0, \infty)$. It turns out that this equation is quadratic in α, and its only positive solution is given by

$$\alpha = \frac{\sqrt{x_0 y_0\left(x_0^2 - 2x_0 y_0 + y_0^2 + 1\right)} + x_0^2 y_0 - x_0 y_0^2}{x_0^2(-y_0) + x_0 y_0^2 + x_0},$$

for

- $0 < x_0 < 2$ and at the same time $0 < y_0 < x_0 + 1/x_0$, or
- $2 \leq x_0 < \infty$ and at the same time

$$\text{either } 0 < y_0 < \frac{x_0}{2} - \frac{\sqrt{x_0^2 - 4}}{2}, \text{ or } \frac{\sqrt{x_0^2 - 4}}{2} + \frac{x_0}{2} < y_0 < \frac{x_0^2 + 1}{x_0}.$$

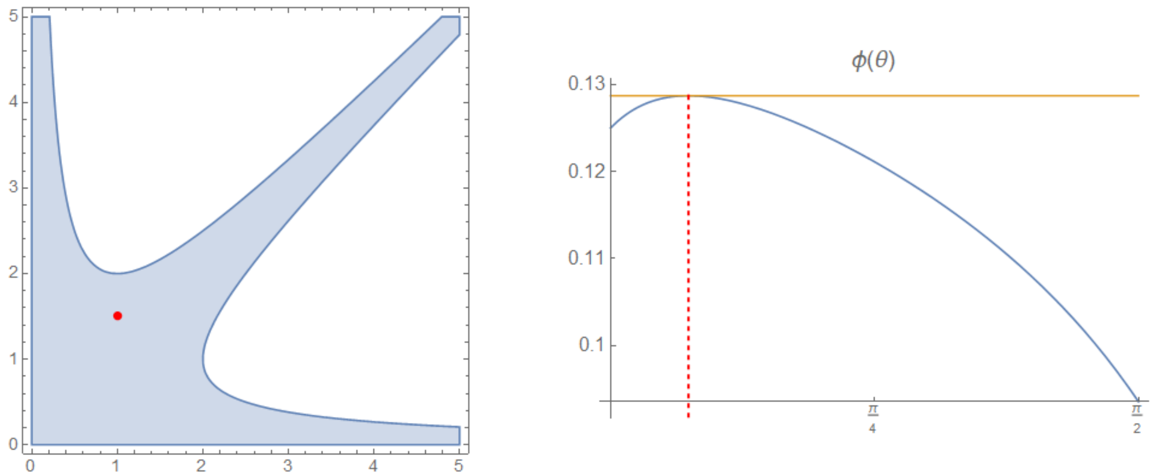

FIGURE 1. *Left panel: region R where a single critical point of φ exists in $(0, \pi/2)$ (coloured region), and a choice of the point $\boldsymbol{x} = (x_0, y_0)^{\mathsf{T}} = (1, 3/2)^{\mathsf{T}} \in R$ (red dot). Right panel: function $\varphi(\theta)$ corresponding to the chosen $\boldsymbol{x}$ (blue curve), its maximal value (orange line), and the critical point of φ (red dashed line). Obviously, function φ is quasi-concave on its domain, and minimized at $\theta = \pi/2$ (since $|x_0| < |y_0|$).*

It is tedious, yet straightforward to verify that this region, say $R \subset (0, \infty) \times (0, \infty)$, is equal to the complement of the union of the two regions given by (8) and (9) in $(0, \infty) \times (0, \infty)$. The part of region R that lies inside the rectangle $(0, 5) \times (0, 5)$ is depicted in Figure 1. Outside R, there is no positive solution to equation $\psi'(t) = 0$ over $t \in (0, \infty)$.

Consequently, function $\varphi(\theta)$ has a single critical point in the interval $\theta \in (0, \pi/2)$ for $\boldsymbol{x} \in R$, and no critical point in the interval $\theta \in (0, \pi/2)$ if $\boldsymbol{x} \in (0,\infty)\times(0,\infty)\setminus R$. Overall, it follows that for any $\boldsymbol{x} \in (0,\infty)\times(0,\infty)$ function φ is quasi-concave on $(0, \pi/2)$, and (10) holds true as asserted. Combine this with (3) and (4) to obtain

$$\begin{aligned} D(\boldsymbol{x}; P) &= \min\{1 - F_x(x_0), 1 - F_x(y_0)\}/2 = (1 - F_x(\max\{x_0, y_0\}))/2 \\ &= 1/4 - \arctan(\max\{x_0, y_0\})/(2\pi) \quad \text{for all } \boldsymbol{x} = (x_0, y_0)^{\mathsf{T}} \in (0,\infty)\times[0,\infty). \end{aligned}$$

Obvious symmetry considerations provide that the above formula holds also generally for any $\boldsymbol{x} = (x_0, y_0)^{\mathsf{T}} \in \mathbb{R}^2 \setminus \{0\}$ — writing $\|\boldsymbol{x}\|_\infty = \max\{|x_0|, |y_0|\}$ we conclude that

$$D(\boldsymbol{x}; P) = \begin{cases} 1/4 - \arctan(\|\boldsymbol{x}\|_\infty)/(2\pi) & \text{for all } \boldsymbol{x} \in \mathbb{R}^2 \setminus \{0\}, \\ 1/2 & \text{for } \boldsymbol{x} = 0. \end{cases}$$

This proves formula (6) in [1] for dimension $d = 2$. The result for general positive integers d follows analogously.

References

[1] Nagy, S. (2019). The halfspace depth characterization problem. *Springer Proc. Math. Stat.*, 10 pp. To appear.