Contents

Chapter 1

Projection-Based Evaluation of Tukey Depth from Weighted Student Differences

1.1 Definition and Notation

In [1], the distribution of a weighted difference of two independent Student variables is derived in the univariate setting ($d = 1$). For a given angle θ and independent Student variables t_1 and t_2 with degrees of freedom ν_1 and ν_2, respectively, one considers the random variable

$$v = t_1 \sin\theta - t_2 \cos\theta.$$

Although the original derivation employs the difference, the substitution of a plus sign is justified by the symmetry of the Student distribution. Moreover, this modification allows the extension of the range of θ from the usual $[0, \pi/2]$ to the full circle $[0, 2\pi)$ without affecting the final density.

Following the approach in [1], the cumulative distribution function (CDF) of v is given by

$$F_v(v_0) = \int_0^1 \Pr\Big\{T_{\nu_1+\nu_2} \le v_0\, \varphi(u)\Big\}\, \frac{u^{\nu_1/2-1}(1-u)^{\nu_2/2-1}}{B(\nu_1/2, \nu_2/2)}\, du,$$

where

$$\varphi(u) = \sqrt{\frac{\sin^2\theta}{u} + \frac{\cos^2\theta}{1-u}},$$

and $T_{\nu_1+\nu_2}$ denotes a Student variable with $\nu_1 + \nu_2$ degrees of freedom. Note that the squared trigonometric terms eliminate any sign ambiguity, ensuring the expression remains valid for any $\theta \in [0, 2\pi)$.

1.2 The Projection Method and Computation of Tukey Depth

In the multivariate setting, the classical definition of the Tukey (halfspace) depth of a point $x \in \mathbb{R}^p$ with respect to a probability measure F is given by

$$\mathrm{TD}(x) = \inf_{u \in S^{p-1}} F\Big(\{y \in \mathbb{R}^p : \langle y, u\rangle \ge \langle x, u\rangle\}\Big),$$

i.e. the smallest probability mass in any closed halfspace that contains x.

Our investigation begins by studying the distribution of the weighted sum

$$v = t_1 \sin\theta + t_2 \cos\theta,$$

whose CDF is expressed as

$$F_v(v_0) = \int_0^1 \Pr\Big\{T_{\nu_1+\nu_2} \le v_0\, \varphi(u)\Big\}\, \frac{u^{\nu_1/2-1}(1-u)^{\nu_2/2-1}}{B(\nu_1/2, \nu_2/2)}\, du,$$

with

$$\varphi(u) = \sqrt{\frac{\sin^2\theta}{u} + \frac{\cos^2\theta}{1-u}},$$

and where $T_{\nu_1+\nu_2}$ is a Student variable with $\nu_1+\nu_2$ degrees of freedom. Notice that by considering the complementary probability $1 - F_v(v_0)$, we obtain the mass contained in the halfspace corresponding to the projection direction determined by θ.

To connect this univariate result to the multivariate Tukey depth, note that if we were to compute the depth by considering all possible projection directions, we would have

$$\mathrm{TD}(x) = \inf_{\theta\in[0,2\pi)} \{1 - F_v(v_0(\theta))\},$$

where $v_0(\theta)$ denotes the threshold in the projected one-dimensional space corresponding to the halfspace boundary for direction θ.

Under the assumption of elliptical symmetry, it can be shown that the infimum is achieved when the projection is aligned with the point x. That is, by choosing the optimal projection direction

$$u = \frac{x}{\|x\|},$$

the scalar projection of x becomes exactly $\|x\|$. Hence, the Tukey depth simplifies to

$$\mathrm{TD}(x) = 1 - F_v\big(\|x\|\big) = 1 - \mathbb{E}\Big[G\Big(\|x\|\,\varphi(U)\Big)\Big],$$

where

$$G(t) = \Pr\{T_{\nu_1+\nu_2} \le t\},$$

and U is a Beta random variable with density

$$f_U(u) = \frac{u^{\nu_1/2-1}(1-u)^{\nu_2/2-1}}{B(\nu_1/2, \nu_2/2)}.$$

This formulation connects our analysis of the weighted sum v to the classical Tukey depth definition by demonstrating that, under optimal projection, the centrality of x is measured by the probability mass in the halfspace determined by $x/\|x\|$. Consequently, the computation of $\mathrm{TD}(x)$ reduces to evaluating a univariate expectation.

1.3 Numerical Evaluation and Monte Carlo Approximation

The integral representation of $F_v(v_0)$ can be recast in expectation form. Define

$$G(t) = \Pr\{T_{\nu_1+\nu_2} \le t\},$$

and let U be a Beta random variable with density

$$f_U(u) = \frac{u^{\nu_1/2-1}(1-u)^{\nu_2/2-1}}{B(\nu_1/2, \nu_2/2)}, \quad 0 \le u \le 1.$$

Then the CDF can be written as

$$F_v(v_0) = \mathbb{E}\Big[G\big(v_0\,\varphi(U)\big)\Big].$$

In practice, this one-dimensional integral is computed either by numerical quadrature or by Monte Carlo simulation. In the quadrature approach, one employs standard routines to evaluate

$$\int_0^1 G\Big(v_0\,\varphi(u)\Big)\,f_U(u)\,du,$$

while in a Monte Carlo approximation one would generate N independent samples $u_1, \ldots, u_N$ from the Beta distribution and approximate the expectation by

$$\widehat{F_v}(v_0) = \frac{1}{N}\sum_{i=1}^{N} G\Big(v_0\,\varphi(u_i)\Big).$$

This reduction to a one-dimensional expectation is particularly attractive in view of its computational tractability.

1.4 Visualization

For the purpose of illustration, we now specialize to the case where $\nu_1 = \nu_2 = v$. Under this simplification, the density of U becomes

$$f_U(u) = \frac{u^{v/2-1}(1-u)^{v/2-1}}{B(v/2, v/2)},$$

and the depth formula reduces to

$$\mathrm{TD}(x) = 1 - \int_0^1 G\Big(\|x\|\,\varphi(u)\Big)\,f_U(u)\,du,$$

with the optimal projection determined by $x/\|x\|$.

To provide a concrete visual interpretation, consider contour plots of the Tukey depth in $\mathbb{R}^2$. In these plots, the depth at a point x is computed as above, with the scalar projection given by its Euclidean norm. Figure 1.1 shows three panels corresponding to different values of the degrees-of-freedom parameter v: $v = 1$, $v = 5$, and $v = 100$. In each case the depth is calculated over a grid of points, and the contour plots illustrate how the depth function changes with v and provide insight into the centrality of points in the plane as measured by Tukey depth.

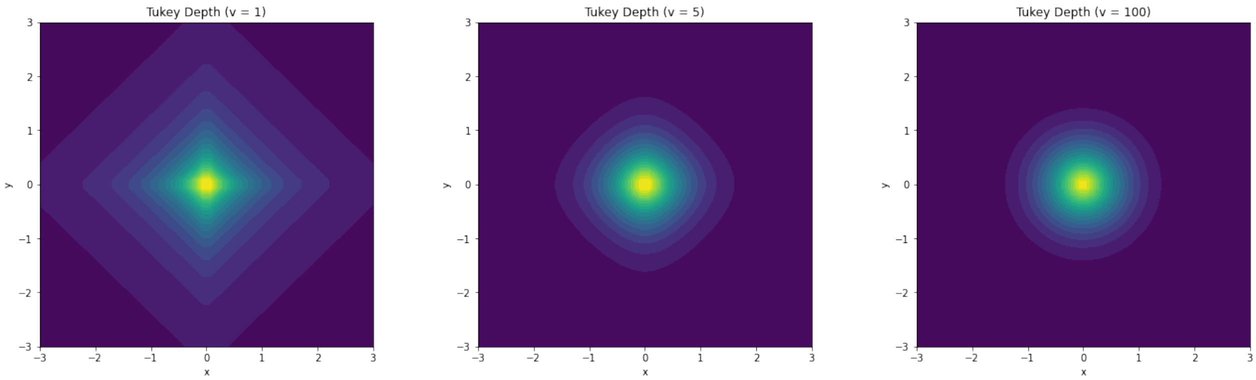

Figure 1.1: Contour plots of the Tukey Depth for three values of v.

Bibliography

[1] Harold Ruben. On the distribution of the weighted difference of two independent student variables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 22(1):188–196, 1960.