A CHARACTERIZATION OF SYMMETRIES OF CONDITIONAL DISTRIBUTIONS

HENGRUI LUO

Abstract. This paper I provided a characterization of Hogg-Randles-Wolfe pairing transformation that can be used to construct symmetric properties about certain class of statistics.

1. Hogg-Wolfe Pairing Transformations

In an earlier paper [1] Wolfe generalized the result of Hogg and Hollander and propose a sufficient (and actually necessary) condition for a specific statistic $U(\boldsymbol{X})$ to possess symmetric distribution. We restated his result below in a sufficient and necessary fashion.

Theorem 1. *(Wolfe's theorem.* [1] *Theorem 2.2) Let $\boldsymbol{X} = (X_1, X_2, \cdots X_N)$ be a sample*[1]*.*

The conditional distribution $U(\boldsymbol{X}) \mid V(\boldsymbol{X}) = v$ is symmetric about μ for every $v \in$ $\mathrm{supp} V(\boldsymbol{X})$

if and only if following conditions are satisfied.

(1) There exists a one-to-one transformation $g : \mathbb{R}^N \to \mathbb{R}^N$ such that $U(\boldsymbol{x}) - \mu = \mu - U(g(\boldsymbol{x}))$ for all $\boldsymbol{x} \in \mathbb{R}^N$ and $V(\boldsymbol{x}) = V(g(\boldsymbol{x}))$ for all $\boldsymbol{x} \in \mathbb{R}^N$.

(2) There exists a distribution equivalence $g(\boldsymbol{X}) \overset{d}{=} \boldsymbol{X}$.

Proof. (Necessity) If the conditional distribution $U(\boldsymbol{X}) \mid V(\boldsymbol{X}) = v$ is symmetric about μ for every $v \in \mathrm{supp} V(\boldsymbol{X})$, then we take $g = id$ and the conditions (1) (2) are readily verified.

[1] which is not necessarily i.i.d. or with additional assumptions.

(Sufficiency) Using the equal in distribution technique, we have

$$\begin{aligned}(U(\boldsymbol{X})-\mu,V(\boldsymbol{X})) &\underset{\text{by (2)}}{\overset{d}{=}} (U(g(\boldsymbol{X}))-\mu,V(g(\boldsymbol{X})))\\ &\underset{\text{by (1)}}{\overset{d}{=}} (\mu-U(g(\boldsymbol{X})),V(g(\boldsymbol{X})))\\ &\underset{\text{by (2)}}{\overset{d}{=}} (\mu-U(\boldsymbol{X}),V(\boldsymbol{X}))\end{aligned}$$

□

The proof is basically word-by-word rephrase of [1]. The case that $V(\boldsymbol{X})$ being a degenerated statistics will lead to Theorem 2.1 of [1] or Theorem 1.3.16 of [4]. Then we call this transformation $g:\mathbb{R}^N\to\mathbb{R}^N$ the *Hogg-Wolfe pairing transformation* for reasons we are to explain below. The existence and uniqueness of this pairing transformation g is of interest.

This result has various implications in nonparametric statistics[2]. For example the Wilcoxon rank statistics under null distribution for two samples $\boldsymbol{X},\boldsymbol{Y}$ we can actually see that the $g:(X_1,X_2,\cdots X_N)\mapsto(X_N,X_{N-1},\cdots X_1)$ will produce the symmetry stated in the result above. In this sense the g served as a "pairing" between two samples $\boldsymbol{X},\boldsymbol{Y}$. One interesting observation is that it is usually the case that $g^2=id$, for example pp.21~22 of [4]. And this is actually not a coincidence but leads to a characterization of the pairing transformation g as we shall see below.

Theorem 2. *Assume that the distribution admits a density function, then the pullback operator $\tau_g: f\mapsto f\circ g^{-1}$ induced by the pairing transformation g that satisfies (1)(2) must be a unitary operator on the space of all density functions endowed with inner product $\langle f_1,f_2\rangle:=\oint f_1(\boldsymbol{x})f_2(\boldsymbol{x})d\boldsymbol{x}$.*

Proof. Without loss of generality we can assume the density function is continuous, the case we have discrete density function is analogous. Consider the transformation $\boldsymbol{Y}:=g(\boldsymbol{X})$, we obtain by the transformation formula $f_{\boldsymbol{Y}}(\boldsymbol{y})=\sum_{\boldsymbol{x}\in g^{-1}(\boldsymbol{y})}f_{\boldsymbol{X}}(\boldsymbol{x})$ that

$$F_{\boldsymbol{Y}}(\boldsymbol{y}_0)=\oint_{\boldsymbol{y}\le\boldsymbol{y}_0}f_{\boldsymbol{Y}}(\boldsymbol{y})d\boldsymbol{y},\forall\boldsymbol{y}_0\in\mathbb{R}^N$$

[2]Thanks to a comment by Prof.Critchlow.

where the notation $\boldsymbol{y} \leq \boldsymbol{y}_0$ defines the integration domain $\left\{ \boldsymbol{y} \in \mathbb{R}^N : y_i \leq y_{0i}, \forall i = 1, 2, \cdots N \right\}, \boldsymbol{y}_0 = (y_{01}, y_{02}, \cdots y_{0N})$. And substitute the transformation formula above we have

$$
\begin{aligned}
F_{\boldsymbol{Y}}(\boldsymbol{y}_0) &= \oint_{\boldsymbol{y} \leq \boldsymbol{y}_0} f_{\boldsymbol{Y}}(\boldsymbol{y}) d\boldsymbol{y} \\
&= \oint_{\boldsymbol{y} \leq \boldsymbol{y}_0} \sum_{\boldsymbol{x} \in g^{-1}(\boldsymbol{y})} f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{y} \\
&= \oint_{\boldsymbol{y} \leq \boldsymbol{y}_0} \left(f_{\boldsymbol{X}} \circ g^{-1} \right) (\boldsymbol{y}) d\boldsymbol{y}
\end{aligned}
$$

since the pairing transformation must be one-to-one by (1) and dimension preserving property in (2). (2) requires further that $F_{\boldsymbol{Y}}(\boldsymbol{y}_0) = Pr(\boldsymbol{Y} \leq \boldsymbol{y}_0) = Pr(\boldsymbol{X} \leq \boldsymbol{y}_0) = F_{\boldsymbol{X}}(\boldsymbol{y}_0)$ and thus

$$
\oint_{\boldsymbol{y} \leq \boldsymbol{y}_0} \left(f_{\boldsymbol{X}} \circ g^{-1} \right) (\boldsymbol{y}) d\boldsymbol{y} = \oint_{\boldsymbol{x} \leq \boldsymbol{y}_0} f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x}, \forall \boldsymbol{y}_0 \in \mathbb{R}^N
$$

The dummy variable in the integral is not relevant and we can write it in form of an integral equality

$$
\oint_{\boldsymbol{u} \leq \boldsymbol{y}_0} \left(f_{\boldsymbol{X}} \circ g^{-1} - f_{\boldsymbol{X}} \right) (\boldsymbol{u}) d\boldsymbol{u} = 0, \forall \boldsymbol{y}_0 \in \mathbb{R}^N
$$

Therefore by Portmanteau Theorem of probability [5], we can assert that $f_{\boldsymbol{X}} \circ g^{-1} = f_{\boldsymbol{X}}$ with probability one. But both of these two functions shall integrate to 1 so if we define $\langle f_1, f_2 \rangle := \oint f_1(\boldsymbol{x}) f_2(\boldsymbol{x}) d\boldsymbol{x}$ then following derivations $\langle f_{\boldsymbol{X}}, f_{\boldsymbol{Y}} \rangle = \langle f_{\boldsymbol{X}} \circ g^{-1}, f_{\boldsymbol{Y}} \rangle = \langle f_{\boldsymbol{X}} \circ g^{-1}, f_{\boldsymbol{Y}} \circ g^{-1} \rangle = \langle \tau_g f_{\boldsymbol{X}}, \tau_g f_{\boldsymbol{Y}} \rangle$ asserts that $\tau_g : f \mapsto f \circ g^{-1}$ is a unitary operator on the space of all density functions endowed with inner product specified above. □

Combine above theorems, we can establish the following geometric characterization of a symmetric distribution

Theorem 3. *Let $\boldsymbol{X} = (X_1, X_2, \cdots X_N)$ be a sample with density functions. The conditional distribution $U(\boldsymbol{X}) \mid V(\boldsymbol{X}) = v$ is symmetric about μ for every $v \in \text{supp} V(\boldsymbol{X})$ iff there exists* *an isometry* $\tau_{g,v}$ *on the inner product space of all conditional density functions endowed with inner product* $\langle f_1, f_2 \rangle := \oint f_1(\boldsymbol{x}) f_2(\boldsymbol{x}) d\boldsymbol{x}$ *for each* $v \in \mathrm{supp} V(\boldsymbol{X})$*.*

And this indicates that the $U(\boldsymbol{X})$ must be chosen from a relatively narrow class of mapping $\eta : \mathbb{R}^N \to \mathbb{R}^m$ whose density function space of image space has plenty of isometries. In other words the restriction of isometries of the density function space will in turn restrict the symmetries of the conditional distribution.

References

[1] Wolfe, Douglas A. "Some general results about uncorrelated statistics." Journal of the American Statistical Association 68.344 (1973): 1013-1018.

[2] Hogg, Robert V. "Certain uncorrelated statistics." Journal of the American Statistical Association 55.290 (1960): 265-267.

[3] Hollander, Myles. "Certain uncorrelated nonparametric test statistics." Journal of the American Statistical Association 63.322 (1968): 707-714.

[4] Randles, Ronald H., and Douglas A. Wolfe. Introduction to the theory of nonparametric statistics. Vol. 1. New York: Wiley, 1979.

[5] Resnick, Sidney I. A probability path. Springer Science & Business Media, 2013.