

# The Mann-Whitney-Wilcoxon random field, with applications to brain mapping

Farzan Rohani<sup>\*1</sup>, Masoud Asgharian<sup>1</sup>, Keith J. Worsley<sup>1,2</sup>

<sup>1</sup>Department of Mathematics and Statistics, McGill University, Montreal, Canada

<sup>2</sup>Montreal Neurological Institute, McGill University, Montreal, Canada

*E-Mail:* rohani@math.mcgill.ca

*Abstract:* Suppose we have two groups of observations at each point in a subset  $S$  of Euclidean space, so that, instead of being scalars, the observations are smooth random fields. We define the Mann-Whitney-Wilcoxon (MWW) random field as the MWW test statistic evaluated at each point in  $S$ . We are interested in the null distribution of the maximum MWW over all points in  $S$ .

The motivation comes from brain mapping, where the observations may be functional magnetic resonance images (fMRI) or anatomical images such as density of grey matter or multiple sclerosis lesions. The two groups may be fMRI measurements made under a control and a task condition, or in the case of anatomical data, a control group and a disease group. The difference between the two groups is thought to be confined to a small number of localised regions in the brain, so we calculate a test statistic  $T(s)$  at each point  $s$  in the brain  $S$ , then determine a threshold so that the  $p$ -value  $P(\max_S T(s) \geq t) = \alpha$ , where  $\alpha$  is a desirable level of significance. The brain regions where  $T(s) \geq t$  are the places where there is a difference between the two groups.

It is commonly assumed that such observations are Gaussian random fields, and there is a well-developed theory for approximations to

$$P(\max_S T(s) \geq t) = \alpha$$

where  $T(s)$  is for example a two-sample T statistic [5]. However there is considerable evidence that even in large samples these approximations are not accurate when the data is not Gaussian. The reason is that the threshold  $t$  is so far out in the tails of the null distribution of  $T(s)$  that the usual Gaussian approximation (by appealing to the Central Limit Theorem) is not accurate. One solution is to base inference on the permutation distribution

of  $\max_S T(s)$ , found by permuting the group labels of the data [4], but this requires time-consuming simulations. Instead we propose replacing  $T(s)$  by the MWW statistic at each point.

In the Gaussian case, the random field theory approximation to the  $p$ -value is the expected Euler characteristic (EC) of the excursion set  $A_t = \{s : T(s) \geq t\}$  [1,2]. We will find a closed form expression for the expected EC of 1D and 2D MWW random fields. Although the resulting test is not fully non-parametric, it depends on a particular smoothness parameter of the data which is estimable from within the groups.

We apply our method to a data set on the difference in multiple sclerosis (MS) lesion density between a group of 212 patients with high disability (EDSS>2) and 213 patients with low disability (EDSS<=2)[3].