FWER

Suppose we want to test m null hypothesis, and suppose we do that with the help of m test statistics: $\hat{\psi}_1, \ldots, \hat{\psi}_m$. We assume that the critical value c is the same for all of them (though in the paper we work with scale-dependent critical values) and we would like to find such $c := c(\alpha)$ that the FWER is controlled at level α.

Denote the set of indices where the null hypothesis is true by $\mathcal{M}_0$. Without loss of generality, we can assume that $\mathcal{M}_0 = \{1, \ldots, m_0\}$ with $m_0 < m$. By definition, the family-wise error rate (FWER) is equal to the probability of at least one Type I error, that is,

$$FWER = \mathbb{P}(\hat{\psi}_1 > c \text{ or } \ldots \text{ or } \hat{\psi}_{m_0} > c).$$

We can always rewrite it as

$$FWER = 1 - \mathbb{P}(\hat{\psi}_1 \leq c \text{ and } \ldots \text{ and } \hat{\psi}_{m_0} \leq c).$$

Hence, to control the FWER at level α, we can pick c such that

$$\mathbb{P}(\hat{\psi}_1 \leq c \text{ and } \ldots \text{ and } \hat{\psi}_{m_0} \leq c) \geq 1 - \alpha.$$

Trivially,

$$\mathbb{P}(\hat{\psi}_1 \leq c \text{ and } \ldots \text{ and } \hat{\psi}_{m_0} \leq c) = \mathbb{P}\Big(\max_{i \in \{1, \ldots, m_0\}} \hat{\psi}_i \leq c \Big)$$

and from here we can proceed in two different (and quite distinct) ways.

The first way is the most commonly used one: we try to interchange the probability and the max somehow, however, in our setting this is not a trivial task. The second way is the one we use in the paper, and the underlying idea as follows. First, irrespective of the number of the true null hypothesis we can always bound $\max_{i \in \{1, \ldots, m_0\}} \hat{\psi}_i$ by $\max_{i \in \{1, \ldots, m\}} \hat{\psi}_i$. Then

$$\mathbb{P}\Big(\max_{i \in \{1, \ldots, m_0\}} \hat{\psi}_i \leq c \Big) \geq \mathbb{P}\Big(\max_{i \in \{1, \ldots, m\}} \hat{\psi}_i \leq c \Big).$$

This bound is quite crude and this is the reason why our test is conservative, but it helps us in controlling FWER. Second, we treat $\max_{i \in \{1, \ldots, m\}} \hat{\psi}_i$ as a random variable, i.e. denote $\Psi := \max_{i \in \{1, \ldots, m\}} \hat{\psi}_i$. In some special cases we know the distribution of Ψ, for example, when the test statistics are independent of each other and each of the follows a uniform distribution on $[a, b]$, then the distribution function of Ψ is given by

$$F_\Psi(y) = P(\Psi \leq y) = \begin{cases} 0 & y \leq a \\ [(y - a)/(b - a)]^n & y \in (a, b) \\ 1 & y \geq b \end{cases}$$

But this is a simple example, here we know the distribution and dependence structure. If we do not know them, then we approximate $\hat{\Psi}$ with for example

CLT (that works even without knowing the data generating structure), and use the approximated version $\hat{\Phi}$ to determine the appropriate value of c. Then we have:

$$\mathbb{P}(\hat{\psi}_1 \leq c \text{ and } \ldots \text{ and } \hat{\psi}_{m_0} \leq c)$$
$$= \mathbb{P}\left(\max_{i \in \{1,\ldots,m_0\}} \hat{\psi}_i \leq c \right)$$
$$\geq \mathbb{P}\left(\max_{i \in \{1,\ldots,m\}} \hat{\psi}_i \leq c \right)$$
$$= \mathbb{P}\left(\hat{\Psi} \leq c \right)$$
$$\approx \mathbb{P}\left(\hat{\Phi} \leq c \right)$$
$$\geq 1 - \alpha$$

with c being equal to the appropriate quantile of the distribution of $\hat{\Phi}$.

Here we never actually use which of the null hypothesis are true and which are not, hence, the described procedure completely coincides with the definition in Dudoit et al. (2003), with FWER being "...defined under the true and typically unknown data generating distribution for $\mathbf{X} = (X_1, ..., X_m)$ and Y. In particular, they depend upon which specific subset $\Lambda_0 \subseteq \{1, ..., m\}$ of null hypotheses is true for this distribution".