and considering the empirical distribution function $\hat{F}_n(\cdot)$, the halfspace depth will be

$$d_{HS}(x, \hat{F}_n) = \min\left(\frac{1}{n}\sum_{i=1}^{n} I(X_i \leq x), \frac{1}{n}\sum_{i=1}^{n} I(X_i \geq x)\right).$$

Consider $\mathbf{T_{0n}} = (\mathbf{T_{0n,1}}, \ldots, \mathbf{T_{0n,p}})$ and $\mathbf{S_{0n}} = (\mathbf{S_{0n,1}}, \ldots, \mathbf{S_{0n,p}})$, a pair of initial location and dispersion estimators. Here we choose for $\mathbf{T_{0n}}$ and $\mathbf{S_{0n}}$ respectively the coordinate-wise median and the median absolute deviation (MAD). For each variable $(X_{1j}, X_{2j}, \ldots, X_{nj})$ $(j = 1, \ldots, p)$, we denote the standardized version of X_{ij} by $Z_{ij} = \frac{X_{ij} - T_{0n,j}}{S_{0n,j}}$. Let F_j a chosen reference distribution for Z_{ij}; here we use the standard normal distribution, i.e., $F_j = \Phi$. Let $\hat{F}_{n,j}$ be the empirical distribution for the standardized values, that is

$$\hat{F}_{n,j}(t) = \frac{1}{n}\sum_{i=1}^{n} I(Z_{ij} \leq t) \qquad j = 1, \ldots, p.$$

We define the proportion of flagged outliers by

$$d_{n,j} = \max\left(\sup_{t \leq -\eta_{\beta,j}} \{d_{HS}(t, \hat{F}_{n,j}) - d_{HS}(t, F_j)\}^+; \sup_{t \geq \eta_{\beta,j}} \{d_{HS}(t, \hat{F}_{n,j}) - d_{HS}(t, F_j)\}^+\right),$$

where $\eta_{\beta,j} = F_j^{-1}(\beta)$ is a large quantile of F_j. Note that, according to (), we are considering the set $C^{\beta}(F_j) = \{x \in \mathbb{R} : d_{HS}(x, F_j) < d_{HS}(\eta_{\beta,j})\}$, which results in the simpler form written above considering the definition of the half-space depth in the univariate case. Here, if we consider the order statistics $Z_{(i),j}$, define $i_- = \min\{i : Z_{(i),j} > -\eta_{\beta,j}\}$ and $i_+ = \max\{i : Z_{(i),j} < \eta_{\beta,j}\}$. Using the definition of half-space depth function in the univariate case, presented above, the previous expression can be written as

$$d_{n,j} = \max\left(\sup_{i < i_-}\{\frac{i}{n} - F_j(Z_{(i),j})\}^+, \sup_{i > i_+}\{F_j(Z_{(i),j}) - \frac{i-1}{n}\}^+\right). \qquad (1)$$

Then, we flag $\lfloor n d_{n,j} \rfloor$ observations with the smallest depth value as cell-wise outliers and replace them by NA's.

0.1 A consistent univariate, bivariate and p-variate filter

Given a sample $\mathbf{X_1}, \ldots, \mathbf{X_n}$ where $\mathbf{X_i} \in \mathbb{R}^{\mathbf{p}}, \mathbf{i} = \mathbf{1}, \ldots, \mathbf{n}$, we first apply the univariate filter described in the previous example to each variable separately. Filtered data are indicated through an auxiliary matrix $\mathbf{U}$ of zeros and ones,