Robust modification to treatment of censoring

In the paper, equation (A5) states that

$$e^{(c)}_{rsl}\hat{p}_{rslc} = n^{(c)}_{rsl}\tilde{p}_{rslc}$$

defines the expected counts $e^{(c)}_{rsl}$. $\tilde{p}_{rslc}$ is said to represent the fitted probabilities from the null (Markov) model. Hence these would typically be defined as

$$\tilde{p}_{rslc} = \frac{1}{n_{rlc}} \sum_i p_{rs}(t_i, z_i, \hat{\theta})$$

where the sum is over all observations in group l, c with previous state, and n_{rlc} is the number of such observations.

However, the implicit assumption in estimating $\hat{p}_{rs}$ by the expression in (A3) is that the censored observations are of a comparable length to non-censored observations, conditional on being in the same time quantile. If this isn't the case, the observed and expected counts will not be consistent under the null model. This will lead to an inflated mean for the null distribution and some loss in power of the test.

For robustness, a modified treatment is used in the function. $\hat{p}_{rs}$ is calculated as before. But then we define

$$\tilde{p}^{*}_{rslc} = \frac{\tilde{n}_{rslc}}{n_{rlc}} \left(\frac{\sum_{j \neq R} \tilde{n}_{rjlc}}{\sum_{j \neq R,C} \tilde{n}_{rjlc}} \right)$$

where

$$\tilde{n}_{rslc} = \sum_i p_{rs}(t_i, z_i, \hat{\theta})$$

where the sum is over all observations in group l, c with previous state, and n_{rlc} is the number of such observations. We then define the expected counts using

$$e^{(c)}_{rsl}\hat{p}_{rslc} = n^{(c)}_{rsl}\tilde{p}^{*}_{rslc}$$

This has the effect of transforming the fitted transition probabilities in the Markov model, in the same way that is assumed for the observed counts. This ensures that the expectation of the observed counts does equal the expected counts if the null model is true.