When estimating an entropy $H[X]$, call the plug-in / nonparametric maximum likelihood estimator (i.e. the estimator you've been using) $\hat{H}[X]$. Call $\hat{\mathcal{X}}$ the *observed* alphabet of X. Then the Miller-Madow estimator of the entropy is

$$\hat{H}_{MM}[X] = \hat{H}[X] + \frac{|\hat{\mathcal{X}}| - 1}{2n} \tag{1}$$

where $|\hat{\mathcal{X}}|$ is the number of observed symbols and n was the number of samples used to estimate $\hat{H}[X]$.

Using the definition from line 368 of your code, we see that transfer entropy can be written as

$$\begin{aligned} TE^{(k)}_{X\to Y} &= H(Y_t|Y^{t-1}_{t-k}) - H(Y_t|Y^{t-1}_{t-k}, X^{t-1}_{t-k}) && (2)\\ &= H(Y_t, Y^{t-1}_{t-k}) - H(Y^{t-1}_{t-k}) - H(Y_t, Y^{t-1}_{t-k}, X^{t-1}_{t-k}) + H(Y^{t-1}_{t-k}, X^{t-1}_{t-k}), && (3)\end{aligned}$$

where $X^{t-1}_{t-k} = (X_{t-k}, X_{t-(k-1)}, \ldots, X_{t-1})$. We would apply the Miller-Madow estimator individually to each of the entropy terms. For example, for the first term, we have

$$\hat{H}_{MM}[Y_t, Y^{t-1}_{t-k}] = \hat{H}_{MM}[Y^t_{t-k}] = \hat{H}[Y^t_{t-k}] + \frac{|\widehat{\mathcal{Y}^{k+1}}| - 1}{2n}, \tag{4}$$

where $|\widehat{\mathcal{Y}^{k+1}}|$ is the number of $(k+1)$-tuples we actually observe (of the 2^{k+1} possible tuples). Doing this for each term, the overall Miller-Madow estimator for the transfer entropy is

$$\begin{aligned} \widehat{TE}^{(k)}_{X\to Y} &= \hat{H}_{MM}(Y_t|Y^{t-1}_{t-k}) - \hat{H}_{MM}(Y_t|Y^{t-1}_{t-k}, X^{t-1}_{t-k}) && (5)\\ &= \hat{H}_{MM}(Y_t, Y^{t-1}_{t-k}) - \hat{H}_{MM}(Y^{t-1}_{t-k}) - \hat{H}_{MM}(Y_t, Y^{t-1}_{t-k}, X^{t-1}_{t-k}) + \hat{H}_{MM}(Y^{t-1}_{t-k}, X^{t-1}_{t-k}). && (6)\end{aligned}$$

One possible problem with this estimator is that it can result in *negative* estimates of entropies. That usually occurs when $\hat{H}$ is very small. I usually record both $\hat{H}$ and $\hat{H}_{MM}$.