

## Appendix A

# Censored Data

As described in Sec. 3.1.5, any simulation exceeding  $T_{cens}$  is terminated. As a consequence, we will have a record of the repetitions where  $T_{sync} < T_{cens}$  but no value of  $T_{sync}$  for the censored results. Despite the incompleteness of the data, we can estimate the cumulative distribution function (CDF) of the synchronization times by using the Kaplan-Meier, K-M, estimator (Kaplan and Meier, 1958).

Assuming that out of  $M$  trials, there are  $K$  where  $T_{sync}^i < T_{cens}$ , the procedure for calculating K-M estimator is as follows (Natrella, 2010):

1. Sort  $T_{sync}^i$  in increasing order, from  $i = 1$  to  $i = K$ .
2. Associate a number  $n_i$  for each  $T_{sync}^i$ .  $n_i$  is the number of trials that take longer than  $T_{sync}^i$  to synchronize.
3. Calculate  $R(T_{sync}^1) = (n_1 - 1) / n_1$ .
4. Calculate  $R(T_{sync}^i) = R(T_{sync}^{i-1}) (n_i - 1) / n_i$ .
5. The CDF is estimated as  $F(T_{sync}^i) = 1 - R(T_{sync}^i)$

In this way, the censored values are counted up to the latest recorded synchronization time,  $T_{sync}^K$ . Note that the CDF is calculated without making any assumption about the form of the probability distribution function (PDF) of the data.

By examining our results we concluded that the data best fits a Weibull distribution for all the performed simulations. By adjusting the K-M estimator to the

CDF of a Weibull distribution,

$$F(T_{sync}) = 1 - e^{-\left(\frac{T_{sync}}{\alpha}\right)^\gamma},$$

we obtain its two parameters,  $\alpha$  and  $\gamma$ . Lastly, we can calculate the mean of  $T_{sync}$  as

$$\langle T_{sync} \rangle = \alpha \Gamma\left(1 + \frac{1}{\gamma}\right),$$

where  $\Gamma$  is the Gamma function.

The accuracy of this method decreases as the number of censored values increases. Therefore, we required a minimum of 10% of the measures to be exactly obtained in order to calculate their average. If this condition is not met then we consider the mean value as undetermined.

Figure A.1 shows how the estimation of censored data fits the uncensored points with high precision for synchronization using nearest neighbor interaction. Each violin plot shows the distribution of  $T_{sync}$  for a given agent speed,  $V$ . The solid line shows the estimated  $T_{sync}$  when setting  $T_{cens} = 10^7$  as in the main text. Note that the estimated mean value can exceed the censoring threshold if enough point were censored.