Analysis of Akaike's Information Criterion for Propagation Delays in a Free-Diffusion Channel

Ethungshan Shitiri†, H. Birkan Yilmaz‡, and Ho-Shin Cho†
†*School of Electronics Engineering, Kyungpook National University, Daegu, South Korea*
‡*Department of Network Engineering, Polytechnic University of Catalonia, Barcelona, Spain*

***Index Terms*—Akaike's Information Criterion; diffusion; molecular communication; propagation delay distribution**

I. Introduction

Information carrying molecular signals propagate in a molecular communication channel either with assisted- or free-diffusion. In the former case, flow is present in the medium whereas in the latter case flow is absent. In both cases, a molecular signal is generally detected when its concentration crosses a certain level. This level may either be a predetermined concentration value or the maximum concentration value. Regardless of the type of level chosen, the propagation delay is defined as the duration it takes for the signal's concentration to attain a specific level. The propagation delay is known to be influenced by a number of factors, such as the number of emitted molecules, the distance, and the diffusion coefficient. These factors introduce noises in the measured concentration, and thereby variations in the propagation delay occur. As one might expect, the propagation delay is random and statistical models are used to characterize its distribution. Recent studies, such as [1], have assumed Normal distribution. However, it is unclear if the assumption holds for the propagation delay that is defined by the maximum concentration of a signal. Motivated by the gap in the literature, we wish to study the distribution of the propagation delay that is related to the maximum concentration, which is commonly referred to as the *peak* [2].

The main contributions of this short paper are summarized here. We conduct preliminary studies to evaluate statistical models that can model the distribution of the propagation delay. We consider two candidate distributions: Burr and Stable distribution. The reason we considered Burr distribution is attributed to its ability to fit various shapes of distributions, while we considered Stable distribution for its richness in skewness and heavy tails [3], [4]. To obtain the data, we perform intensive simulations in a particle-based simulator. For the analysis, we first evaluate the probability plots of the candidate distributions. To further measure the goodness of fit in terms of the information loss, we utilize the Akaike's Information Criterion (AIC) that not only relies on the likelihood of a model's correctness but also on the number of parameters a model uses. We note that an appropriate statistical model will be beneficial for higher layer computations, such as to derive estimators that estimate the clock offset and skews, and thereby achieve symbol synchronization.

II. System Model

We consider an unbounded three-dimensional environment where a distance, d, separates a spherical transmitter nanomachine (Tx) from a spherical receiver nanomachine (Rx). The Tx is assumed to emit G number of molecules as an impulse signal into the environment, which is similar to the instantaneous release of molecules from a vesicle. Flow is absent in the environment, and therefore the emitted signal molecules traverse independently, governed by the dynamics of the Brownian motion. We considered the Tx to be able to reflect the hitting molecules while the Rx is considered to be capable of perfectly absorbing the hitting molecules and thus removing them from the environment. Furthermore, we consider Rx capable of counting the absorbed molecules and identifying the peak concentration. The time it takes to reach the peak, namely the propagation delay, is denoted as T_{delay}.

A. Akaike's Information Criterion

Let n_i denote the number of parameters of the $i-$ th distribution model where $i \in \{Burr, Stable, Normal\}$ and let $\Lambda_{L,i}$ denote the likelihood value of the model. Then the AIC value of the $i-$ th model can be expressed as [5]

$$AIC_i = 2n_i - 2\ln\left(\Lambda_{L,i}\right). \tag{1}$$

III. Simulation Model and Results

The simulations were carried out on the particle-based MolecUlar CommunicatIoN (MUCIN) simulator in MATLAB software [6]. The radius of the nanomachines was set to $4\ \mu m$, distance, d, is set to $5\ \mu m$, diffusion coefficient, D, is set at $79.4\ \mu m^2/s$, and the number of emitted molecules, $G = 8000$ molecules. The concentrations were sampled at regular intervals of $0.001\ s$. We replicated the simulation for 555 replications.

In Fig. 1, we plot the probability plots that provides a comparison between the probabilities of the observed data to the expected probabilities of a distribution model (reference dashed lines). We note that the probability plot maps the probability values on the y-axis to the corresponding data values on the x-axis. We can observe that, for Burr and Stable distributions, most of the data points appear along their reference lines, which suggests that either of them may be a good approximation of the distribution of the observed data. In contrast, for Normal distribution, we can observe that the data

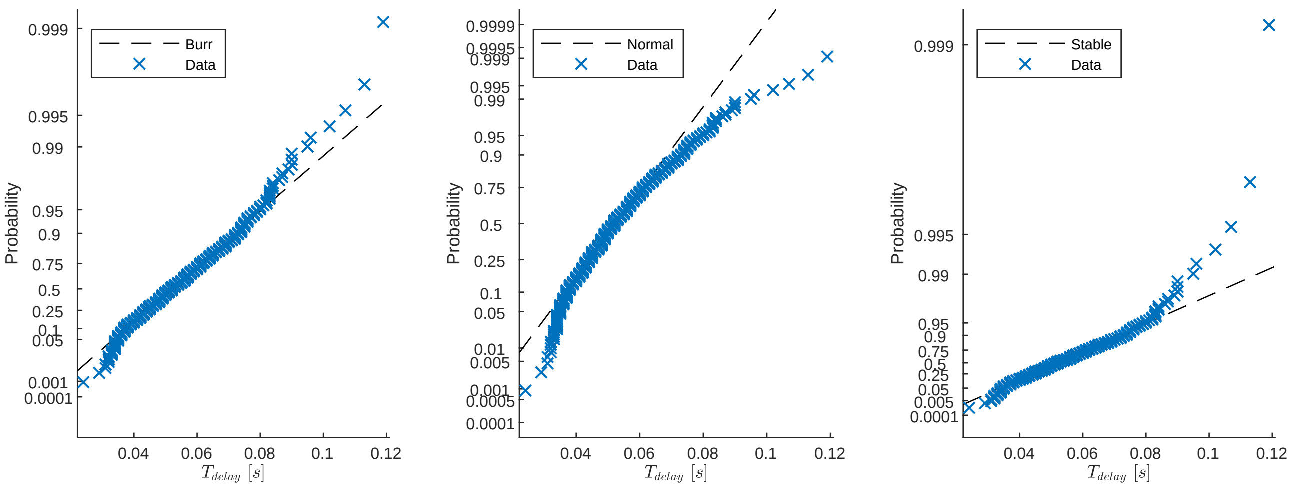

Fig. 1. Probability plots of the estimated models. Each plot represents the probabilities of the observed data with the expected probabilities (dashed lines) of the candidate models

points tend to diverge from the reference line, which suggests that Normal distribution may not be a good approximation.

In Fig. 2, we plot the AIC and the $\ln(\Lambda_{L,i})$ values in the top and bottom, respectively. We can observe that Burr distribution has the lowest information loss as indicated by its low AIC value, meaning it fits the data well. Contrary to common notion of assuming Normal distribution, we can observe that Normal distribution has the highest information loss and therefore it fits the data poorly. The results of AIC is further corroborated by the results of $\ln(\Lambda_{L,i})$. From here, we can infer that Burr distribution has the best estimates of the observed data, while the Normal distribution struggles to estimate the observed data. We also note that Stable distribution, which has four parameters, comes very close to Burr distribution, which has only three parameters and it suggests that the relatively simple Burr distribution describes the data better. We note that, while AIC is sufficient for the analysis, we included $\ln(\Lambda_{L,i})$ to show that AIC does indeed penalize distributions that use a higher number of parameters. However, as we observed here, Stable distribution already has a lower likelihood than Burr distribution and no significant changes are observed in its AIC value.

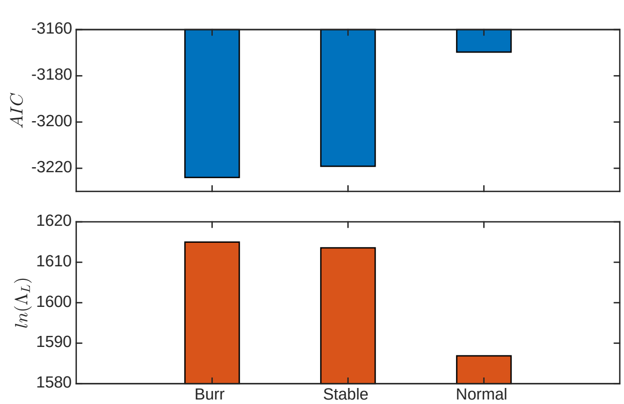

Fig. 2. Akaike's Information Criterion and Log-likelihood values of the estimated models

IV. Conclusions and Future Works

In this short paper, we have shown that alternative statistical models may fit the distribution of propagation delay experienced by a molecular signal much better than Normal distribution. We treated the observed data to probability plot and AIC analysis. Between the two models that we considered, we found that Burr distribution has the lowest information loss and is, therefore, a viable candidate. Future works include a mathematical treatment and analysis to validate the findings presented here.

Acknowledgment

This study was supported by the BK21 PLUS project funded by the Ministry of Education, Korea (21A20131600011) and the National Research Foundation of Korea (NRF) grant funded by the Korean Government (MSIP) (2017R1A2B4002622).

References

[1] L. Lin, C. Yang, M. Ma, S. Ma, and H. Yan, "A clock synchronization method for molecular nanomachines in bionanosensor networks," *IEEE Sensors Journal*, vol. 16, no. 19, pp. 7194–7203, Oct 2016.
[2] H. B. Yilmaz, A. C. Heren, T. Tugcu, and C. Chae, "Three-dimensional channel characteristics for molecular communications with an absorbing receiver," *IEEE Communications Letters*, vol. 18, no. 6, pp. 929–932, June 2014.
[3] R. N. Rodriguez, "A guide to the Burr type XII distributions," *Biometrika*, vol. 64, no. 1, pp. 129–134, April 1977. [Online]. Available: https://doi.org/10.1093/biomet/64.1.129
[4] J. P. Nolan, *Stable Distributions - Models for Heavy Tailed Data*. Boston: Birkhauser, 2018, in progress, Chapter 1 online at http://fs2.american.edu/jpnolan/www/stable/stable.html.
[5] K. P. Burnham and D. R. Anderson, *Model Selection and Multimodel Inference*. Springer-Verlag New York, 2002.
[6] H. B. Yilmaz and C.-B. Chae, "Simulation study of molecular communication systems with an absorbing receiver: Modulation and ISI mitigation techniques," *Simulation Modelling Practice and Theory*, vol. 49, pp. 136 – 150, 2014.