

A Brief Look at Salmonella Run and Tumble

Liane Solomon

Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta, Canada

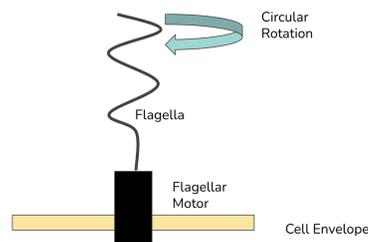


Abstract

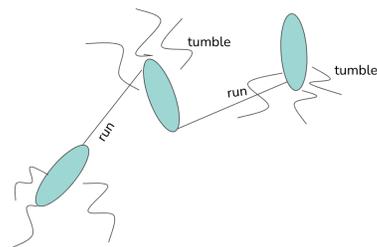
- ▶ A brief biological introduction for salmonella motion in mucus is given
- ▶ Illustrate motivations for modeling such motion
- ▶ Outline the structure of the process model and describe the SDEs
- ▶ Process model is treated as Hidden Markov Model for learning and inference
- ▶ A brief introduction to parameter estimation with incomplete data through Expectation Maximization
- ▶ Will compare impact of different mucosal conditions on salmonella motion
- ▶ Will use additional statistical tools to understand accuracy and things like first passage time

Biological Background and Motivations

- ▶ Salmonella swim with flagella and this creates a distinct movement pattern



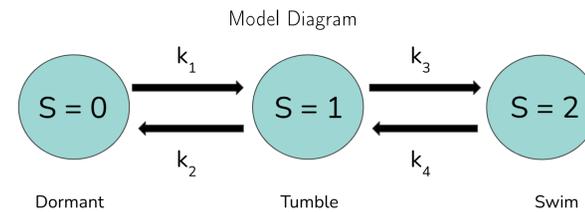
- ▶ Run and Tumble models have been used to model this



- ▶ Flagella rotate synchronously to create forward motion and asynchronously to turn the cell
- ▶ Studies have already shown that various mucosal conditions can impact salmonella motion, in particular in *Rag1^{-/-}* mice, which lack mature T and B lymphocytes
- ▶ Specifically, antibodies that bind to the LPS in salmonella cell walls hinder active motion
- ▶ This is because antibodies act as anchors
- ▶ A model was created to observe the extent of this for Salmonella in mouse GI tracts

Model

- ▶ Cells switch between swimming, tumbling, and dormant
- ▶ Tumble must be used as an intermediate state between swim and dormant
- ▶ This means there are four transition rates



- ▶ Biologically supported by previous run and tumble models and video data
- ▶ Each state experiences some variant of Brownian Motion

SDEs

- ▶ Specific SDEs can then be written to describe each state individually

$$\text{Dormant } (S = 0): dX = \sqrt{2D_0}dW$$

$$\text{Tumble } (S = 1): dX = \sqrt{2D_1}dW$$

$$\text{Swim } (S = 2): dX = v(\phi, \theta)dt + \sqrt{2D_2}dW$$

- ▶ ϕ and θ indicate orientation, uniformly random each time entering the Swim state
- ▶ For simplicity, we can discretized the unit sphere to get a finite number of ϕ and θ angles

Chapman-Kolmogorov Equation

$$\begin{aligned} \frac{\partial}{\partial t} p(n, n_b, \mathbf{x}, t) = & \delta_{n_b,0} D_v \nabla^2 p \\ & + (N - n + 1) k_{\text{on}} p(n - 1, n_b, \mathbf{x}, t) + (n + 1) k_{\text{off}} p(n + 1, n_b, \mathbf{x}, t) \\ & + (n - n_b + 1) a_{\text{on}} p(n, n_b - 1, \mathbf{x}, t) + (n_b + 1) a_{\text{off}} p(n, n_b + 1, \mathbf{x}, t) \\ & - [(N - n) k_{\text{on}} + n k_{\text{off}} + (n - n_b) a_{\text{on}} + n_b a_{\text{off}}] p(n, n_b, \mathbf{x}, t) \end{aligned} \quad (1)$$

Statistical Tools for Parameter Estimation

- ▶ Because of the stochastic nature of the model, in addition to the latent states, statistical tools are needed to estimate parameters
- ▶ Understanding how to estimate parameter values in the case of complete data and the case of incomplete data is necessary

Maximum Likelihood Estimation

- ▶ In the case of complete data, i.e. no latent states, Maximum Likelihood Estimation (MLE) can be used to estimate parameters
- ▶ A likelihood function is function to describe the probability of seeing a certain set of observations, say $X_{1:T} = X_1, X_2, \dots, X_T$, given the parameters θ
- ▶ By optimizing the likelihood function, or equivalently the log-likelihood, we obtain the parameter values that maximize our likelihood of seeing $X_{1:T}$

Expectation Maximization

- ▶ For incomplete data, a more advanced algorithm called Expectation Maximization (EM) is needed as we do not know the complete set of data
- ▶ EM consists of two steps, the E-step and the M-step
- ▶ E-step focuses on optimizing expectations while holding parameters constant
- ▶ M-step focuses on optimizing parameters while holding expectations constant
- ▶ The Forward-Backward algorithm is an algorithm for the E-step which computes two probabilities, $\alpha = P(S_t, X_{1:t})$ and $\beta = P(X_{t+1:T} | S_t)$, by passing through the set of cell position data, $X_{1:t}$, twice (once forward and once backward)
- ▶ These two probabilities are then used to calculate two expectations which represent the probability of being in state S at time t and the joint probability of being in state S_i and time $t - 1$ and state S_j at time t

$$P(S_t = j | X_{1:t}, \theta) = \frac{\alpha_{t,j} \beta_{t,j}}{\sum_i \alpha_{t,i} \beta_{t,i}}$$

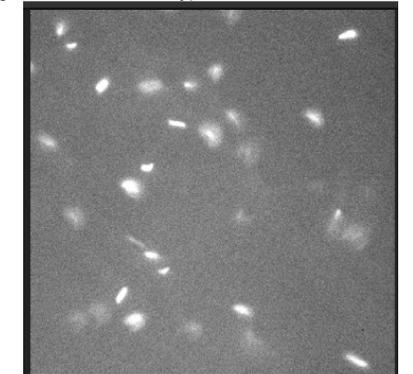
$$P(S_t = j, S_{t+1} = k | X_{1:t}, \theta) = \frac{\alpha_t(k) \phi_{k,j} P(X_{t+1} | S_k) \beta_{t+1}(k)}{\sum_j \sum_k \alpha_t \phi_{t+1,t} P(X_{t+1} | S_{t+1}) \beta_{t+1}}$$

- ▶ Where $\phi_{k,j}$ is the state transition probability matrix
- ▶ These two expectations can then be used to re-estimate the parameters, which makes up the M-step
- ▶ We can iterate over these two steps, recomputing first the expectations then the parameters, until a convergence is reached

Future work

- ▶ In the process of running position data for wild type cells, which means they are a natural strain with no atypical mutations, through EM algorithm to determine parameter values for locations in the GI tract
- ▶ Position data is pulled from microscopy videos via a particle tracking algorithm

Image of Video of Wild-Type cells from Mouse Duodenum



- ▶ Will compare results and evaluate how location in the GI tract impacts motion for wild type cells
- ▶ Confidence of parameter estimation will be examined
- ▶ First Passage Time will be examined
- ▶ Possibly look for population heterogeneity
- ▶ Expand the process model to include population heterogeneity, and use a reducible HMM to estimate the parameter values

References

- [1] A.P. Dempster, N.M. Laird, and D.B. Rubin (1997) Maximum Likelihood for Incomplete Data via the EM Algorithm *Journal of the Royal Statistical Society* Vol. 39, No. 1
- [2] Zoubin Ghahramani (2001) An Introduction to Hidden Markov Models and Bayesian Networks *International Journal of Pattern Recognition and Artificial Intelligence* Vol. 15, No. 1
- [3] Shuichi Nakamura and Tohru Minamino (2019) Flagella-Driven Motility of Bacteria *biomolecules*, 14 July 2019
- [4] L.R. Rabiner and B.H. Juang (1986) An Introduction to Hidden Markov Models *IEEE ASSP Magazine* January 1986
- [5] Holly A. Schroeder, Jay Newby, Alison Schaefer, Babu Subramani, Alan Tubbs, M. Gregory Forest, Ed Miao and Samuel K. Lai (2020) LPS-binding IgG arrests actively motile Salmonella Typhimurium in gastrointestinal mucus *Mucosal Immunology* 13, pages814–823 (2020)