Modeling and Estimating Intratumoral Heterogeneity

Erica M. Rutter, H. T. Banks, and Kevin B. Flores

Center for Research in Scientific Computation,
Department of Mathematics, North Carolina State University

Introduction and Biological Background

- Glioblastoma Multiforme is a deadly primary brain tumor
- Exhibits both strong proliferation and low density migration
- Mathematically modeled assuming 2 separate sub-populations and thus 2 sub-equations

Research Questions

- How can we be sure there are only 2 subpopulations?
- Are we able to recover probability densities independently?
- Can we recover probability densities without assuming anything about the cdf?

Mathematical Model

$$\frac{\partial c(t,x,\boldsymbol{D},\boldsymbol{\rho})}{\partial t} = \nabla \cdot (\boldsymbol{D}\nabla c(t,x,\boldsymbol{D},\boldsymbol{\rho})) + \boldsymbol{\rho} c(t,x,\boldsymbol{D},\boldsymbol{\rho})(1 - c(t,x,\boldsymbol{D},\boldsymbol{\rho}))$$

where

- $c(t,x,\boldsymbol{D},\boldsymbol{\rho})$ is tumor cell population at time t and spatial location x for a subpopulation of the tumor which exhibits growth rate $\boldsymbol{\rho}$ and diffusion $\boldsymbol{D}$
- $\boldsymbol{D}$ and $\boldsymbol{\rho}$ are random variables defined on the compact probability space $\Omega = \Omega_{\boldsymbol{D}} \times \Omega_{\boldsymbol{\rho}}$

Tumor cell population is given by

$$c(t,x) = \mathbb{E}\left[c(t,x,\cdot,\cdot),P\right] = \int_{\Omega} c(t,x,\boldsymbol{D},\boldsymbol{\rho}) dP(\boldsymbol{D},\boldsymbol{\rho})$$

- $P(\boldsymbol{D},\boldsymbol{\rho})$ is the distributions of the parameters
- $c(t,x)$ is aggregate cell population at time t and location x

Inverse Problem

Estimate the probability measure $P(\boldsymbol{D},\boldsymbol{\rho})$ using synthetic data v_{ji}, representing total populations at time t_j and spatial location x_i using M nodes:

Discrete Approximations

$$\hat{P} = \underset{\mathbb{R}}{\operatorname{argmin}} \sum_{j,i} \left[v_{ji} - \left(\sum_{l,k} c(t_j, x_i; D_l, \rho_k) w_l^{M_D} w_k^{M_\rho} \right) \right]^2$$

under the assumption $\sum_{l=1}^{M_D} {w_l}^{M_D} = 1$ and $\sum_{k=1}^{M_\rho} {w_k}^{M_\rho} = 1$.

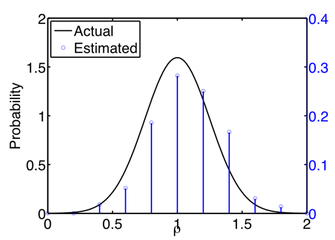
Spline Approximations

$$\hat{P} = \underset{\mathbb{R}}{\operatorname{argmin}} \sum_{i,j} \left[v_{ij} - \sum_{l} a_l \int_{\Omega_\rho} \left(\sum_{k} b_k \int_{\Omega_D} c(t_j, x_i; \boldsymbol{D}, \boldsymbol{\rho}) s_l(\boldsymbol{D}) d\boldsymbol{D} \right) s_k(\boldsymbol{\rho}) d\boldsymbol{\rho} \right]^2$$

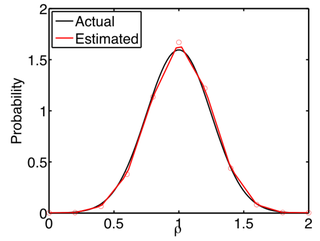
with $\sum_{l=1}^{M_D} a_l \int_{\Omega_D} s_l(\boldsymbol{D}) d\boldsymbol{D} = 1$ and $\sum_{k=1}^{M_\rho} b_k \int_{\Omega_\rho} s_k(\boldsymbol{\rho}) d\boldsymbol{\rho} = 1$

Inverse Problem

Discrete Approximations



Spline Approximations



Choosing Probability Grid Use the Akaike Information Criteria (AIC):

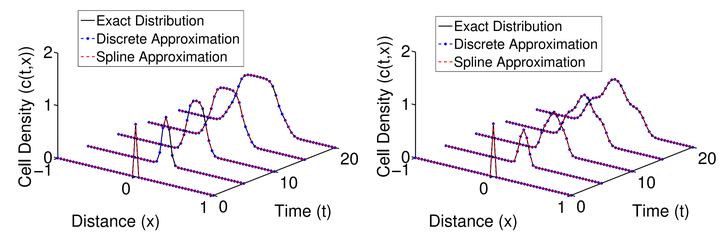
$$AIC = N\nu \ln\left(\frac{\text{RSS}}{N\nu}\right) + N\nu(1+\ln(2\pi)) + 2(M+1)$$

where

- N is the total number of data points
- ν is the number of observables
- RSS is the residual sum of squares error
- M is the number of parameters being estimated (i.e., number of nodes)

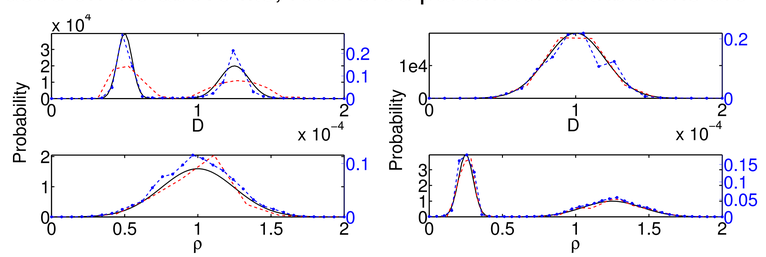
Results

We using synthetic data generated with bigaussian diffusion and normal ρ (left) and bigaussian ρ and normal diffusion (right).



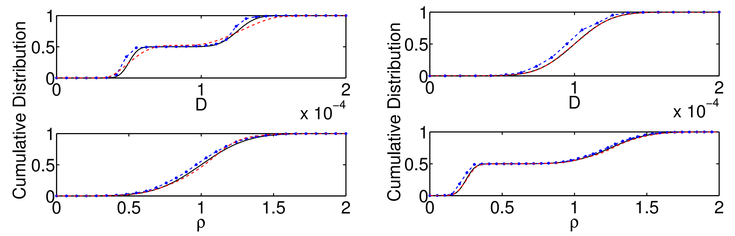
Solutions of the 'best-fit' model using discrete (blue) and spline (red) approximations with exact solution (black).

Based on the above fits, recovered parameter distributions are



The pdf comparisons for the actual distribution (black), the spline approximation (red) and the discrete approximation (blue).

Results



The cdf comparisons for the actual distribution (black), the spline approximation (red) and the discrete approximation (blue).

Treatment Predictions

We consider chemotherapy administered as a modified log-kill hypothesis:

$$\frac{\partial c(t,x)}{\partial t} = \boldsymbol{D}\frac{\partial^2 c(t,x)}{\partial x^2} + \boldsymbol{\rho} c(t,x)(1 - c(t,x)) - r\frac{\boldsymbol{\rho}}{\bar{\rho}} c(t,x)$$

where r represents the strength of chemotherapy.

The estimated tumor burden versus time after initiating treatment. In this case, reaction-diffusion model vastly overestimates efficacy of treatment

Conclusions and Further Directions

Conclusions

- Able to accurately recover probability distributions of parameters from synthetic data up to 5% proportional errors
- Assuming tumor homogeneity may result in overestimating efficacy of treatment

Further Directions

- Structural and practical parameter identifiability
- Uncertainty quantification/sensitivity analysis
- Adaptive meshing of probability nodes
- Test with *in vitro* data

Funding Acknowledgements: NSF grant number DMS-1514929.