Transition Formulas

Mingze Gao

June 29, 2023

Clone specific ODE function (we have 6 clones in total),

$$y'(t) = y(t) \odot K_1(t) + y(t) * K_2(t) - y(t) * K_1^T(t) = f$$
$$y(t_i) = ODESolver(y(t_0), t_0, t_i, f, \theta), t_i \in [0, 3, 10, 17]$$

where $\odot$ is matrix multiplication and $*$ is dot product.

Specifically,

$$K_1(t) = \begin{cases} const^2 * L\_mask \\ [MLP_1(y(t))]^2 * L\_mask \\ mixture\ of\ both \end{cases}$$
$$K_2(t) = \begin{cases} const \\ MLP_2(y(t)) \\ mixture\ of\ both \end{cases}$$
$$T_{i,j}(t) = K_1(t) + diag(K_2(t)), i, j \in P$$

Definitions,

- 1st part of the ODE is inbound transition from other populations, 2nd part of the ODE is self-proliferation and apoptosis process, 3rd part of the ODE is outbound transition to other populations
- Dimension of $y'(t), y(t), K_2(t)$ are $(1, p)$, whilst $K_1(t), L\_mask$ are (p, p)
- $K_1(t)$ is both non-negative and strictly upper triangular matrix confined by square function and a masking term, $K_2(t)$ is a free parameter representing the diagonal line of transition matrix
- MLP_i has 2 layers, Softplus activation, 32 hidden dimensions
- For numerical stability (or maybe to get a less stiffness ODE function), input data is transformed to $y^o_{i,t} = \sqrt[4]{x^o_{i,t}}$, we are using measured real number of cells in population i and time t, and inferred matrix should be $K_i(t) * 4$

Loss function,

$$Loss\_recon = GaussianNLL(y^o_{i,t}, y_{i,t}, \sigma_i^2)\ or\ SmoothL1(y^o_{i,t}, y_{i,t})$$
$$Loss\_pagas = \alpha * \|K_1(t) * (\sim L\_paga)\|_1$$
$$Loss\_upper = \alpha * \|K_2(t) * 4 > 6\|_1$$
$$Loss = Loss\_recon + Loss\_pagas + Loss\_upper$$

Datasets,

- A few simulation data,

  - Given data generated from *K_const*, the model is able to recapitulate the correct dynamics using *const* mode
  - Given raw counts of cells in population i and time t, $x^r_{i,t}$, the model could give corresponding K (although K is meaningless here)
  - Given data generated from *K_const*, fit the data with *mixture* mode, whilst penalize the parameters in MLP_i, the output dynamics K should be similar to *K_const*

- LARRY data at hand
- Weinreb data of hematopoietic differentiation?
- CoSpar?

Issues to be solved,

- The model captured the dynamics well in transformed scale whilst seems over-fitted / under-fitted in original 4^{th} power scale
- Transition rates of certain populations which number of cells equals to 0 in all time points are meaningless, affects visualization
- Variance term σ_i^2 is estimated in transformed scale, how to get correct one in original 4^{th} power scale
- Model fits better without σ_i^2, not sure if it's related to initialization, as how parameters are initialized does affect the performance a lot

Table 1:

	T3		T10		T17	
	Loss	Spearman	Loss	Spearman	Loss	Spearman
const K	28	.545	5781	.881	50250	.973
const K no var	26	.626	2196	.900	8737	.974
K(t)	22	.578	5840	.810	74716	.948
K(t) no var	17	.674	1706	.936	4802	.973
mixture no var	15	.646	2817	.874	51932	.973
mixture diff LR no var	17	.660	360	.954	4019	.973

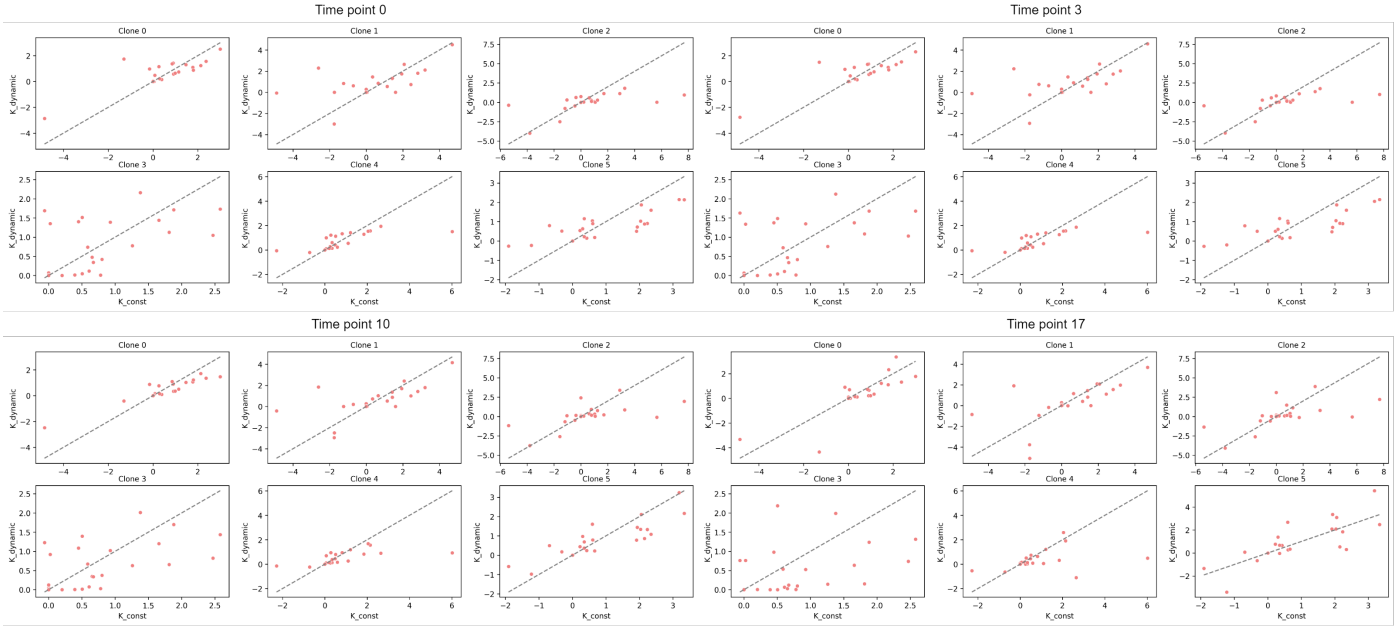
Time point 0
Time point 3
Time point 10
Time point 17
Clone 0
Clone 1
Clone 2
Clone 3
Clone 4
Clone 5
K_dynamic
K_const

Figure 1: