Let s(.) be a statistic that summarize a gene expression distribution P. And s(P) is a cell state proportion vector [a\_i, b\_i, c\_i, d\_i, e\_i] (as described in challenge 1&2) plus a cell count f\_i. From a single cell point of view, a\_i, for example, is the mean probability of the cell being in the progenitor state. Using this statistics s() we can calculate s(P\_0) and s^(P\_i). Let f\_0 denote the cell count in unperturbed condition, and f\_i denote the cell count in the condition i

The final score will be

The rational for this score is that the further the state vector of a condition I from unperturbed condition and the closer the state vector from the desired Q the better. And larger number of cells is preferred.