1. Simulating Random Strings

Currently everything was done by assuming zero developmental noise $p = 0$. So I might use r and $\hat{r}$, q and $\hat{q}$ interchangeably.

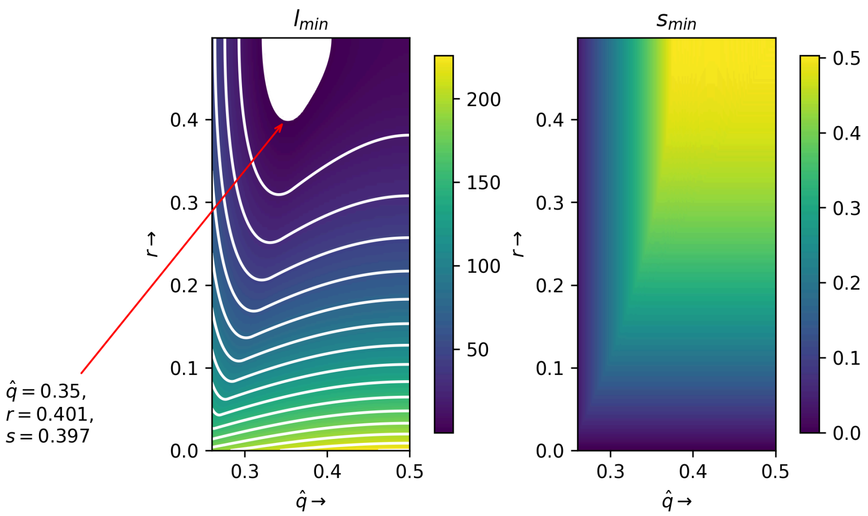


Figure 1: Minimum length of genome required for finite length strings, the white region marks $l_{min} \leq 0$, and the lowest r for zero length genome is found for the string simulations. $r_m = 0.401$, $\hat{q} = 0.35$, $s_m = 0.397$ and $n_0 = 110$ with $\alpha = 0.5$.

Using the parameters obtained above as shown in Figure 1, we use the appropriate $k = \lceil n(1 - H(q, n, t = 2)) \rceil = 19$ to generate 2^k random strings(fixed points) of length $\frac{n_0}{1-\alpha}$. Another string called the *target string*, X^{target} is also generated, with length αn_0. Since $\alpha = 0.5$, we compare the relative hamming distance between the right and left half of the fixed point strings s. The distance r is computed for each fixed point by taking the max relative distance out of X^{target} and left substring or the right substring of the fixed point (Figure 2).

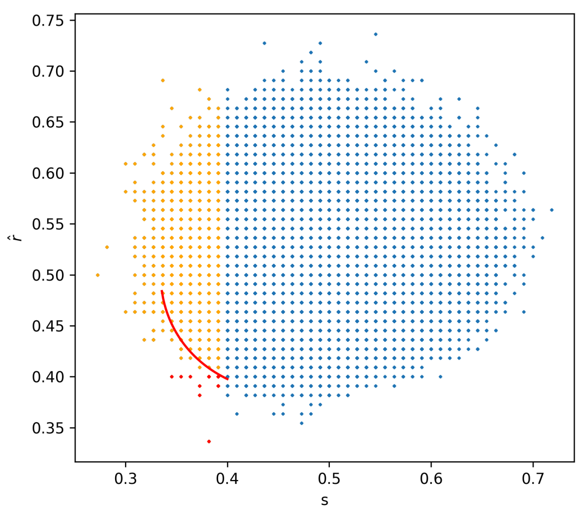


Figure 2: $\hat{r}$ vs s plot for randomly generated strings of total length 220 against a randomly generated target. The curve in red is the solution of the equation $\frac{k}{\alpha n} + H_{t=4}(\{1 - \hat{r} - \frac{s}{2}, \frac{s}{2}, \frac{s}{2}, \hat{r} - \frac{s}{2}\}; n = \alpha n) - 2 = 0$, which represents the points where the expected number of strings is 1.

The scatterplot can be divided into 3 regions by taking two lines $s < s_m$ and $r < r_m$. For our next subset of simulation we pick the string with maximum $s = s_{\text{worst}}$ such that $s_{\text{worst}} \leq s_m$ and with minimum $\hat{r}$. Usually there are approximately 1 or 2 strings. From one of these strings, say X^{t-1} we construct multiple strings by taking the left half of the left substring of X^{t-1} and right half of the right substring of X^{t-1} to construct the left half of the initial condition string, X^t_{ini} and the right half is generated randomly (Figure 3).

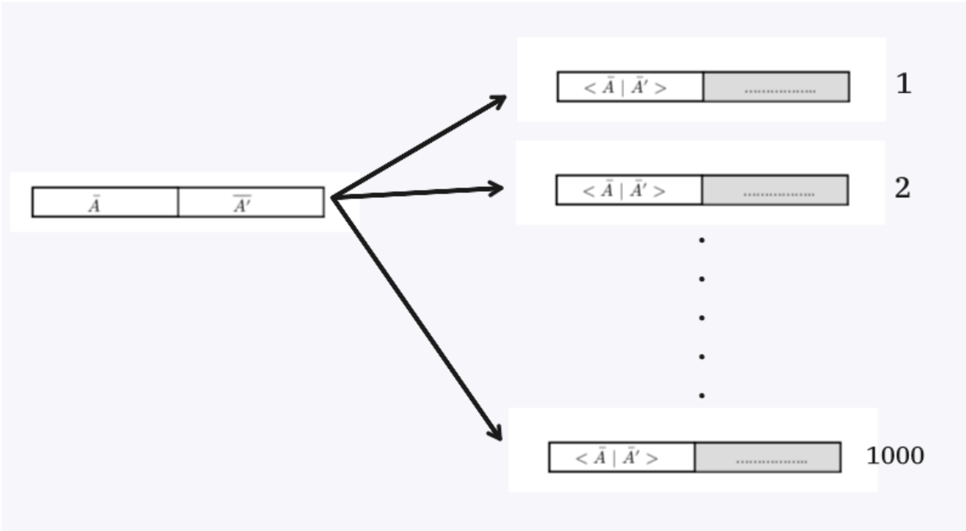
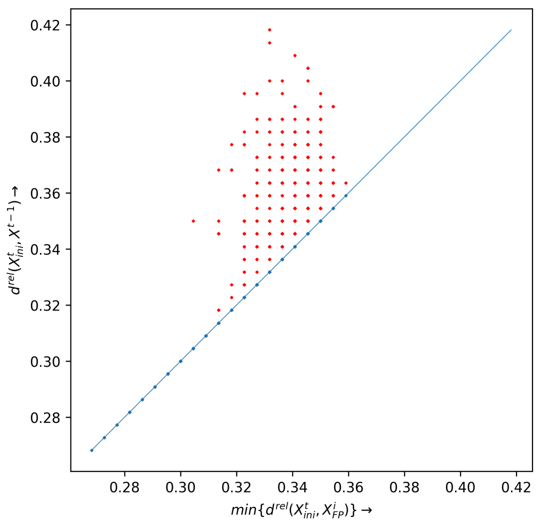


Figure 3: Generating multiple strings from the selected string with maximum possible $s = s_{\text{worst}}$, and minimum r. The grey part is randomly generated for each new constructed string

Now for each of the constructed strings, X^i_{constr} ($i \in \{1, .., 1000\}$), we now compute the minimum relative distance out of all the 2^k fixed points strings to it :$\min_j \left\{ d^{rel}\left(X^j_{FP}, X^i_{constr}\right)\right\}$. Additionally we also compute the distance between each of the constructed strings X^i_{constr} and the string used to construct them X^{t-1}: $d^{rel}(X^i_{constr}, X^{t-1})$. We plot the both of these distance for each of the constructed strings against each other in Figure 4.



The number of constructed strings which are closer to the one of the random fixed point strings than the string they were constructed from is 456 out of a 1000 constructed strings, which is denoted by $y > x$ region.