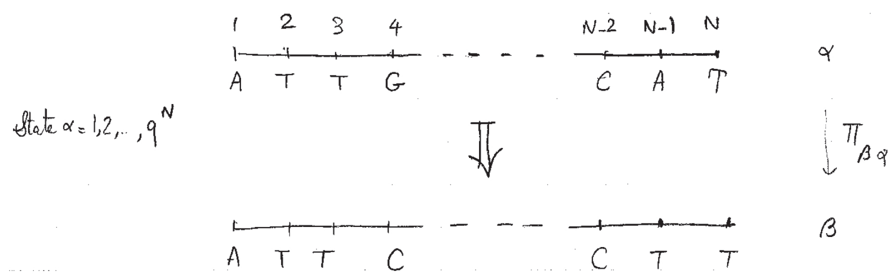
### 5.1.1 Evolving sequence

As organisms reproduce, the underlying genetic information is passed on to subsequent generations. The copying of the genetic content is not perfect, and leads to a diverse and evolving population of organisms after many generations. The changes are stochastic, and are thus appropriately described by evolving probability distributions. After motivating such evolving probabilities in the contexts of DNA and populations, we introduce the mathematical tools for treating them.

Let us consider the evolution of probabilities in the context of the simplified model of $N$ independently distributed sites. We model mutations by assuming that at subsequent time-steps (generations) each site may change its state (independent of the other sites), say from $\alpha$ to $\beta$ with a *transition probability* $\pi_{\beta\alpha}$. The $m \times m$ such elements form the *transition probability matrix* $\pi$. (Without the assumption that the sites evolve independently, we would have constructed a much larger ($m^N \times m^N$) matrix $\mathbf{\Pi}$. With the assumption of independence, this larger matrix is a direct product of transition matrices for individual sites, i.e. $\mathbf{\Pi} = \pi_1 \otimes \pi_2 \otimes \cdots \otimes \pi_N$, with $\pi_i$ a $m \times m$ matrix acting on site $i$.) With the transition probability matrix, we can track the evolution of the probabilities as

$$p_\alpha(\tau+1) = \sum_{\beta=1}^{m} \pi_{\alpha\beta} p_\beta(\tau), \quad \text{or in matrix form} \quad \vec{p}(\tau+1) = \pi\vec{p}(\tau) = \pi^\tau \vec{p}(1), \tag{5.1.1}$$

where the last identity is obtained by recursion, assuming that the transition probability matrix remains the same for all generations, i.e. does not change with time.




Probabilities must be normalized to unity, and thus the transition probabilities are constrained by

$$\sum_\alpha \pi_{\alpha\beta} = 1, \quad \text{or} \quad \pi_{\beta\beta} = 1 - \sum_{\alpha\neq\beta} \pi_{\alpha\beta}. \tag{5.1.2}$$

The last expression formalizes the statement that ithe probability to stay in the same state is the complement of the probabilities to make a change. Using this result, we can rewrite Eq. (5.1.1) as

$$p_\alpha(\tau+1) = p_\alpha(\tau) + \sum_{\beta\neq\alpha} \left[\pi_{\alpha\beta} p_\beta(\tau) - \pi_{\beta\alpha} p_\alpha(\tau)\right]. \tag{5.1.3}$$