

4 选择

选择是指根据个体适应度从种群中挑选个体的过程。其对应的是生物学中的”自然选择”。下面是一些选择操作有关的术语：

选择压力 (selective pressure): 与所有个体的平均选择概率相比，最优个体被选中的概率。

偏差 (bias): 个体标准化的适应度预期期望再生概率之差的绝对值。

个体扩展 (spread): 在进行选择时，比较优秀个体可能会被多次选择。该参数就限制了最多可以被重复选择的个数。

多样性损失 (loss of diversity): 进行选择操作后，未被选择的个体数目占种群个体总数的比例。

选择强度 (selection intensity): 把标准高斯分布用在选择操作后的种群个体平均适应度的期望值。选择强度越大，遗传算法的最优化搜索的收敛速度就越快，但也往往意味着多样性的损失。

选择方差 (selection variance): 把标准高斯分布用在选择操作后的种群个体适应度的方差。

选择操作一般是基于个体的适应度来计算的 (详见上一节)，下面介绍一些经典的选择算子：

1) 轮盘赌选择 (Roulette wheel selection):  
轮盘赌选择是一种有回放的随机抽样选择法。种群的个体被映射到区间的连续片段，每个个体所在片段的长度与其适应度成比例。生成随机数，根据其所落在的片段选择对应的个体，并重复该过程直到获得所需数量的个体。

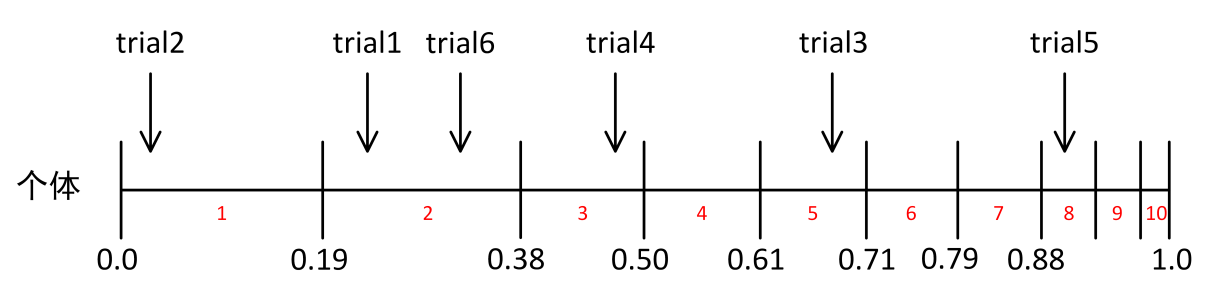


图 1 轮盘赌选择

所选择的个体为：2, 1, 5, 3, 8, 2。  
轮盘赌选择算法实现了零偏差 (bias) 但不保证最小个体扩展 (spread)，因此，上面的例子中，2 号个体被选择了两次 (有可能所选的个体全部都是同一个)。

2) 随机抽样选择 (Stochastic universal sampling):  
随机抽样选择实现了零偏差同时保证最小个体扩展。种群的个体被映射到区间的连续片段，每个个体所在片段的长度与其适应度成比例 (这里跟轮盘赌是相同的)。用一组等距离的指针随机地指向上述区间，每个指针所指的个体即为要选择的个体。指针间的距离是 $1/N_{sel}$  ( $N_{sel}$  为要选择的个体数)，第一个指针的位置由 $[0, 1/N_{sel})$  范围内的随机数给出。

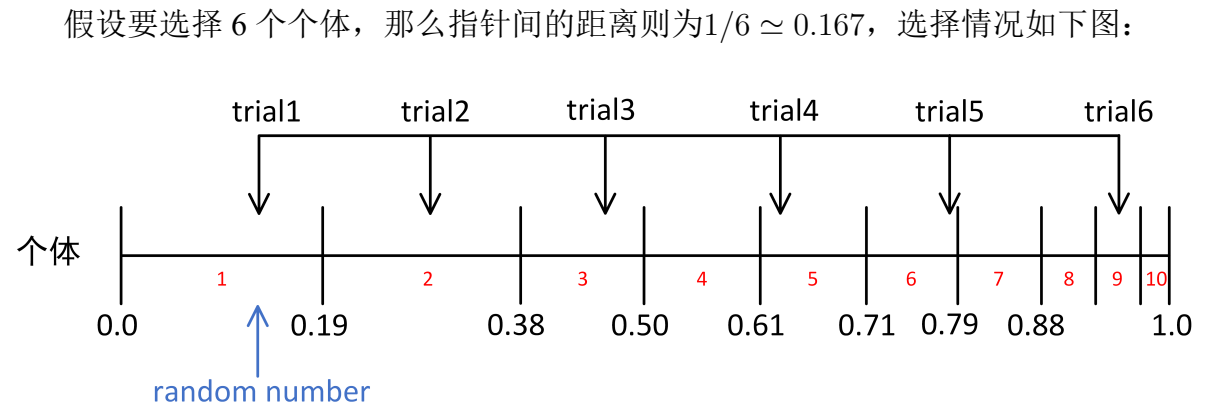


图 2 随机抽样选择

所选择的个体为：1, 2, 3, 5, 6, 9。  
轮盘赌选择和随机抽样选择都属于轮盘赌模型 (或称基于适应度比例的选择模型)，但后者通常能够得到更想要得到的结果，种群多样性也得以保证。  
某些人认为随机抽样选择是不考虑个体的适应度、纯粹依靠随机数的结果来进行个体选择，这种想法是极其错误的。这种选择方法实际上会让遗传算法变为盲目随机搜索，而无法发挥适应度应该起到的指导搜索方向的作用。

3) 锦标赛选择 (Tournament selection):  
该选择算法模仿淘汰赛制，每次从种群中挑选 $T_{our}$  个个体进行比较，从中选出一个最好的个体加入被选集合。重复该操作，直到被选集合的大小达到需要选择的个体数。  
 $T_{our}$  即为竞赛规模。在 [BT95] 中，可以看到锦标赛选择的相关理论分析。  
选择强度：

$$SelInt(T_{our}) \approx \sqrt{2 \left( \ln(T_{our}) - \ln \left( \sqrt{4.14 \cdot \ln(T_{our})} \right) \right)}$$

多样性损失：

$$LossDiv(T_{our}) = T_{our}^{-\frac{1}{T_{our}-1}} - T_{our}^{-\frac{T_{our}}{T_{our}-1}}$$

选择方差：

$$SelVar(T_{our}) \approx \frac{0.918}{\ln(1.186 + 1.328 \cdot T_{our})}$$

4) 截断选择 (Truncation selection):  
截断选择是与前面的几种选择算法比较不一样的算法。它更接近于生物学里的”人工育种”，适用于大规模种群的个体选择。  
在截断选择中，根据适应度对种群个体进行分类。设定截断阈值为 **Trunc**(表示选择作为父母的个体的比例)，低于阈值的个体不会产生后代。”选择强度”这个术语就经常用在截断选择中，”选择强度”和截断阈值之间的关系非常密切。  
在 [BT95] 中可以看到截断选择的相关理论分析。

5) 本地选择 (Local selection):  
在本地选择算法中，每个个体处在一个约束环境中 (称为”本地邻域”)。个体仅与同一邻域内的个体相互作用。邻域是根据种群的结构来定义的，可以是线性的、二维的或者是三位及更加复杂的。邻域之间的距离决定了邻域的大小，而邻域的大小决定了种群个体之间遗传信息的传播速度，即决定了种群是快速繁殖还是维持高度的种群多样性。[VBS91] 的模拟中得出了类似的结果，在较小的邻域中进行的本地选择比在较大的邻域中要好。  
选择算法依赖于适应度的计算，不同的适应度计算方法会使得种群个体的适应度呈现不一样的分布特征，从而影响到选择操作的效果。