Algorithm 1

Algorithm 1 Co-evolutionary Process

Require: Population size $\lambda \in \mathbb{N}$ and strategy spaces $\mathcal{X}$ and $\mathcal{Y}$.
Require: Initial populations $P_0 \in \mathcal{X}^\lambda$ and $Q_0 \in \mathcal{Y}^\lambda$.

```
for each generation number t ∈ ℕ_0 do
    for each interaction number i ∈ [λ] do
        Sample an interaction (x, y) ~ D(P_t, Q_t).
        Set P_{t+1}(i) := x and Q_{t+1}(i) := y.
    end for
end for
```

Theorem 1

Theorem 3. *Given subsets* $A_j \subseteq \mathcal{X}$, $B_j \subseteq \mathcal{Y}$ *for* $j \in [m]$ *where* $A_1 := \mathcal{X}$ *and* $B_1 := \mathcal{Y}$, *define* $T := \min\{t\lambda \mid (P_t \times Q_t) \cap (A_m \times B_m) \neq \emptyset\}$, *where for all* $t \in \mathbb{N}$, $P_t \in \mathcal{X}^\lambda$ *and* $Q_t \in \mathcal{Y}^\lambda$ *are the populations of Algorithm* $\boxed{1}$ *in generation* t. *If there exist* $z_1, \ldots, z_{m-1}, \delta \in (0, 1]$, *and* $\gamma_0 \in (0, 1)$ *such that for any populations* $P \in \mathcal{X}^\lambda$ *and* $Q \in \mathcal{Y}^\lambda$ *with "current level"* $j := \max\{i \in [m-1] \mid |(P \times Q) \cap (A_i \times B_i)| \geq \gamma_0 \lambda^2\}$

(G1) *for* $(x, y) \sim \mathcal{D}(P, Q)$

$$\Pr(x \in A_{j+1}) \Pr(y \in B_{j+1}) \geq z_j,$$

(G2a) *for all* $\gamma \in (0, \gamma_0)$ *if* $|(P \times Q) \cap (A_{j+1} \times B_{j+1})| \geq \gamma\lambda^2$, *then for* $(x, y) \sim \mathcal{D}(P, Q)$,

$$\Pr(x \in A_{j+1}) \Pr(y \in B_{j+1}) \geq (1 + \delta)\gamma,$$

(G2b) *for* $(x, y) \sim \mathcal{D}(P, Q)$,

$$\Pr(x \in A_j) \Pr(y \in B_j) \geq (1 + \delta)\gamma_0,$$

(G3) *and the population size* $\lambda \in \mathbb{N}$ *satisfies for a sufficiently large constant* c', *where* $z_* := \min_{i \in [m-1]} z_i$,

$$\lambda \geq c' \log(m/z_*),$$

then for a constant $c'' > 0$, $\mathbb{E}[T] \leq c''\lambda \left(\lambda^2 m + \sum_{i=1}^{m-1} 1/z_i\right)$.

Theorem 1

Theorem 3. *Given subsets* $A_j \subseteq \mathcal{X}$, $B_j \subseteq \mathcal{Y}$ *for* $j \in [m]$ *where* $A_1 := \mathcal{X}$ *and* $B_1 := \mathcal{Y}$, *define* $T := \min\{t\lambda \mid (P_t \times Q_t) \cap (A_m \times B_m) \neq \emptyset\}$, *where for all* $t \in \mathbb{N}$, $P_t \in \mathcal{X}^\lambda$ *and* $Q_t \in \mathcal{Y}^\lambda$ *are the populations of Algorithm* $\boxed{1}$ *in generation* t. *If there exist* $z_1, \ldots, z_{m-1}, \delta \in (0,1]$, *and* $\gamma_0 \in (0,1)$ *such that for any populations* $P \in \mathcal{X}^\lambda$ *and* $Q \in \mathcal{Y}^\lambda$ *with "current level"* $j := \max\{i \in [m-1] \mid |(P \times Q) \cap (A_i \times B_i)| \geq \gamma_0 \lambda^2\}$

(G1) *for* $(x, y) \sim \mathcal{D}(P, Q)$

$$\Pr(x \in A_{j+1}) \Pr(y \in B_{j+1}) \geq z_j,$$

(G2a) *for all* $\gamma \in (0, \gamma_0)$ *if* $|(P \times Q) \cap (A_{j+1} \times B_{j+1})| \geq \gamma\lambda^2$, *then for* $(x, y) \sim \mathcal{D}(P, Q)$,

$$\Pr(x \in A_{j+1}) \Pr(y \in B_{j+1}) \geq (1+\delta)\gamma,$$

(G2b) *for* $(x, y) \sim \mathcal{D}(P, Q)$,

$$\Pr(x \in A_j) \Pr(y \in B_j) \geq (1+\delta)\gamma_0,$$

(G3) *and the population size* $\lambda \in \mathbb{N}$ *satisfies for a sufficiently large constant* c', *where* $z_* := \min_{i \in [m-1]} z_i$,

$$\lambda \geq c' \log(m/z_*),$$

then for a constant $c'' > 0$, $\mathbb{E}[T] \leq c''\lambda\left(\lambda^2 m + \sum_{i=1}^{m-1} 1/z_i\right)$.

通过定义 level 来追踪目前算法的状态

G1 规定“进化”的概率 z

G2 规定“选择”强度 δ，保证不会“退化”

G3 规定需要的种群大小 λ

Proof 1

通过定义二元函数 g 估计算法的运行状态，并对函数 g 应用 additive drift theorem 得到算法运行时间

Definition 3 (Corus et al. (2018)). *A function* $g : (\{0\} \cup [\lambda^2]) \times [m] \rightarrow \mathbb{R}$ *is called a* level function *if the following three conditions hold*

1. $\forall x \in \{0\} \cup [\lambda^2], \forall y \in [m-1] : \quad g(x, y) \geq g(x, y+1),$
2. $\forall x \in \{0\} \cup [\lambda^2 - 1], \forall y \in [m] : \quad g(x, y) \geq g(x+1, y),$
3. $\forall y \in [m-1] : \quad g(\lambda^2, y) \geq g(0, y+1).$

Proof 1

Definition 3 (Corus et al. (2018)). *A function* $g : (\{0\} \cup [\lambda^2]) \times [m] \to \mathbb{R}$ *is called a* level function *if the following three conditions hold*

1. $\forall x \in \{0\} \cup [\lambda^2], \forall y \in [m-1]: \quad g(x, y) \geq g(x, y+1),$
2. $\forall x \in \{0\} \cup [\lambda^2 - 1], \forall y \in [m]: \quad g(x, y) \geq g(x+1, y),$
3. $\forall y \in [m-1]: \quad g(\lambda^2, y) \geq g(0, y+1).$

定义 Y_t 表示 t 次迭代后 population 的 level
定义 X_t^a 表示 t 次迭代后 population 与 a-level 集合交集的大小

Additive drift 研究变量 $Z_t := g\left(X_t^{(Y_t+1)}, Y_t\right)$ 的变化情况

Proof 1

Definition 3 (Corus et al. (2018)). *A function* $g : (\{0\} \cup [\lambda^2]) \times [m] \to \mathbb{R}$ *is called a* level function *if the following three conditions hold*

1. $\forall x \in \{0\} \cup [\lambda^2], \forall y \in [m-1]: \quad g(x,y) \geq g(x,y+1),$
2. $\forall x \in \{0\} \cup [\lambda^2 - 1], \forall y \in [m]: \quad g(x,y) \geq g(x+1,y),$
3. $\forall y \in [m-1]: \quad g(\lambda^2, y) \geq g(0, y+1).$

Additive drift 研究变量 $Z_t := g\left(X_t^{(Y_t+1)}, Y_t\right)$ 的变化情况

令 $\frac{1}{q_i} = 1 + \frac{4}{\lambda z_i}$ ，有性质 $q_i \leq \Pr_t\left(X_{t+1}^{(Y_t+1)} \geq 1\right)$

$$g_1(k,j) := \frac{\eta}{1+\eta} \cdot ((m-j)\lambda^2 - k)$$

$$g_2(k,j) := \varphi \cdot \left(\frac{e^{-\eta k}}{q_j} + \sum_{i=j+1}^{m-1} \frac{1}{q_i}\right),$$

$$g(k,j) := g_1(k,j) + g_2(k,j)$$

Proof 1

考虑计算 $\Delta_{t+1} := g\left(X_t^{(Y_t+1)}, Y_t\right) - g\left(X_{t+1}^{(Y_{t+1}+1)}, Y_{t+1}\right)$.的期望

case 1: level 不变或变大 $(1-\Pr_t(Y_{t+1} < Y_t))\mathbb{E}_t[\Delta_{t+1} \mid Y_{t+1} \geq Y_t]$

case 2: level 变低 $\Pr_t(Y_{t+1} < Y_t)\mathbb{E}_t[\Delta_{t+1} \mid Y_{t+1} < Y_t]$

$$Z_t := g\left(X_t^{(Y_t+1)}, Y_t\right)$$

$$\frac{1}{q_i} = 1 + \frac{4}{\lambda z_i}$$

$$g_1(k,j) := \frac{\eta}{1+\eta} \cdot ((m-j)\lambda^2 - k)$$

$$g_2(k,j) := \varphi \cdot \left(\frac{e^{-\eta k}}{q_j} + \sum_{i=j+1}^{m-1} \frac{1}{q_i}\right),$$

$$g(k,j) := g_1(k,j) + g_2(k,j)$$

计算 level 降低的概率 $\Pr_t(Y_{t+1} < Y_t) = \Pr_t\left(X_{t+1}^{(Y_t)} < \gamma_0\lambda^2\right)$

基于假设 G2b *for* $(x,y) \sim \mathcal{D}(P,Q)$,

$$\Pr(x \in A_j)\Pr(y \in B_j) \geq (1+\delta)\gamma_0,$$

有 $E\left[X_{t+1}^{(Y_t)}\right] \geq (1+\delta)\gamma_0\lambda(\lambda\text{-}1)$

根据 Markov 不等式 $\Pr_t(Y_{t+1} < Y_t) = \Pr_t\left(X_{t+1}^{(Y_t)} < \gamma_0\lambda^2\right) < e^{-c\lambda} < \frac{48}{(c\lambda)^3} \cdot \frac{z_*}{4m}$

Proof 1

考虑计算 $\Delta_{t+1} := g\left(X_t^{(Y_t+1)}, Y_t\right) - g\left(X_{t+1}^{(Y_{t+1}+1)}, Y_{t+1}\right)$.的期望

case 1: level 不变或变大 $(1-\Pr_t(Y_{t+1} < Y_t))\mathbb{E}_t[\Delta_{t+1} \mid Y_{t+1} \geq Y_t]$

$$\mathbb{E}_t[\Delta_{t+1} \mid Y_{t+1} \geq Y_t] \\ \geq \mathbb{E}_t\left[g\left(X_t^{(Y_t+1)}, Y_t\right) - g\left(X_{t+1}^{(Y_t+1)}, Y_t\right) \mid Y_{t+1} \geq Y_t\right].$$

case 1.1: $X_t^{(Y_t+1)} = \gamma\lambda^2 \geq 1$;

基于假设 G2a *for all* $\gamma \in (0,\gamma_0)$ *if* $|(P \times Q) \cap (A_{j+1} \times B_{j+1})| \geq \gamma\lambda^2$, *then for* $(x,y) \sim \mathcal{D}(P,Q)$,

$$\Pr(x \in A_{j+1}) \Pr(y \in B_{j+1}) \geq (1+\delta)\gamma,$$

$$\mathbb{E}_t[\Delta_{t+1} \mid Y_{t+1} \geq Y_t] \\ \geq \frac{\eta}{1+\eta}(\lambda(\lambda-1)(1+\delta)\gamma - \gamma\lambda^2) > \frac{\eta}{1+\eta}\delta(1-\delta') = \frac{\eta\varphi}{1+\eta},$$

$$Z_t := g\left(X_t^{(Y_t+1)}, Y_t\right)$$

$$\frac{1}{q_i} = 1 + \frac{4}{\lambda z_i}$$

$$g_1(k,j) := \frac{\eta}{1+\eta} \cdot ((m-j)\lambda^2 - k)$$

$$g_2(k,j) := \varphi \cdot \left(\frac{e^{-\eta k}}{q_j} + \sum_{i=j+1}^{m-1} \frac{1}{q_i}\right),$$

$$g(k,j) := g_1(k,j) + g_2(k,j)$$

Proof 1

考虑计算 $\Delta_{t+1} := g\left(X_t^{(Y_t+1)}, Y_t\right) - g\left(X_{t+1}^{(Y_{t+1}+1)}, Y_{t+1}\right)$.的期望

case 1: level 不变或变大 $(1 - \Pr_t(Y_{t+1} < Y_t))\mathbb{E}_t[\Delta_{t+1} \mid Y_{t+1} \geq Y_t]$

$$\begin{aligned}&\mathbb{E}_t[\Delta_{t+1} \mid Y_{t+1} \geq Y_t] \\ &\quad \geq \mathbb{E}_t\left[g\left(X_t^{(Y_t+1)}, Y_t\right) - g\left(X_{t+1}^{(Y_t+1)}, Y_t\right) \mid Y_{t+1} \geq Y_t\right].\end{aligned}$$

case 1.2: $X_t^{(Y_t+1)} = 0,$

根据 $g_2(0, Y_t) - g_2(1, Y_t) = (\varphi/q_{Y_t})(1 - e^{-\eta}) \geq \dfrac{\varphi\eta}{(1+\eta)q_{Y_t}}$

$$\begin{aligned}&\mathbb{E}_t\left[g_2\left(X_t^{(Y_t+1)}, Y_t\right) - g_2\left(X_{t+1}^{(Y_t+1)}, Y_t\right) \mid Y_{t+1} \geq Y_t\right] \\ &\qquad > \Pr_t\left(X_{t+1}^{(Y_t+1)} \geq 1\right)\left(g_2(0, Y_t) - g_2(1, Y_t)\right) \geq \frac{\eta\varphi}{1+\eta}.\end{aligned}$$

$$Z_t := g\left(X_t^{(Y_t+1)}, Y_t\right)$$

$$\frac{1}{q_i} = 1 + \frac{4}{\lambda z_i}$$

$$g_1(k, j) := \frac{\eta}{1+\eta} \cdot \left((m-j)\lambda^2 - k\right)$$

$$g_2(k, j) := \varphi \cdot \left(\frac{e^{-\eta k}}{q_j} + \sum_{i=j+1}^{m-1} \frac{1}{q_i}\right),$$

$$g(k, j) := g_1(k, j) + g_2(k, j)$$

Proof 1

考虑计算 $\Delta_{t+1} := g\left(X_t^{(Y_t+1)}, Y_t\right) - g\left(X_{t+1}^{(Y_{t+1}+1)}, Y_{t+1}\right)$.的期望

case 2: level 变低 $\Pr_t\left(Y_{t+1} < Y_t\right) \mathbb{E}_t\left[\Delta_{t+1} \mid Y_{t+1} < Y_t\right]$

$$\Pr_t\left(Y_{t+1} < Y_t\right) = \Pr_t\left(X_{t+1}^{(Y_t)} < \gamma_0 \lambda^2\right) < e^{-c\lambda} < \frac{48}{(c\lambda)^3} \cdot \frac{z_*}{4m}$$

$$\mathbb{E}_t\left[\Delta_{t+1} \mid Y_{t+1} < Y_t\right] \geq -g(0,1)$$

综合考虑 case 1,2

$$\begin{aligned}
\mathbb{E}_t\left[\Delta_{t+1}\right] &= \left(1 - \Pr_t\left(Y_{t+1} < Y_t\right)\right) \mathbb{E}_t\left[\Delta_{t+1} \mid Y_{t+1} \geq Y_t\right] \\
&\quad + \Pr_t\left(Y_{t+1} < Y_t\right) \mathbb{E}_t\left[\Delta_{t+1} \mid Y_{t+1} < Y_t\right] \\
&\geq \frac{\eta\varphi}{1+\eta} - \frac{48}{(c\lambda)^3} \frac{z_*}{4m}\left(m2\eta\lambda^2 + \frac{4m\varphi}{\lambda z_*} + \frac{\eta\varphi}{1+\eta}\right) \\
&> \frac{\eta\varphi(1-\delta'')}{1+\eta}
\end{aligned}$$

$$Z_t := g\left(X_t^{(Y_t+1)}, Y_t\right)$$

$$\frac{1}{q_i} = 1 + \frac{4}{\lambda z_i}$$

$$g_1(k,j) := \frac{\eta}{1+\eta} \cdot \left((m-j)\lambda^2 - k\right)$$

$$g_2(k,j) := \varphi \cdot \left(\frac{e^{-\eta k}}{q_j} + \sum_{i=j+1}^{m-1} \frac{1}{q_i}\right),$$

$$g(k,j) := g_1(k,j) + g_2(k,j)$$

Proof 1

考虑计算 $\Delta_{t+1} := g\left(X_t^{(Y_t+1)}, Y_t\right) - g\left(X_{t+1}^{(Y_{t+1}+1)}, Y_{t+1}\right)$.的期望

$$
\begin{aligned}
\mathbb{E}_t\left[\Delta_{t+1}\right] &= (1-\Pr{}_t\left(Y_{t+1}<Y_t\right))\mathbb{E}_t\left[\Delta_{t+1} \mid Y_{t+1} \geq Y_t\right] \\
&\quad + \Pr{}_t\left(Y_{t+1}<Y_t\right) \mathbb{E}_t\left[\Delta_{t+1} \mid Y_{t+1}<Y_t\right] \\
&\geq \frac{\eta\varphi}{1+\eta} - \frac{48}{(c\lambda)^3}\frac{z_*}{4m}\left(m2\eta\lambda^2 + \frac{4m\varphi}{\lambda z_*} + \frac{\eta\varphi}{1+\eta}\right) \\
&> \frac{\eta\varphi(1-\delta'')}{1+\eta}
\end{aligned}
$$

运用 Additive drift 分析

$$
\begin{aligned}
\mathbb{E}[T] &\leq \lambda \cdot \frac{(1+\eta)g(0,1)}{\eta\varphi(1-\delta'')} \\
&\leq \frac{\lambda}{(1-\delta'')}\left(\frac{2\lambda^2 m}{\delta(1-\delta')} + \frac{4}{1-(1+\delta)^{-1/2}}\sum_{i=1}^{m-1}\frac{1}{z_i}\right) \\
&\leq c''\lambda\left(\lambda^2 m + \sum_{i=1}^{m-1}\frac{1}{z_i}\right),
\end{aligned}
$$

$$Z_t := g\left(X_t^{(Y_t+1)}, Y_t\right)$$

$$\frac{1}{q_i} = 1 + \frac{4}{\lambda\, z_i}$$

$$g_1(k,j) := \frac{\eta}{1+\eta} \cdot ((m-j)\lambda^2 - k)$$

$$g_2(k,j) := \varphi \cdot \left(\frac{e^{-\eta k}}{q_j} + \sum_{i=j+1}^{m-1}\frac{1}{q_i}\right),$$

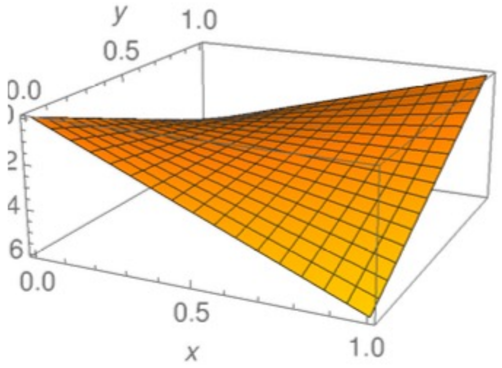
$$g(k,j) := g_1(k,j) + g_2(k,j)$$

Problem 1

Maximin Optimization Problems

$$f(x) := \min_{y \in \mathcal{Y}} g(x, y).$$

求函数 f(x) 的最大值点



Bilinear 函数图像

The Bilinear Problem

$$\textsc{Bilinear}(x, y) := |y|(|x| - \beta n) - \alpha n|x|,$$

对于特殊的 Bilinear 函数，求其最大值点

Problem 1

Maximin Optimization Problems

$$f(x) := \min_{y \in \mathcal{Y}} g(x, y).$$

求函数 f(x) 的最大值点

对于函数 g(x,y) 定义支配关系 $\succeq_g$ ， 如果有 $(x_1, y_1) \succeq_g (x_2, y_2)$ 则满足：

$$g(x_1, y_2) \geq g(x_1, y_1) \geq g(x_2, y_1).$$

当 x 固定时 g(x, y) 越小，(x, y) 在该关系上越大
当 y 固定时 g(x, y) 越大，(x, y) 在该关系上越大

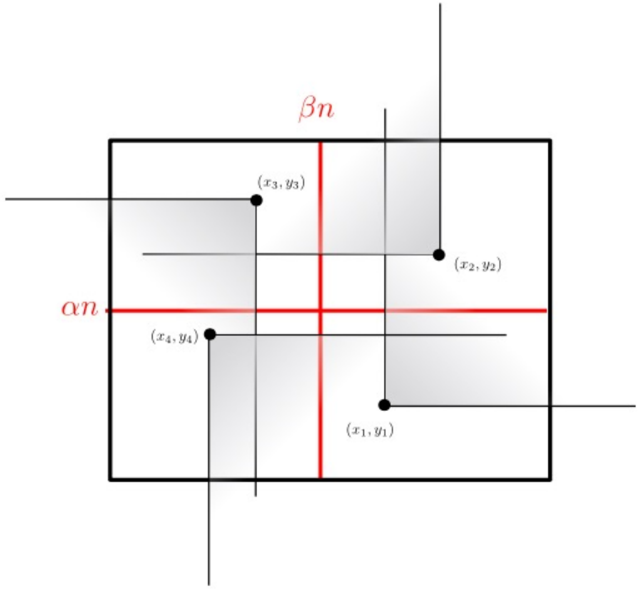
在该关系下最大的点决定了 f(x) 的最大值点

Problem 1

The Bilinear Problem

$$\textsc{Bilinear}(x, y) := |y|(|x| - \beta n) - \alpha n|x|,$$

对于特殊的 Bilinear 函数，求其最大值点



Bilinear 函数支配关系

Lemma 4. *Let* $g := \textsc{Bilinear}$. *For all pairs* $(x_1, y_1), (x_2, y_2) \in \mathcal{X} \times \mathcal{Y}$, $(x_1, y_1) \succeq_g (x_2, y_2)$ *if and only if*

$$\begin{aligned} |y_2|(|x_1| - \beta n) &\geq |y_1|(|x_1| - \beta n) \quad \wedge \\ |x_1|(|y_1| - \alpha n) &\geq |x_2|(|y_1| - \alpha n). \end{aligned}$$

Algorithm 2

Algorithm 2 Pairwise Dominance CoEA (PD-CoEA)

Require: Min-max-objective function $g : \{0,1\}^n \times \{0,1\}^n \to \mathbb{R}$.
Require: Population size $\lambda \in \mathbb{N}$ and mutation rate $\chi \in (0, n]$

for $i \in [\lambda]$ **do**
  Sample $P_0(i) \sim \mathrm{Unif}(\{0,1\}^n)$
  Sample $Q_0(i) \sim \mathrm{Unif}(\{0,1\}^n)$
end for
for $t \in \mathbb{N}$ until termination criterion met **do**
  for $i \in [\lambda]$ **do**
    Sample $(x_1, y_1) \sim \mathrm{Unif}(P_t \times Q_t)$
    Sample $(x_2, y_2) \sim \mathrm{Unif}(P_t \times Q_t)$
    if $(x_1, y_1) \succeq_g (x_2, y_2)$ **then**
      $(x, y) := (x_1, y_1)$
    else
      $(x, y) := (x_2, y_2)$
    end if
    Obtain x' by flipping each bit in x with prob. χ/n.
    Obtain y' by flipping each bit in y with prob. χ/n.
    Set $P_{t+1}(i) := x'$ and $Q_{t+1}(i) := y'$.
  end for
end for

Definition 2

对搜索空间进行划分，定义 $S_0, S_1(k), S_2(k)$

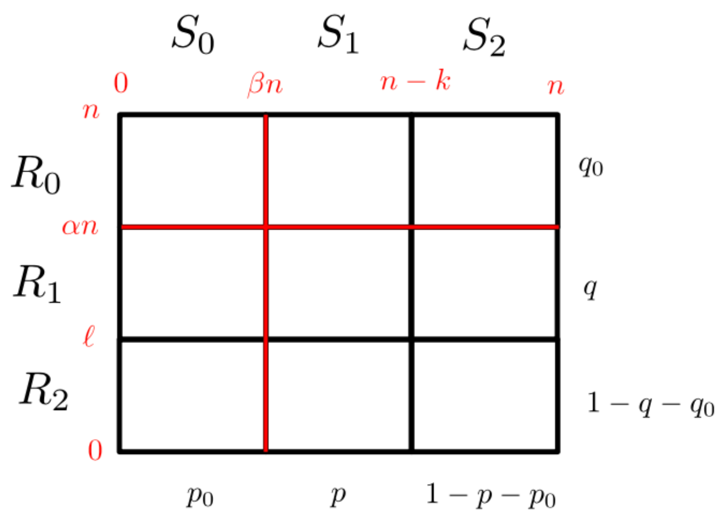
$$S_0 := \{x \in \mathcal{X} \mid 0 \le |x| < \beta n\}$$
$$S_1(k) := \{x \in \mathcal{X} \mid \beta n \le |x| < n-k\},$$
$$S_2(k) := \{x \in \mathcal{X} \mid n-k \le |x| \le n\}.$$

定义 $R_0, R_1(k), R_2(k)$

$$R_0 := \{y \in \mathcal{Y} \mid \alpha n \le |y| \le n\}$$
$$R_1(\ell) := \{y \in \mathcal{Y} \mid \ell \le |y| < \alpha n\}$$
$$R_2(\ell) := \{y \in \mathcal{Y} \mid 0 \le |y| < \ell\}.$$

同时定义概率 $p(C)$ $p_{sel}(C)$，其中 $C \subseteq \mathcal{X} \times \mathcal{Y}$,

$$p(C) := \Pr_{(x,y)\sim \mathrm{Unif}(P\times Q)}((x,y) \in C)$$
$$p_{\mathrm{sel}}(C) := \Pr_{(x,y)\sim \mathtt{select}(P\times Q)}((x,y) \in C).$$

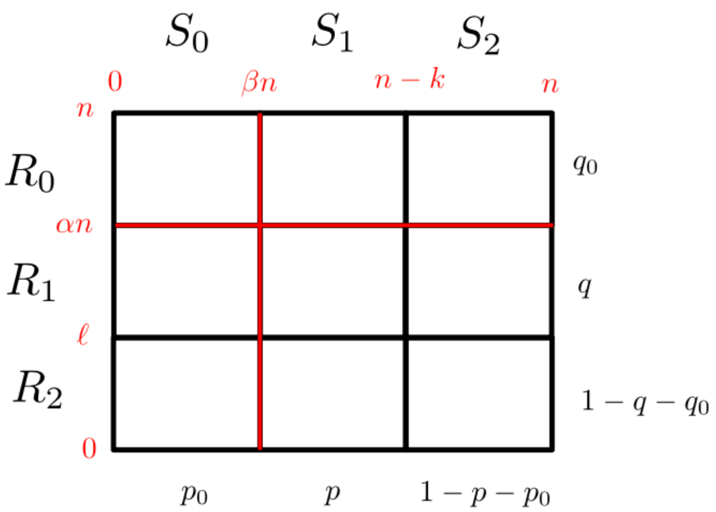


搜索空间的划分

Theorem 2

Theorem 14. *Assume that for a sufficiently large constant c, it holds* $c\log(n) \le \lambda \in \mathrm{poly}(n)$. *Let* $\alpha, \beta, \varepsilon \in (0,1)$ *be three constants where* $\alpha - \varepsilon \ge 4/5$. *Define* $T := \min\{\lambda t \mid (P_t \times Q_t) \cap S_0 \times R_1((\alpha-\varepsilon))n\}$ *where* P_t *and* Q_t *are the populations of Algorithm* $\boxed{2}$ *applied to* BILINEAR $_{\alpha,\beta}$. *Then if the mutation rate* χ *is a sufficiently small constant, and at most* $1/(1-\alpha+\varepsilon)$, *there exists a constant* c_0 *such that for all* $r \in \mathrm{poly}(n)$, *it holds* $\Pr\left(T > c_0 r \lambda^3 n\right) \le (1/r)(1+o(1))$.

1. 规定了 Bilinear Problem 中 α 的范围和算法中突变概率 χ 的范围
2. 应用框架的结论得出期望时间和时间的概率分布



搜索空间的划分

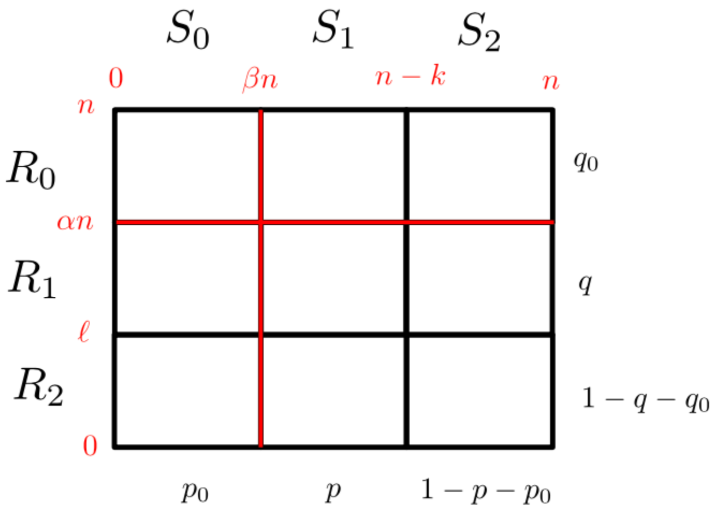
Proof 2

定义 level 集合：

$$(A_0^{(1)} \times B_0^{(1)}), \dots, (A_{(1-\beta)n}^{(1)}, B_{(1-\beta)n}^{(1)}),$$
$$(A_0^{(2)} \times B_0^{(2)}), \dots, (A_{(\alpha-\varepsilon)n}^{(2)}, B_{(\alpha-\varepsilon)n}^{(2)}),$$

Phase 1 for $j \in [0..(1-\beta)n]$ as $A_j^{(1)} := S_0 \cup S_1(j)$ and $B_j^{(1)} := R_2((\alpha-\varepsilon)n)$.

Phase 2 for $j \in [0, (\alpha-\varepsilon)n]$ $A_j^{(2)} := S_0$ and $B_j^{(2)} := R_1(j)$.



搜索空间的划分

Proof 2

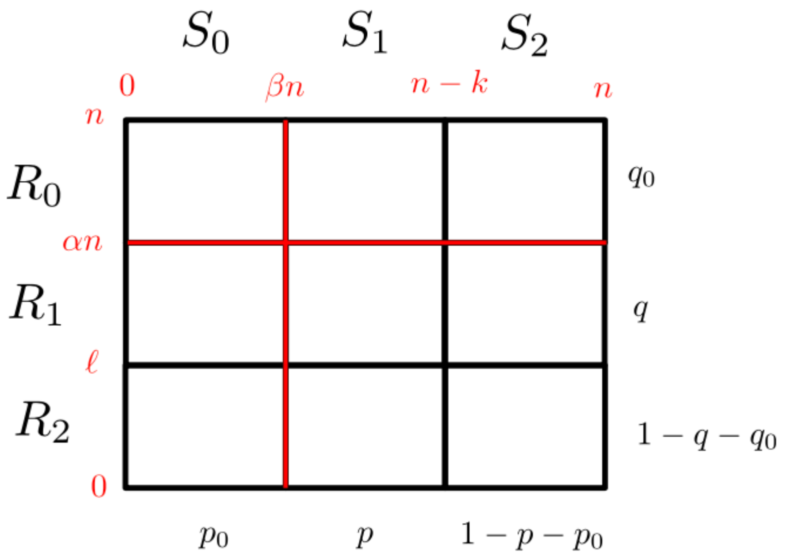
为了便于分析，假设在运行过程中 $p(R_0) = 0$

case 1: 初始解落在 R0 的情况：使用 Chernoff bound 可以计算出概率小于 $e^{-\Omega(n)}$.

case 2: 从 R1∪R2 变异到 R0 的情况：使用 Negative Drift Theorem 可以计算出概率小于 $e^{-\Omega(\lambda)}$

总体上满足 $\Pr\left(\bigvee_{t=0}^{\tau_*}(Q_t \cap R_0) \neq \emptyset\right) \leq \tau e^{-\Omega(n)} + \tau e^{-\Omega(\lambda)}.$

其中 $\tau \leq e^{cn}$ 且 $\tau_* := \min\{T/\lambda - 1, \tau\},$



搜索空间的划分

Proof 2

在 $p(R_0)=0$ 的条件下进行分析，分析 CoEA 算法满足 Theorem 3 的条件 G2a

只证明 phase 1，phase 2 证明类似

定义选择比率 $\varphi := \frac{p_{sel}(S_0 \cup S_1)}{p(S_0 \cup S_1)} \cdot \frac{p_{sel}(R_2)}{p(R_2)}$

根据算法的性质，存在 $\varphi \geq (1+\delta(1-p-p_0))(1+\delta(q+q_0)) \geq 1.$

根据 level 的定义，存在 $(p_0+p)(1-q-q_0) < \gamma_0$

考虑选择到个体 $(x,y) \in (A_{j+1} \times B_{j+1})$ 并且未发生变异的情况

$$
\begin{aligned}
&\Pr\left(x \in A_{j+1}^{(1)}\right)\Pr\left(y \in B_{j+1}^{(1)}\right) \\
&\geq p_{\mathrm{sel}}(A_{j+1}^{(1)})p_{\mathrm{sel}}(B_{j+1}^{(1)})\left(1-\frac{\chi}{n}\right)^{2n} \\
&\geq (1+\delta(1-\sqrt{\gamma_0}))p(A_{j+1}^{(1)})p(B_{j+1}^{(1)})e^{-2\chi}(1-o(1)) \\
&\geq (1+\delta'')\gamma,
\end{aligned}
$$

Theorem 3. *Given subsets $A_j \subseteq \mathcal{X}$, $B_j \subseteq \mathcal{Y}$ for $j \in [m]$ where $A_1 := \mathcal{X}$ and $B_1 := \mathcal{Y}$, define $T := \min\{t\lambda \mid (P_t \times Q_t) \cap (A_m \times B_m) \neq \emptyset\}$, where for all $t \in \mathbb{N}$, $P_t \in \mathcal{X}^\lambda$ and $Q_t \in \mathcal{Y}^\lambda$ are the populations of Algorithm 1 in generation t. If there exist $z_1, \ldots, z_{m-1}, \delta \in (0,1]$, and $\gamma_0 \in (0,1)$ such that for any populations $P \in \mathcal{X}^\lambda$ and $Q \in \mathcal{Y}^\lambda$ with "current level" $j := \max\{i \in [m-1] \mid |(P \times Q) \cap (A_i \times B_i)| \geq \gamma_0\lambda^2\}$*

(G1) *for $(x,y) \sim \mathcal{D}(P,Q)$*

$$\Pr(x \in A_{j+1})\Pr(y \in B_{j+1}) \geq z_j,$$

(G2a) *for all $\gamma \in (0,\gamma_0)$ if $|(P \times Q) \cap (A_{j+1} \times B_{j+1})| \geq \gamma\lambda^2$, then for $(x,y) \sim \mathcal{D}(P,Q)$,*

$$\Pr(x \in A_{j+1})\Pr(y \in B_{j+1}) \geq (1+\delta)\gamma,$$

(G2b) *for $(x,y) \sim \mathcal{D}(P,Q)$,*

$$\Pr(x \in A_j)\Pr(y \in B_j) \geq (1+\delta)\gamma_0,$$

(G3) *and the population size $\lambda \in \mathbb{N}$ satisfies for a sufficiently large constant c', where $z_* := \min_{i \in [m-1]} z_i$,*

$$\lambda \geq c' \log(m/z_*),$$

then for a constant $c'' > 0$, $\mathbb{E}[T] \leq c''\lambda\left(\lambda^2 m + \sum_{i=1}^{m-1} 1/z_i\right)$.

Proof 2

在 $p(R_0) = 0$ 的条件下进行分析，分析 CoEA 算法满足 Theorem 3 的条件 G2b

只证明 phase 1，phase 2 证明类似

定义选择比率 $\varphi := \frac{p_{sel}(S_0 \cup S_1)}{p(S_0 \cup S_1)} \cdot \frac{p_{sel}(R_2)}{p(R_2)}$

根据算法的性质，存在 $\varphi \geq (1+\delta(1-p-p_0))(1+\delta(q+q_0)) \geq 1.$

根据 level 的定义，存在 $(p_0+p)(1-q-q_0) < \gamma_0$

考虑选择到个体 $(x, y) \in (A_{j+1} \times B_{j+1})$ 并且未发生变异的情况

$$\begin{aligned}
&\Pr\left(x \in A_{j+1}^{(1)}\right)\Pr\left(y \in B_{j+1}^{(1)}\right) \\
&\geq p_{\text{sel}}(A_{j+1}^{(1)})p_{\text{sel}}(B_{j+1}^{(1)})\left(1-\frac{\chi}{n}\right)^{2n} \\
&\geq (1+\delta(1-\sqrt{\gamma_0}))p(A_{j+1}^{(1)})p(B_{j+1}^{(1)})e^{-2\chi}(1-o(1)) \\
&\geq (1+\delta'')\gamma,
\end{aligned}$$

Theorem 3. *Given subsets $A_j \subseteq \mathcal{X}$, $B_j \subseteq \mathcal{Y}$ for $j \in [m]$ where $A_1 := \mathcal{X}$ and $B_1 := \mathcal{Y}$, define $T := \min\{t\lambda \mid (P_t \times Q_t) \cap (A_m \times B_m) \neq \emptyset\}$, where for all $t \in \mathbb{N}$, $P_t \in \mathcal{X}^\lambda$ and $Q_t \in \mathcal{Y}^\lambda$ are the populations of Algorithm 1 in generation t. If there exist $z_1, \ldots, z_{m-1}, \delta \in (0,1]$, and $\gamma_0 \in (0,1)$ such that for any populations $P \in \mathcal{X}^\lambda$ and $Q \in \mathcal{Y}^\lambda$ with "current level" $j := \max\{i \in [m-1] \mid |(P \times Q) \cap (A_i \times B_i)| \geq \gamma_0\lambda^2\}$*

(G1) *for $(x, y) \sim \mathcal{D}(P, Q)$*

$$\Pr(x \in A_{j+1})\Pr(y \in B_{j+1}) \geq z_j,$$

(G2a) *for all $\gamma \in (0, \gamma_0)$ if $|(P \times Q) \cap (A_{j+1} \times B_{j+1})| \geq \gamma\lambda^2$, then for $(x, y) \sim \mathcal{D}(P, Q)$,*

$$\Pr(x \in A_{j+1})\Pr(y \in B_{j+1}) \geq (1+\delta)\gamma,$$

(G2b) *for $(x, y) \sim \mathcal{D}(P, Q)$,*

$$\Pr(x \in A_j)\Pr(y \in B_j) \geq (1+\delta)\gamma_0,$$

(G3) *and the population size $\lambda \in \mathbb{N}$ satisfies for a sufficiently large constant c', where $z_* := \min_{i \in [m-1]} z_i$,*

$$\lambda \geq c' \log(m/z_*),$$

then for a constant $c'' > 0$, $\mathbb{E}[T] \leq c''\lambda\left(\lambda^2 m + \sum_{i=1}^{m-1} 1/z_i\right)$.

Proof 2

在 $p(R_0)=0$ 的条件下进行分析，分析 CoEA 算法满足 Theorem 3 的条件 G2a

只证明 phase 1，phase 2 证明类似

case 1: $\gamma_0 \le p(A_j^{(1)})p(B_j^{(1)}) \le \gamma_0(1+\delta)$

证明类似 G2b 的过程，将 γ_0 变为 $\gamma_0(1+\delta)$ 即可

case 2: $\gamma_0(1+\delta) \le p(A_j^{(1)})p(B_j^{(1)})$

$$\begin{aligned} p_{sel}\left(A_j^{(1)}\right) p_{sel}\left(B_j^{(1)}\right) &\ge \varphi p\left(A_j^{(1)}\right) p(B_j^{(1)}) \\ &\ge (1+\delta)\lambda_0 \end{aligned}$$

Theorem 3. *Given subsets $A_j \subseteq \mathcal{X}$, $B_j \subseteq \mathcal{Y}$ for $j \in [m]$ where $A_1 := \mathcal{X}$ and $B_1 := \mathcal{Y}$, define $T := \min\{t\lambda \mid (P_t \times Q_t) \cap (A_m \times B_m) \neq \emptyset\}$, where for all $t \in \mathbb{N}$, $P_t \in \mathcal{X}^\lambda$ and $Q_t \in \mathcal{Y}^\lambda$ are the populations of Algorithm 1 in generation t. If there exist $z_1, \dots, z_{m-1}, \delta \in (0,1]$, and $\gamma_0 \in (0,1)$ such that for any populations $P \in \mathcal{X}^\lambda$ and $Q \in \mathcal{Y}^\lambda$ with "current level" $j := \max\{i \in [m-1] \mid |(P \times Q) \cap (A_i \times B_i)| \ge \gamma_0\lambda^2\}$*

(G1) *for $(x,y) \sim \mathcal{D}(P,Q)$*

$$\Pr(x \in A_{j+1}) \Pr(y \in B_{j+1}) \ge z_j,$$

(G2a) *for all $\gamma \in (0,\gamma_0)$ if $|(P \times Q) \cap (A_{j+1} \times B_{j+1})| \ge \gamma\lambda^2$, then for $(x,y) \sim \mathcal{D}(P,Q)$,*

$$\Pr(x \in A_{j+1}) \Pr(y \in B_{j+1}) \ge (1+\delta)\gamma,$$

(G2b) *for $(x,y) \sim \mathcal{D}(P,Q)$,*

$$\Pr(x \in A_j) \Pr(y \in B_j) \ge (1+\delta)\gamma_0,$$

(G3) *and the population size $\lambda \in \mathbb{N}$ satisfies for a sufficiently large constant c', where $z_* := \min_{i \in [m-1]} z_i$,*

$$\lambda \ge c' \log(m/z_*),$$

then for a constant $c'' > 0$, $\mathbb{E}[T] \le c''\lambda \left(\lambda^2 m + \sum_{i=1}^{m-1} 1/z_i\right)$.

Proof 2

分析 CoEA 算法在 Theorem 3 条件 G1 中 z_j 的范围

只证明 phase 1，phase 2 证明类似

$$\Pr\left(x \in A_{j+1}^{(1)}\right)\Pr\left(y \in B_{j+1}^{(1)}\right)$$
$$\geq \mathrm{p}_{sel}\left(A_j^{(1)}\right)\mathrm{p}_{sel}\left(B_j^{(1)}\right)\left(1-\frac{\chi}{n}\right)^{2n-1}(n-j)\frac{\chi}{n}$$
$$\geq (1+\delta)\lambda_0\left(1-\frac{\chi}{n}\right)^{2n-1}(n-j)\frac{\chi}{n}$$
$$= \Omega(1)$$

Theorem 3. *Given subsets $A_j \subseteq \mathcal{X}$, $B_j \subseteq \mathcal{Y}$ for $j \in [m]$ where $A_1 := \mathcal{X}$ and $B_1 := \mathcal{Y}$, define $T := \min\{t\lambda \mid (P_t \times Q_t) \cap (A_m \times B_m) \neq \emptyset\}$, where for all $t \in \mathbb{N}$, $P_t \in \mathcal{X}^\lambda$ and $Q_t \in \mathcal{Y}^\lambda$ are the populations of Algorithm 1 in generation t. If there exist $z_1, \ldots, z_{m-1}, \delta \in (0,1]$, and $\gamma_0 \in (0,1)$ such that for any populations $P \in \mathcal{X}^\lambda$ and $Q \in \mathcal{Y}^\lambda$ with "current level" $j := \max\{i \in [m-1] \mid |(P \times Q) \cap (A_i \times B_i)| \geq \gamma_0\lambda^2\}$*

(G1) *for $(x,y) \sim \mathcal{D}(P,Q)$*

$$\Pr(x \in A_{j+1})\Pr(y \in B_{j+1}) \geq z_j,$$

(G2a) *for all $\gamma \in (0,\gamma_0)$ if $|(P \times Q) \cap (A_{j+1} \times B_{j+1})| \geq \gamma\lambda^2$, then for $(x,y) \sim \mathcal{D}(P,Q)$,*

$$\Pr(x \in A_{j+1})\Pr(y \in B_{j+1}) \geq (1+\delta)\gamma,$$

(G2b) *for $(x,y) \sim \mathcal{D}(P,Q)$,*

$$\Pr(x \in A_j)\Pr(y \in B_j) \geq (1+\delta)\gamma_0,$$

(G3) *and the population size $\lambda \in \mathbb{N}$ satisfies for a sufficiently large constant c', where $z_* := \min_{i \in [m-1]} z_i$,*

$$\lambda \geq c' \log(m/z_*),$$

then for a constant $c'' > 0$, $\mathbb{E}[T] \leq c''\lambda\left(\lambda^2 m + \sum_{i=1}^{m-1} 1/z_i\right)$.

Proof 2

定义 $H(y)$ 为汉明距离
定义 $R_t(i)$ 为 i 的繁殖率，第 t 代时个体 i 繁殖的后代个数
定义 $P_{mut}(x)$ 为变异函数 χ 为变异比率

Theorem 8 (Negative Drift Theorem for Populations Lehre (2010)). *Given Algorithm 3 on $\mathcal{Y} = \{0,1\}^n$ with population size $\lambda \in \text{poly}(n)$, and transition matrix p_{mut} corresponding to flipping each bit independently with probability χ/n. Let $a(n)$ and $b(n)$ be positive integers s.t. $b(n) \leq n/\chi$ and $d(n) := b(n) - a(n) = \omega(\ln n)$. For an $x^* \in \{0,1\}^n$, let $T(n)$ be the smallest $t \geq 0$, s.t. $\min_{j\in[\lambda]} H(P_t(j), x^*) \leq a(n)$. Let $R_t(i) := \sum_{j=1}^{\lambda} [I_t(j) = i]$. If there are constants $\alpha_0 \geq 1$ and $\delta > 0$ such that*

1) $\mathbb{E}[R_t(i) \mid a(n) < H(P_t(i), x^*) < b(n)] \leq \alpha_0$ *for all* $i \in [\lambda]$

2) $\psi := \ln(\alpha_0)/\chi + \delta < 1$, *and*

3) $\frac{b(n)}{n} < \min\left\{\frac{1}{5}, \frac{1}{2} - \frac{1}{2}\sqrt{\psi(2-\psi)}\right\}$,

then $\Pr\left(T(n) \leq e^{cd(n)}\right) \leq e^{-\Omega(d(n))}$ *for some constant* $c > 0$.

Algorithm 3 Population Selection-Variation Algorithm Lehre (2010)

Require: Finite state space $\mathcal{Y}$.
Require: Transition matrix p_{mut} over $\mathcal{Y}$.
Require: Population size $\lambda \in \mathbb{N}$.
Require: Initial population $Q_0 \in \mathcal{Y}^\lambda$.

```
for t = 0, 1, 2, ... until the termination condition is met do
    for i = 1 to λ do
        Choose I_t(i) ∈ [λ], and set x := Q_t(I_t(i)).
        Sample x' ~ p_mut(x) and set Q_{t+1}(i) := x'.
    end for
end for
```

算法 2 计算集合 P Q 的过程属于算法 3
令 $x^* = \{1\}^n$，$a(n) := (1-\alpha)n$，$b(n) := (1-\alpha+\varepsilon)n$
算法的性质满足 $\alpha_0 \leq 1$

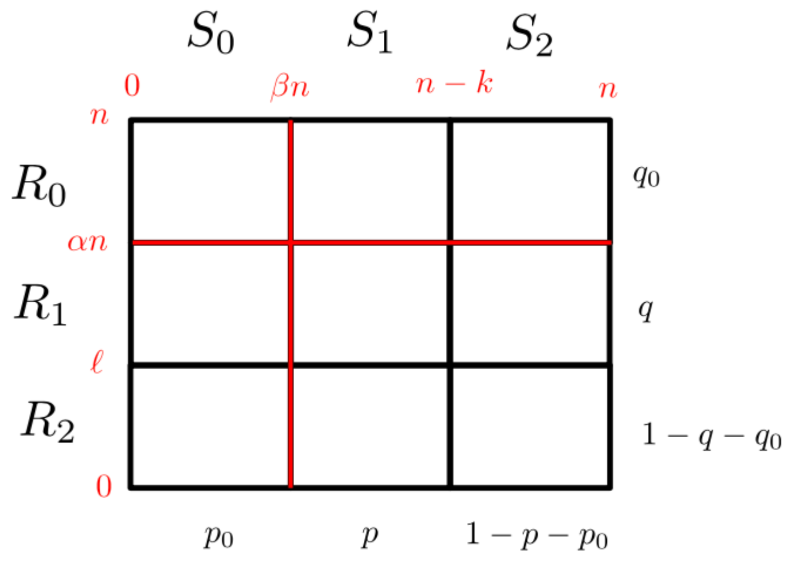
Proof 2

Lemma 23. *If there exists a constant $\delta > 0$ such that*

1) $q_0 \leq \sqrt{2(1-\delta)} - 1$

then

$$\varphi := \frac{p_{sel}(S_0 \cup S_1)}{p(S_0 \cup S_1)} > 1 + \delta(1 - p - p_0).$$



搜索空间的划分

$$\begin{aligned}
&p_{\text{sel}}(S_0 \cup S_1) \\
&= \Pr(x_1 \in S_0 \cup S_1 \wedge x_2 \in S_0 \cup S_1) + \\
&\quad + \Pr(x_1 \in S_0 \cup S_1 \wedge x_2 \notin S_0 \cup S_1) \\
&\qquad \times \Pr((x_1, y_1) \succeq (x_2, y_2) \mid x_1 \in S_0 \cup S_1 \wedge x_2 \notin S_0 \cup S_1) \\
&\quad + \Pr(x_1 \notin S_0 \cup S_1 \wedge x_2 \in S_0 \cup S_1) \\
&\qquad \times (1 - \Pr((x_1, y_1) \succeq (x_2, y_2) \mid x_1 \notin S_0 \cup S_1 \wedge x_2 \in S_0 \cup S_1)) \\
&\geq (p_0 + p)^2 + (p_0 + p)(1 - q_0)^2(1 - p - p_0)/2 + \\
&\quad + (p_0 + p)(1 - (p_0 + p) - q_0^2(1 - p - p_0))
\end{aligned}$$

对第三项进行放缩： $\Pr(x_2 \in S_0 \cup S_1)$
$\times (1 - \Pr(y_2 \in R_0 \wedge (x_1, y_1) \in S_2 \times R_0) - \Pr(x_1 \in S_0 \cup S_1))$

根据 $p(S_0 \cup S_1) = p_0 + p,$

$$\begin{aligned}
\varphi &\geq 1 + (1 - p - p_0)((1 - q_0)^2/2 - q_0) \\
&\geq 1 + (1 - p - p_0)(1/2 - (1 - 2\delta)/2) \\
&= 1 + \delta(1 - p - p_0).
\end{aligned}$$