Lecture 12
Design of Evolutionary Algorithms

Three tasks:
2. Choice / design of variation operators acting in problem representation.
3. Choice of strategy parameters (includes initialization).

ad 1) different “schools”:
(a) operate on binary representation and define genotype/phenotype mapping
   + can use standard algorithm
   – mapping may induce unintentional bias in search
(b) no doctrine: use “most natural” representation
   – must design variation operators for specific representation
   + if design done properly then no bias in search

ad 2) design guidelines for variation operators

a) reachability
   every x ∈ X should be reachable from arbitrary x₀ ∈ X
   after finite number of repeated variations with positive probability bounded from 0
b) unbiasedness
   unless having gathered knowledge about problem
   variation operator should not favor particular subsets of solutions
   ⇒ formally: maximum entropy principle

c) control
   variation operator should have parameters affecting shape of distributions;
   known from theory: weaken variation strength when approaching optimum

ad 2) design guidelines for variation operators in practice

binary search space X = ℤⁿ
variation by k-point or uniform crossover and subsequent mutation

a) reachability:
   regardless of the output of crossover
   we can move from x ∈ ℤⁿ to y ∈ ℤⁿ in 1 step with probability
   \[ p(x, y) = p_m^H(x, y) (1 - p_m)^{n-H(x, y)} > 0 \]
   where H(x,y) is Hamming distance between x and y.
   Since min( p(x,y): x,y ∈ ℤⁿ ) = δ > 0 we are done.
b) **unbiasedness**

don’t prefer any direction or subset of points without reason

⇒ use maximum entropy distribution for sampling!

**properties:**
- distributes probability mass as uniform as possible
- additional knowledge can be included as constraints:
  \( \Rightarrow \) under given constraints sample as uniform as possible

---

**Definition:**

Let \( X \) be discrete random variable (r.v.) with \( p_k = P(X = x_k) \) for some index set \( K \). The quantity

\[
H(X) = -\sum_{k \in K} p_k \log p_k
\]

is called the **entropy of the distribution** of \( X \). If \( X \) is a continuous r.v. with p.d.f. \( f_X(z) \) then the entropy is given by

\[
H(X) = \int_{-\infty}^{\infty} f_X(z) \log f_X(z) dz
\]

The distribution of a random variable \( X \) for which \( H(X) \) is maximal is termed a **maximum entropy distribution**.

---

**Excursion: Maximum Entropy Distributions**

**Knowledge available:**

Discrete distribution with support \( \{ x_1, x_2, \ldots, x_n \} \) with \( x_1 < x_2 < \ldots < x_n < \infty \)

\[
p_k = P(X = x_k)
\]

⇒ leads to nonlinear constrained optimization problem:

\[- \sum_{k=1}^{n} p_k \log p_k \rightarrow \text{max!} \quad \text{s.t.} \quad \sum_{k=1}^{n} p_k = 1\]

**solution:** via Lagrange (find stationary point of Lagrangian function)

\[
L(p, \alpha) = -\sum_{k=1}^{n} p_k \log p_k + \alpha \left( \sum_{k=1}^{n} p_k - 1 \right)
\]

partial derivatives:

\[
\frac{\partial L(p, \alpha)}{\partial p_k} = -1 - \log p_k + \frac{1}{\alpha} \Rightarrow p_k = e^{\alpha - 1}
\]

\[
\frac{\partial L(p, \alpha)}{\partial \alpha} = \sum_{k=1}^{n} p_k - 1 \equiv 0\]

\[
\Rightarrow \sum_{k=1}^{n} p_k = \frac{1}{\alpha} \quad \text{uniform distribution}
\]
Excursion: Maximum Entropy Distributions  Lecture 12

Knowledge available:
Discrete distribution with support \{1, 2, …, n\} with \( p_k = P(X = k) \) and \( E[X] = \nu \)

\[ \Rightarrow \text{leads to nonlinear constrained optimization problem:} \]
\[ - \sum_{k=1}^{n} p_k \log p_k \rightarrow \text{max!} \]
\[ \text{s.t. } \sum_{k=1}^{n} p_k = 1 \quad \text{and} \quad \sum_{k=1}^{n} k p_k = \nu \]

solution: via Lagrange (find stationary point of Lagrangian function)

\[ L(p, \alpha, b) = - \sum_{k=1}^{n} p_k \log p_k + \alpha \left( \sum_{k=1}^{n} p_k - 1 \right) + b \left( \sum_{k=1}^{n} k p_k - \nu \right) \]

\[ \partial L(p, \alpha, b) \bigg/ \partial p_k = 0 \Rightarrow p_k = c a - 1 + b k \]

\[ \partial L(p, \alpha, b) \bigg/ \partial \alpha = 0 \Rightarrow \sum_{k=1}^{n} p_k = U \]

\[ \partial L(p, \alpha, b) \bigg/ \partial b = 0 \Rightarrow \sum_{k=1}^{n} k p_k - \nu = 0 \]

\[ \Rightarrow \sum_{k=1}^{n} p_k = e^a \sum_{k=1}^{n} (e^b)^k = 1 \]

(continued on next slide)

⇒ \( e^a - 1 = \sum_{k=1}^{n} (c^b)^k \Rightarrow p_k = e^a - 1 + b k \)

⇒ discrete Boltzmann distribution

\[ p_k = \frac{q^k}{\sum_{i=1}^{n} q^i} \quad (q = e^b) \]

⇒ value of \( q \) depends on \( \nu \) via third condition: (\( \ast \))

\[ \sum_{k=1}^{n} k p_k = \frac{\sum_{k=1}^{n} k q^k}{\sum_{i=1}^{n} q^i} = \frac{1}{(1-q) \left(1 - q^n \right)} \]

Boltzmann distribution

\( (n = 9) \)

specializes to uniform distribution if \( \nu = 5 \)

(\( \text{as expected} \))

\( \nu = 2 \)

\( \nu = 8 \)

\( \nu = 3 \)

\( \nu = 7 \)

\( \nu = 4 \)

\( \nu = 6 \)

\( \nu = 5 \)
Excursion: Maximum Entropy Distributions

Knowledge available:
Discrete distribution with support \{1, 2, ..., n\} with \(E[X] = \nu\) and \(V[X] = \eta^2\)

\[\begin{align*}
&\Rightarrow \text{leads to nonlinear constrained optimization problem:} \\
&\quad \quad - \sum_{k=1}^{n} p_k \log p_k \to \max! \\
&\quad \quad \text{s.t.} \quad \sum_{k=1}^{n} p_k = 1 \quad \text{and} \quad \sum_{k=1}^{n} k p_k = \nu \quad \text{and} \quad \sum_{k=1}^{n} (k - \nu)^2 p_k = \eta^2
\end{align*}\]

**Note:** Constraints are linear equations in \(p_k\).

Solution: In principle, via Lagrange (find stationary point of Lagrangian function) but very complicated analytically, if possible at all ⇒ consider special cases only

Special case: \(n = 3\) and \(E[X] = 2\) and \(V[X] = \eta^2\)

Linear constraints uniquely determine distribution:

\[\begin{align*}
&\text{I. } p_1 + p_2 + p_3 = 1 \\
&\text{II. } p_1 + 2p_2 + 3p_3 = 2 \\
&\text{III. } p_1 + 0 + p_3 = \eta^2
\end{align*}\]

\[\begin{align*}
&\text{II-}1:\quad p_2 + 7p_3 = 1 \\
&\text{I-III}:\quad p_3 = \frac{\eta^2}{2}
\end{align*}\]

\[\begin{align*}
&\to p = \left(\frac{\eta^2}{2}, 1 - \frac{\eta^2}{2}, \frac{\eta^2}{2}\right)
\end{align*}\]

Excursion: Maximum Entropy Distributions

Knowledge available:
Discrete distribution with unbounded support \{0, 1, 2, ...\} and \(E[X] = \nu\)

\[\begin{align*}
&\Rightarrow \text{leads to infinite-dimensional nonlinear constrained optimization problem:} \\
&\quad \quad - \sum_{k=0}^{\infty} p_k \log p_k \to \max! \\
&\quad \quad \text{s.t.} \quad \sum_{k=0}^{\infty} p_k = 1 \quad \text{and} \quad \sum_{k=0}^{\infty} k p_k = \nu
\end{align*}\]

Solution: Via Lagrange (find stationary point of Lagrangian function)

\[\begin{align*}
L(p, a, b) &= - \sum_{k=0}^{\infty} p_k \log p_k + a \left(\sum_{k=0}^{\infty} k p_k - \nu\right) + b \left(\sum_{k=0}^{\infty} k \cdot p_k - \nu\right)
\end{align*}\]

Partial derivatives:

\[\begin{align*}
&\frac{\partial L(p, a, b)}{\partial p_k} = 1 \log p_k | a \quad \text{and} \quad b k = 0 \Rightarrow p_k = e^{a-1+bk} \\
&\frac{\partial L(p, a, b)}{\partial a} = \sum_{k=0}^{\infty} p_k - 1 = 0 \\
&\frac{\partial L(p, a, b)}{\partial b} = \sum_{k=0}^{\infty} k p_k - \nu = 0 \\
&\sum_{k=0}^{\infty} p_k = e^{a} \sum_{k=0}^{\infty} \left(e^{b}\right)^k = 1
\end{align*}\]

(continued on next slide)
\[ e^{n-1} = \frac{1}{\sum_{k=0}^{\infty} (e^b)^k} \Rightarrow p_k = e^{n-1}b_k = \frac{(e^b)^k}{\sum_{i=0}^{\infty} (e^b)^i} \]

set \( q = e^b \) and insists that \( q < 1 \)

\[ \Rightarrow q = \frac{1}{1 - q} \]

it remains to specify \( q \); to proceed recall that

\[ \sum_{k=0}^{\infty} k p_k = \frac{q}{1 - q} \]

\[ \Rightarrow q = \frac{\nu}{\nu + 1} = 1 - \frac{1}{\nu + 1} \]

\[ \Rightarrow p_k = \frac{1}{\nu + 1} \left(1 - \frac{1}{\nu + 1}\right)^k \]