technische universität dortmund

Computational Intelligence

Winter Term 2010/11

Prof. Dr. Günter Rudolph

Lehrstuhl für Algorithm Engineering (LS 11)

Fakultät für Informatik

TU Dortmund

Three tasks:

1. Choice of an appropriate problem representation.
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 $x \in X$ should be reachable from arbitrary $x_0 \in X$
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
$\Rightarrow$ 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 $X = \mathbb{B}^n$

variation by k-point or uniform crossover and subsequent mutation

a) ***reachability***:
regardless of the output of crossover
we can move from $x \in \mathbb{B}^n$ to $y \in \mathbb{B}^n$ 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): $x,y \in \mathbb{B}^n$ } = $\delta > 0$ 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:
  → under given constraints sample as uniform as possible

Formally:

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. $f_X(\cdot)$ then the entropy is given by

$$H(X) = -\int_{-\infty}^{\infty} f_X(x) \log f_X(x) dx$$

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

Knowledge available:

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

$$p_k = \mathsf{P}\{X = x_k\}$$

⇒ leads to nonlinear constrained optimization problem:

$$-\sum_{k=1}^{n} p_k \log p_k \quad \to \text{max!}$$

$$\text{s.t.} \quad \sum_{k=1}^{n} p_k = 1$$

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

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

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

partial derivatives:

$$\frac{\partial L(p,a)}{\partial p_k} = -1 - \log p_k + a \overset{!}{=} 0 \quad \Rightarrow \quad p_k \overset{!}{=} e^{a-1}$$

$$\frac{\partial L(p,a)}{\partial a} = \sum_{k=1}^{n} p_k - 1 \overset{!}{=} 0$$

$$\Rightarrow \sum_{k=1}^{n} p_k = \sum_{k=1}^{n} e^{a-1} = n\, e^{a-1} \overset{!}{=} 1 \quad \Leftrightarrow \quad e^{a-1} = \frac{1}{n}$$

$$p_k = \frac{1}{n}$$

uniform distribution

■

Knowledge available:

Discrete distribution with support { 1, 2, …, n } with p_k = P { X = k } **and** E[X] = ν

⇒ leads to nonlinear constrained optimization problem:

$$-\sum_{k=1}^{n} p_k \log p_k \quad \to \text{max!}$$

$$\text{s.t.} \quad \sum_{k=1}^{n} p_k = 1 \quad \text{and} \quad \sum_{k=1}^{n} k\, p_k = \nu$$

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

$$L(p,a,b) = -\sum_{k=1}^{n} p_k \log p_k + a\left(\sum_{k=1}^{n} p_k - 1\right) + b\left(\sum_{k=1}^{n} k \cdot p_k - \nu\right)$$

$$L(p,a,b) = -\sum_{k=1}^{n} p_k \log p_k + a\left(\sum_{k=1}^{n} p_k - 1\right) + b\left(\sum_{k=1}^{n} k \cdot p_k - \nu\right)$$

partial derivatives:

$$\frac{\partial L(p,a,b)}{\partial p_k} = -1 - \log p_k + a + b\,k \stackrel{!}{=} 0 \qquad \Rightarrow \quad p_k = e^{a-1+b\,k}$$

$$\frac{\partial L(p,a,b)}{\partial a} = \sum_{k=1}^{n} p_k - 1 \stackrel{!}{=} 0$$

$$\frac{\partial L(p,a,b)}{\partial b} \stackrel{(*)}{=} \sum_{k=1}^{n} k\, p_k - \nu \stackrel{!}{=} 0$$

$$\Rightarrow \quad \sum_{k=1}^{n} p_k = e^{a-1} \sum_{k=1}^{n} (e^b)^k \stackrel{!}{=} 1$$

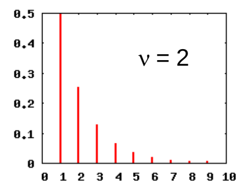
(continued on next slide)

$$\Rightarrow \quad e^{a-1} = \frac{1}{\sum_{k=1}^{n} (e^b)^k} \qquad \Rightarrow \quad p_k = e^{a-1+b\,k} = \frac{(e^b)^k}{\sum_{i=1}^{n} (e^b)^i}$$

⇒ **discrete Boltzmann distribution** $\quad p_k = \dfrac{q^k}{\sum_{i=1}^{n} q^i} \qquad (q = e^b)$

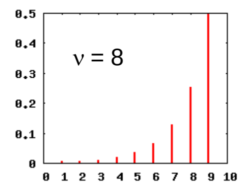
⇒ value of q depends on ν via third condition: (*)

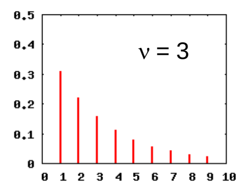
$$\sum_{k=1}^{n} k\, p_k = \frac{\sum_{k=1}^{n} k\, q^k}{\sum_{i=1}^{n} q^i} = \frac{1-(n+1)\,q^n + n\,q^{n+1}}{(1-q)\,(1-q^n)} \stackrel{!}{=} \nu$$



Boltzmann distribution

(n = 9)

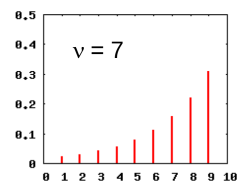


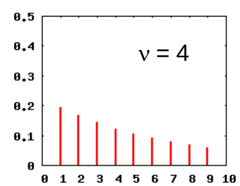


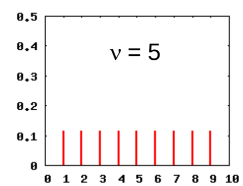
specializes to uniform distribution if ν = 5

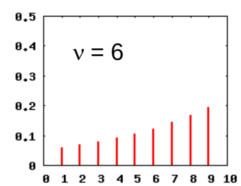
(as expected)

↓









Excursion: Maximum Entropy Distributions — Lecture 11

Knowledge available:

Discrete distribution with support { 1, 2, …, n } with E[X] = ν **and** V[X] = η²

⇒ leads to nonlinear constrained optimization problem:

$$-\sum_{k=1}^{n} p_k \log p_k \quad \to \text{max!}$$

$$\text{s.t.} \quad \sum_{k=1}^{n} p_k = 1 \quad \text{and} \quad \sum_{k=1}^{n} k\, p_k = \nu \quad \text{and} \quad \sum_{k=1}^{n} (k-\nu)^2 p_k = \eta^2$$

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

but very complicated analytically, if possible at all

⇒ consider special cases only

note: constraints are linear equations in p_k

Excursion: Maximum Entropy Distributions — Lecture 11

Special case: **n = 3** and **E[X] = 2** and **V[X] = η²**

Linear constraints uniquely determine distribution:

$$\begin{array}{llllllll}
\text{I.} & p_1 & + & p_2 & + & p_3 & = & 1 \\
\text{II.} & p_1 & + & 2\,p_2 & + & 3\,p_3 & = & 2 \\
\text{III.} & p_1 & + & 0 & + & p_3 & = & \eta^2
\end{array}$$

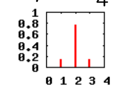
$$\begin{array}{llllll}
\text{II} - \text{I:} & p_2 & + & 2\,p_3 & = & 1 \\
\text{I} - \text{III:} & p_2 & & & = & 1-\eta^2
\end{array} \quad \Big\} \quad p_3 = \frac{\eta^2}{2}$$

insertion in III.

$$p_1 = \frac{\eta^2}{2}$$

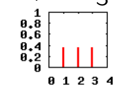
$$\Rightarrow \quad p = \left(\frac{\eta^2}{2}, 1-\eta^2, \frac{\eta^2}{2}\right)$$

$\eta^2 = \frac{1}{4}$



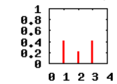
unimodal

$\eta^2 = \frac{2}{3}$



uniform

$\eta^2 = \frac{4}{5}$



bimodal

Excursion: Maximum Entropy Distributions — Lecture 11

Knowledge available:

Discrete distribution with unbounded support { 0, 1, 2, … } and E[X] = ν

⇒ leads to infinite-dimensional nonlinear constrained optimization problem:

$$-\sum_{k=0}^{\infty} p_k \log p_k \quad \to \text{max!}$$

$$\text{s.t.} \quad \sum_{k=0}^{\infty} p_k = 1 \quad \text{and} \quad \sum_{k=0}^{\infty} k\, p_k = \nu$$

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

$$L(p,a,b) = -\sum_{k=0}^{\infty} p_k \log p_k + a\left(\sum_{k=0}^{\infty} p_k - 1\right) + b\left(\sum_{k=0}^{\infty} k \cdot p_k - \nu\right)$$

Excursion: Maximum Entropy Distributions — Lecture 11

$$L(p,a,b) = -\sum_{k=0}^{\infty} p_k \log p_k + a\left(\sum_{k=0}^{\infty} p_k - 1\right) + b\left(\sum_{k=0}^{\infty} k \cdot p_k - \nu\right)$$

partial derivatives:

$$\frac{\partial L(p,a,b)}{\partial p_k} = -1 - \log p_k + a + b\,k \overset{!}{=} 0 \qquad \Rightarrow \quad p_k = e^{a-1+b\,k}$$

$$\frac{\partial L(p,a,b)}{\partial a} = \sum_{k=0}^{\infty} p_k - 1 \overset{!}{=} 0$$

$$\frac{\partial L(p,a,b)}{\partial b} \overset{(*)}{=} \sum_{k=0}^{\infty} k\, p_k - \nu \overset{!}{=} 0$$

$$\Rightarrow \quad \sum_{k=0}^{\infty} p_k = e^{a-1} \sum_{k=0}^{\infty} (e^b)^k \overset{!}{=} 1$$

(continued on next slide)

Excursion: Maximum Entropy Distributions — Lecture 11

$\Rightarrow \quad e^{a-1} = \dfrac{1}{\sum\limits_{k=0}^{\infty} (e^b)^k} \qquad \Rightarrow \quad p_k = e^{a-1+b\,k} = \dfrac{(e^b)^k}{\sum\limits_{i=0}^{\infty} (e^b)^i}$

set $q = e^b$ and insists that $q < 1$ $\quad \Rightarrow \quad \sum\limits_{k=0}^{\infty} q^k = \dfrac{1}{1-q}$ insert

$\Rightarrow \quad p_k = (1-q)\,q^k$ for $k = 0, 1, 2, \ldots$ **geometrical distribution**

it remains to specify q; to proceed recall that $\sum\limits_{k=0}^{\infty} k\,q^k = \dfrac{q}{(1-q)^2}$

Excursion: Maximum Entropy Distributions — Lecture 11

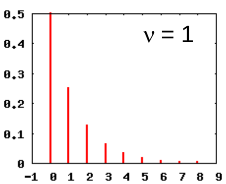
$\Rightarrow$ value of q depends on ν via third condition: (*)

$$\sum_{k=0}^{\infty} k\,p_k = \frac{\sum\limits_{k=0}^{\infty} k\,q^k}{\sum\limits_{i=0}^{\infty} q^i} = \frac{q}{1-q} \overset{!}{=} \nu$$

$$\Rightarrow \quad q = \frac{\nu}{\nu+1} = 1 - \frac{1}{\nu+1}$$

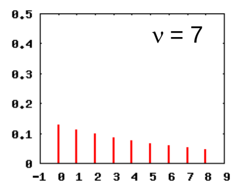
$$\Rightarrow \quad p_k = \frac{1}{\nu+1}\left(1 - \frac{1}{\nu+1}\right)^k$$

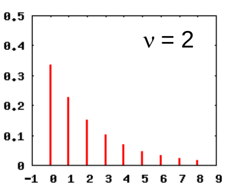
Excursion: Maximum Entropy Distributions — Lecture 11



geometrical distribution

with E[x] = ν





p_k only shown for k = 0, 1, …, 8