

# Computational Intelligence

Winter Term 2014/15

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).

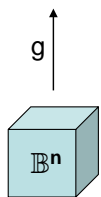
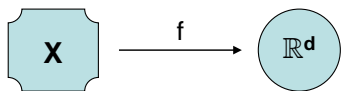
ad 1) different "schools":

- (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 1a) genotype-phenotype mapping

original problem  $f: X \rightarrow \mathbb{R}^d$

scenario: no standard algorithm for search space  $X$  available



- standard EA performs variation on binary strings  $b \in \mathbb{B}^n$
- fitness evaluation of individual  $b$  via  $(f \circ g)(b) = f(g(b))$  where  $g: \mathbb{B}^n \rightarrow X$  is genotype-phenotype mapping
- selection operation independent from representation

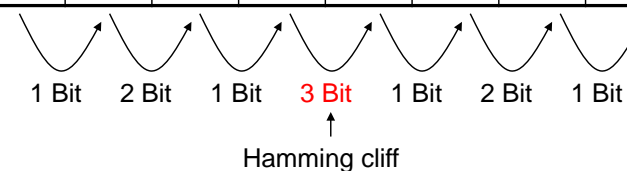
**Genotype-Phenotype-Mapping**  $\mathbb{B}^n \rightarrow [L, R] \subset \mathbb{R}$

- Standard encoding for  $b \in \mathbb{B}^n$

$$x = L + \frac{R - L}{2^n - 1} \sum_{i=0}^{n-1} b_{n-i} 2^i$$

→ Problem: *hamming cliffs*

000	001	010	011	100	101	110	111
0	1	2	3	4	5	6	7



← genotype  
← phenotype

$L = 0, R = 7$   
 $n = 3$

Genotype-Phenotype-Mapping  $\mathbb{B}^n \rightarrow [L, R] \subset \mathbb{R}$ 

- Gray encoding for  $b \in \mathbb{B}^n$

Let  $a \in \mathbb{B}^n$  standard encoded. Then  $b_i = \begin{cases} a_i, & \text{if } i = 1 \\ a_{i-1} \oplus a_i, & \text{if } i > 1 \end{cases}$   $\oplus = \text{XOR}$

000	001	011	010	110	111	101	100	← genotype
0	1	2	3	4	5	6	7	← phenotype

OK, no hamming cliffs any longer ...

⇒ small changes in phenotype „lead to“ small changes in genotype

since we consider evolution in terms of Darwin (not Lamarck):

⇒ small changes in genotype lead to small changes in phenotype!

**but:** 1-Bit-change: 000 → 100 ⇒ ☹

Genotype-Phenotype-Mapping  $\mathbb{B}^n \rightarrow \mathbb{P}^{\log(n)}$  (example only)

- e.g. standard encoding for  $b \in \mathbb{B}^n$

individual:

010	101	111	000	110	001	101	100	← genotype
0	1	2	3	4	5	6	7	← index

consider index and associated genotype entry as unit / record / struct;

sort units with respect to genotype value, old indices yield permutation:

000	001	010	100	101	101	110	111	← genotype
3	5	0	7	1	6	4	2	← old index = permutation

ad 1a) genotype-phenotype mapping

typically required: strong causality

→ small changes in individual leads to small changes in fitness

→ small changes in genotype should lead to small changes in phenotype

**but:** how to find a genotype-phenotype mapping with that property?

**necessary conditions:**

- 1)  $g: \mathbb{B}^n \rightarrow X$  can be computed efficiently (otherwise it is senseless)
- 2)  $g: \mathbb{B}^n \rightarrow X$  is surjective (otherwise we might miss the optimal solution)
- 3)  $g: \mathbb{B}^n \rightarrow X$  *preserves closeness* (otherwise strong causality endangered)

Let  $d(\cdot, \cdot)$  be a metric on  $\mathbb{B}^n$  and  $d_X(\cdot, \cdot)$  be a metric on  $X$ .

$$\forall x, y, z \in \mathbb{B}^n: d(x, y) \leq d(x, z) \Rightarrow d_X(g(x), g(y)) \leq d_X(g(x), g(z))$$

ad 1b) use “most natural“ representation

typically required: strong causality

→ small changes in individual leads to small changes in fitness

→ need variation operators that obey that requirement

**but:** how to find variation operators with that property?

⇒ need design guidelines ...

## 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

$\Rightarrow$  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

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  $\{x_1, x_2, \dots, x_n\}$  with  $x_1 < x_2 < \dots < x_n < \infty$

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

⇒ leads to nonlinear constrained optimization problem:

$$\begin{aligned} & - \sum_{k=1}^n p_k \log p_k \rightarrow \max! \\ \text{s.t.} \quad & \sum_{k=1}^n p_k = 1 \end{aligned}$$

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:

$$\begin{aligned} \frac{\partial L(p, a)}{\partial p_k} &= -1 - \log p_k + a \stackrel{!}{=} 0 & \Rightarrow p_k \stackrel{!}{=} e^{a-1} \\ \frac{\partial L(p, a)}{\partial a} &= \sum_{k=1}^n p_k - 1 \stackrel{!}{=} 0 \end{aligned}$$

}  $p_k = \frac{1}{n}$   
**uniform distribution**

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

**Knowledge available:**

Discrete distribution with support  $\{1, 2, \dots, n\}$  with  $p_k = P\{X = k\}$  and  $E[X] = \nu$

⇒ leads to nonlinear constrained optimization problem:

$$\begin{aligned} & - \sum_{k=1}^n p_k \log p_k \rightarrow \max! \\ \text{s.t.} \quad & \sum_{k=1}^n p_k = 1 \quad \text{and} \quad \sum_{k=1}^n k p_k = \nu \end{aligned}$$

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:

$$\begin{aligned} \frac{\partial L(p, a, b)}{\partial p_k} &= -1 - \log p_k + a + b k \stackrel{!}{=} 0 & \Rightarrow p_k = e^{a-1+bk} \\ \frac{\partial L(p, a, b)}{\partial a} &= \sum_{k=1}^n p_k - 1 \stackrel{!}{=} 0 \\ \frac{\partial L(p, a, b)}{\partial b} &= \sum_{k=1}^n k p_k - \nu \stackrel{!}{=} 0 \end{aligned}$$

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

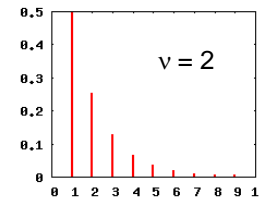
(continued on next slide)

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

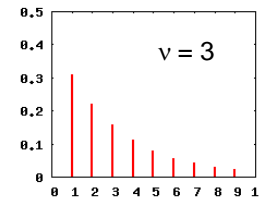
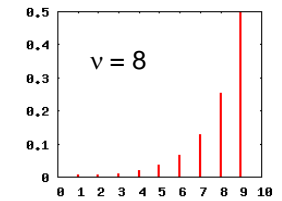
$$\Rightarrow \text{discrete Boltzmann distribution } p_k = \frac{q^k}{\sum_{i=1}^n q^i} \quad (q = e^b)$$

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

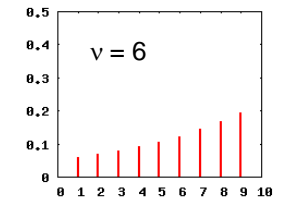
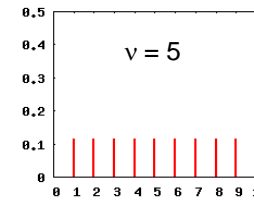
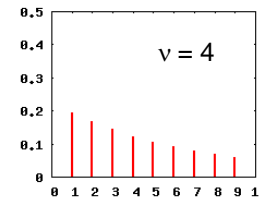
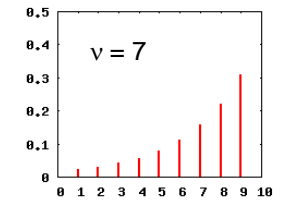
$$\sum_{k=1}^n k p_k = \frac{\sum_{k=1}^n k q^k}{\sum_{i=1}^n q^i} = \frac{1 - (n+1)q^n + nq^{n+1}}{(1-q)(1-q^n)} \stackrel{!}{=} \nu$$



Boltzmann distribution  
(n = 9)



specializes to uniform  
distribution if v = 5  
(as expected)



Knowledge available:

Discrete distribution with support { 1, 2, ..., n } with  $E[X] = \nu$  and  $V[X] = \eta^2$

⇒ leads to nonlinear constrained optimization problem:

$$-\sum_{k=1}^n p_k \log p_k \rightarrow \max!$$

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

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  $p_k$

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

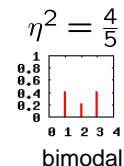
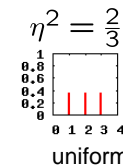
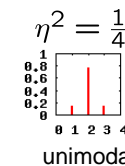
Linear constraints uniquely determine distribution:

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

$$\begin{aligned} \text{II - I: } & p_2 + 2p_3 = 1 \\ \text{I - III: } & p_2 = 1 - \eta^2 \end{aligned}$$

$$\left. \begin{aligned} p_1 &= \frac{\eta^2}{2} \\ p_3 &= \frac{\eta^2}{2} \end{aligned} \right\} \begin{array}{l} \uparrow \\ \text{insertion in III.} \end{array}$$

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



**Knowledge available:**

Discrete distribution with unbounded support  $\{0, 1, 2, \dots\}$  and  $E[X] = \nu$

$\Rightarrow$  leads to infinite-dimensional nonlinear constrained optimization problem:

$$-\sum_{k=0}^{\infty} p_k \log p_k \rightarrow \max!$$

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

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

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

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

partial derivatives:

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

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

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

(continued on next slide)

$$\Rightarrow e^{a-1} = \frac{1}{\sum_{k=0}^{\infty} (e^b)^k} \quad \Rightarrow \quad p_k = e^{a-1+bk} = \frac{(e^b)^k}{\sum_{i=0}^{\infty} (e^b)^i}$$

$$\text{set } q = e^b \text{ and insists that } q < 1 \quad \Rightarrow \quad \sum_{k=0}^{\infty} q^k = \frac{1}{1-q} \quad \text{insert}$$

$$\Rightarrow p_k = (1-q)q^k \quad \text{for } k = 0, 1, 2, \dots \quad \text{geometrical distribution}$$

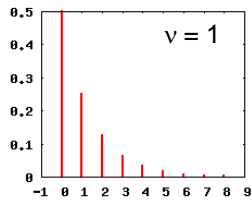
$$\text{it remains to specify } q; \quad \text{to proceed recall that} \quad \sum_{k=0}^{\infty} k q^k = \frac{q}{(1-q)^2}$$

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

$$\sum_{k=0}^{\infty} k p_k = \frac{\sum_{k=0}^{\infty} k q^k}{\sum_{i=0}^{\infty} q^i} = \frac{q}{1-q} \stackrel{!}{=} \nu$$

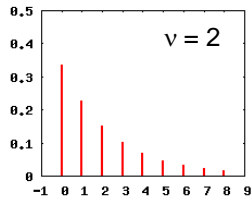
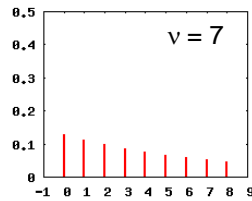
$$\Rightarrow q = \frac{\nu}{\nu+1} = 1 - \frac{1}{\nu+1}$$

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

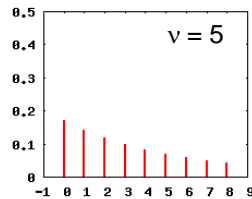
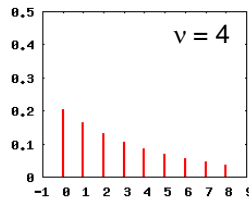
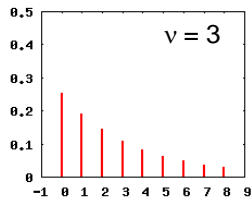
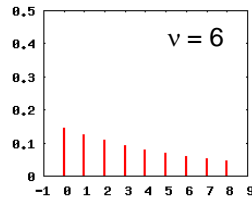


geometrical distribution

with  $E[x] = v$



$p_k$  only shown  
for  $k = 0, 1, \dots, 8$



Overview:

support  $\{ 1, 2, \dots, n \}$   $\Rightarrow$  discrete uniform distribution  
 and require  $E[X] = \theta$   $\Rightarrow$  Boltzmann distribution  
 and require  $V[X] = \eta^2$   $\Rightarrow$  N.N. (not Binomial distribution)

support  $\mathbb{N}$   $\Rightarrow$  not defined!  
 and require  $E[X] = \theta$   $\Rightarrow$  geometrical distribution  
 and require  $V[X] = \eta^2$   $\Rightarrow$  ?

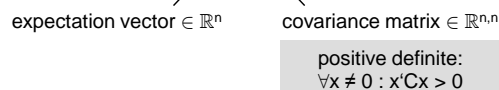
support  $\mathbb{Z}$   $\Rightarrow$  not defined!  
 and require  $E[|X|] = \theta$   $\Rightarrow$  bi-geometrical distribution (discrete Laplace distr.)  
 and require  $E[|X|^2] = \eta^2$   $\Rightarrow$  N.N. (discrete Gaussian distr.)

support  $[a,b] \subset \mathbb{R}$   $\Rightarrow$  uniform distribution

support  $\mathbb{R}^+$  with  $E[X] = \theta$   $\Rightarrow$  Exponential distribution

support  $\mathbb{R}$   
 with  $E[X] = \theta, V[X] = \eta^2$   $\Rightarrow$  normal / Gaussian distribution  $N(\theta, \eta^2)$

support  $\mathbb{R}^n$   
 with  $E[X] = \theta$   
 and  $Cov[X] = C$   $\Rightarrow$  multinormal distribution  $N(\theta, C)$



for permutation distributions ?

$\rightarrow$  uniform distribution on all possible permutations

```

set v[j] = j for j = 1, 2, ..., n
for i = n to 1 step -1
    draw k uniformly at random from { 1, 2, ..., i }
    swap v[i] and v[k]
endfor
    
```

generates permutation uniformly at random in  $\Theta(n)$  time

Guideline:

Only if you know something about the problem *a priori* or  
 if you have learnt something about the problem *during the search*  
 $\Rightarrow$  include that knowledge in search / mutation distribution (via constraints!)

ad 2) design guidelines for variation operators in practice

integer search space  $X = \mathbb{Z}^n$

- a) reachability
- b) unbiasedness
- c) control

- every recombination results in some  $z \in \mathbb{Z}^n$   
 - mutation of  $z$  may then lead to any  $z^* \in \mathbb{Z}^n$  with positive probability in one step

ad a) support of mutation should be  $\mathbb{Z}^n$

ad b) need maximum entropy distribution over support  $\mathbb{Z}^n$

ad c) control variability by parameter

→ formulate as constraint of maximum entropy distribution

ad 2) design guidelines for variation operators in practice

$X = \mathbb{Z}^n$

**task:** find (symmetric) maximum entropy distribution over  $\mathbb{Z}$  with  $E[|Z|] = \theta > 0$

⇒ need *analytic* solution of a  $\infty$ -dimensional, nonlinear optimization problem with constraints!

$$H(p) = - \sum_{k=-\infty}^{\infty} p_k \log p_k \rightarrow \max!$$

s.t.  $p_k = p_{-k} \quad \forall k \in \mathbb{Z},$  (symmetry w.r.t. 0)

$$\sum_{k=-\infty}^{\infty} p_k = 1,$$
 (normalization)

$$\sum_{k=-\infty}^{\infty} |k| p_k = \theta$$
 (control "spread")

$$p_k \geq 0 \quad \forall k \in \mathbb{Z}.$$
 (nonnegativity)

**result:**

a random variable  $Z$  with support  $\mathbb{Z}$  and probability distribution

$$p_k := P\{Z = k\} = \frac{q}{2-q} (1-q)^{|k|}, \quad k \in \mathbb{Z}, \quad q \in (0, 1)$$

symmetric w.r.t. 0, unimodal, spread manageable by  $q$  and has max. entropy ■

**generation of pseudo random numbers:**

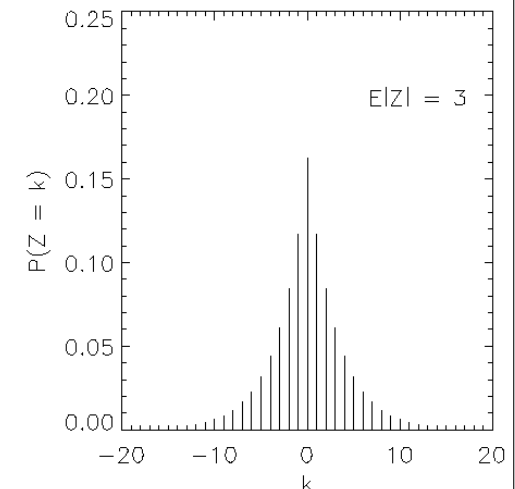
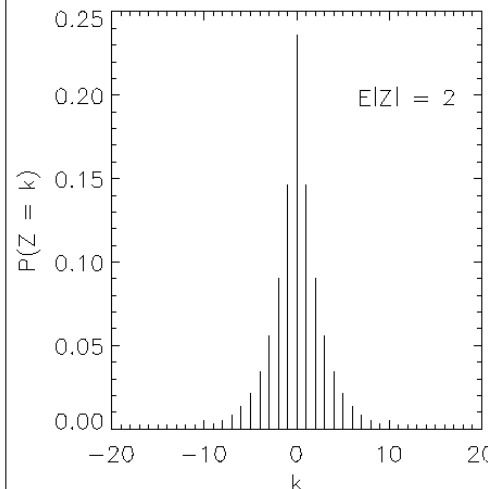
$$Z = G_1 - G_2$$

where

$$U_i \sim U(0, 1) \Rightarrow G_i = \left\lfloor \frac{\log(1 - U_i)}{\log(1 - q)} \right\rfloor, \quad i = 1, 2.$$

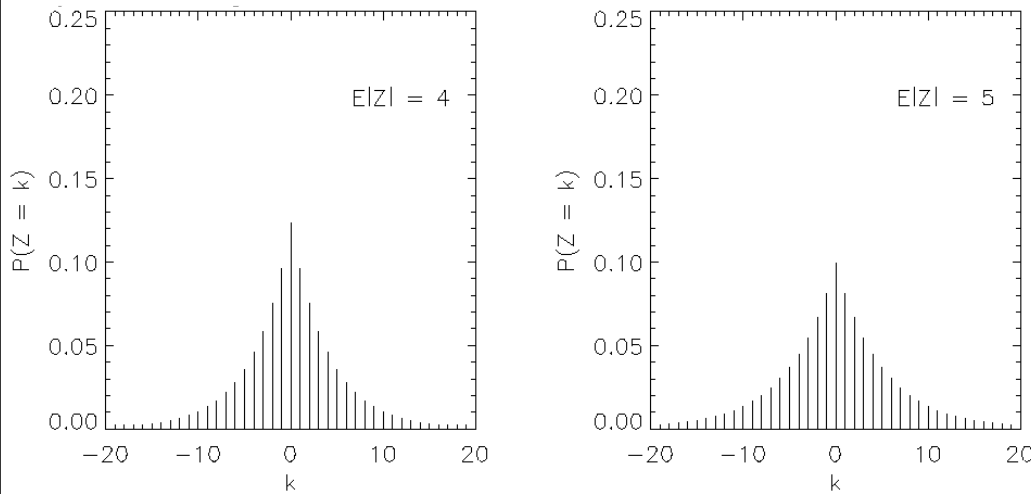
stochastic independent!

**probability distributions for different mean step sizes  $E|Z| = \theta$**





probability distributions for different mean step sizes  $E|Z| = \theta$



How to control the spread?

We must be able to adapt  $q \in (0,1)$  for generating  $Z$  with variable  $E|Z| = \theta$  !

self-adaptation of  $q$  in open interval  $(0,1)$  ?

→ make mean step size  $E[|Z|]$  adjustable!

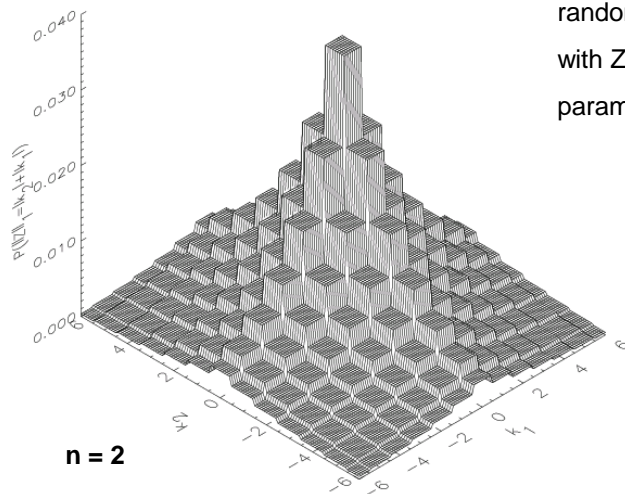
$$E[|Z|] = \sum_{k=-\infty}^{\infty} |k| p_k = \theta = \frac{2(1-q)}{q(2-q)} \Leftrightarrow q = 1 - \frac{\theta}{(1+\theta^2)^{1/2} + 1}$$

→  $\theta$  adjustable by mutative self adaptation

→ get  $q$  from  $\theta$

like mutative step size control of  $\sigma$  in EA with search space  $\mathbb{R}^n$  !

n - dimensional generalization



random vector  $Z = (Z_1, Z_2, \dots, Z_n)$   
with  $Z_i = G_{1,i} - G_{2,i}$  (stoch. indep.);  
parameter  $q$  for all  $G_{1i}, G_{2i}$  equal

n - dimensional generalization

$$P\{Z_i = k\} = \frac{q}{2-q} (1-q)^{|k|}$$

$$P\{Z_1 = k_1, Z_2 = k_2, \dots, Z_n = k_n\} = \prod_{i=1}^n P\{Z_i = k_i\} =$$

$$\left(\frac{q}{2-q}\right)^n \prod_{i=1}^n (1-q)^{|k_i|} = \left(\frac{q}{2-q}\right)^n (1-q)^{\sum_{i=1}^n |k_i|}$$

$$= \left(\frac{q}{2-q}\right)^n (1-q)^{\|k\|_1}$$

⇒ n-dimensional distribution is symmetric w.r.t.  $\ell_1$  norm!

⇒ all random vectors with same step length have same probability!

How to control  $E[\|Z\|_1]$  ?

$$E[\|Z\|_1] = E\left[\sum_{i=1}^n |Z_i|\right] \stackrel{\text{linearity of } E[\cdot]}{=} \sum_{i=1}^n E[|Z_i|] \stackrel{\text{identical distributions for } Z_i}{=} n \cdot E[|Z_1|]$$

by def.
linearity of  $E[\cdot]$ 
identical distributions for  $Z_i$

$$\underbrace{n \cdot E[|Z_1|]}_{= \theta} = n \cdot \frac{2(1-q)}{q(2-q)} \Leftrightarrow q = 1 - \frac{\theta/n}{(1 + (\theta/n)^2)^{1/2} + 1}$$

self-adaptation
calculate from  $\theta$

Algorithm:

- individual :  $(x, \theta) \in \mathbb{Z}^n \times \mathbb{R}_+$
- mutation :  $\theta^{(t+1)} = \theta^{(t)} \cdot \exp(N)$ ,  $N \sim N(0, 1/n)$ .  
 if  $\theta^{(t+1)} < 1$  then  $\theta_{t+1} = 1$   
 calculate new  $q$  for  $G_i$  from  $\theta_{t+1}$   
 $\forall j = 1, \dots, n : X_j^{(t+1)} = X_j^{(t)} + (G_{1,j} - G_{2,j})$
- recombination : discrete (uniform crossover)
- selection :  $(\mu, \lambda)$ -selection

(Rudolph, PPSN 1994)

ad 2) design guidelines for variation operators **in practice**

continuous search space  $X = \mathbb{R}^n$

- a) reachability
- b) unbiasedness
- c) control

leads to CMA-ES