Computational Intelligence

Winter Term 2012/13

Prof. Dr. Günter Rudolph

Lehrstuhl für Algorithm Engineering (LS 11)

Fakultät für Informatik

TU Dortmund

<u>Three tasks:</u>

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 $f: X \to \mathbb{R}^d$

scenario: no standard algorithm for search space X available
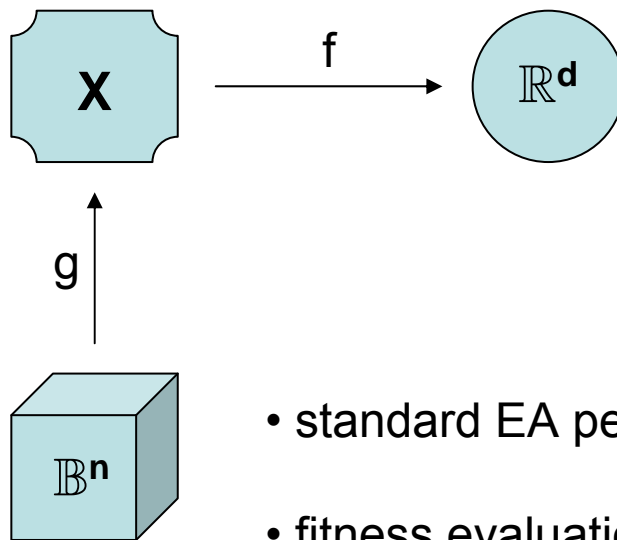
- standard EA performs variation on binary strings $b \in \mathbb{B}^n$

- fitness evaluation of individual b via $(f \circ g)(b) = f(g(b))$

  where $g: \mathbb{B}^n \to X$ is genotype-phenotype mapping

- selection operation independent from representation

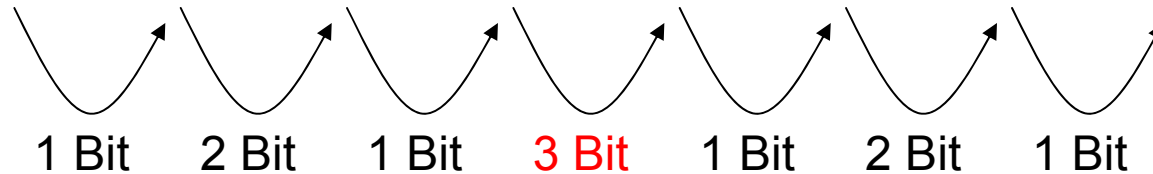**Genotype-Phenotype-Mapping** $\mathbb{B}^n \to [L, R] \subset \mathbb{R}$

● 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	← genotype
0	1	2	3	4	5	6	7	← phenotype

1 Bit 2 Bit 1 Bit 3 Bit 1 Bit 2 Bit 1 Bit

↑
Hamming cliff

L = 0, R = 7

n = 3

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

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

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

$\oplus$ = 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 …

$\Rightarrow$ small changes in phenotype „lead to" small changes in genotype

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

$\Rightarrow$ small changes in genotype lead to small changes in phenotype!

   but: 1-Bit-change: `000 → 100` $\Rightarrow$ ☹

Genotype-Phenotype-Mapping $\mathbb{B}^n \to \mathbb{P}^{\log(n)}$ (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) $g: \mathbb{B}^n \to X$ can be computed efficiently (otherwise it is senseless)

2) $g: \mathbb{B}^n \to X$ is surjective (otherwise we might miss the optimal solution)

3) $g: \mathbb{B}^n \to X$ *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 ...

technische universität
dortmund

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: <u>maximum entropy principle</u>

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): $x,y \in \mathbb{B}^n$ } = δ > 0 we are done.

technische universität
dortmund

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***. ∎

technische universität
dortmund

Knowledge available:

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

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

$\Rightarrow$ leads to nonlinear constrained optimization problem:

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

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

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:

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

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

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

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

uniform distribution

Knowledge available:

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

$\Rightarrow$ leads to nonlinear constrained optimization problem:

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

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

<u>solution:</u> 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:

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

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

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

(continued on next slide)

technische universität
dortmund

$$\Rightarrow \quad e^{a-1} = \frac{1}{\sum\limits_{k=1}^{n} (e^b)^k} \qquad\qquad \Rightarrow \quad p_k = e^{a-1+b\,k} = \frac{(e^b)^k}{\sum\limits_{i=1}^{n} (e^b)^i}$$
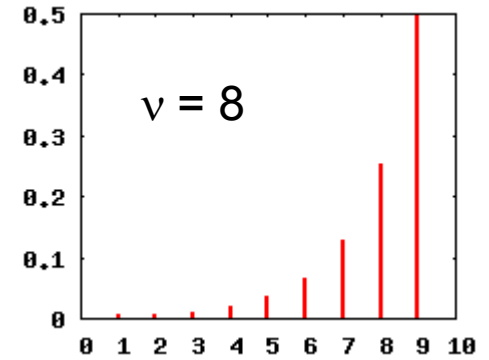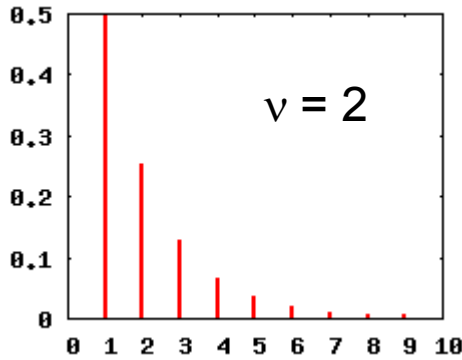
$\Rightarrow$ **discrete Boltzmann distribution** $\qquad p_k = \dfrac{q^k}{\sum\limits_{i=1}^{n} q^i} \qquad (q = e^b)$

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

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

Boltzmann distribution

(n = 9)

specializes to uniform

distribution if $\nu = 5$

(as expected)

$\downarrow$
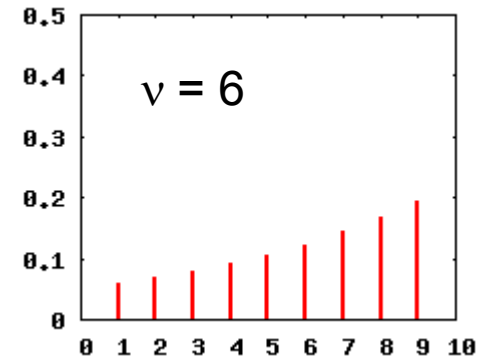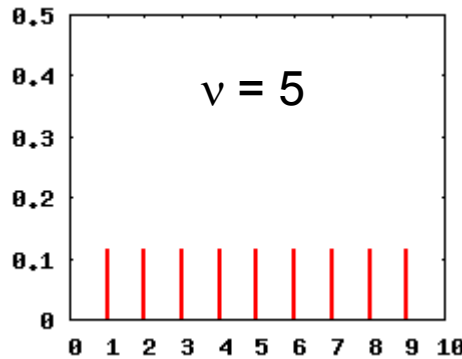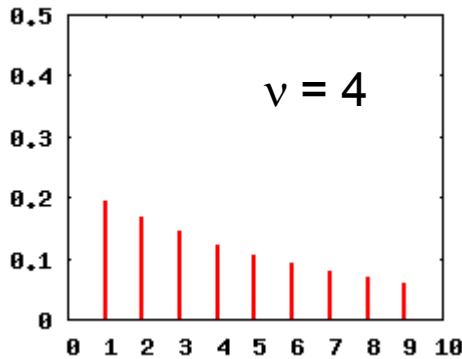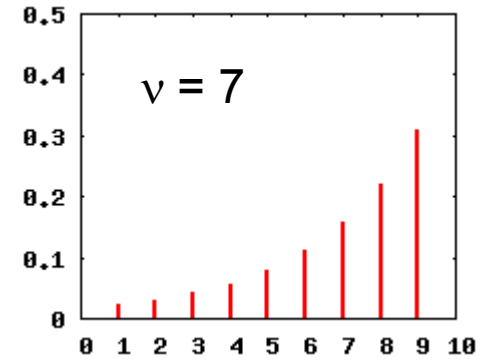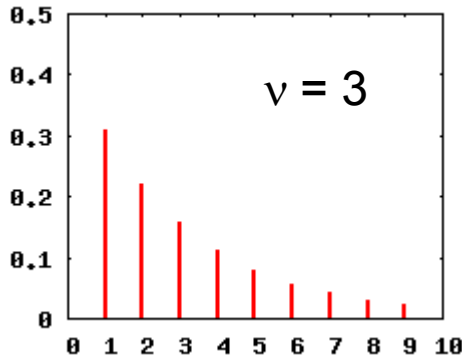
Knowledge available:

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

$\Rightarrow$ leads to nonlinear constrained optimization problem:

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

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

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

but very complicated analytically, if possible at all

$\Rightarrow$ 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:

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

$$
\begin{array}{lrcl}
\text{II} - \text{I:} & p_2 + 2\,p_3 & = & 1 \\
\text{I} - \text{III:} & p_2 & = & 1 - \eta^2
\end{array}
\left.\vphantom{\begin{array}{c}1\\1\end{array}}\right\}
\quad
\begin{array}{c}
p_1 = \dfrac{\eta^2}{2} \\[2mm]
\uparrow \ \text{insertion in III.} \\[2mm]
p_3 = \dfrac{\eta^2}{2}
\end{array}
$$

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

$\eta^2 = \frac{1}{4}$ — unimodal

$\eta^2 = \frac{2}{3}$ — uniform
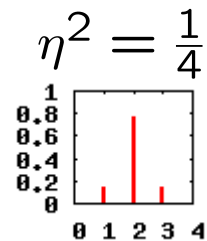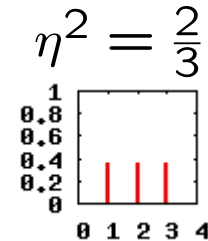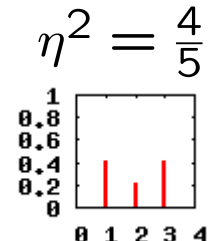
$\eta^2 = \frac{4}{5}$ — bimodal

Knowledge available:

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

⇒ leads to <u>infinite-dimensional</u> nonlinear constrained optimization problem:

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

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

<u>solution:</u> via Lagrange (find stationary point of Lagrangian function)

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

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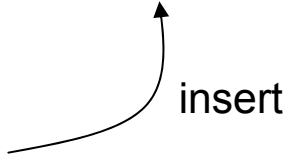
partial derivatives:

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

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

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

technische universität
dortmund

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

insert

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

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

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

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

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

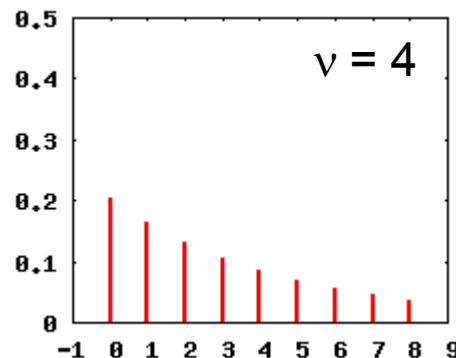$\Rightarrow$ $q \;=\; \dfrac{\nu}{\nu+1} \;=\; 1 - \dfrac{1}{\nu+1}$

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

technische universität
dortmund

geometrical distribution

with $E[\,x\,] = \nu$

p_k only shown
for $k = 0, 1, \ldots, 8$
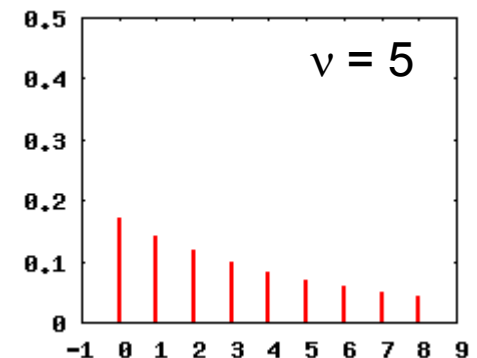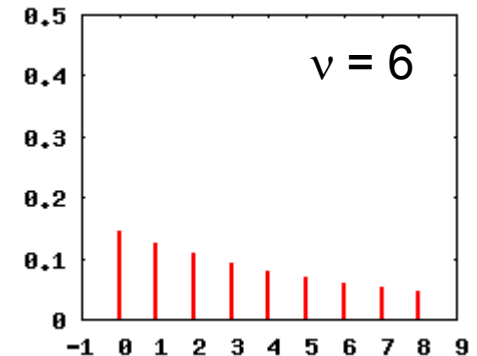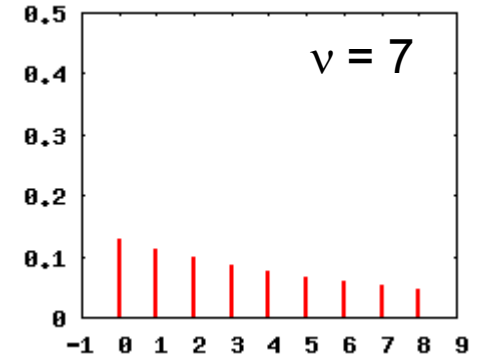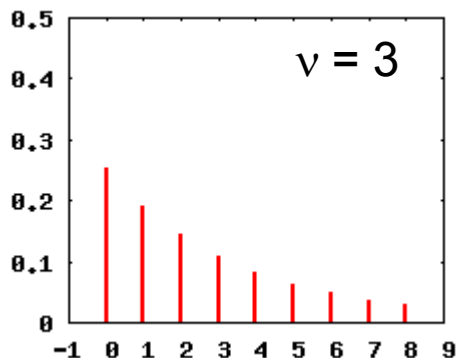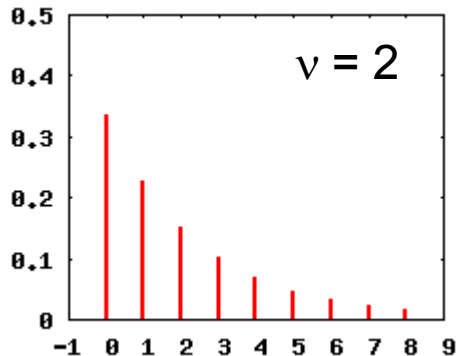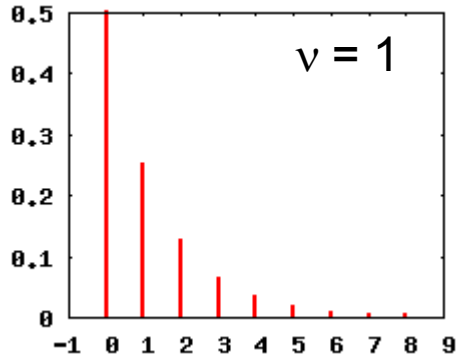
Overview:

support { 1, 2, …, n } $\quad\Rightarrow$ *discrete uniform* distribution

and require E[X] = θ $\quad\Rightarrow$ *Boltzmann* distribution

and require V[X] = η^2 $\quad\Rightarrow$ N.N. (**not** Binomial distribution)

support $\mathbb{N}$ $\quad\Rightarrow$ not defined!

and require E[X] = θ $\quad\Rightarrow$ *geometrical* distribution

and require V[X] = η^2 $\quad\Rightarrow$?

support $\mathbb{Z}$ $\quad\Rightarrow$ not defined!

and require E[|X|] = θ $\quad\Rightarrow$ *bi-geometrical* distribution (*discrete Laplace* distr.)

and require E[|X|^2] = η^2 $\quad\Rightarrow$ N.N. (*discrete Gaussian* distr.)

technische universität
dortmund

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 $N(\theta, C)$

expectation vector $\in \mathbb{R}^n$ covariance matrix $\in \mathbb{R}^{n,n}$

positive definite:
$\forall x \neq 0 : x'Cx > 0$

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

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

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

ad c) control variability by parameter

$\rightarrow$ formulate as constraint of maximum entropy distribution

technische universität
dortmund

ad 2) **design guidelines for variation operators in practice** $X = \mathbb{Z}^n$

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

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

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

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

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

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

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

result:

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

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

symmetric w.r.t. 0, unimodal, spread manageable by *q* and has max. entropy ∎
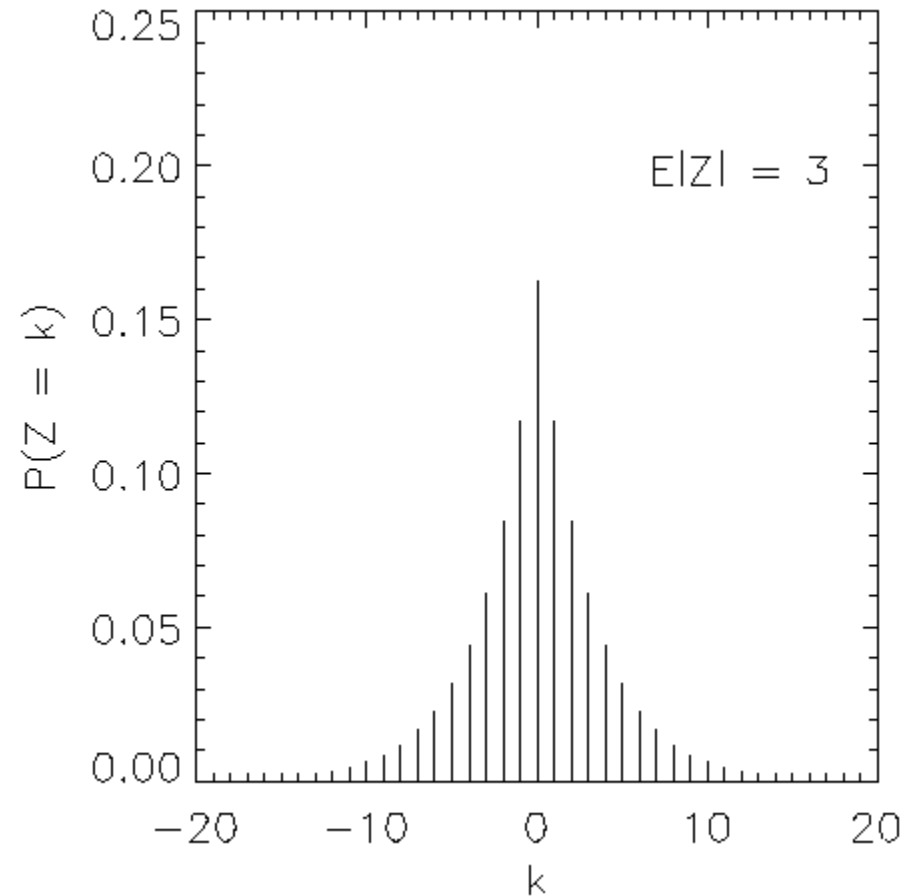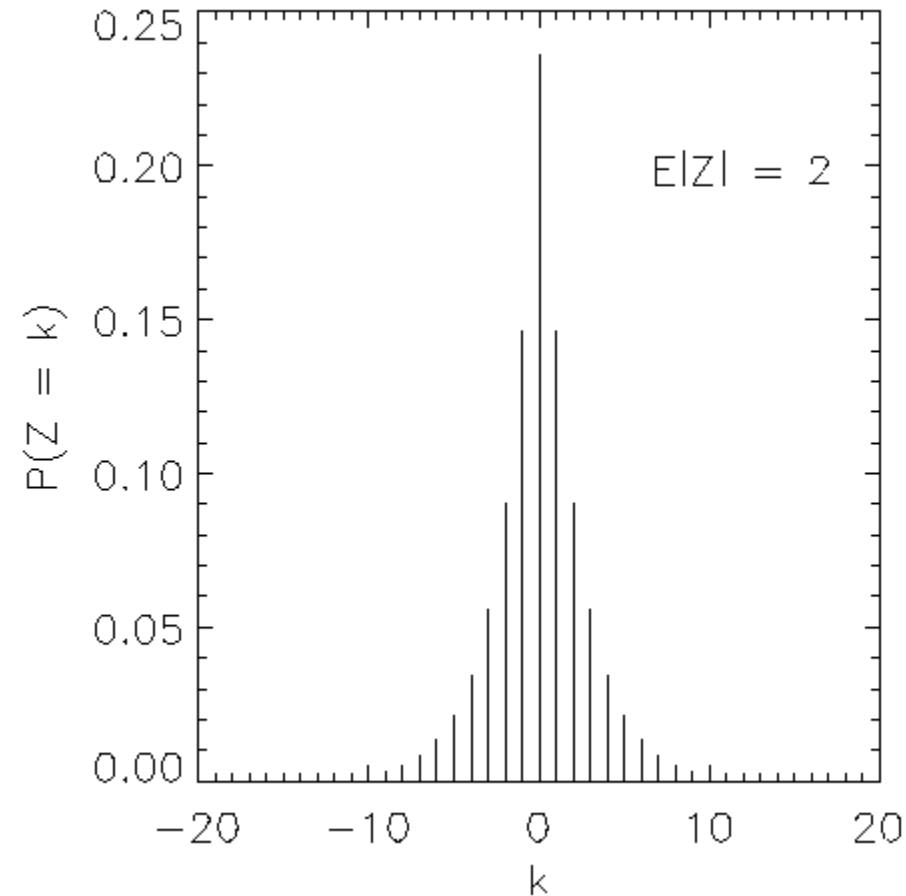
generation of pseudo random numbers: $\boxed{Z = G_1 - G_2}$

where

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

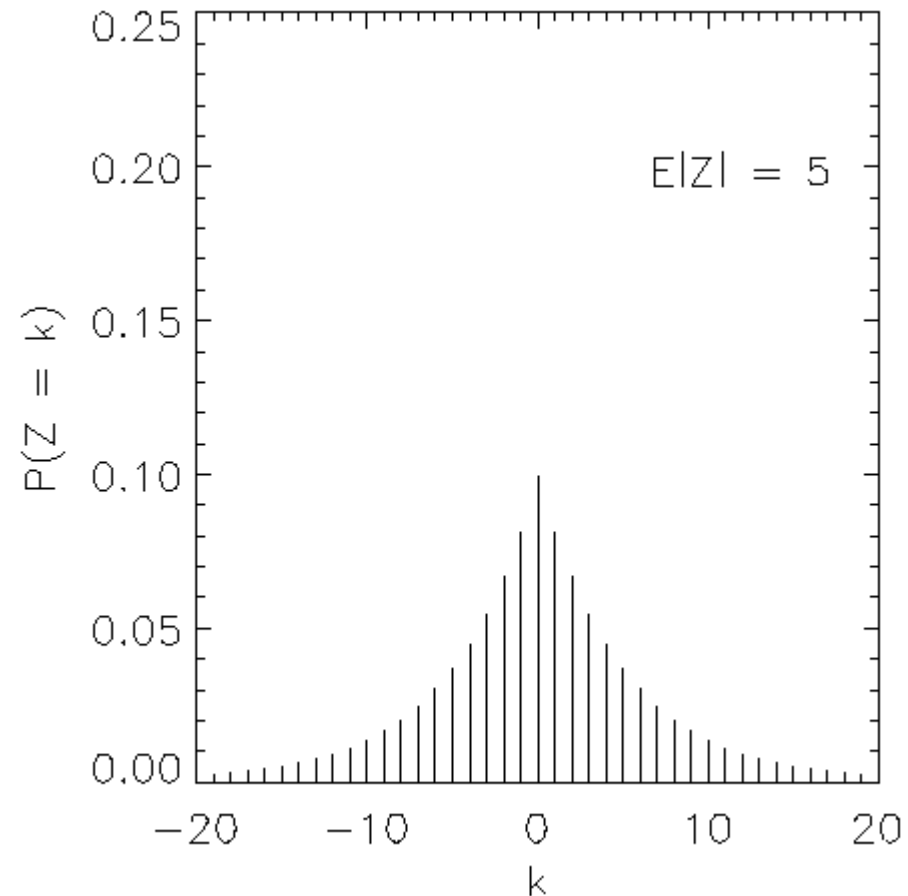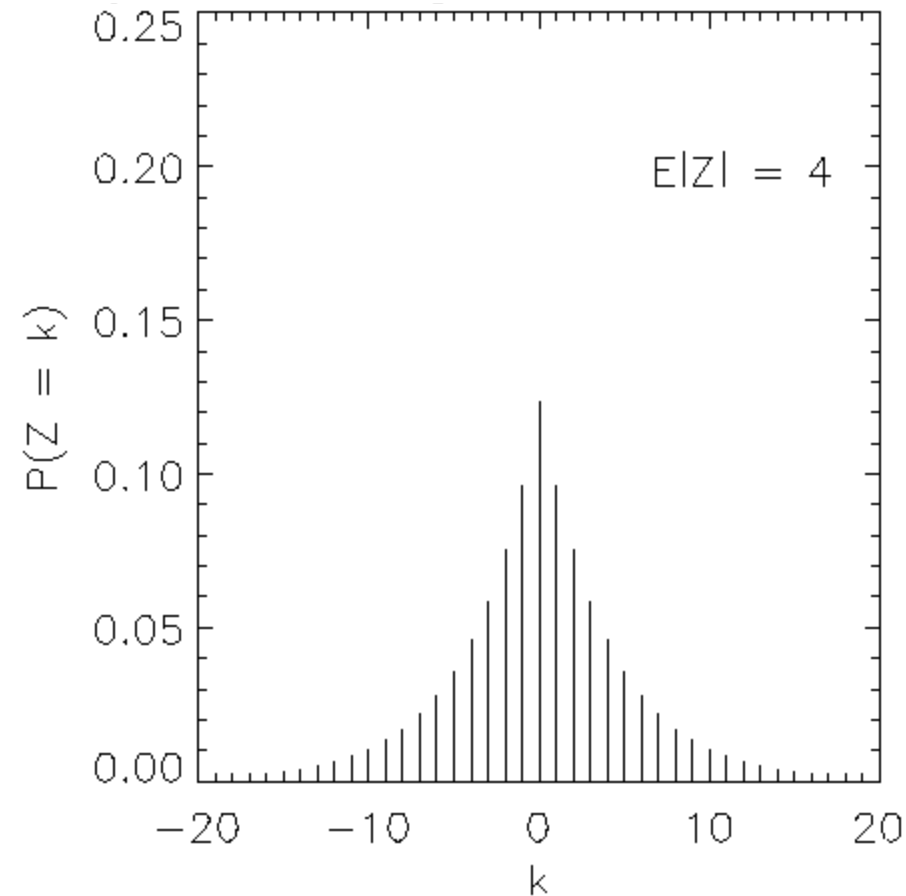stochastic
independent!

probability distributions for different mean step sizes E|Z| = θ

technische universität
dortmund

probability distributions for different mean step sizes E|Z| = θ

How to control the spread?

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

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

$\longrightarrow$ make mean step size $E[\,|Z|\,]$ adjustable!

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

$\in \mathbb{R}_+$

$\in$ (0,1)

$\rightarrow \theta$ adjustable by mutative self adaptation

$\rightarrow$ get q from θ

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

n - dimensional generalization
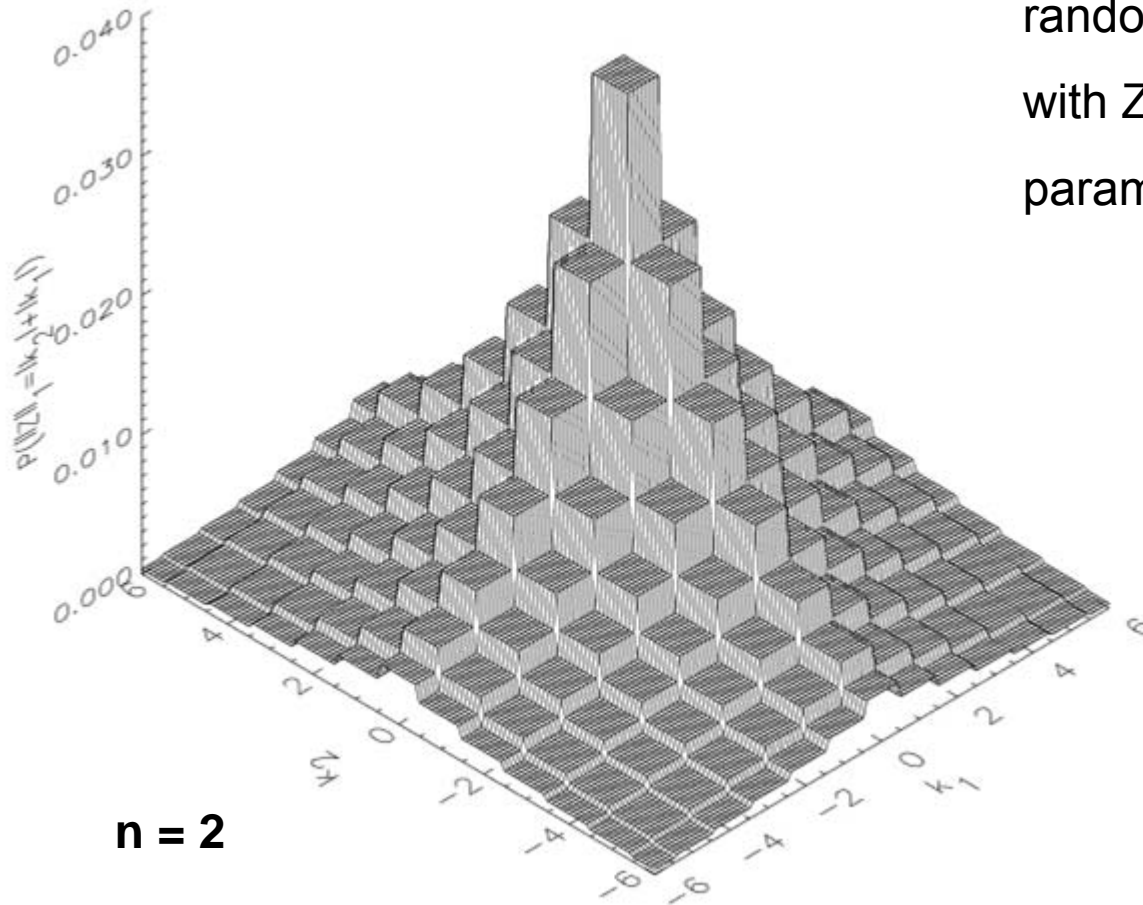
n = 2

random vector $Z = (Z_1, Z_2, ... Z_n)$

with $Z_i = G_{1,i} - G_{2,i}$ (stoch. indep.);

parameter q for all G_{1i}, G_{2i} equal

technische universität
dortmund

n - dimensional generalization

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

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

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

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

$\Rightarrow$ n-dimensional distribution is symmetric w.r.t. ℓ_1 norm!

$\Rightarrow$ all random vectors with same step length have same probability!

technische universität
dortmund

How to control E[ǁ Z ǁ₁] ?

$$E[\,\|Z\|_1\,] \;=\; E\left[\sum_{i=1}^{n}|Z_i|\right] \;=\; \sum_{i=1}^{n}E[|Z_i|] \;=\; n\cdot E[|Z_1|]$$

by def.   linearity of E[·]   identical distributions for Z_i

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

self-adaptation   calculate from θ

Algorithm:

individual : $(x, \theta) \in \mathbb{Z}^n \times \mathbb{R}_+$

mutation : $\theta^{(t+1)} = \theta^{(t)} \cdot \exp(N), \quad N \sim N(0, 1/n).$

if $\theta^{(t+1)} < 1$ then $\theta_{t+1} = 1$

calculate new q for G_i from θ_{t+1}

$\forall j = 1, \ldots, n : X_j^{(t+1)} = X_j^{(t)} + (G_{1,j} - G_{2,j})$

recombination : discrete (uniform crossover)

selection : (μ, λ)-selection

(Rudolph, PPSN 1994)

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

<u>continuous search space</u> X = $\mathbb{R}^n$

a) reachability

b) unbiasedness

c) control

leads to CMA-ES