Wintersemester 2006/07

Fundamente der Computational Intelligence

(Vorlesung)

Prof. Dr. Günter Rudolph

Fachbereich Informatik

Lehrstuhl für Algorithm Engineering

Inhalt

- Konvergenz
- EA im $\mathbb{R}^n$

Typische theoretische Fragestellungen:

Optima erreichbar?

nein

Bis wohin kommt man?

Mit welcher W'keit erreichbar?

ja

Konvergenz?

nein

Zeit bis zum ersten „Treffer"?

ja

Zeit bis Konvergenz?

Zum Begriff „Konvergenz"

$D_k = | f(X_k) - f^* | \geq 0$ ist eine Zufallsvariable

wir betrachten die stochastische Folge $D_0, D_1, D_2, \ldots$

Konvergiert die stochastische Folge $(D_k)_{k \geq 0}$ gegen 0?

Wenn ja, dann offensichtlich „Konvergenz zum Optimum"!

Es existieren viele Arten von **stochastischer Konvergenz**!

→ deshalb hier nur die gebräuchlichsten ...

Notation: $\mathcal{P}^{(t)}$ = Population zum Zeitpunkt $t \geq 0$, $f_b(\mathcal{P}(t)) = \min\{f(x) : x \in \mathcal{P}(t)\}$

Definition

Sei $D_t = |f_b(\mathcal{P}(t)) - f^*| \geq 0$. Wir sagen: Der EA

(a) ***konvergiert vollständig*** zum Optimum, wenn $\forall\ \varepsilon > 0$

$$\lim_{t \to \infty} \sum_{k=1}^{t} P\{D_k > \varepsilon\} < \infty\,;$$

(b) ***konvergiert fast sicher*** oder ***mit W'keit 1*** zum Optimum, wenn

$$P\{\lim_{t \to \infty} D_k = 0\} = 1\,;$$

(c) ***konvergiert in Wahrscheinlichkeit*** zum Optimum, wenn $\forall\ \varepsilon > 0$

$$\lim_{t \to \infty} P\{D_t > \varepsilon\} = 0\,;$$

(a) ***konvergiert im Mittel*** zum Optimum, wenn $\forall\ \varepsilon > 0$

$$\lim_{t \to \infty} E\{D_t\} = 0\,.$$

■

Lemma

- (a) $\Rightarrow$ (b) $\Rightarrow$ (c).
- (d) $\Rightarrow$ (c).
- Falls $\exists\, K < \infty : \forall\, t \geq 0 : D_t \leq K$, dann (d) $\Leftrightarrow$ (c).
- Falls $(D_t)_{t \geq 0}$ stochastisch unabhängige Folge, dann (a) $\Leftrightarrow$ (b). ■

Typische Vorgehensweise:

1. Zeige Konvergenz in W'keit (c). Meistens einfach!
2. Zeige, dass Konvergenz schnell genug (a). Dann folgt auch (b).
3. Folge nach oben beschränkt? Dann folgt (d).

Beispiele: Sei $(X_k)_{k\geq 1}$ Folge unabhängiger Zufallsvariablen.

Verteilung: $P\{X_k = 0\} = 1 - \frac{1}{k}$ $\quad P\{X_k = 1\} = \frac{1}{k}$

1. $P\{X_k > \varepsilon\} = P\{X_k = 1\} = \frac{1}{k} \to 0 \text{ für } t \to \infty$

   $\Rightarrow$ Konvergenz in W'keit (c)

2. $\sum_{k=1}^{\infty} P\{X_k > \varepsilon\} = \sum_{k=1}^{\infty} P\{X_k = 1\} = \sum_{k=1}^{\infty} \frac{1}{k} = \infty$

   $\Rightarrow$ Konvergenz nicht schnell genug! Also <u>keine</u> vollständige Konvergenz!

3. Es gilt: $\forall k \geq 0 : 0 \leq X_k \leq 1.$ Also: Folge beschränkt mit K=1.

   Wg. Konvergenz in W'keit (c) und Beschränktheit $\Rightarrow$ Konvergenz im Mittel (d)

Beispiele: Sei $(X_k)_{k \geq 1}$ Folge unabhängiger Zufallsvariablen.

Verteilung:		(a)	(c)	(d)
$P\{X_k = 0\} = 1 - \frac{1}{k}$	$P\{X_k = 1\} = \frac{1}{k}$	(–)	(+)	(+)
$P\{X_k = 0\} = 1 - \frac{1}{k^2}$	$P\{X_k = 1\} = \frac{1}{k^2}$	(+)	(+)	(+)
$P\{X_k = 0\} = 1 - \frac{1}{k}$	$P\{X_k = k\} = \frac{1}{k}$	(–)	(+)	(–)
$P\{X_k = 0\} = 1 - \frac{1}{k^2}$	$P\{X_k = k\} = \frac{1}{k^2}$	(+)	(+)	(+)
$P\{X_k = 0\} = 1 - \frac{1}{k}$	$P\{X_k = k^2\} = \frac{1}{k}$	(–)	(+)	(–)
$P\{X_k = 0\} = 1 - \frac{1}{k^2}$	$P\{X_k = k^2\} = \frac{1}{k^2}$	(+)	(+)	(–)

Satz:

Sei $D_k = | f(x_k) – f^* |$ für $k \geq 0$ durch einen (1+1)-EA generiert,
$S^* = \{ x^* \in S : f(x^*) = f^* \}$ die Menge der optimalen Lösungen und
$P_m(x, S^*)$ die W'keit, von $x \in S$ durch eine Mutation nach S^* zu gelangen.

Wenn für jedes $x \in S \setminus S^*$ gilt $P_m(x, S^*) \geq \delta > 0$, dann $D_k \to 0$ vollständig.

Beweis:

Für den (1+1)-EA gilt: $P(x, S^*) = 1$ für $x \in S^*$ wg. Selektion des Besseren.

Es reicht also zu zeigen, dass der EA durch Mutation S^* sicher erreicht:

Erfolg in 1. Iteration: $P_m(x, S^*) \geq \delta$.
Kein Erfolg in 1. Iteration $\leq 1 - \delta$.
Kein Erfolg in k. Iteration $\leq (1- \delta)^k$.
$\Rightarrow$ Erfolg in k. Iteration $\geq 1 - (1- \delta)^k \to 1$ für $k \to \infty$.

Wg. $P\{ D_k > \varepsilon \} \leq (1- \delta)^k \to 0$ folgt Konvergenz in W'keit und da $\displaystyle\sum_{k=0}^{\infty} (1-\delta)^k < \infty$
folgt sogar vollständige Konvergenz.

Außerdem: $\forall\, k \geq 0: 0 \leq D_k \leq D_0 < \infty$, also auch Konvergenz im Mittel. ■

Die Bedingung: $\forall\, x \in S \setminus S^*$ gilt $P_m(x, S^*) \geq \delta > 0$

- ist hinreichend, aber nicht notwendig und
- etwa für die globale Mutation mit Mutationsw'keit $p \in (0,1)$ erfüllt, da

$$P_m(x, y) = p^{H(x,y)}\,(1-p)^{n-H(x,y)} > 0 \text{ für alle } x, y \text{ in } S,$$

wobei n = Dimension und H(x,y) der Hamming-Abstand zwischen x und y.

Achtung:

Die gleiche Konvergenzaussage erhält man für reine Zufallssuche:

```
Wähle X_0 ∈ S, setze k = 0.
repeat
  Wähle Y_k zufällig gleichverteilt aus S.
  Falls f(Y_k) < f(X_k)  dann  X_{k+1} = Y_k
                         sonst X_{k+1} = X_k
until Terminierung
```

Folgerung:

Solche Aussagen sind von geringer Bedeutung, weil δ astronomisch klein!

Beispiel:

alle Bits müssen invertiert werden $\Rightarrow p^n = n^{-n} = \delta$

Übergang zum Optimum erfolgt mit W'keit δ (geometrische Verteilung)

$\Rightarrow$ mittlere Zeit bis zum Eintreffen des Ereignisses $1/\delta = n^n$

Sei n = 20 und wir haben einen Tera-Hertz-Rechner (10^{12} Iterationen/Sekunde).

Dann warten wir im Mittel $20^{20} / 10^{12}$ Sekunden = $2^{20} \times 10^8$ Sekunden auf Lösung,

also über 3,325 Millionen Jahre!

$\Rightarrow$ Das ist mathematisch gesehen zwar endlich, praktisch jedoch **unendlich**!

Also:

Stochastische Konvergenzaussagen (auch für Simulated Annealing etc.) sind mit Vorsicht zu genießen!

Aber:

Negative Aussage (keine Konvergenz bzw. kein sicheres Besuchen des Optimums) ist schon von Bedeutung!

Beispiel:

Sicheres Auffinden des Optimums in Zeit t mit W'keit $\gamma > 0$.

$\Rightarrow$ Multistart! W'keit des Mißerfolgs sinkt exponentiell schnell!

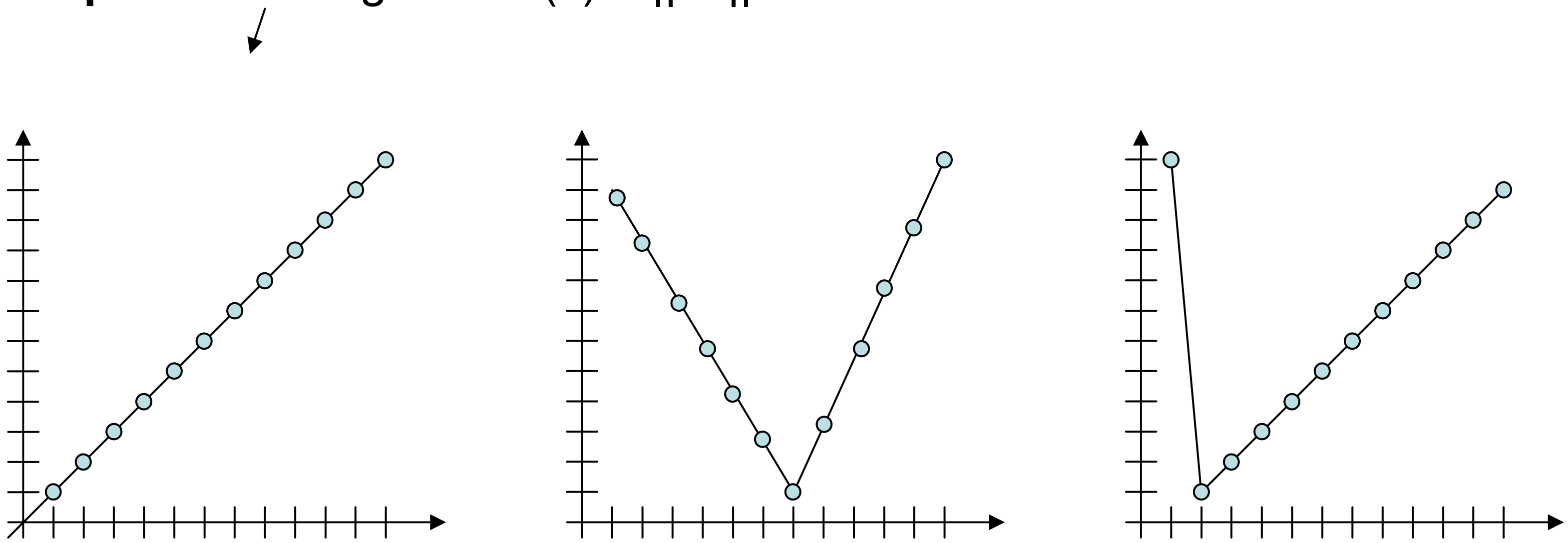
Allerdings spielt Größe von γ eine Rolle!

$$P \{ \text{mindestens 1 Erfolg bis zum k-ten Versuch} \} = 1 - (1 - \gamma)^k$$

"Functions of Unication" (Goldberg/Deb 1993):

Es existiert g: {0,1, … n} → $\mathbb{R}$ mit f(x) = g(||x||), wobei || . || Hamming-Norm.

Beispiel: Counting Ones f(x) = || x ||



γ = 1 γ = 0.5 γ klein

NIAH
(needle-in-a-haystick)

Schematischer Ablauf eines Evolutionären Algorithmus

	GA	ES	EP
`Initialisierung der Population`	x	x	x
`repeat`			
`  Reproduktionsselektion`	x	-	-
`  Rekombination`	x	x	-
`  Mutation`	x	x	x
`  Überlebensselektion`	-	x	x
`until Abbruch`			

Heute: Keine Unterscheidung zwischen GA / ES / EP, da Übergänge fließend

EA im $\mathbb{R}^n$

Selektion: unabhängig von Representation

Rekombination: k-Punkt-Crossover, Uniform Crossover
funktioniert auf allen Produkträumen!

zusätzlich:
Intermediäre Rekombination $z_i = \alpha_i \, x_i + (1-\alpha_i) \, y_i$, $\alpha_i \in (0,1)$

$\alpha_i \equiv \alpha$: auf Verbindungslinie zwischen x und y
sonst innerhalb des durch x und y definierten Hyperrechtecks

auch: Varianten mit > 2 Eltern

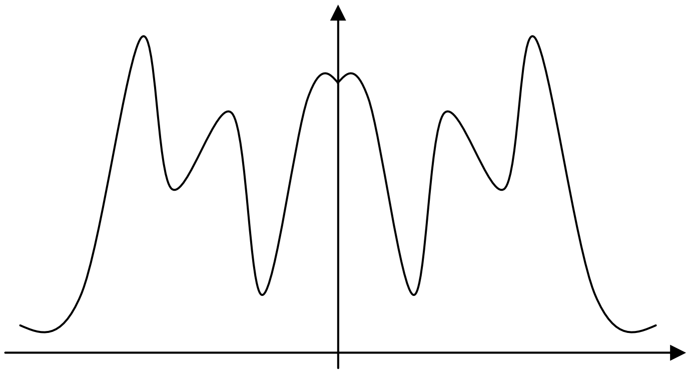
Mutation: additiv via Normalverteilung N(0, C), C Kovarianzmatrix
häufig: $C = \sigma^2 \cdot I_n$ oder individuelle σ_i

auch: Cauchyverteilung („Levy flights“) oder Mixturen

notwendig: Anpassung von C bzw. σ_i während Suche!

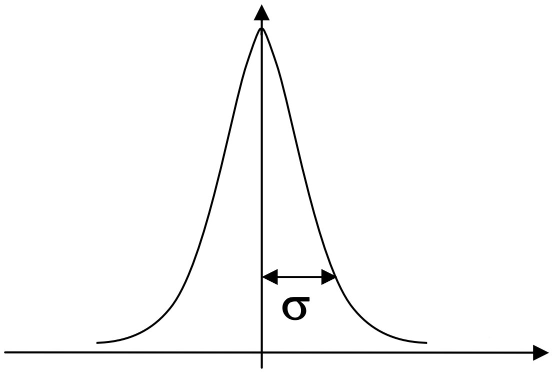
Forderungen an Such- / Mutationsverteilung von m_k

1. Keine Richtung ohne Grund bevorzugen → Symmetrie um 0
2. Kleine Änderungen wahrscheinlicher als große → Unimodal mit Modus 0
3. Steuerbar: Größe der Umgebung, Streuung → Parametrisierbar
4. Leicht erzeugbar
5. …



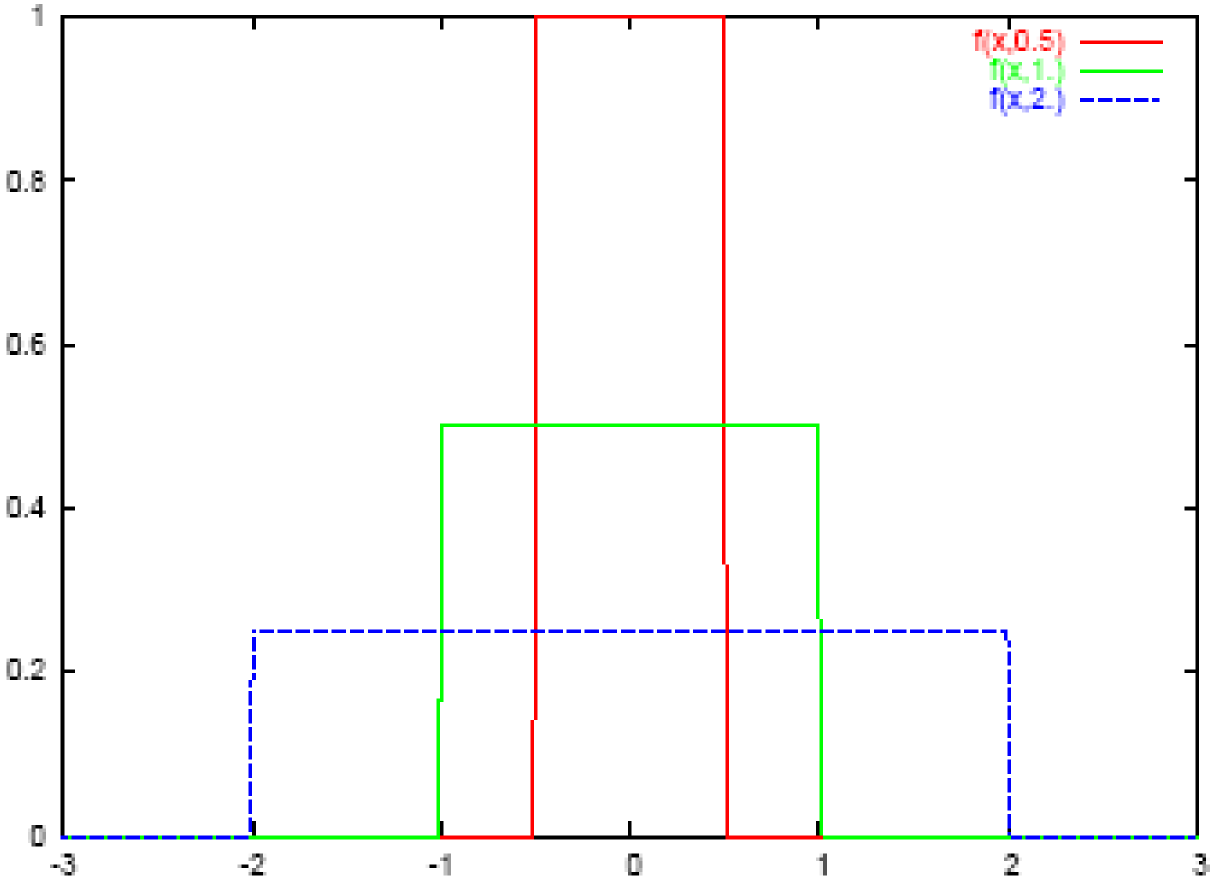
symmetrisch, multimodal

$\Rightarrow$



symmetrisch, unimodal

Gleichverteilung $$f_m(x) = \frac{1}{2\,r} \cdot 1_{(-r,r)}(x)$$

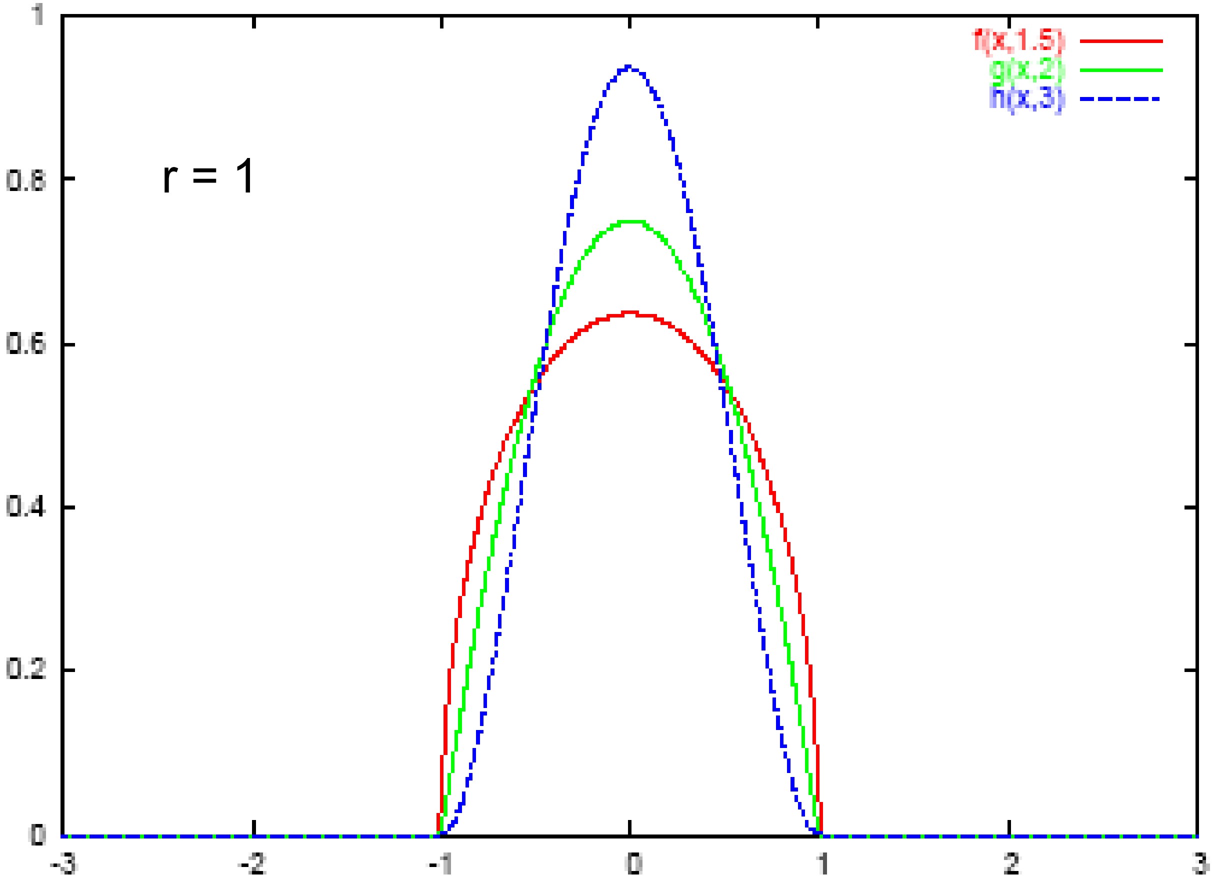


- symmetrisch
- unimodal
- steuerbar → r
- leicht erzeugbar:

  m = r (2 u – 1)

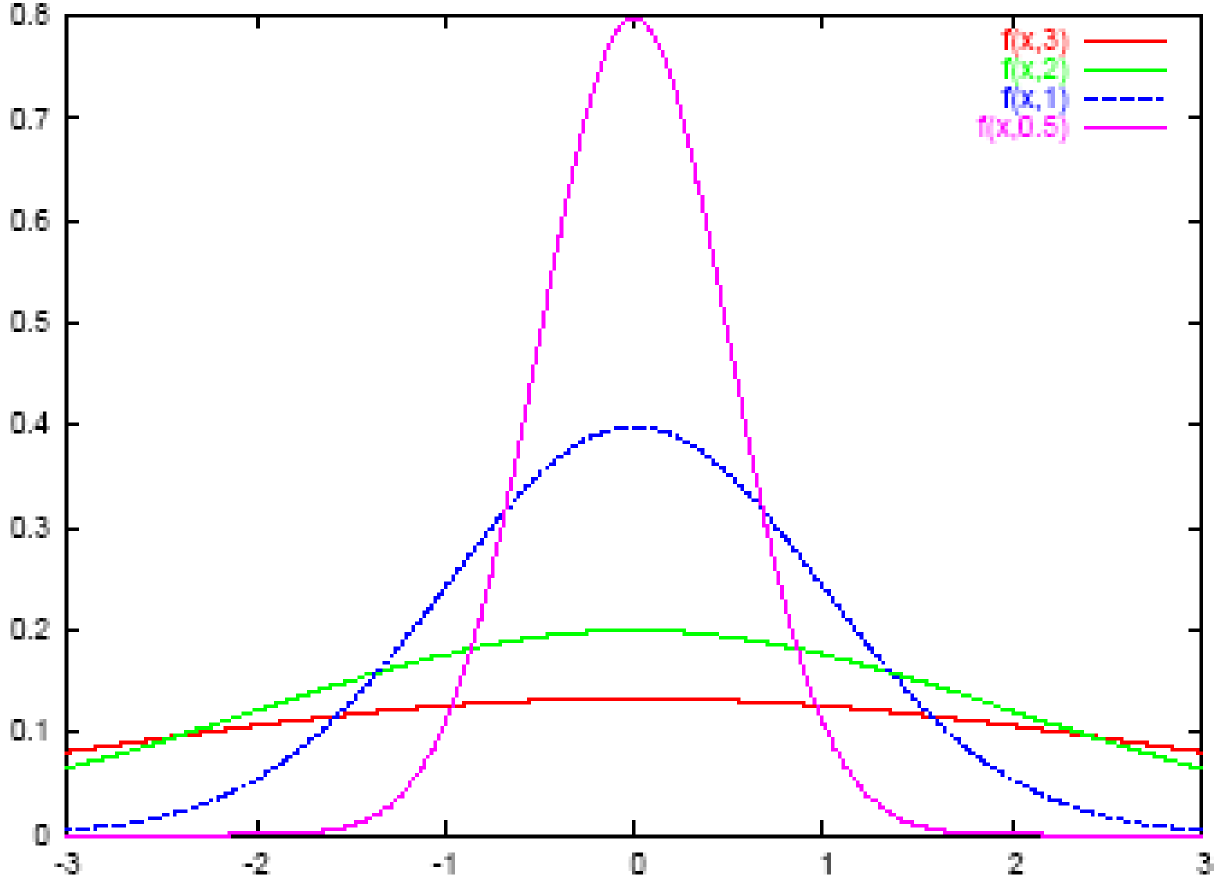
  wobei u ∈ [0,1) gleichverteilt (aus Bibliothek)

Betaverteilung $$f_m(x) = \frac{r^{1-2\,p}}{\sqrt{\pi}} \cdot \frac{\Gamma(p+\frac{1}{2})}{\Gamma(p)}(1-x^2)^{p-1} \cdot 1_{(-r,r)}(x)$$



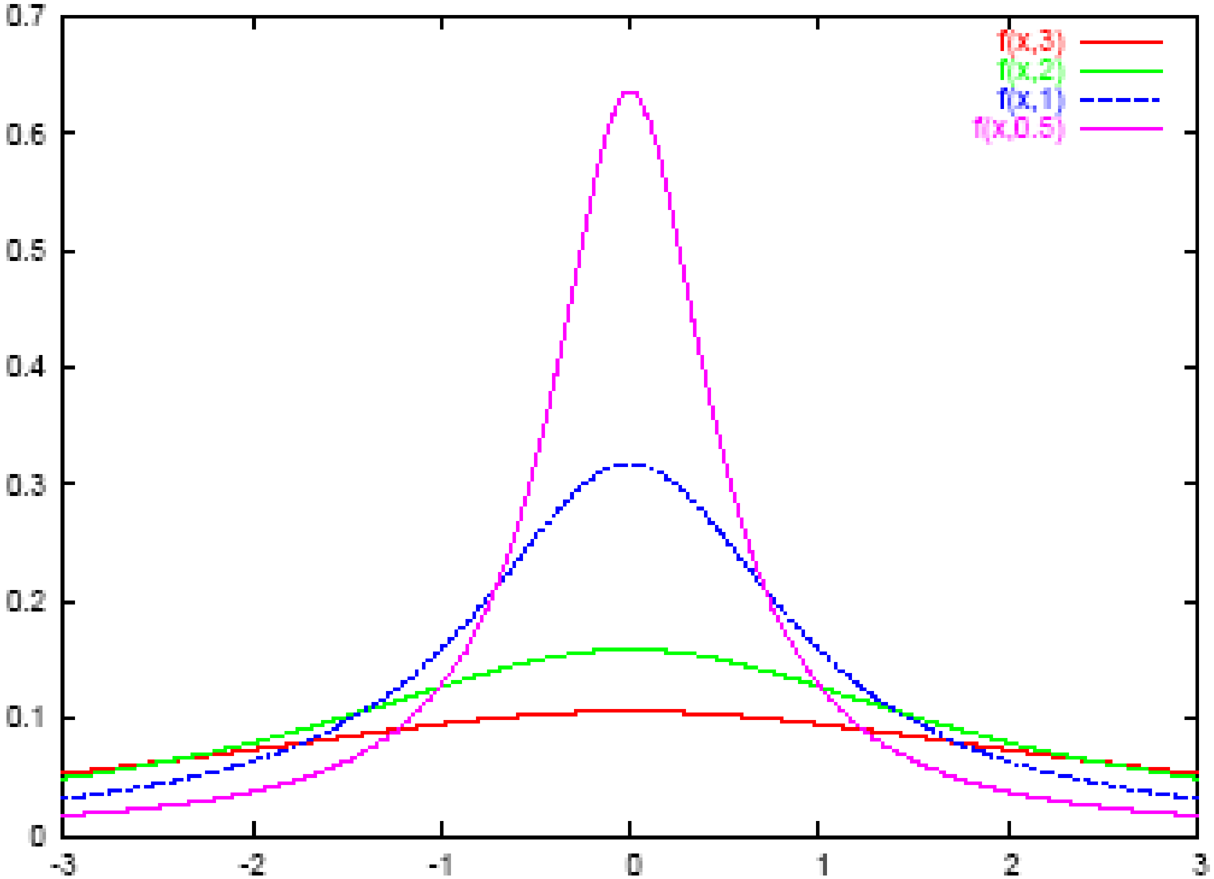
- symmetrisch
- unimodal
- steuerbar → r, p
- leicht erzeugbar (Bibliothek)

Normalverteilung $$f_m(x) = \frac{1}{\sigma\sqrt{2\,\pi}} \exp\left(-\frac{x^2}{2\,\sigma^2}\right)$$



- symmetrisch
- unimodal
- steuerbar → σ
- nicht ganz so leicht erzeugbar (Bibliothek)

Cauchyverteilung $$f_m(x) = \frac{1}{c\,\pi} \cdot \frac{1}{1 + (1 + \frac{x}{c})^2}$$



- symmetrisch
- unimodal
- steuerbar → c
- leicht erzeugbar (Bibliothek)

Besonderheit:

unendliche Varianz

Höherdimensionale Suchräume: Symmetrie? Unimodalität? Steuerbarkeit?

↓

Rotationssymmetrie

Definition:

Sei T eine (n x n)-Matrix mit T‘T = I_n. (I_n: n-dim. Einheitsmatrix)

T heißt ***orthogonale Matrix*** oder ***Rotationsmatrix***. ■

Beispiel:

$$T = \begin{pmatrix} \cos\omega & -\sin\omega \\ \sin\omega & \cos\omega \end{pmatrix}$$

$$y = T'x \quad \Rightarrow \text{Vektor } x \text{ wurde um Winkel } \omega \text{ gedreht}$$

Definition:

n-dimensionaler Zufallsvektor x heißt

sphärisch symmetrisch oder ***rotationssymmetrisch***

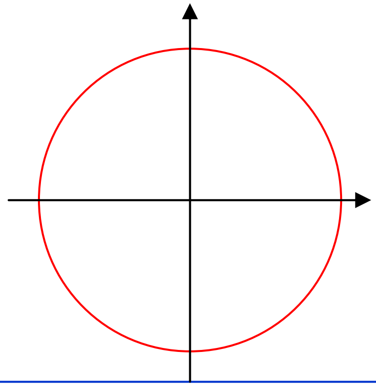
$\Leftrightarrow$ $x \stackrel{d}{=} T'x$ für jede orthogonale Matrix T. ■

$x \stackrel{d}{=} y$ bedeutet: x hat die gleiche Verteilung wie y

Beispiel: Gleichverteilung auf Kreis (Hyperkugel der Dimension n = 2)

u gleichverteilt in [0,1] $\Rightarrow$ $\omega = 2\pi u$

$$x \stackrel{d}{=} \begin{pmatrix} \cos\omega \\ \sin\omega \end{pmatrix}$$



Satz:

Zufallsvektor x rotationssymmetrisch $\Leftrightarrow x \stackrel{d}{=} r\, u^{(n)}$, wobei

r nichtnegative Zufallsvariable und

$u^{(n)}$ Zufallsvektor mit Gleichverteilung auf n-dim. Hyperkugelrand mit Radius 1. ■

Bemerkung:

r und $u^{(n)}$ sind stochastisch unabhängig, $u^{(n)} \stackrel{d}{=} \dfrac{x}{\| x \|}$

Erzeugung von rotationssymmetrischen Zufallsvektoren:

1. Wähle zufällige Richtung $u^{(n)}$
2. Wähle zufällige Schrittlänge r
3. Multiplikation: $x = r\, u^{(n)}$

Beispiel: Multivariate Normalverteilung

Zufallsvektor m erzeugbar via

1. $m = \sigma \cdot (m_1, m_2, \ldots, m_n)$,
   wobei $m_i \sim N(0,1)$ stoch. unabh., oder

2. $m = r \cdot u$, wobei $r \sim \chi_n(\sigma)$, $u \sim U(\partial S_n(1))$.

χ–Verteilung mit n Freiheitsgraden

Gleichverteilung auf Hyperkugelrand

$\partial S_n(r) = \{ x \in \mathbb{R}^n : \| x \| = r \}$ Hyperkugelrand

Beispiel: Multivariate Cauchyverteilung

Zufallsvektor m erzeugbar via

1. $m = \sigma \cdot (m_1, m_2, \ldots, m_n) / m_0$,
   wobei $m_i \sim N(0,1)$ stoch. unabh., oder

2. $m = r \cdot u$, wobei $r/n \sim F_{n,1}$, $u \sim U(\partial S_n(1))$.

F–Verteilung mit (n,1) Freiheitsgraden

Gleichverteilung auf Hyperkugelrand

Achtung:
Zufallsvektor aus n unabh. Cauchy-Zufallsvariablen <u>nicht</u> rotationssymmetrisch!

Optimierung ohne „Anpassen“ (reine Zufallssuche)

$f(x) = \| x \|^2 = x'x \to \min!$ wobei $x \in S_n(r) = \{ x \in \mathbb{R}^n : \| x \| \le r \}$

Z_k ist gleichverteilt in $S_n(r)$

$X_{k+1} = Z_k$ falls $f(Z_k) < f(X_k)$, sonst $X_{k+1} = X_k$

$\Rightarrow V_k = \min \{ f(Z_1), f(Z_2), \ldots, f(Z_k) \}$ bester ZF-Wert bis Iteration k

$$P\{ \| Z \| \le x \} = P\{ Z \in S_n(x) \} = \mathrm{Vol}(S_n(x)) / \mathrm{Vol}(S_n(r)) = (x / r)^n , \quad 0 \le x \le r$$

$$P\{ \| Z \|^2 \le x \} = P\{ \| Z \| \le x^{1/2} \} = x^{n/2} / r^n , \quad 0 \le x \le r^2$$

$$P\{ V_k \le x \} = 1 - (1 - P\{ \| Z \|^2 \le x \})^k = 1 - (1 - x^{n/2} / r^n)^k$$

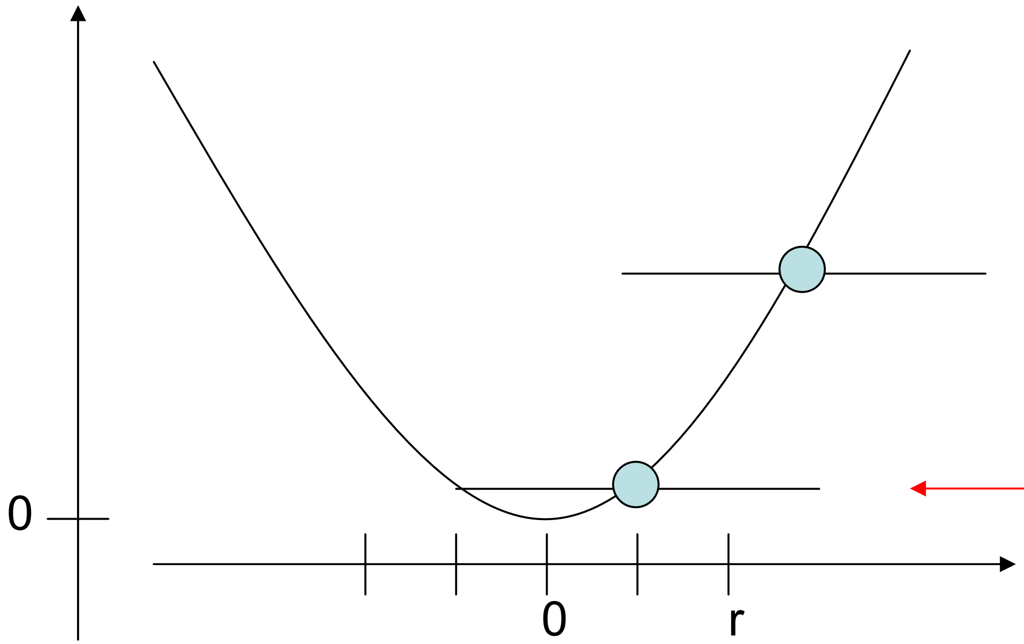
$E[V_k] \to r^2 \, \Gamma(1 + 2/n) \, k^{-2/n}$ für große k

Optimierung ohne „Anpassen“ (lokal gleichverteilt)

$f(x) = \| x \|^2 = x'x \to \min!$ wobei $x \in S_n(r) = \{ x \in \mathbb{R}^n : \| x \| \leq r \}$

Z_k ist gleichverteilt in $[-r, r]$, $n = 1$

$X_{k+1} = X_k + Z_k$ falls $f(X_k + Z_k) < f(X_k)$, sonst $X_{k+1} = X_k$



ohne Anpassung:

$D_k \sim \mathcal{O}(k^{-2/n})$

ab hier:
wie reine
Zufallssuche!

Analyse

(1, λ)-EA mit f(x) = || x ||^2

$$\| Y_k \|^2 = \| X_k + r_k U_k \|^2 = (X_k + r_k U_k)' (X_k + r_k U_k)$$

$$= X'_k X_k + 2 r_k X'_k U_k + r_k^2 U'_k U_k$$

$$= \|X_k\|^2 + 2 r_k X'_k U_k + r_k^2 \underbrace{\|U_k\|^2}_{=1} = \|X_k\|^2 + 2X'_k U_k + r_k^2$$

da das zufällige Skalarprodukt x‘U die gleiche Verteilung hat wie ||x|| B,

wobei Zufallsvariable B betaverteilt mit Parametern (n-1)/2 auf [-1, 1] ist, folgt

$$\| Y_k \|^2 = \| X_k \|^2 + 2 r_k \| X_k \| B + r_k^2$$

da der (1,λ) den besten Wert aus λ Versuchen selektiert, folgt

$$\| X_{k+1} \|^2 = \| X_k \|^2 + 2 r_k \| X_k \| B_{1:\lambda} + r_k^2$$

$$\| X_{k+1} \|^2 = \| X_k \|^2 + 2r_k \| X_k \| B_{1:\lambda} + r_k^2$$

$\downarrow$ bedingte Erwartungswerte auf beiden Seiten

$$E\| X_{k+1} \|^2 = \| X_k \|^2 + 2r_k \| X_k \| E[B_{1:\lambda}] + r_k^2$$

$\downarrow$ Ansatz: $r_k = \gamma \| X_k \|$

$$E\| X_{k+1} \|^2 = \| X_k \|^2 + 2 \gamma \| X_k \|^2 E[B_{1:\lambda}] + \gamma^2 \| X_k \|^2$$

$\downarrow$ wg. Symmetrie von B folgt $E[B_{1:\lambda}] = - E[B_{\lambda:\lambda}] < 0$

$$\begin{aligned} E\| X_{k+1} \|^2 &= \| X_k \|^2 - 2 \gamma \| X_k \|^2 E[B_{\lambda:\lambda}] + \gamma^2 \| X_k \|^2 \\ &= \| X_k \|^2 (1 - 2\gamma E[B_{\lambda:\lambda}] + \gamma^2) \end{aligned}$$

mit Anpassung:

$D_k \sim \mathcal{O}(c^k),\ c \in (0,1)$

Parabelscheitel bei $\gamma^* = E[B_{\lambda:\lambda}]$, also $E\| X_{k+1} \|^2 = \| X_k \|^2 (1 - E[B_{\lambda:\lambda}]^2)$

Praktisches Problem:

Woher bekommen wir $\| X_k \|$ bei $r_k = \| X_k \| \cdot E[\, B_{\lambda:\lambda} \,]$?

Wir wissen aus Analyse: $E\| X_{k+1} \|^2 = \| X_k \|^2 \, (1 - E[B_{\lambda:\lambda}]^2)$

Annahme: r_k war optimal eingestellt

$$\Rightarrow r_{k+1} = \| X_{k+1} \| \; E[\, B_{\lambda:\lambda} \,] \approx \| X_k \| \underbrace{(1 - E[\, B_{\lambda:\lambda} \,]^2)^{1/2} \, E[\, B_{\lambda:\lambda} \,]}_{\text{Konstante!}}$$

$\Rightarrow$ altes r_k mit Konstante multiplizieren: $r_{k+1} = c \cdot r_k$

aber: woher bekommen wir r_0 bzw. $\| X_0 \|$?

$\rightarrow$ r_0 zu groß

$\rightarrow$ r_0 genau richtig

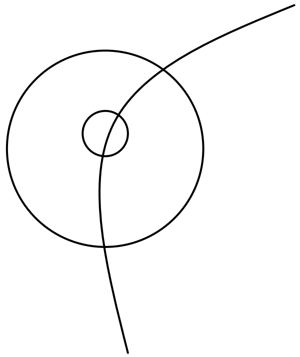
$\rightarrow$ r_0 zu klein

(1+1)-EA mit Schrittweitenanpassung (1/5-Erfolgsregel, Rechenberg 1973)

Idee:

- Wenn viele erfolgreiche Mutationen, dann Schrittweite zu klein.
- Wenn wenige erfolgreiche Mutationen, dann Schrittweite zu groß.

bei infinitesimal
kleinem Radius ist
Erfolgsrate = 1/2



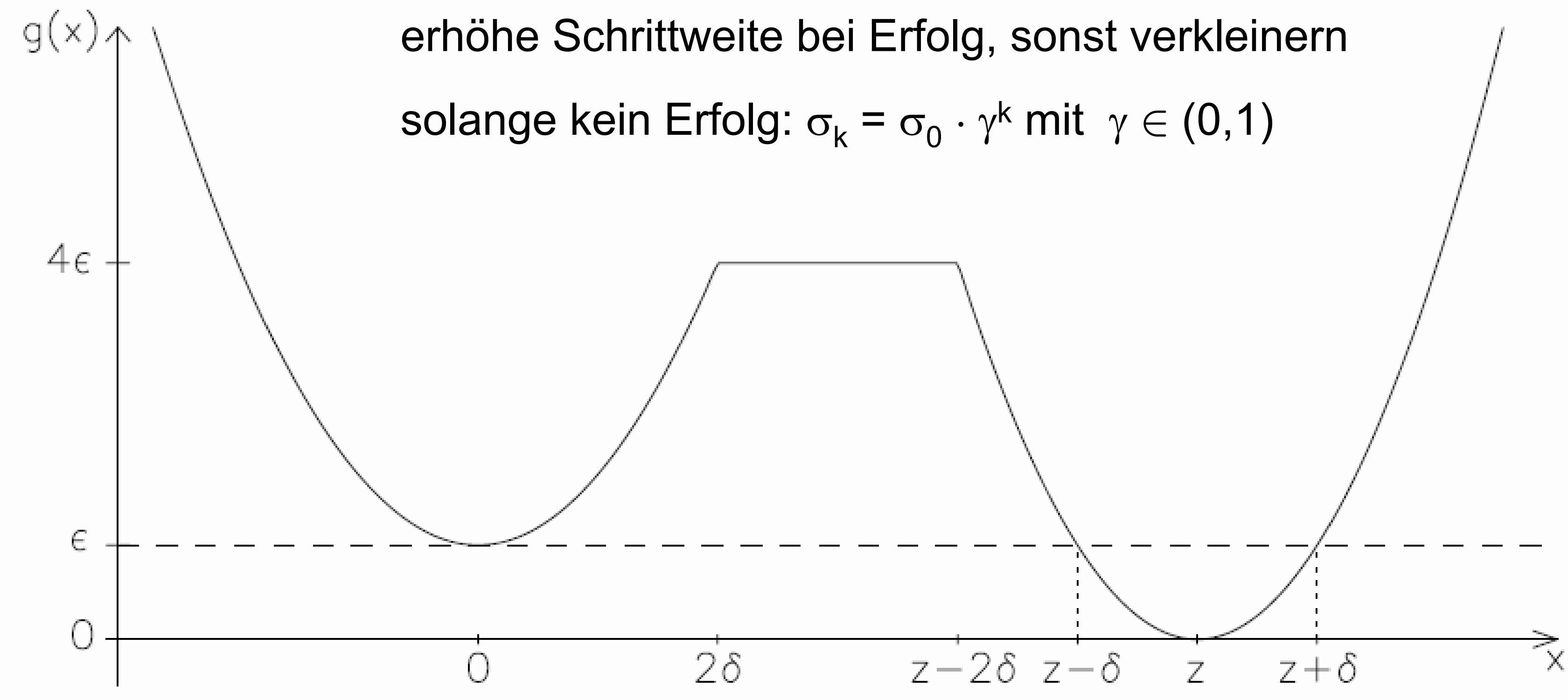
Ansatz:

- Protokolliere erfolgreiche Mutationen in gewissem Zeitraum
- Wenn Anteil größer als gewisse Schranke (z. B. 1/5),
  dann Schrittweite erhöhen,
  sonst Schrittweite verringern

Konvergenzproblematik bei der Schrittweitenanpassung

erhöhe Schrittweite bei Erfolg, sonst verkleinern

solange kein Erfolg: $\sigma_k = \sigma_0 \cdot \gamma^k$ mit $\gamma \in (0,1)$



Annahme: $X_0 = 0$

Frage: Wird lokales Optimum sicher verlassen (Übergang zu [z-δ,z+δ]) ?

Sei q_k Wahrscheinlichkeit, im Schritt k das lokale Optimum zu verlassen.

Kriterium für <u>sicheres</u> Verlassen:

$$1 - \prod_{k=1}^{\infty} (1 - q_k) = 1 \Leftrightarrow \prod_{k=1}^{\infty} (1 - q_k) = 0 \Leftrightarrow \boxed{\sum_{k=1}^{\infty} \log \frac{1}{1 - q_k} = \infty}$$

Kriterium für <u>unsicheres</u> Verlassen:

$$1 - \prod_{k=1}^{\infty} (1 - q_k) < 1 \Leftrightarrow \prod_{k=1}^{\infty} (1 - q_k) > 0 \Leftrightarrow \boxed{\sum_{k=1}^{\infty} \log \frac{1}{1 - q_k} < \infty}$$

Vereinfachung des log-Terms $\longrightarrow$

Lemma:

Sei x ∈ (0,1). Dann gilt: $$x < \log\left(\frac{1}{1-x}\right) < \frac{x}{1-x}$$

Beweis:

Reihenentwicklung $$\log\left(\frac{1}{1-x}\right) - \log(1-x) = \sum_{i=1}^{\infty} \frac{x^i}{i}$$

also: $$0 < x < \sum_{i=1}^{\infty} \frac{x^i}{i} < \sum_{i=1}^{\infty} x^i = \sum_{i=0}^{\infty} x^i - 1 = \frac{x}{1-x}$$

q.e.d.

Hinreichendes Kriterium für unsicheres Verlassen:

$$\sum_{k=1}^{\infty} \log \frac{1}{1-q_k} < \sum_{k=1}^{\infty} \frac{q_k}{1-q_k} < \boxed{\frac{1}{1-q_1} \sum_{k=1}^{\infty} q_k < \infty}$$

Lemma

weil q_k monoton fallend

$p_k = P\{ 0 \rightarrow (z-\delta, z+\delta)\} = P\{ z-\delta < Z < z+\delta \} = F_Z(z+\delta)-F_Z(z-\delta) =$

$= 2\,\delta\, f_Z(z-\delta + \theta \cdot 2\,\delta)$ mit $\theta \in (0,1)$

Mittelwertsatz der Differentialrechnung!

Annahme: Dichte $f_Z(\cdot)$ von Z ist unimodal

dann: $2\,\delta\, f_Z(z+\delta) < p_k < 2\,\delta\, f_Z(z-\delta)$ und deshalb: $q_k = 2\,\delta\, f_Z(z-\delta)$

Z sei normalverteilt

$$f_Z(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right)$$

$$\begin{aligned} p_k \le q_k &= \delta\sqrt{\frac{2}{\pi}}\frac{1}{\sigma_k}\exp\left(-\frac{(z-\delta)^2}{2\sigma_k^2}\right) \\ &= A\,\eta_k \exp(-B\,\eta_k^2) \end{aligned}$$

wobei

$$A = \delta\,(2/\pi)^{1/2},\ B = (z-\delta)^2/2,\ \eta_k = 1/\sigma_k.$$

$$\sum_{k=0}^{\infty} |a_k| < \infty \quad \text{falls} \quad \lim_{k\to\infty} |a_k|^{1/k} = \alpha < \infty$$

Sei $\eta_k = \eta_0\,\beta^k$ mit $\beta = 1/\gamma > 1$

$$\sum_{k=1}^{\infty} \frac{\beta^k}{\exp(B\,\eta_0^2\,\beta^{2k})}$$

konvergiert nach Wurzelkriterium!

$\Rightarrow$ kein sicheres Entkommen von lokalen Optima!