Chapter S:VI

VI. Relaxed Models

- Motivation
- \Box ε -Admissible Speedup Versions of A*
- □ Using Information about Uncertainty of *h*
- □ Risk Measures
- Nonadditive Evaluation Functions
- □ Heuristics Provided by Simplified Models
- Mechanical Generation of Admissible Heuristics
- □ Probability-Based Heuristics

Verwendung einer nicht-zulässigen Schätzfunktion

Idee:

Die Schätzfunktion h schätzt die billigsten verbleibenden Kosten h^* meistens recht gut, aber überschätzt h^* manchmal um nicht mehr als ε .

 \rightarrow A* mit solch einer Schätzfunktion h ist ε -zulässig.

Die Bedingung für die ε -Zulässigkeit von A* wird erfüllt, da zum Zeitpunkt der Terminierung gilt

$$h(n) - h^*(n) \le \varepsilon$$
 für alle $n \in \mathsf{OPEN}$.

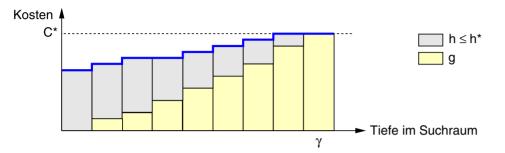
Auch die abgeschwächte Form der Zulässigkeit von h ist oft zu restriktiv.

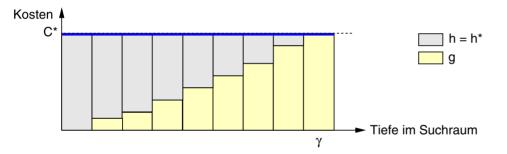
Häufig ist es leichter, eine Schätzfunktion für h^* zu finden, die meistens gut schätzt und h^* manchmal (um deutlich mehr als ε) überschätzt.

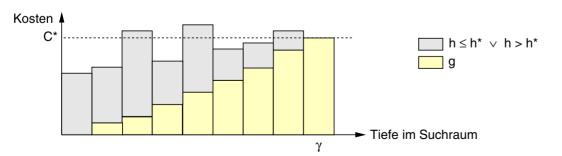
Der Fehler in der Schätzung ist nicht beschränkt, ein großer Fehler aber unwahrscheinlich.

- $\ \, \Box \ \, \text{Schätzfunktionen} \,\, h \,\, \text{mit} \,\, h \leq (1+\varepsilon)h^* \,\, \text{heißen} \,\, \varepsilon\text{-zulässig}.$
- ullet Die Bedingung " $h(n) h^*(n) \le \varepsilon$ für alle $n \in \mathsf{OPEN}$ " ist hinreichend, aber nicht notwendig.

Illustration von unter- und überschätzenden Schätzfunktionen







Beispiel: Suche in "zufälligem" Graphen

Gegeben ist ein Graph mit zufällig gezogenen Kantenkosten. Die minimale Anzahl von Kanten zu einem Zielknoten ist in jedem Knoten bekannt.

- figspace Kantenkosten c(n, n') seien gleichverteilt im Intervall [0; 1].
- □ Für lange Pfade mit N Kanten von einem Knoten n zu einem Ziel in Γ ist $h^*(n)$ wahrscheinlich in der Nähe von $\frac{N}{2}$.
- \Box Die einzige *zulässige* Schätzfunktion für h^* ist $h_1(n) = 0$.
- ullet Die einzige vernünftige Schätzfunktion für h^* ist $h_2(n) = \frac{N}{2}$.

Die Schätzfunktion h_2 führt im Worst-Case zu einer Kostenüberschätzung von $\frac{N}{2}$ und ist damit nicht $(\varepsilon$ -)zulässig. Aber: Dieser Fall ist äußerst unwahrscheinlich.

Beispiel: Suche in "zufälligem" Graphen

Gegeben ist ein Graph mit zufällig gezogenen Kantenkosten. Die minimale Anzahl von Kanten zu einem Zielknoten ist in jedem Knoten bekannt.

- ullet Kantenkosten c(n, n') seien gleichverteilt im Intervall [0; 1].
- □ Für lange Pfade mit N Kanten von einem Knoten n zu einem Ziel in Γ ist $h^*(n)$ wahrscheinlich in der Nähe von $\frac{N}{2}$.
- \Box Die einzige *zulässige* Schätzfunktion für h^* ist $h_1(n) = 0$.
- $oldsymbol{\Box}$ Die einzige vernünftige Schätzfunktion für h^* ist $h_2(n)=rac{N}{2}.$

Die Schätzfunktion h_2 führt im Worst-Case zu einer Kostenüberschätzung von $\frac{N}{2}$ und ist damit nicht $(\varepsilon$ -)zulässig. Aber: Dieser Fall ist äußerst unwahrscheinlich.

→ Algorithmus R^*_{δ} :

- Neben einer Schätzfunktion h für h^* ist auch Wissen über die Unsicherheit der Schätzung vorhanden.
- Das Wissen über die Unsicherheit einer Schätzung ist mit einer Dichtefunktion $\rho_{h^*}(x)$ definiert.

Beschreibung der Schätzunsicherheit durch Dichtefunktionen

Betrachtung von Kostenfunktionen als Zufallsvariable:

Kostenfunktion	Zufallsvariable
$h^*(n)$	h_n^*
$f^*(n) = g^*(n) + h^*(n)$	f_n^*
$f^+(n) = g(n) + h^*(n)$	f_n^+

Sei $\rho_{h_n^*}$ eine Dichtefunktion für die Zufallsvariable h_n^* .

Semantik:

Auf Basis von $\rho_{h_n^*}$ lässt sich die Wahrscheinlichkeit definieren, mit der $h^*(n)$ in einer Umgebung der Kosten x gefunden werden kann.

$$P(h_n^* = x) = \rho_{h_n^*}(x)$$

Beschreibung der Schätzunsicherheit durch Dichtefunktionen (continued)

Sei $\rho_{h_n^*}$ eine Dichtefunktion für die Zufallsvariable h_n^* .

Weiterhin gilt:

1. Aus $\rho_{h_n^*}(x)$ lässt sich eine Dichtefunktion $\rho_{f_n^*}(y)$ für die Zufallsvariable f_n^* ableiten, falls g^* bekannt ist:

$$\rho_{f_n^*}(y) := \rho_{h_n^*}(y - g^*)$$

2. Es sei P_{s-n} der billigste bislang bekannte Pfad von s zu einem OPEN-Knoten n.

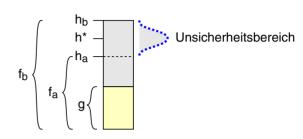
Aus $\rho_{h_n^*}(x)$ lässt sich eine Dichtefunktion $\rho_{f_n^+}(y)$ für die Zufallsvariable f_n^+ für den Fall ableiten, dass P_{s-n} fortgesetzt wird:

$$\rho_{f_n^+}(y) := \rho_{h_n^*}(y - g)$$

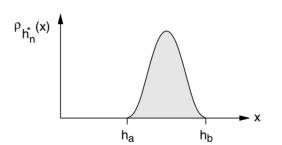
- \Box Die Zufallsvariable f_n^+ mit zugehöriger Dichtefunktion $\rho_{f_n^+}$ ist für jeden Knoten n gegeben.
- \Box Die Zufallsvariable f_n^+ beschreibt die möglichen Kosten eines optimalen Lösungspfades, der als Teilpfad den Zeigerpfad P_{s-n} enthält.
- Wenn von s aus Zielknoten erreichbar sind, enthält die OPEN-Liste immer einen Knoten n, für den $f^+(n) = f^*(n) = C^*$ gilt. (Pearl Lemma 2)

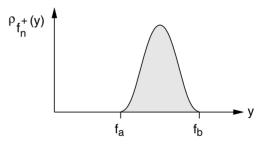
Beschreibung der Schätzunsicherheit durch Dichtefunktionen (continued

Unsicherheitsbereich:

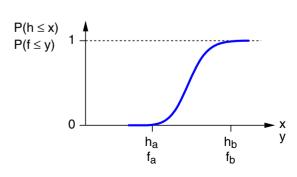


Dichtefunktionen ρ :





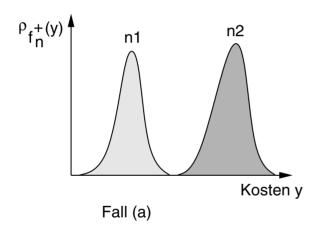
Zugehörige Verteilungsfunktion:

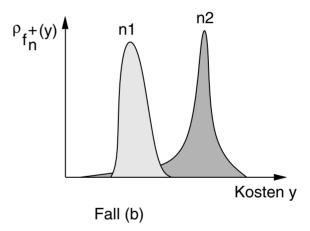


Beschreibung der Schätzunsicherheit durch Dichtefunktionen (continued)

Wie soll aus den Dichtefunktionen $\rho_{f_n^+}$ der Knoten in der OPEN-Liste eine Evaluierungsreihenfolge berechnet werden?

Möglicher Verlauf von zwei Dichtefunktionen:





- (a) Falls sich die Dichtefunktionen nicht überlappen, würde derjenige Knoten gewählt, für den die zugehörige Dichtefunktion ρ_{f^+} den niedrigsten Dichtewert f^+ bzgl. aller anderen Knoten besitzt.
- (b) $f_{n_1}^+$ hat den niedrigeren Erwartungswert; bei n_2 besteht die Möglichkeit, dass die Kosten $f_{n_2}^+$ niedriger als bei n_1 sein können. Ein zulässiger Algorithmus würde n_2 expandieren. Sinnvoller wäre es, n_1 zu expandieren, weil das Ereignis " $f^+(n_2) < f^+(n_1)$ " unwahrscheinlich ist (Anm.: $f^+(n)$ ist Kostenfunktion, nicht Zufallsvariable).
- → Bedingt durch die Unsicherheit können Kosten über- oder unterschätzt werden. D. h., einen Knoten in der OPEN-Liste nicht zu expandieren und als Folge hiervon zu teuer zu terminieren, stellt ein *Risiko* dar.
- → Quantifizierung des Risikos, mit zu hohen Kosten (= zu früh) zu terminieren.

Idee zur Berechnung der Evaluierungsreihenfolge:

Schätzung des Risikos, zu früh zu terminieren mittels eines R für jeden Knoten in der OPEN-Liste.

- \Box Für eine gegebene Kostenhöhe C (eines Zielknotens) bewertet das Risikomaß für jeden Knoten n in der OPEN-Liste, inwieweit C durch Expandieren von n noch verbessert werden kann.
- \square R = R(C). Das Risikomaß ist folglich eine monoton steigende Funktion der Kosten C. Je größer der R(C)-Wert eines Knotens n, desto größer ist das Risiko, eine Verbesserung von C zu verpassen, falls mit C terminiert wird, ohne n zu expandieren.
- $\ \square \ R(C)$ sollte Wissen über die Kostenverteilung des Knotens n verwenden, sollte also auf $\rho_{f_n^+}$ basieren.

Prinzip des Algorithmus R*_δ

Suche fortsetzen, bis der Risikowert R(C) jedes Knotens in der OPEN-Liste unterhalb einer vom Anwender akzeptierten Risikoschwelle δ ist.

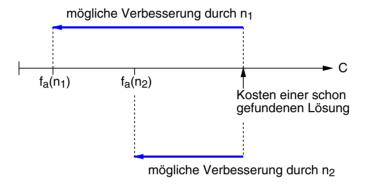
- → Akzeptiert man ein hohes Risiko, so bleiben Knoten mit hohem Risikowert R(C) (= hohes Kostensenkungspotenzial) unexpandiert. Folglich wird eine Kostenunterschätzung unwahrscheinlicher.
 Akzeptiert man nur ein kleines Risiko, so werden selbst Knoten mit geringem Risikowert R(C) (= geringes Kostensenkungspotenzial) expandiert. Folglich wird eine Kostenunterschätzung wahrscheinlicher.
- lacktriangledown D. h, abhängig von einer Risikoschwelle $R(C)=\delta$ lässt sich die Wahrscheinlichkeit einer Kostenunterschätzung (= Wahrscheinlichkeit der Zulässigkeit bzw. Optimalität) steuern.

Bemerkungen und Fragen:

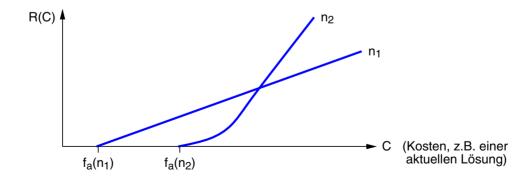
- Die Einhaltung dieses Prinzip durch den Algorithmus R^*_{δ} ist wie später gezeigt wird bei Verwendung bestimmter Risikofunktionen R(C) sichergestellt.
- ullet Welche Eigenschaft muss R(C) erfüllen?

Verbesserungspotenzial einer aktuellen Lösung

Es seien n_1 , n_2 Knoten der OPEN-Liste.



Beispiel für zwei Risikoverläufe R(C) für die Knoten n_1, n_2 :

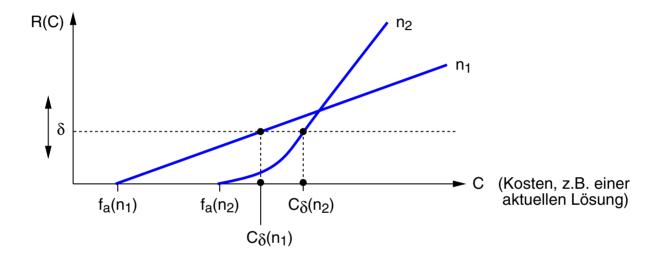


Die Knoten besitzen verschiedene Kostenzufallsvariablen $f_{n_1}^+$ und $f_{n_2}^+$.

- Das Verbesserungspotenzial ist eine statistische Größe, die für einen Knoten n mittels f_n^+ definiert ist.
- \Box Die *Bewertung* des Verbesserungspotenzials hinsichtlich gegebener Kosten C geschieht mit Hilfe eines Risk Measuress R(C).

Risikoschwelle und Risikokosten

Die Risikoschwelle $\delta \geq 0$ definiert für jeden Knoten n der OPEN-Liste seine Risikokosten $C_{\delta}(n)$:



Würde die Suche mit Knoten n_2 und Kosten $C'=C_\delta(n_2)$ terminieren, so läge für n_1 das Riskio R(C') oberhalb der Risikoschwelle δ .

 \rightarrow R*_{δ} wählt in der OPEN-Liste den Knoten n mit niedrigsten Risikokosten $C_{\delta}(n)$. Im obigen Beispiel wäre das der Knoten n_1 .

- Risk Measures und Risikoschwellen müssen im Kontext betrachtet werden: nicht jede Risikoschwelle ist für ein Risikomaß sinnvoll.
- \Box Abhängig von der Kostenzufallsvariablen f_n^+ eines Knotens n kann die Risikoschwelle δ zu verschiedenen Reihenfolgen in der OPEN-Liste führen.
- Die Risikokosten $C_{\delta}(n)$ geben an, wie hoch die Kosten einer Lösung sein dürfen, ohne dass die Risikoschwelle δ für den Knoten n überschritten werden.

Definition 75 (δ-Risikozulässigkeit)

Ein Algorithmus heißt δ -risikozulässig, falls er mit Lösungskosten C terminiert und $R(C) \leq \delta$ für alle Knoten gilt, die sich in der OPEN-Liste befinden.

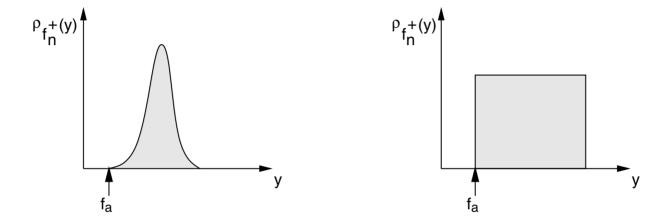
Definition 76 (Algorithmus R^*_{δ})

 R^*_δ ist ein Suchalgorithmus, der identisch ist zu A^* bis auf den Unterschied, dass derjenige Knoten n in der OPEN-Liste zur Expansion gewählt wird, für den die Risikokosten $C_\delta(n)$ am niedrigsten sind.

- \square Eine äquivalente Definition für δ -Risikozulässigkeit ist: Bei Terminierung mit Lösungskosten C müssen alle Knoten n in der OPEN-Liste die Bedingung $C_{\delta}(n) \geq C$ erfüllen. Die erste Definition ist aus dem Blickwinkel des Risikos, die zweite ist aus dem Blickwinkel der Kosten.
- \Box Für $\delta = 0$ ist R^*_{δ} identisch mit A^* . Argumentation:
 - 1. $\delta = 0 \Rightarrow R(C) = 0$
 - 2. Die Lösung zu R(C) = 0 bestimmt die zu akzeptierenden Kosten.
 - 3. Kosten mit Risiko 0 bedeuten, dass von der Dichtefunktion f_n^+ der äußerste linke Wert f_a genommen wird.
 - 4. $f_a \leq g(n) + h^*(n) \Rightarrow \mathsf{R}^*_{\delta}$ ist zulässig im Sinne von A^* .
- \square Mit steigendem δ tendiert R^*_{δ} zur Aufgabe der Zulässigkeit.

Ausgangspunkt sind Dichtefunktionen für die Zufallsvariable f_n^+ eines Knoten n in der OPEN-Liste.

Beispiele:



 f_a (bzw. h_a) ist das kleinste positive Urbild der Dichtefunktion $\rho_{f_n^+}$ (bzw. $\rho_{h_n^*}$).

Risk Measures mit der Struktur $R(C) = \varrho[C - f^+]$

1. Worst-Case-Maß R_1 :

$$R_1(C) = \sup_{\{y \mid \rho_{f_n^+}(y) > 0\}} (C - y) = C - f_a = C - g - h_a$$

2. Risiko einer suboptimalen Terminierung R_2 :

$$R_2(C) = P(C > f_n^+) = P(C - f_n^+ > 0) = \int_{y=-\infty}^{C} \rho_{f_n^+}(y) dy$$

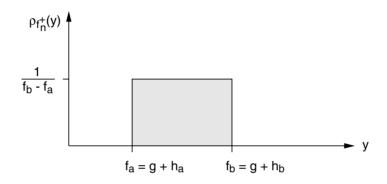
3. Erwartetes Risiko R_3 :

$$R_3(C) = E(\max\{C - f_n^+; 0\}) = \int_{y = -\infty}^{C} (C - y) \rho_{f_n^+}(y) dy$$

- $lue{}$ Die Risk Measures R_1 und R_3 beschreiben Kosten, das Risikomaß R_2 beschreibt eine Wahrscheinlichkeit.
- \square R_1 : Für die durch die Zufallsvariable f_n^+ repräsentierten Kosten wird der kleinstmögliche Wert angenommen. R_1 quantifiziert den maximal möglichen Verlust, falls man sich einer Lösung mit Kosten C zufrieden gibt.
- \square R_2 : Die Wahrscheinlichkeit für das Eintreffen des Ereignisses " $C > f_n^+$ " (eines Verlustes) wird berechnet, falls man sich mit einer Lösung mit Kosten C zufrieden gibt.
- \square R_3 : Für die durch die Zufallsvariable f_n^+ repräsentierten Kosten wird der erwartete Verlust $E(\max\{C-f_n^+;0\})$ berechnet, falls man sich mit einer Lösung mit Kosten C zufrieden gibt.

Beispiel

Gegeben sei eine gleichverteilte Kostenzufallsvariable f_n^+ :



Dichtefunktion:

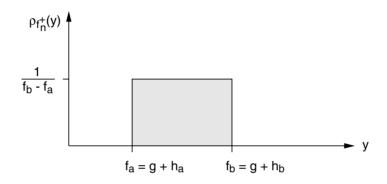
$$\rho_{f_n^+}(y) = \begin{cases} \frac{1}{f_b - f_a} & f_a \le y \le f_b \\ 0 & \text{sonst} \end{cases}$$

Risikomaß:

$$R_1(C) = \sup_{\{y \mid \rho_{f_n^+}(y) > 0\}} (C - y) = C - f_a$$

Beispiel (continued)

Gegeben sei eine gleichverteilte Kostenzufallsvariable f_n^+ :



Dichtefunktion:

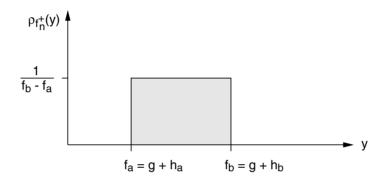
$$\rho_{f_n^+}(y) = \begin{cases} \frac{1}{f_b - f_a} & f_a \le y \le f_b \\ 0 & \text{sonst} \end{cases}$$

Risikomaß:

$$R_2(C) = \int_{y=-\infty}^{C} \rho_{f_n^+}(y) dy = \begin{cases} 0 & C < f_a \\ \frac{(C-f_a)}{(f_b - f_a)} & f_a \le C \le f_b \\ 1 & f_b < C \end{cases}$$

Beispiel (continued)

Gegeben sei eine gleichverteilte Kostenzufallsvariable f_n^+ :



Dichtefunktion:

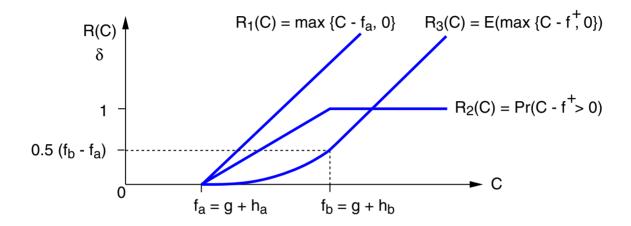
$$\rho_{f_n^+}(y) = \begin{cases} \frac{1}{f_b - f_a} & f_a \le y \le f_b \\ 0 & \text{sonst} \end{cases}$$

Risikomaß:

$$R_3(C) = \int_{y=-\infty}^{C} (C - y) \rho_{f_n^+}(y) dy = \begin{cases} 0 & C < f_a \\ \frac{(C - f_a)^2}{2(f_b - f_a)} & f_a \le C \le f_b \\ C - \frac{f_a + f_b}{2} & f_b < C \end{cases}$$

Beispiel (continued)

Vergleich der Risk Measures R_1 , R_2 und R_3 (f_n^+ gleichverteilt):

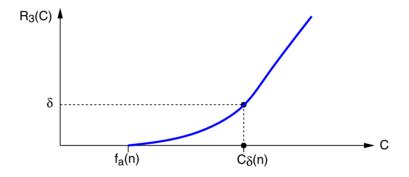


Das Bild zeigt für *einen* Knoten n bei gleichverteilter Kostenzufallsvariable f_n^+ die Berechnung des Risikos mit den Maßen R_1 , R_2 und R_3 .

Beispiel (continued)

Berechnung der Risikokosten C_{δ} für R_3 (f_n^+ gleichverteilt):

Sei δ die Risikoschwelle eines Anwenders. Sie definiert für jeden Knoten der Open-Liste dessen Risikokosten auf Basis der Gleichung $R(C)=\delta$.



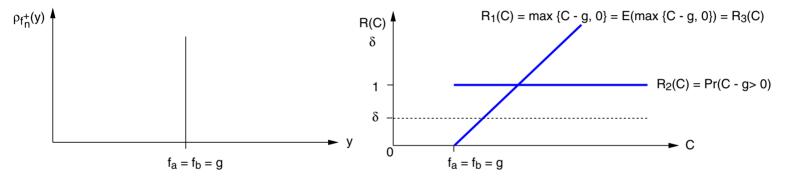
Die Risikokosten $C_{\delta}(n)$ für einen Knoten n der OPEN-Liste berechnen sich aus der Umformung von $R_3(C_{\delta}) = \delta$:

$$C_{\delta} = \begin{cases} f_a & \delta = 0 \\ f_a + \sqrt{2 \cdot (f_b - f_a) \cdot \delta} & 0 < \delta \le \frac{f_b - f_a}{2} \\ \delta + \frac{f_a + f_b}{2} & \frac{f_b - f_a}{2} < \delta \end{cases}$$

Beispiel (continued)

Gegeben sei ein Zielknoten n, d.h., wegen $h(n) = h^*(n) = 0$ gibt es keine Unsicherheit hinsichtlich der Kostenzufallsvariablen f_n^+ .

Verlauf der Kostenzufallsvariablen f_n^+ und der Risk Measures R_1 , R_2 und R_3 :



Risikokosten:

$$C_\delta(n) = \left\{ egin{array}{ll} g(n) + \delta & {
m Risikomaß} \ R_1 \ & g(n) & {
m Risikomaß} \ R_2 \ {
m und} \ \delta < 1 \ & g(n) + \delta & {
m Risikomaß} \ R_3 \end{array}
ight.$$

Damit folgt die δ -Risikozulässigkeit von R^*_{δ} bezüglich R_1 , R_2 und R_3 .

Theorem 77 (δ -Risikozulässigkeit von R^*_{δ})

 R^*_{δ} ist δ -risikozulässig bezüglich der Risk Measures R_1 , R_2 und R_3 .

- Die genaue Form von $\rho_{h^*}(n)$ ist im Allgemeinen unbekannt. Hierzu müssten die Kantenkosten $c(n,n'),n'\in succ(n)$ durch ein vorgebenes probabilistisches Modell generiert worden sein.
- Die Generierung einer guten Schätzung für $C_{\delta}(n)$ ist oft möglich. Hierzu genügt das Wissen über obere und untere Schranken von h^* zusammen mit der oft sinnvollen Annahme einer standardisierten Verteilung dazwischen, wie einer Gleichverteilung, einer Exponentialverteilung oder einer Normalverteilung.
- Das Prinzip der ε -zulässigen Beschleunigung kann für R^*_{δ} genauso angewandt werden, wie für A^* .