

7. Vergleich von Systemkonfigurationen

Simulation wird oftmals zur Entscheidungsunterstützung eingesetzt

Typische Fragestellungen:

1. Was ist die „beste“ Systemkonfiguration?
2. Welche von n Systemkonfigurationen?

Unterschiede hier:

Im ersten Fall viele (u.U. unendlich, überabzählbar viele)

Konfigurationen (\Rightarrow nur einen Teil untersuchen \Rightarrow Optimierung)

Im zweiten Fall wenige (u.U. nur zwei oder drei) Konfigurationen die alle untersucht werden

(\Rightarrow Auswahl aus einer Menge von Resultaten)

Wir behandeln hier das zweite Problem und zeigen einen möglichen Ansatz zur Optimierung in Kap. 8

Warum ist die Auswahl der “besten” Konfiguration ein Problem?

Ergebnisse sind stochastisch \Rightarrow

Verhaltensunterschiede müssen geschätzt werden

Zugehöriges Problem in der Simulation:

Ranking and Selection

Immer noch ein aktuelles Problem siehe Papiere auf den

Winter Simulationskonferenzen

Literatur hier:

- A. M. Law, W. D. Kelton.
Simulation Modeling and Analysis, Kap. 10
- J. Banks et al. Discrete Event System Simulation, Kap. 12
- Diverse Originalartikel (siehe Web-Seite der WSC)

7.1 Vergleich von zwei Systemen

7.2 Vergleich von mehreren Systemen

7.3 Allgemeine Verfahren zur Rangbildung und Auswahl

7.1 Vergleich von zwei Systemen

Zwei Systemkonfigurationen (bzw. Systeme), für die ein Leistungsmaß μ analysiert werden soll
(μ_i Leistungsmaß für Variante $i \in \{1,2\}$)

Ziel: Ermittlung von $\zeta = \mu_1 - \mu_2$
d.h. Bestimmung eines Konfidenzintervalls für ζ

Annahme: Beobachtung von Replikationen:

- n_i Replikationen für das i -te Leistungsmaß werden durchgeführt
- v_{ij} die j -te Beobachtung des i -ten Leistungsmaßes
- v_{ij} ist Realisierung einer ZV V_i mit $E(V_i) = \mu_i$

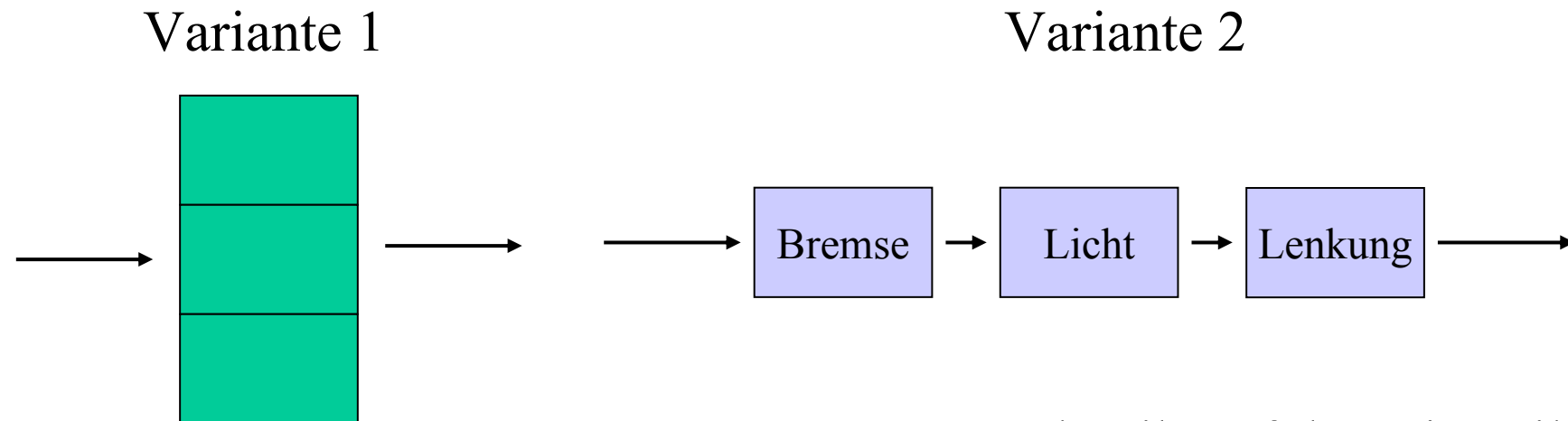
Interpretation des Konfidenzintervalls $\hat{\zeta} \pm \hat{\phi}_\alpha$
($\hat{\phi}_\alpha$ halbe Breite des Konfidenzintervalls für
Signifikanzwahrscheinlichkeit α)

- Falls $\hat{\zeta} + \hat{\phi}_\alpha < 0$ System 2 besser, falls $\hat{\zeta} - \hat{\phi}_\alpha > 0$ System 1 besser

Beispiel: Sicherheitsüberprüfung von PKWs (aus Banks et al. S. 451)

- Überprüfung von Bremsen, Licht und Lenkung
- PKWs kommen nach Poisson-Prozess mit Rate 9.5 PKW/Std. an

Mögliche Systemkonfigurationen (Ziel: Vergleich der Verweilzeiten)



- 3 Mechaniker, jeder führt alle Arbeiten durch
- Dauer der Arbeitsschritte: Bremse 6.5 Min, Licht 6 Min., Lenkung 5.5 Min. jeweils mit Stdabw. 0.5 Min

- 3 Mechaniker, führen jeweils spezifische Arbeiten durch
- Dauer der Arbeitsschritte: Bremse 5.85 Min, Licht 5.4 Min., Lenkung 4.95 Min. jeweils mit Stdabw. 0.5 Min

Annahme: Beobachtungen der Verweilzeiten sind

- (approximativ) normalverteilt
(kann dadurch erreicht werden, dass jede einzelne Beobachtung aus einer Summe von Einzelbeobachtungen erzeugt wird, siehe zentralen Grenzwertsatz)
- und unabhängig

Schätzer für Erwartungswert und Varianz:

$$\tilde{V}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} v_{i,j} \quad \text{und} \quad \tilde{S}_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (v_{i,j} - \tilde{V}_i)^2$$

Falls Varianz beider Konfigurationen gleich (oder zumindest ähnlich) ist, so können die so genannten t-Konfidenzintervalle berechnet werden

$$\tilde{W} \pm \tilde{\phi}_\alpha = (\tilde{V}_1 - \tilde{V}_2) \pm t_{n_1+n_2-2, 1-\alpha/2} \cdot \sqrt{\frac{(n_1-1)\tilde{S}_1^2 + (n_2-1)\tilde{S}_2^2}{n_1+n_2-2}} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Beobachtungen:

- Exakte Resultate, falls
 - Beobachtungen normalverteilt und
 - Varianzen σ_1^2 und σ_2^2 identisch sind

Im Prinzip entspricht das Vorgehen der „normalen“ Berechnung von Konfidenzintervallen für ein System

- Gute Approximation, falls
 - Varianzen ähnlich sind (aus Beobachtungen ableiten) oder
 - $n_1 = n_2$ gilt
- In allen anderen Fällen kann die Überdeckung der ermittelten Konfidenzintervalle sehr schlecht sein (d.h. Wahrscheinlichkeit, dass der wahre Wert im Intervall liegt, ist deutlich kleiner als α)

Ansonsten approximatives Vorgehen nach Welch (1938) verwenden:

Anzahl Freiheitsgrade:

$$\tilde{f} = \frac{\left(\tilde{S}_1^2 / n_1 + \tilde{S}_2^2 / n_2 \right)^2}{\left(\tilde{S}_1^2 / n_1 \right)^2 / (n_1 - 1) + \left(\tilde{S}_2^2 / n_2 \right)^2 / (n_2 - 1)}$$

In der Regel keine ganze Zahl, deshalb zur Bestimmung der Quantile der t-Verteilung interpolieren oder runden

$$\tilde{W} \pm \tilde{\phi}_\alpha = \left(\tilde{V}_1 - \tilde{V}_2 \right) \pm t_{\tilde{f}, 1-\alpha/2} \cdot \sqrt{\frac{\tilde{S}_1^2}{n_1} + \frac{\tilde{S}_2^2}{n_2}}$$

Vergleich der beiden Vorgehensweisen

- Falls beide Methoden anwendbar sind, ist nicht klar, welche schmalere Konfidenzintervalle liefert
- Methoden können auch für den Vergleich von realen Experimentergebnissen und Simulation eingesetzt werden, denn gilt oft $n_1 \neq n_2$
- Die erste Methode kann auch mit gemeinsamen Zufallszahlen eingesetzt werden (siehe Kap. 2.1), Varianz reduziert sich dann um $2 \text{COV}(V_1, V_2)/n$
- Bei der Methode von Welch wird Unabhängigkeit der Beobachtungen vorausgesetzt, damit dürfen keine gemeinsamen Zufallszahlen verwendet werden

Replikationen des Beispielmmodells

- Generierung der Ankunftsabstände mittels Exponentialverteilung mit Rate 9.5
- Generierung der Bedienzeiten $B_i = E(B_i) + 0.5 \cdot Z$ (Z ist eine $N(0,1)$ verteilte ZZ, $E(B_i)$ entspricht dem Erwartungswert des jeweiligen Arbeitsgangs)
- Für Variante 1 wird die Bedienzeit als Summe $B_1 + B_2 + B_3$ berechnet, in der zweiten Variante werden jeweils einzelne Bedienzeiten benutzt
- Experimentaufbau
 - 10 Replikationen für Variante 1 (M1)
 - 10 unabhängige Replikationen für Variante 2 (M2U)
 - 10 Replikationen für Variante 2 mit identischen Zufallszahlen zu Variante 1 (M2A)

Resultate der Replikationen

Repl.	M1	M2U	M2A	M1-M2A
1	29.59	51.62	29.55	0.04
2	23.49	51.91	24.26	-0.77
3	25.68	45.27	26.03	-0.35
4	41.09	30.85	42.64	-1.55
5	33.84	56.15	32.45	1.39
6	39.57	28.82	37.91	1.66
7	37.04	41.30	36.48	0.56
8	40.20	73.06	42.24	-1.04
9	61.82	23.00	60.59	1.23
10	44.00	28.44	41.49	2.51
MW	37.63	43.04		0.37
Var	118.9	244.3		0.42

Experimente:

- Simulationsdauer
16 Modellstunden
pro Replikation
- Start mit 2 PKWs
in der
Warteschlange
- Bestimmung der
mittleren
Verweilzeit

Auswertung für $\alpha = 0.05$:

- Varianzen der Verweilzeiten sind nicht identisch, dadurch liefert die erste Methode nur approximative Ergebnisse, ist aber anwendbar, da $n_1 = n_2$

- Resultate der 1. Methode:

$$\hat{W} = - 5.4 \text{ und}$$

$$\hat{\phi}_\alpha = t_{18,1-\alpha/2} \cdot ((9 \cdot 118.9 + 9 \cdot 244.3) / 18)^{1/2} \cdot (1/5)^{1/2} = 12.66$$

- Resultate der 2. Methode:

$$\hat{W} = - 5.4, \hat{f} = 16.88 \text{ (hier gerundet auf 17) und}$$

$$\hat{\phi}_\alpha = t_{17,1-\alpha/2} \cdot ((118.9/10 + 244.3/10)^{1/2}) = 12.71$$

- Resultate bei Verwendung gemeinsamer Zufallszahlen

$$\hat{W} = 0.37 \text{ und}$$

$$\hat{\phi}_\alpha = t_{9,1-\alpha/2} \cdot (1.74 / 10)^{1/2} = 0.94$$

- Konfidenzintervall mit gemeinsamen Zufallszahlen deutlich kleiner, als mit den anderen beiden Methoden
- Alle Konfidenzintervalle enthalten die 0 !

7.2 Vergleich von mehreren Systemen

- K unterschiedliche Systemkonfigurationen, die miteinander verglichen werden sollen
- Vergleich weiterhin durch Vergleich der Konfidenzintervalle
- Mögliche Fragestellungen
 1. Berechnung der individuellen Konfidenzintervalle für die Leistungsgrößen (K Konfidenzintervalle)
 2. Vergleich der Konfigurationen $k = 2, \dots, K$ mit Konfiguration 1 (K-1 Vergleiche)
 3. Vergleich aller Konfigurationen untereinander ($K(K-1) / 2$ Vergleiche)
 4. Auswahl der besten Konfiguration (bis zu $K(K-1)/2$ Vergleiche)

Ziel: Aussagen mit Konfidenzwahrscheinlichkeit α treffen

Also mit Wahrscheinlichkeit $1-\alpha$

1. Liegen alle wahren Werte in den K ermittelten Konfidenzintervallen
2. Umfassen alle Konfidenzintervalle den korrekten Unterschied zwischen der ersten und k -ten Konfiguration
3. Umfassen alle Konfidenzintervalle die korrekte Differenz zwischen zwei Konfigurationen
4. Wird die beste Konfiguration ausgewählt

Seien α_i und α_{ij} die Konfidenzwahrscheinlichkeiten der einzelnen Konfidenzintervalle

Fragestellung 1 (K unterschiedliche Konfidenzintervalle)

Falls alle Werte unabhängig sind gilt

$$1 - \alpha = \prod_{k=1, \dots, K} (1 - \alpha_i)$$

Wann sind die Werte unabhängig?

- Resultate aus unterschiedlichen Modellen, die mit unabhängigen Zufallszahlen simuliert wurden
- Resultate aus unterschiedlichen Replikationen, die mit unabhängigen Zufallszahlen simuliert wurden

Aber nicht, wenn

- unterschiedliche Resultate aus einem Simulationslauf oder
- Resultate aus Modellen/Replikationen mit gemeinsamen Zufallszahlen analysiert werden

Dann gilt die Bonferroni-Ungleichung $1 - \alpha \geq \max(1 - \sum_{k=1, \dots, K} \alpha_i, 0)$

Fragestellungen 1-4 (abhängigen Beobachtungen)

- Sei C die Anzahl der ermittelten Konfidenzintervalle (K , $K-1$ oder $K(K-1)/2$)
- Setze $\alpha_i = \alpha / C$ ($i = 1, \dots, C$)
- Bestimme individuelle Konfidenzintervalle mit Signifikanzwahrscheinlichkeit α_i (u.U. mit Hilfe gemeinsamer Zufallszahlen)
- Signifikanzwahrscheinlichkeit α gilt mindestens für das gemeinsame Konfidenzintervall

Nachteile der Vorgehensweise:

- Falls C groß, muss α_i klein sein (\Rightarrow breite Konfidenzintervalle oder viele Beobachtungen)
- Vorgegebene Breite führt zu unterschiedlicher Zahl von Beobachtungen
- Auswahl der besten Konfiguration (Fragestellung 4) ohne weitere Annahmen problematisch

7.3 Allgemeine Verfahren zur Rangbildung und Auswahl

Auswahl von m aus K Varianten, so dass

- mit Wahrscheinlichkeit p^* das beste System enthalten ist oder
- mit Wahrscheinlichkeit p^* die m besten Systeme enthalten sind

Notationen:

- Erwartungswert der i -ten Variante μ_i
es gelte $\mu_K \geq \mu_{K-1} \geq \dots \geq \mu_1$
- v_{i1}, v_{i2}, \dots Beobachtungen d.h. Realisierungen der ZVs V_{ij} mit $E(V_{ij}) = \mu_i$
und $\text{Var}(V_{ij}) = \sigma_i^2$
- $p_i = \text{Prob}(V_{ij} > \max_{l \neq i} V_{lj})$

Existierende Ansätze

- Auswahl einer Menge $I \subseteq \{1, \dots, K\}$, so dass die m besten Element mit Wahrscheinlichkeit p^* in I sind (subset selection)
- Auswahl von $I = \{K-m, \dots, K\}$ mit Wahrscheinlichkeit p^* ,
falls $\mu_{K-m} - \mu_{K-m-1} \geq \delta$ (Fall $m=1$ umfasst die Auswahl des besten Systems)

Prozeduren der beschriebenen Art sind ein aktuelles Forschungsthema, welches wir hier nur kurz ansprechen (für Details siehe Papiere auf der Winter Simulation Conference)

Prozedur zur Auswahl einer Menge I,

die mit Wahrscheinlichkeit $1-\alpha$ das beste System enthält

1. Wähle $1 - \alpha$, so dass $1/K < 1 - \alpha < 1$
2. Wähle n_i ($i = 1, \dots, K$) und führe n_i Replikationen für System i aus oder benutze vorhandene Resultate

$\Rightarrow V_{i1}, \dots, V_{in_i}$

3. Sei $t_i = t_{n_i-1, (1-\alpha)^{1/(K-1)}}$ und $W_{ij} = \left(\frac{t_i^2 \tilde{S}_i^2}{n_i} + \frac{t_j^2 \tilde{S}_j^2}{n_j} \right)^{1/2}$

4. Wähle $I = \{k \mid 1 \leq k \leq K \wedge \hat{V}_k \geq \hat{V}_1 - W_{k1} \text{ für alle } 1 \neq k\}$

Algorithmus wählt eine Menge, die Bedingungen erfüllt und von variabler Größe ist (insbesondere zum Screening geeignet!)

Aus der Screening-Prozedur resultiert idealerweise genau ein System, in der Praxis werden aber oft mehrere (im schlechtesten Fall K) Elemente ausgewählt

⇒ Weitere Beobachtungen sind zur Entscheidung notwendig

Alternative: Auswahlprozeduren

Voraussetzung:

- $\mu_K - \mu_{K-1} \geq \delta$,
- ansonsten Auswahl eines Systems i mit $\mu_K - \mu_i < \delta$
- Entscheidung ist mit Wahrscheinlichkeit $1 - \alpha$ korrekt

Prozeduren sind (zwangsläufig) konservativ, da Resultate für die schlechteste Anordnung,

nämlich $\mu_k = \mu_K - \delta$ für alle $k \in \{1, \dots, K-1\}$ gelten müssen

Auswahlprozedur:

1. Wähle $1/K < 1 - \alpha < 1$ und n_i
2. Führe n_i Replikationen für System i aus
3. $h = h(2, (1 - \alpha)^{1/(K-1)}, n_{\min})$ mit $n_{\min} = \min_{i=1, \dots, K} n_i$
(Konstante numerisch zu berechnen oder aus Tafel ablesen)
4. Bestimme Anzahl zusätzlicher Replikationen

$$m_i = \max \left(n_i, \left\lceil \left(\frac{h \tilde{S}_i^2}{\delta} \right)^2 \right\rceil \right)$$

5. Führe $m_i - n_i$ Replikationen für System i durch
6. Auswahl des Systems i mit dem größtem Mittelwert,
d.h. $i = \operatorname{argmax}_{k=1, \dots, K} \hat{V}_k$

- Auswahlprozeduren sind aufwändig, wenn
 - K oder S_i^2 groß sind
 - δ oder α klein sind
- Bei vorgegebener Genauigkeit ist die Reduktion von K die einzige Möglichkeit der Aufwandsreduktion
- Kombination von Screening und Auswahl
(erst Screening zur Reduktion von K , dann Auswahl aus der Restmenge)

Vorgehen ist allerdings nur „statistisch korrekt“, falls

- die Beobachtungen der ersten (Screening-) Phase nicht in der zweiten (Auswahl-) Phase verwendet werden

Ansatz erweiterbar zur Auswahl der m besten Systeme