

Vorlesung

Simulation

technischer

Systeme

Walter Schneider

SS 2005

Inhaltsverzeichnis

1. Was ist und was soll Simulation
 - 1.1 Beispiel
 - 1.2 Definitionen
 - 1.3 Zusammenfassung
2. Simulationskonzepte
 - 2.1 Ereignisorientierung
 - 2.2 Aktivitätsorientierung
 - 2.3 Prozessorientierung
3. Zufallszahlengeneratoren für gleichverteilte Zufallszahlen
 - 3.1 Additiv kongruentielle Methode
 - 3.2 Multiplikativ kongruentielle Methode
 - 3.3 Gemischt kongruentielle Methode
 - 3.4 Mid-Square Methode
4. Test von Zufallszahlen bzw. erhobenen Daten
 - 4.1 Chi-Quadrat-Test auf Gleichförmigkeit
 - 4.2 Chi-Quadrat-Test auf paarweise Unabhängigkeit
 - 4.3 Chi-Quadrat-Test auf Unabhängigkeit von Zufallszahlenfolgen
 - 4.4 Chi-Quadrat-Test auf Normalverteilung
 - 4.5 Chi-Quadrat-Test auf Exponentialverteilung
5. Prozessgeneratoren
 - 5.1 Methode der Inversen Transformation
(Exponentialverteilung, Cauchyverteilung, Logistisch)
 - 5.2 Methode der gespeicherten Datenblätter - Methode der Reproduktion der erhobenen Daten (für beliebige Verteilungen, Poissonverteilung als Beispiel)
 - 5.3 Methode der äquivalenten Transformation
Poissonverteilung -> Exponentialverteilung
Gleichverteilung -> Normalverteilung
 - 5.4 Methode der Reproduktion des stochastischen Prozesses
6. Validitätsprüfung
 - 6.1 Vergleich mit der Realität
 - 6.2 Augenscheinliche Gültigkeitsprüfung
 - 6.3 Sensitivitätsprüfung
 - 6.4 Untermodellprüfung
 - 6.5 Vergleich mit theoretischen Ergebnissen (Warteschlangentheorie)
7. Gründe für fehlerhafte Simulation
 - 7.1 Zu kurze Aufwärmphase
 - 7.2 Simulationsdauer (Dauer der Datenerhebung)
 - 7.3 Fehler bei der Datenerhebung
 - 7.4 Fehler bei systemvereinfachenden Annahmen

8. Varianzreduzierende Maßnahmen
 - Versuchsplanung und Datenanalyse
 - 8.1 1-faktorielle Varianzanalyse
 - 8.2 2-faktorielle Varianzanalyse
 - 8.2.1 2-faktorielle Varianzanalyse mit Wechselwirkung
 - 8.3 3-faktorielle Varianzanalyse
9. Formelzusammenstellung und statistische Tafeln

1.2 Definitionen

Ausgehend von den beiden einleitenden Beispielen können wir Simulation rein intuitiv folgendermaßen beschreiben:

Simulation ist ein effektives Mittel um ein selbst definiertes Modell zu testen, auszuarbeiten und zu manipulieren, ohne dabei direkt die Aktionen am „**Real-System**“ ausprobieren zu müssen. Stunden, Tage, Wochen ja sogar Jahre können in einigen Minuten bzw. Stunden am Computer simuliert werden. In den meisten Fällen ist ein Modell in dem simuliert wird nicht eine ganz genaue Abbildung des aktuellen Systems („Real-System“), sondern eine symbolische Darstellung des Wirklichkeits-Systems. Es wird immer Ergebnisse, Messungen liefern, die auf anderen Wegen unmöglich zu erreichen wären.

Um Simulation exakt definieren zu können, müssen wir die Begriffe System, Modell und Modellierung näher betrachten.

DEFINITION SYSTEM:

-ist eine Gruppe von gegenseitig abhängigen Elementen die zusammenarbeiten, um ein bestimmtes Ziel zu erreichen. (Franta/Banks/Cars)

Die Elemente eines Systems heißen ENTITES.

Nun stellen wir uns die Frage, können wir in einem „Real-System“ Experimente (Untersuchungen) durchführen? Diese Frage wäre eindeutig mit „Ja“ zu beantworten.

Aber beim Beispiel „Tankstelle“ wäre die Untersuchung des Systems in der Realität zu aufwendig (zuerst bauen - eventuell wieder vergrößern oder wieder verkleinern). Es würde sehr viel Geld, Zeit und Ärger investiert werden, bis sich das System nach den gewünschten Eigenschaften verhält.

Sehr oft können die Untersuchungen in der Realität auch zu gefährlich sein (z. B.: Atomstrom, ...) und auch unmöglich (das gesamte System müsste geändert werden). Daher versuchen wir ein System in ein Modell abzubilden.

DEFINITION MODELL:

-ist eine Abbildung (specification) eines Systems derart, dass in ihm das Wesentliche des Systems enthalten ist. (Franta)

DEFINITION MODELLIERUNG:

-heißt der Prozess, eine interne Darstellung (internal representation) und eine Menge von Transformationsregeln zu entwickeln/definieren, die benutzt werden können, um die Reaktion (behavior) des Systems und die Beziehungen der Elemente des Systems vorauszusagen. (Franta)

representation: mit Hilfe von Variablen; diese Variablen charakterisieren den Systemzustand.

transformieren: diese Transformationsregeln verändern den Systemzustand.

Sind diese Regeln zeitabhängig so sprechen wir von dynamischen Modellen (für die lineare Optimierung sind die Transformationsregeln der Simplex).

evaluation: um Fragen an das System zu beantworten muss das Modell gelöst oder ausgewertet werden. Dies kann analytisch (z. B.: Warteschlangentheorie), numerisch oder mittels Simulation (Experimente) erfolgen.

DEFINITION SIMULATION

Ist der Prozess, ein Modell zu entwerfen und Experimente mit diesem Modell durchzuführen, zu dem Zweck entweder die Reaktion (behavior) des Systems zu verstehen oder (innerhalb der Grenzen, die von gewissen Kriterien gesetzt sind) verschiedene Strategien/Alternativen für den Betrieb des Systems (Systemablauf) zu bestimmen. (Banks/Carcar)

WARUM SIMULATION?

In anderen Worten: warum ein Modell entwickeln und untersuchen? Die Antwort ist am leichtesten zu geben im Hinblick auf den Standardablauf wissenschaftlicher Untersuchungen:

1. Ein System beobachten.
2. Eine Hypothese aufstellen, die versucht, die Beobachtungen zu erklären.
3. Eine Vorhersage über das zukünftige Systemverhalten machen auf Grund der Hypothese.
4. Experimente durchführen, um die gemachte Vorhersage zu testen.

Beispiel:

zu 1.

Das System existiert nicht oder ist nur, schwer bzw. mit enormen Kosten zu beobachten.
z. B.: Raumschiffe, Raumkapseln, Technische Einrichtungen, die erst zu errichten sind, ökonomische Systeme

zu 2.

Das System ist so komplex, dass es unmöglich ist, sein Verhalten mit einigen mathematischen Gleichungen zu erklären.
z. B.: ökonomische Systeme, komplexe Warteschlangensysteme

zu 3.

Das entwickelte mathematische Modell ist so komplex, dass es nicht mehr analytisch gelöst werden kann. Somit können keine Vorhersagen gemacht werden.
z. B.: Warteschlangen, Reihenfolgeprobleme, Integrationsprobleme

zu 4.

Wie bei Punkt 1 nur anstelle von „beobachten“, --> „experimentieren“.

Dass es gewiss noch andere Gründe gibt, Simulation einzusetzen als nur den des Auswegs, wenn die gewöhnliche Methode versagt, ist klar. Wir werden im Laufe der weiteren Vorlesung gewiss noch mehr Einblick gewinnen.

1.3 Zusammenfassung

WAS IST SIMULATION?

Simulation stellt mit Hilfe von Simulationssprachen ein Instrumentarium zur Verfügung,

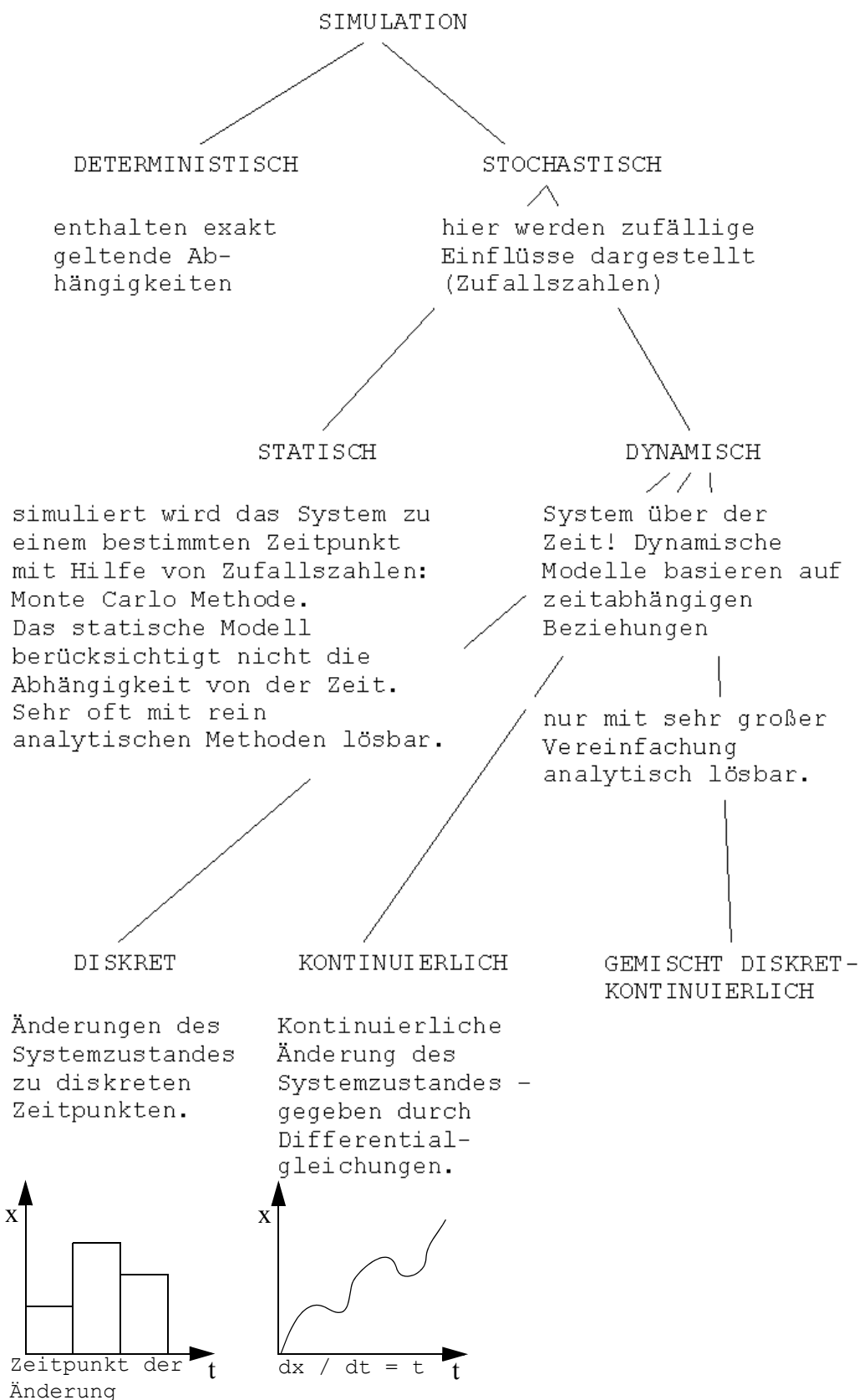
das es ermöglicht in Modellen die Wirklichkeit zu betrachten und über die Zusammenhänge und über das Verhalten der Wirklichkeit Auskunft zu geben.

WAS SOLL SIMULATION?

Simulation soll die **zeitaufwendigen**, **teuren**, oft **gefährlichen** und oft auch **unmöglichen** Untersuchungen über ein System auf den Computer übertragen und dort die Untersuchungen und Messungen durchführen.

2. Simulationskonzepte

Abhängigkeiten, welche zur Änderung des Systemzustandes führen



Bemerkung:

Statistische Modelle, wie zum Beispiel die Modelle der linearen und nichtlinearen Planungsrechnung sind meist auch deterministisch und deshalb unter Umständen leichter mit rein analytischen Methoden lösbar.

Umgekehrt sind dynamische, stochastische Modelle nur in grober Vereinfachung analytisch lösbar (zum Beispiel: Warteschlangen, Lagerhaltung).

Zusätzlich zur Wahl der Zeitvorrückung können drei grundsätzlich verschiedene Simulationskonzepte

- Ereignis-Orientierung
- Aktivitäts-Orientierung
- Prozess-Orientierung

angeführt werden. Jedoch ist keines der drei Konzepte den beiden anderen klar überlegen. Vielmehr treten die Vorteile und die Nachteile je nach dem zu lösenden Problem auf.

2.1 Ereignis-Orientierung (Transaktionsorientiert)



Die Transaktion (Auto, Schiff) ist aktiv - die Aktion (Equipment) ist passiv (Anlegestelle, Verkehrsampel, Zapfsäule).

Jeder zeitabhängige Vorgang vollzieht sich als momenthaftes Geschehen (auf der Simulations - Zeitachse), sozusagen als Ereignis. Ereignisse werden im Modell terminlich festgelegt, und zwar auf jenen Zeitpunkt, an dem - wie im Modell bekannt - alle dafür notwendigen Voraussetzungen erfüllt sein werden. Dabei ist zwischen Primär- und Konditional-Ereignissen zu unterscheiden. Ein Exekutivprogramm legt automatisch die Reihenfolge der geplanten Ereignisse fest, so wie diese in der simulierten Zeit ablaufen.

Das transiente Element (in unseren Beispielen das Schiff oder das Auto) wird durch das System geschleust (GASP, SIMSCRIPT, GPSS).

2.2 Aktivitäts-Orientierung

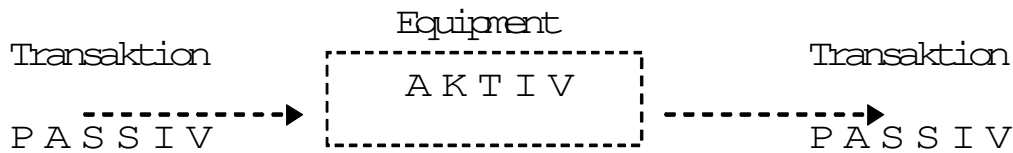
Zeitabhängige Vorgänge vollziehen sich ebenfalls als momenthaftes Geschehen (auf der Simulations - Zeitachse). Diese Ereignisse werden aber nicht direkt im Modell terminisiert. Anstelle dessen werden die Bedingungen angegeben, unter welchen das Ereignis stattfinden kann.

Das Executiv-Programm überprüft nun nach jeder Zeitvorrückung das Vorliegen von Bedingungs-Kombinationen. Sind diese erfüllt, wird das entsprechende Ereignis zum gegenwärtigen Simulationszeitpunkt ausgelöst. Hauptgewicht ist also nicht das nächste Ereignis festzulegen, sondern Aktivitäten durch entsprechende Ereignis-Routinen zu beginnen oder zu beenden.

Vorteile ergeben sich gegenüber der Ereignis - Orientierung

- wenn verschiedene Aktivitäten gleichzeitig beginnen
- wenn Aktivitäten andere Aktivitäten beeinflussen und rückwirkend sind
- wenn eine Aktivität die Länge einer anderen verändert

2.3 Prozess-Orientierung



Die Transaktion (Auto, Schiff) ist passiv - die Aktion (Equipment) ist aktiv (Anlegestellen, Verkehrsampel, Zapfsäule).

Beispiel 2: „Hafen“

Die Anlegestelle entladet die Schiffe.

Die Vorteile der Ereignis-Orientierung (Effizienz der Ereignis-Anordnung) und der Aktivitäts-Orientierung (präzise Darstellung der Aktivitätsbedingungen) sind bei dieser Methode kombiniert. Als Prozess wird dabei eine Folge von Ereignissen (aktiven Phasen) unterbrochen von passiven Phasen (Warten auf Zeitablauf oder Eintreten bestimmter Bedingungen) verstanden. Die Ereignisse sind innerhalb eines Prozesses und mit Ereignissen anderer Prozesse verknüpft. (SIMULA).

- **Direkt:**
Ein Prozess kann einen anderen starten, auslösen und nach Unterbrechung aktivieren. Er kann einen anderen und sich selbst modifizieren, verzögern, unterbrechen und beenden.
- **Indirekt:**
Ein Prozess kann bestimmte Systembedingungen schaffen, auf welche ein anderer Prozess gewartet hat.

Prozesse haben - wie andere Modellkomponenten folgende Eigenschaften:

- Sie sind durch Attribute gekennzeichnet
- Sie existieren in simulierter Zeit, transient, permanent oder zyklisch
- Kopien gleichartiger Prozesse, die gleichzeitig existieren, gehören einer gemeinsamen Klasse an.

Für den Simulationsmodell-Ersteller liegen die Vorteile der Prozess-Orientierung in einer Reduktion des Programm-Overheads (knappere und konzentriertere Darstellung der Modell-Logik), wobei das Exekutivprogramm gegenüber den anderen Methoden komplexere und umfangreichere Funktionen übernimmt.

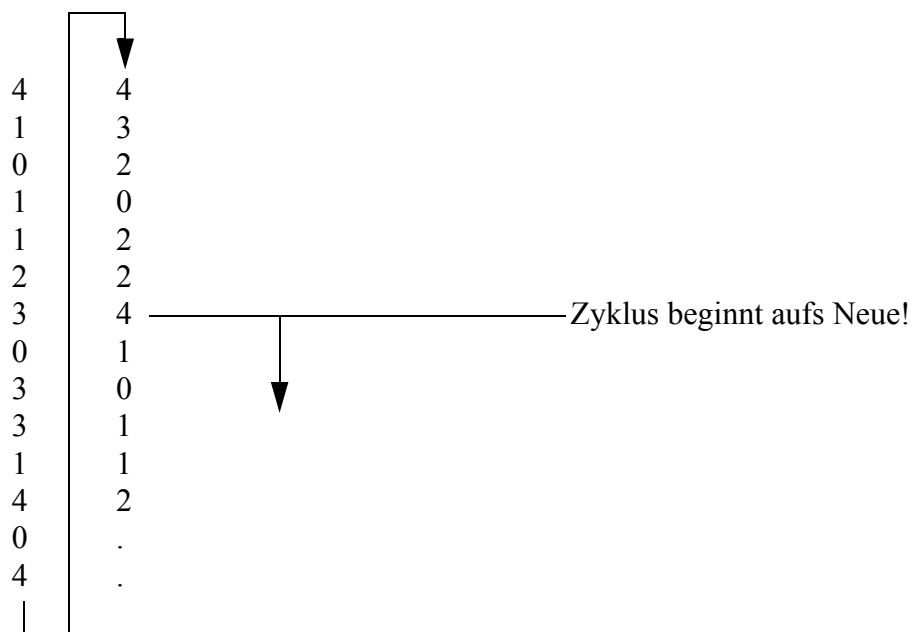
3. Zufallszahlengeneratoren zur Erzeugung gleichverteilter Zufallszahlen

3.1 Additive kongruentielle Methode (Fibonacci-Generator)

$$v_{k+1} = v_k + v_{k-j} \pmod{m} \quad j \dots \text{frei wählbar}$$

Falls $j = 1$ entstehen die Fibonacci-Zahlen (jede Zahl ist Summe der beiden vorangehenden; nach Definition), je nach m entstehen rasch Perioden (Zyklen).

Beispiel: $m = 5$



3.2 Multiplikative kongruentielle Methode

$$v_{k+1} = a * v_k \pmod{m} \quad a > 0, v_1 \dots \text{frei wählbar}$$

Nachteil: falls 0 entsteht kommt er nicht mehr weg. Um diesen Mangel zu beheben wird ein konstanter Term addiert. (Siehe nachfolgendes Kapitel)

3.3 Gemischt kongruentielle Methode

(Linearer Kongruenzgenerator von D.H. Lehmer und A. Rotenberg)

$$v_{k+1} = a * v_k + c \pmod{m} \quad a, c > 0, v_1 \dots \text{frei wählbar}$$

Die Subroutine RANDU:

Wie bestimmt man den Startwert? (multiplikative kongruente Methode). Auf „Binär-Computern“ wähle $m = 2^b \Rightarrow$ größtmögliche Zahlen in interner Darstellung (mit $b =$ Anzahl der Bits pro Wort)

Die Zahlentheorie sagt dann, dass:

$$\text{max. Periode} = 2^{b-2} = m/4$$

Der günstigste Wert für a soll bei $2^{b/2} \pm 3$ liegen. Für eine 16-Bit Maschine wäre $a = 2^8 \pm 3 = 256 + 3 = 259$.

Bei einer 32-Bit Maschine wäre $a = 2^{16} \pm 3 = 65536 + 3 = 65539$.

Mit

IX ... nicht normierte ganze Zufallszahl v_k (Eingabe)
IY ... nicht normierte ganze Zufallszahl v_{k+1} (Ausgabe)
YFL ... normierte Zufallszahl (0,1) (Ausgabe)

schreibt sich die Subroutine RANDU(IX, IY, YFL) für eine 32 Bit Maschine wie folgt:

```
SUBROUTINE RANDU(IX, IY, YFL)
  IY = IX*65539
  IF(IY) 5,6,6
  5  IY = IY*(-1)
  6  YFL = IY
  YFL = YFL / größte ganze Zahl
  RETURN
  END
```

3.4 Mid-Square Methode

Sehr bekannt ist die Mid-Square-Methode. Hier wird eine Zahl von mindestens vier Dezimalstellen quadriert. Von dem Ergebnis bilden die vier mittleren Stellen die Pseudo-Zufallszahl, die zur Berechnung der nächsten Zufallszahl wieder quadriert wird. Beispielsweise ist $3123^2 = 09753129$. Die vier mittleren Stellen sind 7531. Aus ihnen erhält man $7531^2 = 56715961$ und damit die nächste Zufallszahl $7159^2 = 51251281$ usw. Dieses Verfahren muss für die praktische Anwendung durch Störgrößen beeinflusst werden, da es sonst leicht zu kurzen Zyklen führt.

Beispiel:

```
3600  $\Rightarrow$  12 9600 00
9600  $\Rightarrow$  92 1600 00
1600  $\Rightarrow$  02 5600 00
5600  $\Rightarrow$  31 3600 00
3600  $\vdots$ 
bzw. 2500  $\Rightarrow$  06 2500 00
      2500  $\Rightarrow$  06 2500 00
      2500  $\Rightarrow$  06 2500 00
       $\vdots$ 
```

4 Test von Zufallszahlen bzw. erhobenen Daten

4.1 χ^2 -Test auf Gleichförmigkeit

Unterteile Intervall $(0,1)$ in n Klassen. Vergleiche:

B_i ... beobachtete Häufigkeit in Klasse i mit

E_i ... erwartete Häufigkeit in Klasse i

Ermittle:

$$\chi^2 = \sum_{i=1}^n \frac{(B_i - E_i)^2}{E_i} \quad \text{mit } n-1 \text{ Freiheitsgraden}$$

Verwirf H_0 wenn $\chi^2 > \text{Tabellenwert}(n-1, p\%)$

Auch für Test der gleichen Häufigkeit der Ziffern 0 - 9 anwendbar

Beispiel:

Sind folgende 20 Zahlen 6, 0, 6, 2, 3, 4, 7, 2, 9, 0, 5, 2, 6, 4, 1, 6, 7, 5, 7, 9 gleichverteilt?

Man unterteilt das Intervall 0 - 9 in $n = 10$ Klassen, die erwartete Häufigkeit pro Klasse wäre bei 20 Zahlen = 2

Klasse	0	1	2	3	4	5	6	7	8	9
B_i	*	*	*	*	*	*	*	*		*
	*		*		*	*	*	*		*
			*				*	*		
			*				*			
E_i	2	2	2	2	2	2	2	2	2	2
$\frac{(B_i - E_i)^2}{E_i}$	0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	0	0	2	$\frac{1}{2}$	2	0

$\Sigma = 6 = \chi^2$

Für $n - 1 = 9$ Freiheitsgrade gilt bei einer Signifikanz $\alpha = 5\%$

$$P(\chi^2 \leq c) = 0,95$$

$$\downarrow \quad \downarrow$$

$$(6 \leq 16,92)$$

\Rightarrow Die Hypothese: „Die Zahlen sind gleichverteilt“ kann nicht verworfen werden.

4.2 χ^2 -Test auf paarweise Unabhängigkeit

Unterteile die Fläche (1×1) in $n \times n$ Zellen. Untersuche jeweils Paare aufeinanderfolgender Zufallszahlen, die sozusagen Punkte auf dieser Fläche darstellen.

Vergleiche:

B_{ij} ... beobachtete Häufigkeit in Zelle ij mit

E_{ij} ... erwartete Häufigkeit in Zelle ij

Ermittle mit $n^2 - n$ Freiheitsgraden

$$\chi^2 = \sum_{i=1}^n \sum_{j=1}^n \frac{(B_{ij} - E_{ij})^2}{E_{ij}} - \sum_{i=1}^n \frac{(B_{i.} - E_{i.})^2}{E_{i.}}$$

wobei $B_{i.} = \sum_{j=1}^n B_{ij}$ und $E_{i.} = \sum_{j=1}^n E_{ij}$

Verwirf H_0 wenn $\chi^2 >$ Tabellenwert ($n^2 - n, p\%$)

Dieser Test kann auch auf jeweils 3, 4, etc. aufeinanderfolgende Zufallszahlen ausgedehnt werden.

Der Test berücksichtigt paarweise Unabhängigkeit nach Ausschalten eventueller Ungleichmäßigkeiten, die bereits im Test 4.1 getestet wurden.

Beispiel:

Es werden 3x3 je 6 Zufallszahlen im Intervall (0,2) generiert.
Sind die Elemente paarweise unabhängig?

0 0 0 2 0 1	1 2 0 2 2 0	1 1 0 1 2 1
1 2 1 2 0 2	2 1 0 0 0 1	0 2 2 2 2 1
0 1 2 2 2 1	2 0 2 2 0 2	2 0 2 0 1 0

Zählt man in den 9 Feldern die Anzahl 0, 1, 2 so ergibt sich:

Klasse	0	1	2	0	1	2	0	1	2		0	1	2
B_{ij}	4	1	1	2	1	3	1	4	1	$B_{i.}$	7	6	5
E_{ij}	2	2	2	2	2	2	2	2	2	$E_{i.}$	6	6	6
											1/6	0	1/6
													$\Sigma=2/6$
$\frac{(B_{ij} - E_{ij})^2}{E_{ij}}$	2	1/2	1/2	0	1/2	1/2	1/2	2	1/2	$\Sigma=7$			

Klasse	0	1	2	0	1	2	0	1	2		0	1	2
B _{ij}	1	2	3	3	2	1	1	1	4	B _{2.}	5	5	8
E _{ij}	2	2	2	2	2	2	2	2	2	E _{2.}	6	6	6
											1/6	1/6	4/6
											Σ=1		
$\frac{(B_{ij} - E_{ij})^2}{E_{ij}}$	1/2	0	1/2	1/2	0	1/2	1/2	1/2	2	Σ=5			

Klasse	0	1	2	0	1	2	0	1	2		0	1	2
B _{ij}	1	2	3	2	0	4	3	1	2	B _{3.}	6	3	9
E _{ij}	2	2	2	2	2	2	2	2	2	E _{3.}	6	6	6
											0	9/6	9/6
											Σ=3		
$\frac{(B_{ij} - E_{ij})^2}{E_{ij}}$	1/2	0	1/2	0	2	2	1/2	1/2	0	Σ=6			

$$\chi^2 = \left(7 - \frac{2}{6}\right) + (5 - 1) + (6 - 3) = 13\frac{4}{6} = 13,66$$

$n^2 - n = 9 - 3 = 6$ Freiheitsgrade, $p = 0,95 \% >$ Tabellenwert 12,59

$\chi^2 = 13,66 > 12,59 \Rightarrow H_0$: Die Annahme „die Zahlen sind paarweise unabhängig“ muss verworfen werden.

4.3 χ^2 - Test auf Unabhängigkeit von Zufallszahlenfolgen

In verschiedenen Simulationsläufen werden oft verschiedene Folgen von Zufallszahlen desselben Zufallszahlengenerators verwendet. (Achtung: Genau dies soll vermieden werden, verwenden Sie für unterschiedliche Aktivitäten unterschiedliche Generatoren - d.h. Generatoren, welche einen unterschiedlichen Startwert besitzen.)

Untersuche, ob die:

Folge u_1, u_2, u_3, \dots unabhängig ist von der

Folge $u_{\mu+1}, u_{\mu+2}, u_{\mu+3}, \dots$

wobei $\mu \dots$ Verschiebungsintervall der Zahlenfolgen ist (vom Benutzer gewählt).

Die Wertepaare $(u_i, u_{\mu+i})$ bilden Punkte in der Fläche (1×1) , die in n^2 Zellen unterteilt ist. Mit χ^2 - Test wird (wie in 4.2) die gleichförmige Verteilung getestet, das heißt ob die gemeinsame Verteilung das Produkt der individuellen Verteilungen ist.

4.4 χ^2 - Test auf Normalverteilung

Beispiel:

An 6 Arbeitstagen haben wir zu unterschiedlichen Zeiten jeweils während ca 180 Minuten die Bearbeitungszeiten gemessen. Für Maschine 1 ergab sich folgendes Bild (Bearbeitungszeit von jeweils 8 hintereinanderliegenden Jobs in Minuten).

Tag					
Montag	Dienst.	Mittw.	Donnerst.	Freitag	Samstag
23,9	23,8	23,3	24,4	23,3	22,9
22,0	22,8	22,9	22,8	22,9	22,3
23,8	23,5	23,8	23,1	24,1	24,0
21,5	23,1	22,9	22,5	23,0	22,0
24,6	23,4	23,5	22,6	23,3	23,2
21,7	23,0	23,3	21,6	23,2	22,8
23,9	23,3	23,5	23,4	23,8	23,7
23,1	23,0	22,3	22,5	21,8	23,0
184,5	185,9	185,5	182,9	185,4	183,9

Übertragen in eine Häufigkeitstabelle ergibt sich folgendes Bild:

Wert Diagramm (mit Häufigkeit)

statistische Werte

Wert	Häufigkeit	Diagramm	statistische Werte
21.5	1	■	
21.6	0		
21.7	1	■	
21.8	1	■	
21.9	0		
22.0	2	■	
22.1	0		
22.2	0		
22.3	2	■	
22.4	0		
22.5	2	■	
22.6	2	■	
22.7	0		
22.8	3	■	$\sigma_{n-1} = 0,6814897$
22.9	4	■	$\mu = 23.10625$
23.0	4	■	
23.1	3	■	
23.2	2	■	$\sigma_{n-1} = 0.6814897$
23.3	5	■	$\mu = 23,10625$
23.4	2	■	
23.5	3	■	
23.6	0		
23.7	1	■	
23.8	4	■	
23.9	2	■	
24.0	1	■	
24.1	1	■	
24.2	0		
24.3	0		
24.4	1	■	
24.5	0		
24.6	1	■	

Das Diagramm lässt eine Normalverteilung erwarten, wir testen mittels Chi-Quadrat Test auf Normalverteilung.

Beim Chi-Quadrat-Test vergleicht man beobachtete mit erwarteten Häufigkeiten, wobei gefordert ist, dass pro Häufigkeitsklasse mindestens 5 Werte erwartet werden. Wir unterteilen daher unsere 48 Messgrößen in 8 Beobachtungsklassen, die so dimensioniert sein sollen, dass man pro Klasse 6 Beobachtungswerte ($48 \cdot 1/8 = 6$) erwartet.

Aus der Statistik ist bekannt, dass bei $z = 2.00 \Rightarrow D(z) = 95,45$.

Bzw. im Intervall $\mu \pm z\sigma$ liegen $D(z)$ aller Werte. In unserem Beispiel erwarten wir daher im Intervall $(23.11 \pm 2 \cdot 0.6815)$ $D(z) = 95.45\%$ aller Werte.

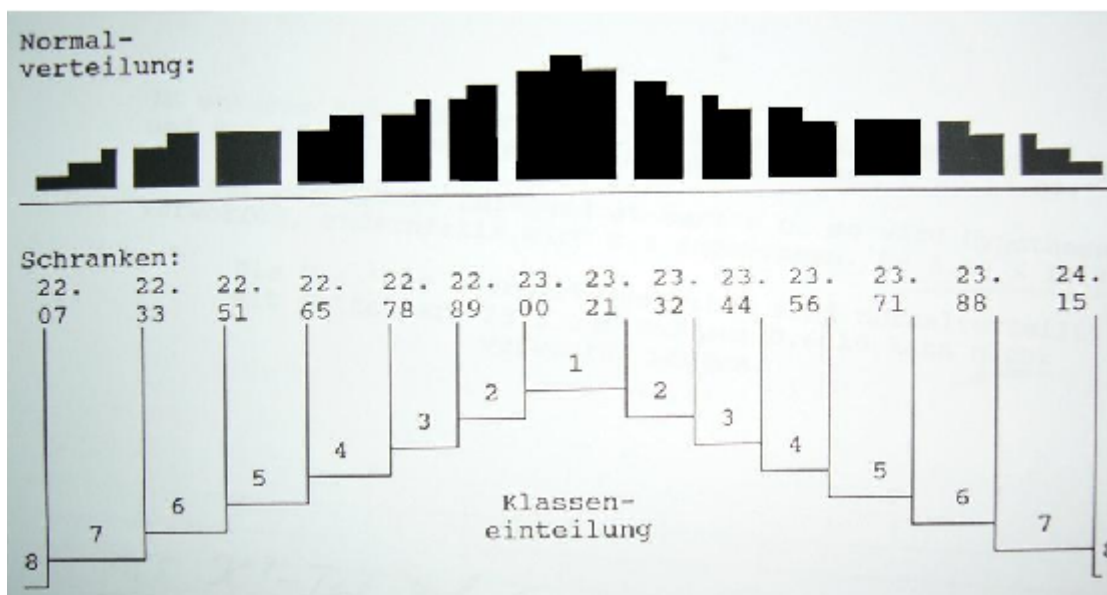
Wir bestimmen nun $D(z)$ derart, dass stufenweise eine Erhöhung um 6 Werte erfolgt und erhalten so jene Klassenbreiten für die gilt, dass pro Klasse 6 Messwerte erwartet werden. In der Tabelle für Normalverteilung ergibt sich für $D(z) = 0.1192$ ein $z = 0.15$ für $D(z) = 0.1271$ ein $z = 0.16$. Durch Interpolation erhalten wir für $D(z) = 0.125$ ein $z = 0.157$.

Berechnung	zu D(z) gehöri- ger Tabellen- wert der Normalvert.
48 * D(z) = 6 → D(z) = 0,125	z = 0,157
48 * D(z) = 12 → D(z) = 0,250	z = 0,319
48 * D(z) = 18 → D(z) = 0,375	z = 0,490
48 * D(z) = 24 → D(z) = 0,500	z = 0,675
48 * D(z) = 30 → D(z) = 0,625	z = 0,887
48 * D(z) = 36 → D(z) = 0,750	z = 1,150
48 * D(z) = 42 → D(z) = 0,875	z = 1,534

Nach der Formel $\mu \pm (z * \sigma_{n-1})$ bestimmen wir nun die entsprechenden Größenklassen der Bearbeitungszeiten:

		Schranken	
Z	Z * σ_{n-1}	$\mu - Z * \sigma_{n-1}$	$\mu + Z * \sigma_{n-1}$
0,157	0,10699	22,999	23,213
0,319	0,21729	22,888	23,323
0,490	0,33392	22,772	23,440
0,675	0,46000	22,646	23,566
0,887	0,60448	22,501	23,710
1,150	0,78371	22,322	23,889
1,534	1,04540	22,060	24,151

Die Schranken zur Veranschaulichung auf der Zahlengeraden:



Wir vergleichen nun die beobachtete Anzahl pro Klasse mit der erwarteten Anzahl und berechnen den Chi-Quadrat Wert:

Klasse	Schranken	Beobachtet Σ		Erwartet	$(B_k - E_k)^2 / E_k$
1	23.00 - 23.21	9	9	6	9/6
2	22.89 - 22.99	4			
	23.22 - 23.32	5	9	6	9/6
3	22.78 - 22.88	3			
	23.33 - 23.44	2	5	6	1/6
4	22.65 - 22.77	0			
	23.45 - 23.56	3	3	6	9/6
5	22.51 - 22.64	2			
	23.57 - 23.71	1	3	6	9/6
6	22.33 - 22.50	2			
	23.72 - 23.89	4	6	6	0
7	22.07 - 22.32	2			
	23.90 - 24.15	4	6	6	0
8	0 - 22.06	5			
	24.16 - ∞	2	7	6	1/6
Chi-Quadrat Wert: $\Sigma (B_k - E_k)^2 / E_k$					38/6 = 6.33

Wir bestimmen nun:

$$P[\text{Chi-Quadrat-Wert} \leq c] = 1 - a$$

In unserem Beispiel ergibt sich mit $a = 5\%$ Irrtumswahrscheinlichkeit und $K-1 = 7$ Freiheitsgraden der Tabellenwert c zu 14.07.

Ist der berechnete Chi-Quadrat-Wert $> c$, so wird Hypothese H_0 verworfen, andernfalls wird sie angenommen. Da $6.33 \leq 14.07 \Rightarrow$

Die H_0 : „Die Bearbeitungszeiten sind normalverteilt mit Mittelwert 23.1 und Varianz 0.6815 kann nicht verworfen werden.“

4.5 χ^2 - Test auf Exponentialverteilung

Schranken und erwartete Häufigkeit in % für eine Exponentialverteilung 1 (bzw. \bar{x}) sind bei Kapitel 5.1 angegeben. Der zugehörige χ^2 - Test verläuft wie vorhin erläutert.

4.6 Zusammenfassung

Aus Erfahrung ist bekannt, dass die multiplikative kongruentielle Methode Zufallszahlen generiert, die die genannten Tests weitgehend erfüllt. Trotzdem soll der Benutzer von Zufallszahlengeneratoren deren Eignung für eine spezielle Anwendung entsprechend den Erfordernissen dieser Anwendung überprüfen. Insbesondere ist es empfehlenswert für verschiedene Prozesse Ankunftsprozess 1, 2, ... Serviceprozess 1, 2, ... jeweils eigene Prozessgeneratoren zu verwenden, um nicht über den Zufallszahlengenerator einen Einfluss des Ankunftsintervalls auf das Serviceintervall zu bekommen.

Für erhobene Daten sind diese Tests genauso zu verwenden. Es ist jedoch vorab zu bestimmen ob überhaupt genügend Daten erhoben wurden. Zur Bestimmung des Stichprobenumfangs von

Daten aus beliebigen Verteilungen siehe die Kapitel 7.2 Simulationsdauer bzw. Umfang der Datenerhebung.

5. Prozessgeneratoren

In einem Simulationsmodell sind Zufallszahlen, die den angenommenen theoretischen Verteilungen gehorchen, zu generieren. Die Zufallszahlen der anderen Verteilungen werden mittels Prozessgenerator generiert.

Folgende Typen werden angewandt:

1. Inverse Transformation
2. Methode der gespeicherten Datenblöcke
3. Methode der äquivalenten Transformation
4. Methode der Reproduktion des stochastischen Prozesses
5. Mischformen

5.1. Methode der inversen Transformation

Es sei U eine in $(0,1)$ gleichverteilte Zufallsvariable und $F(x)$ eine beliebige Verteilungsfunktion, dann besitzt die durch

$$u_0 = F(x_0)$$

beziehungsweise

$$F^{-1}(u_0) = \underbrace{F^{-1}(F(x_0))}_{x_0} \Rightarrow$$

$$x_0 = F^{-1}(u_0)$$

definierte Zufallsvariable die gewünschte Verteilungsfunktion.

Beispiel: $y = \text{Exponentialfunktion mit Mittelwert } 1$

$$U = \underbrace{F(x)}$$

$U = \text{Exponentialfunktion}$

$$U = 1 - e^{-x}$$

$$e^{-x} = 1 - U$$

Wegen: $(b^n = a) \Rightarrow (n = {}^b \log a)$

Beispiel: $(10^2 = 100) \Rightarrow (n = {}^{10} \log 100) \Rightarrow (n = 2)$

Gilt:

$$e^{-x} = (1 - U) \Rightarrow -x = {}^e \log(1 - U)$$

$$-x = \ln(1 - U)$$

$$x = -\ln(1 - U)$$

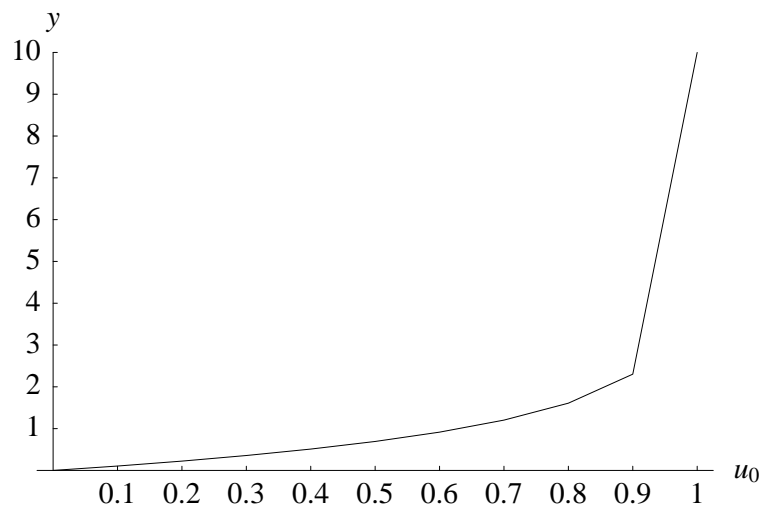
bzw die Variable x_0

$$x_0 = -\ln(1 - u_0)$$

So berechnen sich für $u_0 = 0.0, 0.1, \dots, 1.0$ folgende Werte:

u_0	y
0.0	0.0
0.1	0.105
0.2	0.223
0.3	0.357
0.4	0.511
0.5	0.693
0.6	0.916
0.7	1.204
0.8	1.609
0.9	2.303
1.0	∞

Bzw im Diagramm



Die Kurve hat 2 Mängel:

- Der Wert ∞ verursacht bei jedem Rechner Overflowbedingungen und Abbruch.
- Verwendet man die später geschilderte Methode der gespeicherten Datenblöcke, so sind nur die zu den Werten $0.0, 0.1, 0.2, \dots, 0.9, 1.0$ gehörigen Funktionswerte vorhanden. D.h. besonders im Bereich 0.9 bis 1.0 sind die Werte viel zu ungenau.

Zur Behebung von Mangel a):

Beim Zufallszahlengenerator U werden nur Zahlen im Intervall $(0, 0.99995)$ erlaubt.

Zur Behebung von Mangel b):

Falls wir die Exponentialverteilung nicht durch inverse Transformation ($x_0 = -\ln(1 - u_0)$) generieren, sondern die Methode der gespeicherten Datenblöcke verwenden, so haben sich folgende 24 Stützstellen als ausreichend erwiesen:

Die 24 Stützstellen für die Exponentialverteilung mit Mittelwert 1 lauten (Werte zwischen den Stützstellen werden interpoliert):

0., 0. / 0.1, 0.105 / 0.2, 0.223 / 0.3, 0.357 / 0.4, 0.511 / 0.5, 0.693 / 0.6, 0.916 / 0.6321, 1.0 / 0.7, 1.204 / 0.8, 1.609 / 0.86466, 2.0 / 0.9, 2.303 / 0.93, 2.659 / 0.95021, 3.0 / 0.96, 3.219 / 0.97, 3.507 / 0.98168, 4.0 / 0.99, 4.605 / 0.99329, 5. / 0.99752, 6. / 0.99909, 7.0 / 0.99966, 8. / 0.99988, 9. / 0.99995, 10. /

Diese Stützstellen sind gleichzeitig auch Häufigkeiten mit denen man mittels χ^2 -Test auf Exponentialverteilung testen kann.

Was ist falls die Exponentialverteilung einen Mittelwert $\neq 1$ hat ?

Da X in der Exponentialverteilung für $L(\text{ambda}) * t$ steht mit:

L ... mittlere Anzahl der Ankünfte in einem Zeitintervall der Länge Δt

t ... betrachtetes Zeitintervall der Länge Δt

gilt:

$$L \cdot t = -\ln(1 - u_0)$$

damit ergibt sich für die einzelne exponentialverteilte Zwischenankunftszeit:

$$t = \frac{1}{L} \cdot (-\ln(1 - u_0))$$

Beispiel:

Im Mittel kommt alle 18 Minuten 1 Person an eine Servicestation, wobei die Ankunftsverteilung einer Exponentialverteilung gehorcht. Berechnen Sie einige zugehörige Werte aus dieser Verteilung.

Nehmen wir an, unser Zeitintervall $\Delta t = 1$ Minute, so ergibt sich für:

$$L \text{ ... mittlere Anzahl der Ankünfte pro Minute} = \frac{1}{18}$$

Damit berechnen sich die exponentialverteilten Zwischenankunftszeiten zu:

$$t = \frac{1}{\left(\frac{1}{18}\right)} \cdot (-\ln(1 - u_0))$$

So ergeben sich für $u_0 = 0.0, 0.1, 0.2, \dots, 1.0$ folgende Werte von t:

u_0	$18 \cdot (-\ln(1 - u_0))$	t
0	$18 \cdot 0.0$	0.0
0.1	$18 \cdot 0.105$	1.89
0.2	$18 \cdot 0.223$	4.016
0.3	$18 \cdot 0.357$	6.426
.	.	.
.	.	.
.	.	.

u_0	$18 \cdot (-\ln(1 - 0))$	t
0.99995	$18 \cdot 9.21$	165.786
1.0	$18 \cdot \infty$	∞

Bemerkungen:

1) In 0% der Fälle vergehen 0 Minuten bis eine Person eintrifft, d.h. es kommen niemals zeitgleich 2 Personen an.

2) In 10% der Fälle vergehen 1,89 Minuten bis eine Person eintrifft. In weiteren 10% der Fälle vergehen 4,016 Minuten bis eine Person eintrifft.

Man sieht bereits hier eine Eigenschaft der Exponentialverteilung, es kommen relativ oft kleine Werte vor, die wenigen großen Werte können jedoch ein Simulationsmodell destabilisieren. Aus diesem Grund wird u_0 auf ≤ 0.99995 beschränkt.

Nur zur Erwähnung:

Beispiel: Cauchyverteilung

$$U = \underbrace{F(X)}$$

$$U = \text{Cauchy}$$

$$U = \frac{1}{2} + \frac{1}{\pi} \operatorname{atan} X$$

$$X = \tan \left[\pi \cdot \left(U - \frac{1}{2} \right) \right]$$

Beispiel: Logistisch

$$U = \underbrace{F(X)}$$

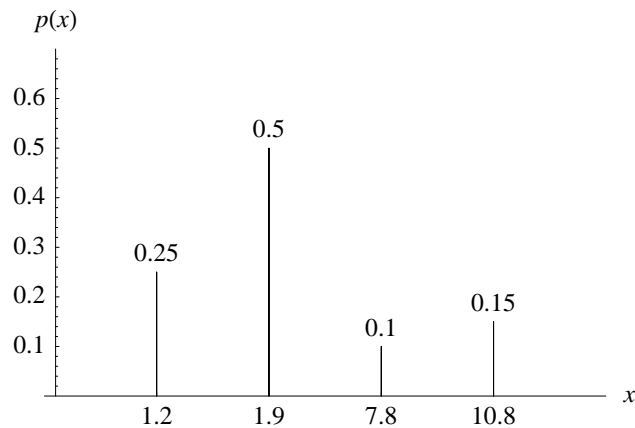
$$U = \text{Logistisch}$$

$$U = (1 + e^{-X})^{-1}$$

$$X = \ln \left(\frac{U}{1 - U} \right)$$

5.2 Die Methode der gespeicherten Datenblöcke

Beispiel: Die folgenden 4 Werte kommen mit folgender Wahrscheinlichkeit vor:



Methode:

Reserviere 100 Speicherplätze

	kumulierte Wahrscheinlichkeit		
Speichere in Adresse	0.00 - 0.24	den Wert	1.2
Speichere in Adresse	0.25 - 0.74	den Wert	1.9
Speichere in Adresse	0.75 - 0.84	den Wert	7.8
Speichere in Adresse	0.85 - 0.99	den Wert	10.1

Generiere Zufallszahl $U(0,1)$

Mittels ersten beiden Ziffern bestimme die Adresse

Inhalt = Zufallszahl

Beachte: Speicherbedarf

Beispiel: Poissonverteilung

Die Wahrscheinlichkeit, dass in einem Zeitintervall der Länge t k -Einheiten (Personen) ankommen, lautet:

$$p_k(t) = \frac{(L \cdot t)^k}{k!} \cdot e^{-Lt}$$

Annahme:

- unser Zeitintervall t (Taktzeit d. Systems) = 10 min.
- die mittlere Anzahl Ankünfte pro Zeiteinheit (= 1 min) sei 0,382 Personen $\Rightarrow L = 0.382$

Dann ergeben sich folgende Wahrscheinlichkeiten:

	$e^{-Lt} = e^{-3.82}$	$\frac{(L \cdot t)^k}{k!}$	
$p_0(t) =$	0,0219278009	*	1 = 0,0219278009
$p_1(t) =$	-''-	*	3.82 = 0.0837641994
$p_2(t) =$	-''-	*	7.2962 = 0.1599896209
$p_3(t) =$	-''-	*	9.2904 = 0.2037201173
$p_4(t) =$	-''-	*	8.872 = 0.194552712
$p_5(t) =$	-''-	*	6.778 = 0.148638272
$p_6(t) =$	-''-	*	4.3156 = 0.0946330332
$p_7(t) =$	-''-	*	2.3551198 = 0.0516425981
$p_8(t) =$	-''-	*	1.124569 = 0.0246593406
$p_9(t) =$	-''-	*	0.477317 = 0.0104665201
$p_{10}(t) =$	-''-	*	0.1823352 = 0.0039982107
			0.9979924253

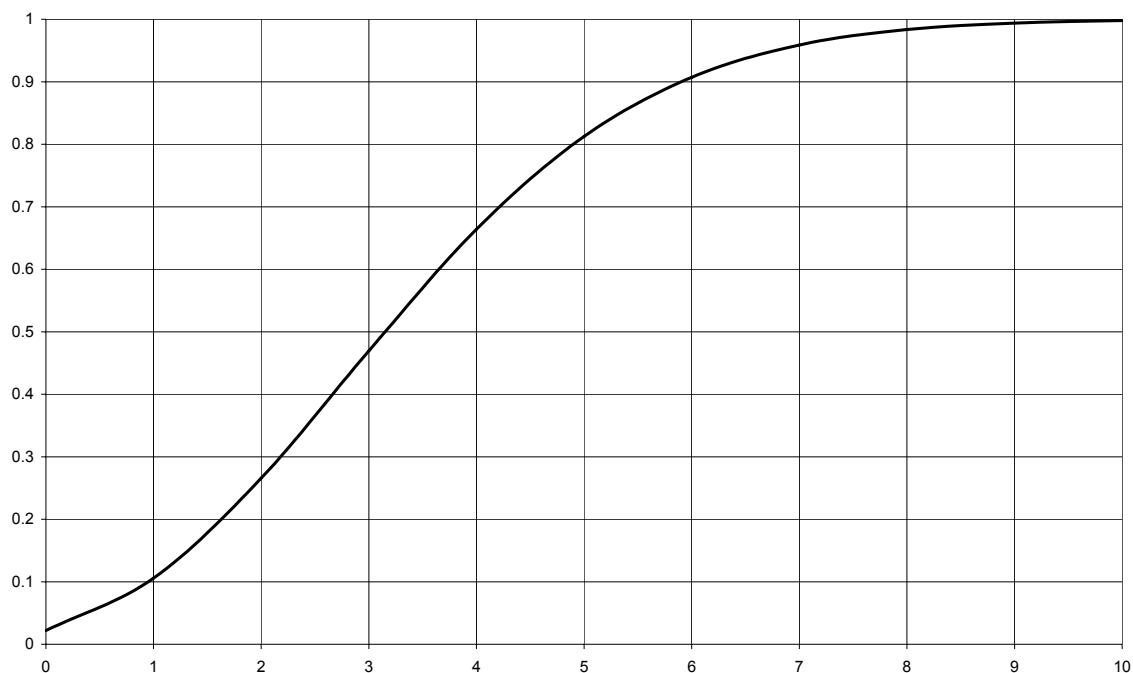
Die Verteilungsfunktion $F(X)$ ergibt sich somit zu:

$$F(x) = \sum_k p_k(t)$$

In unserem Beispiel:

F(0) = 0.0219278009	F(4) = 0.6639544505	F(8) = 0.9835276944
F(1) = 0.1056920063	F(5) = 0.8125927225	F(9) = 0.9939942146
F(2) = 0.2656816212	F(6) = 0.9072257557	F(10) = 0.9979924253
F(3) = 0.4694017385	F(7) = 0.9588683538	...
		F(∞) = 1

Bildet man $F(x)$ auf der Abszisse und x auf der Ordinate ab, so ergibt sich:



Man speichert in die Adresse 0.000 - 0.020 den Wert 0

0.021 - 0.104 den Wert 1

0.105 - 0.265 den Wert 2

⋮

usw.

Generiert die Zufallszahl $U(0,1)$; wobei die ersten beiden Ziffern die Adresse; der Inhalt die Zufallszahl (Anzahl Ankünfte) angeben.

5.3 Methode der äquivalenten Transformation

Zufallszahlen gewisser Verteilungen können generiert werden, indem die Verwandtschaft mit anderen Verteilungen ausgenutzt wird.

Berühmtestes Beispiel:

5.3.1 Poissonverteilung - Exponentialverteilung

Bei poissonverteiletem Input gilt für die Wahrscheinlichkeit, dass in einem Intervall der Länge t 0 Einheiten ankommen:

$$p_0(t) = \frac{(L \cdot t)^0}{0!} \cdot e^{-Lt} = e^{-Lt}$$

Diese Wahrscheinlichkeit ist ident mit der Wahrscheinlichkeit, dass die 1.Einheit nach t Zeiteinheiten ankommt.

$$P(T \geq t) = e^{-Lt}$$

⇒ Die Wahrscheinlichkeit, dass die Zeitspanne bis zum Eintreffen der ersten Einheit $\leq t$ ist

$$= P(T \leq t) = 1 - e^{-Lt} \dots \text{Exponentialverteilung}$$

(dies ist genau die Zeitspanne zwischen 2 Ankünften)

Womit gilt:

Bei poissonverteiletem Input ist die Zeit zwischen zwei Ankünften exponentialverteilt.

Die Anzahl der Ankünfte in einem Zeitintervall der Länge Δt wird nun wie folgt generiert:

Die Zeit zwischen zwei Ankünften = t

$$t = \frac{1}{L} \ln \frac{1}{u_0} \dots \text{Exponentialverteilung (siehe Kapitel 9.1)}$$

Man summiert nun die einzelnen t_i (nach Exponentialverteilung gewonnen) solange auf, bis das Zeitintervall Δt überschritten wird.

Beispiel: $\Delta t = 10$ min

$$t_1 = 3 \text{ min}$$

$$t_2 = 6 \text{ min}$$

$$t_3 = 4 \text{ min}$$

$$\Sigma = 13 \text{ min}$$

im Intervall von 10 min waren 2 Ankünfte

5.3.2 Gleichverteilung - Normalverteilung

a) *Statistische Grundlagen*

Sie Würfeln sechs mal mit einem Würfel (Augenzahl 1 bis 6). Wie groß ist die durchschnittliche Augenzahl?

Es gilt:

$$\mu = \sum_j x_j \cdot f(x_j) \dots \text{Mittelwert einer diskreten Verteilung}$$

Die Zufallsvariable:

X ... Augenzahl beim Wurf eines regelmäßigen Würfels

hat die Wahrscheinlichkeitsfunktion:

$$f(x) = 1/6$$

$$\mu = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3,5$$

Analog ergibt sich die Varianz zu:

$$\sigma^2 = \sum_j (x_j - \mu)^2 \cdot f(x_j) \dots \text{Varianz einer diskreten Verteilung}$$

$$\sigma^2 = (1 - 3,5)^2 \cdot \frac{1}{6} + (2 - 3,5)^2 \cdot \frac{1}{6} + (3 - 3,5)^2 \cdot \frac{1}{6} + (4 - 3,5)^2 \cdot \frac{1}{6} + (5 - 3,5)^2 \cdot \frac{1}{6} + (6 - 3,5)^2 \cdot \frac{1}{6} = 2,91666$$

Für stetige Verteilungen gilt analog:

$$\mu = \int_{-\infty}^{\infty} (x \cdot f(x)) dx$$

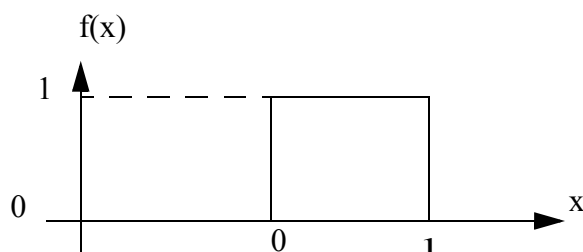
Mittelwert und Varianz einer stetigen Verteilung

$$\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$$

Definiert man im Fall einer Gleichverteilung U im Intervall (0,1) die Wahrscheinlichkeitsfunktion f(x) wie folgt:

$$f(x) = \begin{cases} 1 & \dots \text{für } 0 \leq x \leq 1 \\ 0 & \dots \text{sonst} \end{cases}$$

Also:



Dann ergibt sich für μ

$$\mu = \int_0^1 x \cdot f(x) dx = \int_0^1 (x \cdot 1) dx$$

Wegen:

$$\int x^n dx = \frac{x^{n+1}}{n+1} + C$$

gilt:

$$\int_0^1 x dx = \left. \frac{x^2}{2} \right|_0^1 = \frac{1}{2}$$

Analog gilt für die Varianz σ^2 :

$$\sigma^2 = \int_0^1 \left(\left(x - \frac{1}{2} \right)^2 \cdot 1 \right) dx$$

$$\sigma^2 = \int_0^1 \left(x^2 - x + \frac{1}{4} \right) dx$$

Wegen:

$$\int [f(x) \pm g(x)] dx = \int f(x) dx \pm \int g(x) dx$$

gilt:

$$\sigma^2 = \left. \frac{x^3}{3} - \frac{x^2}{2} + \frac{x}{4} \right|_0^1 = \frac{1}{3} - \frac{1}{2} + \frac{1}{4} = \frac{4}{12} - \frac{6}{12} + \frac{3}{12} = \frac{1}{12}$$

Additionssatz für Mittelwerte

Der Mittelwert einer Summe von Zufallsvariablen, deren Mittelwerte existieren, ist gleich der Summe dieser Mittelwerte.

D.h. ich teile eine Menge von Zufallsvariablen $\{X\}$ in mehrere Teilmengen $\{X_1\} \dots \{X_n\}$, dann gilt:

$$\mu_x = \sum_{i=1}^n \mu_i$$

in unserem Fall:

$$\mu_x = \sum_{i=1}^n \frac{1}{2} = \frac{n}{2}$$

Additionssatz für Varianzen

Die Varianz einer Summe unabhängiger Zufallsvariablen, deren Varianzen existieren, ist gleich der Summe dieser Varianzen.

$$\sigma_x^2 = \sum_{i=1}^n \sigma_i^2$$

in unserem Fall:

$$\sigma_x^2 = \sum_{i=1}^n \frac{1}{12} = \frac{n}{12}$$

b) Die Erzeugung der Normalverteilung aus der Gleichverteilung

Nach dem zentralen Grenzwertsatz (vgl. Kreyszig S 201) gilt:

u_1, \dots, u_n seien unabhängige Zufallsvariable, die alle die gleiche Verteilungsfunktion (also denselben Mittelwert μ und dieselbe Varianz σ) besitzen.

Weiterhin sei $x = \sum_{i=1}^n u_i$.

Dann ist die Zufallsgröße

$$z = \frac{x - \mu_x}{\sigma_x}$$

„asymptotisch normalverteilt“ mit dem Mittelwert 0 und der Varianz 1.

Beweis in L.Schmetterer „Einführung in die math. Statistik“ Springer Verlag, Wien.

In unserem Fall - mit gleichverteilten Zufallszahlen u_i gilt:

$$z = \frac{x - \mu_x}{\sigma_x} = \frac{\sum_{i=1}^k u_i - \frac{k}{2}}{\sqrt{\frac{k}{12}}}$$

Mit: $\frac{1}{\sqrt{\frac{k}{12}}} = \sqrt{\frac{12}{k}}$ ergibt sich:

$$\frac{x - \mu_x}{\sigma_x} = \sqrt{\frac{12}{k}} \cdot \left[\sum_{i=1}^k u_i - \frac{k}{2} \right]$$

$$x = \sqrt{\frac{12}{k}} \cdot \left[\sum_{i=1}^k u_i - \frac{k}{2} \right] \cdot \sigma_x + \mu_x$$

Es wird empfohlen, $k \geq 10$ zu wählen. Bei $k = 12$:

$$\sqrt{\frac{12}{k}} = \sqrt{\frac{12}{12}} = 1$$

spart man sich eine Multiplikation, muss aber in Kauf nehmen, dass die Genauigkeit über $\pm 3\sigma = 99,7\%$ ungenau ist und nur Werte innerhalb von $\pm 6\sigma$ geliefert werden. Dies ist für den praktischen Betrieb voll ausreichend. Im SSP (IBM) wird daher die Formel:

$$x = \left[\sum_{i=1}^{12} u_i - 6 \right] \cdot \sigma_x + \mu_x$$

verwendet. (wobei $6 = 12/2$ ist)

Mit: SUBROUTINE GAUSS (IX,S,AM,V)

IX ... Zufallszahl Randu
S ... Standardabweichung σ_x
AM ... Mittelwert μ_x
V... berechnete normalverteilte Zufallszahl X

5.4 Methode der Reproduktion des stochastischen Prozesses

Gewisse stochastische Prozesse - wie der Bernoulli-Prozess - können leicht im Computer nachgebildet werden. Die Anzahl der Erfolge in einer bestimmten Anzahl von Versuchen wird gezählt und ergibt eine Zufallszahl der binomischen Verteilung.

Beispiel:

Simuliere das Ergebnis $X = x$ (Würfeln einer Sechs trifft bei $n = 10$ Ausführungen x -mal ein“).

Wir unterteilen die Zahlengerade $(0,1)$ in sechs (zwei) Teile

$$\underbrace{0,0 - 0,16 / 0,17 - 0,33 / \dots}_{(0,0-0,833} \quad / \quad \underbrace{0,84 - 1,0}_{/ 0,834 - 1,0)}$$
$$q=5/6 \qquad p=1/6$$

Die Wahrscheinlichkeit des Eintreffens des Ereignisses $= p$ wird durch eine entsprechende Länge auf der Zahlengeraden $(0,1)$, die Gegenwahrscheinlichkeit $(=q)$ durch den Rest der Zahlengeraden wiedergegeben.

Ruft man den Zufallszahlengenerator $U(0,1)$ 10 mal auf und zählt die Anzahl des Eintreffens von Zahlen im Intervall $(0,834;1,0)$ so erhält man die Zufallszahl x der binomischen Verteilung.

Es gilt:

Mittelwert der Binomialverteilung:

$$\mu = n * p$$

Varianz der Binomialverteilung:

$$\sigma^2 = n * p * q$$

Die Wahrscheinlichkeit, dass in einem Zeitintervall der Länge t k Ankünfte stattfinden, erhält man wie folgt:

Annahme:

Δt ... Zeitintervall in dem maximal eine Ankunft stattfindet (beispielsweise 1 min)

t ... 10 min

Frägt man nach der Wahrscheinlichkeit, dass in 10 min zwei Ankünfte stattfinden, gibt es folgende Möglichkeiten der Realisierung:

	Δt_1	Δt_2	Δt_3	Δt_4	Δt_5	Δt_6	Δt_7	Δt_8	Δt_9	Δt_{10}
Möglichkeit 1	p	p	q	q	q	q	q	q	q	q
Möglichkeit 2	p	q	p	q	q	q	q	q	q	q

$$\binom{10}{2} = \frac{10!}{2! \cdot 8!} = 45 \text{ Möglichkeiten}$$

Möglichkeit 1	q	q	q	q	q	q	q	p	q	p
Möglichkeit 2	q	q	q	q	q	q	q	q	p	p

mit: p ... im Intervall Δt_i findet eine Ankunft statt.
 q ... im Intervall Δt_i findet keine Ankunft statt.

Da die einzelnen Ereignisse voneinander unabhängig sind, hat die Form der Realisierung des Ergebnisses $X = x = pqpqq...qq$ die Wahrscheinlichkeit:

$$\underbrace{pp}_{2 \text{ mal}} \underbrace{qqqqqqqq}_{n-2 \text{ mal}} = p^2 \cdot q^{10-2}$$

beziehungsweise allgemein:

$$= p^k \cdot q^{n-k}$$

Da jede der $\binom{10}{2}$ Möglichkeiten dieselbe Wahrscheinlichkeit hat, gilt für 2 Ankünfte:

$$p_2 = \binom{10}{2} p^2 * q^{10-2}$$

bzw. allgemein:

$$p_k = \binom{n}{k} p^k * q^{n-k} \text{ Bernoulli- oder Binomialverteilung}$$

für dieselbe Ankunftsrate, wie bei der Poissonverteilung von Kapitel 5.2 (=0,382 Ankünfte pro Minute):

$$p = 0,382 \text{ bzw. } q = 0,618$$

berechnet sich nach der Formel der Binomialverteilung die Wahrscheinlichkeit, dass $0 \leq x \leq 10$ Ankünfte stattfinden, wie folgt:

$$p_0 = \binom{10}{0} p^0 q^{10-0} = 1 * 1 \quad * 0,0081261484 = 0,00812614$$

$$p_1 = \binom{10}{1} p^1 q^{10-1} = 10 * 0,382 \quad * 0,0131491075 = 0,0502295$$

$$p_2 = \binom{10}{2} p^2 q^{10-2} = 45 * 0,145924 \quad * 0,021276873 = 0,1397162$$

$$p_3 = \binom{10}{3} p^3 q^{10-3} = 120 * 0,055742968 \quad * 0,034428597 = 0,2302982$$

$$p_4 = \binom{10}{4} p^4 q^{10-4} = 210 * 0,0212938138 \quad * 0,055709039 = 0,2491171$$

$$p_5 = \binom{10}{5} p^5 q^{10-5} = 252 * 0,0081342369 \quad * 0,090145152 = 0,184782029$$

$$p_6 = \binom{10}{6} p^6 q^{10-6} = 210 * 0,0031072785 \quad * 0,1458659418 = 0,09518168$$

$$p_7 = \binom{10}{7} p^7 q^{10-7} = 120 * 0,0011869804 \quad * 0,236029032 = 0,03361940$$

$$p_8 = \binom{10}{8} p^8 q^{10-8} = 45 * 0,0004534265 \quad * 0,381924 = 0,00771285$$

$$p_9 = \binom{10}{9} p^9 q^{10-9} = 10 * 0,001732089 \quad * 0,618 = 0,001070431$$

$$p_{10} = \binom{10}{10} p^{10} q^{10-10} = 1 * 0,000661658 * 1 = 0,0000661658$$

Stellt man Bernoulli und Poissonverteilung gegenüber so erhält man:

F(0)	0,0081261484	0,0219278009
F(1)	0,0583557454	0,1056920003
F(2)	0,1980720339	0,2656816212
F(3)	0,4283702956	0,4694017385
F(4)	0,6774874286	0,6639544505
F(5)	0,8622694575	0,8125927225
F(6)	0,9574511395	0,9072257557
F(7)	0,9910705597	0,9588683538
F(8)	0,9988634105	0,9835276944
F(9)	0,9999338415	0,9939942146
F(10)	1,0000000007	0,99799

Das heißt Sie können binomialverteilte Zufallszahlen entweder generieren indem Sie

- den Entstehungsprozess wie eben geschildert nachvollziehen
- die theoretischen Wahrscheinlichkeiten

$$p_k = \binom{n}{k} p^k p^{n-k}$$

ausrechnen und dann mit der Methode der gespeicherten Datenblöcke eine Zahl $U(0,1)$ generieren die einer entsprechenden Wahrscheinlichkeit zugeordnet ist und sie die Anzahl ankommender Personen angibt.

- die Ähnlichkeit einer Verteilung mit einer anderen benutzen, wie soeben bei Binomialverteilung und Poissonverteilung gezeigt.

6. Validitätsprüfung

6.1 Beobachtung und Vergleich mit der Realität

obviously

6.2 Augenscheinliche Gültigkeitsprüfung

Subjektive Glaubwürdigkeit des Modells

6.3 Sensitivitätsprüfung

Vergleich der Modellreaktion auf Änderung einzelner Parameter mit der Wirklichkeit (qualitativ & quantitativ)

Z.B.: Wird die Bedienzeit der Servicestation erhöht, so muss auch die mittlere Schlängellänge von dieser Bedienstation kleiner werden.

6.4 Untermodellprüfung

Überprüfung einzelner Untermodelle mit Theorie bzw. Realität beziehungsweise Punkte 6.2 und 6.3.

6.5 Vergleich mit theoretischen Ergebnissen

Hierzu werden Input- und Serviceverteilung entsprechend den Bedingungen der Warteschlangentheorie verändert (z.B.: Erzeugung eines (M/M/1) Systems) und das Simulationsergebnis wird anschließend mit dem Ergebnis der Warteschlangentheorie verglichen.

6.5.1 Warteschlangentheorie

Notation

Zur allgemeinen Charakterisierung von Warteschlangenmodellen kann die Kendall/Lee/Taha-Notation, vorzugsweise in der Form

$$(a/b/c) : (d/e/f),$$

verwendet werden.

Dabei bedeuten:

a = Ankunftszeit-Verteilung (Arrival time distribution)

b = Bedienzeit-Verteilung (service time distribution)

c = Anzahl parallel geschalteter Kanäle gleicher Arte (number of parallel servers)

d = Auswahlordnung (service discipline)

e = Kapazität von Bedienungs- und Wartesystemen (maximum number allowed in system, in service and waiting), endlich oder unendlich.

f = Kundenreservoir (size of calling source), endlich oder unendlich

Die Verteilungen von a und b werden wie folgt spezifiziert:

M = Poissonverteilte Ankunftsverteilung (= exponentialverteilte Zwischenankunftszeit);
resp poissonverteilter Output (= exponentialverteilte Bedienzeit).

D = konstante Zwischenankunfts- resp. Bedienzeit

U = gleichverteilte Zwischenankunftszeit- resp. Bedienzeit.

Die häufigsten Spezifikationen für d sind:

FIFO = First In, First Out (oder FCFS = First Come, First Served)

LIFO = Last In, First Out (oder LCFS = Last Come, First Served)

SIRO = Service in Random Order

c , e und f werden als natürliche Zahlen spezifiziert.

6.5.1.1 Formeln für (M/M/1)-Systeme

Am einfachsten zu behandeln sind „single queue - single server - Systeme“ (SQ-SS-Systeme) vom Typ (M/M/1): (FIFO/ ∞/∞).

Für sie gelten die nachfolgenden Formeln aus der Warteschlangentheorie:

Bezeichnung	Darstellung
Mittlere Ankunftsrate pro Zeiteinheit (average arrival time)	λ
Mittlere Zwischenankunftszeit (average interarrival time)	$\frac{1}{\lambda}$
Mittlere Bedienungsrate pro Zeiteinheit (average service rate)	μ
Mittlere Bedienzeit (average service time)	$\frac{1}{\mu}$
Mittlere Kapazitätsauslastung (average utilization)	$\rho = \frac{\lambda}{\mu}$
Mittlere Verweilzeit pro Kunde (average time spent in the system)	$W_s = \frac{1}{\mu - \lambda}$
Mittlere Wartezeit pro Kunde (average time spent in the queue) für Wartezeiten = 0	$W_{q \geq 0} = \frac{\lambda}{\mu(\mu - \lambda)}$
Mittlere Wartezeit pro Kunde (average time spent in the queue) für Wartezeiten > 0	$W_{q > 0} = \frac{1}{\mu - \lambda} = W_s$
Mittlere Kundenanzahl im System (average number in the system)	$L_s = \frac{\lambda}{\mu - \lambda}$

Mittlere Kundenanzahl in der Warteschlange (average number in the queue)	$L_q = \frac{\lambda^2}{\mu(\mu - \lambda)}$
Wahrscheinlichkeit, dass n Einheiten im System sind	$p_n = \rho^n (1 - \rho)$

Zur Veranschaulichung des Gesagten ein

6.5.1.2 Musterbeispiel

In der Mensa treffen im Durchschnitt 15 Kunden pro Stunde ein. Die Bedienzeit (Essensausgabe und Geldwechsel) beträgt im Durchschnitt 3 Minuten.

Sowohl Zwischenankunftszeiten als auch Bedienzeiten entsprechen einer Exponentialverteilung (bzw. die Anzahl Ankünfte bzw. Abfertigungen pro Zeiteinheit entsprechen einem Poissonprozess).

- Bestimme mittlere Bedienzeit, Bedienungsrate, Zwischenankunftszeit und Ankunftsrate
- Wie viel Zeit des Tages steht das Verkaufsgeschäft bei einem 8-Stunden Arbeitstag leer?
- Wie viel Zeit muss ein Kunde für den täglichen Besuch der Geschäfte (ohne Hin- und Rückweg) einsetzen?
- Wie lange wartet ein Kunde im Durchschnitt auf die Bedienung (= mittlere Verweilzeit – mittlere Bedienzeit = mittlere Wartezeit beziehungsweise mit der Formel)?
- Wie lange wartet ein Kunde (der warten muss) im Durchschnitt?
- Wie viele Kunden sind im Durchschnitt in der Warteschlange?
- Es treffen in Zukunft doppelt so viele Kunden ein. Wie groß muss die Bedienzeit sein, dass die Wartezeit pro Kunde gleich bleibt?
- Es wird zukünftig doppelt so schnell bedient (zum Beispiel durch Installation einer zweiten parallelen Kassa). Wie muss die Ankunftsrate sein, damit die Wartezeit pro Kunde gleich bleibt? (WICHTIG FÜR SYSTEMVEREINFACHENDE ANNAHMEN).

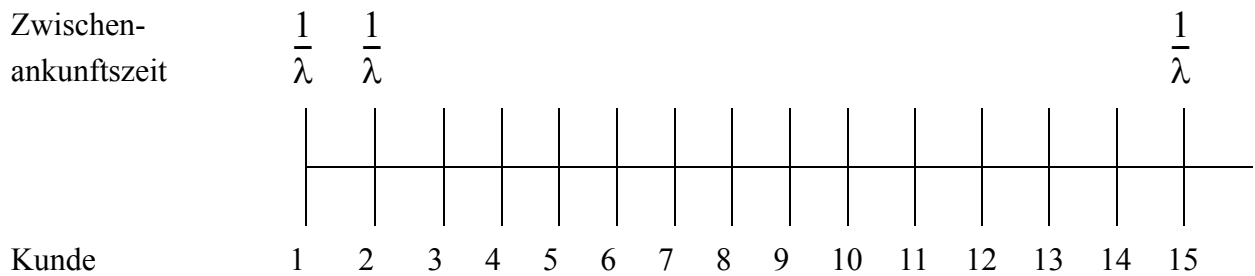
Lösungen der Aufgaben a)-h)

a) Mittlere Bedienzeit = $\frac{1}{\mu} = 3 = 3$ Minuten

Bedienungsrate (Bedienungen pro Zeiteinheit) = $\mu = \frac{1}{3}$

Mittlere Zwischenankunftszeit $\frac{1}{\lambda} = \frac{60}{14} = 4$ Minuten.

Dies entsteht auf Grund folgender Zeitachsenüberlegung:



Ankunftsrate (Ankünfte pro Zeiteinheit) $\lambda = \frac{1}{4}$ (Personen pro Minute)

b) Servicequotient $\rho = \frac{\lambda}{\mu} = \frac{\frac{1}{4}}{\frac{1}{3}} = \frac{3}{4}$. Das heißt, das Geschäft ist $\frac{3}{4} \cdot 8h = 6h$ besetzt

und 2h unbesetzt

c) $W_s = \frac{1}{\mu - \lambda} = \frac{1}{\frac{1}{3} - \frac{1}{4}} = \frac{1}{\frac{1}{12}} = 12$ Minuten

d) Mittlere Wartezeit in der Warteschlange (= mittlere Verweilzeit - mittlere Bedienzeit)

$$W_{q \geq 0} = W_s - \frac{1}{\mu} = \frac{1}{\mu - \lambda} - \frac{1}{\mu} = \frac{\mu - (\mu - \lambda)}{\mu(\mu - \lambda)} = \frac{\lambda}{\mu(\mu - \lambda)} \text{ q.e.d.}$$

$$\dots = \frac{\frac{1}{4}}{\frac{1}{3} \left(\frac{1}{3} - \frac{1}{4} \right)} = \frac{\frac{1}{4}}{\frac{1}{3} \cdot \frac{1}{12}} = \frac{36}{4} = 9 \text{ Minuten}$$

e) $W_{q > 0} = \frac{1}{\mu - \lambda} = \frac{1}{\frac{1}{3} - \frac{1}{4}} = \frac{1}{\frac{1}{12}} = 12$ Minuten

f) $L_q = \frac{\lambda^2}{\mu(\mu - \lambda)} = \frac{\frac{1}{16}}{\frac{1}{3} \left(\frac{1}{3} - \frac{1}{4} \right)} = \frac{\frac{1}{16}}{\frac{1}{3} \cdot \frac{1}{12}} = \frac{36}{16} = 2,25$ Kunden

g) Mittlere Wartezeit 9 Minuten (siehe Aufgabe d)

$$W_{q \geq 0} = 9 \text{ min} = \frac{\lambda}{\mu(\mu - \lambda)}$$

jetzt kommt alle 2 Minuten ein Kunde an \Rightarrow

$$\lambda = \frac{1}{2}$$

$$W_{q \geq 0} = \frac{\frac{1}{2}}{\mu_{neu} \left(\mu_{neu} - \frac{1}{2} \right)}$$

$$\mu_{neu}^2 - \frac{1}{2} \mu_{neu} - \frac{1}{18} = 0$$

$$\mu_{1,2} = -\frac{p}{2} \pm \sqrt{\frac{p^2}{4} - q}$$

$$\mu_{1,2} = \frac{1}{4} \pm \sqrt{\frac{1}{16} + \frac{1}{18}}$$

$$\mu_1 = 0,594$$

$$\mu_2 = \dots$$

$$\frac{1}{\mu} = 1,6835 \text{ min} \Rightarrow \text{Man sieht, obwohl sich die Ankunftsrate verdoppelt hat, muss die}$$

Bedienzeit nicht halbiert werden, um die Wartezeit aufrecht zu erhalten.

h) Mittlere Bedienzeit $\frac{1}{\mu} = 1,5$ Minuten $\Rightarrow \mu = \frac{2}{3}$

$$\text{Mittlere Zwischenankunftszeit} = \frac{1}{\lambda}$$

$$W_{q \geq 0} = \frac{\lambda}{\frac{2}{3} \left(\frac{2}{3} - \lambda \right)} = 9 \rightarrow \lambda = \frac{4}{7} \rightarrow \frac{1}{\lambda} = 1,75 \text{ Minuten}$$

6.5.1.3 Anwendungen der Warteschlangentheorie in der Simulation

a) Wie im Kapitel 6.5 (Validierung) bereits gesagt kann man im Simulationsmodell die Ankunfts- bzw. Serviceverteilung entsprechend dem M/M/1-System verändern und dann das Simulationsergebnis mit dem theoretischen Ergebnis vergleichen.

b) Wie im Kapitel 7.2.1 noch erläutert, kann die Warteschlangentheorie Hilfsmittel sein die Blockgröße abzuschätzen. Als kleiner Vorgriff auf Kapitel 7.2.1 folgendes Beispiel:

Bei exponentialverteilter Zwischenankunfts- und Servicezeit ergibt sich aus der Warteschlangentheorie:

„Die Wahrscheinlichkeit dass 0 Einheiten im System sind := „

$$p_0 = \rho^0 (1 - \rho) = 1 - \frac{\lambda}{\mu}$$

Bei $\lambda = \frac{1}{4}$ und $\mu = \frac{1}{3} \Rightarrow$

$$p_0 = 1 - 0,75 = 0,25$$

Das heißt in $\frac{1}{4}$ der Fälle (Ankünfte) ist das Servicesystem leer (es gibt keinen Einfluss von

Wartenden auf Neuankömmlinge). Bei einer mittleren Zwischenankunftszeit $\frac{1}{\lambda}$ von 2 Minuten

trifft durchschnittlich jeder 4. Ankommende auf ein leeres Servicesystem. Das heißt, dass nach zirka 8 Minuten das Servicesystem im Mittel leer ist. Nimmt man nun eine Blockgröße von 24 Minuten (das Drei- bis Vierfache), so kann man ziemlich sicher sein, dass das Wartesystem 1 mal leer war und damit kein Einfluss eines Wartenden auf die Nachkommenden vorhanden ist (d.h. man schaltet den Einfluss der Autokorrelation aus).

Andere Verteilungen

Bei anderen Verteilungen von Service- und Zwischenankunftszeit ermittelt man aus Vorsimulationen die durchschnittliche Zeitdauer bis das Servicesystem von der mittleren Auslastung auf Null-Auslastungen zurückfällt. Für die Blockgröße nimmt man jetzt das Drei bis Vierfache der ermittelten mittleren Auslastung.

7. Gründe für ein fehlerhaftes Simulationsergebnis

7.1 Fehler bei der Aufwärmphase

Startet man eine Simulation, so ist in der Regel das Gesamtsystem leer, das heißt kein Kunde ist im Servicesystem und es haben sich noch keinerlei Warteschlangen vor den einzelnen Servicestationen gebildet. So ergeben sich zum Beispiel folgende Wartezeiten pro Kunde:

0	
0	
1	
0	„kaltes System“
1	
0	
<hr/>	
2	
1	
3	
1	eingependeltes „warmes“ System
1	
2	
12	

Das heißt über alles gerechnet ergibt sich eine mittlere Wartezeit pro Kunde von $12/12 = 1$ Minute, obwohl sie tatsächlich 1,666 Minuten beträgt.

Um diesen Fehler zu umgehen, lässt man eine „Vorsimulation“ solange laufen, bis man subjektiv (durch laufende Snap-Shots) den Eindruck hat, dass sich das Gesamtsystem in einem mittleren Betriebszustand befindet („aufgewärmt ist“). Danach löscht man die Statistiken, nicht jedoch die Systemzustände.

An die Aufwärmphase schließen dann die einzelnen zu simulierenden Blöcke an.

Beachte: die Aufwärmphase muss nur vor dem ersten Block (Block bzw. Blockung siehe nächster Punkt) erfolgen, danach hat man schon ein „warmes“ System und kann die benötigten Blöcke direkt anschließen.

7.2 Die Simulationsdauer (Dauer der Datenerhebung - Blockgröße und Blockanzahl)

Selbst wenn man die Aufwärmphase verwirft ergibt sich das Problem, ob die Simulationslänge (Dauer) ausreichend war. D.h. ob das Ergebnis aus einer repräsentativen Stichprobe stammt. Hier gibt die Statistik eine Unterstützung, welche bei normalverteilter Grundgesamtheit eine Angabe über den Stichprobenumfang (Simulationsdauer) ermöglicht. Leider stammen die Messdaten einer Simulation in den wenigsten Fällen aus einer derartigen Grundgesamtheit.

In dieser Situation hilft der **zentrale Grenzwertsatz** der Statistik:

Die Mittelwerte von Stichproben derselben Verteilungsfunktion (damit auch der selben Mittelwerte und der selben Varianz) sind normalverteilt.

Vgl. E. KREYSZIG „Statistische Methoden & ihre Anwendungen“

Aus dieser Tatsache folgt:

Die früher behandelten Methoden zur Bestimmung von Konfidenzintervallen für den Mittelwert der Normalverteilung sind gleichzeitig brauchbare Näherungsmethoden bei beliebigen Verteilungen und umfangreichen Stichproben. Die Güte der Näherung wächst mit dem Stichprobenumfang.

D.h. Wir unterteilen einen Simulationslauf in eine Menge von Stichproben (Blöcke) wobei wir aus jedem Block z.B. die mittlere Schlängellänge ermitteln. Für diese mittlere Schlängellänge gilt, dass sie aus Stichproben stammen die jeweils derselben Verteilungsfunktion unterliegen. D.h. diese mittleren Schlängellängen sind nach dem zentralen Grenzwertsatz sehr wohl normalverteilt.

Wir haben nun Messwerte welche einer Normalverteilung unterliegen und könnten ausrechnen wieviele derartige Messwerte wir benötigen um eine fundierte Aussage z.B. über die mittlere Warteschlängellänge zu machen.

7.2.1 Die Blockgröße

Wie groß muss die einzelne Stichprobe (der Block) sein?

Antwort: „Ein Block muss länger sein als der Einfluss der Autokorrelation“.

Zur Veranschaulichung des Einflusses der Autokorrelation folgende Situation: Sie stehen im Supermarkt in einer Warteschlange vor einer Kassa, da ersucht Sie jemand ob sie ihn „reinlassen“, er hat nur einen Artikel gekauft. Es ist klar, dass nicht nur Ihre Wartezeit um die Servicierung dieses Kunden länger wird, sondern auch die aller hinter Ihnen Wartenden. Der Einfluss wie lange solch ein Kunde im Wartesystem nachwirkt, ist durch die Autokorrelation gegeben.

Da die Autokorrelation nur schwer berechenbar ist behilft man sich durch Abschätzung bzw. Vorsimulation. Da die Autokorrelation auf jeden Fall unterbrochen wird, wenn die Warteschlange auf 0 abfällt, versucht man den Abbruch der Warteschlange auf 0 zu ermitteln.

a) durch die Theorie (siehe Warteschlangensysteme)

p_0 = Wahrscheinlichkeit, dass 0 Einheiten im System

$$p_0 = \rho^0(1 - \rho) = 1\left(1 - \frac{\lambda}{\mu}\right)$$

So ergibt sich bei einer mittleren Ankunftsrate $\lambda = 2$ Ankünfte / Minute und einer mittleren Servicerate $\mu = 3$ Servicierungen / Minute

$$p_0 = 1\left(1 - \frac{2}{3}\right) = \frac{1}{3}$$

D.h. in 1/3 der Fälle hat ein ankommender Kunde ein leeres System. D.h. bei einer mittleren Ankunftsrate von 2 Ankünften pro Minute ergibt sich eine Zwischenankunftszeit von 30 sec. D.h. jeder 3. Ankommende hat ein leeres System.

0 sec ... 1. Ankunft, 30 sec. ... 2. Ankunft, 60 sec. ... 3. Ankunft

Nach jeweils 60 sec. ist im Mittel das System leer. Nimmt man einen Block von 180 sec. so ist dies durchaus ausreichend.

b) durch Vorsimulation (bei beliebigen Verteilungen)

Mit Hilfe von Vorsimulation lässt sich ermitteln wie oft das System durchschnittlich leer ist. Daraus lässt sich ebenfalls die Blocklänge ableiten.

c) durch Überlegungen

Annahme: solange in einer Tankstelle die Zapfsäulen besetzt sind fährt kein Kunde in die Tankstelle ein, sondern benützt eine nahegelegene 2. Tankstelle. - Die Ankunftsrate der tankwilligen Autos = $U(60, 480)$. Die Servicerate = $U(100, 300)$.

Wie lange wirkt ein Auto im System?

Ungünstiger Fall: alle Autos kommen im kürzestmöglichen Zeitraum von 60 sec. und benötigen die längste mögliche Tankzeit von 300 sec.

D.h.: Ein Auto kommt zum Zeitpunkt 0 und bleibt bis zum Zeitpunkt 300 sec. In dieser Zeit wirkt dieses Auto auf 5 tankwillige Fahrzeuge (die verlorengehen), danach kann wieder ein Auto zufahren.

In anderen Fällen kann man oft ähnliche Gedankenexperimente machen.

Zusammenfassung: Wir wissen bis jetzt, dass wir nach einer (statistisch zu verwerfenden) Aufwärmphase einzelne Stichproben (Blöcke) ziehen müssen, wobei jede dieser Stichproben länger sein muss als der Einfluss der Autokorrelation. Nun zur Frage wie viele derartige Stichproben (Blöcke) muss man ziehen?

7.2.2 Anzahl Blöcke

Wir wissen, dass die einzelnen Messgrößen der Blöcke (z.B.: die mittlere Schlängellänge eines Blocks) der Normalverteilung gehorchen. **Beachte:** unsere Messgrößen sind Mittelwerte!!! Wir berechnen nun die Anzahl derartiger Messgrößen nach der Formel aus der Stichprobentheorie:

$$n = \left(\frac{cs}{a} \right)^2$$

Mit: s ... Varianz der Messgrößen bzw. deren Schätzung:

$$s^2 = \frac{1}{n-1} \sum_{j=1}^u (x_j - \bar{x})^2 = \frac{1}{n-1} \left[\sum_{j=1}^u x_j^2 - \frac{1}{n} \left(\sum_{j=1}^u x_j \right)^2 \right] = \frac{1}{n-1} \left[\sum_{j=1}^u x_j^2 - n\bar{x}^2 \right]$$

a ... Vorgegebenes Konfidenzintervall

c ... Tabellenwert der t-Verteilung

D.h. Nach einer subjektiv zu bestimmenden Aufräumphase sind n Blöcke, jeder länger als der Einfluss der Autokorrelation, zu simulieren. Siehe dazu auch nachfolgendes Beispiel:

Beachte: Ist die Varianz bekannt (Varianz der Grundgesamtheit) so nimmt man für den Wert c die Tabelle der Normalverteilung. Nimmt man für s eine Schätzung der Varianz aus einer Stichprobe, so nimmt man die Tabelle der t-Verteilung.

Beispiel zu Aufwärmphase, Blockung und Simulationsdauer:

Gegeben eine starr verkettete Fertigungslinie, bestehend aus 4 Fertigungseinheiten (FE's) mit davorliegenden Pufferzonen. Der Einfachheit halber soll sich vor allen FE's eine Pufferzone von 70 Werkstücken (Halbfabrikaten) befinden. Jede Fertigungseinheit besteht aus 2 Maschinen, einer Maschine mit fixer Bearbeitungszeit von 110 Minuten und einer Maschine mit variabler Bearbeitungszeit (jedoch auf allen 4 FE's mit einer mittleren Bearbeitungszeit von 50 Minuten), sodass die mittlere Gesamtarbeitszeit auf allen 4 FE's $110 + 50 = 160$ Minuten

beträgt. Getaktet wird alle 160 Minuten ein Halbfabrikat zur Endbearbeitung aufgelegt. Nachfolgend die Bearbeitungszeiten (in Minuten) der Maschinen mit variabler Bearbeitungszeit:

- Maschine 2 von FE-1, Uniformverteilt:= U(10, 90),
- Maschine 2 von FE-2, Normalverteilt:= N($\mu = 50, \sigma = 10$),
- Maschine 2 von FE-3, Exponentialverteilt:= E($\mu = 50$),
- Maschine 2 von FE-4, Getaktet:= G(50).

Folgende Annahmen (ohne Prüfung auf Richtigkeit) werden getroffen:

- a) Eine Aufwärmphase von 2000 endgefertigten Halbfabrikaten ist ausreichend.
- b) Ein Stichprobenumfang von jeweils ca. 800 endgefertigten Halbfabrikaten ist länger als der Einfluss der Autokorrelation.

Aufgabenstellung:

Machen Sie für obige Annahme eine Vorsimulation in GPSS, bestehend aus Aufwärmphase + 10 Stichproben und berechnen Sie daraus den Stichprobenumfang um eine 90%-ige Sicherheit über die mittlere Durchlaufzeit zu bekommen.

Lösung:

Die Vorsimulation hat nach der Aufwärmphase folgende 10 Durchlaufzeiten für jeweils 800 Jobs ergeben:

Block	Durchlaufzeit	Maximaler Pufferinhalt von				Σ
		FE-1	FE-2	FE-3	FE-4	
1	128191	20	4	6	7	37
2	129512	31	3	4	5	43
3	129344	35	3	8	8	54
4	128377	36	10	8	9	63
5	128000	29	12	7	13	51
6	128000	35	5	4	14	58
7	128000	37	6	6	9	58
8	128000	38	6	4	8	56
9	128000	38	5	4	8	55
10	128000	37	6	5	8	56
11	128090	41	5	6	10	62
12	128370	41	10	7	8	66
13	128065	34	15	6	10	65
14	128000	27	16	7	14	64
15	<u>128000</u>	27	16	7	18	<u>68</u>

$$\bar{x} = 128263,27$$

$$\bar{x} = 57,0666$$

$$\sigma_n = 474,740$$

$$\sigma_n = 8,26613$$

$$\sigma_{n-1} = 491,40288$$

$$\sigma_{n-1} = 8,55625$$

Berechnung des Stichprobenumfangs n ($n = \text{Anzahl Blöcke}$):

Aus der Statistik ergibt sich das Konfidenzintervall für den Mittelwert einer Normalverteilung mit bekannter Varianz σ , zu:

$$KONF\{\bar{x} - a \leq \mu \leq \bar{x} + a\} \text{ mit } a = \frac{(c \cdot \sigma)}{\sqrt{n}}$$

Wobei:

- c ... Tabellenwert der Normalverteilung für gegebene Irrtumswahrscheinlichkeit
- σ ... Varianz der Grundgesamtheit
- n ... Stichprobenumfang

Falls wir statt σ , den Schätzwert s ($= \sigma_{n-1}$) verwenden, nimmt man für c , den entsprechenden Wert der t-Verteilung.

Aus dieser Formel für das Konfidenzintervall lässt sich bei vorgegebenem a , der Stichprobenumfang n berechnen. Es ergibt sich:

Das Problem von a :

a soll in der Größe der Irrtumswahrscheinlichkeit liegen. Es macht keinen Sinn bei 10% Irrtumswahrscheinlichkeit a mit 10% des Messwerts zu wählen.

Beispiel:

Eine Türfabrik produziert Türen mit 2 Meter Höhe. Die Produktionsgenauigkeit ist auf ± 1 cm genau. Es soll die mittlere Türhöhe ermittelt werden. Nimmt man 10% vom Gesamtmaß, so liegt der Mittelwert μ in einem Konfidenzintervall von ± 20 cm. D.h. jeder Messwert kann der Mittelwert sein.

Wir müssen also a auf den Range der Messwerte beziehen. Bei der vorangegangenen Simulation sind jeweils der maximale bzw. minimale Messwert zur Bestimmung des Ranges gekennzeichnet. Wir erhalten:

$$\text{Range} = 129512 - 128000 = 1512$$

Da wir links und rechts vom Mittelwert eine Irrtumswahrscheinlichkeit von 5% (insgesamt also 10%) vorgegeben, fixieren wir $2a$ auf 10% des Ranges, d.h. $a = 10\%$ des halben Ranges.

$$a = 10\% \text{ von } 756 = 75,6$$

Damit berechnet sich der Probenumfang zu:

$$n = \left[\frac{(cs)}{a} \right]^2$$

Mit:

$$c(0,9;14) = 1,76$$

0,9 ... Sicherheitswahrscheinlichkeit

14 ... (n-1) Freiheitsgrade

$s = \sigma_{n-1} = 491,40288$

$a = 75,6$

ergibt sich:

$$n = \left[\frac{(1,76(491,403))}{75,6} \right]^2 = 130,87 = 131 \text{ Blöcke}$$

Wir erwarten, dass mit 90%-iger Sicherheit (bei einer Simulation von 131 Blöcken) die mittlere Durchlaufzeit für 800 Jobs im Intervall $128.263,27 \pm 75,6$ liegt.

D.h. Nach derzeitigem Kenntnisstand wäre eine Simulationsdauer von $131 \cdot 128000 \text{ min} \sim 17 \cdot 10^6 \text{ min}$ bzw. $131 \cdot 800 = 104800 \text{ Jobs}$ notwendig. Wir werden in Folge sehen, dass dies nicht notwendig ist!

Man müsste dies durchführen und könnte dann nach der Simulation von 131 Blöcken **mit festgesetztem n** ein \bar{a} berechnen, welches den Mittelwert in einem entsprechenden Intervall $\pm a$ festgelegt.

Das Problem der Länge der Aufwärmphase:

Betrachtet man in unserer Vorsimulation die Σ der maximalen Pufferinhalte, so sieht man, dass diese kontinuierlich ansteigen (von 37 bei Block 1 auf 68 bei Block 15).

Ausserdem ist zu beobachten, dass Puffer 1 eine fallende, Puffer 2 eine steigende Tendenz aufweist. Um dieses Phänomen genauer zu untersuchen, machen wir Snap's nach einer Aufwärmphase von: 40.000, 60.000, 80.000 bzw. 290.000 und erhalten:

Die Vorsimulation hat nach einer Aufwärmphase von 2000 Jobs folgende 10 Durchlaufzeiten für jeweils 800 Jobs ergeben:

Job	Durchlaufzeit	Maximaler Pufferinhalt von				Σ
		FE-1	FE-2	FE-3	FE-4	
1	128191	20	4	6	7	37
2	129512	31	3	4	5	43
3	129344	35	3	8	8	54
4	128377	36	10	8	9	63
5	128000	29	12	7	13	51
6	128000	35	5	4	14	58
7	128000	37	6	6	9	58
8	128000	38	6	4	8	56
9	128000	38	5	4	8	55
10	128000	37	6	5	8	56
11	128090	41	5	6	10	62
12	128370	41	10	7	8	66
13	128065	34	15	6	10	65
14	128000	27	16	7	14	64
15	<u>128000</u>	27	16	7	18	<u>68</u>

$$\bar{x} = 128263,27$$

$$\bar{x} = 57,0666$$

$$\sigma_n = 474,470$$

$$\sigma_n = 8,26613$$

$$\sigma_{n-1} = 491,40288$$

$$\sigma_{n-1} = 8,55625$$

SNAP nach einer Aufwärmphase von 40.000 Jobs:

Maximaler Pufferinhalt von						
Job	Durchlaufzeit	FE-1	FE-2	FE-3	FE-4	Σ
1	128000	59	6	4	15	84
2	128000	55	7	10	5	77
3	128000	55	6	7	17	85
4	128000	54	5	6	18	83
5	128000	54	5	7	19	85
6	128000	53	6	8	20	87
7	128000	56	4	8	19	87
8	128000	55	8	6	19	88
9	128000	51	7	11	20	89
10	128000	49	4	14	24	91
11	128000	46	7	7	25	85
12	128000	44	7	6	27	84
13	128000	41	9	5	28	83
14	128000	44	7	7	28	86
15	128000	42	7	8	29	86
Σ		758	95	114	313	1280

SNAP nach einer Aufwärmphase von 60.000 Jobs:

Maximaler Pufferinhalt von						
Job	Durchlaufzeit	FE-1	FE-2	FE-3	FE-4	Σ
1	128000	17	23	12	36	88
2	128000	10	25	7	42	84
3	128000	12	22	7	44	85
4	128000	14	20	7	46	87
5	128000	12	19	7	49	87
6	128000	13	16	7	51	87
7	128000	13	12	6	53	84
8	128000	12	12	14	56	94
9	128000	14	6	9	58	87
10	128000	15	7	10	56	88
11	128000	15	5	8	57	85
12	128000	14	4	6	58	82
13	128000	18	5	9	55	87
14	128000	17	5	4	56	82
15	128000	22	4	9	54	89
Σ		218	185	122	771	1296

SNAP nach einer Aufwärmphase von 80.000 Jobs:

Maximaler Pufferinhalt von						
Job	Durchlaufzeit	FE-1	FE-2	FE-3	FE-4	Σ
1	128000	49	7	6	26	88
2	128000	50	4	4	24	82
3	128000	49	5	4	24	84
4	128000	49	7	4	23	83
5	128000	45	11	9	25	90
6	128000	37	12	10	30	89
7	128000	35	10	3	32	80
8	128000	35	9	6	33	83
9	128000	34	7	10	38	89
10	128000	30	8	13	38	89
11	128000	31	6	6	42	85
12	128000	34	3	7	41	85
13	128000	38	3	8	37	86
14	128000	35	10	8	36	89
15	128000	31	15	7	36	89
Σ		582	117	107	485	1291

SNAP nach einer Aufwärmphase von 100.000 Jobs:

Maximaler Pufferinhalt von						
Job	Durchlaufzeit	FE-1	FE-2	FE-3	FE-4	Σ
1	128000	17	5	13	59	94
2	128000	16	4	8	56	84
3	128000	15	9	6	55	85
4	128000	15	9	14	15	53
5	128000	18	9	11	55	93
6	128000	16	8	10	57	91
7	128000	11	11	10	56	88
8	128000	11	6	7	60	84
9	128000	12	3	6	61 <==	82
10	128000	14	7	7	59	87
11	128000	15	8	7	57	87
12	128000	17	5	7	57	86
13	128000	16	8	10	56	90
14	128000	14	6	11	58	89
15	128000	19	7	6	57	89
Σ		226	105	133	818	1282

Was dieses Beispiel gezeigt hat ist, dass wir nach genügend langer Aufwärmphase eine „Pufferatmung“ erzielt haben, welche

- a) nie überläuft (Stauraum ist ja derzeit noch unbegrenzt)
- b) immer genügend Halbfabrikate vor jeder Fertigungseinheit (FE) aufliegen, sodass keine FE warten muss.
- c) dadurch, dass die Varianz der Durchlaufzeiten auf „0“ gefallen ist, können wir mit 100%-iger Sicherheit sagen: „im Mittel wird alle 160 Minuten ein Job fertig“. Zum Nachweis, dass die Varianz = 0 reichen uns die 10 Blöcke leicht.

Hätten wir nur eine einzige Simulation von durchgehend 1.280.000 Minuten gemacht (d.h. ohne Blockung), wäre obige Aussage nicht möglich gewesen.

Da Simulation auch weiterführende Forschung anregen soll, so wäre hier die Frage nach einer Begrenzung des Stauraumes (Fertigungshallen sind nicht unbegrenzt groß) sinnvoll.

Da vor FE-1 kein Stauraum notwendig ist, reicht es hier immer einen Job einzulegen, sobald ein Halbfabrikat von FE-1 serviciert wird, ergibt sich die Länge der Fertigungshalle aus den restlichen maximalen Warteschlangenlängen.

$$1 + 25 + 14 + 61 = 101 \text{ Pufferplätze.}$$

Fragen:

- a) Reicht nicht auch die mittlere Pufferatmung oder gar weniger und welchen Einfluss hat dies auf die Durchlaufzeiten?
- b) Soll man vor FE-3 nicht eine Pufferlänge wesentlich größer als 14 machen, da bei Exponentialverteilung auch Servicezeiten von ∞ möglich sind? Diese Antwort ist die einfachste und lautet...

7.2 Fehler bei der Blockung

- Es werden nicht nur die Statistiken sondern auch die Systemzustände (busy oder free) beziehungsweise Warteschlangenlängen auf 0 gesetzt („kaltes“ System)
- Die Zufallszahlengeneratoren werden immer auf den Anfangszustand gesetzt (jedesmal dieselben Ankunfts- beziehungsweise Servicezeiten).
- Die absolute Simulationszeit wird 0 gesetzt (eventuell unendlich lange Simulation).
- Man nimmt die falsche Varianz (bei der Tankstelle gibt es die Varianz der Ankunftszeiten, der Tankzeiten, der einzelnen Umsätze)

7.3 Fehler bei der Datenerhebung

Annahme: fixe Servicezeiten von 10 Sekunden.

Die Datenerhebung ergab 5 Ankünfte pro Minute. In diesem Fall fehlt die Information ob diese Ankünfte alle am Beginn der Minute stattfinden oder in fixen Zeitintervallen von jeweils 12 Sekunden getaktet sind oder sonst einer Verteilung gehorchen. Es ist sofort ersichtlich, dass im Fall einer gleichzeitigen Ankunft von 5 Personen am Beginn der Minute sofort eine Warteschlange entsteht, die erst abgebaut werden muss, während bei einer getakteten Ankunftsrate von 12 Sekunden keinerlei Warteschlange entsteht.

D.h. die Verteilung der Beobachtungsdaten hat enorme Bedeutung für das Systemverhalten und muss daher unbedingt erfasst werden.

7.4 Fehler bei systemvereinfachenden Annahmen

Mehrere hintereinandergeschaltete Servicierungen aus gleichen oder unterschiedlichen Verteilungen werden zu einer einzigen zusammengefasst.

Hierbei kann ein gänzlich anderes Systemverhalten wie in der Realität entstehen. In diesem Fall ist vorab unbedingt eine I/O-Analyse des zu vereinfachenden Modellbereichs notwendig.

8. Versuchsplanung (Exerimental Design) und Datenanalyse

Bereits das Simulationsziel gibt vor wie man die Simulationsläufe gruppieren soll um möglichst viel Information aus möglichst wenig Simulationsläufen filtern zu können.

Annahme: Wir vermuten, dass folgende 3 Einflussfaktoren auf den Verbrauch eines Fahrzeugs (LKW) wirken:

- Straßenbelag
- Oktanzahl
- Reifenqualität

Dann muss schon das Simulationsexperiment so angelegt werden, dass jeder Straßenbelag mit jeder Oktanzahl mit jeweils jeder Reifenqualität getestet wird. Wir werden dies in unserem letzten Beispiel sehen.

Im umgekehrten Fall kann aber auch getestet werden, ob Auffälligkeiten, die in einem Simulationsergebnis erkennbar sind, statistisch signifikant sind (Datenanalyse).

Wir werden im folgenden Modelle mit bis zu 3 Einflussfaktoren (bzw. Auffälligkeiten) kennenlernen (Näheres dazu in der Literatur über Experimental Design).

Wir nehmen wieder unser Beispiel aus Kapitel 4.4. (Ob die Daten aus einer Datenerhebung oder Ergebnisse einer Simulation sind, spielt für den statistischen Test keine Rolle.)

Sequenz	Tag						Zeilen- summen
	Montag	Dienstag	Mittw.	Donnerst.	Freitag	Samstag	
1	23.9	23.8	23.3	24.4	23.3	22.9	141.6
2	22.0	22.8	22.9	22.8	22.9	22.3	135.7
3	23.8	23.5	23.8	23.1	24.1	24.0	142.3
4	21.5	23.1	22.9	22.5	23.0	22.0	135.0
5	24.6	23.4	23.5	22.6	23.3	23.2	140.6
6	21.7	23.0	23.3	21.6	23.2	22.8	135.6
7	23.9	23.3	23.5	23.4	23.8	23.7	141.6
8	23.1	23.0	22.3	22.5	21.8	23.0	135.7

Analysiert man das Datenmaterial, so sieht man, dass an jedem Tag sequentiell einer „hohen“ Bearbeitungszeit immer eine „niedere“ folgt (pulsierende Bearbeitungszeiten). Dieser Effekt ist auch in den Zeilensummen erkennbar. Da ein derartiger Puls einen starken Einfluss auf die Wartenschlagenlänge hat, untersuchen wir mittels Varianzanalyse, ob dieser Effekt signifikant ist.

1-faktorielle Varianzanalyse

Wir gruppieren das Datenmaterial derart, dass sämtliche vorangehenden „hohen“ Bearbeitungszeiten eine Gruppe, sämtliche „niederen“ Bearbeitungszeiten eine zweite Gruppe bilden. Man erhält somit folgendes Tableau:

i = Sequenz d. Bearbeitung	Tag						Summen
	Montag	Dienstag	Mittw.	Donnerst.	Freitag	Samstag	
i=1: vorangehende Bearbeitungs- zeit	23.9	23.8	23.3	24.4	23.3	22.9	566.1
	23.8	23.5	23.8	23.1	24.1	24.0	
	24.6	23.4	23.5	22.6	23.3	23.2	
	23.9	23.3	23.5	23.4	23.8	23.7	
i=2: nachfolgende Bearbeitungs- zeit	22.0	22.8	22.9	22.8	22.9	22.3	542.0
	21.5	23.1	22.9	22.5	23.0	22.0	
	21.7	23.0	23.3	21.6	23.2	22.8	
	23.1	23.0	22.3	22.5	21.8	23.0	

Wir behaupten nun, dass sich jeder Messwert wie folgt zusammensetzt:

$$y_i = \mu + \alpha_i + e_i$$

D.h. jeder Messwert y_i setzt sich aus dem allgemeinen Mittelwert μ und dem Einflussfaktor Sequenz (= α_i) sowie der allgemeinen Fehlerstreuung e_i zusammen. Mittels Varianzanalyse wird ermittelt, ob der Faktor α_i wirksam ist oder ob er im Rahmen der Fehlerstreuung liegt.

Die H_0 lautet: „ $\alpha_i = 0$ “.

Wir erstellen die Tabelle der Varianzursachen (VU):

Varianz- ursache	Quadratsumme (SQ)	Freiheitsgrade (FG)	Durchschnittsquadrt (D)	F-Wert
Faktor α	$SQ_\alpha = n * \sum_i (y_{i.} - y_{..})^2$	a-1	SQ_α / FG_α	D_α / D_R
Fehler (Rest)	$SQ_R = SQ_T - SQ_\alpha$	$FG_T - FG_\alpha$	SQ_R / FG_R	
Total	$SQ_T = \sum_i \sum_j (y_{ij} - y_{..})^2$	a*n - 1	SQ_T / FG_T	

Mit:

- y_{ij} ... einzelner Messwert (Bearbeitungszeit)
- $Y_{i.}$... Summe der Bearbeitungszeiten der jeweiligen i-ten Reihung (vorangegangene Bearbeitungszeit bzw. nachfolgende Bearbeitungszeit)
- $Y_{..}$... Summe sämtlicher Bearbeitungszeiten
- a ... Anzahl der Elemente einer Sequenz (vorangegangene Bearbeitungszeit, nachfolgende Bearbeitungszeit) = 2
- n ... Anzahl der Beobachtungen pro Gruppe i
- $y_{i.}$... $Y_{i.} / n$
- $y_{..}$... $Y_{..} / (a*n)$

Für unser Beispiel ergibt sich:

$$\begin{aligned}
 a &= 2 \\
 n &= 24 \\
 y_{1.} &= 566.1 / 24 = 23.5875 \\
 y_{2.} &= 542.0 / 24 = 22.5833 \\
 y_{..} &= 1108.1 / 48 = 23.085417 \\
 SQ_{\alpha} &= 24 * ((23.5875 - 23.085417)^2 + (22.5833 - 23.085417)^2) \\
 SQ_{\alpha} &= 24 * 0.5041753 = 12.100208 \\
 SQ_T &= (23.9 - 23.085417)^2 + \dots + (23.0 - 23.085417)^2 = 68.02
 \end{aligned}$$

Eingetragen in der Tabelle der Varianzursachen ergibt sich:

Varianz- ursache	Quadratsumme (SQ)	Freiheitsgrade (FG)	Durchschnittquadrat (D)	F-Wert
Faktor α	$SQ_{\alpha} = 12.100208$	1	12.100208	9.9537131
Fehler	$SQ_R = 55.919792$	46	1.2156477	
Total	$SQ_T = 68.02$	47		

SQ_R ergibt sich als Differenz $SQ_T - SQ_{\alpha}$; analoges gilt für die Freiheitsgrade.

Der entsprechende F-Wert der Tabelle lautet: $F_{(1,46;0.05)} = 4.06$

Da $9.954 > 4.06$ gilt: Die H_0 „Der Einflussfaktor $\alpha = 0$; d.h. es gibt keinen Einfluss der Sequenz“ muss verworfen werden.

D.h. das Pulsieren der Bearbeitungszeiten ist zu berücksichtigen. Betriebsinterne Nachforschungen haben ergeben, dass nach jeder 2. Bearbeitung eine automatische Nachjustierung durchgeführt wird, die genau eine Minute benötigt.

2-faktorielle Varianzanalyse

Obwohl keine Hinweise vorliegen wollen wir prüfen, ob die Tage eventuell Einfluss auf die Bearbeitungszeiten haben. Wir bilden also folgendes Modell:

$$y_{ij} = \mu + \alpha_i + \beta_j + e_{ij}$$

D.h. Auf jeden einzelnen Messwert wirken 2 Faktoren. Zum einen der Einfluss der Sequenz, zum anderen der Einfluss des Tages. Es ergibt sich somit folgende Tabelle der Varianzursachen:

Varianz- ursache	Quadratsumme (SQ)	Freiheitsgrade (FG)	Durchschnittquadrat (D)	F-Wert
Faktor α	$SQ_{\alpha} = b * n * \sum_i (y_{i.} - y_{..})^2$	a-1	SQ_{α}/FG_{α}	D_{α}/D_R
Faktor β	$SQ_{\beta} = a * n * \sum_j (y_{.j} - y_{..})^2$	b-1	SQ_{β}/FG_{β}	D_{β}/D_R
Fehler (Rest)	$SQ_R = SQ_T - SQ_{\alpha} - SQ_{\beta}$	$FG_T - FG_{\alpha} - FG_{\beta}$	SQ_R/FG_R	
Total	$SQ_T = \sum_i \sum_j (y_{ij} - y_{..})^2$	$a*b*n - 1$	SQ_T/FG_T	

Mit:

- y_{ij} ... einzelner Messwert (Bearbeitungszeit)
- Y_i ... Summe der Bearbeitungszeiten der jeweiligen i-ten Reihung (vorangegangene Bearbeitungszeit bzw. nachfolgende Bearbeitungszeit)
- Y_j ... Summe der Bearbeitungszeiten des jeweiligen j-ten Tages
- $Y_{..}$... Summe sämtlicher Bearbeitungszeiten
- a ... Anzahl der Elemente einer Sequenz (vorangegangene Bearbeitungszeit, nachfolgende Bearbeitungszeit) = 2
- b ... Anzahl der Tage = 6
- n ... Anzahl der Beobachtungen pro Gruppe $ij = 4$
- y_i ... $Y_i / (b \cdot n)$
- y_j ... $Y_j / (a \cdot n)$
- $y_{..}$... $Y_{..} / (a \cdot b \cdot n)$

Beachte: Die Aufspaltung des Versuchsplans in einen 2-faktoriellen Versuch erzeugt Untergruppen ij zu jeweils 4 Elementen, was im folgenden Tableau der Messwerte auch schön zu sehen ist.

$i \setminus j$	Tag						Summe Y_i
	$j=1$ Montag	$j=2$ Dienstag	$j=3$ Mittw.	$j=4$ Donn.	$j=5$ Freitag	$j=6$ Samstag	
$i=1$: vorangegangene Bearbeitungszeit	23,9 23,8 24,6 23,9	23,8 23,5 23,4 23,3	23,3 23,8 23,5 23,5	24,4 23,1 23,6 23,4	23,3 24,1 23,3 23,8	22,9 24,0 23,2 23,7	566,1
$i=2$: nachfolgende Bearbeitungszeit	22,0 21,5 21,7 23,1	22,8 23,1 23,0 23,0	22,9 22,9 23,3 22,3	22,8 22,5 21,6 22,5	22,9 23,0 23,2 21,8	22,3 22,0 22,8 23,0	542,0
Summe Y_j	184,5	185,9	185,5	182,9	185,4	183,9	$Y_{..} =$ 1108,1

Für unser Beispiel ergibt sich:

$$a \quad \dots = 2$$

$$b \quad \dots = 6$$

$$n \quad \dots = 4$$

$$y_1 \quad \dots = \frac{566,1}{6 \cdot 4} = 23,5875$$

$$y_2 \quad \dots = \frac{566,1}{6 \cdot 4} = 22,5833$$

$$y_{..} \quad \dots = \frac{1108,1}{48} = 23,085417$$

$$y_{.1\dots} = \frac{184,5}{2 \cdot 4} = 23,0625$$

$$y_{.2\dots} = \frac{185,9}{2 \cdot 4} = 23,2375$$

$$y_{.3\dots} = \frac{185,5}{2 \cdot 4} = 23,1875$$

$$y_{.4\dots} = \frac{182,9}{2 \cdot 4} = 23,8625$$

$$y_{.5\dots} = \frac{185,4}{2 \cdot 4} = 23,1750$$

$$y_{.6\dots} = \frac{183,9}{2 \cdot 4} = 22,9875$$

$$SQ_{\alpha} = 24 \cdot ((23,5875 - 23,085417)^2 + (22,5833 - 23,085417)^2) = 12,100208$$

$$SQ_{\beta} = 8 \cdot ((23,0625 - 23,085417)^2 + \dots + (22,9875 - 23,085417)^2) = 0,8110417$$

$$SQ_T = (23,9 - 23,085417)^2 + \dots + (23,0 - 23,085417)^2 = 68,02$$

Eingetragen in der Tabelle der Varianzursachen ergibt sich:

Varianzursache	Quadratsumme (SQ)	Freiheitsgrade (FG)	Durchschnittsquadrat (DQ)	F-Wert
Faktor α	$SQ_{\alpha} = 12,100208$	1	12,100208	8,986353
Faktor β	$SQ_{\beta} = 0,8110417$	5	0,1622083	0,120465
Fehler	$SQ_R = 55,206878$	41	1,3465092	
Total	$SQ_T = 68,02$	47		

Die entsprechenden F-Werte der Tabelle lauten:

$$F(1,41;0,05) = 4,08$$

$$F(5,41;0,05) = 2,45$$

Damit lauten die entsprechenden H_0 -Hypothesen:

Da $8,986 > 4,08$ gilt: Die H_0 "Der Einflussfaktor $\alpha = 0$; d.h. es gibt keinen Einfluss der Sequenz" muss verworfen werden.

Da $0,120 < 2,45$ gilt: Die H_0 "Der Einflussfaktor $\beta = 0$; d.h. es gibt keinen Einfluss der Tage" kann nicht verworfen werden.

D.h. das Pulsieren der Bearbeitungszeiten ist zu berücksichtigen, der Einfluss der Tage ist nicht zu berücksichtigen.

Nur der Vollständigkeit halber:

Manchmal passiert es, dass 2 harmlose Einflüsse sich gegenseitig hochschaukeln. Wir wollen auch diese Wechselwirkung überprüfen.

2-faktorielle Varianzanalyse unter Berücksichtigung von Wechselwirkungen

Wir behaupten nun, dass sich jeder Messwert wie folgt zusammensetzt:

$$\mu + \alpha_i + \beta_j + \alpha\beta_{ij} + e_{ij}$$

Wobei α_i den Einfluss der Reihung, β_j den des Tages, $\alpha\beta_{ij}$ einen möglichen Einfluss zwischen Tag und Reihung und e_{ij} die allgemeine Fehlerstreuung bedeuten. Mittels Varianzanalyse wird ermittelt ob die Faktoren α_i , β_j bzw. $\alpha\beta_{ij}$ wirksam sind oder ob sie im Rahmen der Fehlerstreuung liegen.

Die H_0 lautet: " $\alpha_i = \beta_j = \alpha\beta_{ij} = 0$ "

Wir erstellen die Tabelle der Varianzursachen (VU):

Varianzursache	Quadratsumme (SQ)	Freiheitsgrade (FG)	Durchschnittsquadrat	F-Wert
Faktor α	$SQ_{\alpha} = b \cdot n \cdot \sum_i (y_{i..} - y_{...})^2$	$a - 1$	$\frac{SQ_{\alpha}}{FG_{\alpha}}$	$\frac{D_{\alpha}}{D_R}$
Faktor β	$SQ_{\beta} = a \cdot n \cdot \sum_j (y_{.j.} - y_{...})^2$	$b - 1$	$\frac{SQ_{\beta}}{FG_{\beta}}$	$\frac{D_{\beta}}{D_R}$
Faktor $\alpha\beta$	$SQ_{\alpha\beta} = n \cdot \sum_i \sum_j (y_{ij.} - y_{i..} - y_{.j.} + y_{...})^2$	$(a - 1) \cdot (b - 1)$	$\frac{SQ_{\alpha\beta}}{FG_{\alpha\beta}}$	$\frac{D_{\alpha\beta}}{D_R}$
Fehler Rest	$SQ_R = SQ_T - SQ_{\alpha} - SQ_{\beta} - SQ_{\alpha\beta}$	$FG_T - FG_{\alpha} - FG_{\beta} - FG_{\alpha\beta}$	$\frac{SQ_R}{FG_R}$	
Total	$SQ_T = \sum_i \sum_j \sum_k (y_{ijk} - y_{...})^2$	$a \cdot b \cdot n - 1$	$\frac{SQ_T}{FG_T}$	

Mit:

- y_{ijk} ... einzelner Messwert (Bearbeitungszeit)
- $Y_{i..}$... Summe der Bearbeitungszeiten der jeweiligen i-ten Reihung (vorangegangene Bearbeitungszeit bzw. nachfolgende Bearbeitungszeit)
- $Y_{.j.}$... Summe der Bearbeitungszeiten des jeweiligen j-ten Tages
- $Y_{ij.}$... Summe der Bearbeitungszeiten der i-ten Reihung am j-ten Tag
- $Y_{...}$... Summe sämtlicher Bearbeitungszeiten
- a ... Anzahl der Elemente einer Sequenz (vorangegangene Bearbeitungszeit, nachfolgende Bearbeitungszeit) = 2
- b ... Anzahl der Tage = 6
- n ... Anzahl der Beobachtungen pro Gruppe $ij = 4$
- $y_{i..}$... $Y_{i..} / (b \cdot n)$
- $y_{.j.}$... $Y_{.j.} / (a \cdot n)$
- $y_{ij.}$... $Y_{ij.} / (n)$
- $y_{...}$... $Y_{...} / (a \cdot b \cdot n)$

Im Versuchsplan ergeben sich folgende Summen:

i\j	Tag						Summe
	j=1 Montag	j=2 Dienstag	j=3 Mittw.	j=4 Donn.	j=5 Freitag	j=6 Samstag	
i=1: vorange- gangene Bearbei- tungszeit	23,9 23,8 24,6 23,9	23,8 23,5 23,4 23,3	23,3 23,8 23,5 23,5	24,4 23,1 22,6 23,4	23,3 24,1 23,3 23,8	22,9 24,0 23,2 23,7	Y_{1..}= 566,1 y _{1..} = 23,5875
Y_{ij}. y _{ij} .	96,2 24,050	94,0 23,500	94,1 23,525	93,5 23,375	94,5 23,625	93,8 23,450	
i=2: nachfol- gende Bearbei- tungszeit	22,0 21,5 21,7 23,1	22,8 23,1 23,0 23,0	22,9 22,9 23,3 22,3	22,8 22,5 21,6 22,5	22,9 23,0 23,2 21,8	22,3 22,0 22,8 23,0	Y_{2..}= 542,0 y _{2..} = 22,5833
Y_{2j}. y _{2j} .	88,3 22,075	91,9 22,975	91,4 22,850	89,4 22,350	90,9 22,725	90,1 22,525	Y_{...}= 1108,1
Y_j. y _j .	184,5 23,0625	185,9 23,2375	185,5 23,1875	182,9 22,8625	185,4 23,1750	183,9 22,9875	y _{...} = 23,085417

Für unser Beispiel ergibt sich:

$$SQ_{\alpha} = 24 \cdot ((23,5875 - 23,085417)^2 + (22,5833 - 23,085417)^2) = 12,100208$$

$$SQ_{\beta} = 8 \cdot ((23,0625 - 23,085417)^2 + \dots + (22,9875 - 23,085417)^2) = 0,8110417$$

$$SQ_{\alpha\beta} = 4 \cdot ((24,050 - 23,5875 - 23,0625 + 23,085417)^2 + \dots + (22,525 - 22,5833 - 22,9875 + 23,085417)^2) = 2,5960417$$

$$SQ_T = (23,9 - 23,085417)^2 + \dots + (23,0 - 23,085417)^2 = 68,02$$

Eingetragen in der Tabelle der Varianzursachen ergibt sich:

Varianzursache	Quadratsumme (SQ)	Freiheitsgrade (FG)	Durchschnittsquadrat	F-Wert
Faktor α	$SQ_{\alpha} = 12,100208$	1	12,100208	8,2952775
Faktor β	$SQ_{\beta} = 0,8110417$	5	0,1622083	0,1112016
Faktor $\alpha\beta$	$SQ_{\alpha\beta} = 2,5960417$	5	0,5192083	0,3559424
Fehler	$SQ_R = 52,512709$	36	1,4586863	
Total	$SQ_T = 68,02$	47		

Die entsprechenden F-Werte der Tabelle lauten:

$$F(1,36;0,05) = 4,11$$

$$F(5,36;0,05) = 2,48$$

Damit lauten die entsprechenden H_0 -Hypothesen:

Da $8,29 > 4,11$ gilt: Die H_0 "Der Einflussfaktor $\alpha = 0$; d.h. es gibt keinen Einfluss der Sequenz" muss verworfen werden.

Da $0,11 < 2,48$ gilt: Die H_0 "Der Einflussfaktor $\beta = 0$; d.h. es gibt keinen Einfluss der Tage" kann nicht verworfen werden.

Da $0,35 < 2,48$ gilt: Die H_0 "Der Einflussfaktor $\alpha\beta = 0$; d.h. Wechselwirkung=0" kann nicht verworfen werden.

3-faktorielle Varianzanalyse

Im folgenden das Beispiel zur Verbrauchsanalyse eines LKW (in zehntel Liter). Wobei in den Zeilen die unterschiedlichen Straßenbeläge (1, ..., 6) in den Spalten die unterschiedlichen Oktanzahlen (a, ..., f) und die Reifenqualitäten (A, ..., F) folgendermaßen angeordnet sind:

D	C	A	B	E	F
E	A	F	C	B	D
B	D	C	F	A	E
F	E	B	D	C	A
C	F	E	A	D	B
A	B	D	E	F	C

Man nennt einen derartige Anordnung wo jeder Einflussfaktor je einmal pro Zeile bzw. Spalte vorkommt auch lateinische Quadrate. Eine rollende Zeilen-(Spalten)vertauschung ergibt unterschiedliche lateinische Quadrate. Man wählt dabei aus diesen vielen lateinischen Quadraten zufällig eines aus. Die Benzinverbräuche in zehntel Liter ergaben sich wie folgt:

Belag\Oktanzahl	a	b	c	d	e	f	Zeilensumme
1	186	187	208	222	296	331	1430
2	213	134	296	265	250	253	1411
3	198	155	272	290	261	310	1486
4	233	184	218	234	248	239	1356
5	245	233	282	248	247	303	1558
6	196	228	242	255	273	274	1488
Spaltensumme	1271	1121	1518	1514	1575	1730	8729

Die Summe der Reifenqualitäten ergibt sich wie folgt:

- A = 1286
- B = 1419
- C = 1511
- D = 1317
- E = 1540
- F = 1656

Es lässt sich beweisen dass:

$$SQ_T = \sum_i \sum_j \sum_k (y_{ijk} - y_{...})^2 = (186 - 242,4722)^2 + (187 - 242,4722)^2 + \dots + (274 - 242,4722)^2 = 68123$$

auch wie folgt berechnet werden kann

$$SQ_T = \sum_i \sum_j \sum_k (y_{ijk})^2 - C$$

mit

$$C = \frac{Y_{...}^2}{m^2} = \frac{8729^2}{36} = 2116540$$

mit

$$\sum_i \sum_j \sum_k (y_{ijk})^2 = 2184663$$

ergibt sich

$$SQ_T = 2184663 - 2116540 = 68123$$

Damit ergibt sich die Tabelle der Varianzursachen wie folgt: