Dokumentation zur IHK-Abschlussprüfung Sommer 2020

Entwicklung eines Softwaresystems

**Geschrieben von: Felix Warschewski**

**Prüflingsnummer 101 20580**

**Tag der Einreichung: 26.06.2020**

**Programmiersprache: C#**

Inhaltsverzeichnis

[1 Eigenhändigkeitserklärung 3](#_Toc43975445)

[2 Änderungen zu Tag 1 4](#_Toc43975446)

[3 Benutzeranleitung 4](#_Toc43975447)

[3.1 Laufzeitumgebung 4](#_Toc43975448)

[3.2 Installation 4](#_Toc43975449)

[3.3 Dateistruktur 4](#_Toc43975450)

[3.4 Dateiformat 4](#_Toc43975451)

[3.5 Aufrufbefehle 5](#_Toc43975452)

[4 Beschreibung der Zufallszahlengeneratoren und Verfahren 6](#_Toc43975453)

[5 Diskussion der Güte-Testverfahren und der Parameter für den LCG 7](#_Toc43975454)

[6 Diskussion der Testbeispiele 9](#_Toc43975455)

[7 Vergleich und Interpretation der Ergebnisse 10](#_Toc43975456)

[8 Zusammenfassung und Ausblick 12](#_Toc43975457)

[9 Programmcode im Anhang 13](#_Toc43975458)

# Eigenhändigkeitserklärung



# Änderungen zu Tag 1

Sämtliche Variablen und Eigenschaften mit dem Datentyp long wurden in double umgeändert. Dies betreffen vor allem die Einstellparameter, sowie Rechnungen und hat den Vorteil, dass sämtliche mathematische Funktionen nicht umgewandelt werden müssen.

Zur Speicherung der Beispiel-LCGs wird eine Konstanten-Klasse verwendet. In dieser werden die Beispielparameter für die LCGs als statische Eigenschaften festgelegt.

Darüber hinaus wurde anstatt einer einfache Verteilungsklasse eine Verteilungsschnittstelle implementiert, welche zwei Unterklassen (Gleichverteilung, Standardnormalverteilung) bereitstellt. Diese Unterklassen besitzen eine Funktion transformiere und in dieser Funktion wird für den Wechsel in die Standardnormalverteilung die Polar-Methode angewendet.

Die Polar-Klasse wird gestrichen und die Funktionalität der Polar-Methode wird in der Funktion „transformiere(x)“ der Standardnormalverteilungsklasse implementiert. Dies hat den Vorteil, dass die Funktion direkt an der richtigen Stelle ist und nicht erst noch über eine Hilfs-Klasse aufgerufen werden muss.

# Benutzeranleitung

## Laufzeitumgebung

Das verwendete Betriebssystem ist Windows 10 64-bit Version. Die Programmiersprache ist C# (Version C# 7.3) und es wurde in der Entwicklungsumgebung Visual Studio 2017 Enterprise Edition geschrieben. Das Zielframework der Konsolen-Applikation ist das .NET Standard 2.0.

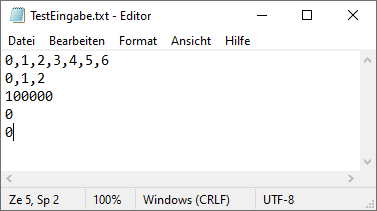
## Installation

Die Installation erfolgt über das Extrahieren der vorhandenen zip-Datei.

## Dateistruktur

## Dateiformat

Mithilfe des args-Parameter in der Main-Methode kann ein Dateipfad übergeben werden. Dieser Pfad sollte auf eine Textdatei zeigen, in der Daten zur Bedienung der Konsolenanwendung stehen.



Die erste Zeile beschreibt die Zufallszahlengeneratoren, welche überprüft werden sollen. Die zweite Zeile beschreibt die Güte-Testverfahren, die angewendet werden sollen. Die dritte Zeile legt fest, wie hoch die Sequenzlänge ist, also wie viele Zufallszahlen für die Prüfungen erstellt werden. Die vierte Zeile legt ein k fest. Bei der seriellen Autokorrelation ist das k die Ordnung der Autokorrelation und steht für den Abstand der verglichenen Paare. Bei dem Sequenz-Up-Down-Test bestimmt das k die Länge einer bestimmten Bitfolge. Gibt man 0 für k an, erstellt die serielle Autokorrelation einen zufälligen Abstand k und bei dem Sequenztest wird über alle k iteriert und letztlich die Differenz ausgegeben. Bei dem Sequenz-Up-Down-Test ist k die Kettenlänge, welche verglichen werden soll. Die letzte Zeile gibt die Art der Verteilung der Zufallszahlen an. Man kann unterscheiden zwischen standardnormalverteilten Zufallszahlen und gleichverteilten Zufallszahlen. Hierbei sind gleichverteilte Zufallszahlen der Standard.

## Aufrufbefehle

Innerhalb der aufzurufenden Textdatei:

Zeile 1 – Generatoren:

1. Ansi-C
2. Minimal Standard
3. RANDU
4. SIMSCRIPT
5. NAG’s LCG
6. Maple’s LCG
7. Datumsbasiert(eigener)

Zeile 2 – Güte-Testverfahren

1. Serielle Autokorrelation
2. Sequenz-Up-Down-Test
3. Eigener Güte-Test

Zeile 3 – Sequenzlänge

Zeile 4 – k

Wenn k = 0 dann werden die allgemeinen Bedingungen übernommen. Siehe 6.4 Dateiformat. Wenn k != 0 dann wird die k-te Ordnung des Verfahrens berechnet und zurückgegeben.

Zeile 5 – Verteilung

1. Gleichverteilung
2. Standardnormalverteilung

# Beschreibung der Zufallszahlengeneratoren und Verfahren

Zu implementieren waren folgende Generatoren und Testverfahren:

* Linearer Kongruenz-Generator mit 6 verschiedenen Einstellungsmöglichkeiten
* Umwandlung gleichverteilter in normalverteilte Zufallszahlen mithilfe der Polar-Methode
* Eigener Zufallszahlengenerator (Datumsbasiert)
* Güte-Testverfahren Serielle Autokorrelation
* Güte-Testverfahren Sequenz-Up-Down-Test
* Eigenes Güte-Testverfahren

Der Lineare Kongruenz-Generator (LCG) errechnet Zufallszahlen auf Basis der Kongruenzrechnung. Hierfür sind 4 Eingabeparameter wichtig. Das Modul m, der Multiplikator a, das Inkrement c und der Startwert x0. Des Weiteren waren sechs zu implementierenden Verfahren für den LCG in Tabellenform gegeben. Diese Verfahren unterscheiden sich in der Wahl ihrer Einstellungsparameter. Der LCG funktioniert, wie eine Folge, wobei jeder Aufruf eine andere und zufällige Zahl erstellen soll.

Der LCG erzeugt hauptsächlich gleichverteilte Zufallszahlen. Daher sollte eine Methode zur Transformierung von gleichverteilte in normalverteilte Zufallszahlen implementiert werden. Diese Methode heißt die Polar-Methode. Hierbei werden euklidische Koordinaten, welche von den verschiedenen Zufallsverfahren generiert werden, in Polarkoordinaten umgewandelt und wenn die beiden Koordinaten als Punkt im Einheitskreis liegen, werden sie zu zwei unabhängigen standardnormalverteilten Zufallszahlen transformiert. Dieses Prinzip der Verteilungstransformation kann auf jedes Verfahren des Linearen Kongruenz-Generators angewendet werden.

Der eigene Zufallszahlengenerator funktioniert auf Basis der Zeit und Laufzeit des Programmes. Es werden die Sekunden, Minuten, Stunden, Tage, Monate als Integer aufsummiert, sowie die aktuelle Laufzeit des Programmes in Millisekunden. Darüber hinaus wird gespeichert, wie viele Zufallszahlen bereits mit einer Instanz des Generators erzeugt wurden. Die Summe wird mit der Anzahl der bereits erzeugten Zufallszahlen multipliziert. Das Ergebnis wird anschließend mit Modulo 1000 gerechnet und durch das Modulo geteilt. So ergibt sich eine Zufallszahl zwischen 0 und 1 mit 3 Nachkommastellen. Das bedeutet sie kann bis 1000 hochskaliert werden ohne generierte Nullen angehängt zu bekommen. Will man höher skalieren benötigt man ein höheres Modul.

Das Güte-Testverfahren mithilfe der seriellen Autokorrelation misst die Abhängigkeit einer Folge von Zahlen. Die Zufallszahlen sollen paarweise möglichst unabhängig sein und im Optimalfall eine Korrelation von p=0 aufweisen. Das Verfahren verwendet eine Sequenz von Zufallszahlen, einen Stichprobenmittelwert von 0.5 in der Gleichverteilung und 0 in der Normalverteilung, sowie die Ordnung k. Je nachdem wie hoch die Ordnung ist, prüft man auf weiter entferntere oder näher aneinanderlegende Paare in der Sequenz. Die Wahl von k kann das Ergebnis des Güte-Tests stark beeinflussen. Daher sollte man zum Testen von bestimmten Generatoren die Verfahren mehrmals mit unterschiedlichen Ordnungen durchlaufen lassen. Wenn man kein k angibt (k=0), wird dies zufällig generiert.

Das nächste Güte-Testverfahren ist der Sequenz-Up-Down-Test. Dieser prüft wie viele Folgen bestimmter Länge von stetig größer oder kleiner werdenden Zahlen in einer Zufallszahlensequenz vorhanden sind. Es wird eine Bit Maske erstellt mit den Werten 0 und 1. Der Wert 0 wird in die Maske eingefügt, sobald ein folgender Wert kleiner ist als der jetzige. Wenn der folgende Wert hingegen größer als der derzeitige ist, wird eine 1 in die Bit Maske eingefügt. Das Ergebnis des Verfahrens gibt einen Aufschluss darüber, wie viele Folgen es mit einer bestimmten Länge gab. Je höher die Zahl bei niedrigen Werten für k (Kettenlänge) ist, desto besser funktioniert Zufallszahlengenerator, da die Werte öfter hin und her springen und sich nicht stetig aufbauen.

Das eigene Güte-Testverfahren erstellt auch eine Bit Maske. Hier werden alle Werte gleich dem Sequenz-Up-Down-Verfahren in die Bit Maske geschrieben. Anschließend wird die Bit Maske aufsummiert und durch die Anzahl der vorhandenen Bits der Bit Maske dividiert. Dies gibt Aufschluss über den Mittelwert der Wechsel der Zahlenfolgen. Wenn der Mittelwert deutlich unter 0.5 liegt, ist das ein Zeichen für sehr wenige Wechsel und die Entstehung längeren Folgen. Das bedeutet die Güte des Generators wäre schlechter, als wenn der Mittelwert über 0.5 liegt. Je höher desto öfter wechselt er innerhalb der Zufallszahlensequenz eine Reihenfolge und desto zufälliger werden die Zahlen generiert.

# Diskussion der Güte-Testverfahren und der Parameter für den LCG

Es werden sechs verschiedene Verfahren für den Linearen Kongruenz-Generator vorgestellt. Diese unterscheiden sich teils stark und haben deswegen auch eine andere Abfolge zur Erzeugung von Zufallszahlen.

Der Ansi-C LCG besitzt vergleichsweise hohe Parameter zu dem Rest der Verfahren. Schon im zweiten Schritt wird das Modul häufig übersprungen. Da das Verfahren so hoch skaliert ist, bekommt es für längere Zahlenfolgen auch immer gleichere Muster heraus und wird bei einer hohen Sequenzlänge schlechter abschneiden als Verfahren mit kleineren Multiplikatoren. Das Ansi-C Verfahren weist auch ein Muster von Geraden und Ungeraden Zufallszahl im dauerhaften Wechsel auf. Dies liegt daran, dass das Modul eine gerade Zahl ist und der Multiplikator ungerade. Durch das ungerade Inkrement springen die Zufallszahlen immer von geraden Zahlen zu ungeraden Zahlen und umgekehrt.

Für das Minimal Standard-Verfahren wird ein ungeraderes Modulo verwendet und ein deutlich kleinerer Multiplikator als bei den meisten anderen Verfahren. Der kleinere Multiplikator sorgt dafür, dass die Zahlen in den ersten Schritten nicht so häufig über das Modul herauskommen. Nichtsdestotrotz sollte durch die exponentiell ansteigende Zahl, vor Abzug des Modulo, das Modul schnell erreicht werden.

Das RANDU -Verfahren fängt mit dem Startwert 1 an und wird mit einem ungeraden Multiplikator multipliziert. Da das Modulo gerade ist, bekommen wir nur ungerade Zahlen heraus. Das bedeutet die mögliche Zahlenmenge des RANDU -Verfahren ist halbiert. Des Weiteren weisen die ersten zehn generierten Zufallszahlen eine steigende Abfolge auf, obwohl das Modulo mehrmals überschritten wird. Dies sollte sehr schlechte Güte-Ergebnisse für kleine Sequenzlängen bis 10 zu Folge haben.

SIMSCRIPT besitzt als einziges Verfahren einen geraden Multiplikator und ein ungerades Modul. Das Verfahren ist von der Höhe der Parameter im Vergleich zu den anderen am meisten ausgeglichen. Daher kann die Vermutung aufgestellt werden, dass die generierten Zufallszahlen über alle Sequenzlängen am besten verteilt ist.

NAG‘s LCG generiert ausschließlich ungerade Zahlen, wobei der zur Verfügung gestellte Zahlenraum mit 2 hoch 59 deutlich über dem der anderen Verfahren liegt. Das gibt eine breitere Streuung der generierten Zufallszahlen, wenn man von der Differenz zu einem Mittelwert ausgeht. Dies könnte ein Nachteil für das Güte-Testverfahren der seriellen Autokorrelation darstellen. Des Weiteren ist der Startwert relativ hoch und das Verfahren hat früh eine gewisse Länge der Zufallszahlen erreicht. Das bedeutet das Verfahren funktioniert für eine geringe Sequenzlänge schon willkürlich, im Gegensatz zum RANDU-Verfahren.

Bei dem Maple’s LCG wird als Modul ein in Relation niedriges Modul gewählt in Kombination mit einem sehr hohen Multiplikator. Dies führt dazu, dass das Modul oft übertroffen wird. Da der Multiplikator beinahe so groß wie das Modul ist, werden vor allem Zahlen generiert, welche das Modul beinahe annähern. Daher haben diese auch im Verhältnis zueinander keine allzu großen Abweichungen und sind relativ konstant. Das kann zur Folge haben, dass das Verfahren eine relativ konstant gute Güte über diverse Sequenzlängen aufzeigt.

Um zwei Beispielverfahren aus der Menge der sechs gegebenen Verfahren zu wählen, verwende ich den Ansi-C LCG und den RANDU LCG. Beide haben ein Modul von 2 hoch 31. Des Weiteren haben beide Verfahren für das Modul einen geraden Wert und für den Multiplikator einen ungeraden Wert. Ohne das ungerade Inkrement des Ansi-C Verfahrens würde dieser auch ausschließlich ungerade Zahlen generieren. Durch das Inkrement werden aber gerade und ungerade Zufallszahlen im Wechsel generiert. Der Ansi-C Generator hat einen deutlich höheren Multiplikator von 1103515245 im Vergleich zu dem RANDU-Multiplikator von 65539. Das Inkrement und der Startwert bei dem Ansi-C ist auch höher. Allein dies gibt eine Aussage darüber, wie schnell der Algorithmus den Wert des Moduls erreicht und überschlägt. Wenn er den Wert des Moduls übertrifft, fängt dieser wieder bei 0 an und zählt sich nach oben. Daher kann man sagen, dass der Ansi-C deutlich schneller sein Modul erreicht und einen Wechsel von größeren und kleineren Zahlen besitzt. Der RANDU hingegen braucht länger und besitzt mehr Folgen von aufeinander größer werdenden Zufallszahlen. Dies sagt aus, dass für das Sequenz-Up-Down-Verfahren der Ansi-C LCG eine bessere Güte für eine kleine Sequenzlänge besitzt. Sobald aber der Parameter x eine gewisse Größe erreicht hat, werden die Sprünge immer größer und das Ergebnis überschlägt öfter das Modul. Dies bedeutet, dass die Werte öfter variieren und die Folgen von aufeinanderfolgenden größer werdenden Zufallszahlen deutlich kleiner werden. Daher sollte man die beiden Zufallszahlengeneratoren mit mehreren beliebig großen Sequenzlängen testen. Es wird also angenommen, je länger die Sequenzlänge desto besser ist die Güte für den RANDU LCG im Up-Down-Test. Der Ansi-C sollte hingegen schon relativ früh eine gute Güte besitzen und keine drastische Verbesserung bei steigender Sequenzlänge erhalten.

Im Endeffekt kann man aber sagen: Je höher die Sequenzlänge, desto besser wird die Güte der Zufallszahlengeneratoren. Das „Lag“ k bzw. die Ordnung k kann bei verschiedenen LCGs die Wirkung auf die Güte variabel beeinflussen.

# Diskussion der Testbeispiele

Die Testbeispiele geben Aufschluss über die Güte der Zufallszahlengeneratoren. Deswegen muss man hier differenzieren und kann nicht nur ein Testverfahren mit einer bestimmten Sequenzlänge und einer bestimmten Ordnung verwenden. Es müssen mehrere unterschiedliche Testvorgänge entstehen, welche sich möglichst unterscheiden.

Aus dem Punkt 5 kann angenommen werden, dass sich die Güte der Testverfahren je nach Sequenzlänge ändern kann. Genauso können Rückschlüsse über einen Zufallszahlengenerator gewonnen werden, indem man die Abstände der verglichenen Zufallszahlen beziehungsweise die Kettenlängen der überprüften Folgen ändert. Die Änderungen kann man mit der Variable „K“ setzen, welche in der Eingabe-Datei gesetzt werden kann.

Zum Vergleich werden folgende Testfälle betrachtet. Alle Testfälle können mit ihren Werten in dem Auslieferungsordner unter „TestsZurDoku“ gefunden werden.

Fall 1:

Alle LCG-Verfahren mit serieller Autokorrelation für eine Sequenzlänge von 10

Fall 2:

Alle LCG-Verfahren mit serieller Autokorrelation für eine Sequenzlänge von 1000

Fall 3:

Alle LCG-Verfahren mit serieller Autokorrelation für eine Sequenzlänge von 100000

Fall 4:

Alle LCG-Verfahren mit Sequenz-Up-Down-Test für eine Sequenzlänge von 10

Fall 5:

Alle LCG-Verfahren mit Sequenz-Up-Down-Test für eine Sequenzlänge von 1000

Fall 6:

Alle LCG-Verfahren mit Sequenz-Up-Down-Test für eine Sequenzlänge von 100000

Fall 7:

Alle LCG-Verfahren mit Sequenzlänge 100000 und einem k von 1

Fall 8:

Alle LCG-Verfahren mit Sequenzlänge 100000 und einem k von 10000

Fall 9:

Alle LCG-Verfahren mit beiden vorgegebenen Testverfahren und einer Länge von 100000 und Standardnormalverteilung

Fall 10:

Eigener Datumsbasierter Zufallszahlengenerator für eine Sequenzlänge von 1000

Fall 11:

Alle LCG-Verfahren auf den eigenen Güte-Test für eine Sequenzlänge von 100000

# Vergleich und Interpretation der Ergebnisse

Um die Ergebnisse der Testfälle besser einordnen zu können, folgt eine Beschreibung ihrer Aussagekraft. Das Güte-Testverfahren der seriellen Autokorrelation fällt besser aus je näher das Ergebnis an null liegt. Wenn das Ergebnis sich signifikant von null unterscheidet, wird der Zufallsgenerator abgelehnt. In dem Fall ist dieser nicht brauchbar, da er Muster aufzeigt und voneinander abhängige Zufallszahlen bildet.

Der Sequenz-Up-Down-Test gibt einen Wert für die Gesamtdifferenz zu dem optimalen Wert aller Kettenlängen zurück. So kann gesagt werden, wie stark ein Verfahren von der optimalen Verteilung abweicht. Wenn der Wert deutlich zu den anderen gewonnen Werten ausschlägt, kann man sagen, dass die Verteilung der Folgen von verschiedenen Kettenlängen besonders ist und nicht optimal. Es gibt dann Ausreißer für verschiedene Kettenlängen und dies ist ein Zeichen dafür, dass Muster gebildet werden.

Der Vergleich der Testergebnisse, sowie der Vergleich folgt zuerst in Reihenfolge der Testfälle. Es wird zu jedem Fall Stellung bezogen und direkte Zusammenhänge interpretiert. Die Zusammenfassung über alle Generatoren und Verfahren folgt im Kapitel 8 „Zusammenfassung und Ausblick“.

Fall 1: Alle LCG-Verfahren mit serieller Autokorrelation für eine Sequenzlänge von 10

Die Folge, die am schlechtesten abschneidet, ist das RANDU-Verfahren. Es wird ein Wert von 1 zurückgegeben. Dies ist der schlechtmöglichste Wert, der erreicht werden kann. Dies liegt daran, dass die Werte für RANDU in dem Intervall 1-10 stetig größer werden. Die generierten Zahlen sehen nicht aus wie Zufallszahlen, sondern wie eine immer größer werdende Folge. Das beste Ergebnis für eine derartig kleine Sequenzlänge liefert das Minimal Standard-Verfahren mit -0,2. Ein negatives Ergebnis kommt zustande, wenn mehr Zufallszahlen unter dem Stichprobenmittelwert von 0.5 liegen als darüber. Das Ergebnis ist das beste von allen sechs Generatoren. Dennoch ist der Wert keineswegs optimal. Dies liegt vor allem daran, dass alle Verfahren auf dem Linearen Kongruenz-Generator basieren und dieser wie eine Folge mit spezifischen Eingabeparametern funktioniert. Der LCG braucht eine bestimmte Menge an Durchläufen, bis er zuverlässige Zufallszahlen ausgibt. In den nächsten Testschritten wird geprüft, wie lange die verschiedenen Verfahren brauchen, um eine gute Güte zu erreichen.

Fall 2: Alle LCG-Verfahren mit serieller Autokorrelation für eine Sequenzlänge von 1000

Die Werte für die Güte der Verfahren hat sich deutlich gebessert. Der einzige Zufallszahlengenerator, welcher eine unbrauchbare Güte aufweist, ist der NAG’s LCG mit einer Güte von -0,42. Dieser Generator besitzt das größte Modulo. Durch die extreme Breite des Zahlenraums, der mithilfe des gewählten Modulo zur Verfügung gestellt wird, treten auch stärkere Abweichungen von einem Mittelwert auf. Nun ist abzuwarten, ob bei mit einer größeren Sequenzmenge sich die Güte des Verfahrens verbessert.

Sehr gute Werte liefern die Generatoren Minimal Standard (0,0006) und Maple’s LCG (-0,0003). Das Maple’s Verfahren nähert, wie in Punkt 4 beschrieben, das Modul an und liefert Zufallszahlen in einem recht konstanten Bereich. Dies reflektiert das starke Güte-Testergebnis mit der seriellen Autokorrelation bei einer Sequenzlänge von 1000. Bei dem Minimal Standard-Verfahren wird genau das gegenteilige Prinzip angewandt. Hier ist der Multiplikator in Relation niedrig und das Modul größer als beim Maple’s LCG. Dies sorgt dafür, dass sich die Zahlen relativ schnell ausgleichen und nicht allzu weit voneinander entfernt liegen.

Fall 3: Alle LCG-Verfahren mit serieller Autokorrelation für eine Sequenzlänge von 100000

Für den dritten Testfall bleibt das NAG’s Verfahren das mit der schlechtesten Güte. Diese hat sich sogar mit zunehmender Sequenzlänge nochmal verschlechtert (-0,5). Dies wird vor allem durch den großen Zahlenraum, welcher durch das große Modulo bereitgestellt wird, hervorgerufen. Die Werte können sich hier deutlicher unterschieden als bei kleineren Zahlenräume.

Die Werte für die Güte der anderen Generatoren haben sich nicht sonderlich verändert. Die besten Gütewerte haben die Verfahren RANDU (0,0018), SIMSCRIPT (-0,0013) und Maple’s (-0,0024).

Das Minimal Standard-Verfahren, welches im vorherigen Testfall die beste Güte aufzeigte, hat sich auf eine Güte von 0,01 verschlechtert. Dies kann bedeuten, dass der Minimal Standard bis 1000 Zufallszahlen eine relativ geringe Abweichung aufweist und danach wieder mehr streut. Also wird es bei diesem Verfahren Schübe geben, in denen die Folge besser zufällig generiert werden bzw. schlechter.

Für das Güte-Testverfahren der seriellen Autokorrelation ist besonders der RANDU-Generator hervorzuheben. Bei einer sehr geringen Sequenzlänge hat dieser noch die schlechtmöglichste Güte und je größer die Zahlen werden, desto besser wird seine Güte. Darüber hinaus kann man über den NAG’s LCG sagen, dass der große Zahlenraum die Werte deutlich weiter streuen lässt und die Differenzen der einzelnen Zahlen zum Mittelwert sehr verschieden ausfallen können. Was den Generator nach dem Prinzip der seriellen Autokorrelation schlecht macht.

Fall 4: Alle LCG-Verfahren mit Sequenz-Up-Down-Test für eine Sequenzlänge von 10

Bei dem Sequenz-Up-Down-Test schneidet der Maple’s Generator am schlechtesten ab (3,35). Es wird die Differenz für die Häufigkeit sämtlicher Kettenlängen mit der Häufigkeit im angenäherten Optimalfall gebildet. Wenn eine hohe Zahl als Güte ausgegeben wird, bedeutet dies, dass der Generator über den ganzen Verlauf der optimalen Häufigkeit verschiedener Kettenlängen relativ schlecht war. Das beste Ergebnis erzielte RANDU mit einer Güte von 0,999. Da man pro Kettenlänge die absolute Differenz auf die Gesamtgüte aufsummiert, besagt das Ergebnis wie sehr die Kurve der Optimal Verteilung approximiert wird. Für eine derartig niedrige Sequenzlänge sind die Ergebnisse nicht wirklich aussagekräftig. Dennoch kann beobachtet werden, das RANDU am besten abschneidet, wobei das RANDU-Verfahren bei der seriellen Autokorrelation am schlechtesten bei niedrigen Sequenzlängen abgeschnitten hat. Die zeigt, dass sich die Verfahren in der Erfassung der Güte deutlich unterscheiden.

Fall 5: Alle LCG-Verfahren mit Sequenz-Up-Down-Test für eine Sequenzlänge von 1000

Für diesen Testfall schneidet der NAG’s LCG am schlechtesten ab (150,15). Am besten verläuft das Testverfahren für den SIMSCRIPT-Generator mit 23,47.

Das NAG’s-Verfahren hat stärkere Ausreißer für die Anzahl mancher Kettenlängen. Also gibt es Stellen an denen das Verfahren besonders häufig oder selten auf oder absteigende Zahlenfolgen erstellt, die an den Stellen der optimalen Verteilung liegen.

Der SIMSCRIPT hingegen hat eine relativ ähnliche Verteilung, wie die optimale Verteilung.

Fall 6: Alle LCG-Verfahren mit Sequenz-Up-Down-Test für eine Sequenzlänge von 100000

Das Maple’s LCG-Verfahren approximiert am besten die optimale Verteilung mit einer Gesamtdifferenz von 242,9. Der NAG’s LCG sticht deutlich mit einer schlechten Approximation hervor. Die Gesamtdifferenz liegt bei 15000. Durch den immensen Zahlenraum und den hohen Startwert, sowie den großen Multiplikator werden die Zufallszahlen vor allem in der zweiten Hälfte des Zahlenraums generiert. Nun muss geprüft werden, ob die extremen Differenzen für kleine Zeichenketten(k) oder für große Zeichenketten(k) auftreten. Dies geschieht in den folgenden Testfällen.

Fall 7 & 8: Alle LCG-Verfahren mit Sequenzlänge 100000 und einem k von 1 und 10000

Der NAG’s LCG weist eine absolute Differenz von über 8000 für k = 1 aufweist. Diese Differenz entsteht dadurch, dass es in dem NAG-Verfahren mehr Zeichenkettenfolgen mit der Länge 1 gibt. Der richtige Differenzwert zwischen optimaler Verteilung und NAG liegt nämlich bei -8000. Dies bedeutet, dass es kaum immer größer werdende oder kleiner werdende Folgen gibt, sondern die Muster durch das springen der Zufallszahlen entstehen. Es muss ein Muster vorhanden sein, sonst wäre der Wert bei der seriellen Autokorrelation nicht derartig schlecht.

Ansonsten sind keine großen Auffälligkeiten bei der Betrachtung verschiedener k mit gleicher Sequenzlänge zu sehen.

Fall 9 & 10 & 11:

Diese Testfälle zeigen die Funktionsweise der Polar-Methode, des eigenen Zufallszahlengenerators und der eigenen Güte-Funktion.

Bei der Anwendung der Standardnormalverteilung auf alle Zufallszahlen ist zu erkennen, dass die Werte relativ ähnlich bleiben. Es werden nur die Zahlenräume verschoben und die Gesamtzusammenhänge der Güte-Ergebnisse sollten sich nicht ändern. Die Zufallszahlen wurden alle mit dem gleichen Verfahren von einem gleichverteilten Intervall [0,1] auf eine Standardnormalverteilung mit dem Intervall [-1,1] verschoben.

Der eigene Zufallszahlengenerator, welcher Datumsbasiert Zahlen generiert, besitzt eine nicht optimale Güte. Dies könnte aber durch die Erweiterung des Modulo verbessert werden oder durch die Erweiterung der Datumsparameter.

Das eigene Güte-Testverfahren liefert den Mittelwert der Wechsel der Zahlenfolgen. Also wie häufig eine Folge mit stetigen Zahlen unterbrochen. Je höher dieser Wert ist, desto besser. In dem Testfall 11 lieferte der NAG’s LCG die besten Werte, da diese am höchsten waren mit ungefähr 0,75. Dies bedeutet, dass es kaum große Zahlenfolgen gibt, die sich stetig aufbauen.

# Zusammenfassung und Ausblick

Zusammenfassend kann man sagen, dass alle Generatoren ab einer bestimmten Sequenzlänge eine gute Güte aufzeigen, außer das NAG’s LCG-Verfahren.

Der NAG’s LCG liefert stetige Wechsel. Dies wurde erfasst aus dem eigenem Güte-Testverfahren und Sequenz-Up-Down. Da der Zahlenraum extrem groß ist, unterscheiden sich die Differenzen zum Mittelwert stark und es kommt eine schlechte Güte heraus. Der NAG’s LCG ist also kein schlechter Generator, welcher nur Muster liefert, sondern die Testverfahren waren eher auf die anderen Generatoren angepasst und nicht auf den LCG.

Ein besonderer Generator ist der RANDU LCG, welcher nur ungerade Zahlen liefert. Er ist anfangs für kleine Sequenzlängen sehr schwach, aber je größer die Sequenzlänge wird, desto eine bessere Güte bekommt er.

Es kann also gesagt werden, dass keine Generatoren verworfen werden sollen, ausgenommen dem NAG LCG. Denn ab einer bestimmten Sequenzlänge funktionieren alle Verfahren sehr gut und können verwendet werden. Für alle Generatoren sollte es eine Mindestsequenzlänge von deutlich über 10 geben, da sie ab meist erst gut funktionieren.

Um die Funktionalität des NAG’s LCG nachzuprüfen, müsste ein weiteres Verfahren benutz werden, was einen differenzierten Blick auf die Güte wirft.

In weiteren Schritten der Programmierung könnte man die Klassenbibliothek um weitere LCG-Schemas erweitern. Damit stellt man eine breitere Funktionalität des LCG-Verfahrens bereit. Darüber hinaus kann man mehr Verteilungen an die Verteilungs-Schnittstelle anbinden. Dies ermöglicht die Transformation von Koordinaten oder Zahlen in andere Verteilungen. Dafür müsste man lediglich eine neue Klasse anlegen mit der dazugehörigen „Transformiere()“-Methode der Verteilungsschnittstelle. Des Weiteren kann man Klassen an die Güte-Tests-Schnittstelle an klemmen. Man kann mehr Testverfahren benutzen um die Güte der verschiedenen Generatoren, welche erzeugt wurden, zu bestimmen. Dies kann man alles mit beliebigen Parametern für die Anzahl der zufallsgenerierten Element (Sequenzlänge) und die Ordnung k der Funktion ausführen, indem die Berechne()-Methode aus dem Interface „GüteTests“ verwendet wird. Die drei Schnittstellen sind Hauptbestandteile um diverse Funktionalitäten zu erweitern. Neue Tests einzuführen, in andere Verteilungen umzuwandeln oder komplett andere Zufallszahlen zu erstellen.

# Programmcode im Anhang

In der Programmierung für das Güte-Testverfahren Sequenz-Up-Down die Differenz zwischen Nopt(k) und N(k) gebildet. Wenn negative Werte für den Algorithmus herauskommen, bedeutet dies, dass häufig der Wert von N(k) größer als der von dem optimalen Wert Nopt(k) gewesen ist. Andererseits wenn öfter kleiner Werte für N(k) herauskommen, als die von Nopt(k), ist die Differenz der beiden Werte positiv. Ein Ergebnis für eine starke Güte eines Zufallszahlengenerators wäre hier, wenn der Wert nahe 0 liegen würden. Wenn der Wert zu weit von 0 abweicht, ist die Güte relativ schlecht für den geprüften Zufallszahlengenerator.

Das Sequenz-Testverfahren wurde so implementiert, wenn man eine Zahl für k mitgibt, gibt er die Differenz der gegebenen Folgenlänge zurück. Wenn für k eine 0 mitgegeben wird, wird die Gesamtdifferenz über alle Folgenlängen zurückgegeben.