Inhaltsverzeichnis

Kapitel II

Stochastische Simulation

Hier ein erster Einblick, siehe auch Grinstead, Snell (1997). Mehr zu diesem Themenkreis (Monte-Carlo-Algorithmen) in eigenen Vorlesungen und etwa bei Müller-Gronbach, Novak, Ritter (2008).

Problemstellung: Gegeben sei eine Zufallsvariable X auf $(\Omega, \mathfrak{A}, P)$ sowie eine Menge $M \in \mathfrak{M}$. Berechne $P(\{X \in M\})$.

1 Die Methode der direkten Simulation

Beispiel 1. Erfolgswahrscheinlichkeit einer Strategie beim Patience-Spiel[1]. Dazu sei P die Gleichverteilung auf der Menge Ω aller Permutationen von $\{1, \ldots, 52\}$ (Kartenverteilungen). Ferner sei $X := 1_A$, wobei A die Menge aller Kartenverteilungen, bei denen die Strategie gewinnt, und $M := \{1\}$. Somit

$$P(\{X \in M\}) = P(A) = |A|/|\Omega|.$$

Hier $|\Omega| = 52! = 8.06 \ldots \cdot 10^{67}$, also Ω sehr groß.

Beispiel 2. Durchgang von Neutronen durch Materie. Dazu seien $(S_t^1)_{t \in \mathbb{N}}$, $(S_t^2)_{t \in \mathbb{N}}$ und $(S_t^3)_{t \in \mathbb{N}}$ Folgen von Zufallsvariablen, so daß $(S_t^1(\omega), S_t^2(\omega), S_t^3(\omega)) \in \mathbb{R}^3$ die Position eines Neutrons nach t Kollisionen beschreibt. Schließlich tritt einer dieser Fälle ein:

- Neutron wird von Abschirmung absorbiert, $X(\omega) := 0$,
- Neutron wird von Abschirmung reflektiert, $X(\omega) := 1$,
- Neutron passiert Abschirmung, $X(\omega) := 2$.

Gesucht ist $P(\{X = 2\})$. Hier X „kompliziert".

[1] Ulam (1946), siehe Los Alamos Science, Special Issue.

Beispiel 3. Ruinwahrscheinlichkeiten als stark vereinfachte Version von Beispiel 2. Betrachte ein einfaches Spiel mit Gewinn oder Verlust ± 1 pro Runde mit Wahrscheinlichkeit p bzw. $1-p$, wobei $p \in]0,1[$. Dazu sei $(Y_j)_{j\in\mathbb{N}}$ iid mit

$$P(\{Y_1 = 1\}) = p, \qquad P(\{Y_1 = -1\}) = 1-p.$$

Dann ist

$$S_t := \sum_{j=1}^{t} Y_j$$

der akkumulierte Gewinn nach t Runden. Ziel sei die Vermehrung eines Startkapitals $a \in \mathbb{N}$ um einen Betrag $b \in \mathbb{N}$. Die Spieldauer wird dann beschrieben durch[2]

$$\tau(\omega) := \inf\{t \in \mathbb{N} : S_t(\omega) = -a \text{ oder } S_t(\omega) = b\}.$$

Gesucht ist die Ruinwahrscheinlichkeit

$$r(a,b,p) := P(\{\omega \in \Omega : \tau(\omega) < \infty \text{ und } S_{\tau(\omega)}(\omega) = -a\})$$

d.h. $P(\{X = 1\})$ für $X := 1_{\{\tau<\infty\}\cap\{S_\tau = -a\}}$. Ausblick: Ruinwahrscheinlichkeit eines Schadenversicherers.

Wir zeigen an dieser Stelle nur $P(\{\tau < \infty\}) = 1$. Setze $d := a + b - 1$. Klar: Spiel endet spätestens nach d Gewinnen in Folge. Betrachte also

$$G_j := \{(X_{(j-1)\cdot d+1}, \ldots, X_{j\cdot d}) = (1,\ldots,1)\}, \qquad j \in \mathbb{N}.$$

Dann

$$\{\tau > \ell \cdot d\} \subseteq \bigcap_{j=1}^{\ell} G_j^{\mathrm{c}},$$

und $G_1^{\mathrm{c}}, \ldots, G_\ell^{\mathrm{c}}$ sind unabhängig mit

$$P(G_j^{\mathrm{c}}) = 1 - p^d < 1,$$

siehe $\boxed{\text{ÜBUNG}}$. Es folgt

$$P(\{\tau > \ell \cdot d\}) \le (1-p^d)^\ell,$$

und weiter

$$0 \le P(\{\tau = \infty\}) \le \lim_{\ell\to\infty} P(\{\tau > \ell \cdot d\}) = 0.$$

Bezeichnung. Sei $(X_i)_{i\in I}$ eine Folge von Zufallsvariablen auf $(\Omega, \mathfrak{A}, P)$. Für $\omega \in \Omega$ wird die Folge $(X_i(\omega))_{i\in I}$ reeller Zahlen als eine *Realisierung* von $(X_i)_{i\in I}$ bezeichnet. Spezialfall: $|I| = 1$, Realisierung einer Zufallsvariable.

[2]Konvention: $\inf \emptyset = \infty$.

Methode der direkten Simulation. Betrachte eine iid Folge von Zufallsvariablen $X'_1, \dots, X'_n$ auf $(\Omega', \mathfrak{A}', P')$, wobei X und X'_1 identisch verteilt sind.

1. „Erzeuge" eine Realisierung $(x'_1, \dots, x'_n)$ von $(X'_1, \dots, X'_n)$.
2. Approximiere $P(\{X \in M\}) = P'(\{X'_1 \in M\})$ durch die relative Häufigkeit

$$\frac{1}{n} \cdot \sum_{i=1}^{n} 1_M(x'_i).$$

Dies entspricht der frequentistischen Interpretation von Wahrscheinlichkeiten, siehe Seite 4.

Nun: Approximation der Verteilungsfunktion F_X von X.

Definition 4. Die *empirische Verteilungsfunktion* $F_n(\cdot\,; x_1, \dots, x_n) : \mathbb{R} \to [0,1]$ zu $x_1, \dots, x_n \in \mathbb{R}$ ist definiert durch

$$F_n(x; x_1, \dots, x_n) := 1/n \cdot \sum_{i=1}^{n} 1_{]-\infty,x]}(x_i).$$

Fragen: Konvergieren die relativen Häufigkeiten $\frac{1}{n} \cdot \sum_{i=1}^{n} 1_M(x'_i)$ gegen $P(\{X \in M\})$? Konvergieren die empirischen Verteilungsfunktionen $F_n(\cdot\,; x'_1, \dots, x'_n)$ gegen F_X? Positive Antworten geben wir in Abschnitt V.2.

Die Mathematische Statistik betrachtet bei diesen Fragen statt Simulationsdaten empirische Daten $x'_1, \dots, x'_n \in \mathbb{R}$, siehe bereits Beispiel I.1.13.

2 Zufallszahlen

Die praktische Durchführung der direkten Simulation (Schritt 1.) beruht auf Zufallszahlengeneratoren, d.h. auf geeigneten Abbildungen $f : K \to K$, wobei $K := \{0, \dots, k-1\}$ und k sehr groß[3], etwa $k = 2^{128} = 3.4 \ldots \cdot 10^{38}$ oder $k = 2^{19937} = 4.3 \ldots \cdot 10^{6001}$.

Man erhält endliche Folgen $v_1, v_2, \ldots \in K$ durch die Initialisierung $v_0 \in K$ und die Iteration $v_\ell := f(v_{\ell-1})$. „Kleine" Abschnitte dieser Folgen „verhalten sich annähernd" wie Realisierungen von iid Zufallsvariablen, die jeweils gleichverteilt auf K sind.

Definition 1. Eine Zufallsvariable U heißt *gleichverteilt* auf $[0,1]$, falls

$$F_U(x) = \begin{cases} 0, & \text{falls } x < 0, \\ x, & \text{falls } 0 \le x \le 1, \\ 1, & \text{falls } x > 1. \end{cases}$$

Bez.: $U \sim \mathbf{U}([0,1])$.

[3]Zum Vergleich: die Anzahl der Atome in Universum liegt in der Größenordnung 10^{79}.

Bemerkung 2. Für $U \sim \mathbf{U}([0,1])$, $x \in \mathbb{R}$ und ein Intervall M mit Endpunkten $0 \le a < b \le 1$ gilt

$$\begin{aligned} P(\{U = x\}) &= 0, \\ P(\{U \in M\}) &= F_U(b) - F_U(a) = b - a. \end{aligned}$$

Setze $u_\ell := v_\ell / k$. „Kleine Abschnitte“ der Folgen $u_1, u_2, \ldots \in [0,1[$ „verhalten sich annähernd“ wie Realisierungen von iid Zufallsvariablen, die jeweils gleichverteilt auf $[0,1]$ sind. Kurz: $u_1, u_2, \ldots$ (gleichverteilte) *Zufallszahlen* in $[0,1]$.

In dieser Vorlesung: Zufallszahlengenerator als black-box. Verwendung etwa in der Form

(i) Initialisierung `init(u0)`,

(ii) Iteration `u := rand()`.

Beispiel 3. Simulation des Spiels aus Beispiel 1.3. Dazu sei

$$T(u) := \begin{cases} 1, & \text{falls } u > 1 - p \\ -1, & \text{falls } u \le 1 - p. \end{cases}$$

Ist $(U'_j)_{j \in \mathbb{N}}$ iid mit $U'_1 \sim \mathbf{U}([0,1])$, dann definiert

$$Y'_j := T(U'_j)$$

eine iid-Folge mit

$$P'(\{Y'_1 = 1\}) = p, \qquad P'(\{Y'_1 = -1\}) = 1 - p.$$

Dies entspricht einmaligen Wurf einer fairen ($p = 1/2$) bzw. unfairen ($p \neq 1/2$) Münze. Die Simulation der Münzwürfe geschieht also durch die Transformation

$$x'_i := T(u'_i)$$

von gleichverteilten Zufallszahlen $u'_i \in [0,1]$.

Die folgenden Abbildungen basieren auf der Wahl $p := 1/2$. Die Abbildungen II.1–II.3 zeigen für eine Simulation von 10 000 Münzwürfen die Entwicklung der relativen Häufigkeit eines Gewinns. Die Abbildung II.4 zeigt eine Entwicklung des akkumulierten Gewinns über 1000 Runden. Die Abbildung II.5 zeigt die Simulation von vier Spielverläufen im Fall $a = 20$ und $b = 60$. Für $a = 20$ und $b = 60$ sowie für $a = b = 100$ zeigen die Abbildungen II.6 und II.7 für eine Simulation von jeweils 10 000 Spielverläufen die Entwicklung des relativen Anteils der Spiele, die im Ruin enden.

Vermutung: für die Ruinwahrscheinlichkeiten $r(a,b,p)$ gilt

$$r(a,b,1/2) = \frac{b}{a+b}.$$

Beweis: $\boxed{\text{Übung}}$.

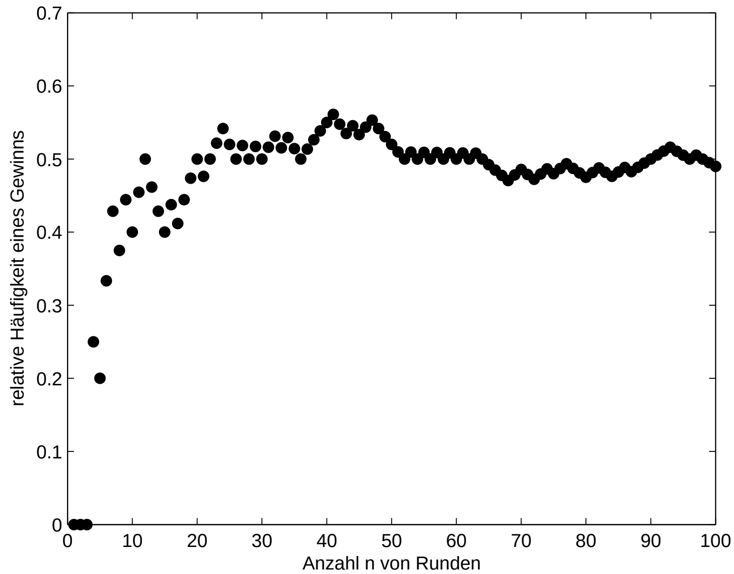

Abbildung II.1: Münzwurf, relative Häufigkeit eines Gewinns, $n \leq 100$

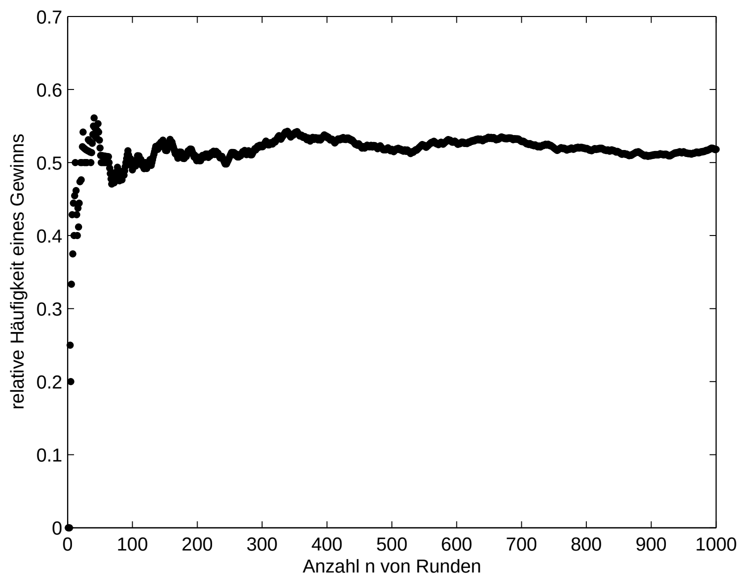

Abbildung II.2: Münzwurf, relative Häufigkeit eines Gewinns, $n \leq 1000$

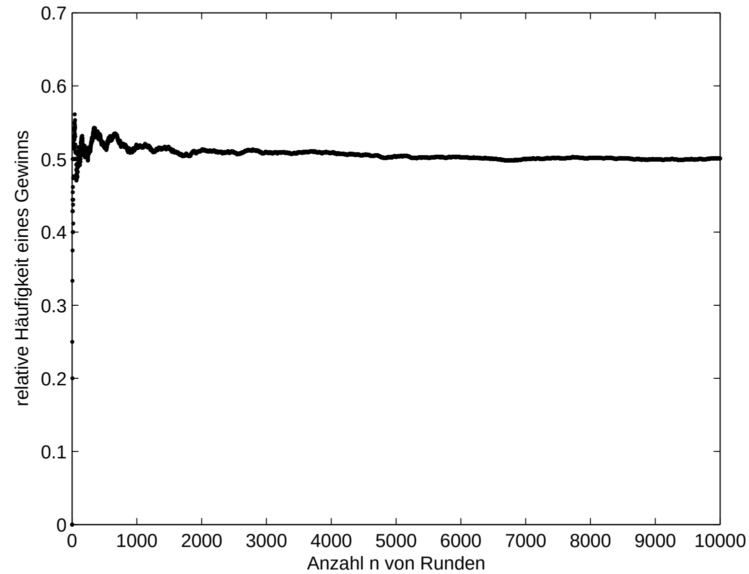

Abbildung II.3: Münzwurf, relative Häufigkeit eines Gewinns, $n \leq 10\,000$

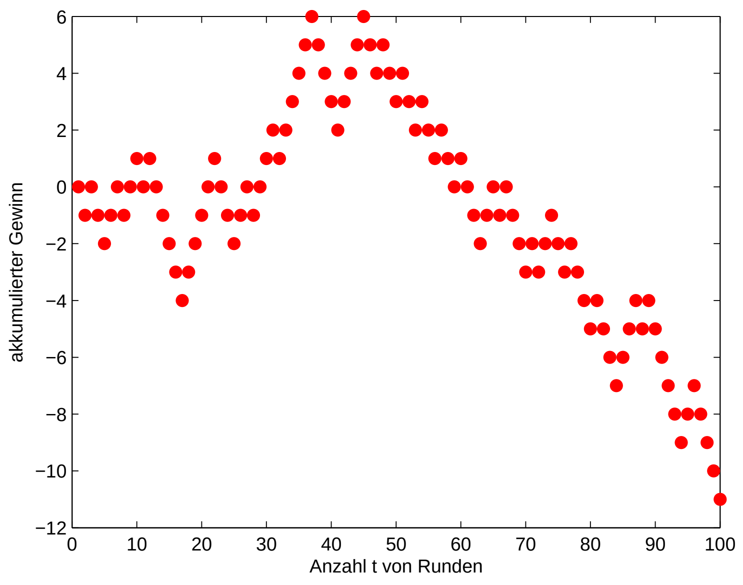

Abbildung II.4: Ruinproblem, akkumulierter Gewinn, $t \leq 100$

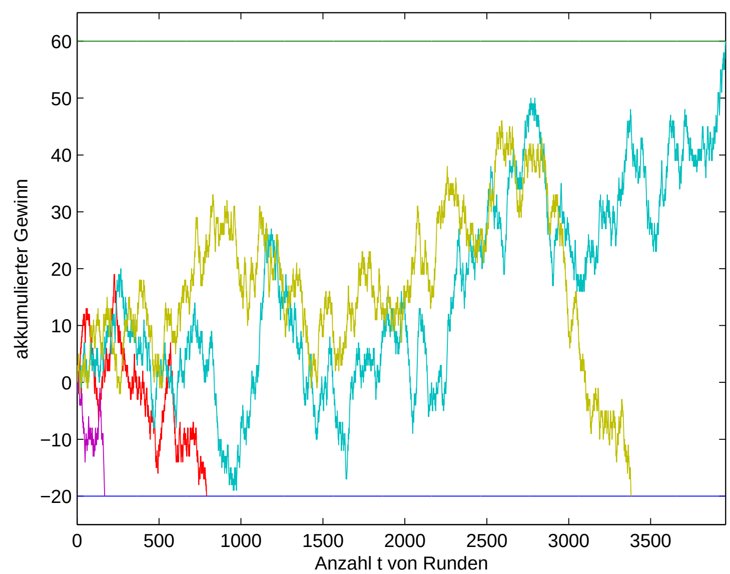

Abbildung II.5: Ruinproblem, $a = 20$, $b = 60$, vier Spielverläufe

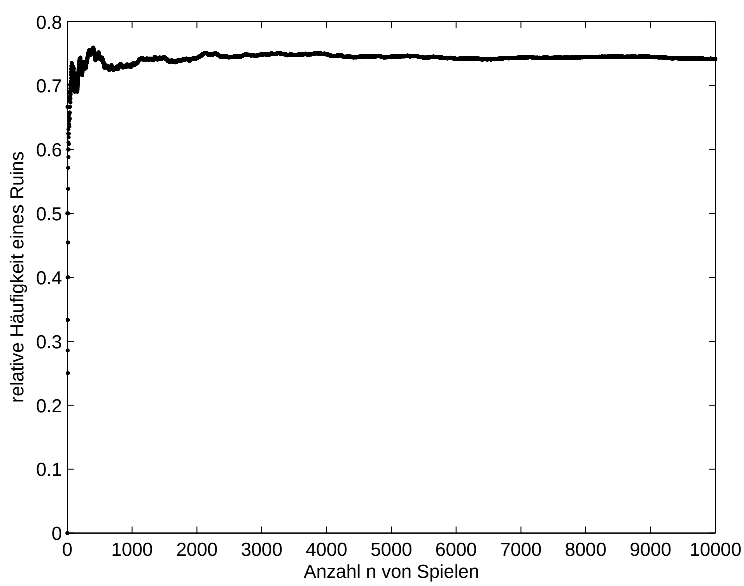

Abbildung II.6: Relative Häufigkeit eines Ruins, $a = 20$, $b = 60$, $n \leq 10\,000$

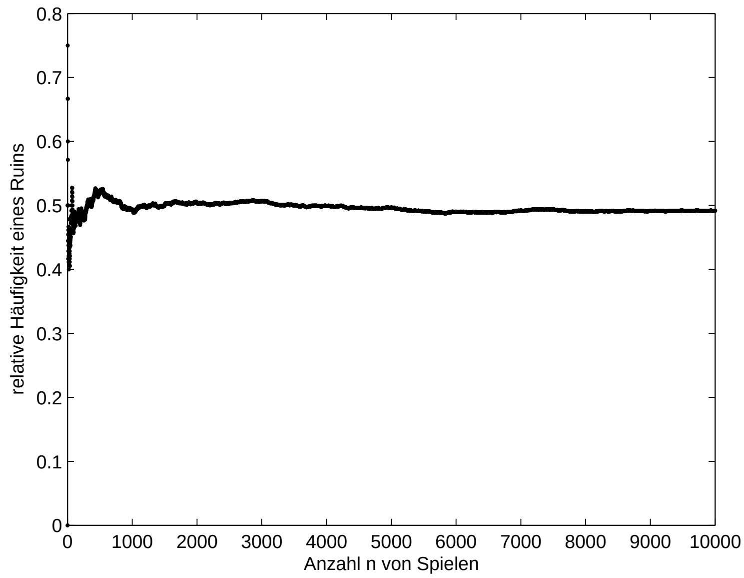

Abbildung II.7: Relative Häufigkeit eines Ruins, $a = b = 40$, $n \leq 10\,000$

3 Die Inversionsmethode

Frage: Simulation „komplizierter" Verteilungen. Eine „universelle" Möglichkeit bietet die Inversionsmethode.

Satz 1. Für jede Verteilungsfunktion F_X gilt

(i) F_X ist monoton wachsend,

(ii) F_X ist rechtsseitig stetig,

(iii) $\lim_{x\to\infty} F_X(x) = 1$ und $\lim_{x\to-\infty} F_X(x) = 0$.

Ferner gilt für alle $x \in \mathbb{R}$

$$P(\{X = x\}) = F_X(x) - \lim_{y\to x-} F_X(y)$$

und

$$P(\{X = x\}) = 0 \quad \Leftrightarrow \quad F_X \text{ stetig in } x.$$

Beweis. $\boxed{\text{ÜBUNG}}$ □

Definition 2. Für $q \in \,]0, 1[$ heißt

$$\inf\{v \in \mathbb{R} : F_X(v) \geq q\}$$

das q-*Quantil* der Verteilungsfunktion F_X. Speziell für $q = 1/2$: z *Median*.

Lemma 3. z genau dann q-Quantil von F_X, wenn

$$F_X(z) \geq q \qquad \text{und} \qquad \forall y \in \mathbb{R}: \; y < z \Rightarrow F_X(y) < q.$$

Beweis. Da F_X rechtsseitig stetig und monoton wachsend, existiert ein $\widetilde{z} \in \mathbb{R}$ mit

$$\{v \in \mathbb{R} : F_X(v) \geq q\} = [\widetilde{z}, \infty[\,.$$

□

Gegeben sei eine Funktion $F : \mathbb{R} \to [0,1]$ mit Eigenschaften (i)–(iii) aus Satz 1 sowie eine Zufallsvariable $U \sim \mathbf{U}([0,1])$.

Definiere $T_F : \mathbb{R} \to \mathbb{R}$ durch

$$T_F(u) := \begin{cases} \inf\{v \in \mathbb{R} : F(v) \geq u\}, & \text{falls } u \in\,]0,1[, \\ 0, & \text{sonst.} \end{cases}$$

Satz 4. $X := T_F \circ U$ ist eine Zufallsvariable mit $F_X = F$.

Beweis. Für $x \in \mathbb{R}$ und $u \in\,]0,1[$ gilt

$$T_F(u) \leq x \Leftrightarrow \inf\{v \in \mathbb{R} : F(v) \geq u\} \leq x \Leftrightarrow F(x) \geq u.$$

Also

$$\{X \leq x\} = \{U \in\,]0,1[\} \cap \{F(x) \geq U\} \cup \{U \notin\,]0,1[\} \cap \underbrace{\{0 \leq x\}}_{\in\{\emptyset,\Omega\}},$$

so daß insbesondere $\{X \leq x\} \in \mathfrak{A}$. Weiter

$$P(\{X \leq x\}) = P(\{F(x) \geq U\}) = F(x).$$

□

Bemerkung 5. Achtung: die Existenz von Zufallsvariablen $U \sim \mathbf{U}([0,1])$ auf geeigneten Wahrscheinlichkeitsräumen bleibt an dieser Stelle offen und wird erst in Abschnitt IV.2 positiv geklärt.

Korollar 6. $F : \mathbb{R} \to [0,1]$ ist genau dann die Verteilungsfunktion einer Zufallsvariable, wenn F die Eigenschaften (i)–(iii) erfüllt.

Beweis. Wende die Sätze 1, 4 und IV.2.1 an. □

Also: Modellierung eines beliebigen Zufallsexperimentes mit reellwertigen Ergebnissen durch Vorgabe der Verteilungsfunktion F.

Inversionsmethode. Simulation einer Zufallsvariable X mit der Verteilungsfunktion $F := F_X$ mittels Transformation

$$T_F(u_1), T_F(u_2), \ldots$$

gleichverteilter Zufallszahlen $u_1, u_2, \ldots$ in $[0,1]$.

Beispiel 7. Das Vorgehen in Beispiel 2.3 entspricht der Inversionsmethode in einer elementaren Situation.

Definitionen und Bezeichnungen