

2. Berechnung von Zufallszahlen

Definition (Stichprobe einer Verteilung)

Eine Folge von Zahlen heißt *Stichprobe (sample) von einer Verteilungsfunktion F* , wenn die Zahlen unabhängige Realisierungen einer Zufallsvariablen mit diesem F sind.

Beispiele

Wenn F die Gleichverteilung auf dem Intervall $[0, 1]$ oder $[0, 1)$ ist, so nennen wir die Stichprobe von F *gleichverteilt*. Bezeichnung $\sim \mathcal{U}[0, 1]$.

Wenn F die Standardnormalverteilung ist, so nennen wir die Stichprobe von F *standard-normalverteilt*. Bezeichnung $\sim \mathcal{N}(0, 1)$.

Die Grundlage wird die Berechnung von Zahlen $\sim \mathcal{U}[0, 1]$ sein. Darauf wird die Berechnung von Zahlen $\sim \mathcal{N}(0, 1)$ aufbauen, und hierauf weitere, z.B. $\sim \mathcal{N}(\mu, \Sigma)$.

2.1 Berechnung gleichverteilter Zufallszahlen

A. Lineare Kongruenz-Generatoren

Wähle $a, b, M \in \mathbb{N}$, $a \neq 0$, $a, b < M$, und definiere für ein $N_0 \in \mathbb{N}$ (‘‘seed’’) eine Folge von Zahlen durch

Algorithmus (Linearer Kongruenz-Generator)

Wähle N_0 .
 Für $i = 1, 2, \dots$ berechne
 $N_i = (aN_{i-1} + b) \bmod M$

Definiere nun $U_i \in [0, 1)$ durch

$$U_i = \frac{N_i}{M}.$$

Diese Zahlen werden als gleichverteilte Zufallszahlen verwendet.

Erste Eigenschaften

- (a) $N_i \in \{0, 1, \dots, M - 1\}$
- (b) Die Folge der N_i ist periodisch mit einer Periode $p \leq M$.
 (denn: Es gibt keine $M + 1$ verschiedenen Zahlen N_i . So müssen zwei Zahlen aus $\{N_0, \dots, N_M\}$ gleich sein, d.h. $N_i = N_{i+p}$ mit $p \leq M$. Es folgt p -Periodizität.)

Literatur: [Knuth: The Art of Computer Programming, Volume 2]

Obige Zahlen U_i sind keine wirklichen Zufallszahlen, sondern deterministisch bestimmte Zahlen. Wir nennen sie *Pseudo-Zufallszahlen*, lassen aber im Folgenden den Präfix “Pseudo” weg. Das Ziel ist, die M, a, b so zu wählen, dass die Zahlen ein brauchbarer Ersatz echter Zufallszahlen sind.

Beispiel

$$M = 244944, \quad a = 1597, \quad b = 51749$$

Weitere brauchbare Parameter a, b, M finden sich in [Press et al.: Numerical Recipes].

Frage: Was sind “gute” Zufallszahlen?

Praktische (und hypothetische) Antwort: Die Zahlen müssen möglichst “alle” Tests erfüllen.

1. Ziel: Verlange eine **große Periode** p , also M so groß wie möglich. Z.B. im Binärrechner $M \approx 2^l$, wobei l die Mantissenlänge ist. Geeignete a, b können mit zahlentheoretischen Methoden hergeleitet werden. [Knuth]

2. Ziel: Die Zahlen müssen wie gewünscht verteilt sein (Dichte f , Erwartungswert μ , Varianz σ^2). Hierzu: **Statistische Tests**. Berechne zunächst mit dem Algorithmus eine größere Anzahl von U_i -Werten. Dann z.B.

- (a) Berechne von dieser Stichprobe den Mittelwert $\hat{\mu}$ und die Varianz $\hat{s}^2$. Es muss gelten $\hat{\mu} \approx \mu$ und $\hat{s}^2 \approx \sigma^2$.
- (b) Werte Korrelationen der U_i mit früheren U_{i-j} aus. Falls starke Korrelationen vorliegen (z.B. wenn es sehr wahrscheinlich ist, dass auf kleine U -Werte wiederum kleine Werte folgen), dann ist der Generator schlecht.
- (c) Schätzung der Dichtefunktion $\hat{f}$ der Stichprobe. Es soll gelten: $\hat{f} \approx f$. Ein einfacher Prototyp eines solchen Tests ist: Teile $[0, 1]$ in äquidistante Stücke

$$k\Delta U \leq U < (k+1)\Delta U,$$

wobei ΔU die Länge der Teilintervalle bezeichnet. Bei insgesamt j vielen Zahlen sei j_k die Anzahl der in das k -te Teilintervall fallenden Zahlen. Die Wahrscheinlichkeit, dass das k -te Teilintervall getroffen wird, ist $\frac{j_k}{j}$. Der Sollwert ist

$$\int_{k\Delta U}^{(k+1)\Delta U} f(x) dx \quad (f = 1 \text{ im Fall der Gleichverteilung}).$$

Eine Näherung des Integrals ist

$$\Delta U f(\bar{U}),$$

wobei $\bar{U} \in [k\Delta U, (k+1)\Delta U]$. Also sollte für einen guten Generator gelten

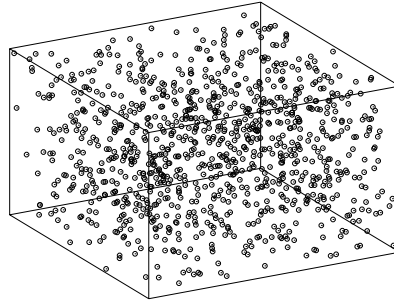
$$\Delta U \hat{f}(\bar{U}) = \frac{j_k}{j} \stackrel{!}{=} \Delta U f(\bar{U}),$$

jedenfalls für kleine ΔU . Man erhält die empirische Dichte

$$\hat{f} = \frac{j_k}{j\Delta U}.$$

3. Ziel: Man benötigt eine günstige **Gitterstruktur**. Bilde Vektoren als m -Tupel durch m aufeinanderfolgende Zahlen:

$$\frac{1}{M}(N_i, N_{i+1}, \dots, N_{i+m-1})$$



Wenn $U \sim \mathcal{U}[0, 1]$, dann sollen diese U -Vektoren den m -dimensionalen Einheits-Würfel möglichst gleichverteilt auffüllen. Die Folgen der Punkte/Vektoren liegen auf $(m - 1)$ -dimensionalen Hyperebenen. Trivialer Fall: Parallele Ebenen durch $U = \frac{i}{M}$, $i = 0, \dots, M - 1$. Schlecht wäre es, wenn alle Punkte auf nur relativ wenig Ebenen fallen. Dann würden in den “breiten” Bereichen zwischen solchen parallelen Ebenen *keine* Zufallszahlen liegen (schlechte Eigenschaft). Analysiere hierzu die “Gitterstruktur” (hier Struktur der Ebenen im $[0, 1]^m$, auf denen alle Punkte “landen”).

Analyse für $m = 2$

$$\begin{aligned} N_i &= (aN_{i-1} + b) \bmod M \\ &= aN_{i-1} + b - kM \quad \text{für } kM \leq aN_{i-1} + b < (k+1)M \end{aligned}$$

Nebenrechnung: Für beliebige Zahlen z_0, z_1

$$\begin{aligned} z_0 N_{i-1} + z_1 N_i &= z_0 N_{i-1} + z_1 (aN_{i-1} + b - kM) \\ &= N_{i-1} (z_0 + az_1) + z_1 b - z_1 kM \\ &= M \underbrace{\left(N_{i-1} \frac{z_0 + az_1}{M} - z_1 k \right)}_{=: c=c(i,k)} + z_1 b \end{aligned}$$

Nach Division durch M erhält man

$$z_0 U_{i-1} + z_1 U_i = c + z_1 b M^{-1}.$$

Dies ist eine Geraden-Gleichung in der (U_{i-1}, U_i) -Ebene. Also liegt hier für festes z_0, z_1 eine *Schar* von parallelen Geraden/Ebenen vor, parametrisiert durch c .

Frage: Gibt es eine Schar solcher Geraden/Ebenen, d.h. ein Tupel (z_0, z_1) , so dass nur *wenige* Geraden/Ebenen den Einheitswürfel schneiden? ($\rightarrow$ worst case).

Für $z_1, z_0 \in \mathbb{Z}$ und $z_0 + az_1 \bmod M = 0$ ist $c \in \mathbb{Z}$, und

$$c = z_0 U_{i-1} + z_1 U_i - z_1 b M^{-1}$$

($z_1 b M^{-1}$ ist hierbei eine konstante Parallelverschiebung.) Wie viele c 's gibt es? Für $0 \leq U < 1$ erhält man eine Abschätzung für die c 's durch ein maximales Zahlenintervall $I_c \subset \mathbb{Z}$, so dass gilt

$$c \in I_c \quad \Rightarrow \quad \text{Gerade berührt oder schneidet den Einheitswürfel.}$$

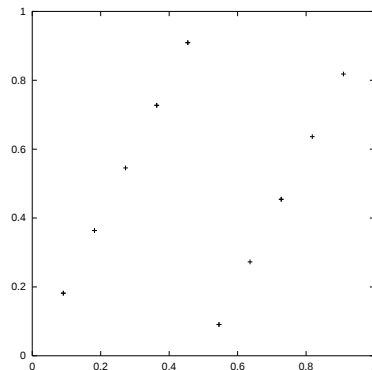
Die Mächtigkeit von I_c ist ein Maß dafür, wie eng der Abstand zwischen den parallelen Ebenen ist. Schlecht ist es, wenn die Mächtigkeit gering ist.

Akademisches Beispiel $N_i = 2N_{i-1} \bmod 11$ (also $a = 2$, $b = 0$, $M = 11$)

Das Tupel $(z_0, z_1) = (-2, 1)$ löst $z_0 + az_1 = 0 \bmod M$

$$\Rightarrow -2U_{i-1} + U_i = c$$

Aus $0 \leq U < 1$ folgt $-2 < c < 1$, außerdem gilt $c \in \mathbb{Z}$. Es bleiben nur $c = -1$ und $c = 0$.
D.h. bei dieser Wahl des Tupels (z_0, z_1) fallen alle Punkte auf nur 2 Geraden.



Figur: U_i über U_{i-1}

Beispiel $N_i = (1229N_{i-1} + 1) \bmod 2048$

Die Bedingung $z_0 + az_1 = 0 \bmod M$

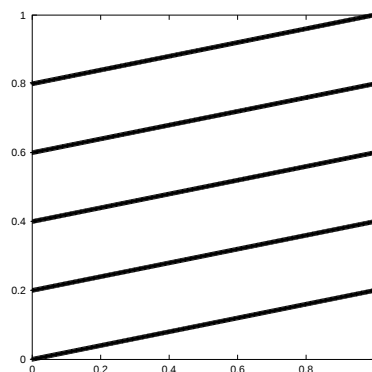
$$\frac{z_0 + 1229z_1}{2048} \in \mathbb{Z}$$

ist erfüllt durch $z_0 = -1$, $z_1 = 5$, denn

$$-1 + 1229 \cdot 5 = 6144 = 3 \cdot 2048.$$

Wegen $c = -U_{i-1} + 5U_i - \frac{5}{2048}$ gilt $-1 - \frac{5}{2048} < c < 5 - \frac{5}{2048}$. Alle Punkte fallen auf 6 Geraden, da $c \in \{-1, 0, 1, 2, 3, 4\}$.

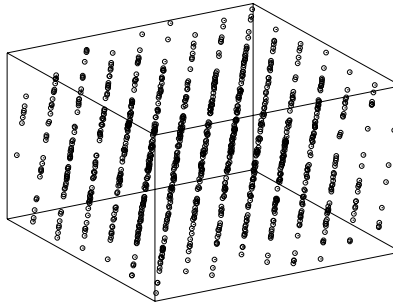
Der vertikale Abstand von zwei benachbarten Geraden ist $\frac{1}{z_1} = \frac{1}{5}$.



Beispiel (RANDU)

$$N_i = aN_{i-1} \bmod M, \text{ mit } a = 2^{16} + 3, M = 2^{31}$$

Die Zufallspunkte im Würfel $[0, 1)^3$ liegen auf nur 15 Ebenen.



Analyse für große m analog.

B. Fibonacci-Generatoren

Es gibt auch andere Klassen von Zufallszahlen-Generatoren, z.B. diejenige der **Fibonacci-Generatoren**. Ein Typ solcher Generatoren ist definiert durch

$$N_{i+1} := N_{i-\nu} - N_{i-\mu} \mod M$$

für geeignete μ, ν (evtl. auch mit “+” oder mit mehr Termen). Literatur: [Knuth]

Beispiel

$$U_i := U_{i-17} - U_{i-5},$$

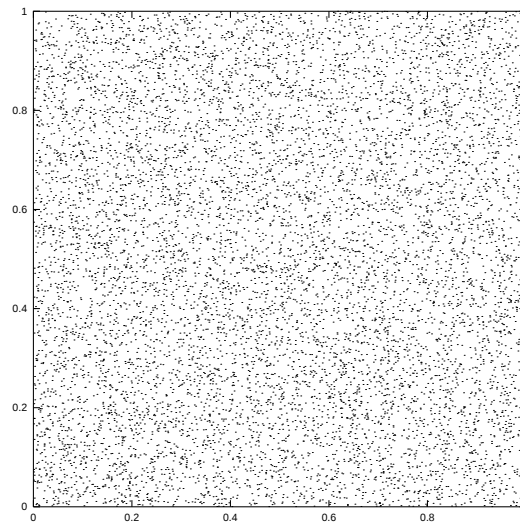
$$\text{falls } U_i < 0 \text{ setze } U_i := U_i + 1.0$$

(einfaches Beispiel mit relativ guten Eigenschaften, jedoch gibt es Korrelationen.)

Algorithmus (einfacher Fibonacci-Generator)

Repeat: $\zeta := U_i - U_j$
 falls $\zeta < 0$, setze $\zeta := \zeta + 1$
 $U_i := \zeta$
 $i := i - 1$
 $j := j - 1$
 falls $i = 0$, setze $i := 17$
 falls $j = 0$, setze $j := 17$

Initialisierung: Setze $i = 17$, $j = 5$, und berechne $U_1, \dots, U_{17}$ mit einem Kongruenz-Generator mit beispielsweise $M = 714025$, $a = 1366$, $b = 150889$.



10000 Zufallszahlen (U_{i-1}, U_i) , mit Fibonacci-Generator erstellt

Professioneller Generator zur Erzeugung von gleichverteilten Zufallszahlen: “Mersenne Twister” von Matsumoto, Nishimura, in: ACM Transactions on Modelling and Computer Simulations **8** (1998), S.3-30.

2.2 Zufallszahlen mit anderen Verteilungen

Die Erzeugung von normalverteilten Zufallszahlen baut auf gleichverteilten Zufallszahlen auf. Zur Erzeugung von Zufallszahlen mit vorgegebenen Verteilungen gibt es Inversions- und Transformations-Methoden.

A. Inversion

Es sei $F(x) := P(X \leq x)$ Verteilungsfunktion, wobei X eine Zufallsvariable und P die mit F verbundene Wahrscheinlichkeit ist.

Satz (Inversion)

Angenommen $U \sim \mathcal{U}[0, 1]$ und F sei stetige und streng monoton wachsende Verteilungsfunktion, dann ist $X := F^{-1}(U)$ nach F verteilt.

Beweis:

$U \sim \mathcal{U}[0, 1]$ heißt $P(U \leq \xi) = \xi$ für $0 \leq \xi \leq 1$. Also

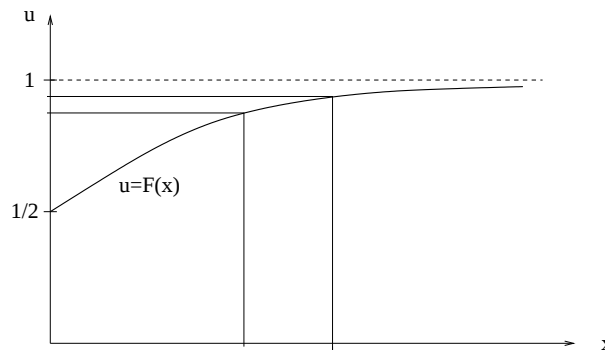
$$P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x).$$

Anwendung

Berechne $U \sim \mathcal{U}[0, 1]$. Werte $F^{-1}(U)$ aus, diese Zahlen haben die gewünschte Verteilung. Dies funktioniert nur numerisch, weil F^{-1} i.A. nicht analytisch bekannt ist.

Zwei Varianten des numerischen Zugangs:

- (a) $F(x) = u$ ist eine nichtlineare Gleichung für x , welche mit den Standardmethoden der Numerik (z.B. Newton-Verfahren) iterativ gelöst werden kann. Schwierigkeit hier: Wegen der großen Auswirkung von Störungen in u auf das zugehörige x in den Bereichen $u \approx 0$, $u \approx 1$ sind die Abbruch-Kriterien für ein Newton-Verfahren knifflig. (Figur: Normalverteilung)



- (b) Konstruiere eine “einfache” Näherung G , so dass $G(u) \approx F^{-1}(u)$, und werte dann $x = G(u)$ aus. Man kann G so konstruieren, dass das asymptotische Verhalten, d.h. die Pole von G , gut wiedergegeben werden. Bei Ausnutzung der Symmetrie der Normalverteilung (zu $(x, u) = (0, \frac{1}{2})$) braucht man nur das Polverhalten bei $u = 1$ zu berücksichtigen. Hierzu Ansatz einer rationalen Funktion für $G(u)$, d.h. das Nennerpolynom hat $u = 1$ als Nullstelle.

B. Transformation

Zunächst der skalare Fall: X sei eine Zufallsvariable. Wie ist ein transformiertes $h(X)$ verteilt?

Satz (skalare Transformation)

Es sei X eine Zufallsvariable mit Dichtefunktion f und Verteilungsfunktion F . Weiter sei

$$h : S \rightarrow B$$

mit $S, B \subseteq \mathbb{R}$, wobei S der Träger von f ist, zudem sei h streng monoton und differenzierbar.

- (a) $Y := h(X)$ ist Zufallsvariable. Die Verteilungsfunktion ist

$$\begin{aligned} F(h^{-1}(y)) & \quad \text{für monoton wachsendes } h \\ 1 - F(h^{-1}(y)) & \quad \text{für monoton fallendes } h \end{aligned}$$

- (b) Falls h^{-1} absolut stetig ist, dann ist für fast alle y

$$f(h^{-1}(y)) \left| \frac{dh^{-1}(y)}{dy} \right|$$

die Dichte von $h(X)$.

Beweis: (schreibe F^X für F)

- (a) $F^Y(y) := P(h(X) \leq y) =$

falls h monoton wachsend:

$$= P(X \leq h^{-1}(y)) = F^X(h^{-1}(y))$$

falls h monoton fallend:

$$= P(X \geq h^{-1}(y)) = 1 - P(X < h^{-1}(y)) = 1 - F^X(h^{-1}(y))$$

- (b) h^{-1} absolut stetig $\Rightarrow$ Dichte von $Y = h(X)$ ist gleich der Ableitung der Verteilungsfunktion (fast überall). Die Auswertung von $\frac{dF(h^{-1}(y))}{dy}$ liefert mit der Kettenregel das gewünschte Ergebnis.

Anwendung

Beginne mit $X \sim \mathcal{U}[0, 1]$ und f als Dichte der Gleichverteilung, also

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

d.h. $S = [0, 1]$. Zu berechnen sind Zufallszahlen Y mit einer vorgegebenen Dichte $g(y)$. Suche also eine Transformation h , so dass

$$f(h^{-1}(y)) \left| \frac{dh^{-1}(y)}{dy} \right| = 1 \left| \frac{dh^{-1}(y)}{dy} \right| \stackrel{!}{=} g(y).$$

Dann ist $h(X)$ wie gewünscht verteilt.

Beispiel (Exponentialverteilung)

Die Exponentialverteilung mit Parameter $\lambda > 0$ hat die Dichte

$$g(y) = \begin{cases} \lambda e^{-\lambda y} & \text{für } y \geq 0 \\ 0 & \text{für } y < 0. \end{cases}$$

d.h. B sind die nichtnegativen reellen Zahlen. Transformation von $[0, 1] \rightarrow B$ mit der monoton fallenden Funktion

$$y = h(x) := -\frac{1}{\lambda} \log x$$

mit der Umkehrfunktion $h^{-1}(y) = e^{-\lambda y}$ für $y \geq 0$. Für dieses h gilt

$$f(h^{-1}(y)) \left| \frac{dh^{-1}(y)}{dy} \right| = 1 \cdot |(-\lambda)e^{-\lambda y}| = \lambda e^{-\lambda y} = g(y).$$

Also ist $h(X)$ exponentialverteilt.

Anwendung

Berechne gleichverteilte $U_1, U_2, \dots$

$$\Rightarrow -\frac{1}{\lambda} \log(U_1), \quad -\frac{1}{\lambda} \log(U_2), \quad \dots \text{ sind exponentialverteilt.}$$

(Die Abstände der Sprungzeiten bei Poisson-Prozessen sind exponentialverteilt.)

Versuch mit Normalverteilung: Suche h so, dass

$$1 \cdot \left| \frac{dh^{-1}(y)}{dy} \right| = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y^2\right).$$

Dies ist eine Differentialgleichung für h^{-1} ohne analytische Lösungsmöglichkeit. Hier hilft die mehrdimensionale Version des Transformationssatzes.

Satz (Transformationssatz im $\mathbb{R}^n$)

Es sei X eine Zufallsvariable im $\mathbb{R}^n$ mit Dichte $f(x) > 0$ auf dem Träger S . Die Transformation $h : S \rightarrow B$, $S, B \subseteq \mathbb{R}^n$ sei umkehrbar und die Umkehrfunktion stetig differenzierbar auf B . Dann hat $Y := h(X)$ die Dichte

$$f(h^{-1}(y)) \left| \frac{\partial(x_1, \dots, x_n)}{\partial(y_1, \dots, y_n)} \right|, \quad y \in B, \quad (2.7)$$

wobei $\frac{\partial(x_1, \dots, x_n)}{\partial(y_1, \dots, y_n)}$ die Determinante der Jacobimatrix von $h^{-1}(y)$ ist.

Beweis: siehe Theorem 4.2 in [Devroye: Non-Uniform Random Variate Generation (1986)]

2.3 Berechnung normalverteilter Zufallszahlen

Anwendung des Transformationssatzes im $\mathbb{R}^2$.

A. Methode von Box und Muller

$S := [0, 1]^2$, Dichte $f = 1$ auf S . Transformation h :

$$\begin{cases} y_1 = \sqrt{-2 \log x_1} \cos 2\pi x_2 =: h_1(x_1, x_2) \\ y_2 = \sqrt{-2 \log x_1} \sin 2\pi x_2 =: h_2(x_1, x_2) \end{cases}$$

Umkehrabbildung h^{-1} :

$$\begin{cases} x_1 = \exp \left\{ -\frac{1}{2}(y_1^2 + y_2^2) \right\} \\ x_2 = \frac{1}{2\pi} \arctan \frac{y_2}{y_1} \end{cases}$$

Für diese Transformation gilt

$$\begin{aligned} \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} &= \det \begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{pmatrix} = \\ &= -\frac{1}{2\pi} \exp \left\{ -\frac{1}{2}(y_1^2 + y_2^2) \right\}. \end{aligned}$$

Absolut genommen ist dies die Dichte einer zweidimensionalen Normalverteilung. Wegen

$$\left| \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} \right| = \left[\frac{1}{\sqrt{2\pi}} \exp \left(-\frac{1}{2}y_1^2 \right) \right] \cdot \left[\frac{1}{\sqrt{2\pi}} \exp \left(-\frac{1}{2}y_2^2 \right) \right]$$

ist die zweidimensionale Dichte das Produkt der eindimensionalen Dichten der Standardnormalverteilung. Folgerung: Die beiden Komponenten y_1, y_2 des Vektors Y sind unabhängig. [z.B. Hesse: Angewandte Wahrscheinlichkeitstheorie, S. 68]

Anwendung

Wenn die beiden Komponenten $x_1, x_2 \sim \mathcal{U}[0, 1]$ dann besteht das Resultat aus zwei unabhängige Zufallszahlen $y_1, y_2 \sim \mathcal{N}(0, 1)$.

Algorithmus (Box-Muller)

- (1) generiere $U_1 \sim \mathcal{U}[0, 1]$ und $U_2 \sim \mathcal{U}[0, 1]$.
- (2) $\theta := 2\pi U_2$, $\rho := \sqrt{-2 \log U_1}$
- (3) $Z_1 := \rho \cos \theta$ ist $\sim \mathcal{N}(0, 1)$
(genauso wie $Z_2 := \rho \sin \theta$).

B. Variante von Marsaglia

Präpariere die Eingangsdaten für Box & Muller derart, dass die trigonometrischen Funktionen entfallen. Mit $U \sim \mathcal{U}[0, 1]$ erhält man $V := 2U - 1 \sim \mathcal{U}[-1, 1]$. Zwei solcher Werte V_1, V_2 definieren einen Punkt im $\mathbb{R}^2$. Definiere

$$\mathcal{D} := \{(V_1, V_2) : V_1^2 + V_2^2 < 1\}.$$

Akzeptiere nur solche Tupel (U_1, U_2) , so dass $(V_1, V_2) \in \mathcal{D}$. Diese verbleibenden Punkte sind auf $\mathcal{D}$ gleichverteilt. Transformation auf $(\text{Radius})^2$ und normierten Winkel:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} V_1^2 + V_2^2 \\ \frac{1}{2\pi} \arg(V_1, V_2) \end{pmatrix}.$$

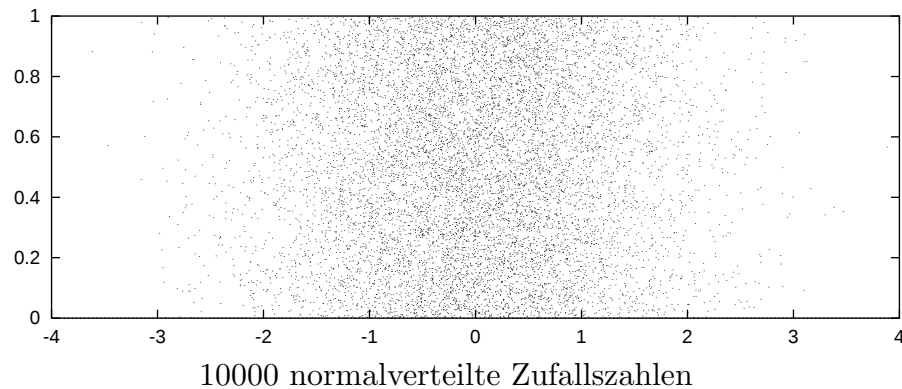
Diese (x_1, x_2) sind in S gleichverteilt ($\longrightarrow$ Übung) und werden als Eingabe für Box-Muller verwendet.

$$\begin{aligned} \cos(2\pi x_2) &= \frac{V_1}{\sqrt{V_1^2 + V_2^2}} \\ \sin(2\pi x_2) &= \frac{V_2}{\sqrt{V_1^2 + V_2^2}} \end{aligned}$$

Algorithmus (Polar-Methode)

- (1) *Repeat*: generiere $U_1, U_2 \sim \mathcal{U}[0, 1]$; $V_i := 2U_i - 1$, ($i = 1, 2$)
bis $W := V_1^2 + V_2^2 < 1$.
- (2) $Z_1 := V_1 \sqrt{-2 \log(W)/W}$ ist $\sim \mathcal{N}(0, 1)$
(ebenso wie $Z_2 := V_2 \sqrt{-2 \log(W)/W}$).

Die Wahrscheinlichkeit, dass $W < 1$, ist das Verhältnis der Flächen, also $\frac{\pi}{4} \approx 0.785\dots$. D.h. in 21% aller Ziehungen von (U_1, U_2) gilt $W \geq 1$. Aber die Einsparung der trigonometrischen Funktionen überwiegt den Nachteil, und Marsaglias Polar Methode ist effizienter als Box-Muller.



2.4 Berechnung korrelierter Zufallsvariablen

Das Ziel sei die Berechnung einer Zufallsvariablen $X = (X_1, \dots, X_n)$ mit vorgegebenen

$$\mu = EX = (EX_1, \dots, EX_n),$$

Kovarianzmatrix

$$\Sigma_{ij} = (\text{Cov}X)_{ij} := E((X_i - \mu_i)(X_j - \mu_j)); \quad \sigma_i^2 = \Sigma_{ii}$$

und den Korrelationen

$$\rho_{ij} := \frac{\Sigma_{ij}}{\sigma_i \sigma_j}.$$

Es gilt die lineare Transformations-Eigenschaft:

$$Z \sim \mathcal{N}(0, I) \Rightarrow AZ \sim \mathcal{N}(0, AA^T)$$

Zur Begründung:

Die Dichtefunktion $f(x_1, \dots, x_n)$ von $\mathcal{N}(\mu, \Sigma)$ ist

$$f(x) = \frac{1}{(2\pi)^{n/2}} \frac{1}{(\det \Sigma)^{1/2}} \exp \left\{ -\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right\}.$$

Angenommen $Z \sim \mathcal{N}(0, I)$ und $x = Az$, $A \in \mathbb{R}^{n \times n}$, wobei z eine Realisierung von Z und I die Einheitsmatrix ist.

$$\exp \left\{ -\frac{1}{2} z^T z \right\} = \exp \left\{ -\frac{1}{2} (A^{-1}x)^T (A^{-1}x) \right\} = \exp \left\{ -\frac{1}{2} x^T A^{-T} A^{-1} x \right\}$$

und $dx = |\det A| dz$ implizieren

$$\frac{1}{|\det A|} \exp \left\{ -\frac{1}{2} x^T (AA^T)^{-1} x \right\} dx = \exp \left\{ -\frac{1}{2} z^T z \right\} dz$$

für beliebige nichtsinguläre Matrizen A . Falls AA^T eine Faktorisierung von Σ ist, also $\Sigma = AA^T$ und folglich $|\det A| = (\det \Sigma)^{1/2}$, dann gilt:

$$AZ \sim \mathcal{N}(0, \Sigma).$$

Folgerung:

$$\mu + AZ \sim \mathcal{N}(\mu, \Sigma).$$

Beispiel: Wähle die Cholesky-Zerlegung von Σ .

Algorithmus (Korrelierte Zufallsvariablen)

- (1) Berechne zu Σ die Cholesky-Zerlegung $AA^T = \Sigma$
- (2) Berechne $Z \sim \mathcal{N}(0, I)$ komponentenweise
mit $Z_i \sim \mathcal{N}(0, 1)$, $i = 1, \dots, n$, z.B. mit
Marsaglia's Polar-Methode
- (3) $\mu + AZ$ hat die gewünschte Verteilung $\sim \mathcal{N}(\mu, \Sigma)$

Beispiel: Wenn $\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$ gefordert ist, dann wird dies durch

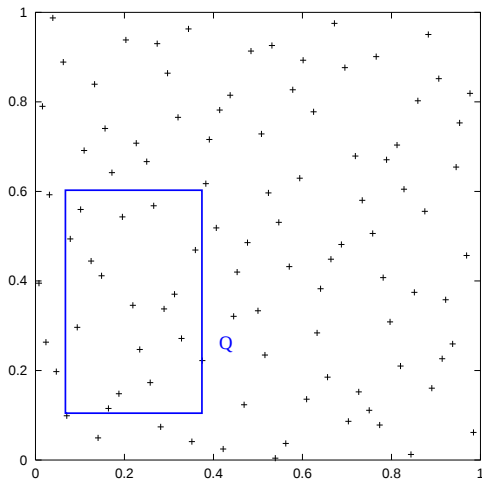
$$\begin{pmatrix} \sigma_1 Z_1 \\ \sigma_2 \rho Z_1 + \sigma_2 \sqrt{1 - \rho^2} Z_2 \end{pmatrix} \text{ erfüllt.}$$

2.5 Zahlenfolgen niedriger Diskrepanz

Ziel: Konstruktion von Punkten derart, dass die Verteilung “ähnlich” ist wie bei Zufallszahlen, deren Nachteile (Auftreten größerer Lücken oder Klumpen) aber vermieden werden.

Idee: Es sei Q ein Quader in $[0, 1]^m$. Ziel einer “gleichmäßigen” Verteilung ist

$$\frac{\# \text{ der } x_i \in Q}{\# \text{ aller Punkte in } [0, 1]^m} \approx \frac{\text{vol}(Q)}{\text{vol}([0, 1]^m)}$$



illustriert im $\mathbb{R}^2$, konkret $[0, 1]^2$

Definition (Diskrepanz)

Die Diskrepanz einer Punktmenge $\{x_1, \dots, x_N\}$ mit $x_i \in [0, 1]^m$ ist

$$D_N := \sup_Q \left| \frac{\# \text{ der } x_i \in Q}{N} - \text{vol}(Q) \right|.$$

Definition (Folge niedriger Diskrepanz)

Eine Folge von Punkten $x_1, \dots, x_N \in [0, 1]^m$ heißt Folge niedriger Diskrepanz, wenn es eine Konstante C_m gibt, so dass für alle N gilt:

$$D_N \leq C_m \frac{(\log N)^m}{N}$$

Kommentar

Der Nenner $\frac{1}{N}$ beschreibt eine relativ schnelle Abnahme von D_N mit der Anzahl der Punkte N , verglichen mit Monte-Carlo-Verfahren, bei denen der Fehler mit der Ordnung $\frac{1}{\sqrt{N}}$ abnimmt.

Allerdings: bei der Abnahme gemäß “niedriger Diskrepanz” haben wir noch die Zunahme des Zählers $(\log N)^m$. $\log N$ wächst aber nur sehr bescheiden, d.h. für niedrige Dimensionen m ist die Abnahme von D_N deutlich schneller als die Abnahme des Fehlers von Monte-Carlo-Verfahren (vgl. Tabelle).

Frage: Gibt es Zahlenfolgen/Punktfolgen niedriger Diskrepanz?

Beispiel: ($m = 1$) **Van der Corput-Folge**

$$\frac{1}{2}, \frac{1}{4}, \frac{3}{4}, \frac{1}{8}, \frac{5}{8}, \frac{3}{8}, \frac{7}{8}, \frac{1}{16}, \dots$$

N	$\frac{1}{\sqrt{N}}$	$\sqrt{\frac{\log \log N}{N}}$	$\frac{\log N}{N}$	$\frac{(\log N)^2}{N}$	$\frac{(\log N)^3}{N}$
10^1	.31622777	.28879620	.23025851	.53018981	1.22080716
10^2	.10000000	.12357911	.04605170	.21207592	.97664572
10^3	.03162278	.04396186	.00690776	.04771708	.32961793
10^4	.01000000	.01490076	.00092103	.00848304	.07813166
10^5	.00316228	.00494315	.00011513	.00132547	.01526009
10^6	.00100000	.00162043	.00001382	.00019087	.00263694
10^7	.00031623	.00052725	.00000161	.00002598	.00041874
10^8	.00010000	.00017069	.00000018	.00000339	.00006251
10^9	.00003162	.00005506	.00000002	.00000043	.00000890

Bildungsgesetz am Beispiel $x_6 = \frac{3}{8}$. Index 6 binär: 110.

radix-invertiert: .011 wird dezimal $\frac{3}{8}$

Definition (Radix-inverse Funktion)

Für $i = 1, 2, \dots$ sei

$$i = \sum_{k=0}^j d_k b^k$$

die Darstellung von i zur Basis b (ganzzahlig ≥ 2), mit $d_k \in \{0, 1, \dots, b-1\}$. Die *Radix-inverse* (*radical-inverse*) *Funktion* ist definiert durch

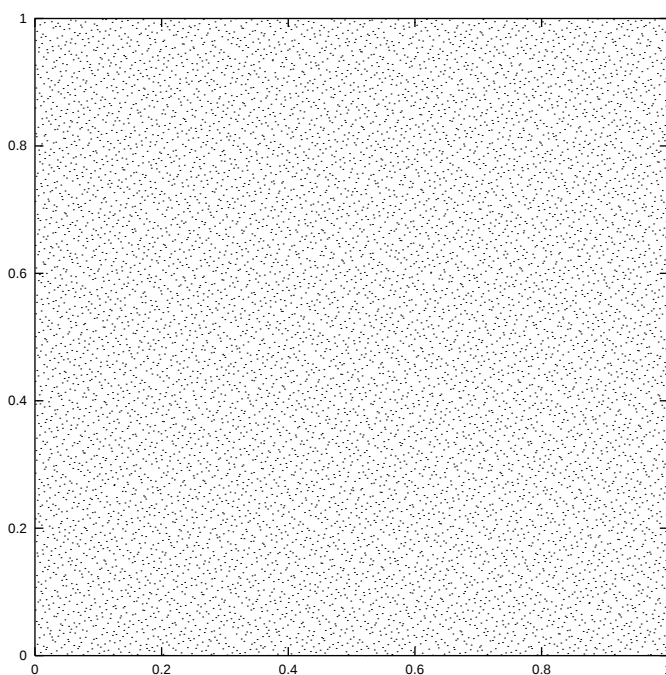
$$\phi_b(i) := \sum_{k=0}^j d_k b^{-k-1}.$$

ein-dimensionales Beispiel: $x_i := \phi_2(i)$ ist die Van der Corput-Folge.

Definition (Halton-Folge)

Es seien $p_1, \dots, p_m$ natürliche Zahlen und paarweise prim. Die *Halton-Folge* ist definiert als Folge der Vektoren

$$x_i := (\phi_{p_1}(i), \dots, \phi_{p_m}(i)), \quad i = 1, 2, \dots$$



Die Abbildung zeigt die ersten 10000 Halton-Punkte mit $m = 2$ und $p_1 = 2, p_2 = 3$

Weitere Folgen niedriger Diskrepanz:

Faure-Folge

Sobol-Folge

Niederreiter-Folge

Die deterministischen Zahlenfolgen niedriger Diskrepanz heißen “Quasi-Zufallszahlen”.