

Monte Carlo-Simulation

Monte Carlo-Methoden

Der Begriff “Monte Carlo-Methode” entstand in den 1940er Jahren, als man im Zusammenhang mit dem Bau der Atombombe die Simulation von Zufallsprozessen erstmals in größerem Stil einsetzte, um die Wechselwirkung von Neutronen mit Materie theoretisch vorherzusagen. Die Bezeichnung ist eine Anspielung auf den für Glücksspiele bekannten Ort, da die Grundlage des Verfahrens Zufallszahlen sind, wie man sie auch mit einem Roulette-Rad erzeugen könnte. Schon damals wurde eine ganze Reihe von grundlegenden Verfahren entwickelt, und heute zählen Monte Carlo (MC)-Methoden zu den wichtigsten numerischen (und auch nicht-numerischen) Verfahren, die sich auf viele naturwissenschaftliche, technische und medizinische Probleme mit großem Erfolg anwenden lassen. Dabei ist es gleichgültig, ob das Problem ursprünglich statistischer Natur war oder nicht, sondern man wendet die Bezeichnung auf alle Verfahren an, bei denen die Verwendung von Zufallszahlen eine entscheidende Rolle spielt. Ein Beispiel dafür—and eine der wichtigsten Anwendungen schlechthin—ist die MC-Integration, d.h. die numerische Berechnung hochdimensionaler Integrale, die ja mit Zufall überhaupt nichts zu tun haben müssen.

Wahrscheinlichkeitstheoretische Grundbegriffe

Zufallsvariablen

Ausgangspunkt für wahrscheinlichkeitstheoretische Betrachtungen sind die Begriffe “(zufälliges) Ereignis” und “Wahrscheinlichkeit”.

Unter einem *Elementarereignis* ω versteht man den möglichen Ausgang einer Messung, Beobachtung usw. (In den Naturwissenschaften verwenden wir dafür gern den Begriff “Experiment”.) Die Gesamtheit der Elementarereignisse bildet den *Ereignisraum* Ω . Teilmengen A von Ω , die i.a. mehr als ein Elementarereignis enthalten können, werden als *Ereignisse* bezeichnet. So entsprechen beim Würfeln mit einem Würfel die sechs möglichen Elementarereignisse dem Auftreten einer bestimmten Augenzahl, während z.B. das Ereignis “gerade Augenzahl” die Tatsache bezeichnet, 2, 4 oder 6 Augen geworfen zu haben. Spezielle Ereignisse sind die leere Menge (das unmögliche Ereignis) und der ganze Ereignisraum (das sichere Ereignis).

Entsprechend der relativen Häufigkeit, mit der sie bei einer großen Anzahl von Wiederholungen des Experiments auftreten, definiert man für Ereignisse *Wahrscheinlichkeiten* P mit den Eigenschaften

$$0 \leq P(A) \leq 1$$

$$P(\emptyset) = 0, \quad P(\Omega) = 1$$

$$\begin{aligned} P(A \cup B) &\leq P(A) + P(B) \\ &= P(A) + P(B) \quad \text{wenn } A \cap B = \emptyset \end{aligned}$$

Eine *Zufallsvariable* ist eine Abbildung

$$x : \Omega \rightarrow R$$

die jedem Elementarereignis einen reellen Zahlenwert $x(\omega)$ zuordnet, z.B. die Körpergröße einer zufällig herausgegriffenen Person. Damit ist auch die Wahrscheinlichkeit definiert, daß x einen bestimmten Wert X annimmt, nämlich mittels Rückführung auf die ursprünglichen Wahrscheinlichkeiten über dem Ereignisraum Ω

$$P(x = X) = P(\{\omega | x(\omega) = X\})$$

Kann die Zufallsvariable nur abzählbar viele diskrete Werte X_1, X_2, \dots annehmen, so definiert man als *Wahrscheinlichkeitsverteilung* von x den Satz von Zahlen (Wahrscheinlichkeiten)

$$p_i = P(x = X_i)$$

für die dann gilt

$$p_i \geq 0$$

$$\sum_i p_i = 1$$

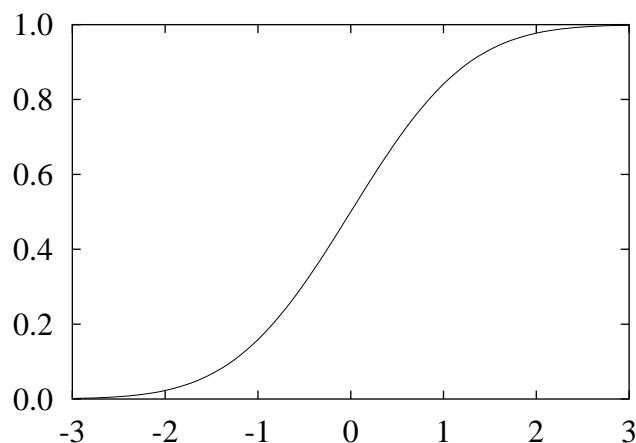
Ein anderer wichtiger Fall ist der, daß die Zufallsvariable kontinuierliche Werte in einem Intervall oder auf der ganzen reellen Achse annehmen kann. Die *kumulative Verteilungsfunktion* $F(X)$ ist dann die Wahrscheinlichkeit, daß x einen Wert kleiner oder gleich X annimmt

$$F(X) = P(x \leq X)$$

mit

$$F(-\infty) = 0 \quad \text{und} \quad F(+\infty) = 1$$

Wenn x nur Werte in einem endlichen Intervall annimmt, dann werden diese Grenzwerte natürlich schon am Rand des Intervalls erreicht. Dazwischen ist $F(X)$ eine monoton wachsende Funktion.



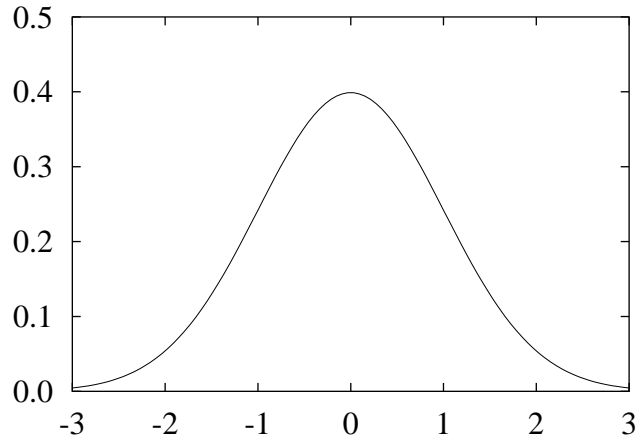
Wenn F differenzierbar ist, dann läßt sich diese Funktion als Integral

$$F(X) = P(x \leq X) = \int_{-\infty}^X dx f(x)$$

einer *Wahrscheinlichkeitsdichte* $f(x)$ darstellen, die

$$f(x) \geq 0$$
$$\int_{-\infty}^{\infty} dx f(x) = 1$$

erfüllt. Für kleine dx ist $f(x) dx$ näherungsweise gleich der Wahrscheinlichkeit, daß die Zufallsvariable einen Wert im Intervall der Größe dx um x annimmt.



Da sich die Behandlung von diskreten und kontinuierlichen Zufallsvariablen hauptsächlich dadurch unterscheidet, ob man über Wahrscheinlichkeiten summiert oder über Wahrscheinlichkeitsdichten integriert, werden im folgenden nur kontinuierliche Variable explizit betrachtet.

Einem Elementarereignis können auch zwei oder mehrere Zufallsvariablen, x, y, \dots , zugeordnet sein (z.B. Orts- und Geschwindigkeitskomponenten eines Gasmoleküls). Im Fall von kontinuierlichen Variablen wird dann das *gleichzeitige* Auftreten bestimmter Werte der einzelnen Variablen durch eine multivariate Wahrscheinlichkeitsdichte $f(x, y, \dots)$ beschrieben. Speziell ist bei zwei Variablen $f(x, y) dx dy$ wieder die Wahrscheinlichkeit, daß die erste Variable einen Wert im Intervall dx um x annimmt und gleichzeitig die zweite Variable einen Wert im Intervall dy um y .

Zwei (oder mehrere) Zufallsvariablen heißen *statistisch unabhängig*, wenn sich ihre gemeinsamen Wahrscheinlichkeiten bzw. Wahrscheinlichkeitsdichten nach dem Muster

$$f(x, y) = g(x) h(y)$$

faktorisieren lassen. Dabei sind $g(x)$ und $h(y)$ die Wahrscheinlichkeitsdichten für das Auftreten von x bzw. y allein (d.h. jeweils ohne Rücksicht auf die andere Variable).

Charakterisierung von Zufallsvariablen

Die Wahrscheinlichkeitsdichte (im diskreten Fall die Wahrscheinlichkeitsverteilung) enthält die gesamte Information über eine Zufallsvariable. Oft ist die Wahrscheinlichkeitsdichte aber modellmäßig nicht bekannt oder nicht im Detail meßbar, und man versucht daher, sie durch abgeleitete, einfacher bestimmbare Größen zu charakterisieren.

Der *Mittelwert* (oder *Erwartungswert*) einer Zufallsvariablen x mit Wahrscheinlichkeitsdichte $f(x)$ ist definiert als

$$\langle x \rangle = \int dx f(x) x$$

Das ist der Wert, der sich ergeben würde, wenn man x sehr oft mißt und die Summe der Meßwerte durch die Anzahl der Messungen dividiert. Für den Mittelwert verwendet man gern das Symbol μ ,

$$\mu = \langle x \rangle$$

Die *Varianz* ist ein Maß für die mögliche Abweichung (Streuung) einer Einzelbeobachtung vom Mittelwert

$$\text{Var}(x) = \sigma^2 = \langle (x - \langle x \rangle)^2 \rangle = \int dx f(x) (x - \langle x \rangle)^2$$

Durch Ausquadrieren des Integranden rechnet man leicht nach, daß auch gilt

$$\sigma^2 = \langle x^2 \rangle - \langle x \rangle^2$$

Der hier verwendete Erwartungswert von x^2 ist ein Beispiel für die *Momente* einer Wahrscheinlichkeitsdichte

$$M_n = \langle x^n \rangle = \int dx f(x) x^n$$

Allgemein ist der *Erwartungswert einer Funktion* $g(x)$ definiert als

$$\langle g(x) \rangle = \int dx f(x) g(x)$$

Die Korrelation zweier Variablen x und y wird durch die *Kovarianz*

$$\text{cov}(x, y) = \langle (x - \langle x \rangle) (y - \langle y \rangle) \rangle = \langle y \rangle - \langle x \rangle \langle y \rangle$$

charakterisiert. Wenn x und y statistisch unabhängig sind, gilt, wie man ebenfalls aus den Definitionen leicht nachrechnet,

$$\langle xy \rangle = \langle x \rangle \langle y \rangle$$

bzw. allgemeiner

$$\langle g(x)h(y) \rangle = \langle g(x) \rangle \langle h(y) \rangle$$

Parallel zur Kovarianz wird auch der *Korrelationskoeffizient*

$$\rho_{xy} = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y}$$

verwendet, für den wegen der Normierung auf die Varianzen von x und y gilt

$$-1 \leq \rho_{xy} \leq 1$$

Oft betrachtet man auch den Korrelationskoeffizienten einer Zufallsvariablen zu verschiedenen Zeitpunkten, z.B. zur Zeit i und eine Zeit k später. Wenn die statistischen Eigenschaften

von x nicht von der Zeit abhängen, dann hängt diese *Zeitkorrelationsfunktion* nur von der Zeitdifferenz k , aber nicht von i selbst ab

$$C_k = \rho_{x_i, x_{i+k}} = \frac{\text{cov}(x_i, x_{i+k})}{\sigma_{x_i} \sigma_{x_{i+k}}} = \frac{\langle x_i x_{i+k} \rangle - \langle x \rangle^2}{\langle x^2 \rangle - \langle x \rangle^2}$$

In der obigen Form ist C_k bei $k = 0$ auf 1 normiert. Da nach einer gewissen Zeit x_{i+k} von x_i statistisch unabhängig ist, gehen Zeitkorrelationsfunktionen in der Regel für $k \rightarrow \infty$ nach Null.

Beispiele

Der einfachste Fall einer kontinuierlichen Zufallsvariablen ist die *Gleichverteilung* über einem Intervall $[a, b]$, d.h. x nimmt alle Werte aus dem Intervall mit der gleichen Wahrscheinlichkeit an

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{wenn } a \leq x \leq b \\ 0 & \text{sonst} \end{cases}$$

[Der Wert von $f(x)$ im Inneren des Intervalls ergibt sich aus der Normierungsbedingung $\int dx f(x) = 1$.]

Eine ganze Reihe von physikalischen Prozessen, darunter die Lebensdauern der Kerne beim radioaktiven Zerfall, werden durch die *Exponentialverteilung*

$$f(x) = \begin{cases} \frac{1}{\lambda} e^{-x/\lambda} & \text{für } x \geq 0 \\ 0 & \text{sonst} \end{cases}$$

beschrieben. Die Zufallsvariable x kann hier alle Werte auf der positiven reellen Achse annehmen; allerdings sind große Werte viel seltener als kleine. Der Erwartungswert von x ist durch den Parameter λ gegeben.

Das wichtigste Modell einer kontinuierlichen Zufallsvariablen ist jedoch die *Gauß-* oder *Normalverteilung*

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{(x - \mu)^2}{2\sigma^2} \right]$$

Hier kann die Zufallsvariable alle Werte zwischen $-\infty$ und $+\infty$ auf der reellen Achse annehmen. Der Erwartungswert von x ist μ , die Varianz σ^2 . Den Standardfall $\mu = 0$, $\sigma = 1$ bezeichnet man als $N(0,1)$ -Verteilung. Die Normalverteilung spielt nicht nur in der Statistik eine zentrale Rolle, auch in der Physik sind z.B. die kartesischen Komponenten der Geschwindigkeit von Gasmolekülen normalverteilt.

Ein Beispiel für eine diskrete Wahrscheinlichkeitsverteilung ist die *Binomialverteilung*

$$B_m(n, p) = \binom{n}{m} p^m (1-p)^{n-m}$$

Sie gibt an, mit welcher Wahrscheinlichkeit ein Ereignis in n unabhängigen Beobachtungen genau m -mal auftritt ($m = 0, 1, 2, \dots, n$). Dabei ist p die Wahrscheinlichkeit, daß das Ereignis in einer einzelnen Beobachtung auftritt.

Stichproben

Sind die Wahrscheinlichkeitsverteilungen oder -dichten von Zufallsvariablen bekannt, so lassen sich daraus im Prinzip alle statistischen Kenngrößen wie Mittelwerte, Varianzen, Korrelationen usw. berechnen—sofern man nämlich in der Lage ist, die entsprechenden mathematischen Ausdrücke analytisch oder numerisch auszuwerten. In der Praxis tritt aber häufig der Fall auf, daß man zwar ein theoretisches Modell für die Verteilung hat, aber die Parameter nicht kennt und daher aus einer Messung bestimmen will, oder man hat überhaupt noch keine Vorstellung von der Form der Verteilung. In diesem Fall kann man die statistischen Kenngrößen auch mit Hilfe einer Stichprobe schätzen.

Wir betrachten ein Zufallsexperiment, dessen Ausgang durch eine Zufallsvariable x beschrieben wird. Unter einer (*unabhängigen*) *Stichprobe vom Umfang n*

$$\{x_1, x_2, \dots, x_n\}$$

versteht man eine n -fache Wiederholung des Experiments, wobei die Ergebnisse in den einzelnen Experimenten einander nicht beeinflussen sollen. Mit anderen Worten, $\{x_1, x_2, \dots, x_n\}$ ist ein Satz von Zufallsvariablen mit denselben statistischen Eigenschaften wie x , und x_i und x_j sind für $i \neq j$ statistisch unabhängig (x_i ist einfach der Wert von x im i -ten Versuch). Die Erzeugung der Stichprobe kann selbst wieder als Zufallsexperiment aufgefaßt werden.

Erwartungswerte und andere statistische Kenngrößen werden nun durch Mittelung über die Stichprobe geschätzt. So ist z.B. der *Stichprobenmittelwert* definiert als

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Man erwartet, daß

$$\bar{x} \approx \langle x \rangle$$

gilt, wobei $\langle x \rangle$ den “wahren” (aus der eventuell unbekanntem theoretischen Verteilung berechneten) Mittelwert von x bezeichnet. Der Stichprobenmittelwert ist tatsächlich *erwartungstreu*, denn der Erwartungswert für das Zufallsexperiment “Stichprobenmittelwert bilden” ist

$$\langle \bar{x} \rangle = \left\langle \frac{1}{n} \sum_i x_i \right\rangle = \frac{1}{n} \sum_i \langle x \rangle = \langle x \rangle$$

Dies ist zunächst aber nur eine Aussage für den Fall, daß man unendlich viele Stichproben erzeugen und daraus jeweils den Stichprobenmittelwert bestimmen könnte.

Der analoge Ansatz für die Varianz

$$(s^2)' = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

unterschätzt aber den wahren Wert um einen Faktor $(n-1)/n$, denn

$$\langle (s^2)' \rangle = \frac{1}{n} \left\langle \sum_i \left(x_i - \frac{1}{n} \sum_j x_j \right) \left(x_i - \frac{1}{n} \sum_k x_k \right) \right\rangle$$

$$\begin{aligned}
&= \frac{1}{n} \left\langle \sum_i x_i^2 - \frac{2}{n} \sum_{ij} x_i x_j + \frac{1}{n^2} \sum_{ijk} x_j x_k \right\rangle = \frac{1}{n} \left[\sum_i \langle x_i^2 \rangle - \frac{1}{n} \sum_{ij} \langle x_i x_j \rangle \right] \\
&= \frac{1}{n} \left[n \langle x^2 \rangle - \frac{1}{n} (n \langle x^2 \rangle + n(n-1) \langle x \rangle^2) \right] = \frac{n-1}{n} (\langle x^2 \rangle - \langle x \rangle^2)
\end{aligned}$$

Wenn man also als Schätzwert

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

verwendet, dann ist diese Größe konstruktionsgemäß erwartungstreu, d.h. $\langle s^2 \rangle = \sigma^2$.

Eine vollkommen analoge Rechnung ergibt für die Varianz des Stichprobenmittelwerts,

$$\sigma_{\bar{x}}^2 = \langle (\bar{x} - \langle x \rangle)^2 \rangle$$

folgende Beziehung zur (wahren) Varianz von x

$$\sigma_{\bar{x}}^2 = \frac{1}{n} \sigma_x^2$$

Die Wurzel daraus bezeichnet man als *Standardfehler des Mittelwerts*. Da σ_x^2 für eine gegebene Zufallsvariable x eine feste Zahl ist, zeigt die letzte Formel, daß auch bei Durchführung nur einer einzigen Stichprobe der Stichprobenmittelwert nahe am wahren Mittelwert liegen wird, wenn nur der Umfang der Stichprobe hinreichend groß ist.

Erzeugung von Zufallszahlen am Computer

Pseudo-Zufallszahlengeneratoren

Sowohl zur Simulation von Zufallsprozessen im eigentlichen Sinn als auch zur Lösung vieler anderer mathematischer Aufgaben ist es wünschenswert, am Computer Stichproben von Zufallsvariablen erzeugen zu können. Im Prinzip könnte man dazu natürlich "echte" Zufallsprozesse wie den Zerfall einer radioaktiven Quelle oder thermisches Rauschen verwenden. Abgesehen von instrumentellen Problemen (und wer hätte schon gern eine Strahlungsquelle auf dem Schreibtisch), würden sich auf diese Weise aber kaum Zufallszahlen in hinreichender Menge und mit der Geschwindigkeit erzeugen lassen, wie sie von vielen Anwendungen verbraucht werden. Außerdem muß es für Testzwecke möglich sein, ein und dieselbe Folge von Zufallszahlen beliebig oft zu reproduzieren.

Aus diesem Grund verwendet man in Simulationen fast ausschließlich *Pseudozufallszahlen*: Das sind Folgen von Zahlen, die zwar durch einen streng deterministischen Algorithmus erzeugt werden, aber "zufällig aussehen", d.h. gewisse statistische Tests auf Zufälligkeit erfüllen. (Daraus folgt auch, daß es für jeden Pseudo-Zufallszahlengenerator eine Anwendung geben kann, bei der er sich als nicht zufällig genug erweist.)

In den meisten Compilern, Interpretern und manchen Script-Sprachen ist wengstens ein (Pseudo-)Zufallszahlengenerator implementiert, der gleichverteilte Zufallszahlen aus dem Intervall $[0, 1)$ liefert. In der Regel ist das ein *linearer Kongruenzgenerator (Modulo-Generator)*, der im einfachsten Fall nach dem folgenden Prinzip arbeitet:

Es werden ganze Zahlen a , c und m sowie ein Startwert (“Seed”) $i_0 < m$ vorgegeben und rekursiv die Folgen

$$i_{n+1} = (a \cdot i_n + c) \bmod m$$

$$\xi_{n+1} = i_{n+1}/m$$

gebildet. Die Zahlen a , c und m sind fest und bestimmen die Eigenschaften des Zufallszahlen-generators. Nach Vorgabe eines (mit gewissen Einschränkungen) beliebigen Startwerts i_0 erhält man durch sukzessives Anwenden der obigen Vorschrift zunächst eine streng deterministische Folge von ganzen Zahlen i_0, i_1, i_2, \dots mit $0 \leq i_n < m$. Die eigentlichen “Funktionswerte” des Generators sind die ξ_n , für die nach Konstruktion $0 \leq \xi_n < 1$ gilt.

Auf Grund der Formeln ist klar, daß die i_n höchstens m verschiedene Werte annehmen können; dann muß sich die Folge wiederholen. Unter bestimmten Voraussetzungen an die Parameter a , c und m kann man erreichen, daß die *Periode* des Generators den Maximalwert m erreicht. Dann sind die ξ_n auf jeden Fall gleichverteilt auf dem Intervall $[0, 1)$. Je nach Wahl von a , c und m erfüllen sie auch gewisse Tests auf Zufälligkeit (statistische Unabhängigkeit). Auf keinen Fall dürfen jedoch für die Parameter des Generators x-beliebige Zahlen eingesetzt werden.

Um sich die Berechnung der Modulo-Funktion zu ersparen, wird häufig $m = 2^r$ gesetzt, wo r die Wortlänge des Computers in Bits ist (bzw. $m = 2^{r-1}$, wenn nicht Zweierkomplement-Darstellung verwendet, sondern das Vorzeichen separat gespeichert wird). In diesem Fall werden vom Ergebnis einer Operation mit ganzen Zahlen, das einen “Overflow”, also mehr als r Binärstellen, produziert, automatisch nur die niedrigsten r Bits gespeichert. Die folgende Tabelle enthält die Parameter für einige einfache Generatoren.

a	c	m	i_0
16807	0	$2^{31} - 1$	ungerade
65539	0	2^{31}	ungerade
69069	1	2^{32}	
1664525	0	2^{32}	ungerade
25214903917	11	2^{48}	

Der Eintrag in der zweiten Zeile (“IBM-Generator”) war seinerzeit auf Großrechnern weit verbreitet, ist aber eigentlich ein Gegenbeispiel, da die resultierenden Zufallszahlen stark korreliert sind. Faßt man nämlich Tripel von aufeinanderfolgenden Funktionswerten zu Koordinaten $(\xi_n, \xi_{n+1}, \xi_{n+2})$ im dreidimensionalen Einheitswürfel zusammen, so kommen alle diese Punkte auf einer geringen Anzahl von Ebenen zu liegen. Solche Korrelationen treten zwar prinzipiell immer auf, bei guten Generatoren aber erst in viel höheren Dimensionen.

Einfache (in Compiler “eingebaute”) Generatoren wie die obigen sind zwar für viele Zwecke ausreichend, bei Anwendungen, die kritisch von der Qualität der Zufallszahlen abhängen, empfiehlt es sich jedoch, einen gut getesteten Generator aus einer Programmbibliothek zu verwenden.

Erzeugung nicht-gleichverteilter Zufallszahlen

Während man davon ausgehen kann, daß am Computer—entweder über den eingebauten Generator oder eine Bibliotheksfunktion—standardmäßig gleichverteilte Zufallszahlen aus dem Intervall $[0, 1)$ zur Verfügung stehen, müssen in der Praxis häufig Zufallsvariablen mit einer anderen Verteilung simuliert werden.

Ist z.B. x eine diskrete Zufallsvariable, die die Werte X_1, X_2, \dots, X_N mit Wahrscheinlichkeiten p_1, p_2, \dots, p_N annimmt, so kann man

$$\xi_0 = 0, \quad \xi_1 = p_1, \quad \xi_2 = p_1 + p_2, \quad \xi_3 = p_1 + p_2 + p_3, \quad \dots, \quad \xi_N = 1$$

setzen und das Einheitsintervall in eine Folge von Teilintervallen

$$[\xi_0, \xi_1), \quad [\xi_1, \xi_2), \quad [\xi_2, \xi_3), \quad \dots, \quad [\xi_{N-1}, \xi_N)$$

zerlegen. Da die Wahrscheinlichkeit, in das Teilintervall $[\xi_{i-1}, \xi_i)$ zu treffen, gleich seiner Länge $\xi_i - \xi_{i-1} = p_i$ ist, "würfelt" man also gleichverteilte Zufallszahlen ξ aus $[0, 1)$ und setzt

$$x = X_i, \quad \text{wenn} \quad \xi_{i-1} \leq \xi < \xi_i$$

Ein weiterer einfacher Fall ist die Simulation einer gleichverteilten (kontinuierlichen) Zufallsvariablen x über dem Intervall $[a, b)$. Hier leistet offenbar die Transformation

$$\xi \rightarrow x = a + \xi(b - a)$$

wobei wieder $\xi \in [0, 1)$ gleichverteilt ist, das Gewünschte.

Die Verallgemeinerung der *Transformationsmethode* für kontinuierliche Zufallsvariablen erhalten wir, indem wir zunächst eine beliebige Transformation $x \rightarrow y(x)$ betrachten, die streng monoton zunehmend (oder streng monoton abnehmend) und differenzierbar sein soll. y ist dann ebenfalls eine Zufallsvariable, und die Wahrscheinlichkeit, daß y in einem Intervall $[Y_1, Y_2)$ liegt, ist

$$P(Y_1 \leq y < Y_2) = P(X_1 \leq x < X_2)$$

wobei X_1 und X_2 die Urbilder von Y_1 und Y_2 sind und der streng monoton zunehmende Fall angenommen wurde. Sind $f(x)$ und $g(y)$ die Wahrscheinlichkeitsdichten von x und y , so ergibt sich daraus

$$\int_{X_1}^{X_2} dx f(x) = \int_{Y_1}^{Y_2} dy g(y) = \int_{x(Y_1)}^{x(Y_2)} dx \frac{dy}{dx} g(y(x))$$

und im Limes $X_2 \rightarrow X_1, Y_2 \rightarrow Y_1$

$$f(x) = \frac{dy}{dx} g(y(x))$$

In der Form

$$\frac{dy}{dx} = \frac{f(x)}{g(y)}$$

kann man dies als Differentialgleichung für jene Transformation ansehen, die aus x eine Zufallsvariable y mit vorgegebener Wahrscheinlichkeitsdichte $g(y)$ macht. (Im Fall einer streng

monoton abnehmenden Transformation ist die linke Seite der letzten Gleichung durch $-dy/dx$ zu ersetzen.)

Soll beispielsweise die Exponentialverteilung mit Hilfe des eingebauten Zufallszahlengenerators simuliert werden, so sind

$$\begin{aligned} f(x) &= 1 & 0 \leq x < 1 \\ g(y) &= \frac{1}{\lambda} e^{-y/\lambda} & 0 \leq y < \infty \end{aligned}$$

also

$$\begin{aligned} \frac{dy}{dx} &= \lambda e^{y/\lambda} \\ d(y/\lambda) e^{-y/\lambda} &= dx \\ -e^{-y/\lambda} &= x + c \end{aligned}$$

Die Integrationskonstante ergibt sich aus der Forderung $y(0) = 0$ zu $c = -1$. Die Transformation

$$x \rightarrow y = -\lambda \ln(1 - x)$$

macht also aus der auf $[0, 1)$ gleichverteilten Zufallsvariablen x eine exponentialverteilte Variable y auf $[0, \infty)$.

Der größte Nachteil der Transformationsmethode, die im wesentlichen darauf hinausläuft, die Gleichung

$$G(y) = F(x)$$

für gegebenes x nach y aufzulösen, besteht darin, daß man häufig entweder schon die Stammfunktionen zu f und g , F und G , nicht analytisch angeben oder die Gleichung $G(y) = F(x)$ nicht nach y auflösen kann. Es gibt aber neben der Transformationsmethode einige andere allgemein anwendbare Verfahren sowie eine Unzahl von Methoden zur Simulation von Stichproben aus speziellen Verteilungen.

Monte Carlo-Integration

Eines der wichtigsten—wenn nicht *das* wichtigste—Monte Carlo-Verfahren ist eine Methode zur numerischen Berechnung von Integralen, insbesondere in hohen Dimensionen: die sogenannte Monte Carlo-Integration.

Angenommen, es sei das Integral einer Funktion $g(x)$ auf dem Intervall $[a, b]$ zu berechnen

$$I = \int_a^b dx g(x)$$

Spaltet man vom Integral die Länge des Integrationsgebietes, $b - a$, ab

$$I = (b - a) \int_a^b dx \frac{1}{b - a} g(x)$$

so kann man den ersten Faktor des Integranden als (korrekt normierte) Wahrscheinlichkeitsdichte

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{wenn } a \leq x \leq b \\ 0 & \text{sonst} \end{cases}$$

einer auf $[a, b]$ gleichverteilten Zufallsvariablen x auffassen. Damit läßt sich aber, abgesehen von einem Vorfaktor, das Integral als Erwartungswert der Funktion g der Zufallsvariablen x interpretieren

$$I = (b-a) \int dx f(x) g(x) = (b-a) \langle g(x) \rangle$$

Kann man das Integral bzw. den Erwartungswert nicht analytisch berechnen, so liegt es nahe, eine Stichprobe $\xi_1, \xi_2, \dots, \xi_n$ der Variablen x zu erzeugen und den Erwartungswert von $g(x)$ durch den Stichprobenmittelwert zu approximieren

$$I \approx (b-a) \overline{g(x)} = (b-a) \frac{1}{n} \sum_{i=1}^n g(\xi_i)$$

Dieser Ausdruck hat große Ähnlichkeit mit konventionellen Formeln für numerische Integration, allerdings werden hier die Stützpunkte *zufällig* im Integrationsgebiet gewählt und nicht äquidistant oder in einer anderen regelmäßigen Anordnung.

Beim obigen, *naive MC-Integration* genannten Verfahren würfelt man die Stützstellen gleichverteilt im Integrationsgebiet. Ein allgemeinerer Fall liegt vor, wenn das Integral von vornherein die Form

$$I = \int dx f(x) g(x)$$

mit einer nicht-konstanten Funktion $f(x) \geq 0$ hat. Man kann dann $f(x)/\int dx f(x)$ als Wahrscheinlichkeitsdichte einer nicht-gleichverteilten Zufallsvariablen x auffassen und das Integral

$$I = \left[\int dx f(x) \right] \int dx \frac{f(x)}{\int dx' f(x')} g(x) = \left[\int dx f(x) \right] \langle g(x) \rangle$$

wieder durch einen Stichprobenmittelwert annähern

$$I \approx \left[\int dx f(x) \right] \frac{1}{n} \sum_{i=1}^n g(\xi_i)$$

wobei die ξ_i jetzt allerdings eine Stichprobe aus der Wahrscheinlichkeitsdichte $f(x)/\int dx f(x)$ sein müssen. Dieses Verfahren nennt man *Importance Sampling*.

Der große Vorteil von MC-Integrationsverfahren besteht darin, daß bei unabhängigen Stichproben vom Umfang n die Varianz des Stichprobenmittelwerts

$$\text{Var}(\bar{g}) = \frac{1}{n} \text{Var}(g)$$

und der Fehler des Integrals daher proportional zu $1/\sqrt{n}$ ist

$$\Delta I = \frac{1}{\sqrt{n}} \left[\int dx f(x) \right] \sqrt{\langle g^2 \rangle - \langle g \rangle^2}$$

Da diese Beziehung unabhängig von der Dimensionalität des Integrationsgebiets ist, sind MC-Verfahren zwar bei niedrig-dimensionalen Problemen konventionellen Integrationsverfahren unterlegen, bei hochdimensionalen Integralen stellen sie jedoch die Methode der Wahl dar.