...ilungen von Zufallszahlen

Wir haben nun eine Reihe von Verfahren kennengelernt, um Pseudo-Zufallszahlen x_i gleichverteilt im Intervall $]0,1]$ zu erzeugen. Wir wollen nun weitere Verfahren kennenlernen, wie man aus diesen gleichverteilten Zufallszahlen neue Sequenzen von Zufallszahlen erzeugen können, die einer vorgegebenen Verteilung folgen.

Der einfachste Fall ist die gleichmäßige Verteilung in einem beliebigen Intervall $]a,b]$, welche wir erzeugen via

$$\tilde{x}_i = a + (b-a) \cdot u_i$$

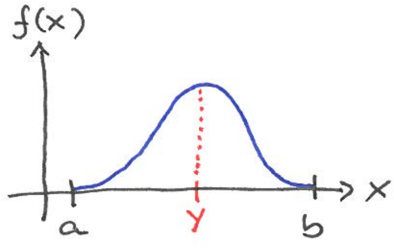
uniform in $]0,1]$ verteilte Zufallszahl.

Für andere, beliebige Verteilungen können wir eines der drei folgenden Verfahren benutzen:

- Inversion der integrierten Verteilung
- von Neumann'sche Verwerfungsmethode
- Metropolis-Algorithmus.

Inversionsmethode

Wie können wir aus uniform verteilten Zufallszahlen u_i neue Zufallszahlen x_i erzeugen, so daß die x_i einer Verteilung $f(x)$ im Intervall $]a, b]$ folgen?



Betrachten wir dazu die Wahrscheinlichkeiten

$$P[x<y] = \frac{\int_a^y f(t)\,dt}{\int_a^b f(t)\,dt} =: F(y) \equiv P[u < F(y)]$$

uniform verteilte Zufallszahl

$$\rightarrow \quad \underline{x = F^{-1}(u)}$$

Diese Inversionsmethode kann immer dann eingesetzt werden, wenn das Integral $F(y)$ einfach invertiert werden kann.

Bsp.:
- Exponentielle Verteilung $f(x) = \lambda e^{-\lambda x}$ kann aus einer gleichverteilten Verteilung erhalten werden via $x = -\frac{1}{\lambda} \ln(1-u)$
- Die Normal-Verteilung $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2}$ kann in einer Dimension <u>nicht</u> einfach integriert werden, wohingegen dies in zwei Dimensionen leicht möglich ist.

  Wir können (für den ein-dimensionalen Fall) allerdings <u>zwei</u> normalverteilte Zufallszahlen x_1 und x_2 von zwei gleichförmig verteilten Zufallszahlen u_1 und u_2 erhalten (Box-Muller Methode):

$$x_1 = \sqrt{-2\ln(1-u_1)} \cdot \sin u_2 \qquad x_2 = \sqrt{-2\ln(1-u_1)} \cdot \cos u_2$$

Verwerfungsmethode (von Neumann)

Gegeben sei wiederum eine beliebige Verteilung $f(x)$, nach welcher wir Zufallszahlen aus einer uniformen Verteilung erzeugen wollen.

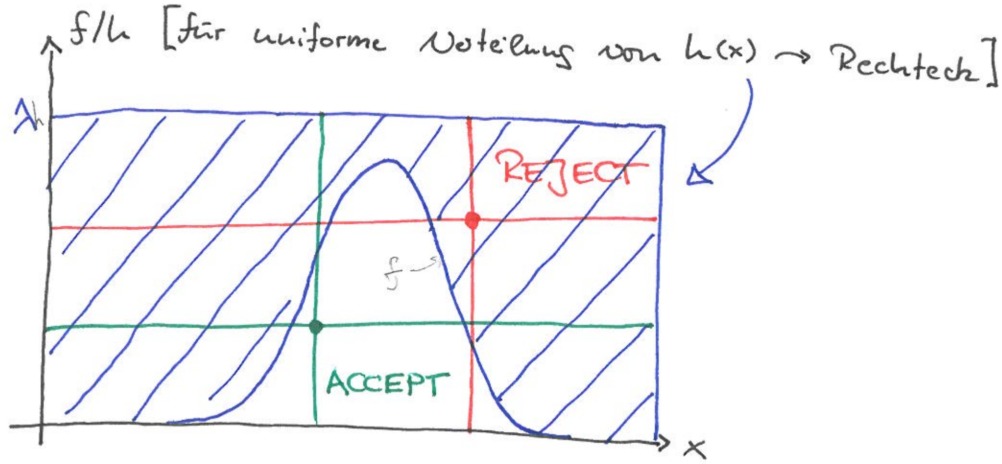
Strategie: Suche nach einer einfachen Verteilung $h(x)$, welche eine obere Schranke der Verteilung $f(x)$ darstellt, d.h.

$$f(x) < \lambda \cdot h(x) \qquad \text{für } \forall x$$

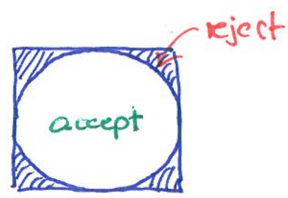
Konstante (λ)
(nötig, da sowohl $f(x)$ als auch $h(x)$ normiert seien)

Verwerfungsmethode:

- Erzeuge eine h-verteilte Zufallszahl x
- Erzeuge eine uniforme Zufallszahl $0 < u \leq 1$
- Akzeptiere x, wenn $u < \dfrac{f(x)}{\lambda \cdot h(x)}$ ist, andernfalls verwerfe x und gehe zurück zum ersten Schritt.



Genau diese Verwerfungsmethode haben wir bereits in der Berechnung der Kreiszahl π benutzt:



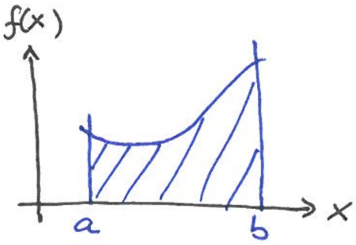
Monte Carlo Integratoren

Als erste Anwendung von Zufallszahlen wollen wir uns noch einmal der Integration einer Funktion im n-dimensionalen Raum zuwenden, d.h. wir betrachten das Integral

$$I = \int d^n x \; f(\vec{x}) \qquad \text{wobei } \vec{x} = (x_1, x_2, \ldots, x_n) \text{ sei.}$$

Der Einfachheit halber betrachten wir zunächst den 1-dimensionalen Fall

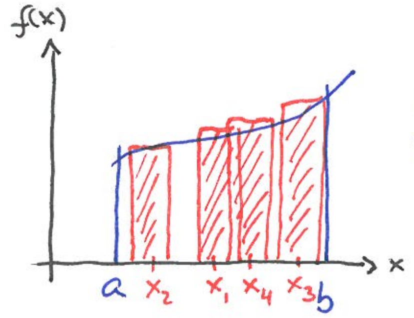
$$I = \int dx \; f(x)$$



Wie können wir nun Zufallszahlen benutzen, um dieses Integral zu berechnen? Die Idee ist, das Integral zu "samplen" über eine Menge von x-Werten $\{x_i\}$. Wähle dazu eine Folge von Zufallszahlen $\{x_i\}$ mit $i = 1, 2, \ldots, N$ gleichverteilt im Intervall $[a, b)$.

Näherung für das Integral:

$$I \approx \underbrace{\frac{b-a}{N}}_{\Omega \text{ Integrationsvolumen}} \cdot \sum_{i=1}^{N} f(x_i)$$



Aus der Gleichverteilung der x_i folgt:

$$\lim_{N \to \infty} \frac{b-a}{N} \sum_{i=1}^{N} f(x_i) = \int_a^b f(x) dx$$

Dieses Vorgehen bezeichnet man auch als "direct sampling".

Wir wollen nun den statistischen Fehler dieser Integration via "direct sampling" bestimmen.
Dazu setzen wir zunächst einmal das Integrationsvolumen $\Omega = \int dx$ auf

$$\Omega = b - a = 1 \qquad \text{etwa durch } \begin{matrix} b=1 \\ a=0 \end{matrix}.$$

Damit wird unsere Abschätzung des Integrals

$$I = \int_0^1 dx\, f(x) \approx \frac{1}{N}\sum_{i=1}^{N} f(x_i) = \bar{I} = \langle f \rangle$$

Mittelwert unserer Abschätzung

Neben dem Mittelwert unserer Abschätzung wollen wir dazu auch noch das 2. Moment der Abschätzung ausrechnen, um schließlich die Varianz zu erhalten. Für das 2. Moment gilt:

$$\begin{aligned}
\left(\frac{1}{N}\sum_{i=1}^{N} f(x_i)\right)^2 &= \frac{1}{N^2}\sum_{i,j} f(x_i)\, f(x_j) \\
&= \frac{1}{N^2}\left[\sum_{i=1}^{N} f(x_i)^2 + \sum_{i\neq j} f(x_i)\, f(x_j)\right] \\
&= \frac{1}{N}\langle f^2\rangle + \frac{N(N-1)}{N^2}\langle f\rangle\langle f\rangle \\
&= \frac{1}{N}\langle f^2\rangle + \langle f\rangle^2 - \frac{1}{N}\langle f\rangle^2 = \langle f\rangle^2 + \underbrace{\frac{1}{N}\left(\langle f^2\rangle - \langle f\rangle^2\right)}_{\text{Varianz}}
\end{aligned}$$

Für die Varianz unser Abschätzung erhalten wir damit

$$\mathrm{Var}\, I = \frac{1}{N}\left(\langle f^2\rangle - \langle f\rangle^2\right)$$

und den statistischen Fehler

$$\Delta = \sqrt{\mathrm{Var}\, I} = \frac{1}{\sqrt{N}} \cdot \sqrt{\langle f^2\rangle - \langle f\rangle^2}$$

Der statistische Fehler skaliert also asymptotisch wie

$$\Delta \sim \mathcal{O}(N^{-1/2})$$

unabhängig von der Dimension des Integrationsvolumens.

Vergleichen wir dieses Skalenverhalten mit jenen der bisherigen Integrationsverfahren:

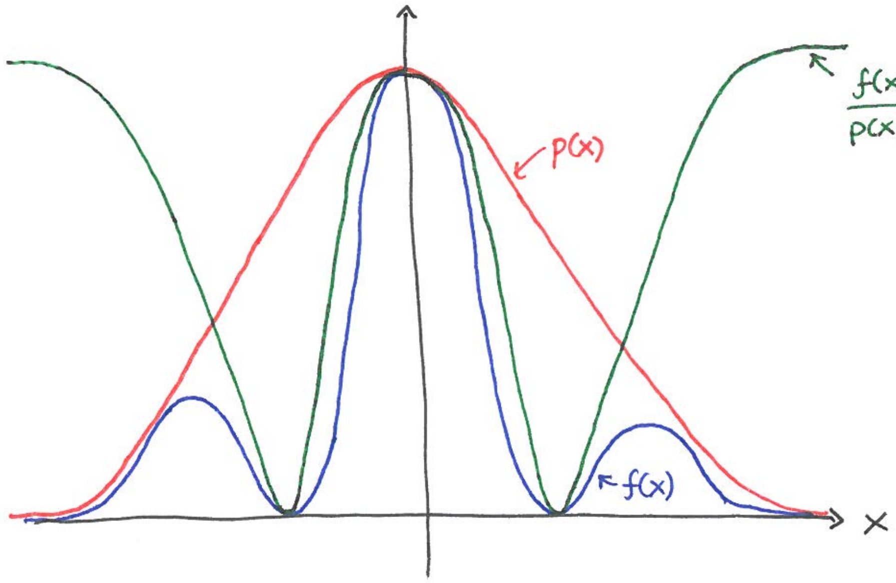
- Riemann-Summe $\quad \mathcal{O}(N^{-1/d})$
- Trapez-Regel $\quad \mathcal{O}(N^{-2/d})$
- Simpson-Regel $\quad \mathcal{O}(N^{-4/d})$

Das heißt, in Dimensionen $d \geq 5$ ist die Monte Carlo-Integration besser als die Trapez-Regel und in Dimensionen $d \geq 9$ besser als die Simpson-Regel.

Monte Carlo Integration eignet sich also insbesondere für höher-dimensionale Integrale.

Importance Sampling

Das soeben eingeführte "direct sampling" hat allerdings auch seine Nachteile: Immer dann, wenn eine Funktion in einer kleinen Region (im Vergleich zum Integrationsvolumen) einen Peak hat und somit eine große Varianz besitzt.



Beim "direct sampling" wird eine Menge "Zeit" in Regionen verschwendet, wo die Funktion klein ist.

Der Ausweg hierzu ist sogenanntes "importance sampling", wobei wir die Zufallszahlen $\{x_i\}$ nicht mehr gleichverteilt sondern gemäß einer Verteilung $p(x)$ mit

$$\int p(x)\,dx = 1$$

erzeugen.

Mit diesen p-verteilten Zufallszahlen führen wir unsere Integration gemäß

$$I = \langle f \rangle = \frac{1}{\Omega} \int f(x)\,dx = \frac{1}{\Omega} \int \frac{f(x)}{p(x)} \cdot p(x)\,dx \approx \frac{1}{N} \sum_{i=1}^{N} \frac{f(x_i)}{p(x_i)}$$

für x_i gemäß Verteilung $p(x)$

durch. Der Fehler wird damit

$$\Delta = \frac{1}{\sqrt{N}} \cdot \sqrt{\mathrm{Var}\, f/p}$$

Wir wollen also die Verteilung $p(x)$ möglichst nah an der zu integrierenden Funktion $f(x)$ wählen.