

VL4: Kapitel 2: Die Monte Carlo Methode

Es sei

$$p : \mathbb{R}^n \rightarrow \mathbb{R}$$

eine ein- oder mehrdimensionale Wahrscheinlichkeitsdichte, also $p \geq 0$, $n \geq 1$ und

$$\int_{\mathbb{R}^n} p(\phi) d^n \phi = 1 .$$

und $F : \mathbb{R}^n \rightarrow \mathbb{R}$ sei eine gegebene Funktion. Wir wollen das folgende Integral, das man auch als Erwartungswert von F interpretieren kann, berechnen¹:

$$I := \int_{\mathbb{R}^n} F(\phi) p(\phi) d^n \phi = \mathbf{E}[F]$$

Dazu erzeugt man N p -verteilte, unabhängige Zufallszahlen, oder, falls $n > 1$, N Vektoren von Zufallszahlen ($\phi_i \in \mathbb{R}^n$)

$$(\phi_1, \phi_2, \dots, \phi_N) \in (\mathbb{R}^n)^N = \mathbb{R}^{nN}$$

und bildet die Monte Carlo Summe (beachten Sie, dass nur das F in der Monte Carlo Summe auftaucht, das p steckt schon in den Zufallszahlen drin)

$$S_N(F) := \frac{1}{N} \sum_{i=1}^N F(\phi_i) \quad (1)$$

etwa mit $N = 10000$ oder $N = 100000$. Grundlage der Monte Carlo Berechnung ist dann die folgende approximative Identität:

$$S_N(F) \stackrel{N \rightarrow \infty}{\approx} I = \int_{\mathbb{R}^n} F(\phi) p(\phi) d^n \phi \quad (2)$$

Bevor wir das ausprobieren, wollen wir uns den theoretischen Hintergrund dazu kurz anschauen. Dazu erinnern wir uns an den zentralen Grenzwertsatz aus der Stochastik-VL:

Theorem (Zentraler Grenzwertsatz): Es seien X_1, X_2, X_3, \dots eine Folge von unabhängigen, identisch verteilten Zufallsvariablen mit

$$\begin{aligned} \mathbf{E}[X_k] &= \mu \\ \mathbf{V}[X_k] &= \sigma^2 \end{aligned}$$

¹ist p eine diskrete Wahrscheinlichkeitsverteilung, so ist das Integral durch eine entsprechende Summe zu ersetzen

Wir betrachten die Mittelwerte

$$S_N := \frac{1}{N}(X_1 + X_2 + \cdots + X_N) \quad (3)$$

die dann Erwartungswert und Varianz

$$\begin{aligned} \mathbf{E}[S_N] &= \mu \\ \mathbf{V}[S_N] &= \frac{\sigma^2}{N} \end{aligned}$$

haben. Wir definieren die normierten Grössen

$$Z_N := \frac{S_N - \mathbf{E}[S_N]}{\sqrt{\mathbf{V}[S_N]}} = \frac{\sqrt{N}}{\sigma} (S_N - \mu)$$

Dann gilt

$$\lim_{N \rightarrow \infty} \text{Prob}[Z_N \in (x, x + dx)] = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \quad (4)$$

Dieses Resultat können wir jetzt auf die Monte Carlo Summe (1) anwenden. Aus der Gleichung (4) folgt:

$$\begin{aligned} \text{Prob}[|Z_N| \geq \alpha] &\stackrel{N \rightarrow \infty}{=} \int_{-\infty}^{-\alpha} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx + \int_{\alpha}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\ &= 2 \int_{-\infty}^{-\alpha} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 2\Phi(-\alpha) \end{aligned} \quad (5)$$

wobei die $\Phi(x)$ -Funktion also definiert ist durch

$$\Phi(x) := \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy \quad (6)$$

Gleichung (5) können wir auch so schreiben:

$$\text{Prob}\left[\left| \frac{\sqrt{N}}{\sigma} (S_N - \mu) \right| \geq \alpha \right] \stackrel{N \rightarrow \infty}{=} 2\Phi(-\alpha)$$

oder

$$\text{Prob}\left[|(S_N - \mu)| \geq \frac{\sigma}{\sqrt{N}} \alpha \right] \stackrel{N \rightarrow \infty}{=} 2\Phi(-\alpha) \quad (7)$$

Als unabhängige, identisch verteilte Zufallsvariablen wählen wir jetzt

$$X_i := F(\phi_i) \quad (8)$$

so dass die S_N 's dann also mit der Monte Carlo Summe

$$S_N = S_N(F) = \frac{1}{N} \sum_{i=1}^N F(\phi_i) \quad (9)$$

übereinstimmen. Wir haben dann

$$\mu = \mathbf{E}[F(\phi_i)] = \mathbf{E}[F(\phi)] = \int_{\mathbb{R}^d} F(\phi) p(\phi) d^d \phi = I \quad (10)$$

das ist also genau das Integral, was wir berechnen wollen, und die Grösse

$$\sigma^2 = \mathbf{V}[F] = \mathbf{E}[F^2] - \mathbf{E}[F]^2 = \int_{\mathbb{R}^d} F(\phi)^2 p(\phi) d^d \phi - \left(\int_{\mathbb{R}^d} F(\phi) p(\phi) d^d \phi \right)^2$$

hat dann einen Einfluss auf den Monte Carlo Error. Die Gleichung (7) lautet dann

$$\text{Prob} \left[|S_N(F) - \mathbf{E}[F]| \geq \alpha \sqrt{\frac{\mathbf{V}[F]}{N}} \right] \stackrel{N \rightarrow \infty}{\cong} 2\Phi(-\alpha) \quad (11)$$

Schauen wir uns Gleichung (11) genauer an. Das $\mathbf{E}[F]$ ist das exakte, das theoretische Resultat, was wir berechnen wollen. Das $S_N = S_N(f)$ ist die Monte Carlo Summe, diese Grösse tun wir also konkret numerisch berechnen indem wir Zufallszahlen ϕ_i simulieren und dann die Summe in (9) berechnen. Wir möchten dann natürlich, dass das S_N das $\mathbf{E}[F]$ möglichst gut ‘trifft’. Da das S_N ja mit Zufallszahlen berechnet wird, kann es natürlich immer mal sein, dass die Zahlen jetzt gerade etwas komisch waren und die Sache nicht so gut hinkommt. Was wir fordern können, ist, dass die Wahrscheinlichkeit, dass so etwas passiert, möglichst klein ist.

Gleichung (11) sagt, dass man schonmal daneben liegen kann, aber die Wahrscheinlichkeit dafür ist kleiner als $2\Phi(-\alpha)$. Wir können dann also etwa fordern:

$$2\Phi(-\alpha) \stackrel{!}{<} 1\%,$$

wir liegen in höchstens 1% aller Fälle daneben, oder etwa

$$2\Phi(-\alpha) \stackrel{!}{<} 0.01\%$$

wir liegen in höchstens 0.01% aller Fälle daneben, und bekommen dann die folgenden Werte für das α :

$$\begin{aligned} 2\Phi(-2.58) &= 1\% \\ 2\Phi(-3.89) &= 0.01\% \end{aligned}$$

Das heisst, mit einer Wahrscheinlichkeit von 99% ist

$$S_N(F) - 2.58\sqrt{\frac{\mathbf{V}[F]}{N}} \leq \mathbf{E}[F] \leq S_N(F) + 2.58\sqrt{\frac{\mathbf{V}[F]}{N}} \quad (12)$$

und mit einer Wahrscheinlichkeit von 99.99% ist

$$S_N(F) - 3.89\sqrt{\frac{\mathbf{V}[F]}{N}} \leq \mathbf{E}[F] \leq S_N(F) + 3.89\sqrt{\frac{\mathbf{V}[F]}{N}} \quad (13)$$

Das heisst, der Monte Carlo Fehler ist proportional zu der Grösse $\sqrt{\mathbf{V}[F]/N}$,

$$\text{Monte Carlo Fehler} \sim \sqrt{\frac{\mathbf{V}[F]}{N}} \quad (14)$$

Schauen wir uns jetzt zwei Beispiele an:

Beispiele:

Beispiel 1) Das Volumen τ_n der n -dimensionalen Einheitskugel lässt sich analytisch berechnen. Man findet:

$$\tau_n := \int_{\mathbb{R}^n} \chi(\|x\| \leq 1) d^n x = \frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2} + 1)} \quad (15)$$

wobei für ganzzahliges m die Formel $\Gamma(m + 1) = m!$ gilt. Überprüfen Sie die Formel (15) mit einer Monte Carlo Simulation für, sagen wir,

$$n \in \{2, 4, 6, 8\}$$

indem Sie auf dem Würfel $[-1, 1]^n$ gleichverteilte Zufallszahlen benutzen.

Lösung 1) Wir schreiben

$$\begin{aligned} \tau_n &= \int_{\mathbb{R}^n} \chi(\|x\| \leq 1) d^n x \\ &= \int_{[-1,1]^n} \chi(\|x\| \leq 1) d^n x \\ &= 2^n \int_{\mathbb{R}^n} \chi(\|x\| \leq 1) \frac{\chi_{[-1,1]^n}(x)}{2^n} d^n x \\ &= \int_{\mathbb{R}^n} F(x) p(x) d^n x \end{aligned}$$

mit dem Integranden

$$F(x) := 2^n \chi(\|x\| \leq 1) \quad (16)$$

und der W'keitsdichte

$$p(x) := \frac{\chi_{[-1,1]^n}(x)}{2^n} \quad (17)$$

Sind also

$$\begin{aligned} x^{(1)} &= (x_1^{(1)}, \dots, x_n^{(1)}) \in [-1, 1]^n \\ &\vdots \\ x^{(N)} &= (x_1^{(N)}, \dots, x_n^{(N)}) \in [-1, 1]^n \end{aligned}$$

N simulierte Zufallsvektoren mit auf dem Intervall $[-1, 1]$ gleichverteilten Zufallszahlen $x_j^{(i)}$, dann haben wir die folgende Monte Carlo Approximation für das τ_n :

$$\tau_n \approx \frac{1}{N} \sum_{i=1}^N F(x^{(i)}) \quad (18)$$

mit dem F gegeben durch (16).

Beispiel 2) Berechnen Sie das Integral (mit $\log 4 = \log_e 4$, Logarithmus zur Basis e)

$$I := \int_0^{\log 4} e^{-x/2} dx$$

- a) analytisch
- b) mit einer Monte Carlo Simulation, indem Sie auf dem Intervall $[0, \log 4]$ gleichverteilte Zufallszahlen benutzen
- c) mit einer Monte Carlo Simulation, indem Sie exponential-verteilte Zufallszahlen benutzen (wäre dann eher eine R-Aufgabe, weniger eine Excel-Aufgabe; R wollen wir hier in der FM2 aber nicht machen).

Was ist jeweils das F und das p in (b) und (c) ?

Lösung 2) a) Das exakte Resultat ist

$$\begin{aligned} I &= -2 e^{-x/2} \Big|_0^{\log 4} = 2(1 - e^{-\frac{\log 4}{2}}) \\ &= 2(1 - e^{-\log(4^{1/2})}) = 2(1 - e^{-\log(2)}) = 1 \end{aligned}$$

b) Wir schreiben

$$\begin{aligned} I &= \int_{-\infty}^{\infty} e^{-x/2} \chi(0 \leq x \leq \log 4) dx \\ &= \log 4 \int_{-\infty}^{\infty} e^{-x/2} \frac{\chi(0 \leq x \leq \log 4)}{\log 4} dx \\ &= \int_{\mathbb{R}} F(x) p(x) dx \end{aligned}$$

mit dem Integranden

$$F(x) := \log 4 \times e^{-x/2} \tag{19}$$

und der W'keitsdichte

$$p(x) := \frac{\chi(0 \leq x \leq \log 4)}{\log 4} \tag{20}$$

Sind also x_1, x_2, \dots, x_N auf dem Intervall $[0, \log 4]$ gleichverteilte Zufallszahlen, dann lautet die Monte Carlo Approximation für I

$$I \approx \frac{1}{N} \sum_{i=1}^N F(x_i)$$

mit dem F gegeben durch (19).

c) Wir schreiben

$$\begin{aligned} I &= \int_0^{\infty} \chi(0 \leq x \leq \log 4) e^{-x/2} dx \\ &= 2 \int_0^{\infty} \chi(0 \leq x \leq \log 4) \frac{1}{2} e^{-x/2} dx \\ &= \int_{\mathbb{R}} F(x) p(x) dx \end{aligned}$$

mit dem Integranden

$$F(x) := 2 \times \chi(0 \leq x \leq \log 4) \quad (21)$$

und der W'keitsdichte

$$p(x) := \frac{1}{2} e^{-x/2} \chi(x \geq 0) \quad (22)$$

Sind also x_1, x_2, \dots, x_N mit Parameter $\lambda = 1/2$ exponential-verteilte Zufallszahlen, dann lautet die Monte Carlo Approximation für I

$$I \approx \frac{1}{N} \sum_{i=1}^N F(x_i)$$

mit dem F gegeben durch (21).