Hochschule RheinMain
Prof. Dr. D. Lehmann

Stochastik II
SS 2023

week7a: Kapitel 4: Monte Carlo Integration
4.1: Theoretischer Hintergrund

Es sei

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

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

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

Für die meisten der gängigsten Wahrscheinlichkeitsverteilungen existieren in R sehr performante Zufallszahlgeneratoren, mit denen man also p-verteilte Zufallszahlen erzeugen kann. Das Kapitel 5 im RSkript_UniGiessen gibt dazu auf 3 Seiten eine sehr kompakte und verständliche Übersicht, das waren diese Sachen:

`http://hsrm-mathematik.de/SS2023/semester4/Stochastik2/W'keitsverteilungen-in-R.pdf`

Es sei jetzt $F : \mathbb{R}^n \to \mathbb{R}$ eine gegebene Funktion. Wir wollen das folgende Integral, das man auch als Erwartungswert von F interpretieren kann, berechnen[1]:

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

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

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

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)

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

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

$$S_N(F) \ \overset{N\to\infty}{\approx} \ I \ = \ \int_{\mathbb{R}^n} F(\phi)\, p(\phi)\; d^n\phi \tag{2}$$

Bevor wir das ausprobieren, wollen wir uns den theoretischen Hintergrund dazu kurz anschauen. Dazu erinnern wir uns an den zentralen Grenzwertsatz aus dem 2. Kapitel:

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

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

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

Wir betrachten die Mittelwerte (die nennen wir hier gleich S_n anstatt M_n, weil das sind dann genau die Mittelwerte aus Gleichung (1) oben)

$$S_N \;:=\; \tfrac{1}{N}\bigl(X_1 + X_2 + \cdots + X_N\bigr) \tag{3}$$

die dann Erwartungswert und Varianz

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

haben. Wir definieren die normierten Grössen

$$Z_N \;:=\; \tfrac{\sqrt{N}}{\sigma}\,\bigl(\,S_N \,-\, \mu\,\bigr)$$

Dann gilt

$$\lim_{N\to\infty} \mathsf{Prob}\bigl[\,Z_N \in (x, x+dx)\,\bigr] \;=\; \tfrac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}\, dx \;\; . \tag{4}$$

Aus der Gleichung (4) folgt:

$$\begin{aligned} \mathsf{Prob}\bigl[\,|Z_N| \,\geq\, \alpha\,\bigr] \;&\overset{N\to\infty}{=}\; \int_{-\infty}^{-\alpha} \tfrac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}\, dx \;+\; \int_{\alpha}^{\infty} \tfrac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}\, dx \\ &=\; 2\int_{-\infty}^{-\alpha} \tfrac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}\, dx \;=\; 2\,\Phi(-\alpha) \end{aligned} \tag{5}$$

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

$$\Phi(x) \;:=\; \int_{-\infty}^{x} \tfrac{1}{\sqrt{2\pi}}\, e^{-\frac{y^2}{2}}\, dy \;\; . \tag{6}$$

Gleichung (5) können wir auch so schreiben:

$$\mathsf{Prob}\Bigl[\,\bigl|\tfrac{\sqrt{N}}{\sigma}\,\bigl(\,S_N \,-\, \mu\,\bigr)\bigr| \,\geq\, \alpha\,\Bigr] \;\overset{N\to\infty}{=}\; 2\,\Phi(-\alpha)$$

oder

$$\mathsf{Prob}\Bigl[\,\bigl|\bigl(\,S_N \,-\, \mu\,\bigr)\bigr| \,\geq\, \tfrac{\sigma}{\sqrt{N}}\,\alpha\,\Bigr] \;\overset{N\to\infty}{=}\; 2\,\Phi(-\alpha) \tag{7}$$

Als unabhängige, identisch verteilte Zufallsvariablen wählen wir jetzt

$$X_i \;:=\; F(\phi_i) \tag{8}$$

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

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

übereinstimmen. Wir haben dann

$$\mu \;=\; \mathsf{E}\big[\,F(\phi_i)\,\big] \;=\; \mathsf{E}\big[\,F(\phi)\,\big] \;=\; \textstyle\int_{\mathbb{R}^d} F(\phi)\,p(\phi)\,d^d\phi \;=\; I \tag{10}$$

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

$$\sigma^2 \;=\; \mathsf{V}[F] \;=\; \mathsf{E}[F^2] - \mathsf{E}[F]^2 \;=\; \textstyle\int_{\mathbb{R}^d} F(\phi)^2 p(\phi)\,d^d\phi \;-\; \big(\int_{\mathbb{R}^d} F(\phi)\,p(\phi)\,d^d\phi\big)^2$$

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

$$\mathsf{Prob}\Big[\;\big|\,S_N(F) - \mathsf{E}[F]\,\big| \;\geq\; \alpha\sqrt{\tfrac{\mathsf{V}[F]}{N}}\;\Big] \;\;\stackrel{N\to\infty}{=}\;\; 2\,\Phi(-\alpha) \tag{11}$$

Schauen wir uns Gleichung (11) genauer an. Das $\mathsf{E}[F]$ 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 $\mathsf{E}[F]$ 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

$$S_N(F) \;-\; 2.58\sqrt{\tfrac{\mathsf{V}[F]}{N}} \;\leq\; \mathsf{E}[F] \;\leq\; S_N(F) \;+\; 2.58\sqrt{\tfrac{\mathsf{V}[F]}{N}} \tag{12}$$

und mit einer Wahrscheinlichkeit von 99.99% ist

$$S_N(F) \;-\; 3.89\sqrt{\tfrac{\mathsf{V}[F]}{N}} \;\leq\; \mathsf{E}[F] \;\leq\; S_N(F) \;+\; 3.89\sqrt{\tfrac{\mathsf{V}[F]}{N}} \tag{13}$$

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

$$\text{Monte Carlo Fehler} \;\sim\; \sqrt{\frac{\mathsf{V}[F]}{N}} \tag{14}$$

Schauen wir uns jetzt ein paar Beispiele an:

Beispiele:

Beispiel 1) Im Kapitel 3 hatten wir das Volumen τ_n der n-dimensionalen Einheitskugel berechnet und hatten die folgende Formel gefunden:

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

Überprüfen Sie die Formel (15) mit einer Monte Carlo Simulation für, sagen wir,

$$n \;\in\; \{2,3,4,5,6,7,8,9,10\}$$

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

$$F(x) \;:=\; 2^n\, \chi(\,\|x\| \leq 1) \tag{16}$$

und der W'keitsdichte

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

Sind also

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

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:

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

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

b1) mit einer Monte Carlo Simulation, indem Sie exponential-verteilte Zufallszahlen benutzen

b2) mit einer Monte Carlo Simulation, indem Sie auf dem Intervall $[0, \log 4]$ gleichverteilte Zufallszahlen benutzen

c) Plotten Sie den Monte Carlo Fehler und die skalierten MC-Fehler

$$\big|\,I \,-\, S_N(F)\,\big|\,, \qquad \sqrt{N}\,\big|\,I \,-\, S_N(F)\,\big|\,, \qquad N\,\big|\,I \,-\, S_N(F)\,\big|$$

als Funktion von N.

Was ist jeweils das F und das p in (b1) und (b2) ?

Lösung 2) a) Das exakte Resultat ist

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

b1) Wir schreiben

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

mit dem Integranden

$$F(x) \;:=\; 2 \,\times\, \chi(\,0 \le x \le \log 4\,) \tag{19}$$

und der W'keitsdichte

$$p(x) \;:=\; \tfrac{1}{2}\,e^{-x/2}\;\chi(x \ge 0) \tag{20}$$

Sind also $x_1, x_2, \cdots, x_N$ 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 (19).

b2) Wir schreiben

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

mit dem Integranden

$$F(x) \;\; := \;\; \log 4 \; \times \; e^{-x/2} \tag{21}$$

und der W'keitsdichte

$$p(x) \;\; := \;\; \frac{\chi(0 \le x \le \log 4)}{\log 4} \tag{22}$$

Sind also $x_1, x_2, \cdots, x_N$ 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 (21).

Ok, schauen wir uns die Sachen jetzt in R an.