TU Chemnitz – Professur Numerische Mathematik
Prof. Dr. Oliver Ernst
Dipl.-Math. Björn Sprungk

19. April 2016

Mathematische Methoden der Unsicherheitsquantifizierung

Sommersemester 2016

2. Übung: Die Monte Carlo-Methode und Konfidenzintervalle

Aufgabe 1

Es sei $(X_n)_{n\in\mathbb{N}}$ eine *i.i.d.* Folge reellwertiger Zufallsvariablen mit Mittelwert μ und Varianz σ^2 und

$$\widehat{\sigma}_N^2 = \frac{1}{N-1}\sum_{n=1}^{N}(X_n - \widehat{\mu}_N)^2, \qquad \widehat{\mu}_N^2 = \frac{1}{N}\sum_{n=1}^{N} X_n.$$

Zeigen Sie, dass

(a) $\widehat{\sigma}_N^2 = \frac{1}{N-1}\sum_{n=1}^{N}(X_n - \mu)^2 - \frac{N}{N-1}(\mu - \widehat{\mu}_N)^2$,

(b) $\widehat{\sigma}_N^2$ ein erwartungstreuer Schätzer für σ^2 ist, also $\mathbf{E}\left[\widehat{\sigma}_N^2\right] = \sigma^2$ gilt,

(c) $\widehat{\sigma}_N^2 \xrightarrow{\mathbf{P}-f.s.} \sigma^2$ für $N \to \infty$ gilt.

Aufgabe 2

Neben asymptotischen Konfidenzintervallen basierend auf dem Zentralen Grenzwertsatz kann man auch die Tschebyscheff-Ungleichung zur Berechnung von Konfidenzgebieten nutzen:

$$\mathbf{P}(|X - \mathbf{E}[X]| \geq c) \leq \frac{\mathrm{Var}(X)}{c^2}, \qquad X \in L^2(\Omega;\mathbb{R}),\ c \in \mathbb{R}.$$

Diese Ungleichung hat den Vorteil, dass sie nicht auf Asymptotiken beruht, sie liefert i. A. aber auch gröbere Intervalle.
Wie würde ein Konfidenzintervall für die Monte Carlo Simulation zum Niveau $1-\alpha$ basierend auf der Tschebyscheff-Ungleichung lauten? Vergleichen Sie es mit dem asymptotischen Konfidenzintervall für $\alpha = 0.1$, 0.05 und 0.01.

Aufgabe 3

Ist das Basiszufallsexperiment beschränkt, d.h., gilt für die Zufallsvariable X, dass $X \in [a,b]$ **P**-fast sicher, so können genauere nicht-asymptotische Schranken mittels der Hoeffding-Ungleichung angegeben werden.

Theorem 1 (Hoeffding-Ungleichung)**.** *Es seien X_n i.i.d. Zufallsvariablen für $n \in \mathbb{N}$ mit $X_n \in [0,1]$ **P**-fast sicher und $\mathbf{E}[X_n] = \mu \in (0,1)$. Dann gilt für $\widehat{\mu}_N = \frac{1}{N}\sum_{n=1}^{N} X_n$, dass*

$$\mathbf{P}\left(\widehat{\mu}_N - \mu \geq \varepsilon\right) \leq \exp(-2N\varepsilon^2).$$

(a) Wie lassen sich aus der Hoeffding-Ungleichung Konfidenzintervalle für $\widehat{\mu}_N$ konstruieren, wenn X_n i.i.d. mit $X_n \in [a,b]$ **P**-fast sicher?

(b) Diskutieren Sie den Hauptunterschied zwischen diesem Resultat und dem entsprechenden Resultat, welches aus der Tschebyscheff-Ungleichung abgeleitet wurde.

Aufgabe 4

Es sei X eine Zufallsgröße mit reellen Werten. Wir wollen die Wahrscheinlichkeit p des Ereignisses $\mathbf{P}(X \leq c)$ für einen bestimmten Wert $c \in \mathbb{R}$ mittels Monte Carlo-Simulationen (bzw. Samples) von X schätzen.

(a) Wie sieht der entsprechende Monte Carlo-Schätzer aus?

(b) Geben Sie eine a priori-Abschätzung für die nötige Anzahl N der Samples, damit die Standardabweichung des Monte Carlo Fehlers zur Berechung von $\mathbf{P}(X \leq c)$ kleiner gleich ε ist.

(c) Wie ändert sich die Abschätzung, wenn die Standardabweichung kleiner gleich $\varepsilon \cdot \mathbf{P}(X \leq c)$ sein soll? Diskutieren Sie das Ergebnis.

(d) Nutzen Sie die Hoeffding-Ungleichung, um zu beweisen, dass, falls $\mathbf{P}(X \leq c) = \frac{1}{2} + \beta$ für $\beta \in (0, \frac{1}{2})$ die Wahrscheinlichkeit, dass wir $X \leq c$ in mehr als der Hälfte aller N Samples beobachten, mindestens $1 - \exp(-N\beta^2)$ ist.

Aufgabe 5

Berechnen Sie mittels Monte Carlo Simulation die durchschnittliche Fläche A_μ eines zufälligen Dreiecks, dessen Eckpunkte einer Gleichverteilung auf dem Einheitsquadrat folgen.

(a) Führen Sie zunächst eine Simulation mit $M = 10^3$ Samples durch und geben Sie zum Konfidenzniveau $\alpha = 0.95\%$

- ein asymptotisches Konfidenzintervall,
- ein Konfidenzintervall mittels Tschebyscheff-Ungleichung und
- ein Konfidenzintervall mittels Hoeffding-Ungleichung

an und quantifizieren Sie zudem den Fehler zwischen asymptotischer und tatsächlicher Verteilung des Monte Carlo-Fehlers mittels der Berry-Esseen-Ungleichung.

(b) Führen Sie nun Monte Carlo Simulationen für $M_j = 2^j\, M_0$, $j = 0, \ldots, 10$ und $M_0 = 100$, durch und stellen Sie die entsprechenden, asymptotischen Konfidenzintervalle sowie den exakten Wert $A_\mu = 11/144$ grafisch dar.

(c) Wie könnten Sie ein zufälliges Dreieck simulieren, dessen Eckpunkte einer Gleichverteilung auf dem Einheitskreis folgen? Testen Sie mittels Monte Carlo-Simulation die Hypothese, dass in diesem Fall der mittlere Flächeninhalt $\frac{3}{2\pi}$ ist.

Hinweis: Nutzen Sie zur Flächenberechnung das Kreuzprodukt (in MATLAB: `cross(x,y)`).