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

11. Juli 2016

Mathematische Methoden der Unsicherheitsquantifizierung

Sommersemester 2016

Exercise Sheet 2: Monte Carlo Method and Confidence Intervals

Exercise 1

Let $(X_n)_{n\in\mathbb{N}}$ be a series of *i.i.d.* real-valued random variables with mean μ and variance σ^2 and

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

Show that

(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) $\widehat{\sigma}_N^2$ is an unbiased estimator of σ^2, i.e. there holds $\mathbf{E}\left[\widehat{\sigma}_N^2\right] = \sigma^2$,

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

Exercise 2

Besides the central limit theorem the Tschebyscheff inequality can be used to derive confidence intervals:

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

This inequality does not depend on asymptotic behavior but only leads to looser bounds compared to the central limit theorem.
What form will confidence intervals of the Monte Carlo method at level $1-\alpha$ based on the Tschebyscheff inequality take? Compare these results with asymptotic confidence intervals for $\alpha = 0.1$,0.05 and 0.01.

Exercise 3

If the random experiment is bounded, i.e. the random variable X fulfills $X\in[a,b]$ **P**-almost surely, then tighter non-asymptotic bounds can be derived from Hoeffding inequality.

Theorem 1 (Hoeffding Inequality)**.** *Let X_n be a series of i.i.d. random variables for $n\in\mathbb{N}$, $X_n\in[0,1]$ **P**-a.s. and $\mathbf{E}[X_n]=\mu\in(0,1)$. Then there holds for $\widehat{\mu}_N = \frac{1}{N}\sum_{n=1}^{N}X_n$*

$$\mathbf{P}\left(\widehat{\mu}_N - \mu \geq \varepsilon\right) \leq \exp(-2N\varepsilon^2) \qquad \forall\epsilon > 0.$$

(a) How can you derive confidence intervals for $\widehat{\mu}_N$ using the Hoeffding inequality if X_n is a series of i.i.d random variables with $X_n\in[a,b]$ **P**-a.s.?

(b) Discuss the difference of the previous main result with the result of the Tschebyscheff inequality.

Exercise 4

Let X be a real-valued random variable. We want to compute the probability p of $\mathbf{P}(X \leq c)$ for a specific $c \in \mathbb{R}$ with Monte Carlo simulation.

(a) What is the Monte Carlo estimator?

(b) Give an a priori lower bound for the required sample size N, such that the standard deviation of the Monte Carlo error of $\mathbf{P}(X \leq c)$ is less than or equal to ε.

(c) How does this lower bound change if the standard deviation should be bound by $\varepsilon \cdot \mathbf{P}(X \leq c)$? Discuss this result.

(d) Use the Hoeffding inequality to prove that the probability of $X \leq c$ in more than half of all N realizations is at least $1 - \exp(-N\beta^2)$ if $\mathbf{P}(X \leq c) = \frac{1}{2} + \beta$ for $\beta \in (0, \frac{1}{2})$.

Exercise 5

Compute the average area A_μ of a random triangle which points are uniformly distributed inside the unit square with Monte Carlo simulation.

(a) Perform a simulation with $M = 10^3$ samples and for the confidence level $\alpha = 0.95\%$ derive

- the asymptotic confidence interval,
- the confidence interval derived from Tschebyscheff inequality and
- the confidence interval derived from Hoeffding inequality.

Furthermore, quantify the error between the asymptotic and real distribution of the Monte Carlo error by using the Berry-Esseen inequality.

(b) Perform Monte Carlo simulations with $M_j = 2^j\, M_0$, $j = 0, \ldots, 10$ and $M_0 = 100$ samples and plot the resulting asymptotic confidence intervals and the correct result $A_\mu = 11/144$.

(c) How can you simulate a random triangle with points uniformly distributed on the unit sphere? Use Monte Carlo simulation to test the hypothesis that the average area is $\frac{3}{2\pi}$.

Tip: Use the cross product to calculate the area of the triangle (in MATLAB: `cross(x,y)`).