Rechnernutzung in der Physik

Institut für Experimentelle Teilchenphysik
Institut für Theoretische Teilchenphysik

Prof. Dr. G. Quast, Prof. Dr. M. Steinhauser
Dr. A. Mildenberger, Dr. Th. Chwalek
http://comp.physik.kit.edu

WS2018/19 – Blatt 08
Prog. bis Di., 15.01.2019

Anwendungen der Monte Carlo-Methode

Aufgabe 15: Zufallszahlen mit beliebiger Verteilungsdichte * — für Bachelor

Ausgehend von im Intervall $[0, 1)$ gleichverteilten Zufallszahlen (`gRandom->Rndm()` in `ROOT` oder `numpy.random.rand()` in `Python`), sollen Zufallszahlen erzeugt werden, die anderen Verteilungsdichten folgen.

Hinweis: Zu dieser Aufgabe gibt es wieder Vorlagen, `PDFs-pyroot.py` bzw. `PDFs.C`[1]. Da nur die Histogramm-Funktionen aus `ROOT` benötigt werden, können Sie die Struktur der `pyroot`-Vorlage auch für Lösungen in `python` verwenden.

a) Streuwinkelverteilung

Erzeugen Sie 10'000 Zufallszahlen mit der Verwerfungsmethode gemäß der Verteilungsdichte

$$f(x) = \frac{3}{8}(1 + x^2) \quad (-1 \leq x \leq 1)$$

und füllen Sie die Zufallszahlen in ein Histogramm.
Schreiben Sie eine allgemeine C++- oder Python-Funktion zur Implementierung der Verwerfungsmethode.

b) Cauchy-Verteilung

Verwenden Sie die Transformationsmethode und füllen Sie ein Histogramm mit 10'000 Zufallszahlen, die einer Cauchy-Verteilungsdichte[2] folgen, die gegeben ist durch

$$f(x) = \frac{1}{\pi}\frac{1}{1 + x^2}\,.$$

c) Zufallszahlen gemäß empirischer Verteilung aus Histogramm

In der Datei „`elefant.dat`“ finden Sie die Einträge in 64 Bins einer vorgegebenen Verteilung. Erzeugen und histogrammieren Sie 10'000 Zufallszahlen, die dieser Verteilung folgen. Beispielcode zum Einlesen der Datei in `C++` oder `Python` ist in den Vorlagen enthalten.

d) Majoranten-Methode („Importance Sampling“) — freiwillig

Füllen Sie ein Histogramm mit 10'000 Zufallszahlen, die gemäß

$$f(x) = \frac{5}{2}\,\sin^2(\pi x)\,\exp(-x),\; 0 \leq x \leq \infty$$

verteilt sind und die wir im Intervall $[0, 2\pi)$ betrachten wollen. Verwenden Sie exponentiell verteilte, mit der Transformationsmethode erzeugte Zufallszahlen als Majorante, $m(x) \propto \exp(-x)$. Sie können die im Aufgabenteil a) erstellte Funktion wiederverwenden bzw. geeignet erweitern, um statt der gleichverteilten Zufallszahlen exponentiell verteilte zu verwenden.

---

[1] Falls Sie in C++ arbeiten und Ihr C++-Makro übersetzen wollen, verwenden Sie „`make SOURCE=PDFs`“.
[2] Bei Physikern ist die Cauchy-Verteilung auch als Lorentz- oder Breit-Wigner-Verteilung bekannt.

Aufgabe 15L: Volumen von Hyper-Kugeln * für Lehramt

Mit Hilfe der Monte Carlo-Methode können recht einfach hochdimensionale, bestimmte Integrale näherungsweise numerisch berechnet werden. In dieser Aufgabe soll das Volumen V_d von Hyper-Kugeln in Abhängigkeit von der Anzahl der Dimensionen d berechnet werden.

Als Basis schauen Sie sich die Scripte `pi.py` aus den Beispielen zur Vorlesung an: darin wird ein einfaches Verfahren zur Bestimmung des Verhältnisses der Flächen eines Viertelkreis und des Einheitsquadrats benutzt: zufällig gewählte Punkte werden auf der Fläche verteilt, und der Anteil der Punkte innerhalb des Kreises wird gezählt.

a) Erweitern Sie diesen Code, so dass Sie mit dem Verfahren das d-dimensionale Kugelvolumen für $d \in [1, 10]$ berechnen können. Arbeiten Sie zunächst mit 10'000 gleichmäßig im Hyperkubus mit Kantenlänge 1 verteilten Punkten. Berechnen Sie aus dem so bestimmten Volumen des Kugelsegments im Einheits-Hyperkubus das Volumen der vollen Hyperkugel.

b) Bestimmen Sie jeweils die statistische Unsicherheit σ_{V_d} der Ergebnisse. Nutzen Sie dazu aus, dass bei N Versuchen die Anzahl N_{in} der innerhalb der Kugel liegenden Punkte einer Binomialverteilung $B(N_{in}; N, p)$ folgt. Die Unsicherheit ergibt sich aus der bekannten Varianz der Binomialverteilung mit der Näherung $p \simeq \hat{p} = N_{in}/N$.

c) Tragen Sie Ihre Ergebnisse für V_d und die Unsicherheit σ_{V_d} in ein Diagramm ein. Die Vorlage `VKugel.py` enthält die Funktion `VolShere(d, r=1.)`, die das mit mathematischen Integrationsmethoden berechnete Volumen von Hyperkugeln angibt. Tragen Sie auch dieses „theoretische" Ergebnis in Ihre Grafik ein.

Aufgabe 16: Vergleich von zwei Verteilungen mit der Bootstrap-Methode *

In dieser Aufgabe sollen zwei Stichproben S_1 und S_2 der Größen n_1 und n_2 darauf hin untersucht werden, ob sie statistisch signifikant unterschiedliche Mittelwerte haben. Die Stichproben finden Sie in den Dateien `sample1.dat` und `sample2.dat`.

Zur Quantifizierung des Unterschieds berechnet man eine Prüfgröße `d`, die auf einen möglichen Unterschied der beiden Proben empfindlich ist, typischerweise den Betrag der Differenz der Mittelwerte μ_1 und μ_2, die man auf die erwartete statistische Unsicherheit normiert:
$d = |\mu_1 - \mu_2|/\sigma$ mit $\sigma = \sqrt{{\sigma_1}^2/n_1 + {\sigma_2}^2/n_2}$.

Mit Hilfe der Bootstrap-Methode lässt sich die erwartete Verteilung von d bestimmen. Zum Test auf Gleichheit wird ein Hypothesentest durchgeführt, bei dem man annimmt, dass die beiden Stichproben aus der gleichen Grundgesamtheit stammen, und bestimmt die Verteilung von d unter dieser Annahme. Der in den ursprünglichen Stichproben beobachtete Wert wird dann mit dieser Verteilung verglichen, d. h. man bestimmt die Wahrscheinlichkeit, auf Grund von statistischen Fluktuationen einen größeren Wert von d zu erhalten als den beobachteten Wert d_{obs}, den sogenannte p-Wert.

Implementieren Sie eine Funktion zur Berechnung der oben definierten Größe d. Bestimmen Sie dann mit der Bootstrap-Methode die Verteilung von d, indem Sie aus der Vereinigungsmenge der beiden Stichproben S_1 und S_2 durch zufälliges Ziehen mit Zurücklegen 10'000 neue Probenpaare S_1' und S_2' der Längen n_1 und n_2 erzeugen. Berechnen Sie für jedes Probenpaar die Prüfgröße d_i für $i = 1, \ldots,$10'000 und zählen Sie, wie oft sich Werte $d_i > d_{obs}$ ergeben. Sind die Proben statistisch signifikant unterschiedlich?

Hinweis: Wenn Sie die ursprünglichen Proben und die durch Bootstrap gewonnenen Verteilung von d grafisch darstellen, bekommen Sie einen sehr klaren Eindruck von der Methodik.

Anmerkung: Sie haben eine Verallgemeinerung des „Student'schen t-Tests" implementiert. Beim eigentlichen t-Test wird zusätzlich davon ausgegangen, dass es sich bei den zu testenden Verteilungen um Gaußverteilungen mit gleicher Standardabweichung handelt.

Hinweis: Mit dem Rechnernamen `fphctssh.physik.uni-karlsruhe.de` können Sie von überall aus mittels ssh/scp Programm per Netzwerk auf einen Poolrechner zugreifen.
