KIT-Fakultät für Physik
Institut für Experimentelle Teilchenphysik
Institut für Theorie der Kondensierten Materie
Prof. Dr. Ulrich Husemann (ETP)
Dr. Nils Faltermann (ETP)
Dr. Andreas Poenicke (TKM)

Praktikum: Computergestützte Datenauswertung

Sommersemester 2021

Übungsblatt Nr. 3 Bearbeitung bis: 12.05., 18 Uhr

Aufgabe 3.1: Berechnung statistischer Größen

In dieser Aufgabe implementieren Sie die Berechnung einiger der statistischen Größen aus der Vorlesung.

Sie haben ein Jupyter-Notebook (`B3A1_start.ipynb`) als Vorlage und eine Datei mit Daten (`numbers.dat`), letztere müssen Sie nach dem Herunterladen von ILIAS wieder umbenennen. In der Vorlage werden aus der Datei die Daten (ganze Zahlen im Bereich 0 bis 9) eingelesen und dann mit Hilfe von `numpy`-Funktionen die statistischen Größen Mittelwert, Varianz und Standardabweichung berechnet und auf dem Bildschirm ausgegeben.

a) Implementieren Sie Ihre eigene Version zur Berechnung von Mittelwert, Varianz, Standardabweichung und Stichprobenvarianz, indem Sie die vorbereiteten Definitionen der Funktionen in der Vorlage ergänzen. Vergleichen Sie Ihre Ergebnisse mit denen, die mit Hilfe der `numpy`-Funktionen in der Vorlage berechnet wurden.

b) Bestimmen Sie nun die Häufigkeit der einzelnen Zahlen in der Datei `numbers.dat`, d. h. füllen Sie einen `numpy array` mit der jeweiligen Häufigkeit, mit der die Zahlen 0 bis 9 vorkommen. Dieses `array` enthält nun statt der ursprünglichen Datenmenge nur noch 10 Zahlen, die eine erheblich komprimiertere Version der ursprünglichen Daten darstellt und außerdem einen sehr viel besseren Überblick erlaubt.

c) Berechnen Sie nun den Mittelwert und die Varianz der Daten direkt aus der Häufigkeitsverteilung. Vergleichen Sie mit dem Ergebnis, das Sie in Aufgabenteil a) erhalten haben.

Aufgabe 3.2: Würfelspiel

Wir wollen in dieser Aufgabe die Überlegungen aus Aufgabe 2.2 vom Übungsblatt 2 weiterführen, eine entsprechende Vorlage `B3A2_start.ipynb` steht für diese Aufgabe bereit. Im Folgenden spielen wir jedoch nicht nur mit einem Würfel, sondern werfen jedes Mal n Würfel gleichzeitig. Dabei betrachten wir jetzt die Augensumme bzw. mittlere Augenzahl.

a) Zunächst spielen wir mit $n = 3$ Würfeln und bei jedem Spiel wird die Summe S der Augenzahlen bestimmt. Führen Sie $N = 1000$ solcher Spiele durch und stellen Sie die Häufigkeitsverteilung der Augensumme S grafisch dar.

b) Verallgemeinern Sie Ihren Code so, dass Sie auch mit einer beliebigen Zahl von n Würfeln gleichzeitig spielen können. Schreiben Sie dazu eine Funktion `wuerfele(n)`, die die mittlere Augenzahl $\bar{s} = S/n$, d.h die Summe S der Augenzahlen geteilt durch die Zahl der Würfel, zurück gibt.

   Stellen Sie die Häufigkeitsverteilung der mittleren Augenzahl für $n = 10$, $n = 50$ und $n = 100$ für jeweils 1000 Spiele dar.

   Hinweis: Zur Darstellung der Häufigkeit sollten Sie die Funktion `hist()` aus `matplotlib.pyplot` verwenden. Beachten Sie, dass `hist(data, b)` als Rückgabewerte die Häufigkeiten in jedem der b

Intervalle („Bins"), die Bingrenzen und die — hier nicht weiter verwendeten — Grafikobjekte der Darstellung zurück gibt.

c) Bestimmen Sie für $n = 100$ den Mittelwert und die Varianz der mittleren Augenzahl. Nutzen Sie dabei die Rückgabewerte von `hist()`. Zeichnen Sie zum Vergleich eine Gauß-Verteilung mit diesem Mittelwert der Varianz und passender Normierung, in die Grafik Ihrer Verteilung ein. Was beobachten Sie?

Aufgabe 3.3: Zentraler Grenzwertsatz

Betrachten Sie die Summe y von n gleichverteilten Zufallszahlen x_i im Intervall [0,1], $y = \sum_{i=1}^{n} x_i$. Verwenden Sie für diese Aufgabe die Vorlage `B3A3_start.ipynb`.

a) Zeigen Sie analytisch, dass eine Gleichverteilung im Intervall [0,1] eine Standardabweichung von $\sigma = \sqrt{1/12}$ hat. Zeigen Sie weiter, dass y einen Erwartungswert von $n/2$ und eine Varianz von $n/12$ hat.

   Verwenden Sie zur Darstellung Ihrer Herleitung ebenfalls ein Jupyer-Notebook, jedoch mit der Zellenoption "Markdown" statt dem sonst üblichen "Code"-Modus. Dieser Modus erlaubt Ihnen die Darstellung Ihrer Rechnung mittels LaTeX. In der Vorlage finden Sie alle benötigten Befehle zur Darstellung der mathematischen Operationen.

b) Nach dem zentralen Grenzwertsatz nähert sich die Verteilung von y für große n einer Gauß-Verteilung an. Dies soll nun überprüft werden. Wir betrachten dazu die Variable

$$z = \frac{(\sum_{i=1}^{n} x_i) - \frac{n}{2}}{\sqrt{n/12}} \tag{1}$$

   mit Mittelwert $\mu = 0$ und Standardabweichung $\sigma = 1$.

   Schreiben Sie ein Programm um für $n = 2$, 10 und 100 jeweils 100 000 Werte von z zu erzeugen und füllen Sie diese in Histogramme. Zeichnen Sie jeweils die Standardnormalverteilung (d.h normierte Gaußverteilung mit $\mu = 0$, $\sigma = 1$) in Ihre Histogramme ein und vergleichen Sie. Achten Sie auf eine ausreichende Anzahl an Bins, damit der Vergleich aussagekräftig ist.

c) Wiederholen Sie Aufgabenteil b) mit einer Summe von exponentiell verteilten Zufallszahlen $y = \sum_{i=1}^{n} \lambda \exp(-\lambda x_i)$ für den Fall $\lambda = 1$. Verwenden Sie die verallgemeinerte Form des zentralen Grenzwertsatzes

$$z = \frac{y - E[x]}{\sqrt{V[x]}}\,, \tag{2}$$

   mit dem Erwartungswert $E[x]$ und der Varianz $V[x]$ der Zufallszahlenverteilung. Erhalten Sie das gleiche Ergebnis wie in Aufgabenteil b) oder gibt es Unterschiede?