

Übungsblatt: Der Zentrale Grenzwertsatz

DSA Rossleben – Kurs 5.1

Bisher haben wir vor allem diskrete Verteilungen mit wenigen, klar unterscheidbaren Werten betrachtet (den Würfel, die Binomial- und die Poissonverteilung). Was passiert, wenn eine solche diskrete Zufallsgröße immer mehr mögliche Werte annehmen kann? Dieser Frage nähern wir uns über Computersimulationen – und stoßen dabei auf den vielleicht wichtigsten Satz der gesamten Wahrscheinlichkeitstheorie: den **zentralen Grenzwertsatz (ZGS)**.

Aufgabe 13.7

Wir untersuchen, was mit der Binomialverteilung $B(k; n, p)$ passiert, wenn n immer größer und p *nicht* kleiner wird.

- a) Wähle $p = 0.3$ fest. Erzeuge für $n = 10, 50, 200, 1000$ jeweils mit `np.random.binomial(n, p, size=10000)` einen Datensatz von 10 000 Stichproben. Plote für jedes n das normierte Histogramm (`density=True`), am besten in vier Teilplots (`plt.subplots`) nebeneinander. Was fällt dir an der *Form* der Histogramme auf, je größer n wird?
- b) Die vier Histogramme aus a) sind auf sehr unterschiedlichen k -Achsen nicht direkt vergleichbar, da sowohl Lage als auch Breite der Verteilung von n abhängen (erinnere dich: $\langle k \rangle = np$ und $\sigma_k^2 = np(1 - p)$). Um die *Form* unabhängig von n vergleichen zu können, *standardisiere* jede Stichprobe k gemäß

$$z = \frac{k - \langle k \rangle}{\sigma_k} = \frac{k - np}{\sqrt{np(1 - p)}}. \quad (1)$$

Plote nun die normierten Histogramme der standardisierten Werte z für alle vier n *gemeinsam in einem einzigen Plot* (z. B. mit `alpha=0.5` und unterschiedlichen Farben, damit man sie unterscheiden kann). Was beobachtest du für wachsendes n ?

- c) Für $n = 10$ kann k nur 11 verschiedene Werte annehmen, für $n = 1000$ dagegen 1001 verschiedene Werte – auf demselben (standardisierten) Bereich, den die Verteilung im Wesentlichen abdeckt. Die „Lücken“ zwischen den möglichen Werten von z werden also immer kleiner. Im Grenzfall $n \rightarrow \infty$ liegt schließlich *jeder* reelle Wert im Bereich der Verteilung – z wird zu einer **kontinuierlichen** Zufallsgröße, beschrieben durch eine **Wahrscheinlichkeitsdichte** statt einzelner Wahrscheinlichkeiten $P(k)$. Die Grenzkurve, der sich die standardisierten Binomialverteilungen annähern, ist die **Standardnormalverteilung**

$$\mathcal{N}(x; 0, 1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right). \quad (2)$$

Ergänze deinen Plot aus b) um diese Kurve.

- d) Der ZGS gilt nicht nur für Summen von Bernoulli-Variablen (also für die Binomialverteilung), sondern für Summen *beliebiger* unabhängiger, identisch verteilter Zufallsgrößen mit endlicher Varianz – ganz gleich, wie diese einzeln verteilt sind! Überprüfe das:

- Ziehe mit `np.random.uniform(0, 1, size=(10000, n))` jeweils 10 000 Stichproben von n gleichverteilten Zufallszahlen und bilde deren Summe S_n (Summation über die richtige Achse).

- Standardisiere die Stichproben von S_n (du benötigst dafür Mittelwert $1/2$ und Varianz $1/12$ einer einzelnen $\mathcal{U}(0, 1)$ -Zufallsgröße).
 - Plote das normierte Histogramm für $n = 1, 2, 5, 30$ zusammen mit $\mathcal{N}(x; 0, 1)$.
- e) Der ZGS ist einer der Hauptgründe, warum die Normalverteilung in der Physik – und insbesondere bei der Beschreibung von **Messunsicherheiten** – eine so herausragende Rolle spielt. Die Grundidee: Eine einzelne Messung (z. B. die Länge eines Tisches mit einem Lineal) wird praktisch nie durch *eine* einzige Fehlerquelle verfälscht, sondern durch eine *Überlagerung vieler kleiner, voneinander unabhängiger Störungen* – Ablesefehler, kleine Temperaturschwankungen, Erschütterungen, Ungenauigkeiten im Material des Lineals, und so weiter. Begründe in eigenen Worten, weshalb der ZGS ein gutes Argument dafür liefert, warum zufällige Messunsicherheiten in der Physik häufig durch eine Normalverteilung modelliert werden – auch wenn man die einzelnen, elementaren Fehlerquellen einer Messung oft gar nicht genau kennt. In welchen Situationen könnte diese Rechtfertigung *nicht* gut funktionieren?