Dr. G. Eichner
Math. Inst.
JLU Gießen

Übungen zu

R1: Grundlagen der Datenanalyse mit R

Blatt 10

SoSe 2024
4. 7. 2024
Abgabe: ≤17. 7. 2024,
14:00 Uhr

1. Realisieren Sie das der Aufgabe 4 auf Blatt 9 Entsprechende für die Wahrscheinlichkeitsfunktionen (!) verschiedener Binomialverteilungen und für deren zugehörige Verteilungsfunktionen! Beachten Sie, dass jene Wahrscheinlichkeitsfunktionen nur auf $\{0, \ldots, n\}$ verschieden von 0 (Null) sind und daher typischerweise als „Stabdiagramme“ veranschaulicht werden sowie dass die zugehörigen Verteilungsfunktionen Treppenfunktionen sind!

   Beides sorgt übrigens dafür, dass es *nicht* sinnvoll ist, die Funktionen zu verschiedenen Verteilungsparametern in ein und dasselbe Koordinatensystem zeichnen zu lassen. (Beachten Sie auch hier die Hinweise zu Aufgabe 3 auf Blatt 9.)

2. Erzeugen Sie sich aus der Normalverteilung für vier von Ihnen selbst gewählte, verschiedene Parameterkombinationen für Erwartungswert und Varianz je eine Stichprobe von $n = 100$ (Pseudo-)Zufallszahlen! Zur grafischen Veranschaulichung der Wirkung der verschiedenen Parameterwerte auf die Verteilung der Zufallszahlen erstellen Sie für jede der vier Stichproben …

   a) ein Histogramm der Daten,

   b) einen Boxplot, wobei diese Boxplots in einem gemeinsamen Koordinatensystem nebeneinander angeordnet sein sollen, sowie

   c) einen Normal-Q-Q-Plot mit überlagerter Soll-Linie.

3. Vergleichen Sie auf dieselbe Art wie in Aufgabe 2 nun *verschiedene* Verteilungen miteinander, und zwar Cauchy-, Exponential-, Log-Normal-, Student's t- und die uniforme Verteilung sowie eine Mischverteilung aus zwei unabhängigen Normalverteilungen, indem Sie aus jeder dieser Verteilungen *eine* Stichprobe vom Umfang $n = 100$ erzeugen und diese Stichproben darstellen wie in Aufgabe 2. Die Parameter der obigen Verteilungen können Sie frei wählen.

   Hinweis: Eine Zufallsvariable Z mit einer Mischverteilung aus zwei unabhängigen Normalverteilungen erhält man z. B., indem drei unabhängige Zufallsvariablen $X \sim \mathcal{N}(\mu_1, \sigma_1^2), Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ und $B \sim \text{Bernoulli}(p)$ erzeugt werden und Z durch $Z := B \cdot X + (1 - B) \cdot Y$ generiert wird. D. h., $Z = X$, falls $B = 1$ und $Z = Y$, falls $B = 0$, wobei $\mathbb{P}(B = 1) = p$. (Memo: $\text{Bernoulli}(p) = \text{Bin}(1, p)$)