Technische Universität Dortmund
Fakultät Statistik
Lehrstuhl Computergestützte Statistik

Uwe Ligges
Marieke Stolte
Luca Sauer
Rudi Zulauf

Übung zur Vorlesung

Computergestützte Statistik

Wintersemester 2022/2023

Übungsblatt Nr. 8

Abgabe ist Dienstag, der 06.12.2022 bis 08:00 Uhr im Moodle

Aufgabe 1 **(4 Punkte)**

Betrachten Sie den Generator für standardnormalverteilte Zufallszahlen durch Addition von gleichverteilten Zufallszahlen (Folie 288).

a) (1 Punkt) Im Skript ist der Spezialfall des Generators angegeben, in dem immer 12 Zufallszahlen verwendet werden. Leiten Sie eine entsprechende Formel her, die k Zufallszahlen verwendet.

b) (1 Punkt) Implementieren Sie Ihre Formel aus a) in einer Funktion `myRNorm1(n, k)` zur Erzeugung von n standardnormalverteilte Zufallsvariablen.

c) (2 Punkte) Überprüfen Sie die Güte des Generators in Abhängigkeit von k. Erzeugen Sie dazu für jedes $k \in \{1, 2, ..., 25\}$ 1000 mal 1000 Zufallszahlen und bestimmen Sie jeweils den p-Wert des Shapiro-Wilk Tests. Visualisieren Sie die Ergebnisse geeignet. Ab welchen Wert für k würden Sie von guten, normalverteilten Zufallszahlen sprechen?

Aufgabe 2 **(4 Punkte)**

Schreiben Sie eine R Funktion `mRbeta`, die mittels Rejection Sampling (Verwerfungsmethode) für gegebene Verteilungsparameter p und q insgesamt n beta-verteilte Zufallszahlen generiert. Verwenden Sie als Einhüllende eine stetige Gleichverteilung mit Parametern 0 und 1. Überprüfen Sie empirisch (z.B. durch Vergleich verschiedener Histogramme mit der wahren Dichtefunktion), ob Ihre Funktion korrekt ist, d.h. ob die erzeugten Zufallszahlen tatsächlich beta-verteilt sind.

Berechnen Sie (analytisch!), wie viele Zufallszahlen Sie im Mittel in Abhängigkeit von p, q und der tatsächlich realisierten Zufallszahl z ziehen müssen, um eine beta-verteilte Zufallszahl zu erzeugen. Achtung: Das Ergebnis ist nicht *schön*.

Hinweis: Nehmen Sie an: $p, q \geq 1$. Dies ist nicht unbeding erforderlich für die Beta-Verteilung, jedoch nimmt die Dichte für $x = 0$ oder $x = 1$ den Wert unendlich an, falls eine der Bedingungen nicht erfüllt ist. Dies erschwert die Bearbeitung dieser Aufgabe ein wenig. Sie dürfen sich gerne auch über diese Spezialfälle Gedanken machen. Weiterhin dürfen Sie gängige Eigenschaften der Beta-Verteilung gerne in der Fachliteratur (Wikipedia oder so) nachschlagen und verwenden.

Aufgabe 3 **(4 Punkte)**

Der Nikolaustag steht vor der Tür. Um herauszufinden, welche von zwei unterschiedlichen Schokoladensorten er dieses Jahr verteilen soll, hat der Nikolaus einige Geschmackstest durchführen lassen. Die Probanden sollten die beiden Sorten dabei mit Schulnoten (also von 1 – sehr gut bis 6 – ungenügend) bewerten. Nach langen Streits, welche der beiden Sorten die bessere ist, bleibt nun nur noch wenig Zeit für die Entscheidung und Sie wurden als Statistiker*in/ Data Scientist zur Klärung angestellt. Sie müssen die erhobenen Daten (`schokolade.RData`) nun bestmöglich nutzen und denjenigen Test finden, der die größte Power besitzt (natürlich unter Einhaltung des Niveaus). Dabei erinnern Sie sich an den lange zurückliegenden Besuch der Computergestützten Statistik und entscheiden sich einen Bootstrap zu verwenden, um so viel wie möglich aus der einen Stichprobe herauszuholen.

a) (1 Punkt) Implementieren Sie eine Funktion `bootstrapTest`, die einen zweiseitigen Test zum Mittelwertvergleich mittels Bootstrapping implementiert. Rückgabe Ihrer Funktion soll der p-Wert des Tests sein.

b) (2 Punkte) Schreiben Sie eine Funktion `bootstrapPower`. Diese erhält als Eingabe zwei numerische Vektoren x und y sowie eine Funktion, die einen statistischen Test ausführt und eine Folge von Verschiebungen δs. Diese Funktion soll den zugehörigen p-Wert zurückgeben. Erzeugen Sie sich mittels Bootstrapping je 200 Stichproben aus x und y, verschieben Sie die Stichprobe y um δ und zählen Sie, wie oft der Test die Nullhypothese ablehnt. Führen Sie dies innerhalb der Funktion für mehrere Werte für δ durch.

c) (1 Punkt) Stellen Sie auf Grundlage Ihrer Funktion aus b) jeweils die Power des t-Tests, des Wilcoxon-Tests und des Bootstrap-Tests dar. Ziehen Sie hierzu für $\pi = 0.25, 0.5, 0.75$, je 100 Beobachtungen aus einer $Bin(6, \pi)$-Verteilung (also `x <- rbinom(100, 6, π)` und `y <- rbinom(100, 6, π)`). Stellen Sie anschließend sicher, dass die Vektoren x und y denselben Mittelwert haben. Nutzen Sie $\delta = -1.25, -1, \ldots, -0.25, 0, 0.25, \ldots, 1.25$. Welchen Test empfehlen Sie? Können Sie mit diesem Test die Nullhypothese verwerfen, dass die beiden Schokoladensorten gleich gut schmecken?