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. 6

Abgabe ist Dienstag, der 22.11.2022 bis 08:00 Uhr im Moodle

Aufgabe 1 **(4 Punkte)**

Jetzt nutzen wir die in der Vorlesung gegebene Rekursionsformel, um einen weiteren Varianz-Algorithmus (`pairwiseVar`) zu erzeugen und zu implementieren. Dieser soll auf den Formeln (32), (33) und (34) der Folien 194 und 198 aufbauen.

a) (1 Punkt) Der gewünschte Algorithmus soll die Eingabe in jedem rekursiven Aufruf möglichst halbieren. Geben Sie eine Formel an, die den rekursiven Zusammenhang beschreibt. Denken Sie auch an einen Rekursionsanfang.

b) (2 Punkte) Implementieren Sie die Rekursionsformel aus a). Denken Sie dabei daran, dass sowohl T als auch S rekursiv bestimmt werden müssen! Daher muss Ihre Funktion eine benannte Liste mit S und T zurückgeben. Dieses Ausgabeverhalten ist unschön – in dieser rekursiven Funktion aber unumgänglich.

c) (1 Punkt) Um das Ausgabeverhalten Ihrer rekursiven Funktion zu verbergen, schreiben Sie eine `Wrapper`-Funktion. Diese Funktion ruft lediglich die rekursive Funktion auf und gibt anstelle der benannten Liste lediglich die Varianz zurück. Wenn Ihr Wrapper länger als 2 Zeilen ist, haben Sie etwas falsch gemacht. Dokumentieren Sie die Wrapper-Funktion! Diese soll vom Benutzer ausgeführt werden. Ihre eigentliche rekursive Funktion bedarf hingegen keiner ausführlichen Dokumenation, da sie nur für interne Aufrufe gedacht ist. Zwei kurze Sätze zu ihrer Funktionsweise sind dennoch angebracht.

Aufgabe 2 **(4 Punkte)**

Erweitern Sie die Varianzfunktionen um die Option der Verschiebung um den Mittelwert von m Beobachtungen. Wie groß muss / sollte mindestens / höchstens m sein? Gehen Sie wie folgt vor:

a) (1 Punkt) Schreiben Sie dazu eine Funktion `makeShiftedVar`, die eine Funktion zur Varianzberechnung sowie m als Eingabe erhält und eine Shifted-Variante der Funktion zurückgibt.

b) (2 Punkte) Bestimmen Sie mit den Varianzfunktionen aus `var.R` sowie ihrer Funktion aus Aufgabe 1 die Varianz von $\frac{10^6+\{1,\ldots,1000\}}{100}$ und $\frac{10^{16}+\{1,\ldots,1000\}}{100}$ jeweils für $m \in \{0, 1, \ldots, 10\}$ und betrachten Sie die Abweichung von der wahren Varianz $\frac{n(n+1)}{12\cdot 100^2}$. Visualisieren Sie das Ergebnis geeignet.

c) (1 Punkt) Wie verändert sich die Qualität der Ergebnisse der einzelnen Algorithmen für $m \to n$? Welche Empfehlung für m geben Sie hier und welche im Allgemeinen? Könnte eine solche Verschiebung die Probleme für große n der Situation aus Aufgabe 1 lösen?

Aufgabe 3 **(4 Punkte)**

Neben dem Bestehen statistischer Tests müssen gute Zufallszahlengeneratoren auch einige weitere Qualitätsmerkmale aufweisen, wie z.B. Effizienz und eine ordentliche Periodenlänge.

Betrachten Sie den LCF-Generator aus dem Material. Zeigen Sie:

(i) Der Generator mit den Einstellungen $a = 5, c = 453816693, m = 2^{31}$ hat volle Periodenlänge.

(ii) Der Generator mit den Einstellungen $a = 698769069, c = 453816693, m = 2^{31}$ hat volle Periodenlänge.

Erzeugen Sie nun mit beiden Generatoren jeweils 10^5 Zufallszahlen. Zeigen Sie, dass die tatsächliche Periode des Generators (ii) kleiner als 10000 ist und erklären Sie das Verhalten.