

Übungen zur Vorlesung

Praktische Optimierung, SoSe 2022

Prof. Dr. Günter Rudolph, Dr. Roman Kalkreuth

<https://ls11-www.cs.tu-dortmund.de/people/rudolph/teaching/lectures/POKS/SS2022/lecture.jsp>**Blatt 3, Block A**

26.04.2022

Abgabe: 03.05.2022**Aufgabe 3.1: Welch-Test und Wilcoxon-Test (6 Punkte)**

In dieser Aufgabe wollen wir uns mit dem Welch-Test und dem Wilcoxon-Test beschäftigen. Dazu betrachten wir erneut die Kompassuche.

Auf der Webseite zur Übung wird eine neue Variante der Kompassuche zur Verfügung gestellt. Diese hat als Abbruchkriterium nicht mehr die Anzahl an Iterationen sondern eine Abweichung vom wahren Optimum. Solange die absolute Abweichung vom wahren Optimum größer als die vorgegebene Toleranz ist, wird die Kompassuche fortgesetzt. Nach spätestens 1000 Iterationen wird die Suche abgebrochen, um eine Endlosschleife zu vermeiden. Die Kompassuche gibt nun die Anzahl an Iterationen aus, die benötigt wird, um nah genug an das Optimum zu kommen.

Erzeugen Sie 500 gleichverteilte (Pseudo-)Zufallszahlen aus dem Intervall $[-10, 10] \times [-10, 10]$. Verwenden Sie dazu `set.seed(1)`. Nutzen Sie diese 500 Zahlen jeweils als Startwert $x^{(0)}$ für die Kompassuche. Verwenden Sie eine Toleranz von 0.001 für die Abweichung vom globalen Optimum.

Sie sollen erneut die Funktion $f(x, y) = x^2 + y^2$, $x, y \in [-10, 10]$ minimieren. Verwenden Sie dabei die folgenden Parametereinstellungen für die Kompassuche:

- (i) $s^{(0)} = 1$ und $\theta = 0.5$,
- (ii) $s^{(0)} = 2$ und $\theta = 0.5$,
- (iii) $s^{(0)} = 1.5$ und $\theta = 0.8$,
- (iv) $s^{(0)} = 0.5$ und $\theta = 0.2$.

Wir wollen die tatsächlich benötigte Zeit analysieren, die für die Kompassuche mit verschiedenen Parametereinstellungen benötigt wird. Diese hängt von der Anzahl der benötigten Iterationen ab. Messen Sie die Laufzeiten für die Kompassuche mit den unterschiedlichen Einstellungen, wobei Sie jede Messung 100 mal wiederholen und die mediane Laufzeit aus den 100 Wiederholungen verwenden sollen, um Messungenauigkeiten zu vermeiden. Sie sollen also $4 \cdot 500 \cdot 100 = 200\,000$ Messungen durchführen und daraus 2000 mediane Laufzeiten berechnen. Verwenden Sie hierzu die Funktion `microbenchmark` aus dem gleichnamigen Paket. Speichern Sie die medianen Laufzeiten in vier Vektoren `t1`, `t2`, `t3` und `t4` für die vier verschiedenen Parametereinstellungen ab und speichern Sie diese als `.RData`-Datei mithilfe der Funktion `save`.

Erstellen Sie in einem ersten Schritt Boxplots der gemessenen Laufzeiten für die vier verschiedenen Kombinationen von Parametern. Jeder Boxplot soll dabei aus 500 Werten erstellt werden, die zu je einer Parametereinstellung gehören. Beschriften Sie die Boxplots sinnvoll und sorgen Sie dafür, dass die y -Achse jeweils die gleichen Wertebereiche abdeckt. Interpretieren Sie die Boxplots im Hinblick auf mögliche Unterschiede bei der Optimierung mit den verschiedenen Parameterkombinationen.

Hinweis 1: Es ist möglich mithilfe von `boxplot` direkt vier Boxplots in einer Grafik zu erzeugen.

Hinweis 2: Es kann sinnvoll sein die Zeitmessung nur einmal durchlaufen zu lassen und anschließend die abgespeicherten Ergebnisse zu verwenden.

Testen Sie nun die folgenden Hypothesen:

- (i) Die Kompasssuche mit $s^{(0)} = 1$ und $\theta = 0.5$ benötigt im Mittel weniger Zeit als die Kompasssuche mit $s^{(0)} = 1.5$ und $\theta = 0.8$.
- (ii) Die Kompasssuche mit $s^{(0)} = 1.5$ und $\theta = 0.8$ benötigt im Mittel weniger Zeit als die Kompasssuche mit $s^{(0)} = 0.5$ und $\theta = 0.2$.
- (iii) Die Kompasssuche mit $s^{(0)} = 2$ und $\theta = 0.5$ benötigt im Mittel weniger Zeit als die Kompasssuche mit $s^{(0)} = 1.5$ und $\theta = 0.8$.
- (iv) Die Kompasssuche mit $s^{(0)} = 1$ und $\theta = 0.5$ benötigt im Mittel weniger Zeit als die Kompasssuche mit $s^{(0)} = 0.5$ und $\theta = 0.2$.

Geben Sie jeweils die Hypothesen H_0 und H_1 an und testen Sie die Hypothesen sowohl mit dem Welch-Test (`t.test`) als auch mit dem Wilcoxon-Test (`wilcox.test`). Nutzen Sie dazu die zuvor von Ihnen gemessenen medianen Laufzeiten. Interpretieren Sie die Ergebnisse. Gibt es große Unterschiede bei den Ergebnissen der beiden Tests? Ergeben sich die gleichen Schlussfolgerungen wie bei der Betrachtung der Laufzeiten mithilfe von Boxplots?

Überprüfen Sie jeweils vorher die Annahmen für die Tests. Verwenden Sie dabei für die Normalverteilungsannahme QQ-Plots (`qqnorm` und `qqline`) und für die Verschiebung der Verteilungsfunktionen empirische Verteilungsfunktionen (`plot.ecdf`). Führen Sie die Tests auch dann durch, wenn Sie zu dem Ergebnis kommen, dass die Annahmen verletzt sind.

Hinweis 3: Informieren Sie sich erst, welche Annahmen die Tests haben und wie QQ-Plots funktionieren, bevor Sie beginnen, die Annahmen zu überprüfen. Es gibt jeweils mehrere Annahmen!

Aufgabe 3.2: Test auf Gleichheit (2 Punkte)

Sie haben in der Vorlesung gelernt, dass ein (sinnvoller) Test auf Gleichheit, also beispielsweise für das Hypothesenpaar

$$H_0: \theta \neq \theta_0 \text{ vs. } H_1: \theta = \theta_0$$

mit $\theta \in \Theta \subseteq \mathbb{R}$ nicht existiert. Begründen Sie, warum ein solcher Test keine höhere Güte unter der Alternativen (Wahrscheinlichkeit eine in Wahrheit nicht zutreffende Nullhypothese zu verwerfen) als das Signifikanzniveau α erreichen kann. Was bedeutet dies inhaltlich?

Hinweis 4: Es kann gezeigt werden (darf ohne Beweis verwendet werden), dass die Gütefunktion stetig ist. Nutzen Sie dies bei Ihrer Argumentation. Beachten Sie außerdem, dass die Alternative in diesem Fall nur aus einem Punkt besteht.

Aufgabe 3.3: Fehler beim Testen (2 Punkte)

- (a) Eine Person formuliert die Hypothese, die sie nachweisen möchte, als H_0 . Ist dies ein sinnvolles Vorgehen? Begründen Sie Ihre Antwort.
- (b) Es wird die Hypothese

$$H_0: \text{ Stichprobe ist normalverteilt vs. } H_1: \text{ Stichprobe ist nicht normalverteilt}$$

zum Niveau $\alpha = 0.05$ getestet. Dabei ergibt sich ein p -Wert von 0.08. Die Nullhypothese kann also nicht abgelehnt werden. Die Person, die den Test durchgeführt hat, ist sich nun sicher, dass Ihre Stichprobe normalverteilt ist. Ist dies sinnvoll? Wie fällt Ihre Antwort bei einem p -Wert von 0.75 aus? Begründen Sie Ihre Antworten.

- (c) Eine Person erhält ein signifikantes Testergebnis. Kann sie sich sicher sein, dass sie damit etwas wissenschaftlich relevantes herausgefunden hat? Begründen Sie Ihre Antwort.