Tutorial 3

Kerndichteschätzer, Poissonregression

Aufgabe 1: Aktienrenditen und Value-at-Risk (VaR)

Im Finanzsektor ist der Value at Risk (VaR) ein Standardmaß, um das Risiko eines Portfolios bzw. einer Investition zu messen. Der Value at Risk gibt an, wie groß der Verlust einer Investition (mit einer bestimmten Wahrscheinlichkeit) in einem bestimmten Zeitraum, z.B. einem Tag, sein kann.

Wenn z.B. eine Aktie einen eintägigen 95 %-VaR von -0.02 aufweist, bedeutet dies, dass die Aktie mit einer Wahrscheinlichkeit von 0.05 innerhalb eines Tages um mehr als 2% an Wert verliert.

- a) Wir lesen zunächst Daten zum Schlusskurs der Amazon Aktie ein:

```
stock_prices = read.csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2023/2023-02-07/big_tech_stock_prices.csv')
amazon_prices = stock_prices[stock_prices$stock_symbol == "AMZN", ]
```

Geben Sie sich per `view(amazon_prices)` den Datensatz aus. Für welchen Tag liegt die erste Beobachtung vor? Und für welchen Tag die letzte?

- b) Um uns den VaR näher anzuschauen, berechnen wir zunächst die Tagesrenditen $y_t = p_t/p_{t-1} - 1$, $t = 2, \dots, 3271$, wobei p_t den Schlusskurs an Tag t bezeichnet. In R können wir y_t wie folgt berechnen:

```
amazon_returns = (amazon_prices$adj_close[2:3271] /
                  amazon_prices$adj_close[1:3270]) - 1
```

- c) Um eine Verteilung für die Tagesrenditen der Amazon-Aktie zu schätzen, könnte man eine Normalverteilung verwenden. Schätzen Sie die Parameter μ und σ der Normalverteilung und plotten Sie die geschätzte Verteilung.

- d) Fügen Sie Ihrem Plot aus c) eine Kerndichteschätzung der Renditen hinzu. Halten Sie die Normalverteilungsannahme für plausibel?
- e) Für den VaR sind insbesondere die Werte sehr weit links auf der x-Achse relevant, d.h. die negativen Renditen. Sind Renditen von weniger als -5% unter der Normalverteilung oder unter dem KDE wahrscheinlicher? Welche praktische Auswirkung hat die Verwendung der Normalverteilung hier hinsichtlich des VaR?

Aufgabe 2: Poissonregression

In dieser Aufgabe beschäftigen wir uns mit dem Zusammenhang der Anzahl ausgeliehener Fahrräder an einer Verleihstation in Washington, D.C., USA und verschiedenen erklärenden Variablen. Der Datensatz enthält die Temperatur (`temperature`, in Grad Fahrenheit), die Windgeschwindigkeit (`windspeed`, in Meilen pro Stunde), ob der Tag auf ein Wochenende gefallen ist (`weekend`) und die Anzahl ausgeliehener Fahrräder (`count`). Wir lesen zunächst die Daten ein:

```
bikes = read.csv("https://tinyurl.com/5bkjupn2")
```

- a) Schätzen Sie mithilfe der `glm()`-Funktion das Poissonregressionsmodell

$$\text{count}_i \sim \text{Po}(\lambda_i), \lambda_i = e^{\beta_0 + \beta_1 \text{temperature}_i + \beta_2 \text{windspeed}_i + \beta_3 \text{weekend}_i}.$$

Beschreiben Sie stichwortartig die geschätzten Effekte der erklärenden Variablen (positiv/negativ, signifikant/insignifikant?).

- b) Geben Sie für die Variablen `temperature` und `windspeed` eine quantitative Interpretation der geschätzten Effekte an.
- c) Erstellen Sie einen Scatterplot von `count` vs. `temperature`. Ergänzen Sie dann mithilfe der `curve()`-Funktion die in a) geschätzte Regressionsfunktion (als Funktion von `temperature`) i) für Wochentage und ii) für Wochenenden, wobei die beiden Regressionsfunktionen verschiedene Farben haben sollen. Für die Windgeschwindigkeit können Sie die mittlere Windgeschwindigkeit über alle Beobachtungen verwenden.
- d) Für kommenden Samstag sind eine Temperatur von 40 Grad Fahrenheit und eine Windgeschwindigkeit von 20 Meilen pro Stunde vorhergesagt. Wie hoch ist die erwartete Anzahl ausgeliehener Fahrräder? Geben Sie die Verteilung der Anzahl ausgeliehener Fahrräder gemäß des Modells aus a) an.
- e) Bestimmen Sie die von dem Modell implizierte Wahrscheinlichkeit, dass kommenden Samstag mehr als 75 (mehr als 100) Fahrräder ausgeliehen werden. Die Wahrscheinlichkeitsfunktion der Poissonverteilung ist in R unter `ppois()` verfügbar.