

Mathematik II für Biologen

Übungsblatt 4 (Abgabe am 08.05.2015)

Aufgabe 12 (Quantil-Quantil-Diagramm, Q-Q-Plot)

(10 Punkte)

Uns liege das Histogramm einer Stichprobe vor, das auf den ersten Blick mehr oder weniger glockenförmig (wie eine Gauß-Kurve) aussieht. Daher stellen wir uns die Frage, ob es Parameter μ und σ gibt, so dass der Plot der Dichte der Normalverteilung,⁷

$$f_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right),$$

das Histogramm gut beschreibt.

Die folgende Darstellungsart, genannt Quantil-Quantil-Diagramm oder kurz Q-Q-Plot (auch normal plot und manchmal leider wenig spezifisch einfach nur Wahrscheinlichkeits-Diagramm), erlaubt es einem, dies zu überprüfen, ohne dafür zunächst die passenden Parameterwerte μ und σ zu bestimmen (bzw. zu raten, zu schätzen oder auszuprobieren). Auch können mit diesem Diagramm Abweichungen von der Gauß-Kurve leichter beurteilt werden.

Dazu definieren wir zunächst für $0 < \alpha < 1$ das (theoretische) α -Quantil $q_\alpha^{(\Phi)}$ für die Dichte der Standardnormalverteilung, $f_{0,1}$, und zwar ist $q_\alpha^{(\Phi)} \in \mathbb{R}$ diejenige Zahl, für die

$$\Phi\left(q_\alpha^{(\Phi)}\right) = \alpha, \quad \text{wobei} \quad \Phi(x) := \int_{-\infty}^x f_{0,1}(x) dx = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx,$$

z.B. ist $q_{0,975}^{(\Phi)} = 1,96$. Im Q-Q-Plot werden dann die Punkte

$$\left(q_{(i-1/2)/n}^{(\Phi)}, x_{(i)}\right), \quad i = 1, \dots, n,$$

in einem zweidimensionalen Diagramm eingetragen, wobei $x_{(i)}$ die der Größe nach geordneten Werte der ursprünglichen Stichprobe sind.

Kurz gesagt, trägt man also die theoretischen Quantile der Normalverteilung gegen die empirischen Quantile der Stichprobe auf.

- Welchen Wert hat die empirische Verteilungsfunktion $F(x)$ für x etwas kleiner als $x_{(i)}$ und für x etwas größer als $x_{(i)}$? Wenn wir versuchen, die (unstetige) Treppenfunktion F durch eine stetige Funktion zu nähern, welchem Wert sollte diese stetige Funktion dann an der Stelle $x_{(i)}$ möglichst nahe kommen?
- Erzeugen Sie mit dem MATLAB-Befehl `qqplot` einen Q-Q-Plot der Daten aus `states.dat` aus Aufgabe 4. Markieren Sie darin die beiden Ausreißerstaaten aus Aufgabe 4d (mit Namen).
- Lassen sich die Daten aus `states.dat` gut durch eine Normalverteilung beschreiben? (Bearbeiten Sie, bevor Sie diese Frage beantworten, zuerst Aufgabe 13.)

⁷Die Funktion $f_{\mu,\sigma}$ ist Ihnen bereits im Wintersemester in Aufgabe 15 begegnet. Sie wird im Laufe des Sommersemesters noch eine Rolle spielen.

Aufgabe 13

(10 Punkte)

Unten wird für die 6 Stichproben aus Aufgabe 1 jeweils der Q-Q-Plot gezeigt. Ordnen Sie die Q-Q-Plots a-f den Histogrammen A-F aus Aufgabe 1 zu, und begründen Sie kurz Ihre Entscheidung. Welche Histogramme lassen sich besser, welche schlechter durch eine Gauß-Kurve beschreiben? Wie sieht man dies den zugehörigen Q-Q-Plots an?

