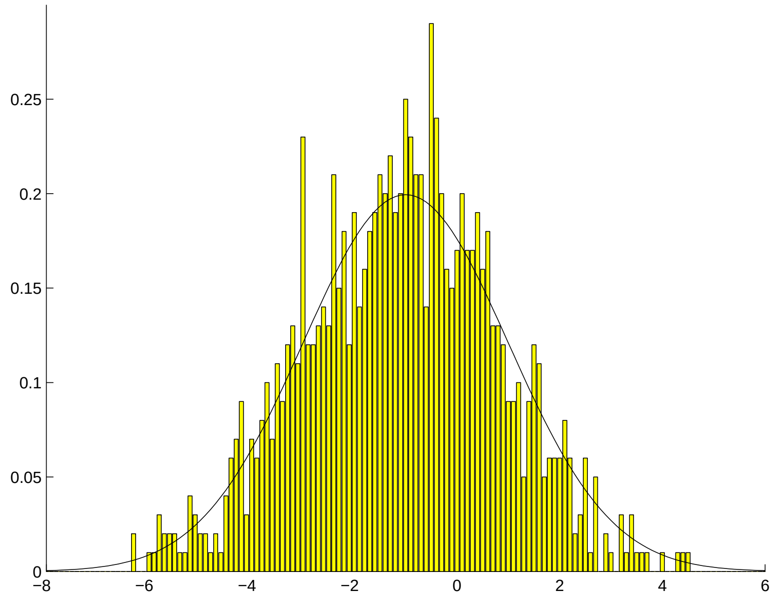
Universität Tübingen
Mathematisches Institut
Dr. Stefan Keppeler

Sommersemester 09

Mathematik II für Biologen

Übungsblatt 3 (Abgabe am 8.5.2009 **vor** der Vorlesung)

Aufgabe 8 (Quantil-Quantil-Diagramm, Q-Q-Plot) (10 Punkte)
Die folgende Abbildung zeigt das Histogramm zu einer Stichprobe $x_1, \ldots, x_n$ der Größe $n = 1000$.



(Die Höhen der Balken des Histogramms wurden dabei so normiert, dass die Gesamtfläche des Histogramms gleich 1 ist.) Auch wenn das Histogramm noch recht unregelmäßig ist, so ist es doch in etwa "glockenförmig". Man mag versucht sein, eine glatte Kurve hindurchzulegen, die das Histogramm beschreibt. Dies ist in der Abbildung bereits geschehen, und zwar ist dort der Graph der Ihnen aus der Mathematik I bekannten Gauß-Funktion

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

für $\mu = -1$ und $\sigma = 2$ eingezeichnet. ($f_{\mu,\sigma}$ ist die sogenannte Dichte der Normalverteilung mit Erwartungswert μ und Standardabweichung σ, die später noch näher behandelt wird.) Augenscheinlich beschreibt diese Kurve das Histogramm recht gut, d.h. es gibt eine geeignetes μ und ein geeignetes σ, für das das Histogramm gut durch $f_{\mu,\sigma}$ angenähert wird. Für andere $\mu \neq -1$ oder $\sigma \neq 2$ ist dies i.a. nicht der Fall.

Die folgende Darstellungsart, genannt Quantil-Quantil-Diagramm (Q-Q-Plot, auch normal plot und manchmal leider wenig spezifisch einfach nur Wahrscheinlichkeits-Diagramm genannt) erlaubt es einem, zu überprüfen, ob eine Stichprobe bzw. deren Histogramm gut durch irgendeine Gauß-Kurve beschrieben werden kann, ohne dass man sich dabei über die geeignete Wahl von μ und σ Gedanken machen muss. Auch können mit diesem Diagramm Abweichungen von der Gauß-Kurve leichter beurteilt werden.

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

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

(Z.B. ist $q_{0.975}^{(\Phi)} = 1.96$. Bem.: $\Phi(\infty) = 1$.) Im Q-Q-Plot werden dann die Punkte

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

in einem zweidimensionalen Diagramm eingetragen, wobei $(x_{(i)})_{i=1,\ldots,n}$ die der Größe nach geordnete ursprüngliche Stichprobe ist.

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

a) Welchen Wert hat die empirische Verteilungsfunktion $F(x)$ für x etwas kleiner als $x_{(i)}$ und für x etwas größer als $x_{(i)}$?

b) 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).

c) Lassen sich die Daten aus `states.dat` gut durch eine Normalverteilung beschreiben? (Bearbeiten Sie, bevor Sie diese Frage beantworten, zuerst die nächste Aufgabe).

d) Schätzen Sie mithilfe des Q-Q-Plots aus Teil b grob die Parameter μ und σ der in Teil c erwähnten Normalverteilung. Wie sind Sie bei der Schätzung vorgegangen und warum? (5 Zusatzpunkte)

Aufgabe 9 (10 Punkte)

Unten sind zu 6 Stichproben der Größe $n = 200$ jeweils das Histogramm und der Q-Q-Plot gezeigt, wobei die Beschriftung der y-Achse weggelassen wurde. Finden Sie die richtigen Paare und begründen Sie kurz Ihre Entscheidung. Welches Histogramm läßt sich am besten durch eine Gauß-Kurve beschreiben? Wie sieht man dies dem zugehörigen Q-Q-Plot an?