Datenanalyse
WS 2001/00

Author(s):
Pruys, H.

Publication Date:
2001

Permanent Link:
https://doi.org/10.3929/ethz-a-004444215

5 Einführung in die Methode der kleinsten Quadrate

Im letzten Kapitel haben wir auf einfache Weise aus einer Anzahl vorgegebener Geschwindigkeitswerte die Temperatur und deren Fehler eines Xenon-Gases bestimmt. Dieses sehr einfache Verfahren war in diesem Fall durchaus angemessen, hat jedoch im allgemeinen einige Schwachpunkte:

- Die Theoriefunktion muß sehr genau bekannt sein, damit man einen analytischen Zusammenhang zwischen der Meßgröße (z.B. v) und der Theoriegröße (z.B. T) herstellen kann. Ferner benötigt man die erste Ableitung der Theoriegröße, um den Fehler abzuschätzen. In der Praxis der Datenanalyse sind beide Vorraussetzungen nicht immer gegeben.
- Das Verfahren macht keine Aussage darüber, ob die Verteilung der Meßwerte überhaupt mit der Theoriefunktion kompatibel ist, da z.B. in unserem Fall nur der Mittelwert der Geschwindigkeiten in die Berechnung der Temperatur eingeht. Wenn dieses Vorgehen blind angewandt wird, kann man also sogar aus den gemessenen Werten einer Radarkontrolle eine Temperatur samt Fehler ausrechnen! Im letzten Kapitel war die einzige Kontrolle, die wir hatten, die graphische Übereinstimmung zwischen der Theoriefunktion und den histogrammierten Meßwerten: ein Verfahren, das nur im ein- und zweidimensionalen Fall einigermaßen funktioniert.
- Bei relativ großen Fehlern der Meßgröße wird die Abschätzung des Fehlers der Theoriegröße über die erste Ableitung beliebig ungenau.

Die in der Vorlesung vorgestellte (siehe insbesondere Gleichung 1.44 und Programm 1.7) Methode der kleinsten Quadrate behebt alle drei Schwachpunkte. Ihr einziger Nachteil ist, daß sie eigentlich nur für Daten anwendbar ist, die gaußverteilt sind. Dieses läßt sich zum Glück meistens wenigstens teilweise durch eine Vorverarbeitung (z.B. Histogrammieren) der Meßwerte erreichen, insbesondere da sich die Poisson- und die Gauß-Verteilung für große Erwartungswerte schnell annähern.
Wir wollen ein Programm schreiben, welches mit der Methode der kleinsten Quadrate erneut die Temperatur, die zu den Meßwerten aus dem letzten Kapitel paßt, berechnet. Wie eben erwähnt müsssen die Daten dafür zuerst geeignet histogrammiert werden, wobei wir nur solche Bins verwenden wollen, die mehr als 10 Einträge

haben, da wir sichergehen wollen, daß der Anzahl der Einträge in jedem verwendeten Bin eine Gauß-Verteiltung entspricht. [1]
Machen wir uns an die Skizzierung des Programmes:

- Einlesen und Histogrammieren der Daten
- Berechnung des $\chi^2(T)$ für einen größeren Temperaturbereich nach folgender Formel (siehe Gleichung 1.44 und Programm 1.7):

$$\chi^2(T) = \sum_{\substack{i=1 \\ y_i \geq 10}}^{n_{bins}} \frac{(y_i - Nf(v_i, T)\Delta v)^2}{y_i}$$

- Bestimmung des Minimums von $\chi^2(T) \Longrightarrow T_0$, $\chi^2_{min} = \chi^2(T_0)$
- Bestimmung des oberen (ΔT_+) und unteren (ΔT_-) Fehlers von T_0, über die Beziehung (vergl. Skript):

$$\chi^2(T_0 + \Delta T_+) = \chi^2(T_0 - \Delta T_-) = \chi^2_{min} + 1$$

Dies können wir sogleich noch weiter in einzelne Schritte zerlegen:

- Einlesen der Daten aus maxwell5.dat
  (Siehe letztes Kapitel)
- Histogrammieren der Geschwindigkeiten
  Wiederum analog zum letzten Kapitel. Man sollte hier genau überlegen, welche Binbreite man wählt. Eine zu kleine Binbreite führt dazu, daß die meisten Daten nicht verwendet werden ($y_i < 10$); eine zu große verschmiert Unterschiede zu stark.
- Auswahl der Bins mit genügend Einträgen
  Hierzu definiere man einen logischen Vektor `ok` , der die Bins angibt, die mindestens 10 Eintäge haben. Dann konstruiere man neue Vektoren, `v_ok` und `y_ok` , die das verkleinerte Histogramm beschreiben.

[1] Dies bedeutet natürlich, daß wir nicht alle gemessenen Daten verwenden, wodurch der Fehler unserer Temperaturbestimmung größer wird, insbesondere, da sich die Maxwellverteilungen zu verschiedenen Temperaturen hauptsächlich im 'Schwanz' der Verteilung unterscheiden. Als verbesserten Algorithmus könnte man die Bins, die zuwenig Einträge haben, geschickt so zusammenfassen, daß keine mit weniger als 10 Einträgen übrigbleiben.

- Berechnung der χ^2-Werte
  Zuerst sollte man die einen Vektor `T` einführen, der in kleinen Schritten den interessanten Temperaturbereich abdeckt. Anschließend definiert man einen Vektor `chi2` gleicher Länge (Befehl `zeros`). Man programmiere eine `for` Schleife, die für jedes Element von `T` einmal durchlaufen wird. Hierzu ist das Kommando `length` gut zu gebrauchen. Innerhalb der Schleife wird für jeden Temperaturwert das entsprechende χ^2 bestimmt. Jedes einzelne Element wird mittels Indizierung angesprochen (z.B. `T(i)`). Man verwende die bereits definierte Maxwell-Funktion und den Befehl `sum` .

- Darstellung von $\chi^2(T)$

- Bestimmung von T_0 und χ^2_{min}
  Dies erreicht man am einfachsten mit der erweiterten Variante von `min` , die sowohl den Minimalwert als auch den Index des Minimalwertes zurückgibt.

- Bestimmung des Fehlers der Temperatur (zwei Varianten):

  1. Graphische Bestimmung
     Man zeichnet zuerst die Linie zu $\chi^2 = \chi^2_{min}+1$ in den χ^2-Plot und bestimme dann die Schnittpunkte mit der Kurve graphisch mittels der Kommandos `ginput` und `disp` .
  2. Rechnerische Bestimmung
     Man bestimme alle Bins, für die `chi2` kleiner ist als χ^2_{min}. Die maximale und die minimale Temperatur, die zu diesen Bins gehört, gibt die obere und untere Grenze des Fehlerbereichs an.

- Ausgabe aller Resultate
  Man sollte die histogrammierten Daten zusammen mit der Maxwell-Verteilung zur errechneten Temperatur in einem Teil der figur darstellen. Im anderen Teil stelle man $\chi^2(T)$ dar. Alle errechneten Werte sollten auch als Text ausgegeben werden.

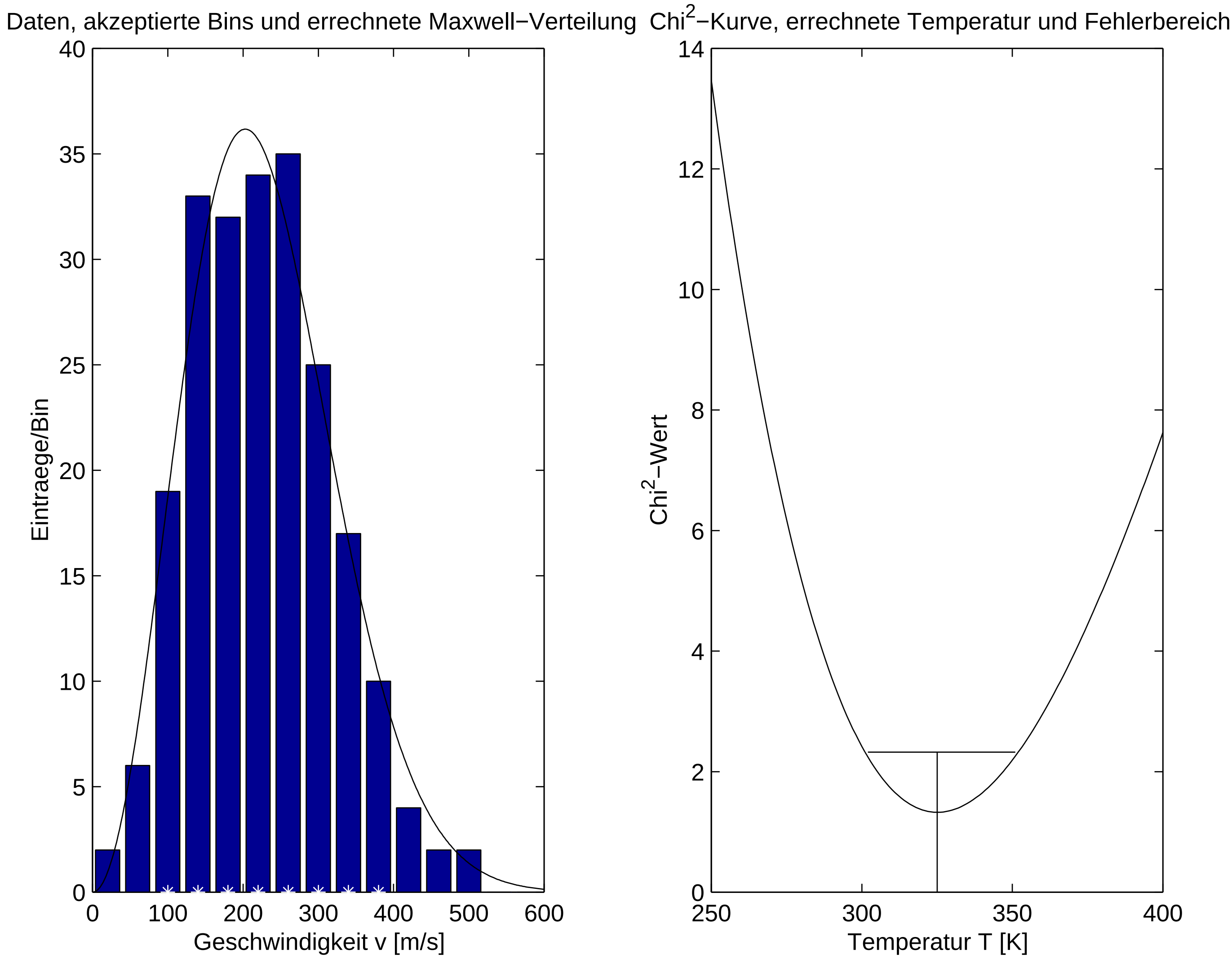

Abbildung 2: Ein mögliches Resultat