

Themenvorschlag Diplomarbeit:
Detektion von Peaks in Funktionen generiert aus DNA-Daten
Betreuer: PD Dr. Martin Grüttmüller

Themengebiet: Optimierung

Eine DNA-Analyse ist ein molekularbiologische Verfahren, welches die DNA verwendet, um Rückschlüsse auf verschiedene Aspekte des Individuums ziehen zu können. DNA-Analysen werden unter anderem durchgeführt, um mit dem Genetischen Fingerabdruck Identitäts- und Verwandtschaftsfragen zu klären, die man unter anderem in der Biologie bei der Konstruktion von Stammbäumen verwendet.

Eine Methode, um einen genetischen Fingerabdruck zu gewinnen ist die Fragmentlängenanalyse. Hier wird die DNA mit Hilfe von Restriktionsenzymen geschnitten. Diese Restriktionsenzyme erkennen spezifische Abschnitte in der DNA. Je nachdem wie oft ein solcher Abschnitt in einem Chromosom vorhanden ist ergeben sich unterschiedlich viele und unterschiedlich lange DNA-Fragmente. Die Anzahl und Länge der DNA-Fragmente versucht man mit Hilfe Gel-Elektrophorese zu bestimmen. Die Fragmente werden dabei in ein Gel gegeben und wandern unter Einfluss eines elektrischen Felds durch das Gel. Je nach Größe und Ladung des DNA-Fragment bewegen sich diese unterschiedlich schnell durch das als Molekularsieb wirkende Gel. Zur Auswertung werden die Fragmente z.B. mit Farbstoffen markiert, die es erlauben automatisch zu zählen, wieviele Fragmente zu welcher Zeit durch das Gel kommen. Unter der idealen Annahme das kurze Fragmente schneller sind als lange Fragmente kann man dann eine Zuordnung Zeit-Länge vornehmen. In der Praxis kann es natürlich vorkommen, dass ein kurzes Fragment länger als braucht als seine Größe erwarten lässt (z.B. wenn es sich mit anderen Fragmenten verharkt, quer durchgeht, verschiedene Fragmente gleicher Länge unterschiedliche Ladung haben, etc.). Zum einen bedeutet das, das die Fragmente einer Länge nicht einen Ausschlag von der Höhe der Anzahl dieser Fragment in der Zählfunktion erzeugen, sondern eine normalverteilte Glockenkurve deren Flächeninhalt der Anzahl entspricht. Zum anderen muss man mit einem Rauschen in den Daten rechnen.

Man kann sich die bei der Gel-Elektrophorese entstehende Funktion als Superposition von Gaußschen Glockenkurven $f(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ mit Erwartungswert μ und Standardabweichung σ vorstellen. In Abbildung 1

ist so eine Superposition exemplarisch dargestellt. Sie wurde erzeugt durch folgende Maple Kommandos:

```
> f:=(x,m,s)->1/(s*sqrt(2*Pi)) * exp(-0.5*((x-m)/s)^2);  
> plot([0.1+f(x,5,1)+f(x,14,1)+f(x,20,1)+f(x,23,1)+f(x,30,1)+f(x,30,1)  
      +f(x,37,1)+f(x,38,1)+f(x,42,1)+f(x,43,1)+f(x,45,1)],x=0..50);
```

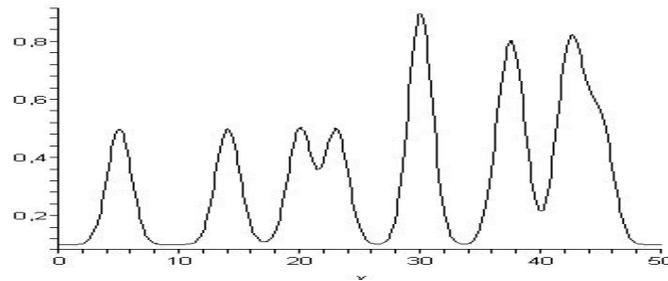


Figure 1: Superposition von 11 Gaußschen Glockenkurven

Das Problem bei der Auswertung ist jetzt zu entscheiden, zu welchem Zeitpunkt μ ein Ausschlag in Form einer Glockenkurve vorlag.

Diese Problemstellung soll mit mathematischen Methoden untersucht werden. Es sind Algorithmen zu entwickeln und zu implementieren, die zu einer gegebenen Superposition (gegeben durch Liste von Wertepaaren) die Parameter der zugrunde liegenden Glockenkurven bestimmen. Diese Algorithmen sollen gegen schon existierende Algorithmen bzgl. Effizienz und Qualität evaluiert werden. Eine denkbare Erweiterung der Aufgabenstellung wäre die Annahme, dass die Fragmente gleicher Länge nicht normalverteilt ankommen, sondern evt. eine rechtsschiefe Verteilung aufweisen.

Quellen: Wikipedia, Ausführungen Dr. Hepperle