

## 4 Induktive Statistik

### 4.1 Konfidenzintervall: Einführung

Grob gesagt wollen wir jedem Ereignis  $\omega$  (in der Regel ein Ergebnis, das wir durch eine Stichprobe vom Umfang  $n$  erhalten) ein Intervall  $I(\omega)$  zuordnen, so dass ein unbekannter Parameter  $\theta$  mit einer gewissen Wahrscheinlichkeit in diesem Intervall liegt, also

$$P(\theta \in I(\omega)) = 1 - \alpha.$$

Die Zahl  $1 - \alpha$  heißt das **Konfidenzniveau**. Man beachte, dass  $I(\omega)$  das “zufällige” Ergebnis ist. Wir können  $I(\omega)$  mit einer unteren und oberen Schranke festlegen. Das Intervall  $I(\omega)$  heißt Konfidenzintervall für den Parameter  $\theta$ .

Das übliche Szenario ist das folgende: Wir ziehen eine Stichprobe vom Umfang  $n$  und berechnen aus den Stichprobenwerten  $x_i$  eine untere  $g_u$  und eine obere  $g_o$  Intervallgrenze. Wir haben also zwei Zufallsvariablen  $g_u(\omega)$  und  $g_o(\omega)$ . Über diese können wir sinnvoll natürlich nur etwas aussagen, wenn wir ihre Verteilung kennen.

**Beispiel 4.1.** Gegeben sei eine  $N(\mu, \sigma^2)$  verteilte Zufallsvariable  $X$ . Wir wollen ein Intervall angeben, das mit W.K.  $1 - \alpha$  den Erwartungswert  $\mu$  enthält. Wir gehen jetzt also davon aus, dass die Z.V.  $N(\mu, \sigma^2)$  verteilt ist und geben dann ein Intervall  $I$  an, das  $\mu$  mit einer gewissen W.K. enthält.

Wähle dazu  $z_{\alpha/2}$  als das  $\alpha/2$ -Quantil der Normalverteilung, also  $\Phi(z_{\alpha/2}) = \alpha/2$ . Damit gilt

$$\begin{aligned} P\left(\frac{X(\omega) - \mu}{\sigma} < z_{\alpha/2}\right) &= \frac{\alpha}{2} \\ P\left(\frac{X(\omega) - \mu}{\sigma} > z_{1-\alpha/2}\right) &= 1 - \left(1 - \frac{\alpha}{2}\right) = \frac{\alpha}{2} \end{aligned}$$

und somit

$$P\left(z_{\alpha/2} \leq \frac{X(\omega) - \mu}{\sigma} \leq z_{1-\alpha/2}\right) = 1 - \alpha.$$

Benutze nun  $z_{\alpha/2} = -z_{1-\alpha/2}$  und erhalte so

$$P(\mu - \sigma z_{1-\alpha/2} \leq X(\omega) \leq \mu + \sigma z_{1-\alpha/2}) = 1 - \alpha.$$

Das bedeutet, mit W.K.  $1 - \alpha$  umfasst das Intervall

$$[X(\omega) - \sigma z_{1-\alpha/2}, X(\omega) + \sigma z_{1-\alpha/2}]$$

den Wert  $\mu$ . Wir können also

$$g_u(\omega) = X(\omega) - \sigma z_{1-\alpha/2}$$

als Z.V. für eine untere Schranke des Konfidenzintervalls auffassen, sowie

$$g_o(\omega) = X(\omega) + \sigma z_{1-\alpha/2}$$

als Z.V. für eine obere Schranke. Wir nennen dies auch ein zweiseitiges Konfidenzintervall. Wenn die Z.V.  $X$  einer anderen Verteilung genügt, müssen wir einfach andere Quantile benutzen!

Um ein einseitiges Konfidenzintervall anzugeben, müssen wir das  $(1 - \alpha)$ -Quantil  $z_{1-\alpha}$  der Normalverteilung suchen und können dann sagen: Mit W.K.  $1 - \alpha$  gilt

$$(-\infty, X(\omega) + \sigma z_{1-\alpha}]$$

den Wert  $\mu$ . Ebenso gilt aber auch, dass

$$[X(\omega) - \sigma z_{1-\alpha}, \infty)$$

den Wert  $\mu$  mit W.K.  $1 - \alpha$  enthält. Man spricht dann von einseitigen Konfidenzintervallen. Schon dieses Beispiel zeigt: Man muss **vorher** festlegen, welches Konfidenzintervall man bestimmen will (einseitig offen nach unten, einseitig offen nach oben oder zweiseitig). Sonst könnte man **nach** dem Experiment ja genau das Intervall herausuchen, das einem passt, was hochgradig unseriös ist!! Außerdem muss  $\alpha$  vorher festgelegt werden!

## 4.2 Konfidenzintervalle für den Erwartungswert

Wir nehmen an,  $X$  ist eine  $N(\mu, \sigma^2)$ -verteilte ZV. Zunächst einmal sei  $\sigma^2$  bekannt. Dann ist der empirische Mittelwert  $\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$  eine  $N(\mu, \sigma^2/n)$ -verteilte Z.V. (die Varianz von  $X_1 + \dots + X_n$  ist  $n\sigma^2$ , die Division durch  $n$  bewirkt eine Division der Varianz durch  $n^2$ ). Wir können also wörtlich das oben beschriebene Verfahren anwenden:

### Zweiseitiges Konfidenzintervall für den Erwartungswert bei bekannter Varianz

- Wähle  $\alpha$  (sollte klein sein) sowie das  $1 - \alpha/2$ -Quantil  $z_{1-\alpha/2}$  der Standardnormalverteilung.
- Berechne  $\bar{X}$ .
- Das Intervall

$$[\bar{X} - \sigma z_{1-\alpha/2}/\sqrt{n}, \bar{X} + \sigma z_{1-\alpha/2}/\sqrt{n}]$$

enthält den Erwartungswert  $\mu$  mit W.K.  $1 - \alpha$ .

### Einseitiges Konfidenzintervall für den Erwartungswert bei bekannter Varianz

- Wähle  $\alpha$  (sollte klein sein) sowie das  $(1 - \alpha)$ -Quantil  $z_{1-\alpha}$  der Standardnormalverteilung.
- Berechne  $\bar{X}$ .

- Das Intervall

$$(-\infty, \bar{X} + \sigma z_{1-\alpha}/\sqrt{n}]$$

enthält den Erwartungswert  $\mu$  mit W.K.  $1 - \alpha$ .

Ebenso kann man natürlich ein einseitiges Konfidenzintervall angeben, das nach oben unbeschränkt ist.

Interessant ist, dass man auch seriös ein Konfidenzintervall angeben kann, wenn man die Standardabweichung nicht kennt. Man muss allerdings immer noch eine Normalverteilungsannahme zugrunde legen. Klar ist, dass die Konfidenzintervalle jetzt größer sein müssen.

Grundlage ist Satz 3.14:

**Zweiseitiges Konfidenzintervall für den Erwartungswert bei unbekannter Varianz**

- Wähle  $\alpha$  (sollte klein sein) sowie das  $(1 - \alpha/2)$ -Quantil  $t_{n-1, 1-\alpha/2}$  der  $t$ -Verteilung mit  $n - 1$  Freiheitsgraden ( $n$  ist die Stichprobengröße).
- Berechne  $\bar{X}$  sowie die empirische Standardabweichung  $\bar{S}$ .
- Das Intervall

$$[\bar{X} - t_{n-1, 1-\alpha/2} \bar{S}/\sqrt{n}, \bar{X} + t_{n-1, 1-\alpha/2} \bar{S}/\sqrt{n}]$$

enthält den Erwartungswert  $\mu$  mit W.K.  $1 - \alpha$ .

Entsprechend kann man auch einseitige Intervalle angeben.

**Zweiseitiges Konfidenzintervall für die Varianz einer normalverteilten Zufallsvariable**

Wir können auch Konfidenzintervalle für  $\sigma^2$  angeben, weil  $(n - 1)\bar{S}^2/\sigma^2$  eine Z.V. ist, die  $\chi^2(n - 1)$ -verteilt ist, sofern wir eine Stichprobe vom Umfang  $n$  einer normalverteilten Z.V. ziehen. Beachte aber hier, dass die  $\chi^2$ -Verteilung nicht symmetrisch ist, wir also das  $(1 - \alpha)$ -Quantil nicht so einfach aus dem  $\alpha$ -Quantil bestimmen können wie bei der  $t$ -Verteilung und der BNormalverteilung.

- Wähle  $\alpha$  sowie das  $\alpha/2$ - und  $(1 - \alpha/2)$ -Quantil der  $\chi^2(n - 1)$ -Verteilung. Diese seien  $c_{n-1, \alpha/2}$  und  $c_{n-1, 1-\alpha/2}$ .
- Berechne  $\bar{S}^2$ .
- Dann gilt

$$P(c_{n-1, \alpha/2} \leq \frac{(n - 1)\bar{S}^2}{\sigma^2} \leq c_{n-1, 1-\alpha/2}) = 1 - \alpha).$$

Also ist

$$\left[ \frac{(n - 1)\bar{S}^2}{c_{n-1, 1-\alpha/2}}, \frac{(n - 1)\bar{S}^2}{c_{n-1, \alpha/2}} \right]$$

ein Konfidenzintervall für  $\sigma^2$  zum Konfidenzniveau  $1 - \alpha$ .

### 4.3 Konfidenzintervall für W.K. eines Bernoulliexperimentes

Unter Benutzung der Approximation der Binomialverteilung durch eine Normalverteilung kann man folgendes Konfidenzintervall für  $p$  in einem Bernoulli-Experiment angeben, wenn dieses Experiment  $n$ -mal wiederholt wird. Dabei sollte  $n > 20$  gelten und der Wert  $k/n$  nicht zu nahe bei 1 oder 0 liegen:

Wenn ein Bernoulli-Experiment mit unbekannter W.K.  $p$  für das Eintreten eines Ereignisses  $n$  mal durchgeführt wird und dabei dieses Ereignis  $k$  mal eintritt, so enthält das Intervall

$$\left[ \frac{k - \frac{1}{2} + \frac{c^2}{2} - c\sqrt{k - \frac{1}{2} - n^{-1}(k - \frac{1}{2})^2 + \frac{c^2}{4}}}{n + c^2}, \frac{k + \frac{1}{2} + \frac{c^2}{2} + c\sqrt{k + \frac{1}{2} - n^{-1}(k + \frac{1}{2})^2 + \frac{c^2}{4}}}{n + c^2} \right]$$

den Wert  $p$  mit W.K.  $1 - \alpha$ . Hierbei ist  $c = z_{1-\alpha/2}$  wieder das  $(1 - \alpha/2)$ -Quantil der Normalverteilung. Wir haben hierbei die Approximation der Binomialverteilung durch die Normalverteilung benutzt.

Für  $k, n - k > 50$  kann man dieses Intervall durch

$$\left[ \frac{k}{n} - \frac{c}{\sqrt{n}} \cdot \sqrt{\frac{k}{n} \left(1 - \frac{k}{n}\right)}, \frac{k}{n} + \frac{c}{\sqrt{n}} \cdot \sqrt{\frac{k}{n} \left(1 - \frac{k}{n}\right)} \right]$$

ersetzen.

### 4.4 Wahl der Stichprobengröße

Bei der Bestimmung eines Konfidenzintervalls gehen neben der Stichprobe auch  $n$  und  $\alpha$  ein. Wenn  $n$  vorgegeben ist, werden die Konfidenzintervalle natürlich immer größer, je kleiner  $\alpha$  wird. Nun kann man ja auch versuchen, zu gegebenem  $\alpha$  ein  $n$  so zu finden, dass das Intervall nicht zu groß wird.

Schauen wir uns dazu das Beispiel eines zweiseitigen Konfidenzintervalls an für den Erwartungswert einer normalverteilten Zufallsvariable bei bekannter Varianz an. Ein Konfidenzintervall zum Niveau  $\alpha$  ist dann

$$[\bar{X} - \sigma z_{1-\alpha/2}/\sqrt{n}, \bar{X} + \sigma z_{1-\alpha/2}/\sqrt{n}]$$

Wenn die Intervalllänge höchstens  $\delta$  sein soll, dann gilt  $\delta \geq 2(z_{1-\alpha/2}\sigma/\sqrt{n})$ , also

$$n \geq \frac{4(z_{1-\alpha/2})^2 \sigma^2}{\delta^2}.$$

### 4.5 Hypothesentest

Bei der Angabe von Konfidenzintervallen sind wir stets von der Gültigkeit einer Hypothese ausgegangen, dass beispielsweise eine normalverteilte Z.V. mit Erwartungswert  $\mu$  vorliegt,

obwohl wir  $\mu$  nicht kennen. Wir haben dann ein Intervall angegeben, das  $\mu$  mit einer gewissen (hohen) W.K. enthält. Grob gesagt würde man sagen, die Annahme über  $\mu$  ist vielleicht nicht richtig, wenn das Konfidenzintervall den Wert  $\mu$  nicht enthält.

Etwas genauer:

Wir haben eine Zufallsvariable  $X$ , die von einem unbekanntem Parameter  $\theta$  abhängt, wobei  $\theta \in \Theta$  in einem Parameterraum  $\Theta$  liegt. Eigentlich sollten wir also  $X_\theta$  schreiben. Wir wollen nun “testen”, ob  $\theta \in \Theta_0$  oder  $\theta \in \Theta_1$  gilt. Dabei müssen  $\Theta_0$  und  $\Theta_1$  natürlich disjunkt sein und es soll  $\Theta = \Theta_0 \cup \Theta_1$  gelten. Wir nennen die Hypothese  $\theta \in \Theta_0$  die (Null)hypothese und die Alternative  $\theta \in \Theta_1$  die Gegenhypothese.

Wenn  $|\Theta_0| = 1$  aus nur einem Element besteht, nennen wir dies eine einfache Hypothese. Die einzige Möglichkeit, etwas über die Hypothese zu erfahren, ist, sich eine konkrete Realisierung  $x$  der Zufallsvariable anzuschauen. Mit  $\mathcal{X}$  bezeichnen wir die Menge aller möglichen Realisierungen von  $X$ . Wir teilen  $\mathcal{X}$  in zwei disjunkte Mengen  $\mathcal{X}_0$  und  $\mathcal{X}_1$  auf. Wenn dann die Realisierung der Z.V.  $X$  in  $\mathcal{X}_0$  liegt, so entscheiden wir uns für  $H_0$ , andernfalls für  $H_1$ . Unsere Entscheidung kann richtig sein, aber auch falsch. So können wir uns für  $H_1$  entscheiden, obwohl  $H_0$  richtig ist (**Fehler erster Art**) oder genau umgekehrt, wir entscheiden uns für  $H_0$ , obwohl  $H_1$  richtig wäre (**Fehler zweiter Art**).

Die Zufallsvariable  $X_\theta$  hat eine Wahrscheinlichkeitsverteilung  $P_\theta$ . Als Gütefunktion  $g$  bezeichnen wir die Abbildung

$$g: \Theta \rightarrow [0, 1] \\ \theta \mapsto P_\theta(x \in \mathcal{X}_1).$$

Die Gütefunktion sollte auf  $\Theta_0$  klein sein, um den Fehler 1. Art klein zu halten, und gleichzeitig muss  $g$  auf  $\Theta_1$  groß sein, um den Fehler 2. Art klein zu halten.

Bei einem statistischen Test müssen wir nun  $\mathcal{X}_0$  und  $\mathcal{X}_1$  angeben. Wir nennen einen solchen Test einen Test zum Signifikanzniveau  $\alpha$  wenn  $g(\theta) \leq \alpha$  für alle  $\theta \in \Theta_0$  gilt. Wir wählen  $\mathcal{X}_1$  so, dass das Eintreten des Ereignisses  $x \in \mathcal{X}_1$  unter der Annahme  $\theta \in \Theta_0$  mit W.K.  $\leq \alpha$  eintritt. Aber genau das haben wir bei der Angabe von Konfidenzintervallen gemacht, zumindest wenn wir eine einfache Hypothese testen wollen.

Der populärste Test ist der sogenannte Gauß-Test. Dabei wollen wir den unbekanntem Erwartungswert einer normalverteilten Zufallsvariablen  $X$  “testen”. Übliche Hypothesen sind

- $H_0: \mu = \mu_0$  versus  $H_1: \mu \neq \mu_0$
- $H_0: \mu \geq \mu_0$  versus  $H_1: \mu < \mu_0$  (einseitiger Test)
- $H_0: \mu \leq \mu_0$  versus  $H_1: \mu > \mu_0$  (einseitiger Test).

Wir testen durch  $n$ -maliges Wiederholen des entsprechenden Zufallsexperimentes. Wir geben im Fall des zweiseitigen Gauß-Tests ein zweiseitiges Konfidenzintervall an (das ist unser Annahmehereich für die Nullhypothese). Im Fall des Tests  $\mu \geq \mu_0$  (als Nullhypothese) versus der Alternative  $\mu < \mu_0$  wählen wir als Annahmehereich ein Intervall  $[z, \infty)$ ,

als  $\mathcal{X}_1$  also  $(-\infty, z)$ . Unter der Annahme  $\mu = \mu_0$  gilt also  $\bar{X} \geq z$  mit W.K.  $1 - \alpha$  und somit gilt  $\bar{X} < z$  mit W.K.  $\alpha$ . Nun soll aber die W.K. für  $\bar{X} < z$  nicht nur unter der Hypothese  $\mu = \mu_0$  einen Wert  $\leq \alpha$  annehmen, sondern für alle  $\mu \geq \mu_0$ . Das ist in diesem Fall aber klar, denn wenn der Erwartungswert  $\mu$  größer wird, sinkt natürlich die W.K., dass  $\bar{X}$ , der empirische Mittelwert, klein wird.

## 5 Numerik

### 5.1 Approximation

Wir betrachten die folgende Situation: Gegeben seien  $n + 1$  sogenannte **Stützstellen**  $(x_i, y_i)$ ,  $x_i, y_i \in \mathbb{R}$   $i = 0, \dots, n$ . Wir nehmen  $x_i \neq x_j$  für  $i \neq j$  an. Dann gibt es genau ein Polynom  $p(x) = \sum_{i=0}^n x^i$  vom Grad  $\leq n$  mit  $p(x_i) = y_i$  für  $i = 0, \dots, n$ . Wir nennen  $p$  das **Interpolationspolynom** bzgl. den Stützstellen  $(x_i, y_i)$ .

Im Prinzip kann man die  $a_i$  aus einem linearen Gleichungssystem bestimmen:

$$\begin{pmatrix} 1 & x_0 & x_0^2 & \cdots & x_0^n \\ 1 & x_1 & x_1^2 & \cdots & x_1^n \\ \vdots & & & & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^n \end{pmatrix} \cdot \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} y_0 \\ y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

#### Lagrange

Eine Möglichkeit ist, die sogenannten **Lagrangepolynome**

$$L_i(x) = \prod_{j=0, j \neq i}^n \frac{x - x_j}{x_i - x_j}$$

zu definieren. Dann gilt

$$p(x) = \sum_{i=0}^n y_i L_i(x).$$

Der Vorteil hier ist, dass wir in einem “pre-processing” die  $L_i$  bestimmen können und dann sofort die Interpolationspolynome daraus ablesen können. Bei der Bestimmung der  $L_i$  muss der Grad aber vorher bekannt sein! Wenn man also evtl. Stützstellen hinzufügen muss, muss man ganz von vorne anfangen zu rechnen. Deshalb ist folgender rekursive Ansatz manchmal vorzuziehen:

#### Neville-Verfahren

Sei  $p_{i,j}(x)$  ein Interpolationspolynom zu den Stützstellen  $(x_k, y_k)$ ,  $k = i - j, \dots, i$ , wobei  $i \geq j$ . Für  $j = 0$  gilt also  $p_{i,0}(x) = y_i$ ,  $i = 0, \dots, n$ . Nun kann man  $p_{i,j}$  aus  $p_{i-1,j-1}$  und  $p_{i,j-1}$  berechnen: Es gilt

$$p_{i,j}(x) = \frac{(x - x_{i-j})p_{i,j-1}(x) - (x - x_i)p_{i-1,j-1}(x)}{x_i - x_{i-j}}. \quad (1)$$

Man rechnet leicht nach, dass das so definierte  $p_{i,j}$  ein Interpolationspolynom für die Stützstellen  $(x_k, y_k)$  mit  $k = i - j, \dots, i$  ist.

Man wendet dieses Verfahren weniger an, um ein Interpolationspolynom konkret zu bestimmen, sondern um das Polynom (ohne es explizit anzugeben) an einer Stelle  $x^*$  auszuwerten: Dann werden in (1) nämlich keine Polynome ausgerechnet, sondern einfach nur Zahlen (denn die Variable  $x$  wird dann zu der konkreten Zahl  $x^*$ ).

## Newton-Interpolation

Ein Beispiel eines “Neville-artigen”-Verfahrens ist die Newton-Interpolation. Wir versuchen hier, Zahlen  $a_i$ ,  $i = 0, \dots, n$  so zu bestimmen, dass

$$p(x) = \sum_{i=0}^n a_i \prod_{j=0}^{i-1} (x - x_j) \quad (2)$$

die Eigenschaft  $p(x_k) = y_k$  für  $k = 0, \dots, n$  hat. Polynome der Form (2) heißen Newton-Polynome.

Wir können genauso wie im Neville-Verfahren die  $p_{i,j}$  definieren, allerdings beschränken wir uns auf die Angabe des Koeffizienten  $\alpha_{i,j}$  von  $x^j$  (beachten Sie, dass die Polynome  $p_{i,j}$  den Grad  $\leq j$  haben). Dann erfüllen die  $\alpha_{i,j}$  die Rekursion

$$\alpha_{i,j} = \frac{\alpha_{i,j-1} - \alpha_{i-1,j-1}}{x_i - x_{i-j}}.$$

Ferner gilt

$$p(x) = \sum_{i=0}^n \alpha_{i,i} \prod_{j=0}^{i-1} (x - x_j),$$

d.h. wir haben mit Hilfe dieser Rekursion auch das Newton-Interpolationspolynom bestimmt: Der Grund dafür ist einfach die schöne Eigenschaft des Newton-Interpolationspolynoms  $p(x)$  zu  $n + 1$  Stützstellen, dass das Newtonpolynom  $q(x)$  nach Hinzufügen einer weiteren Stützstelle  $(x_{n+1}, y_{n+1})$  die Form  $q(x) = p(x) + a_{n+1} \prod_{j=0}^n (x - x_j)$  hat, denn  $q(x_j) = p(x_j)$  für  $j = 0, \dots, n$ . Der Wert  $a_{n+1}$  muss dann so bestimmt werden, dass  $q(x_{n+1}) = y_{n+1}$  gilt. Man kann das Neville-Verfahren sowie die Bestimmung des Newton-Interpolationspolynoms leicht in einem “Dreiecksschema” übersichtlich aufschreiben. Rechnerisch ist es eine simple Rekursion.

### Horner-Schema

Wenn wir ein Polynom  $f(x) = \sum_{i=0}^n a_i x^i$  an der Stelle  $x^*$  auswerten wollen, ist es sinnvoll, das Polynom in der Form

$$f(x) = (\dots (a_n x^* + a_{n-1})x^* + a_{n-2})x^* + \dots)x^* + a_0$$

zu schreiben. Man benötigt dann nur  $n$  Multiplikationen. Ähnlich ist es, wenn das Polynom in der Form (2) gegeben ist. Wir schreiben dann

$$p(x^*) = (\dots (a_n(x^* - x_{n-1}) + a_{n-1})(x^* - x_{n-2}) + \dots + a_1)(x^* - x_0) + a_0.$$



## Fehlerabschätzung

**Satz 5.1.** Sei  $f : [a, b] \rightarrow \mathbb{R}$  eine  $(n + 1)$ -mal stetig differenzierbare Funktion, und seien  $(x_i, f(x_i))$  Stützstellen mit  $x_i \in [a, b]$ , die zu einem Interpolationspolynom  $p$  führen. Dann gibt es zu jedem  $x^* \in [a, b]$  ein  $\zeta \in [a, b]$  so, dass

$$f(x^*) - p(x^*) = \frac{f^{(n+1)}(\zeta)}{(n+1)!} \prod_{i=0}^n (x^* - x_i).$$

## Splines

Die Idee hier ist, dass wir nicht ein Polynom durch alle Stützstellen  $(x_i, y_i)$  suchen, sondern durch je zwei aufeinanderfolgende Stützstellen soll ein Polynom vom Grad  $\leq k$  gehen, wobei dann an den Übergangsstellen gewisse Glattheitseigenschaften gelten (z.B. Stetigkeit, Differenzierbarkeit).

**Definition 5.2.** Eine Spline-Funktion vom Grad  $k$  zu den Stützstellen  $(x_i, y_i)$ ,  $i = 0, \dots, n$ , ist eine  $(k - 1)$ -mal stetig differenzierbare Funktion  $S : [x_0, x_n] \rightarrow \mathbb{R}$ , mit  $S(x_i) = y_i$ ,  $i = 0, \dots, n$  und  $S$  ist auf allen Teilintervallen  $[x_i, x_{i+1}]$ ,  $i = 0, \dots, n - 1$  ein Polynom  $p_i$  vom Grad  $\leq k$ .

Wir müssen also  $n$  Polynome vom Grad  $\leq k$  bestimmen, das bedeutet wir müssen  $n(k + 1)$  Koeffizienten bestimmen. Wie viele Bedingungen haben wir dafür?

- $p_i(x_i) = y_i$ ,  $i = 0 \dots n - 1$
- $p_i(x_{i+1}) = y_{i+1}$ ,  $i = 0, \dots, n - 1$
- $p_i^{(s)}(x_{i+1}) = p_{i+1}^{(s)}(x_{i+1})$ ,  $i = 0, \dots, n - 2$ ,  $s = 1, \dots, k - 1$ .

Das sind  $(n - 1)(k - 1) + 2n$  Forderungen, also fehlen noch  $k - 1$  Bedingungen. Im Fall  $k = 3$  oft:  $p'_0(x_0) = p'_{n-1}(x_n) = 0$ . Auch für größere Werte von  $k$  fordert man manchmal als zusätzliche Bedingungen, dass die Ableitungen am Rand verschwinden.

Sollte die Funktion periodisch und  $x_0$  und  $x_n$  die Intervallgrenzen einer Periode sein, so ist es sinnvoll zu fordern, dass die Ableitungen an den Stellen  $x_0$  und  $x_n$  übereinstimmen, also  $p_0^{(s)}(x_0) = p_{n-1}^{(s)}(x_n)$  für  $s = 1, \dots, k - 1$ .

## 5.2 Numerische Integration

Das Integral ist ja schon als Grenzwert von Ober- und Untersummen definiert, entsprechend kann man die Definition direkt benutzen, um das bestimmte Integral  $\int_a^b f(x)dx$  zu approximieren. Man wählt dazu eine Unterteilung des Intervalls  $[a, b]$ , also  $a = x_0 < x_1 < \dots < x_n = b$  und Zwischenwerte  $z_i \in [x_i, x_{i+1}]$ . Dann gilt

$$\int_a^b f(x)dx \approx \sum_{i=0}^{n-1} f(z_i) \cdot (x_{i+1} - x_i).$$

Man kann beispielsweise  $z_i = \frac{x_i + x_{i+1}}{2}$  wählen, aber auch  $z_i = x_i$  oder  $x_{i+1}$ .

Wir können natürlich die unbekannte Funktion auch durch ein Polynom vom Grad  $n$  approximieren und dieses dann integrieren. Betrachten wir die Situation, dass wir das Intervall  $[a, b]$  äquidistant unterteilen, also  $x_i = a + i \cdot h$ , wobei  $h = \frac{b-a}{n}$ ,  $i = 0, \dots, n$ . Wenn wir mit  $L_i$  die Lagrangepolynome zu diesen Stützstellen bezeichnen, also

$$L_i(x) = \prod_{j=0, j \neq i}^n \frac{x - x_j}{x_i - x_j} = \prod_{j=0, j \neq i}^n \frac{x - a - j \cdot h}{h(i - j)} = \frac{x/h - a/h - j}{i - j},$$

dann gilt (Integration durch Substitution!)

$$\int_a^b L_i(x) dx = h \cdot \int_0^n \prod_{j=0, j \neq i}^n \frac{x - j}{i - j} dx.$$

Wir definieren

$$\alpha_{i,n} = \int_0^n \prod_{j=0, j \neq i}^n \frac{x - j}{i - j} dx.$$

Die  $\alpha_{i,n}$  sind für kleine Werte:

	$i = 0$	$i = 1$	$i = 2$	$i = 3$
$n = 1$	$\frac{1}{2}$	$\frac{1}{2}$		
$n = 2$	$\frac{1}{3}$	$\frac{4}{3}$	$\frac{1}{3}$	
$n = 3$	$\frac{3}{8}$	$\frac{9}{8}$	$\frac{9}{8}$	$\frac{3}{8}$

Da die Lagrangeinterpolation  $p_n(x)$  einfach

$$p_n(x) = \sum_{i=0}^n y_i L_i(x)$$

ist, erhalten wir die Approximation

$$\int_a^b f(x) dx \approx h \cdot \sum_{i=0}^n \alpha_{i,n} y_i$$

und nennen dies die Newton-Cotes-Integrationsformeln. Wenn man diese beispielsweise anwendet, um  $\int_0^1 e^{-x^2} dx$  auszurechnen, erhält man

$n = 1$	0.74718
$n = 2$	0.74718
$n = 3$	0.74699

wobei der exakte Wert 0.74682 ist. In der Praxis unterteilt man zunächst das “große” Intervall in mehrere kleine und wendet auf diese kleinen Intervalle die Newton-Cotes-Formeln für kleine Werte von  $n$  an.

Man kann auch hier eine Fehlerabschätzung beweisen:

**Satz 5.3.** Sei  $f : [a, b] \rightarrow \mathbb{R}$  eine  $(n + 1)$ -mal stetig differenzierbare Funktion, und seien  $x_i, i = 0, \dots, n$  äquidistante Stützstellen (siehe oben) mit  $x_0 = a$  und  $x_n = b$ . Ferner sei  $h = (b - a)/n$ . Dann gilt

$$\left| \int_a^b f(x) dx - h \cdot \sum_{i=0}^n \alpha_{i,n} f(x_i) \right| = \mathcal{O}(h^{n+2}).$$

### 5.3 Nullstellenbestimmung

In der Praxis muss man sehr oft Nullstellen von Funktionen finden, z.B. auch beim Lösen von linearen Gleichungssystemen. Wir betrachten hier zunächst das Newtonverfahren:

**Satz 5.4** (Newton-Verfahren). Sei  $f : [a, b] \rightarrow \mathbb{R}$  eine zweimal stetig differenzierbare Funktion. Sei  $|f'(x)| \geq m$  und  $|f''(x)| \leq M$  für alle  $x \in [a, b]$  mit  $m, M > 0$  (das bedeutet z.B., dass  $f$  in dem Intervall  $[a, b]$  höchstens eine Nullstelle hat, denn sonst müsste es ja ein  $x \in [a, b]$  geben mit  $f'(x) = 0$ , denn zwischen zwei Nullstellen ist ein Maximum oder Minimum). Sei  $x^*$  eine Nullstelle in  $[a, b]$ , und sei  $x^{(0)} \in [a, b]$ . Für  $k \geq 0$  sei

$$x^{(k+1)} = x^{(k)} - \frac{f(x^{(k)})}{f'(x^{(k)})}$$

in  $[a, b]$ . Dann gilt

$$|x^{(k+1)} - x^*| \leq \frac{M}{2m} |x^{(k)} - x^*|^2.$$

**Bemerkung 5.5.** Man sagt auch, dass Newton-Verfahren konvergiert quadratisch gegen die Nullstelle. Nachteile des Newton-Verfahrens: Man muss nahe bei einer Nullstelle starten (sonst hat man evtl. gar keine Konvergenz) und man benötigt Ableitungen.

Man benutzt beim Newton-Verfahren lokale lineare Approximation, d.h. man bestimmt eine Gerade durch den Punkt  $[x^{(k)}, f(x^{(k)})]$  mit Steigung  $f'(x^{(k)})$  und bestimmt dann die Nullstelle dieser Geraden als neuen Punkt  $x^{(k+1)}$ . Anstatt solche “Tangenten” zu wählen, kann man auch Sekanten wählen, die zwei benachbarte Iterationspunkte  $[x^{(k-1)}, f(x^{(k-1)})]$  und  $[x^{(k)}, f(x^{(k)})]$  verbindet und bestimmt die Nullstelle dieser Geraden. Es gilt:

**Satz 5.6** (Sekanten-Verfahren). Sei  $f : [a, b] \rightarrow \mathbb{R}$  eine zweimal stetig differenzierbare Funktion. Sei  $|f'(x)| \geq m$  und  $|f''(x)| \leq M$  für alle  $x \in [a, b]$  mit  $m, M > 0$ . Sei  $x^*$  eine Nullstelle in  $[a, b]$ , und seien  $x^{(0)}, x^{(1)} \in [a, b]$ ,  $x^{(0)} \neq x^{(1)}$ . Für  $k \geq 1$  sei

$$x^{(k+1)} = x^{(k)} - f(x^{(k)}) \frac{x^{(k)} - x^{(k-1)}}{f(x^{(k)}) - f(x^{(k-1)})}.$$

Dann gilt in  $[a, b]$

$$|x^{(k+1)} - x^*| \leq \left(\frac{M}{2m}\right)^{\tau-1} |x^{(k)} - x^*|^\tau$$

mit  $\tau = (1 + \sqrt{5})/2 \approx 1.618$  (Goldener Schnitt).

Ein weiteres Verfahren ist das Bisektionsverfahren, bei dem wir als Voraussetzung nur fordern, dass  $f$  stetig ist. Wir beginnen mit zwei Startwerten  $a_0, b_0 \in [a, b]$  mit  $a_0 < b_0$  und  $f(a_0) \cdot f(b_0) < 0$ , d.h.  $f(a_0)$  und  $f(b_0)$  haben verschiedene Vorzeichen. Wir berechnen nun für  $k \geq 0$  ein neues Intervall  $[a_k, b_k]$  wie folgt: Die Mitte  $c$  des Intervalls  $[a_k, b_k]$  ist  $c = (a_k + b_k)/2$ . Wir erhalten nun  $a_{k+1}$  und  $b_{k+1}$  wie folgt:

- Gilt  $f(c) = 0$ : STOP.
- Haben  $f(a_k)$  und  $f(c)$  verschiedene Vorzeichen, setze  $a_{k+1} = a_k$  und  $b_{k+1} = c$ .
- Haben  $f(b_k)$  und  $f(c)$  verschiedene Vorzeichen, setze  $a_{k+1} = c$  und  $b_{k+1} = b_k$ .

Beachte, dass nicht sowohl  $f(a_k)$  und  $f(c)$  als auch  $f(c)$  und  $f(b_k)$  verschiedene Vorzeichen haben können, weil  $f(a_k)$  und  $f(b_k)$  verschiedene Vorzeichen haben (für  $k = 0$  und dann auch für  $k \geq 1$ ).

Wenn wir  $c$  nicht als Mittelpunkt des Intervalls  $[a_k, b_k]$  wählen sondern als Nullstelle der Sekante durch  $(a_k, f(a_k))$  und  $(b_k, f(b_k))$ , d.h.

$$c = a_k - \frac{f(a_k)(b_k - a_k)}{f(b_k) - f(a_k)},$$

so erhalten wir die **Regula Falsi**.

**Bemerkung 5.7.** Bei den Abschätzungen zur Konvergenz betrachten wir hier die Abweichung der Iterierten von der Nullstelle  $x^*$ . Ein anderes (vlt. sogar in manchen Anwendungen wichtigeres) Maß ist der Defekt  $|f(x^{(k)}) - f(x^*)|$ . Man könnte als Abbruchkriterien für ein Iterationsverfahren wählen, dass sich die Iterierten  $x^{(k)}$  in jedem Schritt kaum noch ändern (man also vermutlich nahe an der Nullstelle  $x^*$  liegt), oder aber, dass die Funktionswerte schon sehr nahe am gewünschten Funktionswert (im Fall einer Nullstellenbestimmung also nahe bei 0) liegen, der Defekt also klein ist.