Numerische Integration und Normalverteilung

Carl-Ferdinand Oliver[*], Tom Springer [†], Jonas Münz [‡]
Max Valentin Orlemann [§], Alexander Kögel [¶]
TES23

28. Oktober 2024

Inhaltsverzeichnis

1 Statistische Grundlagen 1
- 1.1 Normalverteilung . . . 1
- 1.2 Standardnormalverteilung . . . 2
- 1.3 t-Student-Verteilung . . . 2
- 1.4 χ^2-Verteilung . . . 2

2 Trapezregel 3
- 2.1 Herleitung . . . 3
- 2.2 Fehlerabschätzung . . . 5
- 2.3 Vertrauensniveau γ . . . 5
- 2.4 Alternativen zur Trapezregel . . . 5

3 Umsetzung im Programm 5
- 3.1 Programmablauf . . . 5
- 3.2 Konfidenzintervall . . . 5

[*] es23035@lehre.dhbw-stuttgart.de
[†] es23026@lehre.dhbw-stuttgart.de
[‡] es23039@lehre.dhbw-stuttgart.de
[§] es23023@lehre.dhbw-stuttgart.de
[¶] es23031@lehre.dhbw-stuttgart.de

1 Statistische Grundlagen

1.1 Normalverteilung

Die Gaußsche Normalverteilung beschreibt die Verteilung einer stetigen Zufallsvariablen und ist durch zwei Parameter, den Mittelwert μ und die Standardabweichung σ, charakterisiert. Sie wird auch als *Glockenkurve* bezeichnet und ist symmetrisch um den Mittelwert.

Wichtige Eigenschaften

- Die Dichtefunktion der Normalverteilung lautet:

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

- Die Verteilungsfunktion ist gegeben durch:

$$\Phi(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt$$

- Die Normalverteilung ist symmetrisch zum Mittelwert μ, hat ein Maximum bei $x = \mu$ und Wendepunkte bei $x = \mu \pm \sigma$.

- Sie ist normiert, d.h., das Integral über die gesamte Dichtefunktion ergibt 1:

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

Berechnung der Wahrscheinlichkeiten

Für die Berechnung von Wahrscheinlichkeiten, die der Gaußschen Normalverteilung folgen, werden in der Praxis oft Tabellen oder numerische Verfahren verwendet, da die Integrale analytisch nicht lösbar sind. Die Wahrscheinlichkeit für einen Bereich $a \leq X \leq b$ kann mit der Verteilungsfunktion berechnet werden:

$$P(a \leq X \leq b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

Wahrscheinlichkeiten in Abhängigkeit der Standardabweichung

Da die Normalverteilung symmetrisch um den Mittelwert μ ist, lassen sich bestimmte Wahrscheinlichkeitsbereiche in Abhängigkeit von der Standardabweichung σ um μ angeben:

- Etwa 68,27% aller Werte einer normalverteilten Zufallsvariablen liegen im Intervall $[\mu - \sigma, \mu + \sigma]$, also innerhalb einer Standardabweichung um den Mittelwert.

- Etwa 95,45% der Werte befinden sich im Intervall $[\mu - 2\sigma, \mu + 2\sigma]$, also innerhalb von zwei Standardabweichungen.

- Rund 99,73% der Werte liegen im Intervall $[\mu - 3\sigma, \mu + 3\sigma]$, also innerhalb von drei Standardabweichungen.

Diese Bereiche werden auch als *Empirische Regel* oder *68-95-99,7-Regel* bezeichnet und sind besonders nützlich zur Einschätzung, wie wahrscheinlich es ist, dass eine Zufallsvariable in einem bestimmten Bereich um den Mittelwert liegt. Die Wahrscheinlichkeit, dass ein Wert weiter als 3σ vom Mittelwert entfernt liegt, ist sehr gering und beträgt nur ca. 0,27%.
In der Praxis wird die Normalverteilung zur Modellierung vieler realer Phänomene verwendet, wie z.B. Fehlerverteilungen, Körpergrößen oder IQ-Werte, da viele natürliche Prozesse einem normalverteilten Muster folgen.

1.2 Standardnormalverteilung

Eine Normalverteilung lässt sich durch die Standardisierung mit $z = \frac{x-\mu}{\sigma}$ auf die sogenannte Standardnormalverteilung $\mathcal{N}(0, 1)$ zurückführen, die Mittelwert 0 und Standardabweichung 1 hat:

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{z^2}{2}\right)$$

Die Standardnormalverteilung ist ein nützliches Werkzeug in der Statistik, da sie es ermöglicht, beliebige Normalverteilungen durch eine Transformation auf diese Standardform zu bringen. Dies wird durch die Berechnung des *z-Scores* erreicht:

$$z = \frac{X - \mu}{\sigma}$$

Der z-Wert gibt an, wie viele Standardabweichungen ein Wert X vom Mittelwert entfernt ist. Diese Transformation erleichtert die Berechnung von Wahrscheinlichkeiten und statistischen Tests, da die Eigenschaften der Standardnormalverteilung bekannt sind und in Tabellen dokumentiert werden.

1.3 t-Student-Verteilung

Die **t-Student-Verteilung** wird verwendet, wenn die Stichprobengröße klein ist ($n < 30$) und die Varianz der Grundgesamtheit unbekannt ist. Sie ähnelt der Normalverteilung, hat jedoch „schwerere“ Ränder. Dies bedeutet, dass extremere Werte wahrscheinlicher sind als bei der Normalverteilung. Die t-Student-Verteilung hat einen Parameter, die Freiheitsgrade df, der die Form der Verteilung bestimmt. Je größer die Freiheitsgrade, desto mehr nähert sich die t-Verteilung der Normalverteilung an. Bei einer großen Stichprobengröße (ca. $df > 30$) ist der Unterschied zur Normalverteilung vernachlässigbar. In der Statistik wird die t-Verteilung häufig für Hypothesentests verwendet, insbesondere wenn es um kleine Stichproben geht. Sie spielt auch bei der Berechnung von Konfidenzintervallen eine wichtige Rolle, wenn die Varianz der Grundgesamtheit nicht bekannt ist.

1.4 χ^2-Verteilung

Die **χ^2-Verteilung** entsteht als Summe der quadrierten z-Werte aus einer Standardnormalverteilung. Sie wird häufig in der Statistik verwendet, um Varianzen zu schätzen und in Hypothesentests wie dem *Chi²-Anpassungstest* oder dem *Chi²-Test zur Unabhängigkeit*. Die Chi²-Verteilung hat einen Parameter, die Freiheitsgrade *df*, der die Form der Verteilung bestimmt. Sie ist asymmetrisch und nur für positive Werte definiert. Für $df \geq 30$ nähert sich die Chi²-Verteilung zunehmend der Normalverteilung an. Ein häufiges Anwendungsgebiet der Chi²-Verteilung ist die Analyse von Kontingenztafeln, bei denen überprüft wird, ob zwei kategorische Variablen unabhängig voneinander sind.

2 Trapezregel

2.1 Herleitung

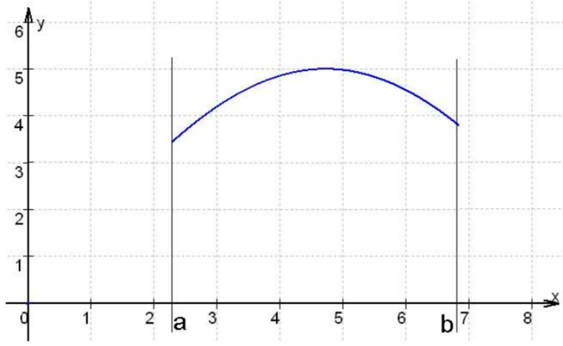

Abbildung 1: Keplersche Fassregel [1]

Wir betrachten eine Funktion f, deren Schaubild im gewünschten Intervall $I = [a, b]$ in Figure 2.1 gezeigt ist.

A. Die zwei Sehnentrapeze

Zur Flächenberechnung zeichnen wir zwei Sehnentrapeze in das gegebene Schaubild, wie in Figure 2.1 zu sehen ist. Für die Fläche eines Trapezes gilt:

$$A_{\text{Trapez}} = \frac{a+c}{2} \cdot h = m \cdot h$$

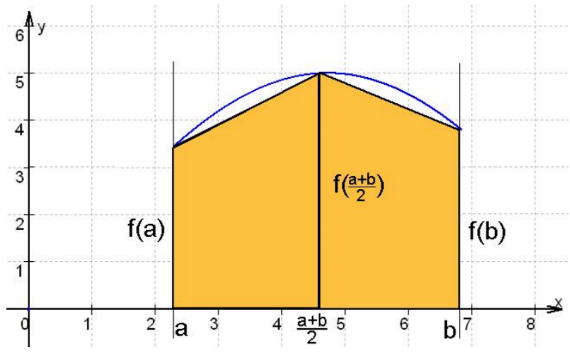

Abbildung 2: Sehnentrapeze für die anstehende Rechnung

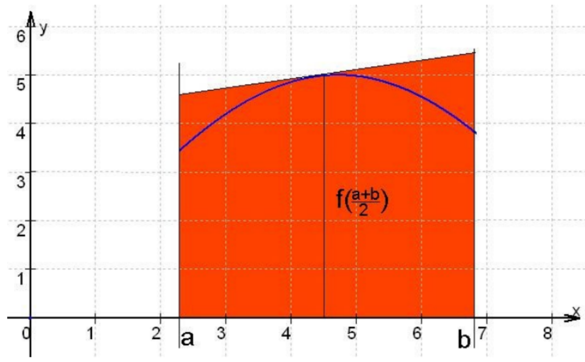

Abbildung 3: Tangententrapez [1]

Übertragen auf die beiden Sehnentrapeze ergibt sich die Gesamtfläche S als:

$$S = \frac{b-a}{2} \cdot \left(\frac{f(a)}{2} + f\left(\frac{a+b}{2}\right) + \frac{f(b)}{2} \right)$$

Dies lässt sich umschreiben zu:

$$S = \frac{b-a}{2} \cdot \left(\frac{f(a)}{2} + f\left(\frac{a+b}{2}\right) + \frac{f(b)}{2} \right)$$

B. Das Tangententrapez

Nun legen wir ein weiteres Tangententrapez in das Schaubild, wie in Figure 2.1 dargestellt. Nach der Trapezformel ergibt sich:

$$T = f\left(\frac{a+b}{2}\right) \cdot (b-a)$$

Da wir doppelt so viele Sehnentrapeze wie Tangententrapeze haben, gewichten wir die Flächen entsprechend. Die Gesamtfläche $I[f]$ nähert sich durch:

$$I[f] \approx A = \frac{1}{3} \cdot (2S + T)$$

Dies lässt sich weiter vereinfachen zu:

$$A = \frac{b-a}{6} \cdot \left(f(a) + 4 \cdot f\left(\frac{a+b}{2}\right) + f(b) \right)$$

Die Keplersche Fassregel

Ist die Funktion f auf dem Intervall $I = [a, b]$ stetig, so gilt die Keplersche Fassregel:

$$I[f] \approx A = \frac{b-a}{6} \cdot \left(f(a) + 4 \cdot f\left(\frac{a+b}{2}\right) + f(b) \right)$$

Dabei ist zu beachten, dass die Keplersche Fassregel nur dann gute Näherungswerte liefert, wenn sich die Funktion im betrachteten Intervall durch eine Parabel annähern lässt. Daher ist es ratsam, das Intervall $[a, b]$ in viele kleine, gleich große Teilintervalle zu unterteilen und die Keplersche Fassregel auf jedes Teilintervall anzuwenden. Diese Methode wird im nächsten Abschnitt besprochen, in dem die summierten Regeln behandelt werden. [1]

2.2 Fehlerabschätzung

2.3 Vertrauensniveau γ

2.4 Alternativen zur Trapezregel

3 Umsetzung im Programm

3.1 Programmablauf

3.2 Konfidenzintervall

Literatur

[1] J. Gehrke. *Skript zur Mathematik 4 Mathematische Anwendungen*. Stand: 2022, ZDM und SKT an der DHBW Stuttgart, Abgerufen am 28.10.2024. 2022.