Das Gaußsche Fehlerintegral Φ

Ac 2015-2018

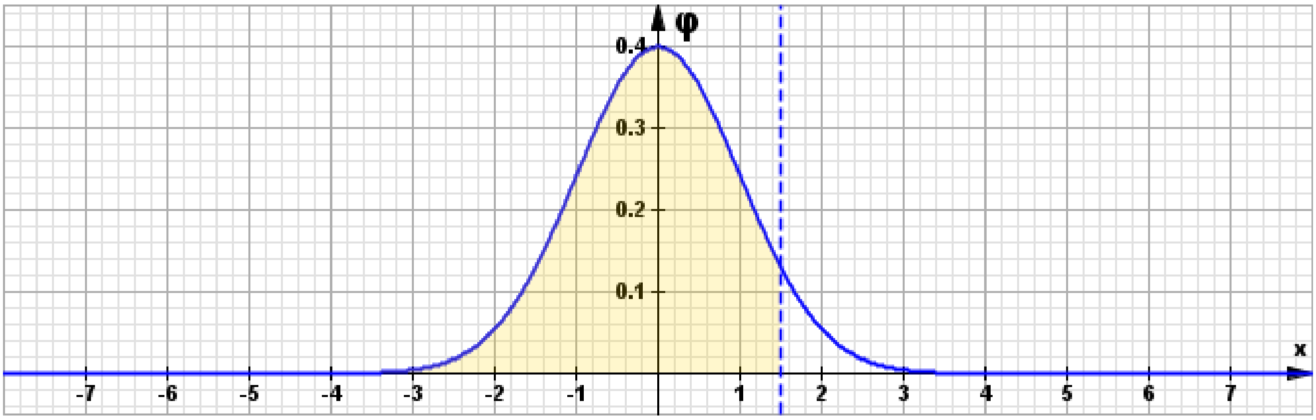
Das Gaußsche Fehlerintegral Φ ist definiert als das Integral über der Standard-Normalverteilung

$\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-0,5x^2}$ in den Grenzen $-\infty$ bis x , also

$$\boxed{\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-0,5t^2} dt \quad \text{(Gaußsches Fehlerintegral)}}$$

Dieses Integral stellt den Flächeninhalt zwischen der Standard-Normalverteilung im Intervall $]-\infty ; x]$ und der x-Achse dar und gibt die Wahrscheinlichkeit P dafür an, dass ein normalverteilter Merkmalswert X mit dem Erwartungswert $\mu = 0$ und der Standardabweichung $\sigma = 1$ höchstens den Wert x annimmt, kurz : **P(X ≤ x)** .

Beispiel: P(X ≤ 1,5)



Die schraffierte Fläche beträgt ca. A = 0,9332 , und dies entspricht der Wahrscheinlichkeit P(X ≤ 1,5).

Da die Normalverteilung keine Stammfunktion besitzt, muss man eine solche Fläche approximativ bestimmen. Hinzu kommt noch das Problem, dass die linke Grenze im negativen Unendlichen liegt.

Das letztgenannte Problem kann man jedoch durch die Tatsache lösen, dass die gesamte zwischen Normalverteilung und x-Achse liegende Fläche exakt den Wert 1 besitzt, was durch den wahrscheinlichkeitstheoretischen Zusammenhang bedingt ist. Die Fläche im Intervall $]-\infty ; 0]$ besitzt demnach den Wert 0,5. Man muss also nur im Intervall [0 ; x] approximieren und 0,5 zum Ergebnis addieren . Ist $x < 0$, so muss man den approximierten Flächenwert von 0,5 subtrahieren .

Wie kann man nun im Intervall **[0 ; x]** approximieren ?

Die gängigsten nummerischen Integrationsverfahren sind Trapezverfahren und Romberg-Verfahren, dazu noch das recht komplizierte Gauß-Kronrod-Verfahren.

Eine weitere Methode besteht darin, das **Taylorpolynom** der Normalverteilung zu ermitteln und dieses dann zu integrieren. So erhält man eine ganzrationale Näherungsfunktion als Stammfunktion, welche sich leicht berechnen lässt.

Die (Integral-)Funktion, welche den Bereich auf das Intervall [0 ; x] einschränkt, heißt **$\Phi_0(x)$** .

Gesucht ist also

$$\boxed{\Phi(x) = 0,5 + \Phi_0(x) \quad \text{mit} \quad \Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int_{0}^{x} e^{-0,5t^2} dt}$$

Das Taylorpolynom wird lediglich vom Integranden $e^{-0,5t^2}$ ermittelt (Entwicklung in der Umgebung von $x_o = 0$), so dass wir nach der Integration dieses Polynoms nur noch mit dem konstanten Faktor vor dem Integralzeichen multiplizieren müssen !

Das allgemeine Taylorpolynom für die Approximation an einer Stelle x_0 lautet:

$$T(x) = f(x_0) + f'(x_0)(x-x_0) + \frac{f''(x_0)}{2!}(x-x_0)^2 + \frac{f'''(x_0)}{3!}(x-x_0)^3 + \ldots + \frac{f^{(n)}(x_0)}{n!}(x-x_0)^n + R_n$$

Für das Restglied R_n gilt: $R_n = \dfrac{(x-x_0)^{n+1}}{(n+1)!} \cdot f^{(n+1)}(x_0 + \vartheta(x-x_0));\quad 0<\vartheta<1$

Wir ermitteln jetzt das Taylorpolynom von $f(x) = e^{-0,5x^2}$ der Stelle $x_0 = 0$:
Zunächst werden Ableitungen benötigt. Wir beschränken uns auf die ersten 8 Ableitungen .

$$f'(x) = e^{-0,5x^2} \cdot (-0,5 \cdot 2x) = -x \cdot e^{-0,5x^2} \qquad f''(x) = -x \cdot (-x \cdot e^{-0,5x^2}) + (-1) \cdot e^{-0,5x^2} = (x^2-1) \cdot e^{-0,5x^2}$$

$$f'''(x) = (x^2-1) \cdot (-x \cdot e^{-0,5x^2}) + 2x \cdot e^{-0,5x^2} = (-x^3+3x) \cdot e^{-0,5x^2}$$

Ebenso erhält man

$$f^{(4)}(x) = (x^4 - 6x^2 + 3) \cdot e^{-0,5x^2} \qquad f^{(5)}(x) = (-x^5 + 10x^3 - 15x) \cdot e^{-0,5x^2}$$

$$f^{(6)}(x) = (x^6 - 15x^4 + 45x^2 - 15) \cdot e^{-0,5x^2} \qquad f^{(7)}(x) = (-x^7 + 21x^5 - 105x^3 + 105x) \cdot e^{-0,5x^2}$$

$$f^{(8)}(x) = (x^8 - 28x^6 + 210x^4 - 420x^2 + 105) \cdot e^{-0,5x^2}$$

Mit $f(0)=1$, $f'(0) = 0$, $f''(0) = -1$, $f'''(0) = 0$, $f^{(4)}(0) = 3$, $f^{(5)}(0) = 0$, $f^{(6)}(0) = -15$, $f^{(7)}(0) = 0$, $f^{(8)}(0) = 105$ erhalten wir das Taylorpolynom:

$$f(x) = e^{-0,5x^2} = 1 - \frac{1}{2!}x^2 + \frac{3}{4!}x^4 - \frac{15}{6!}x^6 + \frac{105}{8!}x^8 \mp \ldots$$

Die Stammfunktion dieses Polynoms ist eine Näherung für das gesuchte Integral. Daher gilt:

$$\int_0^x e^{-0,5t^2}\,dt = x - \frac{1}{2!\cdot 3}x^3 + \frac{3}{4!\cdot 5}x^5 - \frac{15}{6!\cdot 7}x^7 + \frac{105}{8!\cdot 9}x^9 \mp \ldots$$

Dies lässt sich vereinfachen, damit man eine Summenformel angeben kann:

$$\int_0^x e^{-0,5t^2}\,dt = x - \frac{1}{2\cdot 3}x^3 + \frac{3}{24\cdot 5}x^5 - \frac{15}{720\cdot 7}x^7 + \frac{105}{40320\cdot 9}x^9 \mp \ldots$$

$$= x - \frac{1}{2\cdot 3}x^3 + \frac{1}{8\cdot 5}x^5 - \frac{1}{48\cdot 7}x^7 + \frac{1}{384\cdot 9}x^9 \mp \ldots$$

$$= x - \frac{1}{2^1\cdot 1\cdot 3}x^3 + \frac{1}{2^2\cdot 2\cdot 5}x^5 - \frac{1}{2^3\cdot 6\cdot 7}x^7 + \frac{1}{2^4\cdot 24\cdot 9}x^9 \mp \ldots$$

$$= x - \frac{1}{2^1\cdot 1!\cdot 3}x^3 + \frac{1}{2^2\cdot 2!\cdot 5}x^5 - \frac{1}{2^3\cdot 3!\cdot 7}x^7 + \frac{1}{2^4\cdot 4!\cdot 9}x^9 \mp \ldots$$

Die Summenformel (gültig für x in einer nicht zu großen Umgebung von 0) :

$$\int_0^x e^{-0,5t^2}\,dt = x - \frac{1}{2^1\cdot 1!\cdot 3}x^3 + \frac{1}{2^2\cdot 2!\cdot 5}x^5 - \frac{1}{2^3\cdot 3!\cdot 7}x^7 + \frac{1}{2^4\cdot 4!\cdot 9}x^9 \mp \ldots = \sum_{i=0}^{\infty}(-1)^i \cdot \frac{x^{2i+1}}{i!\cdot 2^i\cdot(2i+1)}$$

Daher gilt auch für das Gaußsche Fehlerintegral $\Phi_0(x)$:

$$\boxed{\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \cdot \int_0^x e^{-0{,}5t^2}\,dt = \frac{1}{\sqrt{2\pi}} \cdot \sum_{i=0}^{\infty} (-1)^i \cdot \frac{x^{2i+1}}{(2i+1)\cdot 2^i \cdot i!}}$$

Anmerkung: Das so genannte Restglied (zur Bestimmung des maximalen Approximationsfehlers) ist außerordentlich unhandlich, da sich nur schwer eine geschlossene Form für die (n+1)-te Ableitung von f finden lässt. Es soll daher nicht weiter betrachtet werden .

Hier eine Umsetzung dieses Summenalgorithmus in JAVA:

```java
public static double summeGaussFehlerIntPhi0(double x, int nmax) {
      // Summe((-1)^i*x^(2i+1)/(2i+1)/i!/2^i,i,0,n) ist für kleine x eine
      // Approximation für integral(e^(-t^2/2),t,0,x)
      // Der folgende Algorithmus ist sehr schnell, da Fakultät und Potenz 2^(i+1)
      // nicht immer wieder ab i=0 berechnet werden müssen
      int i = 0, vz = 1;
      long zweiHochI = 1, ZweiIplusEins = 1;
      double fak = 1; // double wichtig, da long nur bis 20! rechnet
      double summand = 1, summe = x, xPotenz = x;
      while (i < nmax || summand > 1e-10) {
            i++;   vz = -vz;    fak = fak * i;
            zweiHochI = zweiHochI * 2;  ZweiIplusEins = ZweiIplusEins + 2;
            xPotenz = xPotenz * x * x;
            summand = vz * xPotenz / ZweiIplusEins / zweiHochI / fak;
            // if (i>20) System.out.println("i = "+i+"  fak = "+fak);
            summe += summand;
      };
      return summe / Math.sqrt(2*Math.PI);
}
```

Wir prüfen den Φ_0 -Algorithmus für das Beispiel x = 1 mit verschiedenen n-Werten :

Berechnet wird also: $$\frac{1}{\sqrt{2\pi}} \cdot \int_0^1 e^{-0{,}5t^2}\,dt \approx \frac{1}{\sqrt{2\pi}} \cdot \sum_{i=0}^{n} (-1)^i \cdot \frac{1}{(2i+1)\cdot 2^i \cdot i!}$$

Zum Vergleich: Näherungswert auf 20 Dezimalen) **0.34134474606854298129**

n	Näherungswert(wxMaxima)	Näherungswert(Java 7; 17 Dezimalen)
0	0.39894228040143268	0.39894228040143270
1	0.33245190033452725	0.33245190033452726
...	...	...
9	0.34134474606363487	0.34134474606363490
10	0.34134474606874730	0.34134474606874730
11	0.34134474606853515	0.34134474606853515
12	0.34134474606854330	0.34134474606854330
13	0.34134474606854300	0.34134474606854304
14	0.34134474606864300	0.34134474606854304

Wie man sieht, sind 14 Nachkommastellen richtig, und zwar ab n = 12.
Genauer wird es nicht mehr, da der Konvergenzradius der Taylorreihe begrenzt ist.
Offensichtlich war x=1 schon recht weit „draußen" gewählt !
Betrachtet man x = 0,5 , so erhält man mit 0.19146246127401308 (Maxima sowie Java)

Untersuchungen zeigen, dass obiger Algorithmus für ca. |x| < 8,5 recht brauchbar ist !!

Zusätzliche Betrachtung:

Wie kann man Wahrscheinlichkeiten **P(X ≤ x)** für <u>beliebige Normalverteilungen</u> (mit μ, σ) berechnen ?

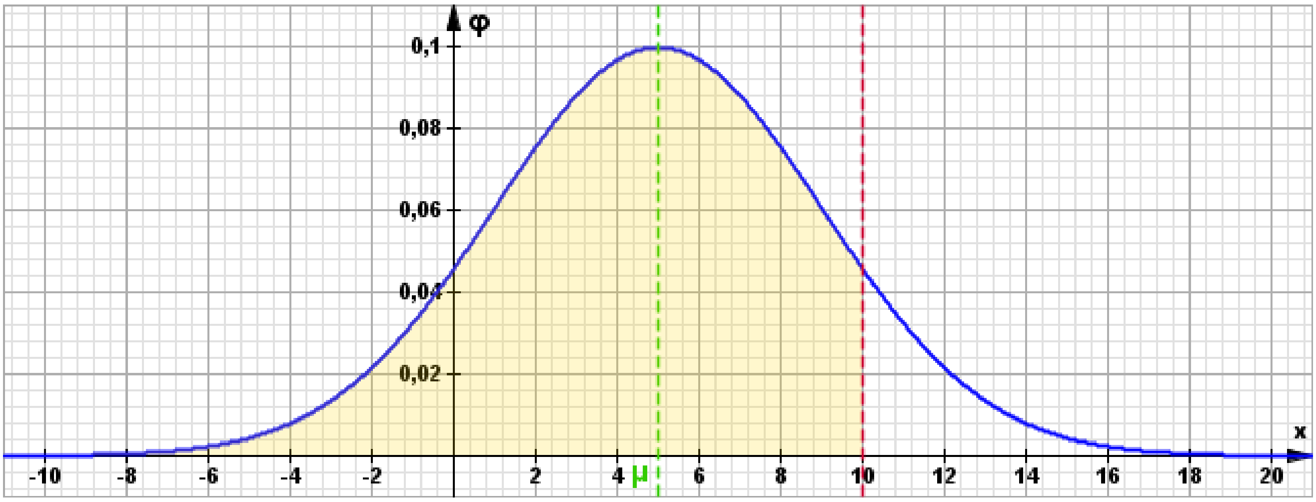
Zu berechnen ist $$\boxed{P(X \le x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-0,5(\frac{t-\mu}{\sigma})^2} dt}$$

Glücklicherweise kann man hier die obige Vorarbeit verwenden. Es gilt nämlich:

$$\mathbf{\Phi\ (x\ ;\ \mu\ ;\ \sigma) = \Phi\ (\ (x - \mu)\ /\ \sigma\ )}$$

Man rechnet also einfach mit der Φ(x) - Approximation, ersetzt jedoch x durch (x - μ) / σ .

Grafische Darstellung mit μ = 5 und σ = 4 und x = 10 :



Die schraffierte Fläche beträgt ca. A = 0,89435 , und dies entspricht der Wahrscheinlichkeit P(X ≤ 10).

Die Fehlerfunktion „erf (x)“ :

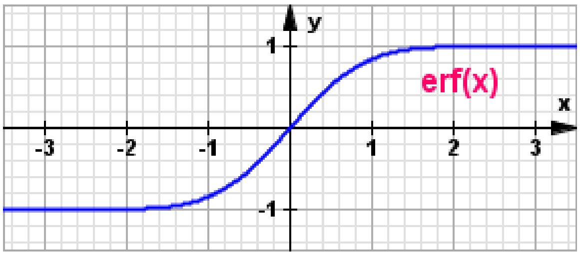
In der Stochastik verwendet man u.a. eine sog. „**error function**“ namens **erf (x)**. Diese ist nicht identisch mit dem hier betrachteten Gaußschen Fehlerintegral, es gibt aber einen Zusammenhang .

erf(x) gibt den prozentualen Anteil derjenigen Messwerte an, die zwischen 0 und x liegen (also zu groß sind), aber um weniger als x vom wahren Wert abweichen.

Definition:

$$erf(x) = \frac{2}{\sqrt{\pi}} \cdot \int_0^x e^{-t^2} dt = \frac{2}{\sqrt{\pi}} \cdot \sum_{i=0}^{\infty} (-1)^i \cdot \frac{x^{2i+1}}{(2i+1) \cdot i!}$$

Grafik für erf (x) :



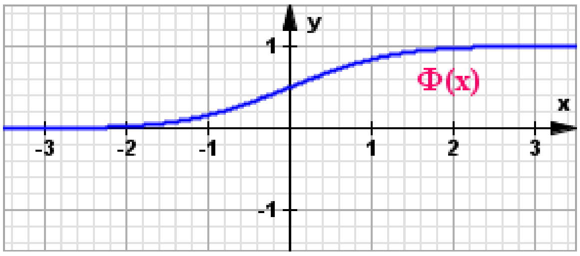
erf (x) ist eine ungerade Funktion, also gilt:
erf (-x) = - erf (x)

Zwischen der Fehlerfunktion und dem Fehlerintegral ist folgender Zusammenhang gegeben:

$$\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \cdot \int_0^x e^{-0,5t^2} dt = \frac{1}{2} \cdot erf(\frac{x}{\sqrt{2}})$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^x e^{-0,5t^2} dt = \frac{1}{2} \cdot [1 + erf(\frac{x}{\sqrt{2}})]$$

Grafik für das Fehlerintegral Φ (x) :



Zur nummerischen Berechnung von **erf (x)** gibt es Näherungsformeln (Fehler maximal: $1{,}2 \cdot 10^{-7}$) :

```
t = 1 / (1 + 0.5 * |x|)
expo = -|x|² - 1.26551223 + 1.00002368*t + 0.37409196*t² + 0.09678418*t³ - 0.18628806*t^4 +
       0.27886807*t^5 - 1.13520398*t^6 + 1.48851587*t^7 - 0.82215223*t^8 + 0.17087277*t^9
erf(x) = t * e^expo - 1   ,  falls x >= 0
erf(x) = 1 - t * e^expo   ,  falls x < 0
```

Wolfram-research gibt folgende Kettenbruchentwicklung (große x) für erf(x) an:

$$erf(x) = \frac{2}{\sqrt{\pi}} \cdot \int_0^x e^{-t^2} dt = 1 - \cfrac{\frac{1}{\sqrt{\pi}} \cdot e^{-x^2}}{x + \cfrac{1}{2x + \cfrac{2}{x + \cfrac{3}{2x + \cfrac{4}{x + \ldots}}}}}$$

Für kleine x-Werte kann man mit der obigen Taylor-Reihenentwicklung von erf(x) rechnen .