A simple algorithm for the cumulative distribution function of the Gaussian distribution

Frank Recker*

October 28, 2012

Abstract

We give a simple algorithm for the value of the cumulative distribution function of the Gaussian distribution.
Key Words: Gaussian distribution, cumulative distribution function, CDF
MSC2010 classification: 60-04, 68-04

Contents

1 Introduction

The Gaussian Integral

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} \mathrm{e}^{-\frac{t^2}{2}} \,\mathrm{d}t \tag{1}$$

is ubiquitous in statistical applications. There are libraries such as the GNU scientific library [1] which offer this calculation as a function. However, sometimes one cannot use these libraries due to technical or legal constraints. In this paper, we describe a simple algorithm for this calculation. Example implementations can be found at the github site [2] of the author.

*Frank Recker, General Reinsurance AG, Theodor-Heuss-Ring 11, 50668 Köln, Germany, e-mail: `frank.recker@arcor.de`

2 The Algorithm

The algorithm for computing $\phi(x, n)$ for $x \in \mathbb{R}$ and $n \in \mathbb{N} = \{1, 2, \dots\}$ is defined as follows:

Algorithm 1. *Let* $d_n = 0$ *and*

$$d_j = -\frac{x^2}{2j}\left(d_{j+1} + \frac{1}{2j+1}\right)$$

for all $j = n-1, \dots, 1$. *Finally let*

$$\phi(x, n) = \frac{1}{2} + \frac{x}{\sqrt{2\pi}}\left(d_1 + 1\right).$$

The algorithm computes an approximation of the Gaussian cumulative distribution function as defined in Equation (1). Theorem 2 formalizes this.

Theorem 2. *Let* $x \in \mathbb{R}$. *Define* $\Phi(x)$ *as in Equation (1) and* $\phi(x, n)$ *for all* $n \in \mathbb{N}$ *as in Algorithm 1. Then we have*

$$\lim_{n \to \infty} \phi(x, n) = \Phi(x).$$

Furthermore if $n \geq \frac{x^2}{2}$ *then*

$$|\phi(x, n) - \Phi(x)| < \frac{1}{\sqrt{2\pi}} \frac{|x|^{2n+1}}{(2n+1)2^n n!}.$$

Table 1 shows the values of $\phi(x, n)$ and the maximum error due to Theorem 2 for some x and n. The values were calculated with the code taken from [2]. Note that all calculations are done with floating point numbers and hence rounding errors will occur.

x	n	$\phi(x, n)$	error bound
1.96	1	1.2819268695868082	no bound
1.96	2	0.7812851592193613	0.28848977918213764
1.96	10	0.9749960638553972	7.014638266104427e-6
1.96	200	0.9750021048517796	1.23e-321
5	1	2.4947114020071637	no bound
5	10	-1169.2649270406318	no bound
5	30	0.9285538915764981	9.958422559186228e-2
5	50	0.9999997133453642	4.5497179496632544e-12
5	200	0.9999997133486902	1.5200212487901728e-158

Table 1: Some values and error bounds for $\phi(x, n)$

3 The Proof

First, we derive a power series for $\Phi(x)$.

Lemma 3. *Let $\Phi(x)$ as in Equation (1). Then*

$$\Phi(x) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k+1}}{(2k+1)2^k k!}. \tag{2}$$

Proof. We have $\Phi(0) = \frac{1}{2}$ and hence we can rewrite Equation (1) as

$$\Phi(x) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_0^x \mathrm{e}^{-\frac{t^2}{2}} \, \mathrm{d}t. \tag{3}$$

Next we use the power series for the exp-function:

$$\mathrm{e}^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}.$$

This series converges absolutely for all $x \in \mathbb{R}$. We insert this into Equation (3) and get

$$\Phi(x) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_0^x \sum_{k=0}^{\infty} \frac{\left(-\frac{t^2}{2}\right)^k}{k!} \, \mathrm{d}t = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_0^x \sum_{k=0}^{\infty} \frac{(-1)^k t^{2k}}{2^k k!} \, \mathrm{d}t.$$

For convergent power series we can swap integration and summation. Hence we get

$$\Phi(x) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \sum_{k=0}^{\infty} \int_0^x \frac{(-1)^k t^{2k}}{2^k k!} \, \mathrm{d}t = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \sum_{k=0}^{\infty} \left[\frac{(-1)^k t^{2k+1}}{(2k+1)2^k k!} \right]_{t=0}^{t=x}$$

and therefore

$$\Phi(x) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k+1}}{(2k+1)2^k k!}$$

as stated in the lemma. □

Next, we will show that $\phi(x, n)$ is in fact just a partial sum of the series given in Equation (2).

Lemma 4. *Let $x \in \mathbb{R}$, $n \in \mathbb{N}$ and $\phi(x, n)$ as defined in Algorithm 1. Then we have*

$$\phi(x, n) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \sum_{k=0}^{n-1} \frac{(-1)^k x^{2k+1}}{(2k+1)2^k k!}. \tag{4}$$

Proof. For brevity define

$$\phi'(x,n) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \sum_{k=0}^{n-1} \frac{(-1)^k x^{2k+1}}{(2k+1)2^k k!}.$$

For all $k \in \mathbb{N}$ let $a_k = \frac{1}{2k+1}$ and $b_k = -\frac{x^2}{2k}$. Furthermore let $a_0 = 1$ and $b_0 = \frac{x}{\sqrt{2\pi}}$. Its easy to show by complete induction that

$$\phi'(x,n) = \frac{1}{2} + \sum_{k=0}^{n-1} \left(a_k \prod_{j=0}^{k} b_j \right).$$

Writing this in a Horner-scheme style gives

$$\phi'(x,n) = \frac{1}{2} + b_0 \left(a_0 + b_1 \left(a_1 + b_2 \left(\cdots + b_{n-2} \left(a_{n-2} + b_{n-1} a_{n-1}\right) \ldots \right)\right)\right).$$

Algorithm 1 just evaluates this expression from the inner brackets to the outer one (d_j is the value of the term $b_j(a_j + b_{j+1}(\ldots))$ for all $j = 1, \ldots, n-2$). $\square$

Now we can prove Theorem 2.

Proof of Theorem 2. From Equations (2) and (4) we get

$$\lim_{n \to \infty} \phi(x,n) = \Phi(x).$$

The series $\phi(x,n)$ is alternating. Let $n \geq \frac{x^2}{2}$. We compare the terms of the series for $k = n$ and $k = n-1$:

$$\left| \frac{\frac{(-1)^n x^{2n+1}}{(2n+1)2^n n!}}{\frac{(-1)^{n-1} x^{2n-1}}{(2n-1)2^{n-1}(n-1)!}} \right| = \left| \frac{(2n-1)x^2}{(2n+1)2n} \right| \leq \left| \frac{2n-1}{2n+1} \right| < 1.$$

Hence, the series for $\phi(x,n)$ is alternating and the absolute values of the terms are monotone decreasing. Calculus shows in this case that $|\phi(x,n) - \Phi(x)|$ is bound by the absolute value of the n-th term of the series which is

$$\frac{1}{\sqrt{2\pi}} \frac{|x|^{2n+1}}{(2n+1)2^n n!}.$$

$\square$

References

[1] The Gaussian Distribution in the GNU Scientific Library, `http://www.gnu.org/software/gsl/manual/html_node/The-Gaussian-Distribution.html`

[2] Implementations of Algorithm 1 on github, `https://github.com/frecker/gaussian-distribution`