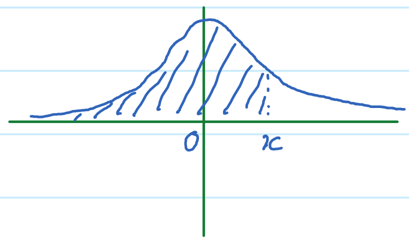
Lecture 4

Monday, January 13, 2020

An well-known function in Probability and Statistics is the cummulative distribution function of the normal distribution.

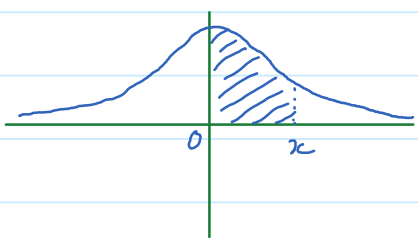
$$\phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$



which is the area under the bell curve from $-\infty$ to x. Due to symmetry, to area only left under the curve from $-\infty$ to 0 is equal to $1/2$.

Therefore, to compute $\phi(x)$ we only need to compute

$$F(x) = \int_{0}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dx$$



The function $e^{-t^2/2}$ has an antiderivative that is not an elementary function. Thus, one can't compute $F(x)$ using the fundamental theorem of Calculus. For this reason, people record a table of values of ϕ to look up when needed. The following picture is an example.

The (approximate) value of $F(2.92)$ is found on the row starting with 2.9 and the column 0.02, which is 0.49825 This table record values of $F(x)$ for x in between 0 and 4, with precision 10^{-5} We will see that with the help of a minimal calculator, we can create a table like this (even with more data and better precision) ourselves.

X	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.00000	0.00399	0.00798	0.01197	0.01595	0.01994	0.02392	0.02790	0.03188	0.03586
0.1	0.03983	0.04380	0.04776	0.05172	0.05567	0.05962	0.06356	0.06749	0.07142	0.07535
0.2	0.07926	0.08317	0.08706	0.09095	0.09483	0.09871	0.10257	0.10642	0.11026	0.11409
0.3	0.11791	0.12172	0.12552	0.12930	0.13307	0.13683	0.14058	0.14431	0.14803	0.15173
0.4	0.15542	0.15910	0.16276	0.16640	0.17003	0.17364	0.17724	0.18082	0.18439	0.18793
0.5	0.19146	0.19497	0.19847	0.20194	0.20540	0.20884	0.21226	0.21566	0.21904	0.22240
0.6	0.22575	0.22907	0.23237	0.23565	0.23891	0.24215	0.24537	0.24857	0.25175	0.25490
0.7	0.25804	0.26115	0.26424	0.26730	0.27035	0.27337	0.27637	0.27935	0.28230	0.28524
0.8	0.28814	0.29103	0.29389	0.29673	0.29955	0.30234	0.30511	0.30785	0.31057	0.31327
0.9	0.31594	0.31859	0.32121	0.32381	0.32639	0.32894	0.33147	0.33398	0.33646	0.33891
1.0	0.34134	0.34375	0.34614	0.34849	0.35083	0.35314	0.35543	0.35769	0.35993	0.36214
1.1	0.36433	0.36650	0.36864	0.37076	0.37286	0.37493	0.37698	0.37900	0.38100	0.38298
1.2	0.38493	0.38686	0.38877	0.39065	0.39251	0.39435	0.39617	0.39796	0.39973	0.40147
1.3	0.40320	0.40490	0.40658	0.40824	0.40988	0.41149	0.41308	0.41466	0.41621	0.41774
1.4	0.41924	0.42073	0.42220	0.42364	0.42507	0.42647	0.42785	0.42922	0.43056	0.43189
1.5	0.43319	0.43448	0.43574	0.43699	0.43822	0.43943	0.44062	0.44179	0.44295	0.44408
1.6	0.44520	0.44630	0.44738	0.44845	0.44950	0.45053	0.45154	0.45254	0.45352	0.45449
1.7	0.45543	0.45637	0.45728	0.45818	0.45907	0.45994	0.46080	0.46164	0.46246	0.46327
1.8	0.46407	0.46485	0.46562	0.46638	0.46712	0.46784	0.46856	0.46926	0.46995	0.47062
1.9	0.47128	0.47193	0.47257	0.47320	0.47381	0.47441	0.47500	0.47558	0.47615	0.47670
2.0	0.47725	0.47778	0.47831	0.47882	0.47932	0.47982	0.48030	0.48077	0.48124	0.48169
2.1	0.48214	0.48257	0.48300	0.48341	0.48382	0.48422	0.48461	0.48500	0.48537	0.48574
2.2	0.48610	0.48645	0.48679	0.48713	0.48745	0.48778	0.48809	0.48840	0.48870	0.48899
2.3	0.48928	0.48956	0.48983	0.49010	0.49036	0.49061	0.49086	0.49111	0.49134	0.49158
2.4	0.49180	0.49202	0.49224	0.49245	0.49266	0.49286	0.49305	0.49324	0.49343	0.49361
2.5	0.49379	0.49396	0.49413	0.49430	0.49446	0.49461	0.49477	0.49492	0.49506	0.49520
2.6	0.49534	0.49547	0.49560	0.49573	0.49585	0.49598	0.49609	0.49621	0.49632	0.49643
2.7	0.49653	0.49664	0.49674	0.49683	0.49693	0.49702	0.49711	0.49720	0.49728	0.49736
2.8	0.49744	0.49752	0.49760	0.49767	0.49774	0.49781	0.49788	0.49795	0.49801	0.49807
2.9	0.49813	0.49819	0.49825	0.49831	0.49836	0.49841	0.49846	0.49851	0.49856	0.49861
3.0	0.49865	0.49869	0.49874	0.49878	0.49882	0.49886	0.49889	0.49893	0.49896	0.49900
3.1	0.49903	0.49906	0.49910	0.49913	0.49916	0.49918	0.49921	0.49924	0.49926	0.49929
3.2	0.49931	0.49934	0.49936	0.49938	0.49940	0.49942	0.49944	0.49946	0.49948	0.49950
3.3	0.49952	0.49953	0.49955	0.49957	0.49958	0.49960	0.49961	0.49962	0.49964	0.49965
3.4	0.49966	0.49968	0.49969	0.49970	0.49971	0.49972	0.49973	0.49974	0.49975	0.49976
3.5	0.49977	0.49978	0.49978	0.49979	0.49980	0.49981	0.49981	0.49982	0.49983	0.49983
3.6	0.49984	0.49985	0.49985	0.49986	0.49986	0.49987	0.49987	0.49988	0.49988	0.49989
3.7	0.49989	0.49990	0.49990	0.49990	0.49991	0.49991	0.49992	0.49992	0.49992	0.49992
3.8	0.49993	0.49993	0.49993	0.49994	0.49994	0.49994	0.49994	0.49995	0.49995	0.49995
3.9	0.49995	0.49995	0.49996	0.49996	0.49996	0.49996	0.49996	0.49996	0.49997	0.49997
4.0	0.49997	0.49997	0.49997	0.49997	0.49997	0.49997	0.49998	0.49998	0.49998	0.49998

Source: https://www.itl.nist.gov/div898/handbook/eda/section3/eda3671.htm

First, let us simplify the function

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-t^2/2}\, dt \overset{s = t/\sqrt{2}}{=\!=\!=} \frac{1}{\sqrt{\pi}} \int_0^{x/\sqrt{2}} e^{-s^2}\, ds$$

The problem becomes: how to compute the values of the function

$$\psi(x) = \int_0^x e^{-t^2}\, dt$$

when x is given.

Let's make the problem more specific: given an error tolerance $\varepsilon = 10^{-4}$, compute approximately the values of $\varphi(x)$ when $0 \le x \le 2$ with error not exceeding ε.

Put $f(t) = e^{-t^2}$. To approximate the integral of f, one needs to approximate f by polynomials. Note that one can integrate a polynomial easily. Let's try to approximate f by its Taylor polynomials about $t_0 = 0$. We need to compute the higher derivatives of f:

$$f'(t) = -2t\, e^{-t^2}$$
$$f''(t) = (4t^2 - 2)\, e^{-t^2}$$
$$f'''(t) = 8t\, e^{-t^2} - 2t(4t^2-2)\, e^{-t^2} = (-8t^3 + 12t)\, e^{-t^2}$$
$$\cdots$$

It is easy to see that the derivatives of f get complicated very quickly. This makes it hard to find the Taylor polynomial of f in the usual way. We know how to compute the Taylor polynomials of e^t:

$$e^t = \underbrace{1 + t + \frac{t^2}{2!} + \frac{t^3}{3!} + \cdots + \frac{t^n}{n!}}_{p_n(t)} + \underbrace{\frac{t^{n+1}}{(n+1)!} + \cdots}_{R_n(t)}$$

Now replace t by $-t^2$:

$$e^{-t^2} = 1 + (-t^2) + \frac{(-t^2)^2}{2!} + \frac{(-t^2)^3}{3!} + \cdots + \frac{(-t^2)^n}{n!} + \cdots$$
$$= \underbrace{1 - t^2 + \frac{t^4}{2!} - \frac{t^6}{3!} + \cdots + \frac{(-1)^n t^{2n}}{n!}}_{q_{2n}(t)} + \underbrace{\cdots}_{r_{2n}(t)}$$

We have found the $(2n)$'th Taylor polynomial of e^{-t^2}, which is

$$\underbrace{q_{2n}(t)}_{\text{Taylor poly of } e^{-t^2}} = \underbrace{p_n}_{\text{Taylor poly. of } e^t}(-t^2)$$

The error term is $\underbrace{r_{2n}(t)}_{\substack{\text{error term} \\ \text{of } e^{-t^2}}} = \underbrace{R_n(-t^2)}_{\substack{\text{error term} \\ \text{of } e^{t}}}$

Thus,

$$\psi(x) = \int_0^x f(t)\,dt = \underbrace{\int_0^x q_n(t)\,dt}_{\text{computable}} + \underbrace{\int_0^x r_n(t)\,dt}_{\text{error term}}$$

The first integral on RHS is the approximate value of $\psi(x)$. It is the integral of a polynomial:

$$\psi(x) \approx \int_0^x q_n(t)\,dt = \left(\frac{t}{1} - \frac{t^3}{(3)1!} + \frac{t^5}{(5)3!} - \frac{t^7}{(7)4!} + \cdots + \frac{(-1)^n t^{2n+1}}{(2n+1)n!}\right)\Bigg|_0^x$$

$$= \frac{x}{1} - \frac{x^3}{(3)1!} + \frac{x^5}{(5)3!} - \frac{x^7}{(7)4!} + \cdots + \frac{(-1)^n x^{2n+1}}{(2n+1)n!}$$

This can be computed by hand (or pocket calculator) since it only involves multiplication, division, addition and subtraction.

What value of n should we pick? It depends on how small we want the error term to be. We need to make sure that the error term

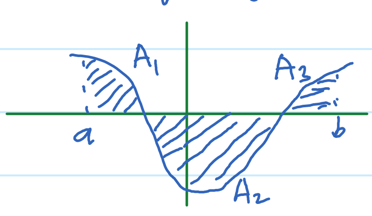
$$\left|\int_0^x r_{2n}(t)\,dt\right|$$

is always below $\varepsilon = 10^{-4}$ when $x \in [0,2]$.

When we are trying to estimate an integral, the following inequality is helpful:

$$\left|\int_a^b f(x)\,dx\right| \le \int_a^b |f(x)|\,dx \qquad (a<b)$$

This inequality can be explained by picture:



The left hand side is equal to $|A_1 - A_2 + A_3|$

The right hand side is $A_1 + A_2 + A_3$.

The inequality $|A_1 - A_2 + A_3| \le A_1 + A_2 + A_3$

is true (by triangle inequality).

Now apply the above property for $f = r_{2n}$:

$$\left| \int_0^x r_{2n}(t)\,dt \right| \le \int_0^x |r_{2n}(t)|\,dt$$

We want to make sure that

$$\int_0^x |r_{2n}(t)|\,dt \le \varepsilon = 10^{-4} \qquad \forall x \in [0,2].$$

It is not very helpful to use Lagrange's theorem for r_{2n} because it would require us to find $f^{(2n+1)}$. It is easier to use Lagrange's theorem for R_n and then use the relation

$$r_{2n}(t) = R_n(-t^2).$$

By Lagrange's theorem for function e^t, which we did last time, one can write $R_n(t)$ as

$$R_n(t) = \frac{e^c}{(n+1)!} t^{n+1}$$

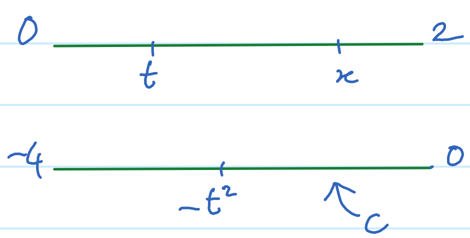
for some c in between 0 and t.

Now replace t by $-t^2$:

$$r_{2n}(t) = R_n(-t^2) = \frac{e^c}{(n+1)!}(-t^2)^{n+1}$$

for some c in between 0 and $-t^2$.

We are considering $t \in [0,2]$ only. Thus, $-t^2$ is in between 0 and -4.



Thus c is in between -4 and 0.

Take the absolute value of r_{2n}:

$$|r_{2n}(t)| = \frac{e^c}{(n+1)!} \left|(-t^2)^{n+1}\right| = \frac{e^c}{(n+1)!} t^{2(n+1)}$$

$$\leq \frac{e^0}{(n+1)!} 2^{2(n+1)}$$

$$= \frac{4^{n+1}}{(n+1)!}$$

Therefore, $$\int_0^x |r_{2n}(t)|\, dt \leq \int_0^x \frac{4^{n+1}}{(n+1)!}\, dt$$

$$= \frac{4^{n+1}}{(n+1)!} x$$

$$\leq 2 \frac{4^{n+1}}{(n+1)!}$$

To make sure that $\int_0^x |r_{2n}(t)|\, dt < \varepsilon$

we only need to choose a big n such that

$$2 \frac{4^{n+1}}{(n+1)!} < \varepsilon = 10^{-4}$$

By calculator, we see that $n = 16$ will do it.

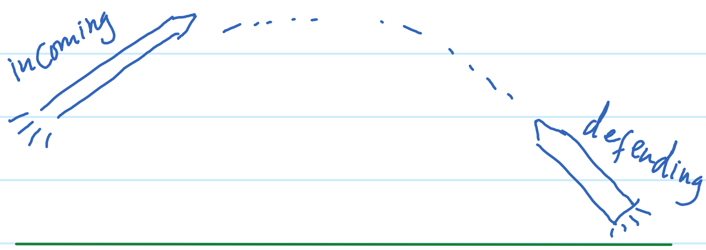
Error from digitalizing numbers on computer

The error we have seen so far comes from Taylor approximation. This is a mathematical error. approximating an arbitrary function by a polynomial naturally causes an error. Mathematical error in this case can be controlled thanks to Lagrange's theorem (via choosing sufficiently big n).

There is another type of error, relatively small, which we have less control over. It is the error caused by the way computers store numbers in its memory.

With finite memory, computers can store/represent numbers by a finite amount of bits. Thus, chopping or rounding is necessary, causing error. This type of error is not caused by math, but by the limitation of computers Although it is hard to control, understanding this type of error is important so that we know how to minimize it.

A classic example of a disaster caused by rounding error is the Patriot missle failure during the gulf war 1991



The US army detected an incoming missle of the enemy and planned to shoot a missle to hit it in the air. The internal clock of the US missle used 24 bits to represent time. The unit time of the system is 1/10 second. In the binary system, 1/10 is a number with infinite number of digits after the dot

Due to truncation, there is a slight different between the true 1/10 s and the 1/10 s stored in the system. We can see that over a long period of time, the error between the true time and the system's time will be magnified significantly, unless the system is restarted over some amount of time.

In reality, the system was running for about 100 hours. As a result, the difference between the real time and the system's time was about 0.34(s). The missle was launched 0.34 seconds earlier or later than the planned time, causing it to miss the enemy's missle, resulting in heavy casualty (28 deaths). One can read the following article for more detail:

http://www-users.math.umn.edu/~arnold//disasters/patriot.html