Improvement of the Matlab program proposed in viXra:2103.0018

Joh. J. Sauren & Aloys J. Sipers

Department of Engineering
Zuyd University of Applied Sciences
NL-6419 DJ Heerlen, The Netherlands
hans.sauren@zuyd.nl aloys.sipers@zuyd.nl

Abstract
In this article, the Matlab program proposed in the article viXra:2103.0018 is improved. Further, the constant d3 depends on the constants d2 and a3. Three theorems are stated on the generating functions for the constants d2 and a3. The first two theorems provide analytical expressions for these generating functions, whereas the third theorem relates them.

At the time of writing the communication mentioned in the title, it was unknown to the first author that mathematical analytical expressions for both of $d_2 -$ and $d_3 -$constants indeed do exist. These analytical expressions were first derived by the English statistician L. H. C. Tippett[1]. A further account of these expressions for both of $d_2 -$ and $d_3 -$constants can be found in the SAS/QC 15.1 User's Guide[2]. The constant d_3 depends on the constants d_2 and a_3. In this communication three theorems are stated on the generating functions for the constants d_2 and a_3. The first two theorems provide analytical expressions for these generating functions, whereas the third theorem relates them. The $d_2 -$ and $d_3 -$constants depend on the variable n and are expressed in this User Guide as

$$d_2(n) := \int_{-\infty}^{+\infty} \left[1 - \left(1 - \Phi(x)\right)^n - \left(\Phi(x)\right)^n\right] dx$$

In the above expression $\Phi(x)$ denotes the standard normal cumulative distribution function with average μ = 0, and standard deviation σ = 1 i.e.

$$\Phi(x) := \int_{-\infty}^{x} \frac{e^{-\frac{\xi^2}{2}}}{\sqrt{2\pi}} d\xi$$

Likewise, the constant $d_3(n)$ is analytically expressed as[3]

$$d_3(n) := \sqrt{2 \cdot a_3(n) - (d_2(n))^2}$$

in which

$$a_3(n) := \int_{-\infty}^{+\infty} \int_{-\infty}^{y} \left[1 - \Phi(y)^n - \left(1 - \Phi(x)\right)^n + \left(\Phi(y) - \Phi(x)\right)^n\right] dx\, dy$$

In the following, a short program in Matlab is proposed that calculates both constants with substantial gain of speed and improvement of accuracy compared to the Matlab program published in “On the computation of the principal constants d_2 and d_3 used to construct control limits for control charts applied in statistical process control”, see **viXra:2103.0018**.

Matlab programs

The Matlab (R2021 a) programs proposed here were run using an Intel® Core™ i7 9850H CPU @ 2.60 GHz processor and 32 GByte RAM.

Program 1

In this program, the $d_2 -$ and $d_3 -$constants are calculated according to the definitions given above for $2 \leq n \leq k$ using symbolic definition for both x and y. To calculate the integrals, the function vpaintegral[4] is used and the tic()-toc()-function measures the execution time of the code.

```
clear
tic();
k = 9;
d2 = zeros(1, k);
d3 = zeros(1, k);
i = (1 : k);
for n = 2 : k
syms x y
d2(n) = vpaintegral(1-(1-normcdf(x)).^n - (normcdf(x)).^n, [-Inf, Inf], 'AbsTol', 1e-8);
f(x,y) = 1 - (normcdf(y)).^n - (1 - normcdf(x)).^n + (normcdf(y) - normcdf(x)).^n;
Fxy = vpaintegral(vpaintegral(f, x, [-inf, y], 'AbsTol', 1e-8), y, [-inf, inf], 'AbsTol', 1e-8);
d3(n) = sqrt(2*Fxy-d2(n).^2);
end
T = [i' d2' d3'];
T(1,:) = [];
fprintf('n |   d2    |    d3  ')
fprintf('%i | %1.5f | %1.5f \n', T')
toc()
```

The results of this code are:

```
n |    d2     |     d3
2 | 1.12838 | 0.85250
3 | 1.69257 | 0.88837
4 | 2.05875 | 0.87981
5 | 2.32593 | 0.86408
6 | 2.53441 | 0.84804
7 | 2.70436 | 0.83321
8 | 2.84720 | 0.81983
9 | 2.97003 | 0.80783
Elapsed time is 26.030876 seconds.
```

However, the execution time of computation of both constants can appreciably be improved by approximately a factor of 3 (around 9 s compared to around 26 s for $k = 9$) using the Matlab code below. This code employs the symmetry properties of the integral definitions of both $d_2 -$ and $d_3 -$constants derived below: in contrast to the integral definitions above, now integration interval values of $[0, +\infty)$ and $[-y, +y]$ are used rather than $(-\infty, +\infty)$ and $(-\infty, +y]$ as above in program 1.

Program 2

```
clear
tic();
k = 9;
d2 = zeros(1, k);
d3 = zeros(1, k);
i = (1 : k);
for n = 2 : k
syms x y
d2(n) = 2.*vpaintegral(1-(normcdf(-x)).^n - (normcdf(x)).^n, [0, Inf], 'AbsTol', 1e-8);
f(x,y) = 1 - (normcdf(y)).^n - (normcdf(-x)).^n + (normcdf(y) - normcdf(x)).^n;
Fxy = 2.*vpaintegral(vpaintegral(f, x, [-y, y], 'AbsTol', 1e-8), y, [0, inf], 'AbsTol', 1e-8);
d3(n) = sqrt(2*Fxy-d2(n).^2);
end
T = [i' d2' d3'];
T(1,:) = [];
fprintf('n |   d2    |    d3     ')
fprintf('%i | %1.5f | %1.5f  \n', T')
toc()
```

Using the Matlab code given above, the following figures result:

```
n |   d2    |   d3
2 | 1.12838 | 0.85250
3 | 1.69257 | 0.88837
4 | 2.05875 | 0.87981
5 | 2.32593 | 0.86408
6 | 2.53441 | 0.84804
7 | 2.70436 | 0.83321
8 | 2.84720 | 0.81983
9 | 2.97003 | 0.80783
Elapsed time is 9.061348 seconds.
```

Compared to literature values, the results of the calculation of both constants using either Matlab code proposed above fully agree[5]:

n	d2	d3
2	1.128	0.8525
3	1.693	0.8884
4	2.059	0.8798
5	2.326	0.8641
6	2.534	0.8480
7	2.704	0.8332
8	2.847	0.8198
9	2.970	0.8078

Lemma
Let

$$f_n(x,y) \coloneqq 1 - \Phi(y)^n - \Phi(-x)^n + \left(\Phi(y) - \Phi(x)\right)^n$$

and its generating function

$$F(x,y,t) \coloneqq \sum_{n=2}^{\infty} f_n(x,y) \cdot t^n$$

Then

$$F(x,y,t) = \frac{t^2 \cdot \Phi(x) \cdot \Phi(-y)}{1-t} \cdot \frac{2 - t \cdot \left(\Phi(y) + \Phi(-x)\right)}{\left(1 - t \cdot \Phi(-x)\right) \cdot \left(1 - t \cdot \Phi(y)\right) \cdot \left(1 - t \cdot \left(\Phi(y) - \Phi(x)\right)\right)}$$

Proof

$$\sum_{n=2}^{\infty} f_n(x,y) \cdot t^n = \sum_{n=2}^{\infty} \left[1 - \Phi(y)^n - \Phi(-x)^n + \left(\Phi(y) - \Phi(x)\right)^n\right] \cdot t^n$$

$$= \sum_{n=2}^{\infty} \left[t^n - \Phi(y)^n \cdot t^n - \Phi(-x)^n \cdot t^n + \left(\Phi(y) - \Phi(x)\right)^n \cdot t^n\right]$$

$$= \frac{t^2}{1-t} - \frac{t^2 \cdot \Phi(y)^2}{1 - t \cdot \Phi(y)} - \frac{t^2 \cdot \Phi(-x)^2}{1 - t \cdot \Phi(-x)} + \frac{t^2 \cdot \left(\Phi(y) - \Phi(x)\right)^2}{1 - t \cdot \left(\Phi(y) - \Phi(x)\right)}$$

$$= \frac{t^2 \cdot \Phi(x) \cdot \Phi(-y)}{1-t} \cdot \frac{2 - t \cdot \left(\Phi(y) + \Phi(-x)\right)}{\left(1 - t \cdot \Phi(-x)\right) \cdot \left(1 - t \cdot \Phi(y)\right) \cdot \left(1 - t \cdot \left(\Phi(y) - \Phi(x)\right)\right)}$$

Theorem 1
Let

$$a_3(n) \coloneqq 2 \cdot \int_0^{\infty} dy \cdot \int_{-y}^{y} f_n(x,y) \cdot dx$$

and its generating function

$$\alpha_3(t) \coloneqq \sum_{n=2}^{\infty} a_3(n) \cdot t^n$$

Then

$$\alpha_3(t) = \frac{2 \cdot t^2}{1-t} \cdot \int_0^{\infty} dy \cdot \int_{-y}^{y} \frac{2 - t \cdot \left(\Phi(y) + \Phi(-x)\right)}{\left(1 - t \cdot \Phi(-x)\right) \cdot \left(1 - t \cdot \Phi(y)\right) \cdot \left(1 - t \cdot \left(\Phi(y) - \Phi(x)\right)\right)} \cdot \Phi(x) \cdot \Phi(-y) \cdot dx$$

Proof

$$\alpha_3(t) = \sum_{n=2}^{\infty} a_3(n) \cdot t^n$$

$$= \sum_{n=2}^{\infty} 2 \cdot \int_0^{\infty} dy \cdot \int_{-y}^{y} f_n(x,y) \cdot dx \cdot t^n$$

$$= 2 \cdot \int_0^{\infty} dy \cdot \int_{-y}^{y} \left[\sum_{n=2}^{\infty} f_n(x,y) \cdot t^n\right] \cdot dx$$

$$= 2 \cdot \int_0^{\infty} dy \cdot \int_{-y}^{y} F(x,y,t) \cdot dx$$

Using the previous lemma:

$$\alpha_3(t) = 2 \cdot \int_0^{\infty} dy \cdot \int_{-y}^{y} \left[\frac{t^2 \cdot \Phi(x) \cdot \Phi(-y)}{1-t} \cdot \frac{2 - t \cdot \big(\Phi(y) + \Phi(-x)\big)}{\big(1 - t \cdot \Phi(-x)\big) \cdot \big(1 - t \cdot \Phi(y)\big) \cdot \Big(1 - t \cdot \big(\Phi(y) - \Phi(x)\big)\Big)}\right] \cdot dx$$

$$= \frac{2 \cdot t^2}{1-t} \cdot \int_0^{\infty} dy \cdot \int_{-y}^{y} \frac{2 - t \cdot \big(\Phi(y) + \Phi(-x)\big)}{\big(1 - t \cdot \Phi(-x)\big) \cdot \big(1 - t \cdot \Phi(y)\big) \cdot \Big(1 - t \cdot \big(\Phi(y) - \Phi(x)\big)\Big)} \cdot \Phi(x) \cdot \Phi(-y) \cdot dx$$

□

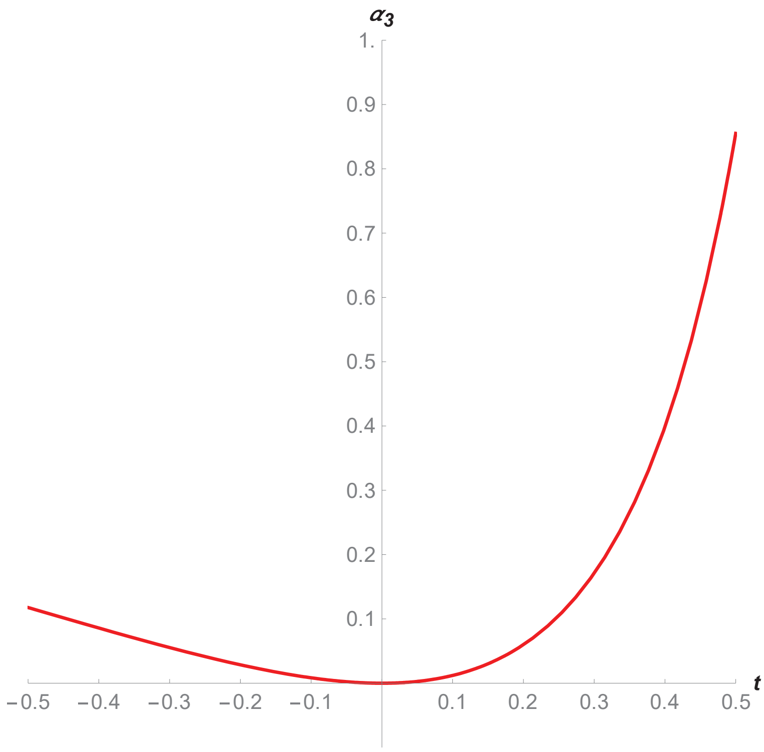

Figure 1: Generating function $\alpha_3(t)$ for the constants $a_3(n)$.

Theorem 2

Let

$$d_2(n) := 2 \cdot \int_0^\infty f_n(x,x) \cdot dx$$

and its generating function

$$\delta_2(t) := \sum_{n=2}^{\infty} d_2(n) \cdot t^n$$

Then

$$\delta_2(t) = 2 \cdot t^2 \cdot \frac{2-t}{1-t} \cdot \int_0^\infty \frac{\Phi(x) \cdot \Phi(-x)}{\big(1 - t \cdot \Phi(x)\big) \cdot \big(1 - t \cdot \Phi(-x)\big)} \cdot dx$$

Proof

$$\delta_2(t) = \sum_{n=2}^{\infty} d_2(n) \cdot t^n$$

$$= \sum_{n=2}^{\infty} 2 \cdot \int_0^\infty f_n(x,x) \cdot dx \cdot t^n$$

$$= 2 \cdot \int_0^\infty \left[\sum_{n=2}^{\infty} f_n(x,x) \cdot t^n \right] \cdot dx$$

$$= 2 \cdot \int_0^\infty F(x,x,t) \cdot dx$$

Using the previous lemma for $y = x$:

$$\delta_2(t) = 2 \cdot \int_0^\infty \left[\frac{t^2 \cdot \Phi(x) \cdot \Phi(-x)}{1-t} \cdot \frac{2 - t \cdot \big(\Phi(x) + \Phi(-x)\big)}{\big(1 - t \cdot \Phi(-x)\big) \cdot \big(1 - t \cdot \Phi(x)\big) \cdot \Big(1 - t \cdot \big(\Phi(x) - \Phi(x)\big)\Big)} \right] \cdot dx$$

$$= 2 \cdot t^2 \cdot \frac{2-t}{1-t} \cdot \int_0^\infty \frac{\Phi(x) \cdot \Phi(-x)}{\big(1 - t \cdot \Phi(x)\big) \cdot \big(1 - t \cdot \Phi(-x)\big)} \cdot dx$$

□

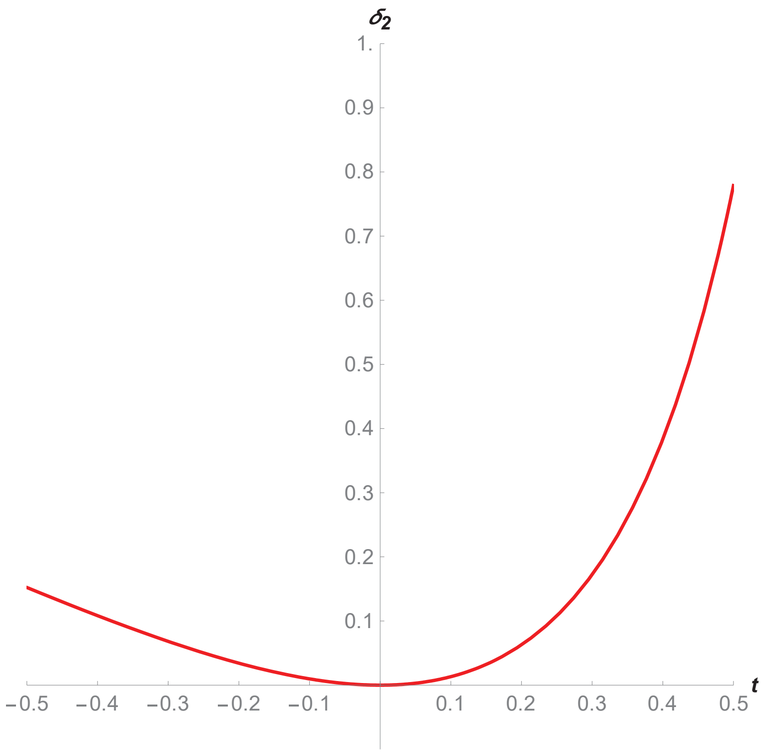

Figure 2: Generating function $\delta_2(t)$ for the constants $d_2(n)$.

Theorem 3

Let

$$A_3(v,t) := 4\cdot\int_0^v d\tilde{v}\cdot\int_0^\infty F(u-\tilde{v},u+\tilde{v},t)\cdot du$$

Then

(1) $\quad \alpha_3(t) = \lim_{v\to\infty} A_3(v,t)$

and

(2) $\quad \delta_2(t) = \frac{1}{2}\cdot\frac{\partial}{\partial v}A_3(v,t)\Big|_{v=0}$

Proof

Using the transformation $x = u - \tilde{v}, y = u + \tilde{v}$:

(1) $$2\cdot\int_0^\infty dy\cdot\int_{-y}^{y} F(x,y,t)\cdot dx = 4\cdot\int_0^\infty d\tilde{v}\cdot\int_0^\infty F(u-\tilde{v},u+\tilde{v},t)\cdot du$$

$$= \lim_{v\to\infty} 4\cdot\int_0^v d\tilde{v}\cdot\int_0^\infty F(u-\tilde{v},u+\tilde{v},t)\cdot du = \lim_{v\to\infty} A_3(v,t) = \alpha_3(t)$$

(2) $$\frac{1}{2}\cdot\frac{\partial}{\partial v}A_3(v,t) = \frac{1}{2}\cdot\frac{\partial}{\partial v}\left[4\cdot\int_0^v d\tilde{v}\cdot\int_0^\infty F(u-\tilde{v},u+\tilde{v},t)\cdot du\right] = 2\cdot\int_0^\infty F(u-v,u+v,t)\cdot du$$

Substituting $v = 0$:

$$\frac{1}{2}\cdot\frac{\partial}{\partial v}A_3(v,t)\Big|_{v=0} = 2\cdot\int_0^\infty F(u,u,t)\cdot du = \delta_2(t)$$

□

Acknowledgement

The authors acknowledge the support of Ad Klein and Charlotte Creusen of the Department of Engineering of Zuyd University of Applied Sciences.

Literature

1) Tippett, L. H. C. (1925). "On the Extreme Individuals and the Range of Samples Taken from a Normal Population." *Biometrika* 17:364–387.
2) http://documentation.sas.com/doc/en/pgmsascdc/9.4_3.4/qcug/qcug_functions_sect066.htm#qcug_functionstipp_l25
3) http://documentation.sas.com/doc/en/pgmsascdc/9.4_3.4/qcug/qcug_functions_sect031.htm
4) https://de.mathworks.com/help/symbolic/vpaintegral.html
5) https://support.minitab.com/en-us/minitab-express/1/control-charts/how-to/variables-data-in-subgroups/xbar-r-chart/methods-and-formulas/unbiasing-constants-d2-d3-and-d4/