

Mathematical approach of PSA level test during prostate cancer analysis

Peter Fesus

University of Debrecen, Doctoral School of Information Science and Technology
`peter.fesus@pro-filter.hu`

Abstract

The most common test for the detection of prostate cancer is based on the PSA level of blood. This inexpensive monitoring method can be repeated quite often compared to the empirical propagation velocity of the disease. The disadvantage of this test is that besides tumour other factors may also increase the PSA level (e.g. inflammation), therefore the number of false positive results is high [1].

In this article we keep the series of one's measured PSA levels as a stochastic process and determine the coefficients of a quadratic equation, which suitable to estimate the expected value of PSA level at a given time. We give formulas for the speed and acceleration of the process, as well as counting the time when a measured value will doubles, or the time when the PSA level will exceed a critical value. The article shows both the closed version and the recursive version of this second-level polynomial [2].

In the end we create the fundamental algorithms of a hypothetical IT service (called PRO-FILTER) modelling the mathematical results¹.

1. Introduction

The prostate cancer is one of the most common male tumorous disease. The disease is generally indolent, plus the detection and treatment at an early stage results high proportion of cured patients or improved quality of life.

Those materials, which pointing to the existence of cancer called *tumour markers*. The tumour markers are verified by laboratory tests, or are tried to isolate the high-risk cases with screening – depending on the type of cancer.

For screening those laboratory method are suitable, which are cost-effective, easy to implement, and sufficiently sensitive and specific – that helps to distinguish the clinically significant cases from the insignificants.

The prostate cancers are often used to indicate with the following tests.

¹Special thanks to Mr. Andor Dobó, and Dr. Károly Nagy MD for their invaluable help.

1. The **PSA level test** measures the presence of prostate-specific antigen (this antigen is a tumour marker) from blood sample. This procedure carried out cheaply (10-12 EUR in 2014), therefore currently this is the most widely used screening. The investigation has the disadvantage that beyond cancer other reasons also raise the PSA level, so the process could over diagnose – about 20% of the positive results are false-positive.
2. During the digital rectal examination (**DRE test**) the urologist checks the prostate through the rectum and looking for differences.
3. The **PCA3 level test** is a genetic test that measures the level of the PCA3 gene. This gene (which is also a tumour marker) is only detectable in the prostate tissue and therefore the sensitivity of the process exceeds the monitoring of PSA level test. The false-positive alerts only about 10% of the total positive tests. However, the procedure is also costly (235-250 EUR in 2014).

In case of symptoms of prostate cancer the general protocol starts with a DRE test and simultaneously a PSA level test. According to the PSA level – and beyond the emergency surgery – the next step could be a prostate biopsy, an active follow of the patient's PSA level with annual tests, or the end of the protocol in a negative case. This protocol is generally over secured. This causes a lot of unnecessary prostate biopsy. The average cost of such a treatment is around 250-280 EUR in 2014. This raise to a very high amount in a population yearly if we think about the 20% false positive alert of PSA level test. Not to mention the additional risks of such an invasive treatment, especially in case of an elder man.

2. The task

Through a man's life in general the PSA level of blood varies according to Fig. 1.

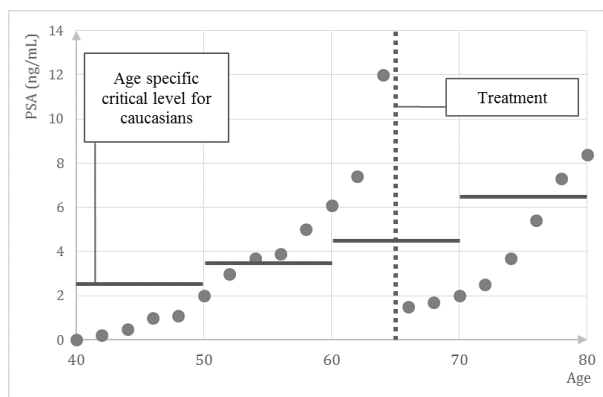


Figure 1

In this paper we consider only the beginning of this process until the first treatment, because our vision is a pre-screening aid service for doctors. Further investigation can be found in [3].

The primary goals are, that after a series of PSA tests to calculate sufficient estimations for the following questions:

1. What is the estimated value of the PSA level in an arbitrary time t ?
2. At what t_c time reaches the estimated PSA level a predetermined critical value c ?
3. What is t_d time when the estimated PSA level reaches the double of the last measured value? (In oncology the prostate cancer process is considered booted, when the PSA level doubling time is less than one year.)
4. How will change the current velocity of change of PSA level?

Further goals are to determine the accuracy of the estimations and reduce the number of false positive results of the PSA level test.

3. Stochastic approach with closed formula

Let $t_1 < t_2 < t_n$ the date of the PSA level tests and p_1, p_2, p_n the measured PSA level for each date ($n \in \mathbb{N}, n \geq 3$).

Consider the PSA level as stochastic, random variable which depends from the time continuously.

Let $F(t)$ is the expected value of the PSA level at a given t time. Consider the Taylor-series of $F(t)$ up to the third (quadratic) member.

$$F(t) \approx F(0) + \frac{F'(0)}{1!}t + \frac{F''(0)}{2!}t^2$$

and name $b = F(0), v = F'(t), a = F''(t)$. With these rename look the

$$F(t) \approx p(t) = \frac{a}{2}t^2 + vt + b$$

quadratic polynomial, which approximates $F(t)$ well near t .

An estimation can be created for the a, v and b using the *least square method* on (t_i, p_i) pairs ($1 \leq i \leq n$). The least square method on a quadratic formula works well if there are at least three points, as we originally assumed that $n \geq 3$. In our case this means that we should minimise the

$$P(a, v, b) = \sum_{i=1}^n (p(t_i) - p_i)^2 = \sum_{i=1}^n \left(\frac{a}{2}t_i^2 + vt_i + b - p_i \right)^2$$

formula. This one could solve with the knowledge of that the minimum value is where the a, v and b directional partial derivatives of the above formulas are 0.

$$\frac{\partial P(a, v, b)}{\partial a} = 0 \quad \frac{\partial P(a, v, b)}{\partial v} = 0 \quad \frac{\partial P(a, v, b)}{\partial b} = 0$$

The equations with the partial derivatives are the following:

$$\frac{\partial P(a, v, b)}{\partial a} = b \sum_{i=1}^n t_i^2 + v \sum_{i=1}^n t_i^3 + \frac{a}{2} \sum_{i=1}^n t_i^4 - \sum_{i=1}^n p_i t_i^2 = 0$$

and

$$\frac{\partial P(a, v, b)}{\partial v} = b \sum_{i=1}^n t_i + v \sum_{i=1}^n t_i^2 - \frac{a}{2} \sum_{i=1}^n t_i^3 - \sum_{i=1}^n p_i t_i = 0$$

as well as

$$\frac{\partial P(a, v, b)}{\partial b} = b + v \sum_{i=1}^n \frac{t_i}{n} - \frac{a}{2} \sum_{i=1}^n \frac{t_i^2}{n+1} - \sum_{i=1}^n \frac{p_i}{n} = 0$$

To solve this equation system with three equations and three unknown variables introduce the following auxiliary variables

$$\begin{aligned} A_0 &= \sum_{i=1}^n \frac{p_i}{n} & A_1 &= \sum_{i=1}^n p_i t_i & A_2 &= \sum_{i=1}^n p_i t_i^2 \\ B_1 &= \sum_{i=1}^n t_i & B_2 &= \sum_{i=1}^n t_i^2 & C_2 &= \sum_{i=1}^n t_i^3 \\ & & & & D_2 &= \sum_{i=1}^n t_i^4 \end{aligned}$$

Using these auxiliary variables we can calculate the value of a on this way:

$$a = 2 \frac{\frac{B_1}{n} B_2 - C_2 (A_1 - A_0 B_1) - A_2 + A_0 B_2}{\frac{B_2}{n} B_2 - D_2 - \frac{B_1}{n} B_2 - C_2 \left(\frac{B_1}{n} B_2 - C_2 \right)}$$

Using the value of a , we can describe the value of v like this:

$$v = -\frac{1}{\frac{B_1}{n} B_1 - B_2} \left(\left(\frac{B_1}{n} B_2 - C_2 \right) \frac{a}{2} + A_1 - A_0 B_1 \right)$$

and finally from a and v we can compute the value of b :

$$b = A_0 - v \frac{B_1}{n} - \frac{a}{2} \frac{B_2}{n}$$

4. Recursive formula

Adding to the known (t_i, p_i) coordinate pairs a new (t_{n+1}, p_{n+1}) one $(1 \leq i \leq n, n \geq 3)$ recursive formulas can be used for simpler and smarter model. This recursive approach is a natural improvement based on the computational method of the A_0, A_1, \dots, D_2 auxiliary variables.

In the beginning we should determine the $a = a_3, v = v_3,$ and $b = b_3$ coefficients with the above described closed version of least square method from the first three given points. When the new coordinate pair is added to the series $(n > 3)$, then we can calculate recursively for each steps the $a_4, a_5, \dots, a_n, v_4, v_5, \dots, v_n,$ and b_4, b_5, \dots, b_n coefficients. Of course this approach uses the indexed version of the auxiliary variables as the following

$$\begin{aligned}
 A_{0,n+1} &= \frac{nA_{0,n} + p_{n+1}}{n+1} & A_{1,n+1} &= A_{1,n} + p_{n+1}t_{n+1} & A_{2,n+1} &= A_{2,n} + p_{n+1}t_{n+1}^2 \\
 B_{1,n+1} &= B_{1,n} + t_{n+1} & B_{2,n+1} &= B_{2,n} + t_{n+1}^2 & C_{2,n+1} &= C_{2,n} + t_{n+1}^3 \\
 & & & & D_{2,n+1} &= D_{2,n} + t_{n+1}^4
 \end{aligned}$$

From these in each steps the

$$a = a_{n+1}$$

the

$$v = v_{n+1}$$

and the

$$b = b_{n+1}$$

values are easily computable. This version has the advantage that in computer implementation, we can re-use the A_0, A_1, \dots, D_2 auxiliary variables in each steps to store the actual values of $A_{0,i}, A_{1,i}, \dots, D_{2,i}$ $(3 \leq i \leq n + 1)$.

Due to the very sensitive nature of the least square method from a certain n , we can use the averages of the previous a_i, v_i and b_i coefficients like this $(3 \leq i \leq n + 1)$:

$$a = \overline{a_{n+1}} = \frac{a_3 + a_4 + \dots + a_{n+1}}{n - 1}$$

and

$$v = \overline{v_{n+1}} = \frac{v_3 + v_4 + \dots + v_{n+1}}{n - 1}$$

and

$$b = \overline{b_{n+1}} = \frac{b_3 + b_4 + \dots + b_{n+1}}{n - 1}$$

Because the least square method assumes homogenous data series this can compensate the distorting effect of a very different (t_i, p_i) coordinate pair.

5. Conclusions

With the knowledge of these parameters we can count the t_c , and t_d times. The

$$t_{c1,c2} = \frac{-v \pm \sqrt{v^2 - 2a(b - c)}}{a}$$

formula counts the t_{c1} and t_{c2} values, and examining the results only the

$$t_c = t_{c1} = \frac{-v + \sqrt{v^2 - 2a(b - c)}}{a}$$

will satisfy the requirements from a time value. This t_c result based on the c critical value, which depends from the age and race of the person.

This t_c time shows when the estimated PSA level will exceed the c critical value ($c > 0$). In the international literature the European standard for the critical values are in Table 1.

Age	Asian	African	Caucasian
40-49	2.0	2.0	2.5
50-59	3.0	4.0	3.5
60-69	4.0	4.5	4.5
70-79	5.0	5.5	6.5

Table 1

The

$$t_{d1,d2} = \frac{-v \pm \sqrt{v^2 - 2a(b - 2p_n)}}{a}$$

equation gives the exact values for the doubling time. Similar to previous reasons only the

$$t_d = t_{d1} = \frac{-v + \sqrt{v^2 - 2a(b - 2p_n)}}{a}$$

value will suitable as a time value.

This time indicates when the estimated PSA level will exceed the double of the last measured p_n PSA value. If the $t_d - t_n$ difference is shorter than a year then the tumorous process is considered booted. (This booted phase can occurs between the (t_i, p_i) pairs ($1 \leq i \leq n, n \geq 3$), when $p_j > 2p_i$ and $t_j - t_i < 1$ ($1 \leq i < j \leq n, n \geq 3$).) In the following in this paper we concentrate only for those cases when the booted phase is not reached before the time of the last PSA level test.

The current velocity of the change countable with the

$$\dot{p}(t) = v + at$$

formula. Here the acceleration of the PSA level raising is described with the

$$\ddot{p}(t) = a$$

equation.

6. Future steps

Based on this mathematical model a web based service is being planned under the name of PRO-FILTER, which will help doctors to evaluate the process of PSA changing and allow them to plan carefully the time of their next actions. The very basic computer implementation of this formula in PHP language is:

```
function estimationPSA($n, $t, $p, &$amp;a, &$amp;v, &$amp;b) {
    for ($i = 1; $i <= $n; $i++) {
        $A0 = $A0 + $p[$i] / $n;
        $A1 = $A1 + $p[$i] * $t[$i];
        $A2 = $A2 + $p[$i] * pow($t[$i], 2);
        $B1 = $B1 + $t[$i];
        $B2 = $B2 + pow($t[$i], 2);
        $C2 = $C2 + pow($t[$i], 3);
        $D2 = $D2 + pow($t[$i], 4);
    }
    $a = 2 * (((($B1 * $B2 / $n - $C2) / ($B1 * $B1 / $n - $B2)) * ($A1 - $A0 * $B1)) -
        ($A2 - $A0 * $B2)) / (($B2 * $B2 / $n - $D2) - ((($B1 * $B2 / $n - $C2) /
        ($B1 * $B1 / $n - $B2)) * ($B1 * $B2 / $n - $C2)));
    $v = (-1 / ($B1 * $B1 / $n - $B2)) * ((($B1 * $B2 / $n - $C2) * $a / 2) +
        ($A1 - $A0 * $B1));
    $b = $A0 - $v * $B1 / $n - ($a / 2) * ($B2 / $n);
}
```

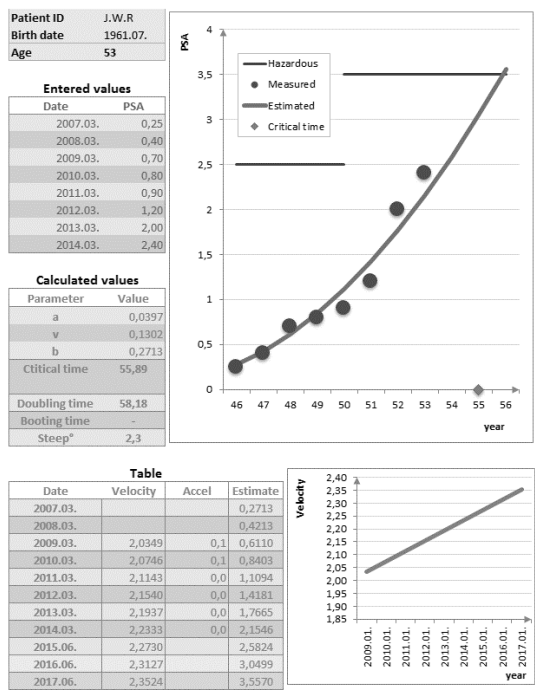


Figure 2

This code is using the closed equation without any investigation of variables or pre-conditions. Taking care on the values a , v and b parameters needs further improvements, The improved version of this algorithm produces the output on Fig. 2.

On this output we can see the entered values, the approximation polynomial with the special times and a few estimated PSA values ahead in time. The medical interpretation of the calculated value is currently under investigation.

This is the fundament of the algorithms in PRO-FILTER. There are more discussion about the additional parameters, and conditions in [3].

References

- [1] DR. NAGY KÁROLY MD: A prosztatata rosszindulatú betegségének rövid összefoglalása (Budapest, 31/03/2012, Manuscript)
- [2] DOBÓ ANDOR: Prosztaták terjedésének matematikai leírása, (Budapest, 07/03/2012, Manuscript)
- [3] PETER FESUS: Discussion of mathematical approach of PSA level test during prostate cancer analysis (Budapest, 31/01/2014, Manuscript)