IB Mathematics Higher Level Exploration on Statistics

Deriving an Alternative Data Approximation Method to Be Used in Non-typical Circumstances

Personal code: fzn106

Contents

1 Introduction

In both physics and mathematics practical work has as much value to humanity as theoretical does. In my school career there have been countless labs, practicals and experiments. I have come to notice that during all of these labs, practicals and experiments in any field, if the lab, practical or experiment is quantitative, the data gathered, almost always, has to be approximated into some kind of functional dependence to be of further use. This made me wonder how do these approximations work; how does a calculator know what the best fit line is? Further more, I've also noticed that in all of these practicals, the method of approximation that my classmates and I are told to use (and the one that is most commonly used) is the so called Least Square Method (LSM). What does it imply? What logic does it use to approximate? When does it work best, and when does it not work at all? I've grown concerned about the validity of my practical results that I have used in the past. What if the approximation had large errors? How big can those errors be?

In this exploration I aim to suggest an alternative method/algorithm of approximation, one that performs better in certain situations, in which the LSM does not. I will do so by deriving the formula for the optimal parameter(s) of a given functional dependence. Then I will use self-written software to (via the formula derived) calculate trials in which it is shown that the method that is suggested in this IA, indeed works better in given circumstances when compared to LSM.

2 Mathematical Background

2.1 Vectors and m-dimensional space

During the IB Math HL program, students learn about vectors. 2D-Vectors usually have only magnitude and direction, however, as we know a vector can be 3D, or even 4D. Essentially, so long that a single value has multiple parameters (components) it is a vector in the same number of dimensions as there are parameters. In this IA, all vectors will be denoted with the variable $\mathbf{c}$, be it with some indexes or accents, as long as a variable has $\mathbf{c}$ in its core, it is a vector. Note that this is just a notation that I will use only in this paper. In addition, I will be using herizontal vector notation like so: $\mathbf{c} = (c_1, c_2, \dots c_m)$. Where the vector $\mathbf{c}$ is in real coordinate space[1]; This allows several real variables to be treated as one, single variable. The notation $\boldsymbol{R}^m$ describes the dimension of space. i.e. $\boldsymbol{R}^1$ is one-dimensional space, $\boldsymbol{R}^2$ two-dimensional, and so on.[2]

2.2 Partial Derivatives

When it comes to functions with multiple variables, their derivative is taken differently. The derivative of such functions is called a partial derivative. Denoted using a stylized symbol ∂, it is the derivative of a function with multiple variables, with respect to one of them, when others are held constant [3] [4]. The following example will explain this further:

Find the local minimum of function z with multiple variables:

$$z = f(x, y) = 5(2x - 3)^2 + 4(4y + 1)^2$$

We see that the minimum of this function occurs when $x = \frac{3}{2}$,$y = -\frac{1}{4}$.

(Expanding this function gives)

$$z = 20x^2 + 64y^2 - 60x + 32y + 49$$

[1] Kelley (1995)
[2] Weisstein (2014)
[3] Matthews (1998), pages 45-48
[4] Stewart (2008a), pages 878-891

Now according to the normal method of finding a minimum of a function, we equate, in this case, its partial derivatives (one with respect to x, the other to y) to zero and solve for x and y.

$$\begin{aligned}\frac{\partial f(x,y)}{\partial x} &= 40x - 60 = 0 \rightarrow x = \frac{60}{40} = \frac{3}{2}\\ \frac{\partial f(x,y)}{\partial y} &= 128y - 32 = 0 \rightarrow y = -\frac{32}{128} = -\frac{1}{4}\end{aligned}$$

Partial derivatives gave the answer as well.

2.3 Multiple Integrals

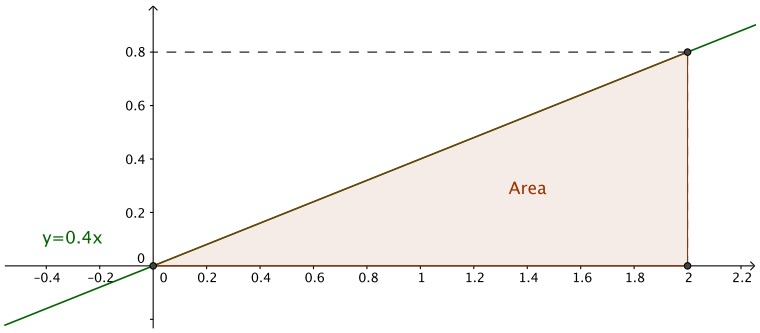
This exploration invokes the use of multiple integrals. These are basically integrals of other integrals, where any function of multiple variables is in question.[5] Here, each integral has a separate variable of integration, even though they stem from the same function. i.e. a double (two-multiple) integral of some two-variable function $g(x, y)$ is denoted as such:

$$\iint g(x,y)\mathrm{d}x\mathrm{d}y$$

Likewise, this can be applied to a function with m variables like so:

$$\int \cdots \int g(x, y, \ldots, m)\mathrm{d}x\mathrm{d}y \ldots \mathrm{d}m$$

Yet again let's employ an example:
Find the area of the shaded region



Obviously,

$$A = \frac{1}{2} \cdot 2 \cdot 0.8 = 0.8 \text{ units}^2$$

Granted, the following method is very unnecessary (and this problem doesn't even require a multiple integral), but for the sake of explanation, let's employ multiple integrals as well.

$$A = \underbrace{\int_0^2}_{\text{(against } x\text{)}} \underbrace{\int_0^{0.4x}}_{\text{(against } y\text{)}} \mathrm{d}x\mathrm{d}y = \int_0^2 \mathrm{d}x \cdot [y]_0^{0.4x} = \int_0^2 \mathrm{d}x \cdot [0.4x - 0] = \int_0^2 \mathrm{d}x \cdot 0.4x = 0.4\int_0^2 x\mathrm{d}x =$$

$$0.4 \cdot \left[\frac{x^2}{2}\right]_0^2 = 0.4 \cdot \left[\frac{2^2}{2} - \frac{0^2}{2}\right] = 0.4 \cdot 2 = 0.8.$$

And, again, the answer is the same.

[5] Stewart (2008b)

2.4 Lebesgue Measure

3 Formulation of the Problem

A function is one of the most known mathematical notions. An important task which has practical applications, is the approximation of a function or relationship based on some information known about the function or relationship in question. This information may either be determinate or statistical. An example of a determinate piece of information about function $f(x)$ is its range (or the possible values this function may have) on a given interval $[\alpha, \beta]$. Example of a statistical information may be the law of distribution of random errors ξ_i in approximate values $\tilde{y}_i = f(x_i) + \xi_i$ of the function, which in turn can describe a certain physical process (change of temperature over time, for example). In practice, n number of points x_i can be obtained as results of some kind of physical experiment. Where in this case, the approximation of function $f(x)$ only makes sense if this function is described by a finite number $m < n$ of parameters (coefficients) c_j, where the true values of said parameters will be denoted as $\dot{c}_j,\ j = 1, 2, \ldots, m$. For the sake of convenience, these parameters will be written as an m-dimensional vector; $\mathbf{c}$ storing the possible parameters and $\dot{\mathbf{c}}$ storing the true parameters.

This Internal Assessment will focus on the estimation of the parameters of a function

$$y = f(\dot{\mathbf{c}}, x), \quad \dot{\mathbf{c}} \in \boldsymbol{R}^m, \quad x \in [\alpha, \beta], \quad \dot{\mathbf{c}} = (\dot{c}_1, \dot{c}_2, \ldots, \dot{c}_m) \tag{3.0.0}$$[6]

based on its approximate values

$$\tilde{y}_i = f(x_i) + \xi_i, \quad i = 1, 2, \ldots, n, \tag{3.0.1}$$

when additionally it is also known, that: (1) vector $\dot{\mathbf{c}} = (\dot{c}_1, \dot{c}_2, \ldots, \dot{c}_m)$ belongs to a given limited set D, like for example a parallelepiped in $\boldsymbol{R}^m$; (2) ξ is a limited continuous random value; the median of which ($Med(\xi)$) is equal to zero.

Judging by references in scientific works that I read while researching for this exploration, the most popular linear model of a studied relationship is

$$f(\dot{\mathbf{c}}, x) = \sum_{j=1}^{m} \dot{\mathbf{c}}_j \phi_j(x), \tag{3.0.2}$$

specifically in polynomial form, when

$$\phi_j(x) = x^{j-1}, \quad j = 1, 2, \ldots, m; \quad \phi_1(x) = 1. \tag{3.0.3}$$

There are countless papers dedicated to the approximation of functions based on their approximate values (in practice - experimental data). Usually, in such papers the consensus is to use a certain condition. This condition is to assume that the mathematical expectancy of error is equal to zero [7].

$$E(\xi) = 0 \tag{3.0.4}$$

However, in this IA this condition will not be used. Here, instead of the condition of mathematical expectancy of error ξ being equal to zero, I will assume that the median of the same error ξ being equal to zero,

$$Med(\xi) = 0 \tag{3.0.5}$$

specifically when the approximation method of the parameters of the function is based on ideas of the *LSM*.

I justify my interest to the condition $Med(\xi) = 0$ by the case when the traditional condition $E(\xi) = 0$

[6] Here, because the vector $\dot{\mathbf{c}}$ has m parameters (components), it will also be in m-dimensional space.
[7] Plackett (1950)

is unachievable. This happens when measurements are taken close to one of the natural limits of the physical relationship being measured. An example of such natural limit is the inability of some magnitude, such as weight, to be negative. In this case, the absolute value of the error made, can only be large (with respect to other errors made) in the *same sign*, either positive or negative. Figure 1 shows an example of this graphically.

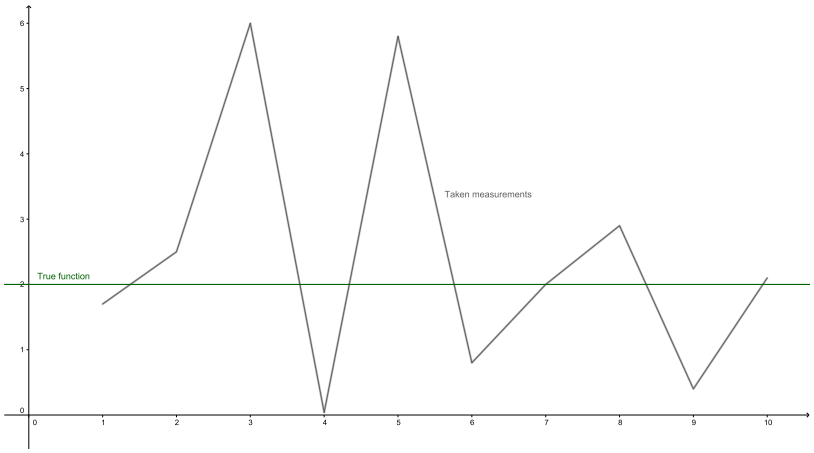


Figure 1: Graphical representation of a case where $E(\xi) = 0$ does not work effectively.

I want to bring attention to the fact, that the argument that this kind of measurement could be withdrawn by human intervention, is invalid for 2 reasons: (1) Any such withdrawal usually leads to loss of information. (2) In cases where the experiment requires high-capacity data collection human intervention may not be possible.

What is meant is not only error that was produced by a faulty measurement, but also any error caused by some factor that was either omitted or unaccounted for in function $f(x)$. Even though both conditions $E(\xi) = 0$ and $Med(\xi) = 0$ are not special cases of each other, it could be argued that from a point of view of solving practical problems, the condition $Med(\xi) = 0$ is the more broad of the two (as in, it is easier to meet). The only requirement for meeting this condition is that the probability of ξ being positive is 50%. Hence the condition $Med(\xi) = 0$ allows for some comparatively large random values of error ξ to be on one side of the true function and not on the other, without the approximation to be significantly affected by those large values, unlike the condition $E(\xi) = 0$. With condition $Med(\xi) = 0$ the approximation can account for large peaks in values of ξ.

As mentioned before, the aim of this exploration is to create an algorithm, capable of estimating the parameters of a functional dependence whose structure is known, accurately, and to securely give an estimate to the accuracy of the calculated approximate values of the functional dependence.

Its clear that the quality of the solution of this problem is dependant of an array of factors, which include: (1) the ratio between the n number of measurements $\tilde{y}_i$ and the m number of estimated parameters c_j. (2) the intensity of error ξ_i. This means that to confirm my theoretical reasoning, quite an ambitions computational experiment is required. I will proceed with the necessary calculations using custom software.

4 Algorithm

Approximation of a functional dependence taking this new condition $Med(\xi) = 0$ in mind, has been looked at in mathematics.[8] It is believed that in this case it is necessary to minimize the sum of absolute values of deviations of the modelled dependence $f(\mathbf{c}^*, x)$ from the unknown true function $f(\dot{\mathbf{c}}, x)$, where $\mathbf{c}^*$ is the found optimal value of vector $\mathbf{c}$. This method is referred to as the Least

[8] Balk (2010)

Absolute Deviations (LAD). However, through my research I have found no methods of estimating LAD's accuracy. What value does an optimization method have if there is no way to determining the error it made? In addition, LAD does not presume the existence of priori limitations on the vector $\dot{\mathbf{c}}$. And I must ask the question: What happens if the vector of parameters $\mathbf{c}^*$, providing the minimum of the sum of modulus of errors, does not belong to the set D?

It is clear, that in any separately taken case (run of an algorithm), the factual accuracy of the model solution (when the true function is known) cannot serve as either a comparative evaluation of two competing algorithms, nor criteria of effectiveness of any given algorithm. This is beacuse the intention of the algorithm itself is to be accurete in practice, i.e. when the true functionn is not known. One separately taken case when compared to the true function may give a good result however there is no way to tell if that result was accedental to that one separatly taken case. It is also clear, that if all, or close to all errors ξ_i have the same sign (the condition $Med(\xi) = 0$, although, the condition $E(\xi) = 0$ as well, allow this, be it with a small probability), then neither method will give any good solutions. So, when estimating the effectiveness of a method, it is necessary to rely on average results of some number of random solutions. In conjunction with this, I will take as a measure of accuracy of constructed approximation $f(\mathbf{c}^*, x)$ to the true function $f(\dot{\mathbf{c}}, x)$, the mathematical expectation[9]

$$E(\rho(\mathbf{c}^*, \dot{\mathbf{c}})) = \int_D P(\mathbf{c})\rho(\mathbf{c}^*, \dot{\mathbf{c}})\mathrm{d}c_1\mathrm{d}c_2 \dots \mathrm{d}c_m, \quad \mathbf{c} = (c_1, c_2, \dots, c_m) \tag{4.0.1}$$[10]

of proximity (distance) $\rho(\mathbf{c}^*, \dot{\mathbf{c}})$ of function $f(\mathbf{c}^*, x)$ from $f(\dot{\mathbf{c}}, x)$ where in the role of distance $\rho(\mathbf{c}^*, \dot{\mathbf{c}})$, one could take on of the functions

$$\rho_1(\mathbf{c}^*, \mathbf{c}) = \sum_{j=1}^{m} |cj^* - c_j| \qquad (\text{ taken from LAD }) \tag{4.0.2}$$

$$\rho_2(\mathbf{c}^*, \mathbf{c}) = \sum_{j=1}^{m} \left(c_j^* - c_j\right)^2 \qquad (\text{ taken from LSM }). \tag{4.0.3}$$

During my research I've found that formula (4.0.1) was considered in the article *'Reggression-type Problems under Zero Median Noise'* by Peter Balk (referenced earlier as (Balk, 2010)). There, Balk used a proximity function that is not considered in this exploration. The reason for me using the proximity functions (4.0.2) and (4.0.3) instead of the one used in (Balk, 2010), is that I will be comparing the method derived here to the already established methods: LSM and LAD. Using a different proximity function, to the ones in those methods, will lead to faulty results. Looking ahead, I say that I will suggest a more constructive algorithm than the one occurring in (Balk, 2010).

The probability density function $P(\mathbf{c})$, ($\mathbf{c} \in D$ where $\mathbf{c}$ is any vector that could be the unknown true vector $\dot{\mathbf{c}}$), which appears in the m-multiple integral (4.0.1), can be constructed on the basis of the formula of the binomial distribution of a random value [citation needed]. In fact, let's say: $\mathbf{c} \in D$ is one of the vectors which claims that it is the unknown true vector $\dot{\mathbf{c}}$ from function (3.0.2 on page 4); q_i are the elements of the sequence

$$q_1(\mathbf{c}) = \tilde{y}_1 - f(\mathbf{c}, x_1),\ q_2(\mathbf{c}) = \tilde{y}_2 - f(\mathbf{c}, x_2), \dots,\ q_n(\mathbf{c}) = \tilde{y}_n - f(\mathbf{c}, x_n); \tag{4.0.4}$$

where q is a discreet random value, that can assume values

$$r = r(\mathbf{c}) = \sum_{i=1}^{n-1} \delta_i(\mathbf{c}), \tag{4.0.5}$$

where

$$\delta_i = \delta_i(\mathbf{c}) = \begin{cases} 1, \text{if} & q_i(\mathbf{c})q_{i+1}(\mathbf{c}) < 0 \\ 0, \text{if} & q_i(\mathbf{c})q_{i+1}(\mathbf{c}) \geq 0 \end{cases} \tag{4.0.6}$$

[9] Ross (2007)
[10] Balk (2010)

In meaningful terms, δ_i is either 0 or 1 depending on if there has been a transition of sign between two consecutive elements of the sequence (4.0.5 on the previous page). (If the product of those two elements is negative, then one of those elements is negative; hence the transition of sign). Then the value of r is the number of transitions of sign of the elements of (4.0.4 on the preceding page) where $r \in [0, n-1]$. If it truly happens that $\mathbf{c} = \dot{\mathbf{c}}$ (the case that is being looked for), then the values of q_i would be nothing but the errors ξ_i, and by the condition $Med(\xi) = 0$ the probabilities p_r of r transitions of sign could be written as

$$p_r = \frac{\binom{n-1}{r}}{2^{n-1}}, \quad r = 0, 1, \ldots, n-1. \tag{4.0.7}$$[11]

Lets move on from discussing the question of the transition of sign with some one vector $\mathbf{c}$, to the analysis of this situation with regards to the whole set D. Say that there exists a partitioning (splitting) of set D into a family of sub-sets $D_1, D_2, \ldots, D_n$ such that the elements (in this case vectors $\mathbf{c}$) of sub-set D_r provide the same number of transitions of sign $r-1$ of elements of (4.0.4 on the previous page). Note that here the sub-sets D_r will take the form of m-dimensional polyhedrons [12] as they are in $\boldsymbol{R}^m$. In this way, the *a priori* probability[13] that the unknown true vector $\dot{\mathbf{c}} \in D_r$ in given by (4.0.7). However, in any separate case, some of the sub-sets D_r could be empty, meaning that with some vectors of parameters $\mathbf{c} \in D$, the number of transitions of sign of (4.0.4 on the previous page) will not equal the given value of r (any one can make sure of this by trying to draw a line which would crossed a given poly-line by guess, in such way that all of the endpoints of this poly-line turned out to be on either side of the guessed line). The a priori probabilities p_r that the unknown true vector $\dot{\mathbf{c}}$ is in one of the non-empty sub-sets D_r can be recalculated to the *posteriori* probabilities[14] p_r^*. Let's define I as the set of the numbers of all the non-empty sub-sets D_r. Then

$$p_r^* = \begin{cases} \frac{p_r}{\sum\limits_{s \in I} p_s}, & r \in I. \\ 0, & r \in \{0, 1, \ldots, n-1\} \setminus I. \end{cases} \tag{4.0.8}$$

To clarify, here the a priori probabilities suggested a chance that the true vector $\dot{\mathbf{c}}$ could be in sub-sets D_r which are empty. This is obviously not possible. The equation (4.0.8) nullifies those cases, and redistributes those probabilities proportionally across the experimentally possible cases.

Because in our case the vectors $\mathbf{c}$, are continuous values, and not discreet ones, then to use formula (4.0.1 on the previous page) we must proceed from estimates p_r^* of the probabilities of event $\dot{\mathbf{c}} \in D_r$ to estimates $p_r^*(\mathbf{c}), \mathbf{c} \in D_r$, of a probability density function. As there is no information suggesting the distribution of vectors $\mathbf{c}$ in the bounds of each sub-set D_r, it is logical to assume uniform distribution across each sub-set. Then for all $r \in I$

$$P_r^*(\mathbf{c}) = \frac{p_r^*}{\mu(D_r)}, \quad \mathbf{c} \in D_r, \tag{4.0.9}$$

where $\mu(D_r)$ is the Lebesgue measure (explained in section 2 on page 2) of sub-set D_r in $\boldsymbol{R}^m$.

Therefore (4.0.1 on the previous page) becomes

$$E(\rho(\mathbf{c}^*, \dot{\mathbf{c}})) = \sum_{r \in I} \frac{p_r^*(\mathbf{c})}{\mu(D_r)} \int\limits_{D_r} \rho(\mathbf{c}^*, \dot{\mathbf{c}}) \mathrm{d}c_1 \mathrm{d}c_2 \ldots \mathrm{d}c_m \tag{4.0.10}$$

(where $\int_{D_r} (\bullet)$ is a laconic notation of an m-multiple integral)

[11] (Balk, 2010)

[12] In geometry, a polyhedron is a solid in three (in this case m) dimensions with flat polygonal faces, straight edges and sharp vertices. A polyhedron is said to be convex, if any two points within it can be connected by a line segment with this segment also being within the polyhedron. Cromwell (1999)

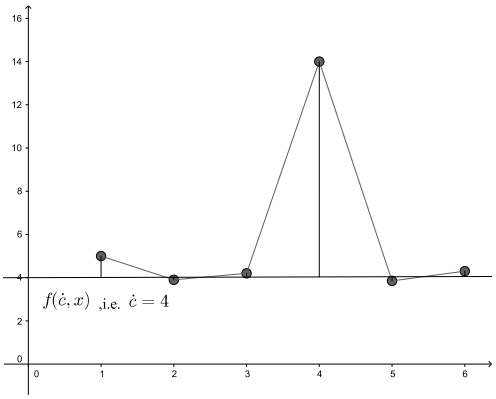
[13] An a priori probability is a probability that is derived purely by deductive reasoning. Mood et al. (1974)

[14] The posteriori probabilities of an event is the ratio of the number of outcomes in which a specified event occurs to the total number of trials, not in a theoretical sample space but in an actual experiment. In a more general sense, empirical probability estimates probabilities from experience and observation.Mood et al. (1974)

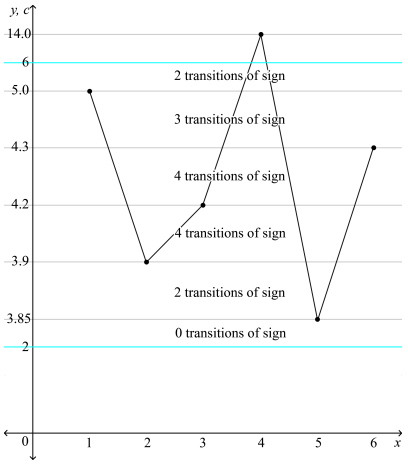
In the general case, such as this one, the limits of integration of each univariate integral depend on other variables, based on which the integration of the external integral relative to the given univariate integral is carried out [15]. (Refer to the example in section 2 on page 2). This significantly complicates the calculation of these multiple integrals. Yet again, looking ahead, I want to note that in the algorithm created, integration is carried out on multidimensional parallelepipeds, when the limits of integration of each integral remain constant.

4.1 Illustrative example

For it to be clear, that the following bulky equations lead to success, let's consider an illustrative example, where $m = 1$ (i.e. the vector $\dot{\mathbf{c}}$ is a scalar value), $y = f(\mathbf{c}; x)$ is a function with one parameter, where $\dot{\mathbf{c}}$ is a scalar that has to be found. In this case $n = 6$ and $D = \{\mathbf{c} : 2 \leq \mathbf{c} \leq 6\}$. In this way, the following graphical representations form.



(a) Graph of proposed example



(b) Illustration of transitions of sign q_i between each $\tilde{y}_i$

Figure 2

Table 1: Table of values for this example

i	x_i	$f(\dot{\mathbf{c}}, x_i)$	ξ_i	$\tilde{y}_i = f(\dot{\mathbf{c}}, x_i) + \xi_i$
1	1	4	1.0	5.0
2	2	4	-0.1	3.9
3	3	4	0.2	4.2
4	4	4	10	14.0
5	5	4	-0.15	3.85
0	6	4	0.3	4.3

For this example I will use proximity function (4.0.3 on page 6) also used in LSM. I am doing this as I will later compare the two methods.
As we already know, we can calculate the priori probabilities using (4.0.7 on the preceding page), where $r = 0, 1, 2, 3, 4, 5$ like so:

$$p_0 = \frac{1}{32};\ p_1 = \frac{5}{32};\ p_2 = \frac{10}{32};\ p_3 = \frac{10}{32};\ p_4 = \frac{5}{32};\ p_5 = \frac{1}{32}$$

But can only have certain values: $r = 0, 2, 3, 4$. And ,so

$$s = p_0 + p_2 + p_3 + p_4 = \frac{1}{32} + \frac{10}{32} + \frac{10}{32} + \frac{5}{32} = \frac{26}{32}.$$

[15] Stewart (2008b)

So, to calculate the posteriori probabilities

$$\begin{aligned}
p_0^* &= p_0 : \frac{26}{32} = \frac{1}{32} : \frac{26}{32} = \frac{1}{26} \\
p_1^* &= 0 \\
p_2^* &= p_2 : \frac{26}{32} = \frac{10}{32} : \frac{26}{32} = \frac{10}{26} \\
p_3^* &= p_3 : \frac{26}{32} = \frac{10}{26} \\
p_4^* &= p_4 : \frac{26}{32} = \frac{5}{32} : \frac{26}{32} = \frac{5}{26} \\
p_5^* &= 0
\end{aligned}$$

Calculating $E(\rho(\mathbf{c}^*, \mathbf{c}))$

$$\begin{aligned}
E(\rho(\mathbf{c}^*, \mathbf{c})) = &\int_2^{3.85} \underbrace{(\mathbf{c}^* - \mathbf{c})^2}_{\rho(\mathbf{c}^*, \mathbf{c})} \cdot \underbrace{\overbrace{\frac{1}{26}}^{p_0^*}}_{\mu(D_0)} \mathrm{d}\mathbf{c} \; + \int_{3.9}^{4.3} (\mathbf{c}^* - \mathbf{c})^2 \cdot \frac{\frac{5}{26}}{4.3 - 3.9} \mathrm{d}\mathbf{c} \; + \int_{4.3}^{5.0} (\mathbf{c}^* - \mathbf{c})^2 \cdot \frac{\frac{10}{26}}{5.0 - 4.3} \mathrm{d}\mathbf{c} \\
&+ 1.05 \cdot \left(\int_{3.85}^{3.9} (\mathbf{c}^* - \mathbf{c})^2 \cdot \frac{10}{26} \mathrm{d}\mathbf{c} + \int_{5.0}^{6.0} (\mathbf{c}^* - \mathbf{c})^2 \cdot \frac{10}{26} \mathrm{d}\mathbf{c} \right) \\
&= (\mathbf{c}^*)^2 \cdot \mathbf{c} - \mathbf{c}^* \cdot \mathbf{c}^2 + \frac{1}{3}\mathbf{c}^3
\end{aligned}$$

Taking into account that later we will want to find the derivative $\frac{\mathrm{d}E(\rho(\mathbf{c}^*, \mathbf{c}))}{\mathrm{d}\mathbf{c}^*}$, we can omit the summand $\frac{1}{3}\mathbf{c}^3$ as it is a constant, and the derivative of a constant is zero. Let's continue.

$$\begin{aligned}
E(\rho(\mathbf{c}^*, \mathbf{c})) = \; & 0.045 \; \cdot \left[(\mathbf{c}^*)^2 \cdot \mathbf{c} - \mathbf{c}^* \cdot \mathbf{c}^2 \right]_2^{3.85} \\
+ \; & 7.323 \; \cdot \left[(\mathbf{c}^*)^2 \cdot \mathbf{c} - \mathbf{c}^* \cdot \mathbf{c}^2 \right]_{3.85}^{3.9} \\
+ \; & 0.481 \; \cdot \left[(\mathbf{c}^*)^2 \cdot \mathbf{c} - \mathbf{c}^* \cdot \mathbf{c}^2 \right]_{3.9}^{4.3} \\
+ \; & 0.549 \; \cdot \left[(\mathbf{c}^*)^2 \cdot \mathbf{c} - \mathbf{c}^* \cdot \mathbf{c}^2 \right]_{4.3}^{5.0} \\
+ \; & 0.366 \; \cdot \left[(\mathbf{c}^*)^2 \cdot \mathbf{c} - \mathbf{c}^* \cdot \mathbf{c}^2 \right]_{5.0}^{6.0} \\
= \; & 1.361 \; \cdot (\mathbf{c}^*)^2 - 12.061 \mathbf{c}^*
\end{aligned}$$

$$\begin{aligned}
\frac{\mathrm{d}E(\rho(\mathbf{c}^*, \mathbf{c}))}{\mathrm{d}\mathbf{c}^*} &= 2.722\mathbf{c}^* - 12.061 = 0 \\
\mathbf{c}^* &= \frac{12.061}{2.722} = 4.431
\end{aligned}$$

Seeing that $\dot{\mathbf{c}} = 4.0$, we finally calculate ρ using this method

$$\rho = |4.431 - 4.0| = 0.431$$

Error made by this method is 0.431. Now let's see what kind of error will the LSM give:

Using Least Square Method

$$\phi(\mathbf{c}^*) = \sum_{i=1}^{6} (\tilde{y}_i - \mathbf{c}^*)^2 \longrightarrow min \qquad \text{(Sum of the squares of errors has to be minimised)}$$

$$\phi(\mathbf{c}^*) = \sum_{i=1}^{6} (\tilde{y}_i^2 - 2\tilde{y}_i\mathbf{c}^* + (\mathbf{c}^*)^2) \qquad \text{(Expanding the polynomeal)}$$

$$\frac{\mathrm{d}\phi(\mathbf{c}^*)}{\mathrm{d}\mathbf{c}^*} = \sum_{i=1}^{6} (0 - \cancel{2}\tilde{y}_i + \cancel{2}\mathbf{c}^*) = 0 \qquad \text{(Finding the derivative and equating it to zero)}$$

$$\sum_{i=1}^{6} \mathbf{c}^* = \sum_{i=1}^{6} \tilde{y}_i \qquad \text{(Re-arranging the result)}$$

$$6 \cdot \mathbf{c}^* = \sum_{i=1}^{6} \tilde{y}_i = 35.25$$

$$\mathbf{c}^* = \frac{35.25}{6} = 5.875 \qquad \text{(Solving for } \mathbf{c}^*\text{)}$$

And so the error $\rho = 1.875$

LSM gives a bigger error because in it deviations are squared [16], and therefore large derivations are weighted more heavily. Because this example shows, that in the case of significant non-compliance to condition $E(\xi) = 0$, my method is notably superior to LSM, and because of the limited format of this IA, I will not include further solutions that used LSM. Back to the theory.

Let the proximity $\rho(\mathbf{c}^*, \mathbf{c})$ in (4.0.10 on page 7), be taken in the form (4.0.3 on page 6). Then the following simplification takes place.

$$E(\rho(\mathbf{c}^*, \dot{\mathbf{c}})) = \sum_{r\in I} \frac{p_r^*}{\mu(D_r)} \int\limits_{D_r} \sum_{j=1}^{m} \left(c_j^* - c_j\right)^2 \mathrm{d}c_1\mathrm{d}c_2 \ldots \mathrm{d}c_m =$$

$$expanding \left(c_j^* - c_j\right)^2$$

$$\sum_{r\in I} \frac{p_r^*}{\mu(D_r)} \sum_{j=1}^{m} \left[(c_j^*)^2 \underbrace{\int\limits_{D_r} \mathrm{d}c_1\mathrm{d}c_2 \ldots \mathrm{d}c_m}_{=\mu(D_r)} - 2c_j^* \int\limits_{D_r} c_j \ \mathrm{d}c_1\mathrm{d}c_2 \ldots \mathrm{d}c_m + \int\limits_{D_r} c_j^2 \ \mathrm{d}c_1\mathrm{d}c_2 \ldots \mathrm{d}c_m \right] =$$

$$\sum_{j=1}^{m} \left(\sum_{r\in I} \frac{p_r^*}{\mu(D_r)} \left[(c_j^*)^2\mu(D_r) - 2c_j^* \int\limits_{D_r} c_j \ \mathrm{d}c_1\mathrm{d}c_2 \ldots \mathrm{d}c_m + \int\limits_{D_r} c_j^2 \ \mathrm{d}c_1\mathrm{d}c_2 \ldots \mathrm{d}c_m \right] \right) =$$

$$\sum_{j=1}^{m} \left((c_j^*)^2 \cancel{\mu(D_r)} \underbrace{\sum_{r\in I} \frac{p_r^*}{\cancel{\mu(D_r)}}}_{\sum_{r\in I} p_r^* = 1} - 2c_j^* \sum_{r\in I} \frac{p_r^*}{\mu(D_r)} \int\limits_{D_r} c_j \ \mathrm{d}c_1\mathrm{d}c_2 \ldots \mathrm{d}c_m + \sum_{r\in I} \frac{p_r^*}{\mu(D_r)} \int\limits_{D_r} c_j^2 \ \mathrm{d}c_1\mathrm{d}c_2 \ldots \mathrm{d}c_m \right) \tag{4.1.1}$$

[16] Plackett (1950)

And finally:

$$E(\rho(\mathbf{c}^*,\dot{\mathbf{c}})) = \sum_{j=1}^{m}\left((c_j^*)^2 - 2c_j^*\sum_{r\in I}p_r^*\frac{\int\limits_{D_r} c_j\ \mathrm{d}c_1\mathrm{d}c_2\ldots\mathrm{d}c_m}{\mu(D_r)} + \underbrace{\sum_{r\in I}p_r^*\frac{\int\limits_{D_r} c_j^2\ \mathrm{d}c_1\mathrm{d}c_2\ldots\mathrm{d}c_m}{\mu(D_r)}}_{\text{derivative of this with respect to } c_1^*,c_2^*,\ldots,c_m^*=0}\right) \tag{4.1.2}$$

In order to find the vector c^*, which provides the minimum of function (4.1.2) and which I will take as the optimal estimate of the unknown vector $\dot{c}$, it is necessary (like in the case with a simple one-parameter function) to take the first partial derivative $\frac{\partial E}{\partial c_j^*}$ of function (4.1.2) with respect to each variable c_j^* [17]. After that, equate these derivatives to zero and solve the resulting system of linear equations. In this case, this system splits into m separate linear equation with one variable:

$$2c_j^* - 2\sum_{r\in I}p_r^*\frac{\int\limits_{D_r} c_j\mathrm{d}c_1\mathrm{d}c_2\ldots\mathrm{d}c_m}{\mu(D_r)} = 0, \quad j = 1,2,\ldots,m. \tag{4.1.3}$$

These equations have the following solutions:

$$c_j^* = \sum_{r\in I}p_r^*\bar{c}_{(j,r)}, \quad j = 1,2,\ldots,m, \tag{4.1.4}$$

where

$$\bar{c}_{(j,r)} = \frac{\int\limits_{D_r} c_j\mathrm{d}c_1\mathrm{d}c_2\ldots\mathrm{d}c_m}{\mu(D_r)}. \tag{4.1.5}$$

5 Algorithm trials

[17] Stewart (2008a)

6 Conclusion

7 Appendix A

7.1 Method Description of Computation of Theoretical Algorithm

Here I will make use of the product notation:

$$\prod_{i=m}^{n} x_i = x_m \cdot x_{m+1} \cdot \ \cdots \ \cdot x_{n-1} \cdot x_n.$$

This essentially works in much of the same way the familiar summation notation we use in class(using the Greek capital letter sigma), except that instead of summing values, it multiplies them (and uses Greek capital letter pi).[18]

From the point of view of practical realisation of my algorithm, that is based of formulae (4.1.4 on page 11),(4.1.5 on page 11), I can ask two questions: (1) Is it possible to derive a simple method of constructing the sets D_r? (2) Is it possible to derive a simple method of calculation the integrals of those sets D_r, which appear in the RHS of (4.1.5 on page 11). What a geometrical construction (Figure omitted) has shown me, is that even in the case where $m = 2$ and $n = 10$, the regions D_r have a relatively complicated geometry (some multi-peaked star-like shapes). However, this problem can be solved with the principal of the 'Gordian knot' - refuse to work with directly with sets D_r, but instead, to use their point (grid) approximation (even without describing these sets concretely) and use an analog of (4.1.4 on page 11).

The essence of this suggested method of thinking is as follows. Define $k(1), k(2), \ldots, k(m)$ as some quite large real values, and a m-dimensional parallelepiped

$$W = \{c : c_j^{(\min)} \leq c_j \leq c_j^{(\max)}, \quad j = 1, 2, \ldots, m\} \subset \boldsymbol{R}^m \tag{7.1.1}$$

which contains the set D (if the set D itself is a parallelepiped, then $W = D$). Cover this parallelepiped with a dense grid Γ with

$$L = \prod_{t=1}^{m} k(t) \tag{7.1.2}$$

number of nodes $\mathbf{c}^{(l)} = (c_1^{(l)}, (c_2^{(l)}, \ldots, c_m^{(l)}), \quad l = 1, 2, \ldots, L$, where each coordinate $c_j^{(l)}, \quad j = 1, 2, \ldots, m$, could independently from all the other $m - 1$ coordinates (note that $m - 1$ does not refer to sets of coordinates) assume one of $k(j)$ of values

$$c_j = c_j^{(\min)} + \frac{c_j^{(\max)} - c_j^{(\min)}}{k(j)}(t - \frac{1}{2}), \quad t = 1, 2, \ldots, k(j). \tag{7.1.3}$$

Parallel to grid Γ implement the system A of parallelepipeds $W_l \subset \boldsymbol{R}^m, \quad l = 1, 2, \ldots, L$, the centres of which are their nodes with sides

$$h_j = \frac{(c_j^{(\max)} - c_j^{(\min)})}{k(j)} : \tag{7.1.4}$$

$$W_l = \{\mathbf{c} = (c_1, c_2, \ldots, c_m) : c_j^{(l)} - \frac{1}{2}h_j \leq c_j \leq c_j^{(l)} + \frac{1}{2}h_j, \quad j = 1, 2, \ldots, m\}. \tag{7.1.5}$$

These parallelepipeds form the surface of the parallelepiped W. Now we have to chose from the defined L nodes those, the ones who belong to the set D. Let L_0 be the quantity of the chosen nodes. Renumber the chosen nodes, by assigning the first L_0 numbers l to them.

For each $l = 1, 2, \ldots, L_0$ find elements of $q_i(c^{(l)}), \quad i = 1, 2, \ldots, n$ of sequence (4.0.4 on page 6), and based on their values, using (4.0.5 on page 6), calculate the number r of transition of sign of

[18] I.S.U.M.D (1995)

elements of (4.0.4 on page 6). Let $J_r, \quad r = 0, 1, \ldots, n-1$ be the set of numbers l of chosen nodes $c^{(l)}$ providing r transitions of sign, and $|J_r|$ - the number of nodes giving that many transitions. In this way, without directly constructing the sub-sets D_r, I approximate each of them with the union.[19]

$$\tilde{D}_r = \bigcup_{l \in J_r} W_l. \tag{7.1.5}$$[20]

It is clear, that the assumption, that all vectors $\mathbf{c}$, belonging to the parallelepiped W_l, will provide the same number of transitions of sign as the center $\mathbf{c}^{(l)}$ of the parallelepiped, will only be incorrect for those parallelepipeds W_l, and do not completely belong to some one sub-set D_r. But with a dense grid Γ, and respectively, quite small-sized parallelepipeds W_l, the percent of such 'debatable' parallelepipeds compared to their general number L_0 will be quite small (so, in figure (omitted), when $m = 2$, these 'debatable' parallelepipeds are positioned along the lines, and the 'non-debatable' - in the areas). In addition, the error resulting from classification of such 'debatable' parallelepipeds W_l completely to some set D_r, will get (mostly) redeemed/repaid. Summing up all these arguments, it can be said, that the approximations of sets D_r, with the union of relatively 'small' parallelepipeds will give a good approximation of the mathematical expectancy (4.0.10 on page 7) of error in estimation of parameters of a studies functional dependence:

$$E(\rho_2(\mathbf{c}^*, \mathbf{c})) \approx \sum_{r \in I} \frac{p_r^*}{\mu(\tilde{D}_r)} \int\limits_{\tilde{D}_r} \sum_{j=1}^{m} (c_j^* - c_j)^2 \mathrm{d}c_1 \mathrm{d}c_2 \ldots \mathrm{d}c_m =$$

(integral with limit D_r can be split as the sum of integrals with limit W_l)

$$\sum_{r \in I} \frac{p_r^*}{\mu(\tilde{D}_r)} \sum_{l \in J_r} \int\limits_{W_l} \sum_{j=1}^{m} (c_j^* - c_j)^2 \mathrm{d}c_1 \mathrm{d}c_2 \ldots \mathrm{d}c_m = \tag{7.1.6}$$

(the sum $\sum_{j=1}^{m}$ is facrotised out)

$$\sum_{j=1}^{m} \left[\sum_{r \in I} \frac{p_r^*}{\mu(\tilde{D}_r)} \left(\sum_{l \in J_r} \int\limits_{W_l} ((c_j^*)^2 - 2c_j^* c_j + c_j^2) \right) \mathrm{d}c_1 \mathrm{d}c_2 \ldots \mathrm{d}c_m \right] =$$

(expanding $((c_j^*)^2 - 2c_j^* c_j + c_j^2)$)

$$\sum_{j=1}^{m} \left[\sum_{r \in I} \frac{p_r^*}{\mu(\tilde{D}_r)} \left(\sum_{l \in J_r} \left[(c_j^*)^2 \mu(W_{(l)}) - 2c_j^* \int\limits_{W_l} c_j \mathrm{d}c_1 \mathrm{d}c_2 \ldots \mathrm{d}c_m + \int\limits_{W_l} c_j^2 \mathrm{d}c_1 \mathrm{d}c_2 \ldots \mathrm{d}c_m \right] \right) \right]. \tag{7.1.7}$$

If in a multiple integral, the limits of integration for each variable are given by other variables, then this multiple integral is equal to the product of included in it, one-dimensional integrals [21]. Now, if we define coordinates of the edges of W_l as $c_j^{(l)} - 0.5h_j = a(j,l)$ and $c_j^{(l)} + 0.5h_j = b(j,l)$ then it is not hard to be sure that

$$\frac{1}{\mu(W_l)} \int\limits_{W_l} c_j \mathrm{d}c_1 \mathrm{d}c_2 \ldots \mathrm{d}c_m = \frac{1}{\prod\limits_{s=1}^{m} h_s} \left(\int\limits_{b(j,l)}^{a(j,l)} c_j \mathrm{d}c_j \right) \prod_{\substack{s=1 \\ (s \neq j)}}^{m} \int\limits_{a(s,l)}^{b(s,l)} \mathrm{d}c_s = c_j^{(l)}. \tag{7.1.8}$$

[19] Smith et al. (2015)

[20] Here, $\tilde{D}_r$ is the arbitrary union of all sets W_l such that l is an element of J_r.

[21] Stewart (2008b)

If in addition we accept that

$$\sum_{r\in I} p_r^* = 1, \tag{7.1.9}$$

$$\sum_{l\in J_r} \mu(W_l) = \mu(\tilde{D}_r), \tag{7.1.10}$$

$$\mu(W_l) \equiv \prod_{j=1}^{m} h_j, \tag{7.1.11}$$

then the simplification chain (7.1.7 on the previous page) could be continued

$$E(\rho_2(\mathbf{c}^*,\mathbf{c})) \approx \sum_{j=1}^{m}\left[\sum_{r\in I}\frac{p_r^*}{\mu(\tilde{D}_r)}\left(\sum_{l\in J_r}\left[(c_j^*)^2\mu(W_l) - 2c_j^*\mu(W_l)c_j^{(l)} + \int\limits_{W_l} c_j^2 \mathrm{d}c_1\mathrm{d}c_2\ldots\mathrm{d}c_m\right]\right)\right] =$$

$$\left(\text{ middle step: } \sum_{l\in J_r}^{m}(c_j^*)^2\mu(W_l) = (c_j^*)^2\underbrace{\sum_{l\in J_r}\mu(W_l)}_{=\mu(\tilde{D}_r)} = (c_j^*)^2\mu(\tilde{D}_r)\ \right)$$

$$\sum_{j=1}^{m}\left[(c_j^*)^2\sum_{r\in I}p_r^* - 2c_j^*\sum_{r\in I}\frac{p_r^*}{\mu(\tilde{D}_r)}\sum_{l\in J_r}\mu(W_l)c_j^{(l)} + \sum_{r\in I}\frac{p_r^*}{\mu(\tilde{D}_r)}\sum_{l\in J_r}\int\limits_{W_l} c_j^2\mathrm{d}c_1\mathrm{d}c_2\ldots\mathrm{d}c_m\right] =$$

$$\left(\text{ middle step: } \sum_{r\in I}(c_j^*)^2\frac{p_r^*}{\cancel{\mu(\tilde{D}_r)}}\cancel{\mu(\tilde{D}_r)} = (c_j^*)^2\underbrace{\sum_{r\in I}p_j^*}_{=1} = (c_j^*)^2\ \right)$$

$$\sum_{j=1}^{m}\left[(c_j^*)^2 - 2c_j^*\sum_{r\in I}p_j^*\frac{1}{\sum\limits_{l\in J_r}\mu(W_l)}\sum_{l\in J_r}\mu(W_j)c_j^{(l)} + \sum_{r\in I}\frac{p_r^*}{\mu(\tilde{D}_r)}\sum_{l\in J_r}\int\limits_{W_l} c_j^2\mathrm{d}c_1\mathrm{d}c_2\ldots\mathrm{d}c_m\right] =$$

$$\sum_{j=1}^{m}\left[(c_j^*)^2 - 2c_j^*\sum_{r\in I}p_j^*\frac{1}{|J_r|\prod\limits_{s=1}^{m}h_s}\sum_{l\in J_r}\left(\prod_{s=1}^{m}h_s\right)c_j^{(l)} + \sum_{r\in I}\frac{p_r^*}{\mu(\tilde{D}_r)}\sum_{l\in J_r}\int\limits_{W_l} c_j^2\mathrm{d}c_1\mathrm{d}c_2\ldots\mathrm{d}c_m\right] =$$

$$\sum_{j=1}^{m}\left[c_j^*)^2 - 2c_j^*\sum_{r\in I}\frac{p_j^*}{|J_r|}\sum_{l\in J_r}c_j^{(l)} + \sum_{r\in I}\frac{p_r^*}{\mu(\tilde{D}_r)}\sum_{l\in J_r}\int\limits_{W_l} c_j^2\mathrm{d}c_1\mathrm{d}c_2\ldots\mathrm{d}c_m\right]. \tag{7.1.12}$$

In order to find vector $\mathbf{c}^*$, which provides the minimum estimate mathematical expectation (7.1.12) of error in the approximation of unknown true parameters of the functional dependence in question, and therefore also of vector $\dot{\mathbf{c}}$, we take the same steps as in the case when we calculated the minimum of function (4.1.2 on page 11). Taking into account, that the last, third, summand in (7.1.12) does not contain any variables c_j^* (which are being optimised), its derivative is again equal to zero, we find its partial derivatives:

$$\frac{\partial E(\rho_2(\mathbf{c}^*,\mathbf{c}))}{\partial c_j^*} \approx 2c_j^* - 2\sum_{r\in I}\frac{p_r^*}{|J_r|}\sum_{l\in J_r}c_j^{(l)}, \quad j = 1,2,\ldots,m. \tag{7.1.13}$$

Equating these partial derivatives (7.1.13) to zero we get, and therefore conclude with:

$$c_j^* = \sum_{r\in I}\frac{p_r^*}{|J_r|}\sum_{l\in J_r}c_j^{(l)}, \quad j = 1,2,\ldots,m. \tag{7.1.14}$$

This equation (7.1.14 on the preceding page) is the one used in calculations of approximate values of the optimised vector $\mathbf{c}^*$ in my custom software, the trails of which are earlier in this document.

Now, let's compare previousy suggested in "Balk (2010)" equations (4.1.4 on page 11) and (4.1.5 on page 11) with the suggested equation (7.1.14 on the previous page). The superiority of the later is shown in that: 1. it does not require the direct construction of sub-sets D_r , which in fact are quite complex convex polyhedrons in $\boldsymbol{R}^m$. Therefore 2. this equation does not require the expansive calculation of multiple integrals found in (4.1.4 on page 11) and (4.1.5 on page 11). In "Balk (2010)" there is one more problem at hand: even a polygon, not to mention a polyhedron, can not be defined only by its vertexes (those could be connected with line segments in not just one way). The algorithm proposed in this IA to estimate the optimal parameters $\mathbf{c}^*$, only requires the regular (evenly distributed) surface of set D, for the nodes of which the only thing left to do is evaluate the elements of the sequence (4.0.4 on page 6).

References

Balk, P. (2010). Regression-type problems under zero median noise. 81(1):142–145.

Cromwell, P. R. (1999). *Polyhedra*. Cambridge University Press.

I.S.U.M.D (1995). *Summation and Product Notation*.

Kelley, J. L. (1995). *General topology*. Springer.

Matthews, P. C. (1998). *Vector calculus*, pages 45–48. Springer.

Mood, A. M., Graybill, F. A., and Boes, D. C. (1974). *Introduction to the theory of statistics*. McGraw-Hill, 3rd edition edition.

Plackett, R. L. (1950). Some theorems in least squares. *Biometrika.*, 37(1-2):149–157.

Ross, S. M. (2007). 2.4 expectation of a random variable. In *Introduction to probability models (9th ed)*. Academic Press.

Smith, D., Eggen, M., and Andre, R. S. (2015). *A transition to advanced mathematics*. Cengage Learning.

Stewart, J. (2008a). *Calculus: early transcendentals*, pages 878–891. Thomson Brooks/Cole.

Stewart, J. (2008b). *Calculus: early transcendentals*, pages 950–1021. Thomson Brooks/Cole.

Weisstein, E. W. (2014). Dimension.