The Cramer-Rao inequality to go beyond the $\sqrt{N}$-limit of the standard least-squares method in track fitting.

Gregorio Landi$^a$, Giovanni E. Landi$^b$

$^a$ Dipartimento di Fisica e Astronomia, Universita’ di Firenze and INFN
Largo E. Fermi 2 (Arcetri) 50125, Firenze, Italy

$^b$ ArchonVR S.a.g.l.,
Via Cisieri 3, 6900 Lugano, Switzerland.

October 30, 2019

Abstract

The Cramer-Rao-Frechet inequality is reviewed specializing it to track fitting. A diffused opinion attributes to this inequality the limitation of the resolution of the track fits with the number N of observations. It turns out that this opinion is incorrect, weighted least squares method is not subjected to that N-limitation. In a previous publication, simulations with a simple Gaussian model produced interesting results: a linear growth of the peaks of the distributions with the number $N$ of observations, much faster than the $\sqrt{N}$ of the standard least squares. These results could be considered a violation of a well known $1/N$-rule for the variance of an unbiased estimator, frequently reported as the Cramer-Rao-Frechet bound. To clarify this point beyond any doubt, it would be essential a direct proof of the consistency of those results with this inequality. Unfortunately, such proof is lacking or very difficult to find. Hence, the Cramer-Rao-Frechet developments are applied to prove the efficiency (optimality) of the simple Gaussian model and the consistency of its results. The inequality remains valid even for irregular models supporting the results of realistic models with similar growths.

Keywords: Cramer-Rao Bound, Least Squares Method, Track fitting, Silicon Microstrip Detectors.

Contents

1 Introduction

2 The standard CRF-inequality
   2.1 The CRF inequality
   2.2 Efficient estimators
   2.3 The standard mean with an i.i.d. Gaussian model

3 CRF inequality for heteroscedastic models
   3.1 The weighted mean in heteroscedastic Gaussian models
   3.2 The weighted mean in a homoscedastic model

*Corresponding author. Gregorio.Landi@fi.infn.it
1 Introduction

Very recently criticisms and doubts were raised against our paper (ref. [1]) on track fitting. The contested results were principally those of the following fig. 1. This figure illustrated (Monte Carlo) simulations of a very simple tracker with N detecting layers of identical technology (silicon microstrips) crossed by a set of parallel straight tracks of minimum ionizing particles (MIPs). The reported estimator was the track direction.

![Figure 1: The simple Gaussian model of ref.[1]. Distributions of the fits to the track directions for tracker models with N=2 to N=13 detecting layers. The first distributions are centered on zero, the others are shifted of N-2 steps of fixed amplitude. The blue distributions are the results of the standard least-squares. The red distributions are obtained with the weighted least-squares.](image)

The hit uncertainties were approximated with Gaussian distributions with two type of hit quality (heteroscedastic model). One of very good quality when the MIP crosses the detector near to the borders...
(20% of strip width), and the other of bad quality (80% of the strip width) around the strip center. A slight random mechanical misalignment was supposed with mean value around a half strip (30 \(\mu m\)). Random misalignments much larger than this are always present in real trackers. This misalignment (rotations and translations) is simulated as a randomization of the hit position on the strips producing a binomial distribution of the hit quality. The collection of \(N\) hits (a track) is fitted with two different least squares method, one with the standard least squares and the other with the weighted least-squares with the weights given by the hit quality. The track direction is the illustrated estimator. With standard least-squares, we define the usual least-squares where all the used observations have identical variances (homoscedastic models). The standard least-squares, built for homoscedastic models, when applied to heteroscedastic models gives results almost insensitive to the presence of the good hits, and the maximums of the empirical probability density functions (PDFs) show a growth very near to \(\sqrt{N}\). The blue lines of fig.1. Instead, weighted least-squares, consistent for the heteroscedastic models, reports a surprising fast increase, almost linear, of the maximums of the simulated PDFs. Translating these results on mean variances, the standard fit shows a \(1/N\) trend and the weighted fit shows a trend of the mean variances as \(1/N^2\). This trend \(1/N^2\) was considered a violation of the Cramer-Rao-Frechet (CRF) bound as reported in many books and documentations. Our arguments about the different hypotheses used to prove the \(1/N\) rule and those of our approach had no effect. To settle any type of contrast, a complete demonstration is required to prove our consistency with the CRF-inequality. It is evident that this long demonstration could not be inserted in ref. [1]. The main part of ref. [1] deals with physical models with Cauchy-like tails and (in theory) with infinite variance. Hence, any finite bound on the variance given by the CRF-inequality is automatically satisfied by them. However, the strong similarity of the physical models with the Gaussian model could extend them the suspects of possible inconsistencies. Thus, to give a strong support to our results we review the CRF inequality to show the inappropriate extension of the rule \(1/N\) to our approach. For this, we will apply the CRF developments to heteroscedastic models and to the least squares method for track fitting (for straight tracks and curved tracks) with a special attention to our Gaussian model. The last two points are very difficult to find in literature. Surely other authors did similar developments and we apologize to not citing them. In any case, the enormous statistical samples of tracks (of the order of \(10^{12}\) tracks per year) renders very important careful studies of any detail of their fitting methods. During our analysis of the track fitting we had to read carefully many papers (in any case a small portion of those published) on this subject and we found inaccurate fitting strategies. Probably the reason of this inaccuracy is connected to the weak growth as \(\sqrt{N}\) of the standard least-squares (the Kalman filter is essentially identical) that is negligibly modified by any improvement. Another element of inaccuracy is a frequent claim of a "measure" the hit resolutions. In reality, those parameters are not a "measure", but an application of an equation valid only for homoscedastic models, inappropriate for the physics of the trackers (surely always heteroscedastic). Instead, our fitting methods of ref. [1][2][3] open very different possibilities. The heteroscedastic models allow drastic improvements of the quality of the track fitting. In spite of the possible doubts, they are perfectly consistent with the CRF inequality. Instead, the CRF inequality demonstrates the inefficiency of the standard least squares for heteroscedastic models. The lucky model of ref. [1] is a first order tool to insert heteroscedasticity in track fitting and able to produce consistent improvements.

## 2 The standard CRF-inequality

The developments of this section follow as far as possible ref. [4] and ref. [6]. As usual it is considered a \(N\)-times repetition of random observations \(\{x_1,x_2,\ldots,x_N\}\) described by an identical probability density function (PDF) \(f(x,\theta)\), where \(\theta\) is a (scalar) parameter. These random variables \(x = \{x_1,x_2,\ldots,x_N\}\) are indicated as independent and identically distributed or i.i.d. random variables. We will use this notation in the following, without commenting the absolute physical inconsistency of this assumption.
The probability of the set \( x \) of i.i.d. random variables (likelihood function) is:

\[
L(x, \theta) = f(x_1, \theta)f(x_2, \theta) \ldots f(x_N, \theta).
\]  

(1)

As usual \( f(x, \theta) \) is positive, normalized and supposed differentiable with respect to the \( \theta \)-parameter. Let us define the random variable \( U(x, \theta) \):

\[
U(x, \theta) = \frac{\partial \ln L(x, \theta)}{\partial \theta} = \sum_{i=1}^{N} \frac{\partial \ln f(x_i, \theta)}{\partial \theta}.
\]  

(2)

The following developments require to differentiate the integrals of \( f(x, \theta) \) and invert the order of integration and differentiation. The functions that allow these operations are called regular models and they imply strict analytical conditions. A necessary condition of regular models is the independence of the range of \( f(x, \theta) \) from the \( \theta \)-parameter. The normalization of the \( f(x, \theta) \) is:

\[
\int L(x, \theta)dx = \int f(x_1, \theta)f(x_2, \theta) \ldots f(x_N, \theta)dx_1dx_2 \ldots dx_N = 1.
\]  

(3)

Differentiating eq. (3) with respect to \( \theta \) and inverting the integral with the differentiation gives:

\[
0 = \int \frac{\partial L(x, \theta)}{\partial \theta}dx = \sum_{i=1}^{N} \int \frac{\partial f(x_i, \theta)}{\partial \theta}dx_i = \int \frac{\partial \ln L(x, \theta)}{\partial \theta}L(x, \theta)dx = \sum_{i=1}^{N} \int \frac{\partial \ln f(x_i, \theta)}{\partial \theta}f(x_i, \theta)dx_i.
\]  

(4)

The second line of eq. (4) is given by a property of the logarithmic derivative and it implies that the mean value of \( U(x, \theta) \) is zero. Instead, its mean square is different from zero:

\[
\int U(x, \theta)^2dx = \sum_{i=1}^{N} \int \left( \frac{\partial \ln f(x_i, \theta)}{\partial \theta} \right)^2 f(x_i, \theta)dx_i = N \int \left( \frac{\partial \ln f(x, \theta)}{\partial \theta} \right)^2 f(x, \theta)dx.
\]  

(5)

The first term of eq. (5) is called Fisher information of the sample, the last integral is the Fisher information contained in a single observation.

For the i.i.d. random variables the Fisher information is N-times the Fisher information of a single observation.

This originates the \( 1/N \) factor in the CRF-bound. It will evident that for non i.i.d (heteroscedastic) random variables this factorization disappears. We underline the i.i.d. assumption in eq. (5) because often this condition is a standard assumption (as in ref. [4]), and it can only be recovered with an accurate re-reading of the inter demonstration and doing another demonstration without the i.i.d assumption.

If \( f(x, \theta) \) is twice differentiable with respect to \( \theta \), the last term of eq. (4) gives the following identity:

\[
0 = \int \left( \frac{\partial \ln f(x, \theta)}{\partial \theta} \right)^2 f(x, \theta)dx + \int \frac{\partial^2 \ln f(x, \theta)}{\partial \theta^2} f(x, \theta)dx.
\]  

(6)

The Fisher information \( i(\theta) \) for a single observation becomes:

\[
i(\theta) = \int \left( \frac{\partial \ln f(x, \theta)}{\partial \theta} \right)^2 f(x, \theta)dx = -\int \frac{\partial^2 \ln f(x, \theta)}{\partial \theta^2} f(x, \theta)dx.
\]  

(7)

The last form of \( i(\theta) \) is particularly effective for Gaussian PDF.
2.1 The CRF inequality

At this point is easy to obtain the CRF-inequality for these i.i.d. random variables. Defining an estimator \( T(X) = T(x_1, x_2, \ldots, x_n) \) and its mean value:

\[
\int T(x)L(x, \theta)dx = \tau(\theta),
\]

differentiating eq. 8 in \( \theta \) we obtain:

\[
\tau(\theta)' = \int T(x)\frac{\partial \ln L(x, \theta)}{\partial \theta}L(x, \theta)dx = \int [T(x) - \tau(\theta)]\frac{\partial \ln L(x, \theta)}{\partial \theta}L(x, \theta)dx = \int \hat{T}(x)\frac{\partial \ln L(x, \theta)}{\partial \theta}\frac{L(x, \theta)}{\sqrt{L(x, \theta)}}dx \quad \hat{T}(x) = T(x) - \tau(\theta).
\]

The form of the last integral is allowed by the positivity of \( L(x, \theta) \). The subtraction of \( \tau(\theta) \) from \( T(x) \) does not modify the integral for eq. 4. The Cauchy-Schwarz inequality gives:

\[
[\tau(\theta)']^2 \leq \int \hat{T}(x)^2 L(x, \theta)dx \int \left[ \frac{\partial \ln L(x, \theta)}{\partial \theta} \right]^2 L(x, \theta)dx
\]

or in its final form:

\[
\int \hat{T}(x)^2 L(x, \theta)dx \geq \frac{[\tau(\theta)']^2}{N i(\theta)}.
\]

The variance of \( T(x) \) can not be lower than the right side of eq. 11. This is the well-known CRF inequality that creates many doubts to our readers. When the CRF inequality becomes an identity, the estimator is defined efficient.

2.2 Efficient estimators

The Cauchy-Schwarz inequality becomes an equality if and only if \( \hat{T}(x) \) and \( U(x, \theta) \) are linearly dependent:

\[
T(x, \theta) - \tau(\theta) = a(\theta)U(x, \theta)
\]

Hence, if an efficient estimator exists, it is a function of the model. The logarithmic functions contained in \( U(x, \theta) \) allow to find efficient estimators only in exponential models.

2.3 The standard mean with an i.i.d. Gaussian model

The PDF \( f(x, \theta) \), the mean value \( \tau(\theta) \) and the Fisher information \( i(\theta) \) are:

\[
f(x, \theta) = \frac{\exp[-(x - \theta)^2/2\sigma^2]}{\sqrt{2\pi}\sigma} \quad \tau(\theta) = \int T(x)L(x, \theta)dx = \theta
\]

\[
i(\theta) = -\int \frac{\partial^2 \ln f(x, \theta)}{\partial \theta^2} f(x, \theta)dx = \frac{1}{\sigma^2}
\]

\[
U(x, \theta) = \frac{1}{\sigma^2} \sum_{j=1}^{N} x_j - \frac{N}{\sigma^2} \theta \quad \frac{1}{N i} U(x, \theta) = T^c(x) - \theta \quad T^c(x) = \frac{\sum_{j=1}^{N} x_j}{N}
\]

The last of eq. 13 gives the efficient estimator of \( \theta \) for this model as defined by eq. 12. Its unbiased form \( \hat{T}^c(x) = T^c(x) - \theta \), inserted in eq. 11 gives the identity \( \sigma^2/N = \sigma^2/N \).
3 CRF inequality for heteroscedastic models

It is well-known by long time that different observations have always non identical PDF. The universe would be very simple if the observations were i.i.d., in this case the precision of a measure could be improved at any level by the simple repetition and averaging the results. Gauss knew very well this problem and in his paper of 1823 extended his least squares method to handle observations of different PDF. For our misfortune the standard books of mathematical statistics do not cover this argument. But to complete our confutation of critics moved to ref. [1], we have to venture in this field. As in the previous section, we suppose a set of observations obtained by an array of different detectors as in a tracker. Even if all the detectors are built with identical technology, differences are always present. Each detector is anisotropic for its construction, optimized to the measure of positions. The signal spreads in a different recognizable way for each hit point. Moreover, each observation has a quantum mechanical interaction that renders unrepeatable the observations. However, a PDF for each observation can be defined (ref. [3]) even if very different from the form required for the CRF-inequality. However, the likelihood of a set of heteroscedastic observations becomes:  

\[ L_{1,2,\ldots,N}(x, \theta) = f_1(x_1, \theta) f_2(x_2, \theta) \ldots \ldots f_N(x_N, \theta) \]  

(14)

The evident difference is an index attached to each PDF. Instead of the awful PDFs of ref. [3][2], we will use an easy model. The simple model of ref. [1] is a heteroscedastic Gaussian model (independent non identically distributed Gaussian random variables), hence, it has all the analytical properties required for the CRF-inequality. Again \( U_{1,2,\ldots,N}(x, \theta) \) is defined as:  

\[ U_{1,2,\ldots,N}(x, \theta) = \frac{\partial \ln L_{1,2,\ldots,N}(x, \theta)}{\partial \theta} = \sum_{i=1}^{N} \frac{\partial \ln f_i(x_i, \theta)}{\partial \theta} \]  

(15)

Even the Fisher information acquires a set of indices,  

\[ \int U_{1,2,\ldots,N}(x, \theta)^2 L_{1,2,\ldots,N}(x, \theta) dx = \sum_{j=1}^{N} \int \left( \frac{\partial \ln f_j(x_j, \theta)}{\partial \theta} \right)^2 f_j(x_j, \theta) dx_j = \sum_{j=1}^{N} i_j(\theta) \]  

(16)

but now all the PDF of the observations are different as the Fisher information of each observation. The factorization of eq. 5 disappears with its N factor and its supposed generality. All the other equations are formally identical up to the first line of eq.10 ( a part an additional index to the PDFs ). With the new Fisher information the CRF-inequality for an unbiased estimator \( \tilde{T}_{1,2,\ldots,N}(x) \) becomes:  

\[ \int \tilde{T}_{1,2,\ldots,N}(x)^2 L_{1,2,\ldots,N}(x, \theta) dx \geq \frac{[\tau_{1,2,\ldots,N}(\theta)]^2}{\sum_{j=1}^{N} i_j(\theta)} \]  

(17)

Supposing a large set of possible PDFs ( as those of ref. [3] ) the estimator \( T_{1,2,\ldots,N}(x, \theta) \) must keep the track of the types and the order of the PDFs contained. It must be substituted with \( T_{j_1,j_2,\ldots,j_N}(x, \theta) \), similarly for the likelihood function \( L_{j_1,j_2,\ldots,j_N}(x, \theta) \) and the Fisher information. In this extended form the CRF- inequality becomes:  

\[ \int \tilde{T}_{j_1,j_2,\ldots,j_N}(x)^2 L_{j_1,j_2,\ldots,j_N}(x, \theta) dx \geq \frac{[\tau_{j_1,j_2,\ldots,j_N}(\theta)]^2}{i_{j_1,j_2,\ldots,j_N}(\theta)} \]  

(18)

The efficient estimators have the variance identical to the CRF-bound. Again the logarithmic functions of eq.15 limit the definition of efficient estimators to exponential PDFs.
3.1 The weighted mean in heteroscedastic Gaussian models

Let us consider the standard mean \( T^a_{1\ldots,N}(x) = \sum_{i=1}^{N} x_i/N \) as an estimator in heteroscedastic Gaussian models. The PDFs \( f_j(x, \theta) \), the mean value \( \tau_{1\ldots,N}(\theta) \) and the Fisher information due to the j-PDF \( i_j(\theta) \) are:

\[
f_j(x, \theta) = \frac{\exp[-(x-\theta)^2/2\sigma^2]}{\sqrt{2\pi}\sigma} \quad \tau_{1\ldots,N}(\theta) = \int T^a_{1\ldots,N}(x) L_{1\ldots,N}(x, \theta) dx = \theta \quad i_j(\theta) = - \int \frac{\partial^2 \ln f_j(x, \theta)}{\partial \theta^2} f_j(x, \theta) dx = \frac{1}{\sigma_j^2} \int \left[ \sum_{i=1}^{N} x_i - \theta \right]^2 L_{1\ldots,N}(x, \theta) dx = \frac{\sum_{j=1}^{N} \sigma_j^2}{N^2}
\]

With the variance of \( T^a \) the CRF-inequality is:

\[
\sum_{j=1}^{N} \frac{\sigma_j^2}{N^2} > \frac{1}{\sum_{j=1}^{N} \frac{1}{\sigma_j}}
\]

The equality is impossible for the assumption of the heteroscedasticity (the identity of the \( \sigma_j \) is excluded). Let us extract from eq. [15] an efficient estimator defined (if it exists) as usual:

\[
T_{1\ldots,N}(x) - \tau_{1\ldots,N}(\theta) = a(\theta)U_{1\ldots,N}(x, \theta)
\]

With \( a(\theta) = 1/\left[ \sum_{j=1}^{N} 1/\sigma_j^2 \right] \) and \( \tau_{1\ldots,N} = \theta \) it is:

\[
a(\theta) \sum_{i=1}^{N} \frac{\partial \ln f_j(x_i, \theta)}{\partial \theta} = \frac{1}{\sum_{j=1}^{N} 1/\sigma_j^2} \sum_{i=1}^{N} \frac{x_i - \theta}{\sigma_i^2} \quad \Rightarrow \quad T_{1\ldots,N}^r(x) = \sum_{i=1}^{N} \frac{x_i}{\sigma_i^2} \sum_{j=1}^{N} \frac{1}{\sigma_j^2} > N^2
\]

This estimator is proportional to \( \partial \ln L_{1\ldots,N}(x, \theta)/\partial \theta \) and transforms the Cauchy-Schwarz inequality in an identity. It is easy to verify, without the CRF inequality, that the variance of the standard mean is always greater than the variance of the weighted mean.

\[
\frac{\sum_{j=1}^{N} \sigma_j^2}{N^2} > \frac{1}{\sum_{j=1}^{N} \frac{1}{\sigma_j}} \quad \sum_{j=1}^{N} \sigma_j^2 \sum_{j=1}^{N} \frac{1}{\sigma_j} > N^2
\]

The second equation is the Cauchy-Schwarz inequality for the two vectors \( \{\sigma_j\} \) and \( \{1/\sigma_j\} \). The equality is impossible because \( \sigma^2 \) and \( 1/\sigma^2 \) can never be proportional excluding the trivial case of identical \( \sigma_j \). Heteroscedasticity excludes this case. Thus, in heteroscedastic models, the weighted average has always minor variance compared to the standard mean for regular and irregular models.

Hence, each combination of observations from our Gaussian model satisfies the CRF inequality. No \( 1/N \) factor is present in the equation for a different Fisher information of each single observation. The case of identity of all the observations is excluded \textit{a priori}. Moreover, the mean of the standard variances of a set of random sequences \( \{j_1, j_2, \ldots, j_N\} \), not all identical, is higher than the mean of the corresponding weighted averages. The absence of \( 1/N \) factors in the variance of the weighted average allows fast decreases with \( N \) that is impossible to the variance of the standard mean.

3.2 The weighted mean in a homoscedastic model

To prove another aspect of the CRF-inequality, we use the weighted mean outside its definition model (heteroscedastic) and we use it in a homoscedastic model. In this case the standard mean is the efficient estimator. The CRF-inequality states that the variance of the weighted mean is greater than the variance of the standard mean. It is easy to illustrate this fact (the reverse of the last equations).

\[
\int (T_{1\ldots,N}^r(x) - \theta)^2 L(x, \theta) dx = \sigma^2 \frac{\sum_j 1/\sigma_j^4}{(\sum_j 1/\sigma_j^2)^2} > \frac{\sigma^2}{N} \quad \Rightarrow \quad N \sum_j \frac{1}{\sigma_j} > (\sum_j \frac{1}{\sigma_j^2})^2
\]
The last inequality is the Cauchy-Schwarz inequality. The equality is excluded. Thus an efficient estimator in a model, outside its model has always a greater variance compared to the efficient estimator of the new model.

4 CRF-bound for heteroscedastic least squares of straight tracks

The application of the CRF-inequality to heteroscedastic least-squares fit for straight tracks adds other complications. Two parameters must be handled. The first parameter is the constant $\beta$: the impact point of the track in the reference plane. The second is $\gamma$: the angular shift of the track in the plane distant $y_i$ from the reference plane. Also here, the heteroscedasticity eliminates any easy rule with the increase of the number ($N$) of the observation planes for two different reasons. One is again the differences of the $\sigma_i$; the new one is due to the constraint of the fixed length of the tracker. Each insertion of a new detecting plane requires a repositioning of all the others. We will study our Gaussian model with the condition $\sum_{j=1}^N y_j = 0$, this introduces some simplification on standard least squares equations. The likelihood function is similar to that of eq. [14] but now, the order of the PDFs must be conserved. The parameter $y_j$ orders the measuring planes:

$$L_{1,2,...,N}(x, \beta, \gamma) = f_1(x_1, \beta, y_1 \gamma) f_2(x_2, \beta, y_2 \gamma) \ldots f_N(x_N, \beta, y_N \gamma)$$ (25)

$$f_j(x, \beta, y_j \gamma) = \frac{\exp[-(x - \beta - y_j \gamma)^2/2\sigma^2_j]}{\sqrt{2\pi}\sigma_j}$$

$$\gamma_{1,2,...,N}(\beta, \gamma) = \int T_{1,2,...,N}(x) L_{1,2,...,N}(x, \beta, \gamma) dx = \gamma$$

The function $U_{1,2,...,N}(x, \theta)$ is the 2x1 matrix $U_{1,2,...,N}(x, \beta, \gamma)$:

$$U_{1,2,...,N}(x, \beta, \gamma) = \left(\begin{array}{c} \frac{\partial \ln L_{1,2,...,N}(x, \beta, \gamma)}{\partial \beta} \\ \frac{\partial \ln L_{1,2,...,N}(x, \beta, \gamma)}{\partial \gamma} \end{array} \right)$$ (27)

The Fisher information $I_{1,2,...,N}$ becomes the 2x2 matrix:

$$I_{1,2,...,N} = \int dx L_{1,2,...,N}(x, \beta, \gamma) \left(\begin{array}{cc} \frac{\partial^2 \ln L_{1,2,...,N}(x, \beta, \gamma)}{\partial \beta^2} & \frac{\partial^2 \ln L_{1,2,...,N}(x, \beta, \gamma)}{\partial \beta \partial \gamma} \\ \frac{\partial^2 \ln L_{1,2,...,N}(x, \beta, \gamma)}{\partial \gamma \partial \beta} & \frac{\partial^2 \ln L_{1,2,...,N}(x, \beta, \gamma)}{\partial \gamma^2} \end{array} \right)$$ (28)

$$I_{1,2,...,N} = \left(\begin{array}{cc} \sum_{j=1}^N \frac{1}{\sigma^2_j} & \sum_{j=1}^N \frac{y_j}{\sigma^2_j} \\ \sum_{j=1}^N \frac{y_j}{\sigma^2_j} & \sum_{j=1}^N \frac{y_j^2}{\sigma^2_j} \end{array} \right)$$ (29)

Its inverse is:

$$I^{-1}_{1,2,...,N} = \frac{1}{\text{Det} \{I_{1,2,...,N}\}} \left(\begin{array}{cc} \sum_{j=1}^N \frac{y_j^2}{\sigma^2_j} - \sum_{j=1}^N \frac{y_j}{\sigma^2_j} \\ -\sum_{j=1}^N \frac{y_j}{\sigma^2_j} & \sum_{j=1}^N \frac{1}{\sigma^2_j} \end{array} \right)$$ (30)

The CRF inequality assumes the form of a matrix inequality, we limit to relations among variances:

$$\text{Diag.} \left(\begin{array}{cc} \text{Var}(\beta, \beta) & \text{Cor}(\beta, \gamma) \\ \text{Cor}(\gamma, \beta) & \text{Var}(\gamma, \gamma) \end{array} \right) \geq \text{Diag.} \cdot I^{-1}_{1,2,...,N}$$ (31)
where $\text{Var}(\beta, \beta)$ and $\text{Var}(\gamma, \gamma)$ are the variances of $\beta$ and $\gamma$:

$$
\text{Var}(\beta, \beta) = \int \left[ T^\beta_{i,1,2,..,N}(x) - \beta \right]^2 L_{1,2,..,N}(x, \beta, \gamma) \, dx \geq \frac{\sum_{i=1}^N y_i^2 / \sigma_i^2}{\sum_{i=1}^N \gamma_i^2 / \sigma_i^2} \frac{1 / \sigma_i^2}{(\sum_{i=1}^N y_i / \sigma_i)^2}
$$

(32)

and:

$$
\text{Var}(\gamma, \gamma) = \int \left[ T^\gamma_{i,1,2,..,N}(x) - \gamma \right]^2 L_{1,2,..,N}(x, \beta, \gamma) \, dx \geq \frac{\sum_{i=1}^N \gamma_i^2 / \sigma_i^2}{\sum_{i=1}^N y_i^2 / \sigma_i^2} \frac{1 / \sigma_i^2}{(\sum_{i=1}^N \gamma_i / \sigma_i)^2}
$$

(33)

The Gaussian model is efficient because its variances are equal to the diagonal of $\mathbf{I}^{-1}$ (here and in the following the indications 1, 2, $\ldots$, $N$ of the used PDFs will be subtended). The unbiased efficient estimators of $\beta$ and $\gamma$ for the model of eq. [25] are given by $\mathbf{I}^{-1} \mathbf{U}(\mathbf{x}, \beta, \gamma)$ (similar to those for a single parameter):

$$
\mathbf{U}(\mathbf{x}, \beta, \gamma) = \left( \begin{array}{c} T_1(\mathbf{x}) \\ T_2(\mathbf{x}) \end{array} \right) - \mathbf{I} \left( \begin{array}{c} \beta \\ \gamma \end{array} \right)
$$

$$
\mathbf{I}^{-1} \mathbf{U}(\mathbf{x}, \beta, \gamma) = \mathbf{I}^{-1} \left( \begin{array}{c} T_1(\mathbf{x}) \\ T_2(\mathbf{x}) \end{array} \right) - \left( \begin{array}{c} \beta \\ \gamma \end{array} \right)
$$

(34)

The last equation allows the extraction of the formal expressions of the unbiased efficient estimators for this model (and are identical to those reported everywhere for example in ref. [5]). The mean values of products of logarithmic derivatives are obtained by the mean values of double derivative of a single logarithmic function (as in eq. [6]), and due to eq. [28] gives:

$$
\int \mathbf{U}(\mathbf{x}, \beta, \gamma) \cdot \mathbf{U}'(\mathbf{x}, \beta, \gamma) L(\mathbf{x}, \beta, \gamma) \, dx = \mathbf{I}
$$

$$
\int \mathbf{I}^{-1} \mathbf{U}(\mathbf{x}, \beta, \gamma) \cdot \mathbf{U}'(\mathbf{x}, \beta, \gamma) \mathbf{I}^{-1'} L(\mathbf{x}, \beta, \gamma) \, dx = \mathbf{I}^{-1}
$$

(35)

The second equation uses the first equation giving the Fisher information and the symmetry $\mathbf{I}^{-1} = \mathbf{I}^{-1'}$. The diagonal terms are the variances of the efficient estimators for $\beta$ and $\gamma$. It is evident the privileged position of the Gaussian PDFs in the CRF-inequality.

### 4.1 The standard least-squares equations with heteroscedastic likelihood

We defined standard least squares equations the estimators (for $\beta$ and $\gamma$) obtained with a homoscedastic model. It is easy to prove that the standard least squares estimators are efficient estimators for a Gaussian model with identical $\sigma$s. The extension of eqs. [12] to the case for two parameters shows immediately this property. The condition $\sum_{i=1}^N y_i = 0$ simplifies the forms of the estimators. In the case of identical $\sigma$s the CRF-inequality states that any other estimator has a greater variance compared to the efficient estimators. It is evident that the efficiency of an estimator is strictly bound to the likelihood model of its definition (in a given model the efficient estimator of a parameter is unique). Outside this model the estimators are no more efficient. We could derive the forms of the standard least squares estimators as the efficient estimators for the Gaussian model with identical $\sigma_j$, but the usual derivation is faster. The estimators of $\beta$ and $\gamma$ of the standard least squares are:

$$
\sum_{i=1}^N (x_i - \beta - \gamma y_i) = 0 \quad \Rightarrow \quad \widehat{T}_\beta(x, \beta, \gamma) = \frac{\sum_{i=1}^N x_i}{N} - \beta
$$

$$
\sum_{i=1}^N (x_i - \beta - \gamma y_i) y_i = 0 \quad \Rightarrow \quad \widehat{T}_\gamma(x, \beta, \gamma) = \frac{\sum_{i=1}^N x_i y_i}{\sum_{i=1}^N y_i^2} - \gamma
$$

(36)

For an easy check of the forms of the estimators, we report the least squares equations that give their expressions (after inserting $\sum y_i = 0$). The heteroscedastic likelihood $L(x, \beta, \gamma)$ destroys their efficiency.
(optimality) and the new variances are different from those in the homoscedastic model ($\sigma^2/N$ and $\sigma^2/\sum_j y_j^2$):

$$
\int T_i^\beta(x, \beta, \gamma)^2L(x, \beta, \gamma)dx = \frac{\sum_{i=1}^N \sigma_{x_i}^2}{N^2} \\
\int T_i^\gamma(x, \beta, \gamma)^2L(x, \beta, \gamma)dx = \frac{\sum_{i=1}^N \sigma_{x_i}^2}{(\sum_{i=1}^N y_i^2)^2},
$$

(37)

and, due to the differences from the variances of the efficient estimators, the CRF inequalities of eq. [32] and eq. [33] impose to them to be greater. We will extensively prove this property as a prototype of demonstration [7] that will be used even for the momentum variances.

First of all, it is required the joint variance-covariance matrix of $\hat{T}_i^\beta \gamma$ with the efficient estimators $I^{-1}U(x, \beta, \gamma)$. This matrix is positive semi-definite (symmetrical with the variances in the diagonal). Let us define with $C$ this matrix, with $T$ the column vector $T = \{T_i^\beta, T_i^\gamma\}$ and with $< >$ the average on the likelihood of eq. [25] The matrix $I^{-1}$ is the variance-covariance matrix of the efficient estimators (eq. [35])

$$
C = \begin{pmatrix}
<T T' > \\
<1^{-1} U T' > \\
< T U I^{-1} > \\
I^{-1}
\end{pmatrix} \geq 0
$$

(38)

It is easy to show this general property of the covariance of the unbiased estimator $\hat{T}_i^\beta$ with the efficient estimators. By definition, the mean value of an unbiased estimator is always zero as any of its derivative:

$$
\frac{\partial}{\partial \beta} \int (T_i^\beta(x) - \beta)L(x, \beta, \gamma)dx = 0 = -1 + \int (T_i^\beta(x) - \beta) \frac{\partial}{\partial \beta} L(x, \beta, \gamma)dx.
$$

(39)

The derivative of the likelihood is the first component of $U(x, \beta, \gamma)$ of eq. [27] and this correlation becomes equal to one. The derivative in $\gamma$ of the mean value of $\hat{T}_i^\beta$ has no one and the sought correlation is zero. Similarly for the other estimator $\hat{T}_i^\gamma$. Hence the matrix $< T U' >$ is:

$$
<T U' > = \mathbb{I}_2 = \begin{pmatrix}
1 & 0 \\
0 & 1
\end{pmatrix} = <U T' > \Rightarrow C = \begin{pmatrix}
<T T' > \\
<1^{-1} > \\
I^{-1} \\
1^{-1}
\end{pmatrix} \geq 0
$$

(40)

The $I^{-1}$ is a symmetric matrix. The positivity of $C$ is conserved even with transformation $A'CA$ with $A$ any real matrix of four lines and two columns (few details about the positive definite matrices are reported in the appendix):

$$
(\mathbb{I}_2 - \mathbb{I}_2)C = \begin{pmatrix}
\mathbb{I}_2 \\
-\mathbb{I}_2
\end{pmatrix} = < T T' > - I^{-1} \geq 0 \Rightarrow < T T' > \geq I^{-1}
$$

(41)

The last of the eq. [41] is eq. [33] as a matrix relation the equality is eliminated being evidently impossible. For the definition of positive definite matrices it is straightforward to extract the relations among the variances. The eqs. [31] to [41] will be extended the estimators of the momentum reconstructions.

Even if it was demonstrated with a heteroscedastic Gaussian model, eqs. [41] remain always greater also for heteroscedastic irregular models with equivalent set of $\{\sigma_j\}$. The efficient estimators of heteroscedastic models continue to be efficient even for homoscedastic models, converging to them when all the $\sigma_j$ become identical.

All this set of demonstrations for heteroscedastic Gaussian model is addressed to each given chain of PDFs with a defined order of $\sigma_j$. Their variances of eq. [32] and eq. [43] are strongly modified by the differences among the $\{\sigma_j\}$ without any evident $N$ limitation. Instead, the variances of the standard least squares are weakly modified and they save explicit $N$ limitations. For example the variances of the $\beta$ parameter has averages $\approx \sigma^2_{\beta}/N$. Figure [1] illustrates this limits for the $\gamma$ estimator (the blue distributions). Due to the heavy modifications of the variances of the chains of PDFs, other PDFs become essential. For
Figure 2: The Gaussian model. A faster increase in the maximums compared to that of fig.1. The blue straight lines connect the maximums of the first distributions to the last red distribution. The color code is that of fig.1.

example those that select the chains among the possible set of chains. These PDFs are introduced by the properties of the measuring devices. As we said above, a binomial PDF suffices in the disordered tracker detector for the models of refs. [1, 2], just a mean disorder of half strip ($\approx 30 \mu m$) is enough for this. The use of standard least squares method in heteroscedastic models has enormous simplifications, because it does not need the knowledge of the $\sigma_j$ for each observation. Unfortunately, it amounts to a large loss of resolution. Our models of ref. [1] illustrate in evident way the amplitude of this loss. In any case the lucky model of ref. [1] can be an economic way to introduce effective $\sigma_j$.

For our pleasure, in fig. 2 we invented a faster growth than that of the Gaussian model of ref. [1]. The quality of the detectors increases with the number $N$ of detecting layers, each inserted layer is better than the previous one. The bleu straight line shows an evident deviation from a linear growth with an almost parabolic growth. The blue distributions are practically insensible to these changes. With cleaver selection of the hit quality many types of growth can be implemented.

4.2 Deviations from optimality

Equations [1] proved the non-optimality of the standard least-squares in the case of it use with a different (heteroscedastic) likelihood. The standard least squares gives the efficient estimators for a homoscedastic Gaussian model, the transportation outside its condition of efficiency (optimality) destroys its optimality and produces an increase of variances with a degradation of the resolution as illustrated by the simulations of ref. [1, 2, 3]. Even our heteroscedastic model suffers of a similar non-optimality loss outside its range of validity. We underlined that the orders and the values of the sets of $\{\sigma_j\}$ define the model, each deviation from the order and values implies an increase of the variances compared to the optimality condition. We used this property to prove the essential connection of the appropriate $\sigma_j$ with the hit $j$. If the ordered set of $\{\sigma_j\}$ is randomized, destroying the correlation with the set of the hits, the distributions
of the estimators change drastically becoming worse than the standard least squares. Instead, the differences of the heteroscedastic efficient estimator minus a suboptimal estimator with few eliminated PDFs rapidly converge as the eliminated PDFs are reduced. This rapid convergence is not observed in the case of the standard least-squares.

5 Inequality for Momentum estimators

As stated above the CRF-inequality can be extended to case of the momentum reconstruction. It will be studied the case of a high-momentum charged-particle moving in a tracker with a homogeneous magnetic field and the path of the particle is orthogonal to the field direction. Here, the circular path can be studied the case of a high-momentum charged-particle moving in a tracker with a homogeneous magnetic field and the path of the particle is orthogonal to the field direction. Here, the circular path can

5.1 The momentum estimators in homoscedastic Gaussian model

We use this approach to define our expressions for the estimators. These forms will be used in heteroscedastic models. Our equation will be expressed in compact way with matrices. The function $\mathcal{L}(x, \beta, \gamma, \eta)$ is the likelihood for this case, $\beta$ and $\gamma$ are just defined, $\eta$ is the curvature of the track, proportional to $1/p$ with $p$ the track momentum:

$$\mathcal{L}(x, \beta, \gamma, \eta) = f_1(x_1, \beta, \gamma, \eta) f_2(x_2, \beta, \gamma, \eta) \cdots f_N(x_N, \beta, \gamma, \eta)$$

$$f_j(x, \beta, \gamma, \eta) = \frac{\exp[-(x - \beta - \gamma y_j - \eta y_j^2)/2\sigma^2]}{\sqrt{2\pi}\sigma}$$

The index $j$ orders the PDFs $f_j$ as the tracker layers. The equivalent of eq. [27]is:

$$V(x, \beta, \gamma, \eta) = \begin{pmatrix} \frac{\partial \ln \mathcal{L}(x, \beta, \gamma, \eta)}{\partial \beta} \\ \frac{\partial \ln \mathcal{L}(x, \beta, \gamma, \eta)}{\partial \gamma} \\ \frac{\partial \ln \mathcal{L}(x, \beta, \gamma, \eta)}{\partial \eta} \end{pmatrix}$$

The unbiased estimators for $\beta, \gamma, \eta$ are:

$$V(x, \beta, \gamma, \eta) = \begin{pmatrix} T_\beta(x) \\ T_\gamma(x) \\ T_\eta(x) \end{pmatrix} - R \begin{pmatrix} \beta \\ \gamma \\ \eta \end{pmatrix}$$

$$\mathbf{R}^{-1}V(x, \beta, \gamma, \eta) = \mathbf{R}^{-1} \begin{pmatrix} T_\beta(x) \\ T_\gamma(x) \\ T_\eta(x) \end{pmatrix} - \begin{pmatrix} \beta \\ \gamma \\ \eta \end{pmatrix}$$

The vector $\mathbf{R}^{-1}V(x, \beta, \gamma, \eta)$ contains the efficient estimators of this model, the matrix $\mathbf{R}$ is its the Fisher information (with the condition $\sum_j y_j = 0$ and $j = (1,2,\ldots,N)$):

$$\mathbf{R} = \begin{pmatrix} N & 0 & \sum_j y_j^2 \\ 0 & \sum_j y_j^2 & \sum_j y_j^3 \\ \sum_j y_j^2 & \sum_j y_j^3 & \sum_j y_j^4 \end{pmatrix} \frac{1}{\sigma^2}$$

For eq. [35] extended to this case, it is:

$$\int V(x, \beta, \gamma, \eta) \cdot V'(x, \beta, \gamma, \eta) \mathcal{L}(x, \beta, \gamma, \eta) \, dx = \mathbf{R}$$

$$\int \mathbf{R}^{-1}V(x, \beta, \gamma, \eta) \cdot V'(x, \beta, \gamma, \eta)\mathbf{R}^{-1} \mathcal{L}(x, \beta, \gamma, \eta) \, dx = \mathbf{R}^{-1}.$$
The matrices \( \mathbf{R} \) and \( \mathbf{R}^{-1} \) are symmetric and positive semi-defined, and the second line of the last equation is the variance-covariance matrix of the efficient estimators. The variance-covariance matrix of a generic vector \( T \) of unbiased estimators of \( \beta, \gamma, \eta \) with the efficient estimators is:

\[
\mathbf{K} = \begin{pmatrix}
<T T' > & < T V' \mathbf{R}^{-1} > \\
< \mathbf{R}^{-1} V T' > & \mathbf{R}^{-1}
\end{pmatrix} \geq 0
\] (47)

As defined, the symbol \( <> \) means an integral with the likelihood \( \mathcal{L} \), the cross correlations \( < V T' > \) follow the rules of eq. [50] giving now the \( 3x3 \) identity matrix \( \mathbb{I}_3 \) reducing the off-diagonal elements of \( \mathbf{K} \) to \( \mathbf{R}^{-1} \). With a transformation of the type \( \mathbf{A}' \mathbf{K} \mathbf{A} \) with a \( 6x3 \) real matrix \( \mathbf{A} \) the fundamental inequality is obtained:

\[
(\mathbb{I}_3 - \mathbb{I}_3) \begin{pmatrix}
<T T' > & \mathbf{R}^{-1} \\
\mathbf{R}^{-1} & \mathbf{R}^{-1}
\end{pmatrix} \begin{pmatrix}
\mathbb{I}_3 \\
\mathbb{I}_3 - \mathbb{I}_3
\end{pmatrix} = < TT' > - \mathbf{R}^{-1} \geq 0
\] (48)

The equality is obtained only for the efficient estimators of this model. Therefore, any other set of estimators have a greater variance in this model. The equality is obtained only for the efficient estimators of this model. Any heteroscedastic model has greater variance than the efficient model. The reason is connected to the modification that the likelihood \( \mathcal{L} \) introduces in the variances. These modifications always increase the variances. In this model, the parameter \( \sigma \) disappears in the expressions for the estimators. The form \( \mathbf{R} \) of the Fisher information depends from \( N \), this limits the growth with the numbers of detecting layers as illustrated in ref. [2].

### 5.2 The momentum estimators in heteroscedastic Gaussian model

In its essence this model can be obtained from the corresponding homoscedastic model with a set of formal differences.

\[
L(x, \beta, \gamma, \eta) = f_1(x_1, \beta, \gamma, \eta)f_2(x_2, \beta, \gamma, \eta) \cdots f_N(x_N, \beta, \gamma, \eta)
\]

\[
f_j(x, \beta, \gamma, \eta) = \frac{\exp[-(x - \beta - \gamma y_j - \eta y_j^2)/2\sigma_j^2]}{\sqrt{2\pi}\sigma_j}
\] (49)

The index \( j \) orders the PDFs as the detecting layers and selects a different \( \sigma_j \) for this observation. The other differences are in the definitions of the vector \( T_1, T_2, T_3 \) that are different from the previous \( T_a, T_b, T_c \) and the Fisher information \( \mathbf{I} \).

\[
U(x, \beta, \gamma, \eta) = \begin{pmatrix}
T_1(x) \\
T_2(x) \\
T_3(x)
\end{pmatrix} - \mathbf{I} \begin{pmatrix}
\beta \\
\gamma \\
\eta
\end{pmatrix} = \begin{pmatrix}
\frac{\Sigma_j y_j/\sigma_j^2}{} \\
\frac{\Sigma_j y_j^2/\sigma_j^2}{}
\end{pmatrix}
\] (50)

The Fisher information becomes:

\[
\mathbf{I} = \begin{pmatrix}
\Sigma_j 1/\sigma_j^2 & \Sigma_j y_j/\sigma_j^2 & \Sigma_j y_j^2/\sigma_j^2 \\
\Sigma_j y_j/\sigma_j^2 & \Sigma_j y_j^2/\sigma_j^2 & \Sigma_j y_j^3/\sigma_j^2
\end{pmatrix}
\] (51)

The efficient unbiased estimators for the likelihood \( L(x, \beta, \gamma, \eta) \) are contained in the vector \( \mathbf{I}^{-1}U(x, \beta, \gamma, \eta) \). Even in this case the Fisher information has no obvious \( N \)-dependence. The variance-covariance matrix for any vector \( T \) of unbiased estimators can be transformed as previously described giving the inequality:

\[
(\mathbb{I}_3 - \mathbb{I}_3) \begin{pmatrix}
<T T' > & \mathbf{I}^{-1} \\
\mathbf{I}^{-1} & \mathbf{I}^{-1}
\end{pmatrix} \begin{pmatrix}
\mathbb{I}_3 \\
\mathbb{I}_3 - \mathbb{I}_3
\end{pmatrix} = < TT' > - \mathbf{I}^{-1} \geq 0
\] (52)

Here, the symbol \( <> \) indicates an integral with the likelihood of this model \( L(x, \beta, \gamma, \eta) \). The equality to zero is obtained only when the vector \( T \) is \( \mathbf{I}^{-1}U(x, \beta, \gamma, \eta) \). For any other unbiased estimator, that difference is greater than zero (in the sense of the positive definite matrices).
6 Conclusions

The Cramer-Rao-Frechet inequality is applied to a set of heteroscedastic Gaussian model to prove the absence of any limitation introduced by this fundamental inequality. Instead, the inequality shows that, in heteroscedastic models, the estimators of the standard least-squares have always greater variances compared to that given by the weighted least-squares. The variances of the standard least-squares conserve approximate $1/N$ form that imply limitations in the growth of the distributions with the number $N$ of detecting layers. No similar limitations are present in the weighted least squares that are dominated by the differences of the weights $1/\sigma_j^2$. For identical $\sigma_j$ the estimators converge to those of the standard least-squares. The inequality is calculated also for the momentum estimators. Again the appropriate heteroscedastic estimators have a lower variances than those of the standard least squares. Given the loss of resolution implied by greater variances it must be avoid the use of standard least squares with heteroscedastic data. The lucky model may be helpful in silicon detectors.

7 Appendix

Let us add few details about variance-covariance matrix or about the cross-covariance matrix. For their definitions these matrices are positive semi-defined (they are symmetric with the positive variances in the principal diagonal). If the matrix $C$ is a positive semi-definite matrix it must be a symmetric $m \times m$ matrix $C' = C$, and for any real vector $z$ it must be $z'Cz \geq 0$. From these definitions it follows that if $A$ is a real matrix of $m \times n$ with $n < m$ the matrix $K$ given by $A'CA$ is again a $n \times n$ positive semi-definite matrix. It is easy to show that the diagonal elements of $K$ are given by $z'Bz$ where each $z$ is a column of $A$ and the off-diagonal terms are symmetric. A positive semi-definite matrix implies an infinite number of inequalities ($\infty^m$), our interest is to find the minimal one with useful dimensions. The eqs. 41 48 52 are addressed to this. It is easy to show that the identity matrices $I_2$ and $I_3$ select the useful blocks of the matrices. As an example we can elaborate the $C$ matrix ($4 \times 4$). In general one can use a $4 \times 2$ matrix of the form $(I_2, aI_2)$ with $a$ any real constant and search the minimum in $a$:

$$
\begin{align*}
(\begin{array}{l}
I_2 \\
aI_2
\end{array}) C
(\begin{array}{l}
I_2 \\
aI_2
\end{array}) &= <TT'> + (2a + a^2)I^{-1} \geq 0 \\
\frac{\partial}{\partial a} \left(<TT'> + (2a + a^2)I^{-1}\right) &= (2 + 2a) = 0 \Rightarrow B = <TT'> - I^{-1} \geq 0
\end{align*}
$$

(53)

The solution $a = -1$ gives the last inequality. The inequality for each couple of variances is obtained with $z'Bz$ where $z$ is $(1,0)$ for the first two and $(0,1)$ for the second two.

References

[1] Landi G.; Landi G. E. Beyond the $\sqrt{N}$-limit of the least squares resolution and the lucky-model arXiv:1808.06708[physics.ins-det] https://arxiv.org/abs/1808.06708.

[2] Landi, G.; Landi G. E. Optimizing momentum resolution with a new fitting method for silicon-strip detectors INSTRUMENTS 2018, 2, 22 arXiv:1806.07874[physics.ins-det] https://arxiv.org/abs/1806.07874

[3] Landi G.; Landi G. E. Improvement of track reconstruction with well tuned probability distributions JINST 9 2014 P10006. arXiv:1404.1968[physics.ins-det] https://arxiv.org/abs/1404.1968

[4] Ivchenko G. and Medvedev Yu. MATHEMATICAL STATISTICS-Moscow-URSS-1990

13
[5] Olive, K.A.; Agashe, K.; Amsler, C.; Antonelli, M.; Arguin, J.-F.; Asner, D.M.; Baer, H.; Band, H.R.; Barnett, R.M.; Basaglia, T.; et al. Particle Data Group. Chin. Phys. C 2014, 38, 090001.

[6] Gowan G. STATISTICAL DATA ANALYSIS Oxford University Press, Oxford, 1998

[7] Takeshi, Amemiya ADVANCED ECONOMETRICS Harvard University Press, Cambridge Massachusetts 1985