The method of the likelihood and the Fisher information in the construction of physical models

E.W. Piotrowski¹, J. Sładkowski², J. Syska², S. Zając²

¹ Institute of Mathematics, The University of Białystok, Pl-15424 Białystok, Poland
² Institute of Physics, University of Silesia, Uniwersytecka 4, 40-007 Katowice, Poland

Received XXXX, revised XXXX, accepted XXXX
Published online XXXX

PACS 03.65.Bz, 03.65.Ca, 03.65.Sq, 02.50.Rj

Corresponding author; e-mail jacek.syska@us.edu.pl, Phone +xx-xx-xxxxxxx, Fax +xx-xx-xxx

The subjects of the paper are the likelihood method (LM) and the expected Fisher information (FI) considered from the point of view of the construction of the physical models which originate in the statistical description of phenomena. The master equation case and structural information principle are derived. Then, the phenomenological description of the information transfer is presented. The extreme physical information (EPI) method is reviewed. As if marginal, the statistical interpretation of the amplitude of the system is given. The formalism developed in this paper would be also applied in quantum information processing and quantum game theory.

1 Introduction. The subjects of the paper are the likelihood method (LM) and the expected Fisher information (FI) considered from the point of view of the construction of the physical models which originate in the statistical description of phenomena. The master equation case and structural information principle are derived. Then, the phenomenological description of the information transfer is presented. The extreme physical information (EPI) method is reviewed. As if marginal, the statistical interpretation of the amplitude of the system is given. The formalism developed in this paper would be also applied in quantum information processing and quantum game theory.

The subjects of the paper are the likelihood method (LM) and the expected Fisher information (FI) considered from the point of view of the construction of the physical models which originate in the statistical description of phenomena. The master equation case and structural information principle are derived. Then, the phenomenological description of the information transfer is presented. The extreme physical information (EPI) method is reviewed. As if marginal, the statistical interpretation of the amplitude of the system is given. The formalism developed in this paper would be also applied in quantum information processing and quantum game theory.
tion (EPI). Frieden began the construction of the physical models by obtaining the kinetic term from the \( F \). Then he postulated the variational (scalar) information principle (IP) and internal (structural) one. Each of the IP has the universal form, yet its realization depends on the particular phenomenon. The variational IP leads to the dispersion of all data gives the structural IP from the analysis of the Taylor expansion of the information (entropy) transfer in the process of measurement when this transfer from the structural to the kinematical degrees of freedom takes place. We will discuss it in Section 3. Despite the differences we still call this approach the Frieden one and the method the EPI estimation.

Finally, with the interpretation of \( I \) as the kinematical term of the theory, the statistical proof on the impossibility of the derivative of the wave mechanics and field theories for which the rank \( N \) of the field is finite from the classical mechanics, was given \( \text{[7]} \).

2 The master equations vs structural information principle The maximum likelihood method (MLM) is based on the set of \( N \) likelihood equations \( \text{[1]} \):

\[
S(\Theta) \bigg|_{\Theta=\hat{\Theta}} = \frac{\partial}{\partial \Theta} \ln P(\Theta) \bigg|_{\Theta=\hat{\Theta}} = 0 ,
\]

where the MLM’s set of \( N \) estimators \( \hat{\Theta} \equiv (\hat{\theta}_n)_N \) is its solution. These \( N \) conditions for the estimates maximize the likelihood of the sample.

2.1 The master equations. Yet, we could approach to the estimation differently. So, after the Taylor expanding of \( P(\Theta) \) around the true value of \( \Theta \) and integrating over the

---

2 Let for a system the square distance between two of its states denoted by \( q(x) \) and \( q'(x) = (q + dq)(x) \) could be written in the Euclidean form \( ds^2 = \int_X dx dq dq' \) where \( X \) is the space of the positions \( x \) (in one measurement). Supposing that the states are described by the probability distributions we compare it with the distance \( ds^2 = \frac{1}{2} \int_X dx \frac{\partial^2}{\partial q \partial q} \) in the space \( S \) with the Fisher-Rao metric \( [6] \), where \( p(x) \) and \( p' = (p + dp)(x) \) are the probability distributions for these two states. Then we notice that these two formulas for \( ds^2 \) coincide if \( q(x) = \sqrt{p(x)} \). In this way the notion of the \( \text{amplitude} \) \( q \) appears naturally from the Riemannian, Fisher-Rao metric of the space \( S \). It satisfies the condition: \( J_X dx \frac{p(x)}{dx} q(x)|_{x=\hat{x}} = 1 \).

Now, let the sample of dimension \( N \) be "collected" by the system and \( q_n \) are the real amplitudes and \( p_{x_n}(x_n) \) are the probability distributions with the property of the shift invariance \( p_{x_n}(x_n) = p_{x_n}(x_n|\theta_n) = p_{y_n}(y_n|\theta_n) \) where \( x_n \equiv y_n - \theta_n \). Assuming that the data are collected independently which gives the factorization property \( P(y) \equiv P(y|\Theta) = \prod_{n=1}^N p_{y_n}(y_n|\theta_n), \) where \( \theta_n \) has no influence on \( y_n \) for \( m \neq n \), and using the chain rule \( \partial/\partial \Theta_n = \partial/\partial(y_n-\theta_n) \partial(y_n-\theta_n)/\partial \theta_n = -\partial/\partial \theta_n \), the transition from the statistical form of \( FI \) given by Eq. \( \text{[1]} \) to its kinematical form given below by the first possibility:

\[
I = 4N \int dx_n \left( \frac{\partial q_n}{\partial x_n} \right)^2 \quad \text{or} \quad 4N \int dx \frac{\partial \psi_n^*}{\partial x} \left( \frac{\partial \psi_n}{\partial x} \right) \quad \text{(2)}
\]

is performed \( \text{[8]} \). The second possibility is achieved \( \text{[9]} \) as follows: Using the amplitudes \( q_n \), the wave function \( \psi_n \) could be constructed as \( \psi_n = \frac{1}{\sqrt{2}}(q_{n-1} + i q_n) \), where \( n = 1, 2, ..., N/2 \) with the number of the real degrees of freedom being twice the complex ones. With this, the first kinematical form of \( I \) in Eq. \( \text{[2]} \) could be rewritten in the second form, where additionally the index \( n \) has been dropped from the integral as the range of all \( x_n \) is the same. Finally, the total probability law for all data gives \( p(x) = N x \prod_{n=1}^N p_{x_n}(x_n|\theta_n)P(\theta_n) = \frac{1}{N} \sum_{n=1}^N q_n^* \psi_n \), where we have chosen \( \prod_{n=1}^N P(\theta_n) = \frac{1}{N} \text{[3]} \). All of these lead to \( p(x) = \sum_{n=1}^{N/2} q_n^* \psi_n \). With the shift invariance condition and the factorization property, \( I \) does not depend on the parameter set \( (\theta_n)_N \), and the wave functions \( \psi_n \) do not depend on these parameters, i.e. (expected) positions, also \( \text{[3]} \). From this summary of the Frieden method we notice that although the wave functions \( \psi_n \) are complex the quantum mechanics is really the statistical method.

---

3 In reality the Fourier transformation forms the type of the entanglement between the variables of the system.
sample space we obtain:
\[
\int dN \mathbf{x} \left( P(\hat{\Theta}) - P(\Theta) \right) = \int dN \mathbf{x} \left( \sum_{n=1}^{N} \frac{\partial P(\Theta)}{\partial \theta_n} (\hat{\theta}_n - \theta_n) \right) + 2 \sum_{n, n'} \frac{\partial^2 P(\Theta)}{\partial \theta_n \partial \theta_{n'}} (\hat{\theta}_n - \theta_n)(\hat{\theta}_{n'} - \theta_{n'}) + \cdots
\]
(4)

where the notation \( P(\Theta) \equiv P(\hat{\Theta}) \) has been used. When the integration over the whole sample space is performed then neglecting the higher order terms and using the normalization condition \( \int dN \times P(\Theta) = \int dN \times P(\hat{\Theta}) = 1 \) we see that the LHS of Eq. (4) is equal to zero. Hence we obtain the result which for the locally unbiased estimators and postulating the zeroing of the integrand at the RHS, takes for particular \( n \) and \( n' \) the following microscopic form of the master equation:

\[
\frac{\partial^2 P(\Theta)}{\partial \theta_n \partial \theta_{n'}} (\hat{\theta}_n - \theta_n)(\hat{\theta}_{n'} - \theta_{n'}) = 0 .
\]
(5)

When the parameter \( \theta_n' \) has the Minkowskian index \( \nu \) then \( P = \prod_{n=1}^{N} p_n(x_n^\nu) \) and Eq. (5) leads, after the transition to the Fisherian variables (as in Footnote 2), to the form of the equation of conservation of flow:

\[
\frac{\partial p_n(x_n^\nu)}{\partial t_n} + \sum_{i=1}^{3} \frac{\partial p_n(x_n^\nu)}{\partial x_n^{\mu}} \dot{v}_n^{\mu} = 0 , \quad \dot{v}_n^{\mu} = \frac{\theta_n^{\nu} - \theta_n}{\theta_n^{\nu} - \theta_n},
\]
(6)

where \( t_n \equiv x_n^0. \) Here \( \theta_n^{\nu} \) and \( \theta_n^{\nu} \) are the expected position and time of the system, respectively and index \( n \) could be omitted.

2.2 The information principle. The EPI method.

Using \( \ln P \) instead of \( P \) and after the Taylor expanding of \( \ln P(\hat{\Theta}) \) around the true value of \( \Theta \) and integrating with the \( dN \times P(\Theta) \) measure over the sample space we obtain (instead of Eq. (4)) the equation of the EPI method of the model estimation:

\[
\int dN \mathbf{x} P(\Theta)(\ln P(\hat{\Theta})/P(\Theta)) - R_3 - \sum_{n=1}^{N} \frac{\partial \ln P(\Theta)}{\partial \theta_n} (\hat{\theta}_n - \theta_n))
\]

\[
= \frac{1}{2} \int dN \times P(\Theta) \sum_{n, n'=1}^{N} \frac{\partial^2 \ln P(\Theta)}{\partial \theta_n \partial \theta_{n'}} (\hat{\theta}_n - \theta_n)(\hat{\theta}_{n'} - \theta_{n'})
\]
(7)

with \( P(\Theta) \equiv P(\hat{\Theta}) \) has the form of the modified relative entropy. Let us define the (observed) structure \( \pi_E \) of the system as follows:

\[
\pi_E \equiv \ln \frac{P(\hat{\Theta})}{P(\Theta)} - R_3 .
\]
(8)

When we define \( \tilde{Q} \) as

\[
\tilde{Q} = \int dN \times P(\Theta) \left( \frac{\partial}{\partial \theta_n} - \sum_{n=1}^{N} \frac{\partial \ln P(\Theta)}{\partial \theta_n} (\hat{\theta}_n - \theta_n) \right)
\]
(9)

then we obtain the structural equation of the IP form:

\[
- \tilde{Q} = \tilde{T}
\]
(10)

\[
\equiv \frac{1}{2} \int dN \times P(\Theta) \sum_{n, n'=1}^{N} \frac{\partial^2 \ln P(\Theta)}{\partial \theta_n \partial \theta_{n'}} (\hat{\theta}_n - \theta_n)(\hat{\theta}_{n'} - \theta_{n'}).
\]

Now, let us postulate the validity of Eq. (10) on the microscopic level:

\[
\Delta_{LHS} \equiv 2 \sum_{n=1}^{N} \frac{\partial \ln P(\Theta)}{\partial \theta_n} (\hat{\theta}_n - \theta_n) - 2 \frac{\pi_E}{N}
\]

\[
= \sum_{n, n'=1}^{N} \pi_{E_{n n'}} (\hat{\theta}_n - \theta_n)(\hat{\theta}_{n'} - \theta_{n'}) \equiv \Delta_{RHS} .
\]

(11)

As the inverse of the covariance matrix the (observed) Fisher information matrix:

\[
\pi_F = -\frac{\partial^2 \ln P(\Theta)}{\partial \theta_n \partial \theta_{n'}}
\]
(12)

is symmetric and positively defined. It follows that there is an orthogonal matrix \( U \) such that \( \Delta_{RHS} \) in Eq. (11) and hence \( \Delta_{LHS} \) also, could be written in the normal form:

\[
\sum_{n, n'=1}^{N} \pi_{E_{n n'}} (\hat{\theta}_n - \theta_n)(\hat{\theta}_{n'} - \theta_{n'})
\]

\[
\equiv \Delta_{LHS} = \sum_{n=1}^{N} m_n \tilde{v}_n^2 ,
\]
(13)

where \( \tilde{v}_n \) are some functions of \( \hat{\theta}_n \)s and \( m_n \) are the elements (obtained for \( \Delta_{LHS} \)) of the positive diagonal matrix \( \pi_F \) which because of the equality (13) has to be equal to the diagonal matrix obtained for \( \Delta_{RHS}, \) i.e.:

\[
\pi_F = D^T U^T \pi_F U D .
\]
(14)

Here \( D \) is the scaling diagonal matrix with the elements \( d_n = \sqrt{\lambda_n} \) and \( \lambda_n \)s are the eigenvalues of \( \pi_E. \) Finally, we could rewrite Eq. (14) in the form of the Frieden structural microscopic IP:

\[
\mathcal{E} + \pi_F = 0 ,
\]
(15)

where

\[
\mathcal{E} = -U (D^T)^{-1} \pi_F D^{-1} U^T .
\]
(16)

The existence of the normal form (13) is a very strong condition which makes the whole Frieden analysis possible. Two particular assumptions lead to the simple physical cases. When the structure \( \pi_F = 0 \) then from Eq. (11) we obtain a form of the ”master equation” (compare with Eq. (5)) with:

\[
\pi_F = \text{diag} \left( 2 \frac{\partial \ln P(\Theta)}{\partial \theta_n} \right) , \quad \hat{\theta}_n = \sqrt{\theta_n - \theta_n}, \quad \pi_F = 0
\]
(17)
and \( d_n = \sqrt{2 \frac{\partial \ln P}{\partial \theta_n}/\lambda_n} \). If we instead suppose that the distribution is regular \(^2\) then with \( \frac{\partial d_n P}{\partial \theta_n} = 0 \) for all \( n = 1, \ldots, N \), we see from Eq. (11) that \( (d_n = \sqrt{2/\lambda_n}) \):

\[
\mathbf{F}_n = (2 \delta_{nn'}) , \quad \hat{\nu}_n = \frac{\mathbf{F}_n}{N}.
\]

After integrating Eq. (15) with the measure \( d^N P(\Theta) \) we recover the integral structural IP (postulated previously in \(^3\) although in a different form and interpretation but) in exactly the same form as in \(^7\):

\[
Q + I = 0 ,
\]

where \( I \) is the Fisher information channel capacity:

\[
I = \int d^N x P(\Theta) \sum_{n,n' = 1}^N (\delta F)_{nn'}
\] (20)

and \( Q \) is the SI:

\[
Q = \int d^N x P(\Theta) \sum_{n,n' = 1}^N (\delta \Phi)_{nn'} .
\] (21)

The IP given by Eq. (19) is the structural equation of many current physical models.

3 The information transfer As \( I \) is the infinitesimal type of the Kulback-Leibler entropy \(^3\) which in the statistical estimation is used as a tool in the model choosing procedure, hence the conjecture appears that \( I \) could be (after imposing the variational and structural IPs) the cornerstone of the equation of motion (or generating equation) of the physical system. These equations are to be the best from the point of view of the IPs what is the essence of the Frieden’s EPI. The inner statistical thought in the Frieden method is that the probing (sampling) of the space-time by the system (even when not subjected to the real measurement) is performed by the system alone, that using its proper field (and connected amplitudes) of rank \( N \), which is the size of the sample, probes with its kinematical "Friedenian" degrees of freedom the position space accessible for it. The transition from the statistical form (1) of the FI to its kinematical representations (2) is given in Footnote 2.

Let us consider the following informational scheme of the system. Before the measurement takes place the system has \( I \) of the system which is contained in the kinematical degrees of freedom and \( Q \) of the system contained in the structural degrees of freedom, as in Figure 1a. Now, let us "switch on" the measurement during which the transfer of the information \( (TI) \) follows the rules (see Figure 1b):

\[
J \geq 0 \text{ hence } \delta I = I' - I \geq 0 , \quad \delta Q = Q' - Q \leq 0 ,
\]

where \( I', Q' \) are the FI and SI after the measurement, respectively and \( J \) is the TI. We postulate that in the measurement the TI is ideal at the "point" \( q \), i.e. we have that \( Q = Q' + J = Q + \delta Q + J \) hence \( \delta Q = -J \). This means that the whole change of the SI at the "point" \( q \) is transferred. On the other hand at the "point" \( i \) the rule for the \( T'I \) is that \( I' \leq I + J \) hence \( 0 \leq \delta I = I' - I \leq J \). Therefore, as \( J \geq 0 \) we obtain that \( |\delta I| \leq |\delta Q| \), the result which is sensible as in the measurement information might be lost. In the ideal measurement we have obtained that \( \delta Q = -\delta I \).

Previously we postulated the existence of the additive total physical information \( (TPI) \) \(^7\):

\[
K = Q + I .
\] (23)

In \(^8\) the intuitive condition that \( K \geq 0 \) was chosen leading to the following form of the structural IP:

\[
\kappa Q + I = 0
\] (24)

or

\[
Q + I = 0 \text{ for } \kappa = 1 ,
\] (25)

which we now derived in Eq. (19). In the case of Eq. (24) we obtain that \( K \) is equal to

\[
K = Q + I = 0 \text{ for } \kappa = 1 .
\] (26)

The structural (internal) IP in Eq. (24) \(^7\) is operationally equivalent to the one postulated by Frieden \(^3\), hence it has at least the same predictive power. For the short description of the difference between both approaches see \(^7\). The other IP is the scalar (variational) one. It has the form \(^7\):

\[
\delta K = \delta(Q + I) = 0 \Rightarrow K = Q + I \text{ is extremal (27)}
\]

The principles (24) and (27) are the cornerstone of the Frieden EPI. They form the set of two differential equations for the amplitudes \( q_n \) which could be consistent, leading for \( \kappa = 1 \) or \( 1/2 \) to many well known models of the field theory and statistical physics \(^3\), as we mentioned in the Introduction. It could be instructive to recalculate them again using the new interpretation of \( K \) \(^8\).

Finally, let us notice that in the concord with the postulated behavior of the system in the measurement, we have obtained \( \delta I \leq J = -\delta Q \) from which it follows that:

\[
K' = I' + Q' \leq (I + J) + (Q - J) = I + Q = K
\]

\[
\Rightarrow K' \leq K .
\] (28)

In the case of \( \delta I = -\delta Q \) we obtain \( K' = K \). Then it means that the \( TPI \) remains unchanged in the ideal measurement \( (\delta I = -\delta Q) \) which, if performed on the intrinsic level of sampling the space by the system alone (and not by the observer), could possibly lead to the variational IP given by Eq. (27).

4 Conclusions The overwhelming impression of the EPI is that the \( TPI \) is the ancestor of the Lagrangian \(^3\). As the Fisherian statistical formalism and hence the IPs lie here really as the background of the description of the physical system it is not the matter of interpretation only especially that the structural IP \( (\text{which is very close to the approach used in } \text{[7,8]}) \) was derived in Section 2.2.
Figure 1 Panel: (a) The system before the measurement: $Q$ is the $SI$ of the system contained in the structural degrees of freedom and $I$ is the $FI$ of the system contained in the kinematical degrees of freedom. (b) The system after the measurement: $Q'$ is the $SI$ and $I'$ is the $FI$ of the system; $\delta Q = Q' - Q \leq 0$ and $\delta I = I' - I \geq 0$ as the transfer of the information ($TI$) in the measurement takes place with $J \geq 0$. In the ideal measurement $\delta I = -\delta Q$.

Finally, let us at least mention that there are other fields of science were EPI could be used, e.g. the econophysics [3][9][10]. We envisage that the formalism developed in this paper would be applied in quantum information processing and quantum game theory also [9][10].

Acknowledgements This work is supported by L.J.CH.

This work was supported in part by the scientific network Laboratorium Fizycznych Podstaw Przetwarzania Informacji sponsored by the Polish Ministry of Science and Higher Education (decision no 106/E-345/BWSN-0166/2008) and by the University of Silesia grant.

References

[1] R.A. Fisher, Phil. Trans. R. Soc. Lond. 222, 309 (1922). R.A. Fisher, Statistical methods and scientific inference, 2nd edn. (London, Oliver and Boyd, 1959).
[2] Y. Pawitan, In all likelihood: Statistical modelling and inference using likelihood, (Oxford Univ. Press, 2001).
[3] B.R. Frieden, Found. Phys. 16, 883 (1986); Phys. Rev. A 41, 4265 (1990). B.R. Frieden, B.H. Soffer, Phys. Rev. E 52, 2247 (1995). B.R. Frieden, Phys. Rev. A 66, 022107 (2002). B.R. Frieden, A. Plastino, A.R. Plastino and B.H. Soffer, Phys. Rev. E 66, 046128 (2002); Phys. Lett. A 304, 73 (2002). B.R. Frieden, Science from Fisher information: A unification, (Cambridge Univ. Press, 2004).
[4] C.R. Rao, Bulletin of the Calcutta Math. Soc. 37, 81-91 (1945).
[5] B. Efron, Ann. Statistics 3, 1189 (1975). A.P. Dawid, Ann. Statistics 3, 1231 (1975); Ann. Statistics 5, 1249 (1977). S. Amari, Ann. Statistics 10, 357 (1982). H. Nagaoka and S. Amari, Technical Report METR 82-7, Dept. of Math. Eng. and Instr. Phys, Univ. of Tokyo, (1982). A. Fujisawa and H. Nagaoka, Phys. Lett. A 201, 119 (1995); J. Math. Phys. 40, 4227 (1999). K. Matsumoto, PhD thesis, Univ. of Tokyo, (1998). S. Amari, H. Nagaoka, Methods of information geometry, (Oxford Univ. Press, 2000).
[6] I. Bengtsson, K. Życzkowski, Geometry of quantum states, (Cambridge Univ. Press, 2006).
[7] J. Syska, Phys. Stat. Sol.(b), 244, No.7, 2531-2537 (2007)/DOI 10.1002/pssb.200674646.
[8] D. Mrozikiewicz, Master thesis, (in Polish) unpublished, Inst. of Physics, Univ. of Silesia, Poland, (2008).
[9] E. W. Piotrowski, J. Sładkowski, Int. J. Theor. Phys. 42, 1089 (2003).

E.g. in [3] the scalar field case is analyzed for which $Q$ is connected with the rest mass of the particle.
[10] E. W. Piotrowski, J. Sładkowski and J. Syska, presented at the SIGMA PHI 2008 conference, to appear in Central European Journal of Physics.