EVIDENCE INTEGRALS

W. VON DER LINDEN, R. FISCHER AND V. DOSE
Max-Planck-Institut für Plasmaphysik, EURATOM Association
D-85740 Garching b. München, Germany ¶

Abstract. Evidence integrals are a key ingredient of quantified maximum entropy (QME). They allow to evaluate hyper-parameters which in turn determine the amount of regularization, the noise level, confidence intervals etc. Consequently, the exact evaluation of these multi-dimensional integrals is of central importance to the theory. Since the conventional 'steepest descent' (SD) approximation fails frequently a more refined approach is needed. Using diagrammatic techniques we derive a correction factor to the SD result which is readily incorporated in existing QME codes and which provides significantly improved results.

Key words: Quantified MaxEnt, evidence integral, regularization parameter

1. Introduction

From the theoretical point of view QME [1,2] is a straight forward application of Bayesian probability theory. The numerical implementation, however, is not that simple since multi-dimensional *evidence integrals* are required. We consider an inferential problem summarized by $\mathcal{I}$ and characterized by certain *hyper-parameters* λ, such as the regularization parameter α, the width of the preblur [2], and the noise level. In QME the *prior* probability $p(f|\lambda\mathcal{I}) \propto \exp(\alpha S)$ for the image f depends on the entropy S. We restrict the discussion to a Gaussian *likelihood* function $p(D|f\lambda\mathcal{I}) \propto \exp(-\frac{1}{2}\chi^2)$, where χ^2 measures the discrepancy between data D and fit. Consequently, the joint posterior of the image and the hyper-parameters is of the form $p(\lambda f|D\mathcal{I}) = c(\lambda)\exp\phi(f,\lambda)$. According to the Bayesian sum-rule the evidences for the data $p(D|\mathcal{I})$ and for the hyper-parameters $p(\lambda|D\mathcal{I})$, respectively, are proportional to the multi-dimensional evidence integrals in image space

$$I = \int_0^\infty e^{\phi(f,\lambda)} \mathcal{D}f = \int_0^\infty e^{\phi(f,\lambda)} \prod_i \frac{df_i}{\sqrt{f_i}} \quad . \tag{1}$$

In general such integrals cannot be determined analytically. Several numerical approaches to 'add up' have been proposed and analyzed by Skilling[3]; so far,

¶e-mail: wvl@ibmop5.ipp-garching.mpg.de

however, with limited success. The commonly employed SD comprises three approximations. a) ϕ is expanded to second order about its maximum f^* (Gaussian approximation), b) it is assumed that $\exp(f)$ is sharply peaked at positive f^* so that the lower integration limits can be extended to $-\infty$, and c) the integration measure is replaced by a constant and extracted from the integral

$$I^{\mathrm{SD}} := \frac{e^{\phi(f^*,\lambda)}}{\prod_i \sqrt{f_i^*}} \int_{-\infty}^{\infty} e^{\Delta f^T (A+B) \Delta f} df^N \;=\; \frac{e^{\phi(f^*,\lambda)} \pi^{N/2}}{\prod_i \sqrt{f_i^*}} \det^{-\frac{1}{2}}(A+B) \qquad . \tag{2}$$

We introduced the abbreviation $\Delta f = f - f^*$ and split the Hessian $-\frac{1}{2}\nabla\nabla^T\phi\big|_{f^*}$ into the diagonal matrix A and the residual trace-zero matrix B. The SD approximation has the advantage of providing an analytic expression for the evidence integrals that can be evaluated with little numerical effort. While the Gaussian approximation has proven to be reliable, tampering with the integration limits and measure has the following negative impact. Due to the divergent prefactor $(\prod_i f_i^{*\,-1/2})$ and the modified integration limits, images with values close to zero are preferred which is accomplished by less regularization, i.e. small α. For the same reason confidence intervals are in general strongly overestimated.

2. Diagrammatic Approach

To overcome these shortcomings we propose a diagrammatic approach providing a correction factor $\mathcal{C} = I/I^{\mathrm{SD}}$ to the SD result which can be implemented into existing QME codes with little effort. The impatient reader who is merely interested in a reliable correction formula for his QME code can directly jump to Eq.13,15. We start out by writing the Taylor expansion of the exponent ϕ as

$$\phi = \phi\big|_{f^*} - \Delta f^T B \Delta f + \sum_i \left[R(f_i, f_i^*) - A_{ii} \Delta f_i^{\,2} \right] \qquad . \tag{3}$$

Only the second order terms, to which the Gaussian approximation is restricted, introduce correlations between image points f_i and f_j. Higher order contributions originate from the entropy and are 'diagonal' in f, expressed by the special form $\sum_i R(f_i, f_i^*)$. The evidence integral is recast into

$$I(B) = e^{\phi(f^*,\lambda)} \int_0^{\infty} e^{-\Delta f^T B \Delta f} \prod_i e^{R(f_i, f_i^*) - A_{ii}\Delta f_i^{\,2}} \mathcal{D}f \qquad . \tag{4}$$

The analytic evaluation of this integral is hampered by the integration limits and the image-space measure. The correction factor reads

$$\mathcal{C} \;:=\; \frac{I(B)}{I^{\mathrm{SD}}} = \frac{I(B)}{I(0)} \det^{\frac{1}{2}}(\mathbb{1} + BA^{-1}) \prod_i \kappa_i \tag{5}$$

$$\kappa_i \;:=\; \sqrt{\frac{A_{ii} f_i^*}{\pi}} \int_0^{\infty} \mathcal{D}f_i \; e^{R(f_i, f_i^*) - A_{ii}(f_i - f_i^*)^2} \tag{6}$$

The SD approximation becomes progressively reliable as $f_i^* \to \infty$ and $\mathcal{C} \to 1$ in this limit. The one-dimensional integrals in Eq.6 can be determined easily by numerical means or even analytically as we will see below. The difficult part is

$$\frac{I(B)}{I(0)} = \int_0^\infty e^{-\Delta f^T B \Delta f} \underbrace{\prod_i \frac{e^{R(f_i,f_i^*)-A_{ii}\Delta f_i{}^2}}{\int_0^\infty \mathcal{D}f'_i \, e^{R(f'_i,f_i^*)-A_{ii}\Delta f_i{}^2}}}_{\rho(f)} \mathcal{D}f =: \langle e^{-\Delta f^T B \Delta f} \rangle_\rho \quad (7)$$

The distribution function $\rho(f)$ is positive definite and normalized to one. Furthermore, due to the product form of $\rho(f)$, $\langle \Delta f_i \Delta f_j \rangle_\rho = \langle \Delta f_i \rangle_\rho \langle \Delta f_j \rangle_\rho$, $\forall i \neq j$. The particular form of Eq.7 prompted us to employ diagrammatic techniques [1]. The Taylor expansion of the exponential yields an N-th order contribution

$$\langle \Big(\sum_{\substack{ij \\ =1}}^{L} \Delta f_i B_{ij} \Delta f_j \Big)^N \rangle_\rho = \langle \sum_{\substack{i_1,j_1,\ldots,i_N,j_N \\ =1}}^{L} \Delta f_{i_1} B_{i_1 j_1} \Delta f_{j_1} \cdots \Delta f_{i_N} B_{i_N j_N} \Delta f_{j_N} \rangle_\rho \quad (8)$$

For a graphical representation we introduce N vertical pairs of sites labeled i and j and enumerate from 1 to N, as illustrated in fig.1 for $N = 7$. Sites (circles)

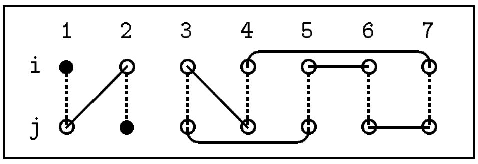

Figure 1. Graphical representation of the 7-th order term as explained in the text.

with coordinates (ν) stand for factors Δf_ν, where ν is a compound index for i_l or j_l, respectively. All vertical pairs are connected by dashed lines which for the sites (i_l) and (j_l) carry a factor B_{i_l,j_l}. Next we split Δf_ν into expectation value $\Delta_\nu :=< \Delta f_\nu >_\rho$ and fluctuation $\delta_\nu := \Delta f_\nu - \Delta_\nu$ which will simplify the diagrammatic rules since $< \delta_\nu >= 0$. Consequently, the N-th order product in Eq.8 decomposes further into a sum over 4^N terms consisting of all possible distributions of Δ-s and δ-s on the 2N sites. Graphically we represent expectation values (Δ) by full circles and fluctuations (δ) by open circles. All possible arrangements of open and full circles, respectively, have to be generated and summed over. A typical arrangement is depicted in fig.1. The Δ-s in the product can be taken out of the expectation value in Eq.8 leaving behind expectation values of products of δ factors only. So far, the entire derivation is exact. Next we propose a pair-fluctuation approximation (PA) which corresponds to a partial summation of diagrams to infinite order. [2] In PA higher order correlation functions are decomposed into all

[1] An exact but much less efficient approach is suggested by Monte Carlo importance sampling.
[2] A systematic improvement to higher order correlations is straight forward, although it becomes increasingly cumbersome.

possible products of pair-fluctuations

$$\langle \delta_{\nu_1} \delta_{\nu_2} \cdots \delta_{\nu_{2k}} \rangle \;=\; \sum_{\mathcal{P}} \langle \delta_{\mathcal{P}_1^{(1)}} \delta_{\mathcal{P}_2^{(1)}} \rangle \cdots \langle \delta_{\mathcal{P}_1^{(k)}} \delta_{\mathcal{P}_2^{(k)}} \rangle \qquad . \tag{9}$$

The sum runs over all $(2k-1)!!$ different complete pairings of the indices $\nu_1, \ldots, \nu_{2k}$. Expectation values for an odd number of δ-s vanish in PA since $\langle \delta \rangle = 0$. The PA is exact for a Gaussian ρ with unrestricted support like in Eq.2. Due to the product-form of Eq.7 it is also exact in the general case for δ-s with unequal indices. The PA approach is closely related to Wick's theorem [4] used in quantum-field theory. Graphically, pair-fluctuations $\langle \delta_\nu \delta_{\nu'} \rangle$ are represented by bonds (solid lines) connecting the open circles at ν and ν'. The complete pairing of δ-s demands that precisely two bonds are attached to each open circle, while there are no bonds connected to full circles. The diagrams are generally composed of *connected diagrams*, in which all sites are successively connected by solid or dashed lines, respectively. E.g. the diagram in fig.1 consists of 2 connected diagrams, one of order 2 and one of order 5. The order of a connected diagram is the number of B-factors, or rather dashed lines. Since $B_{\nu\nu} = 0$ and $< \delta_\nu \delta_{\nu'} >= 0$ for $\nu \neq \nu'$ there are no bonds between vertical nearest neighbor sites. Generally the N-th order term can be decomposed into a set of connected diagrams, specified by $\{n_1, \ldots, n_N\}$, with n_l being the number of connected diagrams of order l. We introduce the constraint $\sum_l l\ n_l = N$ to enforce the correct expansion order N. All connected diagrams of order N contribute one and the same factor $\mathcal{F}_N := \langle (\Delta^T B \Delta)^N \rangle_\rho^c$ to the product Eq.8 so that we merely have to count the number of topologically distinct connected diagrams. The superscript c indicates that only connected diagrams are taken into account. The number of ways to group N different objects into subsets defined by the arrangement $\{n_1, \ldots, n_N\}$ is $N!/\prod_{l=1}^N (n_l!\, l!^{n_l})$. E.g. for the integers $(1, 2, 3, 4)$ there are 3 possibilities to form the arrangement specified by $\{n_1 = 0, n_2 = 2, n_3 = 0, n_4 = 0\}$, namely (12)(34), (13)(24), and (14)(23). Eventually, Eq.7 is recast into the remarkable expression

$$\begin{aligned} \langle e^{-\Delta f^T B \Delta f} \rangle_\rho &= \sum_{N=0}^{\infty} \sum_{\substack{n_1, \ldots, n_\infty \\ =0}}^{\infty} \delta(\sum_l l n_l - N) \frac{\prod_l (-1)^{l n_l}}{\prod_l (n_l! l!^{n_l})} \prod_{l=1}^{\infty} (f_l)^{n_l} \\ &= \exp\left(\langle e^{-\Delta f^T B \Delta f} \rangle_c - 1 \right) \qquad , \end{aligned} \tag{10}$$

according to which it is sufficient to evaluate connected diagrams only. It is harmless to extend the number of terms in the second sum to n_∞ due to the delta function constraint. Obviously, connected diagrams form either one closed loop containing only open circles or one open loop with full circles at the ends. An example of a closed loop of order 5 and an open loop of order 2 has already been encountered in fig.1 . To generate all closed loops of order N we start at site i_1 and place a bond to the upper or lower site of one of the remaining $(N-1)$ vertical pairs. With the alternative site of this pair we proceed to form a bond to one of the now remaining $(N-2)$ vertical pairs and so on until the loop is closed. In total there are therefore $(N-1)!2^{N-1}$ possible closed loops contributing

$$\mathcal{F}_N^c = (N-1)!\mathrm{tr}[(B\tilde{A}^{-1})^l]/2 \tag{11}$$

to Eq.8. The definition of $\tilde{A}$ is given in Eq.14, where we exploit the fact that the integral-correction vanishes if the regularization parameter is such that $\tilde{A}+B$ is not positive definite. The definition of $\tilde{A}$ accelerates the numerical procedure by enforcing positivity from the outset. Open loops are constructed from closed loops by cutting one of the N bonds which yields a total number of $N!2^{N-1}$. Their contribution to Eq.8 amounts

$$\mathcal{F}_l^o = l!\Delta^T (B\tilde{A}^{-1})^l \tilde{A}\Delta \tag{12}$$

Combining Eq.11,12,10, and 5 we obtain the desired correction factor

$$\begin{aligned} \mathcal{C} &= \sqrt{\frac{\det(\mathbb{1} + BA^{-1})}{\det(\mathbb{1} + B\tilde{A}^{-1})}} e^{\Delta^T((\mathbb{1}+B\tilde{A}^{-1})^{-1}-\mathbb{1})\tilde{A}\Delta} \prod_i \kappa_i & (13) \\ \tilde{A}_{\nu\nu'} &:= \delta_{\nu,\nu'} \cdot \max[(2\langle\Delta f_\nu^2\rangle)^{-1}, A_{\nu\nu}] & (14) \end{aligned}$$

The expectation values of Δf_i and ${\Delta f_i}^2$, entering $\tilde{A}_{\nu\nu'}$ and Eq.13, are easily determined numerically. An even simpler and yet fairly reliable approach is the Gaussian approximation, i.e. $R(\Delta, f^*) = 0$ in Eq.3 and Eq.6. In this case κ_i can be expressed in terms of the modified Bessel functions $I_{\pm 1/4}$ [5] while Δ_i and ${\Delta f_i}^2$ follow from κ_i upon differentiation with respect to f_i^* and A_{ii}, respectively

$$\begin{aligned} \kappa_i &= \sqrt{\frac{z\,\pi}{2}} e^{-z} \Big(I_{1/4}(z) + I_{-1/4}(z) \Big) \Big|_{z=A_{ii}\, {f_i^*}^2/2} \\ A_{ii}\langle {\Delta f_i}^2 \rangle &= z \left(1 - \frac{I'_{1/4}(z) + I'_{-1/4}(z)}{I_{1/4}(z) + I_{-1/4}(z)} \right) \Big|_{z=A_{ii}{f_i^*}^2/2} \qquad (15) \\ A_{ii}\langle \Delta f_i \rangle &= (f_i^*)^{-1} \Big(1/4 - A_{ii}\langle {\Delta f_i}^2 \rangle \Big) \end{aligned}$$

In summary the diagrammatic approach treats the diagonal part of ϕ exactly, while the off-diagonal part is approximated by partial summation of diagrams to infinite order. This approximation includes all terms up to second order completely.

3. Putting the integral correction to the test

In order to assess the importance of the proper treatment of the evidence integrals and the reliability of the integral correction formula proposed in the present paper we studied a variety of problems. One example will be given below others can be found elsewhere in these proceedings [7]. We compared the results obtained by the diagrammatic approach (PA) with those of the steepest-descent approximation (SD). In both cases α is determined upon maximizing the evidence $P(\alpha|D\mathcal{I})$. A conceptually different approach was proposed recently by Strauss, Wolpert and Wolf[6] (SWW) in which α is marginalized over, making evidence integrals obsolete as long as no further hyper-parameters or confidence intervals are required. The PA results agree remarkably well with those obtained by the SWW approach in all cases where the SD approximation failed badly. As a specific example we present here an ill-posed inversion problem encountered in Quantum Monte Carlo simulations. Without digressing into the underlying physics the inversion problem reads

$$g_m = \frac{1}{\pi} \int_0^\infty \frac{\omega}{1 - e^{-\beta\omega}} (e^{-\tau_m \omega} + e^{-(\beta - \tau_m)\omega}) f(\omega)\ d\omega \quad . \tag{16}$$

The parameters β and τ_m, $m = 1, \ldots, 40$ are known. Quantum Monte Carlo simulations provide values for g_m with a statistical error of about 1%. QME is employed to infer $f(\omega)$, which is displayed in fig.2 . Obviously SD strongly overfits the data.

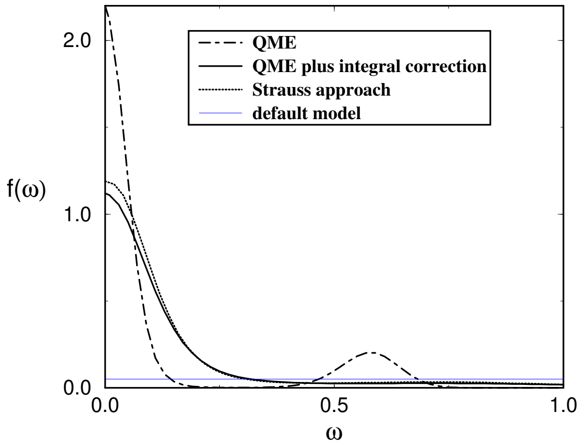

Figure 2. *Comparison of QME results obtained by different approximations.*

In fact, the associated regularization parameter is an order of magnitude too small. SD would not just lose a 'beauty contest' but it yields the **wrong** physics. We note in closing that confidence intervals are overestimated correspondingly.

References

1. S. F. Gull, "Developments in maximum entropy data analysis," in *Maximum Entropy and Bayesian Methods*, J. Skilling, ed., pp. 53–71, Kluwer Academic, Dordrecht, 1989.
2. J. Skilling, "Quantified maximum entropy," in *Maximum Entropy and Bayesian Methods*, P. F. Fougère, ed., pp. 341–350, Kluwer Academic, Dordrecht, 1990.
3. J. Skilling, "How to add up," in *Maximum Entropy and Bayesian Methods 1993*, G. Heidbreder, ed., Kluwer Academic, Dordrecht, 1995.
4. R. D. Mattuck, *A Guide to Feynman Diagrams in the Many-Body Problem*, Dover Publications, Inc., New York, 1992.
5. I. Gradshteyn and I. Ryzhik, "formulae 3.462, 9.240, 9.215," in *Tables of Integrals, Series, and Products*, Academic Press, San Diego, 1980.
6. C. E. M. Strauss, D. H. Wolpert and D. R. Wolf, "Alpha, evidence, and the entropic prior," in *Maximum Entropy and Bayesian Methods*, G. Heidbreder, ed., Kluwer Academic, Dordrecht, 1995.
7. R. Fischer, W. von der Linden and V. Dose, "On the importance of α marginalization," *In these proceedings*, 1995.