Institute of Physics Publishing
doi:10.1088/1742-6596/36/1/011

Journal of Physics: Conference Series **36** (2006) 55–59
The 12th Central European Workshop on Quantum Optics

Likelihood and entropy for statistical inversion

Z. Hradil, J. Řeháček

Department of Optics, Palacky University, 17. listopadu 50, 772 00 Olomouc, Czech Republic

E-mail: hradil@optics.upol.cz

Abstract. Two celebrated statistical principles- Principle of Maximum Likelihood and Principle of Maximum Entropy are merged establishing a novel estimation scheme for statistical inversion.

1. Introduction

The role of entropy as uncertainty measure in communication and information theory was recognized by Shannon. His definition $S = -\sum_n p_n \log p_n$ is unique in the sense that fulfills reasonable demands put on the information measure associated with a probability distribution p_n. In particular, the uniform distribution provides the largest uncertainty. The importance of entropy for physical and technical practice was first noticed by Jaynes [1], who proposed a variational method known as the principle of Maximum Entropy (MaxEnt): The inferred probability distribution should fulfill the given constraints and simultaneously maximize the Shannon entropy. This gives the most unbiased solution of the problem consistent with the given observations. Such inferences appeared to be extremely useful in many applications covering the fields of statistical inference, communication problems or pattern recognition [2, 3, 4].

But entropy is not the only important functional in probability theory. The entropic measure known as Kullback-Leibler divergence [5] or relative entropy $E(\{p_i\}|\{q_i\}) = \sum_i p_i \log(p_i/q_i)$ bears striking resemblance to the Shannon entropy, however it posses rather different interpretation. It quantifies the *distance* in the statistical sense between two different distributions p_i and q_i. Provided that one party (p_i in our notation) are the sampled relative frequencies, the principle of minimum relative entropy coincides with the maximum likelihood (MaxLik) estimation problem [6, 7]. At present there are many examples of successful application of MaxLik estimation technique for solving inverse problems, see e.g. [8].

Though both the celebrated principles, MaxEnt and MaxLik, rely on the notion of entropy, their usage and interpretation differ substantially. The former one provides the most pessimistic guess consistent with the data, while the later one provides the most optimistic fit to the given data. [2, 6]. Both methods have their drawbacks: MaxLik can deal with noisy data in realistic experiments but it usually requires a certain cut-off in the parameter space. Otherwise the solution may appear us under-determined; instead of a unique answer there can exist a convex set of most likely states. MaxEnt principle removes this ambiguity by selecting the most unbiased solution. However, since realistic data are noisy, the corresponding constraints may appear to be inconsistent. The purpose of this contribution is to merge both these concepts into a single estimation procedure providing the most likely and most unbiased solution without any cut-offs.

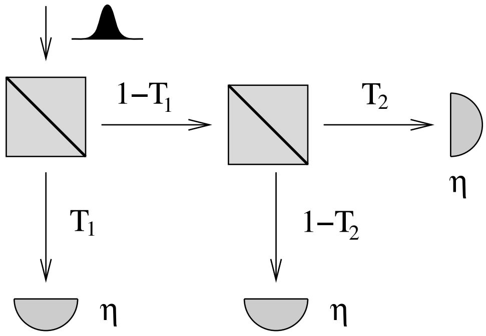

Figure 1. An optical multiport; μ is the quantum efficiency of used detectors; T_j are transmission coefficients.

2. Inverse problems

Let us consider a state $\mu_i \geq 0$, $i = 1 \ldots N$, which is to be inferred from the observed relative frequencies f_j. For simplicity we will assume that the dependence of the corresponding theoretical probabilities p_j on μ_i is linear and positive,

$$p_j = \sum_i c_{ji}\mu_i, \quad c_{ji} \geq 0, \quad j = 1 \ldots M. \tag{1}$$

When $N > M$ the problem is said to be under-determined. Typically, this happens when the parameter space is infinite, $i = 1 \ldots \infty$.

An interesting example of an under-determined problem in optics is the reconstruction of the photon content of a light pulse, see figure 1. The input pulse is linearly split into several parts that get detected by a common detectors with yes/no response. Let us denote j the outcome of a single run. Having d detectors there are 2^n of such possible results. Repeating now the experiment many times with identical light pulses, the probability p_j of detecting result j can be expressed in the form (1), where μ_j represents the photon-number statistics of the light source. The corresponding coefficients c_{ji} can be found in [9]. Of course, this problem is always under-determined since the number of photons is not bounded from above.

3. Maximum-likelihood inversion

MaxLik solution is obtained by minimizing the relative entropy (or log-likelihood) between data and theory $E(f_j|\mu_j)$. Numerically, this can be done using the Expectation-maximization algorithm [10, 11],

$$\mu_i^{n+1} = \mu_i^n R_i, \quad R_i = \frac{\sum p_j}{\sum \mu_i^n \sum_j c_{ji}} \sum_j \frac{f_j c_{ji}}{p_j}, \tag{2}$$

where n labels successive iterations. In well- and over-determined case, $M \geq N$, the MaxLik solution is unique due to the convexity of E. In under-determined case, the likelihood exhibits a plateau of maximum-likely states. In the latter case, the numerical result strongly depends on the starting point of iterations: Let us impose an additional constraint on the MaxLik solution to remove this ambiguity.

4. Maximum-entropy inversion

Taking the observed data as constraints

$$p_j(\mu_i) = f_j, \tag{3}$$

the entropy of the inferred state, $E = -\sum_i \mu_i \log \mu_i$, will be maximized by choosing

$$\mu_i = \exp\big(\sum_j \lambda_j c_{ji}\big), \tag{4}$$

where the Lagrange multipliers λ_j are determined by the constraints. Notice that the MaxEnt solution is always unique. The problem here is that noisy data can yield inconsistent constraints, meaning that the system of equations (3) together with (4) has no solutions. Being aware of this subtlety we will in the following use the concept of entropy to get a unique MaxLik reconstruction.

5. MaxEnt assisted MaxLik inversion

Having found a maximum of the likelihood functional in section 3, we still do not know whether this solution is unique or not. Provided a closed set of such states exists, we would like to maximize the entropy functional over it. In this way we will get the least biased maximum-likelihood guess.

Notice that due to the convexity of the relative entropy all states belonging to the maximum likely set must generate the same theoretical probabilities $p_j(\{\mu^{\mathrm{ML}}\}_k) = p_j(\{\mu^{\mathrm{ML}}\}_l)$, $\forall j, k, l$. We will take those probabilities as constraints of the new optimization problem: Maximize entropy $E(\mu_i)$ subject to constraints

$$p_j(\mu_i) = p_j(\mu_i^{\mathrm{ML}}), \quad j = 0 \dots M, \tag{5}$$

where μ_i^{ML} is a maximum likely state.

Now two distinct cases arise. Provided the measurement is complete, $\sum_j p_j = 1$, one can directly use the MaxEnt solution (4) in (5) yielding the set of nonlinear equations

$$\sum_i e^{\sum_{j'} \lambda_{j'} c_{ij'}} c_{ij} = \sum_i c_{ij} \mu_i^{\mathrm{ML}} \quad , \tag{6}$$

which are to be solved for the Lagrange multipliers λ_j generating the maximum-entropy maximum-likelihood estimate via (4).

In the more general case of an incomplete measurement, $\sum_j p_j \neq 1$, when some output channels are not observed, only the ratios of probabilities can be determined from data. As a consequence the log-likelihood $\sum_j f_j \log p_j$ that has led to (2) must be replaced by $\sum_j f_j \log(p_j / \sum p_j)$ and the constraints (5) assume a more general form

$$p_j(\mu_i) = \alpha p_j(\mu_i^{\mathrm{ML}}), \quad j = 0 \dots M. \tag{7}$$

Here α is a normalization constant to be optimized over. Denoting $\mu(\alpha)$ the α-dependent solution of (7), we look for α_{opt} maximizing the entropy, $\alpha_{\mathrm{opt}} = \arg\max_\alpha E[\mu(\alpha)]$. Numerical implementation of this problem will be discussed elsewhere.

6. Example

The proposed approach combines good features of maximum-likelihood and maximum-entropy methods. From the set of density matrices that are most consistent with the observed data in the sense of maximum likelihood we select the least biased one. At the same time the positivity, and thus also physical soundness, of the result is guaranteed.

Let us illustrate the MaxEnt assisted MaxLik inversion on the measurement of the photon-number statistics of a light pulse. The results of a numerical simulation are summarized in figure 2. For it we have chosen a mixture of two Poissonian pulses with the mean intensities 1

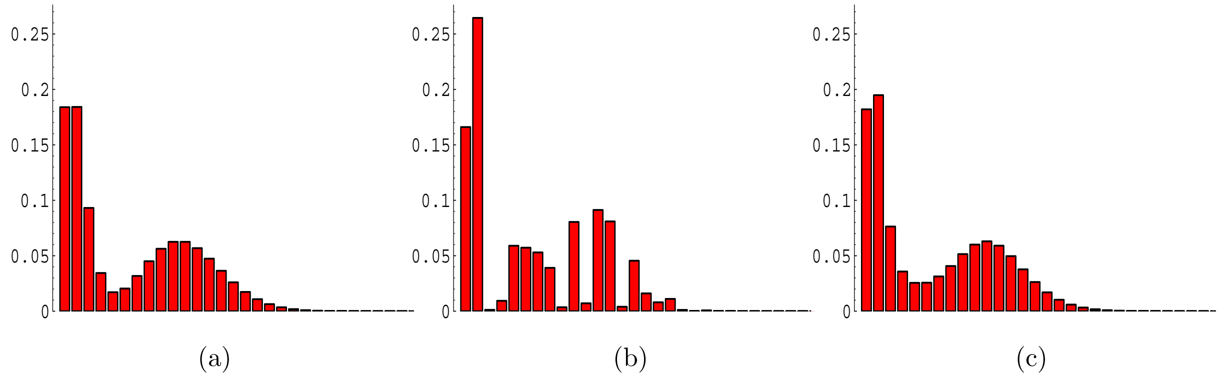

Figure 2. Simulation of an indirect photon-number measurement with the help of an optical multiport: (a) the true photon-number statistics; (b) a MaxLik solution obtained via the expectation-maximization algorithm from a randomly chosen starting point; (c) the maximum likely statistics having the largest entropy.

and 10 photons per pulse, respectively, see figure 2(a). The simulated multiport measurement device had four output channels with the transmission coefficients (see figure 1) $T_1 = 0.208$, $T_2 = 0.566$, $T_3 = 0.286$, and $T_4 = 0.084$. Such a device yields $M = 2^4 = 16$ possible independent outcomes. The goal was to reconstruct the photon-number statistics of the input pulse up to the photon number $n = 50$, so the parameter space had dimension $N = 50$. Notice that this inversion is strongly under-determined $N \gg M$.

Figure 2(b) shows a typical result of the maximum-likelihood estimation of section 3. The initial guess for the expectation-maximization algorithm was chosen in random. As can be expected, the family of maximum-likely states contains states that are far from the true state. Different initial guesses lead to very different answers.

To get a unique reconstruction the simulated data were processed using the method proposed in the previous section. Since we assume that all the output results are observed, the measurement is complete and the corresponding probabilities sum to unity $\sum_{j=1}^{16} p_j = 1$. Inserting the maximum-likely distribution of figure 2(b) on the right-hand-side of (6) and numerically solving the resulting set of nonlinear equations we get Lagrange multipliers representing the photon-number distribution shown in figure 2(c). This is the maximum-likely photon-number statistics having the largest entropy. Comparing the three panels of figure 2 one can say that the use of entropy as an additional criterion for selecting a unique MaxLik reconstruction is well justified.

7. Conclusion

We have demonstrated the utility of the maximum-entropy principle for tomographically incomplete quantum state reconstruction schemes. Although the entropic principles cannot be directly applied to noisy experimental data due to the possible inconsistency of constraints involved, they can be used to remove the ambiguity of maximum likelihood estimation. The proposed method could find applications in quantum homodyne detection and other related infinite-dimensional problems suffering from the lack of experimental data.

Acknowledgments

This work was supported by Project MSM6198959213 of the Czech Ministry of Education and Project FT-TA2/059 of the Czech Ministry of Industry and Trade.

References

[1] Jaynes E T 1963 Information Theory and Statistical Mechanics *1962 Brandeis Summer Lectures* vol 3, edited by Ford K W (Benjamin, New York) p 181

[2] Frieden B R 1983 *Probability, Statistical Optics, and Data Testing*, (Springer-Verlag, Berlin)

[3] Bužek V and Drobny G 2000 *J. Mod. Optics* **47** 2823

[4] Bužek V 2004 Quantum tomography from incomplete data via MaxEnt principle in *Quantum state tomography* edited by Paris M and Řeháček J Lect. Notes. Phys. vol **649** (Springer, Berlin) p 189

[5] Kullback S and Leibler R A 1951 *Ann. of Math. Stat.* **22** 79

[6] Fisher R A 1925 *Proc. Camb. Phi. Soc.* **2**2 700

[7] Helstrom C W 1976 *Quantum Detection and Estimation Theory* (Academic Press, New York)

[8] Hradil 2004 Quantum tomography from incomplete data via MaxEnt principle in *Quantum state tomography* edited by Paris M and Řeháček J Lect. Notes. Phys. vol **649** (Springer, Berlin) p 189

[9] Řeháček J, Hradil Z, Haderka O, Perina J Jr. and Hamar M 2003 *Phys. Rev. A* **67** 061801.

[10] Dempster A P, Laird N M and Rubin D B 1977 *J. R. Statist. Soc. B* **39** 1

[11] Vardi Y and Lee D 1993 *J. R. Statist. Soc. B* **55** 569