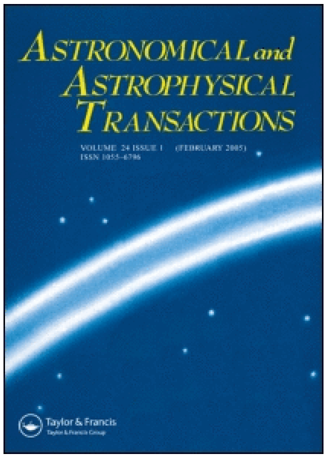

Astronomical & Astrophysical Transactions
The Journal of the Eurasian Astronomical Society
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713453505

The image randomness test for inverse problems

V. Yu. Terebizh [a]; V. V. Biryukov [a]
[a] Crimean Station of Sternberg Astronomical Institute, Nauchny, Crimea

Online Publication Date: 01 January 1995
To cite this Article: Terebizh, V. Yu. and Biryukov, V. V. (1995) 'The image randomness test for inverse problems', Astronomical & Astrophysical Transactions, 6:1, 37 - 57
To link to this article: DOI: 10.1080/10556799508232058
URL: http://dx.doi.org/10.1080/10556799508232058

Astronomical and Astrophysical Transactions, 1994,
Vol. 6, pp. 37–57
Reprints available directly from the publisher
Photocopying permitted by license only

THE IMAGE RANDOMNESS TEST FOR INVERSE PROBLEMS

V. Yu. TEREBIZH and V. V. BIRYUKOV

Crimean Station of Sternberg Astronomical Institute,
334413 Nauchny, Crimea

(Received December 31, 1992)

An Image Randomness Test (IRT) is applied to obtain efficient object estimates in inverse problems, specifically, in the image restoration. According to the IRT, an observed image can be considered as one of the typical random image simulations for a feasible object estimate. The likelihood is not a comprehensive estimate from the IRT point of view; in addition, the Maximum Likelihood principle contradicts the IRT. A Mean Likelihood requirement is proposed for this purpose together with the statistical independence test.

A connection of the concepts introduced with the information theory by Shannon is shown. The limiting accuracy and resolution are discussed for simple observational conditions. Numerical simulations show that the stability of the inverse solution for the method described is significantly increased.

Exact and asymptotic expressions for the information and entropy of a Poisson random variable are given in Appendices, as well as some examples of the parameter estimation with the mean information requirement.

KEY WORDS Data analysis, image restoration

1 INTRODUCTION

The statistical approach to the image restoration problem is based on the following main propositions (Helstrom, 1969, 1970; Terebizh, 1990a, 1991, 1992):

1. An unknown object is (in one-dimensional version, for notation simplicity) a deterministic set of parameters $S \equiv (S_1, \ldots, S_n)$. The parameters correspond usually to the mean counts in detector pixels for an ideal image forming system, but they may include some structural parameters of the object as well.

2. An observed image $N \equiv (N_1, \ldots, N_m)$ is a random simulation of intensity counts for a set of pixels of a real imaging system. The stochastic nature of the counts is caused not only by the background, but also by the photon noise. This means, in particular, that the inverse problem cannot be formulated in terms of integral equation theory.

3. The object S and the image N are connected by a stochastic image formation model, which should be described by the investigator as simply as allowed by the physical meaning of the considered problem. The ultimate goal of the model developing is an explicit expression for the conditional probability $f(N|S)$ to obtain the set N for any object S.

4. The image restoration problem is formulated as searching for the *statistical estimate* $S^* \equiv (S_1^*, \ldots, S_n^*)$ of the true object S on the basis of the observed image N, the model $f(N|S)$, and *a priori* information, which may include, for example, the non-negativity condition for the object, the Point Spread Function (PSF) shape $\{h_{jk}\}$, and the mean background $\{\gamma_j\}$. Being a function of random array N, any estimate $S^*(N)$ is a multidimensional random variable. Its bias and variances relative to the true object S can be considered as a quality measure of the estimate, i.e., of the inverse problem solution.

5. For a wide area of applications, an adequate quality measure is the mean square scatter:
$$\Omega_{ik} = \mathbf{E}[(S_i^* - S_i)(S_k^* - S_k)], \tag{1}$$
where $\mathbf{E}$ denotes the mathematical expectation. Due to the well-known *information inequality*, the scattering of an arbitrary estimate cannot be less than some boundary value that depends on the probability density $f(N|S)$. This means that the accuracy of the restoration (and, in general, of any inverse problem solution) is limited by a method-independent value which is determined by the nature itself. We denote as the *boundary estimate* such an estimate which attains the lower limit in the information inequality, and as the *efficient estimate* such one that has the least scatter in the chosen estimate class. The boundary estimate exists when, and only when, the density $f(N|S)$ belongs to the *exponential family*. As one can easily see that if the boundary estimate exists, it is also the efficient one (the opposite statement is incorrect). Generally speaking, the purpose of the investigator is not to search for just the boundary estimate, but to search for the efficient one, or very close to it practically. The reason is that the efficient estimation gives the accuracy of the solution of the inverse problem which cannot be improved in principle.

A detailed description of the statistical approach to inverse problems and of a number of model and real cases can be found in the above references and in Terebizh and Biryukov (1990, 1991) and Terebizh *et al.* (1991).

It seems that in the frame of the approach described the most important point is the search for a concrete way of estimating that can give the limiting accuracy for a wide range of observational conditions. From the practical point of view, only this problem is of any interest, of course.

The Maximum Likelihood (ML) principle has been considered in our previous papers as the most promising way to obtain the efficient estimates. The ML-estimate is defined by the following requirement (Fisher, 1912):
$$\hat{S}(N) = \arg\max_{S \in U} f(N|S), \tag{2}$$

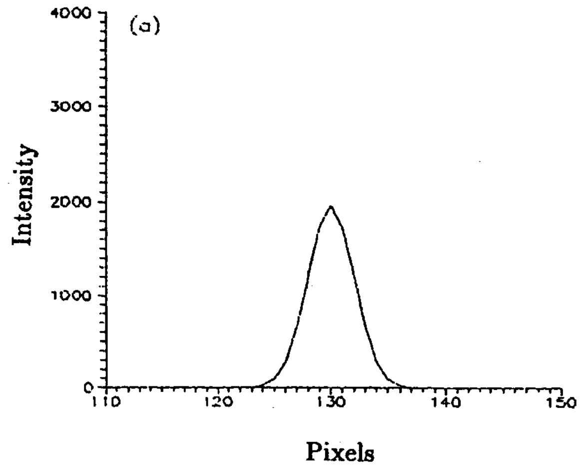

Figure 1 Examples of restoration of a Gaussian object (a). The blurred and noised image (b) was processed with maximum likelihood method (c), and mean information method (d, e).

where, as usually, $\arg\max f(x)$ means an argument value at which the function $f(x)$ is maximum in a given domain. The minimum *a priori* information about the object corresponds to the definition of the domain $U: \{S \geq 0\}$. Note that the ML-estimates are marked by a caret, and the arbitrary estimates are marked by an asterisk.

The reason for considering the ML-estimates as promising ones is the theorem which states that the ML-estimate coincides with the boundary estimate, if the latter exists at all (e.g., Borovkov, 1984). Model simulations and analytical examples that we considered earlier show the closeness of the ML- and efficient estimates to each other. At the same time, they are close not always, and, perhaps, the most evident way to show this fact is connected with the well-known phenomenon of the *instability of inverse solutions.*

Figure 1(a) shows an initial object *t8* as a Gaussian density distribution with the standard deviation $\sigma_{ob} = 2$ pixels and the total flux $F = 10^4$ counts. Figure 1(b) illustrates the result of random blurring with a Gaussian PSF at $\sigma_{\rm PSF} = 3$ pixels and adding a random Poisson background with the mean level $\gamma = 100$ counts/pixel (simulation number *t8*_13). The ML-restoration of the blurred image is shown in Figure 1(c); the non-negativity of the estimate was considered as a sole *a priori* information about the object. We see a hardly oscillating curve instead of a smooth original, and such an effect is typical not only of the ML-estimation, but of other approaches to the inverse problems as well, unless special precautions are introduced to suppress the instability.

It is known for a long time (see, e.g., Phillips, 1962) that the instability of inverse solutions is caused by the fact that a restoration method "tries to explain" details of the observed image, including random fluctuations at all scales. As far as a very large deviation in the object space is required to produce a very tiny fluctuation

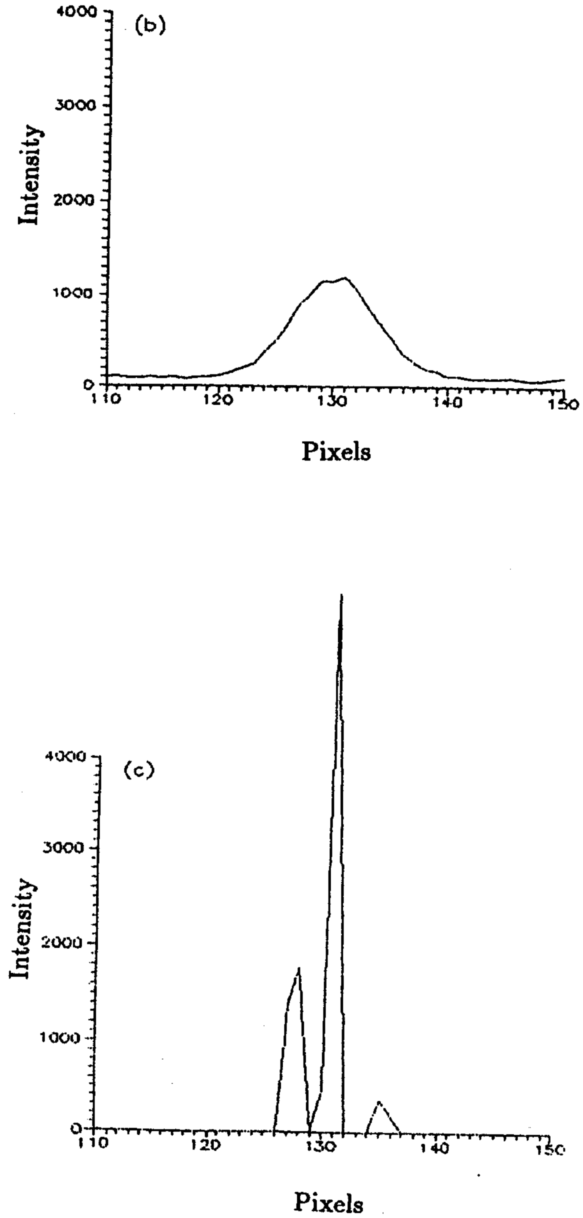

Figure 1 (continued)

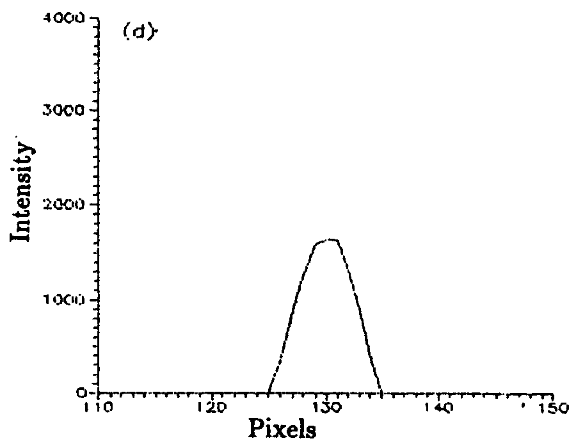

Figure 1 (continued)

in the image space under the smoothing process, the estimate of the object is so broken. It should be noticed that large estimate oscillations are random only to such an extent that fluctuations of the image are random.

Thus, one should take into account under restoration not all but only statistically significant details of the observed picture. This goal is attained in different ways by various image restoration methods (see Section 8). Usually the investigator minimizes some quadratic "misfit" between the data and the model to a statistically insignificant level, and then chooses the smoothest solution (in some definite sense) from all "feasible" solutions (Phillips, 1962; Tikhonov, 1963; Twomey, 1965; Frieden, 1972, 1979; Bryan and Skilling, 1980; Skilling and Bryan, 1984). The misfit can be treated as the χ^2-statistics (Ables, 1974; Lucy, 1974). Using the χ^2-test leads to some technical problems like the subjectivity of a bin set, but the main difficulty is that the universal measure of deviation in the χ^2-test does not account completely for a concrete shape of the probability density $f(N|S)$.

The purpose of this paper is to discuss a general criterion to search for an *efficient* estimate in the discussed above strict meaning. An explicit formulation of this intuitively evident requirement allows to reveal the insufficiency of the ML-approach and to use some new ways to increase the accuracy of the solutions.

2 THE IMAGE RANDOMNESS TEST

Let us consider for simplicity the case when the imaging system is a linear one, and the event statistics follows the Poisson law (see a detailed discussion in Terebizh, 1990a; Snyder, 1990; Terebizh *et al.*, 1991). Then the mean number of image counts at pixel j can be written as follows:

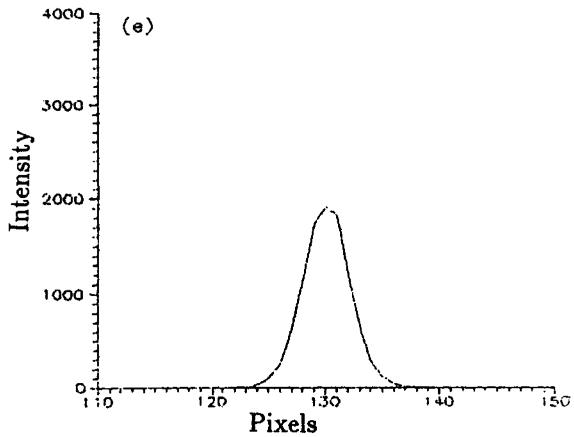

Figure 1 (continued)

$$\lambda_j(S) = \sum_{k=1}^{n} h_{jk} S_k + \gamma_j, \quad j = 1, 2, \ldots, m, \tag{3}$$

and the observed image counts $\{N_j\}$ are simulations of mutually independent Poisson random variables $\{\zeta_j\}$ with the corresponding mean values $\{\lambda_j(S)\}$. We will call, for brevity, the vector $\lambda(S) = \{\lambda_j(S)\}$ the *mean projection* of the object S onto the image space; then the image N should be considered as one of the possible *random projections* of the object S. The analogous notion of the projections can be used for both the true object S and any its estimate S^*.

It was assumed in (3) that all pixels of the detector have equal sensitivities. The effects of detector inhomogeneity together with other effects which are important for real data processing were discussed consistently by Snyder *et al.* (1993).

Let us formulate the Image Randomness Test (IRT) as a certain requirement that continues points 1–5 in Introduction:

6. Only such estimate $\tilde{S}(N)$ of the object S are feasible for which the observed image N cannot be distinguished statistically from the typical image simulations corresponding to $\tilde{S}(N)$.

In the case considered we should specify the IRT as follows: *only such estimates $\tilde{S}(N)$ of the object S are permitted, for which the observed image counts $\{N_j\}$ can be considered as a set of statistically independent one-dimensional Poisson simulations with mean values $\{\lambda_j[\tilde{S}(N)]\}$*. We denote the estimates satisfying the IRT by tilde.

One may conventionally decompose the IRT into the requirement of the independence of the random variables that give rise to the image N (the *white noise test*) and the requirement of the *Poisson distribution* for these variables with the mean value vector equal to the mean projection $\{\lambda_j(\tilde{S})\}$ of the estimate $\tilde{S}$. The scheme in Figure 2 clarifies the IRT.

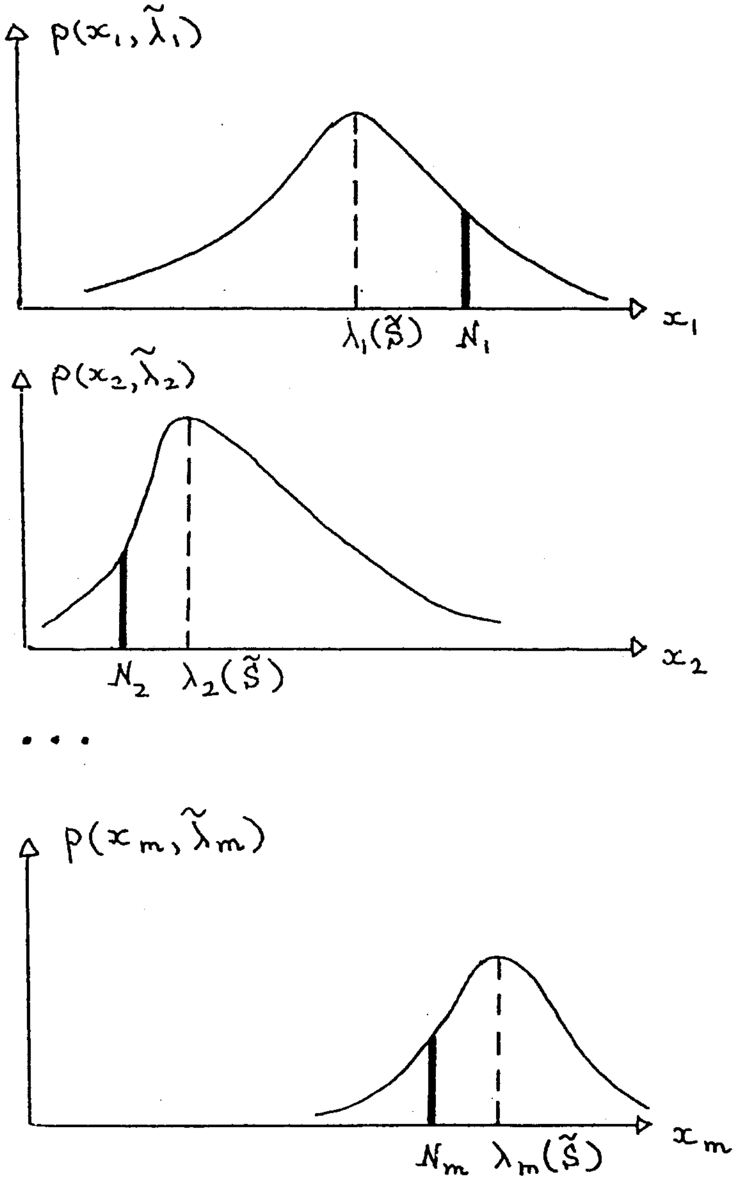

Figure 2 The marginal probability densities $p[x_j, \lambda_j(\tilde{S})]$ for some object estimate $\tilde{S}$ (thin lines) and really observed intensity counts N_j at different pixels. Thick bars represent the marginal likelihoods $p[N_j, \lambda_j(\tilde{S})]$.

3 SHOULD THE LIKELIHOOD BE MAXIMUM

Since Fisher (1912) introduced the Maximum Likelihood (ML) principle as a special estimating method for unknown parameters (in an implicit form, it was used as early as in the XVIII century), the ML-estimates became the most studied and widely used ones. It was mentioned above that, under some conditions, the ML-estimates coincide with the boundary estimates. Especially strong results concern an asymptotic domain when the investigator has a large number of simulations of the same random variable. Unfortunately, such situations are rare in image restoration; we usually have only one image of the object. For this reason, we concentrate on the case of a single frame, though multi-frame restoration does not introduce any principal difficulties.

If some set of one-dimensional patterns is used for the parameter estimation (in the present context, the set of intensities in different pixels), the likelihood L is a product of the marginal distribution densities:

$$L \equiv f(N|S) = \prod_{j=1}^{m} p[N_j, \lambda_j(S)], \tag{4}$$

where the marginal densities are now defined by the Poisson law:

$$p(k,\lambda) = e^{-\lambda}\frac{\lambda^k}{k!}, \; k = 0, 1, \ldots; \; \lambda \geq 0. \tag{5}$$

Equations (2)–(5) completely define, in principle, the ML-estimate $\tilde{S}(N)$. The maximization in (2) is a separate problem which is not considered here.

Let us assume for a while (see Figure 2) that the numbers of image counts which are larger or less than the corresponding mean values are equal to one another within the statistical accuracy (of course, this fact is insufficient to claim that the sample consists of a set of mutually independent random variables, but it does not complicate the problem just from the start). Then we have to check only the agreement of the deviations $r_j \equiv N_j - \lambda_j(\tilde{S})$ with the Poisson law. For this purpose one can use the set $\{r_j\}$ itself, but it seems more adequate to consider a product of the marginal probabilities $p(N_j, \lambda_j)$. Indeed, if $\{N_j\}$ are situated near their mean values within natural fluctuations, the product will be large. At the same time, excessively strong fluctuations will result in decreasing the product, i.e., the likelihood L.

Thus, being a function of the random array N, the likelihood itself is a random variable (statistics). Its expected value should be not far from the maximum value $L_{\max}(N) = f(N|\tilde{S})$, but it is unreasonable to require the observed value L to be strictly equal to $L_{\max}$. Just this requirement suggests that an estimate of the object must fit all details of the image, statistically significant or not, and, consequently, it will be almost definitely strongly oscillating. A much more attractive requirement is that the observed value L is equal to its expectation value:

$$f(N|\tilde{S}) = \mathbf{E}[f(N|\tilde{S})], \tag{6}$$

or deviates from the mathematical expectation by a statistically insignificant value. The function $\mathbf{E}[f(N|\tilde{S})]$ is given in Section 4 and Appendix A.

Like equation (2), which defines the maximum likelihood estimate, equation (6) defines the *mean likelihood estimate* $\tilde{S}(N)$. Some examples of such estimates can be found in Appendix B. It can be seen from these examples that, for simple one-dimensional distributions, $\tilde{S}$ is unique and efficient. At the same time, for multidimensional distributions, which are the most interesting for us, the requirement (6) defines not a single estimate but some class of them. Therefore, when the likelihood L is increased, one does not have to attain the maximum point $L_{\max}$, but it is better to stop after having reached the "feasibility layer", i.e., a certain value which satisfies equation (6) within a natural statistical deviation.†

4 CONNECTION WITH INFORMATION THEORY

There are two notions of *information* in the probability theory; the first definition was introduced by Edgeworth (1908, 1909) and Fisher (1922) for the estimation of parameters, and the second one, by Shannon (1948, 1949) for the needs of communication theory. The Fisher *information matrix* $I_{ik}(S)$ plays the main role in the expression for the lower scatter boundary of any estimate, and just with the Fisher information matrix is connected the name of the corresponding inequality (note, by the way, that along with the term *information inequality* there are widely used the terms *Rao-Cramer inequality* and *Freche inequality*). There were a few attempts to connect the image restoration and information theories; as one can see below, our way has some specific features. To distinguish the notation I for Fisher's information used in our previous publications and the information by Shannon, we denote the latter by J.

Let us remind briefly the main notions. Assume that τ is a continuous random variable with a uniform distribution on the interval $[0,1]$. If we measure information in *bits*, then the information contained in the communication "the simulation of τ is in the interval $[x, x+\varepsilon]$" will be $J = -\log_2 \varepsilon$. This value corresponds simply to the number of the first digits in a binary representation of τ which one should communicate to know its position with the accuracy of ε (e.g., Wiener, 1961). Consider then a discrete random variable ξ which can take the values $0, 1 \dots k, \dots$ with probabilities $p(0), p(1), \dots, p(k), \dots$, whose sum equals unity. We may imagine simulating ξ as simulating τ in the unit interval that includes all subintervals of the lengths $p(0), p(1), \dots$. τ being in the interval $p(n)$ is equivalent to the event that the simulation of ξ is equal to n. As stated above, one has the information $J_n = -\log_2 p(n)$ about the simulation of τ. It is more convenient to pass from

†Generally speaking, one should introduce the probability density $\varphi_L(x)$ of the statistics L and some significance level α to obtain a first type error in the sense of a standard pure significance test, and the corresponding critical value L_α. Then the *critical estimate* $\tilde{S}_\alpha$ is calculated instead of the *mean likelihood estimate*, as usually. Since the Shannon information J considered below is a function of the likelihood L, the same way is applicable to the estimation with J. We will return to this more complicated approach later.

binary logarithms to natural ones; the corresponding unit of information is nat $= \log_2 e \simeq 1.443$ bit.

If the intensity count at the pixel j is N_j, then the amount of information resulting from this pattern is given by

$$J(N_j, \lambda_j) = -\ln p(N_j, \lambda_j), \tag{7}$$

where $p(k, \lambda)$ is now defined by the Poisson law (5). According to our model, the counts in different pixels are mutually independent random variables, therefore the information due to the whole image simulation is

$$\begin{aligned} J(N, \lambda) &= \sum_{j=1}^{m} J(N_j, \lambda_j) = -\sum_{j=1}^{m} \ln p(N_j, \lambda_j) \\ &= -\ln \prod_{j=1}^{m} p(N_j, \lambda_j) = -\ln f(N|S) = -\ln L. \end{aligned} \tag{8}$$

Thus, the complete information in the image N is the logarithm of the likelihood L taken with the opposite sign.

It is clear now that the arguments of Section 3 in favour of the mean, but not the maximum, likelihood are equivalent to the fact that, when considering a typical simulation, we expect to get not the minimum, but only some close to the mean amount of information. Thus, the mean likelihood principle can be treated also as a requirement of equal values of sample information in the image and its entropy.

Note, by the way, that it is convenient to keep in the expressions for information and entropy the terms that are not dependent on the object; we then can measure information in absolute units, and to compare different images.

As to the mean value of information, it is equal, according to the definition by Shannon, to the *entropy* H of a random variable:

$$H(\lambda) \equiv \mathbf{E}[J(N, \lambda)] = \sum_{j=1}^{m} H_p(\lambda_j), \tag{9}$$

where

$$H_p(\lambda_j) = -\sum_{k=0}^{\infty} p(k, \lambda_j) \cdot \ln p(k, \lambda_j) \tag{10}$$

is the Poisson entropy that corresponds to one pixel. Exact and asymptotic expressions for $H_p(\lambda)$ are given in Appendix A.

Similarly to the above relation, one may require approximately equal values of the sample information and the entropy not only for the whole image, but also for different subsystems of pixels, in particular, for adjacent parts of the image. We call for brevity this requirement the *local mean information principle*.

Unlike the ML-estimate, the MI-estimate is non-unique for a multi-dimensional case. We simply narrow the class of "feasible" (in terms by Skilling and Bryan, 1984) solutions by subsequent requirements like the MI. The above local MI-principle also narrows this class. Other requirements may include a strict estimate of the total object flux, the white noise test, and, of course, the available *a priori* information.

Let us stress, in order to avoid misunderstandings, that we do not discuss here any maximization of the entropy, and the meaning of entropy in our case is different from that in the well-known maximum entropy method.

5 THE WHITE NOISE TEST

It was supposed above that before testing data with the aid of the statistics L, an independence was checked for the set of random intensities which generates the observed image $\{N_j\}$. L itself can say nothing about the "mixing" of the deviation set $\{r_j\}$. It is possible, for example, to shift $\{N_j\}$ in such a way that all new counts are on the same side from the corresponding mean values, but L does not change its value at all (see Figure 2). Evidently, insofar as the likelihood is concerned, the new sample will be as feasible as an old one, but the new sample is completely unsatisfactory in terms of the white noise test.

We come to the conclusion that L alone is an insufficient statistics for the IRT. Nevertheless, as a product of marginal probabilities, it accounts for some important properties of the observed image, so it is possible to use L as earlier, but this time not in the maximum sense and together with some white noise statistics.

Another way is to find out a single statistics that simultaneously checks both the mixing and the Poisson nature of the sample. As a matter of fact, the considered problem is the same as the well-known problem of testing the sequences of random numbers produced by a computer. There are powerful relevant methods, but it is out of our possibilities to consider them here.

It should be noticed that the above general testing methods are extremely labour-consuming. Since image restoration makes high demands to computers as well, it is reasonable to consider briefly some simple white noise tests which can be added to the likelihood statistics to check the IRT.

The sign test. Perhaps, in the case when $\lambda_j \gg 1$, the simplest test compares the number of positive and negative deviations between $\{r_j \equiv N_j - \lambda_j\}$. Let ν_m be the number of non-negative deviations in a sequence of length m. The mean value and the variance of ν_m for a completely random sequence (null-hypothesis) are $m/2$ and $m/4$, correspondingly, so, for a rather long image the statistics

$$\theta_1 = \frac{2\nu_m - m}{\sqrt{m}} \tag{11}$$

is distributed approximately as a Gaussian random variable with an integral distribution function,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-t^2/2)\, dt. \tag{12}$$

Therefore, the quantity

$$q_1 = 2[1 - \Phi(|\theta_1|)] \tag{13}$$

is the significance level for a two-sided test, that is for too little q_1 (say, $q_1 < 0.05$) one should reject the null-hypothesis.

For rather small λ_j, the asymmetry of the Poisson density can be taken into account by calculating, instead of r_j, the following deviations:

$$\omega_j = \sum_{k=1}^{N_j} p(k, \lambda_j) - 1/2. \tag{14}$$

The number of the series test. This test is as simple as the previous one. Let us call the series a sequence of deviations r_j (or ω_j) of the same sign. If γ_m is the number of series in a sequence of $m/2$ non-negative and $m/2$ negative deviations, then we expect for a white noise: $\mathrm{E}(\gamma_m) = m/2$, and for the variance, $\mathrm{D}(\gamma_m) = m/4$. Therefore, now we may use the statistics

$$\theta_2 = \frac{2\gamma_m - m}{\sqrt{m}}, \tag{15}$$

which is quite analogous to θ_1. The significance level is now $q_2 = 2[1 - \Phi(|\theta_2|)]$.

The distribution of the series lengths. According to the above point, the mean series length is equal to 2 when negative and positive deviations are equally probable. At the same time, there are series of different lengths $n = 1, 2 \ldots$ in a white noise; their probabilities are 2^{-n}. This relation can be used to test the observed deviations, and a restoration algorithm can proceed, for example, from removing excessively long series.

Further information concerning the testing of random sequences can be found in the special literature, e.g., Kendall and Stuart (1966, 1969), Yermakov and Mikhailov (1982).

6 THE ACCURACY AND RESOLUTION OF THE RESTORED OBJECTS

Evidently, the accuracy and the resolution are connected notions that are interesting from both practical and theoretical points of view. The information inequality gives an explicit expression for the natural accuracy limit (Terebizh, 1991), but the calculations are very complicated for concrete situations. For that reason we investigated the limiting resolution for any object shape both by numerical simulations and in the frame of the pattern recognition theory (Terebizh, 1990b, 1993). Some particular simplest cases will be considered in this Section; they can be used then as a starting point in studying more complex objects and for estimating the quality of the restored images in the course of their processing.

It is well known that the notion of the "limiting resolving power" has no unique definition; it depends on a concrete situation, on *a priori* information, etc. As regards the possibility to describe the resolving power by a single parameter, we can best refer to Wetherell (1980): "The field of image quality analysis, in common with other fields of intellectual endeavor, is beset by an affiction the present author likes to call "*unimania*" – the belief that highly complicated processes can be compared fully and accurately using a single one-real-number merit function. In optics, this takes the form of overreliance on merit functions such as Strehl definition, rms wave-front error, and limiting resolution".

Perhaps, in image restoration theory one can obtain a complete information concerning the resolving power of some imaging system only by calculating the *Frequency Restoration Function*, which is similar to the *Modulation Transfer Function* in optics, and one can describe by a single parameter only the simplest situations like those studied below.

Let us consider first case A, when the limiting resolution $\rho_{\min}$ means a minimum shift of the object as a whole, which can be found from its blurred and noisy image (this is a standard problem for spectroscopy and astrometry). The probabilities of the so-called first and second type errors should be given in advance; we accept their values of the order of 0.05–0.20. Denote by S the total flux of the object, by B the background flux inside the blurred image, and by

$$\Psi = \frac{S}{\sqrt{S+B}} \tag{16}$$

the signal-to-noise ratio for the whole image. Then it can be found from general expressions that the ratio of $\rho_{\min}$ to the PSF width Δ is given approximately by $\rho_{\min}/\Delta \simeq 0.5 \cdot \Psi^{-1}$ (case A). This relation is shown schematically in Figure 3.

Consider then case B, when two alternative objects are a single point-like source of flux S, and a double source with point-like components of flux $S/2$ each. This is the limiting resolution problem in a classic Rayleigh's (1964, p.420) meaning. Now $\rho_{\min}$ is the least separation between the components which can be detected at a given reliability level from the blurred and noisy image, when a single star-like source is the only alternative. Under this definition, the "resolution" is not so high as in the case A. Avoiding a less significant dependence on the PSF shape, we may write approximately: $\rho_{\min}/\Delta \simeq 2 \cdot \Psi^{-1/2}$ (case B).

The last of the particular cases (C) concerns the ability to distinguish a double object with point-like components and an extended object with a Gaussian brightness distribution. Denote by S the flux of each object and by ρ, the separation of the components; let $\rho/2$ be the standard deviation of the Gaussian density. The definition of $\rho_{\min}$ is similar to that in case B. It follows from equation (30) by Terebizh (1990b) that $\rho_{\min}/\Delta \simeq 1.5 \cdot \Psi^{-1/4}$ (case C). If the background level is equal to zero, then the signal-to-noise ratio is $\Psi = S^{1/2}$, and we have for case C: $\rho_{\min}/\Delta \simeq S^{-1/8}$. The latter relation was derived by Lucy (1992a,b).

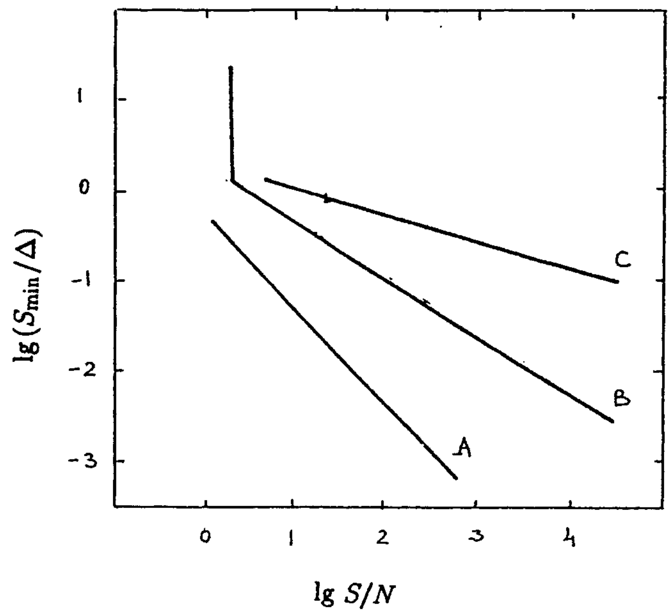

Figure 3 The [limiting resolution – S/N ratio] relation for an object's shift detection (A) and double structure detection when an alternative object is a single star (B) or a Gaussian extended source (C).

7 NUMERICAL SIMULATIONS

Let us return to the model case *t8* illustrated in Figure 1, and try to restore the same blurred image *t8_13* under the mean information requirement. The results of two independent algorithms, which started from the same initial approximation, are shown in Figures 1(d) and 1(e). The calculations were performed using the methods of multidimensional constrained optimization (e.g., Bertsekas, 1982).

First, we may infer that the MI-estimates are much smoother than the ML one. Just this stabilization of the inverse solution was expected on the basis of the above arguments. The fact that the MI-estimates are somewhat dependent on the initial approximation and the method of restoration employed was also expected, because now we have a "layer" of feasible solutions instead of a unique ML-solution. Further constraints (the white noise test, *a priori* information, etc.) should be used to narrow the class of MI-solutions; it is natural that we studied first a pure effect of the mean information requirement.

Similar results follow from processing other simulations of the object *t8*. Two of them, *t8_1* and *t8_13*, are compared in Figure 4 with respect to variations of information and entropy. Both image simulations (say, N_1 and N_{13}) give the values of the information $J(N_1, S)$ and $J(N_{13}, S)$, respectively, within $\pm$ one standard deviation from the mean value, that is the Shannon entropy $H(S)$, but $J(N_1, S)$ is

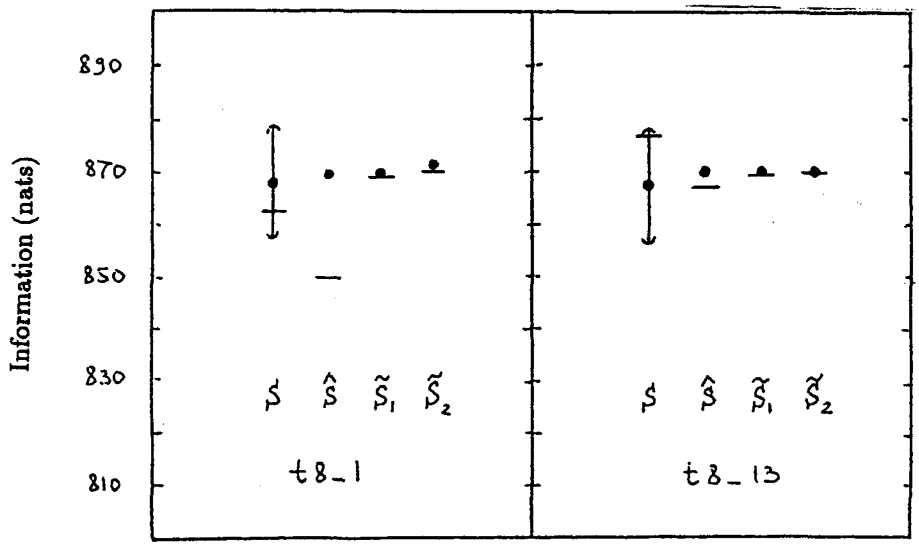

Figure 4 The variations of entropy (•) and information (–) for the simulations $t8\_1$ and $t8\_13$ which were restored using different methods. The true object (S), the ML-estimate ($\hat{S}$), and two independent MI-estimates ($\tilde{S}_1, \tilde{S}_2$) are shown. Bars correspond to ± one standard deviation of information in a random image pattern.

less than the entropy, and $J(N_{13}, S)$ is greater than $H(S)$. One can expect that the ML restoration decreases significantly the estimate information (i.e., increases the likelihood L), so the final difference between the information and the entropy can be very large (as in the case of $t8\_1$). For MI-estimates, by definition, this difference should be less than the standard deviation. Since the entropy of a Poisson random variable is a slow function of λ (see equation (A8) and Figure 5), the variations of the entropy are rather small. At a first approximation, we may consider the entropy even as a fixed value $H[\lambda(S)] \simeq H(N)$. On the contrary, the variations of the information are quite large . For the MI-esttimates the information increases if the initial information value is less than the entropy, and decreases in the opposite case. Thus, to obtain a stable solution one sometimes has to reduce the likelihood L.

A more extensive investigation of the MI-estimation along with the white noise test will be published elsewhere.

8 DISCUSSION

As a matter of fact, the Poisson distribution of image intensities is a consequence of *a priori* information which can be either inapplicable to the considered case, or simply not available for the investigator. Then we have to use a general scheme which does not assume any knowledge of photo-event statistics (Terebizh, 1990a).

Strictly speaking, the intensities at different pixels are in this case dependent on one of other random variables. Nevertheless, the general expression for the probability density $f(N|S)$ does not differ drastically from the Poissonian one, so even for a rather small total flux the numerical results are very close to each other. For that reason the above consideration is valid, in practice, for a wider class of conditions as well.

The same is true for the linearity requirement of the image forming system.

As mentioned above, the instability phenomenon is typical of inverse problems. It is interesting to compare the way proposed here with that in the now most widely used approaches: the Maximum Entropy (MEM) and Regularization methods.

Both methods define the image formation model by an equation which is similar to (3):

$$N_j = \sum_{k=1}^{n} h_{jk} \cdot S_k + \xi_j, \quad j = 1, 2, \ldots, m, \tag{17}$$

but now it contains not the *mean*, but *random* values: the observed image N and the background ξ. It was shown earlier (Terebizh, 1991, 1992) that equation (17) is not adequate to a real stochastic model of the image formation (in particular, it does not account for the photon noise), so we consider here only a part of the whole problem that concerns the instability phenomenon. It is clear that any "good" solution should give a not too large misfit vector

$$\delta_j = N_j - \sum_{k=1}^{n} h_{jk} S_k, \tag{18}$$

and the total normalized misfit

$$D(N, S) = \sum_{1}^{m} \delta_j^2 / \sigma_j^2, \tag{19}$$

where σ_j^2 is the datum variance in the j-th pixel.

The first version of MEM (Frieden, 1972) proceeds from the minimization of (18); the version widely used now, proposed by Bryan and Skilling (1980) and Skilling and Bryan (1984), requires that $D(N, S)$ should be close to that expected for a random simulation value, i.e., a requirement like the mean likelihood condition. This requirement itself defines only some domain in the object space which contains all feasible solutions. The choice of a real solution in this domain is based on the maximization of an "entropy" expression:

$$W(S) = -\sum_{1}^{n} (S_k/S) \ln(S_k/S), \quad S = \sum_{1}^{n} S_k, \tag{20}$$

or some similar expression (there are several alternative versions).

Of course, the choice of a quadratic measure for the misfit is not as natural as the likelihood (4), but a much more important step is the optimization of some functional like the "entropy", which is based on intuitive reasons. Just here subjective requirements are introduced in an implicit way (Terebizh, 1991).

In the regularization method (Tikhonov, 1963; Tikhonov and Arsenin, 1977) the misfit $D(N, S)$ is minimized under the condition that the solution power:

$$P = \sum_{1}^{n} S_k^2, \tag{21}$$

is less than some given value. This is equivalent to the minimization, with respect to S, of the functional

$$F(N, S, \alpha) = D(N, S) - \alpha \cdot P(S), \tag{22}$$

where α is the Lagrange multiplier. Similarly to the MEM, one defines here some feasible domain, and then chooses the solution with a least power instead of the largest-"entropy" one. If the power is really known from *a priori* information, this procedure seems quite natural, but this case should be considered as an exception to the rule in practice. Instead of the power, some other stabilization functionals can be used (Phillips, 1962).

In connection with a danger to introduce at some step subjective motives, one should clearly distinguish the general approach to inverse problems from accounting for *a priori* information.

It seems that this accounting can be done in a natural way only in a statistical approach, which introduces a stochastic model of the image formation instead of equations like (17); the inverse problem is then formulated as an estimation problem for an unknown set of parameters. One can incorporate the available information to a general scheme at different levels. For example, the Poisson distribution appears when the explicit law (5) for probabilities has to be defined, and the non-negativity of the estimate searched for is accounted for by considering a corresponding domain in the object space.

If an object S is known to be randomly chosen from some ensemble with the probability density $w(S)$, one may use the Bayes (1763) approach, and to introduce a two-dimensional density $w(S) \cdot f(N|S)$. It should be stressed that the Bayesian way is allowed only if the object was really chosen randomly from the given ensemble; this situation is extremely rare in practice.

As shown above, the IRT uses some *inner* resources to stabilize the inverse solution. If we have *a priori* information about the object (besides the non-negativity, of course) of both deterministic and stochastic nature, the quality of the solution will be higher. The nature of such information can be very diverse. In particular, sometimes we may consider as known the upper boundary of the spatial frequency interval where the bulk of power is concentrated. The limitation of the total power is also possible. Even stronger is the strict requirement of the smoothness of the desired solution expressed as a bound on second-order deviations. Just this requirement was introduced in a pioneering investigation of Phillips (1962).

Acknowledgements

We are grateful to O. K. Cherbunina and Yu. G. Cherbunin for valuable discussions.

References

Ables, J. G. (1974) *Astr. Astrophys. Suppl.* **15**, 383.

Bayes, T. (1763) *Reprint: Biometrika* **45**, 293, 1958.

Bertsekas, D. P. (1982) *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, New York.

Borovkov, A. A. (1984) *Mathematical Statistics* (in Russian), Nauka, Moscow.

Bryan, R. K. and Skilling, J. (1980) *Mon. Not. RAS* **191**, 69.

Edgeworth, E. Y. (1908) *J. Roy. Statist. Soc.* **71**, 381; 499; 651.

Edgeworth, E. Y. (1909) *J. Roy. Statist. Soc.* **72**, 81.

Fisher, R. A. (1912) *Messenger of Mathematics* **41**, 155.

Fisher, R. A. (1922) *Phil. Trans. Roy. Soc. London* **A 222**, 309.

Frieden, B. R. (1972) *J. Opt. Soc. Amer.* **62**, 511.

Frieden, B. R. (1979) In: *Picture Processing and Digital Filtering*. ed. T. S. Huang, Springer, Berlin, 179–249.

Helstrom, C. W. (1969) *J. Opt. Soc. Amer.* **59**, 164.

Helstrom, C. W. (1970) *J. Opt. Soc. Amer.* **60**, 659.

Jaynes, E. T. (1957a) *Phys. Rev.* **106**, 620.

Jaynes, E. T. (1957b) *Phys. Rev.* **108**, 171.

Kendall, M. G. and Stuart, A. (1966) *The Advanced Theory of Statistics*. Vols 2, 3. Griffin, London.

Lucy, L. B. (1974) *Astron. J.* **79**, 745.

Lucy, L. B. (1992a) *Astron. J.* **104**, 1260.

Lucy, L. B. (1992b) *Astron.Astrophys.* **261**, 706.

Phillips, D. L. (1962) *J. ACM* **9**, 84.

Rayleigh (Lord: Strutt, J. W.) (1964) *Scientific Papers*, Vol. 1, Dover, New York.

Shannon, C. (1948) *Bell Syst. Techn. J.* **27**, 379; 623.

Shannon, C. (1949) *Proc. IRE* **37**, 10.

Skilling, J. and Bryan, R. K. (1984) *Mon. Not. RAS* **211**, 111.

Snyder, D. L. (1990) In: *The Restoration of the HST Images and Spectra*, ed. R. L. White and R. J. Allen, Proc. of the Workshop at the Space Tel. Inst., August 1990, p. 56.

Snyder, D. L., Hammoud, A. M., and White, R. L. (1993) *J. Opt. Soc. Amer.* **A10**, 1014.

Terebizh, V. Yu. (1990a) *Astrofizika* **32**, 327.

Terebizh, V. Yu. (1990b) *Astrofizika* **33**, 409.

Terebizh, V. Yu. (1991) *Astron. Astrophys. Trans.* **1**, 3.

Terebizh, V. Yu. (1992) *Astrophys. Space Sci.* **193**, 269.

Terebizh, V. Yu. (1993) *Astron. Astrophys.* **270**, 543.

Terebizh, V. Yu. and Biryukov, V. V. (1990) *Astrofizika* **33**, 305.

Terebizh, V. Yu. and Biryukov, V. V. (1991) *Astrofizika* **34**, 233.

Terebizh, V. Yu., Cherbunina, O. K., Cherbunin, Yu. G., and Biryukov, V. V., (1991) *Astrofizika* **34**, 91.

Tikhonov, A. N. (1963) *Trans. Soc. Acad. Sci.* **153**, No. 1.

Tikhonov, A. and Arsenin, V. (1977) *Solutions of Ill-Posed Problems*, Winston and Sons, Washington, D.C.

Twomey, S. (1965) *J. Franklin Inst.* **179**, 95.

Wetherell, W. B. (1980) *The Calculation of Image Quality*, in: Applied Optics and Optical Engineering, ed. R. R. Shannon, J. C. Wyant, Vol. 8, Academic Press, New York, p. 176.

Whittaker, E. T. and Watson G.N. (1927) *A Course of Modern Analysis*, Cambridge Univ. Press, London, Ch. 12, Ex. No. 17.

Wiener, N. (1961) *Cybernetics*, New York-London, Ch. 3.

Yermakov, S. M. and Mikhailov, G. A. (1982) *Statistical Simulations* (in Russian), Nauka, Moscow.

Appendix A The information and entropy of the Poisson random variable

Let ξ be a discrete random variable that can be equal to $0, 1, \ldots$ with probabilities $p(0), p(1), \ldots$. According to Shannon's (1948, 1949) definitions, the information of the communication that ξ has value n is $J(n) = -\ln p(n)$, and the entropy H of the random variable ξ is equal to the mean value of information $J(n)$:

$$H = -\sum_{k=0}^{\infty} p(k) \cdot \ln p(k), \tag{A1}$$

where we use natural logarithms, and the corresponding information unit is nat $= \log_2 e \simeq 1.443$ bit.

For the Poisson distribution (5) we have:

$$J(n, \lambda) = \lambda - n \cdot \ln \lambda + \ln n!, \tag{A2}$$

so in this case the entropy is

$$H_p(\lambda) = \lambda(1 - \ln \lambda) + e^{-\lambda} \cdot \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \ln k!. \tag{A3}$$

The sum on the right-hand side can be transformed to an integral, if we use Euler's gamma-function, $k! = \Gamma(k+1)$, and then the well-known Binet representation for logarithm of the gamma-function (e.g., Whittaker and Watson, 1927):

$$\ln \Gamma(z) = \int_0^1 \left(\frac{t^{z-1} - 1}{t - 1} + 1 - z \right) \frac{dt}{\ln t}, \quad (\mathrm{Re} z > 0). \tag{A4}$$

The final expression is

$$H_p(\lambda) = \lambda(1 - \ln \lambda) - \int_0^1 \left(e^{-\lambda x} - 1 + \lambda x \right) \frac{dx}{x \ln(1 - x)}. \tag{A5}$$

The function $H_p(\lambda)$ is shown in Figure 5.

Equation (A5) can be easily expanded into series for small λ:

$$H_p(\lambda) = \lambda(1 - \ln \lambda) + c_2 \lambda^2 - c_3 \lambda^3 + c_4 \lambda^4 - \ldots, \tag{A6}$$

where the first coefficients are:

$$c_2 = \frac{1}{2} \ln 2 \simeq 0.346574, \quad c_3 = \frac{1}{6} \ln \frac{4}{3} \simeq 0.047947,$$

$$c_4 = \frac{1}{24} \ln \frac{32}{27} \simeq 0.007079. \tag{A7}$$

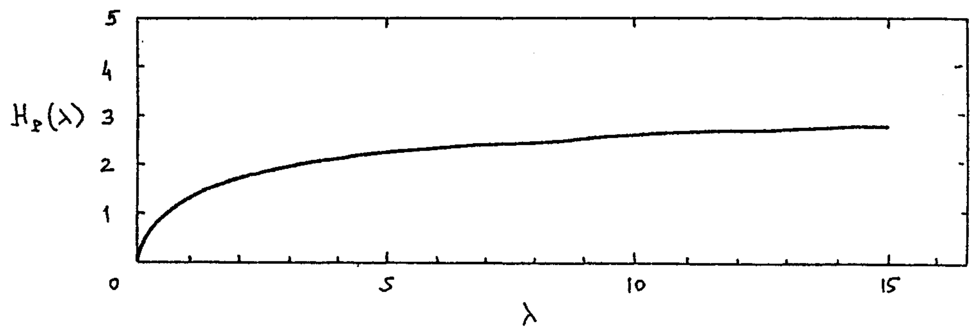

Figure 5 The entropy of Poisson random variable with the mean value λ.

Let us also give an asymptotic expression for large λ:

$$H_p(\lambda) = \ln\sqrt{2\pi e\lambda} + \ldots, \quad \lambda \gg 1, \tag{A8}$$

where the first term corresponds to the entropy of the Gaussian variable with variance λ (even at $\lambda = 3$, the first term provides the relative accuracy better than 2%).

For applications to parameters estimation according to the mean information value (Appendix B), one should know both Shannon's entropy and the variance of the information:

$$\mathbf{D}(J) = \mathbf{E}(J^2) - [\mathbf{E}(J)]^2, \tag{A9}$$

where $\mathbf{E}$ means the mathematical expectation value. An explicit expression for $\mathbf{D}(J)$ is too complicated, so we give here only an asymptotic one:

$$\mathbf{D}_p[\lambda] \simeq \frac{1}{2} + \frac{1}{4\lambda}, \quad \lambda \gg 1. \tag{A10}$$

Note that standard deviation of the Poisson information is practically constant for $\lambda \gg 1$, and equal to $1/\sqrt{2} \simeq 0.71$ nat, that is, nearly to 1 bit. Therefore, the set of m independent Poisson variables with $\lambda_j > 2$ has the standard deviation of Shannon's information equal approximately to $\sqrt{m/2}$.

Appendix B The mean information estimations

Let us consider one-dimensional random variable ξ with the probability density $f(x,a)$ which depends on unknown parameter a. Shannon's information which is connected with a simulation X is $J(X,a) = -\ln f(X,a)$, and its mean value, the entropy of ξ, is equal to:

$$H(a) = \mathbf{E}[J(\xi,a)] = -\int f(x,a)\ln f(x,a)\,dx. \tag{B1}$$

Let us define the Mean Information (MI) estimate of the parameter a as a solution of the equation

$$J(X,a) = H(a) \tag{B2}$$

with respect to a; we denote the solution by $\tilde{a}(X)$. This is, of course, some random variable. Consider a few examples of MI-estimates.

Exponential distribution. We have in this case: $f(x,a) = a^{-1} \cdot \exp(-x/a)$, $J(X,a) = \ln a + X/a$, $H(a) = 1 + \ln a$, and the MI-estimate is $\tilde{a}(X) = X$. It coincides with the ML-estimate; both estimates are unbiased and efficient.

The geometrical distribution is a discrete one:

$$f(k,a) = \frac{a^k}{(1+a)^{1+k}}, \quad k = 0,1,2,\ldots, \tag{B3}$$

where $a > 0$ is the mean value of ξ. Let n be a random simulation of ξ. Then the information is $J(n,a) = n \cdot \ln a - (1+n) \cdot \ln(1+a)$, the entropy $H(a) = -a \cdot \ln a + (1+a) \cdot \ln(1+a)$, and the MI-estimate is again equal to the ML-estimate: $\tilde{a} = \hat{a} = n$. Both estimates are unbiased and efficient.

The Gaussian distribution. Let us assume that the variance of ξ is known in the density

$$f(x,a) = (\sigma\sqrt{2\pi})^{-1} \cdot \exp[-(x-a)^2/2\sigma^2], \quad -\infty < x < \infty \tag{B4}$$

and one searches for an estimate of a. We have: $J(X,a) = (X-a)^2/2\sigma^2 + \ln(\sigma\sqrt{2\pi})$ and $H(a) = \ln(\sigma\sqrt{2\pi e})$, so there are two mean information solutions: $\tilde{a}_{1,2}(X) = X \pm \sigma$. Both are biased estimates, and it is needed to make the mean arithmetic value to have an unbiased MI-estimate.

The Poisson distribution. An analogous result follows for the Poisson density (5) which can be considered, for large λ, as an approximation to the Gaussian density with $\sigma^2 = a = \lambda$. If $\lambda \gg 1$, and a simulation of ξ is n, then $\tilde{\lambda}(n) \simeq n \pm \sqrt{n}$, and the relative bias is $\simeq \lambda^{-1/2}$. In the general case, the asymptotically unbiased and efficient estimate can be obtained by averaging the two MI-estimates.

The only purpose of these examples is to show the reasonable meaning of MI-estimates. A more extensive study of the MI-requirement on the estimates will be given elsewhere.