IOPscience iopscience.iop.org

Home Search Collections Journals About Contact us My IOPscience

The maximum entropy configuration in the case of randomly fluctuating constraints

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

1989 J. Phys. A: Math. Gen. 22 L483

(http://iopscience.iop.org/0305-4470/22/11/006)

View the table of contents for this issue, or go to the journal homepage for more

Download details:
IP Address: 129.252.86.83
The article was downloaded on 01/06/2010 at 06:42

Please note that terms and conditions apply.

J. Phys. A: Math. Gen. 22 (1989) L483–L488. Printed in the UK

LETTER TO THE EDITOR

The maximum entropy configuration in the case of randomly fluctuating constraints

R Lieu

Astrophysics Group, Physics Department, Imperial College, Prince Consort Road, London SW7 2BZ, UK

Received 27 January 1989

Abstract. An attempt is made to crystallise the meaning of the maximum entropy state, when a system is subject to constraints which carry random uncertainties. Such a state is well defined, and there exists a simple statistic which allows it to be determined unambiguously, provided it is assumed that one's state of knowledge of the system fluctuates in close resemblance to the quantum uncertainty principle. Without loss of generality, the senario is verified, for a variety of error-carrying constraints, by results from a large number of 'random die samples'.

Maximum entropy has been proposed as the only logically consistent method of inference, when available knowledge is insufficient (Jaynes 1957). Traditionally, such knowledge takes the form of some constraints, which are well defined, but which do not permit a complete specification of the system. For this class of problems the standard way of arriving at the maximum entropy solution, subject to the given constraints, is the Lagrange method of undetermined multipliers. However, modern developments in the fields of data analysis, quantum mechanics and statistical mechanics render it more customary to have information on all the variables of the system in question, although the 'insufficiency' arises because such information is derived from experimental conditions which are inevitably subject to fluctuations†.

The problem of finding the maximum entropy estimate in the presence of poorly defined constraints has been a matter of controversy for some time. A number of prominent researchers in the field (Burch *et al* 1983, Jaynes 1984) have agreed that the correct statistic to extremise is $S - \lambda\chi^2$, but the difficulty is in finding the appropriate value of λ (Lieu *et al* 1987a). Here we attempt to demonstrate that, in the majority of circumstances, λ assumes a unique value, which can be determined by a re-examination of the underlying axioms of maximum entropy theory.

Consider first the common problem of estimating a distribution of relative proportions (or probabilities) $\{p_i,\ i = 1, 2, \ldots, r\}$; $\Sigma p_i = 1$, subject to some prior knowledge $\{\bar{p}_i \pm \sigma_i,\ i = 1, 2, \ldots, r\}$, where $\{\bar{p}_i\}$ are the constraints and $\{\sigma_i\}$ are their associated random errors (not necessarily Gaussian). The rationale of maximum entropy then proceeds as follows. Imagine that there are many possible 'versions' of the distribution,

† See Erickson and Smith (1988). The following common physical situations serve to illustrate: (1) electronic gain fluctuations randomly shift the pixel positions of a microchannel plate, resistive anode photon detector; (2) during long spectroscopic observation of an astrophysical plasma, interim departures from thermal equilibrium cause variations in the relative population of each atomic level, according to the instantaneous mean temperature.

each produced by randomly throwing a large number of counts N into the r channels, and working out $p_i = n_i/N$, $i = 1, 2, \ldots, r$. The likelihood of forming a particular distribution $\{p_i, i = 1, 2, \ldots, r\}$ depends on the product of two quantities: (i) the number of distinct 'versions' which lead to this distribution; (ii) the probability of complying with the constraints when errors are taken into account. The essential mathematical factor for (i) is well known, namely $\exp(NS)$, where

$$S = -\sum_{i=1}^{r} p_i \log p_i. \tag{1}$$

The expression for (ii) is more subtle. It is *not*, as has been suggested (Jaynes 1984), proportional to $\exp(-\chi^2)$, where

$$\chi^2 = \sum_{i=1}^{r} (p_i - \bar{p}_i)^2/2\sigma_i^2 \tag{2}$$

for the following reasons. Quite apart from the fact that the experimental errors may not be Gaussian distributed, the final product $\exp(NS - \chi^2)$ is infinitely biased towards the entropy function, and thus renders the constraints obsolete.

In order to arrive at the correct factor for (ii), it must be realised that, when forming a version of the distribution, each random event of the N counts is a fictitious measurement (or probing) of the distribution. Such fictitious measurements *cannot* lead to a reduction in the error of the actual experiment. Therefore, we must introduce a likelihood function for the conceptual experiment, through which the probabilities $\{p_i\}$ at each repitition are allowed to *fluctuate* in a random way as governed by our state of knowledge. Such an undertaking distinguishes the present work from previous works by other authors. Although reminiscent of the quantum uncertainty principle, it is the only way to avoid absurd logical consequences.

The remaining argument is quite simple. If we are prepared to accept this new ground rule, then there will be N trial measurements performed on a system of r fluctuating probability variables $\{p_i = \bar{p}_i \pm \sigma_i, i = 1, 2, \ldots, r\}$. Each individual variable will formally be measured N/r times, and the mean will be distributed around $\bar{p}_i$ like a Gaussian, with standard deviation $\sigma_i(r/N)^{1/2}$, irrespective of the error distribution in the original constraint. The probability of agreement between the constraints and the distribution is proportional to $\exp(-N\chi^2/r)$, where χ^2 is given by (2). Assembling now the two factors (i) and (ii), mentioned prior to equation (2), it is found that the overall probability of forming a particular distribution is proportional to $\exp[N(S - \chi^2/r)]$. The most likely distribution is then obtained by maximising $S - \chi^2/r$, subject to the constraint $\Sigma p_i = 1$. Note that the relative scaling between S and χ^2 is now independent of N, as it should be.

Previous works on this problem (see Lieu (1988) and Lieu *et al* (1987b)) have erroneously arrived at the statistic $S - \chi^2$ (i.e. $r = 1$). Although the experimental uncertainties were handled in exactly the same spirit there, the number of 'measurements' on each p_i was assumed to be N, instead of the correct value N/r.

So far the discussion has been concerned with constraints on the individual probabilities. However, the same reasoning can be applied to any number of constraints, each involving any combination of the probabilities, provided that (i) the constraints deal with mutually exclusive sets of probabilities and (ii) the total number of constraints does not exceed r, the degrees of freedom of the system. The following illustrative examples, all of which concern the familiar 'loaded die', will serve to clarify these concepts, without loss of generality.

The first problem is one of estimating the probabilities of the individual faces of a normal six-sided die, given a constraint on the mean

$$\bar{i} = \sum_{i=1}^{6} i p_i . \tag{3}$$

If $\bar{i}$ is known precisely to be 4.5, the maximum entropy estimate of the p_i can be obtained via the method of Lagrange multipliers. The result is shown in the right-hand column of table 1(a). If $\bar{i}$ is known to within a margin of random error σ (the standard deviation), the maximum entropy estimate of the p_i would accordingly be obtained by maximising $S - \chi^2/r$, subject to $\Sigma p_i = 1$, where

$$S = \sum_{i=1}^{6} -p_i \log p_i \qquad \chi^2 = (\bar{i} - 4.5)^2/2\sigma^2 \tag{4}$$

and $r = 6$. Note that this χ^2 is *not* the same as that given in (2), because the available information here involves the single data value $\bar{i}$, and not the six individual probabilities. The resultant estimated distribution, for two values of σ, are shown in the right-hand columns of table 1(b) and 1(c).

Table 1. This six probabilities of the typical loaded die, subject to the constraint that the mean $\bar{i} = \Sigma i p_i$ must lie within a certain interval: (a) $4.4 < \bar{i} < 4.6$; (b) $4.25 < \bar{i} < 4.75$; (c) $4.0 < \bar{i} < 5.0$; (d) $1.5 < \bar{i} < 2.5$. In each case, the left-hand column, marked Data, gives the statistically averaged results of a large number of randomly generated samples, and the right-hand column, marked Theory, gives the maximum $S - \chi^2/r$ distribution. In (a) the theory is worked out in the limit $\sigma = 0$ (which is identical to using the Lagrange method of undetermined multipliers). This is because the standard deviation σ, though finite, is small; (b) also gives, for comparison, corresponding estimates obtained by maximising $S - \chi^2$, the previous statistic (Lieu 1988) which is now believed to be incorrect.

	(a)		(b)			(c)		(d)	
	Data	Theory	Data	Theory	Theory ($S - \chi^2$)	Data	Theory	Data	Theory
p_1	0.0630	0.0543	0.0650	0.0581	0.0550	0.0728	0.0677	0.409	0.436
p_2	0.0774	0.0788	0.0806	0.0826	0.0794	0.0903	0.0921	0.251	0.239
p_3	0.104	0.114	0.107	0.117	0.115	0.118	0.125	0.153	0.126
p_4	0.154	0.165	0.159	0.167	0.166	0.170	0.170	0.0939	0.0844
p_5	0.264	0.240	0.259	0.237	0.239	0.248	0.231	0.0575	0.0635
p_6	0.338	0.348	0.329	0.338	0.345	0.302	0.314	0.0352	0.0511
$\bar{i}$	4.490	4.50	4.455	4.456	4.492	4.335	4.347	2.245	2.254

In the limit $\sigma \to 0$ the $S - \chi^2/r$ procedure can be shown to yield results in complete agreement with those obtained by the Lagrange method of undetermined multipliers. When $\sigma > 0$ the validity of $S - \chi^2/r$ is demonstrated by computer simulation experiments. In a typical experiment, approximately 100 000 sets, each containing six positive numbers, arbitrarily adding to a total of unity, are generated†. From this 'universal set' of dice, it is always possible to select a 'subset' on the sole criterion that $\bar{i}$ must lie within a certain interval (which defines a 'square box' error distribution‡; the

† The computer routine works as follows: (i) first it generates five independent random numbers (each between 0 and 1); (ii) then their positions are marked on the number line; (iii) between 0 and 1 these numbers define six intervals, the lengths of which form a required set.

‡ The fact that $S - \chi^2/r$ can handle non-Gaussian random errors is an indication of its generality. In fact, it should not come as a surprise because of the central limit theorem; see earlier discussion in text.

standard deviation σ is related to the length L of the box by $\sigma = L/\sqrt{12}$). If an average over this subset (usually containing 10 000 loaded dice) is performed, one obtains a 'mean set' $\{p_i,\ i=1,2,\ldots,6\}$ which should represent the most typical loaded die consistent with the constraint on $\bar{i}$.

Simulated data for the case $L=0.2$ can be found in the left-hand column of table 1(*a*). Since σ is very small, $S-\chi^2/r$ is not distinguished from the Lagrange multiplier method, and indeed agreement between theory and data is reasonable. When σ is no longer small, the constraints $4.25<\bar{i}<4.75$ ($\sigma=0.144$) and $4.0<\bar{i}<5.0$ ($\sigma=0.289$) yielded data to be found in table 1(*b*) and 1(*c*). Here deviation from the $\sigma=0$ results is obvious, and $S-\chi^2/r$ can be seen to perform well. In particular, it predicts a value for $\bar{i}$ which agrees closely with that of the typical loaded die. The fact that $\bar{i}<4.5$ in both these cases is because nature has taken advantage of the error σ in its affinity for the state of maximum entropy. More precisely, the probability distribution tends to uniformity as much as the error would permit, which implies that $\bar{i}$ shifts to the equilibrium value of 3.5.

As further examples, table 1(*d*) shows the results, theory and experiment, of the case when $1.5<\bar{i}<2.5$ ($\sigma=0.289$). Again agreement between them is reasonable. Although the data mean value for $\bar{i}$ is 2.0, the typical loaded die has $\bar{i}$ closer to 3.5, namely $\bar{i}=2.25$. This is another clear example of entropy at work. Table 2 gives results for a 20-face die, with constraint $12.5<\bar{i}<13.5$, or $7.5<\bar{i}<9.5$. Owing to the large number of p_i, only the mean values of $\bar{i}$ are listed. Note the value of r in $S-\chi^2/r$ is now $r=20$, which illustrates the scaling with degrees of freedom of the system. Turning now to multiple constraints, tables 3 and 4 show the simulated data, and theoretical best estimates of the probabilities of the loaded die, given the information (and associated standard errors) on *two* of its probabilities, i and j, i.e. $p_i=\alpha_1\pm\sigma$, $p_j=\alpha_2\pm\sigma_2$. In applying $S-\chi^2/r$ to this situation, we define

$$\chi^2=(p_i-\alpha_1)^2/2\sigma_1^2+(p_j-\alpha_2)^2/2\sigma_2^2 \tag{5}$$

and $S=-\Sigma\, p\log p$, $r=$ number of faces as before. Table 5 shows the case when *three* of the six faces are quoted to within errors. In each of the above examples, $S-\chi^2/r$ explain the average behaviour of the random samples reasonably well. Table 6 illustrates the case of constraints with overlapping degrees of freedom. It concerns the maximum entropy solution, subject to information on one of the probabilities p_1 ($p_1=\alpha_1\pm\sigma_1$), and on the mean $\bar{i}$ ($\bar{i}=\alpha_2\pm\sigma_2$). A naive approach would maximise

$$S-\chi^2/r=\sum_{i=1}^{6}-p_i\log p_i-(p_1-\alpha_1)^2/12\sigma_1^2-(\bar{i}-\alpha_2)^2/12\sigma_2^2 \tag{6}$$

but unfortunately it leads to a distribution which is quite unlike the simulated data. As mentioned before, the correct application of $S-\chi^2/r$ must involve independent

Table 2. The value $\bar{i}=\Sigma i p_i$ of the most typical 20-face die (marked $1,2,\ldots,20$ on successive faces), given the constraint: (*a*) $7.5<\bar{i}<9.5$ ($\sigma=0.577$); (*b*) $11.5<\bar{i}<13.5$ ($\sigma=0.577$). Note the tendency towards the global maximum entropy state $\bar{i}=10.5$. Also shown, for comparison, are the results of maximum $S-\chi^2$.

		Data	Theory	Theory $S-\chi^2$
(*a*)	$\bar{i}$	8.864	8.839	8.520
(*b*)	$\bar{i}$	12.148	12.161	12.480

Table 3. The six probabilities of the typical loaded die, given that two of its faces are known to within random errors of the 'square box' type: $0.35 < p_6 < 0.65$ ($\sigma_6 = 0.0866$), and (*a*) $0.16 < p_1 < 0.24$ ($\sigma_1 = 0.0231$); (*b*) $0.07 < p_1 < 0.17$ ($\sigma_1 = 0.0289$); (*c*) $0.3 < p_5 < 0.5$ ($\sigma_5 = 0.0577$).

	(*a*)		(*b*)		(*c*)	
	Data	Theory	Data	Theory	Data	Theory
p_1	0.195	0.198	0.114	0.120	0.0587	0.0566
p_2	0.0923	0.0929	0.110	0.111	0.0585	0.0566
p_3	0.0919	0.0929	0.109	0.111	0.0581	0.0566
p_4	0.0911	0.0929	0.110	0.111	0.0589	0.0566
p_5	0.0919	0.0929	0.109	0.111	0.357	0.363
p_6	0.438	0.431	0.448	0.438	0.409	0.411
$\bar{i}$	4.106	4.083	4.335	4.296	4.824	4.845

Table 4. This table lists values of p_8, p_{20} and $\bar{i} = \Sigma i p_i$ for the most typical 20-sided loaded die which satisfies the constraints $0.07 < p_8 < 0.13$ ($\sigma_8 = 0.0173$), and $0.1 < p_{20} < 0.2$ ($\sigma_{20} = 0.0289$). As before, errors are random, and of the 'square box' type.

	Data	Theory	Theory $S-\chi^2$
p_8	0.0939	0.0952	0.0997
p_{20}	0.132	0.131	0.149
$\bar{i}$	11.197	11.209	11.373

Table 5. The six probabilities of the typical loaded die, given that three of its faces are known to within random errors of the 'square box' type: $0.13 < p_4 < 0.27$ ($\sigma_4 = 0.0404$), $0.3 < p_5 < 0.5$ ($\sigma_5 = 0.0577$), and $0.35 < p_6 < 0.65$ ($\sigma_6 = 0.0866$).

	Data	Theory
p_1	0.0368	0.0285
p_2	0.0368	0.0285
p_3	0.0364	0.0285
p_4	0.166	0.182
p_5	0.337	0.350
p_6	0.387	0.383
$\bar{i}$	4.891	4.945

constraints, i.e. constraints which deal with mutually exclusive sets of probabilities. In the present case, the correct procedure would be to decouple p_1 from $\bar{i}$, i.e. to maximise

$$S - \frac{\chi^2}{r} = \sum_{i=1}^{6} -p_i \log p_i - (p_1 - \alpha_1)^2/12\sigma_1^2 - (\langle i \rangle - \alpha)^2/12\sigma^2 \tag{7}$$

where

$$\langle i \rangle = \sum_{i=2}^{6} i p_i \qquad \alpha = \alpha_2 - \alpha_1 \qquad \sigma = (\sigma_2^2 - \sigma_1^2)^{1/2}. \tag{8}$$

Agreement between theory and data is now restored (table 6).

Table 6. The six probabilities of the typical loaded die, given knowledge of one of its faces, and of the mean, as follows: $0.1 < p_1 < 0.3$ ($\sigma_1 = 0.0577$); $4.0 < \bar{i} < 5.0$ ($\sigma_2 = 0.289$). The errors quoted are random errors, and of the 'square box' type. Note the correct procedure in applying $S - \chi^2/r$; see text.

	Data	Theory
p_1	0.151	0.160
p_2	0.0638	0.0552
p_3	0.0855	0.0872
p_4	0.127	0.138
p_5	0.223	0.217
p_6	0.349	0.343
$\bar{i}$	4.256	4.223

In conclusion, the statistic $S - \chi^2/r$ is demonstrated, via formal mathematical reasoning, and via computer simulation experiments, to be the central limit of any distribution function, which is subject to randomly varying constraints. It non-trivially extends the canonical meaning of entropy. In the course of its derivation, the assumption that our state of knowledge of the system fluctuates like the uncertainty principle was found to be inevitable. The statistic is expected to serve a growing need in the various disciplines of mathematical physics.

Note added in proof. The author now understands that two papers lending much support to the application of $S - \chi^2/r$ have been published by Brand and Le Cäer (1988) and Zhigunov *et al* (1988).

References

Brand R A and Le Cäer G 1988 *Nucl. Instrum. Methods Phys. Res* B **34** 272
Burch S F, Gull S F and Skilling J 1983 *Comp. Vis. Graph Im. Proc.* **23** 113
Erickson G J and Smith C R (eds) 1988 *Maximum Entropy and Bayesian Methods in Science and Engineering* vols 1, 2 (Deventer: Kluwer)
Jaynes E T 1957 *Phys. Rev.* **106** 620
—— 1984 *4th Ann. Workshop on Bayesian/Maximum Entropy Methods in Geophysical Inverse Problems, Calgary*
Lieu R 1988 *J. Phys. A: Math. Gen.* **21** L63
Lieu R, Hicks R B and Bland C J 1987a *Mon. Not. R. Astron. Soc.* **229** L49
—— 1987b *J. Phys. A: Math. Gen.* **20** 2379
Zhigunov V P, Kostina T B and Spiridonov A A 1988 *Nucl. Instum. Methods Phys. Res.* A **273** 362