Pick a light ray-any light ray

1990 J. Phys. A: Math. Gen. 23 2413

(http://iopscience.iop.org/0305-4470/23/12/022)

J. Phys. A: Math. Gen. 23 (1990) 2413-2420. Printed in the UK

Pick a light ray—any light ray

Adam D Helfer
Department of Mathematics, University of Missouri, Columbia, MO 65211, USA

Received 23 October 1989

Abstract. We study the statistics of distributions on the celestial sphere in a Lorentz-invariant setting. There is no Lorentz-invariant fiducial measure on the sphere, so the usual information-theoretic analysis cannot be applied. Nevertheless we show that information theory can be used to analyse the distributions and simultaneously to provide a fiducial measure on the sphere. The most important application is to the cosmic microwave background. We do not reach any unusual conclusions about this, but provide a new justification for an *ad hoc* feature of the standard analysis.

1. Introduction

One of the most significant advances this century in our understanding of the foundations of physics was the application of information theory to statistical physics (Jaynes 1957). Generally, the task of statistical physics is to predict the probability distribution of states of a physical system as a function of a few observable parameters. Information theory gives a general prescription for doing so. The central idea is that there is a quantitative measure, the information-theoretic entropy S_I, of the lack of information in any probability distribution. Maximising this entropy subject to the known constraints on the distribution leads to one's best (i.e. least prejudiced) guess for the distribution subject to the known data. This approach has had two sorts of consequences: first, it has given a broad justification for the usual partition function formalism independent of integrating the microphysical equations of motion or of arguments about weak coupling of the elements of an ensemble to each other or to an ambient heat bath; and second, it has allowed considerable generalisation.

In principle, the information-theoretic approach can be applied to any system for which the measure on the space of states (and of course the observables as functions of the states) are known. It is however not always a trivial matter to find the correct measure†. There are several ways it can be determined. In the original work of Shannon and Weaver (1949), the distributions considered were on finite sets of points; it was natural to weight each point equally. In classical statistical mechanics, the measure is defined on the phase space and is determined by the symplectic form (Poisson bracket). In some cases, the measure is the volume form determined by the metric on a Riemannian manifold. Lastly, the measure may be determined by invariance under a group action.

(In fact, in many physical applications, the measure is determined by invariance under a group action. For a finite set of points, the group is the group of permutations; for phase space, it is the group of canonical transformations; for a symmetric space, it is the group of isometries.)

† This is sometimes called the *reparametrisation problem.*

0305-4470/90/122413+08$03.50

Let us consider the situation generally. Assume a manifold X on which a Lie group G acts. We suppose that G acts *transitively*, i.e. that given any two points in X, there is at least one element of G taking the first to the second. It is well known mathematically that in this case there can be, up to an overall normalisation constant, at most one invariant measure. For example, X might be the circle and G its group of rotations; the measure is then $d\theta/2\pi$. Or X might be the plane and G the group of Euclidean motions; then the measure is just $dx\, dy$.

The second example raises an interesting point. If we had taken G to be the group of translations (that is, if we had excluded the rotations), that still would have been enough of a constraint to force the measure to be $dx\, dy$. (Because the translations act transitively on the plane.) In other words, the problem of finding an invariant measure may be *overdetermined.* It is just luck that an invariant measure exists and no inconsistency arises from this, or is it a manifestation of some deeper principle?

Other authors have realised that there is nothing mathematically to prevent inconsistencies from arising (Jaynes 1973). The prevailing attitude seems to be, though, that in realistic problems, with finite-dimensional X and G, an invariant measure always does exist. It is the purpose of this paper to give a counterexample, and to develop, in spite of the absence of an invariant measure, an information-theoretic statistical analysis for it. So we shall give an example of a physical system with a natural group action for which *no* invariant measure exists, and show that *information theory can be used to select a measure.* The set of states will be the set of light rays at a fixed event in space-time. These may be thought of as the points on the celestial sphere of an observer at that event; however, we shall *not* make a particular choice of observer, so the relevant group is the Lorentz group rather than the rotation group.

The most important application of this is to the cosmic microwave background. The existence and uniformity of this background are the primary evidence for the Big Bang cosmology. The fact that it is so nearly uniform has obscured a subtle question: how should the anisotropy best be measured? The difficulty is that there is no relativistically-invariant notion of a uniform distribution on the celestial sphere: if one observer sees the sphere uniformly illumined, another, boosted, observer will see an anisotropic luminosity density which is brighter in the direction of the boost and dimmer behind. For small boosts, the density is approximately the superposition of a uniform and a dipole distribution. This is what is observed for the microwave background. Its luminosity density contains a dipole moment corresponding to a boost of a few hundred km s^{-1} which is interpreted as the speed of the Earth relative to the Friedmann-Robertson-Walker background (Wilkinson 1986). If there were a frame in which the luminosity were uniform, then of course we would agree that that is the Friedmann-Robertson-Walker frame, that it is the correct one for measuring the isotropy, and that in fact the anisotropy is zero. But if no such frame exists (and this is the case), then there is an issue of principle as to which frame one should use to measure the anisotropy (Liang and Sachs 1980). Should one use a frame in which the dipole moment vanishes, or rather one in which the weighted combination of the higher multipole moments is minimised? This is a more primitive problem than that of choosing a statistic—such as RMS departure from mean—for the anisotropy. We shall see that, according to information theory, the correct thing to do is to work in a frame in which the dipole moment is zero.

Astronomers have long realised that analysis of the cosmic microwave background depends on the Lorentz frame used, and have subtracted the dipole moment on this account. Because of the small boost involved, this is for practical purposes equivalent to working in the frame in which the dipole moment is zero. Our results provide an independent and less ad-hoc justification for doing so. Also the arguments in this paper apply even to very inhomogeneous distributions, when it is less clear *a priori* that going to the zero-dipole-moment frame is the correct thing to do. If in the future neutrino astronomy were developed to a state where we could measure the cosmic neutrino background and it were found to be anisotropic, our analysis would be the correct one for quantifying the anisotropy. Similar comments apply to gravity waves.

A careful treatment of the microwave, neutrino or gravity background would require discussion of a number of technical points (the fact that not just the direction, but the energy of the particles is detected, and other effects). For the present, though, we are interested rather in the issue of principle, so it will be helpful to consider a simpler situation. This is just that an imperfectly known distribution of massless particles is incident on some event in space-time. We observe the directions from which the particles come, but nothing else. We are to make the best inferences possible about the distribution from our data.

Conventions. The metric is denoted g_{ab} and has signature $+---$. If v^a is a vector, then $v^2 = v^a v_a$.

2. Measures on $\mathscr{S}$

Let $\mathscr{S}$ be the set of light rays at an origin O. Physically, there is no measure on $\mathscr{S}$ invariant under the Lorentz group G, because if one observer at O sees $\mathscr{S}$ uniformly illumined, another, boosted, observer at O will detect a non-uniform illumination. In order to make clear what depends on the observer and what does not, we will introduce a formalism which is Lorentz invariant. Although this involves a few technicalities, it is worth the effort, since it makes later calculations transparent and also because similar ideas apply more generally, as will be discussed in the last section. Our goal is to show that there is a 'twisted' two-form on $\mathscr{S}$ which is invariant under G and in terms of which any two-form, i.e. any smooth measure, can be written.

Let $\mathscr{N}$ be the set of non-zero null vectors. A vector $v^a(l^a)$ at l^a in $\mathscr{N}$ can be regarded as an infinitesimal perturbation of l^a to $l^a + \varepsilon v^a(l^a)$. Since the perturbed point must also be on $\mathscr{N}$, we have $(l^a + \varepsilon v^a(l^a))^2 = 0$ to first order in ε, i.e. $v^a(l^a) l_a = 0$.

Now consider $\mathscr{S}$. Any point in $\mathscr{S}$ may be represented by a non-zero null vector l^a. The same point is also represented by λl^a, where λ is any non-zero real number. We will write

$$l^a \sim n^a \qquad \text{if } l^a = \lambda n^a \text{ for some } \lambda \neq 0$$

and say l^a and n^a are *equivalent* in this case. Now suppose $v^a(l^a)$ is a tangent vector to $\mathscr{S}$ at the point l^a represents. Again, we can regard the vector as determining an infinitesimally displaced point $l^a + \varepsilon v^a(l^a)$. We know from the above that for this to be null we must have $v^a(l^a) l_a = 0$. We also know that if $v^a(l^a)$ is proportional to l^a, then $l^a + \varepsilon v^a(l^a) \sim l^a$ so in this case $v^a(l^a)$ represents the zero tangent vector. Thus $v^a(l^a)$ is determined by only modulo addition of multiples of l^a. Lastly, we must make sure that $v^a(l^a)$ is compatible with the freedom to multiply l^a by λ:

$$(\lambda l^a) + \varepsilon v^a(\lambda l^a) \sim l^a + \varepsilon v^a(l^a).$$

This implies that $v^a(\lambda l^a) = \lambda v^a(l^a)$. To summarise: *a tangent vector to $\mathscr{S}$ at l^a is represented by a quantity $v^a(l^a)$ satisfying $v^a(l^a) l_a = 0$, homogeneous of degree one in*

l^a, *and defined modulo addition of multiples of* l^a. This gives a *Lorentz-invariant* characterisation of the tangent vectors to $\mathscr{S}$, since we did not need to choose any coordinates. We will drop the explicit dependence of v^a on l^a, but it should be remembered that tangent vectors at l^a are homogeneous of degree one in l^a.

We turn now to two-forms at l^a on $\mathscr{S}$. A two-form is represented as $\nu_{ab}\,\mathrm{d}l^a \wedge \mathrm{d}l^b$ where $\nu_{ab} = -\nu_{ba}$. Given any two tangent vectors v^a and w^a at l^a, the two-form gives a measure of the signed area of the infinitesimal parallelogram they determine,

$$2\nu_{ab}v^a w^b.$$

There are several restrictions on ν_{ab} if this is to be well defined. First, ν_{ab} must be homogeneous of degree -2 in l^a, since the area is just a number (i.e. homogeneous of degree zero) and v^a and w^a are each homogeneous of degree one. Next, since v^a and w^a are only defined up to multiples of l^a, the area must be insensitive to this freedom, which means we must have

$$\nu_{ab}l^a = 0.$$

Lastly, we note that since $v^a l_a = w^a l_a = 0$, the addition of a term of the form $l_a q_b - q_a l_b$ to ν_{ab} does not effect the two-form. To summarise: *a two-form on* $\mathscr{S}$ *is represented by a quantity* $\nu_{ab}\,\mathrm{d}l^a \wedge \mathrm{d}l^b$, *homogeneous of degree* -2 *in* l^a, *satisfying* $\nu_{ab}l^a = 0$ *and defined modulo addition of terms of the form* $l_a q_b - q_a l_b$.

The twisted two-form is a quantity $\nu_{ab}\,\mathrm{d}l^a \wedge \mathrm{d}l^b$ which we shall write as μ for short with the following properties: (a) it is Lorentz-invariant; (b) for any $\phi(l^a)$ homogeneous of degree -2, the combination $\phi(l^a)\mu$ is a two-form on $\mathscr{S}$, and every two-form in $\mathscr{S}$ arises this way; (c) $\int_{\mathscr{S}} (t_a l^a)^{-2}\mu = 1$ for any timelike vector t^a with $t^2 = 1$. These properties are all that are necessary in this paper, but we give an explicit formula for μ for the sake of concreteness.

Let

$$k_a = \varepsilon_{abcd} l^b\,\mathrm{d}l^c \wedge \mathrm{d}l^d. \tag{1}$$

We have $v^a k_a = 0$ only for $v^a l^b \varepsilon_{abcd} = 0$, i.e. only for v^a proportional to l^a. Using the fact that l^a is null, it is not hard to show that this implies that k_a must itself be proportional to l_a,

$$k_a = l_a \nu_{cd}\,\mathrm{d}l^c \wedge \mathrm{d}l^d \tag{2}$$

for some ν_{bc}. We can find ν_{ab} explicitly by combining (1) and (2):

$$\nu_{cd} = v^a l^b \varepsilon_{abcd}(v^a l_a)^{-1} \tag{3}$$

for any vector v^a with $v^a l_a \neq 0$, and this is independent of v^a. The reader not fluent with the tensor algebra leading to (2) can take (3) as a definition and verify that the right-hand side is independent of v^a (up to the addition of terms of the form $l_c q_d - q_c l_d$). Now we write

$$\mu = (1/4\pi)\nu_{ab}\,\mathrm{d}l^c \wedge \mathrm{d}l^d$$

for short. By construction, μ is Lorentz-covariant. It is clear from (3) that $\nu_{ab}l^a = 0$ and so $\phi(l^a)\mu$ is a two-form on $\mathscr{S}$ for any $\phi(l^a)$ homogeneous of degree -2. Moreover, since all two-forms on a two-manifold are proportional and μ is not zero, any two-form on $\mathscr{S}$ is a multiple of μ and thus can be written as $\phi(l^a)\mu$ with $\phi(l^a)$ homogeneous of degree -2. Lastly, an explicit calculation shows that $\int_{\mathscr{S}} (t_a l^a)^{-2}\mu = 1/t^2$ if t^a is timelike.

3. Distributions and information theory on $\mathscr{S}$

Suppose there are some massless particles which pass through the event O. The relative density of particles coming in along the ray l^a defines is given by $p(l^a)$, a function homogeneous of degree -2, with $\int_C p(l^a)\mu$ the fraction of particles piercing the subset C of the celestial sphere $\mathscr{S}$. An observer at O will in general have only imperfect knowledge of $p(l^a)$; we wish to know what is the best inference for $p(l^a)$ given limited data.

If there were a natural measure $d^2l = m(l^a)\mu$ on $\mathscr{S}$, we could apply the usual formalism of information theory. We would maximise the entropy $-\int_{\mathscr{S}} p(l^a)\log(p(l^a)/m(l^a))\mu$ subject to the constraints imposed by the observed data. However, there is no natural measure on $\mathscr{S}$, and we must develop a new technique. (One might think of using the 'observer at O's' frame to define a measure on $\mathscr{S}$. This is unsatisfactory, for two observers with different frames might both measure the same moment, say, of $p(l^a)$, and if they used different measures they would infer different $p(l^a)$'s.)

We may turn the problem around, and ask instead, is there a least prejudiced observer? An observer with timelike tangent t^a sees the celestial sphere equipped with a measure $\mu_t = (t_a l^a)^{-2}\mu$ and interprets $p_t = p(l^a)(t_a l^a)^2$ as the number density of massless particles per steradian. He or she would assign an entropy

$$S_t = -\int_{\mathscr{S}} p_t \log p_t \mu_t = -\int_{\mathscr{S}} p \log p_t \mu$$

to the distribution. Suppose for the moment the distribution $p(l^a)$ is known. In general, the information-content of the distribution can be regarded as having two contributions: one intrinsic to $p(l^a)$; and the other coming from t^a. In general, there is no way to split the information into two parts and so recover the 'pure information content' of t^a. It is reasonable however to regard $S_{t'} - S_t$ as the *difference* in information due to t^a and t'^a, and to call an observer with maximum entropy a *least prejudiced observer*, and the negative of this entropy the *irreducible information* in the distribution. Remarkably, it turns out that *there is a unique choice of* t^a *which maximises* S_t.

The symbol D_a will be used to denote the covariant derivative intrinsic to the hyperboloid $t^2 = 1$; it is defined by taking $\partial/\partial t^a$ and then projecting all components orthogonal to t^a. Thus

$$D_a S_t = -\int_{\mathscr{S}} 2(t_a l^a)^{-1} p l_a \mu + 2t_a.$$

This means that t^a is a critical point of the entropy iff

$$\int_{\mathscr{S}} (t_a l^a)^{-1} p l_a \mu = t_a.$$

This is the requirement that $p(l^a)$ have no dipole moment in the frame defined by t^a. So the frame of highest entropy is the frame in which the distribution has no dipole moment. This analysis is the one relevant to the microwave background, since, to good approximation, we may regard $p(l^a)$ as known. We conclude that the correct measure of its anisotropy, according to information theory, is its entropy in the frame in which it has no dipole moment. This is the frame that has been used for analysis for other reasons: in this frame, the 'net velocity' of the microwave background is zero.

Whenever $p(l^a)$ is smooth, there is a unique frame in which its dipole moment vanishes. This can be established by representation-theory arguments about the Lorentz group, or more prosaically by calculating the Hessian

$$D_a D_b S_t = 2\int_{\mathscr{S}} (t_a l^a)^{-2} p[(l_a - t_a t\cdot l)(l_b - t_b t\cdot l)]\mu + 2(g_{ab} - t_a t_b)$$

and checking that it is negative-definite, and that S_t diverges to $-\infty$ as t^a goes to infinity on the hyperboloid $t^2 = 1$. A similar argument applies to not-too-pathological non-smooth distributions.

Let us now consider the case where $p(l^a)$ is only imperfectly known. Suppose an observer measures the moment

$$E = \int_{\mathscr{S}} \eta p \mu \tag{4}$$

of a function $\eta(l^a)$ (homogeneous of degree zero). We want to determine what his or her best guess for p is, based on this datum. We interpret this to mean finding what distribution p and vector t^a maximise the entropy S_t subject to (4). The answer can be found by standard techniques (Lagrange multipliers). We define the partition function

$$Z = Z(\beta, t^a) = \int_{\mathscr{S}} \mathrm{e}^{-\beta\eta}(t\cdot l)^{-2}\mu$$

and determine β and t^a implicitly from the simultaneous equations

$$E = -\partial_\beta \log Z$$

$$t_a = -(\tfrac{1}{2})\partial_{t^a} \log Z.$$

Then

$$p(l^a) = \mathrm{e}^{-\beta\eta}(t\cdot l)^{-2}Z^{-1}.$$

It can be shown that (as long as $\eta(l^a)$ is not constant) there will always be a unique solution to these equations. If we think of E as the energy, then the equation relating it to the inverse temperature β is the standard one. What is interesting here is that there is a Lagrange equation for the vector t^a too, and it formally is its own inverse temperature.

4. Operational interpretation

The absence of a Lorentz-invariant measure on $\mathscr{S}$ means it is impossible to pick light rays through O 'at random' in a relativistically invariant fashion. We would like to give another, operational, perspective on this.

In order to phrase this properly, let us recall that a probability space is a collection of three things: a set X; a family of subsets of X (the space of events—events in the sense of probability theory, not relativity); and a measure defined on the family. For us, X is a smooth manifold, the family of sets is determined by the open sets in the manifold, and the measure is a smooth measure. The measure of any open set is the relative probability that a point chosen 'at random' will lie in that set. One talks of picking points in X 'at random'. This is something of an abuse, since each point has measure zero and therefore zero chance of being chosen. What one really means is, picking the location of a point up to a small but non-zero uncertainty.

Let us try to construct a method of 'picking light rays at random', and see what goes wrong. Do the following.

1. Pick 'at random' a future-pointing unit vector t^a. (Here 'at random' refers to the natural measure on the hyperboloid of future-pointing unit vectors, which is induced by the metric on Minkowski space.)

2. Pick 'at random' (with respect to μ_t) a light ray.

3. Repeat steps 1 and 2 as often as desired, until a distribution of light rays is built up on $\mathcal{S}$.

In practice, this involves two limiting procedures. The first occurs for both steps 1 and 2; it is the fact that one does not really pick the vector t^a or the light-ray with perfect accuracy. This is only a technicality for us. The second limit is the heart of the matter, though. This is that the measure on the hyperboloid of future-pointing unit vectors is not normalisable. (Just as the usual measure $\mathrm{d}x$ on the real line is not normalisable: $\int_{-\infty}^{\infty} \mathrm{d}x = \infty$.) If one demands that the points should be evenly distributed over the hyperboloid to some given accuracy, then, no matter what the accuracy is, one needs infinitely many points on the hyperboloid. This means that, even to achieve a model where the vectors are evenly distributed on the hyperboloid to finite accuracy, some limiting procedure must be involved. The content of our observation that no Lorentz-invariant measure exists on $\mathcal{S}$ is that, *the distribution derived by applying steps 1–3 at best depends on the limiting procedure involved in step 1, and at worst may not exist.* (That is, the limiting distribution on $\mathcal{S}$ may not exist, even though the one on the hyperboloid does.)

Briefly, it is impossible to give an operational sense to 'picking points at random' relativistically on $\mathcal{S}$.

5. Discussion

Information theory sets itself the problem of finding the most reasonable probability distribution that can be inferred from limited measurements. The data required (the set, the collection of subsets, and the measure on them) form a probability space in the sense of Kolmogorov. The measure is what we have called a fiducial measure; it represents the probabilities one would assign in the absence of any information. It is generally determined from physical considerations in the particular problem.

The relativistically invariant analysis of the celestial sphere shows that it is possible to generalise this approach to a case where a fiducial measure is not available. Although one is unable to assign a probability distribution on $\mathcal{S}$ in the absence of information, it turned out that, as soon as any information was given, it was again possible to apply information theory.

Suppose more generally that a Lie group G acts on a manifold X (both finite-dimensional), and that the action is transitive. Let x be a point in X, and G_x the subgroup of G which fixes x. If G_x is compact, then there is a measure on X which is invariant under G and unique up to normalisation. (The construction is straightforward; one picks a volume form μ_0 at some x_0, then defines μ_x by averaging over the different possible ways of using the action of G to carry x_0 to x. If X is not orientable, one must work with the measure which is the absolute value of the volume form.) For example, if X is the plane and G is the group of Euclidean motions, then G_x is O(2), the group of rotations and reflections about x. This is compact, and so there is a measure on the plane invariant under G. On the other hand, if $X = \mathscr{S}$ and G is the Lorentz group, then G_x is the group of null rotations about x. The group G_x is itself isomorphic to the group of Euclidean motions of the plane (excluding reflections), and so is non-compact.

In general, it turns out (essentially because all smooth volume-forms are proportional) that there is a 'twisted' measure on X invariant under G, unique up to normalisation. If one can also find a family of preferred measures (corresponding to the μ_r for $\mathscr{S}$), then the analysis of this paper can be carried out. The essential physical input is the family of preferred measures.

One might hope to carry out this analysis for Wiener measures, since these are not translationally invariant. However, any naïve adaptation of our approach seems impossible. The difficulty is that the Wiener measures based at different origins are very far from being multiples of each other.

Acknowledgements

Some of this work was done at the Universities of Pittsburgh and Michigan, and I am grateful for their hospitality.

References

Jaynes E T 1957 *Phys. Rev.* **106** 620-630

Jaynes E T 1973 *Found. Phys.* **3** 477-493

Liang E P T and Sachs R K 1980 Cosmology *General Relativity and Gravitation* vol 2, ed A Held (New York: Plenum) pp 329-57

Shannon C and Weaver W 1949 *The Mathematical Theory of Information* (Urbana IL: University of Illinois Press)

Wilkinson D T 1986 *Science* **232** 1517-22