Maximum entropy arguments in relativistic statistical mechanics

1970 J. Phys. A: Gen. Phys. 3 121

(http://iopscience.iop.org/0022-3689/3/2/003)

Maximum entropy arguments in relativistic statistical mechanics

K. A. JOHNS and P. T. LANDSBERG

Department of Applied Mathematics and Mathematical Physics, University College, Cardiff

MS. received 2nd October 1969

Abstract. The special relativistic statistical mechanics of systems of non-interacting particles is investigated by means of a maximum entropy argument. Particular attention is paid here to ensuring that the constraints required for this process satisfy the principle of covariance. The results of the preceding paper are used to determine the distribution of probabilities in any inertial frame of reference. Assuming an invariant entropy and using an entropy maximization technique, the probability distribution which is found differs from the usual one. This is discussed.

1. Introduction

The work of an earlier paper (Landsberg and Johns 1970) on the relative significance in special relativity of time-based and ensembled-based probability distributions is here extended by considering certain effects of the Lorentz transformation of the time-based probabilities. In § 2 of this paper a maximum entropy argument is considered, and the constraints to be applied in this procedure are formulated so that, in keeping with the principle of covariance, they apply equally in all inertial frames of reference. The eighteen independent constraints which are found in this way are examined and interpreted in § 3.

The actual process of entropy maximization is undertaken in § 4, using a statistical definition of entropy which ensures its Lorentz invariance and its compatibility with non-relativistic theory. The eighteen constraints produce eighteen Lagrange multipliers, which are evaluated in the inertial frame I_0 in which the system of interest is, on average, at rest. These multipliers are grouped into two four-vectors and one symmetric second-order tensor in a way that at once enables them to be transformed under a Lorentz transformation to another inertial frame of reference. An expression for the probability of occurrence, Π_i, of a state i in a general inertial frame I is found, and in § 5 a discussion is given of the fundamental points behind its deduction. In particular, the statistical and thermodynamic grounds for regarding entropy as Lorentz invariant are considered, and the conclusion is reached that the usual statistical arguments are unsound and could, with benefit, be discarded in favour of a single thermodynamic argument.

2. Constraints for maximum entropy arguments in special relativistic statistical mechanics

In non-relativistic statistical mechanics, a well-known procedure used to determine the distribution over a set of discrete states of a system is to maximize the statistically defined entropy. In doing this, certain constraints are applied. In particular, for the grand canonical distribution, these constraints are

$$\sum_i \Pi_{i0} = 1 \tag{1}$$

$$\sum_i \Pi_{i0} E_{i0} = \bar{E}_0 \tag{2}$$

$$\sum_i \Pi_{i0} N_{i0} = \bar{N}_0 \tag{3}$$

where Π_{i0} is a probability for the system to be in state i, and E_{i0} and N_{i0} are respectively the energy and particle numbers of the system in that state, all quantities being judged in a frame I_0.

In relativistic theory the principle of covariance demands that constraints apply equally in *all* frames of reference I (i.e. for all values of the velocity $\boldsymbol{w}$ which the frame I_0 has in I), though with appropriately transformed values for quantities which are not invariant. It is also necessary to put momentum on the same footing as energy, and demand that it also satisfy a constraining condition. We shall therefore introduce a tensorial notation in which the energy–momentum four-vector, $\{c\boldsymbol{P}_i, E_i\}$ is denoted by the symbol P_i^μ. With this and other quantities the superfix μ or ν (taking values 1–4) denotes the tensorial component, and the suffix i denotes the state of the system; as before the suffix 0 indicates that the quantities to which it applies are measured in the frame I_0 in which the system is, on average, at rest.

The constraints which are to apply in all frames of reference are therefore

$$\sum_i \Pi_i = 1 \tag{4}$$

$$\sum_i \Pi_i P_i^\mu = \{c\bar{\boldsymbol{P}}, \bar{E}\} \tag{5}$$

and

$$\sum_i \Pi_i N_i = \bar{N}. \tag{6}$$

It is, of course, impossible to apply these constraints simultaneously to each of an infinity of frames of reference. Therefore, for convenience, we seek to find a finite set of constraints applicable in one frame, which is exactly equivalent to (4), (5) and (6) applied in all frames. This can easily be done if these equations are re-expressed solely in terms of tensorial quantities, including four-vectors and scalars. In particular, Π_i should be rewritten (Landsberg and Johns 1970)

$$\Pi_i = \frac{E_i}{\gamma E_{i0}}\,\Pi_{i0} = \frac{cP_i^4}{w^4 P_{i0}^4}\,\Pi_{i0}. \tag{7}$$

Here w^4 is the fourth component of the four-velocity w^μ, which is equal to $\{\gamma\boldsymbol{w}, \gamma c\}$.

Thus (4), (5) and (6) become

$$\sum_i \Pi_{i0}\frac{P_i^4}{P_{i0}^4} = \frac{w^4}{c} \tag{8}$$

$$\sum_i \Pi_{i0}\frac{P_i^\mu P_i^4}{P_{i0}^4} = \frac{w^4}{c}\{c\bar{\boldsymbol{P}}, \bar{E}\} \tag{9}$$

$$\sum_i \Pi_{i0}\frac{P_i^4}{P_{i0}^4}N_i = \frac{w^4}{c}\bar{N}. \tag{10}$$

Constraints (8), (9) and (10) apply equally for all frames of reference I. N_i and $\bar{N}$ are, of course, Lorentz invariant and may equally well be written as N_{i0} and $\bar{N}_0$, but it is essential to retain Π_{i0} and P_{i0} in these equations. This will be discussed further in § 5. In the cases of (8) and (10) the only non-invariant quantities involved are the fourth components of the four-vectors P_i^μ and w^μ. The linearity of the Lorentz transformation ensures that these equations hold in all frames if, and only if, similar constraints are applied to *all four* components in any *one* frame of reference (replacement of superfix 4 in the numerator by μ). Thus, for a single frame of reference only, two constraints (with eight independent components) are

$$\sum_i \Pi_{i0} \frac{P_i^\mu}{P_{i0}^4} = \frac{w^\mu}{c} \tag{11}$$

and

$$\sum_i \Pi_{i0} \frac{P_i^\mu}{P_{i0}^4} N_i = \frac{w^\mu}{c} \bar{N}. \tag{12}$$

By use of (7), the dependence of these equations on Π_{i0} rather than Π_i can be removed. Hence we have as the final form of the constraints (8) and (10)

$$\sum_i \Pi_i \frac{P_i^\mu}{P_i^4} = \frac{w^\mu}{w^4} \tag{13}$$

and

$$\sum_i \Pi_i \frac{P_i^\mu}{P_i^4} N_i = \frac{w^\mu}{w^4} \bar{N}. \tag{14}$$

Equation (9) involves a slightly different approach, since the quantities $c\bar{\boldsymbol{P}}$ and $\bar{E}$ do not form a four-vector in the case of a *confined* system (as defined in our earlier work, Landsberg and Johns 1967). However, energy and momentum densities do form components of the symmetric energy–momentum–stress tensor $T^{\mu\nu}$. In particular

$$\{c\bar{\boldsymbol{P}}, \bar{E}\} = T^{\mu 4} V$$

where V is the volume of the system in the general frame I. Since $w^4/c = \gamma$ and $V = V_0/\gamma$, equation (9) becomes

$$\sum_i \Pi_{i0} \frac{P_i^\mu P_i^4}{P_{i0}^4} = T^{\mu 4} V_0. \tag{15}$$

By the same reasoning as before, (15) holds for all frames of reference I if, and only if, a similar condition (e.g. (16) below) holds for *all* components of $T^{\mu\nu}$. Thus in any *one* frame of reference only, there are ten further independent constraints given by

$$\sum_i \Pi_{i0} \frac{P_i^\mu P_i^\nu}{P_{i0}^4} = T^{\mu\nu} V_0 \tag{16}$$

or, using Π_i instead of Π_{i0},

$$\sum_i \Pi_i \frac{P_i^\mu P_i^\nu}{P_i^4} = T^{\mu\nu} V. \tag{17}$$

3. Interpretation of the constraints

There are therefore *eighteen* independent constraints, given by equations (13), (14) and (17), which must be applied in any one inertial frame of reference to ensure that the *five* constraints (4), (5) and (6) apply in all such frames. It is interesting to see these results expressed in a non-tensorial form. For instance, (13) leads to

$$\sum_i \Pi_i = 1, \qquad \sum_i \Pi_i \boldsymbol{u}_i = \boldsymbol{w}$$

which ensure that the average of the velocity $\boldsymbol{u}_i$, of the system in frame I, is in fact equal to the velocity $\boldsymbol{w}$ of I_0 in I. Similarly (14) leads to

$$\sum_i \Pi_i N_i = \bar{N}, \qquad \sum_i \Pi_i \boldsymbol{u}_i N_i = \boldsymbol{w}\bar{N}.$$

The latter results appears to be new. From (17), with μ or ν equal to 4, we obtain familiar constraints on energy and momentum, namely

$$\sum_i \Pi_i E_i = \bar{E}$$

and

$$\sum_i \Pi_i \boldsymbol{P}_i = \bar{\boldsymbol{P}}.$$

The other six independent components of (17) involve the stresses in the system, and are in general best expressed by equation (17) itself. Only for a homogeneous isotropic system, considered in frame I_0, do they take an interesting and simplified form:

$$\sum_i \Pi_{i0} \frac{P^\mu_{i0} P^\nu_{i0}}{P^4_{i0}} = \begin{cases} pV_0, & (\mu, \nu = 1, 2, 3;\ \mu = \nu) \\ 0, & (\mu, \nu = 1, 2, 3;\ \mu \neq \nu). \end{cases}$$

Since P^μ_{i0}/P^4_{i0} is equal to the μ-component of the velocity in I_0 of the system in state i, this result is in fact the familiar statistical expression for the pressure p. The zero for $\mu \neq \nu$ indicates that no lateral stresses exist in an isotropic system.

4. Entropy maximization

The conventional non-relativistic expression for the entropy S is

$$S = -k \sum_i \Pi_{i0} \ln \Pi_{i0}. \tag{18}$$

It is clear from (7) that in general any attempt to replace Π_{i0} by Π_i in this equation will not only destroy the Lorentz invariance of S, but will replace it with a transformation entirely different from the formal Lorentz transformations of tensors and four-vectors. Therefore, we use the fact that I_0 is in this instance a preferred frame of reference, and define the entropy in *all* frames by equation (18). Now, before S can be maximized over all probabilities Π_{i0}, it is necessary to express the appropriate constraints (13), (14) and (17) in the variables of I_0. Thus they become

$$\sum_i \Pi_{i0} \frac{P^\mu_{i0}}{P^4_{i0}} = \frac{w^\mu_0}{w^4_0} = \{0, 0, 0, 1\} \tag{19}$$

$$\sum_i \Pi_{i0} \frac{P^\mu_{i0}}{P^4_{i0}} N_{i0} = \frac{w^\mu_0}{w^4_0} \bar{N}_0 = \{0, 0, 0, \bar{N}\} \tag{20}$$

and

$$\sum_i \Pi_{i0} \frac{P^\mu_{i0} P^\nu_{i0}}{P^4_{i0}} = T^{\mu\nu}_0 V_0. \tag{21}$$

We may now introduce Lagrange multipliers α_0^μ and β_0^μ, which have four components, and $\eta_0^{\mu\nu}$ which is symmetric in μ and ν, and thus has ten independent components. By means of the usual arguments, we find at once for the equilibrium distribution

$$\Pi_{i0} = \exp\left(-1 - \frac{\alpha_0^\mu P_{i0}^\mu}{P_{i0}^4} - \frac{\beta_0^\mu P_{i0}^\mu N_{i0}}{P_{i0}^4} - \eta_0^{\mu\nu}\frac{P_{i0}^\mu P_{i0}^\nu}{P_{i0}^4}\right). \tag{22}$$

(Summations are made over the repeated superscripts μ, ν, using the special relativity metric with signature -2.)

From equations (18)–(22), S may be expressed as

$$S = k(1+\alpha_0^4+\beta_0^4\bar{N}_0+\eta_0^{\mu\nu}T_0^{\mu\nu}V_0). \tag{23}$$

The Lagrange multipliers may be readily identified if the system is assumed to be isotropic. From *reflections* of spatial-coordinate axes,

$$\alpha_0^\mu = -\alpha_0^\mu, \qquad \beta_0^\mu = -\beta_0^\mu \qquad (\mu = 1, 2, 3)$$

and

$$\eta_0^{\mu\nu} = -\eta_0^{\mu\nu} \qquad (\nu \neq \mu).$$

Thus

$$\alpha_0^\mu = \beta_0^\mu = 0 \qquad (\mu = 1, 2, 3) \tag{24}$$

$$\eta_0^{\mu\nu} = 0 \qquad (\nu \neq \mu). \tag{25}$$

Similarly, by *rotations* of these axes,

$$\eta_0^{11} = \eta_0^{22} = \eta_0^{33}. \tag{26}$$

Since the system is now assumed isotropic in frame I_0, $T_0^{\mu\nu}$ reduces to a diagonal form, with elements given by

$$T_0^{11} = T_0^{22} = T_0^{33} = p$$

and

$$T_0^{44} = E_0/V_0.$$

By use of these results, and also (23), (25) and (26), the expression for S becomes

$$S = k(1+\alpha_0^4+\beta_0^4\bar{N}_0+3\eta_0^{11}pV_0+\eta_0^{44}E_0). \tag{27}$$

The remaining unidentified multipliers can be determined if (27) is compared with the familiar thermodynamic expression for the entropy,

$$S = \frac{1}{T_0}(-\mu_0 N_0+pV_0+E_0) \tag{28}$$

where T_0 is the temperature and μ_0 the chemical potential in I_0. Thus

$$\alpha_0^4 = -1, \qquad \beta_0^4 = -\frac{\mu_0}{kT_0}$$

$$\eta_0^{11} = \eta_0^{22} = \eta_0^{33} = \frac{1}{3kT_0}, \qquad \eta_0^{44} = \frac{1}{kT_0}.$$

Thus the Lagrange multipliers are

$$\alpha_0^\mu = \{0, 0, 0, -1\}$$

$$\beta_0^\mu = \left\{0, 0, 0, -\frac{\mu_0}{kT_0}\right\}$$

and $\eta_0^{\mu\nu}$ has diagonal elements

$$\left\{\frac{1}{3kT_0}, \quad \frac{1}{3kT_0}, \quad \frac{1}{3kT_0}, \quad \frac{1}{kT_0}\right\}$$

with all off-diagonal elements being zero. It is now only necessary to apply a Lorentz transformation from I_0 to another frame I to obtain coefficients α^μ, β^μ and $\eta^{\mu\nu}$ in any desired coordinate system. Transforming P_{i0}^μ to P_i^μ in the same way, the exponent in equation (22) remains invariant, and with the use of (7) the final expression for the probability Π_i appropriate to frame I becomes

$$\Pi_i = \frac{cP_i^4}{w^4 P_{i0}^4} \exp\left(-1 - \frac{\alpha^\mu P_i^\mu}{P_{i0}^4} - \frac{\beta^\mu P_i^\mu N_i}{P_{i0}^4} - \frac{\eta^{\mu\nu} P_i^\mu P_i^\nu}{P_{i0}^4}\right) \tag{29}$$

$$= \frac{E_i}{\gamma E_{i0}} \exp\left\{(\mu_0 N_{i0} - E_{i0} - \tfrac{1}{3}\boldsymbol{u}_{i0} \cdot \boldsymbol{P}_{i0}) \frac{1}{kT_0}\right\}. \tag{30}$$

5. Discussion

Two points are worth making about these results. The first is the peculiar status of the fourth (i.e. time) coordinate of the various four-vectors involved. This is easily understood since the whole theory described here is formed from time-based, rather than space-based, probabilities. The second point is the preference given to the frame I_0, in that P_{i0}^4 rather than P_i^4 appears in the denominators in (29). As already mentioned, this arises because I_0 was used in equation (18) for the definition of S in order that this work be consistent with non-relativistic theory, and with the Lorentz invariance of entropy. It is notable, however, that this invariance had to be introduced into these calculations, and did not arise from them. It cannot, therefore, be said that the work presented here in any way supports the argument that the Lorentz invariance of entropy can be proved by statistical means.

This view, as given by von Mosengeil (1907) has already been criticized by van Kampen (1969). As was shown earlier (Landsberg and Johns 1970, § 3), Lorentz-invariant single-ensemble-based probabilities are not in general compatible with time-based probabilities. We have shown here that the use of the latter does not automatically bring about the Lorentz invariance of entropy, which was introduced into our argument as a postulate. The desired proof of its invariance can already be obtained from thermodynamics, if one assumes that the gradual acceleration of a system from one inertial frame to another is a reversible process which cannot alter the value ascribed to the entropy by any observer. Thus our assumption of equation (18) as a *definition* of the entropy, and the special status thus given to frame I_0, appear as perfectly natural procedures.

We have earlier shown (Landsberg and Johns 1970, § 3) that the Lorentz transformations of the mean energy and momentum of an *inclusive* system (as defined by Landsberg and Johns 1967) are derivable from the probability transformation (7) by putting $\boldsymbol{u}_{i0}$ permanently equal to zero (i.e. $\boldsymbol{u}_{i0} = \boldsymbol{0}$ for all i). This, however, constitutes no proof that one must put $\boldsymbol{u}_{i0} = \boldsymbol{0}$ in equation (30) (which applies to confined systems) in order to obtain the probabilities Π_i for inclusive systems. Indeed, this procedure would be in contradiction to the fact that it is really the same physical entity in the same state i which is being considered. Clearly then, one or other of the possible forms of Π_i must be discarded. Since the constraints on the motion of the whole system which give rise to the disputed term $(\frac{1}{3}\boldsymbol{u}_{i0}\,.\,\boldsymbol{P}_{i0})$, are valid even when, in the inclusive case, only the external behaviour of the system is considered, it is evident that it is erroneous to put $\boldsymbol{u}_{i0}$ equal to zero in this case. We therefore have in equation (30) a new and general expression apparently applicable to both types of system.

It is interesting to compare this result with that obtained earlier (Landsberg and Johns 1967). The arguments given there are effectively based on an ensemble, or 'best estimate', approach and thus use Lorentz invariant probabilities and a smaller number of constraints. The probability Π_i obtained in that paper therefore lacks the extra term in the exponent, as do the expressions obtained by other authors (e.g. Børs 1965, Pathria 1966), and may perhaps be described as conventional.

The extra term arises in the same form for a canonical ensemble. It clearly leads to discrepancies with many accepted formulae; yet it results from accepted relativistic principles. In particular, it is based on:

(i) The relativistic transformations which gave rise to the transformation (7) in the preceding paper.

(ii) The principle of covariance as applied to the constraints in § 2 of this paper. In addition, it is based on an *ad hoc* assumption which has not been fully integrated with relativity theory.

(iii) Entropy is Lorentz invariant and its maximization leads to equilibrium distributions.

Our considerations therefore seem to cast doubt on at least one of these assumptions.

References

Børs, A., 1965, *Proc. Phys. Soc.*, **86**, 1141–3.
van Kampen, N. G., 1969, *J. Phys. Soc. Japan (Suppl.)*, **26**, 316–21.
Landsberg, P. T., and Johns, K. A., 1967, *Nuovo Cim.*, **52B**, 28–44.
—— 1970, *J. Phys.* A: *Gen. Phys.*, **3**, 113–20.
von Mosengeil, K., 1907, *Ann. Phys., Lpz.*, **22**, 867–904.
Pathria, R. K., 1966, *Proc. Phys. Soc.*, **88**, 791–9.