Point Processes and the Analysis of Collision Data

David S. Newman, Retired (Boeing Applied Mathematics and Statistics Group)
Seattle, Washington, USA

Heavy ion collisions have been investigated since the 1960's. The current series of experiments at the RHIC facility at Brookhaven began in 2000. Each pair of colliding nuclei will yield from several hundred to over 1000 particles, and a typical experimental run will involve tens of thousands of collisions. The current generation of detectors identifies and measures the momentum of the majority of the charged particles emitted in a collision. It is natural to regard the observed particles in a single collision as a point process, and a series of collisions as an independent (but not identically distributed) sample from that process.

While point process methods are implicit in the method heavy-ion physicists use for pion intensity interferometry, their approach to analysis would be improved by their explicit use. The potential relevance of point processes to more precise single particle collision data analysis and more generally to collision modelling is discussed briefly.

1. INTRODUCTION

At very high energies, collisions between leptons, hadrons and/or heavy ions of given species exhibit substantial variability in the outcome from collision to collision, even at fixed energy and impact parameter. There is stochastic variation not only in the angular distribution and momenta of the observed particles, but also in the numbers and proportions of emerging particle species. Currently used modeling and statistical practices do not adequately address this variation and are therefore unable to take advantage of it. This paper outlines how the use of point processes [1, 2] can clarify and possibly remedy this situation.

Once the raw experimental data from collisions has been "decoded" by track separation and particle identification algorithms, the most natural probabilistic and statistical description of the possible observed outcomes of a collider experiment is that of a point process. This framework accounts for all the variability in the number and mix of particles produced as well as the momenta of the individual particles, and represents the data completely, except for instrumentation issues.

Validation of theory with data from collider experiments is generally linked to quantum field theories via differential cross sections, which correspond to the joint probability density of collisions with a fully specified set of particles as the outcome. In the general theory, collisions with different particle mix outcomes are treated separately. In earlier years, experimental particle physics emphasized statistical models (e.g., the Fermi model) of the nucleus. An important refinement of Fermi's model made in 1960 by Goldhaber et al [3] when they introduced pion intensity interferometry, but statistical modeling was soon overshadowed by the discovery of partonic interactions. As the Standard Model developed and its experimental consequences were probed more deeply, increasing collision energies made partonic models more reliable, and statistical models were ignored. But recently the need for greater precision in reporting of results is creating new demands on data analysis. Theoretical results expressed only as differential cross sections for completely specified outcomes do not specify a complete probability distribution for the numbers and species of emerging particles. Partonic models further approximate collisions in the form "x + y ➔ z + anything." But they provide no specific guidance for the treatment of the "anything" in data analysis, so that stochastic dependence on any details of the "anything" is ignored. In practice there is no standard method for calculating cross sections for all possible particle mix outcomes (or all such outcomes with probabilities above some small threshold) from theory. This situation makes the use of point process based data analysis more appealing and perhaps essential.

2. BACKGROUND ON HEAVY ION COLLIDER EXPERIMENTS

Statistical models were not ignored in heavy-ion physics, and have developed extensively since Goldhaber et al's seminal paper. In addition to cross-section based partonic and more "fundamental" models, heavy-ion collision modeling requires a variety of statistical transport process, hydrodynamic flow, and other statistical models of collective behavior. Pion intensity interferometry plays a crucial role in linking these intermediate statistical models with the identified data as it presents itself, i.e. as a point process.

The current series of heavy ion experiments at the RHIC (Relativistic Heavy Ion Collider) facility at Brookhaven began in 2000. In addition to exploring the properties of hadronic and nuclear matter, these experiments are searching for new phenomena, especially the elusive quark-gluon plasma (QGP), at collision energies of up to 200 Gev per nucleon. Each pair of colliding nuclei will yield from several hundred to over 1000 particles, and a typical experimental run will involve tens of thousands of collisions. These experiments are also expected to shed light on such diverse phenomena as phase transitions in subnuclear matter, the early history of the "big bang," stellar evolution, neutron stars, and other astronomical and cosmological phenomena.

Some of the most recently developed detectors — for example, the STAR detector at RHIC — are capable of identifying and measuring the momenta of virtually all the charged particles emitted in a collision. This capability presents an opportunity to address new issues, both in theory and experimentally, which have received little attention in the past. It should be possible, for example, to determine if the occurrence of one phenomenon affects the probability of others, i.e., their occurrence or non-occurrence is statistically dependent. The momentum phase space distribution of mesons (especially pions) plays a key role in other aspects of the analysis. The point process structure is implicit in the current method of pion data analysis, but the failure to recognize it explicitly has led to some misunderstandings which will be discussed below.

Many of the pions produced in heavy-ion collisions are at low momentum. The phase space-time region in which they are produced is small enough so that even if they are assumed (as is generally done) to be produced independently, i.e. by unrelated partonic events, identical pions must be represented by a single, symmetrized multiparticle wave function, since they are bosons. As is the case with photons, this fact makes intensity interferometry — also known as HBT interferometry, for the work of Hanbury-Brown and Twiss [4] — possible. Its introduction to nuclear collisions is due to Goldhaber et al, as noted earlier.

From the observed momentum distribution, pion HBT interferometry reconstructs the position space-time distribution of pion generation, which describes the hadronic "freeze-out" space-time surface, and the equation of state of the ultra-dense matter which precedes the hadronic state [5]. A realistic treatment of this reconstruction requires the support of the other statistical models cited earlier. This freeze-out surface geometry is a vital prediction target for models that attempt to represent the early stages of the collision, especially those concerned with how the phase transition from a pre-hadronic state (whether QGP or some other non-nuclear state) to observed hadrons develops.

Space does not permit a more detailed discussion of the method used by the heavy-ion community for HBT interferometry, how it relates to point processes, and how it could be improved by an explicit recognition of the connection. This material is available in an extended draft of this paper, and will be presented elsewhere.

3. POINT PROCESSES AND COLLISION DATA

Point processes developed from a number of converging ideas, all of which may be thought of as generalizations of a Poisson process. Important early contributions were made by physicists working on the analysis of cosmic ray showers. Subsequent developments by Bogoliubov, together with original mathematical discoveries prompted by very different fields of application, form the core of the subject. A brief history of the subject is given in [1]. With the exception of the work of Bénard and Macchi [6], which made the point process basis of quantum optics explicit, it appears that there has been little or no contact between more recent mathematical developments and the physics community since the 1950's.

In heavy ion experiments, many important comparisons of data to theoretical predictions are based on momentum distributions of specific particles which can be directly related to cross-section calculations. As we shall see, these distributions are not probability distributions in the ordinary sense. In point process terminology, they are *first moment distributions*. The distinction is extremely important, for reasons related to our comments about the highly variable outcomes of high energy collisions.

In the simplified case where only a single particle species is of interest, and all collisions occur at the same energy and impact parameter, collisions are not only independent but identically distributed. The probability distribution $\{p_n\}$ for the number of particles produced in a given collision is the *multiplicity distribution*, with $p_n \geq 0$ for $n \geq 0$, and $\sum_{n=0}^{\infty} p_n = 1$. As noted earlier, it is reasonable to assume that collisions are statistically independent of one another. Given *exactly n* identical particles, the 3-momenta $(\mathbf{k}_1,..., \mathbf{k}_n)$ of these particles have a joint probability density $P_n^{(n)}(\mathbf{k}_1,...,\mathbf{k}_n)$. The double-index notation used here is that of Zimanyi and Csörgö [7]; they and several other authors refer to it as the *n*-particle "exclusive" probability distribution for fixed multiplicity *n*, the "exclusive" referring to the absence of other particles which may influence the distribution. These joint probability densities are symmetric under permutations of the particle indices $1,\ldots,n$, reflecting the symmetry or antisymmetry of underlying field operators for bosons or fermions respectively.

The momentum distribution used in collider data analysis is *not* $P_1^{(1)}(\mathbf{k})$. This momentum distribution is estimated from data by creating a 3-dimensional histogram, i.e., by defining a collection of *B* (3-dimensional) momentum bins $\{\Delta\mathbf{k}_b,\ldots,b=1,\ldots,B\}$ and counting the numbers of particles observed in each from a series of collisions. If one passes to the limit as the number of collisions approaches infinity and the bins shrink to zero in volume, the resulting histogram approaches the *moment density function* for the point process population from which the collisions form an independent sample. In terms of the multiplicity distribution and probability densities defined above it has the definition

$$N_1(\mathbf{k}) = \sum_{n=1}^{\infty} n p_n P_1^{(n)}(\mathbf{k}) \ , \qquad (1)$$

where

$$P_1^{(n)}(\mathbf{k}) = \int_{K^{(n-1)}}\!\!\ldots\int P_n^{(n)}(\mathbf{k},\mathbf{k}_2,\ldots,\mathbf{k}_n)\, \mathbf{dk}_2\ldots\mathbf{dk}_n \ , \qquad (2)$$

which is the marginal probability density for the momentum of one particle, given that exactly n are present. $K^{(n-1)}$ is 3*(n-1)* dimensional momentum space. Note that in equation (1) the number of particles is indefinite, reflecting the fact that the "binning" in the sample estimate of $N_1(\mathbf{k})$ is performed over collisions with a variable number of particles. It is a moment density because

$$\int_K N_1(\mathbf{k})\,\mathbf{dk} = \langle N \rangle = \sum_{n=0}^{\infty} np_n \qquad (3)$$

is the first *population* moment (or mean, or average) of *N*, the multiplicity, i.e., the random number of particles in a single collision. For comparison with data from a finite sample, $\int_{\Delta\mathbf{k}_b} N_1(\mathbf{k})\,\mathbf{dk}$ is estimated, for each *b*, by the number of particles observed in $\Delta\mathbf{k}_b$ in all collisions, divided by the number of collisions - in other words the *sample* mean number of particles, by bin. This is simply an application of the method of moments for estimation: the *sample* mean *estimates* the *population* mean. In fact one does not need to bin one's data. To estimate $N_1(A) = \int_A N_1(\mathbf{k})\,\mathbf{dk}$, where A is *any* subset of K, simply divide the number of particles observed in A in all collisions by the number of collisions. The standard theorems - the law of large numbers and the central limit theorem in particular - all apply to this estimate ($\hat{N}_1(A)$ in the notation of many statistics texts) of $N_1(A)$. When random variables are replaced by observed values, the estimat*or* becomes an estimat*e*. Please pardon the picky statistician terminology, but this basic sample/population conceptual framework is so quickly ignored!

Higher moment densities can also be defined, and are in fact used in heavy-ion data analysis for pion interferometry. The second *factorial* moment density for the *population* is defined by

$$N_2(\mathbf{k}_1,\mathbf{k}_2) = \sum_{n=2}^{\infty} n(n-1)\,p_n P_2^{(n)}(\mathbf{k}_1,\mathbf{k}_2), \qquad (4)$$

where

$$P_2^{(n)}(\mathbf{k}_1,\mathbf{k}_2) = \int_{K^{(n-2)}} P_n^{(n)}(\mathbf{k}_1,\mathbf{k}_2,\ldots,\mathbf{k}_n)\,\mathbf{dk}_3\ldots\mathbf{dk}_n\ , \qquad (5)$$

the marginal density for two particles, given that *n* are present. The factorial rather than the ordinary moment density is used, to avoid counting particles paired with themselves. The estimator corresponding to (4) is

$$\hat{N}_2(A_1,A_2) = \frac{1}{C}\sum_{j=1}^{C}\int_{A_1}\int_{A_2}\left(\sum_{\substack{i,i'=1\\ i\neq i'}}^{N_j}\delta(\mathbf{k}_1-\mathbf{K}_{ij})\delta(\mathbf{k}_2-\mathbf{K}_{i'j})\right)\mathbf{dk}_1\mathbf{dk}_2.$$

Note that the inner sum has $N_j(N_j-1)$ terms. The integral of the population moment (4) over momentum pair space $K^{(2)}$ is $\langle N(N-1)\rangle = \sum_{n=2}^{\infty} n(n-1)\,p_n$.

Higher factorial moment densities are similarly defined. One occasionally needs third or fourth moments in practice in conventional sample statistics. Some heavy-ion HBT investigators have suggested that these may contain valuable information. These moments are defined by

$$N_m(\mathbf{k}_1,\ldots,\mathbf{k}_m) = \sum_{n=m}^{\infty}[n]_m\, p_n P_m^{(n)}(\mathbf{k}_1,\ldots,\mathbf{k}_m), \qquad (6a)$$

and $\int_{K^m} N_m(\mathbf{k}_1,\ldots,\mathbf{k}_m)\mathbf{dk}_1\ldots\mathbf{dk}_m = \sum_{n=m}^{\infty}[n]_m\, p_n = \langle [N]_m \rangle$. (6b) The notation $[n]_m = n(n-1)\ldots(n-m+1) = n!/(n-m)!$, widely used in numerical analysis, has been adopted, and $P_m^{(n)}$ is the marginal probability density for the momenta of the "first *m*" particles, given that exactly *n* are present, generalizing (2) and (5). $\langle [N]_m \rangle$ is the m^{th} factorial moment of the random multiplicity *N*, which has probability distribution $\{p_n\}$. From (6) it is clear that in the m^{th} factorial moment density, the number of particles present is at least *m* but otherwise indefinite.

It is worthwhile seeing how these formulas simplify for a Poisson process. Since the individual particles in a collision are now completely independent of one another, the basic probability densities factorize:

$P_n^{(n)}(\mathbf{k}_1,\ldots,\mathbf{k}_n) = \prod_{i=1}^{n} P_1^{(1)}(\mathbf{k}_i)$. The number of events has a Poisson distribution, $p_n = e^{-\lambda}\lambda^n/n!$. Then the first moment density is, from (1),

$$N_1(\mathbf{k}) = \sum_{n=1}^{\infty} np_n P_1^{(n)}(\mathbf{k})$$

$$= P_1^{(1)}(\mathbf{k})\sum_{n=0}^{\infty} ne^{-\lambda}\lambda^n/n! = \lambda P_1^{(1)}(\mathbf{k}),$$

so that when integrated over K one has $\langle N \rangle = \lambda$. This Poisson process is, in general, inhomogeneous (not constant in infinite momentum space), and the first moment density is just the probability density of position for a single particle times the mean number of particles per collision. In this case — *and only in this case* —the first moment density can be identified with the probability density. The higher moment densities simplify in like manner.

Returning to the general case, the factorial moment densities $N_m(\mathbf{k}_1,\ldots,\mathbf{k}_m)$ are called "inclusive distributions" in many places in the heavy-ion literature [e.g. 7,8]. They have the following probabilistic interpretation [1, p.133] which does not appear to be widely appreciated: $N_m(\mathbf{k}_1,\ldots,\mathbf{k}_m)\mathbf{dk}_1\ldots\mathbf{dk}_m$ is the probability that there is a single particle in each of the *m* momentum space infinitesimal volume elements $\{[\mathbf{k}_1,\mathbf{k}_1+\mathbf{dk}_1),\ldots,[\mathbf{k}_m,\mathbf{k}_m+\mathbf{dk}_m)\}$, *allowing for the presence of other particles elsewhere*. It describes the distribution of *m* of the particles in a sampled event. But it is not a probability distribution, because the infinitesimal events are not mutually exclusive. The term *inclusive* is appropriate since an indefinite number of additional particles, in addition to the *m* with momenta specified by N_m, may be present. Because of this, sample moment densities cannot be used directly as a basis for likelihood inference. It may be possible to approximate the likelihood using the inverse of (6a) which expresses $P_n^{(n)}$ in terms of N_m, see [1] for details.

Because these infinitesimal events are not mutually exclusive, dividing the function N_m by $\langle [N_m] \rangle$ in order to make it integrate to one does not create a probability distribution for the particles in question. Unfortunately, this practice is seen frequently in both theoretical and experimental physics. While the normalization technically creates a probability distribution, its interpretation is unclear. The implications of a careful treatment of factorial moment densities in statistical and quantum physics are a subject for future research, and could have far-reaching implications.

The point process framework generalizes to more than one particle type. When only a single particle species is of interest, the outcome of a single collision is characterized by its multiplicity and its vector of momentum vectors, i.e. the pair of quantities n and $(\mathbf{k}_1,\ldots,\mathbf{k}_n)$. For S particle species, one needs S copies of this structure to characterize a collision. If $n_s, s=1,\ldots,S$ are the multiplicities of the individual species, denote the n_s-vector of momentum vectors of the s^{th} species by $\bar{\mathbf{k}}_s = (\mathbf{k}_{1,s},\ldots,\mathbf{k}_{n_s,s}), s=1,\ldots,S$, which is a $3n_s$ dimensional vector. The joint multiplicity distribution takes the form $p(n_1,\ldots,n_S) = p(\tilde{n})$, and the joint densities of momentum vectors given $\tilde{n} = (n_1,\ldots,n_S)$ are denoted by $P_{\tilde{n}}^{(\tilde{n})}(\bar{\mathbf{k}}_1,\ldots\bar{\mathbf{k}}_S)$. The latter are symmetric *within* the momentum components for a single particle species, reflecting the symmetry (for bosons) or antisymmetry (for fermions) of underlying field operators, as noted earlier. But symmetry *between* species is not required.

This general point process formulation appears quite complex, but models based on quantum field theory and statistical physics will "flow through" to the point process framework for data analysis. For single-hadron and lepton collisions and for "partonic events" (i.e. jets) which can be identified within more complex collisions, conservation of momentum and energy may reduce the number of degrees of freedom substantially, and other particles in large n events may perhaps be ignored, up to a point. But looking at marginal distributions of the particles in a collision which are of interest, averaged over those which are not, will give a clearer picture of what is happening than informal approximations.

Special techniques may be used to carefully identify jets and other specific phenomena within a collision. It may be possible to formalize these techniques within the point process context by using marginal and conditional probability arguments, or if not, to validate them as good exploratory methods *post hoc* using point process based statistical inference. Generally, for collisions in which large numbers of identical particles are produced by similar processes, further reductions in the complexity of the point process description occur where n-body interactions can be summarized by pairwise functions of the interaction. Bénard and Macchi [6] show that n-particle wave functions of identical bosons and fermions can be described in this way in many situations, a result that is likely to be very useful in particle interaction modeling.

References

[1] D.J. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes*. New York: Springer-Verlag, 1988. Second edition, Vol. 1, 2002. Page and chapter references are to the first edition. This text has the best coverage of methods and techniques which will prove relevant to physicists. It is much more readable than most treatises on this subject, much of it being written in the style of William Feller's classic work on probability. There is little on statistical inference here, but that is true of almost all books on the subject, and most treatments are extremely abstract and axiomatic. Still, some may find [1] "too mathematical." In the new edition some of the "heavy mathematics" is postponed to Vol. 2. For a lighter mathematical treatment [2] is recommended as a starter, but it does not cover enough of the material physicists will need.

[2] D.R. Cox and V. Isham, *Point Processes.* New York: Chapman & Hall, 1980 (now published by CRC Press).

[3] G. Goldhaber, S. Goldhaber, W. Lee and A. Pais, *Phys. Rev.* **120** (1960), 300

[4] R. Hanbury Brown and R.Q. Twiss, *Phil. Mag.* **45** (1954) 663; *Nature* **177** (1956) 27; *ibid* **178** (1956) 1046; *Proc. Roy. Soc.* **A 242** (1957) 300; *ibid* **243** (1957) 291; **248** (1958) 199; and **248** (1958) 222.

[5] U. Heinz and U. A. Wiedemann, *Phys. Rep.* **319** (1999) 145, nucl-th/9901094

[6] O. Macchi, *C. Rendus Acad. Sci.* série A **272** (1971), 437; C. Bénard, *Phys. Rev.* A **2** (1970), 2140; O. Macchi, *IEEE Transactions on Information Theory* **17** (1971), 2; C. Bénard and O. Macchi, *J. Math. Phys.* **14** (1973), 155; C. Bénard, *J. Math. Phys.* **16** (1975), 710; O. Macchi, *Advances in Applied Probability* **7** (1975), 83

[7] J. Zimanyi and T. Csörgö, *Heavy Ion Phys.* **9** (1999), 241, hep-ph/9705432

[8] I.V. Andreev, M. Plümer and R.M. Weiner, *Intl. Jour. of Mod. Phys.* **A 8** (1993), 4577