Nuclear Instruments and Methods in Physics Research A 821 (2016) 142–150

Contents lists available at ScienceDirect

Nuclear Instruments and Methods in Physics Research A

journal homepage: www.elsevier.com/locate/nima

# Statistical-noise reduction in correlation analysis of high-energy nuclear collisions with event-mixing

R.L. Ray*, P. Bhattarai

*Department of Physics, The University of Texas at Austin, Austin, TX 78712, USA*

ARTICLE INFO

*Article history:*
Received 5 February 2016
Received in revised form
10 March 2016
Accepted 16 March 2016
Available online 17 March 2016

*Keywords:*
Correlation
Event-mixing
Noise reduction
Error propagation

ABSTRACT

The error propagation and statistical-noise reduction method of Reid and Trainor for two-point correlation applications in high-energy collisions is extended to include particle-pair references constructed by mixing two particles from all event-pair combinations within event subsets of arbitrary size. The Reid–Trainor method is also applied to other particle-pair mixing algorithms commonly used in correlation analysis of particle production from high-energy nuclear collisions. The statistical-noise reduction, inherent in the Reid–Trainor event-mixing procedure, is shown to occur for these other event-mixing algorithms as well. Monte Carlo simulation results are presented which verify the predicted degree of noise reduction. In each case the final errors are determined by the bin-wise particle-pair number, rather than by the bin-wise single-particle count.



## 1. Introduction

In correlation analysis of binned data the quantity of interest is the covariance of observable $x$ between arbitrary bins $m$ and $n$, given by

$$\overline{(x_m-\bar{x}_m)(x_n-\bar{x}_n)} = \overline{x_m x_n} - \bar{x}_m\bar{x}_n \quad (1)$$

where $x$ is the bin content and over-lines indicate averages over independent measurements. For example, high-energy collisions between atomic nuclei $x$ could represent the number of sub-atomic particles produced and detected for each collision, or *event*, within bins defined by particle 3-momentum. For example, such bins could be constructed using the transverse momentum ($p_t$) (component of 3-momentum perpendicular to the direction of the colliding beams), the azimuth angle ($\phi$) in the plane transverse to the beam, the pseudorapidity ($\eta$) where $\eta = -\log(\tan\theta/2)$ and $\theta$ is the polar angle relative to the beam direction, the azimuthal angle difference ($\phi_1-\phi_2$) for arbitrary particles 1 and 2, and the pseudorapidity difference ($\eta_1-\eta_2$). Quantity $x_m x_n$ is the number of particle-pairs in 2D bin ($m$,$n$) and $\overline{x_m x_n}$ is calculated by averaging product $x_m x_n$ over all collision events in the event collection. Pairs of particles from the same event are referred to as *sibling* pairs. Averages $\bar{x}_m$ and $\bar{x}_n$ are calculated using all events in the collection where quantity $\bar{x}_m\bar{x}_n$ is the *reference*. For a given set of measured events the covariance will have a unique numerical value. The goal of this paper is to calculate the statistical error in this quantity by following and extending the error propagation and error reduction method of Reid and Trainor [1].

In practical applications measured quantities $x_m$, $x_n$ and $x_m x_n$ are affected by experimental inefficiency, acceptance and contamination. Inefficiency and acceptance losses in single particle counts and some contamination effects are readily corrected via ratio $\overline{x_m x_n}/(\bar{x}_m\bar{x}_n)$. However, two-particle inefficiencies, which occur when signals in the detectors from two particles are unresolved, cannot be corrected this way. Such inefficiencies are due to finite detector resolution and, if uncorrected, produce significant artifacts in the correlations for heavy-ion collisions [2,3]. The conventional correction method [3] involves removing particle pairs whose signals (e.g. induced ionization, secondary particle showers, etc.) fall within the resolution limits of the detector, and then removing pairs of particles from mixed-events which would have the same, relative locations in the detectors. Corrected results are obtained by increasing the minimum required separation between the detected signals of two nearby particles from zero until ratio $\overline{x_m x_n}/(\bar{x}_m\bar{x}_n)$ stabilizes. The practical consequence of this correction procedure is that the reference $\bar{x}_m\bar{x}_n$ must be calculated by constructing uncorrelated pair counts in 2D bin ($m$,$n$) by averaging over pairs of particles where each particle in a pair is selected from different collision events, referred to as *mixed-events*.

The statistical error of the covariance in Eq. (1) equals the standard deviation of the distribution of covariance values corresponding to independent, statistically equivalent event samples

* Corresponding author.
*E-mail address:* ray@physics.utexas.edu (R.L. Ray).

http://dx.doi.org/10.1016/j.nima.2016.03.058
0168-9002/

(event collections) of underlying parent distributions for quantities $x_m$, $x_n$ and $x_m x_n$. Analytical calculations of this error therefore represent $x_m$, $x_n$ as random, event-wise fluctuating quantities relative to the parent distribution. The sibling and mixed-event pair-numbers are similarly represented.

In Ref. [1] Reid and Trainor derived a practical mixed-event method for calculating the reference in which random, event-wise fluctuations (noise) in $x_m$ and $x_n$, which are common to both sibling and mixed-event pair numbers, cancel in the covariance, significantly reducing the errors. For large data volumes summing the total number of mixed-event pairs can be computationally demanding. Various event-mixing algorithms have been developed by the heavy-ion community to reduce the necessary computation time while retaining sufficient statistical accuracy. One such method was discussed in Ref. [1]. The choice of the reference and the event-mixing method strongly affects the statistical errors in the final correlation measurement. In this paper the statistical noise reduction method of Reid–Trainor will be extended and applied to other, practical event-mixing algorithms. The consequences of these event-mixing choices, or references, for the statistical uncertainties in the correlations will be quantified.

The present application is for ultra-relativistic heavy-ion collisions such as those measured by the STAR experiment [2,4] and by the experiments at the Large Hadron Collider (LHC) [5,6]. The methods presented here are also directly applicable to correlation analysis of multi-particle production from any type of particle collision. The present event-mixing technique for constructing an uncorrelated reference distribution has analogs in, for example, cosmology and acoustics. Measurements of the relative distance correlation between galaxies within an angular patch of the sky require an uncorrelated reference distribution. The latter can be constructed from cross-correlated pairs of galaxies observed in different sky patches or from randomly generated distributions of galaxies [7,8]. In acoustical analysis of multiple, independent time series the autocorrelation, or time-lag dependence, for each time series must be referenced to a cross correlation between two independent time series having the same lag time [9]. In these two examples angular patches of sky or individual time series correspond to collision events and binned numbers of galaxies or acoustical amplitudes correspond to binned number of particles in the present analysis.

This paper is organized as follows. In Section 2 the Reid–Trainor procedure is derived and extended. In Section 3 their method is applied to other event-mixing algorithms. Monte Carlo studies are discussed in Section 4. Conclusions are given in Section 5.

## 2. Generalized Reid–Trainor event-mixing

In Ref. [1] the event collection was separated into pairs of events with similar, total number of detected particles, or *multiplicity*. Event-mixing was only applied between the two events in each pair. For arbitrary 2D bin $(m,n)$ $(m \neq n)$ the total number of sibling pairs of particles in the collection is given by the sum

$$\mathcal{S}_{mn} = \sum_{g=1}^{N_g} \sum_{j=1}^{2} (\overline{m}+\mu_j)(\overline{n}+\nu_j) \tag{2}$$

and the mixed-event pair sum is given by

$$\mathcal{M}_{mn} = \sum_{g=1}^{N_g} \sum_{j' > j=1}^{2} [(\overline{m}+\mu_j)(\overline{n}+\nu_{j'})+(\overline{m}+\mu_{j'})(\overline{n}+\nu_j)] \tag{3}$$

where indices $j$ and $g$ denote events and event groups, respectively. Variables $\overline{m}$ and $\overline{n}$ are the parent distribution number of particles in bins $m,n$ (bins are denoted with subscripts) and fluctuations are represented with random variables $\mu$ and $\nu$ as in Ref. [1]. Event group index $g$ is suppressed in the notation for fluctuations $\mu$ and $\nu$. The number of pairs of events in the collection is $N_g = N_{\text{events}}/2$ where $N_{\text{events}}$ is the number of events in the collection. Note that averages of random variables $\mu$ and $\nu$ within an event collection only vanish in the $N_{\text{events}} \to \infty$ limit. For large event numbers the summations in Eqs. (2) and (3) are approximately given by

$$\begin{aligned}\mathcal{S}_{mn} &\approx N_{\text{events}}(\overline{mn}+\overline{\mu\nu})\\ \mathcal{M}_{mn} &\approx N_{\text{events}}(\overline{mn})\end{aligned} \tag{4}$$

where $\overline{\mu\nu} = (N_{\text{events}}^{-1})\sum_j \mu_j \nu_j$ is non-zero if the fluctuations in bins $m$ and $n$ are correlated. For relativistic heavy-ion collisions $\overline{\mu\nu}/(\overline{mn}) \lll 1$ [2,5,6] and for the purpose of calculating statistical errors the small contributions of $\overline{\mu\nu}$ can be neglected. The large event number limits are defined as $\overline{\mathcal{S}}_{mn} = \overline{\mathcal{M}}_{mn} = N_{\text{events}}(\overline{mn})$.

From the above discussion the correlation quantity of interest is $\mathcal{S}_{mn} - \mathcal{M}_{mn}$ and we calculate the number of correlated pairs per reference pair given by [2,5,6]

$$\frac{\mathcal{S}_{mn} - \mathcal{M}_{mn}}{\mathcal{M}_{mn}} = \frac{\mathcal{S}_{mn}}{\mathcal{M}_{mn}} - 1 \equiv \mathcal{R}_{mn} - 1. \tag{5}$$

The statistical error in $(\mathcal{S}_{mn} - \mathcal{M}_{mn})/\mathcal{M}_{mn}$ equals the statistical error in $\mathcal{R}_{mn}$, denoted by $\Delta\mathcal{R}_{mn}$ and is given by

$$\left(\frac{\Delta\mathcal{R}_{mn}}{\overline{\mathcal{R}}_{mn}}\right)^2 = \left(\frac{\Delta\mathcal{S}_{mn}}{\overline{\mathcal{S}}_{mn}}\right)^2 + \left(\frac{\Delta\mathcal{M}_{mn}}{\overline{\mathcal{M}}_{mn}}\right)^2 - \frac{2\Delta(\mathcal{S},\mathcal{M})_{mn}}{\overline{\mathcal{S}}_{mn}\overline{\mathcal{M}}_{mn}} \tag{6}$$

where $(\Delta\mathcal{R}_{mn})^2$, etc. are variances, $\Delta(\mathcal{S},\mathcal{M})_{mn}$ is a covariance, and $\overline{\mathcal{R}}_{mn} = \overline{\mathcal{S}}_{mn}/\overline{\mathcal{M}}_{mn}$. Simplifying Eq. (6) and averaging over event collections yield

$$(\Delta\mathcal{R}_{mn})^2 = \overline{\mathcal{R}}_{mn}^2 \frac{[\Delta(\mathcal{S}_{mn} - \mathcal{M}_{mn})]^2}{\overline{\mathcal{M}}_{mn}^2} = \frac{\langle[(\mathcal{S}_{mn} - \mathcal{M}_{mn}) - \langle\overline{\mathcal{S}}_{mn} - \overline{\mathcal{M}}_{mn}\rangle]^2\rangle}{\overline{\mathcal{M}}_{mn}^2} = \frac{\langle(\mathcal{S}_{mn} - \mathcal{M}_{mn})^2\rangle}{\overline{\mathcal{M}}_{mn}^2} \tag{7}$$

where angle-brackets represent the average over independent event collections, $[\Delta(\mathcal{S}_{mn} - \mathcal{M}_{mn})]^2$ is the variance of difference $(\mathcal{S}_{mn} - \mathcal{M}_{mn})$, $\overline{\mathcal{R}}_{mn} = 1$, and $\langle\overline{\mathcal{S}}_{mn} - \overline{\mathcal{M}}_{mn}\rangle = 0$.

The key result of Ref. [1] was to show that event-pair-wise mixing eliminates contributions of single-particle fluctuations, leaving only those contributions from fluctuations in the number of pairs. Using Eqs. (2) and (3) the variance in the numerator of Eq. (7) simplifies to

$$\begin{aligned}\langle(\mathcal{S}_{mn} - \mathcal{M}_{mn})^2\rangle &= \left\langle \left[\sum_{g=1}^{N_g}(\mu_1\nu_1 + \mu_2\nu_2 - \mu_1\nu_2 - \mu_2\nu_1)\right]^2 \right\rangle\\ &= \left\langle \sum_{j=1}^{N_{\text{events}}} \mu_j^2\nu_j^2 + \sum_{j=1,\text{odd}}^{N_{\text{events}}} (\mu_j^2\nu_{j+1}^2 + \mu_{j+1}^2\nu_j^2) \right\rangle\end{aligned} \tag{8}$$

where averages over products of bin-wise fluctuations from different events vanish (see Appendix A). Carrying out the above event averaging gives

$$\begin{aligned}\langle(\mathcal{S}_{mn} - \mathcal{M}_{mn})^2\rangle &= N_{\text{events}}\langle\sigma_\mu^2\sigma_\nu^2\rangle + N_g\langle\sigma_\mu^2\sigma_\nu^2 + \sigma_\mu^2\sigma_\nu^2\rangle\\ &= 2N_{\text{events}}\langle\sigma_\mu^2\sigma_\nu^2\rangle \xrightarrow{\text{Poisson}} 2N_{\text{events}}\langle\overline{mn}\rangle \approx 2N_{\text{events}}\overline{mn}\end{aligned} \tag{9}$$

where $\sigma_\mu^2$ and $\sigma_\nu^2$ are the variances in bins $m$ and $n$, respectively. These variances result from the following averages:

$$\begin{aligned}\frac{1}{N_{\text{events}}}\sum_{j=1}^{N_{\text{events}}} \mu_j^2\nu_j^2 &= \sigma_\mu^2\sigma_\nu^2\\ \frac{1}{N_{\text{events}}}\sum_{j=1}^{N_{\text{events}}} \mu_j^2\nu_{j+1}^2 &= \sigma_\mu^2\sigma_\nu^2.\end{aligned}$$

These factorized results are valid if the single-particle fluctuations in bins $m$ and $n$ are uncorrelated, consistent with the above