# Removing centrality fluctuations

**Wojciech Broniowski**

Jan Kochanowski U. & Institute of Nuclear Physics PAN

NA61 theory meeting (webinar), 22 June 2017
arXiv:1704.01532 and arXiv:1706.02862

research w/ **Adam Olszewski**

NATIONAL SCIENCE CENTRE
POLAND

# Introduction

1. Passing between various types of moments
   (textbooky, but apparently not commonly known)
2. Composing moments in superposition (or *compound*) models
   (also known in other fields)
3. Algebraic interpretation of the strongly intensive mesures

---

4. Partial correlations as another efficient method to remove centrality fluctuations

# Centrality fluctuations

- . . . want to separate them from physical quantities (long story)
- More generally:
  separate the effects of fluctuations of external (control) variates from the correlations between physical (measurement) variates

# Lecture on cumulants

# Definition of moments

Standard:

$$M^X(t) \equiv \mathop{\mathbb{E}}_X e^{tX} = 1 + \sum_{i=1}^{\infty} \mu_i'^X \frac{t^i}{i!}, \quad \mu_i'^X = \mathop{\mathbb{E}}_X X^i$$

Cumulant [Fischer 1930]: $K^X(t) \equiv \log M^X(t) = \sum_{i=1}^{\infty} \kappa_i^X \frac{t^i}{i!}$

## Relations:

$$\mu_m'^X = \sum_{k=1}^{m} B_{m,k}(\kappa_1^X, \ldots, \kappa_{m-k+1}^X)$$

$$\kappa_m^X = \sum_{k=1}^{m} (-1)^k (k-1)! B_{m,k}(\mu_1'^X, \ldots, \mu_{m-k+1}'^X)$$

$B_{m,k}(x_1, \ldots, x_{m-k+1})$ – partial exponential Bell polynomials

# Explicitly

$$\begin{aligned}
\kappa_1 &= \mu_1' \\
\kappa_2 &= \mu_2' - {\mu_1'}^2 \\
\kappa_3 &= \mu_3' - 3\mu_2'\mu_1' + 2{\mu_1'}^3 \\
\kappa_4 &= \mu_4' - 4\mu_3'\mu_1' - 3{\mu_2'}^2 + 12\mu_2'{\mu_1'}^2 - 6{\mu_1'}^4 \\
\kappa_5 &= \mu_5' - 5\mu_4'\mu_1' - 10\mu_3'\mu_2' + 20\mu_3'{\mu_1'}^2 + 30{\mu_2'}^2\mu_1' - 60\mu_2'{\mu_1'}^3 + 24{\mu_1'}^5 \\
\kappa_6 &= \mu_6' - 6\mu_5'\mu_1' - 15\mu_4'\mu_2' + 30\mu_4'{\mu_1'}^2 - 10{\mu_3'}^2 + 120\mu_3'\mu_2'\mu_1' - 120\mu_3'{\mu_1'}^3 \\
&\quad +30{\mu_2'}^3 - 270{\mu_2'}^2{\mu_1'}^2 + 360\mu_2'{\mu_1'}^4 - 120{\mu_1'}^6 \\
&\ldots
\end{aligned}$$

(probabilistic interpretation via set partitions, connected components)

# Factorial and factorial cumulant moments

Factorial moments $f_i^X = \mathbb{E}_X\, X(X-1)\ldots(X+1-i)$:

$$F^X(t) = \mathbb{E}_X (1+t)^X = 1 + \sum_{i=1}^{\infty} f_i^X \frac{t^i}{i!} = M^X[\log(1+t)] = e^{K^X[\log(1+t)]}$$

Factorial cumulant:

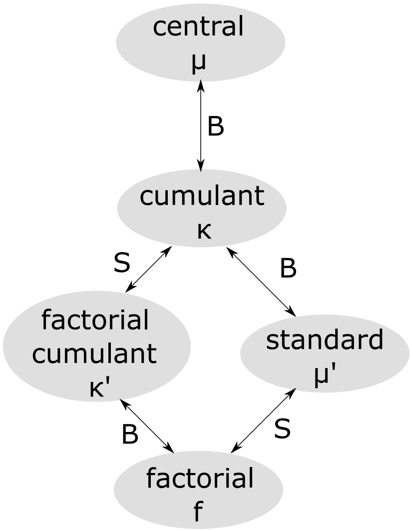
$$G^X(t) \equiv K^X[\log(1+t)] = \log\left[F^X(t)\right] = \sum_{i=1}^{\infty} \kappa_i'^X \frac{t^i}{i!}$$

$$f_m^X = \sum_{k=1}^{m} s(m,k)\mu_m'^X, \quad \mu_m'^X = \sum_{k=1}^{m} S(m,k) f_m^X$$

$$\kappa_m'^X = \sum_{k=1}^{m} s(m,k)\kappa_m^X, \quad \kappa_m^X = \sum_{k=1}^{m} S(m,k)\kappa_m'^X$$

$s(m,k)$ - signed Stirling numbers of the first kind (cycle numbers)
$S(m,k)$ - Stirling numbers of the second kind (partition numbers)

# All are related



# Superposition approach

T

# Superposition (compound) model

The number of particles $N$ is composed via production from $S$ sources; the $j$-th source produces $n_j$ particles:

$$N = \sum_{j=1}^{S} n_j$$

$S$ and $n_j$ are, by assumption, independent from one another.

$$e^{K^N(t)} = \mathop{\mathbb{E}}_{N} e^{tN} = \mathop{\mathbb{E}}_{S,\{n_i\}} e^{\sum_{j=1}^{S} n_j} = \mathop{\mathbb{E}}_{S,\{n_i\}} \prod_{j=1}^{S} e^{tn_j} = \mathop{\mathbb{E}}_{S} \prod_{j=1}^{S} \mathop{\mathbb{E}}_{n_j} e^{tn_j}$$
$$= \mathop{\mathbb{E}}_{S} e^{SK^n(t)} = e^{K^S[K^n(t)]}$$

**Composition formula:**

$$K^N(t) = K^S[K^n(t)]$$

# Composition formulas for generating functions

| type of moments | composition formula |
| --- | --- |
| central | $C^N(t) = [C^n(t)]^{\mu^S} C^S [\log C^n(t) + \mu^n t]$ |
| | $= [C^n(t)]^{\mu^S} C^S [K^n(t)]$ |
| standard | $M^N(t) = M^S [\log M^n(t)] = M^S [K^n(t)]$ |
| cumulant | $K^N(t) = K^S[K^n(t)]$ |
| factorial | $F^N(t) - 1 = F^S [F^n(t) - 1] - 1$ |
| | $F^N(t) = M^S [G^n(t)]$ |
| factorial | $G^N(t) = G^S[e^{G^n(t)} - 1]$ |
| cumulant | $= G^S [F^n(t) - 1] = K^S [G^n(t)]$ |

# Explicit formulas

$$P^N(t) = Q^S[R^n(t)], \quad P, Q, R = M, K, F-1, G$$

**Faà di Bruno's formula →**

$$P_m = \sum_{k=1}^{m} Q_k B_{m,k}(R_1, \ldots, R_{m-k+1})$$

$$\begin{aligned}
P_1 &= Q_1 R_1 \\
P_2 &= Q_2 R_1^2 + Q_1 R_2 \\
P_3 &= Q_3 R_1^3 + 3 Q_2 R_2 R_1 + Q_1 R_3 \\
& \ldots \\
P_6 &= Q_6 R_1^6 + 15 Q_5 R_2 R_1^4 + Q_4 \left(20 R_3 R_1^3 + 45 R_2^2 R_1^2\right) \\
& + Q_3 \left(15 R_2^3 + 60 R_1 R_3 R_2 + 15 R_1^2 R_4\right) \\
& + Q_2 \left(10 R_3^2 + 15 R_2 R_4 + 6 R_1 R_5\right) + Q_1 R_6 \\
& \ldots
\end{aligned}$$

# Why are some moments better than other?

Bernoulli trial: $F^n(t) - 1 = pt \to$

$$f_m^N = p^m f_m^S, \quad \kappa_m'^N = p^m \kappa_m'^S$$

(detector acceptance)

Poisson: $G(t) = \beta t \to$

$$f_m^N = \beta^m \mu_m'^S, \quad \kappa_m'^N = \beta^m \kappa_m^S$$

(thermal production)

Poisson + Bernoulli = Poisson ($\beta \to p\beta$)
(seen by substituting the latter equations to the former)

# Strongly-intensive measures

# Two kinds of particles, one kind of sources

$$N_a = \sum_{j=1}^{S} n_{a,j}, \quad N_b = \sum_{j=1}^{S} n_{b,j}$$

$$P^{N_a,N_b}(t_a, t_b) = Q^S[R^{n_a,n_b}(t_a, t_b)]$$

Compare Mclaurin expansions on both sides →

# Overdetermined set of equations

$$
\begin{aligned}
P_{01} &= Q_1 R_{01} \\
P_{10} &= Q_1 R_{10} \\
&\dots \\
P_{20} &= Q_2 R_{10}^2 + Q_1 R_{20} \\
P_{11} &= Q_2 R_{10} R_{01} + Q_1 R_{11} \\
P_{02} &= Q_2 R_{01}^2 + Q_1 R_{02} \\
&\dots \\
P_{30} &= Q_3 R_{10}^3 + 3 Q_2 R_{10} R_{20} + Q_1 R_{30} \\
P_{21} &= Q_3 R_{10}^2 R_{01} + Q_2 \left(2 R_{10} R_{11} + R_{20} R_{01}\right) + Q_1 R_{21} \\
P_{12} &= Q_3 R_{01}^2 R_{10} + Q_2 \left(2 R_{01} R_{11} + R_{02} R_{10}\right) + Q_1 R_{12} \\
P_{03} &= Q_3 R_{01}^3 + 3 Q_2 R_{01} R_{02} + Q_1 R_{03} \\
&\dots
\end{aligned}
$$

# Relations

- Rouché - Capelli theorem → at rank $n$ there are $n$ relations inolving the $P$ and $R$ moments
- Interesting ones: autonomous, form-invariant (not a trivial problem to find them)

$$n = 1: \quad Q_1 = \frac{P_{01}}{R_{01}} = \frac{P_{10}}{R_{10}} \rightarrow \frac{P_{01}}{P_{10}} = \frac{R_{01}}{R_{10}}$$

$$n = 2: \quad Q_2 \quad = \quad \frac{P_{20} - Q_1 R_{20}}{R_{10}^2} = \frac{P_{11} - Q_1 R_1}{R_{01} R_{10}} = \frac{P_{02} - Q_1 R_{02}}{R_{01}^2}$$

$$\rightarrow \quad \text{two relations}$$

# Strongly-intensive quantities

## Rank-2: $\Delta$ and $\Sigma$

$$\frac{1}{Q_1}\left(\frac{P_{01}P_{20}}{P_{10}} - \frac{P_{10}P_{02}}{P_{01}}\right) = \frac{R_{01}R_{20}}{R_{10}} - \frac{R_{10}R_{02}}{R_{01}} \qquad (\Delta)$$

$$\frac{1}{Q_1}\left(\frac{P_{01}P_{20}}{P_{10}} - 2P_{11} + \frac{P_{10}P_{02}}{P_{01}}\right) = \frac{R_{01}R_{20}}{R_{10}} - 2R_{11} + \frac{R_{10}R_{02}}{R_{01}} \qquad (\Sigma)$$

## Rank-3 (Mrówczyński)

$$\frac{P_{01}P_{30}}{P_{10}^2} - \frac{3P_{21}}{P_{10}} + \frac{3P_{12}}{P_{01}} - \frac{P_{10}P_{03}}{P_{01}^2} = \frac{R_{01}R_{30}}{R_{10}^2} - \frac{3R_{21}}{R_{10}} + \frac{3R_{12}}{R_{01}} - \frac{R_{10}R_{03}}{R_{01}^2}$$

(autonomous, form-invariant)

# Cumulants of the difference of scaled variates

$\hat{X} = X/\langle X \rangle$

$$\hat{N}_- = \hat{N}_a - \hat{N}_b, \quad \hat{n}_- = \hat{n}_a - \hat{n}_b, \quad P^{\hat{N}_-}(t) = Q^S\left[R^{\hat{n}_-}(t/Q_1)\right]$$

$$\begin{aligned} Q_1 P_2 &= R_2 \qquad (\Sigma) \\ Q_1^2 P_3 &= R_3 \qquad (\text{Mrow}) \end{aligned}$$

$$Q_1\left[\frac{P_4}{3P_2^2} - \frac{P_5}{10P_2P_3}\right] = \frac{R_4}{3R_2^2} - \frac{R_5}{10R_2R_3}$$

$$Q_1\left[\frac{7P_3P_6 - P_2P_7}{70P_3^3+70P_2P_4P_3-21P_2^2P_5} - (1-a)\frac{P_4}{3P_2^2} - a\frac{P_5}{10P_2P_3}\right]$$
$$= \frac{7R_3R_6 - R_2R_7}{70R_3^3+70R_2R_4R_3-21R_2^2R_5} - (1-a)\frac{R_4}{3R_2^2} - a\frac{R_5}{10R_2R_3}$$

# Summary 1

- Compound model results: can read off the formulas, e.g., for detector efficiency corrections, to all ranks
- Strongly intensive autonomous, form-invariant measures appear algebraically
- Overdetermined system of equations appears whenever we have more types of measured particles than types of sources
- Rank-3 measure should be tested on the data in exactly the same way as the rank-2 measures

# Partial correlations

T

# Partial correlations

$$
\begin{aligned}
\mathrm{c}(A,B) &= \langle AB\rangle - \langle A\rangle\langle B\rangle \\
\mathrm{v}(A) &= \mathrm{c}(A,A)
\end{aligned}
$$

**The partial covariance:**

$$
\mathrm{c}(X,Y\bullet Z) = \mathrm{c}(X,Y) - \frac{\mathrm{c}(X,Z)\mathrm{c}(Z,Y)}{\mathrm{v}(Z)} \simeq \mathrm{c}(X,Y|Z)
$$

$$
C(X,Y) = \frac{\mathrm{c}(X,Y)}{\langle X\rangle\langle Y\rangle}
$$

# Constraints at the level of sources

## Superposition model

(two kinds of sources, one kind of particles)

$$N_A = \sum_{i=1}^{S_A} m_i, \quad A = F, B$$

$$\begin{aligned} \langle S_A \rangle \langle m \rangle &= \langle N_A \rangle \\ C(S_A, S_{A'}) &= C(N_A, N_{A'}) - \delta^{AA'} \frac{\omega(m)}{\langle N_A \rangle} \equiv \overline{C}(N_A, N_{A'}) \end{aligned}$$

## Sources constrained:

$$C(S_F, S_B \bullet S_C) = \overline{C}(N_F, N_B) - \frac{\overline{C}(N_F, N_C)\overline{C}(N_B, N_C)}{\overline{\mathrm{v}}(N_C)} \simeq C(S_F, S_B | S_C)$$

# Constraints at the level of sources

## Superposition model

(two kinds of sources, one kind of particles)

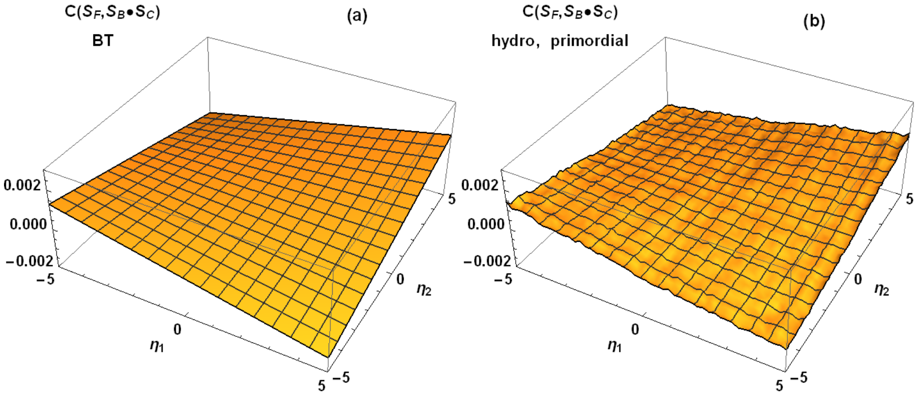
$$N_A = \sum_{i=1}^{S_A} m_i, \quad A = F, B$$

$$\begin{aligned} \langle S_A \rangle \langle m \rangle &= \langle N_A \rangle \\ C(S_A, S_{A'}) &= C(N_A, N_{A'}) - \delta^{AA'} \frac{\omega(m)}{\langle N_A \rangle} \equiv \overline{C}(N_A, N_{A'}) \end{aligned}$$

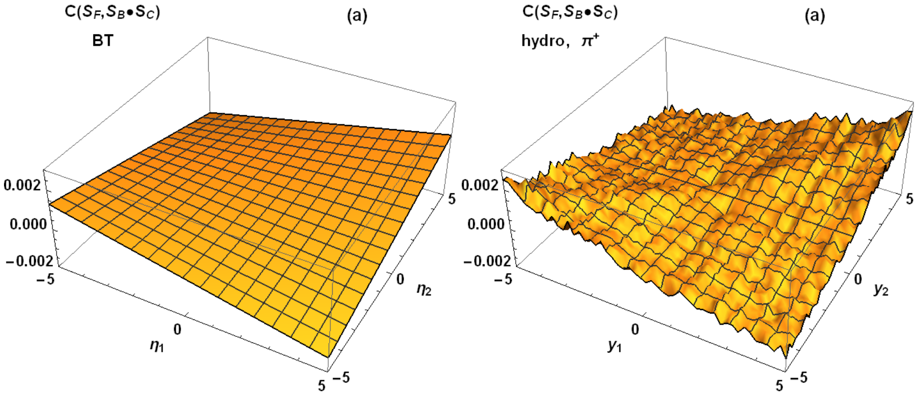
## Particles constrained:

$$C(S_F, S_B \bullet N_C) = \overline{C}(N_F, N_B) - \frac{\overline{C}(N_F, N_C)\overline{C}(N_B, N_C)}{\mathrm{v}(N_C)} \simeq C(S_F, S_B | S_C)$$

# Check on simulated data



# Check on simulated data



# Summary 2

- Use partial correlation techique in 2-particle fluctuation studies!
- Remove autocorrelations (justification via the superposition approach)
- $\rightarrow$ centrality constraints at the level of the initial sources
- Arbitrary number of external constraints (various responses of the detector) may be used

Thanks!