Multivariate Skewness for Finite Mixtures

Nicola Loperfido
Università degli Studi di Urbino "Carlo Bo“,
Dipartimento di Economia, Società e Politica
Via Saffi 42, Urbino (PU), ITALY
e-mail: nicola.loperfido@uniurb.it

Outline

- Finite mixtures
- Third moment
- Multivariate skewness
- Decathlon data

Mixtures: problem

Are the data skewed because they come from different populations?

Mixtures: quotation

"I have at present been unable to find any general condition among the moments, which would be impossible for a skew curve and possible for a compound, and so indicate compoundness. I do not, however, despair of one being found". (Pearson, 1895)

Mixtures: blood pressure

Blood pressure data are skewed.

- Platt (1963): hypertension is an illness present in a genetically defined subpopulation.
- Pickering (1968): hypertension is a labeling for those in the upper tail of the population.

Mixtures: definition

A probability density function $f(\cdot)$ is a finite mixture if it can be represented as a weighted average of several probability density functions:

$$f(x) = \sum_{i=1}^{g} \pi_i \cdot f_i(x)$$

Mixtures: special cases

- Normal: each component is normally distributed.
- Two-component: there are only two components.
- Location: components only differ in location.

Mixtures: options

- Components
- Densities
- Parameters

Mixtures: advantages

- Flexibility
- Interpretability
- Tractability

Mixtures: inference

Maximum likelihood overestimates the number of components when they are erroneously assumed to be symmetric.

↓

How can we choose between different models?

Kullback-Leibler divergence

Let p and q be two probability density functions. The Kullback-Leibler divergence of p from q is

$$J(p,q)=\int p(x)\ln\frac{p(x)}{q(x)}d(x)$$

Kronecker product

$$A = \{a_{ij}\} \in \Re^{p} \times \Re^{q} \qquad B = \{b_{ij}\} \in \Re^{r} \times \Re^{s}$$

$$A \otimes B = \{a_{ij} B\} \in \Re^{p \cdot r} \times \Re^{q \cdot s} \qquad i = 1, \ldots, p \qquad j = 1, \ldots, q$$

Third moment: definition

$$M_3(x) = E\left\{x \otimes x \otimes x^T\right\} \in \Re^{p^2} \times \Re^p$$

Kullback-Leibler approximation

Let p and q be two densities with identical means, identical covariances and possibly different third cumulants $K_{3,p}$ and $K_{3,q}$. Then their Kullback-Liebler divergence is approximately $||K_{3,p} - K_{3,q}||^2/12$, where $||.||$ denotes the Euclidean distance (Lin *et al*, 1999).

Third moment: standardization

The third standardized moment (cumulant) of x is the third moment of $z=\Sigma^{-1/2}(x-\mu)$.

Tensors

A real tensor is a multidimensional array of real values identified by a vector of subscripts:

$$\mathrm{A} = \left\{ a_{i_1 \ldots i_p} \right\} \in R^{n_1} \times \ldots \times R^{n_p}$$

Third-order tensors

$$\mathrm{A} = \{a_{ijk}\} \in R^{n_1} \times R^{n_2} \times R^{n_3}$$

$$\Updownarrow$$

$$A = \begin{pmatrix} A_1 \\ \cdots \\ A_{n_1} \end{pmatrix} \qquad A_i \in R^{n_2} \times R^{n_3}$$

Tensor rank

The rank of a $n_1 \times n_2 \times n_3$ third-order tensor A is the smallest number r satisfying

$$A = \sum_{i=1}^{r} v_i \otimes w_i \otimes u_i \quad v_i \in R^{n_1} \quad w_i \in R^{n_2} \quad u_i \in R^{n_3}$$

Symmetric tensor rank

A $p^2 \times p$ third moment $M_{3,x}$ has symmetric tensor rank k if there are k p-dimensional real vectors v_1, …, v_k satisfying

$$M_{3,x} = v_1 \otimes v_1^T \otimes v_1 + \ldots + v_k \otimes v_k^T \otimes v_k .$$

Third moment: special cases

The symmetric tensor rank of the third moment is easily recovered when the underlying distribution is either a shape mixture of skew-normals or a location normal mixture.

Mardia's skewness: definition

$$x, y \in \Re^p \quad E(x) = \mu \quad V(x) = \Sigma \quad x, y \;\; i.i.d.$$

$$\beta_{1,p}^{M} = E[(x - \mu)^T \Sigma^{-1} (y - \mu)]^3$$

Mardia's skewness: standardization

Mardia's skewness is the trace of the product of the third standardized moment and its transpose

$$\Downarrow$$

$$\beta_{1,p}^{M}(x)=tr\left[M_{3,z}M_{3,z}^{T}\right]$$

Mardias skewness: application

The Kullback-Liebler divergence between a symmetric, standardized random vector and another standardized random vector is approximately proportional to the Mardia's skewness of the latter.

Vectorial skewness: definition

$$\gamma_V = \mathrm{E}\left(z^T z \cdot z\right) = \mathrm{E}\begin{bmatrix} Z_1\left(Z_1^2 + \ldots + Z_p^2\right) \\ \ldots \\ Z_d\left(Z_1^2 + \ldots + Z_p^2\right) \end{bmatrix}$$

Vectorial skewness: standardization

$$\gamma_V = M_{3,z}^T vec\left(I_p\right)$$

Vectorial skewness: clustering

When data come from a mixture of two weakly symmetric distributions with different means and proportional covariances, the projection of the standardized data onto the direction of a vectorial skewness consistently estimates the best linear discriminant projection.

Directional skewness: definition

$$\beta_{1,p}^{D}(z) = \max_{c \in \Re_0^p} \beta_1\left(c^T z\right)$$

Directional skewness: standardization

Directional skewness is a function of the third standardized moment

$$\beta_{1,p}^{D}(x) = \sup_{c^T c=1} \left[\left(c^T \otimes c^T \right) M_{3,z} c \right]^2$$

Directional skewness: fit

Let $\beta^{M}_{2,p}(q)$ and $\beta^{D}_{2,p}(q)$ be the total and directional skewness of the p-dimensional density q. Then the smallest Kullback-Liebler divergence of q from a location mixture of two symmetric densities is

$\beta^{M}_{2,p}(q) - \beta^{D}_{2,p}(q)$ approximately .

$$\Downarrow$$

The fit to the data X of a location mixture of two symmetric densities might be assessed by the difference $b^{M}_{2,p}(X) - b^{D}_{1,p}(X)$.

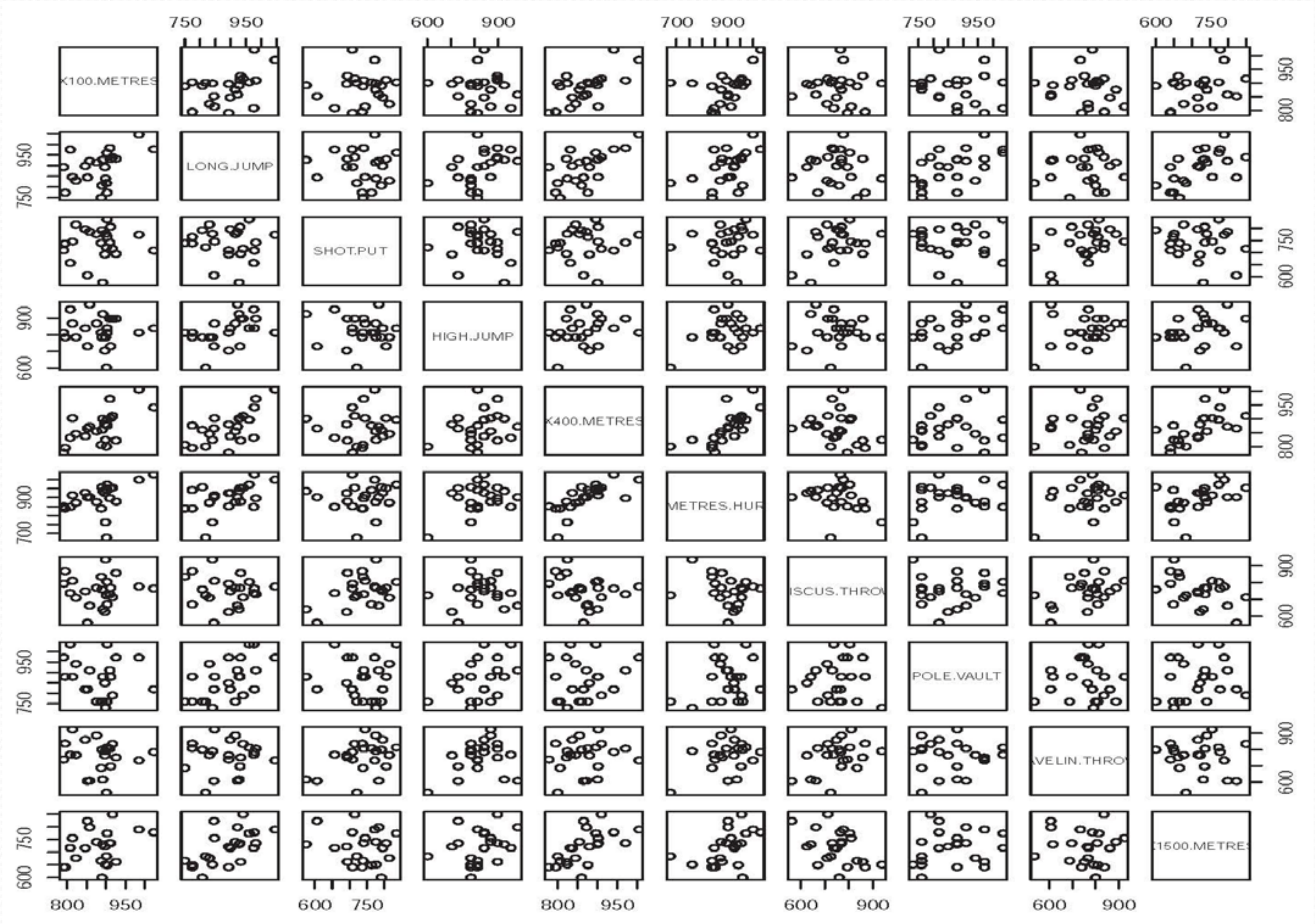
Decathlon data: description

- Units. The 23 athletes who scored points in all 10 events of the Olympic decathlon in Rio 2016.

- Variables. Performances in each event, converted into decathlon points using IAAF scoring tables.

- Source. The official website of the International Association of Athletics Federations (IAAF).

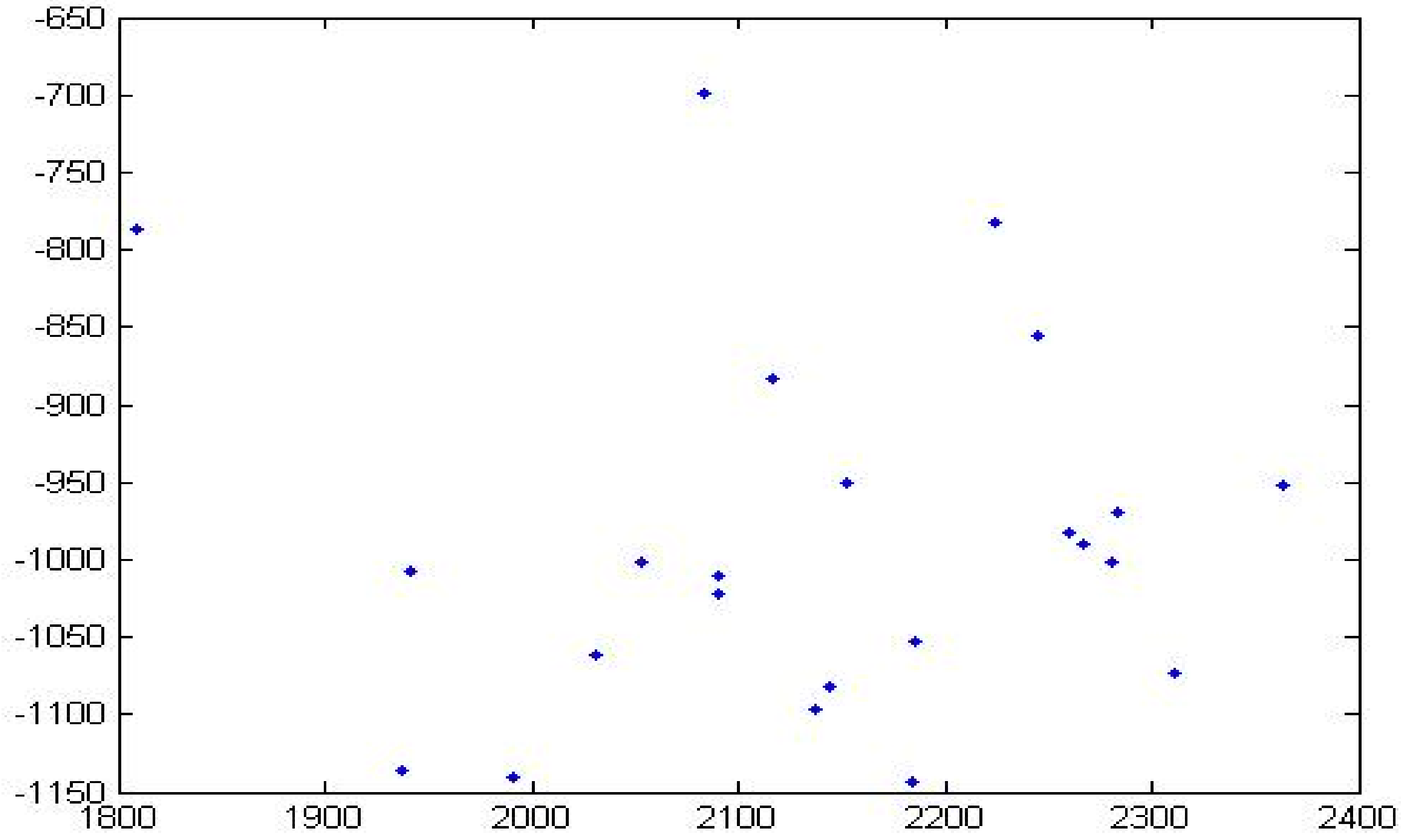
Decathlon data: original variables



Decathlon data: summaries

- Data are slightly nonnormal, with low to moderate levels of skewness and kurtosis.
- The multiple scatterplot does not show any particular features, as for example outliers.
- The number of variables is quite large with respect to the number of units.

Decathlon data: principal components



Decathlon data: summary PCA

- The first two principal component account for about 55% of the total variation.

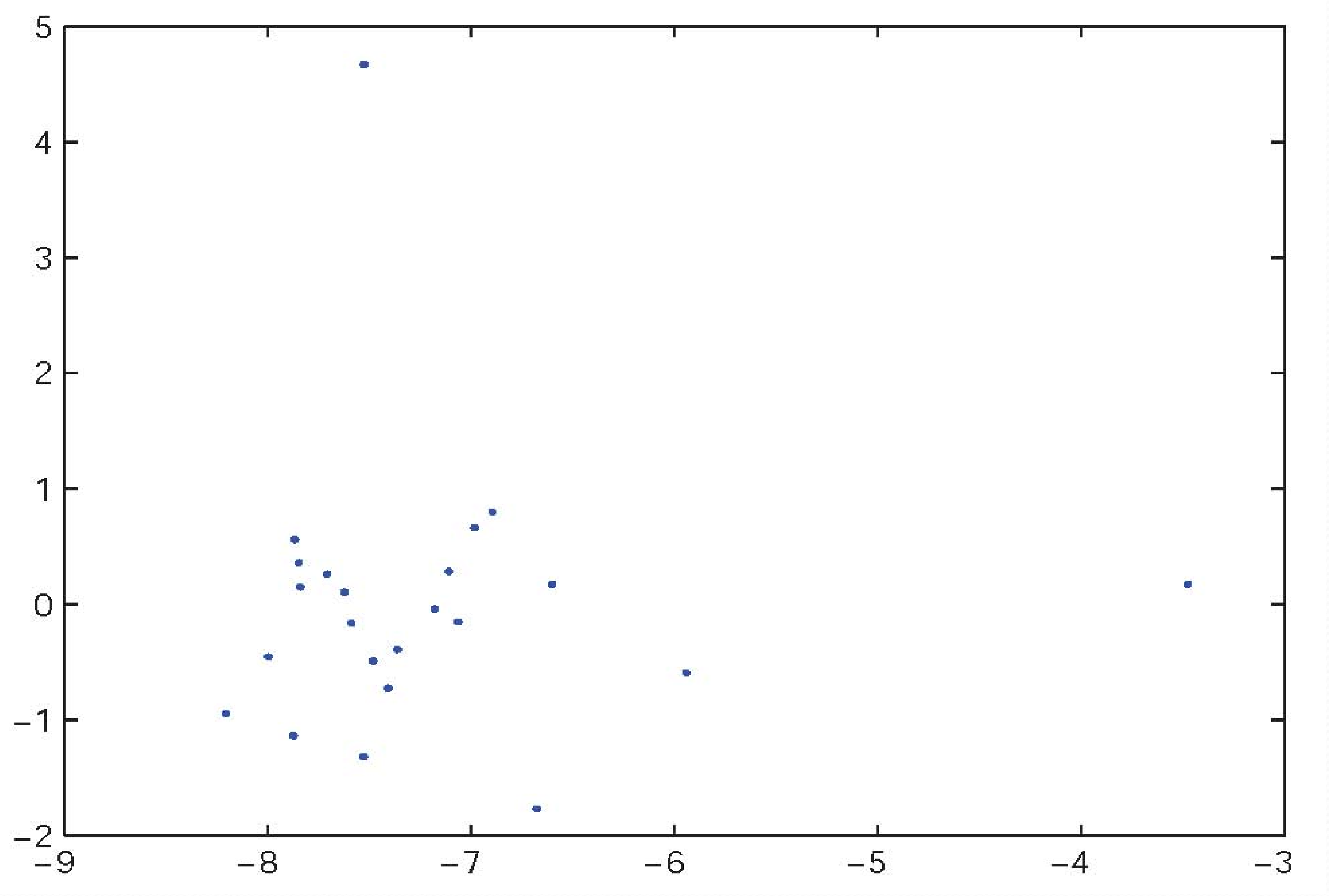
- Their joint distribution is approximately normal.

- They do not clearly suggest the presence of outliers.

Decathlon data: Skewed components

We computed the two most skewed and mutually orthogonal projections of the Decathlon data using the R package MaxSkew.

Decathlon data: skewed components

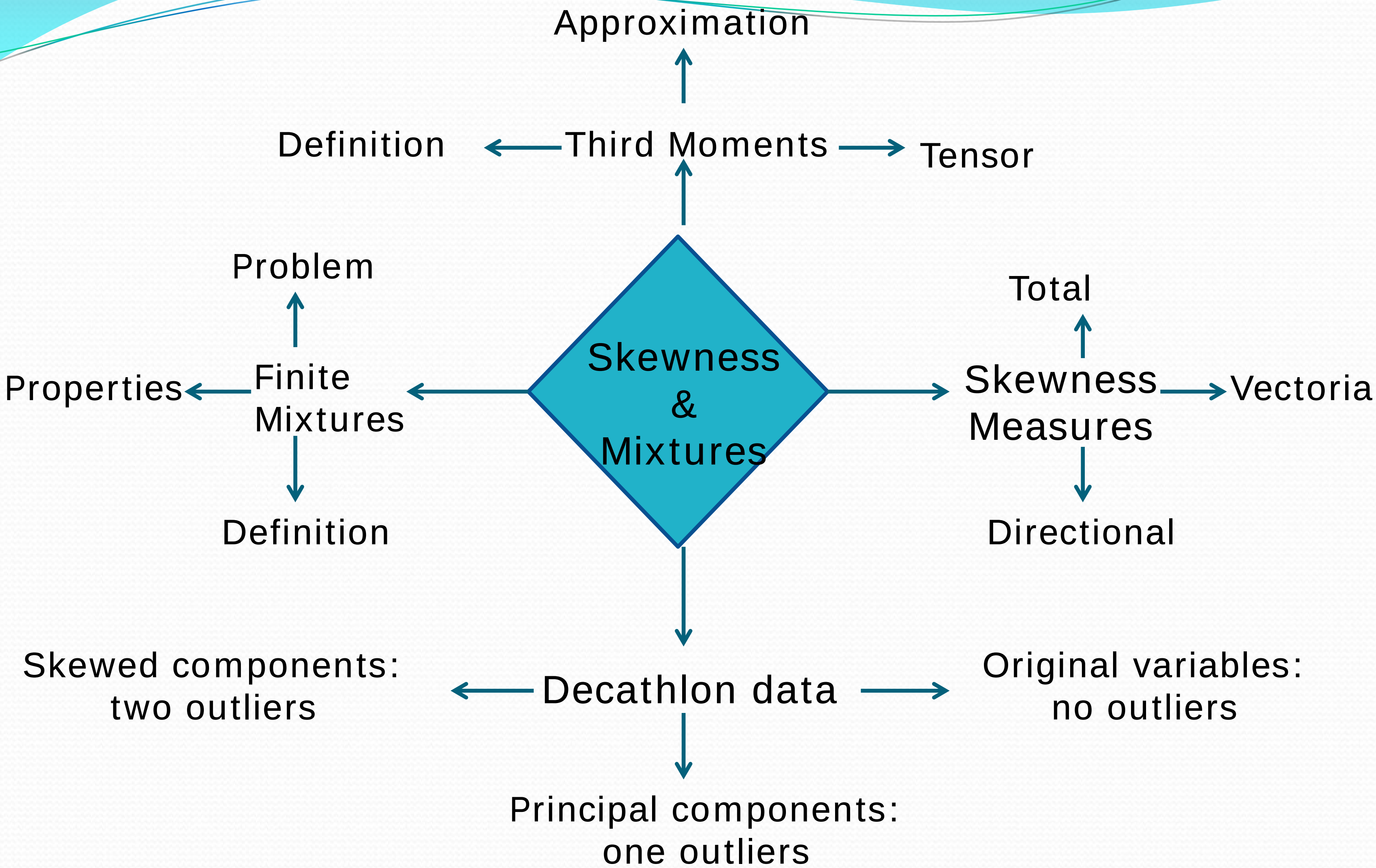


Decathlon data: outliers

- Karl Robert Saluri (EST). He scored lowest due to lower-than-average performances in nearly all events.
- Jeremi Taiwo (USA). He obtained an about average score due to a very unusual pattern of performances.

Decathlon data : conclusion

The scatterplot of the first two most skewed components clearly hints for the presence of two outliers.


Approximation
Definition
Third Moments
Tensor
Problem
Total
Properties
Finite Mixtures
Skewness & Mixtures
Skewness Measures
Vectorial
Definition
Directional
Skewed components: two outliers
Decathlon data
Original variables: no outliers
Principal components: one outliers

Future research

- Package Multiskew
- Fourth moment
- Tensor approach

Essential references

- Loperfido, N. (2015). Singular value decomposition of the third multivariate moment. *Lin. Alg. Appl.* **473**, 202-216.
- Malkovich, J.F. and Afifi, A.A.(1973). On tests for multivariate normality. *J. Amer. Statist. Ass.* **68**, 176-179.
- Mardia, K.V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika* **57**, 519-530.
- McLachlan, G. and Peel, D. (2000). Finite Mixture Models. John Wiley and Sons Inc, New York.
- Mori T.F., Rohatgi V.K. and Székely G.J. (1993). On multivariate skewness and kurtosis. *Theory Probab. Appl.* **38**, 547-551.

Thank you for your attention