Projection pursuit: theory, applications and challenges

LinStat 2024, Poprad-Slovakia

Nicola Loperfido

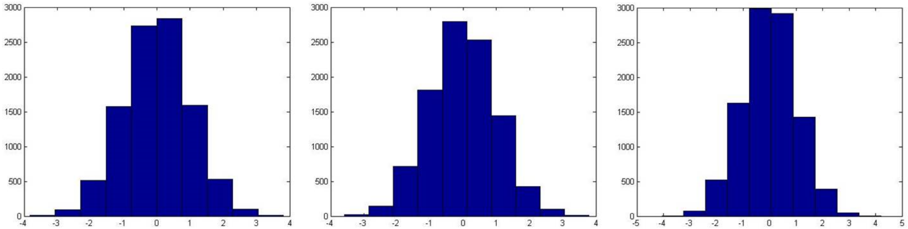
Univ. of Urbino "Carlo Bo" (Italy)
nicola.loperfido@uniurb.it

September 2-6, 2024

Outline

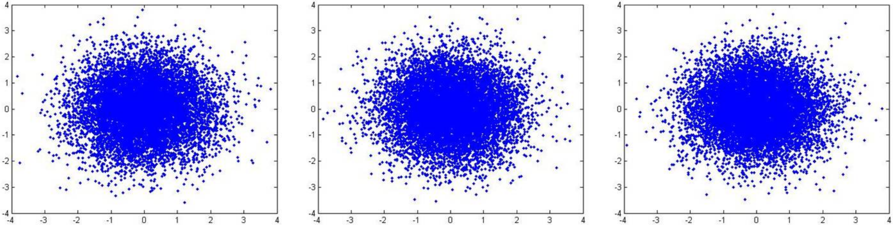
- Motivating example
- Basic concepts
- Statistical applications
- Tensor algebra
- Decathlon data

Motivating example: histograms



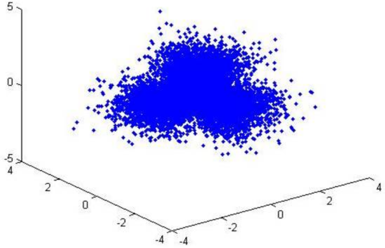
Univariate data appear to be normal.

Motivating example: scatterplots



Bivariate data appear to be uncorrelated and normal.

Motivating example: 3D plot



Trivariate data suggest the presence of three clusters.

Motivating example: distribution

The data were generated from the following distribution:

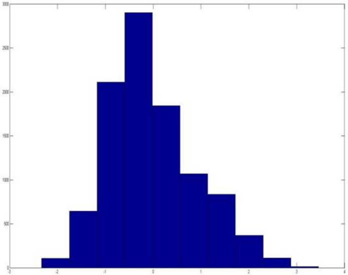
$$f(z_1, z_2, z_3) = 2\phi(z_1)\,\phi(z_2)\,\phi(z_3)\,\Phi(z_1 z_2 z_3)$$

$$\Downarrow$$

$$\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim \begin{pmatrix} Z_1 \\ Z_3 \end{pmatrix} \sim \begin{pmatrix} Z_2 \\ Z_3 \end{pmatrix} \sim N_2\left\{\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right\}$$

Motivating example: skewness

The projection which maximizes skewness clearly hints that data are skewed, but rightly does not suggest the presence of clusters.



Motivating example: problem

Low-dimensional views of the original data might be misleading.

$\Downarrow$

How can we detect meaningful features in high-dimensional datasets?

Projection Pursuit: Definition

Projection pursuit is a multivariate statistical technique aimed at finding interesting low-dimensional data projections.

Projection Pursuit: Normality

Most projections are approximately normal.

$$\Downarrow$$

We look for nonnormal data projections.

Projection Pursuit: Optimization

Each PP analysis focuses on a given data feature.

$$\Downarrow$$

There is a data function (PP index) to be optimized.

Projection Pursuit: Visualization

A picture is worth a thousand words.

$$\Downarrow$$

A graph is worth a thousand figures.

Projection Pursuit: Cleaning

Once a structure is found, it should be removed in order to facilitate the search for other structures.

Projection Pursuit: Features

- **Clusters.** Are there units closer to each other than to other units?
- **Outliers.** Are there units far away from all other units?
- **Shapes.** Do units reminds of a geometric figure?

Projection Pursuit: Indexes

- **Moments.** Skewness, kurtosis, ...
- **Entropies.** Shannon, Gini, ...
- **Distances.** Euclidean, Kolmogorov, ...

Projection pursuit: skewness

$$\beta_{1,p}^{D}(\mathbf{x}) = \underset{\mathbf{c}\in\mathbb{S}^{p-1}}{\arg\max}\, \beta_1\left(\mathbf{c}^\top\mathbf{x}\right) = \underset{\mathbf{c}\in\mathbb{S}^{p-1}}{\arg\max}\, \frac{\mathrm{E}^2\left\{\left(\mathbf{c}^\top\mathbf{x}-\mathbf{c}^\top\boldsymbol{\mu}\right)^3\right\}}{\left(\mathbf{c}^\top\boldsymbol{\Sigma}\mathbf{c}\right)^3}.$$

Projection Pursuit: example

$$f\left(z_1, z_2, z_3\right) = 2\phi\left(z_1\right)\phi\left(z_2\right)\phi\left(z_3\right)\Phi\left(z_1 z_2 z_3\right)$$

$$\Downarrow$$

$$\beta_{1,3}^{D}\left(\mathbf{z}\right) = \mathrm{E}^2\left\{\left(\frac{Z_1 + Z_2 + Z_3}{\sqrt{3}}\right)^3\right\} = \frac{4}{3}\mathrm{E}^2\left(Z_1 Z_2 Z_3\right)$$

Statistical applications: Cluster analysis

Nonnormality might be due to clustering.

$$\Downarrow$$

The least normal projection might be the one which best separates clusters.

Statistical applications: Independent components

Observed data are linear functions of nonnormal, independent sources.

$$\Downarrow$$

Projection pursuit recovers the independent sources.

Statistical application: Discriminant analysis

A unit is to be assigned to one of several groups, with little or no knowledge about them.

$\Downarrow$

The PP index might be the probability of correct classification along a given direction.

Statistical applications: Principal components

A serious drawback of principal components is their lack of robustness.

$$\Downarrow$$

Robustness might be achieved via PP techniques, using an appropriate index.

Statistical applications: Multivariate regression

Approximate the regression function with a linear combination of smooth functions of projected regressors.

$$\Downarrow$$

Use PP to estimate the directions onto which regressors should be projected.

Statistical applications: Normality testing

A random vector is normal iff all its univariate projections are normal.

$$\Downarrow$$

The union-intersection principle naturally leads to projection pursuit for normality testing.

Statistical applications: Canonical correlations

The dependence between most linear projections is itself linear.

$$\Downarrow$$

It is convenient to look for nonlinear associations between linear projections.

Finding outliers might be difficult when the data matrix is large.

$$\Downarrow$$

Data are projected onto the direction where the outliers are best separated from the bulk of the data.

Statistical applications: Portfolio selection

Risk-adverse investors prefer high skewness and small kurtosis.

$$\Downarrow$$

Optimal portfolios should maximize both expectation and skewness, under variance and kurtosis constraints.

Tensor: definition

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

$$\mathcal{A} = \left\{ a_{i_1 \dots i_p} \right\} \in \mathbb{R}^{n_1} \times \dots \times \mathbb{R}^{n_p}.$$

Tensor: symmetry

A tensor $\mathcal{A} = \left\{ a_{i_1 \ldots i_p} \right\} \in \mathbb{R}^{n_1} \times \ldots \times \mathbb{R}^{n_p}$ is symmetric if it is invariant under permuting indices:

$$a_{i_1 \ldots i_p} = a_{i_{\sigma(1)} \ldots i_{\sigma(p)}} \quad 1 \leq i_1, \ldots, i_p \leq n$$

where $\sigma(1)$, ..., $\sigma(p)$ is a permutation of the first p integers.

Tensor: example 1

$$f\left(z_1,z_2,z_3\right)=2\phi\left(z_1\right)\phi\left(z_2\right)\phi\left(z_3\right)\Phi\left(z_1z_2z_3\right),\ \mathrm{E}\left(Z_1Z_2Z_3\right)=\alpha\neq 0$$

$$\Downarrow$$

$$\mathcal{M}_{3,\mathbf{z}}=\left\{\mathrm{E}\left(Z_iZ_jZ_h\right)\right\}=\begin{cases}\alpha & \neq i,j,h\\ \\ 0 & elsewhere\end{cases},\ i,j,h\in\{1,2,3\}$$

Tensor: unfolding

- Tensor unfolding is the process which rearranges the tensor elements into a matrix.
- Each row of the resulting matrix contains the tensor elements identified by the same value of the unfolding index.
- Elements in the same row are arranged beginning with those having the smallest values of the first other indices.

Tensor: skewness

The coskewness of a p-dimensional random vector $\mathbf{x}$ with mean $\boldsymbol{\mu}$ and finite third-order moments is the $p \times p^2$ matrix

$$cos\left(\mathbf{x}\right) = \mathrm{E}\left\{ \left(\mathbf{x} - \boldsymbol{\mu}\right) \otimes \left(\mathbf{x} - \boldsymbol{\mu}\right)^{\top} \otimes \left(\mathbf{x} - \boldsymbol{\mu}\right)^{\top} \right\},$$

that is the unfolding of the symmetric tensor containing all third-order moments of $\mathbf{x} - \boldsymbol{\mu}$.

Tensor: example 2

$$f\left(z_1, z_2, z_3\right) = 2\phi\left(z_1\right)\phi\left(z_2\right)\phi\left(z_3\right)\Phi\left(z_1 z_2 z_3\right), \ \mathrm{E}\left(Z_1 Z_2 Z_3\right) = \alpha$$

$$\Downarrow$$

$$cos\left(\mathbf{z}\right) = \mathrm{E}\left(\mathbf{z} \otimes \mathbf{z}^{\top} \otimes \mathbf{z}^{\top}\right) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & \alpha & 0 & \alpha & 0 \\ 0 & 0 & \alpha & 0 & 0 & 0 & \alpha & 0 & 0 \\ 0 & \alpha & 0 & \alpha & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Tensors: eigenvectors

Let $\mathcal{A}$ be a symmetric tensor of order p and size n. Also, let $\mathbf{A}$ be the matrix obtained by unfolding $\mathcal{A}$ along one of its modes. A scalar $\lambda \in \mathbb{C}$ and a normalized vector $\mathbf{x} \in \mathbb{C}_0^n$ are an eigenvalue and the corresponding eigenvector of $\mathcal{A}$ if they satisfy $\mathbf{A}\mathbf{x}^{\otimes(p-1)} = \lambda\mathbf{x}$.

Tensors: skewness

The projection of the random vector $\mathbf{x}$ achieving maximal skewness is $\mathbf{v}^{\top}\mathbf{\Sigma}^{-1/2}\mathbf{x}$, where $\mathbf{v}$ is the dominant eigenvector of the standardized coskewness of $\mathbf{x}$:

$$cos\left(\mathbf{z}\right)\left(\mathbf{v}\otimes\mathbf{v}\right)=\lambda_{\max}\mathbf{v},\ \mathbf{z}=\mathbf{\Sigma}^{-1/2}\left(\mathbf{x}-\boldsymbol{\mu}\right),\ \ \|\mathbf{v}\|=1.$$

Tensors: example 3

$$f(z_1, z_2, z_3) = 2\phi(z_1)\phi(z_2)\phi(z_3)\Phi(z_1 z_2 z_3), \ \mathrm{E}(Z_1 Z_2 Z_3) = \alpha$$

$$\Downarrow$$

$$cos(\mathbf{z})(\mathbf{1}_3 \otimes \mathbf{1}_3) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & \alpha & 0 & \alpha & 0 \\ 0 & 0 & \alpha & 0 & 0 & 0 & \alpha & 0 & 0 \\ 0 & \alpha & 0 & \alpha & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = 2\alpha \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

Tensors: remarks

- Tensor eigenvectors associated with the same eigenvalue might not constitute a linear space.
- Eigenpairs of real, simmetric tensors might not be real.
- The number of normalized tensor eigenvectors requires further investigation.

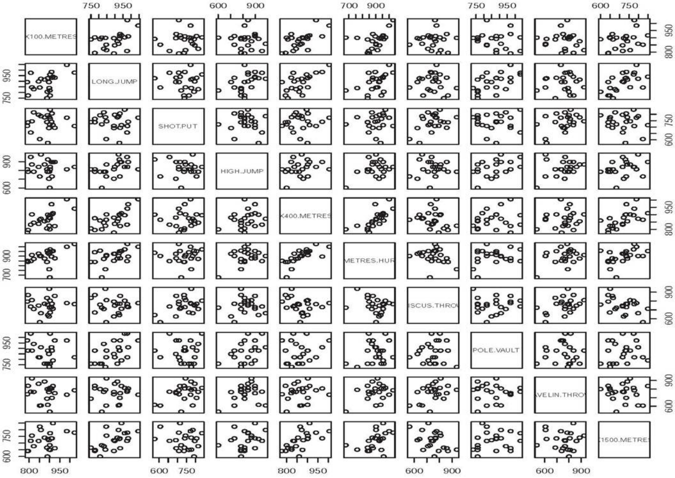
Decathlon data

- **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).

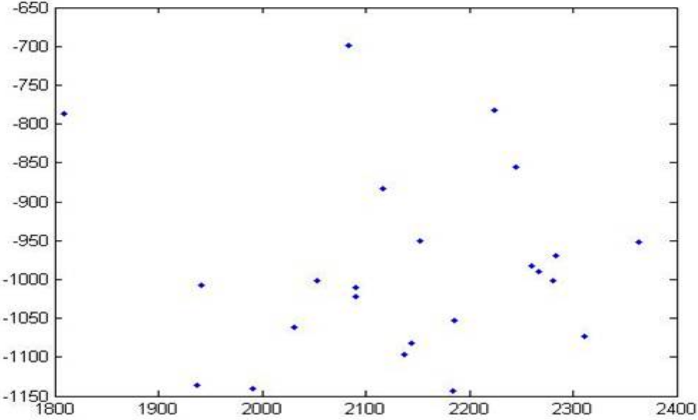
Decathlon data: remarks

- There are just a little more units than variables.
- Decathletes usually have stylized performance patterns.
- Decathlon points have been questioned for their reliability.

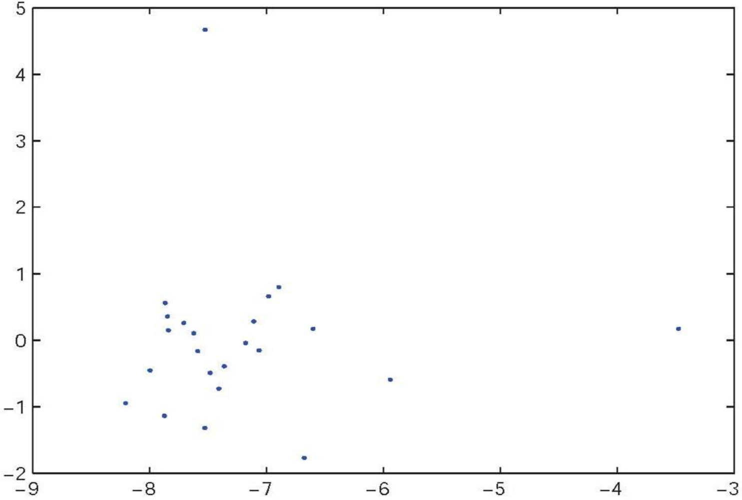
Decathlon data: original variables



Decathlon data: principal components



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: summary

- Scatterplots do not show any clear pattern.
- Principal components do not detect all outliers.
- Skewness maximizations satisfactorily detect outliers.

Ongoing research

- Statistical inference.
- Efficient computing.
- Tensor algebra.

Essential References

- Huber, P. J. (1985). Projection Pursuit (with discussion). The Annals of Statistics 13, 435-525.
- Jones, M. C. and Sibson, R. (1987). What is Projection Pursuit? (with discussion). Journal of the Royal Statistical Society, Series A 150, 1-38.
- Loperfido, N. (2018). Skewness-Based Projection Pursuit: a Computational Approach. Computational Statistics and Data Analysis 120, 42-57.
- Loperfido, N. (2024). Tensor Eigenvectors for Projection Pursuit. *TEST*, 33, 453-472.
- Sturmfels, B. (2016). Tensors and their eigenvectors. *Notices Amer. Math. Soc.* **63**, 604-606.

Thank you for your attention!