# Algebraic models for multilinear dependence

Jason Morton

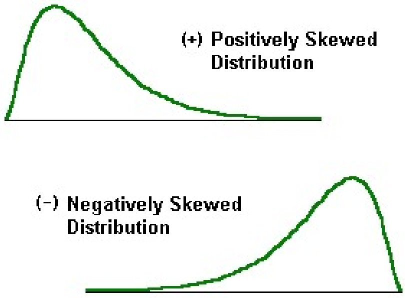
Stanford University

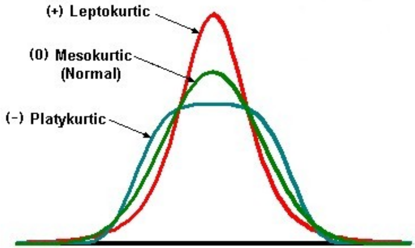
February 21, 2009
NSF Tensor Workshop

Joint work with Lek-Heng Lim of U.C. Berkeley

# Univariate cumulants

Mean, variance, skewness and kurtosis describe the shape of a univariate distribution.



# Covariance matrices

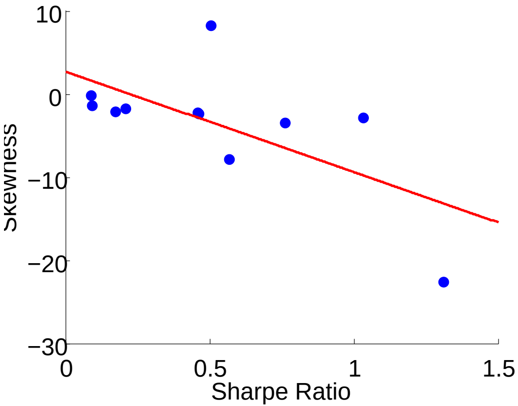
The covariance matrix partly describes the dependence structure of a multivariate distribution.

- Principal Component Analysis
- Factor models
- Risk–bilinear form computes variance $h^\top \Sigma h$ of holdings

But if the variables are not multivariate Gaussian, not the whole story.

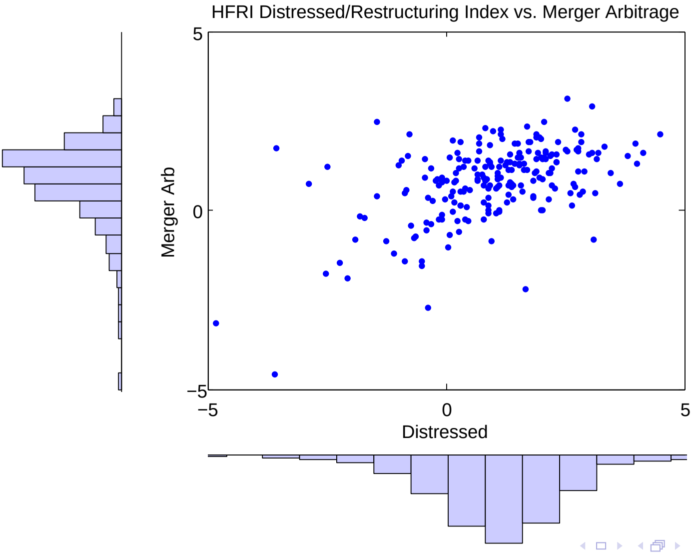
- This is one point of view on the financial crisis; too much reliance on a quadratic, Gaussian perspective on risk.
- Exploited by trading skewness and kurtosis risk for apparent reduction in variance.

# Sharpe Ratio ($\frac{\mu-\mu_f}{\sigma}$) vs Skewness

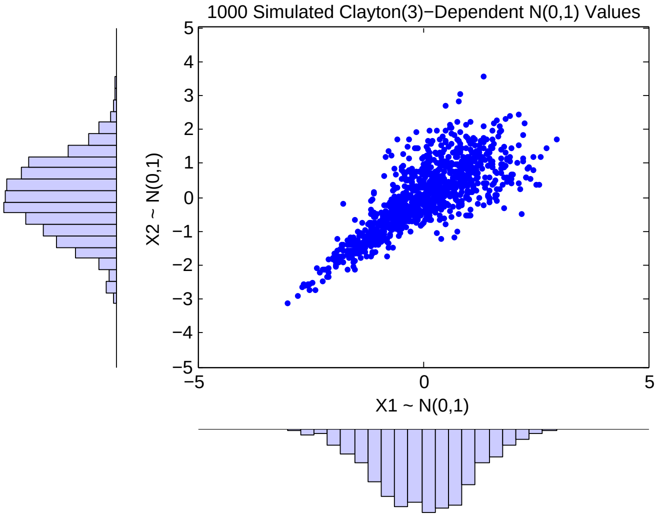



Hedge Fund Research Indices daily returns

# Non-multivariate Gaussian returns are common;



# Even if marginals normal, dependence might not be



# Covariance matrix analogs: multivariate cumulants

- The cumulant tensors are the multivariate analog of skewness and kurtosis.
- They describe higher order dependence among random variables.
- The covariance matrix lets us optimize wrt variance; the cumulant tensors let us optimize wrt skewness, kurtosis, ...

1. Definitions: tensors and cumulants
2. Properties of cumulant tensors
3. Low multilinear rank model (subspace variety)
4. Quasi-Newton algorithm on Grassmannian
5. Multi-moment portfolio optimization
6. Dimension reduction

1 Introduction

2 Definitions

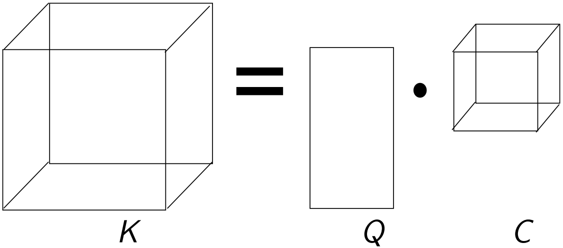
3 Properties

4 Principal Cumulant Component Analysis

5 Algorithm

6 Applications

# Symmetric multilinear matrix multiplication




If $Q$ is a $p \times r$ matrix, $C$ an $r \times r \times r$ tensor, make a $p \times p \times p$ tensor $K = (Q, Q, Q) \cdot C$ or

$$K = Q \cdot C$$

$$\kappa_{\ell mn} = \sum_{i,j,k=(1,1,1)}^{(r,r,r)} q_{\ell i} q_{mj} q_{nk} c_{ijk}.$$

# Moments and Cumulants are symmetric tensors

Vector-valued random variable $\mathbf{x} = (X_1, \ldots, X_n)$.
Three natural $d$-way tensors are:

- The $d$th non-central moment $s_{i_1,\ldots,i_d}$ of $\mathbf{x}$:

$$S_d(\mathbf{x}) = \left[\mathbb{E}(x_{i_1} x_{i_2} \cdots x_{id})\right]_{i_1,\ldots,i_d=1}^{p}.$$

- The $d$th central moment $M_d = S_d(\mathbf{x} - \mathbb{E}[\mathbf{x}])$, and
- The $d$th cumulant $\kappa_{i_1 \ldots i_d}$ of $\mathbf{x}$:

$$K_d(\mathbf{x}) = \left[\sum_{A_1 \sqcup \cdots \sqcup A_q = \{i_1,\ldots,i_d\}} (-1)^{q-1}(q-1)! s_{A_1} \ldots s_{A_q}\right]_{i_1,\ldots,i_d=1}^{p}.$$

$s_{i_1,\ldots,i_d} = \sum_B \prod_{b \in B} \kappa_b$
and $\kappa_{ijk\ell} = m_{ijk\ell} - (m_{ij} m_{k\ell} + m_{ik} m_{j\ell} + m_{i\ell} m_{jk})$

# Measuring useful properties.

For univariate $x$, the cumulants $K_d(x)$ for $d = 1, 2, 3, 4$ are

- expectation $\kappa_i = \mathbb{E}[x]$,
- variance $\kappa_{ii} = \sigma^2$,
- skewness $\kappa_{iii}/\kappa_{ii}^{3/2}$, and
- kurtosis $\kappa_{iiii}/\kappa_{ii}^2$.

The tensor versions are the multivariate generalizations

$$\kappa_{ijk}$$

they provide a natural measure of non-Gaussianity.

# Alternative Definitions of Cumulants

- In terms of log characteristic function,

$$\kappa_{\alpha_1 \cdots \alpha_d}(\mathbf{x}) = (-i)^d \frac{\partial^d}{\partial t_{\alpha_1} \cdots \partial t_{\alpha_d}} \log \mathbb{E}(\exp(i\langle \mathbf{t}, \mathbf{x}\rangle))\Big|_{\mathbf{t}=\mathbf{0}}.$$

- In terms of Edgeworth series,

$$\log \mathbb{E}(\exp(i\langle \mathbf{t}, \mathbf{x}\rangle)) = \sum_{\alpha=0}^{\infty} i^{|\alpha|} \kappa_\alpha(\mathbf{x}) \frac{\mathbf{t}^\alpha}{\alpha!}$$

where $\alpha = (\alpha_1, \ldots, \alpha_d)$ is a multi-index, $\mathbf{t}^{\boldsymbol{\alpha}} = t_1^{\alpha_1} \cdots t_d^{\alpha_d}$, and $\alpha! = \alpha_1! \cdots \alpha_d!$.

See [Fisher 1929, McCullagh 1984,1987] for definitions and properties.

1 Introduction

2 Definitions

3 Properties

4 Principal Cumulant Component Analysis

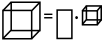
5 Algorithm

6 Applications

# Properties of cumulants: Multilinearity

- Multilinearity: if $\mathbf{x}$ is a $\mathbb{R}^r$-valued random variable and $A \in \mathbb{R}^{p \times r}$

$$K_d(A\mathbf{x}) = A \cdot K_d(\mathbf{x}),$$

  where $\cdot$ is the multilinear action  .
- This makes factor models work: $\mathbf{y} = A\mathbf{x}$ implies $K_d(\mathbf{y}) = A \cdot K_d(\mathbf{x})$;
- Covariance factor model: $K_2(\mathbf{y}) = AK_2(\mathbf{x})A^\top$.
- Independent Component Analysis finds an $A$ to approximately diagonalize $K_d(\mathbf{x})$.

# Properties of cumulants: Independence

Independence:

- If $\mathbf{x}_1, \ldots, \mathbf{x}_p$ are random variables mutually independent of $\mathbf{y}_1, \ldots, \mathbf{y}_p$, we have
  $K_d(\mathbf{x}_1 + \mathbf{y}_1, \ldots, \mathbf{x}_p + \mathbf{y}_p) = K_d(\mathbf{x}_1, \ldots, \mathbf{x}_p) + K_d(\mathbf{y}_1, \ldots, \mathbf{y}_p)$.
- $K_{i_1,\ldots,i_d}(\mathbf{x}) = 0$ whenever there is a partition of $\{i_1, \ldots, i_d\}$ into two nonempty sets $I$ and $J$ such that $\mathbf{x}_I$ and $\mathbf{x}_J$ are independent.
- Why we want to diagonalize in independent component analysis
- Exploitable in other sparse cumulant techniques (breaks rotational symmetry)

# Properties of cumulants: Vanishing and Extending

- Gaussian: If $\mathbf{x}$ is multivariate normal, then $K_d(\mathbf{x}) = 0$ for all $d \geq 3$.
  - Why one might not have heard of them: for Gaussians, the covariance matrix does tell the whole story.
- Marcinkiewicz Theorem: There are no distributions with a bound $D$ so that

$$K_d(\mathbf{x}) \begin{cases} \neq 0 & 3 \leq d \leq D, \\ = 0 & d > D. \end{cases}$$

  - Parametrization is trickier when $K_2$ doesn't tell the whole story.

# Making cumulants useful, tractable and estimable

Cumulant tensors are a useful generalization, but too big. They have $\binom{\#vars+d-1}{d}$ quantities, too many to

- estimate with a reasonable amount of data,
- optimize, and
- store.

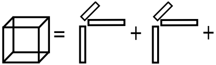
Needed: small, implicit factor models analogous to Principal Component Analysis (PCA)

PCA: eigenvalue decomposition of a positive semidefinite real symmetric matrix. We need a tensor analog.

But, it isn't as easy as it looks...

# Tensor decomposition

Three possible generalizations are the same in the matrix case but not in the tensor case. For a $p \times p \times p$ tensor $K$,

| Name | minimum $r$ such that |
|---|---|
| Tensor rank | $K = \sum_{i=1}^{r} u_i \otimes v_i \otimes w_i$   not closed |
| Border rank | $K = \lim_{\epsilon \to 0}(S_\epsilon)$, Tensor rank$(S_\epsilon) = r$ closed but hard to represent; defining equations unknown. |
| Multilinear rank | $K = A \cdot C, C \in \mathbb{R}^{r \times r \times r}, A \in \mathbb{R}^{p \times r}$, closed and understood. |

# Geometric perspective

- Secants of Veronese in $S^d(\mathbb{R}^p)$ and rank subsets — difficult to study.
- Symmetric subspace variety in $S^d(\mathbb{R}^p)$ — closed, easy to study.
- We take the long skinny matrix to be orthonormal.
  - Stiefel manifold $O(p, r)$ is set of $p \times r$ real matrices $Q$ with orthonormal columns.
  - Grassmannian $Gr(p, r)$ is set of equivalence classes [Q] of $O(p, r)$ under right multiplication by $O(r)$.
- Parametrization of $S^d(\mathbb{R}^n)$ via

$$Gr(p, r) \times S^d(\mathbb{R}^r) \to S^d(\mathbb{R}^p).$$

1 Introduction

2 Definitions

3 Properties

4 Principal Cumulant Component Analysis

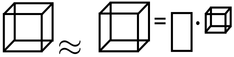
5 Algorithm

6 Applications

# Multilinear rank factor model

Let $\mathbf{y} = Y_1, \ldots, Y_n$ be a random vector. Write the $d$th order cumulant $K_d(\mathbf{y})$ as a best $r$-multilinear rank approximation in terms of the cumulant $K_d(\mathbf{x})$ of a smaller set of $r$ factors $\mathbf{x}$:

$$K_d(\mathbf{y}) \approx Q \cdot K_d(\mathbf{x})$$



where

- $Q$ is orthonormal, and $Q^\top$ projects to the factors
- The column space of $Q$ defines the $r$-dim subspace which best explains the $d$th order dependence.
- In place of eigenvalues, we have the core tensor $K_d(\mathbf{x})$, the cumulant of the factors, analogous to the covariance matrix of the factors in the $r \times r$ case.

Have model, need loss and algorithm.

# Principal cumulant component analysis 1

Factors/principal components that account for variation in each cumulant separately

$$\min_{Q \in \mathrm{O}(p,r),\ C_d \in \mathrm{S}^d(\mathbb{R}^r)} \|\hat{K}_d(\mathbf{y}) - Q \cdot C_d\|^2,$$

Minimize over

- $C_d \approx \hat{K}_d(\mathbf{x})$ a NOT-necessarily-diagonal small $(r \times r \times r)$ symmetric tensor.
- $Q$ an orthonormal matrix

# Principal cumulant component analysis 2

Or, factors/principal components that account for variation in all cumulants simultaneously

$$\min_{Q \in \mathrm{O}(p,r),\ C_d \in \mathrm{S}^d(\mathbb{R}^r)} \sum_{d=1}^{\infty} \alpha_d \|\hat{K}_d(\mathbf{y}) - Q \cdot C_d\|^2,$$

$C_d \approx \hat{K}_d(\mathbf{x})$ not-necessarily-diagonal.

- Appears intractable: optimization over infinite-dimensional manifold
$$\mathrm{O}(p,r) \times \prod_{d=1}^{\infty} \mathrm{S}^d(\mathbb{R}^r).$$
- Reduces to optimization over a single Grassmannian (set of $r$-dim spaces in $p$-dim space) of dimension $r(p-r)$,
$$\max_{Q \in \mathrm{Gr}(p,r)} \sum_{d=1}^{\infty} \alpha_d \|Q^{\top} \cdot \hat{K}_d(\mathbf{y})\|^2.$$
- In practice $\infty = 3$ or 4.

1 Introduction

2 Definitions

3 Properties

4 Principal Cumulant Component Analysis

5 Algorithm

6 Applications

# ALS / Quasi-Newton

- Alternating Least Squares is commonly used for minimizing

$$\Psi(X, Y, Z) = \left\|(X^\top, Y^\top, Z^\top) \cdot T\right\|^2$$

  for $T \in \mathbb{R}^{l \times m \times n}$ cycling between $X, Y, Z$ and solving a least squares problem at each iteration.

- What if $T = K$ is symmetric and

$$\Phi(X) = \left\|X^\top \cdot K\right\|^2?$$

- Better: Quasi-Newton methods, L-BFGS on Grassmannian.

1 Introduction

2 Definitions

3 Properties

4 Principal Cumulant Component Analysis

5 Algorithm

6 Applications

# Mean-variance portfolio optimization

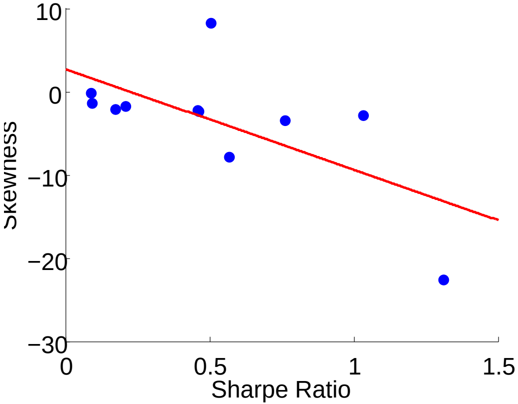
Markowitz mean-variance portfolio optimization defines risk to be variance. For portfolio holdings $h$, solve

$$\min h^\top K_2(\mathbf{x}) h \qquad s.t. \qquad h^\top \mathbb{E}[\mathbf{x}] > \underline{r}$$

Evidence indicates that investors optimizing variance with respect to the covariance matrix accept unwanted skewness and kurtosis risk.

- Extreme example: selling out-of-the-money puts looks safe and uncorrelated
- Many strategies take on this type of risk

# Sharpe Ratio ($\frac{\mu-\mu_f}{\sigma}$) vs Skewness




Hedge Fund Research Indices daily returns

# Muti-moment portfolio optimization

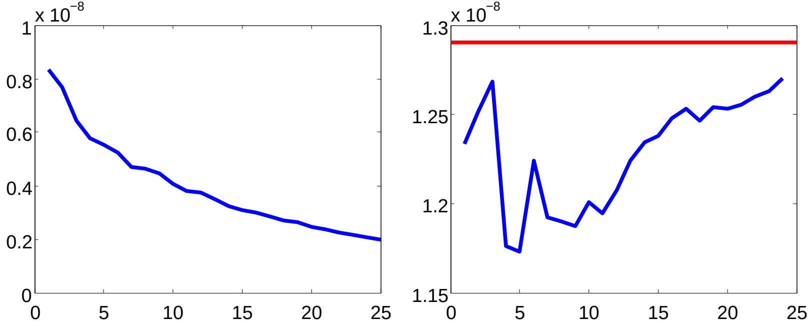
So, take skewness and kurtosis into account in the objective.

- Need to use skewness $K_3$ and kurtosis $K_4$ tensors.
- Use low multilinear rank model to
  - Regularize and
  - Make optimization computable with many assets

To do this,

- Choose an $r$
- Approximate cumulant $K_d \approx Q \cdot C_d$
- For holdings $h$, Multilinear forms $h^\top \cdot K_d \approx h^\top Q \cdot C$ give variance, skewness and kurtosis

# Regularization and optimal number of factors

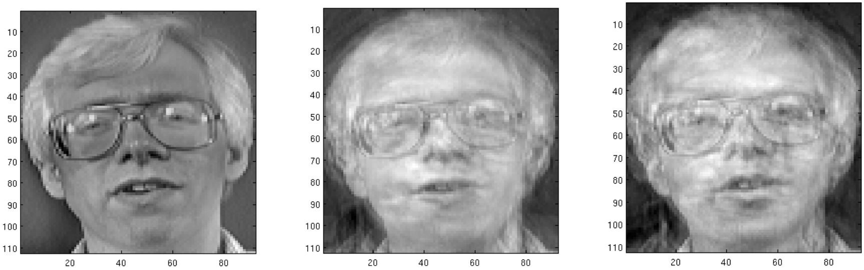



Reconstruction and generalization error $\times$ number of factors for a 50-stock portfolio.

# Dimension reduction

To use PCCA for dimension reduction,

- Compute PCCA approximation $K_d(\mathbf{y}) \approx Q_d \cdot K_d(\mathbf{x})$
- Ignore the cumulant of the factors $K_d(\mathbf{x})$, keep the projector $Q_d$
- In PCCA2 (one $Q$ for all $d$), done
- PCCA1: combine $[Q_2 : Q_3 : Q_4]$ and orthogonalize



# Conclusion

- Introduced cumulant tensors, which generalize skewness, kurtosis, and the covariance matrix,
- Showed they have the expected properties and that we can build factor models from them,
- Used multilinear rank to get around the difficulties of generalizing covariance factor models,
- Estimated out-of sample higher-order portfolio statistics and optimized with them, and
- Performed dimension reduction incorporating higher-order statistics.

# References


- B. Bader and T. Kolda, MATLAB Tensor Toolbox Version 2.2, `http://csmr.ca.sandia.gov/~tgkolda/TensorToolbox/`, January 2007.
- P. Comon, "Independent component analysis: a new concept?," *Signal Processing*, **36** (1994), no. 3, pp. 287–314.
- R.J. Davies, H.M. Kat, and S. Lu, "Fund of hedge funds portfolio selection: a multiple-objective approach," (2006), *Cass Business School Research Paper*.
- L. De Lathauwer, B. De Moor, and J. Vandewalle, "An introduction to independent component analysis," *Journal of Chemometrics* **14** (2000), no. 3, pp. 123–149.
- R.A. Fisher, "Moments and product moments of sampling distributions," *Proceedings of the London Mathematical Society*, **30** (1929), pp. 199–238.
- D.G. Kaiser, D. Schweizer, and L. Wu, "Strategic hedge fund portfolio construction that incorporates higher moments," 2008.

# References


- J.M. Landsberg and J. Morton, *The Geometry of Tensors: Applications to complexity, statistics and engineering*, Book draft.
- J. Marcinkiewicz, "Sur une propriete de la loi de Gauss," Math. Z. 44, (1938) 612-618.
- J.M. Mendel, "Tutorial on higher-order statistics (spectra) in signal processing and system theory: theoretical results and some applications," *Proceedings of the IEEE*, **79** (1991), no. 3, pp. 278–305.
- P. McCullagh, *Tensor methods in statistics*, Chapman and Hall, 1987.
- J. Nocedal and S. Wright, Numerical Optimization (2nd ed.), Berlin, New York: Springer-Verlag, 2006.

# References


- C.L. Nikias and J.M. Mendel, Signal processing with higher-order spectra," *Signal Processing*, **10** (1993), no. 3, pp. 10–37.
- M. Rubinstein, E. Jurczenko, and B. Maillet, *Multi-moment asset allocation and pricing models*, Wiley Finance, 2006.
- F. Samaria and A. Harter. Parameterisation of a Stochastic Model for Human Face Identification. In IEEE Workshop on Applications of Computer Vision, Sarasota (Florida), December 1994. Database of Faces courtesy of AT&T Laboratories.
- B. Savas and L.-H. Lim, "Best multilinear rank approximation of tensors and symmetric tensors with quasi-Newton methods on Grassmannians," *preprint*, October 2008.
- A. Swamia, G.B. Giannakisb, G. Zhou, "Bibliography on higher-order statistics," *Signal Processing*, **60** (1997), no. 1, pp. 65–126.

# References


- M. Turk and A. Pentland. Face Recognition Using Eigenfaces. Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, pp. 586-591, 1991.
- J. Weyman, Cohomology of vector bundles and syzygies, Cambridge University Press, 2003.

End

jason@math.stanford.edu