Majorization and the Lorenz order in statistics, applied probability, economics and beyond

Christian Kleiber

Universität Basel

Vienna, 2017-01-13

Introduction

How can we measure

inequality, variability, diversity, disorder ('chaos'), ...?

Numerous proposals in

- statistics
- economics
- physics
- biology/ecology
- ...

Many parallel developments.

Outline

1. Introduction
2. Majorization
3. Schur convexity
4. Lorenz order
5. Selected applications
   - Taxes and incomes
   - Condorcet jury theorems
   - Portfolio allocation and value at risk
6. Some new results
   - Lorenz ordering of beta distributions
   - Spectra of correlation matrices
   - Schur properties of win-probabilities
7. Concluding remarks

Majorization

Given two vectors

$$\mathbf{x} = (x_1, \ldots, x_n), \qquad \mathbf{y} = (y_1, \ldots, y_n)$$

of equal length n with

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$$

define **majorization** as

$$\mathbf{x} \geqslant_M \mathbf{y} \quad :\Longleftrightarrow \quad \sum_{i=1}^{k} x_{(i:n)} \geqslant \sum_{i=1}^{k} y_{(i:n)}, \quad k = 1, \ldots, n-1.$$

Here $x_{(1:n)} \geqslant x_{(2:n)} \geqslant \cdots \geqslant x_{(n:n)}$ (decreasing rearrangement).

Majorization

Basic properties best explained in terms of income (re)distribution.

Examples.

$$(1,1,1,1) \leqslant_M (2,1,1,0) \leqslant_M (3,1,0,0) \leqslant_M (4,0,0,0)$$

Note: ordering irrelevant, also have

$$(1,1,1,1) \leqslant_M (0,2,1,1) \leqslant_M (1,0,0,3) \leqslant_M (0,4,0,0)$$

More generally

$$(\bar{x}, \bar{x}, \dots, \bar{x}) \leqslant_M (x_1, x_2, \dots, x_n) \leqslant_M (x_1 + x_2 + \cdots + x_n, 0, \dots, 0)$$

Majorization

Interpretation. comparison of income distributions

- identical total incomes
  (majorization describes distributive aspects)
- identical size of populations

Transition from x to y is result of finitely many "Robin Hood transfers":

Majorization and transfers. The following are equivalent

- $x \geqslant_M y$
- $y = T_1 T_2 \cdots T_m x$, with T_i matrix representing 'elementary transfers',
  $T = \epsilon I + (1 - \epsilon) P$ (P 'elementary' permutation matrix)

Majorization

Some pioneers.

- R. F. Muirhead (1903)
- M. O. Lorenz (1905)
- H. Dalton (1920)
- I. Schur (1923)
- G. H. Hardy, J. E. Littlewood and G. Pólya (1929, 1934)

Majorization

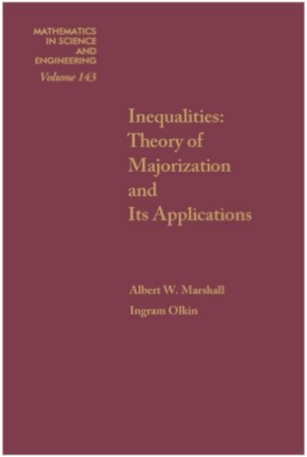
Majorization

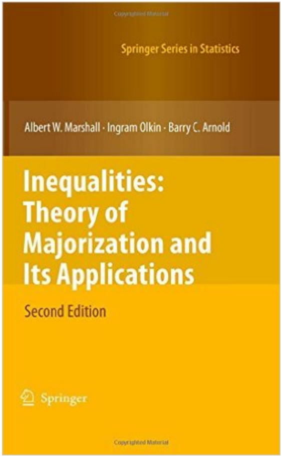
Some references.

- PM Alberti and A Uhlmann (1981). *Stochasticity and Partial Order*, Verlag der Wissenschaften.
- BC Arnold (1987). *Majorization and the Lorenz Order*, Springer-Verlag.
- R Bhatia (1997). *Matrix Analysis*, Springer-Verlag.
- GH Hardy, JE Littlewood and G Pólya (1934). *Inequalities*, Cambridge.
- AW Marshall and I Olkin (1979). *Inequalities: Theory of Majorization and Its Applications*, Academic Press.

  [2nd ed. 2011, with BC Arnold.]
- JM Steele (2004). *The Cauchy-Schwarz Masterclass*, Cambridge.

Majorization





Majorization and Schur convexity

Schur functions

- g Schur convex iff $x \geqslant_M y \quad \Rightarrow \quad g(x) \geqslant g(y)$
- g Schur concave iff $x \geqslant_M y \quad \Rightarrow \quad g(x) \leqslant g(y)$

Unfortunate terminology . . . a *monotonicity* property.

HLP characterization (1934)

The following are equivalent:

- $x \geqslant_M y$
- $y = Px$, P doubly stochastic matrix
- $\sum_i h(x_i) \geqslant \sum_i h(y_i)$ for all (continuous) convex functions h

Not every analytic inequality is a consequence of the Schur convexity of some function, but enough are to make familiarity with majorization/Schur convexity a nece[s]sary part of the required background of a respectable mathematical analyst. *(Arnold 1987)*

Majorization and Schur convexity

How to recognize Schur concave/convex functions?

Schur's criterion (1923)

Continuously differentiable g, permutation symmetric, is Schur convex (concave) if, for all i, j,

$$(x_i - x_j)\left(\frac{\partial g(x)}{\partial x_i} - \frac{\partial g(x)}{\partial x_j}\right) \quad \geqslant (\leqslant) \quad 0$$

Remark on terminology: (convexity connection)

Why 'convex'? For f convex, composite function

$$g(x) := \sum_i f(x_i)$$

is Schur convex. Also have various representations involving doubly stochastic matrices, specific convex functions, etc.

Majorization and Schur convexity

Examples: Classical inequality measures are Schur convex in incomes

- Gini

$$G \;=\; 2 \cdot \text{ concentration area}$$

- coefficient of variation (squared)

$$CV^2 \;=\; \frac{1}{n}\sum_i \left(\frac{x_i}{\bar{x}} - 1\right)^2$$

- Theil

$$T \;=\; \frac{1}{n}\sum_i \frac{x_i}{\bar{x}} \log \frac{x_i}{\bar{x}}$$

- Atkinson

$$A_\epsilon \;=\; 1 - \left\{ \frac{1}{n}\sum_i \left(\frac{x_i}{\bar{x}}\right)^{1-\epsilon} \right\}^{1/(1-\epsilon)}$$

The Lorenz order

Majorization not sufficiently general for many tasks:

- identical population size?
- identical total incomes?

Suggestion of Max Otto Lorenz (1905):

Lorenz curve

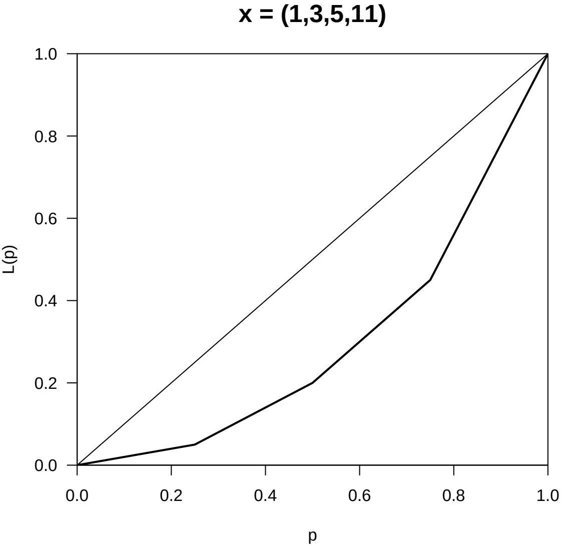
For $\mathbf{x} = (x_1, \ldots, x_n)$, $x_i \geqslant 0$, $\sum_{i=1}^{n} x_i > 0$, define *Lorenz curve* via linear interpolation of $\quad$ ($x_{i:n}$ increasingly ordered)

$$L\left(\frac{k}{n}\right) = \frac{\sum_{i=1}^{k} x_{i:n}}{\sum_{i=1}^{n} x_{i:n}}, \quad k = 0, 1, \ldots, n.$$

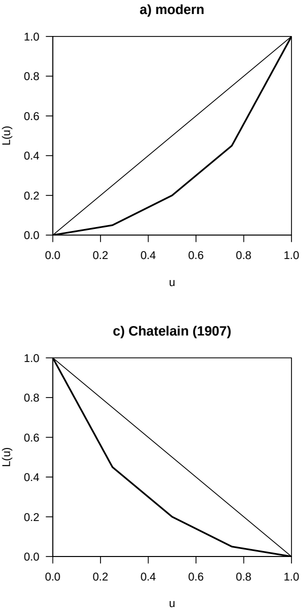
Interpretation:

"poorest $\frac{k}{n} \cdot 100\%$ possess $\frac{\sum_{i=1}^{k} x_{i:n}}{\sum_{i=1}^{n} x_{i:n}}$ of total income"

The Lorenz order



The Lorenz order



The Lorenz order

Lorenz curve (Pietra 1915, Piesch 1967, Gastwirth 1971)

For non-negative X with $0 < E(X) < \infty$, set

$$L_X(u) \;=\; \frac{1}{E(X)} \int_0^u F_X^{-1}(t)\, dt\,, \quad u \in [0,1].$$

Properties.

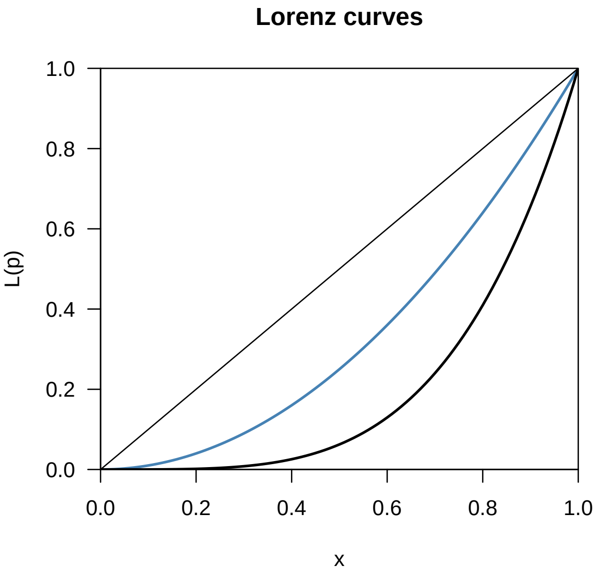
- L continuous on $[0,1]$, with $L(0) = 0$ and $L(1) = 1$,
- L monotonically increasing, and
- L convex.

Lorenz order

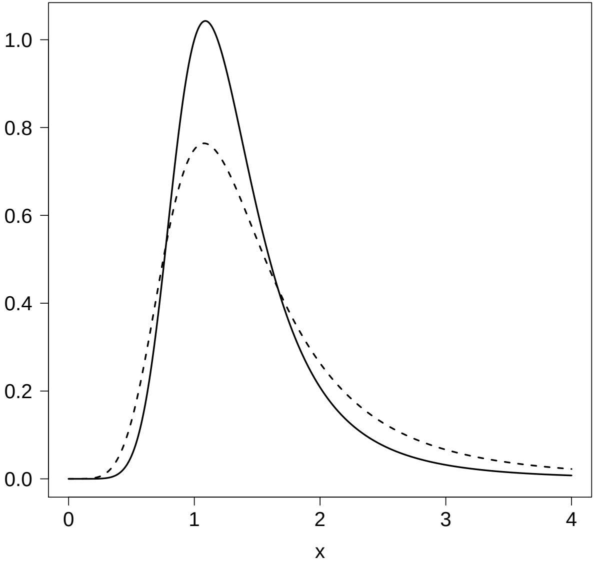
X_1 more unequal (... or more spread out ... or more variable) than X_2 in the Lorenz sense, if $L_1(u) \leqslant L_2(u)$ for all $u \in [0,1]$. Notation:

$$X_1 \geqslant_L X_2 \quad :\Longleftrightarrow \quad L_1 \leqslant L_2.$$

The Lorenz order



The Lorenz order



Applications of majorization and the Lorenz order

'Random' paper in statistical distribution theory:

Kochar and Xu (*J Mult Anal* 2010) show for exponential distribution:

Suppose $X_i \sim \text{Exp}(\lambda_i)$ independent.
If $(1/\lambda_1, \ldots, 1/\lambda_n) \geqslant_M (1/\lambda_1^*, \ldots, 1/\lambda_n^*)$, then

$$\sum_{i=1}^{n} X_{\lambda_i} \quad \geqslant_L \quad \sum_{i=1}^{n} X_{\lambda_i^*}$$

Nice: Majorization *and* Lorenz order!

Remark. Since 2000 dozens (hundreds?) of papers on distributional inequalities for linear combinations, order statistics etc from heterogeneous populations. Many involve majorization.

Applications of majorization and the Lorenz order

- Mathematics, statistics, actuarial science
  - eigenvalues and diagonal elements of matrices
  - distributions of quadratic forms
  - power functions of tests in multivariate analysis
  - inequalities for special functions
  - distributions of aggregate losses (= random sums)
  - value at risk
  - ...
- Social sciences
  - tax progression and income redistribution
  - Condorcet jury theorems
  - "fair representation" in parliaments
  - ...

Applications of majorization and the Lorenz order

- Often variations on the main theme:
  - majorization of transformations (logarithms, ...)
  - weak majorization (super- or submajorization)
  - ...
- Especially Lorenz ordering results often require background on further stochastic orders to exploit interrelations
  - there are **hundreds** of stochastic orders in statistics, economics, reliability theory, actuarial science, ...
  - Examples include

    stochastic dominance (of various orders), convex order, increasing convex/concave order, star-shaped order, mean residual life (or mean excess) order, hazard rate order, likelihood ratio order, excess wealth order, total time on test, superadditive order, ...

Applications: Taxes and incomes

Framework. Given

- vector of incomes $\mathbf{x} = (x_1, \ldots, x_n)$
- tax schedule $t(x)$
  Call $\{1 - t(x)\}\, x$ after-tax income ("residual income")

Goal. Comparison of before- and after-tax incomes wrt. inequality.
Majorization not applicable because

$$\sum_i x_i \quad \neq \quad \sum_i \{1 - t(x_i)\}\, x_i$$

Use Lorenz order instead.

Question. What does a 'Lorenz-equalizing' tax look like?

Applications: Taxes and incomes

Theorem (Eichhorn, Funke, Richter, *J Math Econ* 1984)

$$x \quad \geqslant_L \quad \{1 - t(x)\}\, x$$

iff

- $t(x)$ increasing and
- $\{1 - t(x)\}\, x$ increasing.

Interpretation. Income tax is inequality-reducing iff

- progressive and
- incentive preserving

Applications: Condorcet jury theorems

Framework. Jury of n 'experts' faces binary decision.

- Suppose $X_i \in \{0, 1\}$ decision of expert i and $p_i = P(X_i = 1)$, $i = 1, \ldots, n$. Call p_i competence/ability of expert i.
- Consider number of correct decisions

$$S := \sum_{i=1}^{n} X_i$$

  If all experts equally competent ($p_i \equiv p$) and independent,

$$P(S \geqslant k) = \sum_{i=k}^{n} \binom{n}{i} p^i (1-p)^{n-i},$$

  a binomial probability.
- Decision is via **majority voting**.
  To avoid ties, set $n = 2m + 1$, hence $k = m + 1$.

Applications: Condorcet jury theorems

Applications: Condorcet jury theorems

Setting of classical CJT.

- two alternatives
- common preferences
  (one alternative is superior in the light of full information)
- independent decisions
- homogeneous competences
- decision rule is simple majority voting

Applications: Condorcet jury theorems

Classical CJTs. (Boland, *JRSS D* 1989)

Non-asymptotic CJT

Under majority voting with $p > 1/2$ ("experts") have

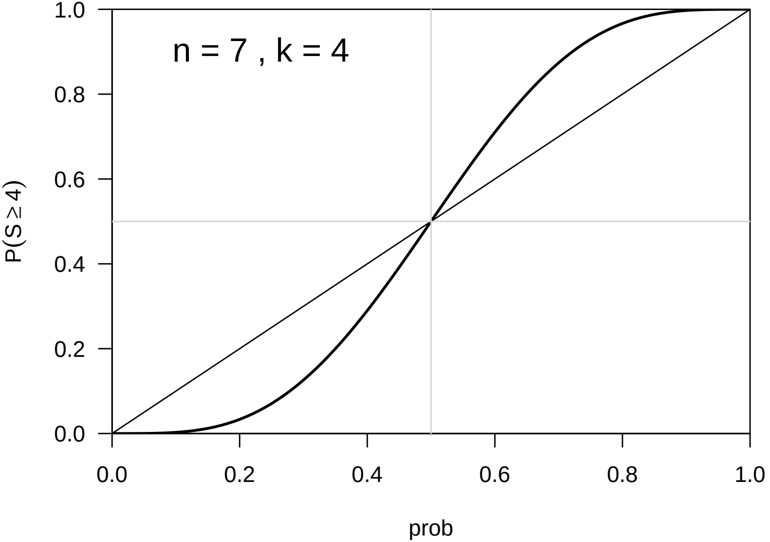
$$P(S \geqslant m+1) \quad > \quad p$$

Proof: use Beta integral representation of binomial probabilities

$$P(S \geqslant m+1) \;=\; \frac{1}{B(m+1, m+1)} \int_0^p t^m (1-t)^m \, dt$$

NB. There is also an **asymptotic CJT**, but not needed here.

Applications: Condorcet jury theorems



Applications: Condorcet jury theorems

Extensions of basic version.

- supermajority voting (also called special majority voting)
- heterogeneous experts
- dependent experts (“opinion leaders”)
- juries of different sizes
- direct vs indirect majority voting ($\rightarrow$ US presidential elections)

Applications: Condorcet jury theorems

Framework. Jury J characterized by vector of probabilities (“competences”)

$$\mathbf{p} \;=\; (p_1, \ldots, p_n) \quad \in \; [0,1]^n$$

Question. Given 2 juries J_1 und J_2 of equal size, with competences $\mathbf{p_1}$ and $\mathbf{p_2}$, when will J_1 do better?

Need conditions for

$$P(S_1 \geqslant m+1) \quad \geqslant \quad P(S_2 \geqslant m+1) \quad \text{for} \;\; \mathbf{p_i} \;\in\; \mathcal{P} \;\subseteq [0,1]^n$$

- New problem: distribution of sums of *independent, but not identically distributed* Bernoulli variables
- Goal: stochastic comparisons with e.g. binomial distribution
- Classical paper: Hoeffding (*Ann Math Stat* 1956)

Applications: Condorcet jury theorems

In Hoeffding (1956) purely probabilistic point of view.

Sums of heterogeneous Bernoullis arise in many contexts

- CJTs
- reliability of "k out of n" systems (unequal default probabilities)
- portfolios of credit risks
- ...

Applications: Condorcet jury theorems

Point of reference. average competence $\bar{p}$

Hoeffding's inequality (Hoeffding 1956)

Suppose $k > 0$ with $\bar{p} \geqslant k/n$. Then

$$P(S \geqslant k) \quad \geqslant \quad \sum_{i=k}^{n} \binom{n}{i} \bar{p}^i (1 - \bar{p})^{n-i}$$

This gives

Boland's CJT (Boland 1989)

Suppose $n \geqslant 3$, $\bar{p} \geqslant 1/2 + 1/(2n)$. Then

$$P(S \geqslant m + 1) \quad > \quad \bar{p}$$

Applications: Condorcet jury theorems

Generalization of Hoeffding's inequality:

Gleser's inequality (*Ann Prob* 1975)

Let $\mathbf{p}_1 \geqslant_M \mathbf{p}_2$. Then

$$P(S \leqslant k \mid \mathbf{p}_1) \quad \leqslant \quad P(S \leqslant k \mid \mathbf{p}_2), \quad k \leqslant \lfloor n\bar{p} - 2 \rfloor$$

This gives

CJT under heterogeneity

Let $n \geqslant 7$ and $\bar{p} \geqslant 1/2 + 5/(2n)$. If $\mathbf{p}_1 \geqslant_M \mathbf{p}_2$ then

$$P(S \geqslant m + 1 \mid \mathbf{p}_1) \quad \geqslant \quad P(S \geqslant m + 1 \mid \mathbf{p}_2)$$

Note: need large $\bar{p}$ for superiority of majority voting!

Applications: Condorcet jury theorems

Further generalization of Hoeffding's inequality:

Boland and Proschan's inequality (*Ann Prob* 1983)

Let $\mathbf{p}_1 \geqslant_M \mathbf{p}_2$. Then

$$P(S \leqslant k \mid \mathbf{p}_1) \quad \leqslant \quad P(S \leqslant k \mid \mathbf{p}_2), \quad \text{all } \ p_i \in [(k-1)/(n-1), 1]^n$$

This gives

CJT under heterogeneity

Let $p_i \in [1/2, 1]^n$ with $\mathbf{p}_1 \geqslant_M \mathbf{p}_2$. Then

$$P(S \geqslant m+1 \mid \mathbf{p}_1) \quad \geqslant \quad P(S \geqslant m+1 \mid \mathbf{p}_2)$$

This differs from the Gleser version!

Can be generalized to supermajority voting.

Applications: Condorcet jury theorems

Visualization via Lorenz curves

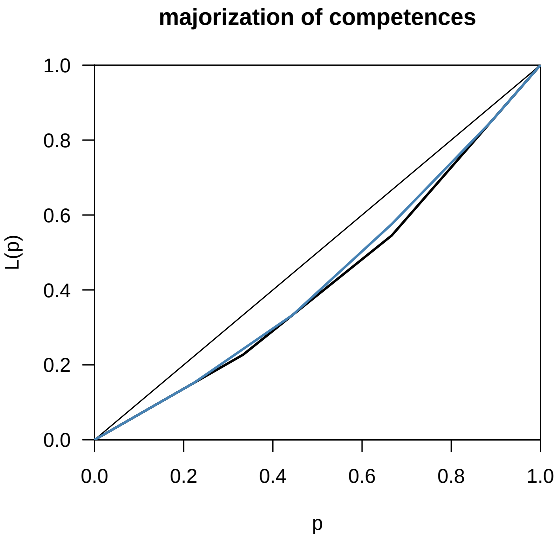
$$L\left(\frac{k}{n}\right) = \frac{\sum_{i=1}^{k} x_{i:n}}{\sum_{i=1}^{n} x_{i:n}}, \quad k = 0, 1, \ldots, n,$$

where $x_{i:n}$ ith smallest income $\rightarrow$ consider probabilities as incomes

Example: $n = 9$, $\bar{p} = 0.6$

```
p1 <- c(1.0, 1.0, 1.0, 0.7, 0.7, 0.7, 0.5, 0.5, 0.5)
p2 <- c(1.0, 0.9, 0.9, 0.8, 0.8, 0.6, 0.6, 0.5, 0.5)
```

Applications: Condorcet jury theorems



Portfolio allocation and value at risk

Conventional wisdom in portfolio allocation:

Diversification reduces risk.

Q. Really ...?

Schur properties of VaR (Ibragimov, *Quant Fin* 2009)

Consider portfolios $Y_a = \sum_i a_i Y_i$ and $Y_b = \sum_i b_i Y_i$, and $\alpha < \frac{1}{2}$. Then

- $a \geqslant_M b \implies VaR_\alpha(Y_a) \geqslant VaR_\alpha(Y_b)$ for Y_i light-tailed.
- $a \geqslant_M b \implies VaR_\alpha(Y_a) \leqslant VaR_\alpha(Y_b)$ for Y_i (very) heavy-tailed.

Applications: Lorenz ordering of beta distributions

Consider beta distribution $\beta(p, q)$

$$f(x) = \frac{1}{B(p,q)}\, x^{p-1}\, (1-x)^{q-1}, \qquad x \in [0,1].$$

Q. Let $X_i \sim \beta(p_i, q_i)$, $i = 1, 2$. When do we have $X_1 \geqslant_L X_2$?

Many applications: Order statistics, reliability, actuarial science, ...

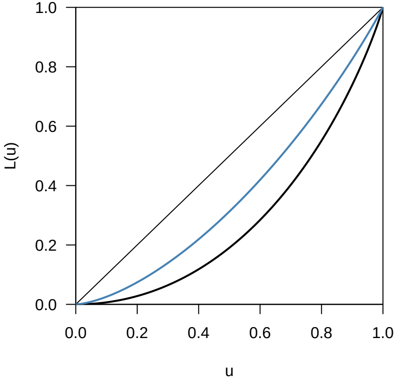
Partial results:

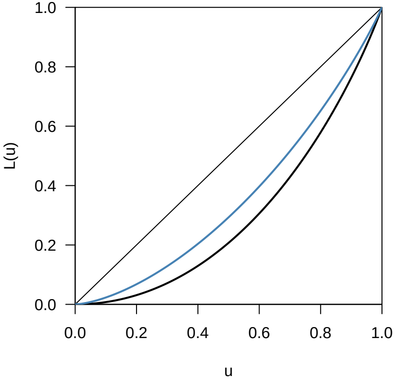
- $X_1 \geqslant_L X_2$ implies $p_1 \leqslant p_2$ and $p_1/p_2 \leqslant q_1/q_2$
- $\beta(p, q) \geqslant_L \beta(q, p) \Longleftrightarrow p \leqslant q$
- Let $X_i \sim \beta(p_i, p_i)$, $i = 1, 2$. Then $X_1 \geqslant_L X_2 \Longleftrightarrow p_1 \leqslant p_2$.
- $p_1 \leqslant p_2$ and $q_1 \geqslant q_2$ imply $X_1 \geqslant_L X_2$.

Tools: relations for tailweight, log-concavity, beta-gamma algebra.

Remark. Can be translated into (obscure?) inequalities for regularized incomplete beta function.

Applications: Lorenz ordering of beta distributions





$\beta(1,3) \geqslant_L \beta(2,2)$ (proof!) $\beta(1,2) \geqslant_L \beta(2,3)$ (no proof ...)

Applications: Spectra of correlation matrices

Q: How to compare correlation matrices of time series models?

Consider AR(1) process

$$y_t \;=\; \rho y_{t-1} + \varepsilon_t$$

and (auto)correlation matrix

$$R_\rho = (\rho^{|i-j|})_{i,j=1,\dots,T}$$

Obvious: process is more persistent for larger ρ.

Can say more: Spectra of correlation matrices are ordered

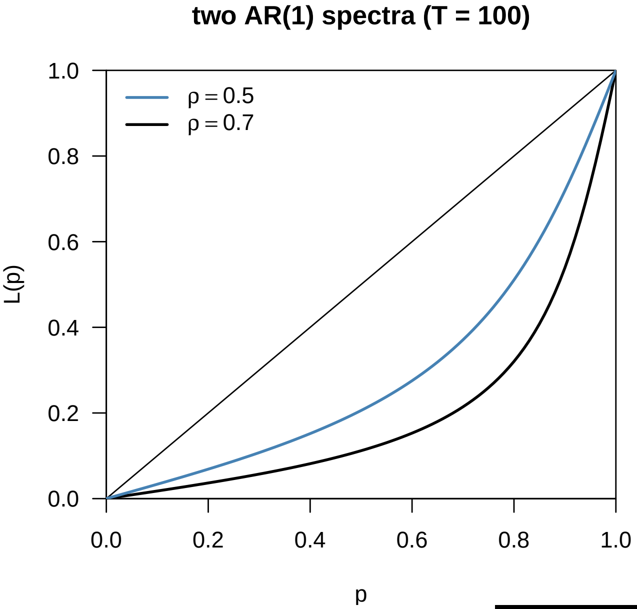
$$\rho_1 \;\leqslant\; \rho_2 \quad\Longrightarrow\quad \lambda(R_{\rho_1}) \;\leqslant_M\; \lambda(R_{\rho_2})$$

Further examples:

- MA(1) processes
- equicorrelation matrices $(1-\rho)I + \rho 11^\top$

Ingredients: Majorization inequalities for Schur products.

Applications: Spectra of correlation matrices



Applications: Win-probabilities

Consider random variables $X_1, \dots, X_k$, independent.

Win-probability for 'treatment' X_k is

$$W^U(k; 1, \dots, k-1) = P\left(X_k > \max_{1 \leqslant j \leqslant k-1} X_j\right) = \int_{\mathbb{R}} f_k(x) \prod_{j=1}^{k-1} F_j(x)\, dx$$

Example: Let $k = 3$ and $X_j \sim \mathsf{Exp}(\lambda_j)$, independent.

With $\rho_i = \lambda_i / \lambda_3$, $i = 1, 2$, have

$$W^U(3; 1, 2 \mid \rho) = 1 - \frac{1}{\rho_1 + 1} - \frac{1}{\rho_2 + 1} + \frac{1}{\rho_1 + \rho_2 + 1}$$

This is Schur-concave in $\rho = (\rho_1, \rho_2)^\top$. Thus

$$\rho \geqslant_M \tau \implies W^U(\ldots \mid \rho) \leqslant W^U(\ldots \mid \tau)$$

Applications: Win-probabilities

Remarks:

- works for $k > 3$
- works for Pareto
- works for Weibull with common shape
- similar for W^L 'lower win (lose?) probability'
- related to stress-strength models in reliability

Concluding remarks

Majorization has many applications, not only in mathematics.

Classical problem: (majorization)

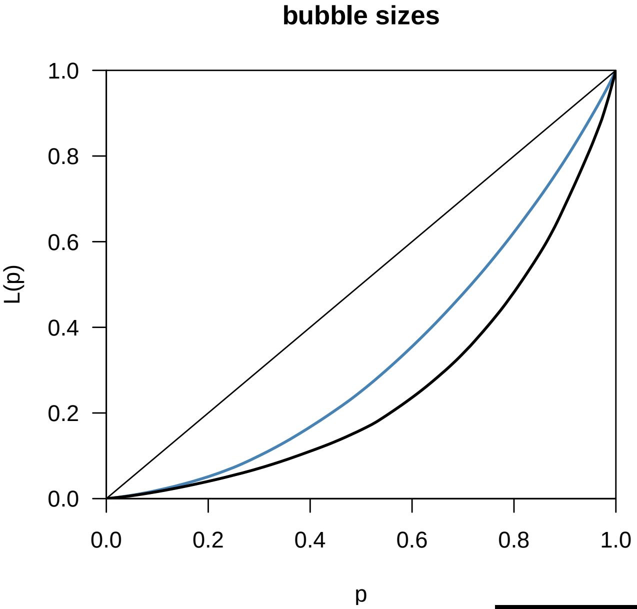
$$a \geqslant_M b \quad \Longrightarrow \quad f(a) \geqslant (\leqslant) f(b)$$

Open problem: (Lorenz order)

$$a \geqslant_L b \quad \text{?} \quad f(a) \geqslant (\leqslant) f(b)$$

- Lorenz order is less widely known but potentially more useful
- Lorenz curve is useful for visualizing majorization inequalities
  ... and for hypothesizing theorems (!)
- many majorization and Lorenz ordering results remain to be discovered

Applications: Chemistry



Applications: Schur-Horn theorem

Problem. Relation between eigenvalues λ_i and diagonal elements a_{ii} of a symmetric matrix A?

Note $\mathrm{tr}(A) = \sum_j \lambda_j$, hence majorization meaningful.

Schur (1923) shows

$$(a_{11}, a_{22}, \ldots, a_{nn}) \quad \leqslant_M \quad (\lambda_1, \lambda_2, \ldots, \lambda_n)$$

This implies **Hadamard's inequality**:
For any real, symmetric matrix

$$\prod_i a_{ii} \geqslant \prod_i \lambda_i$$

Applications: Schur-Horn theorem

But there is more:

Schur-Horn theorem. Suppose $a, b \in \mathbb{R}^n$ with $a \leqslant_M b$.
Then there exists a real, symmetric matrix A with diagonal a and eigenvalues b.

Recent abstract version: majorization of sequences implies existence of compact operator with suitable eigenvalues, etc.

Applications: Credit risks

Framework. n credit risks X_i described by sizes a_i, $i = 1, \dots, n$, and (possibly distinct) default probabilities p_i.

Quantities of interest:

- number of defaults $\sum_i X_i$, $X_i \sim \text{Bin}(1, p_i)$
- aggregate losses $\sum_i a_i X_i$, $X_i \sim \text{Bin}(1, p_i)$

Result on number of defaults.
If $\mathbf{p_{(1)}} \geqslant_M \mathbf{p_{(2)}}$ and risks independent, then

$$\text{Var}\left(\sum_i X_i \mid \mathbf{p_{(1)}}\right) \quad \leqslant \quad \text{Var}\left(\sum_i X_i \mid \mathbf{p_{(2)}}\right)$$

Proof: variance is Schur concave in p

Can also use Hoeffding etc bounds ... but they provide lower bounds on probabilities.

Applications: Credit risks

Result on aggregate losses.
This requires assumption on a_i s. Suppose a_i decreasing in p_i.
Assume

$$a_i p_i \quad \approx \quad const. =: a$$

hence consider

$$\sum a_i X_i \;=\; a \sum \frac{1}{p_i} X_i, \quad \text{wlog } a = 1$$

If $\mathbf{p_{(1)}} \geqslant_M \mathbf{p_{(2)}}$ and risks independent, then

$$\mathsf{Var}\left(\sum_i a_i X_i \mid p_{(1)} \right) \quad \geqslant \quad \mathsf{Var}\left(\sum_i a_i X_i \mid p_{(2)} \right)$$

Proof: variance is Schur concave in p

Majorization and Schur convexity

Axiomatic approach to inequality measurement.

For a scalar measure of inequality I, require (at least) the following properties:

- $I(x) = I(\lambda x)$ for $\lambda > 0$ (homogeneity of degree 0)
- for $x \geqslant_M y$ must have $I(x) \geqslant I(y)$ (Schur convexity)
- $I((x, x)) = I(x)$ (population principle)