On Models & Theory

Elchanan Mossel

MIT

IMAG 10th Anniversary MSM Meeting
Mar 24, 2017

All models are bad

- All models are bad
- but some are useful (Box and Norman, 87)
- Which ones?

Couplas in finance

- David X. Li, a ”mathematical genius”, ”invented” the Gaussian coupla function (2000) to ”compute” the correlation in risk:

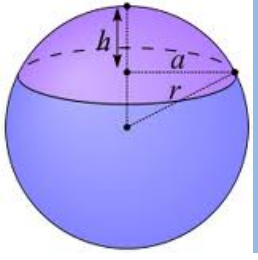
$$P[X \leq u, Y \leq v], \quad (X, Y) \sim N(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix})$$

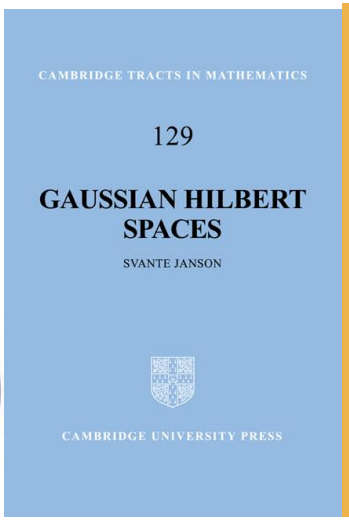
Couplas in finance

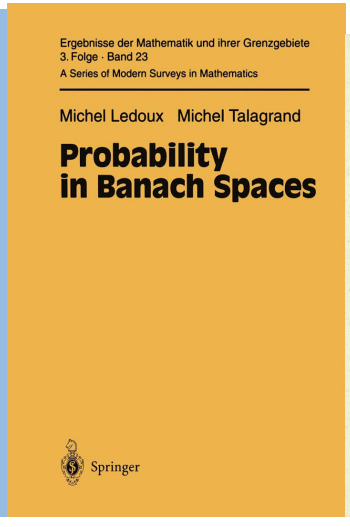
- David X. Li, a ”mathematical genius”, ”invented” the Gaussian coupla function (2000) to ”compute” the correlation in risk:

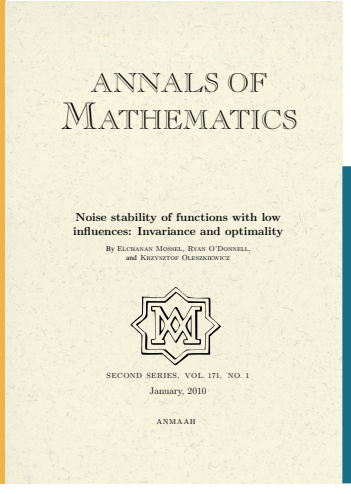
$$P[X \leq u, Y \leq v], \quad (X, Y) \sim N(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix})$$

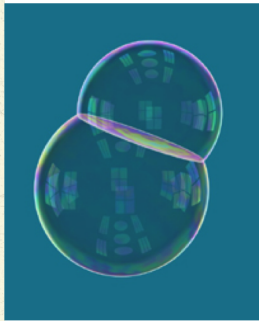
- Did not consult the leading mathematicians in the area
- Borell, Janson, Ledoux, M ...









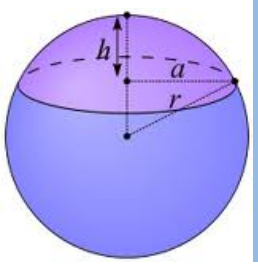


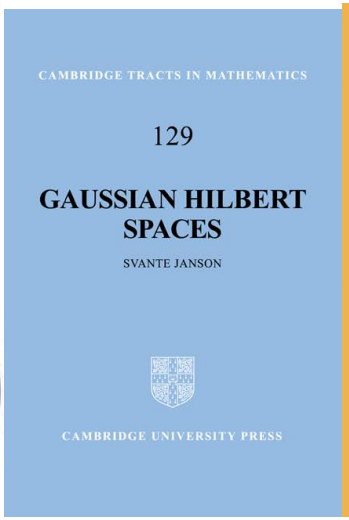
Couplas in finance

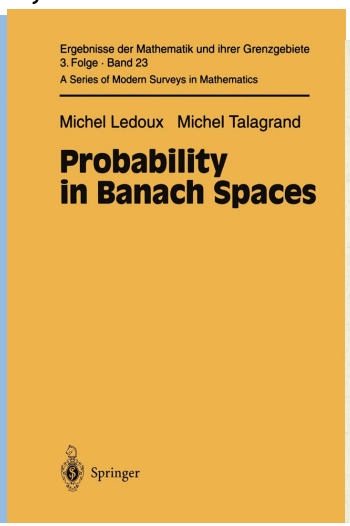
- David X. Li, a ”mathematical genius”, ”invented” the Gaussian coupla function (2000) to ”compute” the correlation in risk:

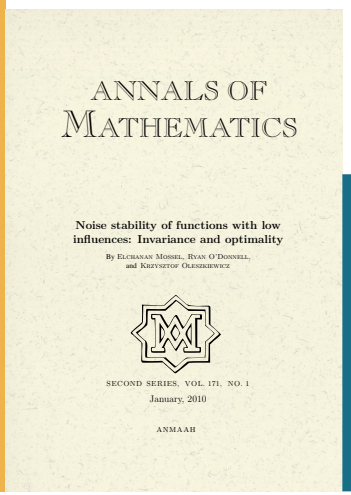
$$P[X \leq u, Y \leq v], \quad (X, Y) \sim N(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix})$$

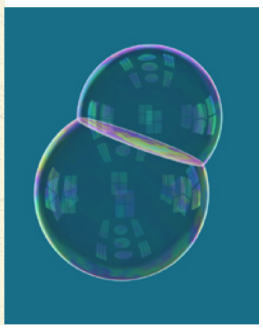
- Did not consult the leading mathematicians in the area
- Borell, Janson, Ledoux, M ...







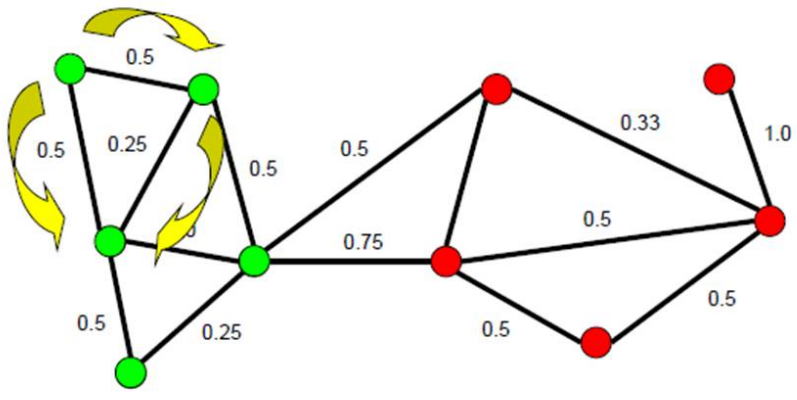




- Lesson 1: Danger in using math you do not understand.

Cascades in Social Networks

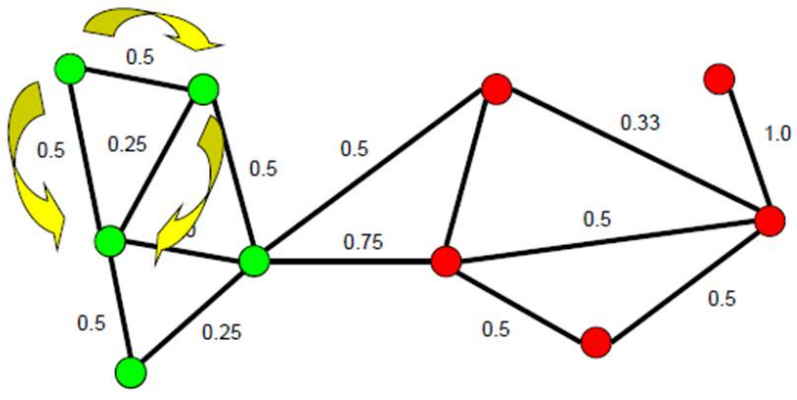
- Kempe, Kleinberg and Tardos 03,05, Mossel-Roch-05:
- Elegant model of spread on social networks.
- $+$ Assuming the model, rigorous proof for excellent performance for finding the most influential sets.



Courtesy of S.Roch

Cascades in Social Networks

- Kempe, Kleinberg and Tardos 03,05, Mossel-Roch-05:
- Elegant model of spread on social networks.
- $+$ Assuming the model, rigorous proof for excellent performance for finding the most influential sets.

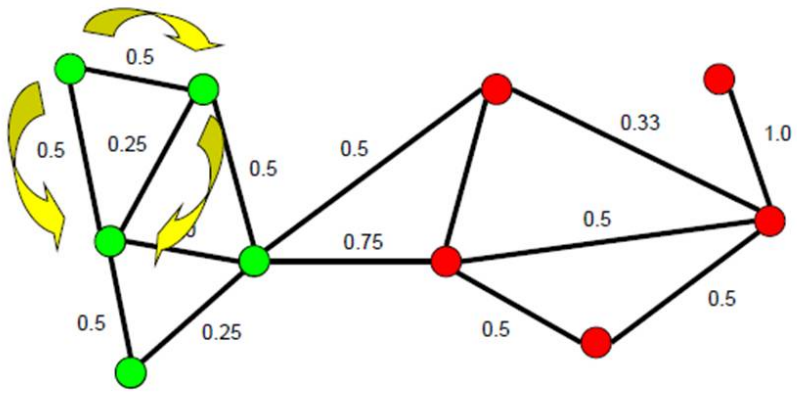


Courtesy of S.Roch

- Consulting experience: a terrible model.

Cascades in Social Networks

- Kempe, Kleinberg and Tardos 03,05, Mossel-Roch-05:
- Elegant model of spread on social networks.
- $+$ Assuming the model, rigorous proof for excellent performance for finding the most influential sets.



Courtesy of S.Roch

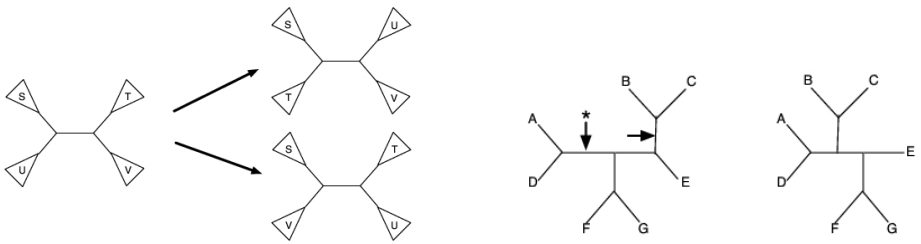
- Consulting experience: a terrible model.
- Lesson 2: Beautiful models are not enough. Area expertise is needed.

Phylogenetic Mixtures

- What happens if you try to infer evolution making reasonable assumption A.
- While in fact a different reasonable assumption B. holds.

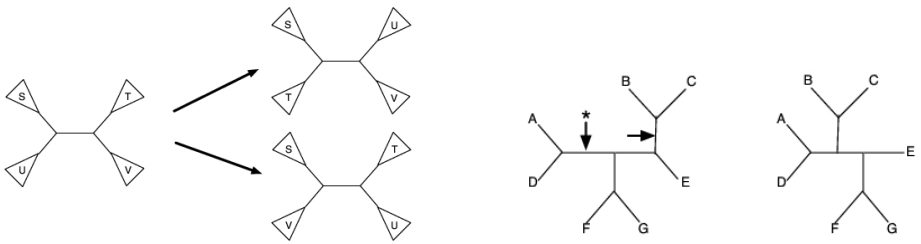
Phylogenetic Mixtures

- What happens if you try to infer evolution making reasonable assumption A.
- While in fact a different reasonable assumption B. holds.
- M-Vigoda (Science 05): You may conclude that
  - A is true and
  - your evolution recovery is accurate (excellent ”posterior” or ”p-value”).



Phylogenetic Mixtures

- What happens if you try to infer evolution making reasonable assumption A.
- While in fact a different reasonable assumption B. holds.
- M-Vigoda (Science 05): You may conclude that
  - A is true and
  - your evolution recovery is accurate (excellent ”posterior” or ”p-value”).



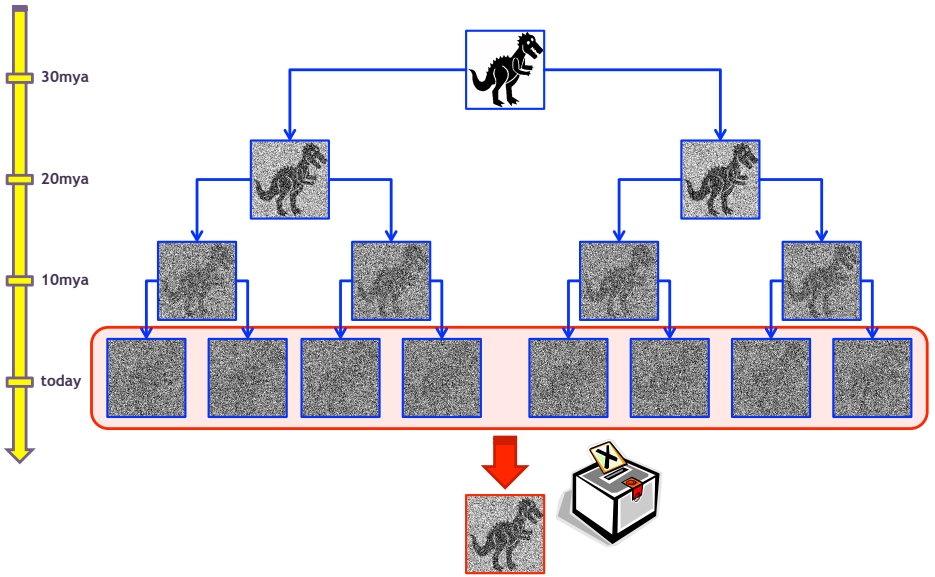
- Lesson 3: Important to explicitly state modeling assumption, so experts can evaluate these assumptions.

A success story: Phylogenetic reconstruction - a success story

- Steel, M, Roch (starting $\sim$ 2002) ...
- Theory provides excellent practical predictions:
- Information theory bounds, phase transitions, gene trees vs. species trees etc.
- Why?

A success story: Phylogenetic reconstruction - a success story

- Steel, M, Roch (starting $\sim$ 2002) ...
- Theory provides excellent practical predictions:
- Information theory bounds, phase transitions, gene trees vs. species trees etc.
- Why?
- Good feedback between specialist in biology, phylogenetic software developers, theoreticians.



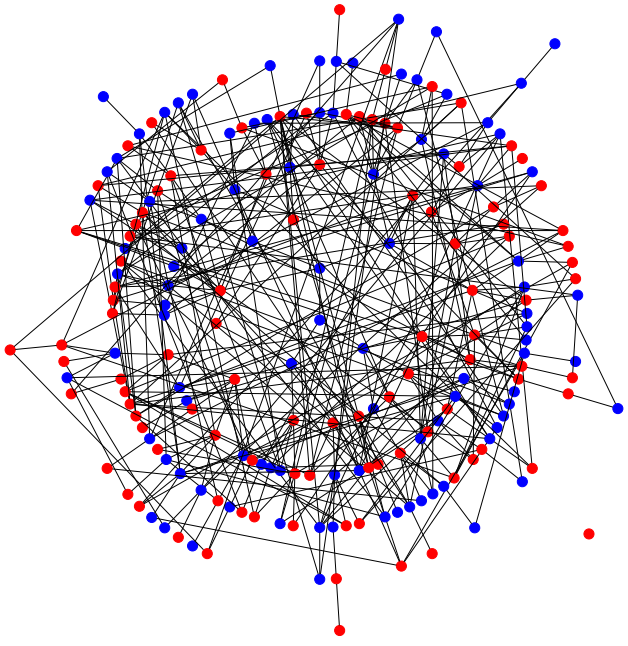
Block Models and non backtracking walks

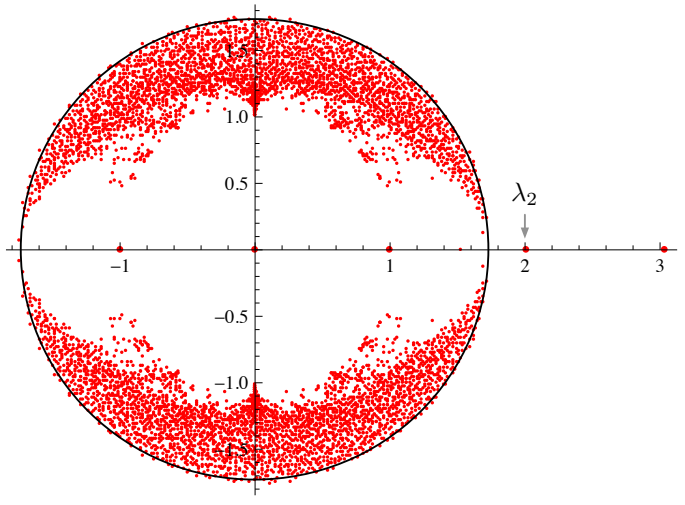
- Simple (but really terrible) model of communities in graphs.
- Lets us ask very precise mathematical questions: phase transitions, optimal algorithms? (Decelle,Moore,Zdeborova,Zhang ... 2011)
- Lead to new practical, fast algorithms ($+$ M-Neeman-Sly, Bordenave,Lelarge,Massouille, ...)
- Why?

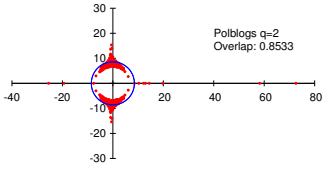
Block Models and non backtracking walks

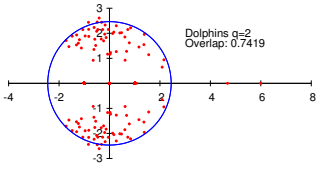
- Simple (but really terrible) model of communities in graphs.
- Lets us ask very precise mathematical questions: phase transitions, optimal algorithms? (Decelle,Moore,Zdeborova,Zhang ... 2011)
- Lead to new practical, fast algorithms (+ M-Neeman-Sly, Bordenave,Lelarge,Massouille, ...)
- Why?
- A **bad model** but a **simple** one.
- Bridging communities: data science, stat. physics, math and computer science.

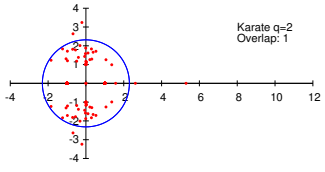
Block Models and non backtracking walks

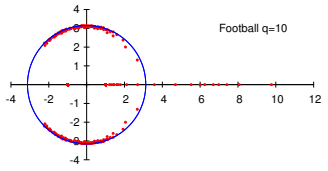


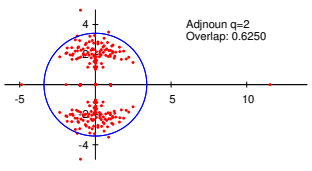


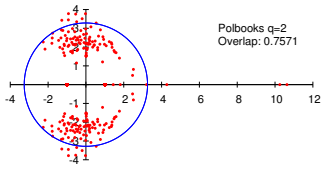












Conclusion

- Lessons: Know your math, be an expert in the area, transparent.
- Integration between different disciplines increasingly necessary.
- The simpler the better.