

PSI

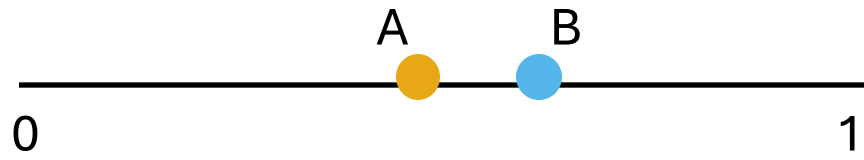
Center for Nuclear Engineering and Sciences
Center for Energy and Environmental Sciences

MaxEnt_Disaggregation

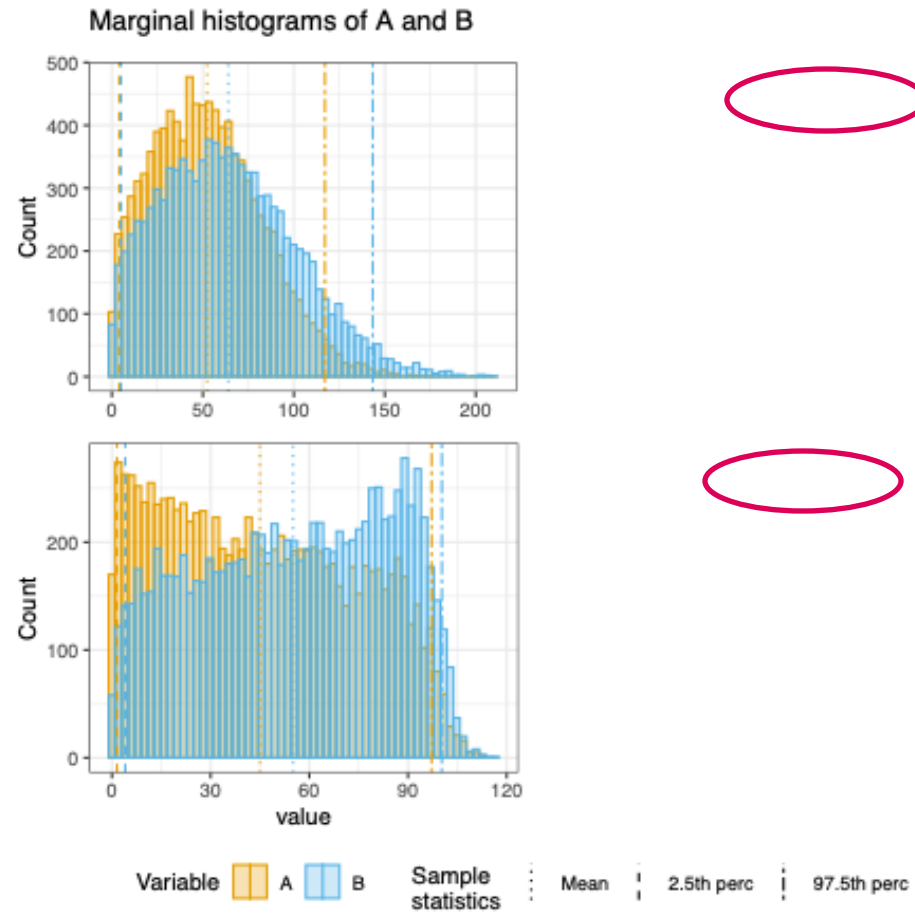
Divide *your data* and Rule *the correlations!*

Arthur Jakobs, Simon Schulte, Rick Lupton
BrightCon Grenoble October 2025

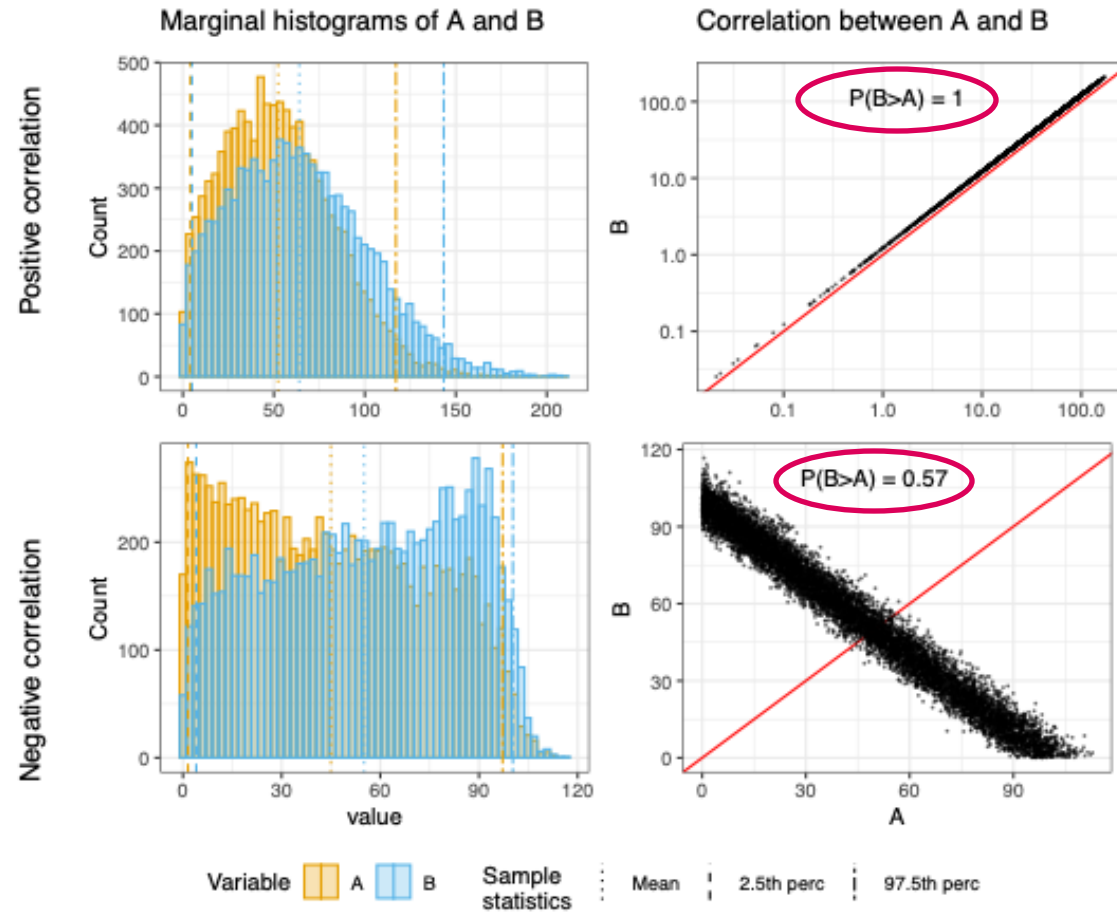
Question: $B > A$?



Question: $B > A$?



Question: Uncertainty B + A ?

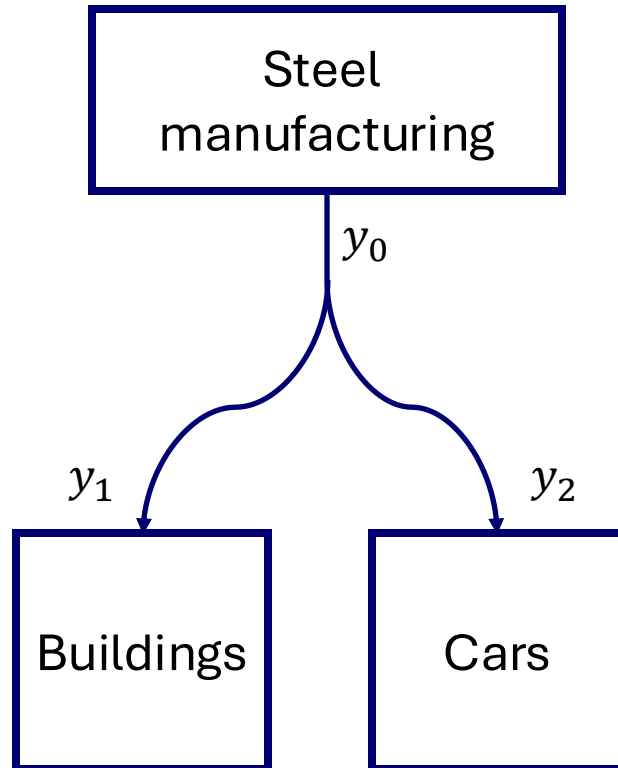


Sources of correlations in uncertainty



- Many **real-life sources** of **correlations** (e.g. mass balances):
 - Need to be modelled individually if information is present
- **Statistical correlations** due to data manipulation:
 - Disaggregation (uncertainty on shares)

The disaggregation problem



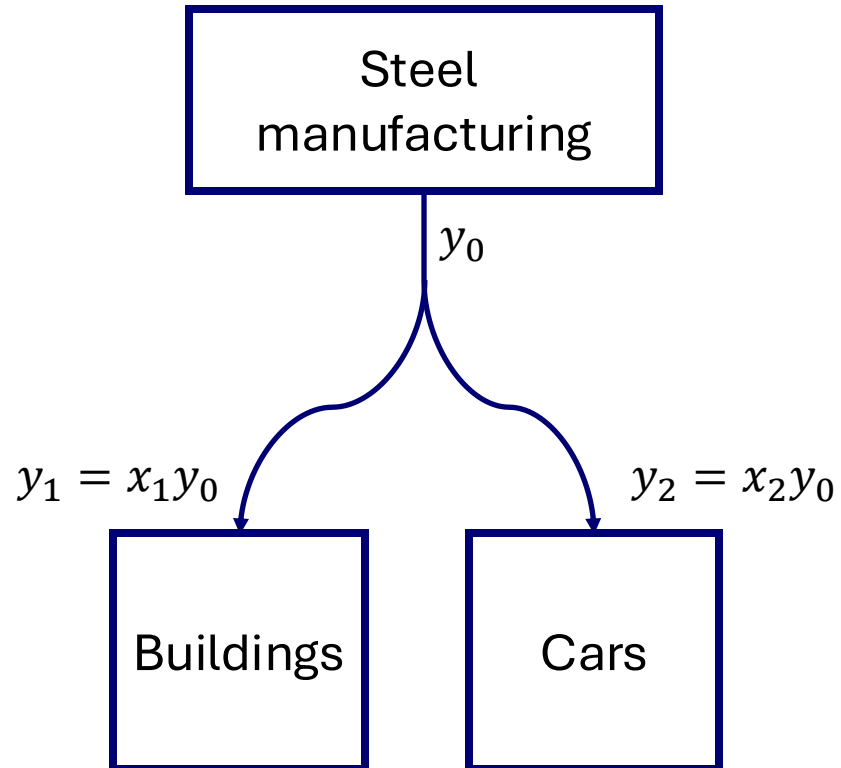
- **HAVE:** Total steel manufactured: y_0
- **WANT:**
Amount of steel going to different sectors: $y_1, \dots, y_k$
- **KNOW:** $y_0 = \sum_{i=1}^k y_i$
- **PROXY:** e.g. relative monetary output of sectors x_i
- **RESULT:** $y_i = x_i y_0$ where: $\sum_{i=1}^k x_i = 1$, shares of y_0

The disaggregation problem in IE research



- **MFA** (see previous example, or Streeck et al. 2023 doi: 10.1111/jiec.13380)
- **Input-output analysis: allocation** of emissions to end-use sectors
(Schulte et al. 2024 doi: 10.5194/essd-16-2669-2024)
- **LCA: allocation** of inputs and emissions to co-products (often hidden in the background database, might be a foreground issue as well)
- **LCA** correlated **sampling of market shares**
(Kim et al. 2025 doi: 10.1111/jiec.70036)

Choosing the right distribution



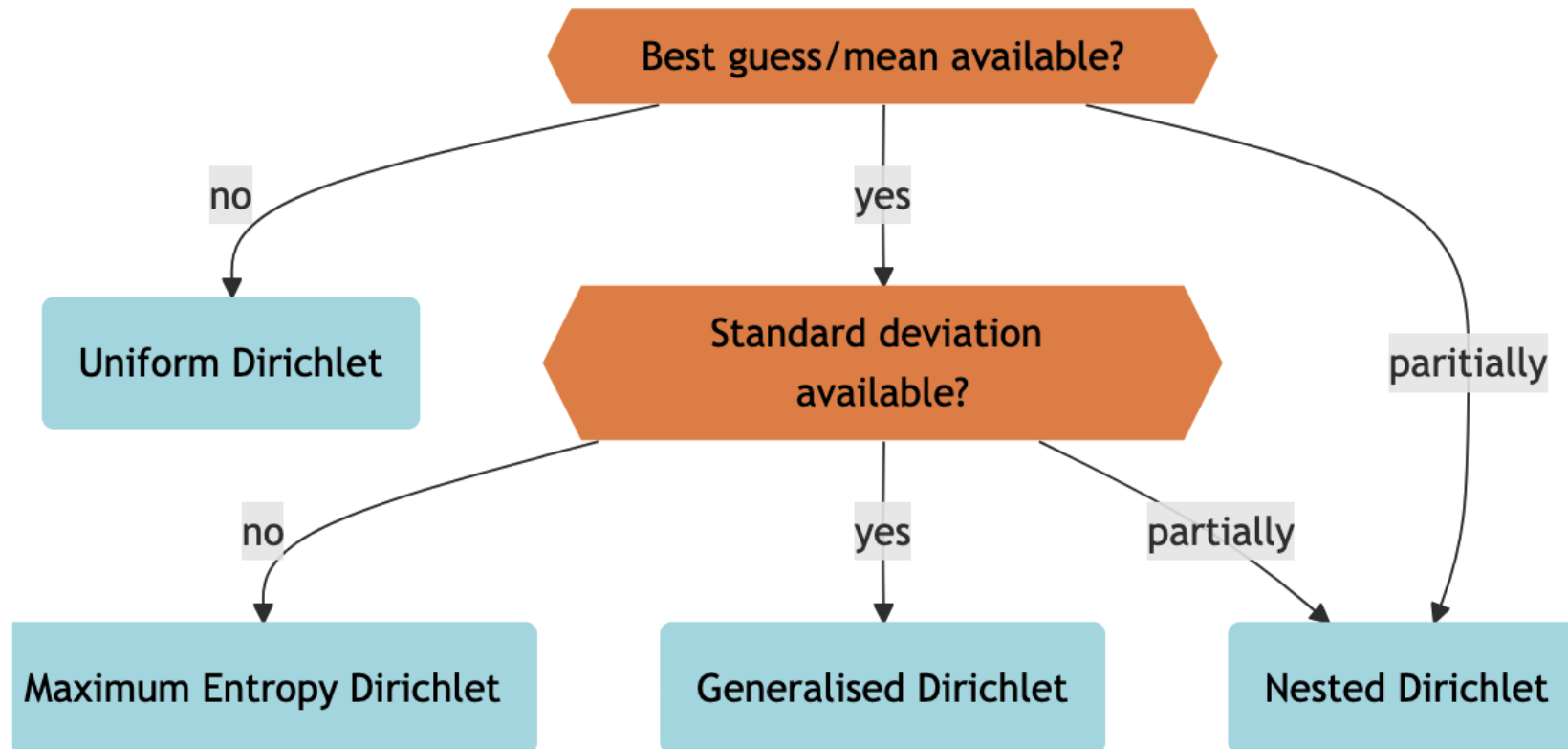
- Need distributions to sample y_0, x_1, x_2
- Choice of distribution has a big influence on results.
- Information on uncertainty can vary a lot! Do we have:
 - Best guess value y_0
 - Uncertainty for y_0 : σ_0
 - Best guesses for shares (proxy data): x_1, x_2
 - Uncertainty estimates for these shares: σ_1, σ_2

Maximum entropy principle

- Only use the **available** information, **no** implicit assumptions
 - Choose the distribution that **maximises** the **Entropy**
- (**Jaynes 1957**, doi: 10.1103/PhysRev.106.620) given the constraints.

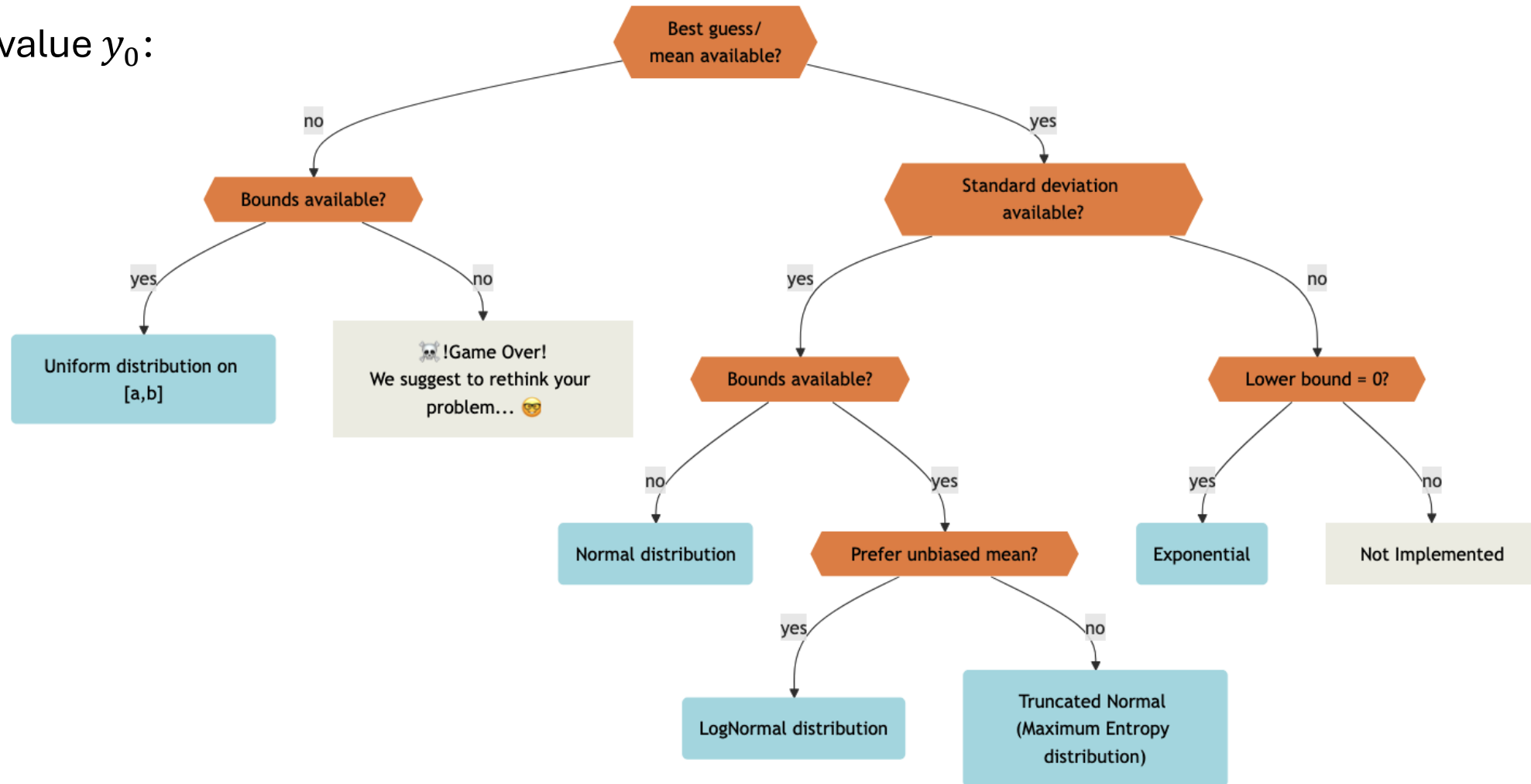
Maximum Entropy distributions for disaggregation

Shares $x_1, \dots, x_k$:



Maximum Entropy distributions for disaggregation

Total value y_0 :



Maximum Entropy distributions for disaggregation



Schulte, S., **Jakobs, A.**, Lupton, R. (2025). When Correlation Matters: A Practical Guide to Dealing with Uncertainty in the Case of Data Disaggregation. **Under Review** in the *Journal of Industrial Ecology*



Preprint here!



maxent_disaggregation

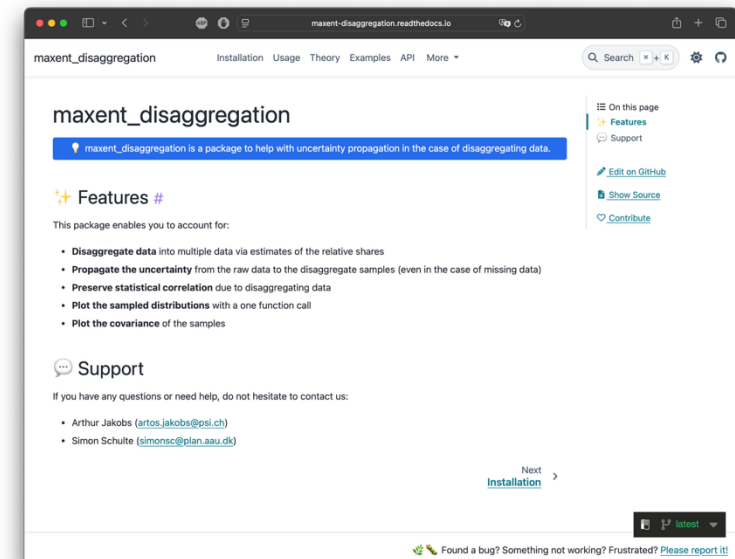
- Samples shares, totals, and disaggregate values
- Handles partial or missing information
- Maximises Entropy
- Easy to use (plug in what you know)
- Available in **Python** AND **R**
- Fully Documented



Available for Python



Available for R





Let's see how it works! 

→ Open *UseCaseExample.ipynb*