Uncertainty correlation and Monte Carlo sampling in LCA

LCAXV, Vancouver, Canada
07 October 2015

Guillaume Bourgault, Ph.D
Project Manager

ecoinvent
Outline

• What is Monte Carlo and why we use it?
• What correlation are we tackling today?
• How beta and Dirichlet distributions can help?
• Ignoring correlation overestimates uncertainty
What is Monte Carlo?

\[ y = x_1 + x_2 \]
What is Monte Carlo?

- $y$ VS $x_1$, Spearman corr = 0.8888
- $y$ VS $x_2$, Spearman corr = 0.4397
Why Monte Carlo?

- What is the uncertainty of the model output?
  - “All models are wrong, but some are useful”. How useful is your model?
  - Decision makers appreciate the information

- What parameter of the model drives the output uncertainty?
  - Learn something about the model
  - Recommendations on model use
  - Guide data collection
  - Restructure the model: more parameters necessary?
Why Monte Carlo?

- Simpler models: analytical tools
  - Cumbersome to impossible for large models
  - Taylor series expansion: ill-adapted to large uncertainties in LCIA

- Advantage of Monte Carlo
  - Variables are sampled on their entire domain
  - All variables change at once: captures interactions between parameters

- Drawback of Monte Carlo
  - Computationally expensive (time and memory)
  - Requires coding skills
Correlation in LCI

\[ \text{NG + C + H} = 1 \]

- **Electricity consuming activity**: 1 kWh
  - **Natural gas**: 0.13 kWh ± X
  - **Coal**: 0.68 kWh ± Y
  - **Hydro**: 0.19 kWh ± Z

Trust in Transparency!
Correlation in LCI

Spearman corr = -0.528

Spearman corr = -0.731

Spearman corr = -0.113
Correlation in LCIA

Fate factor

\[ FF_{ij} = \text{mass transported from source cell (i) to all other cells (j)}. \]

\[ 1 \geq FF_{ij} \quad \forall \ j \geq 0 \]

\[ 1 \geq \sum_{j=1}^{n} FF_{ij} \geq 0 \]
Beta and Dirichlet distributions

- Lognormal
  - 3 degrees of freedom: minimum (often zero), average, variance
  - Only has a lower bound
  - skewed to higher values
  - VERY skewed for high variance

- Beta
  - Upper and lower bound
  - 4 degrees of freedom: min, max, $\alpha$, $\beta$
  - Choose min, max, average, percentile $97.5 \rightarrow \alpha, \beta$
  - Skewed to the right, left, or symmetrical
Beta and Dirichlet distributions

\[ \alpha + \beta \downarrow, \sigma \uparrow \]
Beta and Dirichlet distributions

a = 0.5
b = 0.5
Beta and Dirichlet distributions

![Beta and Dirichlet distributions graphs]

- $\alpha = 2.0$
- $\beta = 4.0$

- $\alpha = 4.0$
- $\beta = 2.0$
Beta and Dirichlet distributions

- Dirichlet is a multi-variate beta
- Inputs of Dirichlet:
  - Vector 1 x N. Example: [0.13, 0.68, 0.19]
  - Scaling value. Example: 10 and 100
  - M iteration. Example: 1000
- Output of Dirichlet:
  - Parameters for N beta distributions
  - N samples of M iterations
Beta and Dirichlet distributions

- Each iteration sums to 1

```
Out[65]:
array([[ 0.22502527,  0.59061152,  0.18436321],
       [ 0.00830886,  0.7870208 ,  0.20467033],
       [ 0.1394482 ,  0.58871838,  0.27183342],
       ...
       [ 0.06549313,  0.85525442,  0.07925245],
       [ 0.10132079,  0.87050497,  0.02817424],
       [ 0.02473596,  0.45133565,  0.52392839]])
```
Beta and Dirichlet distributions

![Graphs showing Beta and Dirichlet distributions with different scaling values.](#)
Beta and Dirichlet distributions

- Shuffled sample

\[
\text{Out[72]}:
\begin{array}{ccc}
0.11361161 & 0.61693219 & 0.28884694 \\
0.110666555 & 0.87386014 & 0.40323989 \\
0.08209619 & 0.62743918 & 0.10454758 \\
\ldots \\
0.01507658 & 0.57664339 & 0.33551239 \\
0.07420796 & 0.61165905 & 0.23939245 \\
0.13773725 & 0.74982074 & 0.10817512
\end{array}
\]
Beta and Dirichlet distributions

Spearman corr = 6.46e-3

Spearman corr = 0.0411

Spearman corr = -0.0556
Beta and Dirichlet distributions

![Graphs showing Beta and Dirichlet distributions with different scalings.](1)

1. Trust in Transparency!
Beta and Dirichlet distributions

- Electricity production, ecoinvent v3.1, IPCC 2007 GWP 100 scores, Austria
  - Natural gas: 0.84362 kg CO$_2$-eq/kWh
  - Hard coal: 1.0022 kg CO$_2$-eq/kWh
  - Hydro, run-of-river: 0.0044188 kg CO$_2$-eq/kWh
Beta and Dirichlet distributions

![Histograms of Beta and Dirichlet distributions](image)

- For scaling = 10:
  - Mean: 0.799
  - Standard deviation: 0.112

- For scaling = 100:
  - Mean: 0.793
  - Standard deviation: 0.0382
Take home message

• Ignoring correlation leads to uncertainty overestimation
• Larger the uncertainty of the parameters → larger overestimation
• No rule of thumb to predict if the difference is noticeable
• Safer to test if the effect is significant in the setting at hand
• Ask your software provider the detail of the algorithm
• Ask a statistician friend!
Thank you for your attention!
Questions?

Guillaume Bourgault
Project Manager

bourgault@ecoinvent.org  www.ecoinvent.org
Why not divide by sum?
Why not divide by sum?

- With Dirichlet: mean = 0.791, st.dev = 0.116
- Shuffled sampling: mean = 0.791, st.dev = 0.169
- Shuffled sampling + division by sum: mean = 0.796, st.dev = 0.0965