

saeSim: Simulation Tools for Small Area Estimation

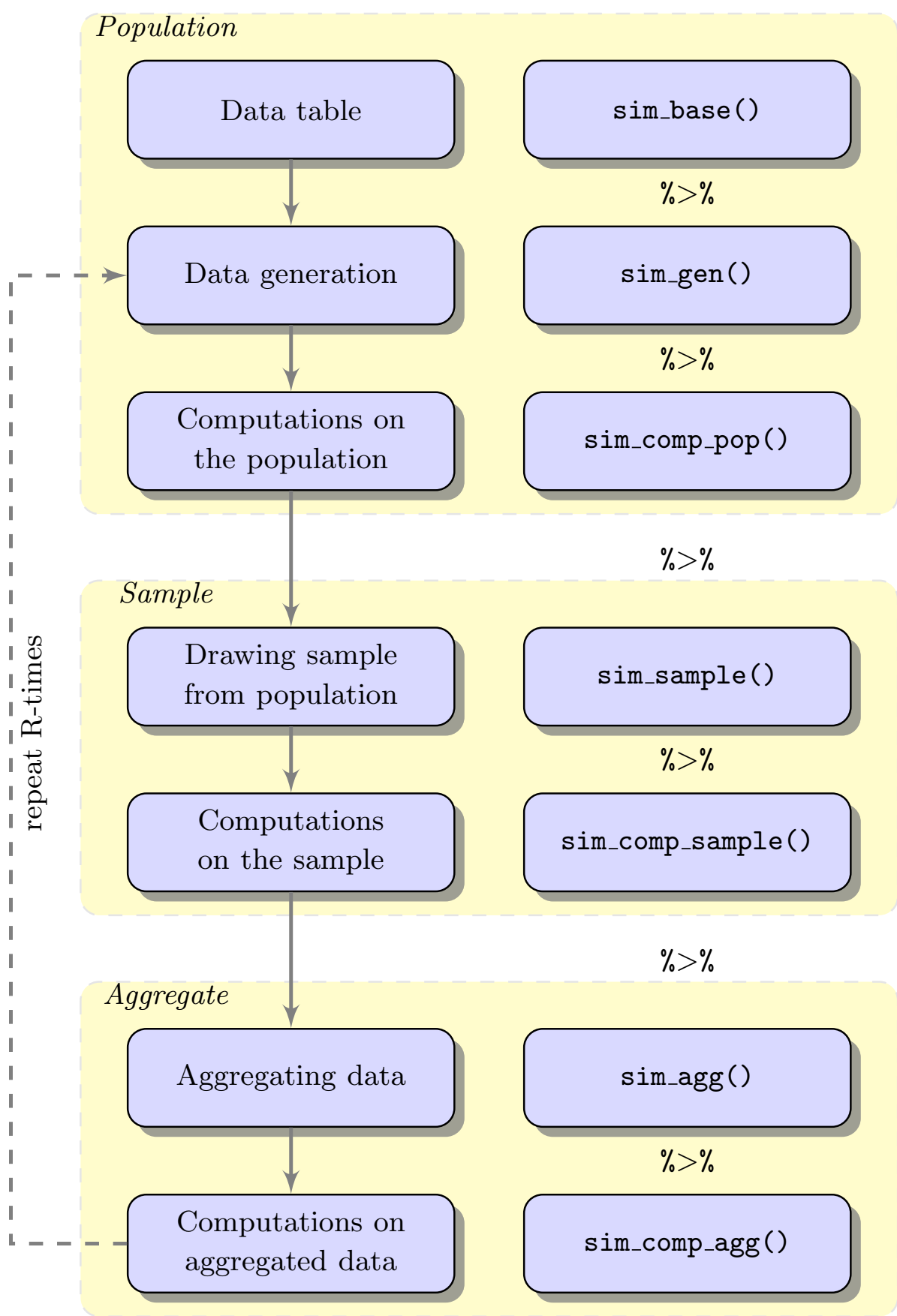
Sebastian Warnholz & T. Schmid

Sebastian.Warnholz@fu-berlin.de



Introduction

- Unified markup/tools for the composition of simulation studies in the context of small area estimation.
- Small area estimation summarises the development and application of statistical methods to report statistics for small groups. *Small* refers to the small number of sampled units.
- Model and design based simulation studies have been used to introduce new methods to the field.
- The package highlights a specific way to map a simulation study into R, namely in terms of a pipeline where a data frame is modified in each step. Also frequently used tools are ready to use which make the composition of such studies considerably more easy.
- With this package the composition of a simulation study is reduced to *chaining the steps together*.



Data Generation

Predefined setups and functions to generate random numbers as random effects including spatial correlation. Also a simple interface to univariate random number generators in R.

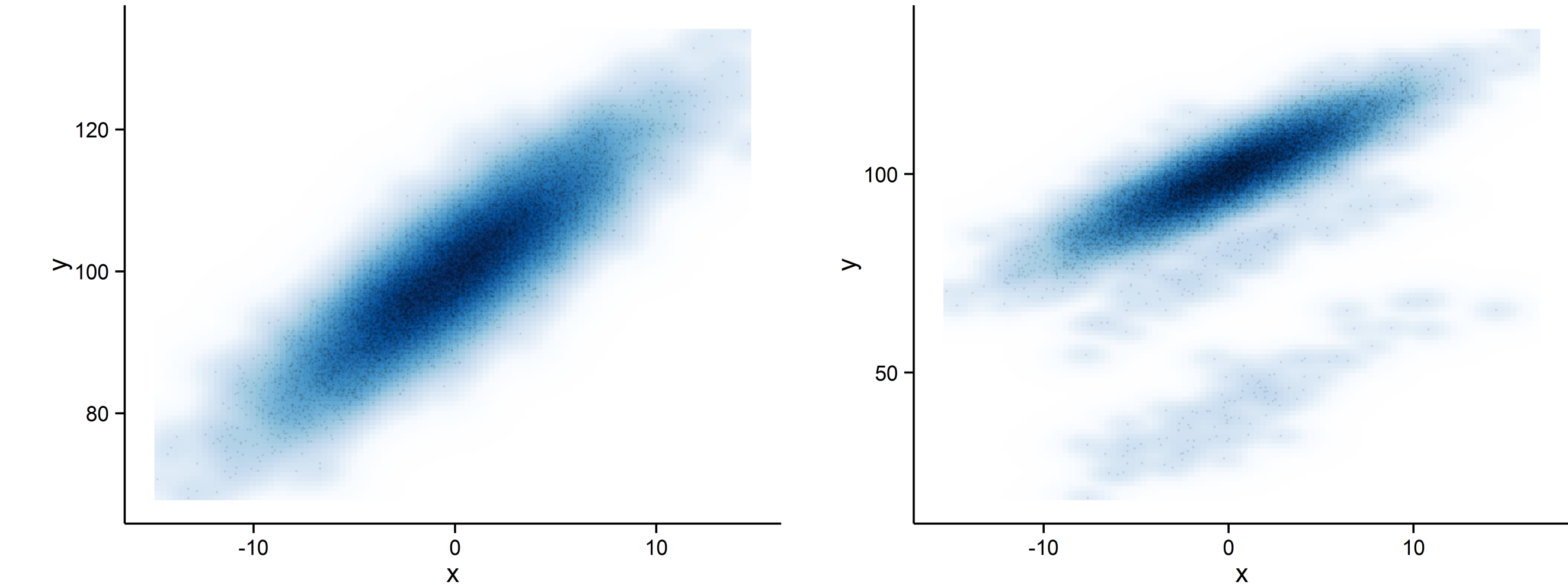
```
setup <- sim_base() %>%
  sim_gen_x() %>% # x ~ N(0, 16) iid
  sim_gen_e() %>% # e ~ N(0, 16) iid
  sim_gen_generic(rtf, df = 2, groupVars = "idD", name = "v") %>%
  sim_resp_eq(y = 100 + 2 * x + v + e)
setup

## data.frame [10,000 x 6]
##
##   idD idU      x      e      v      y
## 1    1    1  3.225797  0.7579929 1.477616 108.68720
## 2    1    2 -1.440586 -1.7312849 1.477616  96.86516
## 3    1    3  1.262536  3.0967488 1.477616 107.09944
## 4    1    4 -1.793524 -1.1042322 1.477616  96.78633
## 5    1    5  1.144710  0.2524723 1.477616 104.01951
## 6    1    6  0.874534 -3.8347551 1.477616  99.39193
## .. ... ..
```

Outliers

Easy introduction of outlier contaminated data.

```
autoplot(setup)
autoplot(setup %>% sim_gen_vc())
```



Heteroscedasticity

Different variances in each domain? For each observation? Maybe a random slope?

Sampling

Sampling schemes like simple random sampling in domains and cluster sampling.

Parallel computations

Simulation studies are embarrassingly parallel. For parallel computations we utilize parallelMap which makes it easy to switch between different parallel back ends in R (multicore, socket, mpi, BatchJobs).

Conclusions

- Remark 1
- Remark 2

References

Alfons, A., Templ, M. & Filzmoser, P. (2010), ‘An object-oriented framework for statistical simulation: The R package simFrame’, *Journal of Statistical Software* **37**(3), 1–36.
URL: <http://www.jstatsoft.org/v37/i03/>

Bache, S. M. & Wickham, H. (2014), *magrittr: A Forward-Pipe Operator for R*. R package version 1.5.
URL: <http://CRAN.R-project.org/package=magrittr>

Warnholz, S. & Schmid, T. (2015), *saeSim: Simulation Tools for Small Area Estimation*. R package version 0.7.0.
URL: <http://CRAN.R-project.org/package=saeSim>

Wickham, H. & Francois, R. (2015), *dplyr: A Grammar of Data Manipulation*. R package version 0.4.1.
URL: <http://CRAN.R-project.org/package=dplyr>

Timo Schmid

Department of Economics
Freie Universität Berlin
D-14195 Berlin, Germany
Timo.Schmid@fu-berlin.de



Sebastian Warnholz

Department of Economics
Freie Universität Berlin
D-14195 Berlin, Germany
Sebastian.Warnholz@fu-berlin.de