Lab 8

Antonio Jurlina

11/14/2020

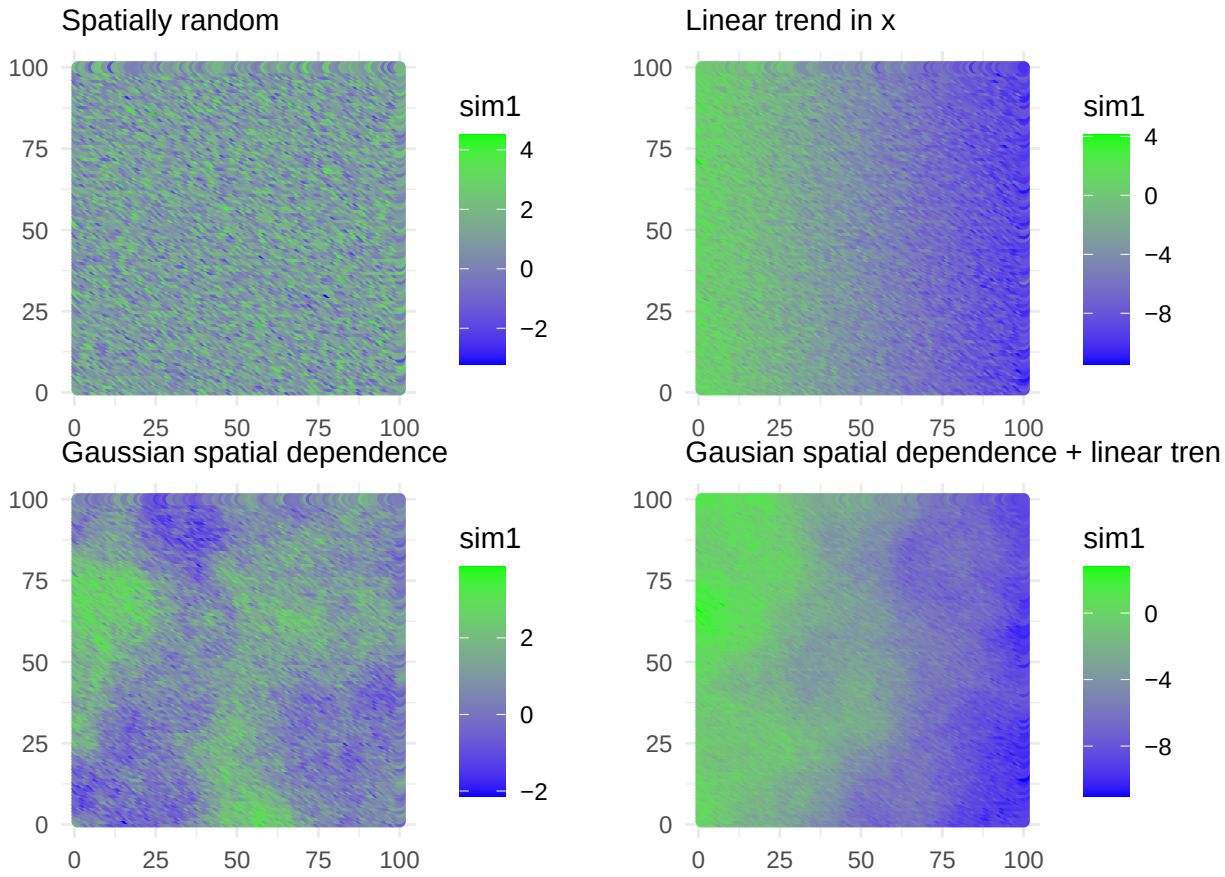

Figure 1: Simulated models

Table 1: Sampling results

output	sample size	type	mean1	sd1	mean2	sd2	mean3	sd3	mean4	sd4
sample mean	50	random	1.003	0.130	0.984	0.154	0.980	0.144	0.987	0.129
percent error	50	random	10.156	8.125	12.343	9.260	11.915	8.215	10.616	7.299
sample mean	150	random	1.009	0.082	0.995	0.081	1.007	0.078	0.993	0.080
percent error	150	random	6.658	4.812	6.344	4.985	6.152	4.807	6.511	4.584
sample mean	50	stratified	0.989	0.137	1.019	0.154	1.020	0.155	0.995	0.146
percent error	50	stratified	11.247	7.873	12.868	8.649	11.281	10.756	11.905	8.342
sample mean	150	stratified	1.005	0.082	1.004	0.074	0.998	0.077	1.002	0.086
percent error	150	stratified	6.903	4.490	5.753	4.584	6.157	4.636	6.502	5.622
sample mean	50	regular	1.008	0.155	0.982	0.158	1.017	0.131	1.022	0.130
percent error	50	regular	11.980	9.807	12.025	10.281	10.479	7.944	10.089	8.516
sample mean	150	regular	0.992	0.083	0.995	0.076	0.982	0.080	0.999	0.089
percent error	150	regular	6.723	4.899	6.003	4.662	6.828	4.512	7.180	5.162

Table 1 shows the results across all sampling approaches, across all 4 models. Means (and standard deviations), 1 through 4, refer to models 1 through 4. Specifically, Model 1 is spatially random, Model 2 has a linear trend in the x direction, Model 3 has a Gaussian spatial dependence and Model 4 combines the linear trend and the Gaussian dependence. Means (and standard deviations), 1 through 4, represent the values calculated over 100 repetitions of the listed sampling approaches.

In the case of simple random sampling, as the sample size increases from 50 to 150, Model 1 stops having the smallest sampling percent error (7.95%) and Model 4 overtakes (with 1.4% error). Overall, increased sample size reduces percent error across all 4 models.

In the case of stratified random sampling, as the sample size increases from 50 to 150, Model 3 sampling retains the smallest overall percent error (3.7% with sample size 50 and 6.33% with sample size 150). Overall, increased sample size reduces percent error across almost all models, except for Model 3 where it increases. This indicates that this sampling approach might not be ideal for Model 3.

In the case of systematic sampling, as the sample size increases from 50 to 150, Models 1,2 and 4 dramatically reduce their percent error while once again Model 3 sees an increase in percent error with a larger sample size. This makes Model 3 drop from having smallest percent error at sample size 50 to having the largest error at sample size 150.

Overall, notwithstanding the effects of sample size, the best approach to sampling from the Model 1 distribution is systematic and for Models 2, 3, and 4 it is simple random sampling. This is based on sampling percent errors. Each error is actually the average error across a 100 simulations (with the standard deviations of these errors reported as well).

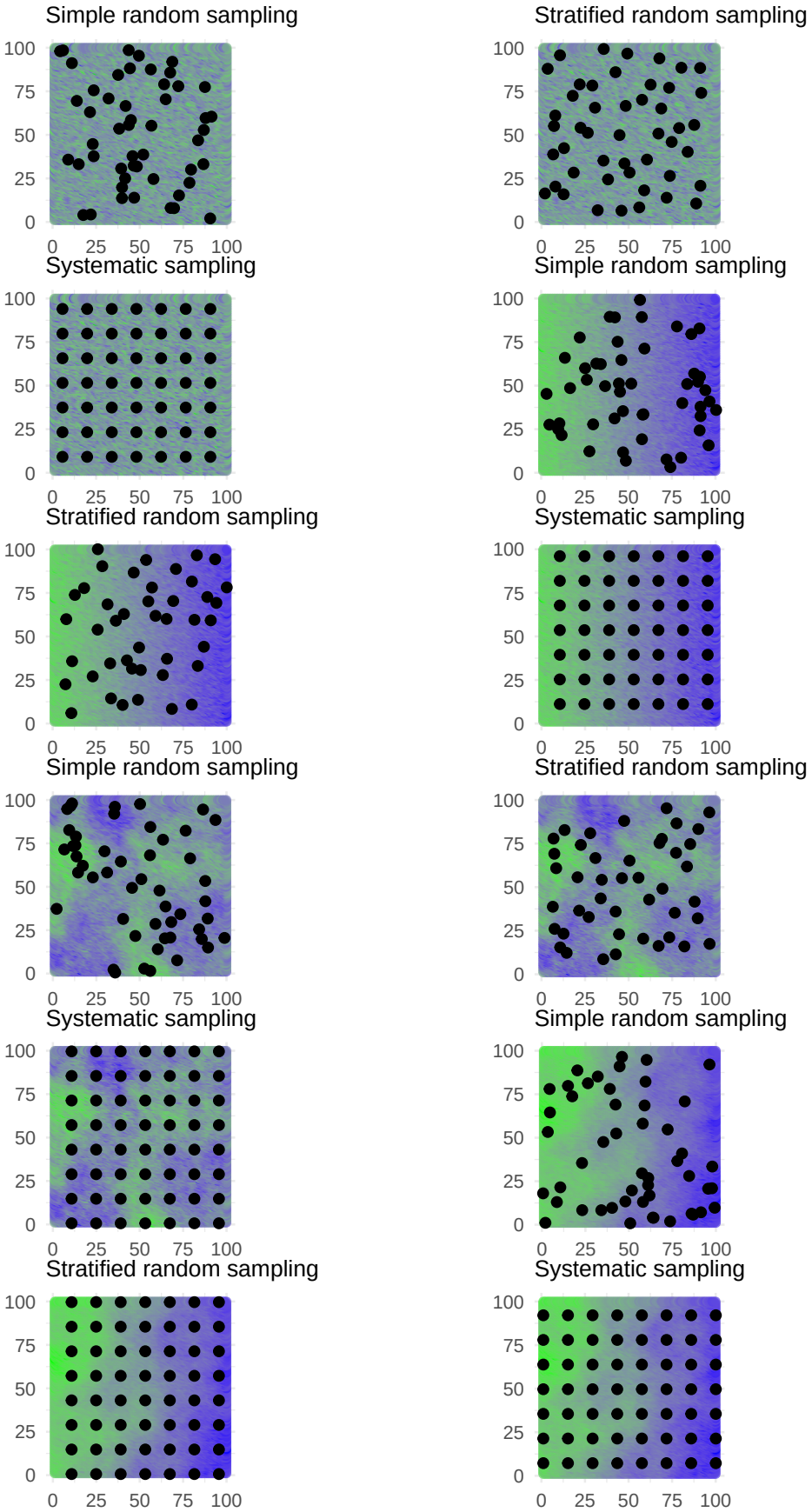

Figure 2: Sampling approaches

```
## R version 3.6.2 (2019-12-12)
## Platform: x86_64-apple-darwin15.6.0 (64-bit)
## Running under: macOS Sierra 10.12.6
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/3.6/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/3.6/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] knitr_1.29      gridExtra_2.3   sf_0.9-5        sp_1.4-2
##  [5] gstat_2.0-6     ggthemes_4.2.0  forcats_0.5.0   stringr_1.4.0
##  [9] dplyr_1.0.2     purrr_0.3.4     readr_1.3.1     tidyr_1.1.2
## [13] tibble_3.0.3    ggplot2_3.3.2   tidyverse_1.3.0
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_1.0.5         lubridate_1.7.9    lattice_0.20-41    FNN_1.1.3
##  [5] class_7.3-17       zoo_1.8-8          assertthat_0.2.1   digest_0.6.25
##  [9] R6_2.4.1           cellranger_1.1.0   backports_1.1.9    reprex_0.3.0
## [13] evaluate_0.14      e1071_1.7-3        highr_0.8          httr_1.4.2
## [17] pillar_1.4.6       rlang_0.4.7        readxl_1.3.1       rstudioapi_0.11
## [21] blob_1.2.1         rmarkdown_2.3      labeling_0.3       munsell_0.5.0
## [25] broom_0.7.0        compiler_3.6.2     modelr_0.1.8       xfun_0.16
## [29] pkgconfig_2.0.3    htmltools_0.5.0    tidyselect_1.1.0   intervals_0.15.2
## [33] fansi_0.4.1        spacetime_1.2-3    crayon_1.3.4       dbplyr_1.4.4
## [37] withr_2.2.0        grid_3.6.2         jsonlite_1.7.0     gtable_0.3.0
## [41] lifecycle_0.2.0    DBI_1.1.0          magrittr_1.5       units_0.6-7
## [45] scales_1.1.1       KernSmooth_2.23-17 cli_2.0.2          stringi_1.4.6
## [49] farver_2.0.3       fs_1.5.0           xml2_1.3.2         ellipsis_0.3.1
## [53] xts_0.12-0         generics_0.0.2     vctrs_0.3.4        tools_3.6.2
## [57] glue_1.4.2         hms_0.5.3          yaml_2.2.1         colorspace_1.4-1
## [61] classInt_0.4-3     rvest_0.3.6        haven_2.3.1
```