

tour_optim_documentation

Huize Zhang

2/3/2020

Tourr: overview

In general, a tour procedure involves two major stages: generating and interpolating stage. Given a current projection basis and its corresponding index value, the generator will try to find a better basis with higher index value through optimisation. The better basis found is called the target basis and the interpolator will interpolate the current and target basis geodesically. Currently there are three optimisers in the tourr package: `search_better()`, `search_better_random()` and `search_geodesic()`.

- `search_better()` randomly generates a nearby basis and compute its index value. If the new index value is larger than the current index value the new basis will be accepted and used as the current basis for the next iteration. This process will iterate until convergence.
- `search_better_random()` modifies on top of `search_better()` to include a random component when the difference between the new index value and the current index value is less than a prescribed epsilon. This stochastic component allows the optimisation to jump out of the local maximum and search for other possibilities
- `search_geodesic()` uses a two-step procedure to find the basis with better index value. Five candidate bases are first generated randomly and the basis with the largest index value is chosen as the promising direction for the target basis. This step is called the direction search. The algorithm then optimises the basis (angle) along the most promising direction over a 90 degree window to find the target basis. This step is called the line search. This two-step procedure completes one iteration in the algorithm and the algorithm ceases when the percentage change on the index value exceeds 0.001 or the number of iteration exceeds a pre-determined `max.tries`. Note that the current basis and index value should be updated in each iteration once a better target basis is found and this has yet to be implemented in the `search_geodesic` function.

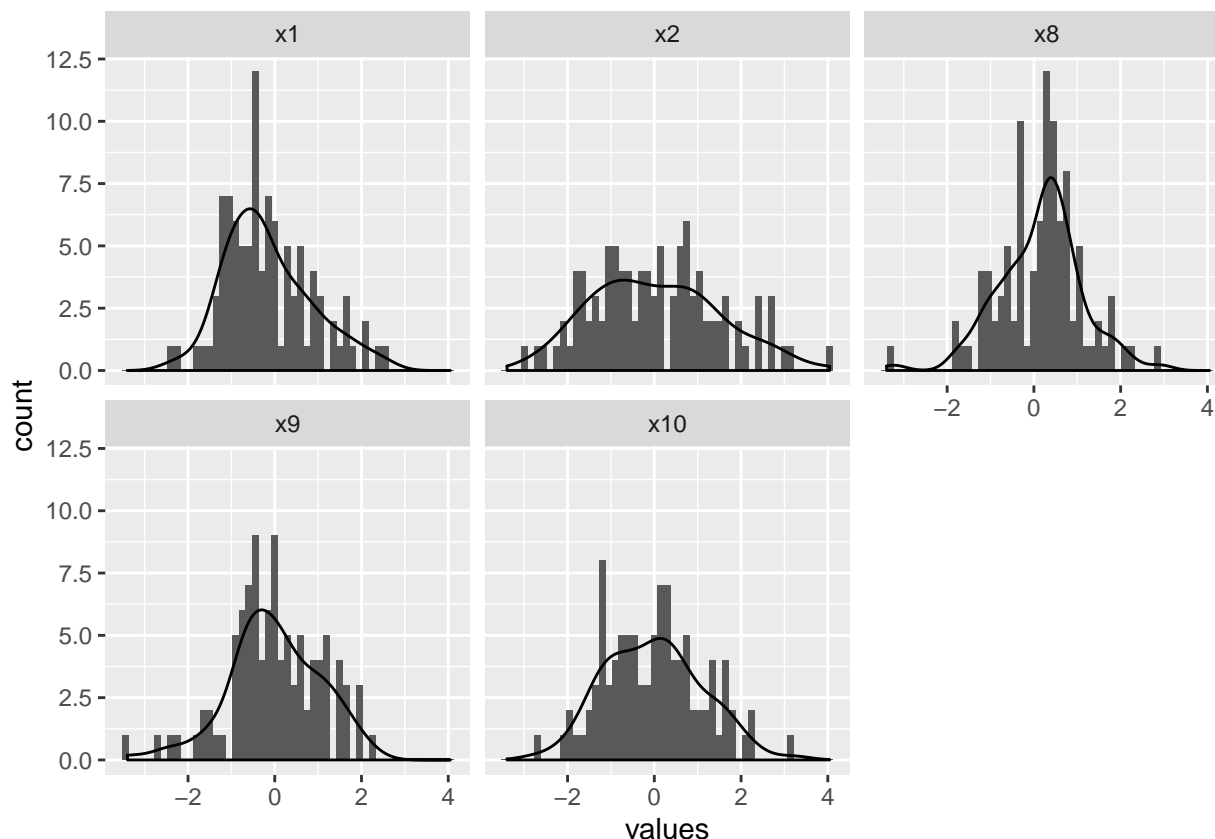
[add the graph]

Global Object

A global object that records all the bases, index value and auxiliary information would be useful to analyse the optimisation path of the tour and this has now been implemented in the tourr package. Along with the display the animated tour visualisation, the `animate_dist()` function would also result in a tibble object that contains all the information mentioned above.

Demo

Simulated data is generated using 4 random variables (`x1`, `x8`, `x9`, `x10`) and a bi-modal distribution generated from two normal distributions with mean at -1 and 1 (`x2`) and variance of 1. The histogram of the five variables are presented below.



Using `search_geodesic()`, the global object generated from the tour is presented below.

```
## # A tibble: 6 x 6
##   basis          index_val tries info          col  method
##   <list>          <dbl> <dbl> <chr>          <chr> <chr>
## 1 <dbl[,1] [5 x 1]>    0.747     1 <NA>          x2    geodesic
## 2 <dbl[,1] [5 x 1]>    0.670     1 direction_search x2    geodesic
## 3 <dbl[,1] [5 x 1]>    0.750     1 direction_search x2    geodesic
## 4 <dbl[,1] [5 x 1]>    0.715     1 direction_search x2    geodesic
## 5 <dbl[,1] [5 x 1]>    0.725     1 direction_search x2    geodesic
## 6 <dbl[,1] [5 x 1]>    0.715     1 direction_search x2    geodesic
```

Visualisation tools could help to visually inspect the bases from the tour. Using principle component analysis, we obtain the 2D projection of the 5D basis. The next three visualisations plot the first two principle components and allow us to draw some insights on the `search_geodesic` routine.

- Colored by the stage in which the basis is generated, direction search bases are further away from the interpolating bases than the line search bases.
- Color the interpolating bases on top of the direction search and line search bases allows us to see the location of the interpolating bases better
- Magnify the interpolating bases provides visual insight on the path of how the bases are updated.

