

mSigHdp vignette

ML and SGR

06/01/2022

This vignette is to introduce how to use mSigHdp (Liu et al., 2021)

```
devtools::install_github("steverozen/mSigHdp",ref="master")
```

Toy dataset

To illustrate the basic features of the mSigHdp package, we consider a toy data set of categorical count data with **10 samples** (rows) and **96 categories** (columns). The **10 samples** are PCAWG platinum tumors

```
library(mSigHdp)
toy_data <- ICAMS::ReadCatalog("toy_data.csv")
```

Run RunhdpXparallel to extract signatures from toy data set. We set the num.child.process and CPU.cores to 4 for a quick run. We recommend num.child.process and CPU.cores set to 20 for real data. There are three checkpoint arguments: checkpoint.chlist to checkpoint the chlist to “initial.chlist.Rdata” in the current working directory. checkpoint.l.chain to checkpoint the sample chain to current working directory, posterior.checkpoint to checkpoint the posterior sampling after every 10 posterior samples collected.

```
## calling extract_components_from_clusters 2022-01-06 16:45:11
## [1] "Performing divisive hierarchical clustering"
## extract_sigs_from_clusters time:
## user.self 0.16
## sys.self 0.012
## elapsed 0.1640000000000001
## user.child 0
## sys.child 0
## extracting components 2022-01-06 16:45:12
## extracting signatures exposures 2022-01-06 16:45:12
## Using existing out.dir vignettes
## Writing signatures
## Writing exposures
## Warning in dir.create(paste0(out.dir, "/Diagnostic_Plots"), recursive = T):
## 'vignettes/Diagnostic_Plots' already exists
## Writing HDP diagnostics
```

We can also extend burn-ins from checkpoints.

```
load("burnin.checkpoint.1000123.Rdata")
retval <- mSigHdp::ExtendBurnin(hdplist = burnin.output$hdplist,
                                seedNumber = 123,
                                burnin = 100)
```

```
## hdp_burnin time:
## user.self 0.632
## sys.self 0
## elapsed 0.638
## user.child 0
## sys.child 0
```

We can also continue Gibbs sampling from current checkpoints

```
load("posterior.checkpoint.1000123.Rdata")
retval <- hdp::hdp_posterior_sample(post.input = ans, post.n = 10, post.space = 5, seed=100123)
```

Extend Gibbs Sampling

Results

There are several outputs in the out.dir (in this vignettes, we saved outputs to data-raw/vignettes)

-extracted.signatures.csv and extracted.signatures.pdf

These two files are profiles and visualization of extracted signatures.

-extracted.signatures.post.samp.number.csv

A csv file with two columns. The first column contains the signature names correspond to extracted.signatures.csv, the second column contains the number of posterior samples extracted the corresponding signature.

-inferred.exposures.csv

Exposures inferred from HDP. Because there is no sparsity constraints of assignment in HDP. We recommend to use a signature assignment program to re-analyze the exposure, e.g. mSigAct, MutationalPatterns..

-inferred.exposure.count.pdf and inferred.exposure.proportion.pdf

Barplots of the counts and exposures of extracted signatures in each tumor

-low.confidence.signatures.csv

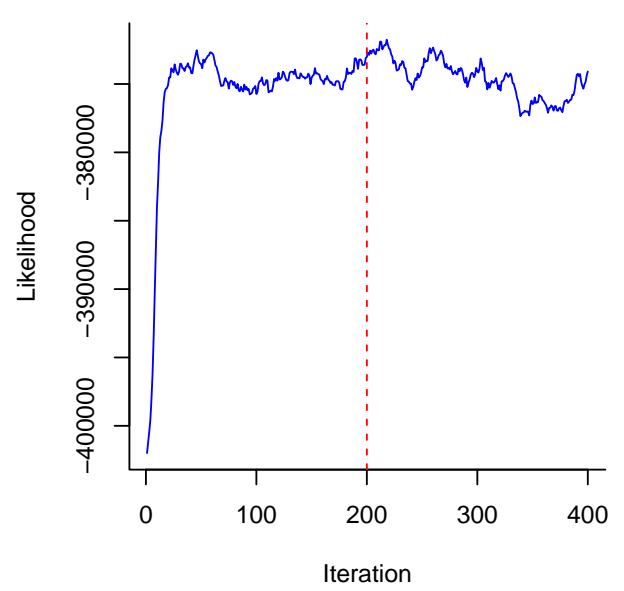
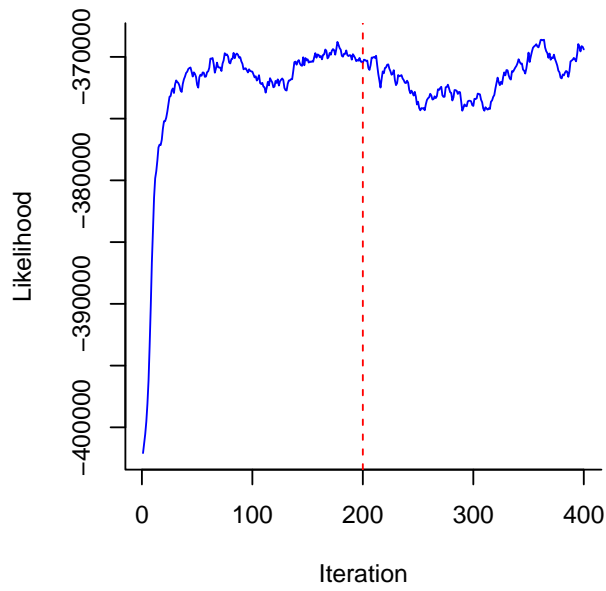
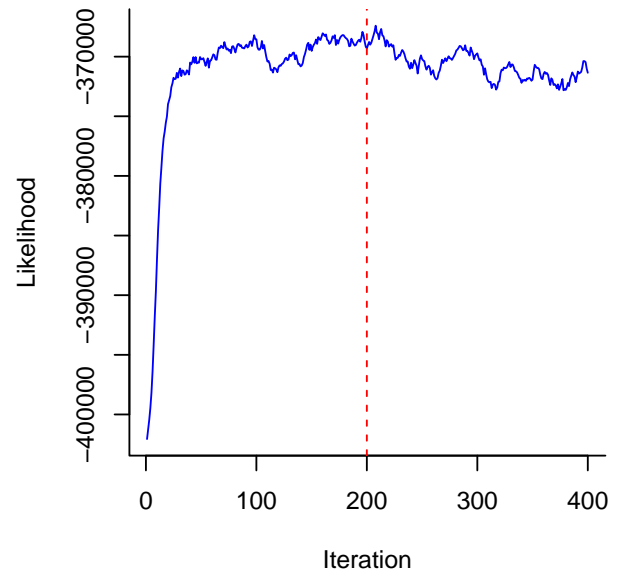
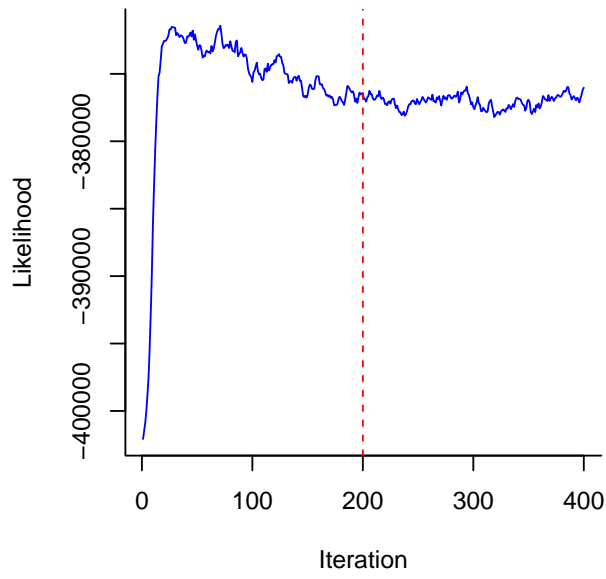
The profiles of extracted signatures with low confidence (i.e. the proportion of posterior samples with these signatures is less than pre-defined high.confidence.prop in RunHdpParallel).

-low.confidence.signatures.post.sample.number.csv

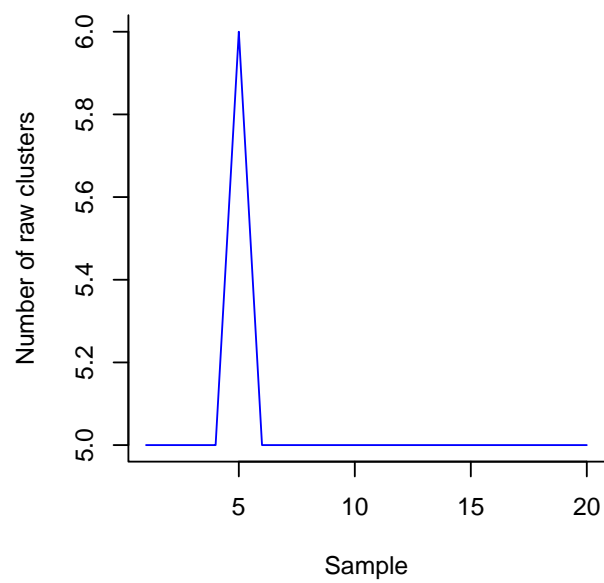
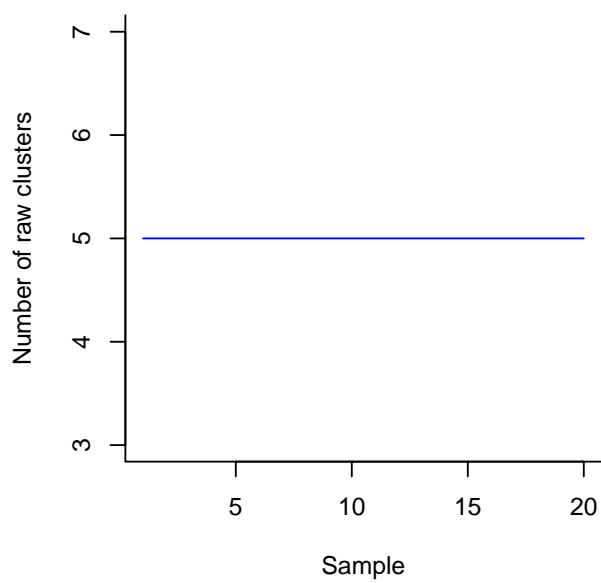
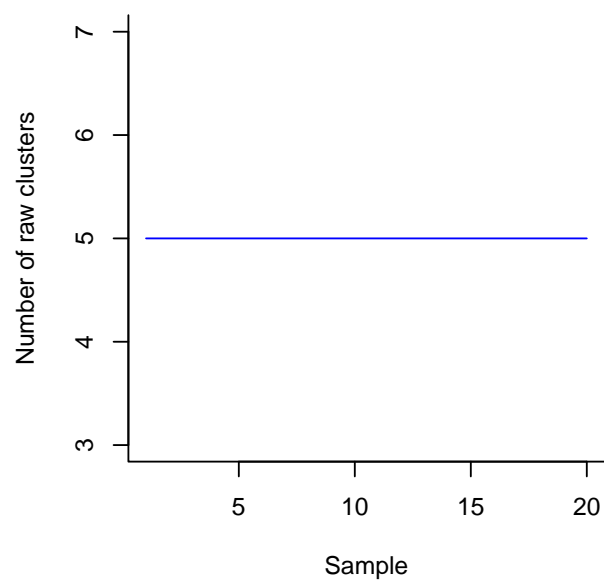
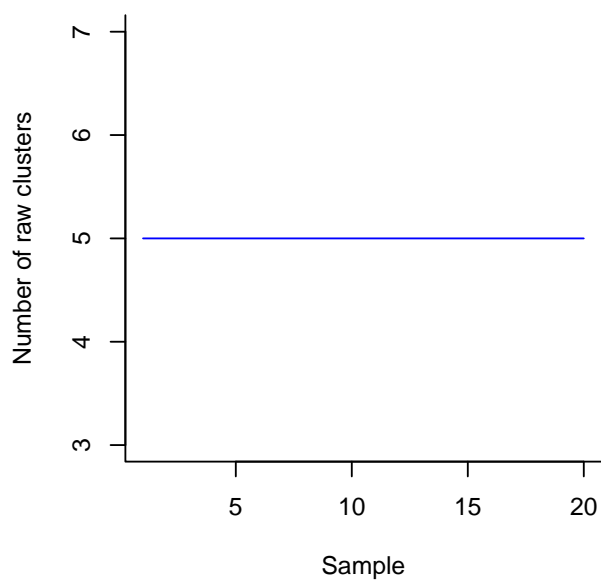
A csv file with two columns. The first column contains the signature names correspond to low.confidence.signatures.csv, the second column contains the number of posterior samples extracted the corresponding signature.

-Diagnostic_Plots folder contains five diagnostic plots:

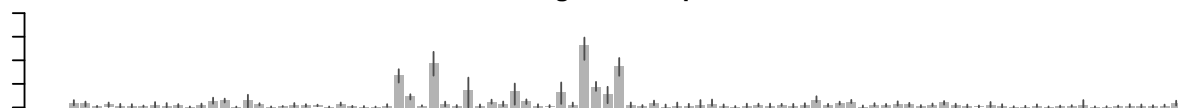
-The diagnostic plot of likelihood. To check if the likelihood is converged for each chain.



-The diagnostic plot of number of clusters. To check if the likelihood is converged for each chain.



Signature hdp.1



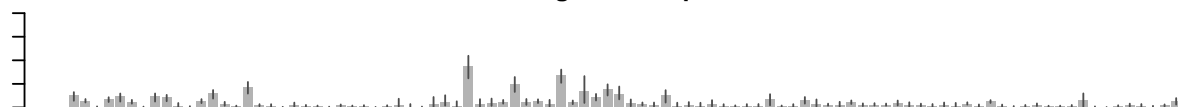
Signature hdp.2



Signature hdp.3



Signature hdp.4



Posterior samples with raw clusters of hdp.1

