

snpMatrix's snpStats Compatibility Mode

Hin-Tak Leung

January 19, 2015

1 Introduction

Starting with snpMatrix 1.19.0.20 (Jan 2015), a “snpStats Compatibility Mode” is introduced. When it is on, some routines within snpMatrix using compiled code are swapped to their snpStats-compatible versions. The number of affected routines is 22 out of 59 at the time of introduction (Jan 2015) and roughly a 3rd, and hopefully will go lower with time. The full list is in the appendix and will be continually updated.

```
> get.ss.mode()
```

```
[1] "normal"
```

```
> set.ss.mode("compat")
```

```
[1] "compat"
```

```
> get.ss.mode()
```

```
[1] "compat"
```

```
> set.ss.mode("normal")
```

```
[1] "normal"
```

```
> get.ss.mode()
```

```
[1] "normal"
```

Compatibility mode is mainly controlled by `set.ss.mode()` and `get.ss.mode()`. `set.ss.normal()` and `set.ss.compat()` are aliases to `set.ss.mode("normal")` and `set.ss.mode("compat")` respectively. There is a 3rd option, `set.ss.strict()` which is planned to be equivalent to `set.ss.mode("normal")` plus *some* checks and abort. However, such checks are very time-consuming and not currently automatically done. We'll see if there is any genuine need for it. A routine `is.uncertain()` is provided for testing data with uncertain calls manually if suspected.

There is an accompanying vignette, “Compatibility Tests”, which tests for details of agreement.

2 Namespaces

In compatibility mode, classes and namespaces retain their `snpMatrix` names. There are also differences as noted below.

2.1 Classes

| | <u>snpMatrix Classes</u> |
|----------------------------------|-------------------------------------|
| <code>XSnpmatrix</code> | <code>X.snp.matrix</code> |
| <code>SnpMatrix</code> | <code>snp.matrix</code> |
| <code>ImputationRules</code> | <code>snp.reg.imputation</code> |
| <code>GlmTestsScore</code> | <code>snp.tests.glm.score</code> |
| <code>GlmTests</code> | <code>snp.tests.glm</code> |
| <code>GlmEstimates</code> | <code>snp.estimates.glm</code> |
| <code>SingleSnpTests</code> | <code>snp.tests.single</code> |
| <code>SingleSnpTestsScore</code> | <code>snp.tests.single.score</code> |

2.1.1 The `X.snp.matrix` class

The `Female` slot was renamed `diploid` on Mar 15 2011 between `snpStats` 1.1.10 and 1.1.11. `snpMatrix` is not making major incompatible changes that way.

This, in particular, means that `read.snps.long()` and `read.plink()` take a `female` argument, rather than a `diploid` argument; and the internal routine for guessing the gender from heterozygosity is `.guessSex()`, rather than `.guessPloidy()`.

2.1.2 The `snp.tests.glm` class

The `snp.tests.glm` class contains a `test.names` slot, but `snp.names` and `var.names` slots in compatibility mode. In the future `snp.names` is likely to be changed (back) to `test.names`.

2.1.3 The `snp.reg.imputation` class

The slot `coefficients` is missing in compatibility mode.

2.2 data files

Bundled data sets with similar names are not expected to have identical contents.

For the time being, at least the `families` data set contain one difference:

| <u>snpMatrix name</u> | |
|-----------------------|----------------------|
| <code>pedfile</code> | <code>pedData</code> |

2.3 Defaults

2.3.1 `col.summary`

`col.summary()` defaults to `uncertain = FALSE`, i.e. not showing uncertain data, even in compatibility mode.

2.3.2 GLM

GLM-related routines have different defaults in `glm.test.control()`. `R2Max` is 0.999 and not adopting the current `snpStats` value of 0.99 (which was lowered on Mar 7 2012, between 1.5.4 and 1.5.5).

This affects `snp.lhs.estimate()`, `snp.lhs.tests()`, `snp.rhs.estimate()` and `snp.rhs.tests()`.

3 Enhancements/Regressions

GLM tests in `snpMatrix` can handle single SNPs; this capability was lost at some point in the past in `snpStats`. The capability is retained in compatibility mode.

Likewise, `mvtests()` in `snpMatrix` can handle smaller (single unit) input, compared to `snpStats`.

A Routines swapped in compatibility mode

The R routines and the corresponding C APIs affected by the Compatibility Mode setting directly are listed below. There are one fewer R routines than C APIs as `col.summary` calls either `X_snp_summary` or `snp_summary`. Note that many are indirectly affected: the constructor `new('X.snp.matrix',)` calls `.forceHom`, and `pool` calls `pool2`, for example. Some generics are only affected for inputs with specific classes, `pool2` (for `snp.tests.glm.score`) and `summary` (for `snp.reg.imputation`).

```
.forceHom
snp.lhs.tests
snp.rhs.tests
pool2.snp.tests.glm.score
snp.lhs.estimate
snp.rhs.estimate
coerce(snp.estimate.glm, 'snp.tests.glm')
snp.imputation
impute.snps
summary.snp.reg.imputation
ld
single.snp.tests
snp.pre.multiply
snp.post.multiply
col.summary
row.summary
xxt
snp.cor
ibsCount
ibsDist
tdt.snp
```

```
force_hom
snp_lhs_score
snp_rhs_score
pool2_glm
snp_lhs_estimate
snp_rhs_estimate
wald
snp_impute
impute_snps
r2_impute
ld
score_single
snp_pre
snp_post
X_snp_summary
snp_summary
row_summary
xxt
corsm
ibs_count
ibs_dist
score_tdt
```