Dwueng Chwuan Jhwueng <djhwueng@umail.iu.edu>

[bomeara/BMhyb] Variance between hybrid descendants (#14)

1 message

Paul Bastide <notifications@github.com>   Fri, Oct 6, 2017 at 4:27 PM
Reply-To: bomeara/BMhyb
<reply+013c45959d634bf0374872ca46d11c6443beaf905948d7d692cf0000000115ef007b92a169ce0fb2b6e8@reply.github.com>
To: bomeara/BMhyb <BMhyb@noreply.github.com>
Cc: Tony Jhwueng <djhwueng@umail.iu.edu>, Mention <mention@noreply.github.com>

Variance between hybrid descendants

Hi again, @bomeara and @djhwueng

This might be related to #13.

I tried a network a little more sophisticated, with an hybrid having several descendants, here R and Y.

```
## Underlying tree
t1 <- 0.3; t2 <- 0.4; t3 <- 0.3;
phy <- read.tree(text = paste0("((R:", t3, ",Y:", t3, "):", t1 + t2, ",X:", t1 + t2 + t3, ");"))
## Network
don_recp <- expand.grid(c("X"), c("Y", "R"))
network <- list(phy = phy,
                flow = data.frame(donor = don_recp[,1],
                                   recipient = don_recp[,2],
                                   gamma = rep(gamma, 2),
                                   time.from.root.donor = rep(t1, 2),
                                   time.from.root.recipient = rep(t1, 2)))
network$flow$donor <- as.character(network$flow$donor)
network$flow$recipient <- as.character(network$flow$recipient)
## Plot
PlotNetwork(network$phy, network$flow)
axis(1, at = c(0, t1, t1+t2, t1+t2+t3), labels = c("0", "t1", "t1+t2", "t1+t2+t3"))
```

I tried to respect your format for the flow matrix, using your description here. Is this network correctly defined ?

I then tried to compute the associated variance matrix.

```
> sigma2 = 1
> x <- c(sigma.sq = sigma2, mu = 0, SE = 0)
> actual.params <- c("sigma.sq", "mu", "bt", "vh", "SE")

> GetVModified(x, network$phy, network$flow, actual.params)
     R    Y    X
R 0.85 0.70 0.15
Y 0.70 0.85 0.15
X 0.15 0.15 1.00
```

In this matrix, if the network is correctly defined and my computations right, I think that Cov[R,Y] is not correct. I think it should be:

```
Cov[Y,R] = sigma^2 * [(gamma^2 + (1-gamma)^2)*t1 + t2] = 0.55 \neq 0.70
```

What do you think about it ? Did I make a mistake somewhere ?

I did not dive into your code very deep, but from what I understood of your algorithm, you are modifying all the couple (recipient, donors) one by one (browsing through your flow matrix), but never the couples (recipient1, recipient2), when there are several descendants from a single event, as it is the case here.
In the example above, the function indeed gives Cov[R,Y]=0.70, which seems like the non-actualized variance one would get from the underlying tree.

But it's possible I misunderstood something, please correct me if I'm wrong !

Thanks again !

Session infos:

```
> sessionInfo()
R version 3.4.2 (2017-09-28)
Platform: x86_64-pc-linux-gnu (64-bit)
```

```
Running under: Ubuntu 16.04.3 LTS

Matrix products: default
BLAS: /usr/lib/openblas-base/libblas.so.3
LAPACK: /usr/lib/libopenblasp-r0.2.18.so

locale:
[1] LC_CTYPE=fr_FR.UTF-8        LC_NUMERIC=C              LC_TIME=fr_FR.UTF-8
LC_COLLATE=fr_FR.UTF-8
[5] LC_MONETARY=fr_FR.UTF-8     LC_MESSAGES=fr_FR.UTF-8   LC_PAPER=fr_FR.UTF-8        LC_NAME=C
[9] LC_ADDRESS=C                LC_TELEPHONE=C            LC_MEASUREMENT=fr_FR.UTF-8
LC_IDENTIFICATION=C

attached base packages:
[1] stats      graphics  grDevices utils     datasets  methods   base

other attached packages:
[1] BMhyb_1.5.1 ape_4.1

loaded via a namespace (and not attached):
 [1] Rcpp_0.12.12          subplex_1.4-1        msm_1.6.4            mvtnorm_1.0-6
 [5] lattice_0.20-35       tidyr_0.7.1          corpcor_1.6.9       prettyunits_1.0.2
 [9] assertthat_0.2.0      digest_0.6.12        foreach_1.4.3       R6_2.2.2
[13] plyr_1.8.4            phytools_0.6-20      coda_0.19-1         httr_1.3.1
[17] ggplot2_2.2.1         progress_1.1.2       rlang_0.1.2.9000    uuid_0.1-2
[21] lazyeval_0.2.0        curl_2.8.1           data.table_1.10.4   taxize_0.9.0
[25] phangorn_2.2.0        Matrix_1.2-11        RNeXML_2.0.7        combinat_0.0-8
[29] splines_3.4.2         stringr_1.2.0        igraph_1.1.2        munsell_0.4.3
[33] compiler_3.4.2        numDeriv_2016.8-1    geiger_2.0.6        pkgconfig_2.0.1
[37] mnormt_1.5-5          tibble_1.3.4         gridExtra_2.2.1     TreeSim_2.3
[41] expm_0.999-2          quadprog_1.5-5       codetools_0.2-15    XML_3.98-1.9
[45] reshape_0.8.7         viridisLite_0.2.0    dplyr_0.7.2         MASS_7.3-47
[49] crul_0.3.8            grid_3.4.2           nlme_3.1-131        jsonlite_1.5
[53] gtable_0.2.0          magrittr_1.5         scales_0.5.0        stringi_1.1.5
[57] reshape2_1.4.2        viridis_0.4.0        bindrcpp_0.2        scatterplot3d_0.3-40
[61] phylobase_0.8.4       xml2_1.1.1           fastmatch_1.1-0     deSolve_1.20
[65] iterators_1.0.8       tools_3.4.2          rncl_0.8.2          ade4_1.7-8
[69] bold_0.5.0            glue_1.1.1           purrr_0.2.3         maps_3.2.0
[73] plotrix_3.6-6         parallel_3.4.2       survival_2.41-3     colorspace_1.3-2
[77] bindr_0.1             animation_2.5        clusterGeneration_1.3.4
```

—

You are receiving this because you were mentioned.
Reply to this email directly, view it on GitHub, or mute the thread.