Sat PM Notes from R Workshop

Luke Harmon

June 20, 2010

1 Phylogenetic Independent Contrasts (PICs)

We can easily calculate independent contrasts in r. First, we read in a tree and some data. I will pull out the data columns and give them names so that r can match the tree tip labels and the data tip names.

```
> require(ape)
> anoleTree <- read.nexus("anolis.ssd.23tree.nex")
> anoleExt <- read.csv("anolis_extended.csv")
> anoleSVL <- anoleExt[, 2]
> names(anoleSVL) <- anoleExt[, 1]
> anoleSSD <- anoleExt[, 3]
> names(anoleSSD) <- anoleExt[, 1]
> anoleEM <- anoleExt[, 4]
> names(anoleEM) <- anoleExt[, 1]
```

Now we can calculate standardized independent contrasts for SVL and SSD:

```
> svlpic <- pic(anoleSVL, anoleTree)
> ssdpic <- pic(anoleSSD, anoleTree)
> plot(svlpic, ssdpic)
```
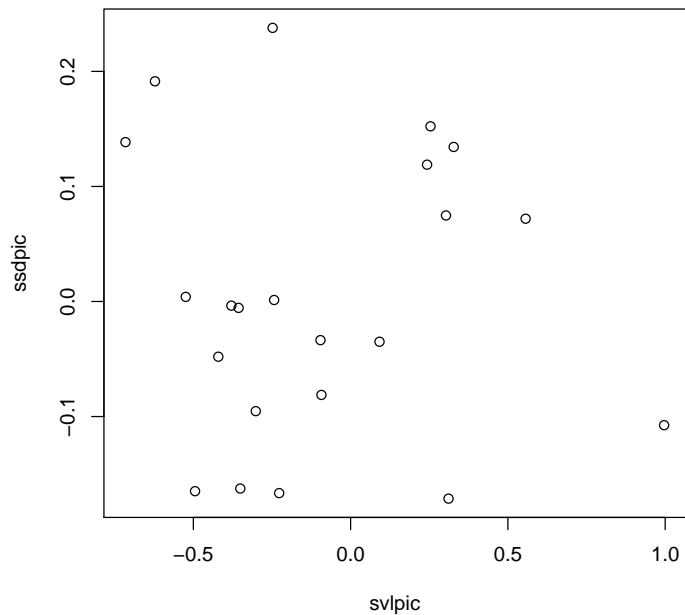
Finally, we can test for a relationship between the PICs for the two traits, forced through the origin:

```
> picModel <- lm(ssdpic ~ svlpic - 1)
> summary(picModel)

Call:
lm(formula = ssdpic ~ svlpic - 1)

Residuals:
     Min       1Q   Median       3Q      Max
-0.16903 -0.09361 -0.00729  0.10981  0.23602

Coefficients:
       Estimate Std. Error t value Pr(>|t|)
svlpic -0.00739    0.06181   -0.12    0.906

Residual standard error: 0.1236 on 21 degrees of freedom
Multiple R-squared: 0.0006803,      Adjusted R-squared: -0.04691
F-statistic: 0.0143 on 1 and 21 DF,  p-value: 0.906
```
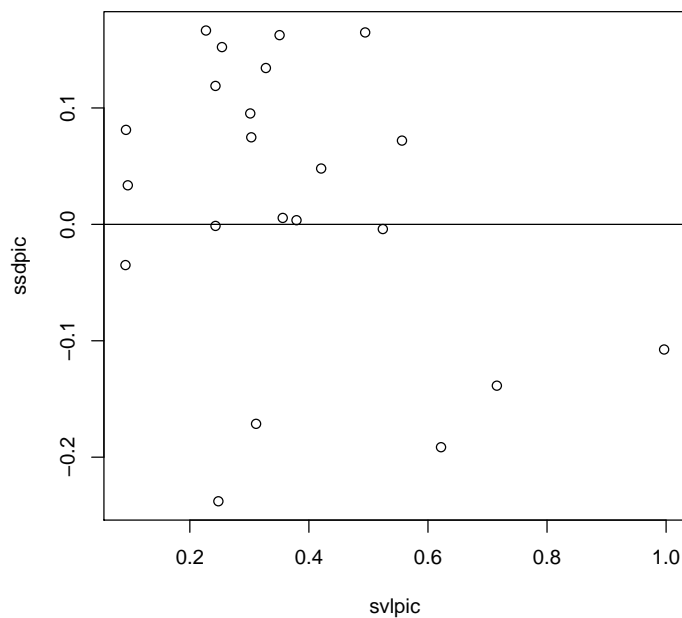
Here is some code to "positivize" the SVL contrasts, and make a plot of these. This is a pretty standard thing to do for PICs and won't affect the statistical analysis.

```
> negSVL <- svlpic < 0
> svlpic[negSVL] <- -svlpic[negSVL]
> ssdpic[negSVL] <- -ssdpic[negSVL]
> plot(svlpic, ssdpic)
> lines(x = c(0, 0), y = c(-5, 5))
> lines(x = c(-5, 5), y = c(0, 0))
```


We can also use the contrasts to estimate σ^2, the Brownian motion rate parameter. This is very close to the ML estimate - in fact, they differ only by a factor of $n/(n-1)$.

```
> require(geiger)
> sigsq_pic <- mean(svlpic^2)
> sigsq_ml <- fitContinuous(anoleTree, anoleSVL)

Fitting  BM model:

> sigsq_ml$Trait1$beta * 23/22

[1] 0.1818564
```

A similar approach can be used to estimate the entire matrix R, the evolutionary variance-covariance matrix for multiple traits at once. This uses the sum of squares and cross products matrix of the standardized contrasts.

3

```
> ic.sigma(anoleTree, cbind(anoleSVL, anoleSSD))
```

```
              anoleSVL       anoleSSD
anoleSVL  0.181856365 -0.001344008
anoleSSD -0.001344008  0.014600556
```

Now I will run some examples of Phylogenetic Generalized Least Squares (PGLS). First, we can recreate the analysis above that we did using contrasts. Note that the p-value is the same.

```
> require(nlme)
> dd <- data.frame(anoleSSD, anoleSVL, anoleEM)
> pglsModel1 <- gls(anoleSSD ~ anoleSVL, correlation = corBrownian(1,
+     anoleTree), data = dd)
> summary(pglsModel1)
```

```
Generalized least squares fit by REML
  Model: anoleSSD ~ anoleSVL
  Data: dd
        AIC       BIC    logLik
  -24.55529 -21.42172 15.27764


Correlation Structure: corBrownian
 Formula: ~1
 Parameter estimate(s):
numeric(0)

Coefficients:
                  Value  Std.Error    t-value  p-value
(Intercept)  0.17318366 0.25760081  0.6722947   0.5087
anoleSVL    -0.00739179 0.06181269 -0.1195837   0.9059


 Correlation:
         (Intr)
anoleSVL -0.978


Standardized residuals:
       Min         Q1        Med         Q3        Max
-1.2250549 -0.3573358  0.1215432  0.9040009  2.0663295


Residual standard error: 0.1236467
Degrees of freedom: 23 total; 21 residual
```

But we can do more with PGLS - like fitting more complicated models with multiple predictors:

```
> pglsModel2 <- gls(anoleSSD ~ anoleEM, correlation = corBrownian(1,
+     anoleTree), data = dd)
> summary(pglsModel2)
```

```
Generalized least squares fit by REML
  Model: anoleSSD ~ anoleEM
  Data: dd
        AIC       BIC    logLik
  -21.46686 -15.63436 17.73343

Correlation Structure: corBrownian
 Formula: ~1
 Parameter estimate(s):
numeric(0)

Coefficients:
                        Value  Std.Error   t-value p-value
(Intercept)        0.11474522 0.04689739  2.446730  0.0256
anoleEMgrass-bush  0.02109545 0.04812551  0.438342  0.6667
anoleEMtrunk       0.00805994 0.05721053  0.140882  0.8896
anoleEMtrunk-crown 0.10343085 0.04453172  2.322633  0.0329
anoleEMtrunk-ground 0.18178214 0.04678216  3.885715  0.0012
anoleEMtwig        -0.04990531 0.04851957 -1.028561  0.3181

 Correlation:
                    (Intr) anlEMg- anlEMtr anlEMtrnk-c anlEMtrnk-g
anoleEMgrass-bush   -0.541
anoleEMtrunk        -0.466  0.485
anoleEMtrunk-crown  -0.477  0.508   0.418
anoleEMtrunk-ground -0.533  0.646   0.468   0.520
anoleEMtwig         -0.592  0.551   0.477   0.474      0.539

Standardized residuals:
       Min         Q1        Med         Q3        Max
-0.8913676 -0.3534415  0.1937266  0.9775636  2.2893745

Residual standard error: 0.07848604
Degrees of freedom: 23 total; 17 residual

> anova(pglsModel2)

Denom. DF: 17
            numDF   F-value p-value
(Intercept)     1 17.275916    7e-04
anoleEM         5  7.030979    1e-03

> pglsModel3 <- gls(anoleSSD ~ anoleSVL + anoleEM, correlation = corBrownian(1,
+     anoleTree), data = dd)
> summary(pglsModel3)

Generalized least squares fit by REML
  Model: anoleSSD ~ anoleSVL + anoleEM
```

```
  Data: dd
        AIC       BIC    logLik
  -16.30971 -10.12900 16.15485

Correlation Structure: corBrownian
 Formula: ~1
 Parameter estimate(s):
numeric(0)

Coefficients:
                         Value Std.Error    t-value p-value
(Intercept)          0.22879729 0.3803100  0.6016073  0.5559
anoleSVL            -0.02406834 0.0796093 -0.3023306  0.7663
anoleEMgrass-bush   -0.00222814 0.0916425 -0.0243134  0.9809
anoleEMtrunk        -0.00852798 0.0804253 -0.1060361  0.9169
anoleEMtrunk-crown   0.08893802 0.0662798  1.3418580  0.1984
anoleEMtrunk-ground  0.16462087 0.0743923  2.2128754  0.0418
anoleEMtwig         -0.07130126 0.0865764 -0.8235647  0.4223

 Correlation:
                    (Intr) anlSVL anlEMg- anlEMtr anlEMtrnk-c anlEMtrnk-g
anoleSVL            -0.992
anoleEMgrass-bush   -0.872  0.842
anoleEMtrunk        -0.720  0.682  0.766
anoleEMtrunk-crown  -0.759  0.723  0.798   0.705
anoleEMtrunk-ground -0.801  0.763  0.868   0.742   0.784
anoleEMtwig         -0.854  0.817  0.859   0.758   0.780       0.825

Standardized residuals:
       Min         Q1        Med         Q3        Max
-0.9328721 -0.3272332  0.1773186  0.9921334  2.2200491

Residual standard error: 0.08067145
Degrees of freedom: 23 total; 16 residual

> anova(pglsModel3)

Denom. DF: 16
            numDF   F-value p-value
(Intercept)     1 16.352573  0.0009
anoleSVL        1  0.033595  0.8569
anoleEM         5  6.666758  0.0016

> pglsModel4 <- gls(anoleSSD ~ anoleSVL * anoleEM, correlation = corBrownian(1,
+     anoleTree), data = dd)
> summary(pglsModel4)
```

```
Generalized least squares fit by REML
  Model: anoleSSD ~ anoleSVL * anoleEM
  Data: dd
       AIC       BIC    logLik
  -6.93665 -1.764012 16.46833

Correlation Structure: corBrownian
 Formula: ~1
 Parameter estimate(s):
numeric(0)

Coefficients:
                                Value Std.Error    t-value p-value
(Intercept)                  0.5642067 0.8484164  0.6650115  0.5197
anoleSVL                    -0.0922728 0.1766093 -0.5224688  0.6117
anoleEMgrass-bush           -1.7736903 1.4012093 -1.2658283  0.2317
anoleEMtrunk                -0.0325667 1.2809828 -0.0254232  0.9802
anoleEMtrunk-crown           1.0408742 1.4116578  0.7373418  0.4763
anoleEMtrunk-ground         -0.3515696 2.1902669 -0.1605145  0.8754
anoleEMtwig                 -0.6017230 1.0455342 -0.5755173  0.5765
anoleSVL:anoleEMgrass-bush   0.4463891 0.3423834  1.3037697  0.2189
anoleSVL:anoleEMtrunk       -0.0087573 0.2949895 -0.0296868  0.9768
anoleSVL:anoleEMtrunk-crown -0.2378074 0.3254682 -0.7306623  0.4803
anoleSVL:anoleEMtrunk-ground 0.1148927 0.5248773  0.2188943  0.8307
anoleSVL:anoleEMtwig         0.1189717 0.2349086  0.5064595  0.6225

 Correlation:
                             (Intr) anlSVL anlEMg- anlEMtr anlEMtrnk-c
anoleSVL                     -0.998
anoleEMgrass-bush            -0.616  0.614
anoleEMtrunk                 -0.657  0.657  0.355
anoleEMtrunk-crown           -0.548  0.553  0.359   0.342
anoleEMtrunk-ground          -0.346  0.349  0.161   0.216   0.245
anoleEMtwig                  -0.848  0.841  0.543   0.530   0.440
anoleSVL:anoleEMgrass-bush    0.529 -0.528 -0.994  -0.294  -0.317
anoleSVL:anoleEMtrunk         0.594 -0.596 -0.311  -0.996  -0.310
anoleSVL:anoleEMtrunk-crown   0.492 -0.498 -0.327  -0.303  -0.997
anoleSVL:anoleEMtrunk-ground  0.296 -0.300 -0.131  -0.182  -0.221
anoleSVL:anoleEMtwig          0.796 -0.790 -0.514  -0.495  -0.411
                             anlEMtrnk-g anlEMtw anlSVL:nlEMg- anlSVL:nlEMtr
anoleSVL
anoleEMgrass-bush
anoleEMtrunk
anoleEMtrunk-crown
anoleEMtrunk-ground
anoleEMtwig                    0.239
```

```
anoleSVL:anoleEMgrass-bush   -0.125        -0.470
anoleSVL:anoleEMtrunk        -0.196        -0.472   0.256
anoleSVL:anoleEMtrunk-crown  -0.229        -0.392   0.290        0.276
anoleSVL:anoleEMtrunk-ground -0.998        -0.196   0.100        0.166
anoleSVL:anoleEMtwig         -0.212        -0.995   0.446        0.441
                             anlSVL:nlEMtrnk-c anlSVL:nlEMtrnk-g
anoleSVL
anoleEMgrass-bush
anoleEMtrunk
anoleEMtrunk-crown
anoleEMtrunk-ground
anoleEMtwig
anoleSVL:anoleEMgrass-bush
anoleSVL:anoleEMtrunk
anoleSVL:anoleEMtrunk-crown
anoleSVL:anoleEMtrunk-ground  0.207
anoleSVL:anoleEMtwig          0.366             0.172


Standardized residuals:
       Min          Q1          Med          Q3          Max
-0.75458125 -0.25233693  0.04831227  0.81728991  1.81794169


Residual standard error: 0.08549308
Degrees of freedom: 23 total; 11 residual

> anova(pglsModel4)

Denom. DF: 11
                 numDF    F-value  p-value
(Intercept)          1  14.560088   0.0029
anoleSVL             1   0.029912   0.8658
anoleEM              5   5.935982   0.0067
anoleSVL:anoleEM     5   0.649232   0.6683
```