Getting Started with nonnest2

Edgar C. Merkle and Dongjun You

2016-01-21

nonnest2 Package Overview

Package **nonnest2** was designed to implement Vuong's (1989) theory of non-nested model comparison for many classes of R models. It has been tested most thoroughly with structural equation models (of class `lavaan`) and with count regression models (including models of class `glm`, `hurdle`, `zeroinfl`). Functionality is available for many other classes, including `lm`, `glm.nb`, `clm`, `mlogit`, `nls`, `polr`, and `rlm`. Users are cautioned that, while we believe the results to be correct for these models (Vuong's theory is generally applicable to models fit via ML), we have not tested all combinations of models.

The package can be installed from CRAN, or the development version is available via github:

```
install.packages("nonnest2")

# to install the development version, run
# library(devtools)
# install_github("qpsy/nonnest2")
```

Analysis

In this section, we highlight the package's functionality using a series of examples. Given two models for comparison, there are different procedures depending on the relationship between the models. Analysts can usually discern whether two models are nested or non-nested, and different arguments are used in these two situations. For non-nested models, there is a further distinction between *overlapping* and *non-overlapping* models. Roughly, overlapping models are those that, under particular populations, share exactly the same density (with each model also possessing some unique densities under other populations). Non-overlapping models do not share any densities.

Because it is often difficult to tell whether or not non-nested models are overlapping, we typically utilize a two-step testing procedure described by Vuong (1989). In the first step, we test whether or not the two models are distinguishable from one another (this is a possibility if models are overlapping). In the second step, we test whether or not the two models' fits are equal. The examples below further illustrate the testing procedure along with the ideas of overlapping models and distinguishability.

Indistinguishable Models

Consider the two non-nested confirmatory factor analysis models below, specified via `lavaan` syntax.

```
m1 <- ' visual  =~ x1 + x2 + x3 + x4
        textual =~ x4 + x5 + x6
        speed   =~ x7 + x8 + x9 '
fit1 <- cfa(m1, data=HolzingerSwineford1939)

m2 <- ' visual  =~ x1 + x2 + x3
        textual =~ x4 + x5 + x6 + x7
```

```
        speed   =~ x7 + x8 + x9 '
fit2 <- cfa(m2, data=HolzingerSwineford1939)
```

Users familiar with the dataset will recognize that each model has an extra, unnecessary free loading: `m1` has a free loading from the visual factor to x4, while `m2` has a free loading from the textual factor to x7. The estimates of these extra loadings will both be close to 0, but they will not be exactly zero. Thus, while these two models will provide different fits to the data, the fits will be very similar. Further, if we fit the models to the entire population (assuming that the population values of these loadings equal zero), then the fits would be exactly the same. This illustrates the notion of *indistinguishability*: two models providing the same fit to a population of interest, but not necessarily to an observed sample. This differs from the idea of *equivalence*, which says that the models provide exactly the same fits to the sample data as well as to the population data.

We now use nonnest2 to compare the two models via Vuong tests.

```
vuongtest(fit1, fit2)
```

```
##
## Model 1
##   Class: lavaan
##   Call: lavaan::lavaan(model = m1, data = HolzingerSwineford1939, model.type = "cfa", ...
##
## Model 2
##   Class: lavaan
##   Call: lavaan::lavaan(model = m2, data = HolzingerSwineford1939, model.type = "cfa", ...
##
## Variance test
##   H0: Model 1 and Model 2 are indistinguishable
##   H1: Model 1 and Model 2 are distinguishable
##     w2 = 0.006,   p = 0.479
##
## Non-nested likelihood ratio test
##   H0: Model fits are equal for the focal population
##   H1A: Model 1 fits better than Model 2
##     z = 0.432,   p = 0.333
##   H1B: Model 2 fits better than Model 1
##     z = 0.432,   p = 0.6672
```

We see that we obtain two tests, the *variance test* and the *non-nested likelihood ratio test*. The former test informs us of the models' distinguishability, while the latter test compares the fits of two distinguishable models. Focusing on the variance test here, we obtain a small test statistic and a large p-value. These results imply that, as we suspected, the two models are indistinguishable in the focal population.

We can also use Vuong's theory to obtain confidence intervals for the difference in the models' AIC or BIC statistics:

```
icci(fit1, fit2)
```

```
##
## Model 1
##   Class: lavaan
##   Call: lavaan::lavaan(model = m1, data = HolzingerSwineford1939, model.type = "cfa", ...
##   AIC: 7518.240
##   BIC: 7599.796
```

```
##
## Model 2
##   Class: lavaan
##   Call: lavaan::lavaan(model = m2, data = HolzingerSwineford1939, model.type = "cfa", ...
##   AIC: 7519.391
##   BIC: 7600.948
##
## 95% Confidence Interval of AIC difference (AICdiff = AIC1 - AIC2)
##    -6.372 < AICdiff < 4.070
##
## 95% Confidence Interval of BIC difference (BICdiff = BIC1 - BIC2)
##    -6.372 < BICdiff < 4.070
```

Based on this output, we see that the AIC and BIC statistics are close but lower for `fit1` than for `fit2`. The 95% confidence intervals overlap with 0, implying that the model fits are sufficiently close that neither can be preferred over the other. The confidence intervals for AIC and BIC differences are exactly the same because both models have the same numbers of parameters; this would not typically occur for other models.

Distiguishable and Non-Nested models

Now consider the following structural equation models, using Bollen's political democracy data. The first model is the original (classic) model, while the second model estimates different residual covariance parameters. The models are non-nested due to the differing residual covariances.

```r
m1 <- '
      # latent variable definitions
         ind60 =~ x1 + x2 + x3
         dem60 =~ y1 + a*y2 + b*y3 + c*y4
         dem65 =~ y5 + a*y6 + b*y7 + c*y8

      # regressions
        dem60 ~ ind60
        dem65 ~ ind60 + dem60

      # residual correlations
        y1 ~~ y5
        y2 ~~ y4 + y6
        y3 ~~ y7
        y4 ~~ y8
        y6 ~~ y8
'

fit1 <- sem(m1, data=PoliticalDemocracy)

m2 <- '
      # latent variable definitions
         ind60 =~ x1 + x2 + x3
         dem60 =~ y1 + a*y2 + b*y3 + c*y4
         dem65 =~ y5 + a*y6 + b*y7 + c*y8

      # regressions
        dem60 ~ ind60
        dem65 ~ ind60 + dem60
```

```
      # residual correlations
        y1 ~~ y3 + y5
        y2 ~~ y6
        y3 ~~ y7
        y4 ~~ y8
        y5 ~~ y7
'

fit2 <- sem(m2, data=PoliticalDemocracy)
```

We now apply the Vuong tests to the two models.

```
vuongtest(fit1, fit2)
```

```
##
## Model 1
##  Class: lavaan
##  Call: lavaan::lavaan(model = m1, data = PoliticalDemocracy, model.type = "sem", ...
##
## Model 2
##  Class: lavaan
##  Call: lavaan::lavaan(model = m2, data = PoliticalDemocracy, model.type = "sem", ...
##
## Variance test
##   H0: Model 1 and Model 2 are indistinguishable
##   H1: Model 1 and Model 2 are distinguishable
##     w2 = 0.233,   p = 0.028
##
## Non-nested likelihood ratio test
##   H0: Model fits are equal for the focal population
##   H1A: Model 1 fits better than Model 2
##     z = 1.166,   p = 0.122
##   H1B: Model 2 fits better than Model 1
##     z = 1.166,   p = 0.8781
```

We now see that the variance tests indicates that the models are distinguishable (at least, at $\alpha = .05$). This allows us to move on to the second test, which compares the fits of the two models. Based on this second test, we conclude that the two models have equal fit in the population of interest. Note that *fit* is defined here in terms of Kullback-Leibler distance (between each candidate model and the true model), and we make no assumption about either of the two candidate models being the true model.

Nested models

Finally, `vuongtest()` includes a `nested` argument, using Vuong's theory to compare nested models. The resulting test statistics, which are very similar to the statistics described above, can be viewed as robust versions of the traditional likelihood ratio test (LRT) for nested models. This is because Vuong's theory makes no assumption about either candidate model being the true model, whereas the traditional LRT assumes that the full model is the truth.

We first fit a third political democracy model, removing the equality constraints on some of the factor loadings. As a result, `fit1` is nested in the model below.

```
m3 <- '
      # latent variable definitions
        ind60 =~ x1 + x2 + x3
        dem60 =~ y1 + y2 + y3 + y4
        dem65 =~ y5 + y6 + y7 + y8

      # regressions
        dem60 ~ ind60
        dem65 ~ ind60 + dem60

      # residual correlations
        y1 ~~ y5
        y2 ~~ y4 + y6
        y3 ~~ y7
        y4 ~~ y8
        y6 ~~ y8
'


fit3 <- sem(m3, data=PoliticalDemocracy)
```

We now use `vuongtest()` with the `nested=TRUE` argument.

```
vuongtest(fit1, fit3, nested=TRUE)
```

```
##
## Model 1
##   Class: lavaan
##   Call: lavaan::lavaan(model = m1, data = PoliticalDemocracy, model.type = "sem", ...
##
## Model 2
##   Class: lavaan
##   Call: lavaan::lavaan(model = m3, data = PoliticalDemocracy, model.type = "sem", ...
##
## Variance test
##   H0: Model 1 and Model 2 are indistinguishable
##   H1: Model 1 and Model 2 are distinguishable
##     w2 = 0.021,   p = 0.384
##
## Robust likelihood ratio test of distinguishable models
##   H0: Model 2 fits as well as Model 1
##   H1: Model 1 fits better than Model 2
##     LR = 2.054,   p = 0.384
```

Both tests can be viewed as alternatives to the traditional likelihood ratio test, and they both indicate that the model fits are equal. This implies that the reduced model (`fit1`) should be preferred. We can also compare to the traditional LRT:

```
anova(fit1, fit3)
```

```
## Chi Square Difference Test
##
##        Df    AIC    BIC  Chisq Chisq diff Df diff Pr(>Chisq)
```

```
## fit3 35 3157.6 3229.4 38.125
## fit1 38 3153.6 3218.5 40.179     2.0543        3     0.5612
```

which provides similar results. In particular, the Vuong likelihood ratio test statistic is equal to the traditional likelihood ratio test statistic, though the p-values differ. This is because the null distributions differ: the traditional LRT uses a chi-square null distribution, whereas the Vuong LRT uses a weighted sum of chi-square distributions. The latter distribution converges to the traditional chi-square distribution when the full model is the true model.