LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Simultaneous Inference in General Parametric Models

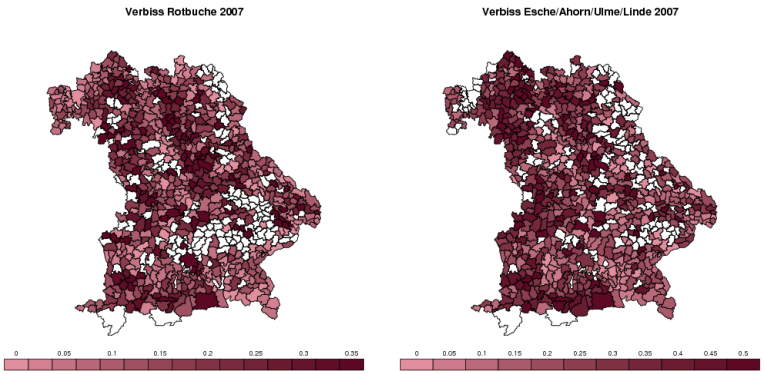
Torsten Hothorn
Institut für Statistik
(in collaboration with Frank Bretz, Novartis, and Peter Westfall, Texas Tech)

Introduction

Introduction

Introduction

Introduction



Model

$\mathcal{M}((\mathbf{Z}_1, \ldots, \mathbf{Z}_n), \theta, \eta)$ is a (semiparametric) model with

- n observations $(\mathbf{Z}_1, \ldots, \mathbf{Z}_n)$
- elemental parameters $\theta \in \mathbb{R}^p$ and
- other (random or nuisance) parameters η.

We are interested in linear functions $\vartheta := \mathbf{K}\theta$ defined by a constant matrix $\mathbf{K} \in \mathbb{R}^{k,p}$.

Estimation

$\hat{\theta}_n \in \mathbb{R}^p$ is an estimate of θ and $\mathsf{S}_n \in \mathbb{R}^{p,p}$ is an estimate of $\mathrm{cov}(\hat{\theta}_n)$ with

$$a_n \mathsf{S}_n \xrightarrow{\mathbb{P}} \Sigma \in \mathbb{R}^{p,p}$$

for some positive, nondecreasing sequence a_n.

A multivariate central limit theorem is assumed:

$$a_n^{1/2}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}_p(0, \Sigma).$$

We write $\hat{\theta}_n \overset{a}{\sim} \mathcal{N}_p(\theta, \mathsf{S}_n)$.

These assumptions are fulfilled for most of the models commonly in use.

Distribution of ϑ

By Theorem 3.3.A in Serfling (1980), the linear function $\hat{\vartheta}_n = \mathbf{K}\hat{\theta}_n$, i.e., an estimate of our parameters of interest, also follows an approximate multivariate normal distribution

$$\hat{\vartheta}_n = \mathbf{K}\hat{\theta}_n \overset{a}{\sim} \mathcal{N}_k(\vartheta, \mathsf{S}_n^\star)$$

with covariance matrix $\mathsf{S}_n^\star := \mathbf{K}\mathsf{S}_n\mathbf{K}^\top$ for any fixed matrix $\mathbf{K} \in \mathbb{R}^{k,p}$

Therefore, we simply assume

$$\hat{\vartheta}_n \overset{a}{\sim} \mathcal{N}_k(\vartheta, \mathsf{S}_n^\star) \text{ with } a_n \mathsf{S}_n^\star \xrightarrow{\mathbb{P}} \Sigma^\star := \mathbf{K}\Sigma\mathbf{K}^\top \in \mathbb{R}^{k,k}$$

A Statistic and its Distribution

Consider the multivariate statistic

$$\mathbf{T}_n := \mathbf{D}_n^{-1/2}(\hat{\vartheta}_n - \vartheta)$$

where $\mathbf{D}_n = \text{diag}(\mathsf{S}_n^\star)$ is the diagonal matrix given by the diagonal elements of $\mathsf{S}_n^\star$.

By Slutzky's Theorem, this statistic is again asymptotically normally distributed

$$\mathbf{T}_n \overset{a}{\sim} \mathcal{N}_k(0, \mathbf{R}_n)$$

where

$$\mathbf{R}_n = \mathbf{D}_n^{-1/2}\mathsf{S}_n^\star\mathbf{D}_n^{-1/2} \in \mathbb{R}^{k,k}$$

is the correlation matrix of the k-dimensional statistic $\mathbf{T}_n$.

General Linear Hypothesis

Consider the null hypothesis

$$H_0 : \vartheta := \mathbf{K}\theta = \mathbf{m}.$$

Classically, F- or χ^2-statistics are used to test H_0. However, a rejection of H_0 does not give further indication about the nature of the significant result. Therefore, one is often interested in the individual null hypotheses

$$H_0^j : \vartheta_j = \mathbf{m}_j.$$

Testing the hypotheses set $\{H_0^1, \ldots, H_0^k\}$ simultaneously thus requires the individual assessments while maintaining the familywise error rate.

A Maximum-Type Statistic

An alternative test statistic for testing H_0 is

$$\max(|\mathbf{T}_n|)$$

Can we approximate it's distribution under H_0 efficiently?

We have to find a good approximation of $\mathbb{P}(\max(|\mathbf{T}_n|) \leq t)$ for some $t \in \mathbb{R}^+$.

Null-Distribution and a Global Test

$$\mathbb{P}(\max(|\mathbf{T}_n|) \leq t) \cong \int_{-t}^{t} \cdots \int_{-t}^{t} \varphi_k(x_1, \ldots, x_k; \mathbf{R})\, dx_1 \cdots dx_k =: g(\mathbf{R}, t)$$

where φ_k is the k-dimensional normal density function.

$\mathbf{R}$ is not known but $g(\mathbf{R}, t)$ is a continuous function of $\mathbf{R}$ and converges as $\mathbf{R}_n \xrightarrow{\mathbb{P}} \mathbf{R}$. The integral can be approximated by quasi-randomized Monte-Carlo methods (Genz, 1992, Genz and Bretz, 1999).

The resulting global p-value for H_0 is then

$$p_{\text{global}} = 1 - g(\mathbf{R}_n, \max|\mathbf{t}|)$$

when $\mathbf{T} = \mathbf{t}$ has been observed.

Simultaneous Inference

But what about the partial hypotheses $H_0^1, \ldots, H_0^k$?

It's simple!

The multiplicity adjusted p-value for the jth individual two-sided hypothesis

$$H_0^j : \vartheta_j = \mathbf{m}_j, j = 1, \ldots, k,$$

is given by

$$p_j = 1 - g(\mathbf{R}_n, |t_j|),$$

where $\mathbf{t} = (t_1, \ldots, t_k)$ denote the observed test statistics (single-step procedure).

Reject each H_0^j at familywise error rate α when $p_j \leq \alpha$.

Simultaneous Confidence Intervals

A simultaneous $(1 - 2\alpha) \times 100\%$ confidence interval for ϑ is given by

$$\hat{\vartheta}_n \pm q_\alpha \text{diag}(\mathbf{D}_n)^{1/2}$$

where q_α is the (approximate) $1 - \alpha$ quantile of the distribution of $\max(|\mathbf{T}_n|)$.

Examples: Linear Regression

$\mathbf{Z}_i = (Y_i, \mathbf{X}_i), i = 1, \ldots, n$, with response Y_i and exploratory variables $\mathbf{X}_i = (X_{i1}, \ldots, X_{iq})$

Model:

$$Y_i = \beta_0 + \sum_{j=1}^{q} \beta_j X_{ij} + \sigma\varepsilon_i,$$

with elemental parameters $\theta = (\beta_0, \beta_1, \ldots, \beta_q)$ estimated via

$$\hat{\theta}_n = \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbf{Y} \sim \mathcal{N}_{q+1}\left(\theta, \sigma^2 \left(\mathbf{X}^\top \mathbf{X}\right)^{-1}\right).$$

Now

$$\hat{\vartheta}_n = \mathbf{K}\hat{\theta}_n \sim \mathcal{N}_k(\mathbf{K}\theta, \sigma^2 \mathbf{K} \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{K}^\top)$$

and

$$\mathbf{T}_n = \mathbf{D}_n^{-1/2} \hat{\vartheta}_n \sim t_{q+1}(n - q, \mathbf{R}) \quad \text{exact inference possible!}$$

Predicting Body Fat

Garcia et al. (2005) describe a linear model for total body fat prediction.

Aim: Based on $p = 9$ simple measurements (circumferences of elbow, knee etc) we want to estimate a simple (!) formula to predict the total body fat obtained for $n = 71$ healthy German women by means of Dual Energy X-Ray Absorptiometry.

Problem: Variable selection!

Linear Model Fit

```
R> data("bodyfat", package = "mboost")
R> lmod <- lm(DEXfat ~ ., data = bodyfat)
R> summary(lmod)

              Estimate Std. Error t value  Pr(>|t|)
(Intercept) -69.028276   7.516860 -9.1831 4.184e-13 ***
age           0.019962   0.032213  0.6197  0.537767
waistcirc     0.210487   0.067145  3.1348  0.002644 **
hipcirc       0.343513   0.080373  4.2740 6.852e-05 ***
elbowbreadth -0.412369   1.022907 -0.4031  0.688259
kneebreadth   1.757984   0.724952  2.4250  0.018286 *
anthro3a      5.742295   5.207524  1.1027  0.274492
anthro3b      9.866431   5.657864  1.7438  0.086224 .
anthro3c      0.387430   2.087463  0.1856  0.853376
anthro4      -6.574395   6.489177 -1.0131  0.314999
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Multiple R-squared: 0.923,        Adjusted R-squared: 0.912
F-statistic: 81.3 on 9 and 61 DF,  p-value: <2e-16
```


Parameters of Interest

```
R> library("multcomp")
R> K <- cbind(0, diag(length(coef(lmod)) - 1))
R> rownames(K) <- names(coef(lmod))[-1]
R> lmod_glht <- glht(lmod, linfct = K)
R> K
             [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
age             0    1    0    0    0    0    0    0    0     0
waistcirc       0    0    1    0    0    0    0    0    0     0
hipcirc         0    0    0    1    0    0    0    0    0     0
elbowbreadth    0    0    0    0    1    0    0    0    0     0
kneebreadth     0    0    0    0    0    1    0    0    0     0
anthro3a        0    0    0    0    0    0    1    0    0     0
anthro3b        0    0    0    0    0    0    0    1    0     0
anthro3c        0    0    0    0    0    0    0    0    1     0
anthro4         0    0    0    0    0    0    0    0    0     1
```


F-Test

```
R> summary(lmod_glht, test = Ftest())
         General Linear Hypotheses

Linear Hypotheses:
                   Estimate
age == 0            0.01996
waistcirc == 0      0.21049
hipcirc == 0        0.34351
elbowbreadth == 0 -0.41237
kneebreadth == 0    1.75798
anthro3a == 0       5.74230
anthro3b == 0       9.86643
anthro3c == 0       0.38743
anthro4 == 0       -6.57439

Global Test:
      F DF1 DF2    Pr(>F)
1 81.35   9  61 1.387e-30
```


Maximum Test

```
R> summary(lmod_glht)
        Simultaneous Tests for General Linear Hypotheses

Fit: lm(formula = DEXfat ~ ., data = bodyfat)

Linear Hypotheses:
                   Estimate Std. Error t value Pr(>|t|)
age == 0            0.01996    0.03221   0.620   0.9959
waistcirc == 0      0.21049    0.06714   3.135   0.0213 *
hipcirc == 0        0.34351    0.08037   4.274   <0.001 ***
elbowbreadth == 0 -0.41237    1.02291  -0.403   0.9998
kneebreadth == 0    1.75798    0.72495   2.425   0.1316
anthro3a == 0       5.74230    5.20752   1.103   0.8948
anthro3b == 0       9.86643    5.65786   1.744   0.4783
anthro3c == 0       0.38743    2.08746   0.186   1.0000
anthro4 == 0       -6.57439    6.48918  -1.013   0.9295
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)
```

ANOVA

Model:

$$Y_{ij} = \mu + \gamma_j + \varepsilon_{ij}$$

Overparameterized, usually the elemental parameters are $\theta = (\mu, \gamma_2 - \gamma_1, \gamma_3 - \gamma_1, \ldots, \gamma_q - \gamma_1)$.

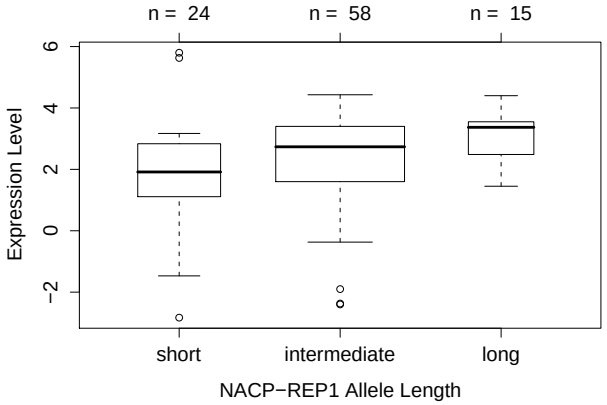
Dunnett many-to-one comparisons:

$$\begin{aligned} \mathbf{K}_{\text{Dunnett}} &= (0, \text{diag}(q)) \\ \vartheta_{\text{Dunnett}} &= \mathbf{K}_{\text{Dunnett}}\theta = (\gamma_2 - \gamma_1, \gamma_3 - \gamma_1, \ldots, \gamma_q - \gamma_1) \end{aligned}$$

Tukey all-pair comparisons:

$$\begin{aligned} \mathbf{K}_{\text{Tukey}} &= \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & -1 \end{pmatrix} \\ \vartheta_{\text{Tukey}} &= \mathbf{K}_{\text{Tukey}}\theta = (\gamma_2 - \gamma_1, \gamma_3 - \gamma_1, \gamma_2 - \gamma_3) \end{aligned}$$

Genetic Components of Alcoholism



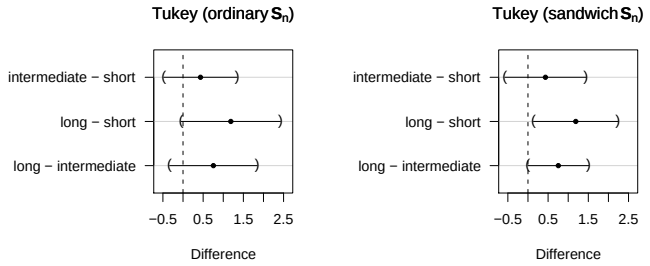
Genetic Components of Alcoholism

```
R> data("alpha", package = "coin")
R> amod <- aov(elevel ~ alength, data = alpha)
R> confint(glht(amod, linfct = mcp(alength = "Tukey")))
         Simultaneous Confidence Intervals

Multiple Comparisons of Means: Tukey Contrasts


Fit: aov(formula = elevel ~ alength, data = alpha)

Estimated Quantile = 2.3714
95% family-wise confidence level


Linear Hypotheses:
                       Estimate lwr      upr
intermediate - short == 0  0.43415 -0.47561  1.34391
long - short == 0          1.18875 -0.04498  2.42248
long - intermediate == 0   0.75460 -0.33118  1.84038
```


Genetic Components of Alcoholism

```
R> amod_glht_sw <- glht(amod, linfct = mcp(alength = "Tukey"),
+                       vcov = sandwich)
R> confint(amod_glht_sw)
         Simultaneous Confidence Intervals

Multiple Comparisons of Means: Tukey Contrasts


Fit: aov(formula = elevel ~ alength, data = alpha)

Estimated Quantile = 2.3718
95% family-wise confidence level


Linear Hypotheses:
                         Estimate   lwr        upr
intermediate - short == 0  0.4341523 -0.5713432  1.4396478
long - short == 0          1.1887500  0.1376593  2.2398407
long - intermediate == 0   0.7545977 -0.0005049  1.5097003
```

Genetic Components of Alcoholism



Generalized Mixed Models

Model:

$$\mathbb{E}(\mathbf{Y}_i) = h(\mathbf{X}_i\theta + \mathbf{Z}\mathbf{b}_i)$$

for the n_i observations in group i with random effects $\mathbf{b}_i$.

We are interested in inference about $\mathbf{K}\theta$.

For example in a logistic mixed model, in confidence intervals for the predicted probabilities in $\hat{\vartheta}_n = \mathbf{X}\hat{\theta}_n$

$$\left(\left(1 + \exp\left(-\left(\hat{\vartheta}_n - q_\alpha \text{diag}(\mathbf{D}_n)^{1/2}\right)\right)\right)^{-1},$$
$$\left(1 + \exp\left(-\left(\hat{\vartheta}_n + q_\alpha \text{diag}(\mathbf{D}_n)^{1/2}\right)\right)\right)^{-1}\right).$$

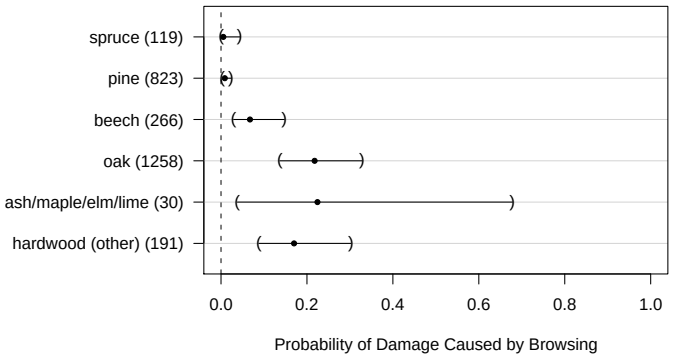
Dear Browsing in Frankonia

```
R> mmod <- lmer(damage ~ species - 1 + (1 | lattice / plot),
+               data = trees513, family = binomial())
R> K <- diag(length(fixef(mmod)))
```


Dear Browsing in Frankonia

```
R> ci <- confint(glht(mmod, linfct = K))
R> ci$confint <- 1 - binomial()$linkinv(ci$confint)
R> ci$confint[,2:3] <- ci$confint[,3:2]
R> ci
         Simultaneous Confidence Intervals

Fit: glmer(formula = damage ~ species - 1 + (1 | lattice/plot), data = trees513,
    family = binomial())

Estimated Quantile = 2.6057
95% family-wise confidence level


Linear Hypotheses:
                              Estimate  lwr       upr
spruce (119) == 0             0.0053819 0.0006415 0.0436233
pine (823) == 0               0.0087864 0.0032629 0.0234403
beech (266) == 0              0.0673833 0.0293407 0.1472677
oak (1258) == 0               0.2178359 0.1370749 0.3280881
ash/maple/elm/lime (30) == 0 0.2244547 0.0382305 0.6781659
hardwood (other) (191) == 0   0.1699804 0.0883225 0.3021162
```

Dear Browsing in Frankonia



Odds-Ratios

Agresti et al. (2008) propose simultaneous confidence intervals for odds-ratios. Simultaneous Wald intervals can be derived from a logistic regression model:

```
R> resp <- cbind(succ = c(13, 27, 22, 9),
+                fail = c(87, 86, 87, 87) - c(13, 27, 22, 9))
R> trt <- as.factor(c("Coenzyme", "Remacemide", "Combination", "Placebo"))
R> mod <- glm(resp ~ trt, family = binomial())
R> exp(confint(glht(mod, mcp(trt = "Tukey")))$confint)
                            Estimate       lwr       upr
Combination - Coenzyme    1.9266272 0.7105033  5.224314
Placebo - Coenzyme        0.6568047 0.2003154  2.153566
Remacemide - Coenzyme     2.6049544 0.9828680  6.904068
Placebo - Combination     0.3409091 0.1131985  1.026683
Remacemide - Combination  1.3520801 0.5669658  3.224393
Remacemide - Placebo      3.9661017 1.3444018 11.700344
attr(,"conf.level")
[1] 0.95
attr(,"calpha")
[1] 2.564786
attr(,"error")
[1] 6.103516e-05
```


Multivariate Time Series

Haufe et al. (NIPS 2008) investigate "spatial causal discovery in multivariate time series" by vector autoregressive models and aim to identify non-vanishing coefficients in these models, for example fitted using Ridge regression.

Multiple tests for this variable selection problem perform as good as a group Lasso approach.

Mixture Models

Leisch and Hothorn (in preparation) aim to identify

- non-zero parameters in components of a mixture model (component-wise variable selection) and
- parameters that are equal in two or more components of a mixture model.

Once an estimate of the variance-covariance matrix of all parameters is available the presented theory and computational infrastructure in **multcomp** can be applied.

References

Alan Agresti, Matilde Bini, Bruno Bertaccini, and Euijung Ryu. Simultaneous confidence intervals for comparing binomial parameters. Biometrics, 64, 1270–1275, 2008.

Ada L. Garcia, Karen Wagner, Torsten Hothorn, Corinna Koebnick, Hans-Joachim F. Zunft, and Ulrike Trippo. Improved prediction of body fat by measuring skinfold thickness, circumferences, and bone breadths. Obesity Research, 13(3):626–634, 2005.

Alan Genz. Numerical computation of multivariate normal probabilities. Journal of Computational and Graphical Statistics, 1:141–149, 1992.

Alan Genz and Frank Bretz. Numerical computation of multivariate t-probabilities with application to power calculation of multiple contrasts. Journal of Statistical Computation and Simulation, 63:361–378, 1999.

Stefan Haufe, Klaus-Robert Müller, Guido Nolte, and Nicole Krämer. Sparse causal discovery in multivariate time series. JMLR: Workshop and Conference Proceedings, to appear.

Torsten Hothorn, Frank Bretz and Peter Westfall. Simultaneous inference in general parametric models. Biometrical Journal, 50(3), 2008.

Robert J. Serfling. Approximation Theorems of Mathematical Statistics. John Wiley & Sons, New York, 1980.