

Informative hypotheses evaluation

Rebecca M. Kuiper

Department of Methodology & Statistics
Utrecht University

Table of Contents

Example Null hypothesis (H_0) vs Informative hypothesis (H_m)

Generalized Order-Restricted Information Criterion (GORIC)

GORIC weights

Examples with R code

Failsafe/Safeguard hypotheses: Unconstrained & Complement

GORICA

GORIC(A) in JASP

End & Extra

ANOVA Example: Comparisons of 3 Means

Example Palmer and Gough (P&G)

Palmer and Gough (2007) examined the difference in the attribution of importance of defective education as an explanation for criminal behavior between three types of “offenders”:
(1) non-offenders, (2) property offenders, and (3) person offenders.

ANOVA

$$y_j = \mu_1 d_{1j} + \mu_2 d_{2j} + \mu_3 d_{3j} + \epsilon_j,$$

where μ_i ($i = 1, 2, 3$) is the mean in group i ,
 d_{ij} is a dummy variable denoting group membership,
 ϵ_j is an error term and $\epsilon_j \sim N(0, \sigma^2)$.

$(n_i = 20, 20, 31)$

Hypotheses of interest

Example Palmer and Gough (P&G)

Palmer and Gough (2007) expect that

- non-offenders (1) attribute more importance to defective education for explaining crime than the other two offenders (2 & 3):
i.e., $\mu_1 > \mu_2$ and $\mu_1 > \mu_3$,
- property offenders (2) attribute more importance to defective education than person offenders (3):
i.e., $\mu_2 > \mu_3$.

This leads to the theory-based hypothesis:

$$H_1 : \quad \mu_1 > \mu_2 > \mu_3.$$

Note that ‘<’ denotes “smaller than” and ‘>’ denotes “larger than”.

ANOVA Example: Comparisons of 3 Means

Simple, hypothetical example

Examine the difference in happiness
between three types of “treatments”:
(1) new treatment, (2) current treatment, and (3) no treatment.

Theory-based hypothesis:

$$H_1 : \quad \mu_1 > \mu_2 > \mu_3,$$

where “>” denotes “larger than”.

Testing the null hypothesis

Example 3 means

Test H_0 with ANOVA F test:

$$H_0 : \quad \mu_1 = \mu_2 = \mu_3,$$

$$H_a : \quad \text{not } H_0.$$

Then, reject or not-reject ('accept') H_0 .

But, often not interested in H_0 !

Cannot say anything about $H_1 : \mu_1 > \mu_2 > \mu_3$.

Conclusions w.r.t. hypothesis of interest $H_1 : \mu_1 > \mu_2 > \mu_3$

Example 3 means

With post-hoc test (e.g., Bonferroni) and on basis of the sample means, we could say a bit more.

But, what if

- $\mu_1 \neq \mu_2, \mu_1 \neq \mu_3, \mu_2 = \mu_3,$
- $\bar{y}_1 > \bar{y}_2,$

which implies $\mu_1 > \mu_2 = \mu_3$.

But also not interested in this.

Furthermore, inconsistent results are possible (like in P&G example):

$\mu_1 \neq \mu_3, \mu_1 = \mu_2,$ and $\mu_2 = \mu_3,$

which is logically impossible, since latter two imply $\mu_1 = \mu_3$.

Using planned contrasts would overcome this, but still = restrictions then.

NHST

Example Palmer and Gough

```
PandG_data <- read.table("Data_PalmerAndGough.txt",
                          header=TRUE)
PandG_data$group <- factor(PandG_data$group)
pairwise.t.test(PandG_data$Importance, PandG_data$group,
                p.adj = 'bonferroni')

##
## Pairwise comparisons using t tests with pooled SD
##
## data: PandG_data$Importance and PandG_data$group
##
##      1      2
## 2 0.191 -
## 3 0.011 1.000
##
## P value adjustment method: bonferroni
```


GORIC

Example Palmer and Gough, that is, 3 means

$$H_0 : \mu_1 = \mu_2 = \mu_3,$$

$$H_1 : \mu_1 > \mu_2 > \mu_3,$$

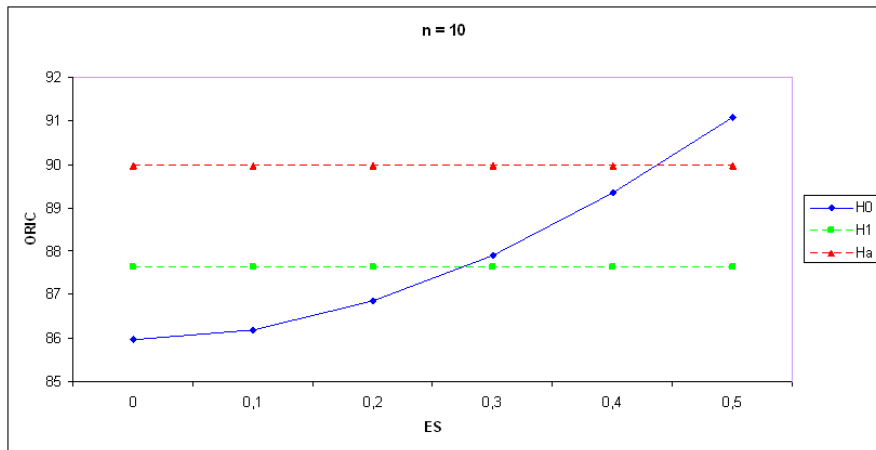
$$H_u : \mu_1, \mu_2, \mu_3.$$

GORIC

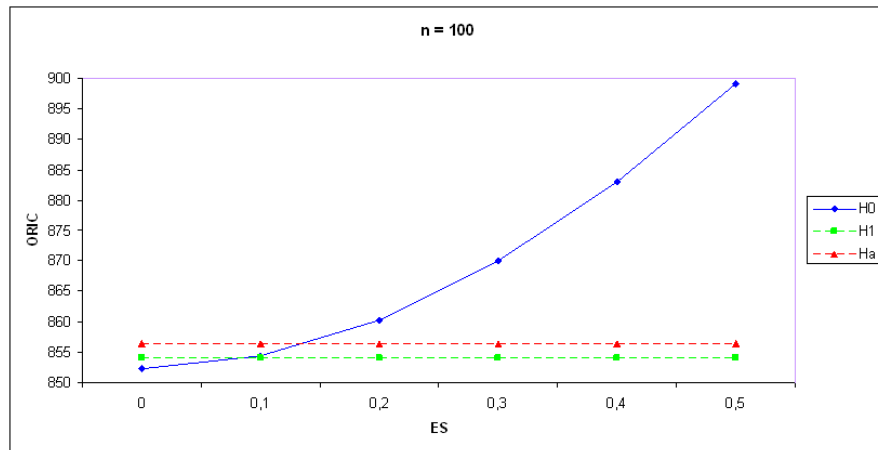
Model	Fit	Complexity	GORIC	GORIC weights
H_0	-196.36	2.00	396.71	0.02
H_1	-191.89	2.81	389.41	0.75
H_u	-191.89	4.00	391.79	0.23

Confirmatory methods (e.g., GORIC) have more “power” than their exploratory counterparts (e.g., AIC; cf. one- vs two-sided testing).

Confirmation more power: 1 data set. GORIC values for 3 groups, effect size ES , and $n = 10$ observations per group



Confirmation more power: 1 data set. GORIC values for 3 groups, effect size ES , and $n = 100$ observations per group



Information criteria (ICs)

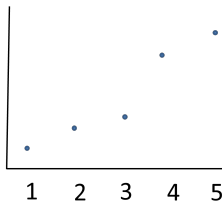
IC, like AIC and GORIC, balances fit and complexity.

Describe data as good as possible (fit)
with fewest number of parameters (simplicity / non-complexity).

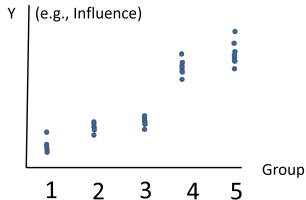
Intermezzo: Balance Fit and Complexity (1/6)

Example 5 means

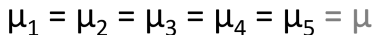
Data for 5 groups – Oversimplified representation



Data for 5 groups – Better representation (scatter of data points)



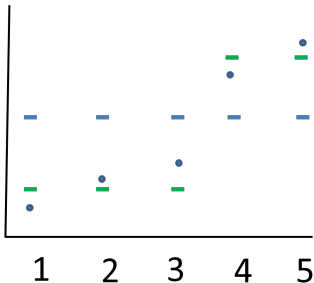
Example 5 means



Fit: Bad fit

Intermezzo: Balance Fit and Complexity (3/6)

Example 5 means

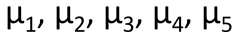


$$\mu_1 = \mu_2 = \mu_3 = \mu_{\text{low}}$$

$$\mu_4 = \mu_5 = \mu_{\text{high}}$$

parameters: 2 mean parameters (i.e., μ_{low} and μ_{high})
Fit: Better than with one parameter, even good fit.

Example 5 means

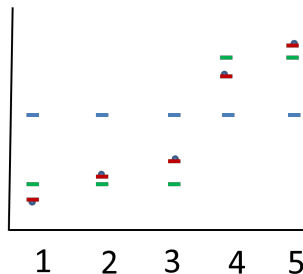


Fit: Better than with two parameters, even best possible fit.

◀ ◻ ▶ ◀ ◻ ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡ ↺ 🔍 ↻ 21/109

Intermezzo: Balance Fit and Complexity (5/6)

Example 5 means



$$\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu$$

$$\mu_1 = \mu_2 = \mu_3 = \mu_{\text{low}} \text{ \& }$$

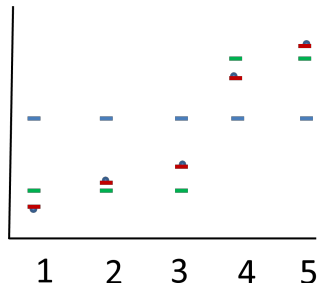
$$\mu_4 = \mu_5 = \mu_{\text{high}}$$

$$\mu_1, \mu_2, \mu_3, \mu_4, \mu_5$$

In this example: **Two means** may be best trade-off between fit & complexity.

Intermezzo: Balance Fit and Complexity (6b/6)

Order-restrictions: From “Example 5 means” to ‘Example 2 means’

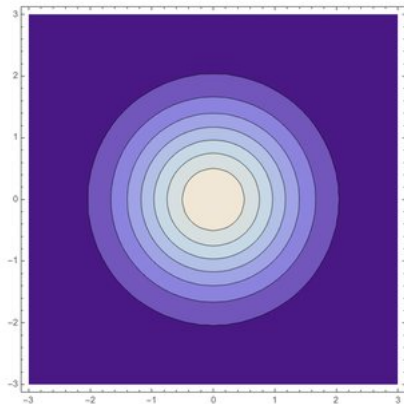
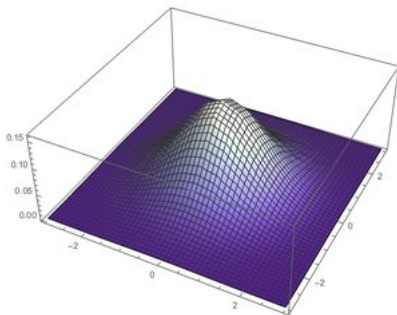


In theory-based hypotheses, we also incorporate **order-restrictions** (e.g., $\mu_1 < \mu_2$). Then, helpful to look at likelihood using **contour plots** (not scatter plot, as done here).

For ease, I will next use **2 means** (instead of 5).

Intermezzo: Contour plot (1/2)

Example 2 means

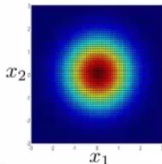
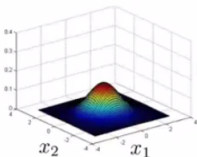


Intermezzo: Contour plot (2/2)

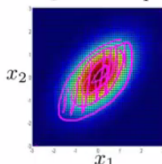
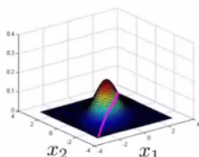
Example 2 means

Multivariate Gaussian (Normal) examples

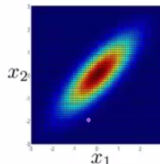
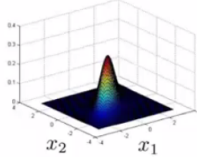
$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$



$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Sigma = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$$



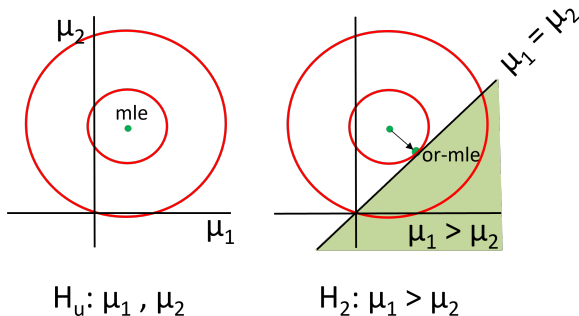
$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Sigma = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}$$



Andreas

Idea fit: order-restricted maximum likelihood (or-ml)

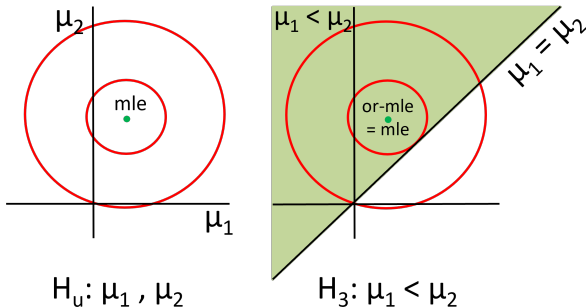
Example 2 means



mle not in H_2 : Find highest likelihood in allowable (= green) space.
The resulting estimated mean $\hat{\mu} = (\hat{\mu}_1, \hat{\mu}_2)$ is referred to as or-ml.

Note: if in agreement, or-mle = mle

Example 2 means



mle in H_3 : highest likelihood is in allowable (= green) space.
Hence, or-mle = mle.

Idea fit

Example Palmer and Gough, that is, 3 means

$$H_0 : \mu_1 = \mu_2 = \mu_3,$$

$$H_1 : \mu_1 > \mu_2 > \mu_3,$$

$$H_2 : \mu_1 > \mu_2 < \mu_3,$$

$$H_3 : \mu_1 < \mu_2 < \mu_3,$$

$$H_u : \mu_1, \mu_2, \mu_3.$$

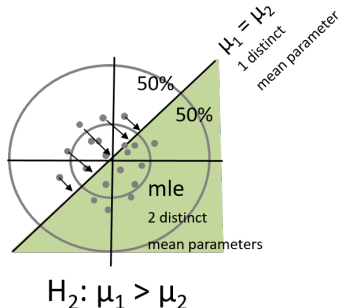
```
## group1 group2 group3
## 11.95  9.75  8.77
```

GORIC

Model	Fit	Complexity	GORIC	GORIC weights
H_0	-196.36	2.00	396.71	0.01
H_1	-191.89	2.81	389.41	0.56
H_2	-192.34	3.19	391.05	0.25
H_3	-196.36	2.81	398.34	0.01
H_u	-191.89	4.00	391.79	0.17

Idea complexity: penalty (PT)

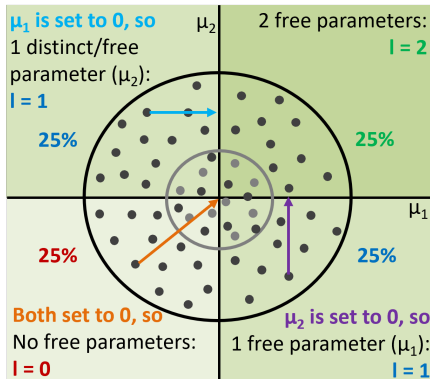
Example 2 means



Note: sampling from distribution with $\mu_1 = \mu_2 = 0$.
 complexity $H_2 = PT_2 = 1 + 0.5 \times 1 + 0.5 \times 2 = 2.5$.
 PT = number of expected distinct/free parameters.

Extra: Levels and level probabilities

Example 2 means



$H_4: \mu_1 > 0, \mu_2 > 0$

Level (l): number of distinct/free parameters.

Level probability (LP_l): probability that there are l levels.

$PT_4 = 1 + 0.25 \times 0 + (0.25 + 0.25) \times 1 + 0.25 \times 2 = 2.$

GORIC

$$IC_m = -2 \text{ fit}_m + 2 \text{ complexity}_m$$

Broad type of restrictions

More or less: any linear restriction.

e.g., the interaction $H_1 : \mu_1 - \mu_2 < \mu_3 - \mu_4$.

Note

If no inequalities ($<$ and/or $>$), then (G)ORIC = AIC.

Reference:

Kuiper, R.M., Hoijtink, H. and Silvapulle, M.J. (2011). An Akaike type information criterion for model selection under inequality constraints. *Biometrika*, 98, 495-501.

Hence, smallest value is best.

GORIC

Example Palmer and Gough

- H_0 : $\mu_1 = \mu_2 = \mu_3$ (include only when of interest)
- H_1 : $\mu_1 > \mu_2 > \mu_3$
- H_2 : $\mu_2 > \mu_1 > \mu_3$ (if competing theory/hypothesis)
- H_u : μ_1, μ_2, μ_3 .

GORIC

Model	Fit	Complexity	GORIC
H_0	-196.36	2.00	396.71
H_1	-191.89	2.81	389.41
H_2	-193.70	2.81	393.03
H_u	-191.89	4.00	391.79

Hands-on/Demo (1a): GORIC

Let's practice.

- Go to <https://github.com/rebeccakuiper/Tutorials>:
 1. Click on green button called Code.
 2. Download zip (last option in list).
 3. Unzip it on your machine (that folder is now your working dir.).
- Start Rstudio. Optional: make project.
- Open 'Tutorial_GORIC_restriktor_ANOVA.html', 'PalmerAndGough_and_Lucas.R', and/or 'Hands-on_1_GORIC_Unc_ANOVA_restriktor.R' ('Hands-on files').
- Install packages and load them.
- Read and inspect data.
Use **Data_PalmerAndGough.txt** and/or Data_Lucas.txt.
- Run model (`lm()`).
- Specify hypotheses (make up your own).
Note: Use names used in the model.
- Run `goric()`.
- Inspect and interpret output.
Note: 'GORIC weights' will be explained next.



Table of Contents

Generalized Order-Restricted Information Criterion (GORIC)

GORIC weights

Examples with R code

GORICA

GORIC(A) in JASP

End & Extra

Illustration of the GORIC weights (w_m)

Example Palmer and Gough

- H_0 : $\mu_1 = \mu_2 = \mu_3$ (include only when of interest)
- H_1 : $\mu_1 > \mu_2 > \mu_3$
- H_2 : $\mu_2 > \mu_1 > \mu_3$ (if competing theory/hypothesis)
- H_u : μ_1, μ_2, μ_3 .

GORIC

Model	Fit	Complexity	GORIC	GORIC weights
H_1	-191.89	2.81	389.41	0.68
H_2	-193.70	2.81	393.03	0.11
H_u	-191.89	4.00	391.79	0.21

H_1 is $0.68 / 0.11 \approx 6.11$ times more supported than H_2 .

Hands-on/Demo (1b): GORIC weights

Let's practice.

- If needed: Start Rstudio again (optional: make project) and then also load packages again.
- Read and inspect data.
- Run model (`lm()`).
- Specify hypotheses (make up your own).
Note: Use names used in the model.
- Run `goric()`.
- Inspect and interpret output: Focus on GORIC weights.

Another illustration: one-way ANOVA

Lucas: 5 groups

Lucas (2003) investigated difference between female and male leadership w.r.t. influence of the leader.

Five experimental groups:

1. a randomly selected male leader
2. a randomly selected female leader
3. male leader selected via task
4. female leader selected via task
5. female leader selected via task + institutionalized female leadership via movie

(Two informative) hypotheses of interest

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ (include only when of interest)

$H_1 : \mu_5 = \mu_3 > \{\mu_1, \mu_4\} > \mu_2$

$H_2 : \mu_3 > \mu_1 > \mu_4 = \mu_5 > \mu_2$

$H_u : \mu_1, \mu_2, \mu_3, \mu_4, \mu_5$ (included as failsafe)

Descriptive statistics of Lucas' Data

Group	Mean Influence	SD	n
1 (male, random)	2.33	1.86	30
2 (female, random)	1.33	1.15	30
3 (male, selected)	3.20	1.79	30
4 (female, selected)	2.23	1.45	30
5 (female, selected+)	3.23	1.50	30

GORIC output

Lucas

$$H_1 : \mu_5 = \mu_3 > \{\mu_1, \mu_4\} > \mu_2$$

$$H_2 : \mu_3 > \mu_1 > \mu_4 = \mu_5 > \mu_2,$$

$$H_u : \mu_1, \mu_2, \mu_3, \mu_4, \mu_5.$$

GORIC

Model	Fit	Complexity	GORIC	GORIC weights
H_1	-278.05	3.46	563.02	0.90
H_2	-281.76	3.14	569.79	0.03
H_u	-278.05	6.00	568.10	0.07

GORIC weights (w_m)

Lucas

GORIC

Model	Fit	Complexity	GORIC	GORIC weights
H_1	-278.05	3.46	563.02	0.90
H_2	-281.76	3.14	569.79	0.03
H_u	-278.05	6.00	568.10	0.07

H_1 is $0.9 / 0.03 \approx 29.59$ times more supported than H_2 .
Thus, there is quite some evidence that H_1 is the best of this set.

Note:

H_1 is $0.9 / 0.07 > 1$ times more supported than H_u .
Hence, H_1 not weak and thus comparable to the other
(weak and non-weak) informative hypotheses (as explained in a bit).

Table of Contents

Generalized Order-Restricted Information Criterion (GORIC)

GORIC weights

Examples with R code

Failsafe/Safeguard hypotheses: Unconstrained & Complement

GORICA

GORIC(A) in JASP

End & Extra

Include “unconstrained” hypothesis

If set of hypotheses does not contain a reasonable/good one:
Select the best of set of weak hypotheses.
E.g.: $w_1 = .8$ and $w_2 = .2$.

Prevent choosing a weak hypothesis

Include unconstrained hypothesis H_u (or H_a):

$$H_1 : \mu_1 > \mu_2 > \mu_3,$$

$$H_2 : \mu_2 > \mu_1 > \mu_3$$

$$H_u : \quad \mu_1, \mu_2, \mu_3, \\ \text{(i.e., no restrictions).}$$

H_u highest fit but also most complex, thus failsafe/safeguard.

E.g.: $w_1 = .08$, $w_2 = .02$, and $w_u = .90$.

What if another hypothesis is true?

What if another informative hypothesis is true?

The failsafe hypothesis will be preferred over the informative hypotheses in the set, if the sample size is large enough.

E.g.: $w_1 = .08$, $w_2 = .02$, and $w_\mu = .90$.

H_1 vs H_u
Palmer & Gough

What if only one informative hypothesis:

$$H_1: \mu_1 > \mu_2 > \mu_3,$$

$$H_u : \quad \mu_1, \mu_2, \mu_3.$$

GORIC

Model	Fit	Complexity	GORIC	GORIC weights
H_1	-191.89	2.81	389.41	0.77
H_u	-191.89	4.00	391.79	0.23

H_1 is $0.77 / 0.23 \approx 3.27$ times more supported than H_{μ} .

BUT: H_{II} includes H_1 .

So, support for H_μ contains support for H_1 .

H_1 vs H_u ctd.

Palmer & Gough

GORIC

Model	Fit	Complexity	GORIC	GORIC weights
H_1	-191.89	2.81	389.41	0.77
H_u	-191.89	4.00	391.79	0.23

Unconstrained is always true, so will always receive support.
Support for H_{ii} contains support for H_1 (if any).

Now: Fit H_1 and H_u the same;
thus, GORIC weights are based on only complexity.

Consequently, $0.77 / 0.23 \approx 3.27$ is an upper bound.

H_1 vs H_u : Upper bound

Based on simulation

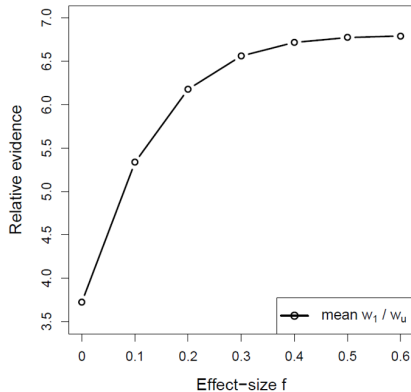
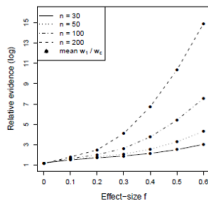


Figure: IC weights can have an upper bound, when informative hypothesis has maximum fit (i.e., is fully in agreement with the data).

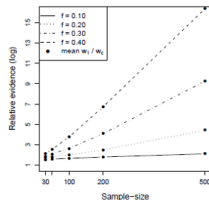
Alternative failsafe: Complement of H_m

Based on simulation

vs complement

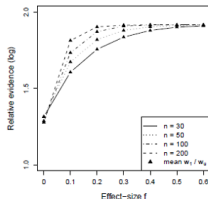


(a)

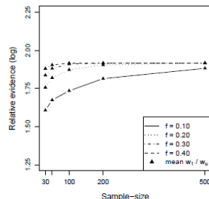


(b)

vs unconstrained



(c)



(d)

Note: unconstrained vs complement

H_1 contains 1 ordering of means:

1. $\mu_1 > \mu_2 > \mu_3$

H_c contains 5 orderings of means:

2. $\mu_1 > \mu_3 > \mu_2$
3. $\mu_2 > \mu_1 > \mu_3$
4. $\mu_2 > \mu_3 > \mu_1$
5. $\mu_3 > \mu_1 > \mu_2$
6. $\mu_3 > \mu_2 > \mu_1$

H_u combines H_1 and H_c , and thus consists of all 6 possible orderings.

R code: complement

Palmer & Gough

$$H_1 : \quad \mu_1 > \mu_2 > \mu_3,$$

$$H_c : \quad \text{not } H_1.$$

GORIC

```

H1 <- 'group1 > group2 > group3'
# vs its complement (default in case of one hypothesis)
#
# GORIC (using goric function in restriktor package)
library(restriktor)
set.seed(123) # Set seed value
goric.PandG_1c <- goric(fit.PandG,
                        hypotheses = list(H1))
    
```


H_1 vs H_c

Palmer & Gough

H_1 : $\mu_1 > \mu_2 > \mu_3$,
 H_c : not H_1 ,
 complement of H_1 .

GORIC

Model	Fit	Complexity	GORIC	GORIC weights
H1	-191.89	2.81	389.41	0.79
complement	-192.34	3.69	392.05	0.21

H_1 is $0.79 / 0.21 \approx 3.73$ times more supported than its complement,
 that is, any other hypothesis (i.e., any other ordering).

Note on use of safeguard hypotheses (2/2)

My advise and default in goric function:

- If multiple informative, theory-based hypotheses:
Goal: Comparison of informative, theory-based hypotheses.
Safeguard: Unconstrained.
- If one informative, theory-based hypothesis (H_m):
Goal: Compare H_m with all other possibilities / theories.
Safeguard: Its complement.

Extra: complement & 'Heq' (3/4)

Palmer & Gough

R output (partly)

```
## restriktor (0.6-15): generalized order-restricted information criterion:
##
## Results:
##      model      loglik  penalty   goric  loglik.weights  penalty.weights
## 1      Heq   -193.703    3.000  393.406         0.123         0.452
## 2      H1   -191.893    3.500  390.786         0.753         0.274
## 3 complement -193.703    3.500  394.406         0.123         0.274
##      goric.weights  goric.weights_without_heq
## 1          0.188
## 2          0.698          0.859
## 3          0.114          0.141
##
## Conclusion:
## - The order-restricted hypothesis H1 is the best in the set, as it has the h
## * 'H1' is 6.109 times more supported than 'complement'.
```


Table of Contents

Generalized Order-Restricted Information Criterion (GORIC)

GORIC weights

Examples with R code

GORICA

GORIC(A) in JASP

End & Extra

GORICA

Similarities with GORIC

- Form: $GORICA_m = -2 \text{ fit} + 2 \text{ complexity}$.
- Broad type of restrictions.

Differences compared to GORIC

- Uses asymptotic expression of the likelihood (is a normal):
can therefore be easily applied to all types of statistical models.
Disadvantage: might work less well in case of small samples.
- Does not need data set; mle's and their covariance matrix suffice.
- Can leave out nuisance parameters (i.e., not part of hypotheses).

Note

In case of normal linear models and/or not too small samples:
GORICA weights = GORIC weights.

R code: GORICA

Palmer & Gough

$H_1 : \mu_1 > \mu_2 > \mu_3,$

$H_c : \text{not } H_1.$

GORICA: type = "gorica" (not default for lm objects)

```

H1 <- 'group1 > group2 > group3'
# vs its complement (default in case of one hypothesis)
#
# GORICA (using goric function in restriktor package)
library(restriktor)
set.seed(123) # Set seed value
gorica.PandG_C <- goric(fit.PandG,
                        hypotheses = list(H1),
                        type = "gorica") # needed if lm object
    
```

GORICA

Palmer & Gough

$$H_1 : \mu_1 > \mu_2 > \mu_3,$$

$$H_c : \text{not } H_1.$$

GORIC

Model	Fit	Complexity	GORICA	GORICA weights
H1	-1.96	1.81	7.55	0.79
complement	-2.39	2.69	10.15	0.21

H_1 is $0.79 / 0.21 \approx 3.67$ times more supported than its complement, that is, any other hypothesis.

Note: GORIC weights are the same (ratio may differ a bit).

Hands-on/Demo (3): GORICA

Let's practice.

- If needed: Start Rstudio again (optional: make project) and then also load packages again.
- Optional: Open 'Tutorial_GORIC_restriktor_ANOVA.html', 'PalmerAndGough_and_Lucas.R', and/or '**Hands-on_3_GORICA_UncAndCompl_ANOVA_restriktor.R**'.
- Read and inspect data.
Use Data_PalmerAndGough.txt and/or **Data_Lucas.txt**.
- Run model (lm()).
- Specify hypotheses (make up your own). E.g., for Lucas:

$$H_1 : \mu_5 = \mu_3 > \{\mu_1, \mu_4\} > \mu_2$$

Note: Use names used in the model (or overwrite those).

- Run goric(); now, add (since lm object):
type = "gorica"
- Inspect and interpret output.

Table of Contents

Generalized Order-Restricted Information Criterion (GORIC)

GORIC weights

Examples with R code

GORICA

GORIC(A) in JASP

End & Extra

GORIC in JASP: screenshot

Palmer & Gough

Descriptives

T-Tests

ANOVA

Mixed Models

Regression

Frequencies

Factor

Bain

SEM

R (Beta)

▼ Order Restricted Hypotheses

Enter each restriction of one hypothesis on a new line, e.g.,
`factorLow == factorMid`
`factorMid < factorHigh`
 where 'factor' is the factor (or covariate) name and 'Low'/'Mid'/'High' are the factor level names.
 Click on the 'plus' icon to add more hypotheses.
 Click the information icon for more examples.

Syntax settings

☐ Include intercept
☐ Show available coefficients

Set for all models

☐ Model summary
☐ Marginal means
☐ Informed hypothesis tests

Model 1

+

group1 > group2 > group3

Ctrl + Enter to apply. Click on the blue button above for help on the restriction syntax

☐ Summary for Model 1
☐ Marginal means for Model 1
☐ Informed hypothesis tests for Model 1

Results

ANOVA

ANOVA - Importance

Cases	Sum of Squares	df	Mean Square	F	p
group	123.960	2	61.960	4.554	0.014
Residuals	925.418	68	13.609		

Note. Type III Sum of Squares

Order Restricted Hypotheses

Model Comparison

Model Comparison Table

Model	Log-likelihood	Penalty	GORIC	Weight	Weights ratio
Model 1	-191.893	2.814	389.415	0.789	3.731
Complement	-192.338	3.686	392.048	0.211	1.000

Note. Weights ratios indicate the relative weight for each model against the "complement" model. GORIC = Generalized Order-Restricted Information Criterion (Kuiper, Hoijtink, & Silvapulle, 2011).

GORIC and GORICA in JASP

GORIC or GORICA in the following statistical models (as part of that model)

The screenshot displays the JASP software interface with three statistical models: ANOVA, ANCOVA, and Repeated Measures ANOVA. In each model's settings panel, the 'Order Restricted Hypotheses' option is highlighted with a red circle. The ANOVA and ANCOVA panels also show the 'GORIC' label. The Repeated Measures ANOVA panel shows the 'GORICA' label. The right-hand side of the interface shows the 'ANCOVA' model selected, with a list of variables on the left and a list of factors on the right. The 'Display' section on the right includes options for 'Descriptive statistics', 'Estimates of effect size', 'Work-Selfie maximum p-ratio', and 'Model'. The 'Order Restricted Hypotheses' option is also highlighted with a red circle in the 'Display' section.

ANOVA

ANOVA

Cases	Sum of Squares	df	Mean Square	F	p
Note: Type III Sum of Squares					

ANCOVA

ANCOVA

Cases	Sum of Squares	df	Mean Square	F	p
Note: Type III Sum of Squares					

Repeated Measures ANOVA

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	p
Note: Type III Sum of Squares					

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	p
Note: Type III Sum of Squares					

The End
GORIC(A)

Thanks for listening!

Are there any questions?

Websites

<https://github.com/rebeccakuiper/Tutorials>
www.uu.nl/staff/RMKuiper/Extra1
www.uu.nl/staff/RMKuiper/Extra2
informative-hypotheses.sites.uu.nl/software/goric/

E-mail

r.m.kuiper@uu.nl

Complement not always higher weight

1. H_m does not have maximum fit (but much lower penalty).
2. Support for H_m is less when evaluating it against its complement (than H_u).

Explanation:

Against H_μ , you choose H_m 'sooner' because of low penalty for H_m .

1. Only include hypotheses with sound theoretical and/or empirical basis.
Often a null hypothesis is not of interest.
2. Keep the number of hypotheses included as small as possible.
3. This is a subjective endeavor, aim for inter-peer / inter-subjective agreement.

Note on possibilities multiple studies

- Update GORIC(A) values & weights.
More data collected: (re-)calculate.
- Update hypotheses.
First data set (or a part of it) generates one or more hypotheses.
Other data set (or part) used to determine evidence / support.
Download 'Tutorial_GORIC_restriktor_UpdateHypo.html' and/or
'Hands-on_4_GORIC_UpdateHypo_restriktor.R' from
<https://github.com/rebeccakuiper/Tutorials>.
- Aggregate evidence for hypotheses.
Aggregate the support for theories (diverse designs allowed).
Bear in mind: Meta-analysis aggregates parameter estimates or effect sizes which need to be comparable (often same designs required).

Download 'Tutorial_GORIC_restriktor_evSyn.html' and/or
'Hands-on_4_GORIC_evSyn_restriktor.R' from
<https://github.com/rebeccakuiper/Tutorials>.

Extra material (1/2)

- Logistic Regression Modeling
 - Article: doi.org/10.1037/met0000406
- GORICA on SEM
 - Article: www.tandfonline.com/doi/full/10.1080/10705511.2020.1836967.
 - R scripts: github.com/rebeccakuiper/GORICA_in_SEM.
- GORICA on cross-lagged panel model (CLPM)
 - Article: doi.org/10.1111/bjep.12455.
 - R scripts: github.com/rebeccakuiper/GORICA_in_CLPM.
- GORICA on Random-Intercept CLPM (RI-CLPM)
 - Article: Sukpan, C. and Kuiper, R.M. (2023). How to evaluate causal dominance hypotheses in lagged effects models.
 - R scripts: github.com/rebeccakuiper/GORICA%20in%20RI-CLPM.
- GORICA for EffectLiteR
 - R scripts: <https://github.com/rebeccakuiper/Tutorials/tree/main/GORICA%20for%20EffectLiteR>.

Note: On github site, go to Code (green button) and download zip.

- Note:** On github site, go to Code (green button) and download zip.

Note on GORIC weights vs BF and PMPs

ratio GORIC weights ($w_m/w_{m'}$) \sim Bayes factor ($BF_{mm'}$).

GORIC weight (w_m) \sim posterior model probability (PMP).

$1 - w_m$ = conditional error probability.

Like PMP, w_m depends on set of hypotheses.

Note on conditional error probability (1/3)

using PMPs and GORIC(A) weights

H_m	weights
H_1 : Sex Match	.04
H_2 : Gender Role Match	.81
H_3 : Sex Mismatch	.01
H_4 : Gender Role Mismatch	.00
H_u :	.14

The conditional error probability for preferred hypotheses H_2 is $1 - .81 = .19$.

Note on conditional error probability (2/3)

using PMPs and GORIC(A) weights

What if we compare many hypotheses?

H_m	weights
H_1 : Sex Match	.013
H_2 : Gender Role Match	.270
H_3 : Sex Mismatch	.003
H_4 : Gender Role Mismatch	.000
H_5 : Lets try this one too	.180
...	...
H_{12} : Don't miss something	.040
H_u :	.047

The conditional error probability for H_2 becomes $1 - .27 = .77$ (was .19).

Namely, included additional hypotheses may obtain some support. Hence, the conditional error probabilities becomes larger.

Note on conditional error probability (3/3)

using PMPS and GORIC(A) weights

What does conditional error probability reflect when hypotheses overlap? Notably, when hypotheses overlap they can also share/divide support (like H_u does with any hypothesis).

My advise:

Only use error probabilities if one hypothesis versus its complement.
(Or when you are sure that there is no overlap in hypotheses;
and they preferably do cover all possibilities)

H_m	weights
H_2 : Gender Role Match	.82
H_c : complement H_2	.18

Population Values in Simulated Data Sets when $k = 3$

Values of Population Means (μ_i) and Standard Deviations (σ_i) Used to Simulate 1000 Data Sets

$k = 3$	Case Number ('ES')	μ_1	μ_2	μ_3	ES (Cohen)
	0	0.00	0.00	0.00	
	1	-0.10	0.00	0.10	.1
	2	-0.20	0.00	0.20	.2
	3	-0.30	0.00	0.30	.3
	4	-0.40	0.00	0.40	.4
	5	-0.50	0.00	0.50	.5
		σ_1	σ_2	σ_3	
	0 - 5	1.00	1.00	1.00	

NB in Case 0 $H_{30C} : \mu_1 = \mu_2 = \mu_3$ is true
in Cases 1 to 5 $H_{31C} : \mu_1 < \mu_2 < \mu_3$ is true.

Results Confirmation ($k = 3$ and $n = 50$)

$H_{30C} : \mu_1 = \mu_2 = \mu_3$

$H_{31C} : \mu_1 < \mu_2 < \mu_3$

$H_{32C} : \mu_1 = \mu_2 < \mu_3$

$H_{33C} : \mu_1 < \mu_2 > \mu_3$

$H_{34C} : \mu_1, \mu_2, \mu_3$

'ES'	Method	'Prior'	H_{30C}	H_{31C}	H_{32C}	H_{33C}	H_{34C}
0	ORIC		0.662	0.065	0.119	0.120	0.034
0	BMS	3	0.865	0.012	0.075	0.036	0.012
0	BMS	2	0.774	0.022	0.124	0.058	0.022
0	BMS	1	0.656	0.047	0.147	0.120	0.033
2	ORIC		0.148	0.506	0.311	0.029	0.006
2	BMS	3	0.299	0.299	0.371	0.031	0.001
2	BMS	2	0.252	0.328	0.381	0.040	0.000
2	BMS	1	0.125	0.494	0.321	0.063	0.000
5	ORIC		0.000	0.955	0.044	0.001	0.000
5	BMS	3	0.000	0.887	0.112	0.001	0.000
5	BMS	2	0.001	0.890	0.107	0.002	0.000
5	BMS	1	0.000	0.928	0.070	0.002	0.000

Results Confirmation ($k = 3$ and $n = 50$)

$H_{30C} : \mu_1 = \mu_2 = \mu_3$

$H_{31C} : \mu_1 < \mu_2 < \mu_3$

$H_{32C} : \mu_1 = \mu_2 < \mu_3$

$H_{33C} : \mu_1 < \mu_2 > \mu_3$

$H_{34C} : \mu_1, \mu_2, \mu_3$

1 – Type I error (here):

preferring the true null hypothesis,
given the set of hypotheses (!).

'ES'	Method	'Prior'	H_{30C}	H_{31C}	H_{32C}	H_{33C}	H_{34C}
0	ORIC		0.662	0.065	0.119	0.120	0.034
0	BMS	3	0.865	0.012	0.075	0.036	0.012
0	BMS	2	0.774	0.022	0.124	0.058	0.022
0	BMS	1	0.656	0.047	0.147	0.120	0.033
2	ORIC		0.148	0.506	0.311	0.029	0.006
2	BMS	3	0.299	0.299	0.371	0.031	0.001
2	BMS	2	0.252	0.328	0.381	0.040	0.000
2	BMS	1	0.125	0.494	0.321	0.063	0.000
5	ORIC		0.000	0.955	0.044	0.001	0.000
5	BMS	3	0.000	0.887	0.112	0.001	0.000
5	BMS	2	0.001	0.890	0.107	0.002	0.000
5	BMS	1	0.000	0.928	0.070	0.002	0.000

Results Confirmation ($k = 3$ and $n = 50$)

$$H_{30C} : \mu_1 = \mu_2 = \mu_3$$

$$H_{31C} : \mu_1 < \mu_2 < \mu_3$$

$$H_{32C} : \mu_1 = \mu_2 < \mu_3$$

$$H_{33C} : \mu_1 < \mu_2 > \mu_3$$

$$H_{34C} : \mu_1, \mu_2, \mu_3$$

1 – Type II error (here):
preferring the true hypothesis,
given the set of hypotheses (!).

'ES'	Method	'Prior'	H_{30C}	H_{31C}	H_{32C}	H_{33C}	H_{34C}
0	ORIC		0.662	0.065	0.119	0.120	0.034
0	BMS	3	0.865	0.012	0.075	0.036	0.012
0	BMS	2	0.774	0.022	0.124	0.058	0.022
0	BMS	1	0.656	0.047	0.147	0.120	0.033
2	ORIC		0.148	0.506	0.311	0.029	0.006
2	BMS	3	0.299	0.299	0.371	0.031	0.001
2	BMS	2	0.252	0.328	0.381	0.040	0.000
2	BMS	1	0.125	0.494	0.321	0.063	0.000
5	ORIC		0.000	0.955	0.044	0.001	0.000
5	BMS	3	0.000	0.887	0.112	0.001	0.000
5	BMS	2	0.001	0.890	0.107	0.002	0.000
5	BMS	1	0.000	0.928	0.070	0.002	0.000

