Theory-based hypothesis evaluation using information criteria

for one and multiple studies

R. M. Kuiper

Department of Methodology & Statistics
Utrecht University

Table of Contents

Confirmatory methods

Most researchers are able to specify "order-restricted" / "informative" / "theory-based" hypotheses.

Use prior knowledge and/or expertise in hypothesis.

Confirmation

Compare only prespecified hypotheses
including order restrictions ($<$, $>$, but also $=$).
Limited set.

ANOVA Example: Comparisons of 3 Means

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".

Confirmatory methods

Methods to evaluate theory-based hypotheses

- Hypothesis testing: Fbar ($\bar{F}$) test
  (renders p-value and can test only one theory-based hypothesis)
- (Confirmatory) Bayesian model selection (BMS)
- Confirmatory model selection using information criteria: GORIC and GORICA

Note: 'model' refers to hypothesis.

Confirmation more power than exploration.

- Kuiper, R. M., and Hoijtink, H. (2010). Comparisons of Means Using Exploratory and Confirmatory Approaches. *Psychological Methods, 15(1)*, 69–86.
- Kuiper, R.M., Nederhof, T., and Klugkist, I. (2015). Properties of hypothesis testing techniques and (Bayesian) model selection for exploration-based and theory-based (order-restricted) hypotheses. *British Journal of Mathematical and Statistical Psychology, 68(2)*, 220 – 245.

Table of Contents

Information criteria (ICs)

IC, like GORIC, balances fit and complexity.

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

GORIC: Lowest value is best

GORIC is like AIC expected distance from the truth (KL-distance).
Hence, smallest value is best.

Example GORIC value

Palmer & Gough (2007) Data

$$
\begin{aligned}
H_1 &: \quad \mu_1 > \mu_2 > \mu_3, \\
H_2 &: \quad \mu_2 > \mu_1 > \mu_3, \\
H_u &: \quad \mu_1,\ \mu_2,\ \mu_3.
\end{aligned}
$$

GORIC

Model	Fit	Complexity	GORIC
H_1	-191.89	2.81	389.41
H_2	-193.70	2.81	393.03
H_u	-191.89	4.00	391.79

H_1 has lowest GORIC:
has best balance between fit and complexity
(and is not weak).

Table of Contents

Interpretation: GORIC weights

GORIC values

GORIC values cannot be interpreted, only compared:
Smallest is best.

GORIC weights (w_m) and ratios ($w_m/w_{m'}$)

w_m quantifies how much H_m is more supported than others in set.
$w_m/w_{m'}$ quantifies relative support of H_m vs $H_{m'}$.
The bigger, the better.

Reference:

Kuiper, R.M., Hoijtink, H. and Silvapulle, M.J. (2012). Generalization of the order restricted information criterion for multivariate normal linear models. *Journal of Statistical Planning and Inference, 142*, 2454-2463.

Example GORIC weights (w_m)

Palmer & Gough (2007) Data

$$
\begin{aligned}
H_1 &: \quad \mu_1 > \mu_2 > \mu_3, \\
H_2 &: \quad \mu_2 > \mu_1 > \mu_3, \\
H_u &: \quad \mu_1,\ \mu_2,\ \mu_3.
\end{aligned}
$$

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 $.68/.11 \approx 6.1$ times more supported than competing hypothesis H_2.

Table of Contents

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 (highest fit but also most complex).
E.g.: $w_1 = .08$, $w_2 = .02$, and $w_u = .90$.

If at least one informative hypothesis not weak ($w_1 > w_u$ or $w_1/w_u > 1$), then (and only then) compare informative hypotheses.

Example Unconstrained

Palmer & Gough (2007) Data

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

If at least one informative hypothesis not weak ($w_1 > w_u$ or $w_1/w_u > 1$), then compare informative hypotheses.

Hence: H_u is only a failsafe not another hypothesis of interest.

Download 'Guidelines_output_GORIC.html' from `https://github.com/rebeccakuiper/Tutorials`.
On github site, go to Code (green button) and download zip.

Alternative safeguard: Complement of H_m

Alternatively (in case of one hypothesis of interest)

Evaluate hypothesis of interest against its complement;
that is, all other possible hypotheses.

More powerful than against the unconstrained
if H_m has maximum fit.

Reference:
Vanbrabant, L., Van Loey, N., and Kuiper, R. M. (2020). Evaluating a Theory-Based Hypothesis Against Its Complement Using an AIC-Type Information Criterion With an Application to Facial Burn Injury. Psychological Methods, 25(2), 129-142.
https://doi.org/10.1037/met0000238

Example H_1 vs H_c

Palmer & Gough (2007) Data

$$
\begin{aligned}
H_1 &: \quad \mu_1 > \mu_2 > \mu_3, \\
H_c &: \quad \text{not } H_1.
\end{aligned}
$$

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 $.79/.21 \approx 3.8$ times more supported than its complement, that is, any other hypothesis.

Table of Contents

GORICA

Similarities with GORIC

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

GORICA: All statistical models

Therefore, GORICA: asymptotic expression for GORIC.
Can be used for all types of statistical models.

Reference:
Altınışık, Y., Van Lissa, C. J., Hoijtink, H., Oldehinkel, A. J., and Kuiper, R. M. (2021). Evaluation of inequality constrained hypotheses using a generalization of the AIC. *Psychological Methods, 26(5)*, 599–621.
`https://doi.org/10.1037/met0000406`

Example GORICA

Palmer & Gough (2007) Data

$$H_1: \quad \mu_1 > \mu_2 > \mu_3,$$
$$H_c: \quad \text{not } H_1.$$

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 $.79/.21 \approx 3.8$ times more supported than its complement, that is, any other hypothesis.

Note: GORIC weights are the same.

Some literature

- Logistic Regression Modeling
  - Article: https://doi.org/10.1037/met0000406
- GORICA on SEM
  - Article: https://www.tandfonline.com/doi/full/10.1080/10705511.2020.1836967.
  - R scripts: https://github.com/rebeccakuiper/GORICA_in_SEM.
- GORICA on cross-lagged panel model (CLPM) – Article: https://doi.org/10.1111/bjep.12455.
  - R scripts: https://github.com/rebeccakuiper/GORICA_in_SEM.
- GORICA on Random-Intercept CLPM (RI-CLPM)
  - Article: Chuenjai Sukpan and Rebecca M. Kuiper (submitted 2023). How to evaluate causal dominance hypotheses in lagged effects models.
  - R scripts: https://github.com/Chuenjai/Causal-dominance.
- GORICA on CTmeta
  - Article: https://doi.org/10.1080/10705511.2020.1823228.
  - R scripts: https://github.com/rebeccakuiper/GORICA_on_CTmeta.
- GORICA on Meta-analysis
  - Article: https://doi.org/10.3390/e24111525.
  - R scripts: https://github.com/rebeccakuiper/GORICA_on_MetaAn.

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

Table of Contents

Possibilities multiple studies (1/2)

- 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.

  – html tutorial: https://github.com/rebeccakuiper/Tutorials/blob/main/Tutorial_GORIC_restriktor_UpdateHypo.html
  – R script tutorial: https://github.com/rebeccakuiper/Tutorials/blob/main/Hands-on%20files/Hands-on_4_GORIC_UpdateHypo_restriktor.R

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

Possibilities multiple studies (2/2)

- 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).

  – html tutorial: https://github.com/rebeccakuiper/Tutorials/blob/main/Tutorial_GORIC_restriktor_AggrSupport.html
  – R script tutorial: https://github.com/rebeccakuiper/Tutorials/blob/main/Hands-on%20files/Hands-on_5_GORICA_CombEv_restriktor.R

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

GORIC(A) for Multiple Studies: Aggregating support (= evidence synthesis)

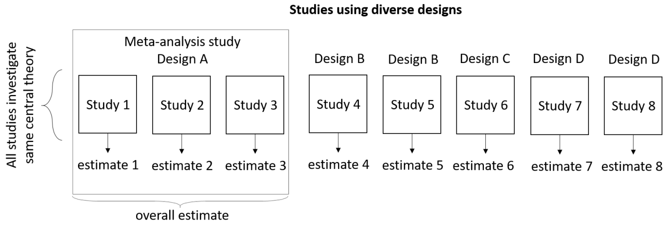
References:

– Kuiper, R.M., Buskens, V.W., Raub, W., and Hoijtink, H. (2013). Combining statistical evidence from several studies: A method using Bayesian updating and an example from research on trust problems in social and economic exchange. *Sociological Methods and Research*, 42 (1), (pp. 60-81) (22 p.).

– Van Lissa, C.J., Clapper, E.-B., and Kuiper, R.M. (submitted 2023). Aggregating evidence from conceptual replication studies using the product Bayes factor. 10.31234/osf.io/nvqpw

– Kuiper, R.M., and Clapper, E.-B. (to be submitted in 2023). GORIC Evidence Aggregation: Combining Statistical Evidence for a Central Theory from Diverse Studies using an AIC-type Criterion. 10.31234/osf.io/qv76x

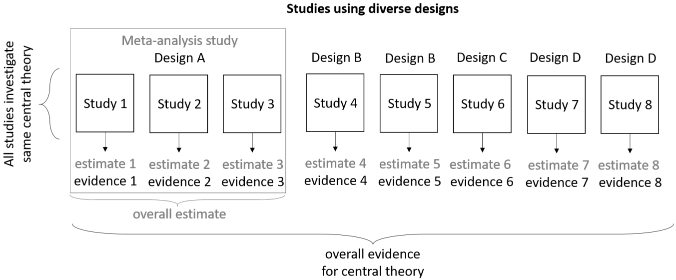
Current best practice



Current best practice is meta-analysis and Bayesian updating.

- Not applicable for diverse research designs.
- Not applicable for incomparable estimates.

Need for new methodology: Evidence Synthesis



Note: All studies do investigate the same, central theory (using diverse designs).

Trust Example: Meta-Analysis versus Evidence Synthesis

Study	Type of model
1	univariate regression
2	univariate regression
3	probit regression
4	three-level logistic regression
	Same design? e.g., same set of predictors?

Conceptual replications!

	Meta-Analysis	Evidence Synthesis
Effect size not required		✓
Deal with diverse designs		✓
Main results	Estimate of effect size	Evidence for hypotheses
Check:		same theoretical relationships?

Reference:

Kuiper, R.M., Buskens, V.W., Raub, W., and Hoijtink, H. (2013). Combining statistical evidence from several studies: A method using Bayesian updating and an example from research on trust problems in social and economic exchange. *Sociological Methods and Research*, 42 (1), (pp. 60-81) (22 p.).

Trust Example: Meta-Analysis versus Evidence Synthesis

Study	Type of model
1	univariate regression
2	univariate regression
3	probit regression
4	three-level logistic regression
	Same design? e.g., same set of predictors?

Conceptual replications!

	Meta-Analysis	Evidence Synthesis
Effect size not required		✓
Deal with diverse designs		✓
Main results	Estimate of effect size	Evidence for hypotheses
Check:		same theoretical relationships?

Reference:
Kuiper, R.M., Buskens, V.W., Raub, W., and Hoijtink, H. (2013). Combining statistical evidence from several studies: A method using Bayesian updating and an example from research on trust problems in social and economic exchange. *Sociological Methods and Research*, 42 (1), (pp. 60-81) (22 p.).

Trust Example: Meta-Analysis versus Evidence Synthesis

Study	Type of model
1	univariate regression
2	univariate regression
3	probit regression
4	three-level logistic regression
	Same design? e.g., same set of predictors?

Conceptual replications!

	Meta-Analysis	Evidence Synthesis
Effect size not required		✓
Deal with diverse designs		✓
Main results	Estimate of effect size	Evidence for hypotheses
Check:		same theoretical relationships?

Reference:

Kuiper, R.M., Buskens, V.W., Raub, W., and Hoijtink, H. (2013). Combining statistical evidence from several studies: A method using Bayesian updating and an example from research on trust problems in social and economic exchange. *Sociological Methods and Research*, 42 (1), (pp. 60-81) (22 p.).

Trust Example: Meta-Analysis versus Evidence Synthesis

Study	Type of model
1	univariate regression
2	univariate regression
3	probit regression
4	three-level logistic regression
	Same design? e.g., same set of predictors?

Conceptual replications!

	Meta-Analysis	Evidence Synthesis
Effect size not required		✓
Deal with diverse designs		✓
Main results	Estimate of effect size	Evidence for hypotheses
Check:		**same theoretical relationships?**

Reference:
Kuiper, R.M., Buskens, V.W., Raub, W., and Hoijtink, H. (2013). Combining statistical evidence from several studies: A method using Bayesian updating and an example from research on trust problems in social and economic exchange. *Sociological Methods and Research*, 42 (1), (pp. 60-81) (22 p.).

Example: 4 studies regarding one concept

Study	Type of study	Number of observations n	Type of model
1	survey	895 transactions	univariate regression
2	experiment	348 decisions by 40 subjects	univariate regression
3	experiment	1249 decisions by 125 subjects	probit regression
4	experiment	2160 decisions by 144 subjects	three-level logistic regression
Study	**Outcome y (trust)**		**scale y**
1	effort invested in management		ratio
2	effort invested in management		ratio
3	choice of vignettes		dummy
4	trustfulness		dummy
Study	**Predictor x_1 (past / previous experience)**		**scale x_1**
1	existence relationship with supplier		dummy
2	type of relationship with supplier		interval
3	bought a car from The Autoshop before		dummy
4	number of times a trustee honored trust in the past		ratio
Study	**some of the other predictors**		
1	transaction characteristics, expected future transactions, network embeddedness		
2	transaction characteristics, expected future transactions, network embeddedness		
3	expected future transactions, network embeddedness		
4	future interactions, network embeddedness		

One-Parameter Example: Hypotheses of interest

Main central theory

Previous experience has a positive effect on trust.
For simplicity, only one relationship here, could have been more.

Study-specific hypothesis

$\beta_1 > 0$
Here, for each study the same hypothesis..

Set of central theories

$$H_0 : \textit{no effect},$$
$$H_> : \textit{positive effect},$$
$$H_< : \textit{negative effect}.$$

Note 1: Central hypotheses for all studies, not w.r.t. average parameter.
Note 2: In practice, I would not include H_0...

One-Parameter Example: Hypotheses of interest

Main central theory

Previous experience has a positive effect on trust.
For simplicity, only one relationship here, could have been more.

Study-specific hypothesis

$\beta_1 > 0$
Here, for each study the same hypothesis..

Set of central theories

$$H_0 : \textit{no effect},$$
$$H_> : \textit{positive effect},$$
$$H_< : \textit{negative effect}.$$

Note 1: Central hypotheses for all studies, not w.r.t. average parameter.
Note 2: In practice, I would not include H_0...

One-Parameter Example: Hypotheses of interest

Main central theory

Previous experience has a positive effect on trust.
For simplicity, only one relationship here, could have been more.

Study-specific hypothesis

$\beta_1 > 0$
Here, for each study the same hypothesis..

Set of central theories

$$H_0 : \textit{no effect},$$
$$H_> : \textit{positive effect},$$
$$H_< : \textit{negative effect}.$$

Note 1: Central hypotheses for all studies, not w.r.t. average parameter.
Note 2: In practice, I would not include H_0...

Example: Trust (y) & previous experience (x_1)

Not full data set (and probit regression), so use

- GORICA (not GORIC) using *goric* function in R package *restriktor*

Input:

- parameter estimates and their covariance matrix

t	$\hat{\beta}_1$	$\hat{\sigma}_{\beta_1}$
1	0.090	0.029
2	0.140	0.054
3	1.090	0.093
4	1.781	0.179

Note: Here, one parameter (β_1); thus, cov. matrix $\hat{\beta}_1$ = variance $\hat{\beta}_1 = \hat{\sigma}^2_{\beta_1}$ (not $\hat{\sigma}_{\beta_1}$)

One-Parameter Example: results per study

using GORICA

Results per study (not aggregated yet)!

Table: GORICA weights ($w_{t,m}$) for Hypothesis H_m in Study t

	$w_{t,m}$			
m / t	1	2	3	4
0	0.013	0.052	0.000	0.000
$>$	**0.979**	**0.916**	**1.000**	**1.000**
$<$	0.008	0.032	0.000	0.000

Note: Weight is at max 1.

So, now on forehand already clear.... but no quantification yet.

One-Parameter Example: Results & Conclusions

using GORICA

Table: Overall GORICA weights ($w^1_{t,m}$) for Hypothesis H_m in Study t

	$w^1_{t,m}$			
m / t	1	2	3	4
0	0.013	0.001	0.000	0.000
$>$	**0.979**	**0.999**	**1.000**	**1.000**
$<$	0.008	0.000	0.000	0.000

- $w^1_{4,>} = 1$ => full support for $H_>$
  $w^1_{4,0} = w^1_{4,<} = 0$ => no support for H_0 and $H_<$
- Support for $H_>$ ($w^1_{4,1}$) is highest: favor $H_>$ over H_0 and $H_<$.

One-Parameter Example: Results & Conclusions

using GORICA

Table: Overall GORICA weights ($w^1_{t,m}$) for Hypothesis H_m in Study t

	$w^1_{t,m}$			
m / t	1	2	3	4
0	0.013	0.001	0.000	0.000
$>$	**0.979**	**0.999**	**1.000**	**1.000**
$<$	0.008	0.000	0.000	0.000

- $w^1_{4,>} = 1$ => full support for $H_>$
  $w^1_{4,0} = w^1_{4,<} = 0$ => no support for H_0 and $H_<$
- Support for $H_>$ ($w^1_{4,1}$) is highest: favor $H_>$ over H_0 and $H_<$.

One-Parameter Example: Results & Conclusions
using GORICA

Table: Overall GORICA weights ($w^1_{t,m}$) for Hypothesis H_m in Study t

	$w^1_{t,m}$			
m / t	1	2	3	4
0	0.013	0.001	0.000	0.000
$>$	**0.979**	**0.999**	**1.000**	**1.000**
$<$	0.008	0.000	0.000	0.000

- $w^1_{4,>} = 1$ => full support for $H_>$
  $w^1_{4,0} = w^1_{4,<} = 0$ => no support for H_0 and $H_<$
- Support for $H_>$ ($w^1_{4,1}$) is highest: favor $H_>$ over H_0 and $H_<$.

One-Parameter Example: Results & Conclusions

using GORICA

Table: Overall GORICA weights ($w^1_{t,m}$) for Hypothesis H_m in Study t

	$w^1_{t,m}$			
m / t	1	2	3	4
0	0.013	0.001	0.000	0.000
$>$	**0.979**	**0.999**	**1.000**	**1.000**
$<$	0.008	0.000	0.000	0.000

- $w^1_{4,>} = 1$ => full support for $H_>$
  $w^1_{4,0} = w^1_{4,<} = 0$ => no support for H_0 and $H_<$
- Support for $H_>$ ($w^1_{4,1}$) is highest: favor $H_>$ over H_0 and $H_<$.

One-Parameter Example: Results & Conclusions

using GORICA

Table: Overall GORICA weights ($w^1_{t,m}$) for Hypothesis H_m in Study t

	$w^1_{t,m}$			
m / t	1	2	3	4
0	0.013	0.001	0.000	0.000
$>$	**0.979**	**0.999**	**1.000**	**1.000**
$<$	0.008	0.000	0.000	0.000

- $w^1_{4,>} = 1$ => full support for $H_>$
  $w^1_{4,0} = w^1_{4,<} = 0$ => no support for H_0 and $H_<$
- Support for $H_>$ ($w^1_{4,1}$) is highest: favor $H_>$ over H_0 and $H_<$.

One-Parameter Example: Results & Conclusions

using GORICA

Table: Overall GORICA weights ($w^1_{t,m}$) for Hypothesis H_m in Study t

	$w^1_{t,m}$			
m / t	1	2	3	4
0	0.013	0.001	0.000	0.000
$>$	**0.979**	**0.999**	**1.000**	**1.000**
$<$	0.008	0.000	0.000	0.000

- $w^1_{4,>} = 1$ => full support for $H_>$
  $w^1_{4,0} = w^1_{4,<} = 0$ => no support for H_0 and $H_<$
- Support for $H_>$ ($w^1_{4,1}$) is highest: favor $H_>$ over H_0 and $H_<$.

Possible types of sets of studies

- Conceptual replications of same authors, done as a robustness check.
- Searching for direct and indirect/conceptual replications in the literature.
- Using multiple $N = 1$ studies.
- Using different cohorts, where one can measure the variables in their own way.
- Using different subpopulations, possibly using different operationalisations.
- ...

Possible type of sets of studies (1/5)

- Conceptual replications of same authors, done as a robustness check.

  This was done in the Trust example of Buskens and Raub.

  Note: There, the central hypotheses regard one parameter of interest (in each study), but one can compare (absolute values of) multiple parameters or multiple effect sizes.

  For some examples, download 'Tutorial_GORIC_restriktor_evSyn.html' from `https://github.com/rebeccakuiper/Tutorials`.

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

Possible type of sets of studies (2/5)

- Searching for direct and indirect/conceptual replications in the literature.

  E.g., using the central hypothesis that the absolute strength of the relationship of communication competence (C) with willingness to communicate in a second language (WTC) is greater than the absolute strength of the relation of communication anxiety (A) with WTC, which is greater than the absolute strength of the relation of motivation (M) with WTC, that is, $|C| > |A| > |M|$;
  where C, A, and M are operationalized differently in the studies.

  – 'Article': Example in bachelor thesis of Martijn Sips
  – R scripts: `https://github.com/rebeccakuiper/Tutorials/tree/main/Examples%20evSyn/Example%20WtC`

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

Possible type of sets of studies (3/5)

- Using multiple $N = 1$ studies.

  – Article: Klaassen et al. (2018).
  All for one or some for all? Evaluating informative hypotheses using multiple $N = 1$ studies. *Behavior Research Methods*. 50, 2276–2291.
  https://link.springer.com/article/10.3758/s13428-017-0992-5.

Possible type of sets of studies (4/5)

- Using different cohorts, where one can measure the variables in their own way.

  So, possibly using different operationalisation of the same constructs.

  – Article: Zondervan-Zwijnenburg et al. (2020).
  Parental Age and Offspring Childhood Mental Health: A Multi-Cohort, Population-Based Investigation. *Child Development*. 91(3), 964–982.

Possible type of sets of studies (5/5)

- Using different subpopulations, possibly using different operationalisations.

  E.g., the Municipal Health Services (Dutch acronym: GGD) studied the positive consequences of corona on loneliness (a), mental health (b), and stress (b); conditional on sex, age, and health.
  Central hypothesis: $a < b < c$.

  – Article: in progress
  – R scripts: `https://github.com/rebeccakuiper/Tutorials/tree/main/Examples%20evSyn/Example%20corona%20GGD`

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

Table of Contents

Software

- GORIC(A)
  - R function *goric* in R package *restriktor*
  - JASP

- GORIC(A) evidence synthesis
  - R function *evSyn* in R package *restriktor*
  - Interactive web application (Shiny app) available from my site (see below).

Websites

https://github.com/rebeccakuiper/Tutorials
www.uu.nl/staff/RMKuiper/Software
www.uu.nl/staff/RMKuiper/Websites%20%2F%20Shiny%20apps
informative-hypotheses.sites.uu.nl/software/goric/

Promo: Publish in special issue

Applying GORIC(A)?

If you collected data or will collect,
if you have one or more a-priori hypotheses,
then you can apply the GORIC(A).

FYI: Special issue

Evaluation of Theory-Driven Hypotheses: No Hypothesis, No G(L)ORIc
`https://www.mdpi.com/journal/mathematics/special_issues/97C26430R3`

Contact: r.m.kuiper@uu.nl

Please contact me (r.m.kuiper@uu.nl), if you want to explore the possibilities (and obtain a fee waiver).

The End

Thanks for listening!

Are there any questions?

Websites

https://github.com/rebeccakuiper/Tutorials
www.uu.nl/staff/RMKuiper/Software
www.uu.nl/staff/RMKuiper/Websites%20%2F%20Shiny%20apps
informative-hypotheses.sites.uu.nl/software/goric/

E-mail

r.m.kuiper@uu.nl

Generalized Order-Restricted Information Criterion

GORIC

'IC' = -2 fit + 2 complexity

Fit = Maximized order-restricted log likelihood

Maximized log likelihood based on parameters in agreement with H_m.

Complexity = Penalty

Represents: Expected number of distinct parameters.
Here, expected number of distinct mean values plus 1 (because of the unknown variance term).
Details: Function of level probabilities.

Idea complexity

loose interpretation

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

contains 1 ordering of three means, 1-2-3.
Thus, not complex (i.e., parsimonious).

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

contains 2 orderings of three means: 1-2-3 and 1-3-2.
Thus, more complex (less parsimonious).

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

contains all six possible orderings of three means.
Thus, is most complex one (least parsimonious).

Example GORIC

Palmer & Gough (2007) Data

$$
\begin{aligned}
H_0 &: \quad \mu_1 = \mu_2 = \mu_3, \\
H_1 &: \quad \mu_1 > \mu_2 > \mu_3, \\
H_2 &: \quad \mu_1 > \mu_2 < \mu_3, \\
H_3 &: \quad \mu_1 < \mu_2 < \mu_3, \\
H_u &: \quad \mu_1, \ \mu_2, \ \mu_3.
\end{aligned}
$$

GORIC

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

GORIC

$IC_m = -2\ fit_m + 2\ complexity_m$

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.

References:

- 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.
- Kuiper, R. M., Klugkist, I., and Hoijtink, H. (2010). A Fortran 90 Program for Confirmatory Analysis of Variance. *Journal of Statistical Software, 34(8)*, 1–31.

GORICA

GORIC: Normal linear models

GORIC can easily be applied to normal linear models
(e.g., ANOVA models or regression models).

GORIC: Other statistical models

In case of other statistical models (e.g., a SEM model), more cumbersome to calculate maximized order-restricted log likelihood and thus GORIC.

GORICA: All statistical models

Therefore, GORICA: asymptotic expression for GORIC.
Can be used for all types of statistical models.

Reference:
Altınışık, Y., Van Lissa, C. J., Hoijtink, H., Oldehinkel, A. J., and Kuiper, R. M. (2021). Evaluation of inequality constrained hypotheses using a generalization of the AIC. *Psychological Methods, 26(5)*, 599–621.
https://doi.org/10.1037/met0000406

GORICA

Similarities with GORIC

- Form: $GORICA_m = -2\ fit + 2\ complexity$.
- 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.