Guidelines interpretation GORIC(A) output

Rebecca M. Kuiper and Leonard Vanbrabant

22 August 2024

Contents

Goal GORIC(A)

The information criteria GORIC and GORICA - referred to as GORIC(A) - are AIC-type information criteria and can evaluate one or more informative, theory-based hypotheses. Hence, you do not have to specify a null hypothesis nor (only) equality restrictions. You can (also) compare the size and/or ordering of mean parameters or of (standardized) regression-type parameters. You could, for example, evaluate the hypothesis $H_m : \mu_1 > \mu_2 > \mu_3$ (a simple ordering of mean parameters) or $H_m : \beta_1 - \beta_2 > \beta_3 - \beta_4$ (a possible representation of an interaction effect in terms of regression parameters). The goal of the GORIC(A) is to select the best from a set of candidate hypotheses/models.

GORIC(A) output

The GORIC(A) is an information criterion and, therefore, balances fit and complexity. **Fit** denotes the compatibility of the hypothesis with the data, expressed by the *maximum log likelihood* part. **Complexity** reflects the size of the hypothesis in terms of (expected) number of parameters, expressed by the *penalty* part.

Stated otherwise, GORIC(A) selects the hypothesis that describes the data best with the fewest number of parameters, out of a set of candidate hypotheses.

GORIC(A) values

The hypothesis with the smallest GORIC(A) value is the preferred one in the set of candidate hypotheses; since that hypothesis then has the smallest (Kullback–Leibler) distance to the truth. The GORIC(A) values themselves are not interpretable; only the differences between the values can be inspected (i.e., the smaller, the better). To improve the interpretation, GORIC(A) weights can be computed.

GORIC(A) weights and ratios

A **GORIC(A) weight (w_m)** represent the relative likelihood of a hypothesis (H_m) given the data and the set of hypotheses. The hypothesis with the largest GORIC(A) weight is the preferred one in the set of candidate hypotheses. Additionally, the **ratio of two GORICA weights ($w_m/w_{m'}$)** can be used to quantify their relative support. This leads to conclusions like Hypothesis H_1 is w_1/w_2 more supported than the competing hypothesis H_2.

I am not in favor of cut-off points, but when you feel you need them to label the height/strength of the weights and/or their ratios, you can use benchmarks as proposed and discussed in 'Guidelines_GORIC-benchmarks' on https://github.com/rebeccakuiper/Tutorials.

Hypotheses sets

The set of hypotheses should consist of the hypotheses of interest (reflecting one or more theories from the literature or ones based on expertise), possibly with a failsafe hypothesis (to prevent from selecting a hypothesis which is the best of the worst). I will distinguish different types of set of hypotheses: The first distinction is based on the number of informative hypotheses: one or more; The second distinction is based on the choice of safeguard hypothesis. In the next section, I will describe these type of sets and give guidelines for interpretation the results from each of them. I will also remark on how to specify your hypothesis such that the GORIC(A) output helps you even more. Subsequently, I will discuss two special cases situations one should be aware of:

- The case of equal fit, which can happen in case of overlapping hypotheses. This is discussed together with finding support for the boundary of hypotheses;

- The case of just below maximum fit, which can occur when there is at least one equality restriction in (one of) the hypothesis under investigation.

Consequently, be aware of specifying overlapping hypotheses and hypotheses containing one or more equality restrictions.

For an overview of what type of hypotheses to include in the set and which failsafe hypothesis to use, see the decision trees in 'Decision trees - Choice hypotheses.pdf' on https://github.com/rebeccakuiper/Tutorials.

Interpretation output

General

In general, the GORIC(A) output is interpreted as follows:

- The hypothesis with the **highest GORIC(A) weight** is the preferred one: the best from the set.

- The GORIC(A) weight of that hypothesis denotes its relative strength in the set (given the data). The *GORIC(A) weights reflect the (un)certainty regarding that hypothesis being the best* - but be aware that it is conditional on the other hypotheses in the set.

- Additionally, one can compare the strength of that hypothesis versus a competing hypothesis using the ratio of their GORIC(A) weights.

As discussed later on, you need to inspect a bit more output to check for special cases (like support for overlap of the hypotheses of interest). Generally good advise - as will become clear later on - **before interpreting the ratios of GORIC(A) weights, one needs to check the (ratio of) fit/loglik weights of the preferred hypotheses with that of the other hypotheses in the set**. Additionally, you may want to take into account the sample size you have.

For an overview of what type of output to inspect and how to interpret, see the decision tree in 'Decision tree - Interpretation.pdf' on https://github.com/rebeccakuiper/Tutorials.

Next, I will go into more detail for each type of hypothesis set.

One informative hypothesis

Here, I assume you have one informative hypothesis, that is, there are no competing hypotheses of interest. Then, I would advise you to evaluate it with its complement (an option in the software), as opposed to the unconstrained hypothesis. Nevertheless, I will describe both situations next.

vs Complement

The complement of a hypothesis denotes all the other possible hypotheses (i.e., all the other possible orderings); so, excluding the one of interest. The complement acts like a competing hypothesis. Either your (informative) hypothesis or its complement is the best and the ratio of their GORIC(A) weights denotes the relative strength.

Say, your hypothesis of interest has a GORIC(A) weight of w_m and, thus, its complement of $w_c = 1 - w_m$. Then, your hypothesis has $w_m/(1 - w_m)$ more support than its complement. This relative support can go from (approximately) zero to infinity - the first denoting *infinite support for the complement* and the latter *infinite support your hypothesis*. When the ratio is close to 1, then *both hypotheses are (approximately) equally good* (in terms of balancing fit and complexity). Notably, the ratio of GORIC(A) weights together with other output can also denote support for their boundary/border, which is discussed in the Special cases Section.

When inspecting a hypothesis versus its complement, one can interpret $1 - w_m = w_c$ as a error probability. That is, the probability that H_m is not the best one is $w_c * 100\%$. Note that in case of multiple hypotheses, especially in case of overlapping hypotheses, I would not advise on using this (conditional) error probability interpretation, since w_m and thus $1 - w_m$ depends on the other hypotheses in the set.

```
# H1 vs complement
H1 <- "D1 > D2 > D3" # Denoting mu1 > mu2 > mu3

# Apply GORIC #
set.seed(123)
results_1c <- goric(fit, hypotheses = list(H1), comparison = "complement")
results_1c
```

Example

```
restriktor (0.5-85): generalized order-restricted information criterion:

Results:
        model      loglik  penalty     goric  loglik.weights  penalty.weights  goric.weights
1          H1    -123.384    2.832   252.431           0.679            0.698          0.830
2  complement    -124.133    3.668   255.601           0.321            0.302          0.170
---
The order-restricted hypothesis 'H1' has 4.88 times more support than its complement.
```

3

From this, you can conclude that $H_1 : \mu_1 > \mu_2 > \mu_3$ is 0.83 / 0.17 $\approx$ 4.88 time more supported than its complement.

Additionally, one could say that the probability that H_1 is not the best is 17%.

Population information:

In the data generation, I used a ratio of population means of 3:2:1; implying that H_1 is correct. More specifically, I used population mean values of approximately 0.98, 0.65, and 0.33. This implies that Cohen's d is approximately .27; thus, there is a small to medium population effect size (which are in the same order as hypothesized). I then sampled 100 observations, ran an ANOVA (with three groups), and applied the GORIC.

When I would sample more observations, the GORIC(A) weight for H_1 converges to 1 (denoting full support for H_1).

Note: As will become clear in the Special cases Section, one may want to be able to find support for the boundary of these two hypotheses. Then, one should first check the ratio of fit/loglik weights (i.e. ratio of loglik.weights). Since the loglik.weights differ (i.e., the ratio of loglik.weights is not close to 1), there is no support for the boundary of these hypotheses; only support for H_1.

vs Unconstrained (but not advised)

The unconstrained hypothesis denotes all the possible hypotheses (i.e., all the possible orderings); so, including the one of interest. Since it covers the hypothesis of interest, it cannot act like a competing hypothesis. The unconstrained can only be used as a failsafe:

- If the unconstrained is better (i.e., has a higher weight), then your hypothesis of interest is weak, that is, it is not supported by the data.

- If *your hypothesis* is better (i.e., *has a higher weight*), then it is **not weak**. In this case, the ratio of GORIC(A) weights is bounded (as described in the Example section below) and should, therefore, not be interpreted. Stated otherwise, in this case, the ratio of GORIC(A) weights will never go to infinity; as it would do in the case of using the complement. Here, a ratio of GORIC(A) weights equal to its bound denotes full support; even when the sample size is small.
  Therefore, I advise using the *complement* in case of one informative hypothesis; because, then, the resulting ratio of GORIC(A) weights is not bounded, and the *weights reflect the uncertainty there is* (in case of a small sample size).

```
# H1 vs unconstrained (default)
H1 <- "D1 > D2 > D3" # mu1 > mu2 > mu3

# Apply GORIC #
set.seed(123)
results_1u <- goric(fit, hypotheses = list(H1))
results_1u
```

Example

```
restriktor (0.5-85): generalized order-restricted information criterion:
```

Results:

	model	loglik	penalty	goric	loglik.weights	penalty.weights	goric.weights
1	H1	-123.384	2.832	252.431	0.500	0.763	0.763
2	unconstrained	-123.384	4.000	254.767	0.500	0.237	0.237

From this, you can conclude that $H_1 : \mu_1 > \mu_2 > \mu_3$ is more supported than the unconstrained (since 0.763 / 0.237 > 1) and is, therefore, not a weak hypothesis.

When you look at the fit/loglik value, you can see that they are the same for both hypotheses (also reflected by the loglik.weights; leading to a ratio of loglik.weights of 1). The unconstrained has, by definition, the maximum fit. Here, H_1 also has maximum fit, meaning that its restrictions are in agreement with the data. Since both have the same fit, their GORIC(A) weights solely depend on the penalty values and, thus, are bounded: they equal the penalty.weights (and thus not 1 and 0, what you otherwise would find in case of full support). Finding equal fit (or GORIC(A) weights being equal to the penalty weights) means that there is support for the overlap (see also the Special cases Section).

The informative hypothesis H_1 is completely contained in the unconstrained (by definition). Hence, the overlap of the two hypotheses is H_1. Therefore, we can say that H_1 has (full) support: it has the maximum fit and is (by definition) more parsimonious than the unconstrained. Note that, now, we cannot say anything regarding the uncertainty (which we could in the case of using the complement as the failsafe, as done above).

Population information:
In the data generation, I used a ratio of population means of 3:2:1; implying that H_1 is correct. More specifically, I used population mean values of approximately 0.98, 0.65, and 0.33. This implies that Cohen's d is approximately .27; thus, there is a small to medium population effect size (which are in the same order as hypothesized). I then sampled 100 observations, ran an ANOVA (with three groups), and applied the GORIC.

When I would sample more observations, the GORIC(A) weight for H_1 remains the same. When I would have sampled less observations, the fit/loglik values may differ. Then, the GORIC(A) weight of H_1 is less than the penalty weight of H_1.

Note: In case of one informative hypothesis of interest, I would advise on using the complement instead (as done in the previous section). Then, the support for a true hypothesis increases with sample size (and/or effect size). Moreover, you have information regarding the uncertainty of your decision: In the example above, one would conclude that H_1 has (full) support; while using the complement (in the previous example), one would conclude that H_1 is preferred and that there is an uncertainty of 17%, namely a 17% chance that H_1 is not the best.

Multiple informative hypotheses

Let us assume that the literature states two competing informative hypotheses. Then, a safeguard hypothesis is needed if these informative hypotheses do not cover the whole parameter space, that is, do not cover all possible theories/orderings. Namely, when both informative hypotheses are weak hypotheses, GORIC(A) selects the best out of a set of weak hypotheses (i.e., best of the worst). To refrain from this, one should then include a safeguard hypothesis.

Complement of set as failsafe

The complement of a set of hypotheses denotes all the other possible hypotheses (i.e., all the other possible orderings); so, excluding the ones of interest. The **complement acts like another competing hypothesis**. One can also compare its strength to that of the hypotheses of interest.

As always, the hypothesis with the highest GORIC(A) weight is the preferred one.

- If the best hypothesis is the complement of the set, then the hypotheses of interest are weak. One should not compare the strength of the hypotheses of interest (to avoid choosing the best from worst). Bear in mind that one can, if that would be of interest, compare the strength of the complement versus each of the hypotheses of interest by inspecting their ratio of GORIC(A) weights.
  Notably, one can take on an additional exploratory approach to create one or more new hypotheses for future research.

- If the best hypothesis is one of the hypotheses of interest, then it is not weak and can be compared to the other hypothesis/-es of interest (including the complement). Note that one can compare all the non-weak hypotheses to all the hypotheses of interest.
  Be aware of the special cases, as will be discussed in the Special cases Section.

Important: This option is not available in GORIC(A) software:

Bear in mind that **the relative support of two informative hypotheses is independent from the choice of failsafe hypothesis**. So, if the goal is to compare informative hypotheses, one could just as well use the unconstrained (as discussed in the next subsection).

Unconstrained as failsafe

The unconstrained hypothesis denotes all the possible hypotheses (i.e., all the possible orderings); so, including the ones of interest. Since it covers the hypotheses of interest, it cannot act like a competing hypothesis. The **unconstrained can only be used as a failsafe**:

- If the best hypothesis is the unconstrained is, then your hypotheses of interest are weak, that is, they are not supported by the data. One should not compare the strength of the hypotheses of interest (to avoid choosing the best from worst).
  Notably, one can take on an additional exploratory approach to create one or more new hypotheses for future research.

- If the best hypothesis is one of the hypotheses of interest, then it is not weak and can be compared to the other hypothesis/-es of interest. Note that one can compare all the non-weak hypotheses to all the hypotheses of interest.
  Be aware of the special cases, as will be discussed in the Special cases Section.

```
# H1, H2, and unconstrained (default)
H1 <- "D1 > D2 > D3" # mu1 > mu2 > mu3
H2 <- "D1 < D2 < D3" # mu1 < mu2 < mu3

# Apply GORIC #
set.seed(123)
results_12u <- goric(fit, hypotheses = list(H1, H2))
results_12u
```

Example

```
restriktor (0.5-85): generalized order-restricted information criterion:

Results:
            model    loglik  penalty    goric  loglik.weights  penalty.weights  goric.weights
1              H1  -123.384    2.832  252.431           0.500            0.433          0.763
2              H2  -131.346    2.832  268.356           0.000            0.433          0.000
3   unconstrained  -123.384    4.000  254.767           0.500            0.135          0.237
```

```
round(results_12u$ratio.gw, 2)
```

```
                vs. H1   vs. H2  vs. unconstrained
H1                1.00  2871.23               3.22
H2                0.00     1.00               0.00
unconstrained     0.31   892.89               1.00
```

From this, you can conclude that $H_1 : \mu_1 > \mu_2 > \mu_3$ is not a weak hypothesis, since it is > 1 times more supported than the unconstrained. Therefore, H_1 can be compared to its competing hypothesis $H_2 : \mu_1 < \mu_2 < \mu_3$. Note that H_2 is weak 0 / 0.237 < 1 or 0 < 0.237; so, we already know that H_1 is better, but not yet how much better.

From the GORIC weights or, better, their ratios (cf. `results_12u$ratio.gw` - H1 vs. H2), you can conclude that H_1 is 2871 times more supported than H_2. This denotes (almost) full support for H_1.
In comparison, if you would calculate the GORIC(A) weights for the set consisting of solely H_1 and H_2 (thus,

without the unconstrained), you would obtain GORIC weights of 1.000 and .000, respectively; denoting full support for H_1.

Population information:
In the data generation, I used a ratio of population means of 3:2:1; implying that H_1 is correct. More specifically, I used population mean values of approximately 0.98, 0.65, and 0.33. This implies that Cohen's d is approximately .27; thus, there is a small to medium population effect size (which are in the same order as hypothesized in H_1). I then sampled 100 observations, ran an ANOVA (with three groups), and applied the GORIC.
When I would sample more observations, it does not affect the results in this case. In case H_2 did receive some support (i.e., had a GORIC(A) weight larger than 0), increasing the same size would lead to GORIC(A) weights as reported here, where the ratio of GORIC(A) weights for H_1 and the unconstrained is fixed and the ratio of GORIC(A) weights for H_1 and H_2 is infinite.

Note: As will become clear in the Special cases Section, one may want to be able to find support for the boundary of the best hypothesis H_1 and the other hypotheses. Then, one should first check the ratio of fit/loglik weights (i.e. ratio of loglik.weights). Since the loglik.weights of H_1 and its competing (non-overlapping) hypothesis H_2 differ (i.e., their ratio of loglik.weights is not close to 1), there is no support for the boundary of these hypotheses; only support for H_1.

No failsafe

Only if your hypotheses of interest cover all theories/orderings (i.e., cover the whole parameter space), then you do not need a failsafe hypothesis. Namely, the truth is covered in one or more hypotheses.
Be ware of the special cases, like overlapping hypotheses, as will be discussed in the Special cases Section.

```
# H1, H2, and unconstrained (default)
H1 <- "D1 > D2" # # => H1: D1 > D2, D3 # mu1 > mu2, mu3
H2 <- "D1 < D2" # # => H2: D1 < D2, D3 # mu1 < mu2, mu3

# Apply GORIC #
set.seed(123)
results_12 <- goric(fit, hypotheses = list(H1, H2), comparison = "none")
results_12
```

Example

```
restriktor (0.5-85): generalized order-restricted information criterion:

Results:
  model    loglik  penalty    goric  loglik.weights  penalty.weights  goric.weights
1    H1  -123.384    3.500  253.767           0.978            0.500          0.978
2    H2  -127.183    3.500  261.366           0.022            0.500          0.022
---
The order-restricted hypothesis 'H1' has 44.68 times more support than 'H2'.
```
```
round(results_12$ratio.gw, 2)
```

```
   vs. H1 vs. H2
H1   1.00  44.68
H2   0.02   1.00
```

From this, you can conclude that $H_1 : \mu_1 > \mu_2, \mu_3$ is
$0.978 / 0.022 \approx 45$ times more supported than $H_2 : \mu_1 < \mu_2, \mu_3$.

Population information:
In the data generation, I used a ratio of population means of 3:2:1; implying that H_1 is correct. More

specifically, I used population mean values of approximately 0.98, 0.65, and 0.33. This implies that Cohen's d is approximately .27; thus, there is a small to medium population effect size (which are in the same order as hypothesized in H_1). I then sampled 100 observations, ran an ANOVA (with three groups), and applied the GORIC.

When I would sample more observations, the GORIC(A) weights of H_1 and H_2 will converge to 1 and 0, respectively, leading to an infinite support for H_1 versus H_2; stated otherwise, leading to full support for H_1.

Note: As will become clear in the Special cases Section, one may want to be able to find support for the boundary of the best hypothesis H_1 and the other hypotheses. Then, one should first check the ratio of fit/loglik weights (i.e. ratio of loglik.weights). Since the loglik.weights differ (i.e., the ratio of loglik.weights is not close to 1), there is no support for the boundary of H_1 and H_2; only support for H_1.

Note: Hypothesis specification

A hypothesis is considered the best of the set if it has the highest GORIC(A) weight. In that case, all ratios of GORIC(A) weights of that hypothesis versus another one is 1 or higher. In case of decision making, you may want to pre-specify a cut-off value x for the ratio of GORIC(A) weights: if the ratio of GORIC(A) weights is larger than x, then I am willing to choose (to make policy based on) this best hypothesis. It can be hard to pre-specify x; unless there is already previous research perhaps. Alternatively, you could use the benchmark package for this, see 'Guidelines_GORIC-benchmarks.html' on https://github.com/rebeccakuiper/Tutorials. My advise is to create the hypotheses in such a way that finding a ratio of 1 (or higher) is enough evidence for the preferred hypothesis. This can still be hard to do, but in my opinion is easier than specifying x.
I would also advise against overlapping hypotheses and using equality restrictions; for more detail, see the special cases discussed in the Special cases Section.

For example, let us assume that we compare an outcome (e.g., a standardized level of happiness) for three types of treatments: a new treatment (A), the established treatment (B), and a placebo treatment (P). You could evaluate $H_{1g} : \mu_A > \mu_B > \mu_P$ versus its complement H_c. Now, it can be hard to decide to go for Treatment A when H_{1g} is 1 or perhaps 1.1 times more supported than H_c. Then, it could be handy (if possible) to specify the hypothesis in such a way that a ratio of (just over) 1 makes you willing to choose for Treatment A. You could specify a minimum difference between Treatments A and B and additionally state (as a check) that Treatment B (and then also A) does better than the placebo: $H_{1m} : \mu_A - \mu_B > .1, \ \mu_B > \mu_P$.

Note: $H_{1g} : \mu_A > \mu_B > \mu_P$ can be rewritten as $H_{1g} : \mu_A - \mu_B > 0, \ \mu_B > \mu_P$. In H_{1m}, the (minimum) difference/bound/threshold of 0 is replaced by .1, the minimum difference in treatments you would like to find.

Example

```
# General hypothesis vs complement
H1g <- "D1 > D2 > D3" # mu1 >  mu2 > mu3

# Apply GORIC #
set.seed(123)
results_1g <- goric(fit, hypotheses = list(H1g = H1g), comparison = "complement")
results_1g


restriktor (0.5-85): generalized order-restricted information criterion:

Results:
        model     loglik  penalty     goric  loglik.weights  penalty.weights  goric.weights
1         H1g   -123.384    2.832   252.431           0.831            0.698          0.919
2  complement   -124.978    3.668   257.292           0.169            0.302          0.081
---
The order-restricted hypothesis 'H1g' has 11.36 times more support than its complement.
```

```r
round(results_1g$ratio.gw, 2)
```

```
          vs. H1g vs. complement
H1g            1.00           11.36
complement     0.09            1.00
```

From this, you can conclude that $H_{1g} : \mu_1 > \mu_2 > \mu_3$ is 0.919 / 0.081 ≈ 11.4 times more supported than its complement. Since the ratio is larger than 1, H_{1g} is the best.
[Note: Since the fit/loglik weights differ (i.e., the ratio of loglik.weights is not close to 1), there is no support for the boundary of H_{1g} and its complement - as discussed in the Special cases Section.]

You may now wonder whether this support is large enough to start using Treatment A instead of B. Therefore, you could – if possible – make you hypothesis more specific, such that a ratio of 1 or more is sufficient to select the hypothesis:

```r
# More specific hypothesis vs complement
H1m <- "D1 - D2 > .1, D2 > D3" # mu1 - mu2 > .1, mu2 > mu3

# Apply GORIC #
set.seed(123)
results_1m <- goric(fit, hypotheses = list(H1m = H1m), comparison = "complement")
results_1m
```

```
restriktor (0.5-85): generalized order-restricted information criterion:

Results:
        model    loglik  penalty     goric  loglik.weights  penalty.weights  goric.weights
1         H1m  -123.384    2.832   252.431           0.736            0.698          0.865
2  complement  -124.409    3.668   256.154           0.264            0.302          0.135
---
The order-restricted hypothesis 'H1m' has 6.43 times more support than its complement.
```

```r
round(results_1m$ratio.gw, 2)
```

```
          vs. H1m vs. complement
H1m            1.00            6.43
complement     0.16            1.00
```

From this, you can conclude that $H_{1m} : \mu_1 - \mu_2 > .1, \ \mu_2 > \mu_3$ is 0.865 / 0.135 ≈ 6.4 > 1 times more supported than its complement. Since the difference in treatment means is at least as large as pre-specified, you can now convincingly go for Treatment A.
[Note: Since the fit/loglik weights differ (i.e., the ratio of loglik.weights is not close to 1), there is no support for the boundary of H_{1m} and its complement - as discussed in the Special cases Section.]

Population information:
In the data generation, I used a ratio of population means of 3:2.5:1; implying that both H_{1g} and H_{1m} are correct. More specifically, I used population mean values of approximately 0.89, 0.74, and 0.30. This implies that Cohen's d is approximately .25; thus, there is a small to medium population effect size (in the same ordering as hypothesized). I then sampled 100 observations, ran an ANOVA (with three groups), and applied the GORIC.
When I would sample more observations, the GORIC(A) weight of both H_{1g} and H_{1m} will converge to 1, reflecting full support; and the ratio of GORIC(A) weights of the hypothesis of interest versus its complement will go to infinity.

Notes:

- When specifying a bound/treshold for a positive difference (e.g., the .1 in the example above), you should use the minimum difference you would like to find. Notably, if you use a higher bound (in case

of an > restriction) then needed, the complexity does not change, while the fit decreases; which leads to less support for your hypothesis of interest.

- In case you want to inspect an absolute difference, you can make use of the abs() function. E.g., $|\beta_1| > |\beta_2|$ can be reflected in the code - using b1 and b2 - by abs(b1) > abs(b2); or, when using bounds/thresholds, abs(b1) - abs(b2) > .1; and $|\beta_1 - \beta_2| > .1$ can be reflected in the code by abs(b1 - b2) > .1.

Special cases

Equal fit

Overlapping hypotheses

You should be aware when some of the hypotheses overlap. Note that all hypotheses overlap with the unconstrained hypotheses (per definition). Also, competing hypotheses can overlap; e.g., $\beta_1 > \beta_2$, β_3 and $\beta_1 > \beta_2 > \beta_3$ overlap, since the latter is a subset / special case of the first.

When hypotheses overlap and the truth lies in this overlapping part (i.e., in the subset):

- Then, the overlapping hypotheses share support, that is, they divide the support. Consequently, none of them will obtain full support (i.e., a GORIC(A) weight of 1); and their relative support (i.e., ratio of GORIC(A) weights) is even bounded. Bear in mind that the most parsimonious hypothesis will be the best one.

- Then, the fit/loglik values of the overlapping hypotheses will be equal (i.e., the ratio of loglik.weights will be 1), that is, these overlapping hypotheses will have the same maximum log likelihood. In that case, the ratio of GORIC(A) weights will be solely based on the penalty part (and thus equal the so-called penalty weights).

- Then, it does not make sense to interpret the ratio of GORIC(A) weights of these overlapping hypotheses, you just know that there is support for the overlap of the hypotheses.
  If possible and if of interest, you can specify the overlap and evaluate that versus its complement (to obtain the support for the overlapping part). This can be of interest for future research. Notably, if the overlap cannot easily be specified, one can alternatively inspect the best hypothesis versus its complement.

Thus, when hypotheses have the same fit values (and thus when the ratio of their loglik.weights equal 1), you know that there is support for the overlap of the hypotheses. Note that, when one hypothesis is a subset of another, this implies support for the subset. Note further that, in some cases, this implies support for the boundary/border of hypotheses, as will be discussed in a section later on.

In contrast, **when there is support for only one of the overlapping hypotheses**, this implies support for the non-overlapping part. Bear in mind that the GORIC(A) weight itself addresses the support for the complete hypothesis, not the overlap. If possible and if of interest, you can specify the (non-)overlapping part and evaluate that versus its complement (to obtain the support for this (non-)overlapping part). This can be of interest for future research: You then create a more specific hypothesis or more adjusted set of hypotheses to be evaluated in future research which aids in theory building.

```
# H1, H2, and unconstrained (default) - subset true
H1 <- "D1 > D2 > D3"              # mu1 > mu2 > mu3
H2 <- "D1 > D2" # H2: D1 > D2, D3  # mu1 > mu2,  mu3

# Apply GORIC #
set.seed(123)
```

```
results_12u <- goric(fit, hypotheses = list(H1, H2))
results_12u
```

Example: Subset true

restriktor (0.5-85): generalized order-restricted information criterion:

Results:

	model	loglik	penalty	goric	loglik.weights	penalty.weights	goric.weights
1	H1	-123.384	2.832	252.431	0.333	0.548	0.548
2	H2	-123.384	3.500	253.767	0.333	0.281	0.281
3	unconstrained	-123.384	4.000	254.767	0.333	0.171	0.171

```
round(results_12u$ratio.gw, 3)
```

	vs. H1	vs. H2	vs. unconstrained
H1	1.000	1.950	3.216
H2	0.513	1.000	1.649
unconstrained	0.311	0.607	1.000

From this, you can conclude that $H_1 : \mu_1 > \mu_2 > \mu_3$ and $H_2 : \mu_1 > \mu_2$, μ_3 are not weak, since they are > 1 times more supported than the unconstrained. Therefore, H_1 and H_2 can be compared.

From the GORIC weights, we can conclude that H_1 is the best. Before we interpret the ratio of GORIC weights (i.e., `goric.weights`), we need to check the fit/loglik values of H_1 and H_2. These are the same, that is, the ratio of loglik.weights is 1. Hence, the relative support (denoted by the GORIC weight) is bounded and, now, attains the maximum value. Thus, there is support for the overlap of the hypotheses, which is H_1 here (since H_1 is a subset of H_2).

You could choose to investigate the support for this overlap, by evaluating H_1 versus its complement. We already did this in the first example subsection, there we found that H_1 is 4.88 times more supported than its complement (with an error probability of 17%). For future research, you could advise to evaluate H_1 versus its complement.

Population information:
In the data generation, I used a ratio of population means of 3:2:1; implying that H_1 is correct. More specifically, I used population mean values of approximately 0.98, 0.65, and 0.33. This implies that Cohen's d is approximately .27; thus, there is a small to medium population effect size (which are in the same order as hypothesized). I then sampled 100 observations, ran an ANOVA (with three groups), and applied the GORIC.
When I would sample more observations, it does not affect the GORIC(A) weights for H_1, H_2, and the unconstrained.

```
# H1, H2, and unconstrained (default) - non-overlapping part true
H1 <- "D1 > D2 > D3"              # mu1 > mu2 > mu3
H2 <- "D1 > D2" # H2: D1 > D2, D3  # mu1 > mu2,  mu3

# Apply GORIC #
set.seed(123)
results_12u <- goric(fit, hypotheses = list(H1, H2))
results_12u
```

Example: Non-overlapping part true

restriktor (0.5-85): generalized order-restricted information criterion:

Results:

```
          model     loglik  penalty     goric  loglik.weights  penalty.weights  goric.weights
1            H1   -124.316    2.832   254.297           0.164            0.548          0.323
2            H2   -123.384    3.500   253.767           0.418            0.281          0.421
3  unconstrained  -123.384    4.000   254.767           0.418            0.171          0.255
```

```
round(results_12u$ratio.gw, 2)
```

```
               vs. H1  vs. H2  vs. unconstrained
H1               1.00    0.77               1.27
H2               1.30    1.00               1.65
unconstrained    0.79    0.61               1.00
```

From this, you can conclude that $H_1 : \mu_1 > \mu_2 > \mu_3$ and $H_2 : \mu_1 > \mu_2, \mu_3$ are not weak, since since they are > 1 times more supported than the unconstrained. Therefore, H_1 and H_2 can be compared.

From the GORIC weights, we can conclude that H_2 is the best. Since the fit/loglik values of H_2 and H_1 are not the same (i.e., the ratio of loglik.weights is not close to 1), there is no support for their overlap (which would have been their boundary then; more details can be found in a section later on). From the GORIC weights `goric.weights` (or their ratios in `results_12u$ratio.gw` - H1 vs. H2), you can conclude that H_2 is $0.421 / 0.323 \approx 1.3$ more supported than H_1.

Notably, if you would calculate the GORIC(A) weight for the set consisting of solely H_1 and H_2, you would obtain weights of $0.323 / (0.323 + 0.421) \approx 0.434$ and $1 - 0.434 = 0.566$, respectively (denoting the same relative support of course).

Since we know/see that H_1 and H_2 overlap, we can take it one step further. The support for H_2 indicates support for the non-overlapping part of H_2 and H_1. In this case, it is possible to specify the the non-overlapping part, namely $\mu_1 > \mu_2 < \mu_3$. If of interest, you could evaluate that versus its complement (as done below). For future research, you could advise to evaluate this non-overlapping part versus its complement (as done in the next subsection).

Population information:
In the data generation, I used a ratio of population means of 3:2:3; implying that H_2 is correct. More specifically, I used population mean values of approximately 0.62, 0.41, and 0.62. This implies that Cohen's d is approximately .107; thus, there is a small population effect size (in the same ordering as hypothesized in H_1). I then sampled 100 observations, ran an ANOVA (with three groups), and applied the GORIC.
When I would sample more observations, the ratio of GORIC(A) weights for H_2 versus H_1 would go to infinity. Note that the GORIC(A) weight of H_1 would go to 0, but that of H_2 not to 1 since the unconstrained always obtains support (because it also includes H_2 / the true ordering).

```
# non-overlapping part vs  its complement
Hn <- "D1 > D2 < D3" # mu1 > mu2 < mu3

# Apply GORIC #
set.seed(123)
results_n <- goric(fit, hypotheses = list(Hn = Hn), comparison = "complement")
results_n
```

Follow-up exploratory analysis for non-overlapping part

```
restriktor (0.5-85): generalized order-restricted information criterion:

Results:
          model     loglik  penalty     goric  loglik.weights  penalty.weights  goric.weights
1            Hn   -123.384    3.168   253.103           0.718            0.541          0.750
2  complement    -124.316    3.332   255.297           0.282            0.459          0.250
---
```

The order-restricted hypothesis 'Hn' has 2.99 times more support than its complement.

Support for boundary

When hypotheses do not overlap, there can mathematically still be overlap, namely the boundary/border of these hypotheses. This is always the case when evaluating an informative hypothesis versus its complement, but can also be the case for other pairs of hypotheses.
As a very simple example: The overlap/boundary of $\beta_1 > 0$ and its complement (here, $\beta_1 < 0$) is $\beta_1 = 0$.

Consequently, when non-overlapping hypotheses have (approximately) the same fit values (and thus when the ratio of loglik.weights (approximately) equal 1), you know that there is support for the boundary of these hypotheses.

Note: Depending on your interest and specification of your hypotheses, you may or may not be interested in checking the support for a boundary/border. As a very simple example: When evaluating $\beta_1 > 0$ versus its complement (here, $\beta_1 < 0$), you may want to rule out that the support for H_1 is actually support for the boundary $\beta_1 = 0$. However, when evaluating $\beta_1 > 0.1$ versus its complement (here, $\beta_1 < 0.1$), you may not be interested in possible support for the boundary $\beta_1 = 0.1$; since you specified a minimum support to rule out support for the null $\beta_1 = 0$.

```
# H1 vs complement
H1 <- "D1 > D2 > D3" # mu1 > mu2 > mu3

# Apply GORIC #
set.seed(123)
results_1c <- goric(fit, hypotheses = list(H1), comparison = "complement")
results_1c
```

Example: H_1 versus its complement

```
restriktor (0.5-85): generalized order-restricted information criterion:

Results:
        model     loglik  penalty     goric  loglik.weights  penalty.weights  goric.weights
1          H1   -123.450    2.832   252.564           0.483            0.698          0.683
2  complement   -123.384    3.668   254.103           0.517            0.302          0.317
---
The order-restricted hypothesis 'H1' has 2.16 times more support than its complement.
#
coef(fit)
```

```
        D1        D2        D3
0.9342210 0.4515234 0.5260562
```

From the GORIC weights, we can conclude that H_1 is the best. Before interpreting the ratio of GORIC(A) weights, one may want to check the fit/loglik values (or the ratio of their weights), to see whether there is support for a boundary. From the output, you can see that the fit values of H_1 and its complement resemble (-123.45 vs -123.384) and, thus, their loglik.weights are almost the same 0.483 vs 0.517; i.e., their ratio is close to 1). Likewise, you can see that the GORICA weights resemble the penalty weights. This indicates that there is support for the boundary of the hypotheses. Bear in mind that the overlap of a hypothesis and its complement is the boundary of them.
In this example, the boundaries are:
$\mu_1 = \mu_2 > \mu_3$,
$\mu_1 > \mu_2 = \mu_3$, and
$\mu_1 = \mu_2 = \mu_3$.

Now, you can take two approaches: i) do an exploratory evaluation of these hypotheses (as done below) or ii) inspect the sample means (by looking at coef(fit)). Here, based on both approaches, you would conclude that there is support for $\mu_1 > \mu_2 = \mu_3$. This hypothesis can then be proposed to be evaluated in future research. If of interest, one can also evaluate this boundary hypothesis versus its complement. Notably, such a boundary hypothesis contains at least one equality (here, there is one), which you may want to rephrase as an about-equality, as discussed in a next section; and exemplified in the the follow-up analysis in a next section.

Population information:
In the data generation, I used a ratio of population means of 3:2:2; implying that (the boundaries of) H_1 and its complement are both correct (where H_1 is the most parsimonious one). More specifically, I used population mean values of approximately 0.71, 0.482, and 0.48. This implies that Cohen's d is approximately .11; thus, there is a small population effect size. I then sampled 100 observations, ran an ANOVA (with three groups), and applied the GORIC.
When I would sample more observations, the fit values will become more closer and the GORIC(A) weights will converge to the penalty weights.

```
# Check on three boundary hypotheses
Hb1 <- "D1 = D2 > D3" # mu1 = mu2 > mu3
Hb2 <- "D1 > D2 = D3" # mu1 > mu2 = mu3
Hb3 <- "D1 = D2 = D3" # mu1 = mu2 = mu3

# Apply GORIC #
set.seed(123)
results_b123 <- goric(fit, hypotheses = list(Hb1 = Hb1, Hb2 = Hb2, Hb3 = Hb3))
results_b123
```

Follow-up exploratory analysis for boundary

restriktor (0.5-85): generalized order-restricted information criterion:

Results:

	model	loglik	penalty	goric	loglik.weights	penalty.weights	goric.weights
1	Hb1	-126.132	2.500	257.263	0.031	0.258	0.050
2	Hb2	-123.450	2.500	251.900	0.458	0.258	0.725
3	Hb3	-126.570	2.000	257.139	0.020	0.426	0.053
4	unconstrained	-123.384	4.000	254.767	0.490	0.058	0.173

Ratio GORIC-weights:

	vs. Hb1	vs. Hb2	vs. Hb3	vs. unconstrained
Hb1	1.000	0.068	0.940	0.287
Hb2	14.611	1.000	13.734	4.194
Hb3	1.064	0.073	1.000	0.305
unconstrained	3.484	0.238	3.275	1.000

```
# H1, H2, and unconstrained (default) - non-overlapping part true
H1 <- "D1 > D2 > D3" # mu1 > mu2 > mu3
H2 <- "D1 > D2 < D3" # mu1 > mu2 < mu3

# Apply GORIC #
set.seed(123)
```

```
results_12u <- goric(fit, hypotheses = list(H1, H2))
results_12u
```

Example: H_1, H_2, and the unconstrained

```
restriktor (0.5-85): generalized order-restricted information criterion:

Results:
          model    loglik  penalty    goric  loglik.weights  penalty.weights  goric.weights
1            H1  -123.450    2.832  252.564           0.319            0.494          0.477
2            H2  -123.384    3.168  253.103           0.341            0.353          0.364
3  unconstrained  -123.384    4.000  254.767           0.341            0.154          0.159
```

```
round(results_12u$ratio.gw, 2)
```

```
               vs. H1 vs. H2 vs. unconstrained
H1               1.00   1.31             3.01
H2               0.76   1.00             2.30
unconstrained    0.33   0.44             1.00
```

```
round(results_12u$ratio.pw, 2)
```

```
               vs. H1 vs. H2 vs. unconstrained
H1               1.00   1.40             3.22
H2               0.71   1.00             2.30
unconstrained    0.31   0.44             1.00
```

From this, you can conclude that $H_1 : \mu_1 > \mu_2 > \mu_3$ and $H_2 : \mu_1 > \mu_2 < \mu_3$ are not weak, since since they are > 1 times more supported than the unconstrained. Therefore, H_1 and H_2 can be compared.

From the GORIC weights, we can conclude that H_1 is the best. When checking the fit/loglik values (or the ratio of their weights), you can see that the fit values of H_1 and H_2 resemble and, thus, their loglik.weights are almost the same (i.e., the ratio of loglik.weights is close to 1). Likewise, you can see that the GORICA weights of H_1 vs H_2 (in `results_12u$ratio.gw` - H1 vs. H2) resemble the penalty weights of H_1 vs H_2 (in `results_12u$ratio.pw` - H1 vs. H2). This indicates that there is support for the overlap of these hypotheses. In this case, the overlap means $\mu_1 > \mu_2$ (i.e., the part H_1 and H_2 have in common) together with the boundary of $\mu_2 > \mu_3$ and $\mu_2 < \mu_3$, which is $\mu_2 = \mu_3$; hence, $\mu_1 > \mu_2 = \mu_3$. This hypothesis can then be proposed to be evaluated in future research.

If of interest, one can evaluate this new hypothesis versus its complement. Notably, this hypothesis contains an equality, which you may want to rephrase as an about-equality, as discussed in a next section and exemplified in the next subsection.

Population information:
In the data generation, I used a ratio of population means of 3:2:2; implying that (the boundaries of) H_1 and $H2$ are both correct (where both hypotheses are equally parsimonious). More specifically, I used population mean values of approximately 0.71, 0.482, and 0.48. This implies that Cohen's d is approximately .11; thus, there is a small population effect size. I then sampled 100 observations, ran an ANOVA (with three groups), and applied the GORIC.
When I would sample more observations, the fit values will become more closer and the GORIC(A) weights will converge to the penalty weights.

```
# overlap H1 and H2 versus its complement
Hb <- "D1 > D2, -.1 < D2 - D3 < .1" # mu1 > mu2, -.1 < mu2 - mu3 < .1

# Apply GORIC #
set.seed(123)
```

```
results_b <- goric(fit, hypotheses = list(Hb = Hb), comparison = "complement")
results_b
```

Follow-up exploratory analysis for boundary

```
restriktor (0.5-85): generalized order-restricted information criterion:

Results:
        model      loglik  penalty      goric  loglik.weights  penalty.weights  goric.weights
1          Hb    -123.384    2.500    251.767           0.944            0.731          0.979
2  complement    -126.207    3.500    259.413           0.056            0.269          0.021
---
The order-restricted hypothesis 'Hb' has 45.75 times more support than its complement.
```

Notes

Note 1:
A ratio of GORIC(A) weights of 1 means that the hypotheses are equally likely (in terms of balancing fit and complexity). Then, both hypotheses are equally supported. This does not per se mean that there is support for their overlap (or boundary). Only when the penalty values of the hypotheses are the same, this coincides. For example, the penalty of $\beta_1 > 0$ equals that of its complement (here, $\beta_1 < 0$), thus their ratio of GORIC(A) weights will equal 1 when $\beta_1 = 0$.
Bear in mind that when the ratio of GORIC(A) weights equals the ratio of penalty weights (which does not need to be 1), then there is support for the overlap (or boundary). Likewise, one can look for equal fit values (reflected by a ratio of loglik.weights of 1), as done in all the examples.

Note 2:
There are no cut-off values for fit values to resemble or for the ratio of `loglik.weights` being close to 1 (only gut feelings). Since the fit values depend on sample size, it is probably better to look the ratio of `loglik.weights`. Additionally, it could help to check the estimates that belong to those hypotheses (using, e.g., `results_0c$ormle`) and see if they meaningfully differ.
One could also look into log-likelihood benchmarks, as discussed in 'Guidelines_GORIC-benchmarks' on https://github.com/rebeccakuiper/Tutorials.

Just below maxmimum fit

Equality restriction (=)

Even if an equality restriction (e.g., $\beta_1 = \beta_2$, that is, $\beta_1 - \beta_2 = 0$) is true in the population, the probability of finding this in your data is 0. In your data, the relationship will never be exactly equal. Therefore, the fit value (i.e., the maximum log likelihood) of a true hypothesis with equality restriction will never equate the maximum fit (i.e., the maximum log likelihood of the unconstrained), but a value (just) below this maximum. Consequently, a true hypothesis with equality restriction will never obtain a GORIC(A) weight of 1, but one lower than 1.

Hence, be aware of interpreting your results when there is at least one equality restriction in the set. Moreover, only include an equality restriction when you are really interested in it. If possible, I would advise you to specify an about-equality restriction instead, as discussed next, after the example.

```
# Equality versus its complement
He <- "D1 = D2 > 0, D3 > 0" # mu1 = mu2 > 0, mu3 > 0

# Apply GORIC #
set.seed(123)
```

```
results_e <- goric(fit, hypotheses = list(He = He), comparison = "complement")
results_e
```

Example

```
restriktor (0.5-85): generalized order-restricted information criterion:

Results:
        model    loglik  penalty    goric  loglik.weights  penalty.weights  goric.weights
1          He  -139.114    2.000  282.228           0.499            0.818          0.817
2  complement  -139.111    3.500  285.222           0.501            0.182          0.183
---
The order-restricted hypothesis 'He' has 4.47 times more support than its complement.
```

```
round(results_e$ratio.lw, 2) # ratio of loglik weights
```

```
            vs. He  vs. complement
He               1               1
complement       1               1
```

```
round(results_e$ratio.gw, 2) # ratio of GORIC weights
```

```
            vs. He  vs. complement
He            1.00            4.47
complement    0.22            1.00
```

From this, you can conclude that $H_1 : \mu_1 = \mu_2 > 0,\ \mu_3 > 0$ is more supported than its complement. When you look at the fit/loglik part, you see that both fit/loglik values resemble; that is, the ratio of loglik.weights is close to 1 (namely, 1). Thus, you can conclude that there is support for the overlap/boundary of these hypotheses.

Here, the complement receives the highest fit; that is, the estimates are in agreement with the complement. In this case, the fit of the complement is based on the fit for $\mu_1 > 0,\ \mu_2 > 0,\ \mu_3 > 0$ (so, the hypothesis that does not restrict the first two means to be equal). Consequently, the preferred hypothesis H_1 does not have the highest fit/loglik (it has the best fit-complexity balance). Notably, since H_1 contains an equality, it will also never have maximum fit. Moreover, the weight for a true equality hypothesis will never be 1. This could be a reason to evaluate an about-equality hypothesis instead, as discussed in the next section.

Population information:
In the data generation, I used a ratio of population means of 3:3:1.5; implying that (the boundaries of) H_1 and its complement are both correct (where H_1 is the most parsimonious one). More specifically, I used population mean values of approximately 0.70, 0.70, and 0.35. This implies that Cohen's d is approximately .16; thus, there is a small to medium population effect size. I then sampled 100 observations, ran an ANOVA (with three groups), and applied the GORIC.
When I would sample more observations, the fit/loglik values of H_1 and its complement will remain close (with the highest value for the complement) and the results are conclusion-wise the same; but, more importantly, the weight for H_1 will not converge to 1.

About-equality restrictions

Instead of equality restrictions (e.g., $\beta_1 = \beta_2$, that is, $\beta_1 - \beta_2 = 0$) one can specify about-equality restrictions (e.g., $-0.2 < \beta_1 - \beta_2 < 0.2$). In this case, the hypothesis can have a maximum fit and a GORIC(A) weight of 1.
In practice, it can be hard to specify such a range. Notably, when the range is too broad, the hypothesis will per definition have maximum fit. In addition, the broader the range, the lower the fit for the complement. Thus, you should specify the range meaningfully; preferably, based on literature and/or expertise or perhaps data-based.

If you use a **data-based approach** to specify the range, it is best to use a **training set** (when the sample size is large enough). In that case, one uses one part of the data to come up with a range value for the about-equality restriction and the other part of the data to evaluate the about-equality restriction.

When the sample size is not large enough, you could for example use a cut-off value from the literature together with the standard deviation (i.e., the standard errors of the mean difference times the square root of the sample size). Bear in mind that you use the data twice, so only use this when creating a training set is not possible.

```
# Equality versus its complement
Hae <- "-0.2 < D1 - D2 < .2, D1 > 0, D2 > 0, D3 > 0"
# -0.2 < mu1 - mu2 < .2, mu1 > 0, mu2 > 0, mu3 > 0

# Apply GORIC #
set.seed(123)
results_ae <- goric(fit, hypotheses = list(Hae = Hae), comparison = "complement")
results_ae
```

Example

```
restriktor (0.5-85): generalized order-restricted information criterion:

Results:
       model    loglik  penalty    goric  loglik.weights  penalty.weights  goric.weights
1        Hae  -139.111    2.000  282.222           0.926            0.818          0.983
2 complement  -141.645    3.500  290.290           0.074            0.182          0.017
---
The order-restricted hypothesis 'Hae' has 56.48 times more support than its complement.
```

```
round(results_ae$ratio.lw, 2) # ratio of loglik weights
```

```
            vs. Hae vs. complement
Hae            1.00          12.6
complement     0.08           1.0
```

```
round(results_ae$ratio.gw, 2) # ratio of GORIC weights
```

```
            vs. Hae vs. complement
Hae            1.00         56.48
complement     0.02          1.00
```

```
coef(fit)
```

```
       D1        D2        D3
0.8073380 0.8251897 0.3860030
```

From this, you can conclude that $H_1 : -0.2 < \mu_1 - \mu_2 < .2, \mu_1 > 0, \mu_2 > 0, \mu_3 > 0$ is $0.98 \,/\, 0.02 \approx 56.48$ times more supported than its complement. Now, the fit/loglik value of this about-equality hypothesis H_1 is the highest (while it will never be the case for an equality hypothesis).

Note that the fit values of H_1 and its complement are not close: the ratio of loglik.weights is $0.93 \,/\, 0.07 \approx 12.6$ (notably, this ratio changes with the choice of range).

Population information: In the data generation, I used a ratio of population means of 3:3:1.5; implying that H_1 is correct. More specifically, I used population mean values of approximately 0.70, 0.70, and 0.35. This implies that Cohen's d is approximately .16; thus, there is a small to medium population effect size. I then sampled 100 observations, ran an ANOVA (with three groups), and applied the GORIC.

When I would sample more observations, the ratio of loglik weights of H_1 and its complement will go to infinity and, most importantly, the GORIC(A) weight of H_1 will converge to 1.

Note on about-equality vs absolute specification An about-equality restriction should preferably be specified as a range restriction (e.g., $-0.2 < \beta_1 - \beta_2 < 0.2$). In that case, the fit is based on that range restriction; and the penalty then reduces to the penalty for an equality restriction, which make sense since this range restriction is meant as an equality.

Notably, the same applies for using ranges not centered around 0 (e.g., $-0.1 < \beta_1 - \beta_2 < 0.2$).

In an about-equality restriction, one specifies that a (function of) parameter(s) lies in between two values. In theory, one could use the abs() function as well, but that is not advised. If you would use abs(b1 - b2) < .2, an inequality restriction is evaluated where the sign of the expression in abs() is made positive. That is, if b1 - b2 is positive, b1 - b2 < .2 is evaluated; if b1 - b2 is negative, then -(b1 - b2) < .2 is evaluated. In this case, the penalty is based on the inequality, not an equality restriction.

I would advise on
- expressing an about-equality via a range restriction, where you specify a range of values.
- using the abs() function when you are interested in an inequality restriction for the absolute value of one or more parameters / for the strength of one or more predictors (e.g., abs(b1) > abs(b2); or, when using bounds/thresholds, abs(b1) - abs(b2) > .1).

Not highest fit

The GORIC(A), like the AIC or any other information criterion, balances fit and complexity. You are thus looking for a trade-off: The hypothesis that describes the data best with the fewest number of parameters. Consequently, the preferred/best hypothesis does not need to have the highest fit - then, the difference in complexity outweighs the difference in fit.

When the best hypothesis does not have maximum fit, you know that the estimates (i.e., your data) are not fully in agreement with the preferred hypothesis; they are in agreement with the hypothesis which has maximum fit. Per definition, this scenario goes hand in hand with a GORIC(A) weight for the preferred hypothesis being lower than 1.

In such cases, the sample size is not large enough to find convincing evidence. In such a case, the GORIC(A) weights will also denote this uncertainty, since none of the weights will be 1 (and perhaps the weights will even be close together).

When you encounter such a situation, you may want to additionally take on an additional exploratory approach to come up with a few hypotheses for future research (by looking at sample means/estimates taking into account what theoretically makes sense or by specifying some plausible hypotheses which theoretically make sense). Then, in future research, one can evaluate the new set of hypotheses with a larger sample size. When you believe the lower fit is solely because of a small sample or when you cannot create other theoretically meaningful hypotheses, you should suggest to inspect the same hypothesis with a larger sample size in future research.

Example: Correct hypothesis

```
# Equality versus its complement
H1 <- "D1 > D2 > D3" # mu1 > mu2 > mu3

# Apply GORIC #
set.seed(123)
results_c <- goric(fit, hypotheses = list(H1 = H1), comparison = "complement")
results_c

restriktor (0.5-85): generalized order-restricted information criterion:

Results:
        model    loglik  penalty     goric  loglik.weights  penalty.weights  goric.weights
1          H1  -120.398    2.832   246.459           0.419            0.698          0.625
```

```
2   complement   -120.072   3.668   247.481            0.581              0.302              0.375
---
```

The order-restricted hypothesis 'H1' has 1.67 times more support than its complement.

```
round(results_c$ratio.gw, 2)
```

```
            vs. H1 vs. complement
H1             1.0          1.67
complement     0.6          1.00
```

Here, we see that our informative hypothesis H_1 is the preferred one, but it does not have the best fit (ratio of loglik.weights is 0.42 / 0.58 ≈ 0.72 < 1).

Based on the sample size ($N = 100$, that is, approximately 33 observations per group), one could decide to leave the hypothesis like it is and only advise to increase the sample size in future research. Alternatively, you could also go for an additional exploratory approach to come up with one or a few competing hypotheses for future research.

Population information:
In the data generation, I used a ratio of population means of 1.7:1.6:1.5; implying that H_1 is correct. More specifically, I used population mean values of approximately 0.57, 0.53, and 0.50. This implies that Cohen's d is approximately .03; thus, there is a very small population effect size. I then sampled 100 observations, ran an ANOVA (with three groups), and applied the GORIC.
Notably, I sampled multiple times 100 observations (by using a different seed value every time) until I found this situation: Support for the correct H_1, but H_1 does not have highest fit.
When I would sample more observations, the GORIC(A) weight for H_1 converges to 1 (denoting full support for H_1).

Example: Incorrect hypothesis

```
# Equality versus its complement
H1 <- "D1 - D2 > .5, D2 > D3" # mu1 - mu2 > .5, mu2 > mu3

# Apply GORIC #
set.seed(123)
results_i <- goric(fit, hypotheses = list(H1 = H1), comparison = "complement")
results_i
```

```
restriktor (0.5-85): generalized order-restricted information criterion:

Results:
        model     loglik  penalty     goric  loglik.weights  penalty.weights  goric.weights
1          H1   -139.436    2.832   284.536           0.419            0.698          0.625
2  complement   -139.111    3.668   285.558           0.581            0.302          0.375
---
```

The order-restricted hypothesis 'H1' has 1.67 times more support than its complement.

```
round(results_i$ratio.gw, 3)
```

```
            vs. H1 vs. complement
H1             1.0         1.667
complement     0.6         1.000
```

Here, we see that our informative hypothesis H_1 is the preferred one, but it does not have the best fit (ratio of loglik.weights is 0.42 / 0.58 ≈ 0.72 < 1).

Based on your sample size, you could decide to leave the hypothesis like it is and only advise to increase the sample size in future research. Alternatively, you could also go for an additional exploratory approach to

come up with one or a few competing hypotheses for future research.

Population information:
In the data generation, I used a ratio of population means of 3:2:1 and population mean values of approximately 0.98, 0.65, and 0.33. Note that Cohen's d is approximately .27; thus, there is a small to medium population effect size. Note further that the difference between the first two population means is approximately $.33 < .5$. Hence, H_1 is incorrect (and, thus, its complement is correct). I then sampled 100 observations, ran an ANOVA (with three groups), and applied the GORIC.
Notably, I sampled two times 100 observations (by using a different seed value) to find this situation: Support for the incorrect H_1, and the incorrect H_1 does not have the highest fit.
When I would sample more observations, the GORIC(A) weight for H_1 converges to 0 (denoting no support for H_1).

Note on sample size

When the preferred hypothesis does not have the highest fit, this is an indication for having a too small sample - whether the hypothesis is correct or not. In case of a small sample, the GORIC(A) weights will also reflect this uncertainty - which can also be seen (indirectly) at the end of the last section, where I inspect cases for various sample sizes.

Let me also repeat the following:
There are no cut-off values for fit values to resemble or for the ratio of `loglik.weights` being close to 1 (only gut feelings). Since the fit values depend on sample size, it is probably better to look the ratio of `loglik.weights`. Additionally, it could help to check the estimates that belong to those hypotheses (using, e.g., `results_0c$ormle`) and see if they meaningfully differ.
One could also look into log-likelihood benchmarks, as discussed in 'Guidelines_GORIC-benchmarks' on https://github.com/rebeccakuiper/Tutorials.

Remarks Bayesian model selection

All of the above also holds true for model selection using Bayes factors (BFs; comparable to ratio of GORIC(A) weights) and posterior model probabilities (PMPs; comparable to the GORIC(A) weights). However, there is one extra element to take into account in case of Bayesian model selection: prior sensitivity in case of at least one equality restriction. In the R package `bain`, one can do this by checking the results for multiple so-called fraction values; as shown in the first example below.

In terms of conclusions (i.e., there is support for H_1 or not), the output of GORIC(A) and `bain` are (often if not always) the same. In terms of the amount of support, `bain` renders more support for hypotheses with equality restrictions than the GORIC(A) does. This is the case when the equality restrictions are correct but also when they are incorrect; where the first is shown in the first example below and the latter in the second example below.

Example: Prior sensitivity `bain`

```
# Equality versus its complement
H1 <- "D1 = D2 > 0 & D3 > 0"

# Apply bain
set.seed(123)
bain_c_f1 <- bain(fit, H1)
bain_c_f1
```

Bayesian informative hypothesis testing for an object of class lm (ANOVA):

```
   Fit   Com   BF.u   BF.c   PMPa  PMPb  PMPc
```

```
H1 1.628 0.071 22.800 22.800 1.000 0.958 0.958
Hu                                     0.042
Hc                                           0.042
```

Hypotheses:
  H1: D1=D2>0&D3>0

Note: BF.u denotes the Bayes factor of the hypothesis at hand versus the unconstrained hypothesis Hu. BF

```
set.seed(123)
bain_c_f2 <- bain(fit, H1, fraction = 2)
bain_c_f2
```

Bayesian informative hypothesis testing for an object of class lm (ANOVA):

```
    Fit   Com   BF.u   BF.c   PMPa  PMPb  PMPc
H1 1.628 0.101 16.122 16.122 1.000 0.942 0.942
Hu                                     0.058
Hc                                           0.058
```

Hypotheses:
  H1: D1=D2>0&D3>0

Note: BF.u denotes the Bayes factor of the hypothesis at hand versus the unconstrained hypothesis Hu. BF

```
set.seed(123)
bain_c_f3 <- bain(fit, H1, fraction = 3)
bain_c_f3
```

Bayesian informative hypothesis testing for an object of class lm (ANOVA):

```
    Fit   Com   BF.u   BF.c   PMPa  PMPb  PMPc
H1 1.628 0.124 13.164 13.164 1.000 0.929 0.929
Hu                                     0.071
Hc                                           0.071
```

Hypotheses:
  H1: D1=D2>0&D3>0

Note: BF.u denotes the Bayes factor of the hypothesis at hand versus the unconstrained hypothesis Hu. BF

From this, you can see (from BF.c) that the equality hypothesis H_1 is approximately 23, 16, and 13, times more supported than its complement when using a fraction of 1, 2, and 3, respectively.

Population information:
In the data generation, I used a ratio of population means of 3:3:1.5; implying that H_1 is correct. More specifically, I used population mean values of approximately 0.70, 0.70, and 0.35. This implies that Cohen's d is approximately .16; thus, there is a small to medium population effect size. I then sampled 100 observations, ran an ANOVA (with three groups), and applied `bain`.

Comparison `bain` and GORIC(A):
In terms of conclusions (i.e., there is support for H_1 or not), the output of GORIC(A) and `bain` are the same. In terms of the amount of support, `bain` renders more support for equality restrictions than the GORIC(A). In this example, this is good property. But: `bain` rendering more support for equality restrictions happens both when the equality restrictions are true/correct and when they are incorrect; as will be shown in the examples below. Bear in mind that I advise on using about-equality restrictions instead of equality restrictions (notably, the prior sensitivity (of `bain`) occurs for both types of restrictions).

Example: Support for incorrect equality

N = 100

```
# Equality versus its complement
H1 <- "D1 = D2 > 0 & D3 > 0"

# Apply bain
set.seed(123)
bain_i_f1 <- bain(fit, H1)
bain_i_f1
```

Bayesian informative hypothesis testing for an object of class lm (ANOVA):

```
    Fit   Com   BF.u   BF.c   PMPa  PMPb  PMPc
H1 1.393 0.071 19.504 19.504 1.000 0.951 0.951
Hu                                 0.049
Hc                                       0.049
```

Hypotheses:
  H1: D1=D2>0&D3>0

Note: BF.u denotes the Bayes factor of the hypothesis at hand versus the unconstrained hypothesis Hu. BI

```
# Apply GORIC #
set.seed(123)
results_i <- goric(fit, hypotheses = list(H1), comparison = "complement")
results_i
```

restriktor (0.5-85): generalized order-restricted information criterion:

Results:

	model	loglik	penalty	goric	loglik.weights	penalty.weights	goric.weights
1	H1	-139.260	2.000	282.520	0.463	0.818	0.794
2	complement	-139.111	3.500	285.222	0.537	0.182	0.206

The order-restricted hypothesis 'H1' has 3.86 times more support than its complement.

From the `bain` output, you can see (from BF.c) that the incorrect equality hypothesis H_1 is approximately 19.5 times more supported than its complement (when using a fraction of 1). From the GORIC output, you can see that the incorrect equality hypothesis H_1 is approximately 3.8 3.9 times more supported than its complement.

Population information:
In the data generation, I used a ratio of population means of 3:2.5:1; implying that H_1 is correct. More specifically, I used population mean values of approximately 0.89, 0.74, and 0.30. This implies that Cohen's d is approximately .25; thus, there is a small to medium population effect size. I then sampled 100 observations, ran an ANOVA (with three groups), and applied `bain` and the GORIC.
When I would sample more observations, the support for H_1 will go to 0, for both methods; thus, rendering full support for the complement. Note that the GORIC(A) will conclude this sooner (i.e., for smaller samples) than `bain` will:

Below, you can find output for a sample size (N) of 1000 and 10000. In the case of $N = 1000$, GORIC already shows support for the correct complement; while `bain` still favors the incorrect equality restriction. In the case of $N = 10000$, both methods show full support for the complement of H_1.

N = 1000

```
# Equality versus its complement
H1 <- "D1 = D2 > 0 & D3 > 0"
```

```
# Apply bain
set.seed(123)
bain_i_f1_N1000 <- bain(fit, H1)
bain_i_f1_N1000
```

bain

Bayesian informative hypothesis testing for an object of class lm (ANOVA):

```
    Fit    Com   BF.u   BF.c  PMPa   PMPb   PMPc
H1 0.263  0.076  3.458  3.458 1.000  0.776  0.776
Hu                                    0.224
Hc                                           0.224
```

Hypotheses:
  H1: D1=D2>0&D3>0

Note: BF.u denotes the Bayes factor of the hypothesis at hand versus the unconstrained hypothesis Hu. B

```
# Apply GORIC #
set.seed(123)
results_i_N1000 <- goric(fit, hypotheses = list(H1), comparison = "complement")
results_i_N1000
```

GORIC

restriktor (0.5-85): generalized order-restricted information criterion:

Results:
	model	loglik	penalty	goric	loglik.weights	penalty.weights	goric.weights
1	H1	-1346.431	2.000	2696.861	0.045	0.818	0.175
2	complement	-1343.380	3.500	2693.760	0.955	0.182	0.825

The order-restricted hypothesis 'H1' has 0.21 times more support than its complement.

N = 10000

```
# Equality versus its complement
H1 <- "D1 = D2 > 0 & D3 > 0"
```

```
# Apply bain
set.seed(123)
bain_i_f1_N10000 <- bain(fit, H1)
bain_i_f1_N10000
```

bain

Bayesian informative hypothesis testing for an object of class lm (ANOVA):

```
     Fit    Com   BF.u  BF.c  PMPa  PMPb  PMPc
H1 0.000 0.076 0.000 0.000 1.000 0.000 0.000
Hu                                1.000
Hc                                      1.000


Hypotheses:
  H1: D1=D2>0&D3>0


Note: BF.u denotes the Bayes factor of the hypothesis at hand versus the unconstrained hypothesis Hu. BH
```

```
# Apply GORIC #
set.seed(123)
results_i_N10000 <- goric(fit, hypotheses = list(H1), comparison = "complement")
results_i_N10000
```

GORIC

restriktor (0.5-85): generalized order-restricted information criterion:

Results:

	model	loglik	penalty	goric	loglik.weights	penalty.weights	goric.weights
1	H1	-13428.294	2.000	26860.588	0.000	0.818	0.000
2	complement	-13411.461	3.500	26829.922	1.000	0.182	1.000

The order-restricted hypothesis 'H1' has 0.00 times more support than its complement.