

Assessing the p -values of hierarchies [11.01.2025 (logarithm ratio multiplied with n ; closest integer for factorial); 14.01.2025 (number of combinations after each element-specific hierarchy assessment; non-integer factorials); 15.01.2025 (summation form); 21.01.2025 (hierarchy coherence); 23.01.2025 (reversed hierarchy); 14.02.2025 (indices); 22.05.2025 (predicted hierarchy p -values from reversed hierarchy p -values); 10.08.2025 (p -values for binary data)]

Author: Sky Darnos (Germany & Greece).

Sponsors: OSS Capital (US); Janine Hofmann (Swiss).

When looking at hierarchies in observational data, a very obvious way of assessing their statistical significance is to note that, under the null hypothesis, the chance for a neighboring element in the hierarchy to be lower or higher should be exactly 50%, and so after $(n - 1)$ elements in perfect order, the chance for the observed hierarchy to be coincidental would be 0.5^{n-1} . So far, this logic seems inevitable, but the situation becomes trickier when we wish to analyze the hierarchical order relative to every element of the assembly. We are then no longer dealing with independent probabilities, because in each independent hierarchy assessment, we are looking at the same set of numbers.

Ultimately, the odds for a certain hierarchical order can never drop below $1/n!$, because $n!$ is the total number of configurations of n elements in one dimension. We cannot know the probability of a configuration without counting all configurations deemed equally 'ordered', and without knowing the total number of configurations. This means such probabilities cannot be assessed without using factorials in every step of the way.

Configurations left after each individual hierarchy assessment

When the element predicted to be at the top of a hierarchy is indeed at the very top, then $(n - 1)!$ arrangements are left, because the arrangement of the elements below it is not fixed or ordered, and so the chance for this state is:

$$p = \frac{(n - 1)!}{n!}$$

The element that is predicted to be on top might not be exactly on top, and so we can correlate the p -value with how well the observed hierarchy agrees with the predicted hierarchy as follows:

$$p = \frac{\left(n - \frac{0.5^{n-1}}{0.5^m}\right)!}{n!}$$

Here m is the number of times the hierarchical order was right (r) minus the number of times it was wrong (w), where right can mean lower or higher than a specific element, depending on if the elements are arranged in ascending or descending order:

$$m = r - w$$

Odds after assessing the hierarchical order for all elements in the assembly

For two consecutive hierarchy assessments, including only the top 2 elements, we write:

$$p = \frac{\left(n - \left[\frac{0.5^{n-1}}{0.5^{m_1}} + \frac{0.5^{n-2}}{0.5^{m_2}}\right]\right)!}{n!}$$

We can now generalize this to the rest of the elements, and write:

$$p = \left(n - \sum_{i=1}^{k=n} \frac{0.5^{n-k}}{0.5^{m_{n-k}}}\right)! / n!$$

The index of m , which is the success rate for each hierarchy assessment, is $n - k$, which means that the first term is labeled as $n - 1$. It seems more natural to start at $n - n = 0$ or $n - (n - 1) = 1$, but at the end it doesn't matter in which direction we label the terms. It is not like we are forced to start at one side of the hierarchy or the other; the result is the same.

Note that this factorial is no longer the factorial of an integer, and so we need to use the gamma function to find its value. For non-complex integers we can use the Euler gamma function, which is:

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt$$

Here x is our positive real number input, and t is the bound variable used for the integration process.¹

Some calculators do not support non-integer factorials, because they go beyond the most basic definition of a factorial. Luckily the Google calculator does support non-integer factorials.²

After one element in perfect hierarchical order, we get:

$$p = \frac{\left(n - \frac{0.5^{n-1}}{0.5^{n-1}}\right)!}{n!} = \frac{(n-1)!}{n!}$$

If all are in perfect hierarchical order, we get:

$$p = \frac{(n-n)!}{n!} = \frac{1}{n!}$$

When there is only one element, then we get $n! = 1$ arrangement, and a p -value of $p = 1$.

The only problematic case is the perfectly random one, where $m = 0$. In this case we should get $p = 0.5$, but in reality, we get 1 for $n = 1$, $1/2$ for $n = 2$, $-1/6$ for $n = 3$, and so on. It is no wonder that a formula made for assessing the p -value of a hierarchy breaks down when there is no hierarchy. However, p approaches $p = 0.5$ as m drops, and so from that we can extrapolate that the complete absence of any hierarchy implies $p = 0.5$. So, we can define that:

$$\lim_{m \rightarrow 0} f(m) = 0.5$$

This is not very strange, as the gamma function, much like this function, which relies on it, also only applies to non-zero real numbers.

Hierarchy coherence

We can use base-2 logarithms to define a percent measure for how perfect a hierarchy is. For example, the odds for the hierarchy we observe in G -values is 0.00000000001%, while the ‘hierarchy coherence’ is:

$$q_{\log} = \frac{\log_2(9.9601045 \times 10^{15})}{\log_2(28!)} = \frac{53.1450823022344}{97.94419575123734} = 0.54$$

Or 54%. So, it is a hierarchy that is 54% coherent or perfect.

As a general formula this is:

$$q_{\log} = \log_2 \left[\left(n - \sum_{i=1}^{k=n} \frac{0.5^{n-k}}{0.5^{m_{n-k}}} \right)! \right] / \log_2(n!)$$

The result does not change when switching to logarithms of different bases, as long as we use the same type of logarithm both for the numerator and the denominator. However, information theory would suggest we pick base-2 logarithms.

Reversed hierarchies

The p -values one obtains with this formula never exceed 50%. This makes sense, because 50% is random chance. However, when we are looking at a perfectly reversed hierarchy, we still get 50%, but the perfect reversal of a hierarchy seems just as unlikely to happen under the null hypothesis as a perfect agreement with the predicted hierarchy.

So, for values above $p = 0.1$, we have to calculate the odds for the reversed hierarchy as well to see if the reversed hierarchy has a lower p -value than the hierarchy. If it does, then the true p -value of the hierarchy is the p -value of the reversed hierarchy subtracted from 1:

$$p < \bar{p} \Rightarrow p = 1 - \bar{p}$$

When dealing with only one p -value, it seems more sensible to just write down the p -value of the reversed hierarchy when actually dealing with a reversed hierarchy. However, when conducting a meta-analysis, which may contain both data that follows the predicted hierarchy and data that follows the reversed hierarchy more closely, then it is important that all p -values are p -values associated with the predicted hierarchy.

Combining p -values using Fischer's method

When assessing the p -values of hierarchies, $p = 0.5$ corresponds to pure randomness. Now, when combining many such random results using Fischer's method, one would expect to very slowly move up to $p = 1$, which would be problematic, because $p = 1$ represents the inversed hierarchy. In reality this does not happen, because even under the null hypothesis, we don't expect to always get $p = 0.5$, after all, there is only 1 in $n!$ states that corresponds to $p = 0.5$. Instead, we expect to get values around $p = 0.5$, which means that the end result after many individual studies will also be at or around $p = 0.5$.

Note that for combining any number of p -values, we have to first calculate the chi-square score, which we obtain by summing up the logarithms of our individual p -values, and multiply by (-2) . Then we can use an online chi-square score calculator to find the p -value corresponding to our chi-square score and our degrees of freedom, which is double our number of p -values.³

p -values of Bernoulli trials

In binary trials with two possible outcomes, say heads and tails, or 1 and 0, the chance for a certain number of heads x and a certain number of tails y , is given by the binomial probability formula:

$$f(k, n, p) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

Where k is the success rate, n is the number of trials, and p is the probability for a single trial.

While the binomial probability formula will give us the correct probability for having x times heads and y times tails, that is not really what we want to know when we are looking at the probability for a null hypothesis. In that case we want the p -value to be 0.5 or 50%, when the number of heads and tails is equal, say when we get 5 times heads and 5 times tail. What the binomial probability formula tells us instead is that the probability is 24.6% but this probability is for the specific case $x = y = 5$, and not $x = y$ in general. It does not tell us about statistical significance. For that the random distribution $x = y$ should have a p -value of 0.5.

We do not care that x exactly equaling y even after a very large number of trials is pretty unlikely, we are only interested in the fact that $x = y$ is the perfectly random case. This perfectly random case is what gives the strongest support for the null hypothesis, which corresponds to $p = 0.5$.

It is rather obvious that the p -value of obtaining n times heads in a row is 0.5^n . Let us say that a theory predicts that outcome A (or heads) is more likely, and the null hypothesis predicts that both

A and B (or heads and tails) are equally likely. Intuitively we want to simply divide a half to the power of the number of successes, by a half to the power of the number of failures:

$$P = 0.5^s \div 0.5^f$$

However, this gives us $p = 1$ for the perfectly random distribution, which is not the desired $p = 0.5$. For that we should divide by 2. Yet, in the case that $f = 0$, we still expect $p = 0.5^n$, without a division by 2.

For example, with 2 successes and 1 failure, we do not expect the result to be:

$$p = \frac{0.5^2}{0.5^1} = 0.5$$

We expect it to be:

$$p = \frac{0.5^2}{0.5^1 \cdot 2} = 0.25$$

So that after another failure it is:

$$p = \frac{0.5^2}{0.5^2 \cdot 2} = 0.5$$

So, whenever there are both successes and failures, we shall divide by 2. This makes sure that the p -values of binary data collections are 0.5 when there is perfect agreement with the null hypothesis, just as it is the case with the aforementioned formula for the p -value of hierarchies.

Notes and references:

1. Davis, P. J. (1959). "Leonhard Euler's Integral: A Historical Profile of the Gamma Function". *American Mathematical Monthly*. 66 (10): 849–869. doi:10.2307/2309786. JSTOR 2309786. Archived from the original on 7 November 2012. Retrieved 3 December 2016.
2. Link: [<https://calculator.apps.chrome/>].
3. Link: [<https://www.socscistatistics.com/pvalues/chidistribution.aspx>].