PASSING A THRESHOLD IS A MATTER OF INFINITESIMALS

A WARNING FOR α VALUE 'ADDICTS'

Notwithstanding the cautionary notes recently released in multiple ASA papers, I fear that the still many α *values* 'addicts' will continue in their dichotomization attitude due to the difficulty humans face in changing their habits. For the *significant / not significant* decision drawn from the comparison between a stated α value and a sample *p* value, the computational accuracy plays a critical role. For instance, in a unilateral comparison between two independent sample proportions, the observed difference

1. is deemed significant at the given α if the sample *p* is less than α
2. otherwise, it is deemed not significant

However, because *p values* are estimated with approximations, like when one leverages the asymptotic normality of the distribution of the difference, *significant* becomes a fuzzy term when the size of the discrepancy between α and the *p* falls within the numerical approximation entailed by the assumptions and the algorithms involved in computing *p*. Therefore, stating that the predefined α is greater or lesser with respect to the observed *p,* is not a "certain" or "true" statement, inasmuch as its computational approximations spreads a foggy cloud around the discrepancy between the two. The consideration of *coverage probabilities* is commonplace in building confidence intervals around a proportion in small samples. For example, with a sample of 10 cases and a proportion of 0.95, one finds that the distribution $Bin(10, 0.95)$ is far from normal. Therefore, the coverage probability differs from the nominal one taken from the asymptotic normal distribution. However, in the *comparison of proportions* between two samples, questioning the numerical accuracy of *p* is a bit difficult and thus less common. The binomial distribution can be taken as a typical case because it has a finite range, while the asymptotic normal distribution enjoys infinite tails. Therefore, at least when tests boil down to a unilateral or bilateral *tail problem*, it happens that something actually exceeding the distribution's range is included in the calculation of *p.*

This paper is structured as follows

1. *construction of the "true distribution"* of the difference d between two binomial proportions in two independent samples of *10* cases with different expected proportions by
    a. directly computing the differences from the scalar product of the two binomial distributions
    b. computing their joint probability as the product of the two marginal probabilities
2. *evaluation of the probability of a difference d exceeding a stated δ, based on the (cumulative) distribution of the difference*. Therefore, the question to be answered is "what is the probability that in sample's replicates the observed difference d of the two proportions happens to be bigger than δ?"
3. *evaluation of the same under the (cumulative) asymptotic normal approximation*
4. *comparison between (2) and (3)*
5. *Conclusions and recommendations*

The true distribution of the difference between two proportions in independent samples. For building the binomial distribution of a proportion p in a sample of size *n*, you need the $n + 1$ binomial coefficients of a polynomial of degree n. When the sample size does not exceed a few dozen cases, leveraging the Stifel formula is a very effective way to iteratively compute binomial coefficients. For wider samples you can use the dedicated function in your package. The difference of the proportions between two independent samples is computed as a straightforward application of the SQL scalar product of the two tables of binomial coefficients. Now you need the joint distribution of the two sample's proportions. You start with the bivariate binomial distribution and multiply the difference by its probability. Independence implies that the joint probability amounts to the product of the terms of the polynomial $(p_1 + q_1)^n$ and $(p_2 + q_2)^n$. Neither new nor difficult.

Here you find an extract of the difference table for the case of the two samples $(n_1 = 10,\ p_1 = 0.50)$ and $(n_2 = 10,\ p_2 = 0.50)$. So, both samples show an expected proportion of 0.50. In the table, **deg**'s are the polynomial degrees, **coeff**'s are the binomial coefficients, **jointp** is the joint probability **pobs** and **pobs1** are the observable proportions in the first and second sample respectively and **diffp** is the observable proportion difference between the two samples. At the end of the table, where the role of deg's inverts, the table will symmetrically decrease (not shown). The last column, **cum**, is the (cumulative) distribution. An expert's eye would easily recognize the *convolution of the two marginals distributions*. In fact, convolutions span the bidimensional matrix along the principal diagonal in the case of sums and along the symmetric minor diagonal in the case of differences.

deg	coeff	deg1	coeff1	diff	jointp	pobs	pobs1	diffp	cum
0	1	10	1	-10	0.000000954	0	1	-1	0.000000954
0	1	9	10	-9	0.000009537	0	0.9	-0.9	0.00001049
1	10	10	1	-9	0.000009537	0.1	1	-0.9	0.000020027
0	1	8	45	-8	0.000042915	0	0.8	-0.8	0.000062943
1	10	9	10	-8	0.000095367	0.1	0.9	-0.8	0.00015831
2	45	10	1	-8	0.000042915	0.2	1	-0.8	0.000201225
1	10	8	45	-7	0.000429153	0.1	0.8	-0.7	0.000630379
0	1	7	120	-7	0.000114441	0	0.7	-0.7	0.00074482
2	45	9	10	-7	0.000429153	0.2	0.9	-0.7	0.00117397
3	120	10	1	-7	0.000114441	0.3	1	-0.7	0.00128841
2	45	8	45	-6	0.00193119	0.2	0.8	-0.6	0.0032196
3	120	9	10	-6	0.00114441	0.3	0.9	-0.6	0.00436401
0	1	6	210	-6	0.000200272	0	0.6	-0.6	0.00456429
1	10	7	120	-6	0.00114441	0.1	0.7	-0.6	0.00570869
4	210	10	1	-6	0.000200272	0.4	1	-0.6	0.00590897
0	1	5	252	-5	0.000240326	0	0.5	-0.5	0.00614929
1	10	6	210	-5	0.00200272	0.1	0.6	-0.5	0.008152
3	120	8	45	-5	0.00514984	0.3	0.8	-0.5	0.0133018

It is worthwhile to note that the joint probability is a function of the degrees of both polynomials, so that equal differences enjoy different probabilities because they come from different marginal probabilities due to the setting of polynomials' degrees.
Even in the simple case of equal proportions, using the normal asymptotic distribution leads to a different *p* with respect to that of the true distribution. The true distribution shows that under the hypothesis that $diff = 0$, the *p value* of a difference in proportions of more than 0.3 is $1 - 0.9423340 = 0.057666$

diff	cumulative
0.3	0.942340851

The normal asymptotic distribution would say 0.9101437, a 'bad' approximation, indeed. You wouldn't expect such a difference because a variable having a $Bin(0.5, 10)$ distribution is asymptotically normal and because the difference between two independent normal variables is indeed normal. Therefore, the true distribution and its normal asymptotic approximation should provide close estimates of *p* in both computing ways. This doesn't happen, however. The reason why is in a too "easy" switch to the asymptotic distribution when performing tests.

Things get worse when p and p_1 are far away from 0.50, like $Bin(10, 0.95)$ and $Bin(10, 0.90)$. Now the distribution of the difference is no more symmetric. The expected difference is of course 0.5. What is the probability of a difference of a least 0.6?

diff	cumulative
0.6	0.99999442

The probability of a difference beyond 0.6 is virtually 0. The normal approximation would say 0.994742377 with an approximation of around 0.005. However, though apparently small, it could make the passing of a given threshold seem to happen while it doesn't.

Conclusions and recommendations. Beyond the obvious recommendation of stopping dichotomization, I would suggest

1. A bit of caution in using those beautiful asymptotic convergence theorems in that they entail numerical inaccuracies in the calculation of *p*.
2. Follow the ongoing trend of explicit computation of statistical parameters leveraging the immense computing power now available