P-values as random variables

Halsey and colleagues[1] present illustrative simulations for a two-sample t-test, by obtaining random P-values that result from pairs of samples of different sizes N_1, N_2, assuming that the true standardized mean difference between two populations is 0.5. Their graphs and histograms (Figures 3,4 in Halsey et al) underscore substantial randomness in the P-values over repeated samples from the same populations. Indeed, P-values can be viewed as random variables with their respective distributions. For example, P-values derived from common continuous test statistics (such as Student's t) will have the cumulative distribution function (CDF)[2,3] given by

$$F_\gamma(p) = 1 - G_\gamma\left(G_0^{-1}(1-p)\right), \tag{1}$$

where $G_0(\cdot)$ and $G_\gamma(\cdot)$ denote the CDF of the test statistic under the null and the alternative hypotheses and γ is the noncentrality parameter, which in Halsey's et al. experiments is $\gamma = 0.5\sqrt{1/(1/N_1 + 1/N_2)}$. For example, $F_{\gamma=0.5\sqrt{1/(1/30+1/30)}}(0.05/2) = 0.48$, as in Halsey et al.

The probability density function (PDF) of a P-value for a fixed γ follows from differentiating $F_\gamma(\cdot)$, and gives

$$f_\gamma(p) = \frac{g_\gamma\left(G_0^{-1}(1-p)\right)}{g_0\left(G_0^{-1}(1-p)\right)}, \tag{2}$$

where $g_\gamma(\cdot)$ is the density that corresponds to the cumulative distribution $G_\gamma(\cdot)$.

The CDF inverse allows to sample random P-values as

$$P = 1 - G_0\left(G_\gamma^{-1}(U)\right), \tag{3}$$

where U is a uniform (0-1) random number. Thus, empirical histograms shown in Halsey's et al. can be reproduced by generating P-values directly, *without simulating the actual samples and computing the t-statistics*. CDF values, $F_\gamma(p)$, would give the expected proportion of P-values in a histogram that are smaller or equal to 'p'. The plot of the PDF, $f_\gamma(p)$, on top of a histogram would match its shape when the histogram is obtained using a large number of simulations. Furthermore, P-value variability can be assessed visually by simply plotting its density.

References

1. Halsey LG, Curran-Everett D, Vowler SL, Drummond GB. The fickle P value generates irreproducible results. Nat Methods. 2015;12(3):179–185.

2. Kuo CL, Vsevolozhskaya OA, Zaykin DV. Assessing the probability that a finding is genuine for large-scale genetic association studies. PLoS ONE. 2015;10(5):e0124107.

3. Zaykin DV, Zhivotovsky LA. Ranks of genuine associations in whole-genome scans. Genetics. 2005;171(2):813–823.