

Can I Take a Peek?

Continuous Monitoring of Online A/B Tests

Ramesh Johari
Stanford University
ramesh.johari@stanford.edu

Abstract

A/B testing is a hallmark of Internet services: from e-commerce sites to social networks to marketplaces, nearly all online services use randomized experiments as a mechanism to make better business decisions. Such tests are generally analyzed using classical frequentist statistical measures: p-values and confidence intervals. Despite their ubiquity, these reported values are computed under the assumption that the experimenter will not continuously monitor their test—in other words, there should be no repeated “peeking” at the results that affects the decision of whether to continue the test. On the other hand, one of the greatest benefits of advances in information technology, computational power, and visualization is precisely the fact that experimenters can watch experiments in progress, with greater granularity and insight over time than ever before.

We ask the question: if users will continuously monitor experiments, then what statistical methodology is appropriate for hypothesis testing, significance, and confidence intervals? We present recent work addressing this question. In particular, building from results in sequential hypothesis testing, we present analogues of classical frequentist statistical measures that are valid even though users are continuously monitoring the results.

Joint work with Leo Pekelis and David Walsh. (This work was carried out in part with Optimizely, a leading A/B testing platform.)

ACM Classification

Categories: G.3 [Probability and Statistics]

Subject descriptor: *Experimental design*

Keywords

Online A/B testing; sequential hypothesis testing; multiple hypothesis testing



Short Bio

Ramesh Johari is an Associate Professor at Stanford University and the Cisco Faculty Scholar in the School of Engineering, with a full-time appointment in the Department of Management Science and Engineering (MS&E), and courtesy appointments in the Departments of Computer Science (CS) and Electrical Engineering (EE). He is a member of the Operations Research group in MS&E, the Information Systems Laboratory in EE, and the Institute for Computational and Mathematical Engineering. He is faculty director of the Social Algorithms Lab (SOAL). He received an A.B. in Mathematics from Harvard (1998), a Certificate of Advanced Study in Mathematics from Cambridge (1999), and a Ph.D. in Electrical Engineering and Computer Science from MIT (2004). In 2012-2013 he was on leave from Stanford at oDesk (now Elance-oDesk), first as a Consulting Scientist, then as Director of Data Products and Research. He also serves as an advisor to several startups in addition to Elance-oDesk, including Optimizely, a leading web optimization platform.