Rocks v Gold, or Estimating True Lift by Logging Losses

November 24, 2015

1 Notation

Response Rates: R^1_{tw} denotes response rate of test-group winner, *if* they were exposed to an ad, and R^0_{tw} is the response rate of the test-group winner, *if* they were *not* exposed to an ad. Similarly define R^1_{tl} and R^0_{tl} for the test-group losers.

Also let R^0_t, R^1_t denote the overall response rates for the test group, if none and all were exposed to ads, respectively.

For the control group there is no question of winning or losing a bid, so we just use R^1_c and R^0_c to denote the response-rate of a control user, if they were and were not exposed to an ad.

We know control users are not exposed to ads, but it's useful to think of a hypothetical (so-called *counterfactual*) scenario where we ask how would the control user have responded *if* they had seen an ad.

2 Formulation

Since test and control group assignments are completely random, and occur *after* targeting and matching, we can assert that:

$$R_t^0 = R_c^0, \tag{1}$$
$$R_t^1 = R_c^1, \tag{2}$$

where we note that *only* R_c^0 is directly observable, and the rest are hypotheticals (counterfactuals). To see why, note that R_t^0 is the (hypothetical) response rate if a test user were *not* exposed to an ad; because of the auction dynamics there *will* be test users not exposed to ads, however R_t^0 is *not* simply the average response rate of the un-exposed test users, because of selection bias introduced by the auction process. In this sense R_t^0 is not *directly* observable. A similar reason applies to the other counterfactuals.

The *true lift* we want to measure is:

$$L_{true} = (R_{tw}^1 \ / \ R_{tw}^0) - 1, \tag{3}$$

i.e., how much higher is the test group winner's conversion rate when *exposed* to an ad, compared to their response rate *if they had not been exposed* to an ad. So far we have been implicitly using fact (1) combined with a tacit assumption that test winners and losers are equivalent in their (un-exposed) response rates, i.e., $R_{tw}^0 = R_{tl}^0 = R_t^0$, and using $R_c^0 = R_t^0$ as a proxy for R_{tw}^0 in eq (3), resulting in this *naive* lift estimate:

$$L_n = R_{tw}^1 / R_c^0 - 1 \tag{4}$$

However we are suspecting that our assumption that $R_{tw}^0 = R_{tl}^0$ is faulty, and this has lead us to the idea of *logging test losses*, to observe the un-exposed response rate of test-group losers, R_{tl}^0. Below we show how the hypothetical response-rate of *unexposed test-group winners*, R_{tw}^0, can be derived from the observed R_{tl}^0, which allows us to compute true lift. In particular we are hoping to show that $R_{tl}^0 > R_{tw}^0$, which has come to be known as the **rock vs gold hypothesis**, i.e. test losers will tend to have a higher unexposed conversion-rate than test winners, because we are more likely to win ”rocks” (low-converters) rather than ”gold” (high-converters). The aim of this note is to derive a formula for true lift based on the conversion-rate of unexposed test-losers, and also to quantify how large of an effect to expect.

3 Derivation

Note that if w is the win-rate in the test-group (i.e. a fraction w of the test-group actually see an ad), then we can say:

$$R_t^0 = wR_{tw}^0 + (1-w)R_{tl}^0, \tag{5}$$

where R_{tl}^0, the response rate of test-group losers (under no-ad exposure), and is directly observable **if we log losses**. Also we noted in (1) that $R_t^0 = R_c^0$ so we solve (5) for the hypothetical R_{tw}^0 to get:

$$\begin{aligned} R_{tw}^0 &= [R_t^0 - (1-w)R_{tl}^0]/w && (6) \\ &= [R_c^0 - (1-w)R_{tl}^0]/w. && (7) \end{aligned}$$

Now we can use (3) to compute true lift in terms of observable quantities:

$$L_{true} = \frac{wR_{tw}^1}{R_c^0 - (1-w)R_{tl}^0} - 1. \tag{8}$$

It will be useful to consider the *ratios* $F_w = R_{tw}^0/R_c^0$ and $F_l = R_{tl}^0/R_c^0$. These represent the unexposed test-winner and test-loser response rates in units of the control response-rates. In terms of these ratios we re-write (7) as:

$$F_w = [1 - (1-w)F_l]/w, \tag{9}$$

so in terms of the naive lift $L_n = R^1_{tw}/R^0_c - 1$ and the ratio F_l we can write the true lift formula as:

$$\begin{aligned} L_{true} &= (R^1_{tw} \ / \ R^0_{tw}) - 1 && (10)\\ &= (R^1_{tw}/R^0_c) \ / \ (R^0_{tw}/R^0_c) - 1 && (11)\\ &= (1 + L_n) \ / \ F_w - 1 && (12)\\ &= \tfrac{w(1+L_n)}{1-(1-w)F_l} && (13) \end{aligned}$$

If we ignore win-bias, i.e. assume that $R^0_c = R^0_t = R^0_{tl}$, then we have $F_l = 1$ and this formula reduces to our original naive lift estimate (4).

To get a better sense of the factor F_l, solve above to get

$$F_l = \left(1 - \frac{w(1+L_n)}{(1+L_{true})}\right)\frac{1}{1-w}, \tag{14}$$

This form allows us to answer how much larger must the unexposed test-loser response-rate be than the control conversion rate, if we want to turn a naive negative lift into a true positive lift? For example for $L_n = -20\%$ and $w = 5\%$, and $L_{true} = 5\%$, the factor F_l is

$$\begin{aligned} F_l &= (1/0.95) * (1 - 0.05 * (1 - 0.20)/(1 + 0.05)) && (15)\\ &= 1.0125, && (16) \end{aligned}$$

i.e. the test-loser (unexposed) response rate is around 1.25% higher than the control conversion rate. If we repeat the above calculation with $L_{true} = 20\%$, we get $F_l = 1.0175$, i.e. to turn a naive negative lift of -20% into a true positive 20% lift, the test-loser conversion rate need only be 1.75% larger than the control conversion rate. This gives a rough idea of how much effect we should be looking for, when trying to compute true lift by logging losing bids.