Choice of the Randomization Unit in Online Controlled Experiment

Shaojie Deng* Roger Longbotham† Toby Walker‡ Ya Xu§

September 8, 2010
This version: May 10, 2011

Abstract

Controlled experiment has been used widely to support data driven decision making for online businesses. By applying appropriate randomization of the experiment units, causal inference can be established. The choice of the experiment unit for randomization can vary. User and page view are two mostly used units. Moreover, the analysis unit is sometimes different from the experiment unit. There are pros and cons in choosing which experiment unit to use and the choice affects the downstream statistical analysis. In this paper, we compare the two experiment units and highlight the differences in related statistical analysis.

Keywords

Controlled experiment, Experimentation, A/B testing, randomization unit, Variance estimation.

1 Introduction

[Todo] Introduction section.

The following paper is organized as follows. We firt introduce notation in Section 1.1. In Section 2, we first review the two sample t-test for metrics using user as the corresponding analysis unit. We then show delta method should be used when analysis unit is finer than user level and also gives a formula for the bias introduced should we fail to use delta method. In Section 3, we change gear to the case that the randomization unit is page view. In particular, we present the correct variance formula for the two layer randomization framework. In Section 4, we discuss the pros and cons of different randomizaiton units. Section 5 presents empirical results(or simulation results). Section 6 summarizes and concludes.

1.1 Notation

Before going into the next section, we introduce notation and assumptions which will be consistently used through out this paper. Denote n the total number of

*Microsoft, Redmond, WA 98052, USA, email: `alexdeng@microsoft.com`.
†Microsoft, Redmond, WA 98052, USA, email: `roger.longbotham@microsoft.com`.
‡Microsoft, Redmond, WA 98052, USA, email: `toby.walker@microsoft.com`.
§Microsoft, Redmond, WA 98052, USA, email: `yaxu@microsoft.com`.

unique users. Let $X_{i,j}$ be the per-page measurement (e.g. number of clicks on ths page) on user i's j^{th} page view and $X_{i,j}$ has mean μ_i and variance σ_i^2. Denote K_i the total number of page views from user i and $N = \sum_{i=1}^n K_i$ be the total number of page views. We assume for any i, $X_{i,j}, j = 1, \ldots, K_i$ are i.i.d. and uniformly bounded above by some finite constant. But we allow (μ_i, σ_i^2) to differ from user to user. We also assume $K_i, i = 1, \ldots, n$ are i.i.d. and independent of $(\mu_i, \sigma_i^2), i = 1, \ldots, n$. This last assmption may not be true in practice and need to be checked case by case. We have checked this assumption for some key metrics of web experiments using empirical data and this assumption is reasonable.

2 User as Randomization Unit

2.1 User Level Metrics

This is the simple case. Brifely introduce the two sample t-test. Introduce necessary notations for following sections.

2.2 Page Level Metrics

A page level metric can be denoted by:

$$\overline{X} = \frac{\sum_{i=1}^n \sum_{j=1}^{K_i} X_{i,j}}{N}.$$

To estimate the varaince of $\overline{X}$, it is tempting to treat page level metrics $X_{i,j}$, $j = 1, \ldots, K_i$, $i = 1, \ldots, n$, as i.i.d. and $\overline{X}$ under this assumption is an average of i.i.d. samples so the variance of $\overline{X}$ can be easily estimated by

$$\frac{1}{N^2}\Big(\sum_{i=1}^n \sum_{j=1}^{K_i} (X_{i,j} - \overline{X})^2\Big).$$

This estimator, which we call the *naive* estimator, is not consistent because unlike the fixed effect model, where the only randomness is from the noise of $X_{i,j}$. In our model the user effect (μ_i, σ_i^2) are also a random sample from a distribution. Nevertheless, it is true in our model that the user level measurement $(\sum_{i=1}^{K_i} X_{i,j}, K_i), i = 1, \ldots, n$ are i.i.d. By letting $Y_i = \sum_{i=1}^{K_i} X_{i,j}$ and express $\overline{X}$ as $\sum_{i=1}^n Y_i / \sum_{i=1}^n K_i$, it is then a straightforward application of the delta method to get an asymptotically consistent estimator for $\mathbb{V}ar\overline{X}$:

$$\frac{1}{n}\Big\{\frac{1}{\widehat{\mathbb{E}K_i}^2}\widehat{\mathbb{V}arY_i} + \frac{\widehat{\mathbb{E}Y_i}^2}{\widehat{\mathbb{E}K_i}^4}\widehat{\mathbb{V}arK_i} - 2\frac{\widehat{\mathbb{E}Y_i}}{\widehat{\mathbb{E}K_i}^3}\widehat{Cov(Y_i, K_i)}\Big\}$$

where these "hatted" quantities are the sample mean, variance or covariance.

For asymptotic analysis, we will let $n \to \infty$ (so $N \to \infty$ a.s.). To normalize the naive estimator and delta method estimator, we multiply them by n so that they will converge to some nonzero numbers. We introduce the normalized naive estimator

$$\widehat{\sigma_n^2} = n\frac{1}{N^2}\Big(\sum_{i=1}^n \sum_{j=1}^{K_i} (X_{i,j} - \overline{X})^2\Big). \tag{1}$$

and the normalized delta method estimator

$$\widehat{\sigma_d^2} = \frac{1}{\widehat{\mathbb{E}K_i}^2}\widehat{\mathbb{V}arY_i} + \frac{\widehat{\mathbb{E}Y_i}^2}{\widehat{\mathbb{E}K_i}^4}\widehat{\mathbb{V}arK_i} - 2\frac{\widehat{\mathbb{E}Y_i}}{\widehat{\mathbb{E}K_i}^3}\widehat{Cov(Y_i, K_i)} \tag{2}$$

A natural question to ask is how biased is the naive estimator $\widehat{\sigma_n^2}$ relative to the true normalized variance $n\mathbb{V}ar\overline{X}$. This is answered in the following theorem.

Theorem 1. *Let $C = \frac{\mathbb{E}K_i^2}{(\mathbb{E}K_i)^2}$. Then, as $n \to \infty$,*

$$n\mathbb{V}ar\overline{X} \to C\mathbb{V}ar(\mu_i) + \mathbb{E}(\sigma_i^2)/\mathbb{E}(K_i) \tag{3}$$

$$\widehat{\sigma_d^2} \to C\mathbb{V}ar(\mu_i) + \mathbb{E}(\sigma_i^2)/\mathbb{E}(K_i) \tag{4}$$

$$\widehat{\sigma_n^2} \to \frac{1}{\mathbb{E}(K_i)}(\mathbb{V}ar(\mu_i) + \mathbb{E}(\sigma_i^2)). \tag{5}$$

Let $\rho := \mathbb{V}ar(\mu_i)/(\mathbb{V}ar(\mu_i) + \mathbb{E}(\sigma_i^2))$ be the user effect coefficient(variances that explained by between user variation), then

$$\frac{n\mathbb{V}ar(\overline{X})}{\widehat{\sigma_n^2}} \to (\mathbb{E}(K_i)C - 1)\rho + 1. \tag{6}$$

The convergence in (4) and (5) are in probability.

Proof of Theorem 1. (4) follows directly from the property of the delta method. To prove (3), we first apply conditional variance formula by conditioning on $(\mu_i, \sigma_i^2, K_i, i = 1, \dots, n)$. This gives

$$\begin{aligned}\mathbb{V}ar\overline{X} = \mathbb{V}ar\Big(\mathbb{E}\Big(\frac{\sum_{i=1}^n \sum_{j=1}^{K_i} X_{i,j}}{N}\Big| K_i, \mu_i, \sigma_i^2, i = 1, \dots, n\Big)\Big)\\ +\mathbb{E}\Big(\mathbb{V}ar\Big(\frac{\sum_{i=1}^n \sum_{j=1}^{K_i} X_{i,j}}{N}\Big| K_i, \mu_i, \sigma_i^2, i = 1, \dots, n\Big)\Big)\\ =\mathbb{V}ar\Big(\frac{1}{N}\sum_{i=1}^n K_i\mu_i\Big) + \mathbb{E}\Big(\frac{1}{N^2}\sum_{i=1}^n K_i\sigma_i^2\Big).\end{aligned}$$

Let $w_i = K_i/\sum_{i=1}^n K_i = K_i/N$. Since K_i independent of (μ_i, σ_i^2) and $N/n \to \mathbb{E}K_i$ as $n \to \infty$, we can further simplify the right hand. First, by applying iterative expectation(frist conditioning on $w_1, \dots, w_n$), we have

$$n\mathbb{E}\Big(\frac{1}{N^2}\sum_{i=1}^n K_i\sigma_i^2\Big) = \sum_{i=1}^n \mathbb{E}\Big(\frac{n}{N}w_i\sigma_i^2\Big) = \frac{1}{\mathbb{E}K_i}(\sum_{i=1}^n w_i)\mathbb{E}\sigma_i^2 = \frac{\mathbb{E}\sigma_i^2}{\mathbb{E}K_i} \tag{7}$$

where the second equality is by bounded convergence theorm(since $N/n \to \mathbb{E}K_i$ and $\sum w_i\sigma_i^2$ bounded) and the last equation is from $\sum w_i = 1$. Since $(\mu_i, \sigma_i^2), i = 1, \dots, n$ are i.i.d.,

$$n\mathbb{V}ar(\sum_{i=1}^n w_i\mu_i) = n\mathbb{E}(\mathbb{V}ar(\sum_{i=1}^n w_i\mu_i|w_1, \dots, w_n)) + n\mathbb{V}ar(\mathbb{E}(\sum_{i=1}^n w_i\mu_i|w_1, \dots, w_n)) \tag{8}$$

$$= n\mathbb{E}(\sum_{i=1}^n w_i^2\mathbb{V}ar(\mu_i)) + n\mathbb{V}ar((\sum_{i=1}^n w_i)\mathbb{E}\mu_i) = n\mathbb{E}(\sum_{i=1}^n w_i^2)\mathbb{V}ar(\mu_i) \tag{9}$$

where the last equality is from the fact that the second term is 0. By simple algebra, $n\sum_{i=1}^{n} w_i^2 = \frac{\overline{K_i^2}}{\overline{K_i}\times\overline{K_i}}$, where $\overline{K_i^2}$ and $\overline{K_i}$ are sample mean of K_i^2 and K_i, respectively. By strong law of large number, $\overline{K_i^2} \to \mathbb{E}K_i^2$ a.s., $\overline{K_i} \to \mathbb{E}K_i$ a.s., therefore $n\sum_{i=1}^{n} w_i^2 = \frac{\mathbb{E}K_i^2}{(\mathbb{E}K_i)^2}$ a.s. Combine this result with (7) and (9), we've proved (3).

We now turn to the limit of $\widehat{\sigma_n^2}$.

$$\widehat{\sigma_n^2} = n\frac{1}{N^2}\Big(\sum_{i=1}^{n}\sum_{j=1}^{K_i}(X_{i,j}-\overline{X})^2\Big) = \frac{n}{N^2}\Big\{\sum_{i=1}^{n}\sum_{j=1}^{K_i}X_{i,j}^2 - N\overline{X}^2\Big\}$$
$$\to \lim_{n\to\infty}\Big(\frac{n^2}{N^2}\Big)\mathbb{E}\big(\sum_{j=1}^{K_i}X_{i,j}^2\big) - \lim_{n\to\infty}\Big(\frac{n}{N}\Big)(\mathbb{E}\mu_i)^2.$$

The last limit is from $(1/n)\sum_{j=1}^{K_i}X_{i,j}^2 \to \mathbb{E}\big(\sum_{j=1}^{K_i}X_{i,j}^2\big)$ and $\overline{X} \to \mathbb{E}\mu_i$ a.s., both by the strong law of large number. Using bounded convergence theroem and $N/n \to \mathbb{E}K_i$, and also $\mathbb{E}\big(\sum_{j=1}^{K_i}X_{i,j}^2\big) = \mathbb{E}K_i\mathbb{E}X_{i,j}^2 = \mathbb{E}K_i(\mathbb{E}\mu_i^2+\mathbb{E}\sigma_i^2)$, (5) follows. □

3 Page View as Randomizaiton Unit

3.1 A two layer randomization framework

Now we consider the case that randomization is done by page view. Suppose ALL page views are randomly divided into different groups. By ALL page views we mean page views from all users that could show up. Under this framework, it is from the typical marginalization argument that we can treat page view level measurement as i.i.d. i.e., there is no user effect in the analysis because the page views are drawn from all users and no user selection variance is induced in this randomization scheme. Since page view level measurements are i.i.d., all statistical analysis no finer than page view level is therefore straightforward.

The case that is of more interest is the following. We do not want to perform experiment on all users. Instead, we first randomly selected n user from all the users in the universe. n is usually only a small percentage of the total number M of users in the universe so let us assume M is infinity and hence the users are drawn independently. All the page views from these n users are then randomly split into control and treatment. The goal is to make inference by comparing some metrics in control and treatment.

We inherit all the notations from Part I. Suppose we are interested in page view level metrics $\overline{X}_r = \sum_{i=1}^{n}\sum_{j=1}^{{K_i}^{(r)}} X_{i,j}^{(r)}/N_r$ where $r = 1,2$ stands for control and treatment. In Part I, we never consider the variance of both control and treatment together. This is because the control and treatment groups have different users and since randomization is based on user, the metrics of the two groups are therefore independent. As a result the variance of the metrics difference is simply the sum of the two variances of metric in each group. What make things more complicated is that the current framework, control and treatment share the same group of n users. It is the page view, not the user that is randomized into two groups. Due to this very fact, $\overline{X}_1$ and $\overline{X}_2$ are no longer independent.

3.2 Variance formula

In this section we give the asymptotically unbiased estimator for $\mathbb{V}ar(\overline{X}_1 - \overline{X}_2)$ when the page views are split into control and treatment with fixed weights. For simplicity, we first assume control and treatment have the same weights. But we will give the result for general case in the end of this Section. Under the same assumption as in Part I that number of page views K_i are independent of (μ_i, σ_i^2). Then conditioned on K_i, K_i^r follows from $binomial(K_i, 0.5)$ distribution. If we only consider one group, say control. Then the only difference between this framework and that of Section 2 is that now K_i^r follows from a different distribution(from K_i). But note that all the result in Section 2 does not depend on the distribution of K_i. Therefore all results in Section 2 directly apply on $\overline{X}_1$ (or $\overline{X}_2$).

Proposition 2. *Let* $w_i^{(r)} = K_i^{(r)}/\sum_{i=1}^n K_i^{(r)}$. *Assume* $n\mathbb{E}(\sum_{i=1}^n (w_i^{(r)})^2) \to C$, $r = 1, 2$ *as* $n \to \infty$. *Then for* $r = 1, 2$

$$\mathbb{E}(\widehat{\sigma^2_{nr}}) \to \frac{1}{\mathbb{E}(K_i^{(r)})}(\mathbb{V}ar(\mu_i) + \mathbb{E}(\sigma_i^2)) \tag{10}$$

$$\mathbb{E}(\widehat{\sigma^2_{dr}}) \to C\mathbb{V}ar(\mu_i) + \mathbb{E}(\sigma_i^2)/\mathbb{E}(K_i^{(r)}). \tag{11}$$

What Proposition 2 says is exactly that if we apply naive formula or delta method formula to one group, we will get asymptotically unbiased estimator for the right hand side of (10) and (11), respectively.

To analyze $\mathbb{V}ar(\overline{X}_1 - \overline{X}_2)$, we begin by using the basic formula (**??**), with a little bit modification.

$$\begin{aligned}
\mathbb{V}ar(\overline{X}_1 - \overline{X}_2) &= \mathbb{V}ar\Big(\frac{\sum_{i=1}^n \sum_{j=1}^{K_i^{(1)}} X_{i,j}^{(1)}}{N_1} - \frac{\sum_{i=1}^n \sum_{j=1}^{K_i^{(2)}} X_{i,j}^{(2)}}{N_2}\Big) \\
&= \mathbb{V}ar\Big(\mathbb{E}\Big(\frac{\sum_{i=1}^n \sum_{j=1}^{K_i^{(1)}} X_{i,j}^{(1)}}{N_1} - \frac{\sum_{i=1}^n \sum_{j=1}^{K_i^{(2)}} X_{i,j}^{(2)}}{N_2} | K_i^{(r)}, \mu_i^{(r)}, \sigma_i^{(r)}, i = 1, \ldots, n, r = 1, 2)\Big)\Big) \\
&+ \mathbb{E}\Big(\mathbb{V}ar\Big(\frac{\sum_{i=1}^n \sum_{j=1}^{K_i^{(1)}} X_{i,j}^{(1)}}{N_1^2} - \frac{\sum_{i=1}^n \sum_{j=1}^{K_i^{(2)}} X_{i,j}^{(2)}}{N_2^2} | K_i^{(r)}, \mu_i^{(r)}, \sigma_i^{(r)}, i = 1, \ldots, n, r = 1, 2\Big)\Big) \\
&= \mathbb{V}ar\Big(\frac{1}{N_1}\sum_{i=1}^n K_i^{(1)}\mu_i - \frac{1}{N_2}\sum_{i=1}^n K_i^{(2)}\mu_i\Big) + \mathbb{E}\Big(\frac{1}{N_1^2}\sum_{i=1}^n K_i^{(1)}\sigma_i^2 + \frac{1}{N_2^2}\sum_{i=1}^n K_i^{(2)}\sigma_i^2\Big)
\end{aligned} \tag{12}$$

By using the short hand notation $w_i^{(r)}$, we can simplify $n\mathbb{V}ar(\overline{X}_1 - \overline{X}_2)$ into

$$n\mathbb{V}ar\Big(\sum_{i=1}^n (w_i^{(1)} - w_i^{(2)})\mu_i\Big) + n\mathbb{E}\Big(\sum_{i=1}^n (w_i^{(1)}/N_1 + w_i^{(2)}/N_2)\sigma_i^2\Big) \tag{13}$$

Comparing to (7), we see

$$n\mathbb{E}\Big(\sum_{i=1}^n (w_i^{(1)}/N_1 + w_i^{(2)}/N_2)\sigma_i^2\Big) \to \frac{\mathbb{E}\sigma_i^2}{\mathbb{E}K_i^{(1)}} + \frac{\mathbb{E}\sigma_i^2}{\mathbb{E}K_i^{(2)}} = 2\frac{\mathbb{E}\sigma_i^2}{\mathbb{E}K_i^{(r)}} \tag{14}$$

where the last term is because $K_i^{(1)}$ has the same distribution as $K_i^{(2)}$.

By using conditional variance formula for another time and following the exact same argument as in (9) (replace w_i by $(w_i^{(1)} - w_i^{(2)})$), we have

$$n\mathbb{V}ar\Big(\sum_{i=1}^{n}(w_i^{(1)} - w_i^{(2)})\mu_i\Big) = n\mathbb{E}\big(\sum_{i=1}^{n}(w_i^{(1)} - w_i^{(2)})^2\mathbb{V}ar\mu_i\big)$$

$$= \Big(n\mathbb{E}\big(\sum_{i=1}^{n}(w_i^{(1)})^2\big) + n\mathbb{E}\big(\sum_{i=1}^{n}(w_i^{(2)})^2\big) - 2n\mathbb{E}\big(\sum_{i=1}^{n} w_i^{(1)}w_i^{(2)}\big)\Big)\mathbb{V}ar\mu_i \qquad (15)$$

Assuming for $r = 1, 2$, $n\mathbb{E}\big(\sum_{i=1}^{n}(w_i^{(r)})^2\big) \to C$, $\mathbb{E}\big(\sum_{i=1}^{n} w_i^{(1)}w_i^{(2)}\big) \to C_x$ as $n \to \infty$, then

$$n\mathbb{V}ar\Big(\sum_{i=1}^{n}(w_i^{(1)} - w_i^{(2)})\mu_i\Big) \to 2(C - C_x)\mathbb{V}ar\mu_i \qquad (16)$$

Combining this with (14), we have proved the following.

Proposition 3. *Under the framework of this section, assuming for* $r = 1, 2$, $n\mathbb{E}\big(\sum_{i=1}^{n}(w_i^{(r)})^2\big) \to C$, $\mathbb{E}\big(\sum_{i=1}^{n} w_i^{(1)}w_i^{(2)}\big) \to C_x$ *as* $n \to \infty$,

$$n\mathbb{V}ar(\overline{X}_1 - \overline{X}_2) \to 2(C - C_x)\mathbb{V}ar\mu_i + 2\frac{\mathbb{E}\sigma_i^2}{\mathbb{E}K_i^{(r)}} \qquad (17)$$

What is $C - C_x$? As we have seen before, $C = \mathbb{E}(K_i^{(r)})^2/(\mathbb{E}K_i^{(r)})^2$ (replace K_i by $K_i^{(r)}$ in Proposition **??**). A somewhat striking result shows that $C - C_x = \frac{1}{\mathbb{E}K_i^{(r)}}$, which means that from Proposition 2, the naive estimator $\widehat{\sigma^2_{nr}}$ actually gives an asymptoticly unbiased estimate for $n\mathbb{V}ar(\overline{X}_1 - \overline{X}_2)$ simply by multiplying itself by a factor of 2!

Proposition 4.

$$C - C_x = \frac{1}{\mathbb{E}K_i^{(r)}}.$$

Proof.

$$C - C_x = \lim_{n\to\infty} n\sum_{i=1}^{n}(w_i^{(1)})^2 - n\sum_{i=1}^{n} w_i^{(1)}w_i^{(2)} = \lim_{n\to\infty}\left\{\frac{\overline{(K_i^{(1)})^2}}{\overline{K_i^{(1)}} \times \overline{K_i^{(1)}}} - \frac{\overline{K_i^{(1)}K_i^{(2)}}}{\overline{K_i^{(1)}} \times \overline{K_i^{(2)}}}\right\}$$
$$= \frac{\mathbb{E}(K_i^{(1)})^2}{(\mathbb{E}K_i^{(1)})^2} - \frac{\mathbb{E}(K_i^{(1)}K_i^{(2)})}{(\mathbb{E}K_i^{(1)})^2},$$

where the last equality from strong law of large number and $K_i^{(1)}$ and $K_i^{(2)}$ have the same distribution. Note that $K_i = K_i^{(1)} + K_i^{(2)}$ where K_i is the total page views from user i and $K_i^{(1)}$ follows binomial distribution with $p = 1/2$.

$$\mathbb{E}((K_i^{(1)})^2) - \mathbb{E}(K_i^{(1)}K_i^{(2)}) = \mathbb{E}\big(K_i^{(1)}(K_i^{(1)} - K_i^{(2)})\big) = \mathbb{E}\Big(\mathbb{E}\big(K_i^{(1)}(2K_i^{(1)} - K_i)\big)\big|K_i\Big)$$
$$= \mathbb{E}\big(2\mathbb{E}((K_i^{(1)})^2|K_i) - K_i\mathbb{E}(K_i^{(1)}|K_i)\big) = \mathbb{E}\big(\frac{K_i}{2} + \frac{K_i^2}{2} - \frac{K_i^2}{2}\big)$$
$$= \mathbb{E}K_i/2 = \mathbb{E}K_i^{(r)}, \qquad r = 1, 2.$$

Combining the two parts, we have proved $C - C_x = 1/\mathbb{E}K_i^{(r)}$. $\square$

For the general case, suppose control has weight p and treatment weight q where $p+q=1$. Proposition 4 and Proposition 3 can be generalized into the following:

Proposition 5. *Suppose control has weight p and treatment weight q.*

$$
\begin{aligned}
&n\mathbb{E}\big(\sum_{i=1}^{n}(w_i^{(r)})^2\big) \to \frac{\mathbb{E}(K_i^{(r)})^2}{(\mathbb{E}K_i^{(r)})^2} = C_r, r=1,2 \\
&n\mathbb{E}\big(\sum_{i=1}^{n}(w_i^{(1)}w_i^{(2)})\big) \to \frac{\mathbb{E}(K_i^{(1)}K_i^{(2)})}{\mathbb{E}K_i^{(1)}\mathbb{E}K_i^{(2)}} = C_x \\
&n\mathbb{V}ar(\overline{X}_1-\overline{X}_2) \to (C_1+C_2-2C_x)\mathbb{V}ar\mu_i + \mathbb{E}\sigma_i^2\Big(\frac{1}{\mathbb{E}K_i^{(1)}}+\frac{1}{\mathbb{E}K_i^{(2)}}\Big).
\end{aligned}
$$

Moreover,

$$
C_1+C_2-2C_x = \frac{1}{\mathbb{E}K_i^{(1)}}+\frac{1}{\mathbb{E}K_i^{(2)}}.
$$

Therefore,

$$
n\mathbb{V}ar(\overline{X}_1-\overline{X}_2) \to \Big(\frac{1}{\mathbb{E}K_i^{(1)}}+\frac{1}{\mathbb{E}K_i^{(2)}}\Big)\Big(\mathbb{V}ar\mu_i+\mathbb{E}\sigma_i^2\Big).
$$

We now summarize the result into the following theorem.

Theorem 6. *Under the framework of this section, as $n\to\infty$*

$$
n\mathbb{V}ar(\overline{X}_1-\overline{X}_2) \to \Big(\frac{1}{\mathbb{E}K_i^{(1)}}+\frac{1}{\mathbb{E}K_i^{(2)}}\Big)\Big(\mathbb{V}ar\mu_i+\mathbb{E}\sigma_i^2\Big).
$$

Moreover $\widehat{\sigma^2_{n1}}+\widehat{\sigma^2_{n2}}$ is also an asymptotically unbiased estimator for $n\mathbb{V}ar(\overline{X}_1-\overline{X}_2)$.

We will denote $\widehat{\sigma^2_{n1}}+\widehat{\sigma^2_{n2}}$ as Formula P, where P stands for "randomization by page view".

Proof of Proposition 5. The proof is basically similar to the proof of Proposition 4 and 3. Here we show $C_1+C_2-2C_x = \frac{1}{\mathbb{E}K_i^{(1)}}+\frac{1}{\mathbb{E}K_i^{(2)}}$.

To see this, note that $K_i = K_i^{(1)}+K_i^{(2)}$ and $K_i^{(1)}$ follows $Binomial(K_i,p)$.

$$
\begin{aligned}
\mathbb{E}K_i^{(1)} &= p\mathbb{E}K_i \\
\mathbb{E}K_i^{(2)} &= q\mathbb{E}K_i \\
\mathbb{E}\big((K_i^{(1)})^2\big) &= pq\mathbb{E}K_i + p^2\mathbb{E}K_i^2 \\
\mathbb{E}\big((K_i^{(2)})^2\big) &= pq\mathbb{E}K_i + q^2\mathbb{E}K_i^2 \\
\mathbb{E}K_i^{(1)}K_i^{(2)} &= p\mathbb{E}K_i^2 - pq\mathbb{E}K_i - p^2\mathbb{E}K_i^2 = pq\mathbb{E}K_i^2 - pq\mathbb{E}K_i.
\end{aligned}
$$

By definition,

$$
\begin{aligned}
C_1+C_2-2C_x &= \frac{\mathbb{E}(K_i^{(1)})^2}{(\mathbb{E}K_i^{(1)})^2}+\frac{\mathbb{E}(K_i^{(2)})^2}{(\mathbb{E}K_i^{(2)})^2} - 2\frac{p\mathbb{E}K_i^2 - pq\mathbb{E}K_i - p^2\mathbb{E}K_i^2 = pq\mathbb{E}K_i^2 - pq\mathbb{E}K_i}{\mathbb{E}K_i^{(1)}\mathbb{E}K_i^{(2)}} \\
&= \frac{1}{(\mathbb{E}K_i)^2}\Big(\frac{1}{p^2}\mathbb{E}\big((K_i^{(1)})^2\big)+\frac{1}{q^2}\mathbb{E}\big((K_i^{(2)})^2\big)-\frac{2}{pq}\mathbb{E}K_i^{(1)}K_i^{(2)}\Big) \\
&= (q/p+p/q+2)\frac{1}{\mathbb{E}K_i} = (1/p+1/q)\frac{1}{\mathbb{E}K_i}.
\end{aligned}
$$

On the other hand,

$$\frac{1}{\mathbb{E}K_i^{(1)}} + \frac{1}{\mathbb{E}K_i^{(2)}} = (1/p + 1/q)\frac{1}{\mathbb{E}K_i}.$$

Hence $C_1 + C_2 - 2C_x = \frac{1}{\mathbb{E}K_i^{(1)}} + \frac{1}{\mathbb{E}K_i^{(2)}}$. □

4 Discussion

5 Simulation and Empirical Results

5.1 Randomization by User

5.2 Randomization by Page View

In this section, we again use PCR as example. For a fixed n, we first sample p_i, the click through rate for this user from a $Beta(0.1, 0.5)$ distribution. We then sample the total number of page view K_i from some distribution, which we can vary, and then use binomial distribution to split K_i into $K_i^{(1)}$ and $K_i^{(2)}$. We then simulate $\sum_{j=1}^{K_i^{(r)}} X_{i,j}^{(r)}$ from $Binomial(p_i)$. In each simulation run, we can get $\overline{X}_1 - \overline{X}_2$, as well as $\widehat{\sigma_n^2}$, $\widehat{C}$, $\widehat{C_x}$, $\widehat{\mathbb{E}K_i^{(r)}}$. For $\widehat{\sigma_n^2}$, $\widehat{C}$, $\widehat{\mathbb{E}K_i^{(r)}}$, we can actually calculate from both control and treatment and then take the average to get a more accurate estimate. We then repeat this step for 1000 times. After the 1000 simulation run, we can estimate $\mathbb{V}ar\overline{X}_1 - \overline{X}_2$ from the sample variances of the 1000 realizations of $\overline{X}_1 - \overline{X}_2$, which we denote by $\widehat{\sigma_{sim}^2}$ the normalized variances, which is n times the sample variance of $\overline{X}_1 - \overline{X}_2$. We also did bootstrap simulation(100 subsamples) to get an estimate of $SD(\widehat{\sigma_{sim}^2})$. On the other hand, for each of these 1000 simulation run, we can apply Formula P to estimate the normalized variance. We then compare the distribution of these 1000 estimates from Formula P to the 95% confidence interval $(\widehat{\sigma_{sim}^2} - 1.96SD(\widehat{\sigma_{sim}^2}), \widehat{\sigma_{sim}^2} + 1.96SD(\widehat{\sigma_{sim}^2}))$. In all the simulation, we fixed $n = 100,000$.

5.2.1 K_i Poisson case

We first use $Poisson(6)$ to generate K_i, and then $K_i^{(r)}, r = 1, 2$ from the binomial distribution. The left plot in Figure 1 shows that $C - C_x$ is indeed close to $1/\mathbb{E}K_i^{(r)}$. The ratio of the two is normally distributed and concentrated around 1. The right plot shows all the 1000 estimates from Formula P are within the bootstrpped confidence interval.

5.2.2 K_i constant case

In this simulation study, we fixed $K_1 = 5$. Figure 2 shows the performance in this case.

6 Conclusion

References

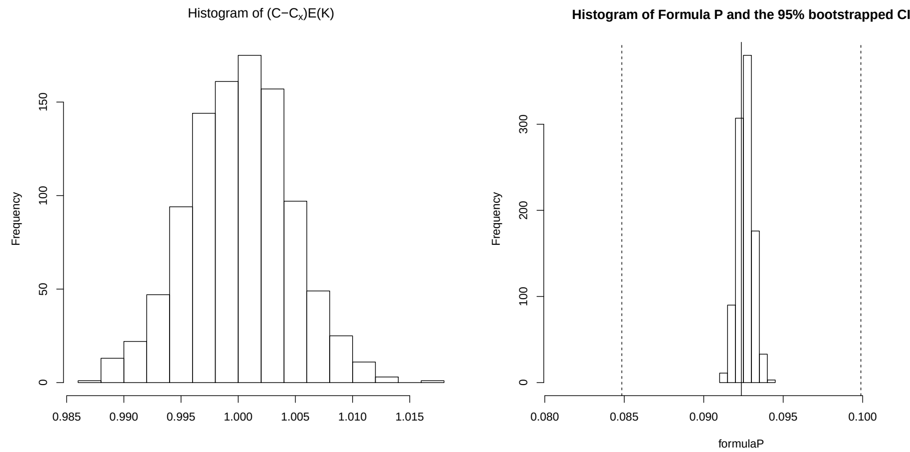

Figure 1: Left: Histogram of $(\widehat{C}-\widehat{C_x})\widehat{\mathbb{E}K_i^{(r)}}$. Right: Histogram of the 1000 estimates from Formula P and the 95% confidence interval form bootstrap. The two dashed lines are lower and upper bound of the confidence interval and the solid line is the sample variance of 1000 realization of $\overline{X}_1 - \overline{X}_2$ multiplied by n

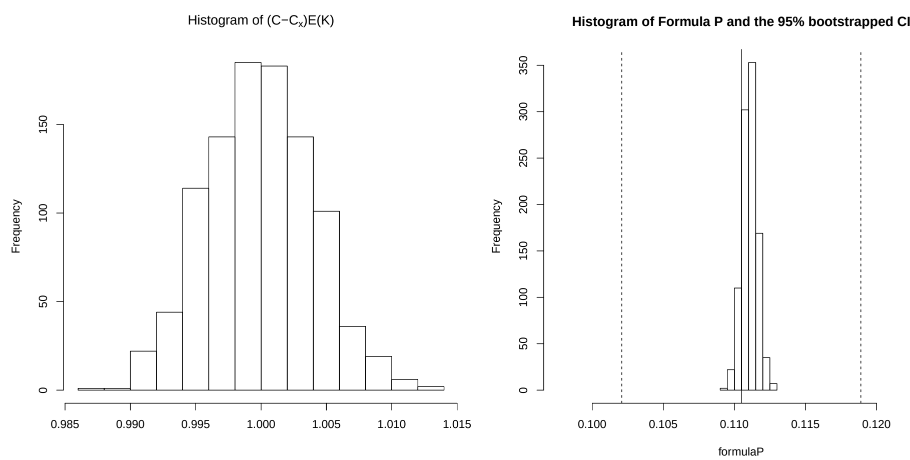

Figure 2: Left: Histogram of $(\widehat{C}-\widehat{C_x})\widehat{\mathbb{E}K_i^{(r)}}$. Right: Histogram of the 1000 estimates from Formula P and the 95% confidence interval form bootstrap. The two dashed lines are lower and upper bound of the confidence interval and the solid line is the sample variance of 1000 realization of $\overline{X}_1 - \overline{X}_2$ multiplied by n