

A FUNDAMENTAL PROBLEM OF HYPOTHESIS TESTING WITH FINITE INVENTORY IN E-COMMERCE

DENNIS BOHLE, ALEXANDER MARYNYCH, AND MATTHIAS MEINERS

ABSTRACT. We draw attention to a problem that is often overlooked or ignored by companies practicing hypothesis testing (A/B testing) in online environments. We show that conducting experiments on limited inventory that is shared between variants in the experiment can lead to high false positive rates since the core assumption of independence between the groups is violated. We provide a detailed analysis of the problem in a simplified setting whose parameters are informed by realistic scenarios. The setting we consider is a 2-dimensional random walk in a semi-infinite strip. It is rich enough to take a finite inventory into account, but is at the same time simple enough to allow for a closed form of the false-positive probability. We prove that high false-positive rates can occur, and develop tools that are suitable to help design adequate tests in follow-up work. Our results also show that high false-negative rates may occur. The proofs rely on a functional limit theorem for the 2-dimensional random walk in a semi-infinite strip.

1. INTRODUCTION

The golden standard for testing product changes on e-commerce websites is large scale hypothesis testing also known as A/B-Testing.

When a given version of a website is modified, it is natural to ask whether or not the modified (new) version of the website performs better than the old one. It is very common to use the following approach based on classic hypothesis testing:

During a fixed time period, the so-called testing phase, whenever customers visit the website, they are displayed one of the two versions of it, where the choice which one they get to see is random. For each version of the website the owner thus collects a sample containing for each customer visiting that website relevant data such as whether or not they bought a good or how much money was spent by the customers, etc. Then a statistical test (A/B test) is applied to evaluate which version of the website performed better.

Typically, these tests rely on the assumption of independent samples. In the present paper, we point out in a quantitative way that in the situation where there is a finite amount of a popular good the independence assumption is not feasible and can often lead to wrong conclusions. The inventory is shared between variants and if a copy of an item is sold it can not be bought by users that enter the experiment later. This implies that users are not independent both inside as well as between the variants. These dependencies could be avoided by randomly splitting on an item level instead of a user level, but this would reduce the choice of the customer and is therefore not a realistic setup.

Date: September 1, 2020.

2010 Mathematics Subject Classification. 62F03, 62E20, 60F17.

Key words and phrases. A/B test, conversion rate, functional limit theorem, 2-dimensional random walk in a semi-infinite strip.

We think it is best to illustrate the dependence problem with a ranking example that we will use throughout the paper. We limit ourselves to only two distinct products (which each should be thought of as a variety of different products grouped into one). Using realistic parameters we show that two different ranking algorithms which perform identical if run independently show a significant difference almost 20% of the time when run in an industry standard A/B experiment. We also show that if there is a difference in performance between the algorithms there are scenarios where the power of a standard A/B test is close to 0.

Example 1.1 (Ranking experiment, take 1). We consider a ranking experiment with two types of goods, one rare good, which is very attractive (good 2), and a second, less attractive good (good 1) available in practically unlimited quantities. In applications, there may be more than two types of goods, but the less attractive ones are labeled as type-1 goods, while the most attractive ones are labeled as type-2 goods. Suppose that in total there are 1 000 goods of type 2 and 1 000 000 goods of type 1.

A website displays the available goods to each visitor. The goods are displayed in a certain order, which depends on the ranking algorithm used. The owner of the website wants to compare two different algorithms. The default algorithm, Algorithm 0, displays the goods such that the type-2 goods have a low ranking and appear late in the list. Thus, only a fraction of the visitors gets to see them. The new algorithm, Algorithm 1, displays the goods such that the type-2 goods have the highest ranking and appear first in the list. Every visitor seeing the goods ranked by Algorithm 1 will see both, type-1 and type-2 goods (as long as they are available). The goal is to find out which of the two algorithms leads to a higher overall conversion rate, i.e., to a higher empirical probability to make a sale.

Suppose that during a test phase, $n = 4\,000\,000$ customers visit the website. Whenever a customer visits the website, a fair coin is tossed. If the coin shows heads, the products are displayed ranked according to Algorithm 1, if the coin shows tails, the products are displayed ranked according to Algorithm 0.

We now make the following model assumptions. We assume that, independent of all other customers, each customer has a chance of 20% of preferring good 1 over good 2 and an 80% chance of preferring good 2 over good 1. When the goods are ranked according to Algorithm 0, the customer first sees type-1 goods. There is a 5% chance that the customer scrolls down and spots a type-2 good (if still available). A customer who sees both goods and has a preference for good 2 will buy good 2 with 10% chance and will not buy at all with 90% chance. A customer who either sees both goods and has a preference for good 1 or sees only good 1 will buy good 1 with a 5% chance and will not make a buy at all with 95% chance. For simplicity, we assume that each customer buys at most one good.

The data collected is a sample $(x_1, y_1, i_1), \dots, (x_n, y_n, i_n)$ where n is the sample size, i.e., the number of customers visiting the website during a certain test period. Here, i_k is either 0 or 1, depending on whether Algorithm 0 or 1 was used to display the goods to the k^{th} customer. Further, $x_k = 1$ or $x_k = 0$ depending on whether good 1 was bought or not and, analogously, $y_k = 1$ or $y_k = 0$ depending on whether good 2 was bought or not. Notice that by our assumption that each customer buys at most one good, we have $x_k + y_k \in \{0, 1\}$. Those (x_k, y_k, i_k) with $i_k = 0$ are assigned to sample 0 and the others to sample 1. We write $n_0 := \sum_{k=1}^n (1 - i_k)$ and $n_1 := \sum_{k=1}^n i_k$ for the corresponding sample sizes. The numbers of sales in each group are $\ell_0 = \sum_{k=1}^n (x_k + y_k)(1 - i_k)$ and $\ell_1 = \sum_{k=1}^n (x_k + y_k)i_k$, the total number of sales is $\ell = \ell_0 + \ell_1$. The empirical probabilities for

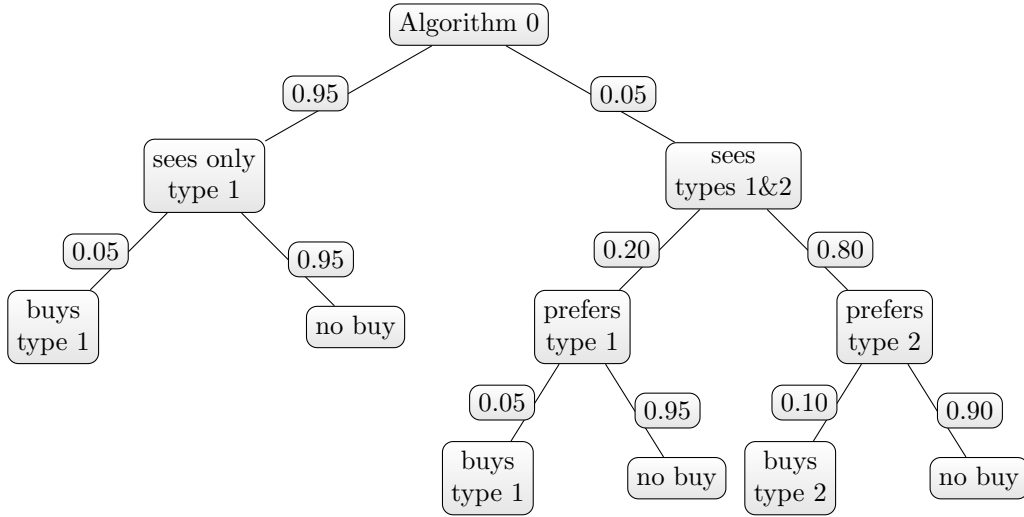


FIGURE 1. Ranking experiment: Algorithm 0

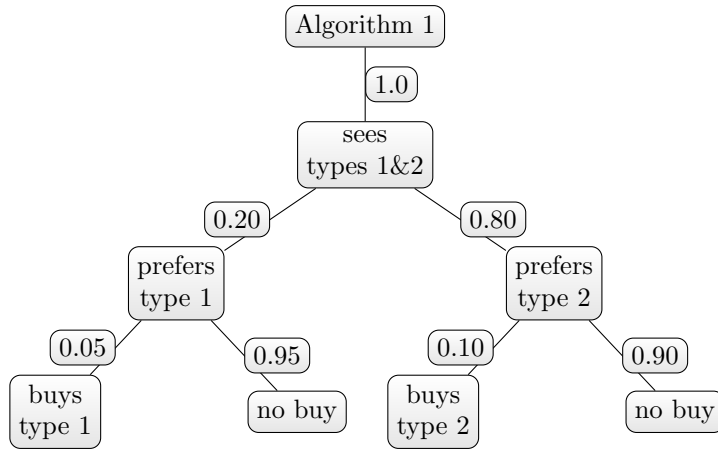


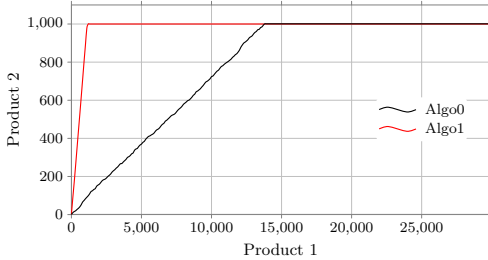
FIGURE 2. Ranking experiment: Algorithm 1

sales in samples 0 and 1 are

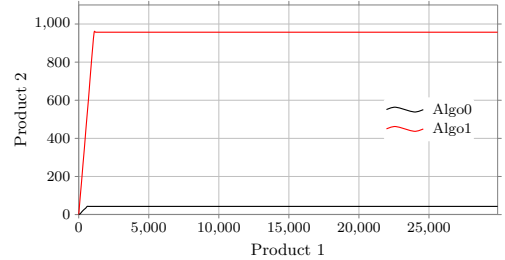
$$p_0 := \frac{1}{n_0} \sum_{k=1}^n (x_k + y_k)(1 - i_k) = \frac{\ell_0}{n_0} \quad \text{and} \quad p_1 := \frac{1}{n_1} \sum_{k=1}^n (x_k + y_k)i_k = \frac{\ell_1}{n_1}.$$

The website owner wants to find out whether Algorithm 1 performs better than Algorithm 0.

It is a common approach to test for the higher probability of a sale by assuming an independent sample and using a G -test or the asymptotically equivalent two-sample chi-squared test. The hypothesis is that the conversion rates are identical in both samples. For simplicity, in the paper



(A) Simulation of sales for Algorithms 0 and 1 run on separate inventories. Both algorithms sell all 1 000 attractive goods. Algorithm 0 additionally sells 199 528 goods of type 1, Algorithm 1 sells 199 325. The differences between the two algorithms are within the fluctuations one expects. Surely, this simulation does not give rise to the conclusion that Algorithm 1 outperforms Algorithm 0. However, Algorithm 1 sells the attractive goods earlier.



(B) Simulation of sales for Algorithms 0 and 1 run on shared inventory. Algorithm 0 sells 99 224 items of good 1 and 43 items of good 2, while Algorithm 1 sells 99 229 items of product 1 and the remaining 957 items of product 2. The huge difference in sales of the attractive good leads to a rejection of the null hypothesis (of both algorithms performing equally well) by the chi-squared test, the p -value in this simulation is $0.0369 \dots < 0.05$.

FIGURE 3. Simulation of sales in Example 1.1 on separate and shared inventory.

at hand, we shall always consider the chi-squared test. The test statistics for the latter is

$$\chi^2 = \sum_{i=0,1} \frac{(\ell_i - \ell \frac{n_i}{n})^2}{\ell \frac{n_i}{n}} + \sum_{i=0,1} \frac{(n_i - \ell_i - \ell \frac{n_i}{n})^2}{(n - \ell) \frac{n_i}{n}}.$$

The hypothesis is rejected if $\chi^2 > q_{1-\alpha}$ where $\alpha \in (0, 1)$ is the significance level and $q_{1-\alpha}$ is the $(1-\alpha)$ -quantile of the chi-squared distribution with one degree of freedom, see [3, Chapter 17].

Throughout the paper, we shall return repeatedly to Example 1.1 and discuss it in the light of our findings.

We shall discuss a variant of this example later on showing that ignoring the dependencies might also lead to too high false negative rates, see Example 3.8 below.

2. MODEL ASSUMPTIONS

We return to the general situation, in which a website offers two types of goods, good 1 and good 2. During a test phase, in which a new website design is used in parallel, the website has n visitors. Suppose that the website has a practically unlimited supply of items of good 1, while there are only $c_n \in \{1, 2, 3, \dots\} =: \mathbb{N}$ items of good 2. Typically, n will be very large and c_n will also be large, but significantly smaller than n . Whenever a user visits the website, a coin with success probability p is tossed. If the coin shows heads, the new website design is displayed, whereas if the coin shows tails, the old design is displayed. We thus observe a sample $((x_1, y_1, i_1), \dots, (x_n, y_n, i_n)) \in (\mathbb{N}_0^2 \times \{0, 1\})^n$ where $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$. Here, x_k and y_k are the numbers of goods of type 1 and 2, respectively, that have been bought by the k^{th} visitor of the website during the test phase, while $i_k = 1$ if the new design has been displayed to the k^{th} visitor, and $i_k = 0$, otherwise. We consider $((x_1, y_1, i_1), \dots, (x_n, y_n, i_n))$ as the realization of a random vector

$((X_1, Y_1, I_1), \dots, (X_n, Y_n, I_n))$. We define $Z_k := \mathbb{1}_{\{X_k + Y_k > 0\}}$ to be the indicator of the event that the k^{th} customer bought something. Further, we set $\mathbf{r}_k := (s_k, t_k) := (x_1, y_1) + \dots + (x_k, y_k)$ and $\mathbf{R}_k := (S_k, T_k) := (X_1, Y_1) + \dots + (X_k, Y_k)$ for $k = 0, \dots, n$ where the empty sum is defined to be the zero vector.

2.1. The classical model assuming independence. Many website owners in e-commerce use the G -test or the chi-squared test in the given situation. This test only uses the information whether or not a good was purchased, that is, the only information from the sample $((X_1, Y_1, I_1), \dots, (X_n, Y_n, I_n))$ used by the test is $(Z_1, I_1), \dots, (Z_n, I_n)$. This amounts to the following model assumptions.

($\chi 1$) There is a sequence (I_1, I_2, \dots) of i.i.d. copies of a Bernoulli variable I with $\mathbb{P}(I = 1) = p = 1 - \mathbb{P}(I = 0) \in (0, 1)$.

($\chi 2$) There are a random variable ζ and $p_0, p_1 \in (0, 1)$ such that

$$\mathbb{P}(Z_k \in \cdot | I_k = i) = \mathbb{P}(\zeta \in \cdot | I = i) = \text{Ber}(p_i)(\cdot) = p_i \delta_1(\cdot) + (1 - p_i) \delta_0(\cdot)$$

for all $k \in \mathbb{N}$ and $i = 0, 1$,

($\chi 3$) The family $((I_k, Z_k))_{k \in \mathbb{N}}$ is independent.

Here, and throughout the paper, for $x \in \mathbb{R}^d$, we write δ_x for the Dirac measure with a point at x . Assumptions ($\chi 1$) through ($\chi 3$) possess the following interpretations.

($\chi 1$): The random variable I_k models the coin toss that is used to decide whether the new or the old website design is displayed to the k^{th} visitor of the website during the test phase.

($\chi 2$): The random variable Z_k is the indicator of the event that the k^{th} visitor bought something.

($\chi 3$): The independence assumption in the context of the low inventory problem is made for simplicity. We question the feasibility of this assumption in the present paper.

2.2. A model incorporating low inventory of a popular good. We propose a simple model in which we keep track of the inventory of a rare good. Throughout the paper, we shall refer to this model as the ‘model incorporating low inventory’. By $c_n \in \mathbb{N}$ we denote the quantity at which the rare good is available. The most important case we consider is where c_n is asymptotically equivalent to a constant times \sqrt{n} . However, as we need this assumption only occasionally, throughout the paper, we only assume that $(c_n)_{n \in \mathbb{N}}$ is a non-decreasing unbounded sequence of integers which is regularly varying with index $\rho \in (0, 1]$ at infinity¹, that is,

$$\lim_{n \rightarrow \infty} \frac{c_{\lfloor nt \rfloor}}{c_n} = t^\rho, \quad t \geq 0 \tag{2.1}$$

and further $c_n = O(n)$ as $n \rightarrow \infty$, which is relevant only if $\rho = 1$. Notice that the case $c_n \sim \text{const} \cdot \sqrt{n}$ is covered. Indeed, in this case, we have $\rho = \frac{1}{2}$.

In the next step, we informally describe the evolution of the process $(\mathbf{R}_k)_{k \in \mathbb{N}_0}$. Let $C_n := \mathbb{N}_0 \times ([0, c_n] \cap \mathbb{N}_0)$, $C_n^\circ := \mathbb{N}_0 \times ([0, c_n) \cap \mathbb{N}_0)$ and $\partial C_n := C_n \setminus C_n^\circ = \mathbb{N} \times \{c_n\}$. At each step, a coin with success probability p is tossed. Depending on whether the coin shows heads or tails, the walk attempts to make one step according to a probability distribution μ_1 or μ_0 , respectively, on \mathbb{N}_0^2 . The step is actually performed if the walk stays in the strip C_n . Otherwise, another independent coin with success probability q is tossed. If the second coin shows heads, the walk moves in each coordinate direction according to the attempted step as far as possible but stops at the boundary of C_n . If the second coin shows tails, the walk stays put. Once the walk is on the boundary of C_n , it moves there according to a one-dimensional random walk in horizontal direction.

¹see [2] for a standard textbook reference

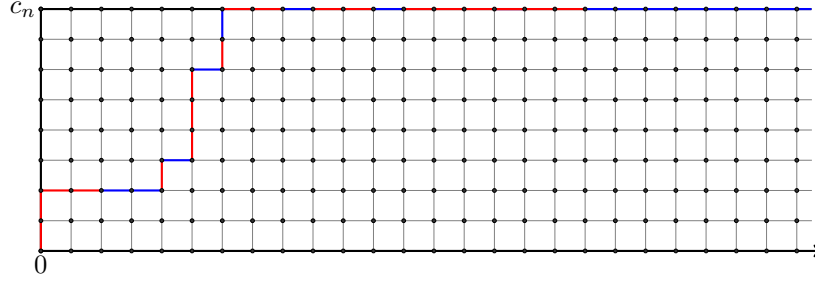


FIGURE 4. The 2-dimensional random walk in a semi-infinite strip. The k^{th} step is drawn in red if $I_k = 1$ and it is drawn in blue, otherwise. The walk moves in the strip until it hits the upper boundary and continues on the boundary as a 1-dimensional random walk. Here, we use the particular model from Example 1.1. Notice that the majority of steps have length zero and are not displayed.

The underlying model assumptions are the following.

- (A1) There are two sequences (I_1, I_2, \dots) and (J_1, J_2, \dots) of i.i.d. copies of Bernoulli variables I and J , respectively, with $\mathbb{P}(I = 1) = p = 1 - \mathbb{P}(I = 0) \in (0, 1)$ and $\mathbb{P}(J = 1) = q = 1 - \mathbb{P}(J = 0) \in [0, 1]$.
- (A2) There are a sequence $(\xi_1, \eta_1), (\xi_2, \eta_2), \dots$ of i.i.d. copies of a random variable (ξ, η) and two probability measures μ_0, μ_1 on \mathbb{N}_0^2 satisfying $\mu_i(\{(a, b)\}) > 0$ for all $(a, b) \in \{0, 1\}^2$ and $i = 0, 1$ such that

$$\mathbb{P}((\xi_k, \eta_k) \in \cdot | I_k = i) = \mathbb{P}((\xi, \eta) \in \cdot | I = i) = \mu_i(\cdot)$$

for all $k \in \mathbb{N}$ and $i = 0, 1$. We set $\mu_p(\cdot) := p\mu_1(\cdot) + (1-p)\mu_0(\cdot) = \mathbb{P}((\xi, \eta) \in \cdot)$.

- (A3) There are a probability measure ν on \mathbb{N}_0 and i.i.d. copies $\theta_1, \theta_2, \dots$ of a random variable θ with $\mathbb{P}(\theta \in \cdot) = \nu(\cdot)$.
- (A4) The sequences $(I_k)_{k \in \mathbb{N}}, (J_k)_{k \in \mathbb{N}}, ((\xi_k, \eta_k))_{k \in \mathbb{N}}$ and $(\theta_k)_{k \in \mathbb{N}}$ are independent; all random variables have finite second moments.
- (A5) Let $k \in \mathbb{N}$. If $\mathbf{R}_{k-1} = (S_{k-1}, T_{k-1}) \in C_n^\circ$, then

$$\mathbf{R}_k = \begin{cases} \mathbf{R}_{k-1} + (\xi_k, \eta_k) & \text{if } T_{k-1} + \eta_k \leq c_n, \\ (S_k, (T_{k-1} + \eta_k) \wedge c_n) & \text{if } T_{k-1} + \eta_k > c_n \text{ and } J_k = 1, \\ \mathbf{R}_{k-1} & \text{if } T_{k-1} + \eta_k > c_n \text{ and } J_k = 0. \end{cases}$$

On the other hand, if $\mathbf{R}_{k-1} \in \partial C_n$, then $T_{k-1} = c_n$. In this case,

$$\mathbf{R}_k = \mathbf{R}_{k-1} + (\theta_k, 0).$$

Finally, define $(X_k, Y_k) := \mathbf{R}_k - \mathbf{R}_{k-1}$.

The interpretations of these assumptions are the following.

- (A1): The random variable I_k models the coin toss that is used to decide whether the new or the old website design is displayed to the k^{th} visitor of the website during the test phase. The random variable J_k models the preference of the k^{th} visitor. If $J_k = 1$, then the user must buy. Users with $J_k = 0$ only buy when they get exactly what they want in the first place.

(A2): The random variable (ξ_k, η_k) can be interpreted as the vector of goods that the k^{th} visitor would buy when visiting the (displayed version of the) website if there was enough supply of these goods.

(A3): The random variable θ_k can be interpreted as the amount of type-1 goods that the k^{th} visitor would buy when visiting the website and finding only goods of type 1 left.

(A4): This is an independence assumption which is made to keep the model as simple as possible.

(A5): The random variable (X_k, Y_k) models what is actually bought by the k^{th} user. This depends on the needs of the user, ξ_k, η_k and θ_k , the remaining amount of the rare good 2 given by $c_n - T_{k-1}$, and the user's preference J_k . If there are enough goods available to meet the needs of the k^{th} user, then the user will buy exactly the needed amounts, namely, ξ_k of good 1 and η_k of good 2. If good 2 is not available at a sufficient quantity, then the user will either buy as much as possible of each of the goods if $J_k = 1$ or nothing at all if $J_k = 0$. If there is nothing left of good 2, the user will buy θ_k of good 1.²

Notice that in both models, \mathbb{P} depends on p , which is not explicit in the notation. While in large parts of the paper, p is fixed, in some places, however, it is important to make the dependence of \mathbb{P} on p explicit. In these places, we write \mathbb{P}_p . Often, this will be \mathbb{P}_0 or \mathbb{P}_1 , which correspond to the situations where only one version of the website is used.

Let us introduce some notation for various characteristics of the above variables which we shall use throughout the paper.

- We set $\mathbf{m}_0 := (m_0^\xi, m_0^\eta) := \mathbb{E}[(\xi, \eta)|I = 0]$, $\mathbf{m}_1 := (m_1^\xi, m_1^\eta) := \mathbb{E}[(\xi, \eta)|I = 1]$ and $\mathbf{m} := (m^\xi, m^\eta) := \mathbb{E}[(\xi, \eta)] = p\mathbf{m}_1 + (1-p)\mathbf{m}_0$. Notice that m^ξ and m^η depend on p even though this is not explicit in the notation.
- The covariance matrices of the probability measures $\mu_i, i = 0, 1$ are denoted by $\mathbf{C}_i, i = 0, 1$, respectively. The covariance matrix of the probability measure μ_p is then

$$\mathbf{C} = p\mathbf{C}_1 + (1-p)\mathbf{C}_0 = \begin{pmatrix} \sigma_\xi^2 & \rho_{\xi\eta} \\ \rho_{\xi\eta} & \sigma_\eta^2 \end{pmatrix}.$$

- We denote by $m^\theta = \mathbb{E}[\theta]$ and $\sigma_\theta^2 = \text{Var}[\theta]$, the mean and the variance of the probability measure ν .
- Finally, we set $p_i := \mu_i(\{(0, 0)^c\}) = \mathbb{P}(\xi + \eta > 0|I = i)$ for $i = 0, 1$ and $p_\theta := \mathbb{P}(\theta > 0)$. The $p_i, i = 0, 1$ and p_θ are the theoretical conversion rates.

Example 2.1 (Ranking experiment, take 2). We return to Example 1.1 and briefly explain how this example fits into the framework of the above model. The number of visitors during the test phase is $n = 4 \cdot 10^6$. The quantity of the attractive good 2 is $c_n = 1000 = \frac{1}{2} \cdot \sqrt{n}$. Website visitors view each version of the website with equal probability, so I_1, I_2, \dots have success probability $p = \frac{1}{2}$.

Further, as can be readily seen from Figure 1,

$$\mu_0 = \frac{96}{100} \left(\frac{1}{20} \delta_{(1,0)} + \frac{19}{20} \delta_{(0,0)} \right) + \frac{4}{100} \left(\frac{1}{10} \delta_{(0,1)} + \frac{9}{10} \delta_{(0,0)} \right) = \frac{948}{1000} \delta_{(0,0)} + \frac{4}{1000} \delta_{(0,1)} + \frac{48}{1000} \delta_{(1,0)}.$$

Analogously, from Figure 2, we deduce

$$\mu_1 = \frac{1}{5} \left(\frac{1}{20} \delta_{(1,0)} + \frac{19}{20} \delta_{(0,0)} \right) + \frac{4}{5} \left(\frac{1}{10} \delta_{(0,1)} + \frac{9}{10} \delta_{(0,0)} \right) = \frac{91}{100} \delta_{(0,0)} + \frac{8}{100} \delta_{(0,1)} + \frac{1}{100} \delta_{(1,0)}.$$

² Notice that according to our model, the two versions of the website have an identical effect on the user once the popular good is sold out. This is a simplifying assumption which excludes situations where for instance the effect of a new banner on the website is investigated. These situations can sometimes be analyzed via classical tests. In any case, we point out that our proofs could be easily modified to deal with the situation where the law of θ_k depends on the value of I_k , but then the results become even more cumbersome.

Moreover, the law of θ is given by $\frac{19}{20}\delta_{(0,0)} + \frac{1}{20}\delta_{(1,0)}$. The variables J_k are irrelevant in the given situation as step sizes here are at most one, hence there will never be the situation where a visitor attempts to buy more of the popular good than what is left. We conclude that the theoretical conversion rates are given by

$$p_0 = \mu_0(\{(0,0)\}^c) = \frac{52}{1000}, \quad p_1 = \mu_1(\{(0,0)\}^c) = \frac{9}{100} \quad \text{and} \quad p_\theta = \frac{5}{100}.$$

We shall see that if $c_n = c\sqrt{n}$, then the chi-squared test will reject the hypothesis with probability tending to 1 as c becomes large. On the other hand, we shall demonstrate that Algorithm 2 does not perform better given the model assumptions (A1) through (A5).

3. TESTING FOR THE HIGHER CONVERSION RATE

We address the question which algorithm, when used alone, leads to the higher conversion rate, where the conversion rate is the total number of sales divided by the total number of visitors. More formally, for $i = 0, 1$, we define

$$N_n^{(i)} := \sum_{k=1}^n \mathbb{1}_{\{I_k=i\}} \quad \text{and} \quad L_n^{(i)} := \sum_{k=1}^n \mathbb{1}_{\{Z_k>0, I_k=i\}}, \quad (3.1)$$

which model the number of visitors of website version i and the number of those visitors who make a purchase. We set $L_n := L_n^{(0)} + L_n^{(1)}$, which is the total number of purchases, and notice that $N_n^{(0)} + N_n^{(1)} = n$. Then

$$C_n^{(i)} := \frac{L_n^{(i)}}{N_n^{(i)}} \quad (3.2)$$

is the empirical conversion rate in group i . We stipulate that $C_n^{(i)} := 0$ on $\{N_n^{(i)} = 0\}$.

If one chooses $p \in \{0, 1\}$, then $C_n^{(p)}$ under \mathbb{P}_p is the empirical conversion rate when only website $i = p$ is used. In view of this, version 1 of the website is better than version 0 if $C_n^{(1)}$ under \mathbb{P}_1 is ‘larger’ than $C_n^{(0)}$ under \mathbb{P}_0 . Here, the term ‘larger’ is not specified a priori, so we need to clarify what we mean by this.

3.1. The chi-squared test statistics in the classical model. In the classical model, i.e., if assumptions $(\chi 1)$, $(\chi 2)$ and $(\chi 3)$ are in force, Z_1, Z_2, \dots are i.i.d. with expectation $\mathbb{E}[Z_1] = \mathbb{P}(Z_1 = 1) = pp_1 + (1-p)p_0$. Hence, by the strong law of large numbers, for $p \in \{0, 1\}$,

$$c^{(p)} := \lim_{n \rightarrow \infty} C_n^{(p)} = \lim_{n \rightarrow \infty} \frac{1}{n} L_n^{(p)} = \mathbb{P}(\xi + \eta > 0) = pp_1 + (1-p)p_0 \quad \text{a. s.}$$

Consequently, testing whether $C_n^{(1)}$ under \mathbb{P}_1 is ‘different’ from $C_n^{(0)}$ under \mathbb{P}_0 can be formulated as follows:

$$H_0 : p_1 = p_0 \quad \text{vs.} \quad H_1 : p_1 \neq p_0.$$

More interest, in fact, would be in the corresponding one-sided test problem. Hence, in this case, it is a classical test problem and widely used tests for this problem are the chi-squared test, the G -test, and Fisher’s exact test. To keep the paper short, we shall always restrict attention to the chi-squared test. For the reader’s convenience, we recall some facts about this test.

The test statistics for the chi-squared test is

$$\chi^2 := \sum_{i=0,1} \frac{(L_n^{(i)} - L_n \frac{N_n^{(i)}}{n})^2}{L_n \frac{N_n^{(i)}}{n}} + \sum_{i=0,1} \frac{(N_n^{(i)} - L_n^{(i)} - (n - L_n) \frac{N_n^{(i)}}{n})^2}{(n - L_n) \frac{N_n^{(i)}}{n}}. \quad (3.3)$$

TABLE 1. Contingency table

| | purchase | no purchase | \sum |
|---------|-------------|-------------------------|-------------|
| group 0 | $L_n^{(0)}$ | $N_n^{(0)} - L_n^{(0)}$ | $N_n^{(0)}$ |
| group 1 | $L_n^{(1)}$ | $N_n^{(1)} - L_n^{(1)}$ | $N_n^{(1)}$ |
| \sum | L_n | $n - L_n$ | n |

If $(\chi 1)$ through $(\chi 3)$ are in force, as $n \rightarrow \infty$, the distribution of χ^2 approaches a chi-squared distribution with 1 degree of freedom [3, Chapter 17]. Write $q_{1-\alpha}$ for the $1 - \alpha$ quantile of the chi-squared distribution with 1 degree of freedom. Then, with significance level of α , the hypothesis is rejected if $\chi^2 > q_{1-\alpha}$.

3.2. The limiting law of the chi-squared test statistics in the model incorporating low inventory. Now suppose that there is a rare but popular good, i.e., suppose that the model assumptions (A1) through (A5) hold. One goal of this paper is to point out in a quantitative way that when (A1) through (A5) instead of $(\chi 1)$ through $(\chi 3)$ are in force, then the chi-squared test may produce too many false positives. In other words, it may fail to hold the specified significance level. This is because the distribution of χ^2 under the null hypothesis is different when (A1) through (A5) rather than $(\chi 1)$ through $(\chi 3)$ are in force. The detailed statement is given in the following theorem.

Theorem 3.1. *Suppose that (A1) through (A5) and (2.1) are in force. Assume additionally that*

$$d_\infty := \lim_{n \rightarrow \infty} \frac{c_n}{\sqrt{n}} \in [0, \infty). \tag{3.4}$$

Then the chi-squared statistics defined by (3.3) satisfies

$$\chi^2 \xrightarrow{d} \left(\mathcal{N} - \frac{d_\infty(p_0 - p_1)\sqrt{p(1-p)}}{m^\eta \sqrt{p_\theta(1-p_\theta)}} \right)^2 \quad \text{as } n \rightarrow \infty \tag{3.5}$$

where \mathcal{N} is a standard normal random variable.

Hence, if one applies the chi-squared test with significance level $\alpha \in (0, 1)$ in the given situation, the test rejects the hypothesis with (asymptotic) probability

$$\mathbb{P} \left(\left(\mathcal{N} - \frac{d_\infty(p_0 - p_1)\sqrt{p(1-p)}}{m^\eta \sqrt{p_\theta(1-p_\theta)}} \right)^2 > q_{1-\alpha} \right) > \mathbb{P}(\mathcal{N}^2 > q_{1-\alpha}) = \alpha.$$

In fact, the probability on the left-hand side tends to 1 as $d_\infty \rightarrow \infty$. We specialize to the situation of Example 1.1.

Example 3.2 (Ranking experiment, take 3). Recall the situation of Example 1.1. Then, see also Example 2.1, we have

$$p = \frac{1}{2}, \quad p_\theta = \frac{1}{20}, \quad p_0 = \frac{52}{1000}, \quad p_1 = \frac{9}{100}, \quad m^\eta = \frac{42}{1000} \quad \text{and} \quad d_\infty = \frac{1}{2}.$$

Consequently,

$$\frac{d_\infty(p_0 - p_1)\sqrt{p(1-p)}}{m^\eta \sqrt{p_\theta(1-p_\theta)}} = \frac{\frac{1}{2} \cdot \frac{-38}{1000} \cdot \frac{1}{2}}{\frac{42}{1000} \cdot \sqrt{\frac{19}{400}}} = -\frac{\sqrt{19} \cdot 5}{21} = -1.037833\dots \tag{3.6}$$

Hence, in the given situation, the chi-squared test rejects the hypothesis with (asymptotic) probability

$$\mathbb{P}((\mathcal{N} + 1.037833)^2 > q_{95\%}) = 0.1795898\dots > 0.05.$$

This becomes worse as d_∞ becomes larger, see Figure 5 below.

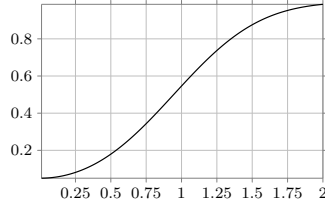


FIGURE 5. False-positive probability as a function of d_∞ with all other parameters fixed as in (3.6).

At first glance, this may occur to be no problem as $p_1 > p_0$, so one is tempted to guess that algorithm 1 performs better than algorithm 0 and what we see above is just the power of the test, which becomes better as d_∞ becomes large. However, we shall argue in Example 3.6 below that the two algorithms perform equally well when used separately.

Next, we show that in the general situation, assuming that (A1) through (A5) hold and that $\frac{c_n}{n} \rightarrow 0$, we show that on the linear scale, the asymptotic empirical conversion rates of the two versions of the website, when used separately, are identical.

Theorem 3.3. *Suppose that (A1) through (A5) hold and that $c_\infty := \lim_{n \rightarrow \infty} \frac{c_n}{n} = 0$. Then, for $p \in \{0, 1\}$,*

$$\lim_{n \rightarrow \infty} C_n^{(p)} = p_\theta \quad \mathbb{P}_p\text{-a. s.} \quad (3.7)$$

Proof. The result is a consequence of Theorem 3.4 below. \square

Hence, in the relevant regime ($\frac{c_n}{n} \rightarrow 0$) the first order of growth of $C_n^{(p)}$ depends only on what happens after the popular good is sold out. According to our model assumptions, the two versions of the website have identical performance once the popular good is sold out. This implies that on the linear scale, there is no difference between the two versions of the website. Hence, we need to make a comparison on a finer scale.

3.3. A joint limit theorem for the group sizes and numbers of purchases in each group.

As χ^2 is a function of $(N_n^{(0)}, N_n^{(1)}, L_n^{(0)}, L_n^{(1)})$, a limit theorem for χ^2 follows from one for the above vector via the continuous mapping theorem [1, Theorem 2.7]. We begin with a strong law of large numbers for the variables $L_n^{(0)}$ and $L_n^{(1)}$ (the corresponding result for $N_n^{(0)}$ and $N_n^{(1)}$ is classical).

Theorem 3.4. *Suppose that (A1) through (A5) are in force and that the limit*

$$c_\infty := \lim_{n \rightarrow \infty} \frac{c_n}{n} \in [0, \infty)$$

exists³. If $c_\infty \in [0, \frac{1}{m^\eta})$, then

$$\frac{1}{n}L_n^{(0)} \rightarrow (1-p)p_\theta + c_\infty(1-p)\frac{1}{m^\eta}(p_0 - p_\theta) \quad \text{and} \quad \frac{1}{n}L_n^{(1)} \rightarrow pp_\theta + c_\infty p\frac{1}{m^\eta}(p_1 - p_\theta) \quad a. s.$$

In particular, in the most relevant case $c_\infty = 0$,

$$\frac{1}{n}L_n^{(0)} \rightarrow (1-p)p_\theta \quad \text{and} \quad \frac{1}{n}L_n^{(1)} \rightarrow pp_\theta \quad a. s.$$

If $c_\infty > \frac{1}{m^\eta}$, then

$$\frac{1}{n}L_n^{(0)} \rightarrow (1-p)p_0 \quad \text{and} \quad \frac{1}{n}L_n^{(1)} \rightarrow pp_1 \quad a. s.$$

We continue with the asymptotic law of the vector $(N_n^{(0)}, N_n^{(1)}, L_n^{(0)}, L_n^{(1)})$, suitably shifted and scaled in the most relevant scenario where c_n is of the order \sqrt{n} .

Theorem 3.5. *Suppose that (A1) through (A5) are in force and suppose in addition to (2.1) that the limit*

$$d_\infty = \lim_{n \rightarrow \infty} \frac{c_n}{\sqrt{n}} \in [0, \infty) \quad (3.4)$$

exists, we have, as $n \rightarrow \infty$,

$$\left(\frac{N_n^{(0)} - (1-p)n}{\sqrt{n}}, \frac{N_n^{(1)} - pn}{\sqrt{n}}, \frac{L_n^{(0)} - n(1-p)p_\theta}{\sqrt{n}}, \frac{L_n^{(1)} - npp_\theta}{\sqrt{n}} \right) \xrightarrow{d} d_\infty(0, 0, (1-p)\frac{1}{m^\eta}(p_0 - p_\theta), p\frac{1}{m^\eta}(p_1 - p_\theta)) + (G_1, -G_1, G_2, G_3) \quad (3.8)$$

where (G_1, G_2, G_3) is a centered Gaussian vector with covariance matrix

$$\mathbf{V}_1 = \begin{pmatrix} p(1-p) & p(1-p)p_\theta & -p(1-p)p_\theta \\ p(1-p)p_\theta & p_\theta(1-p)(1-p_\theta(1-p)) & -p(1-p)p_\theta^2 \\ -p(1-p)p_\theta & -p(1-p)p_\theta^2 & pp_\theta(1-pp_\theta) \end{pmatrix}. \quad (3.9)$$

Notice that the theorem contains a limit theorem for the pure conversion rates $C_n^{(0)}$ and $C_n^{(1)}$ by choosing $p = 0$ and projecting on the third coordinate or by choosing $p = 1$ and projecting on the fourth component, respectively. This gives, with $\mathcal{N}(\mu, \sigma^2)$ denoting a normal random variable with mean μ and variance σ^2 ,

$$\frac{L_n^{(0)} - np_\theta}{\sqrt{n}} \xrightarrow{d} \mathcal{N}\left(d_\infty \frac{p_0 - p_\theta}{m_0^\eta}, p_\theta(1-p_\theta)\right) \quad \text{under } \mathbb{P}_0 \quad (3.10)$$

$$\text{and} \quad \frac{L_n^{(1)} - np_\theta}{\sqrt{n}} \xrightarrow{d} \mathcal{N}\left(d_\infty \frac{p_1 - p_\theta}{m_1^\eta}, p_\theta(1-p_\theta)\right) \quad \text{under } \mathbb{P}_1 \quad (3.11)$$

where $m^\eta = pm_1^\eta + (1-p)m_0^\eta$ with $p = 0$ and $p = 1$, respectively, has been used. Hence, if the two expectations in (3.10) and (3.11) coincide, then the performances of the two websites coincide asymptotically both on the linear scale as well as on the level of fluctuations. The subsequent example demonstrates that this can be the case even if $p_0 \neq p_1$.

³In the applications we have in mind, $c_\infty = 0$ because the quantity of good 2 should be much smaller than the total number of observations. But from a theoretical perspective positive values of c_∞ are also interesting because of the occurrence of different asymptotic regimes. Let us also stress that $c_\infty > 0$ necessitates $\rho = 1$, where the definition of ρ may be recalled from (2.1).

Example 3.6 (Ranking experiment, take 4). Recall the situation of Example 1.1 and the calculations of Example 3.2:

$$p_\theta = \frac{1}{20}, \quad p_0 = \frac{52}{1000}, \quad p_1 = \frac{9}{100}, \quad m_0^\eta = \frac{4}{1000}, \quad m_1^\eta = \frac{8}{100} \quad \text{and} \quad d_\infty = \frac{1}{2}.$$

Consequently,

$$\frac{p_0 - p_\theta}{m_0^\eta} = \frac{\frac{2}{1000}}{\frac{4}{1000}} = \frac{1}{2} = \frac{\frac{4}{100}}{\frac{8}{100}} = \frac{p_1 - p_\theta}{m_1^\eta}.$$

This means that the two limits in (3.10) and (3.11) coincide in the given example even though $p_1 > p_0$.

From Theorem 3.5, we may immediately deduce a limit theorem for χ^2 , which is a preliminary version of Theorem 3.1.

Corollary 3.7. *Suppose that (A1) through (A5) and (2.1) are in force and that the limit*

$$d_\infty = \lim_{n \rightarrow \infty} \frac{c_n}{\sqrt{n}} \in [0, \infty) \quad (3.4)$$

exists, we have the following limit theorem for the chi-squared statistics defined by (3.3)

$$\begin{aligned} \chi^2 \xrightarrow{d} & \frac{(d_2 + G_2 - p_\theta G_1 - (1-p)(d_2 + d_3 + G_2 + G_3))^2}{(1-p_\theta)p_\theta(1-p)} \\ & + \frac{(d_3 + G_3 + p_\theta G_1 - p(d_2 + d_3 + G_2 + G_3))^2}{(1-p_\theta)p_\theta p} \quad \text{as } n \rightarrow \infty \end{aligned}$$

where $(0, 0, d_2, d_3)$ denotes the expectation of the right-hand side in (3.8) and (G_1, G_2, G_3) is the Gaussian vector from Theorem 3.5.

We close this section with another example showing that ignoring the dependencies might also lead to a high false negative rate.

Example 3.8 (Ranking experiment with picky customers). We consider a variant of Example 1.1 in which there are picky customers that will only buy the rare good. This time, we use the former Algorithm 1 from above as the default algorithm displaying the rare goods first. The former Algorithm 0 strategically keeps the rare inventory for later arrival of picky customers by ranking the rare good low. In an experiment with shared inventory, Algorithm 1 sells off the rare good greedily, and the value of Algorithm 0's strategy will not be properly assessed in the model ignoring dependencies. We make this precise in the following.

Again suppose that during a test phase, $n = 4\,000\,000$ customers visit the website. Again, there are $c_n = 1\,000$ rare goods while good 1 is available at sufficient quantities.

The algorithms work as in Example 1.1, but there is a difference in the behavior of the customers. We assume that, independent of all other customers, each customer has a 1% chance of being picky. If not picky, the customer behaves like the customers of Example 1.1. A picky customer, however, will search as long as required to check whether there is something of the rare good left. If the rare good is still available, the picky customer buys one unit with 50% chance. Otherwise, the customer leaves the website. The overviews given in Figure 2 still apply to regular customers, for picky customers and when good 2 is still available, there is a simplified decision tree:

A calculation of the relevant model parameters in this example based on the corresponding parameter values from Example 3.2 gives $p = \frac{1}{2}$ and $d_\infty = \frac{1}{2}$ as before and

$$p_\theta = \frac{99}{100} \cdot \frac{1}{20}, \quad p_0 = \frac{99}{100} \cdot \frac{52}{1000} + \frac{1}{100} \cdot \frac{1}{2}, \quad p_1 = \frac{99}{100} \cdot \frac{9}{100} + \frac{1}{100} \cdot \frac{1}{2}, \quad m^\eta = \frac{99}{100} \cdot \frac{42}{1000} + \frac{1}{100} \cdot \frac{1}{2}.$$

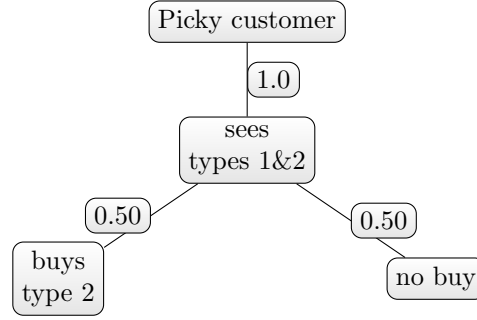


FIGURE 6. Ranking experiment: Picky customer behavior when good 2 is still available

We shall now show that Algorithm 0 performs actually better than Algorithm 1. We have

$$m_0^\eta = \frac{99}{100} \cdot \frac{4}{1000} + \frac{1}{100} \cdot \frac{1}{2} \quad \text{and} \quad m_1^\eta = \frac{99}{100} \cdot \frac{8}{100} + \frac{1}{100} \cdot \frac{1}{2}.$$

Consequently,

$$d_\infty \cdot \frac{p_0 - p_\theta}{m_0^\eta} = \frac{349}{896} > \frac{223}{842} = d_\infty \cdot \frac{p_1 - p_\theta}{m_1^\eta}.$$

In view of (3.10) and (3.11), Algorithm 0 does perform better than Algorithm 1. Now let us calculate the probability that the chi-squared test rejects the hypothesis that Algorithm 0 and Algorithm 1 perform equally well. To this end, we first calculate

$$\frac{d_\infty(p_1 - p_0)\sqrt{p(1-p)}}{m^\eta\sqrt{p_\theta(1-p_\theta)}} = 0.930852\dots$$

According to Theorem 3.1, the chi-squared test (with significance level 5%) rejects the hypothesis with probability 0.1536348\dots. However, it is a standard practice to say that Algorithm 0 is significantly better than Algorithm 1 only when $\chi^2 > q_{1-\alpha}$ and $C_n^{(0)} > C_n^{(1)}$ (with $\alpha \in (0, 1)$ being the significance level). So, asymptotically, the power of the test is $\lim_{n \rightarrow \infty} \mathbb{P}(C_n^{(0)} > C_n^{(1)}, \chi^2 > q_{1-\alpha})$. We shall now calculate this probability in the given situation with $d_\infty = 1/2$ but also as a function of d_∞ to point out that the probability becomes arbitrarily small as d_∞ becomes large. We begin by reformulating the condition $C_n^{(0)} > C_n^{(1)}$. Recall that $C_n^{(i)} = L_n^{(i)}/N_n^{(i)}$ for $i = 0, 1$. Hence,

$$\begin{aligned} C_n^{(0)} > C_n^{(1)} & \text{ iff } L_n^{(0)}N_n^{(1)} - L_n^{(1)}N_n^{(0)} > 0 \\ \text{iff } & (L_n^{(0)} - \frac{np_\theta}{2})(N_n^{(1)} - \frac{n}{2}) - (L_n^{(1)} - \frac{np_\theta}{2})(N_n^{(0)} - \frac{n}{2}) + \frac{np_\theta}{2}(N_n^{(1)} - N_n^{(0)}) + \frac{n}{2}(L_n^{(0)} - L_n^{(1)}) > 0 \\ \text{iff } & \frac{2}{\sqrt{n}} \frac{L_n^{(0)} - np_\theta/2}{\sqrt{n}} \cdot \frac{N_n^{(1)} - n/2}{\sqrt{n}} - \frac{2}{\sqrt{n}} \frac{L_n^{(1)} - np_\theta/2}{\sqrt{n}} \cdot \frac{N_n^{(0)} - n/2}{\sqrt{n}} \\ & - p_\theta \frac{N_n^{(0)} - n/2}{\sqrt{n}} + p_\theta \frac{N_n^{(1)} - n/2}{\sqrt{n}} + \frac{L_n^{(0)} - np_\theta/2}{\sqrt{n}} - \frac{L_n^{(1)} - np_\theta/2}{\sqrt{n}} > 0. \end{aligned}$$

By Theorem 3.5 and Slutsky's theorem, the two terms in the penultimate line tend to 0 in probability as $n \rightarrow \infty$. By Theorem 3.5, the other summands converge in distribution so that in the limit, the above inequality becomes

$$-2p_\theta G_1 + d_\infty \frac{p_0 - p_\theta}{2m^\eta} + G_2 - d_\infty \frac{p_1 - p_\theta}{2m^\eta} - G_3 > 0,$$

which can be simplified to

$$-2p_\theta G_1 + G_2 - G_3 > d_\infty \frac{p_1 - p_0}{2m^\eta}.$$

By Theorem 3.1, χ^2 converges also in distribution. According to Corollary 3.7, we can express the condition $\chi^2 > q_{1-\alpha}$ in the limit as $n \rightarrow \infty$ in the form

$$(d_2 + G_2 - p_\theta G_1 - \frac{1}{2}(d_2 + d_3 + G_2 + G_3))^2 + (d_3 + G_3 + p_\theta G_1 - \frac{1}{2}(d_2 + d_3 + G_2 + G_3))^2 > \frac{1}{2}(1 - p_\theta)p_\theta q_{1-\alpha} \quad (3.12)$$

where $(0, 0, d_2, d_3)$ is the expectation vector in (3.8), i.e., $(0, 0, d_2, d_3) = d_\infty(0, 0, \frac{p_0 - p_\theta}{2m^\eta}, \frac{p_1 - p_\theta}{2m^\eta})$. Since the convergence in Theorem 3.5 is jointly and since the law of (G_1, G_2, G_3) on \mathbb{R}^3 is absolutely continuous with respect to Lebesgue measure, the Portmanteau theorem implies that

$$\lim_{n \rightarrow \infty} \mathbb{P}(\chi^2 > q_{1-\alpha}, C_n^{(0)} > C_n^{(1)}) = \mathbb{P}((3.12) \text{ holds and } -2p_\theta G_1 + G_2 - G_3 > d_\infty \frac{p_1 - p_0}{2m^\eta}).$$

In the given situation (with $d_\infty = \frac{1}{2}$), we have used a Monte Carlo simulation (with $2 \cdot 10^6$ iterations) to estimate the power of the test resulting in an estimate of 0.001933... Further, it is already immediate that the power tends to 0 as d_∞ tends to ∞ since $p_1 > p_0$. To get a better quantitative picture, we have performed Monte Carlo simulations for 200 equidistant values of d_∞ between 0 and 1, each with $2 \cdot 10^6$ iterations. The results of the Monte Carlo simulations are displayed in Figure 7.

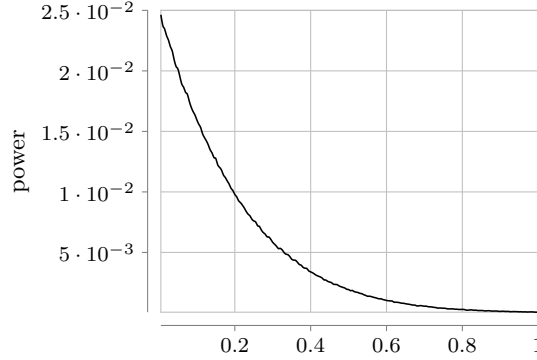


FIGURE 7. Monte Carlo simulation of the power of the test as a function of d_∞ with all other parameters fixed as above.

In particular, for every value $d_\infty > 0$, the estimate of the power of the test is strictly smaller than $0.025 = \frac{\alpha}{2}$.

3.4. Functional limit theorem for the model incorporating low inventory. We deduce Theorem 3.5 from a more general result, namely, a joint functional central limit theorem. To formulate it, we need additional notation.

Henceforth, convergence in distribution of random elements in the Skorohod spaces $D([0, \infty), \mathbb{R}^d)$ and $D((0, \infty), \mathbb{R}^d)$ of \mathbb{R}^d -valued, right-continuous functions with existing left limits is with respect to the standard J_1 -topology and will be denoted by \Rightarrow . To distinguish between convergence in the above two spaces we adopt the following convention. The convergence is in $D([0, \infty), \mathbb{R}^d)$ if the

4. PROOFS

In the model, there are two natural breaks, namely, first the process evolves in the positive quadrant like an unrestricted two-dimensional random walk until the second coordinate of the walk for the first time attempts to step to or beyond the level c_n . This time we call $\tau_1(n)$. Then there is a number of attempts to reach that border until this is eventually achieved at a time we call $\tau_2(n)$. From that time on, the walk keeps the second coordinate fixed at c_n and evolves horizontally as a one-dimensional walk. Crucial both for the proof of the strong law of large numbers, Theorem 3.4, and the joint functional limit theorem, Theorem 3.9, is a sufficient understanding of these times, $\tau_1(n)$ and $\tau_2(n)$.

4.1. Stopping time analysis and the strong law of large numbers. Formally, the two natural stopping times associated with the stochastic process $(\mathbf{R}_k)_{k \in \mathbb{N}_0}$ are defined as follows:

- $\tau_1(n) := \inf\{k \in \mathbb{N} : T_{k-1} + \eta_k \geq c_n\}$, the first attempt to reach the half-plane $\mathbb{N}_0 \times [c_n, \infty)$, and
- $\tau_2(n) := \inf\{k \in \mathbb{N} : T_k = c_n\}$, the first entrance to the horizontal line $\mathbb{N}_0 \times \{c_n\}$.

We stipulate that $\tau_1(0) = \tau_2(0) = 0$. To simplify notation later on, we set $\tau_j(t) := \tau_j(\lfloor t \rfloor)$ for $t \geq 0$ and $j = 1, 2$. The above quantities are stopping times with respect to the natural filtration of $(\mathbf{R}_k)_{k \in \mathbb{N}_0}$. By the strong law of large numbers, we have

$$\frac{\tau_1(n)}{c_n} \rightarrow \frac{1}{\mathbb{E}[\eta]} = \frac{1}{m^\eta} \quad \text{as } n \rightarrow \infty \quad \text{a. s.} \quad (4.1)$$

Note that $\tau_1(n) \leq \tau_2(n)$ and

$$\{\tau_1(n) = \tau_2(n)\} = \{T_{\tau_1(n)-1} + \eta_{\tau_1(n)} = c_n\} \cup \{T_{\tau_1(n)-1} + \eta_{\tau_1(n)} > c_n, J_{\tau_1(n)} = 1\}.$$

We start by proving that $\tau_1(n)$ and $\tau_2(n)$ are uniformly close on the $\sqrt{c_n}$ -scale.

Lemma 4.1. *For arbitrary $T > 0$ we have*

$$\sup_{t \in [0, T]} \frac{|\tau_2(nt) - \tau_1(nt)|}{\sqrt{c_n}} \xrightarrow{\mathbb{P}} 0 \quad \text{as } n \rightarrow \infty.$$

Proof. By the regular variation of $(c_n)_{n \in \mathbb{N}}$ it is enough to prove the claim for $T = 1$. We have the following bound

$$\begin{aligned} \tau_2(n) - \tau_1(n) &= \inf \left\{ k \geq 0 : T_{\tau_1(n)} + \sum_{j=1}^k Y_{\tau_1(n)+j} = c_n \right\} \\ &\leq \inf \left\{ k \geq 0 : T_{\tau_1(n)} + \sum_{j=1}^k \eta_{\tau_1(n)+j} \mathbb{1}_{\{J_{\tau_1(n)+j}=1\}} \geq c_n \right\}, \end{aligned} \quad (4.2)$$

and therefore

$$\begin{aligned} &\sup_{t \in [0, 1]} |\tau_2(nt) - \tau_1(nt)| \\ &\leq \inf \left\{ k \geq 0 : \inf_{t \in [0, 1]} \sum_{j=1}^k \eta_{\tau_1(\lfloor nt \rfloor)+j} \mathbb{1}_{\{J_{\tau_1(\lfloor nt \rfloor)+j}=1\}} \geq \sup_{t \in [n^{-1}, 1]} (c_{\lfloor nt \rfloor} - T_{\tau_1(nt)}) \right\}. \end{aligned} \quad (4.3)$$

Furthermore, $c_m - T_{\tau_1(m)} \leq \eta_{\tau_1(m)}$ and thus

$$\sup_{t \in [n^{-1}, 1]} (c_{\lfloor nt \rfloor} - T_{\tau_1(nt)}) \leq \max_{t \in [n^{-1}, 1]} \eta_{\tau_1(nt)} \leq \max_{m=1, \dots, \tau_1(n)} \eta_m,$$

which in view of (4.3) yields

$$\sup_{t \in [0, 1]} |\tau_2(nt) - \tau_1(nt)| \leq \inf \left\{ k \geq 0 : \min_{m=0, \dots, \tau_1(n)} \sum_{j=1}^k \eta_{m+j} \mathbb{1}_{\{J_{m+j}=1\}} \geq \max_{m=1, \dots, \tau_1(n)} \eta_m \right\}. \quad (4.4)$$

From (4.1) we conclude that for any $a > \frac{1}{m^\eta}$ there exists a random $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, we have

$$\sup_{t \in [n^{-1}, 1]} (c_{\lfloor nt \rfloor} - T_{\tau_1(nt)}) \leq \max_{k=1, \dots, \lfloor ac_n \rfloor} \eta_k.$$

This further implies

$$\sup_{t \in [0, 1]} |\tau_2(nt) - \tau_1(nt)| \leq \inf \left\{ k \geq 0 : \min_{m=0, \dots, \lfloor ac_n \rfloor} \sum_{j=1}^k \eta_{m+j} \mathbb{1}_{\{J_{m+j}=1\}} \geq \max_{m=0, \dots, \lfloor ac_n \rfloor} \eta_m \right\} \quad (4.5)$$

if $n \geq n_0$. Using the assumption $\mathbb{E}[\eta^2] < \infty$ it is not difficult to check that

$$c_n^{-\frac{1}{2}} \max_{k=1, \dots, \lfloor ac_n \rfloor} \eta_k \xrightarrow{\mathbb{P}} 0 \quad \text{as } n \rightarrow \infty. \quad (4.6)$$

Define

$$\tilde{S}_0 := 0 \quad \text{and} \quad \tilde{S}_n = \sum_{k=1}^n \eta_k \mathbb{1}_{\{J_k=1\}}, \quad n \in \mathbb{N}$$

and note that by (4.5) and (4.6) it is enough to prove that for arbitrary $\varepsilon > 0$ there exist $\delta > 0$ such

$$\mathbb{P} \left(\min_{m=1, \dots, \lfloor ac_n \rfloor} (\tilde{S}_{m+\lfloor \varepsilon \sqrt{c_n} \rfloor} - \tilde{S}_m) \geq \delta \sqrt{c_n} \right) \rightarrow 1 \quad \text{as } n \rightarrow \infty. \quad (4.7)$$

Fix $\varepsilon > 0$ and let us show that (4.7) holds for any $\delta \in (0, \varepsilon \mathbb{E} \tilde{S}_1) = (0, \varepsilon q m^\eta)$. To this end, fix arbitrary such δ and write

$$\begin{aligned} \mathbb{P} \left(\min_{m=1, \dots, \lfloor ac_n \rfloor} (\tilde{S}_{m+\lfloor \varepsilon \sqrt{c_n} \rfloor} - \tilde{S}_m) < \delta \sqrt{c_n} \right) &\leq \sum_{m=1}^{\lfloor ac_n \rfloor} \mathbb{P} \left(\tilde{S}_{m+\lfloor \varepsilon \sqrt{c_n} \rfloor} - \tilde{S}_m < \delta \sqrt{c_n} \right) \\ &= \lfloor ac_n \rfloor \mathbb{P} \left(\tilde{S}_{\lfloor \varepsilon \sqrt{c_n} \rfloor} < \delta \sqrt{c_n} \right). \end{aligned}$$

For every $\lambda > 0$ we have by Markov's inequality

$$\mathbb{P} \left(\tilde{S}_{\lfloor \varepsilon \sqrt{c_n} \rfloor} < \delta \sqrt{c_n} \right) = \mathbb{P} \left(e^{-\lambda \tilde{S}_{\lfloor \varepsilon \sqrt{c_n} \rfloor}} > e^{-\lambda \delta \sqrt{c_n}} \right) \leq e^{\lambda \delta \sqrt{c_n}} (\mathbb{E}[e^{-\lambda \eta \mathbb{1}_{\{J=1\}}}]^{\lfloor \varepsilon \sqrt{c_n} \rfloor}).$$

It remains to note that $e^{\lambda \delta \varepsilon^{-1}} (\mathbb{E}[e^{-\lambda \eta \mathbb{1}_{\{J=1\}}}]^{\lfloor \varepsilon \sqrt{c_n} \rfloor}) < 1$ for $\delta \in (0, \varepsilon q m^\eta)$ and sufficiently small $\lambda > 0$. The proof is complete. \square

For the proof of Theorem 3.4 we also need a counterpart of the above lemma for convergence in the almost sure sense.

Lemma 4.2. *It holds that*

$$\lim_{n \rightarrow \infty} \frac{\tau_2(n) - \tau_1(n)}{c_n} = 0 \quad \text{a. s.}$$

Proof. In view of (4.2) for every $\varepsilon > 0$ it holds

$$\begin{aligned} \{\tau_2(n) - \tau_1(n) > \lfloor \varepsilon c_n \rfloor\} &\subseteq \left\{ \sum_{j=1}^{\lfloor \varepsilon c_n \rfloor} \eta_{\tau_1(n)+j} \mathbb{1}_{\{J_{\tau_1(n)+j}=1\}} < c_n - T_{\tau_1(n)} \right\} \\ &= \left\{ \tilde{S}_{\tau_1(n)+\lfloor \varepsilon c_n \rfloor} - \tilde{S}_{\tau_1(n)} < c_n - T_{\tau_1(n)} \right\}. \end{aligned} \quad (4.8)$$

By the classical strong law of large numbers for $(\tilde{S}_k)_{k \in \mathbb{N}_0}$ and in view of (4.1)

$$\lim_{n \rightarrow \infty} \frac{\tilde{S}_{\tau_1(n)+\lfloor \varepsilon c_n \rfloor} - \tilde{S}_{\tau_1(n)}}{c_n} = \varepsilon q m^\eta > 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{T_{\tau_1(n)}}{c_n} = 1 \quad \text{a. s.}$$

Therefore, with probability one, the event in (4.8) occurs only for finitely many n . \square

Proof of Theorem 3.4. Note that for $i = 0, 1$ we can write

$$\begin{aligned} L_n^{(i)} &= \sum_{k=1}^n \mathbb{1}_{\{X_k + Y_k > 0, I_k = i\}} = \sum_{k=1}^{(\tau_1(n)-1) \wedge n} \mathbb{1}_{\{\xi_k + \eta_k > 0, I_k = i\}} \\ &\quad + \sum_{k=\tau_1(n)}^{\tau_2(n) \wedge n} \mathbb{1}_{\{X_k + Y_k > 0, I_k = i\}} + \sum_{k=\tau_2(n)+1}^n \mathbb{1}_{\{\theta_k > 0, I_k = i\}}. \end{aligned} \quad (4.9)$$

From this representation all the claims follow immediately from the classical strong law of large numbers and the fact

$$\lim_{n \rightarrow \infty} \frac{\tau_1(n)}{c_n} = \lim_{n \rightarrow \infty} \frac{\tau_2(n)}{c_n} = \frac{1}{m^\eta} \quad \text{a. s.},$$

which is a consequence of (4.1) and Lemma 4.2. \square

4.2. Proof of Theorem 3.9. Define $X_n \in D([0, \infty), \mathbb{R}^5)$ and $Y_n \in D([0, \infty), \mathbb{R}^2)$ via

$$\begin{aligned} X_n(t) &= \sum_{k=1}^{\lfloor nt \rfloor} \left(\mathbb{1}_{\{\xi_k + \eta_k > 0, I_k = 0\}}, \mathbb{1}_{\{\xi_k + \eta_k > 0, I_k = 1\}}, I_k, \xi_k, \eta_k \right) \\ \text{and } Y_n(t) &= \sum_{k=1}^{\lfloor nt \rfloor} \left(\mathbb{1}_{\{\theta_k > 0, I_k = 0\}}, \mathbb{1}_{\{\theta_k > 0, I_k = 1\}} \right) \end{aligned}$$

for $t \geq 0$. The following proposition is the key ingredient in the proof of Theorem 3.9.

Proposition 4.3. *If the assumptions of Theorem 3.9 hold, then, as $n \rightarrow \infty$,*

$$\left(\frac{X_n(t) - nt((1-p)p_0, pp_1, p, m^\xi, m^\eta)}{\sqrt{n}}, \frac{Y_n(t) - nt((1-p)p_\theta, pp_\theta)}{\sqrt{n}} \right)_{t \geq 0} \Longrightarrow (\mathbf{B}(t))_{t \geq 0},$$

where $(\mathbf{B}(t))_{t \geq 0}$ is a centered Brownian motion with covariance matrix \mathbf{V} as in (3.13).

Proof. The convergence follows from Donsker's invariance principle since $(X_n(t), Y_n(t))$ is the sum of independent identically distributed random vectors in \mathbb{R}^7 with finite second moments. The increment vectors have expectation

$$\begin{aligned} \mathbb{E}[\left(\mathbb{1}_{\{\xi_k + \eta_k > 0, I_k = 0\}}, \mathbb{1}_{\{\xi_k + \eta_k > 0, I_k = 1\}}, I_k, \xi_k, \eta_k, \mathbb{1}_{\{\theta_k > 0, I_k = 0\}}, \mathbb{1}_{\{\theta_k > 0, I_k = 1\}} \right)] \\ = (p_0(1-p), p_1 p, p, m^\xi, m^\eta, p_\theta(1-p), p_\theta p). \end{aligned}$$

The explicit form of the covariance matrix now results from elementary, yet cumbersome calculations. For example the entry at the first row and fourth column of \mathbf{V} can be calculated as follows:

$$\begin{aligned} \text{Cov}[\mathbb{1}_{\{\xi_k + \eta_k > 0, I_k = 0\}}, \xi_k] &= \mathbb{E}[\xi_k \mathbb{1}_{\{\xi_k + \eta_k > 0, I_k = 0\}}] - m^\xi \mathbb{P}(\xi_k + \eta_k > 0, I_k = 0) \\ &= \mathbb{E}[\xi_k \mathbb{1}_{\{\xi_k > 0, I_k = 0\}}] - m^\xi (1-p)p_0 \\ &= \mathbb{E}[\xi_k \mathbb{1}_{\{I_k = 0\}}] - m^\xi (1-p)p_0 \\ &= (1-p)(m_0^\xi - m^\xi p_0). \end{aligned}$$

□

In what follows, we shall frequently use the following two facts (see the Lemma on p. 151 in [1] and [4, Theorem 3.1]):

Fact 1: the addition mapping $+: D([0, \infty), \mathbb{R}) \times D([0, \infty), \mathbb{R}) \mapsto D([0, \infty))$ defined by $(f+g)(x) = f(x) + g(x)$, is continuous with respect to the J_1 -topology at all points (f, g) such that both f and g are continuous;

Fact 2: the composition mapping $\circ: D([0, \infty), \mathbb{R}) \times D([0, \infty), \mathbb{R}) \mapsto D([0, \infty))$ defined by $(f \circ g)(x) = f(g(x))$ is continuous with respect to the J_1 -topology at all points (f, g) such that both f and g are continuous and g is nondecreasing.

Proof of Theorem 3.9. From [5, Corollary 7.3.1], (2.1) and Fact 2, we infer

$$\left(\frac{\tau_1(nt) - \frac{1}{m^\eta} c_{\lfloor nt \rfloor}}{(m^\eta)^{-3/2} \sqrt{c_n}} \right)_{t \geq 0} \implies (-B_5(t^\rho))_{t \geq 0}, \quad (4.10)$$

and, in view of Lemma 4.1, the same relation for $\tau_2(nt)$. By applying Corollary 13.8.1 in [5] and again Lemma 4.1, we can further extend the convergence in Proposition 4.3 to a joint convergence:

$$\begin{aligned} &\left(\frac{X_n(t) - nt(p_0(1-p), pp_1, p, m^\xi, m^\eta)}{\sqrt{n}}, \frac{Y_n(t) - nt((1-p)p_\theta, pp_\theta)}{\sqrt{n}}, \right. \\ &\quad \left. \frac{\tau_1(nt) - \frac{1}{m^\eta} c_{\lfloor nt \rfloor}}{(m^\eta)^{-3/2} \sqrt{c_n}}, \frac{\tau_2(nt) - \frac{1}{m^\eta} c_{\lfloor nt \rfloor}}{(m^\eta)^{-3/2} \sqrt{c_n}} \right)_{t \geq 0} \implies (\mathbf{B}(t), -B_5(t^\rho), -B_5(t^\rho))_{t \geq 0}. \quad (4.11) \end{aligned}$$

As in (4.9), for $i = 0, 1$, we can write

$$\begin{aligned} L_{\lfloor nt \rfloor}^{(i)} &= \sum_{k=1}^{\lfloor nt \rfloor} \mathbb{1}_{\{X_k + Y_k > 0, I_k = i\}} = \sum_{k=1}^{(\tau_1(nt)-1) \wedge \lfloor nt \rfloor} \mathbb{1}_{\{\xi_k + \eta_k > 0, I_k = i\}} \\ &\quad + \sum_{k=\tau_1(nt) \wedge \lfloor nt \rfloor + 1}^{\tau_2(nt) \wedge \lfloor nt \rfloor} \mathbb{1}_{\{X_k + Y_k > 0, I_k = i\}} + \sum_{k=\tau_2(nt) \wedge \lfloor nt \rfloor + 1}^{\lfloor nt \rfloor} \mathbb{1}_{\{\theta_k > 0, I_k = i\}}. \end{aligned}$$

The second summand above is bounded by $\tau_2(nt) - \tau_1(nt) + 1$ and thus the supremum over t in a compact interval divided by $\sqrt{c_n} = O(\sqrt{n})$ converges to zero in probability by Lemma 4.1 as $n \rightarrow \infty$. The behavior of the second and the third summand strongly depends on whether $\lim_{n \rightarrow \infty} \frac{c_n}{n}$ is smaller, larger or equal to $\frac{1}{m^\eta}$.

Given $0 < a < b$ and $i = 1, 2$ we put

$$A_n^{a,b,i} := \{\tau_i(nt) \leq \lfloor nt \rfloor \text{ for all } t \in [a, b]\} \quad \text{and} \quad B_n^{a,b,i} := \{\tau_i(nt) \geq \lfloor nt \rfloor + 1 \text{ for all } t \in [a, b]\}.$$

We first deal with the case $c_\infty \in (0, \frac{1}{m^\eta})$. In this case, we necessarily have $\rho = 1$. By Eq. (4.10), Lemma 4.1 and the uniform convergence theorem for regularly varying functions we obtain for $i = 1, 2$ and arbitrary $0 < a < b$

$$\sup_{t \in [a, b]} \left| \frac{\tau_i(nt)}{c_n} - \frac{t}{m^\eta} \right| \xrightarrow{\mathbb{P}} 0 \quad \text{as } n \rightarrow \infty, \quad (4.12)$$

and therefore

$$\lim_{n \rightarrow \infty} \mathbb{P}(A_n^{a, b, 1} \cap A_n^{a, b, 2}) = 1. \quad (4.13)$$

For $i = 1, 2$, put

$$\widehat{L}_{\lfloor nt \rfloor}^{(i)} = \sum_{k=1}^{\tau_1(nt)-1} \mathbb{1}_{\{\xi_k + \eta_k > 0, I_k = i\}} + \sum_{k=\tau_2(nt)}^{\lfloor nt \rfloor} \mathbb{1}_{\{\theta_k > 0, I_k = i\}}. \quad (4.14)$$

From what have proved above it is clear that it suffices to check (3.15) with $L_{\lfloor nt \rfloor}^{(i)}$ replaced by $\widehat{L}_{\lfloor nt \rfloor}^{(i)}$, $i = 1, 2$. For typographical reasons we shall write convergences of various components in separate formulas keeping in mind that they actually converge jointly in view of (4.11). Firstly,

$$\left(\frac{\sum_{k=1}^{\lfloor c_n t \rfloor - 1} \mathbb{1}_{\{\xi_k + \eta_k > 0, I_k = 0\}} - c_n t p_0 (1-p)}{\sqrt{c_n}}, \frac{\sum_{k=1}^{\lfloor c_n t \rfloor - 1} \mathbb{1}_{\{\xi_k + \eta_k > 0, I_k = 1\}} - c_n t p p_1}{\sqrt{c_n}} \right)_{t \geq 0} \Rightarrow (B_1(t), B_2(t))_{t \geq 0} \quad \text{as } n \rightarrow \infty,$$

and therefore using Fact 2 (continuity of the composition mapping), the continuous mapping theorem and (4.12)

$$\left(\frac{\sum_{k=1}^{\tau_1(nt)-1} \mathbb{1}_{\{\xi_k + \eta_k > 0, I_k = 0\}} - \tau_1(nt) p_0 (1-p)}{\sqrt{c_n}}, \frac{\sum_{k=1}^{\tau_1(nt)-1} \mathbb{1}_{\{\xi_k + \eta_k > 0, I_k = 1\}} - \tau_1(nt) p p_1}{\sqrt{c_n}} \right)_{t > 0} \Rightarrow \left(B_1\left(\frac{t}{m^\eta}\right), B_2\left(\frac{t}{m^\eta}\right) \right)_{t > 0} \quad \text{as } n \rightarrow \infty.$$

Secondly, using Fact 1 (continuity of addition) and convergence of the last components in (4.11) we deduce

$$\left(\frac{\sum_{k=1}^{\tau_1(nt)-1} \mathbb{1}_{\{\xi_k + \eta_k > 0, I_k = 0\}} - p_0(1-p) \frac{1}{m^\eta} c_{\lfloor nt \rfloor}}{\sqrt{c_n}}, \frac{\sum_{k=1}^{\tau_1(nt)-1} \mathbb{1}_{\{\xi_k + \eta_k > 0, I_k = 1\}} - p p_1 \frac{1}{m^\eta} c_{\lfloor nt \rfloor}}{\sqrt{c_n}} \right)_{t > 0} \Rightarrow \left(B_1\left(\frac{t}{m^\eta}\right) - p_0(1-p)(m^\eta)^{3/2} B_5(t), B_2\left(\frac{t}{m^\eta}\right) - p p_1 (m^\eta)^{3/2} B_5(t) \right)_{t > 0} \quad \text{as } n \rightarrow \infty. \quad (4.15)$$

In the same vein,

$$\left(\frac{\sum_{k=1}^{\tau_2(nt)-1} \mathbb{1}_{\{\theta_k > 0, I_k = 0\}} - p_\theta(1-p) \frac{1}{m^\eta} c_{\lfloor nt \rfloor}}{\sqrt{c_n}}, \frac{\sum_{k=1}^{\tau_2(nt)-1} \mathbb{1}_{\{\theta_k > 0, I_k = 1\}} - p p_\theta \frac{1}{m^\eta} c_{\lfloor nt \rfloor}}{\sqrt{c_n}} \right)_{t > 0} \Rightarrow \left(B_6\left(\frac{t}{m^\eta}\right) - p_\theta(1-p)(m^\eta)^{3/2} B_5(t), B_7\left(\frac{t}{m^\eta}\right) - p p_\theta (m^\eta)^{3/2} B_5(t) \right)_{t > 0} \quad \text{as } n \rightarrow \infty \quad (4.16)$$

Replacing $\sqrt{c_n}$ in the denominators by $\sqrt{c_\infty n}$ and summing everything up we get

$$\begin{aligned} & \left(\frac{L_{[nt]}^{(0)} - (1-p)\frac{1}{m^n}(p_0 - p_\theta)c_{[nt]} - nt(1-p)p_\theta}{\sqrt{n}}, \frac{L_{[nt]}^{(1)} - p\frac{1}{m^n}(p_1 - p_\theta)c_{[nt]} - ntpp_\theta}{\sqrt{n}} \right)_{t>0} \\ & \implies \left(\sqrt{c_\infty} \left(B_1\left(\frac{t}{m^n}\right) - p_0(1-p)(m^n)^{3/2}B_5(t) - B_6\left(\frac{t}{m^n}\right) + p_\theta(1-p)(m^n)^{3/2}B_5(t) \right) + B_6(t), \right. \\ & \quad \left. \sqrt{c_\infty} \left(B_2\left(\frac{t}{m^n}\right) - pp_1(m^n)^{3/2}B_5(t) - B_7\left(\frac{t}{m^n}\right) + pp_\theta(m^n)^{3/2}B_5(t) \right) + B_7(t) \right)_{t>0} \quad (4.17) \end{aligned}$$

as $n \rightarrow \infty$. It remains to note that (4.17) holds jointly with

$$\left(\frac{N_{[nt]}^{(0)} - (1-p)nt}{\sqrt{n}}, \frac{N_{[nt]}^{(1)} - pnt}{\sqrt{n}} \right)_{t \geq 0} \implies (-B_3(t), B_3(t))_{t \geq 0},$$

and together this is (3.15).

CASE $c_\infty = 0$. In this case (4.13) still holds but there are significant simplifications. First of all note that in this case we can have $\rho \leq 1$ and thus in (4.12) $\frac{1}{m^n}t$ must be replaced by $\frac{1}{m^n}t^\rho$. Further, in (4.15) and (4.16) upon replacing $\sqrt{c_n}$ in the denominator by \sqrt{n} the limit becomes identical zero and, thus we have the same convergence (4.17) but with $c_\infty = 0$ on the right-hand side.

Finally, we deal with the case $c_\infty > \frac{1}{m^n}$. In this case (4.12) implies

$$\lim_{n \rightarrow \infty} \mathbb{P}(B_n^{a,b,1} \cap B_n^{a,b,1}) = 1.$$

Similar to (4.14), for $i = 1, 2$, we now put

$$\widehat{L}_{[nt]}^{(i)} := \sum_{k=1}^{[nt]} \mathbb{1}_{\{\xi_k + \eta_k > 0, I_k = i\}},$$

and note that now $L_{[nt]}^{(i)}$ can be replaced by $\widehat{L}_{[nt]}^{(i)}$ in (3.16). After this replacement (3.16) is just a part of (4.11). \square

We now turn to the proof of Theorem 3.1. As a first step, we notice that Corollary 3.7 follows from Theorem 3.5 and the continuous mapping theorem. Theorem 3.5, in turn, follows immediately from Theorem 3.9. It thus remains to deduce Theorem 3.1 from Theorem 3.5.

Proof of Theorem 3.1. Our aim is to show how to calculate the distribution of the variable

$$\begin{aligned} \chi_\infty^2 & := \frac{(d_2 + G_2 - p_\theta G_1 - (1-p)(d_2 + d_3 + G_2 + G_3))^2}{(1-p_\theta)p_\theta(1-p)} + \frac{(d_3 + G_3 + p_\theta G_1 - p(d_2 + d_3 + G_2 + G_3))^2}{(1-p_\theta)p_\theta p} \\ & = \frac{(pd_2 - (1-p)d_3 - p_\theta G_1 + pG_2 - (1-p)G_3)^2}{(1-p_\theta)p_\theta(1-p)} + \frac{((1-p)d_3 - pd_2 + p_\theta G_1 - pG_2 + (1-p)G_3)^2}{(1-p_\theta)p_\theta p}, \end{aligned}$$

where $\mathbf{G} := (G_1, G_2, G_3)$ is a centered Gaussian vector with covariance matrix \mathbf{V}_1 . A simple calculation shows that

$$\chi_\infty^2 = \frac{(p_\theta G_1 - pG_2 + (1-p)G_3 - (pd_2 - (1-p)d_3))^2}{(1-p_\theta)p_\theta(1-p)p}.$$

Note that $p_\theta G_1 - pG_2 + (1-p)G_3$ has a centered normal distribution with the variance

$$\begin{aligned} \text{Var}[p_\theta G_1 - pG_2 + (1-p)G_3] &= p_\theta^2 \text{Var}[G_1] + p^2 \text{Var}[G_2] + (1-p)^2 \text{Var}[G_3] - 2pp_\theta \text{Cov}[G_1, G_2] \\ &\quad + 2p_\theta(1-p) \text{Cov}[G_1, G_3] - 2p(1-p) \text{Cov}[G_2, G_3] \\ &= p_\theta(1-p_\theta)p(1-p). \end{aligned}$$

Thus, for \mathcal{N} having the standard normal distribution, we see that

$$\chi_\infty^2 \stackrel{\text{law}}{=} \left(\mathcal{N} - \frac{pd_2 - (1-p)d_3}{\sqrt{pp_\theta(1-p)(1-p_\theta)}} \right)^2 = \left(\mathcal{N} - \frac{d_\infty(p_0 - p_1)\sqrt{p(1-p)}}{m^n \sqrt{p_\theta(1-p_\theta)}} \right)^2.$$

The proof is complete. \square

5. CONCLUSIONS

Starting from the observation that the standard for testing product changes on e-commerce websites is large scale hypothesis testing with statistical tests based on the assumption of independent samples such as the chi-squared test, we have suggested a new model for the samples which incorporates shared inventories. This model introduces new dependencies. Our main result is the calculation of the asymptotic law of the chi-squared test statistics under the new model assumptions in the critical regime where the number of items of a popular good is of the order of the square root of the sample size. Website versions that greedily sell the popular good have an initial advantage in the number of sales of the order of the square root of the sample size, which is the order of the overall random fluctuations. Thus the initial advantage has an impact on the probability of rejecting the hypothesis. We have demonstrated in examples that this may lead to both, arbitrarily high false-positive as well as arbitrarily high false-negative rates. This questions the assumption implicit in the industry standard that dependencies are small enough to be ignored. Moreover, it suggests that the present standard of A/B testing favors algorithms that are designed to be good in competition against others, but not necessarily good when used on separate inventory.

Our work may be extended in the future in several directions. On the one hand, our results may be used to construct tests for the model that keep the significance level. On the other hand, the model may be extended to incorporate more features of real samples such as a priori information about website visitors etc.

6. ACKNOWLEDGEMENTS

A. M. was supported by the Ulam programme funded by the Polish national agency for academic exchange (NAWA), project no. PPN/U LM/2019/1/00004/DEC/1. The authors would like to thank Tanja Matic and Onno Zoeter for many helpful discussions. We are particularly grateful to Onno Zoeter for communicating the essence of Example 3.8.

REFERENCES

- [1] Patrick Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999. A Wiley-Interscience Publication.
- [2] N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular variation*, volume 27 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1987.
- [3] A. W. van der Vaart. *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998.
- [4] Ward Whitt. Some useful functions for functional limit theorems. *Math. Oper. Res.*, 5(1):67–85, 1980.

- [5] Ward Whitt. *Stochastic-process limits*. Springer Series in Operations Research. Springer-Verlag, New York, 2002. An introduction to stochastic-process limits and their application to queues.

DENNIS BOHLE, BOOKING.COM B.V., AMSTERDAM, NETHERLANDS
E-mail address: `dennis.bohle@booking.com`

ALEXANDER MARYNYCH, FACULTY OF COMPUTER SCIENCE AND CYBERNETICS, TARAS SHEVCHENKO NATIONAL UNIVERSITY OF KYIV, UKRAINE
E-mail address: `marynych@unicyb.kiev.ua`

MATTHIAS MEINERS, MATHEMATISCHES INSTITUT, UNIVERSITÄT GIESSEN, GERMANY
E-mail address: `matthias.meiners@math.uni-giessen.de`