Evaluation of delivery time on e-retailer app through Statistical (Bayesian) Inference

Mauricio Nascimento

December 2022

Abstract

The present work intends to use expert knowledge and samples of different sizes to evaluate the probability of delivering an order within promised time for a given e-retailer, testing it against a threshold of 0.9. Although the conclusion was the same for all samples, it was possible to notice that despite the claim to deliver orders below 35 minutes, it is reasonable to expect that less than 90% of the orders will satisfy that condition. It was also possible to observe the contribution of the prior over different sample sizes as well as the decreasing of the evidence value.

1 Introduction

Before the dawn of the Internet, the most common way one would choose to go shopping was to leave his home and effectively go to a shop. After the invention of the Internet, however, more and more convenient retail solutions have emerged, ranging from websites to phone apps. At first, these would compromise on delivering a purchase in a time window of months or weeks.

These time windows have been shrinking over the years, intensified by the Covid Pandemic. Nowadays it's not unusual to have e-retailers focusing on convenience, thus promising to deliver purchases within a few days or even a few minutes. But do they really?

In order to investigate that, a sample of orders, with their respective delivery time, was collected from a e-retailer that promises to deliver purchases below thirty five minutes. Within this sample, given the delivery time of each order it is possible to create a dichotomical variable that indicates if the order was delivered below the promised time or not. Such feature can be described as a random variable that follows the Binomial Distribution with parameters n and θ, where n represents the number of trials and θ describes the probability of success, hence the probability of having an order delivered within the promised delivery time.

Using concepts of Bayesian Inference, the present work aims at estimating a probability distribution for θ using prior knowledge from experts and, after that, update such distribution using the sample. Having the probability distribution, another objective is to infer, at a population level, if the probability of having an ordered delivered below thirty five minutes is greater than 0.90. Finally, having a large dataset of orders available, this work will also evaluate if the conclusion changes as we increase the size of the sample.

2 Dataset

The dataset used for the current work consists of 5000 observations of orders from the evaluated e-retailer. For each order there is information about it's delivery time. Figure1 below shows an example of the dataset.

	order_number	delivery_time_in_minutes
0	27877	23.97
1	5620	35.33
2	72807	77.58
3	14269	28.97
4	84171	26.62

Figure 1: Example of dataset.

3 Methodology

Starting with a sample of 50 orders randomly chosen from the full dataset, the first step is to create a binary variable that indicates if the order was delivered below 35 minutes. In those cases, the variable will receive the value of 1 otherwise, the value will be 0. This set of observations is described by a random variable that follows a Binomial distribution and can be written as $X \sim Bin(n, \theta)$ having n as the number of trials (50 in this case) and θ as the probability of success for each trial, in other words, the probability of having an order delivered below 35 minutes.

The probability mass function for a Binomial distribution can be represented by the equation:

$$f(X|\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x} \tag{1}$$

having x as the number of successes.

Unlike in a Frequentist approach, which assumes that θ is static, a Bayesian approach assumes that θ can be represented by a probability distribution. Moreover, it is possible to chose a distribution such that prior knowledge about the object of study is considered. Since we have access to expert knowledge and a sample of orders, our goal is to use those to infer if θ is greater than 0.90 at population level.

Because θ describes a probability in this case, a Beta Distribution is well suited for a prior distribution. From that we can write:

$$\theta \sim Beta(a, b) \tag{2}$$

$$f(\theta) = \frac{\theta^{a-1}(1-\theta)^{b-1}}{B(a,b)} \tag{3}$$

where $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$

Once defined the likelihood function for the random variable X and the prior for θ it is possible to apply the Bayes Theorem, described by equation (4) below, to update the distribution for θ given the sample.

$$f(\theta|X) = \frac{f(X|\theta)f(\theta)}{f(X)} = \frac{f(X|\theta)f(\theta)}{\int_\Theta f(X|\theta)f(\theta)d\theta} \tag{4}$$

where Θ is the parametric space, such that $\theta \in \Theta, \forall\theta$.

Because the Beta distribution is a conjugated prior for a Binomial likelihood, as Fossaluza demonstrates in Section 3.2 of his class notes[1], it is possible to write the posterior distribution as follows.

$$\begin{gathered} f(\theta|X) \sim Beta(a + \sum_{i=1}^{n} x_i, b + n - \sum_{i=1}^{n} x_i) \implies \\ f(\theta|X) = \frac{\theta^{(a+\sum_{i=1}^n x_i - 1)}(1-\theta)^{(b+n-\sum_{i=1}^n x_i - 1)}}{B((a+\sum_{i=1}^n x_i), (b+n-\sum_{i=1}^n x_i))} \end{gathered} \tag{5}$$

where a and b are the parameters for the prior distribution, n is the sample size and $\sum_{i=1}^n x_i$ is the number of successes in the sample. Therefore, in order to estimate the posterior distribution the parameters for the prior are necessary.

As mentioned before, one of the benefits of the Bayesian approach is the possibility to incorporate external knowledge through the prior distribution. When in touch with business experts regarding e-retailers, such experts claim that if the company promises to deliver orders below thirty five minutes it is reasonable to think that most of the orders are delivered within that time.

From that information it is possible to assume that the expected value for θ should be around 0.75. Moreover it is also likely that the values for that parameter should range from 0.7 to 0.9.

Keeping in mind that the prior for θ is a Beta distribution, we can write:

$$E[\theta|X] = \frac{a}{a+b} = 0.75 \tag{6}$$

$$P(\delta_1 < \theta < \delta_2) = \int_{\delta_1}^{\delta_2} f(\theta)d\theta \geq 1 - \alpha_p \tag{7}$$

where δ_1 and δ_2 are, respectively, the lower and the upper limits for θ and $\alpha_p = 0.01$ is the significance level for the prior considered for this work.

After solving the system described by Equations (6) and (7) above, it is possible to calculate parameters a and b, thus describing a prior distribution from the Beta family that incorporates expert knowledge. Plugging in the parameters for the prior and from the sample in Equation (5) it is possible to estimate the posterior distribution.

Since one of the objectives is to infer if the probability of having an ordered delivered in less than 35 minutes is greater than 0.90, we can enunciate the following hypothesis test:

Hypothesis $\mathbf{H_0}$: $\theta \geq 0.90$

Hypothesis $\mathbf{H_1}$: $\theta < 0.90$

Considering $\alpha_t = 0.01$ as the level of significance for the test a Full Significance Bayesian Test (FBST) can be applied to evaluate H_0 and H_1, since this is a precise hypothesis. The test consists in finding a tangent region such that:

$$T_x = \left\{ \theta \in \Theta : f(\theta|X) \geq \sup_{\theta \in \Theta_0} f(\theta|X) \right\} \tag{8}$$

where Θ is the parametric space and Θ_0 is the parametric space for the null hypothesis. Having the tangent region it is possible to calculate the evidence measure of Pereira-Stern (e-value) as follows:

$$E_v(\Theta_0, X) = 1 - P(\theta \in T_x|X) \tag{9}$$

If the e-value is smaller than α_t the null hypothesis is rejected.

In order to investigate how the sample size affects the conclusion drawn from the inference, the same procedure described above for n=50 was repeated for n=500 and n=5000.

4 Results

With the raw dataset in hand, the first step was to create a binary variable which indicates whether the order was delivered within promised time or not. After doing so the dataset looked as shown in Figure 2 below.

	order_number	delivery_time_in_minutes	within_promised_time
0	27877	23.97	1
1	5620	35.33	0
2	72807	77.58	0
3	14269	28.97	1
4	84171	26.62	1

Figure 2: Dataset after building binary variable.

The proportion of orders delivered within promised time varies according to the sample size and can be seen in Figure 3. However, in all of the cases the proportion is greater than what was expected by experts but smaller than the threshold of 0.9. It is reasonable to assume that the posterior distribution will reflect this behaviour, with more contribution of the sample as we increase the sample size.

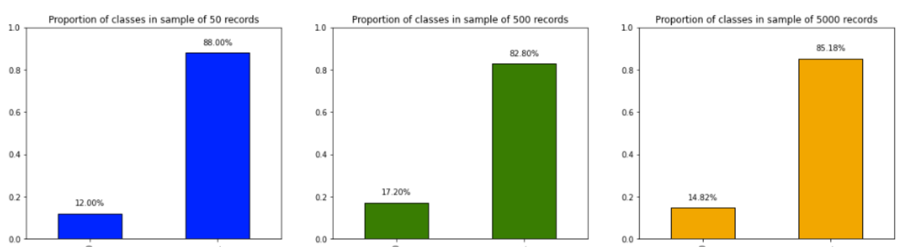

Figure 3: Proportion of classes in samples of different sample sizes.

The following step was to incorporate expert knowledge in the form of a prior distribution for θ, in other words, to find parameters a and b for the prior. Knowing that the expected value for θ should be around 0.75 and could range from 0.7 and 0.9, Equation (7) was used to write a as a function of b and $E[\theta|X]$. Later a computational method was used to find b that satisfies the conditions described in Equation (8), having $\delta_1 = 0.7$ and $\delta_2 = 0.9$. The code for such method can be found in the repository displayed in the appendix.

After doing so, it was found that the distribution $\theta \sim Beta(57, 19)$ was a suitable prior. Having the prior and the sample of orders it was possible to use Equation (5) in order to update our belief in the form of a posterior distribution for θ. Figure 4 shows a comparision between the prior and posterior distributions for different sample sizes.

It is noticeable how, as we increase the sample size, posterior distribution

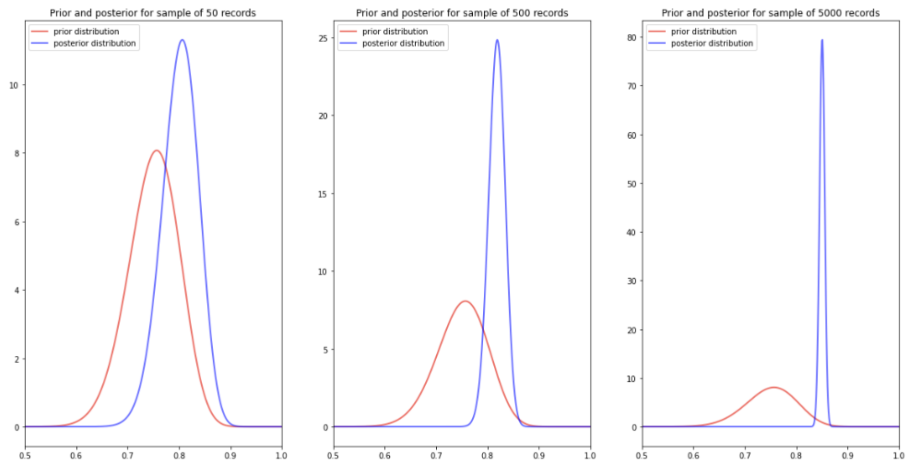

Figure 4: Comparison between prior and posterior for different sample sizes.

gets thinner and taller when compared to the prior. This was expected and indicates less variability regarding parameter θ, thus, more certainty.

The posterior is a more robust distribution for the probability of having an ordered delivered below 35 minuites (θ) since it combines expert knowledge and observations from the sample. The next step was to use such distribution to make inferences about θ, namely, to evaluate H_0 and H_1.

In order to do so, the FBST was applied. The first step was to find the Tangent Region, described by Equation (8). Since the posterior is a distribution of the Beta family, it is possible to write:

$$\sup_{\theta \in \Theta_0} f(\theta|X) = f(\theta_0|X) \tag{10}$$

where θ_0 is the value of θ for H_0. And since $\theta_0 \neq E[\theta|X]$ there is, necessarily, $\theta_0^{'}$ such that

$$f(\theta_0|X) = f(\theta_0^{'}|X) \tag{11}$$

Therefore, to find T_x is to find $\theta_0^{'}$ that satisfies the condition described in Equation (11). Figure 5 below shows the tangent region of the posterior for different samples sizes.

Again it is noticeable that the tangent region gets narrower as we increase the sample size. Because variability decreases with larger samples, probability densities get more concentrated around the expected value, shrinking the length of T_x.

Finally it was possible to estimate the e-value for each sample through Equation (9) and draw a conclusion about H_0 and H_1. Figure 6 compiles the main results from the analysis, including the e-value for each sample.

Because the results for e-value for every sample were smaller than α_t, H_0 should be rejected in every case. Therefore it is possible to infer that, at population level, the probability of having an order delivered by such e-retailer in

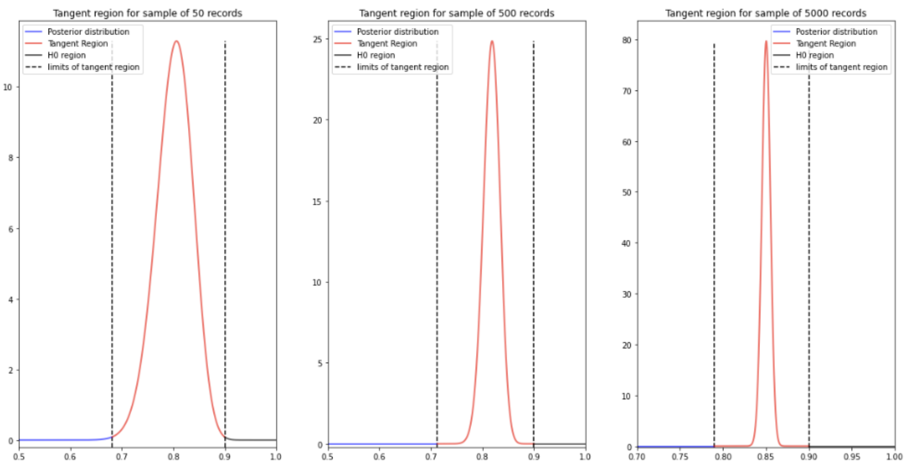

Figure 5: Comparison between prior and posterior for different sample sizes.

	a_posterior	b_posterior	expected_value	tg_region	e_value
50	101	25	0.802	(0.68, 0.9)	1.71e-03
500	471	105	0.818	(0.71, 0.9)	3.61e-09
5000	4316	760	0.850	(0.79, 0.9)	1.11e-16

Figure 6: Compilation of the results of the analysis for each sample.

less than 35 minutes is smaller than 0.9. It is important to notice, however, how the e-value gets considerably smaller as sample size increases. As discussed before, this is due to a lesser variability as sample size is increased. The value of θ_0 would have to be way closer to the expected value of the posterior in the sample of 5000 records than in the sample of 50 records in order for H_0 not to be rejected.

Another interesting aspect to point out is how the expected value approaches the propotion of the class 1 as the sample size increases. This indicates a smaller influence of the prior on the result of the posterior distribution, reinforcing one of the pros of the Bayesian approach, which is the ability to incorporate prior knowledge, specially for small samples.

5 Conclusion

Claiming to deliver orders in ever smaller times is deffinitely an appealing marketing strategy. However, as shown by the present work, for a specific e-retailer, the probability of having an order delivered within the promised time is smaller than 90%, casting doubts over such attractive claim.

The conclusion was the same among samples of different sizes even though the evidence value was lower for smaller samples. It was also possible to realize that the contribution of the prior was more significant on the sample of 50 records whereas on the sample of 5000 records it had almost no influence since the expected value for the posterior was pretty close to the proportion of successes in the sample.

6 Appendix

All of the codes utilized in order to develop the present analysis can be found at this GitHub repository: `https://github.com/mrnascimento/aritgo-inf-bayes`

References

[1] Class notes from Bayesian Inference subject at IME-USP by Victor Fossaluza. `https://vfossaluza.github.io/InfBayes/Bayes.html`.