International Journal of Scientific Reports
Nain A. Int J Sci Rep. 2024 Mar;10(3):75-80
http://www.sci-rep.com
pISSN 2454-2156 | eISSN 2454-2164

DOI: https://dx.doi.org/10.18203/issn.2454-2156.IntJSciRep20240444

Original Research Article

Modeling the number of clothing merchandise returned through online shopping using a mixture of Poisson distribution: an interdisciplinary approach

Anuj Nain*

Department of Statistics, Central University of Rajasthan, Ajmer, Rajasthan, India

Received: 23 January 2024
Revised: 13 February 2024
Accepted: 14 February 2024

***Correspondence:**
Dr. Anuj Nain,
E-mail: anuj.nain.stats@gmaul.com

Copyright: © the author(s), publisher and licensee Medip Academy. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT

Background: Online shopping has witnessed a significant surge in popularity owing to its accessibility and convenience. The rise of online shopping has revolutionized the retail industry, with clothing purchases garnering a dedicated following. However, a persistent challenge in internet purchasing lies in the uncertainty surrounding the fit, appearance, and quality of clothing merchandise. As a result, customers often resort to returning items that fail to meet their expectations or differ from the website's descriptions. In this study, we aim to address this issue by proposing a novel approach to model the number of items returned through online purchases. Specifically, we employ a mixture of Poisson distribution to capture the complexities of the return process and its underlying factors.
Methods: A mixture of Poisson distribution is employed to model the uncertainties in online clothing returns. The approach aims to capture the complexities of the return process and its underlying factors, providing valuable insights for e-commerce businesses to optimize operations and enhance customer satisfaction.
Results: The findings of this study have the potential to shed light on the reasons behind increased return rates, offering e-commerce businesses valuable information to optimize their operations and enhance customer satisfaction.
Conclusions: Poisson distribution is capable of analyzing the hidden patterns in the data, and hence the results of the study are beneficial for various stakeholders, including statisticians, e-commerce firms, and app developers.

Keywords: Online shopping, Item returns, Clothing merchandise, Poisson distribution, Mixture distribution

INTRODUCTION

In recent years, online shopping has experienced an unprecedented surge in popularity, revolutionizing the retail industry by providing consumers with unparalleled accessibility and convenience. When it comes to buying clothes online, the ease of browsing and purchasing a wide array of products from the comfort of one's own home has led to the rapid growth of e-commerce platforms. However, amidst the remarkable advantages of online shopping, a persistent challenge continues to vex consumers and e-commerce businesses - the uncertainty surrounding the fit, appearance, and quality of products. The inability to physically examine and assess products prior to purchase introduces a level of uncertainty that often leaves customers dissatisfied when their received items fail to meet their expectations or differ significantly from the descriptions provided on the website. As a consequence, a considerable number of customers resort to the process of returning purchased items, creating a complex web of logistics and economic implications for e-commerce businesses.

In response to this pressing issue, this research study endeavors to propose a novel approach to model the number of items returned through online purchases. By tackling this problem, we aim to contribute valuable insights that could ultimately help reduce return rates and enhance overall customer satisfaction in the e-commerce industry. Some notable references in the domain of online shopping are to Vyas et al, Menon et al, Mummalaneni, Park et al, Gupta et al, Arora, Chandra et al, Shenbhagavadivu, Warrier, and Mishra et al.[1-10] In probability theory and statistics, a mixture distribution is a type of probability distribution that arises from combining multiple component distributions. It is constructed by mixing or blending these individual distributions with specified weights or proportions. The mixture distribution is capable of capturing multimodality or clustering in data, thus allowing flexible and effective modeling. For details on Mixture distributions, one may refer to Chandra, and Dempster et al.[11,12]

The Poisson distribution is a probability distribution that models the number of events that occur in a fixed interval of time or space, given a known average rate of occurrence. The distribution is commonly used to model rare events that occur independently of each other over a continuous interval, such as the number of phone calls received at a call center in a given hour, the number of accidents in a day, the number of customers arriving at a store in a certain time period, or the number of products returned per month in online purchases. For more details on Poisson process one may refer to Zhao et al.[13] The Poisson process can be used to model the number of products returned through online shopping. When a customer purchases a product online, there is a certain probability that they may return the product if it does not meet their expectations or if it is damaged during shipping.

Assuming that returns occur randomly over time, we can model the number of returns using a Poisson process. The process assumes that the probability of a return occurring in a given time interval is proportional to the length of the interval. The rate parameter λ of the Poisson process represents the expected number of returns per unit time. By using historical data, we can estimate the rate parameter λ and use it to make predictions about future returns. This can help online retailers to optimize their inventory management, supply chain, and customer service processes to minimize returns and improve customer satisfaction.

The hierarchy of our work is as follows. In section 2, we delve into the methodology of our work. Section 2.1 describes the data under study, while section 2.2 provides a brief introduction to the Poisson distribution. In section 2.3, we give a concise overview of the mixture distribution and in section 2.4 we explain how to formulate the mixture of Poisson distribution. Moving on to section 3, we discuss the data analysis and present our results. In section 4 we present the discussions based on study and finally in section 5, the conclusion has been presented.

METHODS

The data under study

The dataset utilized for this research was restricted to clothing merchandise and exclusively collected from a range of popular online shopping platforms, specifically Nyka, Amazon, Myntra, Flipkart, Zara, and Meesho as these apps are more popular among users. The data collection process encompassed various districts within Delhi, extending from July 2020 to June 2023. The variable of interest was the number of items returned per month. The objective is to evaluate the overall market performance of these apps and comparison amongst each other as well.

The Poisson distribution

The Poisson distribution is an important discrete distribution that represents the number of events that occur in a fixed interval of time or space, given a known average rate of occurrence and assuming that the events happen independently and at a constant average rate. Such as the number of phone calls received at a call center in a given hour, the number of accidents in a day, the number of emails received per day, the number of online products ordered per month, and the number of online products returned per month.[13] If X is a random variable following Poisson distribution. The probability mass function is given by the following equation.

$$P(X = x) = e^{-\lambda} \frac{\lambda^x}{x!'} \qquad x = 0, 1, ...$$

Where λ is a constant and represents the average rate. So, in our case, if X represents the number of items returned in a month, then λ represents the average number of items returned per month.

The mixture distribution

A mixture distribution is a probability distribution that arises from combining multiple component distributions. In a mixture distribution, each component contributes to the overall distribution with a certain weight or mixing proportion. Mixture distributions are employed when the data, instead of arising from a single distribution, arise from a combination of several distributions. The empirical density plot of the data gives us an idea of whether we should go for a mixture distribution or not. When the empirical density plot is multimodal it is beneficial to use a suitable mixture distribution.[11,12]

Let X_1 X_2 . . . X_n be an independent and identically distributed random sample from a K-component finite mixture of probability distributions. This mixture distribution is represented as given.

$$f(x;\theta) =; \sum_{k=1}^{K} \pi_k f_k(x;\theta_K) \qquad (1)$$

Subject to $\sum_{k=1}^{K} \pi_k = 1$

where $\Theta = (\pi', \theta') = (\pi_1, \pi_2 ,. . . , \pi_{k-1}, \theta_1, \theta_2, . . . \theta_k)$ is the vector of unknown parameters and $0 < \pi_i \leq 1$. These K distributions may or may not be from the same family. In this paper, we assume that for the mixture density given in (1) the component densities f_k (.) are from the same family (Poisson family). Further, we implement EM Algorithm for parameter estimation.[12]

The mixture of Poisson distribution

Introducing the notion of mixture distribution under the Poisson family we see analogues to the density in (1). The probability mass function comprising a mixture of K Poisson distributions would be represented as given below.[14]

$$f(x; \Theta) = ;\sum_{k=1}^{K} \pi_k \, e^{-\lambda_k} \frac{\lambda_k{}^{x}}{x!} \qquad x = 0, 1, \ldots \qquad (2)$$

Subject to $\sum_{k=1}^{K} \pi_k = 1$

The rate of occurrence is given by $\lambda = E(X) = \sum_{k=1}^{K} \pi_k \lambda_k$. If the empirical density plot of a non-negative count data shows humps or multimode or a heavy-tailed behavior. In that case, a suitable mixture of Poisson distribution is a preferable choice.

RESULTS

This study employs R software to analyze the return patterns of purchased items, with a focus on clothing merchandise. The analysis of the three figures, viz. Figures 1-3, reveals exciting insights into the return patterns of online purchases from different franchises.

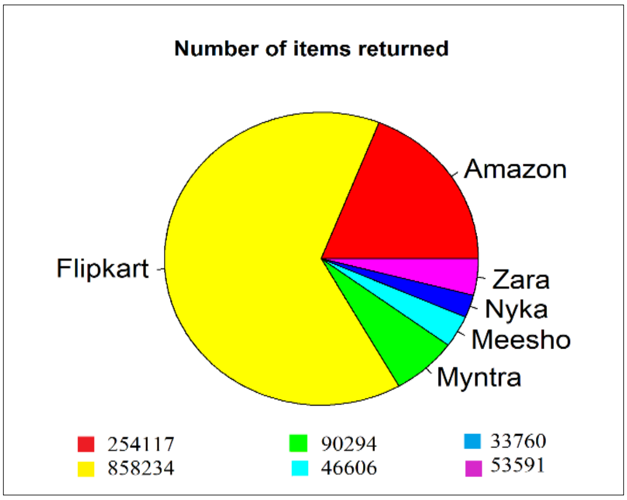

Figure 1: Pie chart of the number of items returned.

In Figure 1, a pie chart is presented, illustrating the absolute number of items returned after online purchases from various franchises. From the chart, it can be observed that Flipkart (858234) has the highest number of products returned among all the listed franchises. This indicates that customers are returning many items purchased from Flipkart.

Figure 2 introduces a more insightful metric by calculating the ratio of returned products to the total number of purchases for each franchise. This ratio provides a better measure to assess the return behavior relative to the overall sales volume. The pie chart in Figure 2 shows that even after accounting for the total number of purchases made on each platform, Flipkart (0.427) still leads with the highest ratio of returned products. This suggests that the return rate for Flipkart is relatively high compared to the other franchises.

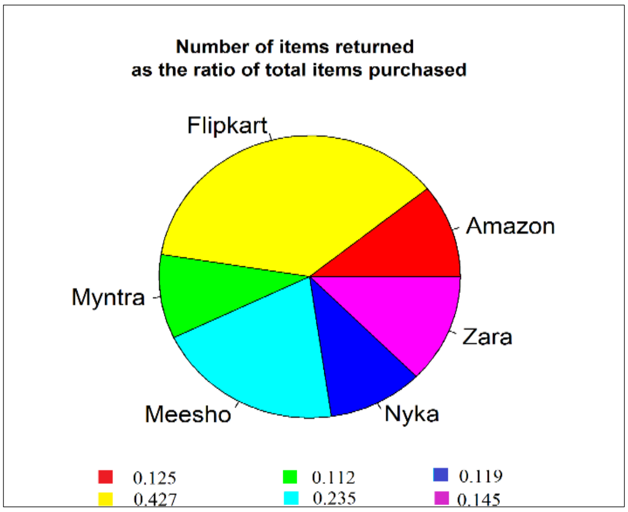

Figure 2: Pie chart of total returns as a ratio of total purchases.

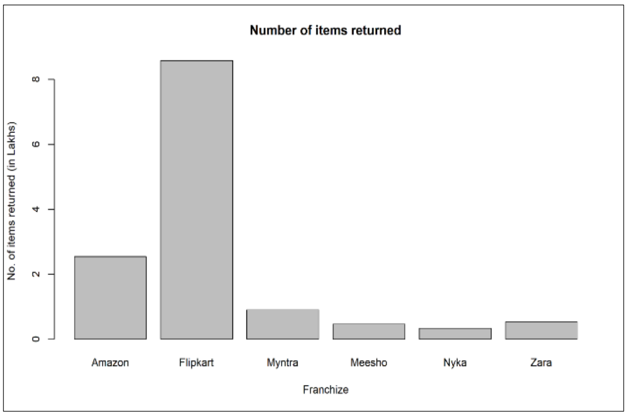

Figure 3: Bar plot of individual franchise returns.

Figure 3 further elaborates on the return patterns by presenting a bar plot that showcases the individual performance of each franchise in terms of the number of products returned. The plot ranks the franchises in descending order based on their return rates. As observed, Flipkart retains the top position with the highest number of returned items. Following Flipkart, Amazon, Myntra, Zara, Meesho, and Nykaa follow in descending order of return frequency.

Overall, the data analysis highlights that Flipkart experiences the highest number of returned products, even when considering the ratio of returns to total purchases. This finding may prompt further investigation into the underlying factors contributing to the relatively higher return rates on Flipkart's platform compared to its competitors. Businesses, especially e-commerce platforms, can use these valuable insights to optimize their operations, improve product descriptions, enhance customer satisfaction, and minimize return rates, ultimately leading to a more successful and efficient online shopping experience.

Figure 4 holds an essential realm from the model-building perspective. This figure shows the empirical density plot of the data. We can see that the density is trimodal as it contains three humps. This implies that a mixture of three distributions would be a better choice to fit the observed data than a single distribution. The descriptive and empirical analysis of our data suggests using a mixture of three Poisson distributions rather than single Poisson distribution. To facilitate the comparison better we fit mixture distribution for K=1, 2, 3 and 4. It is to be noted that by K=1 mean, Single Poisson density. Table 1 represents the summary of fitted models.

The analysis of Table 1 confirms that the model with three Poisson components offers the most suitable fit for the data, as indicated by the lowest values for AIC (454) and BIC (441). This finding aligns with the observations from Figure 4. From Table 1 we estimate of average number of products returned as $\hat{\lambda}$=6200.824 (section 3.4). Figure 5 shows how well the fitted density aligns with the empirical density along with the K-S test of goodness of fit. It can be seen how well the fitted 3-K Poisson density covers the empirical density, representing the accuracy of parameter estimates. Also the K-S test statistic (0.001) is less than the critical value of 0.227 justifying the goodness of fit of 3-K Poisson density.

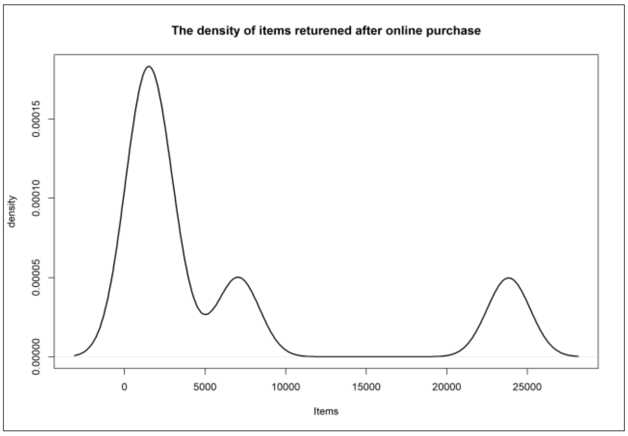

Figure 4: Empirical density plot.

Table 1: Model summary along with parameter estimates.

Number of components in the model (K)	AIC	BIC	Parameter estimates
1	564	542	$\hat{\lambda}$ = 6187.972
2	501	472	$\hat{\lambda}_1$ = 23839.83, $\hat{\pi}_1$ = 0.166 , $\hat{\lambda}_2$ = 2657.60 , $\hat{\pi}_2$ = 0.834
3	454	441	$\hat{\lambda}_1$ = 1537.299 , $\hat{\pi}_1$ = 0.667 , $\hat{\lambda}_2$ = 7058.806 , $\hat{\pi}_2$ = 0.167
			$\hat{\lambda}_3$ = 23839.833 , π_3 = 0.167
4	479	465	$\hat{\lambda}_1$ = 1240.343 , $\hat{\pi}_1$ = 0.495 , $\hat{\lambda}_2$ = 2508.167 , $\hat{\pi}_2$ = 0.148
			$\hat{\lambda}_3$ = 7058.806 , $\hat{\pi}_2$ = 0.167 , $\hat{\lambda}_4$ = 23839.833 , $\hat{\pi}_4$ = 0.190

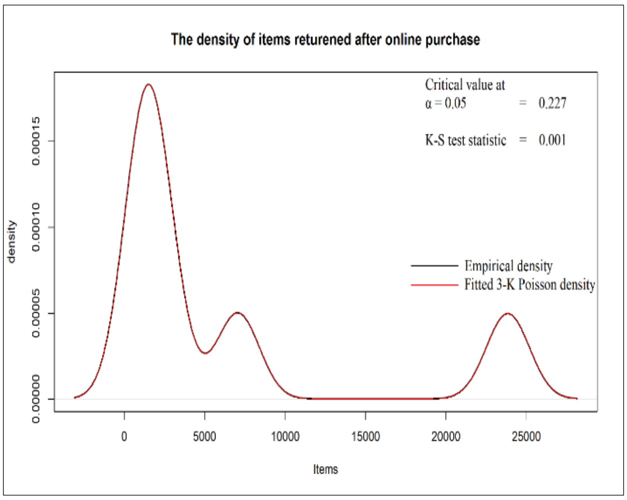

Figure 5: Empirical density and fitted 3-k Poisson density.

DISCUSSION

Based on the findings of the descriptive study and the distribution fitting of the data, several key suggestions can be made for various stakeholders.

Flipkart's market strategies

As observed from the analysis, Flipkart experiences the highest number of product returns among the listed franchises. This indicates a need for Flipkart to review and improve its market strategies. Identifying the reasons behind the higher return rates and addressing them strategically can help optimize their operations and reduce return instances.

Quality of clothing merchandise

The study suggests that improving the quality of clothing merchandise offered by Flipkart could be beneficial. A focus on ensuring accurate product descriptions, high-quality images, and detailed sizing information may help mitigate customer uncertainties, leading to reduced returns.

Understanding subpopulations

The distribution fitting analysis reveals that three distinct clusters or subpopulations dominate the overall density of product returns. With 66.7% ($\hat{\pi}_1$) returns occurring from one cluster having mean 1537.299 ($\hat{\lambda}_1$), 16.7% ($\hat{\pi}_2$) returns from another cluster with mean 7058.806 ($\hat{\lambda}_2$) and 16.7% (π_3) returns from the third cluster with mean 23839.833 ($\hat{\lambda}_3$). The overall estimate of the average number of returns per month $\hat{\lambda}$=6200.824. Recognizing and understanding these subpopulations can provide valuable insights into the different customer segments and their return behaviors.

Targeted strategies

Tailoring marketing and customer service strategies to address the specific needs of each subpopulation can be beneficial. Different segments may have unique preferences or concerns, and catering to these differences can enhance customer satisfaction and reduce returns.

App developing

Pay close attention to the user interface and experience of the app, especially during the product selection process. Ensure that product images are clear and high-quality, and provide detailed product descriptions, including sizing information and material specifications. A user-friendly and informative UI/UX can help customers make more informed decisions, reducing the likelihood of returns.

Statistical modeling

From the point of view of statistical modeling, this research significantly contributes to the field of e-commerce analytics. Modeling the data with a finite mixture of Poisson distribution offers a powerful tool for understanding the complexities of return processes and providing actionable insights to businesses and decision-makers. Understanding the distinct subpopulations and their respective means (λ) and proportions (π) will provide valuable insights into the underlying factors affecting return rates. Furthermore, as statisticians, we propose the application of a finite mixture of Poisson distribution, leveraging both Poisson and mixture distribution techniques, not only for the dataset analyzed in this paper but also for other datasets sharing similar statistical properties.

CONCLUSION

From our perspective, we are the pioneers in conducting this particular study, as no similar research has been previously undertaken. Our work offers a distinctive and unparalleled contribution to the field by providing a comprehensive understanding of the market dynamics surrounding the number of returns in online shoping. This research paper adopts an interdisciplinary approach, employing both fundamental and advanced statistical techniques to analyze the collected data. We consider our study to be at the forefront of exploring this market. However, to achieve a more comprehensive perspective, a future study could collect a larger sample from diverse states across India. Furthermore, our research was confined to clothing merchandise, and conducting further investigations encompassing all product categories would offer a more comprehensive depiction of the overall scenario.

Funding: No funding sources
Conflict of interest: None declared
Ethical approval: This study was approved by the Ethics and Deontology Committee of the Department of Physical Education and Sport Science, University of Thessaly

REFERENCES

1. Vyas P, Sriniwas P. Online Buying Fact or Fiction. Delhi Business Rev. 2002;3(1):2.
2. Menon S, Kahn B. Cross-category effects of induced arousal and pleasure on the Internet shopping experience. J Retail. 2002;78(1):31-40.
3. Mummalaneni V. An empirical investigation of web site characteristics, consumer emotional states and on-line shopping behaviors. J Business Res. 2005;58(4):526-32.
4. Park J, Lennon SJ, Stoel L. Online product presentation: Effects on mood, perceived risk, and purchase intention. Psychol Marketing. 2005;22(9):695-719.
5. Gupta A, Bansal R, Bansal A. Online shopping: A shining future. International Journal of Techno-Management Research. 2013;1(1):1.
6. Arora R. The influence of trust and brand image on purchase intention in online shopping: A comparative study between Amazon and Flipkart. Adarsh J Management Res. 2017;10(2):1-3.
7. Chandra P, Chen J. Taming the Amazon: the domestication of online shopping in Bangalore, India. In Proceedings of the Tenth International Conference on Information and Communication Technologies and Development. 2019;4:1-11.
8. Shenbhagavadivu T, Sharanya M, Priyanka SS, Ajithkumar V, Irshad A. A Study On Online Shopping And Customer Satisfaction On Myntra. Think India J. 2019;22(4):8102.
9. Warrier U, Singh P, Jien CW, Kee DM, Yi GZ, Jiann TW, et al. Factors that lead Amazon. com to a successful online shopping platform. Int J Tourism Hospitality Asia Pasific (IJTHAP). 2021;4(1):7-17.
10. Mishra SJ, Rout D, Kantha R, Majhi B. Customer Perception Towards Amazon and Flipkart: A Comparative Analysis. Int J Res Appl Sci Engineer Technol. 1977;10(5):2147-57.

11. Chandra S. On the mixtures of probability distributions. Scand J Statistics. 1977;4:105-12.
12. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. J Royal Stat Soc Series B (Methodological). 1977;39(1):1-2.
13. Zhao J, Zhang F, Zhao C, Wu G, Wang H, Cao X. The Properties and Application of Poisson Distribution. J Physics: Conference Series. 2020;1550(3):032109.
14. Karlis D, Meligkotsidou L. Finite mixtures of multivariate Poisson distributions with application. J Stat Planning Inference. 2007;137(6):1942-60.

Cite this article as: Nain A. Modeling the number of clothing merchandise returned through online shopping using a mixture of Poisson distribution: an interdisciplinary approach. Int J Sci Rep 2024;10(3):75-80.