Bayesian (Buy Till You Die) Approach to Predict Future Spend of a Customer

Amrita V (21060641004), Ishita Deodhar (21060641020), Sayali Bora (21060641045) and Vidya Mallick (21060641056)

1. Introduction:

Bayesian approach is generally used for estimating the parameters of the posterior distribution. It is considered as the best approach to estimate parameters when the distribution is unknown. This approach is used in various domains. In this project, the Bayesian approach is applied to a B2C company that uses an online platform to sell compact discs and music-related products.

1.1 Background of the industry: B2C companies operate on the internet and sell products to customers online.

1.2 About the Dataset
The dataset contains information about the entire purchase history up to the end of June 1998 of the cohort of 23,570 individuals who made their first-ever purchase at CDNOW in the first quarter of 1997. The data contains 69660 rows and 4 variables.

- Id - Customer Id
- Date - Date Of Purchase
- CDs - Number of CDs purchased
- Price - Price of purchase

1.3 Business Problem: To find the customer lifetime value of CDNOW by studying the purchase rate, dropout probability of customers and by predicting the percentage of monetizing customers.

1.4 Impact of the problem: The major knowledge that has to be obtained by CDNOW is to know the customer lifetime value (CLV). This knowledge would enable them to know their position in the market with respect to profit earned, their monetizing target customer and most importantly this will give them upper hand to make strategic plans for the future in order to sustain their position.

2. Proposed Solution:

2.1 Analysis of the problem:

2.2 Need for Bayesian approach: Prior knowledge λ (transaction rate) to understand the CLV. With posteriors, it is easy to visualize the most plausible values for the parameters and quantify any uncertainty in our estimates.

2.3 Proposed Solution:

We study the purchase rate and dropout probability of customers in a non-contractual setting using the BG/NBD (Beta-Geometric Negative Binomial) model and the BTYD (Bayesian: Buy Till You Die) approach.

It uses historic transaction records to fit a probabilistic model, which then allows to compute quantities of managerial interest on a cohort- as well as on a customer level.
We use the built-in function in BTYDplus to help us change the orders-level data into customer-level data. This function converts the orders data into ITT (Intra-Transaction Time) and total sales data for each customer.

We compute the average order size for each customer using the total sales and the number of orders. Finally, set the prediction horizon (T.star) to one year (365 days) for everyone. If we want to predict future spend across a different length of time, we can just change T.star.

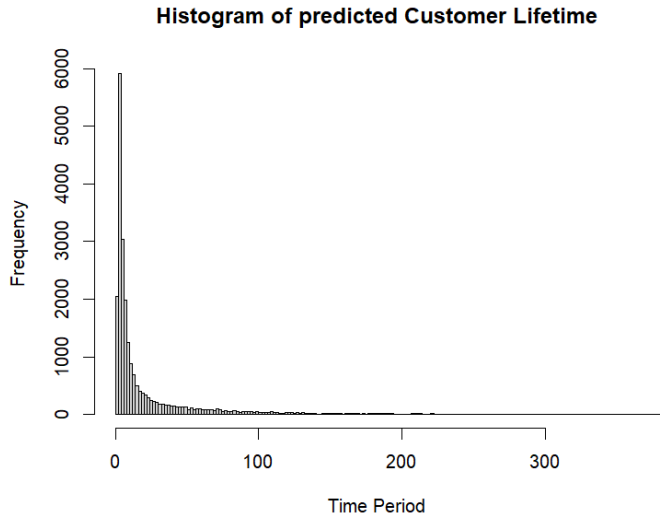
2.4 Background Study:
The assumptions related to BG/ NBD model:

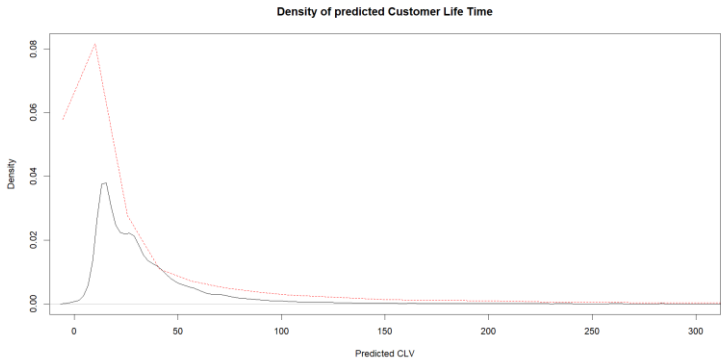
- While active, the number of transactions made by a customer follows a Poisson process with transaction rate λ. This is equivalent to assuming that the time between transactions is distributed exponentially with transaction rate λ. Heterogeneity in λ follows a gamma distribution.
- After any transaction, a customer becomes inactive with probability p. Therefore the point at which the customer “drops out” is distributed across transactions according to a (shifted) geometric distribution. We assume that a customer is alive at the beginning of the observation period. Therefore, a customer cannot “die” before he makes his first transaction.
- Heterogeneity in p follows a beta distribution.
- The transaction rate λ and the dropout probability p vary independently across customers.

3. Implementation
Software: RStudio, package “BTDYPlus”
We predict CLV using the BTYD family of models. We use BG-NBD and model. This model will use each customer’s spending behavior to predict the number of transactions they’ll make in the chosen time period. Once we have the predicted number of transactions, we can compute CLV by multiplying the predicted number of transactions by the average monetary value (spend).

Result:

1. Estimate NBD model parameters, r - 0.3847659 and alpha - 12.0720137.
2. Estimate total transactions during holdout = 2929.82





4. Interpretation:

Histogram: In a customer's lifetime (the customer's time period for the CD store). As time increases, the number of transactions by the customer decreases, thereby the CLV decreases with time.

Density: The BG-NBD model (red line) overestimates the CLV as compared to the actual data. Hence, we can change the prior model to obtain a better fit.

5. Conclusion and Future Scope:

The BG-NBD overestimates the actual CLV. We can implement pareto NBD to find the 20% customers who are contributing the most.

Further prediction of average profit earned can be done by using the gamma-gamma model. This model will give the expected average profit for each customer after modeling the average profit for the mass.

References:

1. GitHub - mplatzer/BTYDplus: R package for Customer Behavior Analysis
2. Bayesian Customer Lifetime Values Modeling using PyMC3 | by Meraldo Antonio | Towards Data Science
3. Bayesian Inference for Predicting the Monetization Percentage in Free-to-Play Games | IEEE Journals & Magazine