Customer Lifetime Value

Yogesh Bhalerao
ybhalerao3@gmail.com
September 2019

In this article, we will look at the mathematics of predicting the customer lifetime value (CLV) in an non-contractual setting (businesses where consumer is not bound by a contract or subscription). The CLV definition that we will be using is "*The present value of the future cash flows attributed to the customer during his/her entire relationship with the company*" (as described in Marketing Metrics Pfeifer et al.), and mathematically expressed as below (See Fader and Hardie)

$$E(CLV) = \sum_t expected\ net\ cash\ flow \mid alive * P(alive) * discount\ factor$$

The remainder of this article discusses how to compute the expected number of purchases for each consumer over a specific period of time, calculating the net cash flow from these purchases, and the probability that the consumer is alive in that specific period of time.

To compute the expected number of purchases, and the probability the consumer is alive, we will be using two models

1) Pareto/Negative Binomial Distribution (Pareto/NBD). The R package "BYTD Models" has implemented this technique, or refer to the python package (Lifetimes)
2) "Counting your customers..." by Makoto Abe which is a Hierarchical Bayes extension to the above Pareto/NBD model. If you have already read the Makoto Abe paper and looking for a python based implementation, you can find my implementation in Jupyter notebook (ported from BYTDPlus R package), please note that the implementation is quite complex and non-trivial

Next we will also look at the Gamma-Gamma model to compute the expected consumer net cash flow per transaction. For a complete working python example, refer to my Jupyter Notebook. The code uses the dataset provided by Olist (the largest department store in Brazilian marketplaces) and hosted by Kaggle.

Finally we will bring all these components together to get to the expected customer lifetime value

Expected number of purchases

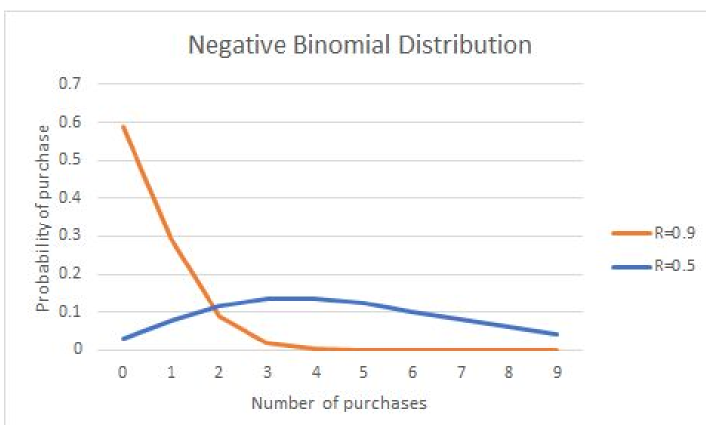
To predict the expected number of purchases, we will look at the math behind the Pareto/Negative Binomial Distribution (Pareto/NBD) model and also at the extension to the Pareto/NBD model by Makoto Abe. The Pareto/NBD model name is derived from the two submodels one based on NBD and the second using "Pareto distribution of second kind". However I personally like the Makoto Abe extension for the following reasons mentioned by the author

1) The Pareto/NBD model assumes that the purchase rate of a consumer and consumer attrition (or lifetime) vary independently across consumer. However there may be scenarios where consumers that are purchasing frequently may have a longer lifetime and hence the purchase rate and lifetime is no longer independent by correlated. In this scenario the parameters estimated by the Pareto/NBD model may be biased. Makoto Abe model relaxes this assumption
2) It is easier to introduce covariates in the Makoto Abe model
3) The purchase rate and consumer attrition parameters can be estimated at the individual level providing greater insights for Marketing in Makoto Abe model

Let's discuss the NBD model intuition and then the underlying details of the Pareto/NBD model. For a detailed discussion on each of the topics below, please refer to the notes by Bruce and Hardie.

Repeat buying behavior

Assume that you have a database of consumer transactions, and using that database if you count the number of consumers that have bought your product say 0, 1, 2, 3, and so on in any calendar year (analysis period), and if that graph looks something like the one shown below, then we can model your highly skewed consumer purchases distribution using a mathematical function called Negative Binomial Distribution (NBD)



Thus if we assume that the purchase probability has a negative binomial distribution then the probability mass function is as given below(refer wikipedia article for more details), if x = number of repeat purchases, p = probability that a consumer will make a purchase in the analysis period, r is the number of customers with 0 purchases, and m = average number of purchases per consumer then the purchase probability P(X = x) is given as

$$P(X = x) = f(x;\ r, p) = \frac{\Gamma(r+x)}{\Gamma(r)x!} p^r (1-p)^x \ for\ x = 0, 1,\ 2,\ 3, ...$$

If we substitute m = $(1-p)^r/p$ in the above equation we get

$$P(X = x) = f(x;\ r, p) = \frac{\Gamma(r+x)}{\Gamma(r)x!} \left(\frac{r}{r+m}\right)^r \left(\frac{m}{r+m}\right)^x \ for\ x = 0, 1,\ 2,\ 3, ... \text{ ------------ } (1)$$

This is a good starting point, however within this one calendar year analysis period, there are new, repeat consumers, so we are dealing with censored data. Hence we will need to relax the restriction on the analysis period.

If the manager would like to compare one year to the next year or previous year, or if analysis period varies (say quarter or even month), then the model is not particularly suited for these insights.

Then there is the question of consumer attrition, how do we deal with attrited consumers? Since the consumer attrition or lifetime is unobserved.

These questions leads us into the Pareto/NBD model which accounts for consumer attrition, flexible analysis period and consumer heterogeneity.

Consumer heterogeneity (NBD Model)

To account for consumer heterogeneity, we assume that each consumer has a fixed purchase rate (**λ**), in which case we can model the probability of the repeat-purchases by each consumer in any given time period by Poisson distribution

$$P(X(t) = x \mid \lambda) = \frac{(\lambda t)^x e^{-\lambda t}}{x!} \ for\ x = 0,\ 1,\ 2,\ 3,\ ...$$

These individual consumer purchase rates can be modeled by a Gamma distribution with shape = r and scale = **α** .

$$g(\lambda | r, \alpha) = \frac{\alpha^r \lambda^{(r-1)} e^{-\lambda\alpha}}{\Gamma(r)}$$

The consumer purchase probability over the entire population can then be derived as below

$$P(X(t) = x \mid r, \alpha) = \int_0^\infty P(X(t) = x|\lambda) g(\lambda|r, \alpha) d\lambda$$

$$P(X(t) = x \mid r, \alpha) = \frac{\Gamma(r+x)}{\Gamma(r)x!} \left(\frac{\alpha}{\alpha+t}\right)^r \left(\frac{t}{\alpha+t}\right)^x \text{ -------------- } (2)$$

Note the similarity between equations 1 and 2, we represent consumer heterogeneity with an NBD model.

Consumer Attrition (Pareto Model)

In our above NBD example, we assumed that after the first purchase, every consumer is always active, meaning that they will make subsequent purchases, though some of them may purchase the next day, the next week, or next month and so on. We assumed that the consumer never attrited.

In a typical subscription business (contractual setting), we do know when the consumer attrites, but in a non-contractual setting (ecommerce marketplace or retail website) it is not always know when and if the consumer attrited.

In such an non-contractual setting, we assume that the consumer is active for some length of time (τ) and inactive for (τ > T) and is exponentially distributed with dropout rate (μ)

$$f(\tau|\mu) = \mu e^{-\mu\tau}$$

As before we model these individual consumer unobserved lifetimes by a Gamma distribution with shape = s and scale = β .

$$g(\mu|s,\beta) = \frac{\beta^s \mu^{(s-1)} e^{-\mu\beta}}{\Gamma(s)}$$

The probability of the consumer being active is obtained by integrating over the entire population

$$f(\tau|s,\beta) = \int_0^\infty f(\tau|\mu) g(\mu|s,\beta) d\mu = \frac{s}{\beta}\left(\frac{\beta}{\beta+\tau}\right)^{s+1}$$

And the cumulative distribution function is

$$F(\tau|s,\beta) = \int_0^\infty F(\tau|\mu) g(\mu|s,\beta) d\mu = 1 - \left(\frac{\beta}{\beta+\tau}\right)^{s}$$

The above equation are similar to the equation for “Pareto distribution of second kind”.

Expected purchases (Pareto/NBD)

For any time interval between (0,t], and an unobserved time (τ), when the consumer becomes inactive. Then if (τ > t) then the expected purchases is λt, and if (τ ≤ t) then the expected purchases are $\lambda\tau$.

Since τ is unobserved, so removing the conditioning on τ, then for a consumer with a purchase rate (λ) and a dropout rate (μ), the expected number of purchases in the time interval (0,t], conditional on λ and μ is given as

$$E[X(t)|\lambda, \mu] = \lambda t P(\tau > t|\mu) + \int_0^t \lambda \tau f(\tau|\mu) d\tau$$
$$= \lambda t e^{-\mu t} + \lambda \int_0^t \mu \tau e^{-\mu \tau} d\tau$$
$$= \frac{\lambda}{\mu} - \frac{\lambda}{\mu} e^{-\mu t}$$

For a randomly chosen individual, the expected number of purchases is integrating over all possible values of λ and μ.

$$E[X(t)|r, \alpha, s, \beta] = \int_0^\infty \int_0^\infty E[X(t)|\lambda, \mu] g(\lambda|r, \alpha) g(\mu|s, \beta) d\lambda d\mu$$
$$= \frac{r\beta}{\alpha(s-1)} \left[1 - \left(\frac{\beta}{\beta+t}\right)^{s-1}\right]$$

The python package lifetimes uses this equation in the function
expected_number_of_purchases_up_to_time

Consumer is Alive (Pareto/NBD)

For a consumer with a purchase rate (λ), dropout rate (μ), and given a purchase history of (x, t_x, T) where x is the number of repeat-purchases (frequency), t_x the time of the last purchase (recency) in the analysis period T. The probability that the consumer is alive at time T is the probability the consumer is active for time period (τ) and inactive for (τ > T).

$$P(\tau > T|\lambda, \mu, x, t_x, T) = \frac{\lambda^x e^{-(\lambda+\mu)T}}{L(\lambda, \mu|x, t_x, T)}$$

For a randomly chosen individual, the probability that any consumer is alive given the past history of that consumer is by integrating over all possible values of λ and μ.

$$P(alive|r, \alpha, s, \beta, x, t_x, T) = \int_0^\infty \int_0^\infty P(\tau > T|\lambda, \mu, x, t_x, T) g(\lambda, \mu|r, \alpha, s, \beta, x, t_x, T)\, d\lambda d\mu$$
$$= \left\{1 + \left(\frac{s}{r+s+x}\right)(\alpha + T)^{r+x}(\beta + T)^s A_0\right\}^{-1}$$

Where A_0 is defined as follows

For $\alpha \geq \beta$

$$A_0 = \frac{{}_2F_1\left(r+s+x; s+1; r+s+x+1; \frac{\alpha-\beta}{\alpha+t_x}\right)}{(\alpha+t_x)^{r+s+x}} - \frac{{}_2F_1\left(r+s+x; s+1; r+s+x+1; \frac{\alpha-\beta}{\alpha+T}\right)}{(\alpha+T)^{r+s+x}}$$

For $\alpha \leq \beta$

$$A_0 = \frac{{}_2F_1\left(r+s+x; r+x; r+s+x+1; \frac{\beta-\alpha}{\beta+t_x}\right)}{(\beta+t_x)^{r+s+x}} - \frac{{}_2F_1\left(r+s+x; r+x; r+s+x+1; \frac{\beta-\alpha}{\beta+T}\right)}{(\alpha+T)^{r+s+x}}$$

And ${}_2F_1(a; b; c; z)$ is the gaussian hypergeometric function

The python package lifetimes uses these equations in the function conditional_probability_alive

Forecasting the Expected Purchases (Pareto/NBD)

To forecast the expected purchases (conditional expectation) in period (T, T+t], is similar to the expected purchases formula

$$E[X(t)|\lambda, \mu,\ alive\ at\ T] = \lambda t P(\tau > T+t|\mu, \tau > T) + \int_T^{T+t} \lambda\tau f(\tau|\mu, \tau > T)d\tau$$

And if the consumer with a known (λ, μ, x, t_x, T)

$$E[X(t)|\lambda, \mu, x,\ tx,\ T] = E[X(t)|\lambda, \mu,\ alive\ at\ T]P(\tau > T|\lambda, \mu, x, t_x, T)$$

For a randomly chosen consumer, we will need to take the expectation over the joint posterior distribution of λ and μ

$$E[X(t)|r, \alpha, s, \beta,\ x,\ tx,\ T] = \int_0^\infty\int_0^\infty \{E[X(t)|\lambda, \mu,\ alive\ at\ T]P(\tau > T|\lambda, \mu, x, t_x, T)g(\lambda, \mu|r, \alpha, s, \beta, x, t_x, T)\}\ d\lambda d\mu$$

$$= P(alive|r, \alpha, s, \beta, x, t_x, T) * \frac{(r+x)(\beta+T)}{(\alpha+T)(s-1)} * \left[1 - \left(\frac{\beta+T}{\beta+T+t}\right)^{s-1}\right]$$

$$= \left\{\frac{\Gamma(r+x)\alpha^r\beta^s}{\Gamma(r)(\alpha+T)^{r+x}(\beta+T)^s}/L(r, \alpha, s, \beta|x, t_x, T)\right\} * \frac{(r+x)(\beta+T)}{(\alpha+T)(s-1)} * \left[1 - \left(\frac{\beta+T}{\beta+T+t}\right)^{s-1}\right]$$

The python package lifetimes uses this equations in the function conditional_expected_number_of_purchases_up_to_time

Makoto Abe extension to Pareto/NBD

We will now look at the Makoto Abe extension to the Pareto/NBD model, In the Pareto/NBD model the consumer purchase rates (λ) and dropout rates (μ) were modeled using a gamma distribution, however in Makoto Abe extension they follow a multivariate lognormal distribution denoted as $\beta_i = [log(\lambda_i),\ log(\mu_i)]$ with mean $\boldsymbol{\beta}_0$ and variance-covariance matrix Γ_0.The priors for these hyper-parameters are $\beta_0 \sim MVN(\beta_{00},\ \Sigma_{00}),\ \Gamma_0 \sim IW(\nu_{00},\ \Gamma_{00})$. Now we will describe the MCMC procedure to estimate the model parameters (λ, μ, z, τ), where z is defined as z=1 if the consumer is active at T and 0 otherwise. All function names refer to the function in the following jupyter notebook. The entire MCMC chain described above is implemented in the run_single_chain routine.

1. Set the initial value of $\beta_i^{(0)}\ \forall i$. Refer to function (set_hyperpriors)
2. For each consumer i,

   2.1. Draw z_i using below equation (refer function draw_z)

$$P[\tau > T|\lambda, \mu, t_x, T] = P[z = 1|\lambda, \mu, t_x, T] = \frac{1}{1+\frac{\mu}{\lambda+\mu}\left[e^{(\lambda+\mu)(T-t_x)}-1\right]}$$

2.2. If $z_i = 0$, generate using a truncated exponential function (refer function draw_tau)

2.3. Generate $\{\beta_i | z_i, \tau_i\}$ using equation (refer function draw_tau)

$$L(x, t_x, T|\lambda, \mu, z, \tau) = \frac{\lambda^x t_x}{\Gamma(x)} \mu^{1-z} e^{-(\lambda+\mu)\{zT+(1-z)\tau\}}$$

3. Generate $\{\beta_0, \Gamma_0 | \beta_i; \forall i\}$ using a standard multivariate normal regression update (refer function draw_level_2)

4. Iterate steps 2 and 3 until converged (refer function draw_level_1)

Forecasting the Expected Purchases (Makoto Abe)

To forecast the expected purchases (conditional expectation) in period (T, T+t], use the individual customer estimated purchase rates draws from the MCMC procedure to estimate the future transactions, please refer to the function DrawFutureTransactions

Consumer is Alive (Makoto Abe)

Similar to forecasting of expected purchases, a consumer's probability of being active in the analysis period can be estimated from the MCMC draws, please refer to the function PAlive, to determine if the consumer is active after analysis period (T, T+t] refer function PActive

Expected consumer net cash flow per purchase incident (Gamma-Gamma model)

Once we have computed the expected number of purchases for a consumer, we now have to compute the net cash flow from these purchases to complete the monetary computation in the CLV formula. One such model to compute the monetary value of consumer with known purchase history is the Gamma-Gamma model.

If the consumer with purchase history (x, t_x, T), and if z_1, z_2, …, z_x be the net cash flow of each purchase incident with $\bar{z}$ be the mean net cash flow, and ζ be the unobserved mean net cash flow of these future purchases.

Next if we assume that $z_i \sim$ gamma(p,ν), and ζ = E(Z_i|p,ν) = p/ν, and $\nu \sim$ gamma(q, □). So for a consumer with x purchases, $\sum_{i=i}^{x} Z_i \sim gamma(px, \upsilon)$ and by the scaling property we get $\bar{z} \sim$ gamma(px, νx). Then the distribution of $\bar{z}$ given x is

$$f(\bar{z} \mid p, q, \gamma; x) = \int_0^\infty f(\bar{z}|p, \nu; x) g(\nu|q, \gamma) d\nu$$

$$= \int_0^\infty \frac{(\upsilon x)^{px} \bar{z}^{px-1} e^{-\nu x \bar{z}}}{\Gamma(px)} \frac{\gamma^q \nu^{q-1} e^{-\gamma\nu}}{\Gamma(q)} d\nu$$
$$= \frac{1}{\bar{z} B(px,q)} \left(\frac{\gamma}{\gamma + x\bar{z}}\right)^q \left(\frac{x\bar{z}}{\gamma + x\bar{z}}\right)^{px}$$

And

$$g(\nu|p, q, \gamma; \bar{z}, x) = \frac{f(\bar{z}|p,\nu;x) g(\nu|q,\gamma)}{f(\bar{z}|p,q,\gamma;x)} = IG(px + q,\ \gamma + x\bar{z})$$

Also $f(\zeta|p, q, \gamma) \sim IG(q,\ p\gamma)$ with $E(Z|p, q, \gamma) = \frac{p\gamma}{q-1}$

Finally to get $E(Z|p, q, \gamma; \bar{z}, x) = \frac{p(\gamma + x\bar{z})}{px+q-1} = \left(\frac{q-1}{px+q-1}\right) \frac{p\gamma}{q-1} + \left(\frac{px}{px+q-1}\right) \bar{z}$

The python package lifetimes uses this equations in the function conditional_expected_average_profit

Computing CLV

Finally we can bring all the above formulae to compute the customer lifetime value. As mentioned at the beginning of the article, the expected consumer lifetime value can be computed using the below formula

$$E(CLV) = \sum_t expected\ net\ cash\ flow \mid alive\ *\ P(alive)\ *\ discount\ factor$$

Given a consumer with purchase history (x, t_x, T), and if $\bar{z}$ the mean net cash flow per purchase, we use the following algorithm

1. Determine the unit of measure for T (in terms of days, weeks, months) called the analysis or calibration period. Assume for now T is in days
2. Next determine the time period for forecasting in months. We typically assume t = 12 (1 year expected CLV) thus we will need to forecast consumer purchases from (0, 12 months)
3. Determine the discount factor to be applied per unit of forecasting time. So if the forecasting is in months, then determine the monthly adjusted discount rate
4. Use the gamma-gamma model to compute the expected consumer net cash flow per purchase (conditional_expected_average_profit function)

5. Since the discount factor is in months, we will compute the expected CLV for the first month, the second month and so on, and report the sum of the expected CLV per month for 12 month
   5.1. Forecast the expected purchase for first month. Use the conditional_expected_number_of_purchases_up_to_time(first 30 days). Then compute the first month expected CLV as (expected consumer net cash flow per purchase) * (first month purchases) / ((1 + discount rate) ^(30/30))
   5.2. Forecast the expected purchase for second month. Because the conditional_expected_number_of_purchases_up_to_time is cumulative, we will need to subtract the first month's purchases from the purchases.
      5.2.1. B = conditional_expected_number_of_purchases_up_to_time(0,60)
      5.2.2. A = conditional_expected_number_of_purchases_up_to_time(0,30)
      5.2.3. Expected second month purchases = B - A
      5.2.4. Then compute the second month expected CLV as (expected consumer net cash flow per purchase) * (first month purchases) / ((1 + discount rate) ^(60/30))
   5.3. Forecast the expected purchase for third month. Because the conditional_expected_number_of_purchases_up_to_time is cumulative, we will need to subtract the second month's purchases from the purchases.
      5.3.1. B = conditional_expected_number_of_purchases_up_to_time(0,60)
      5.3.2. A = conditional_expected_number_of_purchases_up_to_time(0,90)
      5.3.3. Expected third month purchases = B - A
      5.3.4. Then compute the third month expected CLV as (expected consumer net cash flow per purchase) * (first month purchases) / ((1 + discount rate) ^(90/30))
6. To get the expected CLV for 12 month (1 year CLV), we will need to
   6.1. Expected CLV = (first month expected CLV) + (second month expected CLV) + … + (twelfth month expected CLV)

Summary

In this article we discussed the math behind the computation of the expected Customer Lifetime Value. The Pareto/NBD model is currently the most commonly used model in the industry to compute the expected customer lifetime value. The MCMC extensions to the Pareto/NBD model improves the easy of computation and provides additional insights to marketing especially at the individual level.

More recently there have been attempts to apply machine learning to this problem, see google for one such commercial product.

I also found an excellent discussion on "customer lifetime value and customer profitability" in the paper by Pfeifer et. al. I am summarizing some of the author's thoughts below

There is a marked distinction between “customer lifetime value” and “customer profitability”. In some terms we can equate “customer lifetime value” as the firm’s willingness to pay for a consumer relationship (perceived value) and “customer profitability” as the difference between revenue generated and costs associated with the customer relationship during a specified period and typically must equal one of the reported profit figures to the shareholders. Hence the firm will need to decide on what to include in the “net cash flow” when computing “customer lifetime value”, which may be different from the revenue and cost components included while computing the “customer profitability”.

To add a quick note on Acquisition and Retention costs, acquisition and retention costs will always be included in the computation of “Customer Profitability”, but should it be included in the “customer lifetime value”? This will largely depend on how the firm defines looks at prospecting. For e.g., if a firm is planning to spend $100K in prospecting to a 10K consumers and successfully acquires say 1K consumers, then do not include acquisition cost in the “customer lifetime value” and then compare the expected CLV to (($100K/10K)/(1K/10K)) = $100. If you choose, you can apply the same for retention costs, where a retained consumer is treated the same as a new prospect.

Working through the expected Customer Lifetime Value computation, I have advanced my understanding on expected CLV, and hoping that it will also help you to get started.