Customer Lifetime Value in Mobile Games: a Note on Stylized Facts and Statistical Challenges

Arturo Valdivia
Tactile Games
Copenhagen, Denmark
arturo@valdivia.xyz

***Abstract*—We analyze a series of empirical properties exhibited by customer lifetime value data, including zero-inflation, heavy tails, and the varying behaviour of inter-purchase times and purchase size. Rather than focusing on specific models already established in the literature (*e.g.*, *RFM* or *Pareto/NBD*), what is emphasized here are empirical properties of mobile games data which may lead to revisit certain assumptions of existing models.**

***Index Terms*—LTV; Mobile Gaming; Stochastic Modelling**

I. Introduction

The mobile gaming industry has experienced strong sustained growth over the past years, with a particular upward spike during 2020 when mobile games offered a source of entertainment while social distancing, and a distraction from the stress derived from the pandemic. Indeed, during 2020 the industry observed a $+45\%$ increase year-over-year in game downloads, exceeding the $+32\%$ from the year before [1]. Moreover, these figures move to almost $+70\%$ when they are to limited the subset of downloads obtained as result of a marketing campaign. In terms of revenues the gaming industry expects a compound annual growth rate of $+7.2\%$, and forecasts to surpass the $200Bn mark by 2023 [20]. In such a growing and competitive industry, being able to predict how much an individual user will spend in the game (*i.e.*, *customer lifetime value* or *LTV*) is crucial, not only from the game design perspective but also from the financial point of view.

In this context, we analyze a series of LTV properties arising from data in mobile games. Beyond discussing specific models already established in the literature, what is emphasized here are empirical properties of the data which may lead to revisit certain assumptions of existing models. In particular we show that the *inter-purchase times* (*i.e.*, time between consecutive in-app purchases) are not identically distributed, thus questioning whether models like [12], [21], [22], [24] can be applied to mobile games out of the box. On the other hand we also highlight some potential new avenues of research, including hypothesis testing for LTV, and the unexploited fact that, in mobile gaming, the monetary value of the purchases follows a multinomial distribution.

The rest of the content is organized as follows. Section II provides an overview of different approaches for analytical estimation of LTV. Section III present and discuss the main results, and gives a mathematical formalization for the LTV. Concluding remarks are given in Section IV.

978-1-6654-3886-5/21/$31.00 ©2021 IEEE

II. Related Work

Analytical estimation of LTV has been typically approached by means of three types of techniques: by means of heuristics based on expert knowledge; by employing machine learning algorithms; and by calibrating stochastic models. Arguably the most prominent example of a heuristic used in practical customer analytics is the so-called *Recency, Frequency, and Monetary* value (RFM) framework which was introduced in the 1920's —see [2], [23], [29], [30], and Ch. 4 in [28] for an in-depth discussion about the use of heuristics. On the other hand, machine learning methods are increasingly being used to predict LTV in the context of *freemium* products and *free-to-play* mobile games —see for instance [5], [6], [8], [25]–[27] and references therein. Some benefits of such methods include (*i*) the ability to work with relaxed assumptions about the underlying distribution of the data; and (*ii*) the capacity to effectively scale (computationally speaking) when dealing with games having millions of active players on daily basis. However, these methods may face major drawbacks when it comes to model new games for which there is insufficient historical data available. Furthermore, when such models are trained in a supervised learning framework, then one may incur into significant additional development costs (*i.e.*, related to train, validate business-wise and deploy a new model) when the business application requires *simple* modifications. A common scenario of such modifications arises when marketing managers require to assess different profitability scenarios for their marketing investments, say for instance looking at profitability after nine or fifteen months since the campaign start, whereas the original model was initially trained for the fixed horizon of twelve months.

One could say that stochastic models fall somewhere in-between the heuristics and the machine learning approaches. These models start by making some assumptions regarding the purchase patterns and potential attrition of customers. For purchasing patterns, the most natural starting point (mathematically speaking) would be to assume that the purchases take place following a Poisson process with rate λ, and then account for the variation across customers by allowing the parameter λ to be random itself, say Gamma-distributed.

This Poisson-Gamma mixture results in a *negative binomial distribution* (NBD), which is the common name by which such a baseline model for purchasing patterns is referred to. The NBD was derived in the 1920s (as the RFM) by [12], but introduced in the context of marketing later on in the 1950s [9]. The next significant extension of the NBD, known as the *Pareto/NBD model*, was introduced in [24] by combining the NBD with a Gamma mixture of exponentials (or Pareto Type II) as a timing model for customer churn. Since then multiple extensions have been developed with aims of capturing more complex behaviours or increasing the analytical tractability of the model —see [3], [10], [11], [14], [15], [21] and references therein. However it is worth mentioning that little work has been done regarding the empirical validation of the aforementioned approaches. And, to the best of our knowledge, the present paper may be one of the first works addressing such validation in the context of mobile gaming. For our analysis we consider a sample of one million of in-app purchases made by players across multiple puzzle games developed by the company Tactile Games.

III. Customer lifetime value: Empirical Properties

Let us start this section by providing a mathematical formalization for the customer lifetime value: Assume a player installs the game at time $t_0 = 0$. The total amount of money this player has spent in the game up to time $t \geq t_0$ can be then described by means of the quantity

$$V_t := \sum_{n\geq 0} M_n \mathbb{I}_{\{T_n \leq t\}}, \tag{1}$$

where T_n and M_n denote the time at which the n^{th} in-app purchase takes place and its monetary value, respectively. The *indicator function* $\mathbb{I}_{\{T_n \leq t\}}$ is simply defined as 1 if $T_n \leq t$, and 0 otherwise. As a convention, we shall set $T_0 := t_0$ and set $T_n = \infty$ if the player does not pay for an n^{th} time. We refer to the stochastic process $\{V_t\}_{t\geq t_0}$ as the *customer value process*. In these terms, the customer's lifetime value equals V_τ, where τ describes the elapsed time between the install date t_0 and the last session of the player before churning.

Two remarks are in order. On the one hand, it is customary to declare players as churned users if they haven't played the game for two consecutive weeks; and so that is the convention we follow in this analysis as well. On the other hand, practitioners tend to look at specific horizons in order to evaluate LTV and the return on marketing investment. Typically a horizon T such that $T - t_0$ equals one year is considered and, thus, instead of the actual *lifetime* value one works with the censored value up to such horizon, that is, $V_{T\wedge\tau}$ where $T \wedge \tau := \min\{T, \tau\}$. In what follows by LTV we shall implicitly mean $V_{T\wedge\tau}$.

Notice that the formalization in Eq. 1 is given in a general form, without making any assumptions on the underlying distribution of its constituents.

A. High-level properties

Similar to other products with *freemium* business models, the design of *free-to-play* mobile games implies the vast majority of the users will play for free, while only a small fraction of around 10% or less will make at least one in-app purchase —typically with the goal of obtaining virtual goods or gaining access to special content. This *zero-inflation* property is well-known (*cf.* [5]) and applies for our dataset as well. What is more interesting, both mathematically and from the game design perspective, is the *tail* of the LTV distribution, *i.e.*, the behaviour of the high spenders. Following the approach in [7], we analyzed this tail distribution and found that its behaviour is consistent with that of a power-law (or Pareto) distribution —which is known for its *heavy* tail. As an example, when applying this to the players from the US market we found that for players on an Android device (resp., iOS device), the tail values of LTV —beyond the 97% (resp., 96%) percentile— behave like a power-law with parameter $\alpha \simeq 2.89 \pm 0.03$ (resp., $\alpha \simeq 2.49 \pm 0.05$). The high p-value of 0.12 (resp., 0.20) we obtained suggests that a power-law distribution is indeed a plausible benchmark for the behaviour of high spenders. It is important to highlight here that having a parameter $2 < \alpha < 3$ implies that the power-law distribution has finite mean, but an infinite variance. This could suggest that when dealing with LTV, a better numerical stability could be achieved by handling the sub-population of high spenders (*e.g.*, those on the top 3%-4% percentiles) differently. This observation also leads us to hypothesize that conducting hypothesis testing —also referred to as *A/B testing* by the industry— on LTV using non-parametric bootstrap methods (*cf.* [17]) may lead to tests with low power and higher sample size requirements. Needless to say, this may be a crucial limitation for game development since it is customary to drive product optimization by means of A/B testing.

B. Time between consecutive in-app purchases

Consider the n^{th} inter-purchase time as defined by $\Delta T_n := T_{n+1} - T_n$ for $n \geq 1$, where the convention $T_0 := t_0$ implies that the first inter-purchase time ΔT_1 corresponds to the elapsed time (in days) between install and the first in-app purchase if that exists. A natural question to ask is whether the inter-purchase times $\{\Delta T_n\}_{n\geq 0}$ are equally distributed, or if at least they show some regularity. In order to assess the similarity between the inter-purchase times, Table I shows the Kolmogorov-Smirnov distance between ΔT_n and ΔT_{n+1}, and also between ΔT_n and some natural theoritical benchmarks —which are fitted by MLE and then discretized. In addition, we inspect these differences by means of the *empirical probability generating function* (EPGF), which is defined as

$$\varphi_k(t) := \frac{1}{k} \sum_{j=1}^{k} t^{x_j} \tag{2}$$

for every $t \in [0, 1]$, where $x_1, x_2, ..., x_k$ is a sample from a given discrete distribution [19]. As an example, Fig. 1 illustrates how the EPGF looks like for the first $k := 15$ transactions of a group of users, each of which made at least k in-app purchases in the game. This alternative inspection coincides with the reading that the first in-app purchase made by a user may have a different statistical distribution than the

Table I. Kolmogorov-Smirnov distance between the n^{th} inter-purchase time and various benchmark distributions: next inter-purchase $(n+1)$, Exponential (Exp), Weibull, Gamma and Generalized Gamma (Gen. G.).

	n=1	n=2	n=3	n=4	n=5	n=10	n=20
$n+1$	10.8%	5.3%	3.3%	3.1%	2.4%	1.3%	0.8%
Exp	32%	26%	23%	21%	19%	14%	8.5%
Weibull	9.1%	6.8%	6.3%	6.2%	6.2%	5.4%	5.3%
Gamma	7.4%	5.9%	6.0%	6.3%	6.4%	5.3%	4.9%
Gen. G.	5.9%	8.2%	8.4%	8.5%	8.2%	7.2%	6.2%

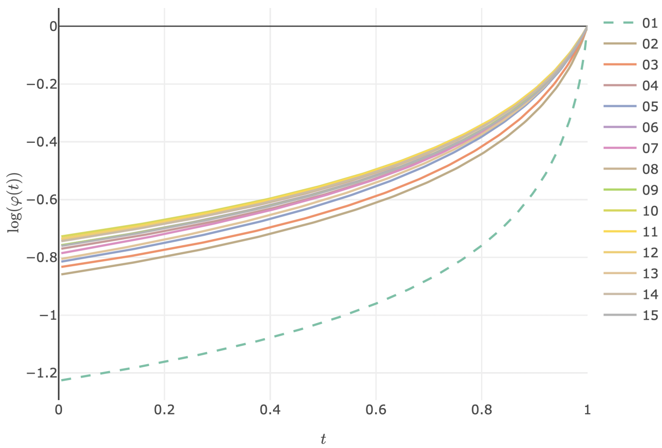

Fig. 1. Logarithm of the EPGF for the time between consecutive transactions. The first in-app purchase shows a consistent different behaviour than the rest.

subsequent consecutive transactions. We also observed that this fact holds when controlling for other co-factors (*e.g.*, country, mobile device type) and varying k.

These results suggest that:

(i) ΔT_1 may have a different distribution than the rest of the subsequent inter-purchase times, and thus could be modelled separately (*cf.* [26]).

(ii) On early transactions (*e.g.*, $1 < n \leq 10$) the inter-purchase time ΔT_n does not appear to be exponentially-distributed, which may lead to invalidating the usage of Poisson processes to model occurrence of transactions (*cf.* [12], [24]) in gaming.

(iii) Outperforming the exponential distribution benchmark we find the Gamma and the Weibull alternatives; this provides evidence in favour of approaches like [13] and [18], respectively.

(iv) Overall, the inter-purchase times $\{\Delta T_n\}_{n\geq 1}$ are not equally distributed, which questions the validity of approaches like [21], [22].

(v) The statistical distance between ΔT_n and ΔT_{n+1} seems to decrease as n increases, suggesting that most engaged paying users have regular playing habits.

C. Expected player lifetime

The expected player lifetime (τ) is another crucial element that determines customer lifetime value. Naturally, the study of τ fits well into the area of statistics known as *survival analysis* [16] where a variety of methods have already found success in other industries. We borrowed some of the well-known parametric distributions used in survival analysis and used them as benchmark for the behaviour of τ and the *player survival probabilities*, defined for every $t \geq t_0$ as $\mathbb{P}(\tau > t)$, *i.e.*, the probability of staying at least t days in the game before churning. The results of this analysis are depicted in Fig. 2. We observe that player survival probabilities deviate significantly from the benchmarks given by the following distributions: Exponential, log-logistic, log-normal, and Gompertz. Interestingly, the Exponential —*i.e.*, the case of a constant hazard rate— and Gompertz distributions massively overestimate $\mathbb{P}(\tau > t)$ at the beginning, *e.g.*, for $t \leq 7$, which is when most of the new players tend to abandon the game. Hence the Gompertz model proposed in [4] does not seem applicable in gaming. Similarly, the Gamma benchmark (*cf.* [21], [22]) may not be enough to describe player survival probabilities for the $t \leq 30$. This region ($t \leq 30$) is crucial since it involves most of the population, and here the two-parameter Gamma is clearly outperformed by its multi-parameter counterparts, the Generalized Gamma and Generalized F. Finally, The deviation from log-logistic behaviour is also worth noticing since this model is typically used in scenarios similar to player engagement —*i.e.*, scenarios where the churn rate increases initially (as the player discovers the new game) and decreases later (as the player engagement grows).

D. Monetary value of in-app purchases

A very specific feature of modelling LTV in mobile games is that, in order to be listed in the main app market places like the App Store or Google Play Store, the in-app purchases offered by the game can only take values within a predetermined set of price points. This directly implies that the random variable M_n in Eq. 1 follow a multinomial distribution. To the best of our knowledge, this very specific feature seems to be currently unexploited by the literature regarding customer lifetime value —while natural approaches like a Multinomial-Dirichlet Bayesian model could be easily be explored.

As with the inter-purchases times, it is natural to inquire if the monetary values $\{M_n\}_{n\geq 1}$ are identically distributed. We see that is not the case by applying a Chi-squared test on each pair (M_n, M_{n+1}) of consecutive transactions: For $n \leq 3$ we obtained a low p-value under 10^{-4} suggesting that the purchases prices are not identically distributed. This fact seems to change as n increases since for next group of comparisons (*e.g.*, $3 < n < 15$) we observed p-value bigger than 0.10, with an average value around 0.74. Table II illustrates the distribution of purchases per price point for the n^{th} in-app purchase —*i.e.*, the distribution of M_n. Notice that higher price points tend to have higher purchase rate as the number of transactions increases, which is consistent with the notion that most users will prefer lower price points at the beginning when they are still in the process of discovering the new game. Although we also observed a small number of players having a very expensive first in-app purchase giving them access to a lot of virtual goods. We attribute this behaviour to a special type of players who prefer to discover the game deeper and

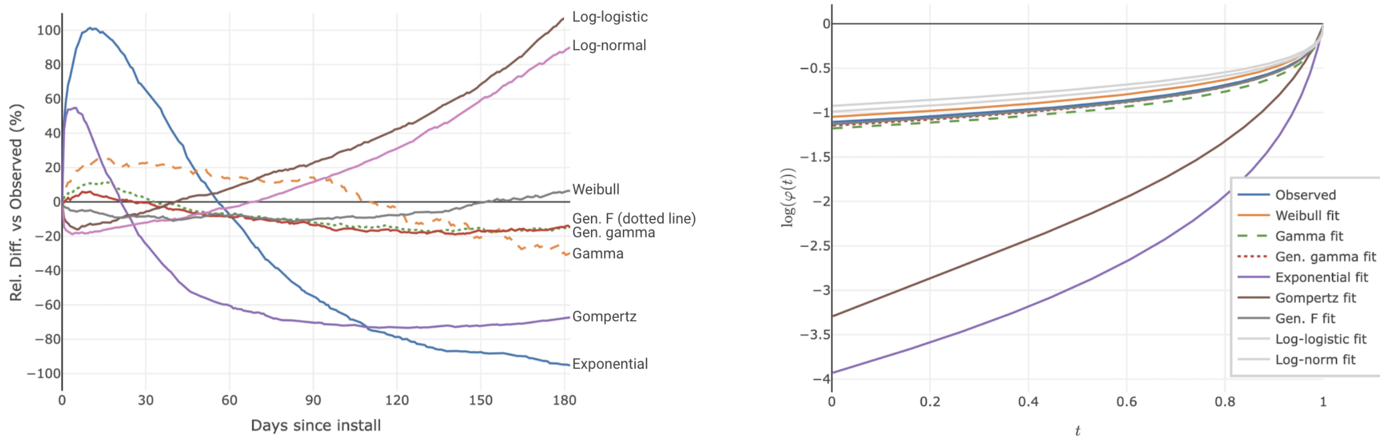

Fig. 2. **Left**: Relative error between the observed player survival probability and theoretical benchmarks. **Right**: Comparison between the logarithm of the EPGF of player survival probabilities and a variety of benchmarks.

Table II. Distribution of purchases per price point for the n^{th} transaction. All values are normalised using the value of the \$1.99 price point in the corresponding row (*e.g.*, a value of 50% means that price point was purchased half as frequently than the \$1.99).

n	\$99.99	\$49.99	\$19.99	\$9.99	\$6.99	\$3.99	\$0.99
1	0.3%	0.7%	5.3%	29.1%	2.5%	6.7%	57%
2	0.2%	0.6%	6.3%	33.3%	4%	7%	12.5%
3	0.3%	0.8%	6.7%	31.1%	3.2%	7.5%	7.4%
4	0.4%	1%	6.8%	33.2%	2.4%	8.4%	4.4%
5	0.4%	1%	7.2%	33.7%	3.2%	7.8%	4.5%

faster than the game economy allows it to the average non-paying player.

IV. Concluding remarks

We have analyzed a series of empirical properties exhibited by customer lifetime values (LTV) using data arising from mobile games. Our results emphasize the challenges one may find when modelling LTV, *e.g.*, handling zero-inflation, heavy tails and non-identically distribution of the constituents in Eq. 1. At the same time some avenues of future research have been suggested, *e.g.*, assessing the power of bootstrap methods for hypothesis testing on LTV, and exploding the unique behavior of $\{M_n\}_{n\geq 1}$ with Bayesian methods. It would be also interesting to further analyze the interaction between different variables in Eq. 1; in particular, the interaction between the inter-purchase time ΔT_n and the monetary value of the last transaction, M_n.

It is worth highlighting that the properties discussed here may help researchers to construct synthetic LTV datasets, which may prove to be valuable since at the time writing this paper no public datasets were known.

Altogether, it is our hope that this work will contribute to the advancement of industry knowledge-sharing regarding LTV and its potential *stylized facts*: a set of properties, common across many games, markets and time periods.

V. Acknowledgements

We thank our colleagues Paolo Burelli and Jeppe Theiss Kristensen for their feedback on early versions of this work.

References

[1] AppsFlyer. State of gaming report. https://www.appsflyer.com/state-of-gaming-november-2020. last visited on 2021-05-18. November 2020.

[2] F.M. Artinger, N. Kozodoi, and J. Runge. Predicting revenues with the multiplier heuristic. *SSRN working paper 3546017*, 2020.

[3] M. Batislam, E.P. Denizel and A. Filiztekin. Empirical validation and comparison of models for customer base analysis. *International Journal of Research in Marketing*, 24(3):201–209, 2007.

[4] A.C. Bemmaor and N. Glady. Modeling purchasing behavior with sudden "death": A flexible customer lifetime model. *Management Science*, 58(5):1012–1021, 2012.

[5] P. Burelli. *Predicting Customer Lifetime Value in Free-to-Play Games*, pages 79–107. 07 2019.

[6] P. P. Chen, A. Guitart, A. Periáñez, and A. Fernández. Customer lifetime value in video games using deep learning and parametric models. *IEEE International Conference on Big Data*, pages 2134–2140, 2018.

[7] A. Clauset, C.R. Shalizi, and M.E.J. Newman. Power-law distributions in empirical data. *SIAM Review*, 51(4):661–703, 2009.

[8] A. Drachen, M. Pastor, A. Liu, D. J. Fontaine, Y. Chang, J. Runge, R. Sifa, and D. Klabjan. To be or not to be... social: Incorporating simple social features in mobile game customer lifetime value predictions. In *Proceedings of the Australasian Computer Science Week Multiconference (ACSW)*, number 40, 2018.

[9] A.S.C. Ehrenberg. The pattern of consumer purchases. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 8(1):26–41, 1959.

[10] P. S. Fader, B. G. S. Hardy, and K. L. Lee. Counting your customers the easy way: An alternative to the pareto/nbd model. *Marketing Science*, 24(2):275–284, 2005.

[11] P.S. Fader and B.G. Hardie. Probability models for customer-base analysis. *Journal of Interactive Marketing*, 23(1):61–69, 2009.

[12] M Greenwood and G. U. Yule. An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. *Journal of the Royal Statistical Society*, 83(2):225–279, 1920.

[13] N. Holtrop and J. E. Wieringa. Timing customer reactivation initiatives. *Marketing Science Institute Working Papers*, 20(120), 2020.

[14] D. Hoppe and U. Wagner. Customer base analysis: The case for a central variant of the bg/nbd model. *Marketing ZFP*, 29(2):75–90, 2007.

[15] P.S. Jerath, K. Fader and B.G. Hardie. New perspectives on customer "death" using a generalization of the pareto/nbd model. *Marketing Science*, 30(5):866–880, 2011.

[16] D.G. Kleinbaum and M. Klein. *Survival Analysis*. Springer, New York, NY, 2 edition, 2005.

[17] J. G. MacKinnon. *Handbook of Computational Econometrics*, chapter 6, pages 183–213. John Wiley Sons, 2009.

[18] B. McShane, M. Adrian, E. T. Eric Bradlow, and P. S. Fader. Count models based on weibull interarrival times. *Journal of Business & Economic Statistics*, 26(3):369–378, 2008.

[19] M. Nakamura and V. Pérez-Abreu. Exploratory data analysis for counts using the empirical probability generating function. *Communications in Statistics - Theory and Methods*, 22(3):827–842.

[20] Newzoo. Global games market report. https://newzoo.com/insights/articles/global-games-market-to-generate-175-8-billion-in-2021-despite-a-slight-decline-the-market-is-on-track-to-surpass-200-billion-in-2023. last visited on 2021-05-18. 2021.

[21] M. Platzer and T. Reutterer. Ticking away the moments: Timing regularity helps to better predict customer activity. *Marketing Science*, 35(5):779–799, 2016.

[22] T. Reutterer, M. Platzer, and N. Schröder. Leveraging purchase regularity for predicting customer behavior the easy way. *International Journal of Research in Marketing*, 38(1):194–215, 2021.

[23] R. Roel. Direct marketing's 50 big ideas. *Direct Marketing*, 50:45–52, May 1988.

[24] D.C. Schmittlein, D.G. Morrison, and R. Colombo. Counting your customers: Who-are they and what will they do next? *Management Science*, 33(1):1–24, 1987.

[25] J. Sifa, R. Runge, C. Bauckhage, and D. Klapper. Customer lifetime value prediction in non-contractual freemium settings: Chasing high-value users using deep neural networks and smote. In *Proceedings 51st Hawaii Int. Conference on System Sciences*, pages 923–932, 2018.

[26] R. Sifa, F. Hadiji, J. Runge, A. Drachen, K. Kersting, and C. Bauckhage. Predicting purchase decisions in mobile free-to-play games. In *Proceedings of the Eleventh AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE-15)*, pages 79–85, 2015.

[27] S. Voigt and O. Hinz. Making digital freemium business models a success: Predicting customers' lifetime value via initial purchase information. *Business Information Systems Engineering*, 38(2):107–118, 2016.

[28] M. Wübben. *Analytical CRM*. Applied Marketing Science. Gabler Verlag, 2009.

[29] M. Wübben and F. Wangenheim. Instant customer base analysis: Managerial heuristics often "get it right". *Journal of Marketing*, 72:82–93, May 2008.

[30] Y. Zhang, E. T. Bradlow, and D. S. Small. Predicting customer value using clumpiness: From rfm to rfmc. *Marketing Science*, 34(2):179–307, 2014.