

UNIVERSIDAD DEL VALLE DE GUATEMALA
Facultad de Ciencias y Humanidades



**Implementación de un modelo estocástico
para el pronóstico del comportamiento de
consumidores en un entorno no contractual.**

Trabajo de graduación en modalidad de Tesis presentado por
Juan Andrés García Porres
para optar al grado académico de Licenciado en Matemática Aplicada

Guatemala,
2020

UNIVERSIDAD DEL VALLE DE GUATEMALA
Facultad de Ciencias y Humanidades



**Implementación de un modelo estocástico
para el pronóstico del comportamiento de
consumidores en un entorno no contractual.**

Trabajo de graduación en modalidad de Tesis presentado por
Juan Andrés García Porres
para optar al grado académico de Licenciado en Matemática Aplicada

Guatemala,
2020

Vo.Bo.:



(f)

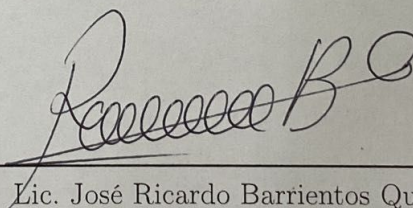
Lic. Cristian Eduardo Valdez Santos

Tribunal Examinador:



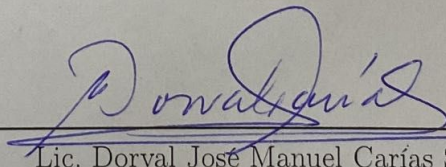
(f)

Lic. Cristian Eduardo Valdez Santos



(f)

Lic. José Ricardo Barrientos Quezada



(f)

Lic. Dorval José Manuel Carías Samayoa

Fecha de aprobación: Guatemala, 17 de febrero de 2020.

La elaboración del presente surgió del interés en profundizar acerca de la aplicación de modelos probabilísticos en el análisis de series de tiempo, ya que a lo largo de la carrera no he tenido la oportunidad de estudiar métodos alternativos afuera de los convencionales enseñados en cursos como Análisis Numérico o Métodos Matemáticos Para La Estadística.

La idea surgió a través de charlas con el Lic. Valdez acerca de métodos alternativos para análisis de comportamiento de una base de clientes así como la aplicación de procesos estocásticos para la elaboración de pronósticos. Agradezco al Lic. Valdez por sus aseorías en la delimitación del presente y la confianza que depositó en mí durante la elaboración de este.

Agradecimientos

A mi mamá, Gladys, por apoyarme incondicionalmente en todo momento de mi desarrollo personal y profesional.

Índice

Prefacio	III
Agradecimientos	IV
Lista de figuras	VI
Lista de cuadros	VII
Resumen	VIII
1 Introducción	1
2 Objetivos	2
2.1 Objetivo general	2
2.2 Objetivos específicos	2
3 Justificación	3
4 Marco teórico	4
4.1 Aspectos generales del valor de tiempo de vida útil de un consumidor	4
4.1.1 Los principales modelos de LTV: retención simple y migración	5
4.2 Procesos estocásticos	7
4.2.1 Clasificación de procesos estocásticos	8
4.3 Reseña de modelos estocásticos <i>Buy 'Till You Die</i> - BTYD	10
4.3.1 Clasificación de entornos de análisis	10
4.3.2 Revisando el concepto de LTV	10
4.3.3 Modelos para entornos contractuales	11
4.3.4 Modelos para entornos no contractuales	12
4.4 Selección de modelos BTYD	14
4.4.1 Pareto/NBD	14
4.4.2 BG/NBD	15
5 Análisis exploratorio de la base de datos	18
6 Implementación del modelo	24
7 Conclusiones	29
8 Recomendaciones	30
9 Bibliografía	31
10 Anexos	32

Lista de figuras

5.1	Distribución de montos de transacción	19
5.2	Distribución de transacciones por cliente	20
5.3	Serie de tiempo del volumen transaccional	20
5.4	Volumen transaccional anual	21
5.5	Serie de tiempo del monto transaccional	21
5.6	Volumen anual	22
5.7	Evolución interanual de transacciones	22
5.8	Evolución interanual de monto transaccional	23
6.1	Tendencia incremental de transacciones semanales	25
6.2	Tendencia acumulada de transacciones semanales	25
6.3	Patrones de consumo para $n = 50$ clientes seleccionados aleatoriamente	27
6.4	Distribuciones de LTV ajustado por modelo	28

Lista de cuadros

4.1	Clasificando bases de clientes	10
4.2	Conjunto de modelos estocásticos para el cálculo de LTV según tipo de base de clientes	14
5.1	Descripción de variables que componen la base de datos	18
5.2	Medidas de tendencia central para la variable <i>sales</i>	18
5.3	Medidas de tendencia central para la cantidad de transacciones por cliente	19
6.1	Estimación de parámetros para los modelos a implementar	26
6.2	Cálculo de $LL(r, \alpha, a, b)$ para el período de calibración	26
6.3	Estimación del tiempo promedio inter-transaccional en semanas para la cohorte de clientes	26
6.4	Cálculo de $E(Y(t) X = x, t_x, r, \alpha, a, b)$ y medidas de error	26
6.5	Medidas de tendencia central para $P(X(t) = x \lambda, p)$ para el período de calibración .	27
6.6	Medidas de tendencia central para el LTV de la cohorte de clientes	27
6.7	Medidas de tendencia central para el LTV de la cohorte de clientes ajustadas a probabilidad de supervivencia	28

El objetivo principal de este trabajo es implementar una metodología para analizar y pronosticar el comportamiento de consumidores en un entorno no contractual, esto es, en donde no se puede discernir *a priori* si un cliente terminó su relación con la compañía o solo hay un hiato en cuanto a sus consumos. La selección de modelos estocásticos implementada es conocida como modelos *Buy 'Till You Die* (BTYD) y a su vez se presenta cómo a través de la implementación de dichos modelos se pueden derivar cantidades que revelan el valor que un cliente representa para la compañía, tales como el valor de tiempo de vida de un cliente (LTV por sus siglas en inglés).

Implementando los modelos BTYD a una cohorte de 6889 clientes de una tienda especializada en ventas al menudeo se logró pronosticar con una precisión mayor al 98 % la cantidad de transacciones en un horizonte de tiempo de un año así como la probabilidad de que un cliente permanezca activo y su valor al capital de la empresa.

Perfilando la cohorte de clientes se encontraron tres subconjuntos con patrones de consumo particulares, el primer grupo que consiste del 11.73 % del total pero únicamente 2.3 % del capital de la empresa. El segundo grupo, que consta únicamente del 4.83 % de la cohorte de clientes pero que componen un 11.70 % del capital total de la empresa. Por último, el tercer grupo, conforma el 83.4 % restante de la cohorte y el 85.9 % del capital.

Este trabajo es una implementación de una selección de modelos estocásticos para el pronóstico del comportamiento de una base de clientes en un entorno no contractual, es decir, en donde no se puede diferenciar si un cliente terminó su relación con la compañía o simplemente decidió retirarse silenciosamente. El caso en particular que se analiza en el presente es el de clientes de una tienda enfocada en ventas al menudeo.

Previo a los avances encabezados por Peter S. Fader [5] con respecto a la implementación de modelos estocásticos para entender los patrones de consumo de los consumidores, las metodologías estaban enfocadas en métodos convencionales, como regresiones. Un ejemplo es el caso de Malthouse y Blattberg [8], los cuales encontraron de manera empírica una forma de hallar el valor futuro de los clientes de una compañía con el fin de examinar la precisión de dichas predicciones, la dependencia en el horizonte de tiempo establecido y las variables predictoras para los clientes en cuestión. Su enfoque se centró en dos supuestos con respecto al comportamiento de los clientes:

- La regla del “20-55”: del 20 % de los mejores (i.e., más rentables) clientes, el 55 % estarán clasificados erróneamente.
- La regla del “80-15”: del 80 % del total de clientes, un 15 % se considera clasificado erróneamente.

Estos supuestos permiten desechar la idea de que los consumidores con un alto valor en el pasado van a seguir siendo valiosos en el futuro y que los clientes que producen poco valor van a continuar con esta tendencia.

A lo largo de la presentación de los conceptos teóricos detrás de la implementación de los modelos estocásticos, se puede notar cómo las ideas propuestas por [8] se presentan en los conceptos fundamentales del modelo y cómo dichas ideas son llevadas más allá gracias al andamiaje teórico que brinda la teoría de probabilidades.

Se asumirá que el lector tiene conocimientos básicos de teoría de probabilidades, en particular, los conceptos de espacio muestral, variable aleatoria, medida probabilidad, espacio de probabilidad, probabilidad condicional y valor esperado.

2.1. Objetivo general

Implementar una metodología para analizar y pronosticar el comportamiento de consumidores en un entorno no contractual.

2.2. Objetivos específicos

- Perfilar y segmentar el comportamiento de consumidores en entornos no contractuales en un caso particular.
- Establecer los factores más influyentes en el comportamiento de consumidores aplicado a un conjunto de datos en particular.
- Determinar una forma de derivar el valor del tiempo de vida útil y el valor del capital de marca de una base de clientes por medio de modelos estocásticos.

Los modelos de series de tiempo suelen clasificarse de dos formas distintas: los modelos basados en observaciones y los basados en parámetros. A pesar de que los modelos basados en observaciones son útiles para pronosticar un solo período, se enfrentan con dificultades al momento de calcular los valores del tiempo de vida útil y el valor de capital de marca de un cliente debido a la propagación del error. La mayoría de los modelos desarrollados por investigadores interesados en el cálculo del valor del capital de un cliente han empleado modelos basados en parámetros con el objetivo de generar pronósticos a múltiples períodos basados en comportamientos clave para la derivación del valor de tiempo de vida útil y el valor de capital de los consumidores.

En lugar de tratar de estimar los efectos de los factores que pueden influenciar el comportamiento del consumidor, los modelos probabilísticos utilizan la noción de la estocacidad, viendo el comportamiento de interés como el resultado de uno o varios procesos probabilísticos. Esto no implica que un consumidor se comporte aleatoriamente, más bien, funge como una manera de tomar en cuenta la incertidumbre concerniente a los factores que influyen en el comportamiento de un consumidor. El proceso probabilístico captura el efecto neto de todas estas influencias no consideradas a priori en el modelo [1].

4.1. Aspectos generales del valor de tiempo de vida útil de un consumidor

La pregunta fundamental en cualquier actividad de mercadeo es: “¿cuál es el propósito de dicha actividad para mejorar el rendimiento de la compañía?” Las principales razones para evaluar cuantitativamente una base de clientes es:

- Mejorar la productividad del mercadeo.
- Mejorar el desarrollo de las relaciones con los clientes.
- Desarrollo de una ventaja competitiva sostenible.

La razón para mejorar la productividad del mercado se deriva de tres factores principales:

- (a) Uno de los principales problemas del mercadeo masivo es la falta de focalización.
- (b) Un departamento de mercadeo necesita asumir la responsabilidad de sus acciones.
- (c) Los esfuerzos de mercadeo masivo son difíciles de enfocar y llevar a cabo, mientras que el análisis de una base de clientes puede llevar a un proceso de focalización más efectivo.

Por otro lado, el desarrollo de las relaciones con los clientes lleva a la fomentación de lealtad a la marca mejorando así los niveles de retención y frecuencia de consumo.

El concepto del desarrollo de una ventaja competitiva sostenible se puede traducir a un esfuerzo integral para establecer y fortalecer continuamente las relaciones con los clientes a través de interacciones personalizadas centradas en agregar valor, intercambios entre la compañía y el consumidor y constancia en el cumplimiento de promesas realizadas al individuo, todo esto centrado en el hecho que muchas veces la compañía cuenta con información de los clientes con las que sus competidores no cuentan [9].

Definición 4.1.1. El **valor de tiempo de vida útil** (LTV por sus siglas en inglés) de un consumidor se define como el valor presente neto de las ganancias asociadas a dicho consumidor luego de restar los costos asociados a mercadeo, ventas, producción y otros servicios [9].

Nota 4.1.1. Existen tres puntos importantes a considerar en la definición anterior:

1. El pronóstico de ventas para un consumidor.
2. El cálculo de los costos incrementales por consumidor.
3. La determinación de la tasa de descuento adecuada en el cálculo del valor presente neto.

Definición 4.1.2. El valor del tiempo de vida útil de un consumidor se puede calcular como:

$$\text{LTV} = \sum_{t=1}^{\infty} \frac{E[\bar{V}_t]}{(1 + \delta)^{t-1}}$$

donde:

- \bar{V}_t es una variable aleatoria representando la ganancia neta brindada por el consumidor en el período t .
- δ la tasa de descuento por unidad de tiempo t [9].

4.1.1. Los principales modelos de LTV: retención simple y migración

Existen dos principales modelos para el cálculo del LTV: retención simple y migración. Los modelos de retención simple asumen que el consumidor se retiró, que está perdido para la compañía. Los modelos de migración reconocen que un consumidor puede migrar entre ser un consumidor o no durante su tiempo de vida como cliente. Generalmente, los modelos de retención simple son aplicables en industrias como las de servicios financieros y servicios de suscripción, mientras que los modelos de migración aplican a industrias como *retail* y ventas por catálogo [9].

Nota 4.1.2. Los modelos de riesgo se utilizan para el cálculo de S_t , la probabilidad que un consumidor haya sobrevivido hasta el período t . Sea T la variable aleatoria representando el tiempo en que el consumidor se retira con función de densidad de probabilidad $f(t)$.

La probabilidad de retiro es $P(\bar{T} < t) = F(t)$, donde $F(t)$ es la función de distribución acumulada

$$F(t) = \int_0^t f(x) dx.$$

La probabilidad de que un consumidor sobreviva después del período t es $S_t = 1 - F(t)$. La función de riesgo es útil, ya que permiten representar la probabilidad de que el consumidor se retire en el instante Δt a través de $h(t) = \frac{f(t)}{S(t)}$. Las funciones de riesgo pueden representarse también por

$$h(t) = \frac{d}{dx} \log(S_t).$$

Por lo tanto, para una función de riesgo dada, existe una relación uno a uno con la función de supervivencia [9].

Nota 4.1.3. A diferencia de los modelos de retención simple, los modelos de migración establecen que un consumidor puede transicionar entre distintos estados a lo largo de su tiempo de vida. Operacionalmente, defínase el **estado de actualidad** j cuando el consumidor realizó compras con j períodos de anterioridad.

Por otro lado, defínase p_j como la probabilidad de que el consumidor haya realizado compras en el período actual dado que su último consumo fue con j períodos de anterioridad. Los modelos de migración se reducen a modelos de retención simple cuando $p_j = 0, j > 1$.

Los cálculos en los modelos de migración pueden generalizarse a través de una cadena de Markov.

El modelo de migración puede representarse con una matriz de transición de dimensión de 3×3 , como se presenta a continuación:

$$\mathbf{P} = \begin{bmatrix} p_1 & 1 - p_1 & 0 \\ p_2 & 0 & 1 - p_2 \\ p_3 & 0 & 1 - p_3 \end{bmatrix} \quad (4.1)$$

Cada elemento de \mathbf{P} representa la probabilidad de que un consumidor migre de un estado a otro en un solo período.

Por la propiedad de Markov, es posible calcular la matriz de transición en un horizonte de t períodos, es decir, \mathbf{P}^t , en donde el (i, j) -ésimo elemento de la matriz es la probabilidad de que el consumidor que empieza en el i -ésimo estado y se encuentre en el j -ésimo estado t períodos después.

Asumiendo que una compañía gana la cantidad de MC si un consumidor compra y la compañía invierte M para asegurar su recompra. El vector de ganancias \mathbf{G} puede escribirse como:

$$\mathbf{G} = \begin{bmatrix} MC - M \\ -M \\ -M \end{bmatrix} \quad (4.2)$$

La primera fila de \mathbf{G} representa la contribución a la ganancia si el consumidor se encuentra en el estado R_1 (recién comprado), la segunda fila representa la contribución a la ganancia si el cliente se encuentra en el estado R_2 (el consumidor no compró pero se invirtió M en mercadeo enfocado a este), el mismo caso para la tercera fila.

Sea Π_1 el vector de ganancias luego de un período, entonces:

$$\Pi_1 = \mathbf{P}\mathbf{G} = \begin{bmatrix} p_1 MC - M \\ p_2 MC - M \\ p_3 MC - M \end{bmatrix} \quad (4.3)$$

Por tanto, el vector de ganancia esperada luego de dos períodos es $\mathbf{P}^2\mathbf{G}$ y en general, $\mathbf{P}^T\mathbf{G}$ luego de T períodos.

Asimismo, el vector del valor presente neto esperado con un factor de descuento d luego de un período es $\mathbf{P}\mathbf{G}/(1+d)$ y $\mathbf{P}^T\mathbf{G}/(1+d)^T$ para T períodos. De esta manera, se tiene que el vector del valor presente neto total desde el período 0 al período T es:

$$\mathbf{V}^T = \sum_{t=0}^T [(1+d)^{-1}\mathbf{P}]^t \mathbf{G} \quad (4.4)$$

Generalizando a un horizonte de tiempo infinito se tiene entonces que:

$$\mathbf{V}^\infty = \lim_{T \rightarrow \infty} \mathbf{V}^T = [\mathbf{I} - (1+d)^{-1}\mathbf{P}]^{-1}\mathbf{G} \quad (4.5)$$

donde \mathbf{I} es la matriz identidad. El vector en la ecuación anterior es de particular utilidad porque cada elemento represente el LTV de un consumidor empezando en los estados R_1, R_2 , etc. El primer elemento de \mathbf{V}^∞ es el valor a largo plazo de un consumidor que inicia en el estado R_1 [9].

4.2. Procesos estocásticos

El desarrollo de los modelos para el pronóstico del comportamiento de clientes se fundamenta en la teoría de los procesos estocásticos, los cuales se definen de la siguiente manera:

Definición 4.2.1. Un **proceso estocástico** $\{X(t) : t \in T\}$ es una colección de variables aleatorias. Las variables X_t toman valores en un conjunto χ llamado el **espacio de estados**. El conjunto T es llamado el **conjunto de índices**, el cual puede ser considerado como el tiempo. El conjunto de índices puede ser discreto o continuo [10].

Definición 4.2.2. Considere un proceso estocástico $X(t)$. Para un tiempo fijo t_1 , $X(t_1) = X_1$ es una variable aleatoria y su función de distribución acumulada es definida como:

$$F_X(x_1; t_1) = P\{X(t_1) \leq x_1\} \quad (4.6)$$

entonces, la función $F_X(x_1; t_1)$ es denotada como la **distribución de primer orden** de $X(t)$ [6].

Nota 4.2.1. Si $X(t)$ es un proceso en tiempo discreto, entonces $X(t)$ está determinado por una colección de funciones de masa de probabilidad:

$$p_X(x_1, \dots, x_n; t_1, \dots, t_n) = P\{X(t_1) = x_1, \dots, X(t_n) = x_n\}. \quad (4.7)$$

Si $X(t)$ es un proceso en tiempo continuo, entonces $X(t)$ está especificado por una colección de funciones de densidad de probabilidad:

$$f_X(x_1, \dots, x_n; t_1, \dots, t_n) = \frac{\partial^n F_X(x_1, \dots, x_n; t_1, \dots, t_n)}{\partial x_1 \dots \partial x_n}. \quad (4.8)$$

La caracterización completa de $X(t)$ requiere conocer todas las distribuciones conforme $n \rightarrow \infty$ [6].

Definición 4.2.3. La **media** de $X(t)$ está definida por:

$$\mu_X(t) = E[X(t)] \quad (4.9)$$

donde $X(t)$ es tratado como una variable aleatoria para un valor fijo de t [6].

Nota 4.2.2. Una medida de dependencia entre las variables aleatorias de $X(t)$ es proveída por la **función de autocorrelación**, definida por:

$$R_X(t, s) = E[X(t)X(s)]. \quad (4.10)$$

Nótese que:

$$R_X(t, s) = R_X(s, t).$$

Y además:

$$R_X(t, t) = E[X^2(t)] [6].$$

Definición 4.2.4. La **función de autocovarianza** de $X(t)$ está definida por:

$$K_X(t, s) = Cov[X(t), X(s)] = E[X(t) - \mu_X(t)][X(s) - \mu_X(s)] \quad (4.11)$$

$$= R_X(t, s) - \mu_X(t)\mu_X(s) \quad (4.12)$$

Nótese que si la media de $X(t)$ es cero, entonces $K_X(t, s) = R_X(t, s)$ [6].

Definición 4.2.5. La **varianza** de $X(t)$ está definida por:

$$\sigma_X^2(t) = Var[X(t)] = E\{[X(t) - \mu_X(t)]^2\} = K_X(t, t) [6]. \quad (4.13)$$

4.2.1. Clasificación de procesos estocásticos

Definición 4.2.6. Un proceso estocástico $\{X(t) : t \in T\}$ es **estacionario**, si para todo n y para todo conjunto de instantes $t_i \in T, i = 1, 2, \dots, n$, se tiene:

$$F_X(x_1, \dots, x_n; t_1, \dots, t_n) = F_X(x_1, \dots, x_n; t_1 + \tau, \dots, t_n + \tau), \forall \tau \quad [6]. \quad (4.14)$$

Nota 4.2.3. La distribución en un proceso estacionario no se ve afectada por un cambio en el origen del tiempo, y por lo tanto, $X(t)$ y $X(t + \tau)$ tendrán las mismas distribuciones para cualquier τ . Por tanto, para la distribución de primer orden se tiene:

$$F_X(x; t) = F_X(x; t + \tau) = F_X(x) \quad (4.15)$$

$$f_X(x; t) = f_X(x) \quad (4.16)$$

$$\mu_X(t) = E[X(t)] = \mu \quad (4.17)$$

$$Var[X(t)] = \sigma^2 \quad [6]. \quad (4.18)$$

Definición 4.2.7. Si la condición en la definición anterior no se cumple para todo n pero se cumple para $n \leq k$, se dice que el proceso $X(t)$ es **estacionario de orden k** . Si $X(t)$ es estacionario de orden 2, entonces se dice que $X(t)$ es **débilmente estacionario**. Si $X(t)$ es débilmente estacionario, entonces se tiene:

$$E[X(t)] = \mu \text{ constante}$$

$$R_X(t, s) = E[X(t)X(s)] = R_X(|s - t|) \quad [6].$$

Definición 4.2.8. En un proceso estocástico $X(t)$, si $X(t_i)$ para $i = 1, 2, \dots, n$ son variables aleatorias independientes tales que para $n = 1, 2, 3, \dots, n$ se tiene:

$$F_X(x_1, \dots, x_n; t_1, \dots, t_n) = \prod_{i=1}^n F_X(x_i; t_i) \quad (4.19)$$

entonces, $X(t)$ es un **proceso estocástico independiente**. Por lo tanto, una distribución de primer orden es suficiente para caracterizar un proceso independiente [6].

Definición 4.2.9. Un proceso estocástico $\{X(t) : t \geq 0\}$ tiene **incrementos independientes** si para $0 < t_1 < t_2 < \dots < t_n$, las variables aleatorias

$$X(0), X(t_1) - X(0), \dots, X(t_n) - X(t_{n-1})$$

son independientes. Si $\{X(t) : t \geq 0\}$ tiene incrementos independientes y $X(t) - X(s)$ tienen la misma distribución que $X(t+h) - X(s+h)$ para todos $s, t, h \geq 0, s < t$, entonces el proceso $X(t)$ se dice que tiene **incrementos independientes estacionarios** [6].

Propiedad 4.2.1. Si $\{X(t) : t \geq 0\}$ es un proceso con incrementos independientes estacionarios y si $X(0) = 0$, entonces:

$$E[X(t)] = \mu_1 t \quad (4.20)$$

$$Var[X(t)] = \sigma_1^2 t \quad (4.21)$$

donde $\mu_1 = E[X(1)]$ y $\sigma_1^2 = Var[X(1)]$ [6].

Demostración. Sea $f(t) = E[X(t)] = E[X(t) - X(0)]$. Entonces, para cualesquiera s, t y por la propiedad de un proceso con incrementos independientes estacionarios se tiene:

$$\begin{aligned} f(t+s) &= E[X(t+s) - X(0)] \\ &= E[X(t+s) - X(s)] + E[X(s) - X(0)] \\ &= E[X(t) - X(0)] + E[X(s) - X(0)] \\ &= f(t) + f(s) \end{aligned}$$

Nótese que la única solución a la ecuación anterior es $f(t) = ct$, con c constante. Dado que $c = f(1) = E[X(1)]$, tenemos

$$E[X(t)] = \mu_1 t$$

De la misma manera, sea $g(t) = \text{Var}[X(t) - X(0)]$. Entonces, $g(t+s) = g(t) + g(s) \Rightarrow g(t) = kt$, con lo cual $k = g(1) = \text{Var}[X(1)]$ y por lo tanto

$$\text{Var}[X(t)] = \sigma_1^2 t$$

□

Nota 4.2.4. De la propiedad anterior se tiene que todos los procesos con incrementos independientes son no estacionarios.

Definición 4.2.10. Un proceso estocástico $\{X(t) : t \in T\}$ es un **proceso de Markov** si:

$$P\{X(t_{n+1}) \leq x_{n+1} | X(t_1) = t_1, \dots, X(t_n) = t_n\} = P\{X(t_{n+1}) \leq x_{n+1} | X(t_n) = t_n\} \quad (4.22)$$

para $t_1 < t_2 < \dots < t_n < t_{n+1}$ [6].

Definición 4.2.11. Un proceso de Markov discreto es denominado una **cadena de Markov** si para un parámetro discreto $\{X_n : n \geq 0\}$ se tiene que para cada n :

$$P(X_{n+1} = j | X_0 = i_0, X_1 = i_1, \dots, X_n = i) = P(X_{n+1} = j | X_n = i) \quad [6]. \quad (4.23)$$

Nota 4.2.5. La propiedad que caracteriza a los procesos de Markov indica que el estado futuro de un proceso depende únicamente del estado presente y no de su historia. Por lo tanto, todo proceso con incrementos independientes es un proceso de Markov.

Propiedad 4.2.2. La distribución de n -ésimo orden de un proceso de Markov $X(t)$ puede expresarse como:

$$F_X(x_1, \dots, x_n; t_1, \dots, t_n) = F_X(x_1; t_1) \prod_{k=2}^n P\{X(t_k) \leq x_k | X(t_{k-1}) = x_{k-1}\} \quad (4.24)$$

Por tanto, todas las distribuciones de orden finito de un proceso de Markov pueden expresarse en términos de sus distribuciones de segundo orden [6].

Demostración. Sea $X(t)$ un proceso de Markov con distribución de n -ésimo orden:

$$F_X(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) = P\{X(t_1) \leq x_1, X(t_2) \leq x_2, \dots, X(t_n) \leq x_n\}$$

Por la propiedad de Markov se tiene que

$$\begin{aligned} F_X(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) &= P\{X(t_n) \leq x_n | X(t_1) \leq x_1, \dots, X(t_{n-1}) \leq x_{n-1}\} \\ &\times P\{X(t_1) \leq x_1, \dots, X(t_{n-1}) \leq x_{n-1}\} \\ &= P\{X(t_n) \leq x_n | X(t_{n-1}) \leq x_{n-1}\} F_X(x_1, \dots, x_{n-1}; t_1, \dots, t_{n-1}) \end{aligned}$$

Aplicando la igualdad anterior para distribuciones de menor orden se tiene que:

$$F_X(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) = F_X(x_1, t_1) \prod_{k=2}^n P\{X(t_k) \leq x_k | X(t_{k-1}) \leq x_{k-1}\}$$

□

4.3. Reseña de modelos estocásticos *Buy 'Till You Die* - BTYD

4.3.1. Clasificación de entornos de análisis

Previo a la reseña de los modelos, es necesario entender los tipos de relaciones que se tienen entre una compañía y sus clientes, las cuales guían la formulación de los modelos. Para ello, considere el siguiente ejemplo tomando como base los enunciados con respecto al tamaño de la base de clientes de dos compañías:

- Basado en las cifras presentadas en un comunicado de prensa, se reporta que Vodafone UK tuvo 11.3 millones de “pagadores mensuales” al final del período del 30 de septiembre de 2013.
- En el reporte de ganancias del tercer trimestre de 2013, el gerente financiero de Amazon comentó que la cantidad de consumidores activos ha excedido la cantidad de 224 millones, considerando a un cliente como activo si ha realizado una orden en los últimos doce meses.

A primera vista, ambos parecen brindar una caracterización lógica de la cantidad de clientes que poseen las compañías, sin embargo, solo el primer enunciado es preciso con respecto a esta cifra. Esto es debido a que estos “pagadores mensuales” deben informar a Vodafone en el momento que deseen terminar su relación con la compañía, y por lo tanto saben el momento exacto en que la relación con el cliente “muere”. Este modelo de negocio es visto como un **entorno contractual** entre la compañía y el consumidor. Por otro lado, en el caso de Amazon la “muerte” de la relación entre ellos y el cliente no es observable, y los clientes no terminan su relación formalmente con ellos, solo se “retiran silenciosamente”, este modelo es denominado como un **entorno no contractual**.

El reto en el análisis de entornos no contractuales es la diferenciación entre los clientes que han terminado su relación con la compañía contra los que simplemente se han tomado un largo intervalo de tiempo entre sus transacciones. Por otro lado, también debe considerarse la frecuencia en la que ocurren las transacciones, estas pueden ocurrir en eventos discretos en el tiempo o sin restricción alguna en la cual el cliente puede comprar o terminar su relación con la compañía. Esto lleva a la siguiente clasificación de las oportunidades de transacciones [4]:

Continuo	Compra de abarrotes	Tarjetas de crédito
	Consultas médicas	Servicios y utilidades
Discreto	Estadías en hoteles	Programas de continuidad
	Recetas médicas	Servicios de suscripción
	Asistencia a eventos y conferencias	Seguros
	No contractual	Contractual

Tabla 4.1: Clasificando bases de clientes

4.3.2. Revisando el concepto de LTV

A diferencia de la sección anterior, que se enfocaba en el cálculo del LTV de manera determinista, este enfoque busca la obtención el mismo valor pero valiéndose de las herramientas brindadas por los modelos estocásticos utilizados para el comportamiento de una base de clientes. Se empieza por definir el valor esperado del LTV en términos del comportamiento transaccional de un cliente.

Definición 4.3.1. Matemáticamente, se puede definir el valor esperado del LTV como:

$$E[\text{LTV}] = \int_0^{\infty} E[v(t)]S(t)d(t) dt \quad (4.25)$$

para $t \geq 0$, con $t = 0$ representando “ahora mismo”, $E[v(t)]$ es el valor esperado (o flujo de efectivo neto) del consumidor en el tiempo t , $S(t)$ es la probabilidad de que el consumidor se haya mantenido al menos hasta el tiempo t y $d(t)$ es el factor de descuento que refleja el valor presente de las ganancias recibidas al tiempo t [4].

Nota 4.3.1. Es importante distinguir entre el valor esperado del LTV para un consumidor que aún no ha sido adquirido y el de uno ya adquirido.

Definición 4.3.2. El **valor de tiempo de vida útil residual** esperado, $E[RLV]$, de un consumidor existente es el valor de la primera transacción que marca el inicio de la relación. Matemáticamente se denota como:

$$E[RLV] = \int_{t'}^{\infty} E[v(t)]S(t|t > t')d(t - t') dt \quad (4.26)$$

donde t' es la “edad” del cliente al momento en que su valor residual es calculado [4].

Nota 4.3.2. Considere un entorno contractual, si se asume que el valor esperado para el flujo de efectivo neto por período contractual es constante (i.e., $E[v(t)] = \bar{v}$), entonces podemos expresar $E[LTV]$ como:

$$E[LTV] = \bar{v}DEL \quad (4.27)$$

donde el valor esperado del tiempo de vida descontando, DEL se define por:

$$DEL = \int_0^{\infty} S(t)d(t) dt \quad (4.28)$$

De manera análoga, se puede expresar el valor residual como:

$$E[RLV] = \bar{v}DERL \quad (4.29)$$

donde DERL se define como:

$$DERL = \int_{t'}^{\infty} \frac{S(t)}{S(t')} d(t - t') dt [4]. \quad (4.30)$$

Nota 4.3.3. En un entorno no-contractual, al asumir que el flujo de efectivo neto por transacción es constante, se tiene que $v(t) = \bar{v}t(t)$, donde $t(t)$ es la tasa de transacciones al tiempo t , con lo que se puede expresar $E[LTV]$ como:

$$E[LTV] = \bar{v}DET \quad (4.31)$$

donde el valor esperado de las transacciones descontadas, DET, se define como:

$$DET = \int_0^{\infty} t(t)S(t)d(t) dt \quad (4.32)$$

para un cliente recién adquirido, DET mide el valor presente de todas las futuras transacciones del cliente.

Por otro lado, $E[RLV]$ se puede expresar a través de:

$$E[RLV] = \bar{v}DERT = \bar{v} \int_{t'}^{\infty} t(t)S(t|t > t')d(t - t') dt [4]. \quad (4.33)$$

4.3.3. Modelos para entornos contractuales

Un modelo útil para datos en tiempo discreto es el modelo beta-geométrico (BG). En un entorno contractual, el modelo se formula asumiendo lo siguiente:

- Al final de cada período contractual, el cliente se mantiene con la compañía con una probabilidad de retención de $1 - \theta$, lo cual es equivalente a decir que la duración de la relación del cliente con la compañía está caracterizada por una distribución geométrica.

- Las diferencias individuales en θ están caracterizadas por una distribución beta con parámetros γ y δ . La función de supervivencia es:

$$S(t|\gamma, \delta) = \frac{B(\gamma, \delta + t)}{B(\gamma, \delta)}, \quad t = 0, 1, 2, \dots \quad (4.34)$$

La tasa de retención para la cosecha de clientes correspondiente al período es:

$$r(t) = \frac{\delta + t - 1}{\gamma + \delta + t - 1} [4].$$

Nota 4.3.4. Sustituyendo 4.34 en 4.28 y resolviendo la integral (la cual se vuelve una suma en tiempo discreto) leva a la siguiente expresión para un cliente que aún no ha sido adquirido:

$$\text{DEL}(\gamma, \delta, d) = {}_2F_1\left(1, \delta; \gamma + \delta; \frac{1}{1+d}\right) \quad (4.35)$$

donde ${}_2F_1(\cdot)$ es la función gaussiana hipergeométrica y d es el factor de descuento apropiado para la representación del valor del dinero en el tiempo.

Similarmente, al sustituir en 4.30 se obtiene la expresión para el valor esperado residual para un consumidor que ha renovado su contrato $n - 1$ veces, evaluado en el momento anterior a conocer si realizará su n -ésima renovación:

$$[\text{DERL}](\gamma, \delta, d; n - 1) = \left(\frac{\delta + n - 1}{\gamma + \delta + n - 1}\right) {}_2F_1\left(1, \delta + n; \gamma + \delta + n; \frac{1}{1+d}\right) [4]. \quad (4.36)$$

El análogo continuo para BG es la distribución de Pareto de segundo tipo (Pareto tipo II), una caracterización de esta distribución la mezcla de gamma con exponencial, es decir:

1. La duración de la relación con cliente con la compañía está dada por una distribución exponencial con parámetro λ .
2. Las diferencias individuales en λ están dadas por una distribución gamma con parámetro de forma s y parámetro de escala β . La función de supervivencia asociada es:

$$S(t|s, \beta) = \left(\frac{\beta}{\beta + t}\right)^s, \quad t \geq 0 \quad (4.37)$$

La tasa de abandono correspondiente está dada por la función de riesgo:

$$h(t) = \frac{s}{\beta + t} [4].$$

Nota 4.3.5. Al sustituir 4.37 en 4.28 y 4.30 y calculando las integrales respectivas se tiene que:

$$\text{DEL}(s, \beta, d_c) = \beta^s d_c^{s-1} \Psi(s, s; \beta d_c) \quad (4.38)$$

$$\text{DERL}(s, \beta, d_c; t') = \Psi(s, s; (\beta + t') d_c) \quad (4.39)$$

donde $\Psi(\cdot)$ es la función hipergeométrica confluyente del segundo tipo y d_c es la tasa de descuento compuesta continuamente [4].

4.3.4. Modelos para entornos no contractuales

Defínase la **tasa de retención** de clientes como la razón de clientes retenidos con respecto al valor al riesgo. Como se estableció previamente, en un entorno no contractual, el abandono de un cliente no es observado y no es observable. La formulación para este modelo (Pareto/NBD) para tiempo continuo bajo este entorno asume lo siguiente:

1. Mientras el cliente está “vivo”, las compras del cliente están caracterizadas por una distribución binomial negativa (NBD) con parámetros r y α .
2. Las “vidas” no observadas del cliente son tratadas como aleatorias y son caracterizadas por una distribución Pareto tipo II con parámetros s y β .

Estos supuestos implican que no es necesario conocer toda la historia transaccional de un consumidor, los únicos parámetros que son necesarios a nivel individual son la frecuencia y la antigüedad. **Notación:** la frecuencia y la antigüedad se denotan por (x, t_x, T) , donde x es el número de transacciones observadas en el intervalo de tiempo $(0, T]$ y $t_x, 0 < t_x \leq T$ es el tiempo de la última transacción [4].

Nota 4.3.6. Al sustituir las caracterizaciones de Pareto/NDB para $t(t)$ y $S(t)$ en 4.32 y 4.33 y resolver las integrales respectivas se tiene:

$$\text{DET}(r, \alpha, s, \beta, d_c) = \frac{r}{\alpha} \beta^s d_c^{s-1} \Psi(s, s; \beta d_c) \quad (4.40)$$

$$\text{DERT}(r, \alpha, s, \beta, d_c; x, t_x, T) = \frac{\Gamma(r+x+1)}{\Gamma(r)} \frac{\alpha^r \beta^s d_c^{s-1}}{(\alpha+T)^{r+x+1}} \frac{\Psi(s, s; (\beta+T)d_c)}{L(r, \alpha, s, \beta | x, t_x, T)} \quad (4.41)$$

donde $L(\cdot | \cdot)$ es la función de verosimilitud del modelo Pareto/NBD [4].

Nota 4.3.7. Para poder calcular LTV y RLV con este modelo, es necesario aumentar las ecuaciones 4.40 y 4.41 para incluir el gasto por transacción. Para esto, se asume que un consumidor solo puede terminar su relación con la compañía inmediatamente después de una transacción y no en cualquier momento, esto se caracteriza con una distribución beta geométrica con parámetros a y b , resultando en las siguientes ecuaciones para el valor esperado condicional de DERT y DET:

$$E(X(T, T+t) | r, \alpha, a, b; x, t_x, T) = \left(\frac{a+b+x-1}{a-1} \right) \left[1 - \left(\frac{\alpha+T}{\alpha+T+t} \right)^{r+x} {}_2F_1 \left(r+x, b+x; a+b+x-1; \frac{t}{\alpha+T+t} \right) \right] \left[1 + \delta_{x>0} \frac{a}{b+x-1} \left(\frac{\alpha+T}{\alpha+t_x} \right)^{r+x} \right]^{-1} \quad (4.42)$$

$$\text{DET} = \sum_{t=1}^{\infty} \left(\frac{1}{1+d} \right)^{i-0.5} \{ E(X(t) | r, \alpha, a, b) - E(X(t-1) | r, \alpha, a, b) \} \quad (4.43)$$

donde el valor esperado para la cantidad de transacciones en el intervalo $(0, t]$ es:

$$E(X(t) | r, \alpha, a, b) = \frac{a+b-1}{a-1} \left[1 - \left(\frac{\alpha}{\alpha+t} \right)^r {}_2F_1 \left(r, b, a+b-1; \frac{t}{\alpha+t} \right) \right] [4] \quad (4.44)$$

Nota 4.3.8. El análogo discreto de este modelo asume lo siguiente:

1. Las transacciones de un cliente están caracterizadas por un proceso de Bernoulli con heterogeneidad β .
2. Las “vidas” no observadas están caracterizadas por una distribución beta geométrica con parámetros γ y δ .

La versión discreta no requiere información de cuándo se dieron las x transacciones, el único requerimiento es la frecuencia y antigüedad de las transacciones calibrado a n períodos discretos,

denotado por (x, t_x, n) .

Las ecuaciones para DET y DERT son como sigue:

$$\text{DET}(\alpha, \beta, \gamma, \delta, d; x, t_x, n) = \left(\frac{\alpha}{\alpha + \beta}\right) \left(\frac{\delta}{\gamma + \delta}\right) \left(\frac{1}{1 + d}\right) {}_2F_1\left(1, \delta + 1; \gamma + \delta + 1; \frac{1}{1 + d}\right) \quad (4.45)$$

$$\text{DERT}(\alpha, \beta, \gamma, \delta, d; x, t_x, n) = \frac{B(\alpha + x + 1, \beta + n - x)B(\gamma, \delta + n + 1)}{B(\alpha, \beta)B(\gamma, \delta)(1 + d)} \frac{{}_2F_1\left(1, \delta + n + 1; \gamma + \delta + n + 1; \frac{1}{1 + d}\right)}{L(\alpha, \beta, \gamma, \delta | x, t_x, n)} \quad (4.46)$$

donde $L(\cdot | \cdot)$ es la función de verosimilitud del modelo BG/BB [4].

Los modelos y sus respectivas aplicaciones en los entornos adecuados se resumen en el siguiente cuadro:

Continuo	Pareto/NBD BG/NBD	Pareto (II)
Discreto	BG/BB No contractual	BG Contractual

Tabla 4.2: Conjunto de modelos estocásticos para el cálculo de LTV según tipo de base de clientes

4.4. Selección de modelos BTYD

4.4.1. Pareto/NBD

Como se estableció en la sección anterior, el modelo Pareto/NBD se basa en los siguientes supuestos:

1. El número de transacciones realizadas por un cliente activo en el período t tiene una distribución de Poisson con media λt .
2. La heterogeneidad en la tasa de transacciones λ a través de los clientes tiene una distribución gamma con parámetro de forma r y parámetro de escala α .
3. Cada cliente tiene una “vida” no observada de largo τ . Este punto en donde el cliente se torna inactivo sigue una distribución exponencial con tasa de abandono μ .
4. La heterogeneidad en las tasas de abandono sigue una distribución gamma con parámetro de forma s y parámetro de escala β .
5. La tasa de transacciones λ y la tasa de abandono μ varían independientemente a través de clientes.

Partiendo de la frecuencia y antigüedad de transacciones $(X = x, t_x, T)$ se pueden derivar las siguientes cantidades:

- $E[X(t)]$, el valor esperado para el número de transacciones en el período t .
- $P(X(t) = x)$, la probabilidad de observar x transacciones en el período t .
- $E(Y(t)|X = x, t_x, T)$, el valor esperado de transacciones en el período $(T, T + t]$ para un cliente con comportamiento observado $(X = x, t_x, T)$.

La función de verosimilitud asociada con este modelo requiere numerosas evaluaciones de la función hipergeométrica gaussiana, lo cual es muy demandante computacionalmente y por lo tanto la precisión para evaluar esta función numericamente varía sustancialmente sobre el espacio de parámetros, causando problemas para optimizar dicha función [5].

4.4.2. BG/NBD

Muchos aspectos del modelo BG/NBD reflejan al modelo Pareto/NBD, la única diferencia yace en cómo y cuándo un cliente se torna inactivo. El modelo de Pareto asume que el abandono puede ocurrir en cualquier momento en el tiempo, independientemente de la ocurrencia de las transacciones, asumiendo que que el abandono ocurre inmediatamente después de una transacción, entonces se puede modelar dicho proceso utilizando una distribución beta-geométrica (BG). Formalmente, el modelo asume lo siguiente:

1. El número de transacciones realizadas por un cliente activo sigue una distribución de Poisson con tasa de transacción λ . Esto es equivalente a asumir que el tiempo entre transacciones está distribuido de manera exponencial con una tasa λ , esto es:

$$f(t_j|t_{j-1}; \lambda) = \lambda e^{-\lambda(t_j - t_{j-1})}, \quad t_j > t_{j-1} \geq 0.$$

2. La heterogeneidad en λ tiene una distribución gamma con la siguiente función de densidad de probabilidad:

$$f(\lambda|r, \alpha) = \frac{\alpha^r \lambda^{r-1} e^{-\lambda\alpha}}{\Gamma(r)}, \quad \lambda > 0. \quad (4.47)$$

3. Después de cualquier transacción, un cliente se torna inactivo con probabilidad p . Por lo tanto, el punto en el que el cliente abandona está distribuido a través de las transacciones con una distribución geométrica con función de masa de probabilidad:

$$P(\text{inactivo inmediatamente después de la } j\text{-ésima transacción}) = p(1-p)^{j-1}, \quad j = 1, 2, 3, \dots$$

4. La heterogeneidad en p sigue una distribución beta con la siguiente función de densidad:

$$f(p|a, b) = \frac{p^{a-1}(1-p)^{b-1}}{B(a, b)}, \quad 0 \leq p \leq 1 \quad (4.48)$$

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

5. La tasa de transacción λ y la probabilidad de abandono p varían independientemente a través de clientes.

Derivación de las funciones que componen el modelo

Considere entonces un cliente que ha realizado x transacciones en el período $(0, T]$ con las transacciones ocurriendo en los tiempos t_1, t_2, \dots, t_x . A partir de esto, la función de verosimilitud a nivel individual puede derivarse de la siguiente manera:

- La verosimilitud de la primera transacción ocurriendo en t_1 es $\lambda e^{-\lambda t_1}$, el componente estándar de la verosimilitud exponencial.
- La verosimilitud de la segunda transacción ocurriendo en t_2 es la probabilidad de permanecer activo en t_1 por el componente estándar de la verosimilitud exponencial, esto es $(1-p)\lambda e^{-\lambda(t_2-t_1)}$.

Esto se repite sucesivamente hasta la transacción x , donde la función de verosimilitud es $(1-p)\lambda e^{-\lambda(t_x-t_{x-1})}$. Por último, la verosimilitud de observar cero transacciones en $(t_x, T]$ es la probabilidad de que el cliente se torne inactivo en t_x en adición a la probabilidad que permanezca activo pero no haya realizado ninguna transacción en este intervalo, esto es $p + (1-p)e^{-\lambda(T-t_x)}$, por tanto, se tiene que la función de verosimilitud es:

$$\begin{aligned} L(\lambda, p|t_1, t_2, \dots, t_x, T) &= \lambda e^{-\lambda t_1} (1-p)\lambda e^{-\lambda(t_2-t_1)} \dots (1-p)e^{-\lambda(t_x-t_{x-1})} \\ &\quad \cdot \{p + (1-p)e^{-\lambda(T-t_x)}\} \\ &= p(1-p)^{x-1} \lambda^x e^{-\lambda t_x} + (1-p)^x \lambda^x e^{-\lambda T}. \end{aligned}$$

Asumiendo que todos los clientes están activos al inicio del período de observación, la verosimilitud de que un cliente realice cero transacciones en el período $(0, T]$ es la función de supervivencia exponencial estándar:

$$L(\lambda|X=0, T) = e^{-\lambda T}$$

Por lo tanto, la función de verosimilitud a nivel individual se puede escribir como:

$$L(\lambda, p|X=x, T) = (1-p)^x \lambda^x e^{-\lambda T} + \delta_{x>0} p(1-p)^{x-1} \lambda^x e^{-\lambda t_x} \quad (4.49)$$

donde $\delta_{x>0} = 1$ si $x > 0$ y 0 en cualquier otro caso.

Sea $X(t)$ la variable aleatoria que denota el número de transacciones ocurriendo en un período de largo t . Con el objetivo de derivar una expresión para $P(X(t) = x)$ es necesario recordar la relación fundamental entre los tiempos entre eventos y la cantidad de eventos, esto es: $X(t) \geq x \iff T_x \leq t$, donde T_x es la variable aleatoria que denota el tiempo de la transacción x . Dado el supuesto al respecto de la naturaleza del proceso de abandono, se tiene que:

$$\begin{aligned} P(X(t) = x) &= P(\text{activo después de la transacción } x) \cdot P(T_x \leq t \wedge T_{x+1} > t) \\ &\quad + \delta_{x>0} \cdot P(\text{inactivo luego de la transacción } x) \cdot P(T_x \leq t) \end{aligned}$$

Partiendo del supuesto que el tiempo entre transacciones está caracterizado por la distribución exponencial, $P(T_x \leq t \wedge T_{x+1} > t)$ es en esencia la probabilidad de Poisson de que $X(t) = x$ y $P(T_x \leq t)$ es la función de distribución de acumulada de la distribución de Erlang, por lo tanto:

$$P(X(t) = x|\lambda, p) = (1-p)^x \frac{(\lambda t)^x e^{-\lambda t}}{x!} + \delta_{x>0} p(1-p)^{x-1} \left[1 - e^{-\lambda t} \sum_{j=0}^{x-1} \frac{(\lambda t)^j}{j!} \right] \quad (4.50)$$

Dado que el número de transacciones sigue un proceso de Poisson, $E[X(t)]$ es simplemente λt si el cliente está activo en t . Para un cliente que se volvió inactivo en $\tau \leq t$, el número esperado de transacciones en el período $(0, \tau]$ es $\lambda \tau$. Sin embargo, la verosimilitud de que un cliente se torne inactivo en τ es:

$$P(\tau > t) = P(\text{activo en } t|\lambda, p) = \sum_{j=0}^{\infty} (1-p)^j \frac{(\lambda t)^j e^{-\lambda t}}{j!} = e^{-\lambda p t}$$

Esto implica que la función de densidad de probabilidad del tiempo de abandono está dada por $g(\tau|\lambda, p) = \lambda p e^{-\lambda p \tau}$. Esto denota una relación explícita con la tasa de transacciones λ , a diferencia

del modelo Pareto/NBD, la cual tiene un proceso de abandono exponencial que es independiente de dicha tasa. De esto se tiene que el valor esperado del número de transacciones en un período de largo t está dado por:

$$E(X(t)|\lambda, p) = \lambda t + P(t\tau > t) + \int_0^t \lambda \tau g(\tau|\lambda, p) d\tau = \frac{1}{p} - \frac{1}{p} e^{-\lambda p t} \quad (4.51)$$

Desarrollo del modelo para un cliente seleccionado aleatoriamente

Todas las expresiones desarrolladas previamente son probabilidades condicionales que dependen de la tasa de transacciones λ y la tasa de abandono p , que son parámetros no observados. Para derivar expresiones equivalentes para un cliente seleccionado aleatoriamente, se deben tomar los valores esperados a nivel individual sobre las distribuciones respectivas para λ y p , como se muestra en las ecuaciones 4.47 y 4.48.

Tomando la expectativa de 4.49 sobre las distribuciones de λ y p resulta en la siguiente expresión para función de verosimilitud para un individuo seleccionado aleatoriamente con historia transaccional ($X = x, t_x, T$):

$$L(r, \alpha, a, b|X = x, t_x, T) = \frac{B(a, b+x)}{B(a, b)} \frac{\Gamma(r+x)\alpha^r}{\Gamma(r)(\alpha+T)^{r+x}} + \delta_{x>0} \frac{B(a+1, b+x-1)}{B(a, b)} \frac{\Gamma(r+x)\alpha^r}{\Gamma(r)(\alpha+t_x)^{r+x}} \quad (4.52)$$

Los cuatro parámetros para el modelo, (r, α, a, b) , pueden estimarse vía el método de máxima verosimilitud. Supóngase que se tiene una muestra de N clientes, donde el cliente i tuvo $X_i = x_i$ transacciones en el período $(0, T_i]$ con la última transacción ocurriendo en el tiempo t_{x_i} . La función muestral para la log-verosimilitud está dada por:

$$LL(r, \alpha, a, b) = \sum_{i=1}^N \ln[L(r, \alpha, a, b|X_i = x_i, t_{x_i}, T_i)] \quad (4.53)$$

esta función puede maximizarse utilizando métodos estándar de optimización numérica.

Tomando la expectativa de 4.50 sobre las distribuciones de λ y p resulta en la siguiente expresión para la probabilidad de observar x transacciones en el período de largo t :

$$P(X(t) = x|r, \alpha, a, b) = \frac{B(a, b+x)}{B(a, b)} \frac{\Gamma(r+x)}{\Gamma(r)x!} \left(\frac{\alpha}{\alpha+t}\right)^r \left(\frac{t}{\alpha+t}\right)^x + \delta_{x>0} \frac{B(a+1, b+x-1)}{B(a, b)} \cdot \left[1 - \left(\frac{\alpha}{\alpha+t}\right)^r {}_2F_1\left(r, b; a+b-1; \frac{t}{\alpha+t}\right)\right] \quad (4.54)$$

Por último, al tomar la expectativa de 4.51 sobre las distribuciones mencionadas, se tiene la siguiente expresión para el valor esperado de transacciones en un período de largo t :

$$E(X(t)|r, \alpha, a, b) = \frac{a+b-1}{a-1} \left[1 - \left(\frac{\alpha}{\alpha+t}\right)^r {}_2F_1\left(r, b; a+b-1; \frac{t}{\alpha+t}\right)\right] \quad (4.55)$$

Nótese que una sola evaluación de la función hipergeométrica gaussiana puede aproximarse fácilmente a través de series polinomiales.

Con el objetivo de implementar el modelo en un análisis predictivo, es necesario obtener el valor esperado para un período futuro de duración t con un comportamiento observado ($X = x, t_x, T$)

$$E(Y(t)|X = x, t_x, T, r, \alpha, a, b) = \frac{\frac{a+b+x-1}{a-1} \left[1 - \left(\frac{\alpha+T}{\alpha+T+t}\right)^{r+x} {}_2F_1\left(r+x, b+x; a+b+x-1; \frac{t}{\alpha+T+t}\right)\right]}{1 + \delta_{x>0} \frac{a}{b+x-1} \left(\frac{\alpha+T}{\alpha+t_x}\right)^{r+x}} \quad [5] \quad (4.56)$$

Análisis exploratorio de la base de datos

La base de datos empleada para este trabajo se obtuvo del sitio web `kaggle.com` [7]. Dicha base consiste de la información transaccional de clientes en una tienda enfocada en ventas al menudeo, la base cuenta con 125,000 observaciones de 3 variables, las cuales se presentan a continuación:

Variable	Tipo	Descripción
<code>cust</code>	Catagórica	Identificador único para el cliente
<code>date</code>	Continua	Fecha de transacción
<code>sales</code>	Continua	Monto de transacción

Tabla 5.1: Descripción de variables que componen la base de datos

En primera instancia, se enlistan las dimensiones de cada una de las variables. La variable `cust` indica que en la base de datos existen 6,889 clientes únicos, mientras que las fechas de transacción datan desde el 16 de mayo de 2011 al 16 de marzo de 2015. Por otro lado, se presentan algunas estadísticas descriptivas para el monto de transacción:

Estadístico	Valor
Media	267.04
Mediana	241.37
Desviación Estándar	148.47
Mínimo	10.07
Máximo	1116.07

Tabla 5.2: Medidas de tendencia central para la variable `sales`

En adición a esto, es de notar que un 44% de las transacciones exceden el valor promedio y que a lo largo de su historia transaccional, 48% de los clientes consume por encima del promedio.

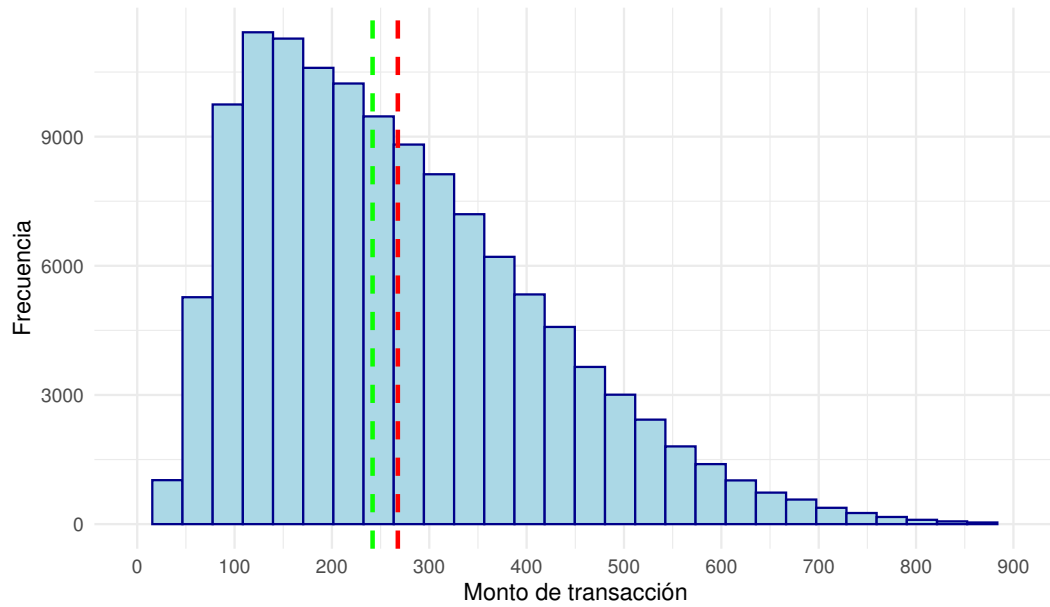


Figura 5.1: Distribución de montos de transacción

Esta tendencia en los consumos de los clientes se ve respaldada por un comportamiento bastante regular en cuanto a la cantidad de transacciones que efectúan, tal y como denotan las medidas de dispersión del volumen transaccional

Estadístico	Valor
Media	18.14
Mediana	18
Desviación Estándar	5.19
Mínimo	4
Máximo	39

Tabla 5.3: Medidas de tendencia central para la cantidad de transacciones por cliente

La división entre clientes que realizaron más transacciones que el promedio y los que no es casi indistinguible, únicamente 47% del total de consumidores realizó más de 18 transacciones.

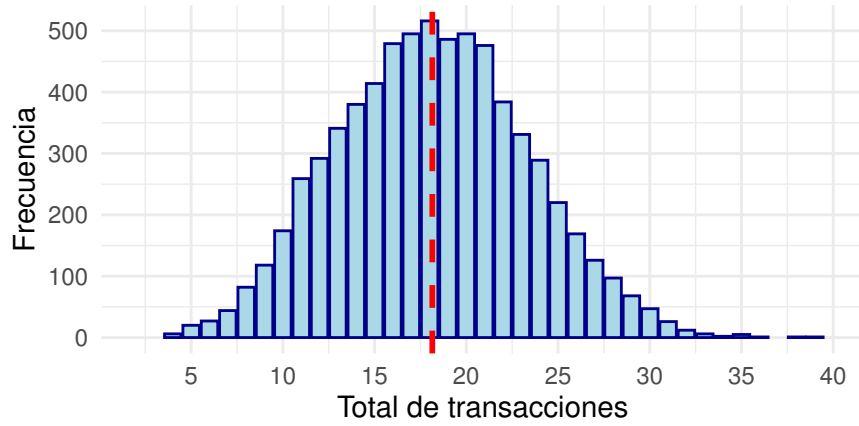


Figura 5.2: Distribución de transacciones por cliente

Ahora bien, la evolución del volumen de transacciones en el tiempo no tiene un patrón definido más allá que un decaimiento cerca de los cierres trimestrales

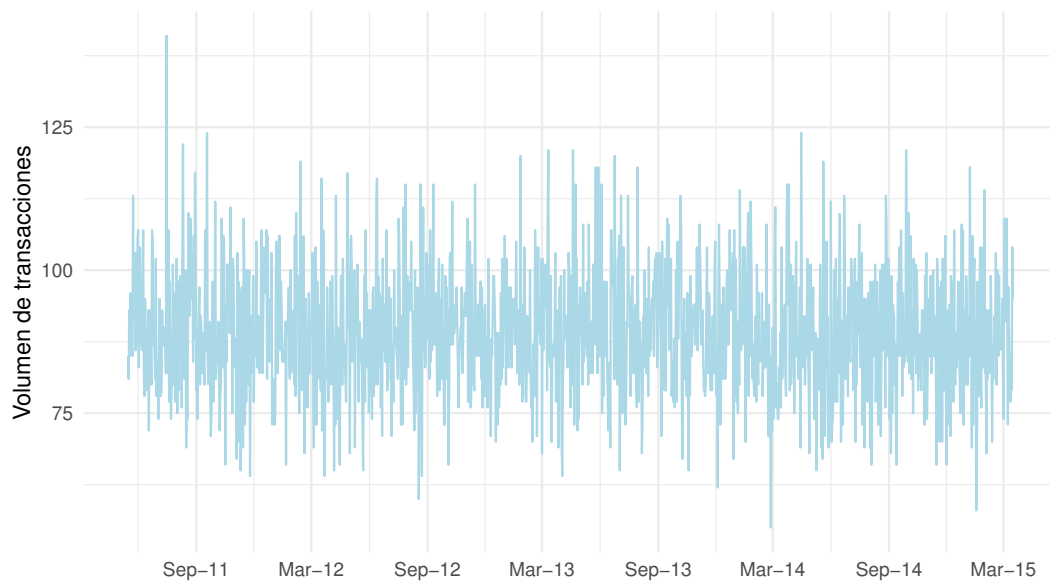


Figura 5.3: Serie de tiempo del volumen transaccional

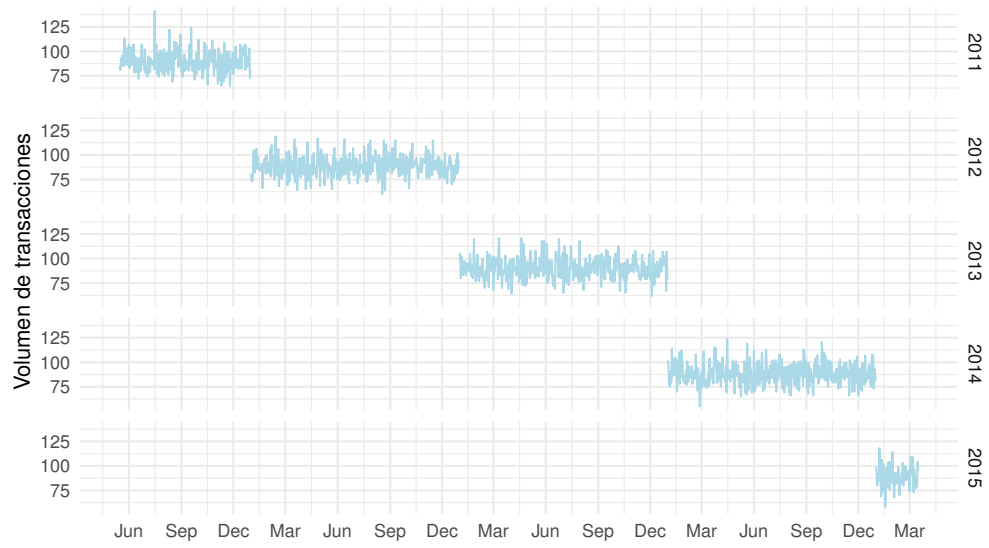


Figura 5.4: Volumen transaccional anual

Esta misma realidad se ve reflejada desde la perspectiva de la evolución en el tiempo del monto total por transacción

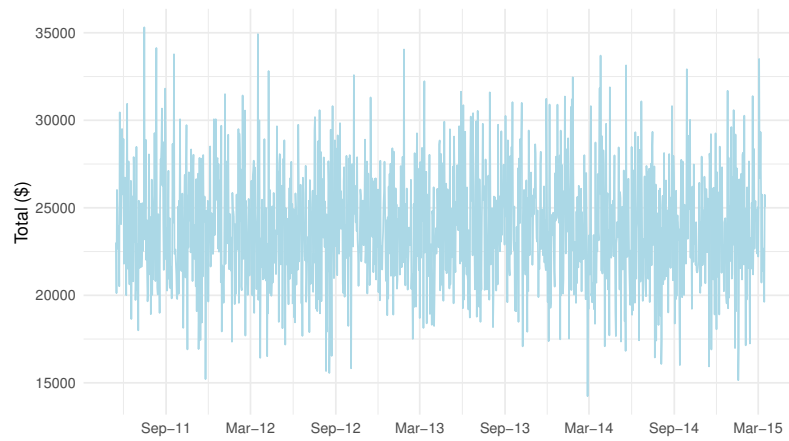


Figura 5.5: Serie de tiempo del monto transaccional

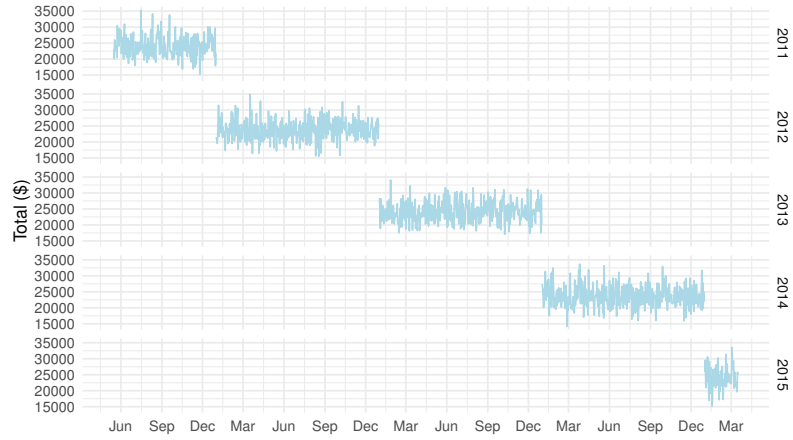


Figura 5.6: Volumen anual

Aunque no es trivial discernir un patrón con respecto a la variación intermensual en base a las figuras anteriores, no es el caso para los cambios interanuales, los cuales denotan mantener la misma tendencia tanto en volumen de transacciones como el monto total.

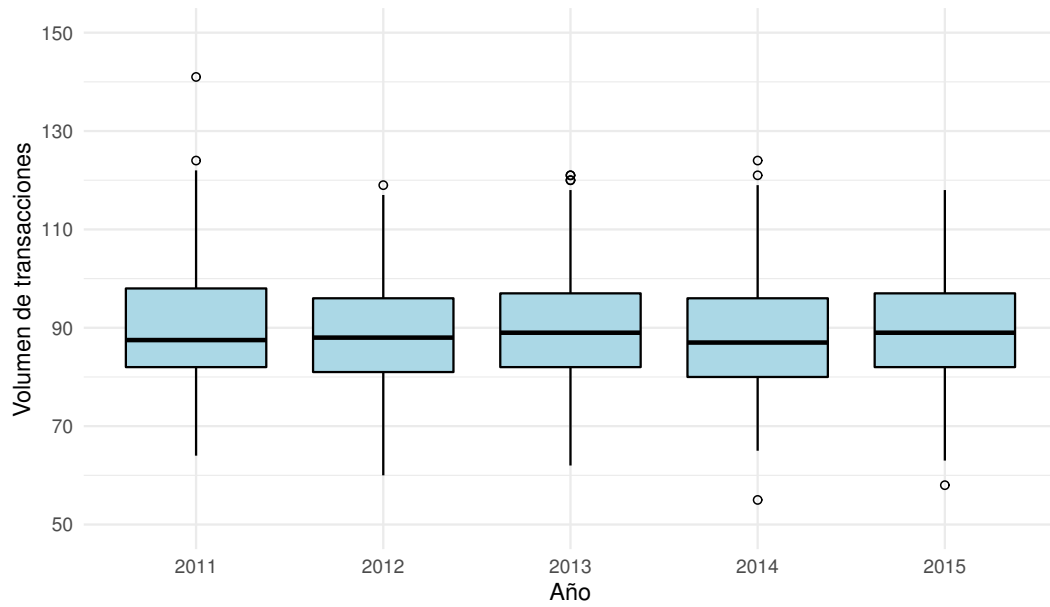


Figura 5.7: Evolución interanual de transacciones

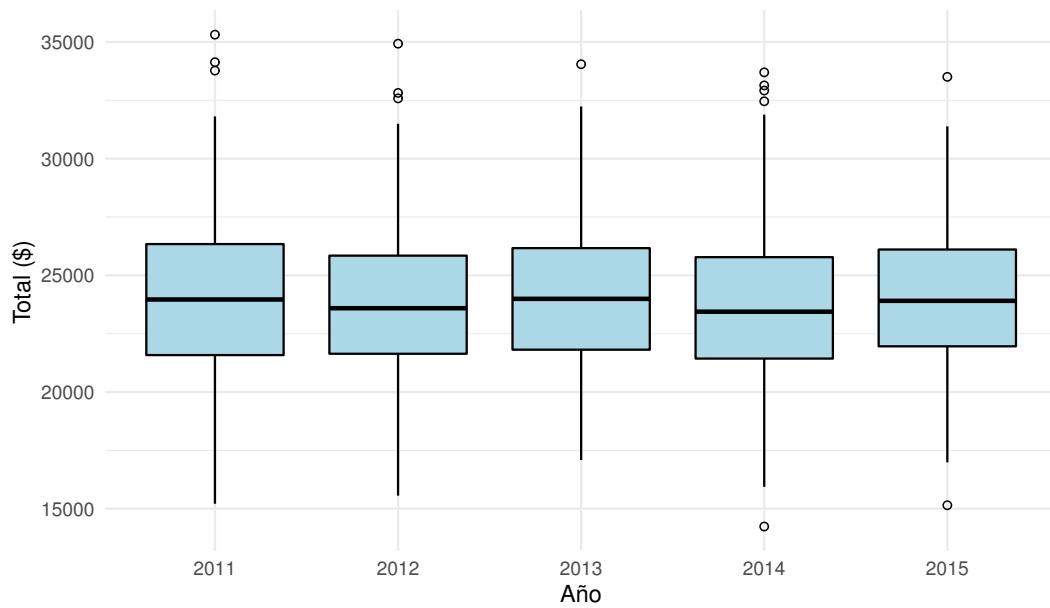


Figura 5.8: Evolución interanual de monto transaccional

Implementación del modelo

Luego del entendimiento obtenido de la base de datos mostrado en 5, se procede a implementar los modelos descritos previamente a través del lenguaje de programación estadística R, en particular, con los paquetes *BTYDplus* [2] y *timeSeries* [3].

Previo a la implementación de los modelos, es necesario definir un período de calibración con el fin de medir la precisión de los pronósticos a realizar contra los datos ya existentes. Dado que se cuentan con 200 semanas de historia para la base de datos dada, se definió un período de calibración de 148 semanas, es decir, se van a comparar los pronósticos de los modelos contra el último año de observaciones.

Habiendo definido el período de calibración, se presentan las tendencias incrementales y acumuladas para el volumen de transacciones

Tendencia incremental de transacciones

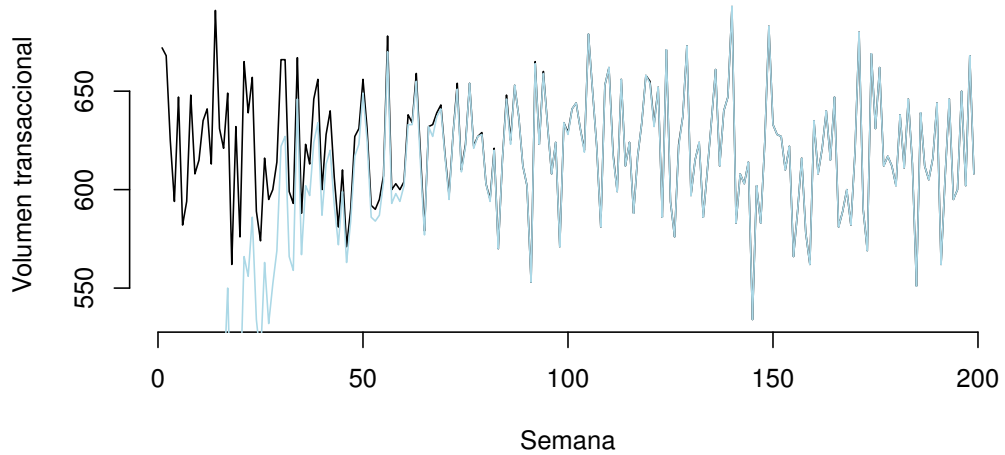


Figura 6.1: Tendencia incremental de transacciones semanales

Tendencia acumulada de transacciones

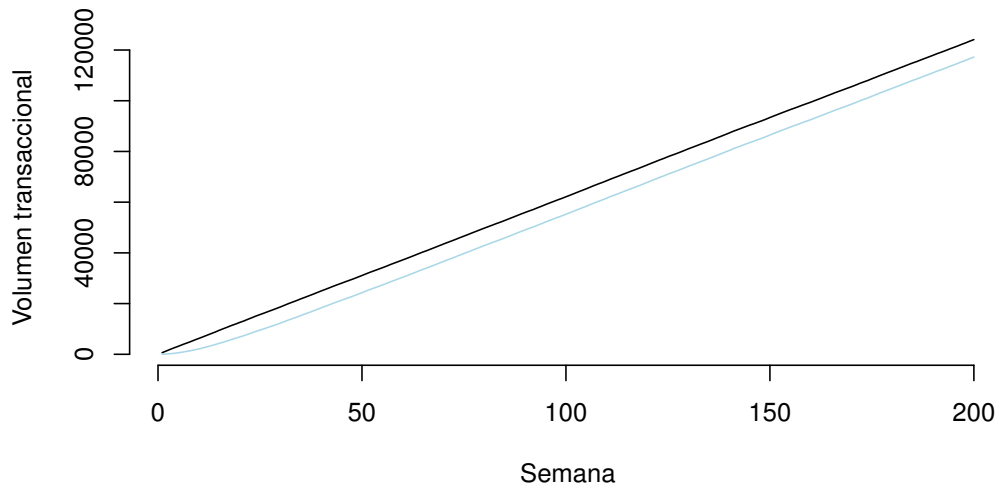


Figura 6.2: Tendencia acumulada de transacciones semanales

Nótese cómo la tendencia acumulada presenta un comportamiento lineal, pero en el transcurso del período, el volumen de transacciones crece con una tasa constante, evidenciado también por la Figura 5.7. La información más valiosa que presentan las figuras anteriores es que no existe una distinción mayor entre las transacciones totales (línea negra) y las transacciones consecutivas (línea celeste), indicando que la cohorte de clientes no tiende a volverse inactiva.

Ahora bien, el primer paso para la implementación del modelo es la estimación de parámetros para las distribuciones, la cual resultó en lo siguiente:

Modelo	Parámetros (r, α, a, b)
BG/NBD	(36.42756,402.66592,0.00045,897.86299)
BG/CNBD-k	(26.42744,402.66465,0.00045,898.09683)
MBG/CNBD	(37.33484,412.43904,0.00797,72.77565)
MBG/CNBD-k	(37.33484,412.43904,0.00797,72.77565)

Tabla 6.1: Estimación de parámetros para los modelos a implementar

La estimación realizada busca maximizar la log-verosimilitud, con lo cual se encontró que para el período de calibración indicado, el modelo MBG/CNBD maximiza el valor

Modelo	$LL(r, \alpha, a, b)$
BG/NBD	-289697.8
BG/CNBD-k	-289697.8
MBG/CNBD	-289697.2
MBG/CNBD-k	-289697.2

Tabla 6.2: Cálculo de $LL(r, \alpha, a, b)$ para el período de calibración

La estimación de parámetros permite hacer ciertas afirmaciones con respecto a la cohorte de clientes, dado que la media de distribución gamma es $\frac{r}{\alpha}$, la estimación para λ y, por lo tanto, la media del tiempo inter-transaccional en semanas $\frac{1}{\lambda}$

Modelo	Estimación de $\frac{1}{\lambda}$
BG/NBD	0
BG/CNBD-k	11.053
MBG/CNBD	11.047
MBG/CNBD-k	11.047

Tabla 6.3: Estimación del tiempo promedio inter-transaccional en semanas para la cohorte de clientes

Nótese que el modelo BG/NBD estima que no hay tiempo entre transacciones dado que este asume el tiempo inter-transaccional promedio es λt , distribuido de manera exponencial. Sabiendo la estimación de tiempo para la frecuencia de transacciones de los clientes, se procede a pronosticar el número de esperado de transacciones para el período de calibración a través de cómo se modeló en la ecuación 4.56 y partiendo del hecho que se observan 32126 transacciones en el período de observación se obtuvo lo siguiente:

Modelo	Transacciones pronosticadas	Error relativo	Error absoluto medio
BG/NBD	32407	0.87 %	1.782
BG/CNBD-k	32407	0.87 %	1.782
MBG/CNBD	32380	0.79 %	1.783
MBG/CNBD-k	32380	0.79 %	1.783

Tabla 6.4: Cálculo de $E(Y(t)|X = x, t_x, r, \alpha, a, b)$ y medidas de error

Es de notar la alta precisión de los pronósticos independientemente del modelo implementado, esto es en virtud de los tiempos entre transacciones que presenta la cohorte de clientes así como sus patrones de consumo, los patrones de consumo para una muestra aleatoria de $n = 50$ clientes se visualiza de la siguiente manera:

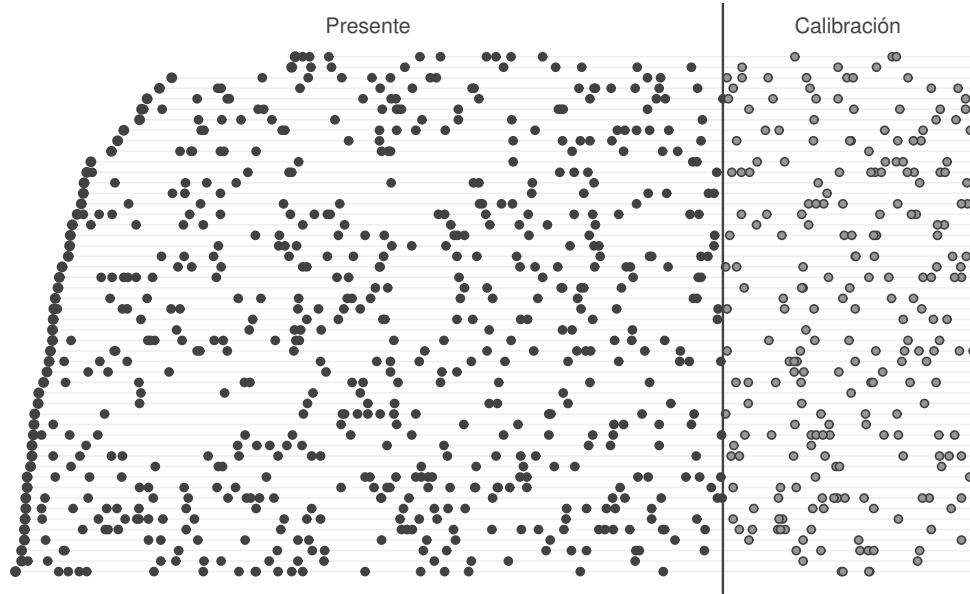


Figura 6.3: Patrones de consumo para $n = 50$ clientes seleccionados aleatoriamente

Asimismo, el pronóstico de transacciones y los patrones de consumo de la cohorte lleva a la predicción de que un cliente permanezca activo luego de cierta cantidad de tiempo, según la ecuación 4.50, la probabilidad de que un cliente permanezca activo en el período de calibración es la siguiente:

Modelo	Mínimo	Máximo	Media	Mediana	Desviación estándar
BG/NBD	0.99721	1.0	0.99999	0.99999	0.00005
BG/CNBD-k	0.99721	1.0	0.99999	0.99999	0.00005
MBG/CNBD	0.18806	0.99992	0.99876	0.99998	0.01485
MBG/CNBD-k	0.18806	0.99992	0.99876	0.99998	0.01485

Tabla 6.5: Medidas de tendencia central para $P(X(t) = x|\lambda, p)$ para el período de calibración

Cabe recalcar que aunque los modelos MBG/CNBD y MBG/CNBD-k pronostican clientes con una probabilidad baja de permanecer activos, la probabilidad media y su desviación estándar permanecen consistentes a lo largo de los cuatro modelos. A su vez, estos modelos identificaron únicamente 18 clientes en toda la cohorte con una probabilidad de mantenerse activos menor o igual a 0.90, donde estos casos se encuentran en un perfil de cliente con 6 o menos transacciones en el período de calibración y un tiempo inter-transaccional de 4.7 semanas.

Por último, se tiene el cálculo del pronóstico del LTV de la cohorte de clientes, para ello es necesario calcular 4.31 el cual a su vez depende 4.43 y 4.44. Definiendo la tasa de descuento como $d = 0.02$ se obtuvo lo siguiente:

Modelo	Mínimo	Máximo	Media	Mediana	Desviación estándar
BG/NBD	0	21841.10	5938.320	5510.554	3416.952
BG/CNBD-k	0	21841.09	5938.320	5510.554	3416.951
MBG/CNBD	0	21790.76	5933.137	5503.938	3413.873
MBG/CNBD-k	0	21790.76	5933.137	5503.938	3413.873

Tabla 6.6: Medidas de tendencia central para el LTV de la cohorte de clientes

Ajustando para la probabilidad de que un cliente permanezca activo se tiene:

Modelo	Mínimo	Máximo	Media	Mediana	Desviación estándar
BG/NBD	0	21841.06	5938.292	5510.550	3416.943
BG/CNBD-k	0	21841.06	5938.291	5510.550	3416.943
MBG/CNBD	0	21784.26	5928.126	5498.578	3413.019
MBG/CNBD-k	0	21784.26	5928.126	5498.578	3413.019

Tabla 6.7: Medidas de tendencia central para el LTV de la cohorte de clientes ajustadas a probabilidad de supervivencia

Las diferencias entre la distribución de los valores es mínima al ajustar a la probabilidad de supervivencia dado que la gran mayoría de clientes presenta patrones de consumo constantes, tal y como se mostró en la Figura 6.1 y en la Tabla 6.5.

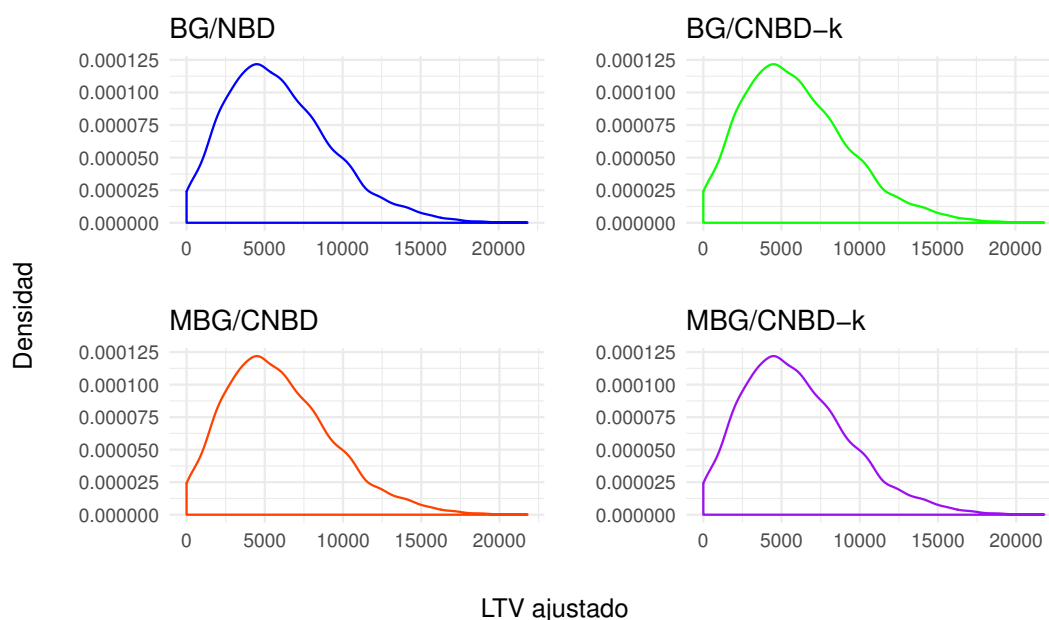


Figura 6.4: Distribuciones de LTV ajustado por modelo

La distribución del LTV a través de la cohorte de clientes indica que la mayoría del valor del capital de la compañía está dado por un grupo selecto de clientes, mientras que el resto de la cohorte aporta en menor manera, aunque componen la mayoría del volumen del capital. Se tiene que únicamente 333 de los 6889 clientes (i.e., 4.83 %) se encuentran dos desviaciones estándar a la derecha en la curva de LTV, componiendo un 11.70 % del total del valor del capital de la empresa.

Tomando únicamente el período de calibración, este subconjunto de la cohorte de clientes realiza en promedio 9 transacciones con montos rondando en el intervalo de \$2557.87 a \$3286.13. Comparando contra el resto de la base (ver Tabla 5.2 y Tabla 5.3), estos clientes realizan la mitad de transacciones que el cliente promedio, pero los montos por transacción rondan en el rango de 9.6 a 12.31 veces la media. Por otro lado, existen 808 (11.73 %) clientes que se encuentran una desviación estándar a la izquierda en la curva de LTV, estos componen únicamente un 2.3 % del total del capital de la compañía. Aunque el monto promedio por transacción de este conjunto de clientes gira alrededor de 0.39 y 1.56 veces la media, su promedio de transacciones para el período de calibración es únicamente de 1.55. El 83.4 % restante conforman el 85.9 % del capital y tienen patrones de consumo que constan de 4.84 transacciones y montos en el rango de \$774.32 y \$1814.94 (2.89 a 6.79 veces la media).

CAPÍTULO 7

Conclusiones

- Se cumple con el objetivo principal, que es implementar una metodología para el análisis del comportamiento de la base clientes por medio de un modelo estocástico.
- Se perfiló el comportamiento de los clientes en un entorno no contractual según el pronóstico de transacciones, describiendo los patrones de consumo de la cohorte de clientes y su aporte al valor de capital de la compañía.
- Se determinó una forma de derivar el valor del tiempo de vida útil y el valor de capital de marca de una base de cliente por medio de modelos estocásticos.
- Se encontró que el tiempo inter-transaccional es un factor influyente en la determinación de la retención de un cliente.

Recomendaciones

- Para perfilar de manera más profunda a la base de clientes, se recomienda complementar con información demográfica.
- En base a los resultados experimentales, se recomienda implementar el modelo con una base de datos distinta con más heterogeneidad en las transacciones así como mayor cantidad de observaciones e historial de transacciones.
- Para profundizar acerca del rendimiento del modelo, se recomienda compararlo contra otros modelos que se especialicen en el cálculo del valor de tiempo de vida útil de la base de datos, en lugar de modelos de regresión o similares.

- [1] Cox, D.R.: *Statistical Analysis Of Time Series: Some Recent Developments*. Scadinavian Journal Of Statistics, Vol. 8, No. 2, páginas 93–108, 1981.
- [2] CRAN, The Comprehensive R Archive Network: *Probabilistic Models for Assessing and Predicting your Customer Base*. <https://cran.r-project.org/web/packages/BTYDplus/BTYDplus.pdf>, 2016. Último acceso el 21 de octubre de 2019.
- [3] CRAN, The Comprehensive R Archive Network: *Rmetrics - Financial Time Series Objects*. <https://cran.r-project.org/web/packages/timeSeries/timeSeries.pdf>, 2017. Último acceso el 20 de octubre de 2019.
- [4] Fader, Peter S., Bruce G. S. Hardie Subrata Sen.: *Stochastic models of buyer behavior*. World Scientific Publishing Co. Pte. Ltd., páginas 165–205, 2015.
- [5] Fader, Peter S., Bruce G.S. Hardie Ka Lok Lee: *“Counting Your Customers” the Easy Way: An Alternative to the Pareto/NBD Model*. Marketing Science Vol. 24, No. 2, páginas 275–284, 2005.
- [6] Hsu, Hwei P.: *Probability, Random Variables And Random Processes*. McGraw-Hill, 1997.
- [7] Kaggle: *Retail Transaction Data*. https://www.kaggle.com/regivm/retailtransactiondata#Retail_Data_Transactions.csv, 2018. Último acceso el 20 de octubre de 2019.
- [8] Malthouse, E. y Blattberg, R.: *Can we predict customer lifetime value?* Journal Of Interactive Marketing Vol. 19, No. 1, páginas 1–16, 2005.
- [9] Robert Blattberg, Byung Do Kim y Scott A. Nelsin: *Database Marketing. Analyzing and Managing Customers*. Springer Science+Business Media, 2008.
- [10] Wasserman, Larry: *All Of Statistics - A Concise Course In Statistical Inference*. Springer-Verlag, 2004.

Derivación de funciones especiales

Una función esencial para la derivación de las funciones $E[X(t)]$ y de $E[Y(t)|X = x, t_x, T]$ es la función hipergeométrica gaussiana, definida por:

$${}_2F_1(a, b; c; z) = \frac{1}{B(b, c-b)} \int_0^1 t^{b-1} (1-t)^{c-b-1} (1-zt)^{-a} dt, \quad c > b \quad (1)$$

Con el fin de obtener una expresión para $E[X(t)]$ para un consumidor seleccionado aleatoriamente, es necesario tomar el valor esperado de 4.51 sobre la distribución de λ y p . En primera instancia, se toma el valor esperado con respecto a λ , resultando en:

$$E[X(t)|r, \alpha, p] = \frac{1}{p} - \frac{\alpha^r}{p(\alpha + pt)^r}.$$

El siguiente paso es tomar el valor esperado sobre la distribución de p , tomando la integral respectiva:

$$\int_0^1 \frac{1}{p} \frac{p^{a-1} (1-p)^{b-1}}{B(a, b)} dp = \frac{a+b-1}{a-1}.$$

Luego, evaluamos:

$$\begin{aligned} \int_0^1 \frac{\alpha^r}{p(\alpha + pt)^r} \frac{p^{a-1} (1-p)^{b-1}}{B(a, b)} dp \\ = \alpha^r \frac{1}{B(a, b)} \int_0^1 p^{a-2} (1-p)^{b-1} (\alpha + pt)^{-r} dp \end{aligned}$$

sea $q = 1 - p$, por lo que se tiene:

$$\left(\frac{\alpha}{\alpha + t} \right)^r \frac{1}{B(a, b)} \int_0^1 q^{b-1} (1-q)^{a-2} \left(1 - \frac{t}{\alpha + t} q \right)^{-r} dq$$

Sustituyendo 1 en la integral anterior se tiene entonces:

$$E[X(t)|r, \alpha, a, b] = \frac{a+b-1}{a-1} \left[1 - \left(\frac{\alpha}{\alpha + t} \right)^r {}_2F_1 \left(r, b; a+b-1; \frac{t}{\alpha + t} \right) \right]. \quad (2)$$

Sea $Y(t)$ la variable aleatoria que denota el número de transacciones realizadas en el período $(T, T + t]$. Dado esto, nos es de interés la esperanza condicional del número de transacciones en el período $(T, T + t]$ con un cliente con historia transaccional $X = x, t_x, T$. Si el cliente se encuentra activo en el período T , de la ecuación 4.51 se puede obtener que:

$$E[Y(t)|\lambda, p] = \frac{1}{p} - \frac{1}{p}e^{-\lambda pt}. \quad (3)$$

En virtud del supuesto que todos los clientes están activos en el momento inicial del período de observación, un cliente no puede retirarse antes de que este haya realizado alguna transacción, por lo tanto:

$$P(\text{activo en } T|X = 0, T, \lambda, p) = 1.$$

En el caso que las transacciones se realizaron dentro del período $(0, T]$, la probabilidad que un cliente con historial $(X = x, t_x, T)$ esté activo en T , condicional en λ y p es simplemente la probabilidad que no se haya retirado en t_x y no haya realizado transacciones en el período $(t_x, T]$, dividido por la probabilidad que no haya realizado transacciones en dicho período. A partir de esto, se tiene entonces:

$$P(\text{activo en } T|X = x, t_x, T, \lambda, p) = \frac{(1-p)e^{-\lambda(T-t_x)}}{p + (1-p)e^{-\lambda(T-t_x)}}.$$

Al multiplicar la expresión anterior por $\frac{(1-p)^{x-1}\lambda^x e^{-\lambda t_x}}{(1-p)^{x-1}\lambda^x e^{-\lambda t_x}}$ se tiene lo siguiente:

$$P(\text{activo en } T|X = x, t_x, T, \lambda, p) = \frac{(1-p)^x \lambda^x e^{-\lambda T}}{L(\lambda, p|X = x, t_x, T)} \quad (4)$$

donde la expresión para $L(\lambda, p|X = x, t_x, T)$ está dada en 4.49. Note además que si $x = 0$, entonces 4 es igual a 1.

Multiplicando 3 con 4 resulta en:

$$E[Y(t)|X = x, t_x, T, \lambda, p] = \frac{p^{-1}(1-p)^x \lambda^x e^{-\lambda T} - p^{-1}(1-p)^x \lambda^x e^{-\lambda(T+pt)}}{L(\lambda, p|X = x, t_x, T)}. \quad (5)$$

Nótese que cuando $x = 0$, esta expresión se reduce a 3, lo cual es derivado del resultado que un cliente debe asumirse activo en el período T incluso si este no realizó transacciones en el período $(0, T]$.

Conforme la tasa de transacciones λ y la probabilidad de abandono p sean no observadas, el valor de $E[Y(t)|X = x, t_x, T]$ para un cliente seleccionado aleatoriamente se computa tomando el valor esperado de 5 sobre la distribución de λ y p tomando en cuenta la historia transaccional dada:

$$E[Y(t)|X = x, t_x, T, r, \alpha, a, b] = \int_0^1 \int_0^\infty (E[Y(t)|X = x, t_x, T, \lambda, p])(f(\lambda, p|r, \alpha, a, b, X = x, t_x, T)) d\lambda dp. \quad (6)$$

Por el teorema de Bayes, la distribución de λ y p está dada por:

$$f(\lambda, p|r, \alpha, a, b, X = x, t_x, T) = \frac{L(\lambda, p|X = x, t_x, T)f(\lambda|r, \alpha)f(p|a, b)}{L(r, \alpha, a, b|X = x, t_x, T)}. \quad (7)$$

Sustituyendo 5 y 7 en 6 se tiene que:

$$E[Y(t)|X = x, t_x, T, r, \alpha, a, b] = \frac{A - B}{L(r, \alpha, a, b|X = x, t_x, T)} \quad (8)$$

donde:

$$A = \int_0^1 \int_0^\infty p^{-1}(1-p)^x \lambda^x e^{-\lambda T} f(\lambda|r, \alpha)f(p|a, b) d\lambda dp = \frac{B(a-1, b+x)}{B(a, b)} \frac{\Gamma(r+x)\alpha^r}{\Gamma(r)(\alpha+T)^{r+x}} \quad (9)$$

y:

$$\begin{aligned}
B &= \int_0^1 \int_0^\infty p^{-1}(1-p)^x \lambda^x e^{-\lambda(T+pt)} f(\lambda|r, \alpha) f(p|a, b) \, d\lambda dp \\
&= \frac{\Gamma(r+x)\alpha^r}{\Gamma(r)B(a, b)} \int_0^1 p^{a-2}(1-p)^{b+x-1} (\alpha + T + pt)^{-(r+x)} \, dp \quad (10)
\end{aligned}$$

si $q = 1 - p$, entonces:

$$\begin{aligned}
&= \frac{\Gamma(r+x)\alpha^r}{\Gamma(r)B(a, b)(\alpha + T + t)^{r+x}} \int_0^1 q^{b+x-1}(1-q)^{a-2} \left(1 - \frac{t}{\alpha + T + t}q\right)^{-(r+x)} \, dq \\
&= \frac{B(a-1, b+x)}{B(a, b)} \frac{\Gamma(r+x)\alpha^r}{\Gamma(r)(\alpha + T + t)^{r+x}} {}_2F_1\left(r+x, b+x; a+b+x-1; \frac{t}{\alpha + T + t}\right) [5]. \quad (11)
\end{aligned}$$