

**UNIVERSIDAD NACIONAL DE INGENIERÍA  
FACULTAD DE INGENIERÍA ECONÓMICA Y CCSS  
Escuela Profesional de Ingeniería Estadística**



**ESTIMACIÓN DEL TIEMPO DE VIDA COMERCIAL DE LOS CLIENTES DE  
UN NEGOCIO: UNA APLICACIÓN DE UN MODELO ERLANG**

**INFORME DE SUFICIENCIA  
PARA OPTAR EL TÍTULO PROFESIONAL DE  
INGENIERO ESTADÍSTICO**

**Por**

**Bach. Hugo Antonio Solano Martínez**

**LIMA - PERÚ  
2011**

**Composición de la Comisión de Sustentación:**

**Prof. Mg. Alipio Ordoñez Mercado**  
**Escuela Profesional de Ingeniería Estadística, Presidente de Comisión**

**Prof. Mg Magen Infante Rojas**  
**Escuela Profesional de Ingeniería Estadística, Asesor**

**Prof. Lic. José Camero Jimenez**  
**Escuela Profesional de Ingeniería Estadística, Revisor**

*Dedicado a  
mi esposa,  
mis padres,  
mi familia,  
amigos,  
amigas  
y profesores*

## RESUMEN

La presente investigación tiene como objetivo la estimación del tiempo de vida comercial de los clientes de un negocio tipo *retail* mediante la implementación de un modelo basado en la distribución de probabilidad Erlang-k. Por ello, se plantea la hipótesis que a través del número de transacciones históricas de cada cliente, se puede estimar su tiempo de vida comercial con el negocio en un periodo de tiempo. Para alcanzar los objetivos, se dispuso de un grupo de clientes de un negocio del sector comercial con sus transacciones históricas agrupadas para las fases de modelado y validación y se realizaron ensayos en diferentes escenarios del espacio parametral. Se estimó la probabilidad de actividad de los clientes y el número de transacciones futuras. Se concluyó que el modelo *Erlang-4* satisface las pruebas de bondad de ajuste y explica favorablemente tanto las transacciones históricas como las estimadas para la estimación del tiempo de vida comercial. Finalmente, la estimación del tiempo de vida comercial de clientes tiene su principal aplicación en el estudio del comportamiento del consumidor, al permitir realizar segmentaciones con la probabilidad de actividad, valor futuro y tiempo de vida comercial, y con ello invertir esfuerzos y dinero en los clientes “más leales”.

Palabras Clave.- Lealtad del cliente, Procesos Estocásticos, Comportamiento del Consumidor, Valor del Cliente.

## ÍNDICE

	Pag.
INTRODUCCIÓN.....	1
OBJETIVO.....	5
JUSTIFICACIÓN.....	6
CAPITULO I La deserción del cliente.....	7
CAPITULO II El problema de investigación.....	26
2.1 Pregunta Principal.....	27
2.2 Preguntas Específicas.....	27
CAPITULO III Modelado estadístico: Distribución Erlang-k.....	28
3.1 Suposiciones.....	28
3.2 La distribución Erlang-k.....	29
3.3 Verosimilitud Individual.....	30
3.4 Verosimilitud Agregada.....	32
3.5 Estimación de Parámetros.....	33
3.6 Distribución de probabilidad de la Frecuencia de Compra.....	34
3.7 Probabilidad de Actividad.....	35
3.8 Número Estimado de Transacciones.....	38
3.8.1 Transacciones Estimadas No Condicionales para CBG/CNBD-k.....	38
3.8.2 Transacciones Estimadas Condicionales para CBG/CNBD-k.....	38
CAPITULO IV Diseño de la investigación y metodología.....	40
4.1 Presentación de Metodología.....	40
4.2 Obtención de datos.....	41
4.3 Tratamiento de datos y estimación de parámetros.....	41
4.4 Aplicación de modelo.....	41
4.5 Aplicación de resultados.....	42
4.6 Especificación de datos.....	42
CAPITULO V Análisis de resultados.....	43
5.1 Análisis Descriptivo.....	44
5.2 Modelado.....	48
CONCLUSIONES.....	57
RECOMENDACIONES.....	59
BIBLIOGRAFIA.....	60
ANEXOS.....	63
Anexo Estadístico.....	63

## ÍNDICE DE CUADROS

		Pag.
Cuadro N° 5.1	Variables en análisis descriptivo.....	43
Cuadro N° 5.2	Distribución de Frecuencia de los Clientes de Retail según Sexo.....	44
Cuadro N° 5.3	Distribución de Frecuencia de los Clientes de Retail según Rango de Edad.....	44
Cuadro N° 5.4	Distribución de Frecuencia de los Clientes de Retail según Departamento de Residencia.....	45
Cuadro N° 5.5	Distribución de Frecuencia de los Clientes de Retail según Tipo de Crédito.....	45
Cuadro N° 5.6	Distribución de Frecuencia de los Clientes de Retail según Línea de Crédito.....	46
Cuadro N° 5.7	Distribución de Frecuencia de los Clientes de Retail según Montos de Compra en el periodo de Modelado.....	47
Cuadro N° 5.8	Valores para los parámetros $(r, \alpha, s, \beta)$ según modelo Erlang-k propuesto.....	49
Cuadro N° 5.9	Valores del estadístico $\chi^2$ según modelo Erlang-k propuesto.....	51
Cuadro N° 5.10	Probabilidades de actividad en el siguiente periodo para modelo Erlang-4.....	53
Cuadro N° 5.11	Distribución de frecuencia de los clientes de retail según montos de compra en el periodo de validación..	53
Cuadro N° 5.12	Distribución de clientes y promedio de compra según grupo de fidelización en el periodo de validación.....	55

## ÍNDICE DE GRÁFICOS

		Pag.
Gráfico N° 5.1	Número de clientes por cantidad de transacciones en periodo de modelado.....	47
Gráfico N° 5.2	Gráfico N° 5.2 Evolución de transacciones en el periodo de modelado y validación.....	48
Gráfico N° 5.3	Distribución de Clientes por Rango de Probabilidad de Actividad Estimada de los clientes (Erlang-4).....	50
Gráfico N° 5.4	Distribución de transacciones acumuladas en el periodo de validación para cada modelo Erlang-k comparado con distribución real.....	51
Gráfico N° 5.5	Distribución de cantidad de clientes por número de transacciones reales y estimadas en periodo de validación para Erlang-4.....	52
Gráfico N° 5.6	Relación entre probabilidad de actividad y monto de compra total estimado del cliente.....	54
Gráfico N° 5.7	Relación entre probabilidad de actividad y línea de crédito asignada al cliente.....	56
Gráfico N° 5.8	Relación entre probabilidad de actividad y edad del cliente.....	56

## ÍNDICE DE FIGURAS

	Pag.
Figura N° 3.1	
Tiempos inter-eventos distribuidos como Erlang-3. $P_3(X(t)=2)$ es la probabilidad de encontrar 6, 7 u 8 eventos de Poisson.....	30



## **INTRODUCCIÓN**

En la globalización y nivel de competencia que vive la industria actualmente, para una empresa de cualquier sector, es de suma importancia ya no preocuparse solamente de su cadena de producción, sino que debe de estar atento al responsable directo de su éxito: el cliente. Por ello, es importante conocer cuánto tiempo pasará hasta que el cliente deje de comprarle, ya sea por la captación por parte de la competencia o por falta de atención de lado de la empresa. En cualquiera de los casos, es inevitable pensar cuánto tiempo estará fidelizado el cliente. Por este motivo se convierte en una necesidad estudiar la fuga de clientes en un negocio. El resultado de este estudio contribuirá en concentrar las inversiones y campañas de marketing al grupo más propenso a continuar como cliente y más importante, dirigirse al más rentable.

Como se conoce por la propia experiencia como cliente de alguna empresa, la relación comercial puede haberse originado a partir de un contrato con términos de inicio y fin. Pero por otro lado están las relaciones comerciales que sólo se conocen a partir de la firma de un contrato que no tiene fecha de caducidad. Es en estos casos dónde la gestión de la empresa se vuelve más complicada por no conocer de antemano quienes siguen siendo o no clientes, ya que puede darse el caso de una absorción por parte de la competencia en cualquier momento. Son en estos casos, de relaciones contractuales sin término fijado, dónde se centra la presente investigación

La presente investigación que se propone tiene como objetivo estimar el tiempo de vida comercial de los clientes de un negocio a partir de la información histórica de compras. Esta vida comercial debe entenderse como el número de compras que puede realizar en un periodo de tiempo determinado, condicionada a una probabilidad de actividad en dicho periodo. Para alcanzar lo propuesto, se presenta esta investigación de tipo cuantitativa, descriptiva y aplicada, prospectiva y de corte transversal que permitirá obtener el modelo que estime el tiempo de vida comercial de los clientes.

Esta investigación inicia con la exploración de la bibliografía relacionada con conceptos que van desde definiciones de *CRM (Customer Relationship Management)*, Lealtad de Cliente hasta el de RFM (Recencia, Frecuencia, Moneda), *Lifetime Value (Valor del Cliente)*, y modelos relacionados con supervivencia, por tratarse de eventos en los cuales se desea saber si un sujeto logrará llegar a un nuevo estado, siempre basado en una distribución de probabilidad. Y justamente hay un grupo especial de distribuciones relacionadas a este fenómeno, la cual es la familia de distribuciones exponenciales, ampliamente usadas en la medición de estos sucesos. Se conoce de un estudio de Ehrenberg (1959) quien propuso la distribución binomial negativa (NBD por sus siglas en inglés Negative Binomial Distribution) en 1959 como un modelo para la repetición de compra. Además también se tiene el modelo más conocido de todos, el modelo de Pareto/NBD (Schmittlein et. al. 1987) y dos de sus variantes, el BG/NBD (Fader et. al., 2005) y el CBG/NBD (Hoppe et.al. ,2007), los cuales son extensiones del modelo NBD pero haciendo suposiciones adicionales respecto al proceso de deserción o fuga y la heterogeneidad entre los consumidor. Finalmente el modelo CBG/CNBD-k, el cual es una nueva variante del modelo CBG/NBD. Este modelo hace diferentes suposiciones respecto a los tiempos de compra, ya que en particular considera en cierta medida la regularidad y por tanto mejorará la calidad de pronóstico considerablemente.

La metodología propone un modelo que incluye una distribución de probabilidad de supervivencia, del tipo exponencial, para estimar mediante el método de estimadores de máxima verosimilitud, los parámetros asociados al número de compras específicamente. La distribución de probabilidad a utilizar es la distribución Erlang. A diferencia de otras distribuciones propuestas para esta clase de modelado, ésta es más cómoda para trabajar que una Log-Normal o Weibull (Platzer, 2008). Por otra parte, la distribución Erlang tiene ventaja sobre la exponencial, debido a que carece de la propiedad "sin memoria". Esto quiere decir que a la distribución Erlang si le interesa el pasado y a partir de él calcula una nueva transacción. La metodología en sí no es limitada por el sólo hecho de calcular una probabilidad de actividad, sino que utilizando está información, predice el tiempo de vida de los clientes en base a sus compras pasadas de un determinado periodo, traducido en una estimación del número de compras futuras para un periodo determinado.

Se estima el modelo Erlang-4 el cual llega a tener el mejor estadístico en la prueba de bondad de ajuste a comparación de los otros modelos con distribución Erlang propuestos. Además se calcularon de manera individual las probabilidades de los clientes de continuar activos en el siguiente periodo. Por otro lado se identificó un grupo de clientes con altos niveles de compras que tienen una probabilidad de menos del 0.20 de continuar activos con el negocio por lo que se deduce que fueron de alguna manera absorbidos por la competencia o dejaron simplemente el negocio. Una relación a tomar en cuenta por el negocio es el nivel de endeudamiento permitido, ya que se encontró una relación directamente proporcional entre las líneas de crédito y las probabilidades de actividad.

La presente investigación está organizada de la siguiente manera: en el capítulo I se hace una reseña del estado del arte. En el capítulo II se describe el problema de la investigación, la pregunta principal y preguntas específicas a responder al alcanzar el objetivo de la investigación. En el capítulo III se describe y detalla todo el marco teórico utilizado para elaborar el modelo de probabilidad, las suposiciones hechas y la teoría

usada para la estimación de parámetros, estimación de probabilidad de actividad y finalmente la estimación de transacciones futuras. En el capítulo IV se describe la metodología usada desde la especificación y obtención de datos, tratamiento, transformación, estimación de parámetros hasta la aplicación del modelo y de los resultados. Luego se presenta el capítulo V de resultados dónde se plasma la aplicación propia del modelo, las relaciones encontradas entre las variables disponibles de los clientes y las obtenidas como la probabilidad de actividad y transacciones futuras estimadas. Finalmente se detallan las conclusiones alcanzadas con la investigación y sus resultados; y posteriormente las recomendaciones que enumeran las mejoras y limitaciones que se encontraron en la elaboración de esta investigación.

## **OBJETIVO**

### **OBJETIVO GENERAL**

Estimar el tiempo de vida comercial de los clientes de un negocio a través del número de transacciones históricas de cada cliente.

### **OBJETIVOS ESPECÍFICOS**

- 1.- Estimar el número de transacciones por cliente para un periodo determinado de tiempo.
- 2.- Estimar la probabilidad de actividad para un periodo determinado de tiempo.
- 3.- Estimar montos de compras futuros en base al número de transacciones estimadas para un periodo determinado de tiempo

## JUSTIFICACIÓN

Para las empresas es de imperiosa necesidad saber hasta cuando pueden contar con un cliente, es decir, cuánto tiempo pasará hasta que el cliente deje de comprarle, ya sea por la captación por parte de la competencia o por falta de atención de lado de la empresa. En cualquiera de los casos, es inevitable pensar cuánto tiempo estará fidelizado el cliente. Por este motivo se convierte en una necesidad estudiar la fuga de clientes en un negocio. Por medio de la cuantificación de este fenómeno, las empresas enfocarán sus recursos al conjunto de los clientes con mayor vida comercial, y de ellos los que entreguen a la empresa mayores réditos. Las empresas se verán beneficiadas ya que conocerán a priori, cuáles son y serán los clientes que tendrán mayor tiempo fidelizados y concentrarán sus esfuerzos y sus campañas de marketing al grupo más importante de ellos y obtener el beneficio económico que traería ese grupo de clientes.

De las múltiples técnicas o modelos estudiados, se resolvió en aplicar el modelo que considera la distribución Erlang-k como la ideal para predecir las futuras transacciones de los clientes. Por otro lado, tener a la mano una variable como la probabilidad que un cliente se mantenga en estado de actividad sirve como herramientas de gestión, dentro de las muchas que hay, para elaborar estrategias de marketing más precisas y personalizadas en el mundo competitivo de hoy.

# **CAPITULO I**

## **La deserción del cliente**

Las organizaciones comerciales en el mundo actual se tienen que mover al ritmo que marca la demanda, y el responsable directo de este fenómeno es el cliente. Es así, que desde la década de 1980, un nuevo término dentro del Marketing es integrado en las estrategias de las empresas. El marketing tradicional y la teoría de las cuatro "P" (Producto, Plaza, Promoción, Precio) se volvieron de alguna manera insuficientes o irrelevantes para poder explicar el marketing que se vive desde hace ya 30 años.

El marketing relacional, es acuñado para tener la capacidad de establecer una nueva conexión entre el producto y el cliente, entre la oferta y la demanda. Este término surge de la mano del profesor Leonard Berry en la Universidad de Texas a principios de los años ochenta.

Según Abad (2011) el marketing relacional es una metodología para mantener las relaciones de las empresas con sus clientes, buscando lograr los máximos ingresos por cliente. Su objetivo es identificar a los clientes más rentables para establecer una estrecha relación con ellos, que permita conocer sus necesidades y mantener una evolución del producto de acuerdo con ellas a lo largo del tiempo. El marketing relacional es un proceso que gestiona los recursos de la empresa para crear la mejor experiencia posible y el máximo valor al cliente.

Para esta especialidad del marketing, la relación es más que la suma de transacciones, se ha vuelto un vínculo entre la empresa y el cliente. Este vínculo tiene 2 pilares: la información del cliente lo más precisa, fidedigna y adecuada posible por un lado; por otro, la comunicación bidireccional, frecuente, continuada e interactiva.

Dentro de los instrumentos de apoyo para la implementación de estrategias desarrolladas del marketing relacional está el CRM, mostrada como herramienta tecnológica, que ayuda en la gestión de las relaciones con los clientes. Los principales beneficios que pueden cuantificar las empresas de la implementación de estrategias de marketing relacional son fundamentalmente: la retención de los clientes y la optimización de las acciones del marketing (Abad, 2011).

Estudios como el de Reicheld y Sasser (1990) han demostrado la pérdida de beneficios que tenían las empresas en el momento en el que descendía su tasa de retención de clientes. Tan sólo un incremento de 5 puntos en la tasa de retención de los clientes, puede repercutir directamente en un aumento de entre un 25% hasta un 100% o más en los beneficios (Abad, 2011).

Por otra parte, a través del marketing relacional, obtendremos un marketing más efectivo, se pueden crear oportunidades de venta cruzada. Además existen muchos otros beneficios como: adquisición de nuevos clientes, incremento de la facturación por una mayor rotación, estabilización de la demanda, mayor probabilidad de compras futuras, mejora de la imagen en el mercado, optimización del inventario, mejora en el proceso productivo, no competir en precios, referencias de terceros, diferenciación del producto, etc. (Abad, 2011).

Una manera de cuantificar el éxito o fracaso del despliegue de las estrategias de marketing relacional es tener la capacidad de medir la lealtad del cliente. Se comenta la relación intrínseca que tiene la rentabilidad de una empresa con la lealtad de sus consumidores. Se



puede afirmar que a mayor lealtad, mayor rentabilidad (Wusst, 2002). Se deben resaltar dos cosas: el logro de la lealtad de los clientes es un objetivo crítico para cualquier organización y la periódica medición de la lealtad de los clientes es un asunto obligado: lo que no se mide, no se conoce, por lo que no se puede controlar y por ende mejorar.

Los factores claves que influyen en la formación de “clientes leales” (Bowen et al., 2006) más importantes y que pueden ser manejados por la organización son:

1. El carácter: Son las cualidades personales que los clientes buscan en el personal de alguna organización.
2. La química: Es la capacidad de los empleados de estar sincronizados o conectados con sus clientes, es decir, cuando el personal comprende y conoce perfectamente las exigencias y necesidades del consumidor.
3. La preocupación: Es un interés genuino acerca del bienestar de los clientes, es decir, el empleado o la organización debe mirar siempre a sus clientes como personas, debe encontrar la forma de solucionar sus problemas y no como una forma más para ganar dinero.
4. La capacidad: Los empleados deben mostrarle y comunicarle indirectamente a sus clientes, sus capacidades.
5. La eficacia de costos: Implica a que clientes deben recibir el valor verdadero por el costo de los servicios y productos.
6. Consultivo: Es el factor más importante, ya que enmarca la relación de cliente-empresa como una sociedad en curso a través del tiempo, es decir, inicia una relación más cercana entre estas dos entidades.

Para investigar el fenómeno de abandono de los clientes de un negocio o empresa, es necesario partir de la definición de la “lealtad del cliente” y “satisfacción”. Se cree que “lealtad” y “satisfacción” tienen el mismo significado. Años atrás, Rodríguez et. al. (2001), señalaban que la

lealtad comienza al momento que el cliente está satisfecho con la compra de un bien o la prestación de un determinado servicio; pero definitivamente estos dos conceptos tiene una gran diferencia: la satisfacción del cliente constituye la antesala a la lealtad, es una forma en que los consumidores pueden llegar a ser leales con la empresa. De hecho, en condiciones de libre competencia un "Cliente Leal" siempre será un "Cliente Satisfecho", pero un "Cliente Satisfecho" no es necesariamente un "Cliente Leal". Por ejemplo, una persona que va a un supermercado y por la gran atención, se sintió satisfecha, es cierto que existe la posibilidad que vuelva a comprar en el mismo lugar, pero no significa que la empresa haya retenido a ese consumidor por siempre, se necesita más que una gran atención para que el consumidor sea leal o fiel hacia aquella empresa.

Una vez reconocida la lealtad del cliente hacia la empresa, ésta reconoce que sus clientes (no todos) pueden permanecer leales o no. Respecto a estos últimos, comenzaron a preocuparse por los motivos que llevan a esos clientes a disminuir el consumo regular que tenían o simplemente dejar de comprar. Esto motivó a organizarse y considerar estrategias para la recuperación de aquellos clientes. Los pasos que siguen las empresas (Begoña et. al., 2001) son:

1. Reconocimiento de la no lealtad del cliente.
2. Análisis de las razones de su abandono.
3. Reacción para solucionar los problemas que llevaron al cliente a dejar la marca.

Estos pasos son pautas para entender porque el cliente dejo de comprar, pero no permite conocer con certeza en qué momento puede pasar lo mismo con otros clientes.

Por otro lado, el análisis del valor de cliente es crítico para una buena estrategia de marketing y el Customer Relationship Management (CRM). Por ello, un componente importante de esta estrategia es la tasa

de retención del cliente (Mutanen, 2006). Dicha tasa tiene un alto impacto en el valor de cliente y entender el verdadero valor de una posible fuga de clientes o *churn* ayudará a la empresa en la administración su CRM. El concepto del valor del cliente es usualmente definido como el ingreso neto total del cliente a lo largo de la relación comercial. Este tipo de análisis del consumidor es hecho bajo diferentes términos: valor del consumidor, valor del cliente, acción del cliente y ganancias del consumidor. La idea central en el concepto de Valor de Vida ó *Lifetime Value* (LTV) es simple y medir el LTV es fácil después de que la relación con el cliente ha acabado. El reto de este concepto es definir y medir el LTV del cliente durante e incluso antes de la etapa activa de la relación con el cliente (Mutanen, 2006). Por ejemplo, Hoekstra et. al. (1999) define el modelo conceptual del LTV como el valor total de las contribuciones directas e indirectas a los gastos generales y ganancias de un cliente individual durante su ciclo entero de vida como cliente, esto es desde el inicio de la relación hasta su fin proyectado. La mayoría de los modelos LTV inician desde una ecuación básica. Sus componentes son:

1. El valor presente neto sobre el tiempo del cliente (ingresos y costos).
2. La tasa de retención o duración del servicio (DS).
3. Factor de descuento.

Los beneficios de entender el valor del cliente son numerosos. La compañía puede medir los ingresos presentes y futuros de los consumidores. La compañía puede también promover la retención del cliente y lealtad, lo cual llevará a una ganancia mayor por cliente. El análisis LTV puede también ayudar a la empresa en personalizar sus productos y servicios. Este entendimiento del valor del consumidor ayuda a la empresa a enfocarse en los clientes más productivos y tratar de otra manera al segmento con un posible impacto negativo a los ingresos. Y por último, el valor del cliente no es un valor fijo, este puede ser influenciado por las estrategias de marketing (Mutanen, 2006).

El enfoque de la fuga de clientes o *churn* sirve para determinar quienes están en riesgo de abandono y si es que vale la pena retenerlos. El análisis de fuga es altamente dependiente de la definición de la fuga de clientes. El sector del negocio y la relación con el cliente afecta los resultados de cómo son detectados los clientes fugados. Por ejemplo, en el negocio de las tarjetas de crédito, los clientes pueden fácilmente empezar a usar otra tarjeta de crédito, por tanto el único indicador para la empresa de la primera tarjeta es la disminución de las transacciones. La fuga de clientes está relacionada con la tasa de retención de clientes y lealtad. Hwang et. al. (2004) define la lealtad del cliente como el indicador del cliente que quisiera estar con la empresa. La fuga describe el número o porcentaje de clientes regulares que abandonan la relación con el proveedor del servicio. Se define la lealtad del cliente como:

$$\text{Lealtad del cliente} = 1 - \text{Tasa de Fuga}$$

Entonces se puede derivar la siguiente definición:

$$\text{Probabilidad de Actividad} = 1 - \text{Probabilidad de Fuga} \quad (1.1)$$

A medida que la información transaccional seguía en aumento, las empresas podían contar con más información a detalle de cada uno de sus clientes a través de sus programas de fidelización. Por medio de una *datawarehouse* ("centro de almacenamiento de datos") y medios informáticos que permitan acceder a la información almacenada, las empresas podían distinguir del universo de clientes, aquellos que puedan considerarse leales o fieles, por su cantidad de transacciones y compras. Para tomar en cuenta esa información de los clientes, una de las más prácticas metodologías, la segmentación RFM, considera que el mejor predictor del futuro es el comportamiento pasado y define:

1. Recencia: Tiempo desde la última compra.
2. Frecuencia: Tiempo entre compras.
3. Moneda: Monto de Compras.

Este tipo de segmentación es la base de cualquier modelo de predicción de comportamiento, porque se miden 3 dimensiones del comportamiento (Segovia et. al., 2005). Para proceder con la segmentación RFM se requiere fechas, montos de transacciones e identificar al cliente. Dentro de los puntos bajos encontramos la confiabilidad de los datos, disponibilidad de ellos y la existencia de gran cantidad de datos, por lo que es necesario grandes espacios virtuales para su almacenamiento. Por otro lado, una de las ventajas que presenta el RFM, está dirigido a la gestión de clientes, ya que balancea las tres dimensiones o características más potentes; asimismo es de fácil comprensión y permite focalizar las medidas de fidelización de clientes (Segovia et. al., 2005).

Dentro de otras metodologías que se ocupan del estudio de la fuga de clientes, las Cadenas de Markov, fueron utilizadas por Marker (1998) quién estudio políticas de retención de clientes en una compañía de seguros, modelando la dinámica de la situación con cadenas de Markov, en donde cada estado representaba el número de periodos que el cliente permanecía activo. Más adelante, Pfeifer et.al (2000), también modelaron mediante cadenas de Markov la relación de un cliente con una compañía de marketing, obteniéndose distintos valores de LifeTime Value para diferentes configuraciones RFM.

Segovia et. al. (Segovia et. al., 2005) caracterizaron el proceso de fuga de clientes, también modelando la dinámica de la situación a través de cadenas de Markov, con estados correspondientes a distintas clasificaciones RFM. Se utilizó la información transaccional, específicamente las variables RFM, de los años 2003 y 2004. Para el análisis se utilizó una parte de la base de datos para entrenar los modelos, para predecir con diferentes formas la matriz de probabilidades de transición y otra parte para las pruebas, que se utilizó para elegir el mejor tipo de clasificación RFM.

Otro método utilizado en el tratamiento de la fuga de clientes es la regresión logística. El caso de la regresión logística binaria o binomial es una forma de regresión la cual es usada en la situación cuando la variable dependiente no es una variable continua pero con un estado que puede o no suceder, o categoría en una clasificación específica. La regresión logística puede ser usada para predecir una salida discreta partiendo de variables continuas y/o categóricas (Mutanen, 2006). Por ejemplo, Buckinx et.al. (2005) ha usado la regresión logística para predecir parcialmente los clientes que desertan en un negocio de *retail*.

Otro método es la Curva de Ajuste. A manera de complemento a la regresión logística, se utiliza para analizar los resultados de la estimación de la regresión. Dicha curva está relacionada con la curva ROC de la teoría de la detección de signos. El ajuste es una medida de un modelo predictivo calculado como el ratio entre los resultados obtenidos con y sin el modelo predictivo. Lo que muestra la curva es la separación de los tipos “fugados” y “no fugados”: todos los clientes fugados son detectados por el modelo. La curva de ajuste ayudará a analizar la cantidad de verdaderos fugados son discriminados en cada subconjunto. Esto será de extrema ayuda en una situación donde el departamento de marketing desea contactar un grupo de clientes. Esto es, una empresa puede contar cuantos clientes contactar si por ejemplo un 25% de clientes con propensión a fugar está por ser contactado. O si la estrategia de marketing tiene un límite de 5,000 contactos, cuantos posibles clientes con fuga son entonces alcanzados (Mutanen, 2006).

Pero citando a Wübben (2008), todas las veces que “un cliente compra de un catalogo, sale de un avión, sale de un hotel o deja la tienda de un *retail*, las empresas no tienen la manera de saber si, y con qué frecuencia el consumidor se relacionará con el negocio en el futuro”.

A finales de la década de los cincuenta, Ehrenberg (1959), publicó el artículo “El patrón en la compra del consumidor”, en que sugiere la distribución binomial negativa (NBD) como el ajuste para la información

agregada de ventas de bienes de consumo no duraderos. Dicha publicación ha sido citada numerosas veces en la literatura del marketing, probando que sus suposiciones son razonables y ampliamente aplicables. Ehrenberg (1959) dispuso de una lógica justificación por escoger esa distribución en particular. El argumenta que cada una de las compras de los clientes sigue una distribución de Poisson y que las tasas de compra asociadas varían entre clientes según la distribución Gamma. La distribución binomial negativa es exactamente la distribución teórica que surge de la mezcla de la Gamma y Poisson.

Las suposiciones que hizo Ehrenberg (1959) fueron:

1. El número de transacciones sigue un proceso Poisson con parámetro  $\lambda$ .
2. La heterogeneidad en  $\lambda$  sigue una distribución Gamma con parámetro de forma  $r$  y el parámetro de la tasa  $\alpha$  a través de los clientes.

La distribución de Poisson está caracterizada por la relación que tiene asociada su media y su varianza y que son iguales al parámetro  $\lambda$ . Es más, puede mostrarse que asumiendo que un número de transacciones distribuidas como Poisson es equivalente a asumir que el tiempo entre dos transacciones exitosas, sigue una distribución exponencial. En otras palabras, el proceso Poisson con tasa  $\lambda$  es el respectivo proceso de contar para un proceso de tiempo con tiempos de espera con distribución exponencial independientemente con media  $1/\lambda$  (Chatfield et.al., 1973).

La distribución exponencial en sí misma, es un caso particular de la distribución Gamma con el parámetro de forma igual a 1. Una propiedad importante de las variables aleatorias distribuidas exponencialmente es que no tienen memoria. Esto significa que cualquier información acerca del tiempo desde el último evento no cambia la probabilidad de un evento ocurrido dentro del futuro inmediato.

$$P(T > s + t | T > s) = P(T > t), \forall s, t \geq 0. \quad (1.2)$$

Para cálculos matemáticos, tal propiedad puede ser aprovechada, porque simplifica algunas derivaciones. Pero aplicado en información de ventas, esto implica que el tiempo de compra no depende de tan lejos en el pasado la última compra tuvo lugar. Una conclusión un poco contraria a la intuición común que más bien sugiere que los bienes de consumo no duraderos se compran con cierta regularidad. Si un cliente compra por ejemplo, un determinado bien, como un paquete de detergente, el o ella esperará hacer la siguiente compra cuando el paquete este cerca de acabarse. Pero la propiedad de "sin memoria" incluso implica, además, que el momento más probable para otra compra es inmediatamente después de una compra que acaba de ocurrir (Morrison, 1988).

Sin embargo, la distribución Poisson ha probado ser un modelo preciso para un amplio rango de aplicaciones, como el decaimiento de partículas radioactivas, la ocurrencia de accidentes o la llegada de clientes a una cola. Pero en todos los casos la propiedad de "sin memoria" soporta los controles de validez. Parece razonable por ejemplo, que el tiempo de llegada de un cliente particular en una cola es absolutamente independiente de la llegada del próximo cliente, ya que no hay interacción el uno con el otro. El hecho que un cliente haya llegado no influye en el tiempo de llegada del otro. Por lo tanto, se puede argumentar que llegadas a líneas de espera son de hecho un proceso sin memoria. Pero como se ha argumentado anteriormente, este no es el caso de las compras de bienes de consumo no duraderos para un cliente individual. La regularidad del consumo de bienes da a lugar a un cierto grado de frecuencia a su compra. Ehrenberg (1959) fue consciente de este defecto pero simplemente requiere que los periodos observados no deben ser demasiados cortos, por lo que las compras realizadas en un periodo no afectan directamente a aquellas hechas en el siguiente.

Ehrenberg (1959) sugiere un método de estimación de los parámetros que sólo requiere el número promedio de compras  $m$  y la proporción de no compradores. Sin embargo, con los programas de hoy y el potencial de las computadoras, el cálculo de la estimación por máxima



verosimilitud ya no es un problema. Este método trata de encontrar los valores de los parámetros, para los cuales la verosimilitud de la data observada es máxima. Se puede mostrar que este método tiene la propiedad a favor de ser asintóticamente insesgado, asintóticamente eficiente y ser un estimador asintóticamente normal. Para su cálculo se requiere el tiempo en el cual se está haciendo el análisis,  $T$ , y el número de transacciones en el tiempo  $T$ .

La capacidad predictiva de este modelo se vio traducida en una sobre-estimación, en el caso de una base de donantes. Una posible causa de aquel bajo rendimiento del modelo NBD es el largo tiempo estudiado en combinación con la suposición que todos los clientes se mantienen activos.

En 1987, Schmittlein, Morrison, y Colombo presentaron el modelo Pareto/NBD a la comunidad de ciencias de marketing (Schmittlein et al., 1987). En este modelo se toma en cuenta un proceso de posible deserción parte de los clientes. Este es actualmente bien conocido y bien estudiado modelo estocástico de compra para escenarios sin contratos y “tuvo incluso mayor atención por parte de investigadores y administradores en los últimos años” (Fader et al., 2005).

Schmittlein et al. trata de abordar el problema del proceso no observable de deserción o fuga. Por muchas razones los consumidores pueden decidir salir de una relación de negocios, es decir, dejar de adquirir un producto o comprar en una tienda. Las razones pueden abarcar desde el cambio en gustos o actitudes personales hasta cambios de índole personal tales como el matrimonio, hijos, enfermedad o mudanzas hasta la más exacta definición de deserción, llamada muerte. Pero sin tener en cuenta la causa real, el problema fundamental en una relación comercial no contractual es que la organización generalmente no será avisada o notificada de la deserción. Por lo tanto, la organización se basa en otros indicadores para evaluar el estado de actividad actual.

Este modelo utiliza la misma técnica que el modelo NBD utilizó para modelar las frecuencias de compras individuales y lo aplica el proceso de deserción. En particular asume algún tipo de proceso estocástico de deserción individual y hace suposiciones con respecto a la heterogeneidad sobre todos los clientes al mismo tiempo.

Las suposiciones del modelo Pareto/NBD son:

1. Mientras este activo, el número de transacciones sigue un proceso Poisson con parámetro  $\lambda$ .
2. La Heterogeneidad en  $\lambda$  sigue una distribución Gamma con parámetro de forma  $r$  y parámetro de tasa  $\alpha$  para todos los clientes.
3. El tiempo de vida de los clientes está distribuido exponencialmente con tasa de muerte  $\mu$ .
4. La Heterogeneidad en  $\mu$  sigue una distribución Gamma con parámetro de forma  $s$  y parámetro de tasa  $\beta$  para todos los clientes.
5. La tasa de compra  $\lambda$  y la tasa de muerte  $\mu$  están distribuidos independientemente.

El modelo BG/NBD fue introducido por Fader et al. (2005) quienes llamaron la atención por la discrepancia entre el alto interés científico del modelo Pareto/NBD, medido en base a las citaciones, y el pequeño número de implementaciones reales. Ellos argumentaron que debido a la matemática compleja y a la carga computacional requerida por el modelo Pareto/NBD es que mantiene a los profesionales alejados de aplicarlo a los datos del mundo real. Por ello, dichos autores introdujeron un modelo alternativo que hace suposiciones ligeramente diferentes acerca de la deserción denominándolo Beta-Geometric/NBD (abbr. BG/NBD). Ellos tuvieron éxito en simplificar las principales expresiones matemáticas del modelo y demostraron que su implementación es actualmente posible incluso con una hoja de cálculo. Es más, ellos mostraron que a pesar del cambio en las suposiciones, la exactitud del ajuste de los resultados y el

poder predictivo individual son muy similares los resultados del modelo Pareto/NBD. Las suposiciones que maneja son:

1. Mientras este activo, número de transacciones sigue un proceso de Poisson con tasa  $\lambda$ .
2. Heterogeneidad en  $\lambda$  sigue una distribución Gamma con parámetro de forma  $r$  y parámetro de tasa  $\alpha$  sobre todos los clientes.
3. Directamente después de cada compra hay una probabilidad constante  $p$  que los clientes se conviertan en inactivos:
4. Heterogeneidad en  $p$  sigue una distribución Beta con parámetros  $a$  y  $b$  sobre todos los clientes.
5. La tasa de transacción  $\lambda$  y la probabilidad de deserción  $p$  están distribuidos independientemente uno del otro.

La historia del comportamiento asumido acerca del proceso de deserción está modificado por Fader et al. en lo que respecta que un cliente no puede desertar en un punto arbitrario en el tiempo, solamente justo después que una compra fue hecha. Esta modificación parece ser razonable porque el consumidor es más probable que tenga una experiencia positiva o negativa sobre el producto o servicio justo después de la compra. Y esta extensión de la satisfacción tendrá una fuerte influencia en las decisiones de compra futuras.

El tiempo de vida medido en el número de transacciones “sobrevividas” resulta en una distribución geométrica. Esta distribución puede ser vista como la versión discreta de la distribución exponencial continua ya que se caracteriza también por no tener memoria. Esto significa que el número de transacciones alcanzadas no afecta a la probabilidad de deserción  $p$  para las siguientes transacciones. Asumir que  $p$  se mantiene constante parece razonable ya que es posible encontrar argumentos a favor de probabilidades de deserción muy tempranas (por ejemplo el cliente está probando el producto) así como una alta probabilidad tardía (por ejemplo el cliente se cansa de cierto producto o servicio y es más probable que lo cambie por uno nuevo).

El último supuesto requiere independencia entre la probabilidad de deserción y la frecuencia de compra. Pero se debe poner atención al hecho que el tiempo de vida real medido en días y no en número de transacciones alcanzadas es, comparado al modelo Pareto/NBD, ya no independiente de la frecuencia de compra. Mientras más frecuente compre un cliente, tendrá más oportunidades de deserción y por la independencia de  $p$  es  $\lambda$  lo más temprano que un cliente desertará. Interesantemente, esta consecuencia de la suposición, parece no jugar un papel importante en la exactitud del modelo. En el desarrollo del modelo se pudo comprobar que el modelo BG/NBD tiene problemas en imitar el modelo Pareto/NBD si la tasa de transacción es muy pequeña.

El modelo que resuelve este problema es el CBG/NBD desarrollado por Hoppe et al. (2007), el cual es una variante del modelo BG/NBD. Lo que hace es agregar una oportunidad de deserción adicional en el tiempo cero. Haciendo esto se resuelve la implicación no realista del modelo BG/NBD de que todos los consumidores que no tiene (nuevas) compras para nada después del tiempo cero son todavía activos. Sus suposiciones son:

1. Mientras este activo, el número de transacciones sigue un proceso de Poisson con tasa  $\lambda$ .
2. La heterogeneidad en  $\lambda$  sigue una distribución Gamma con parámetro de forma  $r$  y tasa  $\alpha$  en todos los clientes.
3. En el tiempo cero y directamente después de cada compra hay una probabilidad constante  $p$  que el cliente se vuelva inactivo.
4. La heterogeneidad en  $p$  sigue una distribución Beta con parámetros  $a$  y  $b$  sobre todos los clientes.
5. La tasa de transacción  $\lambda$  y la probabilidad de fuga están distribuidas independientemente para cada una.

Según Platzer (2008), la meta es trabajar con modelos de probabilidad que permitan pronosticar el comportamiento del consumidor en condiciones sin contratos y a un nivel individual. Asimismo, se deben

usar modelos que puedan incorporar la deserción del cliente para poder modelar la frecuencia de compra. Lo que es más, se enfrenta el problema de las relaciones no contractuales con los clientes, una característica ampliamente común, pero que desafortunadamente añade mayor complejidad a la tarea de predicción. La dificultad surge porque no hay información definitiva respecto al estado de la relación cliente-empresa. Esto significa que es imposible decir cuando un cliente está aún activo y/o fugado. Por el contrario, en un escenario con contrato, como lo es la base de una empresa de telecomunicaciones, es conocido cuando un cliente cancela su contrato y se pierde para siempre. Sin contratos de por medio, como las tiendas por departamento, pasajeros aéreos o donantes, no podemos observar el estado actual de la relación cliente-empresa, sino más bien se basa en otra información tales como el historial de transacciones para hacer juicios apropiados. Por tanto, se debe limitar la investigación a modelos que puedan manipular este tipo de incertidumbre. Lo que es más, de disponer sólo de la historia de las transacciones, lo importante es que los modelos sean capaces de extraer lo más valioso de aquella, y que no se base en incorporar otras covariables, tales como las demográficas, la actividad de la competencia u otras variables exógenas.

Según los hallazgos de Platzer (2008), investigando las suposiciones del modelo NBD y sus implicaciones, señaló que modelando con la distribución binomial negativa es equivalente a asumir total aleatoriedad en los tiempos de transacción. Se encontró que dicha suposición no es cierta cuando en datos de compras de bienes que son consumidos con cierta regularidad. Por otro lado, se demostró que los modelos mostrados, arrojaron resultados mediocres al momento de extrapolar al futuro. Por ello se busca algún tipo de ajuste direccionada a la regularidad en los modelos estocásticos.

¿Cómo puede ser medida la regularidad? Los tiempos observados pueden caer entre patrones totalmente aleatorios hasta patrones con regularidad de reloj. Por ello, una medida de regularidad debe proporcionar apoyo para indicar en donde se encuentran entre esos dos

extremos. Un método común para evaluar la regularidad es ajustar una distribución Gamma a los tiempos inter-transaccionales observados y subsecuentemente revisar los parámetros de forma estimados. Dunn et al. (1983) repitieron la aproximación de H.C.S. Thom de la Estimación por Máxima Verosimilitud del parámetro de forma como sigue:

$$\tilde{r} = \frac{1}{4} Y^{-1} \left( 1 + \sqrt{1 + \frac{4}{3} Y} \right), \quad Y = \log\left(\frac{\text{media aritmética}}{\text{media geométrica}}\right) \quad (1.3)$$

Adicionalmente, Wagner et al. (1986), proporcionan una prueba estadística y una distribución teórica asociada que permite a los especialistas en marketing probar adecuadamente si un proceso observado es Poisson. Si el parámetro de forma para los tiempos inter-transacción es cercano a uno, entonces la suposición de Poisson no necesita ser rechazada para los clientes. Esto resulta del hecho que la correspondiente distribución exponencial iguala a la distribución Gamma con parámetro de forma 1.

Pero el problema surge cuando se aplica datos reales, ya que es necesario por lo menos 5 o más transacciones para cada individuo, caso contrario las estimaciones serían sesgadas. Desafortunadamente, en la mayoría de las veces, tal información no está disponible o no existe. Hoppe et al. (2007) por ejemplo aplicaron esta prueba para la suposición de Poisson en datos de compras de un minorista por catálogo y tuvo que restringirlo sólo al 10% de los clientes con al menos 5 compras. Los cálculos mostraron que la suposición de Poisson tenía que ser rechazada. Por tanto la prueba confirmaba que debía de mantenerse la suposición de NBD.

Por otra parte, Wheat et al. (1990) introdujeron otra medida en el campo del comportamiento del consumidor. Esta medida lo que busca es flexibilizar el problema de la restricción de los registros con transacciones con historia. Ellos también asumieron que los tiempos inter-transaccionales están distribuidos según una distribución Gamma, pero

adicionalmente asumen que todos los clientes comparten el mismo parámetro de forma  $r$ . Para esta medida  $M$ , sólo se requiere 2 tiempos de transacción por cada cliente.

$$M = \frac{T_1}{T_1 + T_2} \tag{1.4}$$

Ellos muestran que bajo esa suposición  $M$  sigue una distribución Beta( $r,r$ ). Por tanto  $M$  está distribuido uniformemente dentro del intervalo (0,1) en el caso los tiempos entre eventos estén distribuidos exponencialmente ( $r=1$ ). La estimación real para  $r$  está dado por:

$$\hat{r} = \frac{1 - 4 \cdot \text{var}(M)}{8 \cdot \text{var}(M)} \tag{1.5}$$

Siendo  $\text{var}(M)$  la estimación de la varianza de  $M$ . Esta estimación de  $r$  sirve otra vez como medida para la regularidad observada, pero no a un nivel individual para la regularidad del cohorte completo. Cuando se comprueba que se debe de rechazar la suposición de que los tiempos entre transacciones sigue una distribución exponencial, se sugiere una distribución alternativa conocida como la distribución Gamma con parámetro de forma  $r$ .

La familia de distribuciones Erlang- $k$  es un caso especial de la distribución Gamma con parámetro de forma restringido a valores enteros positivos. El parámetro de forma  $r$  es puesto después en la  $k$ . Una variable distribuida como Erlang- $k$  puede ser vista como la suma de  $k$  variables independientes e idénticamente distribuidas que sigue una distribución exponencial. Otra interpretación es que el tiempo entre eventos son exponencialmente distribuidos, pero sólo cada  $k$ -ésimo evento es observado o contado, por ello el término proceso de cuenta censurado es usado en tales modelos. (Chatfield et al., 1973).

Por muchas razones la familia de distribuciones Erlang- $k$  parece ser una buena elección para incorporar regularidad. En primer lugar, es

posible modelar en un grado específico de regularidad eligiendo  $k$  de acuerdo a los estimados observados (ecuación de  $r$ ).

En segundo lugar, la distribución Erlang- $k$  es, debido a su relación con el proceso de Poisson, matemáticamente fácil de manejar, en comparación con la Weibull o Log-normal. Y finalmente, es posible describir la historia del comportamiento que resulta en una distribución Erlang- $k$  de los tiempos inter-compras. Incluso si un usuario consume cierto bien de acuerdo a una Poisson, esto es, en tiempos totalmente aleatorios, pero sólo cada  $k$ -ésimo consumo resulta una compra de un nuevo paquete de ese bien, entonces los tiempos de espera observados entre dos compras estarán distribuidos según la distribución Erlang- $k$ .

El modelo que contempla la distribución Erlang- $k$  es el CBG/CNBD- $k$  cuyas suposiciones son:

1. Mientras este activo, las transacciones de los consumidores ocurren con tiempo de espera distribuidos como Erlang- $k$  (parámetro de tasa  $\lambda$ ).
2. Heterogeneidad en  $\lambda$  sigue una distribución Gamma con parámetro de forma  $r$  y parámetro de tasa  $\alpha$  para todos los clientes.
3. En el tiempo cero y directamente después de cada transacción hay una probabilidad constante  $p$  que el consumidor se vuelva inactivo.
4. La heterogeneidad de  $p$  sigue una distribución Beta con parámetros  $a$  y  $b$  sobre todos los clientes.
5. La tasa de transacción  $\lambda$  y la probabilidad de fuga  $p$  están distribuidos independientemente uno del otro.
6. El tiempo de observación de cada individuo empieza con una transacción en el tiempo cero.

La idea de modelar mediante Erlang- $k$  los tiempos entre compras no es nueva del todo en el campo del comportamiento del consumidor. Herniter (1971) también observó periodos muertos dentro de sus histogramas de tiempos entre compras observados. En ese momento, él



fue el primero en sugerir la familia de distribuciones Erlang-k para ajustar dichos histogramas apropiadamente. Pero analizando el ratio de la varianza respecto a la media, es decir, el coeficiente de variación, concluyó que una distribución Erlang-2 provee el mejor ajuste para sus datos.

Dos años después, Chatfield et al. (1973), investigaron esta aproximación en detalle. En primer lugar, ellos derivaron algunos resultados básicos en base a la distribución de probabilidad del tiempo de proceso que corresponde a tiempos entre compras Erlang-2. Ellos acuñaron el término distribución condensada de Poisson. Esto debido a su cercana relación a la distribución de Poisson. Pero opuesto a la Poisson es ahora más pequeña que su media, por ello el término condensada. En segundo lugar, ellos siguieron a Ehrenberg (1959) y también asumieron una mixtura Gamma de las frecuencias de compra. La distribución derivada ha sido llamada consistentemente Distribución Binomial Negativa Condensada (CNBD por sus siglas en inglés).

Es necesario tomar en cuenta que Chatfield (1973) asumieron un tiempo de inicio arbitrario para el proceso de conteo, por ello la distribución de Poisson condensada asume una distribución de conteo asíncrona. Por el contrario, la suposición 6 postula una distribución de conteo síncrono el cual surge cuando el inicio del conteo coincide con un evento (Haight, 1965). Sin embargo, el nombre del modelo intencionalmente contiene el término condensado por dos razones. Por una parte, la distribución de conteo resultante es "condensada" si se examina el coeficiente de variación. Por otra parte, el conteo asíncrono tiene que ser asumido para un periodo objetivo con el fin de mantener la complejidad matemática dentro de los límites.

Schmittlein et al. (1983) demostraron que el modelo CNBD efectivamente supera el rendimiento del modelo NBD significativamente, en particular cuando el número de no compradores es grande.

## **CAPITULO II**

### **El problema de investigación**

En un negocio, los altos directivos saben cuáles son los clientes, en un momento determinado, que les dan mayores réditos a la empresa, pero, ¿qué sucedería si desean saber cuáles son los clientes que continuarán fidelizados en los siguientes meses o años y por tanto, seguirán aportando a la empresa a través de sus compras?

Esta interrogante lleva a analizar a los clientes a través de los datos que se tiene de ellos. Dentro de ellos encontramos las compras, también llamadas transacciones, que han realizado los clientes desde el inicio de la vinculación contractual con la empresa. Estas transacciones conforman una frecuencia con la que el cliente adquiere los productos ofertados por la empresa.

Por tanto, el problema se centra en lo siguiente: ¿mediante el número de transacciones históricas de cada cliente, se puede estimar su tiempo de vida comercial en un negocio?, es decir ¿se puede estimar cuántas transacciones futuras tendrá el cliente en un periodo determinado?

Las preguntas que se propone responder en la presente investigación son:

## **2.1 Pregunta Principal**

¿Mediante el número de transacciones históricas de cada cliente, se puede estimar su tiempo de vida comercial en un negocio?

## **2.2 Preguntas Específicas**

1. ¿Cuántas son las futuras transacciones que puede tener un cliente en un periodo determinado?
2. ¿Cuál es la probabilidad que un cliente siga continúe activo o pase a un estado de inactividad o de fuga en un periodo determinado?
3. ¿Cómo estarían relacionadas las características de mis clientes en relación con esa probabilidad de actividad en un periodo determinado?

## CAPITULO III

### Modelado estadístico: Distribución Erlang-k

Los modelos ampliamente expuestos en líneas anteriores desembocan en la utilización de distribuciones de probabilidad derivadas de la Gamma. Una de ellas es la Erlang-k, en la cual la presente investigación se apoya para alcanzar sus objetivos y que maneja ciertas suposiciones.

#### 3.1 Suposiciones

1. Mientras está activo, las transacciones del consumidor ocurre con una distribución de tiempo de espera Erlang-k (parámetro de tasa  $\lambda$ ).
2. La heterogeneidad en  $\lambda$  sigue una distribución Gamma con parámetro de de forma  $r$  y parámetro de tasa  $\alpha$  sobre los consumidores.
3. En el tiempo cero y directamente después de cada transacción hay una probabilidad constante  $p$  que el cliente se vuelva inactivo.
4. La heterogeneidad en  $p$  sigue una distribución Beta con parámetro  $a$  y  $b$  sobre los clientes.
5. La tasa de transacción  $\lambda$  y la probabilidad de fuga están distribuidos independientemente una de otra.
6. El periodo de observación de cada persona inicia con una transacción en el tiempo cero.

Por otro lado, el vector más importante dentro del modelo Erlang-k es el  $(x, t_x, T)$ . En el está contenido  $x$ , que es el número de transacciones dentro del periodo  $(0, T]$  y  $t_x$  que es tiempo de la última transacción.

### 3.2 La distribución Erlang-k

La distribución Erlang-k con parámetro  $k$  y  $\lambda$  está definido por la función de densidad

$$f(t | k, \lambda) = \lambda^k t^{k-1} e^{-\lambda t} \frac{1}{(k-1)!}, \forall t > 0; k \in \mathbb{N}^+, \lambda > 0. \quad (3.1)$$

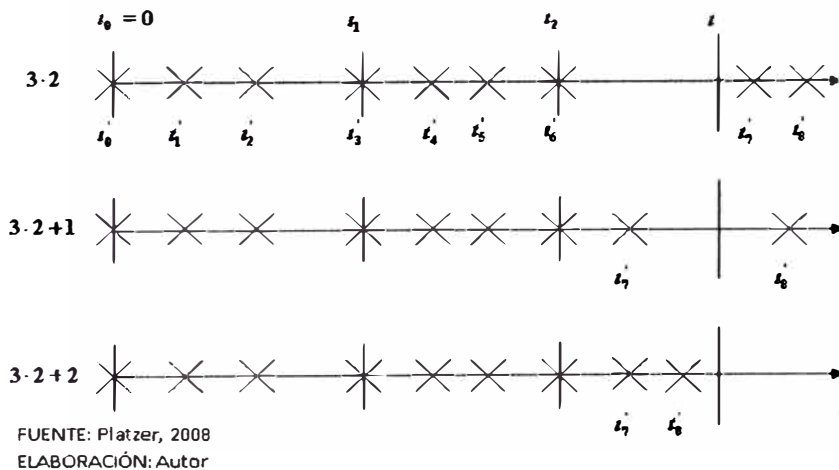
La distribución Erlang-k es una especialización de una distribución Gamma más general, con la restricción de que  $k$  debe ser entero. Si  $k=1$ , entonces estamos tratando con una distribución exponencial. La distribución Erlang-k puede ser vista también como la suma de  $k$  variables aleatorias distribuidas exponencialmente independientes e idénticamente distribuidas con parámetro  $\lambda$ . Por tanto, el correspondiente proceso de conteo de eventos con tiempos de espera con distribución Erlang-k puede ser deducido del proceso de Poisson simple. Bajo la suposición que un evento realmente ocurrió en el tiempo cero la probabilidad de encontrar  $x$  eventos hasta el tiempo  $t$  es

$$P_k = (X(t) = x) = \sum_{j=0}^{k-1} P_p(X(t) = k(x) + j). \quad (3.2)$$

Este resultado es simple si damos una vista a la Figura N° 3.1, la cual muestra la relación entre un proceso de Poisson  $(t'_0, t'_1, t'_2, \dots)$  y los tiempos de Erlang-k  $(t_0, t_1, t_2, \dots)$ . Se considera la ocurrencia de un evento como la  $k$ -ésima realización del correspondiente proceso exponencialmente distribuido  $(t_x = t'_{kx})$ . Por tanto, la probabilidad de encontrar  $x$  eventos hasta el tiempo  $t$ , es la suma de las probabilidades de

encontrar  $kx, kx+1, \dots, kx+k-1$  eventos de Poisson. El hecho de resaltar que se empieza a contar con un evento en el tiempo cero es importante ya que no se está tratando con un proceso sin memoria, como el caso de los tiempos distribuidos exponencialmente. Estar en un proceso sin memoria implica que la chance que un evento ocurra dentro de un futuro cercano se mantiene constante y es independiente del tiempo que ha pasado desde el último evento. Por otra parte, la distribución Erlang- $k$  claramente tiene un pico diferente a cero (para  $k > 1$ ). La ausencia de un proceso sin memoria es la razón porque se tiene postulada la suposición 6 para el modelo. Haight (1965) distinguió entre el proceso de conteo que empieza con un evento en el tiempo cero y los que no. El los llamó síncronos y asíncronos proceso de conteo. Chatfield et al. (1973) estudiaron el conteo asíncrono de eventos Erlang- $k$  y llamaron al resultado proceso de Poisson Condensado.

**Figura N° 3.1**  
**TIEMPOS INTER-EVENTOS DISTRIBUIDOS COMO**  
**ERLANG-3.  $P_3(X(T)=2)$  ES LA PROBABILIDAD DE**  
**ENCONTRAR 6, 7, U 8 EVENTOS DE POISSON.**



### 3.3 Verosimilitud Individual

La verosimilitud de parámetros  $\lambda$  y  $p$  para un patrón de compras particular  $(t_1, t_2, \dots, t_x, T)$  puede ser deducida. Es la verosimilitud de los periodos inter-eventos observados  $(t_1 - t_0, t_2 - t_1, \dots, t_x - t_{x-1})$ , temporaliza la

probabilidad de haber “sobrevivido” el tiempo 0 y las primeras  $x-1$  compras, temporaliza la probabilidad de no observar transacciones dentro de  $(t_x, T]$ . Mientras que lo último puede resultar de un consumidor que deserta inmediatamente después de su última compra ó de un consumidor cuya siguiente transacción simplemente sucede después del tiempo  $T$ .

$$L(\lambda, p | t_1, \dots, t_x, T) = (1-p)f_{\Gamma}(t_1 | k, \lambda) \cdots (1-p)f_{\Gamma}(t_x - t_{x-1} | k, \lambda) \cdot \\ \{p + (1-p)P(X(T - t_x) = 0 | k, \lambda)\}$$

Insertando la función de densidad de probabilidad Erlang-k (3.1) y el resultado (3.2), se tiene lo siguiente:

$$\begin{aligned} L(\lambda, p | t_1, \dots, t_x, T) &= \\ &= (1-p)^x \cdot \frac{\lambda^k t_1^{k-1} e^{-\lambda t_1}}{(k-1)!} \cdots \frac{\lambda^k (t_x - t_{x-1})^{k-1} e^{-\lambda(t_x - t_{x-1})}}{(k-1)!} \cdot \\ &\quad \left\{ p + (1-p) \sum_{j=0}^{k-1} P_p(X(T - t_x) = j | \lambda) \right\} \\ &= (1-p)^x \lambda^{kx} e^{-\lambda t_x} (1/(k-1)!)^x (t_x - t_{x-1})^{k-1} \cdots (t_1 - 0)^{k-1} \cdot \\ &\quad \left\{ p + (1-p) e^{-\lambda(T-t_x)} \sum_{j=0}^{k-1} \frac{\lambda^j (T - t_x)^j}{j!} \right\} \\ &= \tilde{t} \cdot p(1-p)^x \lambda^{kx} e^{-\lambda t_x} + \tilde{t} \cdot (1-p)^{x+1} \lambda^{kx} e^{-\lambda T} \sum_{j=0}^{k-1} \frac{\lambda^j (T - t_x)^j}{j!} \end{aligned} \quad (3.3)$$

Una diferencia importante de este resultado de los métodos de verosimilitud de modelos con tiempos exponenciales es que aun se tiene los tiempos reales de las transacciones  $t_1, \dots, t_x$  (resumido en  $\tilde{t}$ ) en la fórmula final.  $(x, t_x, T)$  no es por tanto una estadística suficiente para la verosimilitud. Pero como se ha visto, no se necesita esos tiempos para la estimación de los parámetros.

### 3.4 Verosimilitud Agregada

Con el fin de tomar en cuenta las suposiciones 2 y 4 sobre la distribución de  $\lambda$  y  $p$ , se necesita juntar la distribución gamma y beta por integración.

$$\begin{aligned}
 L(r, \alpha, a, b | t_1, \dots, t_x, T) = & \\
 \approx & \int_0^1 \int_0^\infty p(1-p)^x \lambda^{kx} e^{-\lambda t_x} f_\Gamma(\lambda | r, \alpha) f_B(p | a, b) d\lambda dp \quad (3.4) \\
 + & \int_0^1 \int_0^\infty (1-p)^{x+1} \left( \sum_{j=0}^{k-1} \frac{(T-t_x)^j \lambda^j}{j!} \right) \lambda^{kx} e^{-\lambda T} f_\Gamma(\lambda | r, \alpha) f_B(p | a, b) d\lambda dp
 \end{aligned}$$

Debido a la suposición 5 se puede resolver esta integral de manera separada y para este propósito usar las siguientes definiciones y resultados de Hoppe et al. (2008):

$$I_\Gamma(i, j, r, \alpha) := \int_0^\infty \lambda^i e^{-\lambda j} f_\Gamma(\lambda | r, \alpha) d\lambda = \frac{\alpha^r \cdot (r)_i}{(j + \alpha)^{r+i}} \quad (3.5)$$

$$I_B(i, j, a, b) := \int_0^1 p^i (1-p)^j f_B(p | a, b) dp = \frac{B(a+i, b+j)}{B(a, b)} \quad (3.6)$$

$B(a, b)$  denota la función Beta y  $(r)_x$  el símbolo de Pochhammer:

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \quad (3.7)$$

$$(r)_x = \frac{\Gamma(r+x)}{\Gamma(r)} \quad (3.8)$$

Además, se puede notar fácilmente, considerando  $\Gamma(a+1) = a\Gamma(a)$ , que

$$B(a+1, b+x) = \frac{a}{b+x} \cdot B(a, b) \quad (3.9)$$

$$(r)_{x+y} = (r+x)_y (r)_x \quad (3.10)$$

Por tanto:



$$\begin{aligned}
L(r, \alpha, a, b | t_1, \dots, t_x, T) &= \\
&= \tilde{t} \cdot (I_B(1, x, a, b) \cdot I_\Gamma(kx, t_x, r, \alpha) + \\
&\tilde{t} \cdot I_B(0, x+1, a, b) \cdot \left( \sum_{j=0}^{k-1} \frac{(T-t_x)^j}{j!} I_\Gamma(kx+j, T, r, \alpha) \right) \quad (3.11)
\end{aligned}$$

$$\begin{aligned}
&= \tilde{t} \cdot \frac{(b)_{x+1}}{(a+b)_{x+1}} \cdot \alpha^r (r)_{kx} \cdot \\
&\left( \frac{a}{b+x} \left( \frac{1}{\alpha+t_x} \right)^{r+kx} + \sum_{j=0}^{k-1} \frac{(T-t_x)^j}{j!} \frac{(r+kx)_j}{(\alpha+T)^{r+kx+j}} \right) \quad (3.12)
\end{aligned}$$

siendo  $\tilde{t}$  igual a  $t_1 \cdot (t_2 - t_1) \cdots (t_x - t_{x-1})$ .

### 3.5 Estimación de Parámetros

Un método de estimación de parámetros bien conocido, el cual es insesgado y eficiente, es la estimación por máxima verosimilitud (EMV). Este método trata de encontrar un conjunto de parámetros  $(r, \alpha, a, b)$  con el cual la verosimilitud alcanza su máximo global para algún conjunto de datos  $(t_{i,1}, \dots, t_{i,x}, T_i)_{i=1 \dots N}$ .

$$\begin{aligned}
(\hat{r}, \hat{\alpha}, \hat{a}, \hat{b}) &= \arg \max_{r, \alpha, a, b} L(r, \alpha, a, b | (t_{i,1}, \dots, t_{i,x}, T)_{i=1 \dots N}) \\
&= \arg \max_{r, \alpha, a, b} \prod_{i=1}^N L(r, \alpha, a, b | (t_{i,1}, \dots, t_{i,x}, T_i)) \quad (3.13)
\end{aligned}$$

Como puede notarse, se puede simplemente disgregar el término acumulativo  $t_i$  para los patrones de tiempo exactos, dado que este factor multiplicativo no tiene efecto en la localización del máximo, esto es, en los parámetros estimados. Por ello, se puede mantener  $(x, t_x, T)$  como información de entrada para los cálculos. Para evitar problemas con precisiones numéricas, es común en realidad optimizar el logaritmo de la verosimilitud, que transforma la multiplicación (de números muy pequeños) en una suma.

$$(\hat{r}, \hat{\alpha}, \hat{a}, \hat{b}) = \arg \max_{r, \alpha, a, b} \sum_{i=1}^N \log(L(r, \alpha, a, b | t_{i,1}, \dots, t_{i,x}, T_i)) \quad (3.14)$$

### 3.6 Distribución de Probabilidad de la Frecuencia de Compra

Se necesita una expresión para  $P(X(t) = x | r, \alpha, a, b)$ , esto es, la distribución de probabilidad de la frecuencia de compra condicional de los parámetros (estimados), y se tomarán en cuenta nuevamente las derivaciones matemáticas de Hoppe et al. (2008). Para un solo consumidor (con un  $\lambda$  y  $p$  dado) la probabilidad de encontrar  $x$  transacciones hasta el tiempo  $t$  puede ser partida en dos casos. O el consumidor simplemente ha tenido  $x$  transacciones y sigue activo en el tiempo  $t$ , ó pudo tener más de  $x$  transacciones pero desertó inmediatamente después de la  $x$ -ésima compra.

$$P(X(t) = x | \lambda, p) = (1 - p)^{x+1} P(X(t) = x) + p(1 - p)P(X(t) \geq x) \quad (3.15)$$

Usando  $P(X(t) \geq x) = 1 - P(X(t) < x)$  y el resultado (3.2), se puede derivar

$$P(X(t) = x | \lambda, p) = (1 - p)^{x+1} \sum_{j=kx}^{kx+k-1} P_p(X(t) = j) + p(1 - p)^x (1 - \delta_{x>0} \sum_{j=0}^{kx-1} P_p(X(t) = j)) \quad (3.16)$$

Se debe observar que se agregó el Delta de Kronecker, el cual es 1 para  $x > 0$  y 0 cualquier otro caso, para considerar correctamente el caso  $x = 0$  para el cual el segundo término de sumatoria simplemente se convierte en la probabilidad de deserción  $p$  en el tiempo cero.

Nuevamente se unen las suposiciones de heterogeneidad:

$$P(X(t) = x | r, \alpha, a, b) = \int_0^1 \int_0^\infty P(X(t) = x | \lambda, p) f_\Gamma(\lambda | r, \alpha) f_B(p | a, b) d\lambda dp$$

$$\begin{aligned}
&= \int_0^1 (1-p)^{x+1} f_B dp \int_0^\infty \left( \sum_{j=kx}^{kx+k-1} \frac{(\lambda t)^j}{(j)!} e^{-\lambda t} \right) f_\Gamma d\lambda \\
&+ \int_0^1 p(1-p)^x f_B dp \int_0^\infty (1 - \delta_{x>0} \sum_{j=0}^{kx-1} \frac{(\lambda t)^j}{(j)!} e^{-\lambda t}) f_\Gamma d\lambda
\end{aligned} \tag{3.17}$$

y se aplica los resultados (3.6) y (3.5)

$$\begin{aligned}
P(X(t) = x | r, \alpha, a, b) &= I_B(0, x+1, a, b) \cdot \left( \sum_{j=kx}^{kx+k-1} \frac{t^j}{j!} I_\Gamma(j, t, r, \alpha) \right) \\
&+ I_B(1, x, a, b) \cdot \left( 1 - \delta_{x>0} \sum_{j=0}^{kx-1} \frac{t^j}{j!} I_\Gamma(j, t, r, \alpha) \right) \\
&= \frac{B(a, b+x+1)}{B(a, b)} \cdot \left( \sum_{j=kx}^{kx+k-1} \frac{t^j}{j!} \frac{\alpha^r(r)_j}{(\alpha+t)^{r+j}} \right) \\
&+ \frac{B(a+1, b+x)}{B(a, b)} \cdot \left( 1 - \delta_{x>0} \sum_{j=0}^{kx-1} \frac{t^j}{j!} \frac{\alpha^r(r)_j}{(\alpha+t)^{r+j}} \right)
\end{aligned} \tag{3.18}$$

Considerando la distribución de probabilidad para la distribución binomial negativa

$$P_{NBD}(X(t) = j) = \frac{t^j}{j!} \frac{\alpha^r(r)_j}{(\alpha+t)^{r+j}}, \tag{3.19}$$

se puede escribir

$$\begin{aligned}
P(X(t) = x | r, \alpha, a, b) &= \\
&\frac{B(a, b+x+1)}{B(a, b)} \left( \sum_{j=kx}^{kx+k-1} P_{NBD}(X(t) = j) \right) \\
&+ \frac{B(a+1, b+x)}{B(a, b)} \cdot \left( 1 - \delta_{x>0} \sum_{j=0}^{kx-1} P_{NBD}(X(t) = j) \right)
\end{aligned} \tag{3.20}$$

### 3.7 Probabilidad de Actividad

Tal como lo señalo Schmittlein et al. (1987), una de las expresiones principales de los modelos de este tipo es la probabilidad de un solo cliente siga aún activo al final del periodo de observación, basado en la

historia transaccional pasada. Esto es, la probabilidad  $P(\tau > T | t_1, \dots, t_x, T, r, \alpha, a, b)$  con  $\tau$  como el tiempo de vida no observado del cliente.

$$\begin{aligned}
 P(\tau > T | t_1, \dots, t_x, T, \lambda, p) &= 1 - P(\tau \leq T | t_1, \dots, t_x, T, \lambda, p) = 1 - \frac{p}{P(X(T - t_x) = 0)} \\
 &= 1 - \frac{p}{p + (1-p) \sum_{j=0}^{k-1} P_p(T - t_x) = j)} \\
 &= \frac{(1-p) \sum_{j=0}^{k-1} P_p(X(T - t_x) = j)}{p + (1-p) \sum_{j=0}^{k-1} P_p(X(T - t_x) = j)} \quad (3.21)
 \end{aligned}$$

Expandiendo este término con  $t(1-p)^x \lambda^{kx} e^{-\lambda t}$ , y comparando el denominador con la ecuación (3.3) se obtiene

$$P(\tau > T | t_1, \dots, t_x, T, \lambda, p) = \frac{t(1-p)^{x+1} \lambda^{kx} e^{-\lambda T} \sum_{j=0}^{k-1} \frac{\lambda^j (T - t_x)^j}{j!}}{L(\lambda, p | t_1, \dots, t_x, T)} \quad (3.22)$$

Construyendo la doble integral

$$\begin{aligned}
 P(\tau > T | t_1, \dots, t_x, T, r, \alpha, a, b) &= \\
 \int_0^1 \int_0^\infty P(\tau > T | t_1, \dots, t_x, T, \lambda, p) f_\Gamma(\lambda | r, \alpha) f_B(p | a, b) d\lambda dp &\quad (3.23)
 \end{aligned}$$

y usando el siguiente resultado de Hoppe and Wagner (2008)

$$f(\lambda, p | t_1, \dots, t_x, T) = \frac{L(\lambda, p | t_1, \dots, t_x, T) f_\Gamma(\lambda | r, \alpha) f_B(p | a, b)}{L(r, \alpha, a, b | t_1, \dots, t_x, T)}, \quad (3.24)$$

lleva a

$$P(\tau > T | t_1, \dots, t_x, T, r, \alpha, a, b) =$$

$$\begin{aligned}
&= \frac{\tilde{t}}{L(r, \alpha, a, b | t_1, \dots, t_x, T)} \cdot \int_0^1 (1-p)^{x+1} f_B(p | a, b) dp \cdot \\
&\quad \int_0^\infty \lambda^{kx} e^{-\lambda T} \sum_{j=0}^{k-1} \frac{(T-t_x)^j}{j!} \lambda^j f_\Gamma(\lambda | r, \alpha) d\lambda \\
&= \tilde{t} \cdot I_B(0, x+1, a, b) \cdot \sum_{j=0}^{k-1} \frac{(T-t_x)^j}{j!} I_\Gamma(kx+j, T, r, \alpha) \\
&\quad / L(r, \alpha, a, b | t_1, \dots, t_x, T) \tag{3.25}
\end{aligned}$$

Comparando el resultado con (3.11), se puede observar que el numerador es realmente uno de los términos de la sumatoria de la función agregada de verosimilitud en el denominador. Y considerando

$\frac{A}{A+B} = (1 + \frac{A}{B})^{-1}$  la fracción puede ser reducida a

$$\begin{aligned}
&P(\tau > T | t_1, \dots, t_x, T, r, \alpha, a, b) = \\
&= (1 + \frac{\tilde{t} \cdot I_B(1, x, a, b) \cdot I_\Gamma(kx, t_x, r, \alpha)}{\tilde{t} \cdot I_B(0, x+1, a, b) \cdot \sum_{j=0}^{k-1} \frac{(T-t_x)^j}{j!} I_\Gamma(kx+j, T, r, \alpha)})^{-1} \tag{3.26}
\end{aligned}$$

Afortunadamente, el término  $\tilde{t}$  se anula y por tanto, aún no se requiere la información del tiempo exacto de las transacciones para llevar los cálculos. Se resuelven las integrales, se extraen los términos en común y se usa la relación  $(r)_{kx+j} = (r)_{kx} \cdot (r+kx)_j$  y lleva a

$$\begin{aligned}
&P(\tau > T | x, t_x, T, r, \alpha, a, b) = \\
&= (1 + \frac{B(a+1, b+x)}{B(a, b+x+1)} \cdot \frac{\alpha^r (r)_{kx}}{(\alpha+t_x)^{r+kx}} \cdot \frac{(\alpha+T)^{r+kx}}{\alpha^r (r)_{kx}} \\
&\quad / \sum_{j=0}^{k-1} \frac{(T-t_x)^j}{j!} (r+kx)_j (\alpha+T)^j)^{-1} \\
&= (1 + \frac{a}{b+x} \cdot (\frac{\alpha+T}{\alpha+t_x})^{r+kx} / \sum_{j=0}^{k-1} \frac{(T-t_x)^j}{j!} \frac{(r+kx)_j}{(\alpha+T)^j})^{-1} \tag{3.27}
\end{aligned}$$

### 3.8 Número Estimado de Transacciones

#### 3.8.1 Transacciones Estimadas No Condicionales para CBG/CNBD-k

Por temas de aproximación no se tiene una forma exacta para  $E(X(t) | r, \alpha, a, b)$ . Platzter (2008), usando algunas modificaciones heurísticas sobre el resultado de Hoppe y Wagner (2007), define

$$G(v_1, v_2, v_3, v_4 | \alpha, t) := 1 - \left(\frac{v_4}{v_4 + t}\right)^{v_1} {}_2F_1(v_1, v_2 + 1; v_3 + \alpha; \frac{t}{v_4 + t}) \quad (3.28)$$

con  ${}_2F_1$  como la función hipergeométrica Gaussiana y fijando

$$E(X(t) | r, \alpha, a, b) = \frac{b}{a-1} \cdot G(r, b, b, \alpha | \alpha, t) \quad (3.29)$$

para el número de transacciones esperadas no condicionada hasta el tiempo  $t$  para el modelo CBG/NBD. Recordando los resultados que lo esperado para conteo asíncrono es simplemente  $1/k$  del correspondiente proceso Poisson (3.29), y que el conteo asíncrono sólo difiere por algunos términos que se vuelven constantes para un horizonte de tiempo largo, simplemente se aproxima el número esperado de transacciones para el modelo CBG/CNBD-k con

$$\hat{E}(X(t) | r, \alpha, a, b) = \frac{1}{k} \cdot \frac{b}{a-1} \cdot G(r, b, b, \alpha | \alpha, t). \quad (3.30)$$

#### 3.8.2 Transacciones Estimadas Condicionales para CBG/CNBD-k

Si se parte del hecho que se está trabajando con la distribución Erlang-k, que a diferencia de la distribución exponencial, carece de la propiedad “sin memoria”, no se puede usar la siguiente relación

$$\begin{aligned}
E(Y(T, T+t) | x, t_x, T, r, \alpha, a, b) &= \\
&= E(X(t) | \tau > T, \lambda, p) \cdot (P(\tau > T | x, t_x, T, \lambda, p)) \quad (3.31)
\end{aligned}$$

como es el caso del modelo CBG/NBD. La Recencia  $(T - t_x)$  influye en el número esperado de transacciones futuras (i.e. el primer término de la multiplicación), y no sólo la probabilidad de seguir activo. Asumiendo que los clientes han sobrevivido la última transacción hace pensar que la siguiente transacción sucederá próximamente. Por tanto, se sub-estimará sistemáticamente las transacciones futuras si se usa esta relación para CBG/CNBD-k. Se procede y se adapta de los resultados de Hoppe and Wagner (2007). Ellos derivaron:

$$\begin{aligned}
E(Y(T, T+t) | x, t_x, T, r, \alpha, a, b) &= \\
&= \frac{a+b+x}{a-1} \cdot G(r+x, b+x, b+x, \alpha+T | \alpha, t) \cdot P(\tau > T | x, t_x, T, r, \alpha, a, b) \quad (3.32)
\end{aligned}$$

para el modelo CBG/NBD. En su fe de erratas (Wagner et al., 2008) de Batislam et al. (2007), se indicó que es posible derivar el resultado para la predicción actualizando los parámetros  $(r, \alpha, a, b)$  a  $(r+x, \alpha+T, a, b+x)$ . Se utiliza el resultado (3.27) para  $P(\tau > T | x, t_x, T, r, \alpha, a, b)$ , y se junta con la expresión de la sección anterior. Adicionalmente, se actualiza los parámetros de  $(r, \alpha, a, b)$  a  $(r+x, \alpha+T, a, b+x)$ , dado que se encuentran  $kx$  eventos no censurados dentro de  $(0, T]$ . Entonces, se puede concluir en:

$$\begin{aligned}
\hat{E}(Y(T, T+t) | x, t_x, T, r, \alpha, a, b) &= \frac{1}{k} \cdot \frac{a+b+x}{a-1} \cdot \\
&\cdot G(r+kx, b+x, b+x, \alpha+T | \alpha, t) \cdot P(\tau > T | x, t_x, T, r, \alpha, a, b) \quad (3.33)
\end{aligned}$$

## **CAPITULO IV**

### **Diseño de la investigación y metodología**

#### **4.1 Presentación de Metodología**

La presente investigación es de tipo cuantitativa, descriptiva y aplicada, prospectiva y de corte transversal que permitirá obtener un modelo que estime el tiempo de vida comercial de los clientes.

La metodología propone un modelo el cual ha sido evaluado con otros modelos semejantes en cuanto a la inclusión de una distribución de probabilidad para estimar los distintos fenómenos aleatorios que suceden, refiriéndose al número de transacciones específicamente (Platzer, 2008).

La metodología en sí no es limitada por el sólo hecho de calcular una probabilidad de actividad, sino, que utilizando esta información predice el tiempo de vida de los clientes en base a sus transacciones pasadas de un determinado periodo. Debe entenderse que el tiempo de vida de los clientes es la cantidad de transacciones estimadas en un periodo de tiempo determinado.

Las distribuciones de la familia de las exponenciales son utilizadas ampliamente en estudios de supervivencia, debido a las características únicas que tienen. La distribución de probabilidad a utilizar es la distribución Erlang. A diferencia de otras distribuciones propuestas para esta clase de modelado, ésta es más cómoda para trabajar que una Log-Normal o Weibull (Platzer, 2008).



Por otra parte, la distribución Erlang tiene ventaja sobre la exponencial, debido a que carece de la propiedad "sin memoria". Esto quiere decir que a la distribución Erlang si le interesa el pasado y a partir de él calcula una nueva transacción.

#### **4.2 Obtención de datos**

Los datos a trabajar en la presente investigación han sido obtenidos de fuentes secundarias proporcionadas por una institución privada, por lo que se reserva el nombre de dicha empresa. El periodo de tiempo considerado para desarrollar las variables requeridas del modelo se inicia en enero del año 2007 hasta febrero del año 2010. Las variables proporcionadas son las mencionadas en el apartado 4.6.

#### **4.3 Tratamiento de datos y estimación de parámetros**

En primer lugar, los datos a trabajar deben estar completos, es decir, contar con el periodo de inicio de actividades para poder, a partir de ahí, realizar las estimaciones de las probabilidades de actividad y de las transacciones. Adicionalmente se va a trabajar con las distribuciones Gamma y Beta para la estimación de los demás parámetros como la tasa de transacción y probabilidad de fuga respectivamente. Dado que se desea trabajar a nivel cliente y no solo a nivel de la población, la versatilidad del modelo da la disponibilidad para poder trabajar en el nivel deseado y poder obtener las estimaciones requeridas.

#### **4.4 Aplicación de modelo**

Teniendo claro los supuestos para el cálculo de los parámetros estimados y la información de transacciones agregada (a nivel de número de meses con y sin transacciones), se pasa a estimar en primer lugar la probabilidad de actividad de cada cliente y las futuras transacciones a través de la distribución de Erlang.

## 4.5 Aplicación de resultados

Lo primero a estimar son los parámetros asociados y que resultan de la aplicación del método de estimación por máxima verosimilitud. Posteriormente teniendo los parámetros estimados, se pasa a estimar la probabilidad de actividad y con ello se podrá estimar el número de transacciones futuras en un periodo determinado. Asimismo teniendo los montos promedios de compra en el periodo de modelado se podrá calcular los montos de compras futuros con la participación de las transacciones futuras. Estos resultados se pueden analizar conjuntamente con las variables disponibles por cada cliente. Adicionalmente, se podrá segmentar a los clientes con estos dos indicadores, la probabilidad de actividad y el número de transacciones estimadas. En general, el negocio que disponga de suficiente información socio-demográficas de sus clientes puede formar perfiles más poderosos de clientes y de esta manera categorizarlos e implementar sus estrategias de fidelización y marketing.

## 4.6 Especificación de datos

La población en estudio son los clientes de una tienda *retail* o minorista que ingresaron a su cartera de clientes en el mes de enero del 2007 y con transacciones entre enero del 2007 y febrero del 2010. El tamaño de la población es de 150 clientes. Para la aplicación del modelo se agrupará desde enero del 2007 hasta julio del 2008 como periodo de modelado y desde agosto del 2008 hasta febrero del 2010 servirá como periodo de validación. Las variables con las que se trabaja la metodología son:

1. Número de meses con y sin transacciones.
2. Monto Total de Compra mensual.
3. Mes de primera y última transacción.
4. Variables socio-demográficas y crediticias

## CAPITULO V

### Análisis de resultados

Se dispone de una población compuesta por 150 clientes de un grupo de clientes que fueron registrados como nuevos clientes de una tienda tipo *retail* en enero 2007. Se cuenta con información transaccional además de la socio-demográfica y financiera. En lo referente a las transacciones se cuenta con los 38 meses posteriores a su incorporación como clientes además de los montos de compras efectuadas. En primer lugar se desarrolla un análisis exploratorio de los datos. Las variables con las que se cuentan, además de las transacciones de cada cliente se muestran en el Cuadro N° 5.1.

**Cuadro N° 5.1**  
**VARIABLES EN ANÁLISIS DESCRIPTIVO**

N°	VARIABLES
1	Sexo
2	Edad
3	Ubicación geográfica
4	Tpo de crédito
5	Línea de crédito
6	Transacciones
7	Montos de Compra

FUENTE: Base estudio

2010

ELABORACIÓN: Propia

Para cada una se realizó el análisis descriptivo correspondiente.

## 5.1 Análisis Descriptivo

Para la variable sexo se obtuvo el resultado que se muestra en el Cuadro N° 5.2. Se puede notar que en este grupo de clientes, la presencia es mayoritariamente masculina. Cerca de las dos terceras partes de los clientes son hombres. El sexo de por sí es una variable importante si se desea posteriormente hacer un perfil de los clientes con los resultados que se alcancen.

**Cuadro N° 5.2**  
**DISTRIBUCIÓN DE FRECUENCIA DE LOS CLIENTES DE RETAIL**  
**SEGÚN SEXO**

Sexo	Cientes	Porcentaje
Femenino	56	37%
Masculino	94	63%
<b>Total</b>	<b>150</b>	<b>100%</b>

FUENTE: Base estudio

2010

ELABORACIÓN: Propia

La siguiente variable analizada es la edad, la cual es determinante en el comportamiento de compra de los clientes de cualquier negocio. Su tratamiento puede dar lineamientos para una segmentación. La descripción de esta variable está contenida en la Cuadro N° 5.3:

**Cuadro N° 5.3**  
**DISTRIBUCIÓN DE FRECUENCIA DE LOS CLIENTES DE RETAIL**  
**SEGÚN RANGO DE EDAD**

Rango de Edad	Cientes	Porcentaje
[18 - 24]	4	3%
[25 - 30]	40	27%
[31 - 35]	32	21%
[36 - 45]	45	30%
[46 - 55]	15	10%
[56 a más]	14	9%
<b>Total</b>	<b>150</b>	<b>100%</b>

FUENTE: Base estudio

2010

ELABORACION: Propia

Se puede observar del cuadro N° 5.3 la concentración de clientes entre las edades de 25 a 45 años. Es un indicador de la juventud de este grupo. Además da indicios de los resultados que podemos encontrar más adelante y como la edad está asociada a la fidelidad. Otra variable analizada es la ubicación geográfica, más precisamente a nivel de departamento. El Cuadro N° 5.4 muestra como están repartidos o distribuidos los clientes en los diferentes departamentos que hay y en los cuales residen. Como era previsible, la mayoría (53%) residen en el departamento de Lima, seguido de Lambayeque (21%). Son puntos donde está desarrollado el *retail* más intensamente, resultando por ello en estos valores. Por otra parte, en el Cuadro N° 5.5, se analiza el tipo de crédito que brinda el *retail* a sus clientes. Este crédito tiene dos clases: cuotas o revolving. En este caso, la gran mayoría de clientes (85%) han sido gestionados a través de un crédito en cuotas.

**Cuadro N° 5.4**  
**DISTRIBUCIÓN DE FRECUENCIA DE LOS CLIENTES DE RETAIL**  
**SEGÚN DEPARTAMENTO DE RESIDENCIA**

Departamento	Clientes	Porcentaje
LIMA	79	53%
LAMBAYEQUE	32	21%
LA LIBERTAD	25	17%
AREQUIPA	8	5%
PIURA	6	4%
<b>Total</b>	<b>150</b>	<b>100%</b>

FUENTE: Base estudio

2010

ELABORACIÓN: Propia

**Cuadro N° 5.5**  
**DISTRIBUCIÓN DE FRECUENCIA DE LOS CLIENTES DE RETAIL**  
**SEGÚN TIPO DE CRÉDITO**

Tipo de Crédito	Clientes	Porcentaje
Cuotas	128	85%
Revolvente	22	15%
<b>Total</b>	<b>150</b>	<b>100%</b>

FUENTE: Base estudio

2010

ELABORACIÓN: Propia

Otro factor importante, no sólo como oferta para el cliente, sino como demanda del mismo, es la línea de crédito que maneja el cliente al momento de ser admitido como tal en la cartera del negocio. El desglose de esta variable se muestra en el Cuadro N° 5.6.

**Cuadro N° 5.6**  
**DISTRIBUCIÓN DE FRECUENCIA DE LOS CLIENTES DE RETAIL SEGÚN LÍNEA DE CRÉDITO**

<b>Rango de Línea de Crédito</b>	<b>Clientes</b>	<b>Porcentaje</b>
<0-1,000]	109	73%
<1,000-2,000]	20	13%
<2,000-3,000]	9	6%
<3,000-4,000]	5	3%
<4,000-5,000]	2	1%
<5,000 a más	5	3%
<b>Total</b>	<b>150</b>	<b>100%</b>

FUENTE: Base estudio

2010

ELABORACION: Propia

Se puede notar que este grupo es en casi sus tres cuartas partes que tiene líneas de crédito por debajo de los S/.1,000. Si bien un resultado del trabajo del área de riesgos para mitigar las posibles pérdidas es otorgar líneas de crédito acordes con el nivel de endeudamiento de cada cliente, esto a veces puede generar resultados inesperados como clientes insatisfechos con las líneas otorgadas lo que puede desembocar en una fidelización parcial o temporal. Las relaciones de estas variables con los resultados del modelo generado brindarán otro punto de vista más nutrido de conocimiento y así dar recomendaciones mejor sustentadas.

Una variable importante, tal vez la más importante, y que puede colaborar en hallar a los clientes más valiosos, es el monto de compra. En este caso, en primer lugar se desarrolla aquel perteneciente al periodo de modelado. El Cuadro N° 5.7 nos muestra cómo están distribuidos los clientes en los diferentes rangos de compra.

**Cuadro N° 5.7**

**DISTRIBUCIÓN DE FRECUENCIA DE LOS CLIENTES DE RETAIL SEGÚN MONTOS DE COMPRA EN EL PERIODO DE MODELAMIENTO**

Rango de Monto de Compra	Clientes	Porcentaje
<=100	11	7%
>100;200]	19	13%
>200;300]	31	21%
>300;400]	17	11%
>400;500]	24	16%
>500;800]	36	24%
>800;1,000]	5	3%
>1,000;1,500]	3	2%
>1,500;2,000]	3	2%
>2,000	1	1%
<b>Total</b>	<b>150</b>	<b>100%</b>

FUENTE: Base estudio

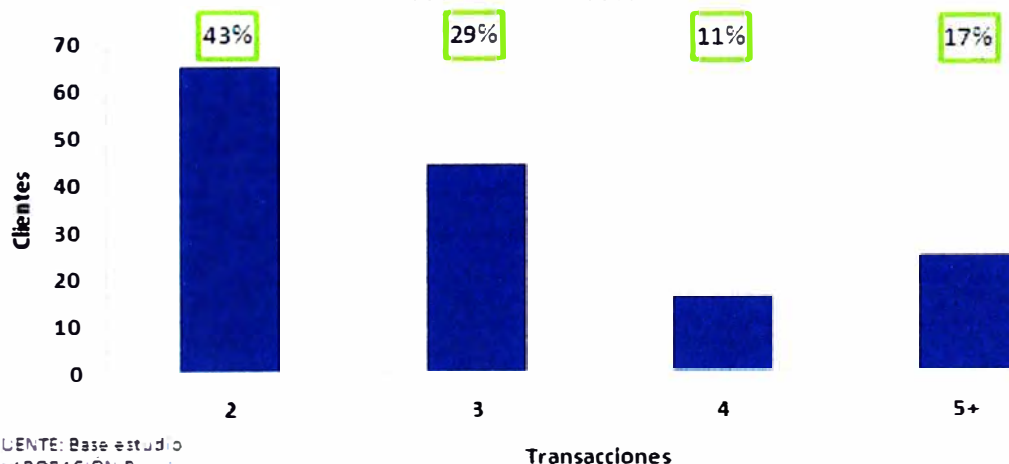
2010

ELABORACIÓN: Propia

Se puede observar que cerca del 60% de los clientes ha realizado compras entre S/.300 y S/.800. Más adelante se analizará los montos de compra para el periodo de validación. Además, se presenta el Gráfico N° 5.1 que muestra la distribución de las transacciones de los clientes en el periodo de modelado.

**Gráfico N° 5.1**

**NÚMERO DE CLIENTES POR CANTIDAD DE TRANSACCIONES EN PERIODO DE MODELAMIENTO**



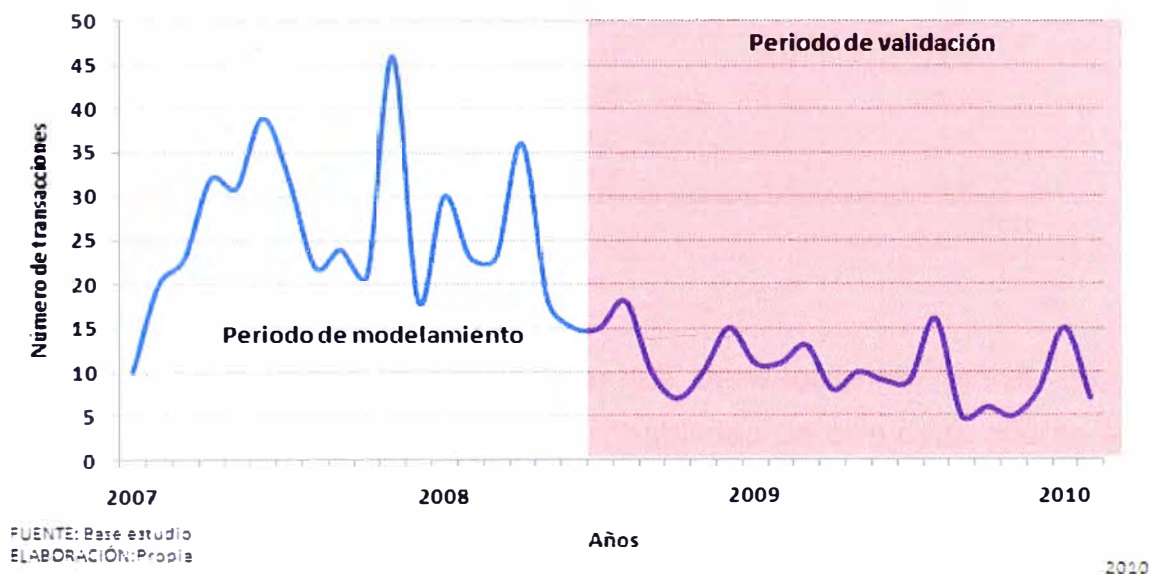
FUENTE: Base estudio  
ELABORACIÓN: Propia

2010

## 5.2 Modelado

Para esta etapa se utilizó los primeros 19 meses de transacciones. Los siguientes 19 meses de observaciones se usó para la validación del modelo. El Gráfico N° 5.2, muestra los periodos mencionados donde se puede observar su comportamiento a través del tiempo y cómo ha ido evolucionando. Se puede notar una caída importante en la cantidad de transacciones en el periodo de validación. Este comportamiento debe reflejarse en las estimaciones de las transacciones mediante el modelo.

**Gráfico N° 5.2**  
**EVOLUCIÓN DE TRANSACCIONES EN EL PERIODO DE MODELAMIENTO Y VALIDACIÓN**



Como se recuerda, se está trabajando en la construcción de un modelo Erlang-k para el cual se requiere estimar el vector de parámetros  $(r, \alpha, s, \beta)$ . Además uno de los pilares de la presente investigación es el concepto de RFM. Por ello, la información disponible de los clientes son las transacciones realizadas en el periodo mencionado. Estas transacciones se transforman para obtener los tiempos de Recencia  $t_x$ , la cantidad de transacciones  $x$  y el periodo de observación  $T$  de cada



cliente. Para iniciar el proceso de estimación de las transacciones de los clientes, se necesita primero estimar los parámetros del vector  $(r, \alpha, s, \beta)$ .

Para ello, se requiere del proceso de estimación por máxima verosimilitud para maximizar la función  $L(r, \alpha, a, b | t_1, \dots, t_x, T)$  y con ello estimar los parámetros respectivos. Los resultados entregados por el software *Mathlab 12.0*, para los valores de  $k$  propuestos con su respectivo vector  $(r, \alpha, s, \beta)$  están en la Cuadro N° 5.8:

**Cuadro N° 5.8**  
**VALORES PARA LOS PARÁMETROS  $(r, \alpha, s, \beta)$  SEGÚN MODELO**  
**ERLANG-K PROPUESTO**

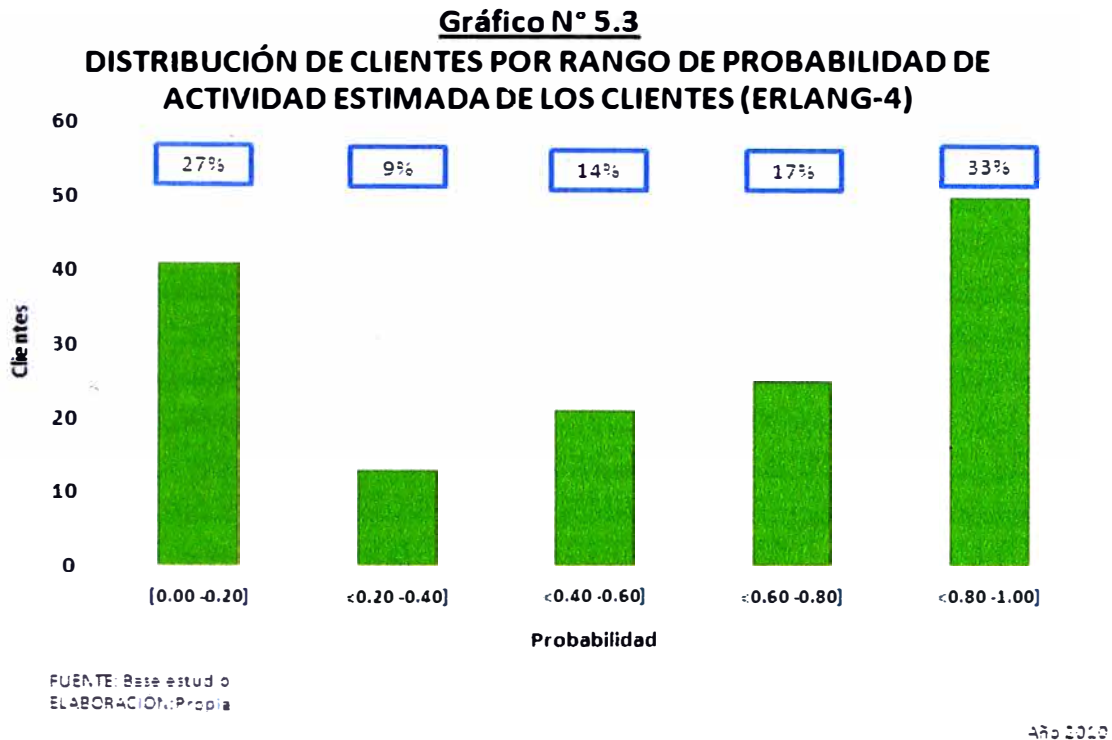
Erlang-k	Parámetros			
	r	$\alpha$	s	$\beta$
1	1,6	5,0	1,0	5,0
2	3,0	5,0	1,0	5,0
3	4,5	5,0	1,0	5,0
4	4,5	3,5	1,1	5,0

FUENTE: Base estudio

2010

ELABORACION: Propia

Para ser capaz de pronosticar las transacciones futuras de los clientes, se debe calcular a priori la probabilidad de que cada cliente siga activo después del periodo de modelado. En base al vector  $(r, \alpha, s, \beta)$  estimado, y los tiempos de Recencia  $t_x$ , la cantidad de transacciones  $x$  y el periodo de observación  $T$  de cada cliente, se estima las probabilidades de cada uno de ellos. Como muestra del cálculo de dichas probabilidades se presenta la distribución de aquellas para el modelo Erlang-4 ( $k=4$ ) y se muestra en el Gráfico N° 5.3



Ya teniendo calculadas las probabilidades de actividad para valor de  $k$ , se puede disponer de ellas para estimar las transacciones del grupo de clientes, además de tener el respectivo vector  $(r, \alpha, s, \beta)$  y los tiempos de Recencia  $t_x$ , la cantidad de transacciones  $x$  y el periodo de observación  $T$  de cada cliente. Con estos *inputs* ó 'entradas' se pasó a calcular las transacciones futuras para los siguientes 19 meses, que es el periodo de validación. Se obtuvieron los resultados para los cuatro valores de  $k$  y se continuó a comparar las distribuciones acumuladas de las transacciones pronosticadas con la distribución acumulada de las transacciones reales y aplicar sobre cada una de ellas una prueba de bondad de ajuste de chi-cuadrado comparando las frecuencias acumuladas esperadas con las observadas para poner a prueba la hipótesis de que ambas distribuciones provienen de una misma población. Se procedió con la prueba de chi-cuadrado a un nivel de significancia de 0.05 y con 18 grados de libertad. El valor crítico es 28.9. Los resultados del contraste de bondad de ajuste se muestran en el Cuadro N° 5.9. De las comparaciones del valor del estadístico con el valor crítico, se descartan los 3 primeros modelos, quedándonos así con el modelo

Erlang-4 al tener el valor de chi-cuadrado menor al valor crítico. Es el que mejor se ajusta a la distribución esperada.

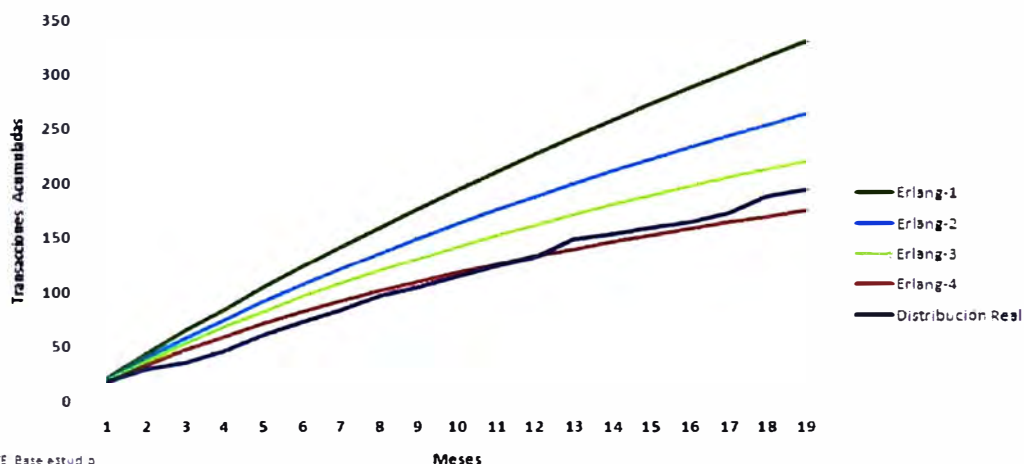
**Cuadro N° 5.9**  
**VALORES DE ESTADÍSTICO  $\chi^2$**   
**SEGÚN MODELO ERLANG-K**  
**PROPUESTO**

Erlang-k	Estadístico $\chi^2$
<b>1</b>	1046,9
<b>2</b>	363,1
<b>3</b>	112,6
<b>4</b>	18,9

FUENTE: Base estudio 2010  
 ELABORACIÓN: Propia

De manera gráfica también se puede llegar a la misma conclusión que en el párrafo anterior. En el Gráfico N° 5.4, donde se muestran las distribuciones pronosticadas de cada modelo Erlang-k y la distribución real, también se puede sacar algunas conclusiones.

**Gráfico N° 5.4**  
**DISTRIBUCIÓN DE TRANSACCIONES ACUMULADAS EN EL PERIODO DE VALIDACIÓN**  
**PARA CADA MODELO ERLANG-K COMPARADO CON DISTRIBUCIÓN REAL**



FUENTE: Base estudio  
 ELABORACIÓN: Propia

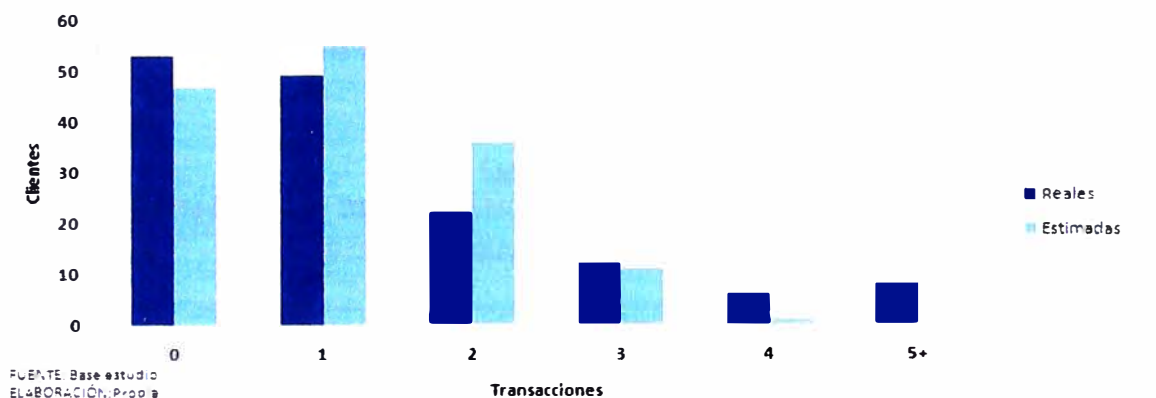
Año 2010

En dicho gráfico se puede apreciar las diferencias entre las transacciones pronosticadas de los modelos Erlang-1, Erlang-2, Erlang-3

y Erlang-4 respecto a la distribución acumulada real. Por otro lado, la distribución que pronostica el modelo Erlang-4, resulta ser el de mejor acoplamiento, casi superpuesta sobre la distribución real. Este hecho está respaldado por la prueba de bondad de ajuste chi-cuadrado.

Además es importante contrastar el poder de predicción del modelo, no solamente a nivel macro sino a nivel del cliente. Para ello, se puede comparar las transacciones reales del periodo de validación con las transacciones estimadas y tener otra medida de la efectividad del modelo Erlang-4. El Gráfico N° 5.5 muestra justamente lo que deseamos contrastar.

**Gráfico N° 5.5**  
**DISTRIBUCIÓN DE CANTIDAD DE CLIENTES POR NÚMERO**  
**DE TRANSACCIONES REALES Y ESTIMADAS EN PERIODO DE**  
**VALIDACIÓN PARA ERLANG-4**



Un hallazgo importante en el cálculo de la probabilidad de mantenerse activo en el siguiente periodo se puede extraer de la Cuadro N° 5.10. Es posible definir un criterio el cual delimite los niveles, o mejor dicho, los rangos de probabilidad que separen a los clientes fugados de los que no. Si se pone como límite de actividad que su probabilidad propiamente dicha sea mayor a 0.20, de la Cuadro N° 5.10 podemos concluir que el 27% de la población no tiene una probabilidad satisfactoria que indique una posible actividad en el siguiente periodo. Como referencia de fuga de clientes, la tasa de *churn* ('deserción') en USA y en

Europa giran alrededor del 32% (Talking Retail, 2008) en el mismo sector *retail*, a la cual pertenece la población.

**Cuadro N° 5.10**  
**PROBABILIDADES DE ACTIVIDAD EN EL SIGUIENTE**  
**PERIODO PARA MODELO ERLANG-4**

Rangos de Probabilidad	Proporción de Muestra
[0.00 -0.20]	27%
<0.20 -0.40]	9%
<0.40 -0.60]	14%
<0.60 -0.80]	17%
<0.80 -1.00]	33%

FUENTE: Base estudio

2010

ELABORACIÓN: Propia

Teniendo los resultados alcanzados hasta el momento podemos hondar más en la relación de estos con las variables disponibles de la población. En el Cuadro N° 5.11 tenemos la distribución de la cantidad de clientes por montos de compra en el periodo de validación.

**Cuadro N° 5.11**  
**DISTRIBUCIÓN DE FRECUENCIA DE LOS CLIENTES DE RETAIL SEGÚN**  
**MONTOS DE COMPRA EN EL PERIODO DE VALIDACIÓN**

Rango de Monto de Compra	Clientes	Porcentaje
<=100	70	21%
>100;200]	39	24%
>200;300]	16	15%
>300;400]	12	15%
>400;500]	5	8%
>500;800]	4	7%
>800;1,000]	2	4%
>1,000;1,500]	-	-
>1,500;2,000]	-	-
>2,000	2	6%
<b>Total</b>	<b>150</b>	<b>100%</b>

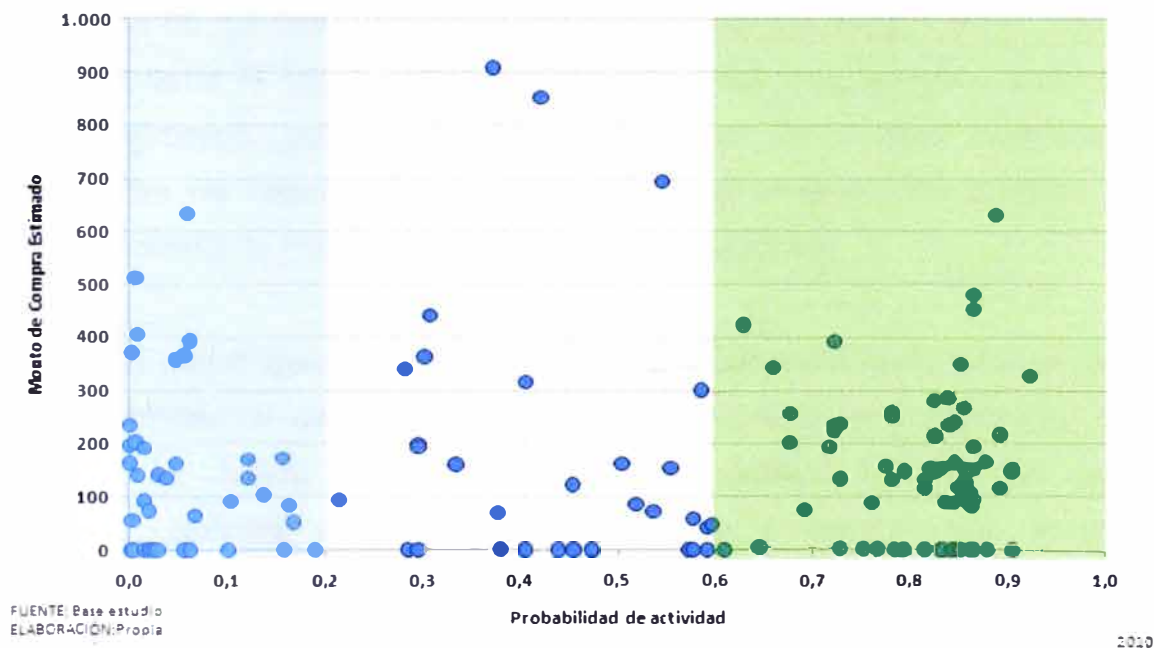
FUENTE: Base estudio

2010

ELABORACIÓN: Propia

En comparación del Cuadro N° 5.11 con Cuadro N° 5.7, se puede observar una disminución importante en los montos de compra. Esto ya se había observado en el Gráfico N° 5.2, dónde hay una caída importante en la cantidad de transacciones en el periodo de validación. Por otro lado, se debe recordar que contamos con las probabilidades de actividad en el periodo de validación. Lo siguiente que se realizará es analizar la relación entre esas probabilidades y los montos de compra del periodo de validación. Para ello se tiene el Gráfico N° 5.6 que cruza esas dos variables.

**Gráfico N° 5.6**  
**RELACIÓN ENTRE PROBABILIDAD DE ACTIVIDAD Y**  
**MONTO DE COMPRA TOTAL ESTIMADO DEL CLIENTE**



En él se ha hecho dos tipos de agrupaciones: la primera sobre los clientes con probabilidad de actividad menor al 0.20 y la segunda con los clientes con probabilidad de actividad mayor a 0.60. El primer grupo son los definidos como clientes en proceso de fuga o fugados. El segundo son los clientes que tienen altas probabilidades de seguir comprándole a la empresa, se les puede llamar fidelizados. Se puede notar que a nivel de montos de compras ambos grupos se comportan similarmente. Si se ahonda en esta comparación tenemos el Cuadro N° 5.12.

**Cuadro N° 5.12**  
**DISTRIBUCIÓN DE CLIENTES Y PROMEDIO DE COMPRA SEGÚN**  
**GRUPO DE FIDELIZACIÓN EN EL PERIODO DE VALIDACIÓN**

<b>Grupo</b>	<b>Clientes</b>	<b>Promedio de Compra</b>
Fugados	18	S/.147
Fidelizados	73	S/.146
<b>Total</b>	<b>91</b>	<b>S/.146</b>

FUENTE: Base estudio

2010

ELABORACIÓN: Propia

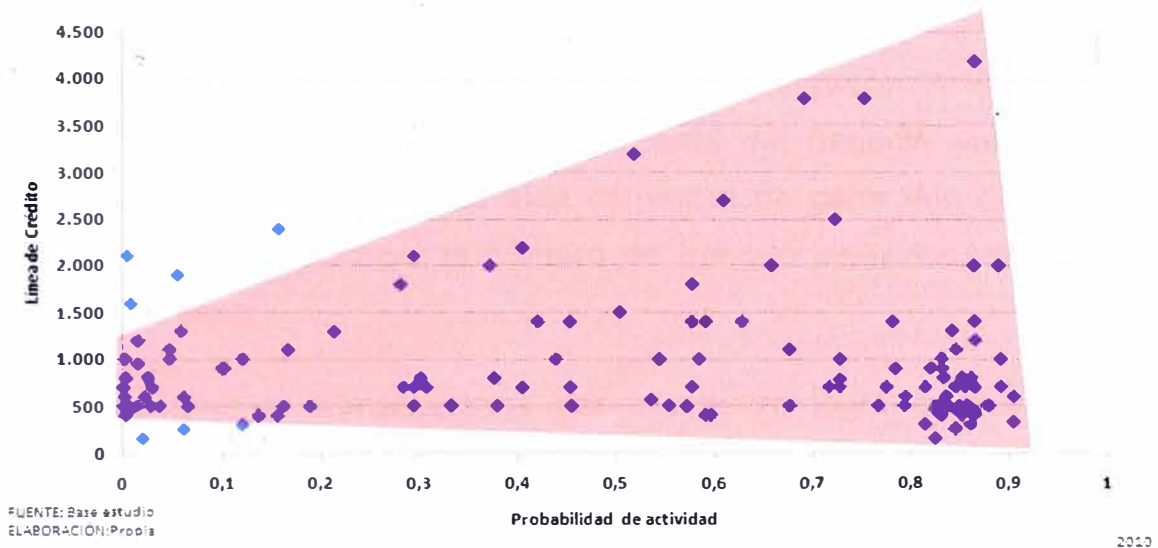
De dicho cuadro se deduce lo siguiente: los cliente fugados al tener un promedio de compra en el periodo de validación mayor al grupo de fidelizados y al promedio total son clientes con compras importantes, de alto valor y posiblemente estén siendo absorbidos por la competencia, a diferencia de los fidelizados que si bien, se han mantenido en el negocio, no se obtiene el retorno esperado de ellos. Es muy positivo que hayan sido fidelizados, pero el negocio no orientó de manera correcta las estrategias de retención o fidelización sobre el grupo de fugados que probablemente le hubieran traído mayores beneficios.

Un cruce igualmente interesante se encuentra en la relación entre la probabilidad de actividad con las líneas de crédito que les asignó a los clientes. El detalle está plasmado en el Gráfico N° 5.7. Se puede observar, gracias al área sombreada, que a medida que crece la probabilidad de actividad también lo hacen las líneas de crédito asignadas. Por ello podemos deducir que la oferta que le entregue el *retail* al cliente es un fuerte aliciente para él de permanecer con el negocio. Por otra parte los clientes con niveles de probabilidad menores a 0.20 están enmarcados en líneas de crédito más bajas. Hay una relación directamente proporcional entre línea de crédito y probabilidad de actividad.

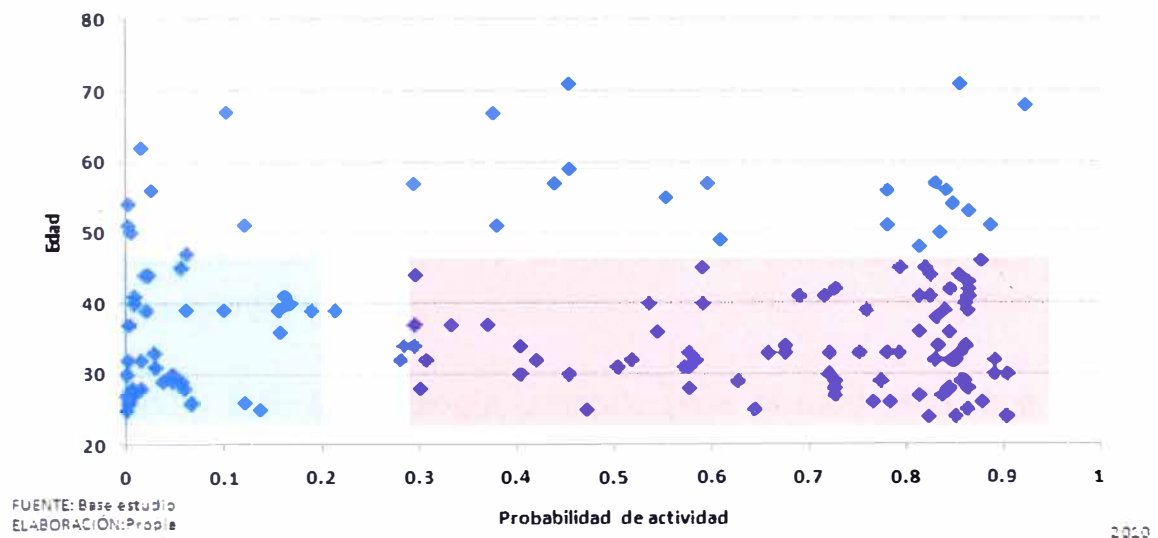
Otro cruce que llama la atención es el de la probabilidad de actividad con las edades de los clientes. Esto se ve reflejado en el Gráfico N°5.8. Se puede observar que la proporción de clientes entre edades de

20 y 45 con probabilidad de actividad mayor a 0.3 es mayor a la proporción de clientes con riesgo a fuga, es decir, con probabilidad menor a 0.20.

**Gráfico N° 5.7**  
**RELACIÓN ENTRE PROBABILIDAD DE ACTIVIDAD Y LÍNEA DE CRÉDITO ASIGNADA AL CLIENTE**



**Gráfico N° 5.8**  
**RELACIÓN ENTRE PROBABILIDAD DE ACTIVIDAD Y EDAD DEL CLIENTE**



Otros cruces con las variables disponibles por cada cliente se encuentran en la sección de Anexos.



## CONCLUSIONES

1. Se alcanzó el objetivo principal de esta investigación. A través de las transacciones históricas de cada cliente del negocio analizado, se logró estimar el tiempo de vida comercial de cada uno de ellos, es decir, se logró estimar el número de transacciones futuras de cada uno de ellos.
2. El número de transacciones por cliente se logró estimar para el periodo de validación. Estos resultados fueron validados por la prueba de bondad de ajuste chi-cuadrado, la cual pasó satisfactoriamente en el análisis realizado.
3. Asimismo se estimó la probabilidad de actividad para el periodo de validación del modelo. Esta probabilidad está evaluada a la par con el número de transacciones estimadas para el mismo periodo por la prueba de chi-cuadrado, que pasó satisfactoriamente.
4. Para cada cliente se estimó los montos de compra, apoyándose en el número de transacciones estimadas para el periodo de validación y en conjunto con el monto promedio de compras del periodo de modelamiento.
5. Respecto a la metodología utilizada para el modelado, que gira en torno a la distribución Erlang, se comprueba su fortaleza en términos de predicción, aprobando satisfactoriamente la prueba de bondad de ajuste  $\chi^2$ , para el caso del modelo Erlang-4. Asimismo la viabilidad de calcular la probabilidad de que un cliente este activo en el

siguiente periodo, se convierte en una poderosa herramienta de segmentación y gestión.

6. En términos de estabilidad, el modelo Erlang-4 propuesto entrega resultados con una menor varianza, generando resultados más robustos.
7. El modelo de estimación del tiempo de vida comercial, se convierte así, en una herramienta importante de apoyo a la hora de decidir cuáles de los clientes de la cartera poseen una mayor tendencia a la fuga.
8. Al identificar de mejor forma a estos clientes es posible: (1) aumentar las utilidades y la rentabilidad del negocio, ya que da la posibilidad de retener clientes en fuga y mantenerlos como clientes captando sus flujos futuros; (2) focalizar los recursos productivos sobre los segmentos que realmente necesitan y merecen acciones de retención fuertes aumentando la eficiencia y efectividad de las políticas de fidelización; (3) generar un mejoramiento y fortalecimiento de la relación entre el cliente y la empresa, al anticiparse a sus necesidades, así como hacer menos sensible al cliente frente a las campañas de marketing de la competencia.
9. De los resultados observados, se pudo concluir que la empresa no orientó de manera correcta sus estrategias de marketing y por ello un grupo de clientes importantes fugaron o están en proceso de fuga.
10. Las líneas de crédito han sido un factor importante en la determinación de la fidelidad de los clientes y esto se ve reflejado en la probabilidad de actividad.

## RECOMENDACIONES

1. Al ser este un modelo de predicción, es susceptible a evoluciones en el estilo de vida y de comportamiento de compra de los clientes. Por ello, se sugiere una actualización de los parámetros cada 3 meses.
2. El negocio debe reconsiderar sus políticas de riesgo ya que al otorgarle al cliente una oferta de línea de crédito poco atractiva, es probable que no se mantenga mucho tiempo con el negocio y termine perteneciendo al grupo de clientes fugados.
3. Deben realizarse pruebas piloto para evaluar la modificación o no en la relación entre las líneas de crédito y las probabilidades de actividad en el periodo siguiente.
4. Tratándose de un sector asociado al sector financiero debe incluirse en la segmentación la relación de la probabilidad de fuga con la situación crediticia en el sistema.
5. Al haberse trabajado de manera general las transacciones, es decir, transacciones sin especificar algún tipo de producto, puede estar pasándose información valiosa. Se invita a ingresar al detalle de cada tipo de transacción que puede realizar cada cliente, para construir un modelo específico para cada tipo de producto.
6. De querer trabajar el modelo a nivel de cartera, se requerirían recursos de cómputo mayor por la complejidad de los cálculos que involucra principalmente la estimación de parámetros.

## BIBLIOGRAFÍA

- Abad, R. (2011). Marketing Relacional, futuro relacional, [en línea]. Microsoft Centro para Empresas y Profesionales: Raúl Abad, Consultor de Marketing Online de Abad Marketing. Dirección URL: <[http://www.microsoft.com/business/smb/es-es/marketing/marketing\\_relacional.msp](http://www.microsoft.com/business/smb/es-es/marketing/marketing_relacional.msp)>. [Consulta: 3 enero 2011].
- Batistam, E.P., Denizel M., and Filiztekin A. (2007). "Empirical validation and comparison of models for customer base analysis". International Journal of Research in Marketing, 24(3), pp. 201-209.
- Bowen, J. y Kurey, M. (2006). Shattering Illusions of Loyalty, [en línea]. Onwallstreet. Dirección URL: <<http://www.onwallstreet.com/news/shattering-illusions-loyalty-529073-1.html?zkPrintable=true>>. [Consulta: 17 febrero 2011].
- Begoña, V., Portabales, Y. y de la Fuente, J. (2001) – Master Executive en Marketing Relacional, CRM y Comercio Electrónico ICEMD – <http://www.icemd.com/>
- Buckinx W., Van den Poel D. (2005) Customer base analysis: partial detection of behaviorally loyal clients in a noncontractual FMCG retail setting. European Journal of Operational Research, 164(1), pp. 252-268.
- Chatfield, C. and Goodhardt, G.J. (1973). "A Consumer Purchasing Model with Erlang Inter-Purchase Time". Journal of the American Statistical Association, 68(344), pp. 828-835.
- Dunn, R., Reader, S., and Wrigley, N. (1983). "An Investigation of the Assumptions of the NBD Model as Applied to Purchasing at Individual Stores". Applied Statistics, 32(3), pp. 249-259.
- Ehrenberg, A. (1959). "The Pattern of Consumer Purchases". Journal of the Royal Statistical Society. Series C (Applied Statistics), 8(1), pp. 26-41.

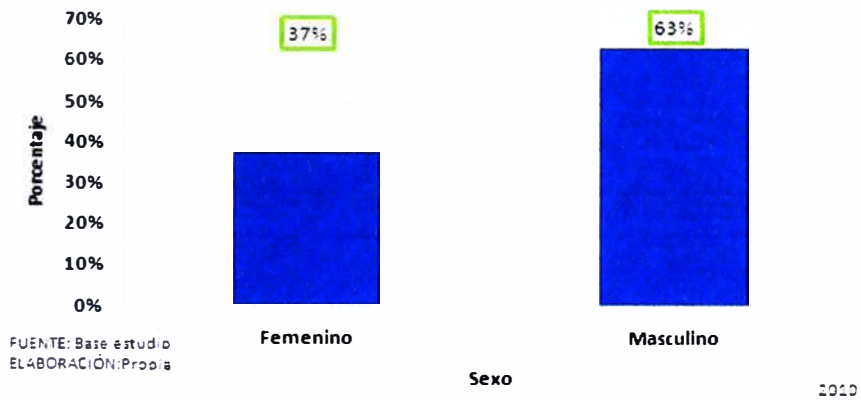
- Fader, P., Hardie B., and Lee, K.L. (2005). Counting Your Customers the Easy Way: An Alternative to the Pareto/NBD Model. *Marketing Science*, 24, pp. 275-284,
- Haight, F.A. (1965). "Counting distributions for renewal processes". *Biometrika*, 52(3-4), pp. 395-403.
- Herniter, J. (1971). "A Probabilistic Market Model of Purchase Timing and Brand Selection". *Management Science*, 18(4), pp. 102-112.
- Hoekstra J., Huizingh E. (1999). The Lifetime Value Concept in Customer-Based Marketing. *Journal of Market Focused Management*, 3, 257-274. DOI: 10.1023/A:1009842805871.
- Hoppe, D. and Wagner U. (2007). Customer Base Analysis: The Case for a Central Variant of the Betageometric/NBD Model. *Marketing - Journal of Research and Management*, 2, pp. 75-90.
- Hwang H., Jung T., Suh E. (2004). An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry. *Science Direct Expert Systems with Applications*, 26(2), pp. 181-188.
- Metropolis International Group Ltd (2008). Supermarket shoppers more likely to switch as home delivery trumps location. Dirección URL: <http://www.talkingretail.com/news/industry-news/supermarket-shoppers-more-likely-to-switch-as-home-delivery-trumps-location> [Consulta: 17 febrero 2011].
- Marker, J. (1998). Studying Policy Retention Rates Using Markov Chains. *Casualty actuarial society* v85.
- Morrison, D.G. and Schmittlein, D.C.(1988). "Generalizing the NBD Model for Customer Purchases: What Are the Implications and Is It Worth the Effort?". *Journal of Business and Economic Statistics*, 6(2).
- Mutanen, T. (2006). "Customer churn analysis". VTT Business from Technology.
- Pfeifer, P., Carraway, R. (2000). Modeling Customer Relationships as Markov Chains. *Journal of interactive marketing* v14, 2, pp. 43-55.
- Platzer, M. (2008). "Stochastic Models of Noncontractual Consumer Relationships". Master Thesis at the Vienna University of Economics and Business Administration.
- Rodríguez, I., Collado, J. y Herrero, A. (2001). La lealtad del consumidor y el Marketing relacional. Una aproximación empírica. Ponencia Presentada al XI Congreso Nacional de ACEDE (Asociación Científica de Economía y Dirección de la Empresa), Septiembre.

- Segovia, C., Aburto, L. y Goic M., (2005). Caracterización del proceso de fuga de clientes de un Retail Banking utilizando información transaccional. VI Congreso Chileno de Investigación Operativa Óptima 2005.
- Schmittlein, D.C. and Morrison, D.G. (1983). "Prediction of Future Random Events With the Condensed Negative Binomial Distribution". *Journal of the American Statistical Association*, 78(382), pp. 449-456.
- Schmittlein D.C., Morrison D.G., and Colombo, R. (1987). Counting your customers: who are they and what will they do next? *Management Science*, 33(1), pp. 1-24.
- Wheat, R.D. and Morrison, D.G. (1990). "Estimating Purchase Regularity with Two Interpurchase Times". *Journal of Marketing Research*, 27(1), pp. 87-93.
- Wagner, U. and Hoppe, D. (2008). "Erratum on the MBG/NBD Model". *International Journal of Research in Marketing*.
- Wagner, U. and Taudes, A. (1986). "A Multivariate Polya Model of Brand Choice and Purchase Incidence". *Marketing Science*, 5(3):219-244.
- Wusst, C. (2002). "La lealtad de los Clientes y su Medición". *ITAM Segmento*, N°16 Año 3.
- Wübben, M. and Von Wangenheim, F. (2008). Instant Customer Base Analysis: Managerial Heuristics Often "Get It Right". *Journal of Marketing*, 72, pp. 82-93.

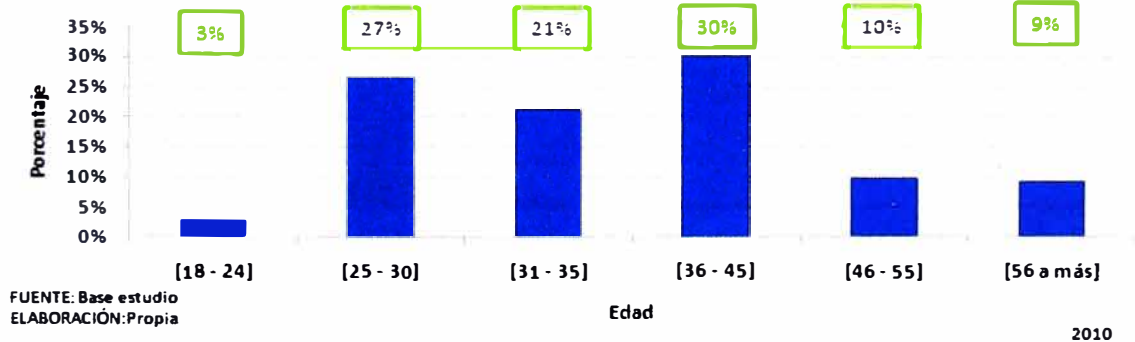
## ANEXOS

### Anexo Estadístico

#### Anexo N° 1 DISTRIBUCIÓN DE CLIENTES POR TIPO DE SEXO

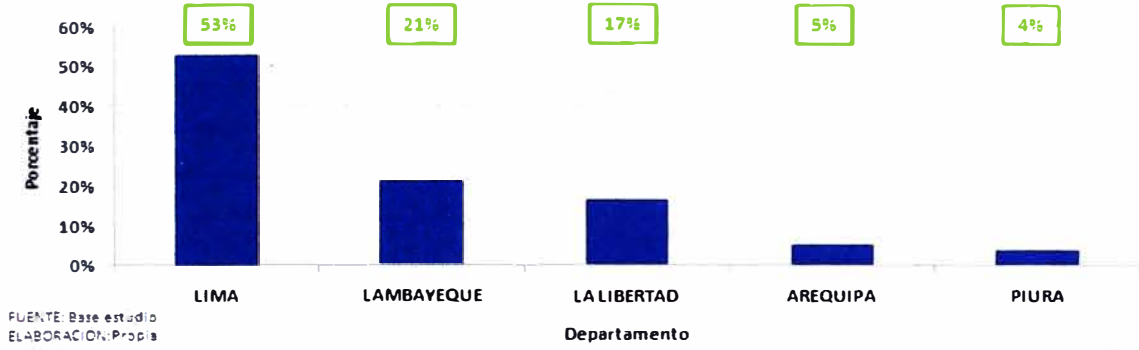


#### Anexo N° 2 DISTRIBUCIÓN DE CLIENTES POR RANGO DE EDAD



### Anexo N° 3

#### DISTRIBUCIÓN DE CLIENTES POR LUGAR DE RESIDENCIA



2010

### Anexo N° 4

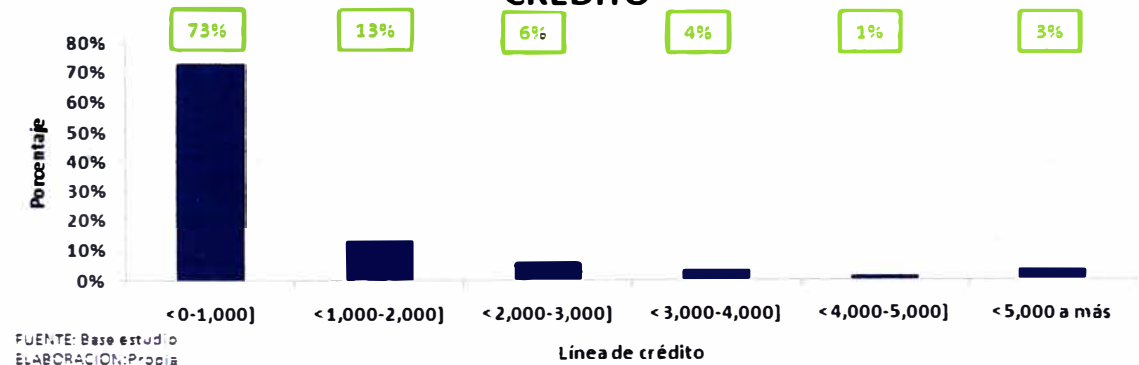
#### DISTRIBUCIÓN DE CLIENTES POR TIPO DE CRÉDITO



2010

### Anexo N° 5

#### DISTRIBUCIÓN DE CLIENTES POR RANGO DE LÍNEA DE CRÉDITO



2010



**Anexo N° 6**  
**DISTRIBUCIÓN DE CLIENTES SEGÚN RANGO DE PROBABILIDAD**  
**Y SEXO**

Rango de Probabilidad	Sexo		Total
	Masculino	Femenino	
[0.00 -0.20]	27	14	41
<0.20 -0.40]	11	2	13
<0.40 -0.60]	15	6	21
<0.60 -0.80]	15	10	25
<0.80 -1.00]	26	24	50
<b>Total</b>	<b>94</b>	<b>56</b>	<b>150</b>

FUENTE: Base estudio

2010

ELABORACIÓN: Propia

**Anexo N° 7**  
**DISTRIBUCIÓN DE CLIENTES SEGÚN RANGO DE PROBABILIDAD Y SEXO**

Rango de Probabilidad	Departamento					Total
	Lima	La Libertad	Lambayeque	Arequipa	Piura	
[0.00 -0.20]	27	3	7	3	1	41
<0.20 -0.40]	9	1	3	-	-	13
<0.40 -0.60]	9	2	7	2	1	21
<0.60 -0.80]	12	8	2	1	2	25
<0.80 -1.00]	22	11	13	2	2	50
<b>Total</b>	<b>79</b>	<b>25</b>	<b>32</b>	<b>8</b>	<b>6</b>	<b>150</b>

FUENTE: Base estudio

2010

ELABORACIÓN: Propia