

ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE CIENCIAS

APLICACIÓN A LA EDUCACIÓN SUPERIOR DEL ECUADOR DE
UN TEOREMA DE LÍMITE CENTRAL PARA MARTINGALAS

PROYECTO PREVIO A LA OBTENCIÓN DEL TÍTULO DE INGENIERÍA
MATEMÁTICA

PAULA MONSERRATTE CASTRO CASTRO
paula.mc4@gmail.com

Director: LUIS ALCIDES HORNA HUARACA
luis.horna@epn.edu.ec

QUITO, OCTUBRE 2014

DECLARACIÓN

Yo PAULA MONSERRATTE CASTRO CASTRO, declaro bajo juramento que el trabajo aquí escrito es de mi autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y que he consultado las referencias bibliográficas que se incluyen en este documento.

A través de la presente declaración cedo mis derechos de propiedad intelectual, correspondientes a este trabajo, a la Escuela Politécnica Nacional, según lo establecido por la Ley de Propiedad Intelectual, por su reglamento y por la normatividad institucional vigente.

Paula Monserratte Castro Castro

CERTIFICACIÓN

Certifico que el presente trabajo fue desarrollado por PAULA MONSERRATTE CASTRO CASTRO, bajo mi supervisión

Luis Alcides Horna Huaraca
Director del Proyecto

AGRADECIMIENTOS

Agradezco a todas las personas que me acompañaron durante la elaboración de este proyecto, a aquellos que me conocen hace tanto tiempo que sabían cuando necesitaba ayuda y a aquellos que casi sin conocerme me la ofrecieron.

Agradezco además a todos los docentes que contribuyeron en mi formación profesional, especialmente al Dr. Luis Horna por su colaboración pero sobre todo por su interés.

DEDICATORIA

A mi familia, por todo su apoyo.
A Carlitos, por su invaluable ayuda.

Paula

Índice de contenido

Índice de figuras	2
Índice de tablas	4
Resumen	5
Abstract	6
1 Introducción	1
Introducción	1
1.1 Nociones preliminares	1
1.1.1 Introducción a la teoría de la medida	1
1.1.2 Introducción a la teoría de las probabilidades	3
1.1.3 Momentos de una variable aleatoria	8
1.1.4 Vectores aleatorios	12
1.1.5 Momentos de un vector aleatorio	14
1.1.6 Convergencia de variables aleatorias	18
2 Ligera historia y desarrollo del Teorema del Límite Central	24
Ligera historia y desarrollo del Teorema del Límite Central	24
2.1 Reseña histórica	24
2.2 Adaptaciones del Teorema del Límite Central	27
2.2.1 Procesos de renovación	27
2.2.1.1 Aplicación de los procesos de renovación	29
2.2.2 Procesos de Wiener	31
2.2.3 Estadístico de Wilcoxon y su aproximación a la distribución normal estándar	32
3 Martingalas y el Teorema del Límite Central	35

Martingalas y el Teorema del Límite Central	35
3.1 Conceptos previos a la definición de las martingalas	35
3.1.1 Esperanza condicional	35
3.1.2 Tiempo de parada	38
3.1.3 Identidad de Wald	42
3.2 Martingalas	53
3.2.1 Origen de las martingalas	53
3.2.2 Definición y primeros resultados	53
3.3 Teorema del Límite Central para Martingalas	56
3.3.1 Ley de los grandes números	56
3.3.2 Teorema del Límite Central	58
4 Aplicación	60
Aplicación	60
4.1 Análisis descriptivo	61
4.2 Aplicación TLC	68
4.2.1 Interpretación de resultados.	75
5 Conclusiones y Recomendaciones	83
Conclusiones y Recomendaciones	83
5.1 Conclusiones	83
5.2 Recomendaciones	84
Anexo A	85
Anexo B	92
Anexo C	97

Índice de figuras

1.1	Función de densidad de X_1	21
1.2	Función de densidad de $X_1 + X_2$	21
1.3	Función de densidad de $X_1 + X_2 + X_3$	22
1.4	Función de densidad χ^2 con 15, 20 y 30 grados de libertad.	23
3.1	Gráfica Ruina Jugador - Pérdida	41
3.2	Gráfica Ruina Jugador - Ganancia	41
4.1	Número de titulados por área del conocimiento	61
4.2	Años promedio de estudio por provincia	62
4.3	Tiempo promedio de estudio hasta la titulación por provincia	63
4.4	Tiempo promedio de estudio por área - Azuay	64
4.5	Tiempo promedio de estudio por área - Guayas	64
4.6	Tiempo promedio de estudio por área - Pichincha	64
4.7	Tiempo promedio de estudio hasta la titulación por área y provincia	65
4.8	Diferencia titulación real y titulación oficial por provincia	66
4.9	Diferencia en años entre titulación real y titulación oficial	67
4.10	Esquema variable Y_i	68
4.11	Frecuencias de las variables tiempo real y tiempo oficial de estudio	70
4.12	Tabla de contingencia Tiempo real vs Tiempo oficial	71
4.13	Valores obtenidos para la variable X_n	72
4.14	Frecuencias de la variable r_i^2	74
4.15	Valores obtenidos para la esperanza condicional	74
4.16	Valores obtenidos para $E(r_i^2 T_i = t)$ en ciencias.	76
4.17	Valores obtenidos para $E(r_i^2 T_i = t)$ en educación.	77
4.18	Valores obtenidos para $E(r_i^2 T_i = t)$ en humanidades y artes.	77
4.19	Valores obtenidos para $E(r_i^2 T_i = t)$ en ciencias sociales.	77
4.20	Valores obtenidos para $E(r_i^2 T_i = t)$ en ingeniería.	78
4.21	Valores obtenidos para $E(r_i^2 T_i = t)$ en agricultura.	78
4.22	Valores obtenidos para $E(r_i^2 T_i = t)$ en salud y servicios sociales.	78

4.23	Valores obtenidos para $E(r_i^2 T_i = t)$ en servicios.	78
5.1	Número y porcentaje de titulados por área del conocimiento	92
5.2	Años promedio de estudio por área del conocimiento y provincia	93
5.3	Años promedio de diferencia en años entre duración oficial de la carrera y tiempo real de estudios	94
5.4	Probabilidad de la intersección entre Tiempo real y Tiempo oficial	95
5.5	Duración oficial de estudios por área del conocimiento	96
5.6	Duración oficial de estudios por universidad	96
5.7	Años promedio de estudio por provincia - carreras 5 años	98
5.8	Tiempo promedio de estudio por provincia - carreras 5 años	99
5.9	Años promedio de estudio por área y provincia - carreras 5 años	100
5.10	Tiempo promedio de estudio por área y provincia - carreras 5 años	101
5.11	Diferencia entre titulación real y oficial por provincia - carreras 5 años	102
5.12	Diferencia promedio entre tiempo real y oficial por provincia - carreras 5 años	103

Índice de cuadros

1.1	Valores obtenidos en la realización de la encuesta	4
3.1	Algoritmo tiempo de paro en el problema ruina del jugador	40
3.2	Parámetros utilizados simulación tiempo de parada	40
3.3	Ruina del jugador - Pérdida	40
3.4	Ruina del jugador - Ganancia	41
3.5	Resultado simulaciones Ganancia - Tiempo de Paro	42
3.6	Frecuencia número de accidentes	46
3.7	Parámetros distribución binomial negativa	46
3.8	Frecuencias teóricas del número de accidentes	47
3.9	Frecuencias teóricas del número de accidentes agrupadas	47
3.10	Prueba bondad de ajuste	47
3.11	Fractil distribución χ^2	47
3.12	Cuantía de los accidentes	48
3.13	Parámetros distribución log –normal	49
3.14	Frecuencias teóricas de la cuantía de los accidentes	49
3.15	Frecuencias teóricas de la cuantía agrupadas	50
3.16	Prueba bondad de ajuste	50
3.17	Fractil distribución χ^2	50
3.18	Distribución de la cuantía de los accidentes	51
3.19	Probabilidades del número de accidentes - algoritmo	51
3.20	Probabilidad del número de accidentes	52
3.21	Distribución acumulada de S	52
3.22	Distribución de S	52
4.1	Clasificación CINE - UNESCO 1997	61
4.2	Resumen carreras Facultad de Ciencias	68
4.3	Algoritmo construcción Esperanza Condicional	72
4.4	Valor de c ciencias	77
4.5	Valor de c educación	77

4.6	Valor de c humanidades y artes	77
4.7	Valor de c ciencias sociales	77
4.8	Valor de c ingeniería	78
4.9	Valor de c agricultura	78
4.10	Valor de c salud	78
4.11	Valor de c servicios	79
4.12	Varianza y desviación estándar por área	79
4.13	Tiempo estudios por área	79
4.14	Varianza y desviación estándar por universidad	80
4.15	Tiempo de estudio por universidad	81
4.16	Varianza de X_i por universidad	81

Resumen

Aplicaciones concretas del concepto de martingala, así como de sus teoremas límites han sido poco desarrolladas, el presente trabajo tiene como objetivo usar el Teorema del Límite Central en martingalas y estudiar el comportamiento de la variable titulación en carreras de tercer nivel del sistema de educación superior del país. Mediante un correcto planteamiento de los datos se obtiene una martingala que representa el tiempo real esperado de estudios, dado el tiempo de duración oficial de la carrera de cada estudiante. Una vez aplicado el TLC se obtiene la distribución, media y varianza de la variable, esta última resulta de particular interés ya que puede ser vista como la brecha entre el tiempo que toma titularse y lo que realmente debería tomar y puede ser utilizada para ranquear grupos de titulados, teniendo así el potencial de convertirse en un indicador de calidad en la gestión de las universidades en cuanto a la titulación de sus estudiantes.

Abstract

Specific applications of martingales and their limit theorems have been poorly developed, this paper's aim is to use the martingale's Central Limit Theorem and to study the behavior of chosen variable, titling in tertiary careers of higher education system in Ecuador. Through a correct approach to the data is obtained a martingale that represents the real-time expected of studies given the official-time of career's duration of each student. Distribution, mean and variance of the variable are obtained with the implementation of CLT, the last one results interesting in our study because it can be seen as the gap between time taken in titling and time really should be employed. It can be used to classify groups of graduates, thus it has the potential to become a quality indicator in the management of universities as regards the titling of their students.

Capítulo 1

Introducción

1.1 Nociones preliminares

Antes de iniciar el estudio del Teorema del Límite Central, TLC, se introducirán de manera breve ciertas definiciones, así como los teoremas más importantes necesarios para una mejor comprensión del mismo.

1.1.1 Introducción a la teoría de la medida

Definición 1 (Álgebra). Sea Ω un conjunto, un álgebra sobre Ω es un conjunto $\mathcal{A} \subset \mathcal{P}(\Omega)$, con $\mathcal{P}(\Omega)$ el conjunto de partes de Ω , tal que:

1. $\emptyset \in \mathcal{A}$.
2. Si $A \in \mathcal{A} \Rightarrow A^C \in \mathcal{A}$.
3. La unión finita de elementos del álgebra pertenece al álgebra, es decir,

$$A, B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}.$$

Definición 2 (σ -Álgebra). Sea Ω un conjunto, un conjunto $\mathcal{A} \subset \mathcal{P}(\Omega)$ se denomina σ -álgebra sobre Ω si y solo si

1. $\emptyset \in \mathcal{A}$.
2. Si $A \in \mathcal{A} \Rightarrow A^C \in \mathcal{A}$.
3. La unión numerable de elementos de \mathcal{A} pertenece al conjunto, es decir

$$\forall i \in I \subset \mathbb{N}. \text{ Si } A_i \in \mathcal{A} \Rightarrow \bigcup_{i \in I} A_i \in \mathcal{A}.$$

Observación 1 (Importancia de la σ -álgebra). *Dentro de un conjunto Ω dotado con una σ -álgebra se definen medidas. En el caso particular de las medidas de probabilidad, la σ -álgebra determina los subconjuntos de Ω que conforman todos los eventos posibles en la realización de un experimento aleatorio.*

Sea Ω un conjunto no vacío, \mathcal{A} una σ -álgebra sobre este conjunto. La dupla (Ω, \mathcal{A}) se denomina espacio medible.

Sea $C \subset \mathcal{P}(\Omega)$ un subconjunto cualquiera, se define la σ -álgebra generada por C , notada $\sigma(C)$, como la intersección de todas las σ -álgebras que contienen a dicho conjunto,

$$\sigma(C) = \bigcap_{i \in I} \mathcal{A}_i$$

donde $\{\mathcal{A}_i\}_{i \in I}$ representa el conjunto de todas las σ -álgebras que contienen a C . Nótese que $\sigma(C)$ es la menor σ -álgebra que contiene a C .

Observación 2 (σ -álgebra de Borel). *Sea $\Omega = \mathbb{R}$ provisto de la topología de las colas derechas*

$$\tau = \{[a, +\infty[\mid a \in \mathbb{R}\} \cup \{\emptyset, \mathbb{R}\}$$

la σ -álgebra generada por τ recibe el nombre de σ -álgebra de Borel en \mathbb{R} , se nota mediante $\mathcal{B}(\mathbb{R})$, y cumple

$$\begin{aligned} \mathcal{B}(\mathbb{R}) &= \sigma(\tau) \\ &= \sigma(\{[a, b[\mid a, b \in \mathbb{R}\}) \\ &= \sigma(\{]-\infty, c] \mid c \in \mathbb{R}\}) \end{aligned}$$

De manera general, sea Ω un conjunto no vacío cualquiera provisto de una topología, la σ -álgebra generada por los abiertos de Ω recibe el nombre de σ -álgebra de Borel.

Definición 3 (Medida). *Sea (Ω, \mathcal{A}) un espacio medible, una función $u : \mathcal{A} \rightarrow \mathbb{R}^+$ que verifique las siguientes propiedades se denomina medida sobre (Ω, \mathcal{A}) .*

1. $u(\emptyset) = 0$.
2. Sea $\{A_n\}_{n \in \mathbb{N}}$ una sucesión de elementos de \mathcal{A} . Si los A_n son disjuntos 2 a 2, se tiene que

$$u\left(\bigcup_{n \geq 1} A_n\right) = \sum_{n \geq 1} u(A_n)$$

Esta propiedad es conocida como σ -aditividad.

- La tripleta (Ω, \mathcal{A}, u) se denomina espacio medido.

- $u(\Omega)$ se conoce como masa total de la medida.
- Si $u(\Omega) < +\infty$, la medida es finita.
- Si $u(\Omega)=1$, la medida recibe el nombre de probabilidad.
- $A \in \mathcal{A}$ es de medida nula si $u(A) = 0$.
- Una propiedad es verdadera casi en todas partes, si el conjunto donde no se verifica es de medida nula.

1.1.2 Introducción a la teoría de las probabilidades

Definición 4 (Evento aleatorio). *Se denomina evento aleatorio o no determinístico a todo fenómeno cuyo resultado es incierto, es decir, así se replique bajo las mismas condiciones el resultado obtenido no es el mismo necesariamente.*

Definición 5 (Espacio muestral). *Se denomina espacio muestral al conjunto que contiene todos los resultados de un experimento aleatorio. Se notará al espacio muestral por Ω .*

Definición 6 (Evento). *Sea \mathcal{A} una σ -álgebra de Ω . Un conjunto $A \in \mathcal{A}$ se denomina evento. Se dice que un evento ocurre si $\omega \in \Omega$, resultado de un experimento aleatorio, es tal que $\omega \in A$.*

Definición 7 (Probabilidad). *Sea (Ω, \mathcal{A}) un espacio medible, se define una probabilidad como una medida de masa 1, es decir $P : \Omega \rightarrow \mathbb{R}^+$ tal que:*

1. $P(A) \geq 0$ para todo $A \in \mathcal{A}$.
2. $P(\Omega) = 1$.
3. Sean A_1, A_2, \dots eventos disjuntos, entonces

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

La tripleta (Ω, \mathcal{A}, P) se conoce como espacio probabilístico.

Nota 1. *Siempre se debe tener en cuenta que:*

$$\forall A \in \mathcal{A}, \quad P(A) + P(A^C) = 1,$$

propiedad sencilla pero frecuentemente usada, constituyéndola en un “truco” probabilístico.

Teorema 1 (Ley de probabilidad total). Sean A_1, A_2, \dots, A_n elementos de Ω , que forman una partición del espacio y $B \subset \Omega$. Entonces,

$$P(B) = \sum_{i=1}^n P(B|A_i)P(A_i)$$

Este teorema muy sencillo es de suma importancia en la obtención de resultados útiles relacionados con la probabilidad condicional y la esperanza matemática condicional. Ha sido especialmente aplicado en la elaboración de encuestas que incluyen preguntas que pueden ser consideradas sensibles, el siguiente ejemplo tomado de [33] ilustra esta aplicación:

Ejemplo 1. Se realiza un estudio sobre consumo de drogas en una determinada institución. Se elige 100 personas al azar y para garantizar confidencialidad se diseña la siguiente estrategia:

Cada persona extrae una bola numerada de una urna.

- *Si el número de la bola está entre 1 y 70 se contesta a la pregunta, ¿ha consumido droga alguna vez?*
- *Si el número es mayor a 70, se pide a la persona que arroje una moneda y que conteste a la pregunta, ¿al lanzar la moneda obtuvo cara?*

En ambos casos se escribe la respuesta “sí” o “no” sin precisar a cual de las dos preguntas dieron respuesta; una vez finalizado el proceso se contabilizan 25 respuestas positivas. Aplicando el Teorema de Probabilidad Total a la probabilidad de respuestas afirmativas se obtiene lo siguiente:

$$P(\text{si}) = P(\text{si}|\text{pregunta delicada})P(\text{pregunta delicada}) + P(\text{si}|\text{pregunta intrascendente})P(\text{pregunta intrascendente})$$

de la realización del experimento se cuenta con la siguiente información:

Probabilidad	Valor
$P(\text{si})$	0,25
$P(\text{pregunta delicada})$	0,7
$P(\text{pregunta intrascendente})$	0,3
$P(\text{si} \text{pregunta intrascendente})$	0,5

Cuadro 1.1: Valores obtenidos en la realización de la encuesta

Reemplazando estos valores

$$0,25 = P(\text{si}|\text{pregunta delicada})0,70 + 0,5 * 0,3$$

Así

$$P(\text{si|pregunta delicada}) \approx 0,1428$$

Se concluye que de las 70 personas que contestaron la pregunta de interés, el 14 % lo hicieron positivamente, es decir 9 personas afirman haber consumido droga alguna vez.

Definición 8 (Variable aleatoria). Sea (Ω, \mathcal{A}, P) espacio probabilístico. Se denomina variable aleatoria a una función $X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ medible, que verifica

$$P_X(B) = P(X^{-1}(B)), \quad B \in \mathcal{B}(\mathbb{R})$$

Es fácil ver que $P_X : \mathcal{B}(\mathbb{R}) \rightarrow \mathbb{R}^+$ es una medida de probabilidad respecto a $\mathcal{B}(\mathbb{R})$, en efecto se tiene que:

1. $P_X(\emptyset) = 0$. En efecto,

$$P_X(\emptyset) = P(X^{-1}(\emptyset)) = P(\emptyset) = 0.$$

2. $P_X(\mathbb{R}) = 1$. En efecto,

$$P_X(\mathbb{R}) = P(X^{-1}(\mathbb{R})) = P(\Omega) = 1.$$

3. Sean $\{B_i\}_{i \in \mathbb{N}}$ tal que $\forall i \in \mathbb{N}, B_i \in \mathcal{B}(\mathbb{R})$ disjuntos dos a dos, entonces

$$P_X\left(\bigcup_{i \in \mathbb{N}} B_i\right) = \sum_{i \in \mathbb{N}} P_X(B_i)$$

La igualdad se sigue ya que,

$$\begin{aligned} P_X\left(\bigcup_{i \in \mathbb{N}} B_i\right) &= P\left(X^{-1}\left(\bigcup_{i \in \mathbb{N}} B_i\right)\right) \\ &= P\left(\bigcup_{i \in \mathbb{N}} X^{-1}(B_i)\right) \\ &= \sum_{i \in \mathbb{N}} P(X^{-1}(B_i)) \\ &= \sum_{i \in \mathbb{N}} P_X(B_i) \end{aligned}$$

La tercera igualdad resulta ya que al ser los B_i disjuntos dos a dos por hipótesis, los $X^{-1}(B_i)$ también lo son, en efecto, para demostrarlo se supone lo contrario, es decir

$$\exists z \in X^{-1}(B_i) \cap X^{-1}(B_j) \text{ con } i \neq j,$$

de donde

$$\begin{aligned}z &\in X^{-1}(B_i) \text{ y } z \in X^{-1}(B_j) \\X(z) &\in B_i \text{ y } X(z) \in B_j \\X(z) &\in B_i \cap B_j\end{aligned}$$

lo que contradice la hipótesis inicial, por tanto se concluye el resultado.

Definición 9 (σ -álgebra generada por una variable aleatoria). *Al ser*

$$X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$

variable aleatoria una función medible, la σ -álgebra generada por X , notada por $\sigma(X)$, está dada por

$$\sigma(X) = \{X^{-1}(A') \in \mathcal{A}, A' \in \mathcal{B}(\mathbb{R})\}.$$

Sean ahora $\{X_i\}_{i \in I}$ una familia de variables aleatorias de (Ω, \mathcal{A}) en $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, la σ -álgebra generada por la familia $\{X_i\}_{i \in I}$ está dada por

$$\sigma(X_i, i \in I) = \sigma\left(\bigcup_{i \in I} \sigma(X_i)\right)$$

Definición 10 (Función de Distribución). *Sea X una variable aleatoria*

$$\begin{aligned}F : \mathbb{R} &\rightarrow \mathbb{R} \\x &\mapsto F(x) = P(X \leq x)\end{aligned}$$

F es la función de distribución de X .

Teorema 2. *Sea F función de distribución de una variable aleatoria X , F cumple con:*

1. *F es una función monótona creciente.*
2. *F es continua por la derecha y tiene límite por la izquierda.*
3. *$\lim_{x \rightarrow -\infty} F(x) = 0$ y $\lim_{x \rightarrow +\infty} F(x) = 1$.*

- *Sea X una variable aleatoria, se dice que es discreta si su conjunto de imágenes es discreto. Formalmente se dice que la variable aleatoria es discreta si existe un conjunto $S = \{x_i\}_{i \in I \subseteq \mathbb{N}}$ tal que*

$$P(X \in S) = 1$$

La función de masa de probabilidad de una variable aleatoria discreta X se define por

$$f(x) = P(X = x),$$

que verifica las siguientes propiedades

1. $\forall i \in I, f(x_i) \geq 0$.
2. $\sum_{i \in I} f(x_i) = 1$.

Dada la función de masa de probabilidad, se determina la función de distribución como

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t).$$

- Una variable aleatoria continua X , cuya función de distribución F es tal que existe una función $f : \mathbb{R} \rightarrow \mathbb{R}$, no negativa, que verifica

$$F(x) = \int_{-\infty}^x f(t)dt,$$

es una variable absolutamente continua con f su función de densidad, la misma que cumple:

1. $\forall x, f(x) \geq 0$.
2. $\int_{-\infty}^{+\infty} f(x)dx = 1$.

Además, si $a, b \in \mathbb{R}$ con $a < b$

$$P(a \leq x \leq b) = \int_a^b f(x)dx.$$

Observación 3. Si F es continua, $F'(x) = f(x)$ c.t.p.

- Toda función de distribución que verifica alguna de las siguientes condiciones se denomina función de distribución singular
 1. F es continua y $F' = 0$ c.t.p.
 2. F es continua y $F' \neq 0$ sobre un conjunto no numerable de medida nula

Un ejemplo de este tipo de distribuciones es la distribución de Cantor, cuya función de distribución es la función de Cantor, que es continua pero no absolutamente continua y su derivada es nula casi en todos los puntos.

Nota 2. De manera general y haciendo mal uso del lenguaje se nombran como distribuciones de probabilidad continuas a aquellas que son absolutamente continuas. La ambigüedad se justifica en cierto modo, ya que las distribuciones continuas que no son absolutamente continuas no son usuales [34], un raro exponente es precisamente la distribución de Cantor.

Teorema 3 (Descomposición de Lebesgue). *Cualquier medida de probabilidad μ sobre \mathbb{R} puede ser descompuesta, en forma única, como*

$$\mu = \mu_{disc} + \mu_{ac} + \mu_{sing}$$

donde las medidas μ_{disc} , μ_{ac} , μ_{sing} son, respectivamente, discreta, absolutamente continua y singular.

1.1.3 Momentos de una variable aleatoria

Definición 11 (Esperanza). *Sea X variable aleatoria, $g : \mathbb{R} \rightarrow \mathbb{R}$ medible. La esperanza de $g(x)$ está dada por:*

- *Caso discreto*

$$E(g(x)) = \sum_x g(x)f(x)$$

Con f función de masa de probabilidad de X .

- *Caso continuo*

$$E(g(x)) = \int_{-\infty}^{\infty} g(x)f(x)dx$$

Con f función de densidad de X .

Propiedad 1. *Sea X variable aleatoria, a y b constantes y g_1, g_2 funciones medibles a valores reales.*

1. *Si $X \geq 0$, entonces $E(X) \geq 0$ casi seguramente.*
2. *$E(aX + b) = aE(X) + b$.*
3. *Si $g_1(X) \leq g_2(X)$, entonces $E(g_1(X)) \leq E(g_2(X))$ casi seguramente.*

Definición 12 (Varianza). *Sea X una variable aleatoria con $E(X) = \mu$. La varianza de X está definida por*

$$Var(X) = E[(X - \mu)^2].$$

Si se nota por $\sigma^2 = \text{Var}(X)$, se define la desviación estándar de X por

$$SD(X) = \sigma = \sqrt{\text{Var}(X)}.$$

Propiedad 2. Sea X variable aleatoria, a y b constantes.

1. $\text{Var}(X) \geq 0$.
2. $\text{Var}(X) = E(X^2) - [E(X)]^2$.
3. $\text{Var}(aX + b) = a^2\text{Var}(X)$.

Unas de las desigualdades más importantes relacionadas con la esperanza y la varianza de una variable aleatoria son las de Chebyshev, la de Jensen y la de Cauchy-Schwarz.

Teorema 4 (Desigualdad de Chebyshev). Sea X variable aleatoria, $E(X) = \mu$ y $\text{Var}(X) = \sigma^2 < +\infty$. Para todo $\epsilon > 0$ se cumple que

$$P(|X - \mu| > \epsilon) \leq \frac{\sigma^2}{\epsilon^2}$$

Si se considera un intervalo para la variable aleatoria X centrado en su media y de radio k -veces su desviación estándar, la desigualdad de Chebyshev proporciona una cota superior para la probabilidad de que la variable no se encuentre dentro de dicho intervalo. En efecto,

$$P(|X - \mu| > k\sigma) \leq \frac{1}{k^2}$$

Teorema 5 (Desigualdad de Jensen). Sea X variable aleatoria, $E(|X|) < +\infty$ y $g : \mathbb{R} \rightarrow \mathbb{R}$ una función convexa, esto es

$$g(tx + (1-t)y) \leq tg(x) + (1-t)g(y)$$

con $t \in [0, 1]$. Se tiene que

$$g(E(X)) \leq E(g(X))$$

Definición 13 (Función generadora de momentos). Sea X variable aleatoria, si existe $b > 0$ tal que para todo $t, |t| \leq b$. Se define la función generadora de momentos como

$$m(t) = E(\exp(tx))$$

Teorema 6. Sea X variable aleatoria con función generadora de momentos $m(t)$, se tiene entonces que

1. $\forall k \in \mathbb{N}, E(|x|^k) < +\infty$

2. $E(x^k) = m^{(k)}(0)$, donde $m^{(k)}$ representa la k -ésima derivada de la función m .

Teorema 7. Sean X, Y variables aleatorias, m_x y m_y sus funciones generadoras de momentos, si

$$m_x(t) = m_y(t), \quad |t| \leq b, \quad \text{con } b > 0$$

entonces las distribuciones de X y Y son las mismas, es decir

$$\forall x, P(X \leq x) = P(Y \leq x)$$

Definición 14 (Función característica). Sea X variable aleatoria, la función característica de X es una función $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ tal que

$$\varphi_X(t) = E(\exp(itx))$$

La función característica posee una variedad de propiedades, las más significativas son las siguientes:

- Sean X y Y dos variables aleatorias. Si $\varphi_X = \varphi_Y$ entonces $F_X = F_Y$, donde F representa la función de distribución de las variables indicadas en cada caso.
- $\forall t \in \mathbb{R}, |\varphi(t)| \leq 1$ y $\varphi(0) = 1$.
- Si $E(|X|^m) < +\infty$, entonces φ_X es diferenciables hasta el orden m y además

$$\varphi_X^{(k)}(0) = i^k E(X^k), \quad k = 1, \dots, m.$$

donde $\varphi_X^{(k)}$ representa la k -ésima derivada de la función.

Observación 4. La relación existente entre la función generadora de momentos y la función característica de la variable aleatoria X , queda determinado por

$$\varphi_X(t) = m_{iX}(t)$$

o por

$$\varphi_X(t) = m_X(it)$$

Observación 5. Existen variables aleatorias para las cuales su función generadora de momentos, $m(t)$, es no definida. En estos casos no se cumple la condición

$$m(t) < \infty \quad \forall t \in \mathbb{R}.$$

A diferencia de $m(t)$, la función característica $\varphi(t)$ siempre está bien definida ([2], p.54).

Ejemplo 2. Sea X variable aleatoria con distribución de Cauchy y función de masa de probabilidad

$$f(x) = \frac{1}{\pi(1+x^2)} \quad -\infty < x < +\infty$$

la función generadora de momentos viene dada por

$$\begin{aligned} m_X(t) &= E(\exp(tx)) \\ &= \int_{-\infty}^{+\infty} \frac{\exp tx}{\pi(1+x^2)} dx \end{aligned}$$

(i) Si $-\infty < x < 0$.

Si se considera únicamente los valores negativos de x , se cumple que $|\exp(tx)| < 1$. Así,

$$\begin{aligned} E(|\exp(tx)|) &= \int_{-\infty}^0 |\exp tx| \frac{1}{\pi(1+x^2)} dx \\ &< \int_{-\infty}^0 \frac{1}{\pi(1+x^2)} dx < \infty \end{aligned}$$

(ii) Si $x > 0$.

Para los valores positivos de la variable se verifica que $|\exp(tx)| > x$. Se sigue que $E(x) = \infty$, en efecto

$$\begin{aligned} E(x) &= \int_0^{+\infty} \frac{x}{\pi(1+x^2)} dx \\ &= \frac{1}{2\pi} \lim_{x \rightarrow \infty} \ln(1+x^2) = \infty. \end{aligned}$$

de donde y por propiedades de la esperanza matemática se tiene que $E(|\exp(tx)|)$ es también infinita.

De (i) y (ii) se concluye que la función generadora de momentos de X no existe. Por otro lado la función característica de una variable aleatoria con distribución de Cauchy está bien definida y viene dada por

$$\begin{aligned} \varphi_X(t) &= E(\exp(itx)) \\ &= \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{\exp(itx)}{1+x^2} dx \\ &= \exp(-|t|) \end{aligned}$$

1.1.4 Vectores aleatorios

Definición 15 (Vector aleatorio). X se denomina vector aleatorio si

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_k \end{pmatrix}$$

donde $\forall i \in \{1, \dots, k\}$, X_i es una variable aleatoria.

Definición 16 (Distribución conjunta). Sean:

- (Ω, \mathcal{A}, P) un espacio probabilístico.
- $X_i : \Omega \rightarrow \mathbb{R}$ función medible para $i = 1, \dots, k$.
- $B_i =]-\infty, x_i]$ con $x_i \in \mathbb{R}$, y $B = B_1 \times B_2 \times \dots \times B_k$.

Se define la distribución conjunta del vector aleatorio $X = (X_1, X_2, \dots, X_k)'$ por

$$\begin{aligned} P(X \in B) &= P(X_1 \in B_1, \dots, X_k \in B_k) \\ &= P(X_1 \leq x_1, \dots, X_k \leq x_k) \\ &= F(x_1, \dots, x_k) \\ &= F(x), \end{aligned}$$

donde $x = (x_1, \dots, x_k)$. Nótese además que

$$F : \mathbb{R}^k \rightarrow [0, 1].$$

- Si $\forall i \in \{1, \dots, k\}$, X_i es una variable aleatoria discreta. La función de frecuencias de $X = (X_1, X_2, \dots, X_k)'$ está dada por

$$f(x_1, \dots, x_k) = P(X_1 = x_1, \dots, X_k = x_k).$$

f es tal que:

1. $\forall x \in \mathbb{R}^k, f(x) \geq 0$.
 2. $\sum_{x_1} \sum_{x_2} \dots \sum_{x_k} f(x_1, \dots, x_k) = 1$.
- Si $\forall i \in \{1, \dots, k\}$, X_i es una variable aleatoria continua. La función de densidad de $X = (X_1, X_2, \dots, X_k)'$ está dada por

$$f : \mathbb{R}^k \rightarrow \mathbb{R}$$

$$x \mapsto f(x)$$

f es tal que:

1. $\forall x \in \mathbb{R}^k, f(x) \geq 0$.
2. $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f(x) dx = 1$.

Observación 6. Sea F función de distribución de un vector aleatorio continuo $X = (X_1, X_2, \dots, X_k)'$ y f su función de densidad, se tiene que

1. $F(x) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_k} f(x_1, x_2, \dots, x_k) dx_1 dx_2 \dots dx_k$.
2. $f(x_1, \dots, x_k) = \frac{\partial^k F(x_1, \dots, x_k)}{\partial x_k \dots \partial x_1}$ c.t.p.

Definición 17 (Distribuciones marginales). Sea

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_k \end{pmatrix}$$

un vector aleatorio, dada la distribución de probabilidad de X , es posible calcular la distribución de probabilidad de cualquier sub-vector de X , esta distribución es conocida como marginal.

- *Caso discreto.*

La distribución marginal de la variable aleatoria X_i con $i = 1, \dots, k$, está dada por

$$f(x_i) = \sum_{x_1} \dots \sum_{\cancel{x_i}} \dots \sum_{x_k} f(x_1, \dots, x_i, \dots, x_k)$$

- *Caso continuo.*

La distribución marginal de la variable aleatoria X_i con $i = 1, \dots, k$, está dada por

$$f(x_i) = \underbrace{\int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty}}_{k-1 \text{ veces}} f(x_1, \dots, x_i, \dots, x_k) dx_1 \dots \cancel{dx_i} \dots dx_k$$

Definición 18 (Distribución condicional). Sean X y Y variables aleatorias, la distribución de probabilidad de X dado $Y = y$ se define por

$$f(x|y) = \frac{f(x, y)}{f_Y(y)}$$

donde $f(x, y)$ es la función de distribución conjunta de (X, Y) y $f_Y(y)$ es la distribución marginal de Y .

Definición 19 (Variables aleatorias independientes). Se dice que las variables aleatorias X_i con $i = 1, \dots, k$ son independientes si y solo si para todo a_i, b_i tal que $a_i < b_i$ los eventos $\{a_i < x \leq b_i\}$ son independientes, es decir

$$P(a_1 < X_1 \leq b_1, \dots, a_k < X_k \leq b_k) = \prod_{i=1}^k P(a_i < X_i \leq b_i)$$

Teorema 8. Las variables aleatorias X_i con $i = 1, \dots, k$ son independientes si y solo si

$$f(x_1, \dots, x_k) = f_1(x_1) \dots f_k(x_k).$$

Donde f_i representa la función de densidad marginal de la variable X_i para todo $i = 1, \dots, k$.

Corolario 1. Las variables aleatorias X_i con $i = 1, \dots, k$ son independientes si y solo si

$$F(x_1, \dots, x_k) = \prod_{i=1}^k F_i(x_i).$$

Donde F_i representa la función de distribución marginal de la variable X_i para todo $i = 1, \dots, k$.

1.1.5 Momentos de un vector aleatorio

Definición 20 (Esperanza de un vector aleatorio). Sean $X = (X_1, \dots, X_k)'$ un vector aleatorio de distribución conjunta F y h una función

$$\begin{aligned} h : \mathbb{R}^k &\rightarrow \mathbb{R} \\ x &\mapsto h(x) \end{aligned}$$

- Si $\forall i \in \{1, \dots, k\}, X_i$ es una variable aleatoria discreta, la esperanza de $h(X)$ está dada por

$$E(h(X)) = \sum_{x_1} \sum_{x_2} \dots \sum_{x_k} h(x_1, x_2, \dots, x_k) f(x_1, x_2, \dots, x_k)$$

Donde $f(x_1, x_2, \dots, x_k)$ representa la función de frecuencias de X .

- Si $\forall i \in \{1, \dots, k\}, X_i$ es una variable aleatoria continua, la esperanza de $h(X)$

está dada por

$$\underbrace{\int_{\mathbb{R}} \dots \int_{\mathbb{R}}}_{k\text{-veces}} h(x_1, x_2, \dots, x_k) f(x_1, x_2, \dots, x_k) dx_1 dx_2 \dots dx_k$$

Teorema 9. Sean X_1, \dots, X_k variables aleatorias.

1. $E(X_1 + X_2 + \dots + X_k) = E(X_1) + E(X_2) + \dots + E(X_k)$.
2. Si X_1, \dots, X_k son independientes, entonces

$$E(X_1 X_2 \dots X_k) = E(X_1) E(X_2) \dots E(X_k).$$

Definición 21 (Función generadora de momentos de vectores aleatorios). Sea $X = (X_1, \dots, X_k)'$ un vector aleatorio, se define la función generadora de momentos de X por

$$m(t) = E(\exp(t'X))$$

donde $t \in \mathbb{R}^k$ es tal que $\|t\| < b$, con $b > 0$.

Teorema 10. Sea $X = (X_1, \dots, X_k)'$ un vector aleatorio, $t = (t_1, t_2, \dots, t_k)' \in \mathbb{R}^k$, $\forall i, j \in \{1, \dots, k\}$ se tiene que

- $E(X_i) = \frac{\partial m(0)}{\partial t_i}$.
- $E(X_i X_j) = \frac{\partial^2 m(0)}{\partial t_i \partial t_j}$.

Definición 22 (Covarianza y correlación). Sean X, Y variables aleatorias tales que $E(X^2) < +\infty$ y $E(Y^2) < +\infty$, es decir son de cuadrado integrable. Se define la covarianza por

$$\text{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y).$$

La correlación está dada por

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}, \quad \text{Var}(X) \neq 0, \text{Var}(Y) \neq 0$$

Observación 7. Si X, Y son variables aleatorias independientes entonces no están correlacionadas, es decir $\rho(X, Y) = 0$. El recíproco no se da.

Ejemplo 3. Sean $X \sim U[-1, 1]$ con $f_X(x) = \frac{1}{2}$ para $-1 \leq x \leq 1$ e Y variables aleatorias, con Y definida por

$$Y = \begin{cases} -1, & |x| \leq \frac{1}{2} \\ 1, & \frac{1}{2} < |x| \leq 1. \end{cases}$$

Con X e Y se ejemplificará el enunciado de la observación anterior. Calculando las esperanzas de las variables aleatorias:

$$\begin{aligned} E(X) &= 0 \\ E(Y) &= 0. \end{aligned}$$

En efecto,

$$\begin{aligned} E(Y) &= (-1)P(Y = -1) + P(Y = 1) \\ &= -P\left(|x| \leq \frac{1}{2}\right) + P\left(\frac{1}{2} < |x| \leq 1\right) \\ &= -P\left(-\frac{1}{2} \leq x \leq \frac{1}{2}\right) + P\left(-1 \leq x < -\frac{1}{2}\right) + P\left(\frac{1}{2} < x \leq 1\right) \\ &= -\int_{-1/2}^{1/2} \frac{1}{2} dx + \int_{-1}^{-1/2} \frac{1}{2} dx + \int_{1/2}^1 \frac{1}{2} dx = 0. \end{aligned}$$

Por otro lado

$$XY = \begin{cases} -X, & |x| \leq \frac{1}{2} \\ X, & \frac{1}{2} < |x| \leq 1. \end{cases}$$

de donde

$$\begin{aligned} E(XY) &= \int_{-1/2}^{1/2} -\frac{x}{2} dx + \int_{-1}^{-1/2} \frac{x}{2} dx + \int_{1/2}^1 \frac{x}{2} dx \\ &= -\frac{x^2}{4} \Big|_{-1/2}^{1/2} + \frac{x^2}{4} \Big|_{-1}^{-1/2} + \frac{x^2}{4} \Big|_{1/2}^1 = 0. \end{aligned}$$

Así

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = 0$$

Como

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}, \quad \text{Var}(X) \neq 0, \text{Var}(Y) \neq 0$$

entonces

$$\rho(X, Y) = 0.$$

Por otro lado X e Y no son independientes, en efecto

$$P(Y = -1) = \frac{1}{2}$$

mientras que

$$P(Y = -1|X = x) = \begin{cases} 1, & |x| \leq \frac{1}{2} \\ 0, & \frac{1}{2} < |x| \leq 1. \end{cases}$$

Ya que $P(Y = -1) \neq P(Y = -1|X = x)$ se concluye el resultado.

Definición 23 (Varianza de un vector aleatorio). Sea $X = (X_1 \dots, X_k)'$ un vector aleatorio, si todas las componentes X_i con $i \in \{1, \dots, k\}$ son de cuadrado integrable, la varianza de X se define por

$$\text{Var}(X) = E[\underbrace{(X - E(X))}_{k \times 1} \underbrace{(X - E(X))'}_{1 \times k}]$$

La varianza de X es una matriz simétrica de dimensiones $k \times k$

$$\text{Var}(X) = \begin{pmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) & \dots & \text{Cov}(X_1, X_k) \\ \text{Cov}(X_2, X_1) & \text{Var}(X_2) & \dots & \text{Cov}(X_2, X_k) \\ \vdots & \ddots & \ddots & \vdots \\ \text{Cov}(X_k, X_1) & \text{Cov}(X_k, X_2) & \dots & \text{Var}(X_k) \end{pmatrix}$$

Sea $Y = (Y_1, \dots, Y_q)'$ vector aleatorio, la covarianza entre X y Y está dada por

$$\begin{aligned} \text{Cov}(X, Y) &= E[(X - E(X))(Y - E(Y))'] \text{ matriz } k \times q \\ &= E(XY') - E(X)[E(Y)]' \end{aligned}$$

Observación 8. Nótese que $\underbrace{\text{Cov}(X, Y)}_{k \times q} \neq \underbrace{\text{Cov}(Y, X)}_{q \times k}$.

Teorema 11. Sea A matriz $q \times k$, $X = (X_1 \dots, X_k)'$ un vector aleatorio cuyas componentes son de cuadrado integrable y $b \in \mathbb{R}^q$. Se tiene que

$$\text{Var}(AX + b) = A\text{Var}(X)A'$$

Definición 24 (Procesos estocásticos). Dado el espacio probabilístico (Ω, \mathcal{A}, P) , una familia de variables aleatorias $\{X_t\}_{t \in T}$ tal que

$$\begin{aligned} X_t : \Omega &\rightarrow E \subset \mathbb{R}^k \\ w &\mapsto X_t(w) \in E \end{aligned}$$

donde T representa el espacio de tiempos y E el espacio de estados.

Propiedad 3 (Propiedad de Markov). Sea $\{X_n\}_{n \geq 0}$ un proceso estocástico con su espacio de estados $E \subset \mathbb{N}$, la igualdad siguiente se conoce como propiedad markoviana

$$P(X_{n+1} = x_{n+1} | X_n = x_n, X_{n-1} = x_{n-1}, X_0 = x_0) = P(X_{n+1} = x_{n+1} | X_n = x_n)$$

En palabras la propiedad de Markov expresa, estando en el instante n , que la probabilidad del futuro no depende del pasado sino del presente. Un proceso estocástico que verifique la propiedad anterior se conoce como Cadena de Markov.

1.1.6 Convergencia de variables aleatorias

Definición 25 (Convergencia casi segura). Sea $\{x_n\}_{n \geq 1}$ y x variables aleatorias, se dice que x_n converge de forma casi segura hacia x si y solo si

$$P\left(\lim_{n \rightarrow \infty} x_n = x\right) = 1.$$

Se nota $x_n \rightarrow_{c.s.} x$.

Definición 26 (Convergencia en probabilidad). Sea $\{x_n\}_{n \geq 1}$ y x variables aleatorias, se dice que x_n converge en probabilidad hacia x si y solo si

$$\forall \epsilon > 0, \quad \lim_{n \rightarrow \infty} P(|x_n - x| > \epsilon) = 0.$$

Se nota $x_n \rightarrow_P x$.

Definición 27 (Convergencia en distribución). Sea $\{x_n\}_{n \geq 1}$ y x variables aleatorias, $\{F_n\}_{n \geq 1}$ y F sus respectivas funciones de distribución. Se dice que x_n converge en distribución hacia x si y solo si para todo x punto de continuidad de F

$$\lim_{n \rightarrow \infty} F_n(x) = F(x) \Leftrightarrow \lim_{n \rightarrow \infty} P(x_n \leq x) = P(x \leq x)$$

Se nota $x_n \rightarrow_D x$.

Nota 3.

$$x_n \rightarrow_{c.s.} x \Rightarrow x_n \rightarrow_P x \Rightarrow x_n \rightarrow_D x.$$

Los tres tipos de convergencia presentados: convergencia casi segura, en probabilidad y en distribución, se dan cuando $n \rightarrow \infty$.

Teorema 12 (De la función continua). Sea $\{x_n\}_{n \geq 1}$, x variables aleatorias y $g : \mathbb{R} \rightarrow \mathbb{R}$ continua.

$$1. x_n \rightarrow_P x \Rightarrow g(x_n) \rightarrow_P g(x)$$

$$2. x_n \rightarrow_D x \Rightarrow g(x_n) \rightarrow_D g(x)$$

Teorema 13 (Método Delta). Sea $\{a_n\}_{n \geq 1}$ una sucesión creciente de constantes, θ constante, g una función a valores reales continuamente diferenciable en θ .

Si $a_n(x_n - \theta) \rightarrow_D N(0, \sigma^2)$, entonces

$$a_n(g(x_n) - g(\theta)) \rightarrow_D N(0, (g'(\theta))^2 \sigma^2).$$

Teorema 14 (Ley de los Grandes Números). Sean $\{x_i\}_{1 \leq i \leq n}$ variables aleatorias i.i.d. tales que $E(x_i) = \mu$, $\forall i = 1, \dots, n$. Entonces

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i \rightarrow_P \mu$$

Teorema 15 (Ley Fuerte de los Grandes Números). Sean $\{x_i\}_{1 \leq i \leq n}$ variables aleatorias i.i.d. tales que $E(x_i) = \mu < \infty$ y $E(x_i^4) = C, \forall i = 1, \dots, n$ donde C es una constante, entonces

$$\frac{1}{n} \sum_{i=1}^n x_i$$

converge a μ casi seguramente.

Teorema 16 (Teorema del Límite Central). Sean $\{x_i\}_{1 \leq i \leq n}$ variables aleatorias i.i.d. con media μ y varianza $\sigma^2 < \infty$, con $\sigma > 0$, sea

$$Y_n = \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^n (x_i - \mu) = \frac{\sqrt{n}(\bar{x}_n - \mu)}{\sigma}.$$

Entonces $Y_n \rightarrow_D Z$, donde $Z \sim N(0, 1)$ si $n \rightarrow +\infty$.

$$\text{Sea } S_n = \sum_{i=1}^n x_i,$$

$$\frac{\sqrt{n}(\frac{1}{n}S_n - \mu)}{\sigma} = \frac{S_n - n\mu}{\sigma\sqrt{n}} \rightarrow_D N(0, 1)$$

En términos prácticos se dice que para un valor de n lo suficientemente grande S_n sigue aproximadamente una distribución normal con media $n\mu$ y varianza $n\sigma^2$.

Observación 9 (Importancia del Teorema del Límite Central). En la práctica el Teorema del Límite central es importante ya que explica porque en muchos campos de aplicación reiteradamente se encuentran distribuciones normales o casi normales[6].

Haciendo un análisis más teórico la importancia del Teorema Central de Límite radica en que relaciona la estadística y la probabilidad.

Demostración. Para esta demostración se supondrá la existencia de las funciones generadoras de momentos de las variables x_i , misma que está dada por

$$\begin{aligned} m(t) &= E(\exp(tx_i)) \\ &= E\left(\sum_{k=0}^{\infty} \frac{(tx_i)^k}{k!}\right) \\ &= \sum_{k=0}^{\infty} \frac{t^k E(x_i^k)}{k!} \\ &= 1 + \frac{t^2}{2} + \frac{t^3 E(x_i^3)}{6} + \dots \end{aligned}$$

Se asume, sin pérdida de generalidad, que $E(x_i) = 0$ y $Var(x_i) = 1$, de donde $Y_n = \frac{S_n}{\sqrt{n}}$, ahora la función generadora de momentos de Y_n está dada por

$$\begin{aligned} m_{Y_n}(t) &= E(\exp(Y_n t)) \\ &= E\left(\exp\left(\frac{t}{\sqrt{n}} \sum_{i=1}^n x_i\right)\right) \\ &= E\left(\exp\left(\frac{t}{\sqrt{n}} x_1 + \dots + \frac{t}{\sqrt{n}} x_n\right)\right) \\ &= E\left(\exp\left(\frac{t}{\sqrt{n}} x_1\right) \dots \exp\left(\frac{t}{\sqrt{n}} x_n\right)\right) \\ &= E\left(\exp\left(\frac{t}{\sqrt{n}} x_1\right)\right) \dots E\left(\exp\left(\frac{t}{\sqrt{n}} x_n\right)\right) \\ &= m^n\left(\frac{t}{\sqrt{n}}\right) \\ &= \left(1 + \frac{t^2}{2n} + \frac{t^3 E(x_i^3)}{6n^{3/2}} + \frac{t^4 E(x_i^4)}{24n^2} + \dots\right)^n \\ &= \left(1 + \frac{t^2}{2n} \gamma_n(t)\right)^n \quad \text{con } \gamma_n(t) = 1 + \frac{t E(x_i^3)}{3\sqrt{n}} + \frac{t^2 E(x_i^4)}{12n} + \dots \rightarrow 1, \text{ si } n \rightarrow +\infty. \end{aligned}$$

De donde $\lim_{n \rightarrow \infty} m_{Y_n}(t) = \exp\left(\frac{1}{2}t^2\right)$, que corresponde a la función generadora de momentos de la distribución normal con media 0 y varianza 1. Ya que la función generadora de momentos de Y_n converge hacia la función generadora de momentos de Z , se concluye el resultado. \square

Nótese que la suposición hecha inicialmente, la existencia de $m(t)$ para todos los

x_i , supone la existencia de $E(x_i^k)$ para todo $1 \leq i \leq n$ y para todo $k \in \mathbb{N}$. Una segunda forma de probar el teorema del límite central suponiendo únicamente la existencia de $E(x_i^2)$ para $1 \leq i \leq n$ fue propuesta por Lindeberg, misma que puede ser encontrada en ([2],p. 134).

A continuación se ilustrará gráficamente el Teorema del Límite Central.

- Distribución Uniforme.

(i) $X_1 \sim U[-1, 1]$

$$f_{X_1}(x) = \begin{cases} \frac{1}{2}, & |x| \leq 1, \\ 0, & |x| > 1. \end{cases}$$

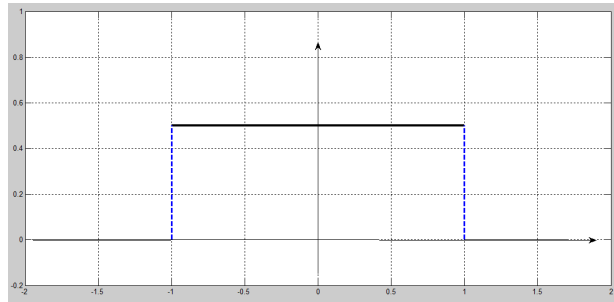


Figura 1.1: Función de densidad de X_1

(ii) $X_i \sim U[-1, 1]$ con $i = 1, 2$.

$$f_{X_1+X_2}(x) = \begin{cases} \frac{2 - |x|}{4}, & |x| \leq 2, \\ 0, & |x| > 2. \end{cases}$$

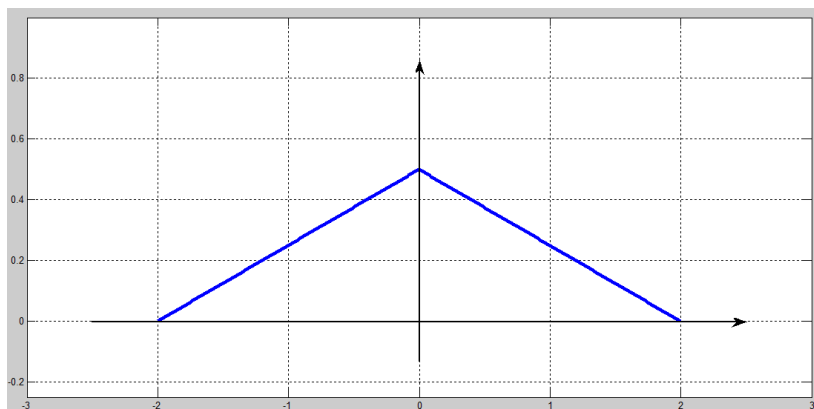


Figura 1.2: Función de densidad de $X_1 + X_2$

(iii) $X_i \sim U[-1, 1]$ con $i = 1, 2, 3$.

$$f_{X_1+X_2+X_3}(x) = \begin{cases} \frac{(3-|x|)^2}{16}, & 1 \leq |x| \leq 3, \\ \frac{3-x^2}{8}, & 0 \leq |x| \leq 1, \\ 0, & |x| > 3. \end{cases}$$

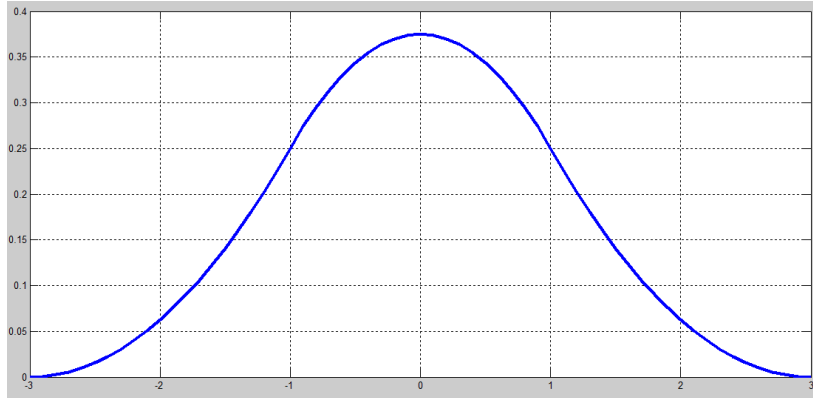


Figura 1.3: Función de densidad de $X_1 + X_2 + X_3$

De manera general para $X_i \sim U[-1, 1]$ con $i = 1, \dots, n$.

$$f_{X_1+\dots+X_n}(x) = \begin{cases} \frac{1}{2^n(n-1)!} \sum_{k=0}^{\frac{n+x}{2}} (-1)^k C_n^k (n+x-2k)^{n-1}, & |x| \leq n, \\ 0, & |x| > n. \end{cases}$$

- Distribución χ^2

Sean x_1, x_2, \dots, x_n variables aleatorias *i.i.d* con distribución normal estándar, se tiene entonces que

$$x_i^2 \sim \chi_1^2, \quad \forall 1 \leq i \leq n \quad (1.1)$$

y

$$\sum_{i=1}^n x_i^2 \sim \chi_n^2 \quad (1.2)$$

Aplicando el teorema del límite central

$$\frac{\sum_{i=1}^n x_i^2 - n}{\sqrt{2n}} \approx N(0, 1)$$

$$\sum_{i=1}^n x_i^2 \approx N(n, 2n)$$

de (1.2) se sigue que

$$\chi_n^2 \approx N(n, 2n).$$

A continuación se presenta la gráfica de la función de densidad de χ^2 con diferentes grados de libertad

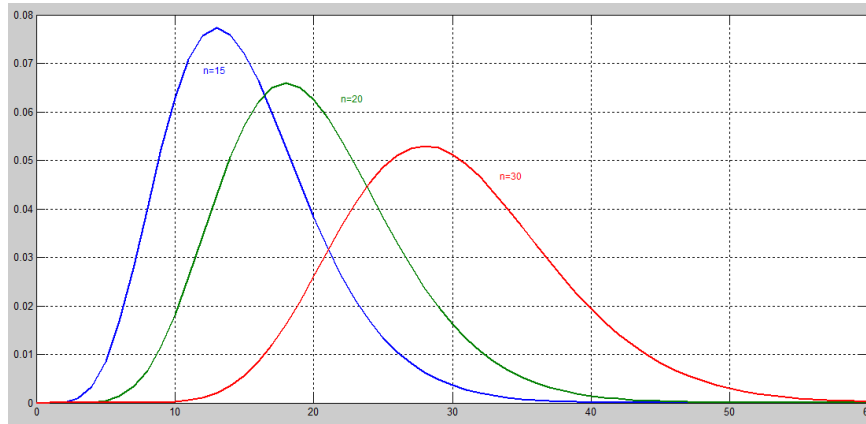


Figura 1.4: Función de densidad χ^2 con 15, 20 y 30 grados de libertad.

Capítulo 2

Ligera historia y desarrollo del Teorema del Límite Central

En este capítulo se abordará brevemente el teorema del límite central desde un punto de vista histórico. Se conocerá en qué siglo apareció la primera versión de este teorema y todos los cambios que se han venido realizando a través de los tiempos, así como los matemáticos responsables de ellos.

En una segunda sección se tratarán las diferentes adaptaciones del teorema central del límite, las hipótesis adicionales que son requeridas en cada una de ellas, así como ejemplos y aplicaciones en cada caso.

2.1 Reseña histórica

La primera versión del teorema central del límite se atribuye al matemático francés Abraham de Moivre al demostrar, en 1738 en la segunda edición de su libro “The Doctrine of Chances”, que la distribución binomial $\mathcal{B}(n, \theta)$ para valores grandes de n y con $\theta = \frac{1}{2}$ se aproxima a lo que hoy conocemos como distribución normal de media $n\theta$ y varianza $n\theta(1 - \theta)$.

Este descubrimiento surgió cuando de Moivre observó que mientras más repeticiones de la experiencia aleatoria, lanzamiento de una moneda, realizaba la gráfica de la distribución obtenida correspondía a una curva suave, específicamente a la curva de la distribución normal.

Años más tarde, en 1810, Laplace enuncia una versión del teorema que generalizaba el resultado obtenido por Moivre. Según [6] el resultado expuesto por Laplace en términos modernos es el siguiente:

Teorema 17. Sean X_1, \dots, X_n variables independientes, idénticamente distribuidas (i.i.d.) y acotadas, discretas o absolutamente continuas. Y sean $\mu = E(X_1)$ y

$\sigma^2 = \text{Var}(X_1) < \infty$ con $\sigma > 0$. Si n es suficientemente grande, vale la siguiente aproximación:

$$P(n\mu + r_1\sqrt{n} \leq X_1 + \dots + X_n \leq n\mu + r_2\sqrt{n}) \approx \frac{1}{\sqrt{2\pi}} \int_{r_1}^{r_2} \frac{1}{\sigma} e^{-\frac{t^2}{2\sigma^2}} dt.$$

Aunque el trabajo realizado por Laplace sienta las bases del teorema, nunca lo enunció de manera formal y consecuentemente tampoco realizó una demostración que lo hiciera válido para una distribución de probabilidad arbitraria.

Mas tarde en 1824 Poisson generalizó el resultado obtenido por Laplace, trabajando primero en la suma y luego en combinaciones lineales de variables aleatorias independientes. Fue mediante el uso de contraejemplos que Poisson observó la importancia de suposiciones iniciales para que el teorema se verifique, en [6] se señala como el más importante de estos contraejemplos al siguiente:

Contraejemplo 1. *La suma de variables aleatorias X_i independientes, idénticamente distribuidas (i.d.d.) con función de densidad*

$$f(x) = \frac{1}{\pi(1+x^2)} \quad -\infty < x < +\infty$$

no se aproxima a una distribución normal, aun cuando la cantidad de sumandos sea muy grande.

La distribución arriba señalada corresponde a la distribución de Cauchy, conocida por no tener un valor esperado definido ([2],p. 37), es decir no verifica las condiciones del teorema del límite central.

Famosos matemáticos como Dirichlet, Bessel y Cauchy también hicieron contribuciones en la demostración de este teorema. Dirichlet realizó la demostración desde el punto de vista de la teoría de la integración ya que consideraba los problemas surgidos en probabilidades como aplicaciones de este campo.

Por su parte Cauchy utilizó funciones características y teoremas de inversión para el problema de aproximar a la distribución normal la combinación lineal de errores aleatorios, las suposiciones adicionales que Cauchy necesitó sobre la función de densidad de estas variables aleatorias fueron que sea simétrica y de soporte compacto¹.

Si bien las demostraciones proporcionadas constituyeron una base importante en el estudio del teorema, no satisfacían aspectos fundamentales, entre ellos:

- No se demostró el teorema en su caso más general, es decir para distribuciones arbitrarias.

¹La definición de soporte compacto se encuentra en el Anexo A.

- No determinaban condiciones necesarias y suficientes bajo las cuales la aproximación a la distribución normal fuera posible.
- No se determinó el orden de convergencia, es decir cuantas variables aleatorias (i.i.d.) deben sumarse para obtener una buena aproximación a la curva normal.

A finales del siglo XIX, Chebyshev, Markov y Lyapunov se encargan de solventar de manera parcial las falencias en las anteriores demostraciones. Chebyshev y Markov demuestran el teorema por el método de momentos².

Por su parte Lyapunov, en 1910, realiza una demostración mediante el uso de funciones características, además de presentar condiciones suficientes para la consecución del teorema, condición que lleva su nombre.

Según [6] el teorema del límite central enunciado por Lyapunov, es el siguiente:

Teorema 18 (Lyapunov). Sean $(X_k)_{k \in \mathbb{N}}$ v.a. independientes cuyos tres primeros momentos existen, y sean: $m_k = E(X_k)$, $\sigma^2 = \text{Var}(X_k) < +\infty$
 $a_k = E((X_k - m_k)^3)$ momento centrado de tercer orden.
 $b_k = E(|X_k - m_k|^3)$ momento centrado absoluto de tercera orden.

Llamemos

$$s_n := \sqrt{\sum_{k=1}^n \sigma^2} \quad B_n := \sqrt[3]{\sum_{k=1}^n b_k}$$

Si se verifica la siguiente condición (condición de Lyapunov):

$$\lim_{n \rightarrow \infty} \frac{B_n}{s_n} = 0$$

entonces

$$\lim_{n \rightarrow \infty} F_n(x) = \Phi(x)$$

donde Φ representa la función de distribución de la distribución normal estándar y F_n es la función de distribución de

$$S_n^* = \frac{\sum_{k=1}^n (X_k - m_k)}{s_n}$$

Años más tarde, en 1922, Lindeberg proporciona una condición de suficiencia más general que la enunciada por Lyapunov. Finalmente, en 1937, Feller demuestra las condiciones necesarias para el cumplimiento del teorema.

²El método de momentos se describe en el Anexo A

2.2 Adaptaciones del Teorema del Límite Central

Debido a su gran versatilidad, el teorema del límite central ha sido usado en muchos campos de la matemática, tales como los procesos de renovación, los procesos de Wiener, en las martingalas, etc. En la presente sección se introducirán los conceptos básicos de algunos de los procesos antes señalados hasta llegar al teorema del límite central para cada caso.

2.2.1 Procesos de renovación

Un proceso de renovación $\{N(t)\}_{t \geq 0}$ es un proceso estocástico que registra el número de ocurrencias de un evento durante un intervalo de tiempo $(0, t]$. Sea X_i la variable aleatoria que representa el tiempo que transcurre desde la ocurrencia del evento $i - 1$ a la ocurrencia del i -ésimo evento, en un proceso de renovación, estas variables aleatorias son independientes idénticamente distribuidas con función de distribución

$$F(x) = P(X_k \leq x), \quad k = 1, 2, \dots,$$

$F(x)$ representa la probabilidad de que el tiempo que transcurra entre los eventos $k - 1$ y k sea menor a cierto valor x . Una condición básica en este tipo de procesos es que $F(0) = 0$. Se define al tiempo de espera hasta la ocurrencia del n -ésimo evento con $n \geq 1$ como

$$S_n = X_1 + X_2 + \dots + X_n$$

por convención $S_0 = 0$.

El experimento de reemplazar bombillas eléctricas es el que le da el nombre al proceso. Supongamos que en el tiempo 0 se instala una bombilla eléctrica, la misma que falla en el tiempo X_1 y es reemplazada por una nueva, la segunda bombilla falla en el tiempo $X_1 + X_2$, tiempo en el cual es reemplazado por una tercera y así sucesivamente. Entendiendo como evento la falla de cada bombilla, la variable $N(t)$ contaría el número de bombillas que fueron cambiadas o en otras palabras el número de renovaciones.

Otro concepto importante dentro de la teoría de la renovación es el de función de renovación, notada por $M(t)$, representa el número esperado de renovaciones en el intervalo $(0, t]$, es decir

$$M(t) = E(N(t)).$$

Claramente existe una conexión entre $N(t) = n$ y S_n , ya que el primero muestra el número de eventos ocurridos en el tiempo $S_n < t$. Más formalmente se tiene que

$$N(t) \geq k \Leftrightarrow S_k \leq t \tag{2.1}$$

Uno de los resultados más representativos dentro de la teoría de renovaciones es el proceso de Poisson, descrito a continuación:

Proceso de Poisson: En un proceso de renovación de este tipo, las variables $\{X_k\}_{k \geq 1}$ que representan el tiempo entre cada ocurrencia, siguen una distribución exponencial con parámetro λ , es decir, su función de distribución viene dada por

$$F(x) = 1 - \exp(-\lambda x), \quad x \geq 0.$$

La distribución exponencial verifica las siguientes propiedades:

$$\begin{aligned} E(X_k) &= \frac{1}{\lambda}, \quad \forall k \geq 1 \\ \text{Var}(X_k) &= \frac{1}{\lambda^2}, \quad \forall k \geq 1 \\ P(X_k > t + s | X_k > t) &= P(X_k > s), \quad \forall k \geq 1 \end{aligned}$$

Esta última se conoce como pérdida de memoria y es una de las propiedades más importantes de la distribución exponencial.

Sea $N(t)$ un proceso de Poisson de tasa λ , en el cual la ocurrencia entre cada evento sigue una distribución exponencial con este parámetro, $N(t)$ representa el número de eventos ocurridos hasta el instante t y sigue una distribución de Poisson con parámetro λt , de donde la función de renovación de proceso es:

$$M(t) = E(N(t)) = \lambda t.$$

Propiedad 4. *Las propiedades más importantes del proceso de Poisson $N(t)$ son:*

1. $N(0) = 0$.
2. $N(t+s) - N(s)$ sigue una distribución de Poisson de parámetro $\lambda(t+s-s) = \lambda t$ con $0 \leq s < t$. Se dice que $N(t)$ tiene incrementos estacionarios, es decir que solo dependen del salto.
3. $N(t_1) - N(t_0), N(t_2) - N(t_1), N(t_3) - N(t_2), \dots, N(t_n) - N(t_{n-1})$ son independientes, donde $0 \leq t_0 \leq t_1 \leq \dots \leq t_n$. Esta propiedad recibe el nombre de incrementos independientes.

A continuación se enunciarán sin demostración algunos de los resultados más importantes dentro de la teoría de la renovación³.

³La demostración de los teoremas expuestos, así como más resultados de la teoría de las renovaciones se encuentran en ([7], p.167-208)

Teorema 19. Sean $\{X_k\}_{k \geq 1}$ variables aleatorias i.i.d. que representan el tiempo entre cada ocurrencia dentro de un proceso de renovación, tales que $E(X_k) = \mu < \infty$ para $k \geq 1$, entonces

$$\lim_{t \rightarrow \infty} \frac{1}{t} M(t) = \frac{1}{\mu}.$$

Si los $\{X_k\}_{k \geq 1}$ representan el tiempo de vida de cierto artículo, el mismo que se espera sea de μ unidades de tiempo, el teorema indica que al largo plazo cada artículo será reemplazado a una tasa $\frac{1}{\mu}$.

Teorema 20 (Teorema de renovación). *Bajo las mismas condiciones del teorema anterior y con $h > 0$, se tiene que*

$$\lim_{t \rightarrow \infty} (M(t+h) - M(t)) = \frac{h}{\mu}$$

Es decir, el número esperado de renovaciones en un intervalo de longitud $h > 0$ es $\frac{h}{\mu}$.

Teorema 21 (Ley fuerte para procesos de renovación). *Sea $N(t)$ un proceso de renovación donde la media del tiempo entre ocurrencia de los eventos está dada por \bar{X} , entonces*

$$\lim_{t \rightarrow \infty} \frac{N(t)}{t} = \frac{1}{\bar{X}}$$

con probabilidad 1.

Teorema 22 (Teorema del límite central para procesos de renovación). *Sea $N(t)$ un proceso de renovación en el cual, $E(X_k) = \mu < \infty$ y $\sigma^2 = E[(X_k - \mu)^2]$ para todo $k \geq 1$. Entonces*

$$\lim_{t \rightarrow \infty} P \left(\frac{N(t) - t/\mu}{\sqrt{t\sigma^2/\mu^3}} < x \right) = \Phi(x),$$

donde $\Phi(x)$ representa la distribución normal estándar.

2.2.1.1 Aplicación de los procesos de renovación

Una de las aplicaciones más importantes de los procesos de renovación, específicamente de los procesos de Poisson, se dio en el campo actuarial y fue desarrollado por Lundberg (1903) y Crámer (1930), se trata del modelo clásico de riesgo conocido también como aproximación de Crámer-Lundberg en honor a sus creadores⁴.

El modelo de riesgo es de gran importancia para todo tipo de entidades financieras, en especial las aseguradoras, que estén interesadas en saber si en algún momento no podrán cumplir con sus obligaciones sean estas: pagos o reclamos.

⁴Para mas referencias véase [8].

Si nos centramos en las aseguradoras el modelo de riesgo estima la probabilidad de ruina, es decir, se estima la probabilidad de que en caso de accidente no pueda reconocer el valor estipulado a un cliente que ha pagado con regularidad el monto impuesto (prima). Una hipótesis aceptable en el planteamiento del problema es el de suponer que la aseguradora inició sus actividades con un presupuesto o capital inicial, además de que cuenta con el capital proveniente de las primas.

Aunque el valor adeudado por la aseguradora a sus clientes en caso de que ellos requieran su servicio es desconocido, se desea saber si podrá solventar estos gastos con la suma de capitales obtenida o caso contrario al no poder cubrir sus obligaciones entra en un estado de ruina.

Modelo clásico de Crámer-Lundberg: Sean $u \geq 0$ el capital inicial con el que cuenta la aseguradora. X_1, X_2, \dots los tiempos entre llegadas de reclamos, $\{N(t)\}_{t>0}$ un proceso de Poisson de tasa $\lambda > 0$, mismo que cuenta el número de reclamos. Z_1, Z_2, \dots variables aleatorias i.i.d. que representan el tamaño o monto de los reclamos con F su función de distribución, la misma que verifica

$$\begin{aligned} F(0) &= 0 \\ E(Z_k) &= \mu > 0. \end{aligned}$$

Sea $c > 0$ un valor de prima constante, se tiene entonces

$$W(t) = u + ct - \sum_{k=1}^{N(t)} Z_k$$

que representa el capital de la aseguradora en el tiempo t , $W(t)$ recibe el nombre de modelo clásico de Crámer-Lundberg.

Se dice que la aseguradora está en ruina si para algún tiempo $t \geq 0$ su capital es negativo, es decir:

$$W(t) < 0, \text{ para } t \geq 0.$$

El momento de ruina se define como

$$T = \text{mín}\{t \geq 0 : W(t) < 0\},$$

mientras que la probabilidad de ruina está dada por

$$\phi(u) = P(W(t) < 0 | W(0) = u), \quad \text{para } t > 0.$$

En [8] se define el coeficiente de ajuste como el valor constante $R > 0$ tal que

$$\frac{\lambda}{c} \int_0^{\infty} \exp(RZ)[1 - F(Z)]dZ = 1.$$

Por otro lado, de existir el coeficiente de ajuste R , para todo $u \geq 0$ se tiene que

$$\phi(u) < \exp(-Ru)$$

esta desigualdad se conoce como la Desigualdad de Crámer-Lundberg y su importancia radica en que su lado derecho puede ser considerado como una medida del riesgo de la aseguradora.

2.2.2 Procesos de Wiener

En 1827, el botánico escocés Robert Brown observó que al suspender partículas de polen en agua, estas adquirirían un movimiento aleatorio. La primera conclusión a la que llegó Brown fue que el movimiento observado se debía a que las partículas de polen tenían vida. Tiempo después realizó nuevamente el experimento, pero esta vez con partículas de minerales, el resultado de su observación fue el mismo. Ante este hecho sus conclusiones cambiaron, señalando que al suspender partículas de pequeñas dimensiones y de todo tipo en soluciones como el agua se observa un movimiento desordenado o aleatorio de las mismas. Posteriormente nombraron a este fenómeno movimiento Browniano.

En 1905, Albert Einstein, en su artículo “Sobre el movimiento requerido por la teoría cinética molecular del calor de pequeñas partículas suspendidas en un líquido estacionario ” logró predecir el movimiento de las partículas descritas por Brown además de darle una interpretación física.

Entre 1920 y 1923 Wiener desarrolló un modelo matemático que describe el movimiento browniano.

Un proceso estocástico $\{W(t)\}$ con $-\infty < t < +\infty$ de segundo orden, es decir

$$E[W(t)^2] < +\infty$$

se dice un proceso de Wiener si verifica lo siguiente:

1. $W(0) = 0$.
2. $W(t) - W(s)$ con $s \leq t$ es normalmente distribuido con media 0 y varianza $\sigma^2(t - s)$, donde σ^2 representa un valor constante positivo.

3. $W(t_2) - W(t_1), W(t_3) - W(t_2), \dots, W(t_n) - W(t_{n-1})$ con $t_1 \leq t_2 \leq t_3 \leq \dots \leq t_n$ son independientes.

Propiedad 5. Las propiedades más importantes de un proceso de Wiener $W(t)$ son:

1. $E[W(t)] = 0$.
2. $E[(W(t_2) - W(t_1))(W(t_4) - W(t_3))] = 0, t_1 \leq t_2 \leq t_3 \leq t_4$.

3.

$$Cov(s, t) = \begin{cases} \sigma^2 \min\{|s|, |t|\}, & st > 0; \\ 0, & st \leq 0. \end{cases}$$

4. $E[(W(t) - W(r))(W(s) - W(r))] = \sigma^2 \min\{s - r, t - r\}$, con $s \geq r$ y $t \geq r$.

5. Un proceso de Wiener es un proceso Guassiano, es decir

$$b_1W(t_1) + b_2W(t_2) \dots + b_nW(t_n)$$

está normalmente distribuido, con b_1, b_2, \dots, b_n constantes.

Con los procesos de Wiener se da un nuevo tipo de teorema limite, el llamado teorema del límite central funcional, el resultado de su aplicación es similar al obtenido con el teorema central del límite ordinario con la consideración de que el lugar de tomar una sucesión de variables aleatorias se toma una sucesión de funciones, hecho del cual se deriva su nombre.⁵

2.2.3 Estadístico de Wilcoxon y su aproximación a la distribución normal estándar

Sean X_1, \dots, X_{n1} y Y_1, \dots, Y_{n2} muestras independientes con F y G continuas sus respectivas funciones de distribución. Se desea probar

$$H_0 : F = G$$

$$H_1 : P(Y > t) \geq P(X > t)$$

Nótese que la hipótesis alternativa puede ser expresada como

$$H_1 : G(t) \leq F(t)$$

representa que Y es estocásticamente mayor a X . Si no se conoce nada sobre la forma de F y G es razonable probar H_0 versus H_1 utilizando el orden de las variables X y Y

⁵Para más referencias véase [9].

en lugar de sus valores, esta información es proporcionada por los rangos R_1, R_2, \dots, R_n con $n = n_1 + n_2$.

A la observación con el menor valor dentro de la muestra combinada se le asigna rango 1, a la siguiente 2 y así consecutivamente. Sean

$$\begin{aligned} R_1 & \text{ rango de } X_1 \\ R_2 & \text{ rango de } X_2 \\ & \vdots \\ R_{n_1} & \text{ rango de } Y_1 \\ & \vdots \\ R_n & \text{ rango de } Y_{n_2} \end{aligned}$$

Se define el estadístico de Wilcoxon como

$$W = \sum_{i=n_1+1}^n R_i$$

es decir, como la suma de los rangos de la segunda muestra. W es tal que si

$$P(W \geq c) = \alpha$$

entonces $F = G$, caso contrario si

$$P(W \geq c) \leq \alpha$$

se dice que Y es estocásticamente menor a X .

Para los valores de $n_1, n_2 > 8$ se utiliza la aproximación a la normal⁶, se define para esto

$$W - \frac{1}{2}n_2(n_2 + 1) = U$$

donde U representa el número de pares (X_i, Y_j) para los cuales $X_i < Y_j$. Sea

$$h_{ij} = \begin{cases} 1, & X_i < Y_j \\ 0, & \text{caso contrario} \end{cases}$$

así

$$\sum_{j=1}^{n_2} \sum_{i=1}^{n_1} h_{ij}$$

⁶Para n_1 y n_2 menores el valor de W está calculado en tablas.

ahora $E(h_{ij}) = P(X < Y) = p$, con $p = \frac{1}{2}$ se tiene que

$$E(U) = \frac{1}{2}n_1n_2$$
$$Var(U) = \frac{1}{12}n_1n_2(n+1)$$

La distribución asintótica de U al estandarizarla es normal. Ya que W no es la suma de variables aleatoria independientes no se puede aplicar el Teorema del Límite Central, pero se establece la siguiente aproximación.

Teorema 23 ([37], p.349). Sean F y G continuas, $0 < P(X < Y) < 1$, se tiene entonces que

$$P\left(\frac{U - E(U)}{\sqrt{Var(U)}} \leq t\right) \rightarrow \phi(t)$$

con $\min(n_1, n_2) \rightarrow \infty$, donde ϕ representa la distribución normal estándar.

Para más referencias y ejemplos sobre el estadístico de Wilcoxon ver ([37],p.345)

Los dos casos presentados anteriormente, son solo dos ejemplos de las adaptaciones del teorema del límite central en diferentes tipos de procesos como los de renovación y los de Wiener, así como aproximaciones a la distribución normal estándar que cumplen ciertos estadísticos aún cuando no se verifican las condiciones del teorema de límite central.

En la literatura matemática se encuentran otras formas del teorema, por ejemplo, el teorema del límite central para variables aleatorias dependientes o el teorema del límite central en martingalas, resultado que será desarrollado a profundidad en el siguiente capítulo.

Capítulo 3

Martingalas y el Teorema del Límite Central

3.1 Conceptos previos a la definición de las martingalas

En la presente sección se tratarán conceptos fundamentales previos a la definición de las martingalas, entre ellos el de esperanza condicional y sus propiedades más importantes así como tiempos de parada.

3.1.1 Esperanza condicional

Sean (Ω, \mathcal{A}, P) espacio probabilístico, X y Y variables aleatorias discretas que toman valores $\{x_1, \dots, x_n\}$ y $\{y_1, \dots, y_m\}$ respectivamente, se tiene que

$$E(X|Y) : \Omega \rightarrow \mathbb{R}$$
$$w \mapsto E(X|Y)(w) = E(X|Y = y_j)$$

es una variable aleatoria respecto a $\sigma(Y)$, donde

$$Y : \Omega \rightarrow \mathbb{R}$$
$$w \mapsto Y(w) = y_j$$

y

$$\sigma(Y) = \{Y^{-1}(A'), A' \in \mathcal{B}(\mathbb{R})\}$$

Propiedad 6. Sea $A \in \sigma(Y)$, entonces

$$\int_A E(X|Y)(w)dP(w) = \int_A X(w)dP(w)$$

Nótese que $E(X|Y)$ también se expresa como $E(X|\sigma(Y))$.

Definición 28 (Sub σ -álgebra). Sea Ω, \mathcal{A} espacio medible, se dice que \mathcal{F} es una sub σ -álgebra de \mathcal{A} si:

1. \mathcal{F} es una σ -álgebra de Ω .
2. $\mathcal{F} \subseteq \mathcal{A}$.

Definición 29 (Esperanza condicionada respecto a una σ -álgebra). Sean (Ω, \mathcal{A}, P) espacio probabilístico, X variable aleatoria tal que $E(X) < \infty$, \mathcal{F} sub σ -álgebra de \mathcal{A} . Existe entonces $E(X|\mathcal{F})$ variable aleatoria que verifica:

1. $E(X|\mathcal{F})$ es \mathcal{F} -medible.
2. $E[E(X|\mathcal{F})] < +\infty$.
3. $\forall F \in \mathcal{F}$,

$$\int_F E(X|\mathcal{F})(w)dP(w) = \int_F X(w)dP(w).$$

Propiedad 7. Sea X como en la definición anterior, la esperanza condicional cumple las siguientes propiedades:

1. Si existe una variable aleatoria \tilde{X} que verifique las condiciones de la definición anterior, se tiene entonces que

$$\tilde{X} = E(X|\mathcal{F}) \text{ c.s.}$$

2. $E[E(X|\mathcal{F})] = E(X)$.
3. Si X es \mathcal{F} -medible, entonces

$$X = E(X|\mathcal{F}) \text{ c.s.}$$

4. Sean $a, b \in \mathbb{R}$, X y Y variables aleatorias, entonces

$$E(aX + bY|\mathcal{F}) = aE(X|\mathcal{F}) + bE(Y|\mathcal{F}) \text{ c.s.}$$

5. Si $X \geq 0$, entonces $E(X|\mathcal{F}) \geq 0$ c.s.

6. Sean \mathcal{F} y \mathcal{G} dos sub σ -álgebras de \mathcal{A} tales que $\mathcal{G} \subseteq \mathcal{F}$, se tiene que

$$E[E(X|\mathcal{F})|\mathcal{G}] = E(X|\mathcal{G}) \text{ c.s.}$$

esta propiedad recibe el nombre de Propiedad de la Torre.

7. Sea Y variable aleatoria \mathcal{F} -medible y acotada, entonces

$$E(XY|\mathcal{F}) = YE(X|\mathcal{F}) \text{ c.s.}$$

8. Si X es independiente de \mathcal{F} , entonces $E(X|\mathcal{F}) = E(X)$

En el caso de vectores aleatorios, la esperanza y la varianza condicional se definen como sigue:

Definición 30 (Esperanza condicional - vector aleatorio). Sea X un vector aleatorio de \mathbb{R}^k , Y un vector aleatorio de \mathbb{R}^q y una función medible

$$h : \mathbb{R}^q \rightarrow \mathbb{R}$$

se define la esperanza condicional de $h(Y)$ dado $X = x$ como

- Caso discreto

$$E(h(Y)|X = x) = \sum_{y_1} \dots \sum_{y_q} h(y_1, \dots, y_q) f(y|x).$$

- Caso continuo

$$E(h(Y)|X = x) = \int_{\mathbb{R}^q} h(y_1, \dots, y_q) f(y|x) dy.$$

Nótese que $E(h(Y)|X)$ es una variable aleatoria, en efecto

$$\begin{aligned} E(h(Y)|X) : \mathbb{R}^k &\rightarrow \mathbb{R} \\ x &\mapsto E(h(Y)|X = x) \end{aligned}$$

donde $E(h(Y)|X = x)$ con $x \in \mathbb{R}^k$ son los posibles valores que toma la variable aleatoria.

Definición 31 (Varianza condicional - vector aleatorio). La varianza condicional de una variable aleatoria X respecto de una variable Y está dada por

$$\text{Var}(X|Y) = E(X^2|Y) - (E(X|Y))^2 \tag{3.1}$$

La varianza de X verifica la siguiente igualdad:

$$\text{Var}(X) = E[\text{Var}(X|Y)] + \text{Var}[E(X|Y)]$$

3.1.2 Tiempo de parada

Sea $\{X_n, n \geq 0\}$ un proceso estocástico con espacio probabilístico (Ω, \mathcal{A}, P) , sea T una variable aleatoria discreta

$$T : \Omega \rightarrow \mathbb{N} \cup \{\infty\}$$

X_T representa en qué estado se encuentra la variable en el tiempo T . La idea básica del tiempo de parada es observar la secuencia de valores que toma la variable en los instantes $\{0, 1, \dots, n\}$, basando la decisión de detenerse en el instante n en la información que se ha obtenido hasta dicho momento.

Definición 32 (Tiempo de parada). *El tiempo de parada respecto al proceso estocástico $\{X_n, n \geq 0\}$ es un tiempo aleatorio para el cual el evento $\{T = n\}$ queda totalmente determinado por la información obtenida de la variable hasta el tiempo n , es decir por los $\{X_0, X_1, \dots, X_n\}$.*

En el contexto de los juegos de azar, si X_n representa la ganancia obtenida hasta el instante n , el tiempo de paro T representa el instante en el cual se detiene el juego. La decisión de detenerse depende de la información obtenida hasta un determinado momento.

Para ilustrar el concepto de tiempo de parada, se utilizará el mismo en el problema de *La Ruina del Jugador*.

Ejemplo 4 (La ruina del jugador). *El jugador empieza el juego con una fortuna inicial de X_0 unidades de dinero, en cada jugada gana 1 unidad con probabilidad p o pierde la misma cantidad con probabilidad $1-p$. El juego termina cuando la fortuna del apostador alcanza un valor de N previamente establecido, donde $0 < X_0 < N$, o cuando el jugador queda en ruina, es decir, cuando se queda sin dinero para seguir apostando ([19], p.11).*

El proceso señalado anteriormente puede ser modelado como una marcha aleatoria. En efecto, sea $\{\Delta_n\}_{n \geq 1}$ la variable aleatoria que representa la ganancia (o pérdida) en cada jugada, su función de probabilidad está dada por:

$$P(\Delta_n = x) = \begin{cases} p, & x = 1 \\ 1 - p, & x = -1 \end{cases}$$

Así

$$X_n = X_0 + \Delta_1 + \Delta_2 + \dots + \Delta_n,$$

donde X_n representa el monto total obtenido hasta el instante n .

El juego anterior se detiene cuando $X_n = 0$ o $X_n = N$, se define el tiempo de parada para el problema de la ruina del jugador como:

$$T = \min_n \{n > 0 : X_n \in A\}$$

donde $A = \{0, N\}$.¹

A continuación se verifica que T define un tiempo de parada, es decir, que el evento $\{T = n\}$ queda determinado por $\{X_0, X_1, \dots, X_n\}$, la información obtenida hasta el n -ésimo momento.

Demostración.

- Para $n = 0$

$$\{T = 0\} = \{X_0 \in A\}$$

- Para $n \geq 1$

$$\{T = n\} = \{X_0 \notin A, X_1 \notin A, \dots, X_{n-1} \notin A, X_n \in A\}$$

Obsérvese que el evento $\{T = n\}$ depende únicamente de los valores que tome la variable en los instantes X_0, X_1, \dots, X_n , como es requerido en la definición.²

□

A continuación se simulará el tiempo de parada para el problema de la ruina del jugador. El algoritmo para hacerlo fue desarrollado en *Matlab*.

```
function [ x, j ] = tiempo_parada( N, p, x0 )
x : Ganancia al terminar el juego 0 o N.
j : Número de iteraciones que toma terminar el juego.
N : El valor objetivo con el cual se gana el juego.
p : Probabilidad de ganar.
x0 : Capital inicial de jugador.
```

¹El tiempo de parada así definido se conoce como tiempo de paro de primera visita. [7]

²Para más referencias sobre tiempos de paro en el problema de la ruina del jugador véase [19].

```

for j=1:500
r=rand();
if p<r
    x0=x0+1;
else
    x0=x0-1;
end
if x0==0
    break
end
if x0==N
    break
end
end
x=x0;
end

```

Cuadro 3.1: Algoritmo tiempo de paro en el problema ruina del jugador

Seguido se graficarán dos escenarios de cómo varía la fortuna del jugador en el tiempo, tanto para el caso en el que gana la apuesta como para el que queda en ruina. Los parámetros utilizados en ambos casos son:

N=	10
p=	0,4
X0=	1

Cuadro 3.2: Parámetros utilizados simulación tiempo de parada

```

>> [y,ny]=tiempo_parada(10,0.4,1)
>> z=[1 2 1 2 3 2 1 2 3 2 3 2 1 0];
>> ny=[0:13];
>> plot(ny,y);grid on;xlabel('No. Apuestas');ylabel('Capital
obtenido');title('Ruina del jugador-Pérdida')

```

Cuadro 3.3: Ruina del jugador - Pérdida

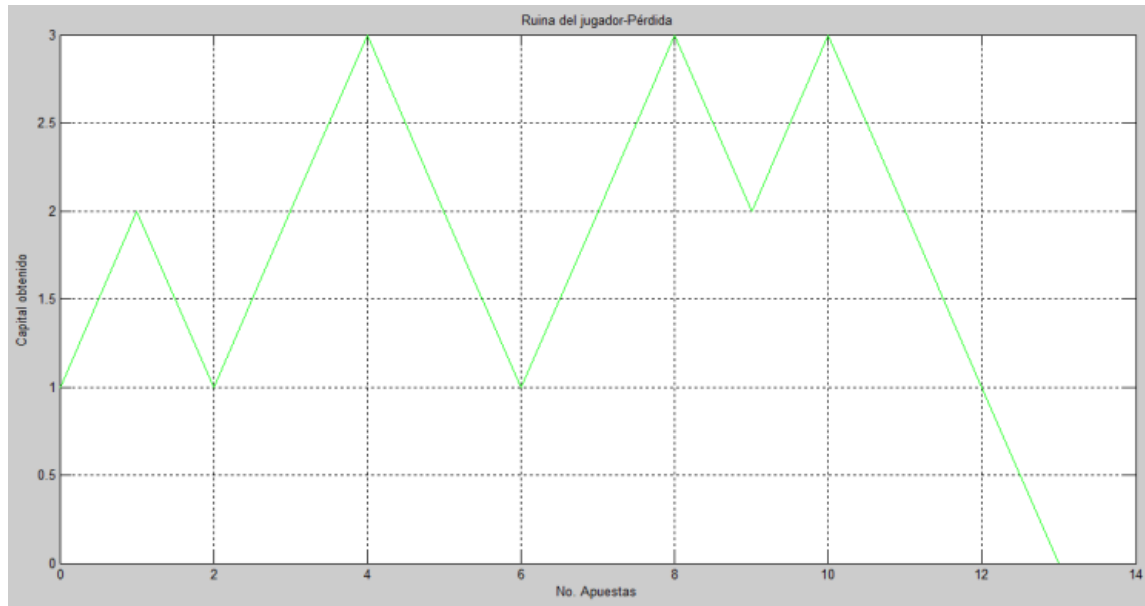


Figura 3.1: Gráfica Ruina Jugador - Pérdida

```
>> [z,nz]=tiempo_parada(10,0.4,1)
>> z=[1 2 3 2 3 4 3 2 3 4 5 4 5 6 7 8 7 8 9 10];
>> nz=[0:19];
>> plot(nz,z);grid on;xlabel('No. Apuestas');ylabel('Capital
obtenido');title('Ruina del jugador-Ganancia')
```

Cuadro 3.4: Ruina del jugador - Ganancia

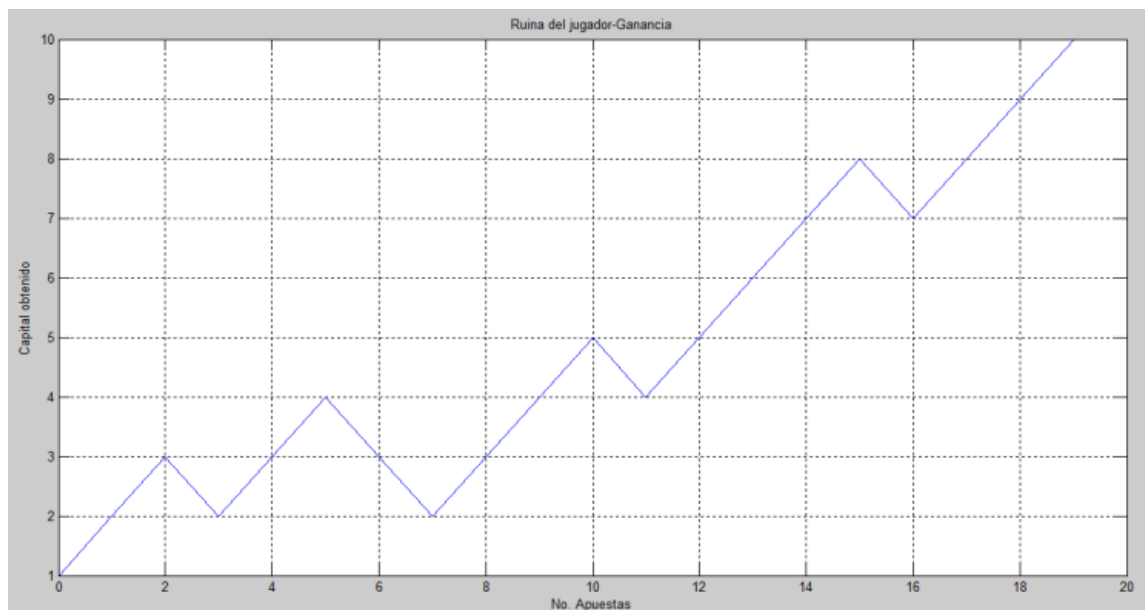


Figura 3.2: Gráfica Ruina Jugador - Ganancia

Las trayectorias graficadas corresponden cada una a una corrida el algoritmo, para cuando se registra una pérdida y una ganancia respectivamente. Al realizar el juego varias veces (50 veces) con el algoritmo y los parámetros establecidos se obtienen los siguientes resultados:

No. Juegos	Ganancia	Tiempo Paro	No. Juegos	Ganancia	Tiempo Paro
1	10	31	26	0	17
2	0	11	27	10	25
3	10	23	28	0	1
4	10	29	29	0	1
5	0	1	30	0	1
6	0	1	31	0	1
7	0	3	32	10	21
8	0	1	33	0	1
9	0	23	34	10	31
10	0	5	35	10	19
11	0	1	36	0	1
12	0	1	37	0	1
13	0	1	38	0	1
14	10	35	39	0	17
15	0	5	40	10	31
16	10	59	41	0	7
17	0	1	42	0	1
18	0	1	43	10	17
19	0	1	44	0	3
20	0	23	45	0	25
21	0	1	46	0	1
22	0	1	47	10	23
23	0	1	48	10	19
24	0	1	49	0	3
25	0	1	50	0	3

Cuadro 3.5: Resultado simulaciones Ganancia - Tiempo de Paro

3.1.3 Identidad de Wald

Antes de introducir la identidad de Wald como tal, es necesario revisar los conceptos de filtración y tiempo de parada respecto a una filtración.

Definición 33 (Filtración). *Sea (Ω, \mathcal{A}, P) espacio probabilístico, una sucesión creciente de sub σ -álgebras de \mathcal{A} , $\{\mathcal{F}_n\}_{n \in \mathbb{N}}$, $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \dots$ se denomina filtración.*

Sea $\{X_n\}_{n \in \mathbb{N}}$ una sucesión de variables aleatorias, si se define \mathcal{F}_t como

$$\mathcal{F}_t = \sigma(\{X_s : 0 \leq s \leq t\}), \quad \forall t \in \mathbb{N}$$

$\{\mathcal{F}_n\}_{n \in \mathbb{N}}$ se dice filtración natural de $\{X_n\}_{n \in \mathbb{N}}$. Una sub σ -álgebra así definida se interpreta como la historia del proceso hasta el tiempo t , es decir \mathcal{F}_t contiene todos los posibles eventos generados por la variable aleatoria durante el intervalo de tiempo $[0, t]$ ([11], p.28).

Definición 34 (Tiempo de parada respecto a una filtración). Sean $T : \Omega \rightarrow \mathbb{N} \cup \{\infty\}$ una variable aleatoria y $\{\mathcal{F}_n\}_{n \in \mathbb{N}}$ una filtración se dice que T es un tiempo de paro respecto a \mathcal{F}_n si el conjunto $A := \{w \in \Omega : T(w) \leq n\}$ es tal que

$$A \in \mathcal{F}_n$$

La variable aleatoria T , tiempo de paro respecto a la filtración $\{\mathcal{F}_n\}_{n \in \mathbb{N}}$, permite saber si hasta la n -ésima experiencia aleatoria, el valor de $T(w)$ es menor al de el n dado. Una de las aplicaciones más importantes de los tiempos de parada es el de la identidad o ecuación de Wald.

Teorema 24 (Identidad de Wald). Sea $\{X_i\}_{i \in \mathbb{N}}$ una familia de variables aleatorias i.i.d. tal que $E(X_i) = E(X) < \infty$. Si T es un tiempo de parada respecto de los $\{X_i\}_{n \in \mathbb{N}}$ y es tal que $E(T) < \infty$, entonces

$$E\left(\sum_{i=1}^T X_i\right) = E(T)E(X)$$

La demostración de este teorema puede ser encontrada en ([15], p.487). Existe una identidad para la varianza, adicional a las hipótesis de la identidad anterior se requiere que $E(X_i^2) = E(X^2) < \infty$ entonces

$$Var\left(\sum_{i=1}^T X_i\right) = E(T)E(X^2)$$

La demostración de la identidad referida por Wald para la varianza se encuentra en ([17], p.166).

El siguiente ejemplo muestra la importancia de la verificación de las hipótesis para la aplicación del teorema

Ejemplo 5.

1. Sean X_1, X_2, \dots variables aleatorias independientes tales que

$$P(X_n = 0) = P(X_n = 1) = \frac{1}{2}, \quad \forall n \geq 1$$

y

$$T = \min\{n \geq 1 : \sum_{i=1}^n X_i = 10\} \text{ tiempo de parada.}$$

Se tiene que

$$E(X_i) = \frac{1}{2}, \quad \forall i \geq 1$$

por otro lado

$$\sum_{i=1}^T X_i = 10$$
$$E\left(\sum_{i=1}^T X_i\right) = 10$$

aplicando la identidad de Wald

$$E(T)E(X_i) \stackrel{1/2}{=} 10$$

de donde

$$E(T) = 20 < \infty$$

2. Sean X_1, X_2, \dots variables aleatorias independientes tales que

$$P(X_n = -1) = P(X_n = 1) = \frac{1}{2}, \quad \forall n \geq 1.$$

Se tiene que

$$E(X_i) = 0, \quad \forall i \geq 1$$

nuevamente

$$\sum_{i=1}^T X_i = 10$$
$$E\left(\sum_{i=1}^T X_i\right) = 10$$

aplicando la identidad de Wald

$$E(T)E(X_i) = 10$$

pero $E(X_i) = 0$, lo que conduce a una contradicción. Para que se cumpla esta identidad es necesario que $E(T) < \infty$, condición que no se verifica.

La identidad de Wald es de gran utilidad en la rama de la matemática actuarial, por

ejemplo, para una aseguradora de autos es importante conocer el valor total a pagar por los autos accidentados que contaban con una póliza. Se observan dos variables aleatorias

$X_i =$ Valor a pagar o cuantía de cada vehículo i accidentado.

$N =$ Total de accidentes reportados.

Se define $S = \sum_{i=1}^N X_i$ como el costo total a cancelar por parte de la aseguradora. Ya que no se conoce el número total de siniestros no se puede saber el valor exacto de S antes de terminar el periodo de estudio, sin embargo se puede conocer su valor esperado, valor que resulta de vital importancia ya que le permitirá a la aseguradora prever una determinada cantidad de dinero antes que culmine el periodo de reclamaciones. Para resolver el problema se supone que:

1. La variable N cuenta el número de accidentes en un determinado periodo, sigue una distribución de probabilidad discreta, la que mejor modela este tipo de procesos es la distribución de Poisson, la binomial negativa o la binomial, donde los parámetros necesarios para emplear cada distribución serán estimados a partir de las observaciones anteriores.
2. Las variables X_i son i.i.d., es decir, los montos de cada uno de los accidentes vehiculares siguen una misma distribución de probabilidad, en la mayoría de los casos esta variable aleatoria se modela con distribuciones de probabilidad continuas entre ellas la distribución exponencial, log-normal, normal, entre otras. Es importante notar que el costo de cada accidente X_i no depende del número total de siniestros N .

A continuación se realizará la simulación para constatar que efectivamente el número de accidentes y la cuantía de los mismos siguen distribuciones discreta y continua respectivamente, los ejemplos siguientes, así como sus simulaciones siguen el esquema planteado en [16].

- Número de accidentes.

Se dispone de los datos de las pólizas de una aseguradora de vehículos, es decir, el número de accidentes que han reportado sus asegurados en un periodo determinado.

Se ajustará la distribución de frecuencias observadas a una distribución binomial negativa, mediante la prueba bondad de ajuste se comprobará si la distribución discreta escogida modela adecuadamente los datos.

No. Accidentes	No. Pólizas
0	6299
1	441
2	50
3	10
4	2
5	0

Cuadro 3.6: Frecuencia número de accidentes

Para emplear la distribución binomial negativa se debe estimar el número de intentos (\hat{n}) y la proporción (\hat{p}), las estimaciones están dadas por³:

$$\hat{n} = \frac{\bar{N}^2}{\bar{N} - s^2}$$

y

$$\hat{p} = \frac{\bar{N} - s^2}{\bar{N}}$$

donde \bar{N} representa la media de los datos y s^2 la varianza de los mismos.

El código siguiente muestra el cálculo de la media, la varianza y los parámetros para la distribución binomial negativa, en R.

```
> accidentes=0:4
> polizas=c(6299,441,50,10,2)
> media=weighted.mean(accidentes , polizas ); media
[1] 0.08512202
> var=sum(( accidentes -media) ^2* polizas )/sum( polizas ); var
[1] 0.1049271
> n=media ^2/( varianza -media) ; n
[1] 0.365853
> p=(varianza -media) /media ; p
[1] 0.2326673
```

Cuadro 3.7: Parámetros distribución binomial negativa

A continuación se calculan las probabilidades teóricas según la distribución binomial negativa con los parámetros antes señalados, se completa con la probabilidad de que más de 4 accidentes sucedan para que la suma de todas ellas sea 1, finalmente con estas probabilidades se calcula las frecuencias teóricas para el número de accidentes.

³La estimación de los parámetros se realiza por el método de momentos.

```

> bin_neg=dnbinom( accidentes ,n,1/(1+p));bin_neg
[1] 0.9263258129 0.0639675678 0.0082456185 0.0012273804
0.0001949413
> u=1-sum(bin_neg)
> bin_neg_completa=c(bin_neg ,u);bin_neg_completa
[1] 9.263258e-01 6.396757e-02 8.245618e-03 1.227380e-03 1.949413e
-04
[6] 3.867921e-05
> fteo=sum(polizas)*bin_neg_completa;fteo
[1] 6300.868179 435.107396 56.086697 8.348641 1.325990
0.263096

```

Cuadro 3.8: Frecuencias teóricas del número de accidentes

Se procede a agrupar las frecuencias teóricas menores a 5, debiendo reducir al mismo número las frecuencias observadas.

```

> fteo1=c(fteo [1:3],sum(fteo [4:6]));fteo1
[1] 6300.868179 435.107396 56.086697 9.937728
> polizas1=c(polizas [1:3],sum(polizas [4:5]));polizas1
[1] 6299 441 50 12

```

Cuadro 3.9: Frecuencias teóricas del número de accidentes agrupadas

Finalmente se realiza la prueba de bondad de ajuste para comprobar si la distribución elegida modela adecuadamente a las observaciones.

No. Accidentes	Frecuencia observada	Frecuencia teórica	Diferencia
0	6299	6300,86	0,00055391
1	441	435,10	0,07980278
2	50	56,08	0,66054666
mayor a 3	12	9,937	0,42796158
Total	6802	6802	1,168865

Cuadro 3.10: Prueba bondad de ajuste

Para los valores usuales de α se calcula el valor del fractil de la distribución χ_1^2 .

α	χ_t^2
0,1	2,70554
0,05	3,84146
0,01	6,6349

Cuadro 3.11: Fractil distribución χ^2

Ya que el valor calculado es menor que los valores tabulados se acepta que el número de accidentes se distribuyen de acuerdo a una binomial negativa.

- Cuantía de los accidentes.

En el último periodo se registraron un total de 200 accidentes, el valor que tuvo que pagar la aseguradora por los mismos, es decir la cuantía, se distribuyó de la siguiente forma:

Cuantía	No. Accidentes
0 - 400	5
400 - 800	48
800 - 1200	64
1200 - 1600	42
1600 - 2000	20
2000 - 2400	12
2400 - 2800	6
2800 - 3200	2
3200 - 3600	1
3600 - 4000	0

Cuadro 3.12: Cuantía de los accidentes

Se ajustará la distribución observada a una distribución log –normal y mediante la prueba de bondad de ajuste se verificará si efectivamente la distribución escogida modela correctamente los datos.

Para emplear esta distribución se deben estimar la media y la desviación estándar del proceso, mismas que están dadas por:

$$\hat{\mu} = \ln \left(\frac{\bar{X}}{\sqrt{1 + a^2}} \right)$$

y

$$\hat{\sigma} = \sqrt{\ln(1 + a^2)}$$

donde

$$a = \frac{s}{\bar{X}}$$

con s la desviación estándar y \bar{X} la media de las observaciones. El siguiente código muestra el cálculo de los parámetros de la distribución log –normal, en R.

```

> cuantia=c(200, 600,1000, 1400,1800,2200, 2600, 3000, 3400);
  cuantia
[1] 200 600 1000 1400 1800 2200 2600 3000 3400
> accidentes=c(5,48,64,42,20,12,6,2,1);accidentes
[1] 5 48 64 42 20 12 6 2 1
> media=weighted.mean(cuantia , accidentes);media
[1] 1200
> varianza=sum(( cuantia –media) ^2* accidentes)/sum(accidentes);
  varianza
[1] 344000
> a=sqrt(varianza)/media;a
[1] 0.4887626
> mu=log(media/sqrt(1+a^2));mu
[1] 6.982969
> sigma=sqrt(log(1+a^2));sigma
[1] 0.4628336

```

Cuadro 3.13: Parámetros distribución log –normal

Nótese que el vector “cuantía” corresponde a los puntos medios de los valores observados, a continuación se calcula las probabilidades teóricas siguiendo una distribución log –normal, se completa la misma hasta que el valor de la probabilidad acumulada sea 1 y finalmente se encuentra las frecuencias teóricas para la cuantía de los accidentes.

```

> lim_inf=c(0,400,800,1200,1600,2000,2400,2800,3200,3600);lim_inf
[1] 0 400 800 1200 1600 2000 2400 2800 3200 3600
> p_acum=plnorm(lim_inf ,mu, sigma);p_acum
[1] 0.00000000 0.01608672 0.25958263 0.59150449 0.80316585
  0.90907960
[7] 0.95809843 0.98040050 0.99062842 0.99540740
> p_acum_completa=c(p_acum ,1);p_acum_completa
[1] 0.00000000 0.01608672 0.25958263 0.59150449 0.80316585
  0.90907960
[7] 0.95809843 0.98040050 0.99062842 0.99540740 1.00000000
> p_log_normal=diff(p_acum_completa);p_log_normal
[1] 0.016086720 0.243495907 0.331921861 0.211661361 0.105913746
  0.049018832
[7] 0.022302074 0.010227919 0.004778975 0.004592604
> fteo=p_log_normal*sum(accidentes);fteo
[1] 3.2173440 48.6991814 66.3843722 42.3322722 21.1827492
  9.8037663
[7] 4.4604148 2.0455839 0.9557951 0.9185208

```

Cuadro 3.14: Frecuencias teóricas de la cuantía de los accidentes

Se procede a agrupar las frecuencias teóricas menores a 5, debiendo reducir al

mismo número las frecuencias observadas.

```
> fteo1=c(sum(fteo [1:2]), fteo [3:6], sum(fteo [7:10])); fteo1
[1] 51.916525 66.384372 42.332272 21.182749 9.803766 8.380315
> accidentes1=c(sum(accidentes [1:2]), accidentes [3:6], sum(accidentes
[7:9])); accidentes1
[1] 53 64 42 20 12 9
```

Cuadro 3.15: Frecuencias teóricas de la cuantía agrupadas

Finalmente se realiza la prueba de bondad de ajuste para comprobar si la distribución elegida modela correctamente a las observaciones.

Cuantía	Frecuencia observada	Frecuencia teórica	Diferencia
0 – 800	53	51,91	0,02261165
800 – 1200	64	66,38	0,08564109
1200 – 1600	42	42,33	0,00260805
1600 – 2000	20	21,18	0,06603936
2000 – 2400	12	9,80	0,49100007
mayor a 2400	9	8,38	0,0458228
Total	200	200	0,414722

Cuadro 3.16: Prueba bondad de ajuste

Se calcula el valor del fractil de la distribución χ_2 con 3 grados de libertad para los valores usuales de α

α	χ_t^2
0,1	6,25139
0,05	7,81473
0,01	11,3449

Cuadro 3.17: Fractil distribución χ^2

Ya que el valor calculado es menor al de los valores tabulados se acepta que la cuantía de los accidentes se distribuyen de acuerdo a una log –normal.

Una vez verificado el tipo de distribuciones que siguen en número de accidentes y el precio que paga la aseguradora por siniestro (cuantía), se retoma la aplicación de la identidad de Wald para conocer el valor esperado del costo total

$$S = \sum_{i=1}^N X_i$$

así,

$$E(S) = E(N)E(X) \quad (3.2)$$

donde $E(N)$ es el valor esperado del total de accidentes y $E(X)$ el valor esperado de la cuantía de los accidentes. Existen diversos métodos para obtener la función de distribución de la variable aleatoria S , uno de ellos es el cálculo mediante convoluciones⁴. Obtenida la distribución de S se puede calcular su esperanza, valor que será contrastado con el obtenido mediante la ecuación (3.2).

- Costo total.

De periodos anteriores una aseguradora de vehículos conoce que el número de accidentes que se reportan sigue una distribución de Poisson con media $\lambda = 0,8$, mientras que la cuantía de los siniestros X sigue la siguiente distribución⁵

X	$P(X = x)$	$P(X \leq x)$
1	0,3	0,3
2	0,7	1

Cuadro 3.18: Distribución de la cuantía de los accidentes

El valor máximo del número de accidentes reportados por un asegurado no es conocido, se empieza entonces el estudio suponiendo un valor relativamente alto, para reducirlo se considerará aquellos valores cuya probabilidad sea inferior a 0,001 como 0.

```
> k=0:14
> pn=dpois(k,0.8);pn
[1] 4.493290e-01 3.594632e-01 1.437853e-01 3.834274e-02 7.668548e-03
[6] 1.226968e-03 1.635957e-04 1.869665e-05 1.869665e-06 1.661924e-07
[11] 1.329540e-08 9.669378e-10 6.446252e-11 3.966924e-12 2.266814e-13
```

Cuadro 3.19: Probabilidades del número de accidentes - algoritmo

Tomando en cuenta la consideración señalada

⁴La descripción de este y otros métodos se encuentra en [16]

⁵Se ha dicho que en la mayoría de los casos la cuantía sigue una distribución continua, por tratarse de un ejemplo se presenta una distribución discreta.

k	$P(N = k)$
0	0,449
1	0,359
2	0,144
3	0,038
4	0,01

Cuadro 3.20: Probabilidad del número de accidentes

Tomando los valores máximos que pueden tomar N y X se observa que la variable aleatoria S estará en un rango de 0 a 8. A continuación se procede a calcular la función de distribución de S , en R.

```
> px=c(0,0.3,0.7)
> pn=dpois(0:4,0.8)
> Fs=aggregateDist("convolution",model.freq=pn,model.sev=px)
> Fs(0:8)
[1] 4.493290e-01 5.571679e-01 8.217328e-01 8.831579e-01 9.609216e-01
[6] 9.784104e-01 9.935911e-01 9.967475e-01 9.9985887e-01
> c(Fs(0),diff(Fs(0:8)))
[1] 0.449328964 0.107838951 0.264564894 0.061425067 0.077763674
0.017488890
[7] 0.015180657 0.003156374 0.001841218
```

Cuadro 3.21: Distribución acumulada de S .

Para una mejor visualización se colocan los resultados anteriores en la siguiente tabla

S	$P(S = s)$	$F(s)$
0	0,449328964	0,449328964
1	0,107838951	0,557167915
2	0,264564894	0,821732809
3	0,061425067	0,883157876
4	0,077763674	0,96092155
5	0,01748889	0,97841044
6	0,015180657	0,993591097
7	0,003156374	0,996747471
8	0,001841218	1

Cuadro 3.22: Distribución de S

Calculando la esperanza de S con la distribución obtenida

$$E(S) = 1,360218$$

Por otro lado empleando la identidad de Wald se tiene que

$$E(S) = \lambda E(X) = 0,8 * 1,7 = 1,36$$

Se observa mediante el ejemplo la utilidad de la identidad planteada por Wald, ya que no siempre es posible, o fácil, calcular la distribución del costo total.

3.2 Martingalas

3.2.1 Origen de las martingalas

La Real Academia Española(RAE) define el término martingala como

“Artificio o astucia para engañar a alguien, o para otro fin.”

En el siglo XVIII nace la martingala como un método para obtener ganancias en los juegos de azar, bajo condiciones como contar con un capital inicial infinito. Consiste en doblar la apuesta cada vez que se registre una pérdida en el juego, al momento de la primera ganancia además de recuperar el capital perdido anteriormente se recibe el equivalente al monto inicial apostado.

La estrategia de la martingala surge con el propósito de generar ganancias al apostar en el lanzamiento de una moneda, se extendió a los juegos de casino, en particular a la ruleta, siglos más tarde Paul Pierre Lévy (1935) aún sin emplear el término martingala, si no como una generalización de la suma de variables aleatorias independientes, introduce formalmente este resultado.

3.2.2 Definición y primeros resultados

Sean (Ω, \mathcal{A}, P) espacio probabilístico, $\{\mathcal{F}_n\}_{n \in \mathbb{N}}$ una sucesión creciente de sub σ -álgebras de \mathcal{A} , $\{X_n\}_{n \in \mathbb{N}}$ una sucesión de variables aleatorias

$$X_n : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$

Se dice que $\{X_n, \mathcal{F}_n\}_{n \in \mathbb{N}}$ es una martingala si verifica:

(i) X_n es medible respecto a \mathcal{F}_n , es decir

$$X_n : (\Omega, \mathcal{F}_n) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$

es tal que

$$\forall A \in \mathcal{B}(\mathbb{R}), X_n^{-1}(A) \in \mathcal{F}_n.$$

$$(ii) \ E(|X_n|) < +\infty$$

$$(iii) \ E(X_n|\mathcal{F}_m) = X_m \quad \forall m < n.$$

Observación 10 (Submartingala, supermartingala y martingalas en reversa). $\{X_n, \mathcal{F}_n\}_{n \in \mathbb{N}}$ es una submartingala si en la definición anterior, la condición (iii), es reemplazada por

$$(iii') \ E(X_n|\mathcal{F}_m) \geq X_m \quad \forall m < n.$$

En cambio, $\{X_n, \mathcal{F}_n\}_{n \in \mathbb{N}}$ recibe el nombre de supermartingala, si se reemplaza (iii) por

$$(iii'') \ E(X_n|\mathcal{F}_m) \leq X_m \quad \forall m < n.$$

Sea ahora $\{\mathcal{F}_n\}_{n \in \mathbb{N}}$ una sucesión decreciente de sub σ -álgebras de \mathcal{A} , manteniendo la condiciones (i) y (ii) de la definición y cambiando (iii) por

$$(iii^*) \ E(X_n|\mathcal{F}_m) = X_m \quad \forall m > n.$$

$\{X_n, \mathcal{F}_n\}_{n \in \mathbb{N}}$ es llamada martingala en reversa.

A continuación se probará la definición de martingala en un ejemplo clásico de esta.

Ejemplo 6. Sean Y una variable aleatoria integrable, $\{\mathcal{F}_n\}_{n \in \mathbb{N}}$ una filtración y $X_n = E(Y|\mathcal{F}_n)$. $\{X_n, \mathcal{F}_n\}_{n \in \mathbb{N}}$ es una martingala.

(i) X_n es medible respecto a \mathcal{F}_n .

En efecto, de la definición de esperanza condicional se tiene que $X_n = E(Y|\mathcal{F}_n)$ es \mathcal{F}_n medible.

(ii) $E(|X_n|) < +\infty$.

Se tiene que:

$$\begin{aligned} E(|X_n|) &= E(|E(Y|\mathcal{F}_n)|) \\ &\leq E(E(|Y| | \mathcal{F}_n)) \\ &= E(|Y|) \\ &< +\infty \end{aligned}$$

La primera desigualdad se obtiene aplicando la desigualdad de Jensen a $|E(Y|\mathcal{F}_n)|$, mientras que la última resulta al ser Y una variable aleatoria integrable.

(iii) $E(X_n|\mathcal{F}_m) = X_m \quad \forall m < n.$

Sea $m < n$,

$$E(X_n|\mathcal{F}_m) = E[E(Y|\mathcal{F}_n)|\mathcal{F}_m]$$

$$\begin{aligned}
&= E(Y|\mathcal{F}_m) \\
&= X_m
\end{aligned}$$

La segunda desigualdad es resultado de aplicar la propiedad de la Torre, ya que $\mathcal{F}_m \subset \mathcal{F}_n$, es decir \mathcal{F}_m es sub σ -álgebra de \mathcal{F}_n .

Teorema 25 (Desigualdad de Doob-Kolmogorov). *Sea $\{X_n, \mathcal{F}_n\}_{n \in \mathbb{N}}$ una martingala, entonces*

$$P[\max_{1 \leq i \leq n} |X_i| \geq \epsilon] \leq \frac{1}{\epsilon^2} E(X_n^2), \quad \epsilon > 0$$

Observación 11. *Si $\{X_n, \mathcal{F}_n\}_{n \in \mathbb{N}}$ es una martingala, nótese que*

$$E(X_n) = E(X_{n+1}) \quad \forall n \geq 1$$

de donde

$$E(X_n) = E(X_1)$$

Si $\{X_n, \mathcal{F}_n\}_{n \in \mathbb{N}}$ es una submartingala o una supermartingala, entonces $E(X_n) \geq E(X_1)$ o $E(X_n) \leq E(X_1)$ respectivamente.

Teorema 26 (Martingala Frenada). *Sea $\{X_n, \mathcal{F}_n\}_{n \in \mathbb{N}}$ y T tiempo de paro respecto a la filtración $\{\mathcal{F}_n\}_{n \in \mathbb{N}}$, entonces*

$$\{X_{T \wedge n}, \mathcal{F}_n\}_{n \in \mathbb{N}}$$

es una martingala y

$$E(X_{T \wedge n}) = E(X_1) \quad n \in \mathbb{N}$$

donde $T \wedge n = \min\{T, n\}$.

Si $\{X_n, \mathcal{F}_n\}_{n \in \mathbb{N}}$ es una submartingala, se tiene entonces que $\{X_{T \wedge n}, \mathcal{F}_n\}_{n \in \mathbb{N}}$ también lo es y además $E(X_{T \wedge n}) \geq E(X_1)$. Similarmente si $\{X_n, \mathcal{F}_n\}_{n \in \mathbb{N}}$ es una supermartingala entonces $\{X_{T \wedge n}, \mathcal{F}_n\}_{n \in \mathbb{N}}$ también lo es y además $E(X_{T \wedge n}) \leq E(X_1)$.

El teorema anterior, así como su demostración se encuentra en ([12], p.143).

Observación 12. *Sean $\{X_n\}_{n \in \mathbb{N}}$ variables aleatorias independientes centradas, \mathcal{F}_n filtración natural de X_n y $S_n = \sum_{i=1}^n X_i$. Se tiene que $\{S_n, \mathcal{F}_n\}_{n \in \mathbb{N}}$ es una martingala. La demostración de este importante resultado se encuentra en ([39], p.99).*

Al ser $\{S_n, \mathcal{F}_n\}_{n \in \mathbb{N}}$ una martingala, se puede aplicar sobre él teoremas de convergencia tales como el teorema del límite central, la ley de los grandes números, entre otros.

Si la variable X_n representa la ganancia obtenida en el instante n , entonces S_n representa la ganancia obtenida durante el intervalo de tiempo $[0, n]$. Si $\{S_n, \mathcal{F}_n\}_{n \in \mathbb{N}}$ es

una martingala se dice que el juego es justo, ya que la ganancia esperada en el instante $n + 1$ dada la ganancia hasta el instante n es nula, es decir

$$\begin{aligned} E(X_{n+1}|\mathcal{F}_n) &= E(S_{n+1} - S_n|\mathcal{F}_n) \\ E(X_{n+1}|\mathcal{F}_n) &= S_n - S_n \\ &= 0. \end{aligned}$$

De igual forma, si $\{S_n, \mathcal{F}_n\}_{n \in \mathbb{N}}$ es una submartingala el juego será favorable

$$\begin{aligned} E(X_{n+1}|\mathcal{F}_n) &= E(S_{n+1} - S_n|\mathcal{F}_n) \\ &\geq S_n - S_n = 0, \end{aligned}$$

finalmente si $\{S_n, \mathcal{F}_n\}_{n \in \mathbb{N}}$ es una supermartingala el juego resulta desfavorable para el jugador

$$\begin{aligned} E(X_{n+1}|\mathcal{F}_n) &= E(S_{n+1} - S_n|\mathcal{F}_n) \\ &\leq S_n - S_n = 0. \end{aligned}$$

Teorema 27 (Teorema de convergencia de martingalas,[1],p.26). *Si $\{S_n, \mathcal{F}_n\}_{n \geq 1}$ es una submartingala acotada en L^1 , entonces S_n converge c.s. a la variable aleatoria X con $E(|X|) < \infty$.*

Corolario 2. *Sea $1 < p < \infty$. Si $\{S_n, \mathcal{F}_n\}_{n \geq 1}$ es una martingala y $\sup_{n \geq 1} E(|S_n|^p) < \infty$, entonces S_n converge en L^p con probabilidad 1.*

Es importante señalar que el resultado del corolario anterior no se aplica con $p = 1$.

Las martingalas son una sucesión de variables aleatorias que aunque tienen la misma esperanza⁶ tienden a crecer, si se tiene control sobre $\sup_{n \geq 1} E(|S_n|^p) < \infty$ se puede esperar que la sucesión no varíe demasiado e inclusive tenga límite.

3.3 Teorema del Límite Central para Martingalas

3.3.1 Ley de los grandes números

El estudio de las martingalas permite generalizar importantes resultados de la teoría de las probabilidades, en especial los teoremas límites tales como la ley de los grandes números y el teorema del límite central. La ley débil de los grandes trata de la conver-

⁶Ver **Observación 11**.

gencia en probabilidad de la suma de variables aleatorias, y dado que se ha demostrado que $\{S_n, \mathcal{F}_n\}_{n \in \mathbb{N}}$ es una martingala el teorema se aplica como sigue.

Teorema 28 (Ley débil de los grandes números en martingalas, [1],p.29).

Sea $\{S_n = \sum_{i=1}^n X_i, \mathcal{F}_n\}_{n \in \mathbb{N}}$ una martingala y $\{b_n\}$ una secuencia de constantes positivas tales que $b_n \rightarrow \infty$ cuando $n \rightarrow \infty$. Entonces, escribiendo $X_{ni} = X_i I(|X_i| \leq b_n)$, $1 \leq i \leq n$, se tiene que $b_n^{-1} S_n \rightarrow_p 0$ cuando $n \rightarrow \infty$, si

$$(i) \sum_{i=1}^n P(|X_i| > b_n) \rightarrow 0,$$

$$(ii) b_n^{-1} \sum_{i=1}^n E(X_{ni} | \mathcal{F}_{i-1}) \rightarrow_p 0, \text{ y}$$

$$(iii) b_n^{-2} \sum_{i=1}^n \{E(X_{ni}^2) - E[E(X_{ni} | \mathcal{F}_{i-1})]^2\} \rightarrow 0.$$

Demostración. Sea

$$\begin{aligned} S_{nn} &= \sum_{i=1}^n X_{ni} \\ &= \sum_{i=1}^n X_i I(|X_i| \leq b_n), \end{aligned}$$

nótese que la probabilidad que los eventos $\frac{S_{nn}}{b_n}$ y $\frac{S_n}{b_n}$ sean distintos es nula. En efecto se tiene que

$$\begin{aligned} P\left(\frac{S_{nn}}{b_n} \neq \frac{S_n}{b_n}\right) &= P\left(\sum_{i=1}^n X_{ni} \neq \sum_{i=1}^n X_i\right) \\ &\leq \sum_{i=1}^n P(X_{ni} \neq X_i) \\ &= \sum_{i=1}^n P(|X_i| > b_n) \rightarrow 0 \quad \text{de (i)}. \end{aligned}$$

De lo anterior, probar el resultado del teorema equivale a probar que

$$\frac{1}{b_n} S_{nn} \rightarrow_p 0, n \rightarrow \infty$$

se debe probar entonces que para todo $\epsilon > 0$ se verifica que

$$P\left(\left|\frac{S_{nn}}{b_n}\right| > \epsilon\right) \rightarrow 0, \quad n \rightarrow \infty$$

para hacerlo se empleará la desigualdad de Chebyshev. De (ii) se sabe que

$$b_n^{-1} \sum_{i=1}^n E(X_{ni} | \mathcal{F}_{i-1}) \rightarrow_p 0,$$

de donde empleando la desigualdad antes citada a $\frac{1}{b_n} S_{nn}$ se tiene que

$$P \left(\left| \frac{1}{b_n} S_{nn} \right| > k\sigma \right) \leq \frac{\sigma^2}{k^2}$$

con $\sigma^2 = b_n^{-2} \sum_{i=1}^n \{E(X_{ni}^2) - E[E(X_{ni} | \mathcal{F}_{i-1})]^2\}$, tomando $k\sigma = \epsilon$, de (iii) se concluye el resultado. \square

3.3.2 Teorema del Límite Central

Al igual que en el teorema del límite central ordinal, el objetivo del TLC en martingalas es probar que la suma de variables aleatorias bajo ciertas condiciones al límite sigue una distribución normal. El TLC en martingalas puede ser visto como una generalización del ordinal, ya que impone condiciones sobre los dos primeros momentos del proceso. Antes de introducir formalmente el teorema de estudio central del presente trabajo, es necesario incluir algunos conceptos y resultados relevantes para su comprensión.

Definición 35.

- **Diferencia de martingalas o martingala-diferencia.**

Sea $\{S_n, \mathcal{F}_n\}_{n \geq 1}$ una martingala de media cero y de cuadrado integrable, se define la diferencia de martingalas como sigue.

$$\begin{aligned} X_n &= S_n - S_{n-1} \quad n \geq 2 \\ X_1 &= S_1 \end{aligned}$$

Si X_n constituye una diferencia de martingalas se tiene que

$$E(X_n | \mathcal{F}_{n-1}) = 0, \quad \forall n \geq 1$$

Los X_i pueden ser vistos como incrementos del proceso.

- **Arreglo de martingalas**

Sea $\{S_{ni}, \mathcal{F}_{ni}\}_{1 \leq i \leq k_n}$ donde $k_n \rightarrow \infty$ si $n \rightarrow \infty$, una martingala de media cero y de cuadrado integrable, la doble secuencia $\{S_{ni}, \mathcal{F}_{ni}\}$ con $1 \leq i \leq k_n$ y $n \geq 1$

se denomina arreglo de martingalas. Tomando los índices adecuadamente una martingala puede ser vista como un arreglo de martingalas ([1],p.53).

- **Condición de Lindeberg**

La condición de Lindeberg de orden $2p$ está dada por

$$\forall \epsilon > 0, \sum_i E[|X_{ni}|^{2p} I(|X_{ni}| > \epsilon)] \rightarrow 0, n \rightarrow \infty \quad (3.3)$$

Teorema 29 (TLC en Martingalas[27],p.48). Sea $\{X_{ni}\}$ un arreglo de diferencias de martingalas, con $1 \leq i \leq k_n$ y $n \geq 1$, donde k_n representan tiempos de parada tales que $\lim_{n \rightarrow \infty} P(k_n < \infty) = 1$. Si se verifican las condiciones

1. $\sum_{i=1}^{k_n} E[X_{ni}^2 I(|X_{ni}| > \epsilon) | \mathcal{F}_{n,i-1}] \rightarrow_p 0, n \rightarrow \infty \quad \forall \epsilon > 0$ y

2. $\sum_{i=1}^{k_n} E(X_{ni}^2 | \mathcal{F}_{n,i-1}) \rightarrow_p c, n \rightarrow \infty$

Entonces $S_n = \sum_{i=1}^{k_n} X_{ni} \rightarrow_D N(0, c)$.

Capítulo 4

Aplicación

En este capítulo se realizará la aplicación práctica del teorema central del límite para martingalas utilizando datos correspondientes a la educación superior del país. La información recopilada corresponde a personas tituladas en universidades y escuelas politécnicas nacionales en carreras de tercer nivel. Debido al extenso número de carreras existente se realizará el estudio agrupando las mismas según el área del conocimiento al que pertenecen. ¹

Área	Subárea
Programas Generales	Programas básicos Programas de alfabetización y de aritmética Desarrollo personal
Educación	Formación de personal docente y ciencias de la educación
Humanidades y artes	Artes Humanidades
Ciencias sociales, educación comercial y derecho	Ciencias sociales y del comportamiento Periodismo e información Educación comercial y administración Derecho
Ciencias	Ciencias de la vida Ciencias físicas Matemática y estadística Informática
Ingeniería, industria y construcción	Ingeniería y profesiones afines Industria y producción Arquitectura y construcción

¹La Clasificación Internacional Normalizada de la Educación, CINE, propuesta por la UNESCO para clasificar los programas educativos de acuerdo a sus contenidos en áreas y subáreas del conocimiento para facilitar la evaluación y comparación de los mismos a nivel nacional e internacional.

Agricultura	Agricultura, silvicultura y pesca Veterinaria
Salud y servicios sociales	Medicina Servicios sociales
Servicios	Servicios personales Servicios de transporte Protección del medio ambiente Servicios de seguridad

Cuadro 4.1: Clasificación CINE - UNESCO 1997

Se decidió trabajar con la información referente a titulados ya que muestra en forma clara el desarrollo académico del país, tomando para este efecto la variable tiempo real de estudios, que comprende, el tiempo en años entre la fecha de inicio de estudios y la fecha de investidura o de graduación, considerando además el tiempo oficial de duración de cada carrera. A continuación se realizará un análisis descriptivo de la variable propuesta seguido por la aplicación del teorema.

4.1 Análisis descriptivo

Para el análisis se consideró a 518.755 titulados en universidades y escuelas politécnicas del país en carreras de tercer nivel, que finalizaron sus estudios en un tiempo comprendido entre los 2 y 30 años y que recibieron su título en el periodo 2000 - 2013.

En base a esta información se construye una serie de tablas y gráficos que muestran la situación actual de la titulación en el país.

1. Titulados por áreas del conocimiento.

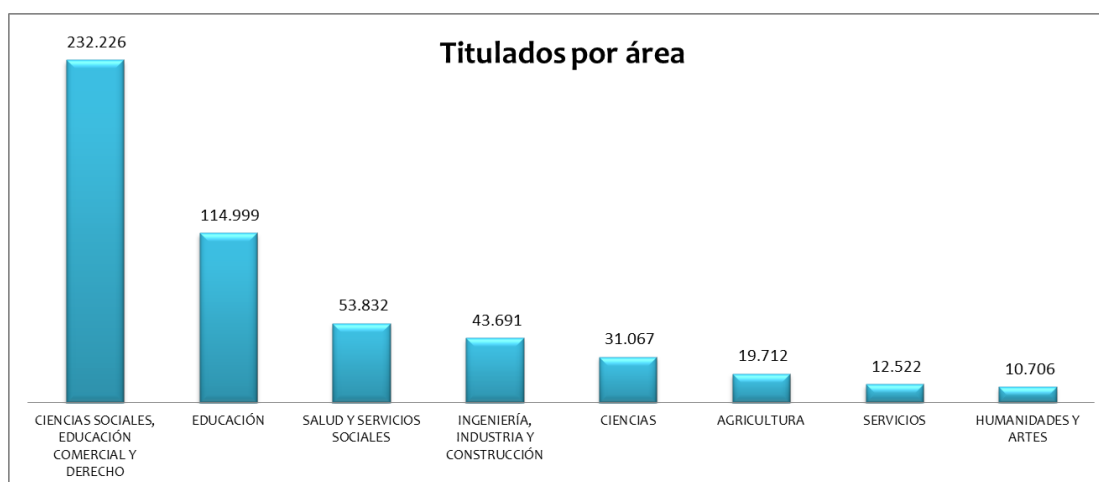


Figura 4.1: Número de titulados por área del conocimiento

2. Tiempo promedio de estudio por provincia.

Provincia	Tiempo Promedio
Azuay	7,43
Bolívar	5,59
Cañar	6,91
Carchi	5,88
Chimborazo	6,88
Cotopaxi	6,29
El Oro	8,75
Esmeraldas	7,63
Galápagos	6,53
Guayas	7,65
Imbabura	7,53
Loja	7,93
Los Ríos	7,66
Manabí	8,11
Morona Santiago	6,78
Napo	5,62
Orellana	6,04
Pastaza	4,72
Pichincha	7,63
Santa Elena	6,39
Santo Domingo	6,80
Sucumbios	6,08
Tungurahua	6,15
Zamora Chinchipe	7,49

Fuente: SENESCYT - SNIIESE 2014

Figura 4.2: Años promedio de estudio por provincia

La provincia con menor tiempo promedio en titular a sus estudiantes es Pastaza, ya que se observa titulados de las carreras Ciencias de la educación básica e inicial que culminan sus estudios en aproximadamente 2 años.

Para una mejor visualización se presenta la información de la tabla anterior en el mapa del Ecuador.

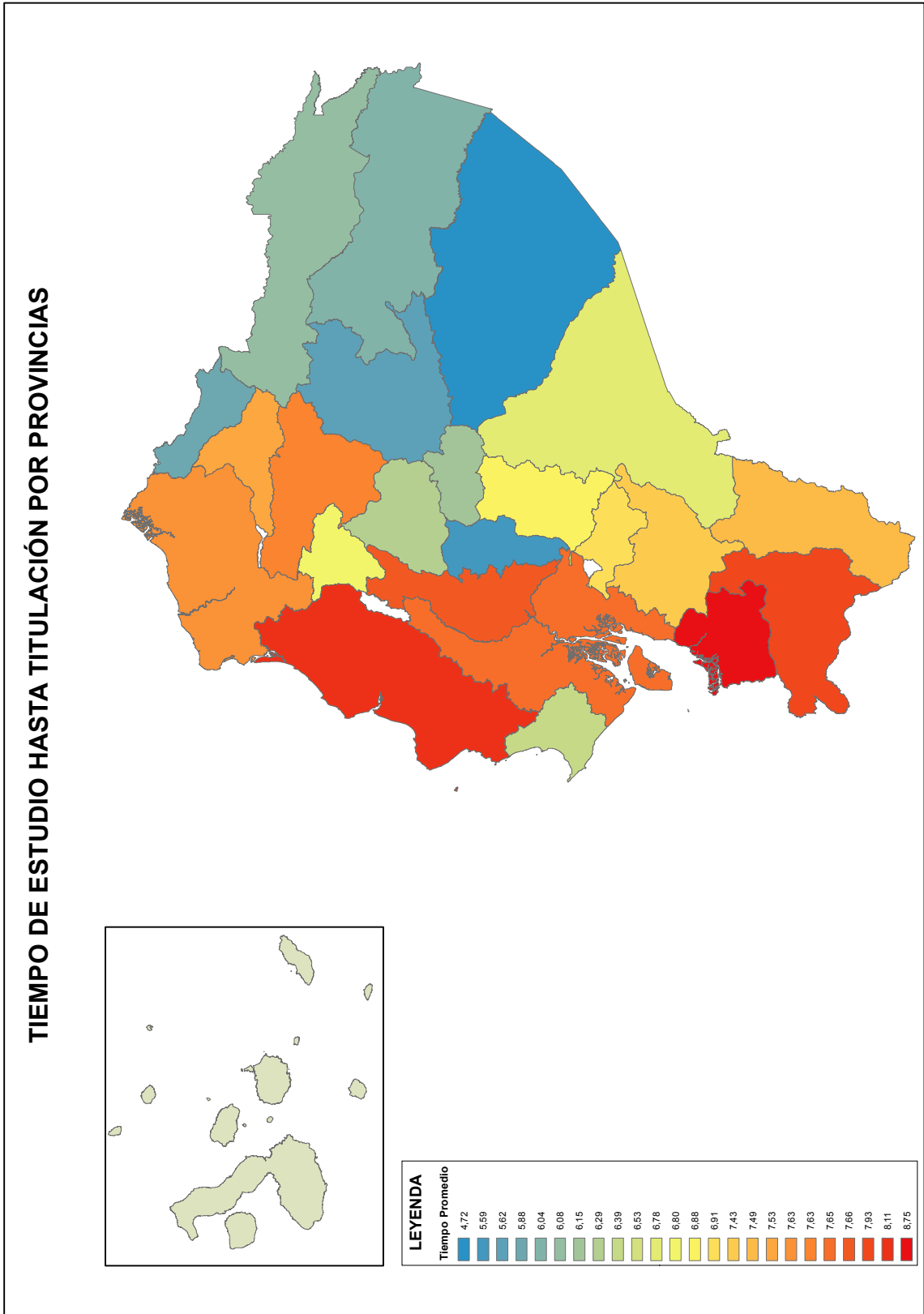


Figura 4.3: Tiempo promedio de estudio hasta la titulación por provincia

3. Tiempo promedio de estudio por área y provincia.

Los siguientes gráficos representan el tiempo promedio de estudio, en años, por área del conocimiento de las 3 principales provincias del país:

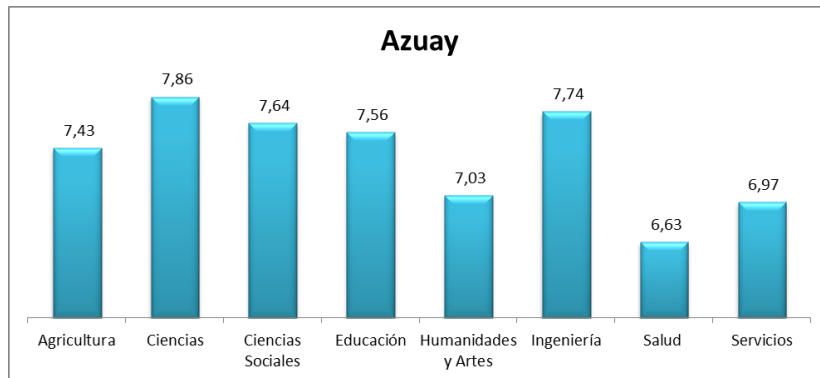


Figura 4.4: Tiempo promedio de estudio por área - Azuay

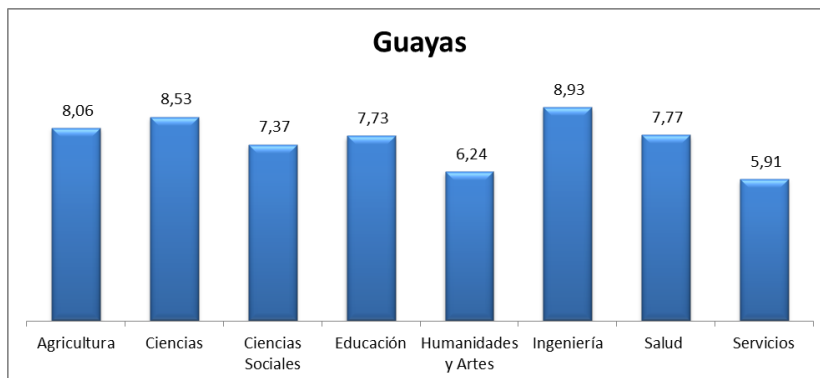


Figura 4.5: Tiempo promedio de estudio por área - Guayas

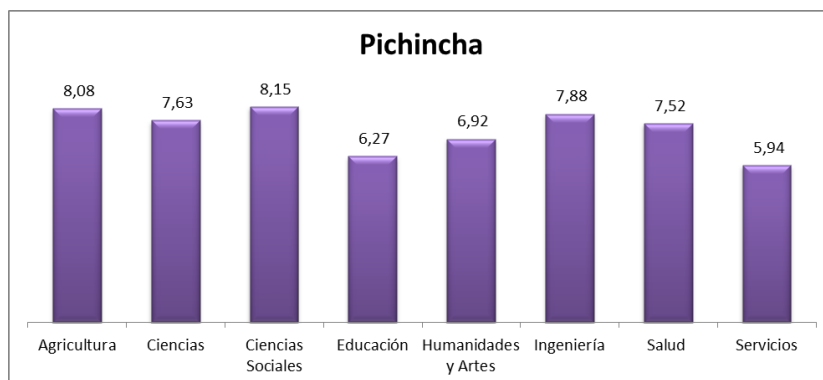


Figura 4.6: Tiempo promedio de estudio por área - Pichincha

En el siguiente mapa se amplía esta información para todas las provincias, para

cada una de ellas una mayor porción en el gráfico de pastel indica una duración promedio mayor del área de conocimiento respectiva.²

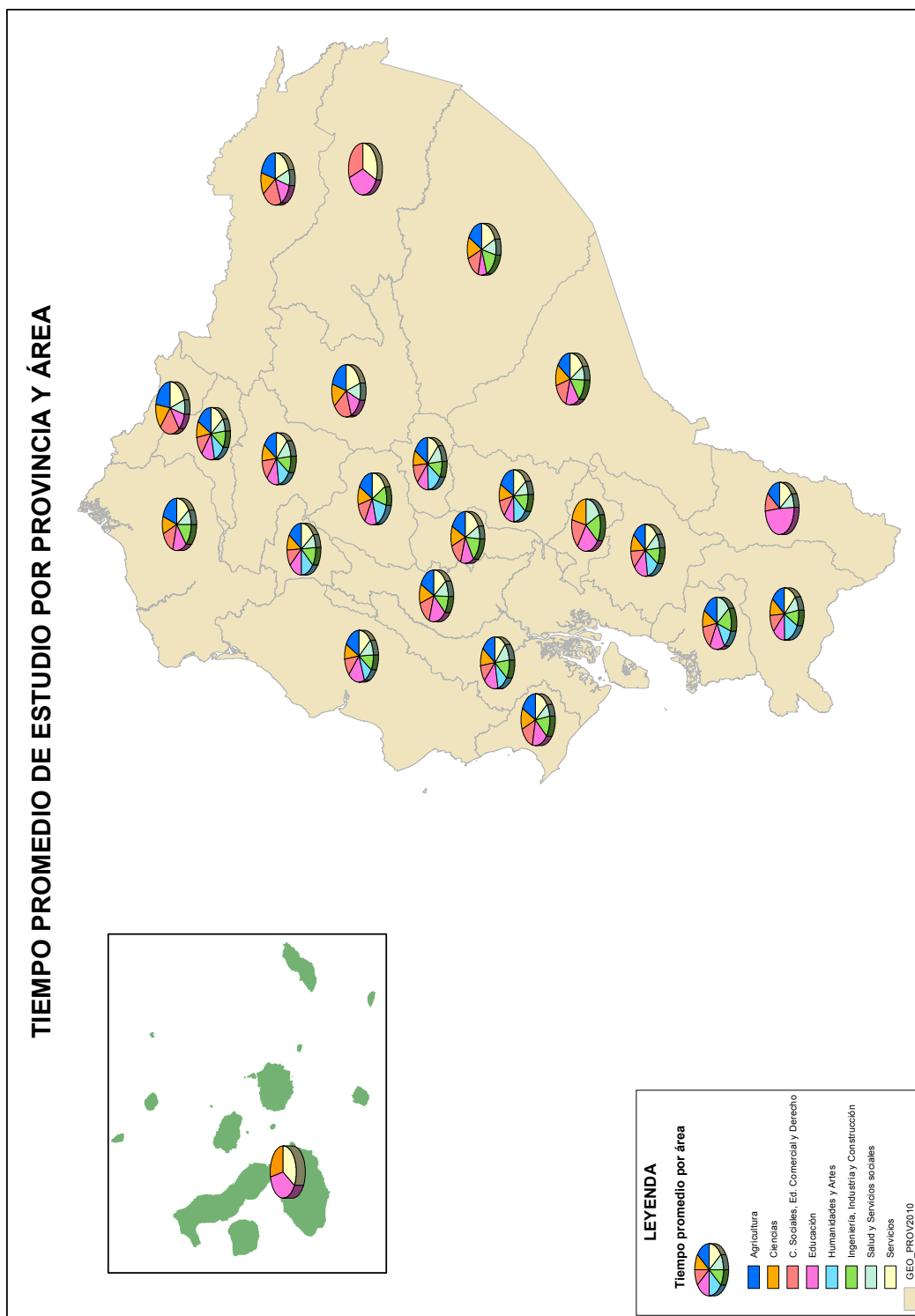


Figura 4.7: Tiempo promedio de estudio hasta la titulación por área y provincia

²La tabla con todos los datos para la realización del mapa se encuentran en el **Anexo B**.

4. Diferencia entre tiempo de titulación real y tiempo de titulación oficial por provincia.

El siguiente cuadro representa en promedio la diferencia, en años, entre el tiempo que le toma titularse a un estudiante y el tiempo que oficialmente dura su carrera, por provincia.

Provincia	Diferencia Promedio
Azuay	1,69
Bolívar	0,05
Cañar	1,10
Carchi	-0,30
Chimborazo	0,95
Cotopaxi	1,10
El Oro	2,62
Esmeraldas	2,19
Galápagos	1,92
Guayas	2,28
Imbabura	2,61
Loja	2,94
Los Ríos	1,76
Manabí	2,50
Morona Santiago	1,49
Napo	0,70
Orellana	0,78
Pastaza	-0,45
Pichincha	2,21
Santa Elena	0,29
Santo Domingo	0,94
Sucumbios	1,27
Tungurahua	-0,19
Zamora Chinchipe	2,77

Fuente: SENESCYT - SNIIESE 2014

Figura 4.8: Diferencia titulación real y titulación oficial por provincia

Existen provincias en las cuales la diferencia es negativa, es decir, los estudiantes se titularon en un tiempo menor al establecido, por ejemplo al tomar cursos de verano o remediales, la inexistencia de proyectos de titulación, entre otros.

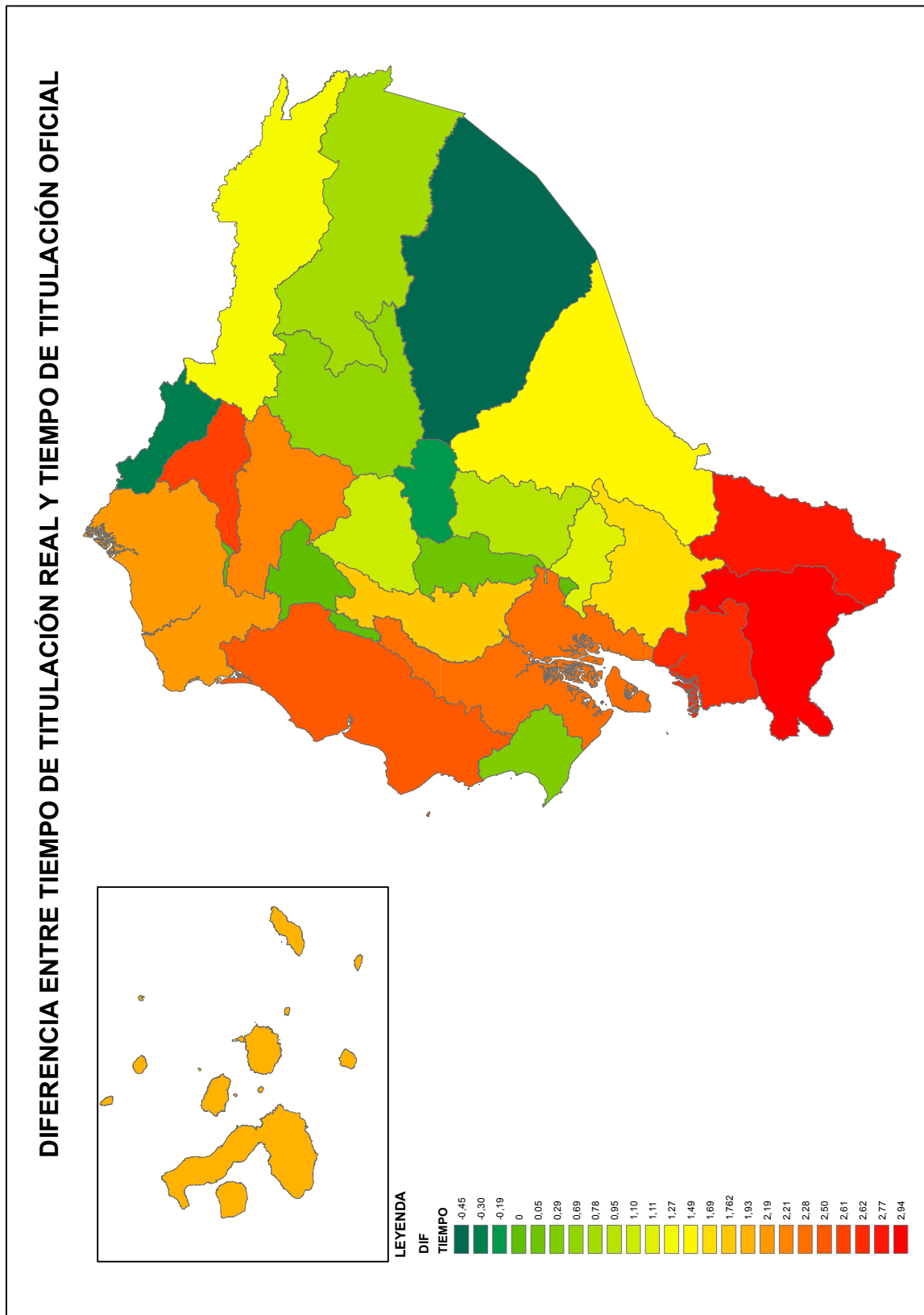


Figura 4.9: Diferencia en años entre titulación real y titulación oficial

5. Información acerca de las carreras de la Facultad de Ciencias, Escuela Politécnica Nacional.

Carrera	No. Titulados	Tiempo promedio estudios	Diferencia entre tiempo real y oficial
Ingeniería Matemática	79	9,21	3,71
Matemática	24	8,78	3,28
Física	66	8,42	2,92
Ciencias Económicas y Financieras	252	7,14	1,64

Fuente: SENESCYT - SNIESE 2014

Cuadro 4.2: Resumen carreras Facultad de Ciencias

4.2 Aplicación TLC

Para iniciar la aplicación se crean las variables $Y_i, T_i, 1 \leq i \leq n$ con:

$Y_i =$ Tiempo real de estudios de la i -ésima persona

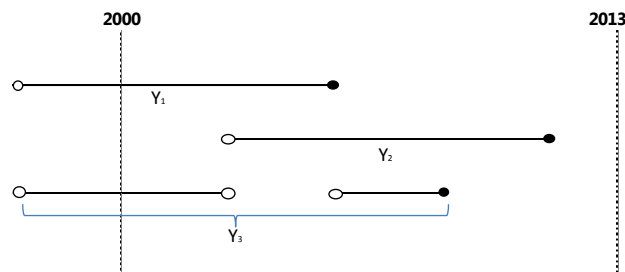


Figura 4.10: Esquema variable Y_i

Observación 13.

- La variable Y_i representa el tiempo comprendido entre la fecha de inicio de estudios y la fecha de incorporación del estudiante i .
- Se considera a los individuos cuyo año de investidura (o graduación) esté comprendido en el periodo 2000-2013.
- Los Y_i son variables aleatorias independientes y acotadas, con un mínimo de 2 años y un máximo de 30 años de estudio.

- La variable Y_i puede tomar valores mayores al establecido como cota superior, llegando inclusive a 71 años, estos valores se excluyen del estudio ya que sus probabilidades de ocurrencia tienden a cero.

T_i = Tiempo oficial de duración de la carrera del i -ésimo estudiante.

Observación 14.

- La variable T_i representa la duración establecida para la carrera del estudiante i ; al igual que la variable anterior los T_i son variables aleatorias independientes.
- Las variables Y_i y T_i se expresan en años, con $1 \leq i \leq n$, donde i representa al i -ésimo titulado y n a la totalidad de ellos (518.755 titulados).

Sean $\mathcal{F}_i = \sigma(T_1, T_2, \dots, T_i)$ la filtración natural de la variable T_i con $i \geq 1$ y

$$X_i = E(Y_i | \mathcal{F}_{i-1}) \quad \forall i \geq 1,$$

que representa el tiempo que se espera le tome al estudiante i obtener su título, teniendo en cuenta la duración de las carreras de los estudiantes $1, 2, \dots, i-1$.

De lo demostrado en el ejemplo 6 se sigue que $\{X_i, \mathcal{F}_i\}_{i \geq 1}$ es una martingala, por tanto de la definición verifica que:

$$(i) \quad E(|X_i|) < \infty, \quad \forall i \geq 1$$

$$(ii) \quad E(X_i | \mathcal{F}_{i-1}) = X_{i-1}, \quad \forall i \geq 1$$

Es decir, se espera que el tiempo de titulación del estudiante i y el $i-1$ sea el mismo, dado que se conoce la duración de las carreras de los estudiantes $1, 2, \dots, i-1$.

Además verifica propiedades como

$$E(X_i) = E(X_1), \quad \forall i \geq 1. \tag{4.1}$$

Lo anterior es una descripción teórica de cómo construir la variable X_i . A continuación se describe el proceso práctico de construcción de la martingala previo a la aplicación del teorema:

1. Se observan las frecuencias individuales de las variables Y_i y T_i , que intervienen en la construcción de la martingala.

Tiempo real de estudios		
Valores	Frecuencia Absoluta	Frecuencia Relativa
2,00	6.335	1,22
3,00	16.722	3,22
4,00	35.302	6,81
5,00	90.997	17,54
6,00	110.104	21,22
7,00	87.354	16,84
8,00	54.166	10,44
9,00	31.202	6,01
10,00	20.034	3,86
11,00	13.808	2,66
12,00	9.759	1,88
13,00	6.977	1,34
14,00	5.186	1,00
15,00	3.967	0,76
16,00	3.450	0,67
17,00	2.931	0,57
18,00	2.650	0,51
19,00	2.353	0,45
20,00	2.213	0,43
21,00	2.001	0,39
22,00	1.828	0,35
23,00	1.624	0,31
24,00	1.618	0,31
25,00	1.403	0,27
26,00	1.296	0,25
27,00	1.095	0,21
28,00	1.011	0,19
29,00	923	0,18
30,00	446	0,09
Total	518.755	100,00

Fuente: SENESCYT - SNIIESE 2014.

Tiempo oficial de estudios		
Valores	Frecuencia Absoluta	Frecuencia Relativa
2,00	345	0,07
2,50	421	0,08
3,00	414	0,08
3,50	672	0,13
3,60	47	0,01
4,00	34.987	6,74
4,10	649	0,13
4,50	56.713	10,93
4,80	78	0,02
5,00	143.121	27,59
5,30	1.579	0,30
5,40	3.507	0,68
5,50	52.856	10,19
5,60	846	0,16
6,00	130.298	25,12
6,50	36.660	7,07
7,00	48.335	9,32
7,50	6.038	1,16
8,00	1.189	0,23
Total	518.755	100,00

Fuente: SENESCYT - SNIIESE 2014.

Figura 4.11: Frecuencias de las variables tiempo real y tiempo oficial de estudio

2. Construcción de una tabla de contingencia entre las variables tiempo real y tiempo oficial de estudios hasta la titulación.

Tiempo estudios	Duración oficial																	Total		
	2,00	2,50	3,00	3,50	3,60	4,00	4,10	4,50	4,80	5,00	5,30	5,40	5,50	5,60	6,00	6,50	7,00		7,50	8,00
2,00	17	7	1	-	-	1,299	-	1,027	-	2,507	21	-	588	220	540	42	66	-	-	6,335
3,00	109	299	54	106	2	1,655	-	1,249	-	7,090	37	-	1,221	216	3,706	534	443	-	1	16,722
4,00	107	51	184	179	5	5,806	73	4,427	10	13,320	149	45	2,031	254	5,958	1,776	893	-	34	35,302
5,00	47	18	75	145	12	8,391	125	16,824	29	29,715	280	369	7,204	116	14,564	7,919	5,148	3	13	90,997
6,00	17	14	34	142	10	5,559	204	11,060	22	30,162	332	1,019	11,006	29	29,776	9,181	11,334	6	197	110,104
7,00	32	31	32	56	6	3,101	85	8,063	6	19,567	255	720	10,025	5	25,779	6,699	10,909	1,523	460	87,354
8,00	6	1	19	21	2	2,061	41	4,373	6	11,375	174	453	7,063	4	16,436	3,876	6,220	1,790	245	54,166
9,00	2	-	7	9	4	1,367	20	2,145	5	6,981	97	254	4,187	2	8,932	2,217	3,827	1,043	103	31,202
10,00	4	-	2	9	2	1,087	11	1,602	-	4,746	79	171	2,525	-	5,659	1,307	2,223	574	33	20,034
11,00	2	-	4	-	-	821	12	1,112	-	3,439	51	116	1,497	-	3,940	800	1,661	335	18	13,808
12,00	-	-	-	3	4	688	8	816	-	2,368	37	52	891	-	2,958	528	1,209	185	12	9,759
13,00	-	-	1	-	-	529	5	568	-	1,857	15	50	595	-	1,960	434	808	136	19	6,977
14,00	-	-	1	-	-	387	5	485	-	1,369	15	42	471	-	1,507	222	584	94	4	5,186
15,00	-	-	-	-	-	355	7	405	-	1,010	6	23	309	-	1,142	186	451	68	5	3,967
16,00	1	-	-	1	-	285	14	352	-	882	6	20	314	-	986	181	348	55	5	3,450
17,00	-	-	-	1	-	197	2	285	-	814	4	24	296	-	846	139	272	42	9	2,931
18,00	-	-	-	-	-	229	3	234	-	736	4	19	250	-	749	108	283	31	4	2,650
19,00	-	-	-	-	-	164	4	224	-	644	3	17	257	-	678	97	239	24	2	2,353
20,00	1	-	-	-	-	156	4	221	-	664	-	17	226	-	600	66	230	19	9	2,213
21,00	-	-	-	-	-	117	2	200	-	616	2	21	235	-	538	68	189	13	-	2,001
22,00	-	-	-	-	-	122	4	167	-	539	-	11	247	-	492	54	177	14	1	1,828
23,00	-	-	-	-	-	128	4	178	-	511	2	7	169	-	414	51	146	13	1	1,624
24,00	-	-	-	-	-	155	3	159	-	478	4	8	212	-	412	34	135	14	4	1,618
25,00	-	-	-	-	-	137	3	121	-	401	3	11	177	-	357	49	129	12	3	1,403
26,00	-	-	-	-	-	81	3	100	-	396	1	11	175	-	343	45	123	15	3	1,296
27,00	-	-	-	-	-	46	3	92	-	307	1	8	213	-	313	13	93	5	1	1,095
28,00	-	-	-	-	-	30	2	94	-	292	-	6	168	-	313	16	85	5	-	1,011
29,00	-	-	-	-	-	25	-	89	-	236	-	10	183	-	276	14	78	10	2	923
30,00	-	-	-	-	-	9	2	41	-	99	1	3	121	-	124	4	32	9	1	446
Total	345	421	414	672	47	34,987	649	56,713	78	143,121	1,579	3,507	52,856	846	130,298	36,660	48,335	6,038	1,189	518,755

Figura 4.12: Tabla de contingencia Tiempo real vs Tiempo oficial

3. Se calcula la probabilidad de la intersección entre las dos variables, es decir

$$P(Y_i = y, T_i = t)$$

tomando para este efecto la definición de probabilidad empírica, es decir, las frecuencias observadas sobre el total de observaciones.

4. Una vez obtenidas las probabilidades de la intersección, se calcula para cada caso las probabilidades condicionales, utilizando la definición:

$$P(Y_i = y|T_i = t) = \frac{P(Y_i = y, T_i = t)}{P(T_i = t)}.$$

5. Con las probabilidades condicionales se calcula la esperanza condicional del tiempo real de estudios, dado la duración oficial de la carrera,

$$E(Y_i|T_i = t) = \sum_y yP(Y_i = y|T_i = t).$$

Así, para cada valor de T_i se tiene una salida de la variable esperanza condicional, en efecto

Tiempo oficial de estudios (Tn)	2,00	2,50	3,00	3,50	3,60	4,00	4,10	4,50	4,80	5,00	5,30	5,40	5,50	5,60	6,00	6,50	7,00	7,50	8,00
Esperanza Condicional (Xn)	4,40	3,60	4,86	5,08	6,58	6,89	7,63	7,00	5,81	7,22	7,05	8,09	8,00	3,48	7,95	7,09	8,14	9,26	7,98

Figura 4.13: Valores obtenidos para la variable X_n

El algoritmo para el cálculo de la esperanza condicional se realizó en *Matlab*:

```
function [E]=EspCond(P,X)
%P: Matriz de distribuciones conjuntas
%X: Valores que toma la variable tiempo real de estudios
%M: Matriz probabilidades condicionales
%E: Vector de esperanzas condicionales
M = conditional(P);
[f,c] = size(M);
for i = 1:c
    E(i) = X * M(1:f,i);
end
```

Cuadro 4.3: Algoritmo construcción Esperanza Condicional

Una vez obtenida la martingala $\{X_i, \mathcal{F}_i\}_{i \geq 1}$ se verifican las hipótesis para la aplicación

del teorema. Se construye para este efecto la diferencia de martingalas

$$r_i = X_i - X_{i-1} \quad \forall i \geq 1$$

que representa la diferencia entre tiempos de estudios esperados de los estudiantes i e $i - 1$. Obsérvese que $E(r_i) = 0, \forall i \geq 1$, la igualdad se sigue de (4.1).

La primera condición para la aplicación del TLC en martingalas es conocida como condición de Lindeberg, de ([38], p.332) se sabe que si las variables aleatorias son independientes y acotadas esta condición se verifica, basta entonces probar que X_i es acotada.

En efecto ya que por hipótesis las variables Y_i son acotadas y positivas existe una constante M positiva y finita tal que

$$Y_i = |Y_i| \leq M, \quad \forall i \geq 1$$

de donde

$$E(Y_i | \mathcal{F}_{i-1}) \leq E(M | \mathcal{F}_{i-1}) = M \text{ c.s} \quad (4.2)$$

Así,

$$\begin{aligned} |X_i| &= |E(Y_i | \mathcal{F}_{i-1})| \\ &\leq E(|Y_i| | \mathcal{F}_{i-1}) \text{ aplicando la desigualdad de Jensen} \\ &\leq E(M | \mathcal{F}_{i-1}) = M \text{ c.s, se sigue de (4.2)} \end{aligned}$$

La definición de convergencia casi segura establece que a un mayor valor de k la probabilidad de que la variable $\sum_{i=1}^k E(r_i^2 | \mathcal{F}_{i-1})$ tome el valor de c es igual a 1.

Así, tomando

$$c = \sum_{i=1}^n E(r_i^2 | \mathcal{F}_{i-1})$$

con n la totalidad de casos de estudio, se sigue que

$$\sum_{i=1}^k E(r_i^2 | \mathcal{F}_{i-1}) \rightarrow_{c.s} c, \quad k \rightarrow \infty$$

de donde,

$$\sum_{i=1}^k E(r_i^2 | \mathcal{F}_{i-1}) \rightarrow_p c, \quad k \rightarrow \infty$$

lo que verifica la segunda condición del TLC para martingalas.

A continuación, a partir de los datos se determinará el valor de c . Para hacerlo es necesario calcular

$$E(r_i^2 | \mathcal{F}_{i-1}), \quad \forall i \geq 1,$$

empleado para ello el procedimiento descrito anteriormente en la construcción de la variable

$$E(Y_i | \mathcal{F}_{i-1}), \forall i \geq 1$$

1. Se crea la variable r_i^2 y se observa la frecuencia de la misma

Diferencia de martingalas		
Valores	Frecuencia Absoluta	Frecuencia Relativa
-	518.735	99,9961
0,0004	1	0,0002
0,0009	1	0,0002
0,0016	2	0,0004
0,0025	2	0,0004
0,0081	1	0,0002
0,0081	1	0,0002
0,0081	1	0,0002
0,0081	1	0,0002
0,0289	1	0,0002
0,0484	1	0,0002
0,0961	1	0,0002
0,1024	1	0,0002
0,1369	1	0,0002
0,2116	1	0,0002
0,3600	1	0,0002
0,5329	1	0,0002
0,6400	1	0,0002
1,2544	1	0,0002
Total	518.755	100,0000

Figura 4.14: Frecuencias de la variable r_i^2

2. Se calcula la probabilidad de la intersección entre las variables r_i^2 y T_i (ver anexo de resultados).
3. Para cada caso se calcula las probabilidades condicionales

$$P(r_i^2 = r | T_i = t).$$

4. Con el algoritmo implementado en *Matlab* se calcula la variable $E(r_i^2 | T_i = t)$

Tempo oficial de estudios (Ti)	2,00	2,50	3,00	3,50	3,60	4,00	4,10	4,50	4,80	5,00	5,30	5,40	5,50	5,60	6,00	6,50	7,00	7,50	8,00	10,00	11,00
Esperanza Condicional	0,000605	0,001498	0,000115	0,000782	0,002015	0,0000002	0,000156	0,000000	0,000365	0,000001	0,000001	0,000001	0,000000	0,000009	0,00000001	0,00000022	0,000026	-	0,00000003	0,000009	0,002151

Figura 4.15: Valores obtenidos para la esperanza condicional

5. Finalmente tras colocar los valores obtenidos para cada titulado, se calcula

$$\sum_{i=1}^n E[r_i^2 | \mathcal{F}_{i-1}] = 3,404$$

de donde $c = 3,404$.

El resultado obtenido al aplicar el teorema es que $\sum_{i=1}^n r_i$ converge hacia una distribución normal centrada con varianza $c = 3,4041$.

4.2.1 Interpretación de resultados.

Antes interpretar los resultados obtenidos, resulta conveniente observar la relación que existe entre las variables que intervienen en la aplicación:

- Y_i , tiempo que le toma al estudiante i titularse.

$X_i = E(Y_i | \mathcal{F}_{i-1})$, tiempo esperado de titulación del i -ésimo estudiante dado que se conoce la duración oficial de las carreras de los estudiantes $1, 2, \dots, i-1$.

$r_i = X_i - X_{i-1}$, diferencia entre el tiempo de titulación esperado del estudiante i y el $i-1$.

- Nótese que las variables Y_i y X_i son positivas y acotadas, mientras que los r_i toman valores negativos o nulos, en efecto

$$X_i \leq X_{i-1}, \quad \forall i \geq 1$$

con $X_0 = X_1$.

- El promedio de las variables Y_i y X_i es el mismo, en efecto

$$\begin{aligned} E(X_i) &= E(E(Y_i | \mathcal{F}_{i-1})) \\ &= E(Y_i) \end{aligned}$$

al calcular el promedio de las variables se obtiene

$$\begin{aligned} \bar{Y} &= 7,52 \\ \bar{X} &= 7,52. \end{aligned}$$

Por otro lado $E(r_i) = 0, \forall i \geq 1$, ya que

$$\begin{aligned} E(r_i) &= E(X_i - X_{i-1}) \\ &= E(X_i) - E(X_{i-1}) = 0 \quad \text{de (4.1)} \end{aligned}$$

Así mismo, al calcular el promedio de esta variable se obtiene $\bar{r} = 0$.

La variable sobre la cual el teorema arroja resultados es la suma de los r_i , esta variable representa la interacción entre tiempo oficial y tiempo real hasta la titulación, considerando simultáneamente la interacción entre estudiantes. De la aplicación misma del teorema se sabe que la variable es centrada, independientemente de si se realiza el análisis con la totalidad de los datos o segmentándolos en grupos de interés, por ejemplo: áreas del conocimiento, universidades, provincias, etc.

Sin embargo, un análisis más específico puede realizarse utilizando el segundo parámetro obtenido del teorema, la varianza c , que puede ser vista como la brecha entre lo que pasa y lo que debería pasar en cuanto a tiempo de titulación de los estudiantes se refiere, en otras palabras c permite medir la variación entre el tiempo establecido y el tiempo que realmente les toma finalizar sus estudios, tanto de la totalidad del grupo como de segmentos específicos del mismo.

El valor de c de un determinado grupo brinda mayor información que el tiempo promedio de estudios del mismo, además que considera la variabilidad de tiempos de las personas que integren el grupo estudiado.

Ya que todas las carreras del sistema educativo están clasificadas según áreas del conocimiento, resulta interesante observar en cual de ellas dicha brecha es mayor y menor.

Al segmentar la base en áreas y aplicar en cada sub base el proceso descrito anteriormente se obtiene el valor de la varianza c para cada caso:

- Ciencias.

Tiempo oficial de estudios (Tn)	2,5	4,0	4,5	5,0	5,3	5,5	6,0	6,5	7,0
Esperanza Condicional	0,9784497	0,0001972	0,0000145	0,0000595	0,0000419	0,0000025	0,0000002	0,0000049	-

Figura 4.16: Valores obtenidos para $E(r_i^2|T_i = t)$ en ciencias.

Varianza c	11,3476
--------------	---------

Cuadro 4.4: Valor de c ciencias

- Educación.

Tiempo oficial de estudios (Tn)	2,00	2,50	3,00	3,50	3,60	4,00	4,10	4,50	5,00	5,50	6,00	6,50	7,00	8,00
Esperanza Condicional	0,0010184	0,0018147	0,0021780	0,0072999	0,0021147	0,0000013	0,0001633	0,0000006	0,0000034	0,0000024	0,0000005	0,0000020	-	0,0000032

Figura 4.17: Valores obtenidos para $E(r_i^2|T_i = t)$ en educación.

Varianza c	3,6994
--------------	--------

Cuadro 4.5: Valor de c educación

- Humanidades y artes.

Tiempo oficial de estudios (Tn)	4,00	4,50	5,00	5,50	6,00	6,50	7,00	8,00
Esperanza Condicional	0,0000096	0,0000029	0,0002158	0,0000402	0,0000004	0,0000141	-	0,0000052

Figura 4.18: Valores obtenidos para $E(r_i^2|T_i = t)$ en humanidades y artes.

Varianza c	0,5883
--------------	--------

Cuadro 4.6: Valor de c humanidades y artes

- Ciencias sociales, educación comercial y derecho.

Tiempo oficial de estudios (Tn)	2,00	2,50	3,00	3,50	4,00	4,50	5,00	5,30	5,40	5,50	5,60	6,00	6,50	7,00	11,00
Esperanza Condicional	0,00782356	0,01420021	0,00012361	0,03371733	0,00000078	0,00000007	0,00001013	0,00000236	0,00000071	0,00000041	0,00001700	0,00000004	0,00000128	-	0,00501172

Figura 4.19: Valores obtenidos para $E(r_i^2|T_i = t)$ en ciencias sociales.

Varianza c	3,1273
--------------	--------

Cuadro 4.7: Valor de c ciencias sociales

- Ingeniería, industria y construcción.

Tiempo oficial de estudios (Tn)	2,00	4,00	4,50	5,00	5,50	6,00	6,50	7,00
Esperanza Condicional	1,3778000	0,0001432	0,0000186	0,0000555	0,0000014	0,0000002	0,0000033	-

Figura 4.20: Valores obtenidos para $E(r_i^2|T_i = t)$ en ingeniería.

Varianza c	6,0286
--------------	--------

Cuadro 4.8: Valor de c ingeniería

- Agricultura.

Tiempo oficial de estudios (Tn)	4,00	4,50	5,00	5,30	5,50	6,00	6,50	7,00	10,00
Esperanza Condicional	0,0017383	0,0000095	0,0000412	0,0000021	0,0000320	0,0000004	0,0000078	-	0,0007080

Figura 4.21: Valores obtenidos para $E(r_i^2|T_i = t)$ en agricultura.

Varianza c	0,3234
--------------	--------

Cuadro 4.9: Valor de c agricultura

- Salud y servicios sociales.

Tiempo oficial de estudios (Tn)	2,00	3,50	4,00	4,50	5,00	5,50	6,00	6,50	7,00	7,50	8,00
Esperanza Condicional	0,0046765	0,0136111	0,0000016	0,0000036	0,0000399	0,0000078	0,0000001	0,0000102	0,0003871	-	0,0000004

Figura 4.22: Valores obtenidos para $E(r_i^2|T_i = t)$ en salud y servicios sociales.

Varianza c	5,5909
--------------	--------

Cuadro 4.10: Valor de c salud

- Servicios.

Tiempo oficial de estudios (Tn)	3,50	4,00	4,50	4,80	5,00	5,30	5,50	6,00	6,50	7,00
Esperanza Condicional	0,006066	0,000023	0,000001	0,014980	0,000162	0,000017	0,000016	0,000002	0,000011	-

Figura 4.23: Valores obtenidos para $E(r_i^2|T_i = t)$ en servicios.

Varianza c	2,2914
--------------	--------

Cuadro 4.11: Valor de c servicios

Se resume los resultados obtenidos en el siguiente cuadro:

Área	Varianza c	Desviación estándar
Agricultura	0,3234	0,56869
Humanidades y artes	0,5883	0,76703
Servicios	2,2914	1,51374
Ciencias sociales, educación comercial y derecho	3,1274	1,76844
Educación	3,6995	1,92340
Salud y servicios sociales	5,5909	2,36451
Ingeniería, industria y construcción	6,0286	2,4553
Ciencias	11,3476	3,3686

Cuadro 4.12: Varianza y desviación estándar por área

Al considerar las interacciones existentes entre tiempos real y oficial de los estudiantes en las distintas áreas, el valor obtenido de c brinda más información al compararlo con el promedio del tiempo de estudios hasta la titulación por áreas:

Área	Tiempo promedio	Desviación estándar	Intervalo de confianza
Servicios	6,4028	2,1457	6,3652 - 6,4404
Humanidades y artes	6,6989	3,1033	6,6402 - 6,7577
Educación	6,9981	4,8974	6,9698 - 7,0264
Salud y servicios sociales	7,1999	3,2769	7,1722 - 7,2276
Ciencias sociales, educación comercial y derecho	7,7660	4,0100	7,7496 - 7,7823
Ciencias	7,8109	3,2847	7,7743 - 7,8474
Ingeniería, industria y construcción	8,0350	3,4409	8,0028 - 8,0673
Agricultura	8,2643	3,6096	8,2139 - 8,3146

Cuadro 4.13: Tiempo estudios por área

La diferencia más significativa entre los dos indicadores se observa en Agricultura, se explica ya que el tiempo de duración oficial promedio de las carreras en esta área es

mayor que en las demás³, esta información no está siendo considerada en el segundo indicador.

Se puede realizar el mismo análisis por universidad, para un primer ejercicio se considerarán las siguientes instituciones de educación superior:

1. Escuela Politécnica Nacional (EPN).
2. Universidad San Francisco de Quito (USFQ).
3. Universidad Central del Ecuador (UCE).
4. Universidad de Cuenca (UC).
5. Universidad Tecnológica Equinoccial (UTE).
6. Universidad de las Américas (UDLA).
7. Escuela Superior Politécnica del Litoral (ESPOL).
8. Universidad de Guayaquil (UG).
9. Escuela Superior Politécnica de Chimborazo (ESPOCH).

Obteniendo los siguientes resultados:

Universidad	Varianza c	Desviación estándar
Universidad de las Américas	0,5374	0,73306
Universidad San Francisco de Quito	0,5780	0,76026
Escuela Superior Politécnica del Litoral	0,5916	0,76916
Universidad Tecnológica Equinoccial	0,5942	0,77083
Universidad de Cuenca	0,6212	0,78818
Universidad de Guayaquil	4,5485	2,13272
Escuela Superior Politécnica de Chimborazo	8,4162	2,90106
Escuela Politécnica Nacional	9,8851	3,14406
Universidad Central del Ecuador	11,2042	3,34727

Cuadro 4.14: Varianza y desviación estándar por universidad

Al ordenar en forma descendente a las Instituciones de Educación Superior (IES) consideradas según los valores de c , se obtiene un ranquin de las mismas. Al ranquear a las IES pero esta vez según el tiempo promedio de estudio por universidad⁴:

³En el anexo se presentan el promedio de los tiempos de duración oficial de la carrera por cada área del conocimiento.

⁴Intervalo de confianza al 95 % de confianza

Universidad	Tiempo promedio	Desviación estándar	Intervalo de confianza
Universidad San Francisco de Quito	5,2481	1,7120	5,2078 - 5,2884
Universidad de las Américas	5,7924	1,5915	5,7442 - 5,8407
Universidad Tecnológica Equinoccial	6,2375	3,0234	6,1917 - 6,2833
Universidad de Cuenca	7,4912	3,2850	7,4386 - 7,5439
Escuela Superior Politécnica de Chimborazo	7,5655	2,6687	7,5184 - 7,6125
Escuela Superior Politécnica del Litoral	7,6338	3,5016	7,5685 - 7,6991
Universidad de Guayaquil	7,9519	4,6376	7,9196 - 7,9842
Escuela Politécnica Nacional	8,0981	2,4918	8,0433 - 8,1528
Universidad Central del Ecuador	9,1696	4,8807	9,1265 - 9,2128

Cuadro 4.15: Tiempo de estudio por universidad

La diferencia más importante entre las posiciones de ambos listados se observa en la Escuela Superior Politécnica del Litoral (ESPOL), que pese a estar ubicada dentro de los primeros lugares en la primera tabla se ubica dentro de los cuatro últimos puestos en la segunda. Al considerar el tiempo de duración oficial de estudios⁵, el valor promedio obtenido por la ESPOL no justifica el cambio observado (como ocurría en el análisis realizado por áreas).

El cambio se explica al analizar la varianza la variable X_i , es decir, del tiempo esperado de estudios dado la duración oficial de la carrera

Universidad	Varianza de X_i	Desviación estándar
Escuela Superior Politécnica del Litoral	0,0106	0,1029
Universidad San Francisco de Quito	0,1028	0,3207
Universidad Tecnológica Equinoccial	0,1292	0,3595
Universidad de las Américas	0,1393	0,3732
Escuela Superior Politécnica de Chimborazo	0,1462	0,3823
Universidad de Cuenca	0,1534	0,3917
Universidad Central del Ecuador	0,2738	0,5233
Universidad de Guayaquil	0,4125	0,6422
Escuela Politécnica Nacional	0,4584	0,6770

Cuadro 4.16: Varianza de X_i por universidad

Se evidencia que si bien los estudiantes de la ESPOL no se titulan dentro de lo establecido, lo hacen en un tiempo muy similar considerando la duración de cada una de sus carreras. Ya que c considera las diferentes interacciones, entre ellas la existente

⁵En los anexos se incluye la tabla con los valores correspondientes al promedio de duración oficial de las carreras de las IES estudiadas.

entre estudiantes, se observas que es esta precisamente la que hace que la Escuela Superior Politécnica del Litoral esté mejor posicionada en la primera tabla.

Capítulo 5

Conclusiones y Recomendaciones

5.1 Conclusiones

1. La importancia de las martingalas radica en que generalizan la suma de variables aleatorias, permitiendo la aplicación de teoremas límites tales como el Teorema del Límite Central, la Ley de los Grandes Números, entre otros.
2. La utilización de esperanzas condicionales permite considerar información extra, estudiar las variables de manera aislada conduce a resultados ciertamente alejados de la realidad ya que en la cotidianidad del proceso dichas variables interactúan entre sí.
3. Un correcto manejo de los datos permite pasar de conceptos con una fuerte base teórica a algo concreto. La aplicación del Teorema del Límite Central en martingalas a la educación superior del Ecuador arroja luces sobre la realidad de la titulación en el país.
4. La varianza c obtenida de la aplicación del Teorema del Límite Central en martingalas a la variable que representa la interacción entre tiempo oficial y tiempo real, considerando simultáneamente la interacción entre estudiantes, puede ser vista como la brecha entre lo que es y lo que debería ser. En otras palabras un menor valor de c representa que entre estudiantes de un grupo determinado no exista mayor variación entre el tiempo establecido para su titulación y lo que realmente les toma.
5. El valor de la varianza puede ser utilizado para ranquear grupos de estudiantes según áreas de conocimiento, universidades, provincias o como se requiera según el análisis que se realice.

6. El ranquin de los grupos hecho con la utilización del valor c considera la interacción entre tiempo real y oficial por lo que brinda mayor información frente a una categorización realizada considerando el promedio de tiempos de estudios hasta la titulación de los integrantes de los diferentes grupos.
7. El valor c considera además la varianza de la variable X_i , es decir, la posición en la que se encuentre ubicado el grupo dentro del ranquin dependerá también de que tan grande o pequeña sea la variación entre sus integrantes.
8. El valor c puede ser utilizado como un indicador de calidad en la gestión de las universidades en cuanto respecta a la titulación de sus estudiantes.

5.2 Recomendaciones

1. Se recomienda tomar como base el estudio realizado por áreas del conocimiento y calcular el valor de c por carreras, evitando así la variabilidad existente entre tiempos oficiales de duración dentro de una misma área. El caso mas representativo se observa en el área de salud y servicios sociales que contiene carreras como medicina y enfermería con duración oficial de 8 años y 2 años respectivamente.
2. A partir de los valores de c obtenidos por universidad se pueden realizar estudios para determinar las razones por las cuales la brecha entre lo que es y lo que debería ser es tan grande en algunos casos, posteriormente en base a esto las instituciones rectoras de la educación superior podrían tomar decisiones con la finalidad de mejorar los tiempos de titulación en el país.

Anexo A

Definiciones adicionales

Capítulo 1.

Definición 36 (Aplicaciones medibles). Sean (Ω, \mathcal{A}) y (Ω', \mathcal{A}') dos espacios medibles, una función o aplicación

$$f : \Omega \rightarrow \Omega'$$

se dice medible si

$$\forall B \in \mathcal{A}', f^{-1}(B) \in \mathcal{A}.$$

Definición 37 (σ -álgebra generada y engendrada por una aplicación f). Sean (Ω, \mathcal{A}) y (Ω', \mathcal{A}') dos espacios medibles, una función o aplicación

$$f : \Omega \rightarrow \Omega'$$

se definen:

- σ -álgebra engendrada por f .

$$f^{-1}(\mathcal{A}') = \{A \in \Omega \mid \exists A' \in \mathcal{A}', f^{-1}(A') = A\}$$

$f^{-1}(\mathcal{A}')$ es la menor σ -álgebra tal que f es medible, es decir si $\exists \mathcal{B}$ tal que f es medible respecto a \mathcal{B} , entonces

$$f^{-1}(\mathcal{A}') \subset \mathcal{B}$$

- σ -álgebra generada por f .

$$\mathcal{A}_f = \{f^{-1}(A') \in \mathcal{A}, A' \in \mathcal{A}'\} \subset \mathcal{A}$$

\mathcal{A}_f es la σ -álgebra más grande tal que f es medible.

Teorema 30 (Fórmula de inversión). Sea X una variable aleatoria con función de

distribución F . Sean a y b puntos de continuidad de F tales que $a < b$, entonces

$$F(b) - F(a) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{e^{-itb} - e^{-ita}}{-it} \varphi(t) dt.$$

Esta igualdad es conocida como fórmula de inversión.

Definición 38 (Convergencia puntual). Sean $\{X_n, n \geq 1\}$ una sucesión de funciones medibles a valores reales definidas sobre $(\Omega, \mathcal{A}, \mu)$ y X función real medible definida sobre Ω , se dice que X_n converge puntualmente a X si para todo $w \in \Omega$ se tiene que

$$\lim_{n \rightarrow \infty} X_n(w) = X(w)$$

Nota 4. La convergencia puntual implica la convergencia casi segura.

Capítulo 2.

Soporte compacto

Se llama soporte de una función $f : \mathbb{R} \rightarrow \mathbb{R}$ a la clausura de $Supp(f)$, donde

$$Supp(f) = \{x \in \mathbb{R} : f(x) \neq 0\}.$$

Ya que el conjunto sobre el cual se trabaja es el de los números reales, se dice que el soporte de f es compacto si $\overline{Supp(f)}$ es cerrado y acotado.

Método de momentos

El método de momentos es un método de estimación de funciones que dependen de un parámetro desconocido $\theta \in \Theta$ que se sabe existe. Para estimar $q(\theta)$, donde q es una función que depende del parámetro, se selecciona un estadístico $T(\mathbf{X})$ tal que

$$\underbrace{q(\theta)}_{\text{desconocido}} = \underbrace{T(x_1, \dots, x_n)}_{\text{conocido}}.$$

con $\mathbf{X} = (X_1, \dots, X_n)$ y $X_1 = x_1, \dots, X_n = x_n$. Sean $m_1(\theta), \dots, m_r(\theta)$ los primeros r momentos teóricos de la población,

$$m_j(\theta) = E_\theta(X^j) \quad \text{con } j = 1, \dots, r.$$

Se define los momentos muestrales como

$$\widehat{m}_j = \frac{1}{n} \sum_{i=1}^n X_i^j \quad \text{con } j = 1, \dots, r.$$

Si la función a estimar se puede expresar como función de los r momentos obtenidos, es decir

$$q(\theta) = g(m_1(\theta), \dots, m_r(\theta)), \quad g \text{ continua.}$$

El método de momentos sostiene que se puede estimar $q(\theta)$ mediante

$$T(\mathbf{X}) = g(\widehat{m}_1, \dots, \widehat{m}_r).$$

Espacios L^p

Sea $(\Omega, \mathcal{A}, \mu)$ espacio medido, se define el espacio de funciones de módulo integrable a

$$\mathcal{L}^1(\mu) = \{f : \Omega \rightarrow \mathbb{R}, f \text{ medible y } \int_{\Omega} |f(w)| d\mu(w) < \infty\}.$$

Análogamente para $1 \leq p < +\infty$ se define

$$\mathcal{L}^p(\mu) = \{f : \Omega \rightarrow \mathbb{R}, f \text{ medible y } \int_{\Omega} |f(w)|^p d\mu(w) < \infty\}.$$

Sea $\|f\|_p = \left(\int_{\Omega} |f(w)|^p d\mu(w) \right)^{\frac{1}{p}}$, entonces

$$\mathcal{L}^p(\mu) = \{f : \Omega \rightarrow \mathbb{R} \text{ medible, } \|f\|_p < \infty\}.$$

Para el caso $p = \infty$ se define $\|f\|_{\infty}$ como sigue

$$\|f\|_{\infty} = \begin{cases} \inf S, & S \neq \emptyset; \\ +\infty, & S = \emptyset. \end{cases}$$

con

$$S = \{\alpha \in \mathbb{R} : \mu(|f|^{-1}(\alpha, +\infty)) = 0\}.$$

Así

$$\mathcal{L}^{\infty}(\mu) = \{f : \Omega \rightarrow \mathbb{R} \text{ medible, } \|f\|_{\infty} < \infty\}$$

Observación 15. $\forall p, \|f\|_p$ no define una norma. En efecto, $\|f\|_p$ no verifica la implicación siguiente

$$\|f\|_p = 0 \Rightarrow f = 0$$

A continuación se definen los espacios L^p , que son aquellos para los cuales $\forall p, \|f\|_p$ define una norma. Sean $f, g : \Omega \rightarrow \mathbb{R}$, la relación de equivalencia \mathcal{R} tal que

$$f \mathcal{R} g \Leftrightarrow f = g \text{ c.t.p.}$$

Se nota al conjunto de todas las clases de equivalencia, es decir al conjunto cociente $\mathcal{L}^p/\mathcal{R}$ como L^p .

Definición 39 (Clase de equivalencia). *La clase de equivalencia de $f : \Omega \rightarrow \mathbb{R}$ se define como*

$$[f] = \{g \in \mathcal{L}^p : f \mathcal{R} g\}$$

En particular el espacio L^2 es un espacio de Hilbert, es decir un espacio completo dotado de un producto interno, mismo que está dado por

$$\langle f, g \rangle = \int_{\Omega} f(w)g(w)d\mu(w).$$

Definición 40 (Convergencia de variables aleatorias en espacios L^p [18].). *Una sucesión de variables aleatorias $\{X_n\}_{n \geq 1}$ converge en L^p hacia X con $1 \leq p < \infty$ si $|X_n|$, $|X|$ están en L^p y*

$$\lim_{n \rightarrow \infty} E(|X_n - X|^p) = 0$$

Capítulo 3.

Desigualdades sobre la martingala $\{S_n, \mathcal{F}_n\}_{n \in \mathbb{N}}$ Existe sobre la martingala $\{S_n, \mathcal{F}_n\}_{n \in \mathbb{N}}$ muchas propiedades útiles, las siguientes desigualdades se toman de [1].

Teorema 31. *Si $\{S_i, \mathcal{F}_i\}_{1 \leq i \leq n}$ es una submartingala, entonces para cada λ real,*

$$\lambda P(\max_{i \leq n} S_i > \lambda) \leq E \left[S_n I(\max_{i \leq n} S_i > \lambda) \right]$$

Corolario 3. *Si $\{S_i, \mathcal{F}_i\}_{1 \leq i \leq n}$ es una martingala, entonces para cada $p \geq 1$ y $\lambda > 0$,*

$$\lambda^p P(\max_{i \leq n} |S_i| > \lambda) \leq E(|S_n|^p)$$

Teorema 32. *Si $\{S_i, \mathcal{F}_i\}_{1 \leq i \leq n}$ es una martingala de media 0, entonces para cada $\lambda > 0$,*

$$\lambda P \left(\max_{i \leq n} |S_i| > 2\lambda \right) \leq \lambda P(|S_n| > \lambda) + E[(|S_n| - 2\lambda)I(|S_n| \geq 2\lambda)] \\ E[|S_n|I(|S_n| > \lambda)]$$

A continuación se presentan definiciones y teoremas adicionales al concepto de TLC en martingalas. Sean $\{S_n, \mathcal{F}_n\}_{n \geq 1}$ y $X_n = S_n - S_{n-1}$ una martingala y una diferencia de martingalas respectivamente:

- **Varianza Condicional**

Sea $\{S_n, \mathcal{F}_n\}_{n \geq 1}$ una martingala de media cero y de cuadrado integrable, se define la varianza condicional por

$$V_n^2 = \sum_{i=1}^n E(X_i^2 | \mathcal{F}_{i-1})$$

V_n^2 es una estimación de la varianza del proceso, que está dada por $E(S_n^2)$, en ciertas ocasiones representa información contenida en el pasado. Si el comportamiento de V_n^2 es muy errático, lo será también el de S_n , en este caso podría no obtenerse el T.L.C ([1], p.54). La varianza condicional para el arreglo de martingalas dado, se expresa por

$$V_{ni}^2 = \sum_{j=1}^i E(X_{nj}^2 | \mathcal{F}_{n,j-1})$$

- **Variación cuadrática**

La variación cuadrática del arreglo de martingalas se define por

$$U_{ni}^2 = \sum_{j=1}^i X_{nj}^2$$

- La varianza condicional y la variación cuadrática guardan la siguiente relación

$$\max_i |U_{ni}^2 - V_{ni}^2| \rightarrow_p 0$$

- **Estabilidad y convergencia débil en L^1**

Si $\{Y_n\}$ es una secuencia de variables aleatorias en un espacio de probabilidad (Ω, \mathcal{F}, P) converge en distribución a una variable aleatoria Y , se dice que la convergencia es estable si para todo punto de continuidad $y \in Y$ y para todos los eventos $\xi \in \mathcal{F}$ el siguiente límite existe

$$\lim_{n \rightarrow \infty} P(\{Y_n \leq y\} \cap \xi) = Q_y(\xi)$$

y $Q_y(\xi) \rightarrow P(\xi)$ cuando $y \rightarrow \infty$.

Supóngase ahora que las variables Y_n, Y son integrables, se dice que Y_n converge hacia Y débilmente en L^1 si

$$E[Y_n I(\xi)] \rightarrow E[Y I(\xi)]$$

A continuación se presentan dos resultados fundamentales para la construcción del T.L.C en martingalas, estos así como sus demostraciones, pueden ser encontrados en [1].

Teorema 33. *Se supone que $Y_n \rightarrow_D Y$, donde todas las variables Y_n pertenecen al mismo espacio (Ω, \mathcal{F}, P) . Entonces $Y_n \rightarrow Y$ (estable) si y solo si existe una variable aleatoria Y' con la misma distribución que Y , tal que para todo $t \in \mathbb{R}$ se tiene*

$$\exp(itY_n) \rightarrow \exp(itY') \text{ débilmente en } L^1$$

cuando $n \rightarrow \infty$

Sean $\{X_{ni}, 1 \leq i \leq k_n, n \geq 1\}$ un arreglo de variables aleatorias definidos sobre el mismo espacio de probabilístico (Ω, \mathcal{F}, P) y $t \in \mathbb{R}$

$$T_n(t) = \prod_{j=1}^{k_n} (1 + itX_{nj})$$

Lema 1. *Sea η^2 una variable aleatoria finita c.s. y suponiendo que*

$$\max_i |X_{ni}| \rightarrow_p 0 \quad (5.1)$$

$$\sum_i X_{ni}^2 \rightarrow_p \eta^2 \quad (5.2)$$

$$T_n(t) \rightarrow 1 \text{ débilmente en } L^1 \quad (5.3)$$

$$(5.4)$$

Entonces $S_{nk_n} \rightarrow_D Z$ (estable), donde Z tiene función característica $\exp(-\frac{1}{2}\eta^2 t^2)$

Teorema 34. *Sea $\{S_{ni}, \mathcal{F}_{ni}, 1 \leq i \leq k_n, n \geq 1\}$ martingala de media cero y de cuadrado integrable, con diferencias X_{ni} , y sea η^2 variable aleatoria finita c.s. Se supone además que*

$$\max_i |X_{ni}| \rightarrow_p 0 \quad (5.5)$$

$$\sum_i X_{ni}^2 \rightarrow_p \eta^2 \quad (5.6)$$

$$E(\max_i X_{ni}^2) \text{ es acotado en } n \quad (5.7)$$

$$\mathcal{F}_{ni} \subseteq \mathcal{F}_{n+1,i}, \quad 1 \leq i \leq k_n, n \geq 1 \quad (5.8)$$

Se tiene entonces que

$$S_{nk_n} = \sum X_{ni} \rightarrow_d Z \text{ (estable)}$$

Donde Z tiene como función característica a

$$\exp\left(-\frac{1}{2}\eta^2 t^2\right)$$

Observación 16. Sea $\{S_{ni}, \mathcal{F}_{ni}, 1 \leq i \leq k_n, n \geq 1\}$ martingala de media cero y de cuadrado integrable, sea $\{X_{ni}\}$ la secuencia de las diferencias de la martingala dada. Se definen

$$V_{ni}^2 = \sum_{j=1}^i E(X_{nj}^2 | \mathcal{F}_{n,j-1})$$

$$U_{ni}^2 = \sum_{j=1}^i X_{nj}^2, \quad 1 \leq j \leq k_n$$

Se supone además que $E(S_{nn_k}^2) = 1$ para todo $n \geq 1$ y sea

$$f_n(t) = \prod_j E[\exp(itX_{nj}) | \mathcal{F}_{n,j-1}], \quad n \geq 1$$

la función característica condicional. Si los X_{ni} son independientes y

$$f_n(t) \rightarrow \exp\left(-\frac{1}{2}t^2\right), \quad \forall t$$

entonces, evidentemente el teorema se verifica.

Anexo B

Anexo de Resultados

Área del conocimiento	N° Titulados	% Titulados
Ciencias sociales, educación comercial y derecho	232.226	44,77%
Educación	114.999	22,17%
Salud y servicios sociales	53.832	10,38%
Ingeniería, industria y construcción	43.691	8,42%
Ciencias	31.067	5,99%
Agricultura	19.712	3,80%
Servicios	12.522	2,41%
Humanidades y artes	10.706	2,06%
Total	518.755	100,00%

Fuente: SENESCYT - SNIIESE 2014

Figura 5.1: Número y porcentaje de titulados por área del conocimiento

Provincia	Agricultura	Ciencias	Ciencias Sociales	Educación	Humanidades y Artes	Ingeniería	Salud	Servicios	Tiempo Promedio
Azuay	7,43	7,86	7,64	7,56	7,03	7,74	6,63	6,97	7,43
Bolívar	7,30	7,57	6,32	5,15	-	7,67	4,89	7,17	5,59
Cañar	-	6,68	7,21	6,48	-	7,05	5,35	-	6,91
Carchi	8,01	6,72	6,34	4,24	-	-	4,42	6,60	5,88
Chimborazo	9,20	8,00	6,83	5,34	7,19	7,94	6,42	7,55	6,88
Cotopaxi	7,03	7,15	6,25	4,81	7,45	7,32	-	7,02	6,29
El Oro	8,88	8,08	9,39	7,41	7,45	11,27	7,61	-	8,75
Esmeraldas	10,84	8,47	8,25	6,99	-	9,48	6,94	7,82	7,63
Galápagos	-	5,69	-	6,69	-	-	-	6,94	6,53
Guayas	8,06	8,53	7,37	7,73	6,24	8,93	7,77	5,91	7,65
Imbabura	9,21	8,13	7,89	7,08	6,60	8,31	6,10	7,80	7,53
Loja	8,20	8,69	8,28	7,19	9,57	8,27	6,99	7,09	7,93
Los Ríos	7,97	8,03	7,50	7,87	-	6,72	6,43	6,55	7,66
Manabí	9,48	7,50	7,74	9,04	5,54	7,65	7,46	7,23	8,11
Morona Santiago	6,05	7,55	7,04	5,41	-	7,01	5,10	6,45	6,78
Napo	6,46	6,01	5,59	4,31	-	-	4,82	6,17	5,62
Orellana	-	-	5,93	6,11	-	-	-	6,21	6,04
Pastaza	5,76	6,81	5,37	2,74	-	6,06	5,06	6,26	4,72
Pichincha	8,08	7,63	8,15	6,27	6,92	7,88	7,52	5,94	7,63
Santa Elena	7,06	7,74	6,47	6,31	-	7,46	4,31	5,51	6,39
Santo Domingo	7,30	6,59	6,82	6,33	6,36	7,85	5,61	7,10	6,80
Sucumbios	7,31	6,28	6,53	5,41	-	-	4,86	6,32	6,08
Tungurahua	7,13	6,22	6,76	5,18	6,56	6,51	6,20	5,23	6,15
Zamora Chinchipe	5,95	-	6,81	22,89	-	-	5,25	6,04	7,49

Fuente: SENESCYT - SNIESE 2014

Figura 5.2: Años promedio de estudio por área del conocimiento y provincia

Provincia	Agricultura	Ciencias	Ciencias Sociales	Educación	Humanidades y Artes	Ingeniería	Salud	Servicios	Tiempo Promedio
Azuay	1,64	1,80	1,83	2,29	2,22	1,87	0,90	1,15	1,69
Bolívar	0,30	0,57	-0,30	0,03	0,00	0,67	0,89	0,49	0,05
Cañar	0,00	0,71	1,26	1,48	0,00	1,06	-0,07	0,00	1,10
Carchi	1,01	1,22	0,23	-1,76	0,00	0,00	-0,58	-0,40	-0,30
Chimborazo	3,13	1,98	0,76	-0,07	1,19	1,79	0,36	1,42	0,95
Cotopaxi	2,03	1,88	1,03	0,18	2,45	1,25	0,00	2,02	1,10
El Oro	3,22	2,08	2,71	2,27	2,48	4,31	2,00	0,00	2,62
Esmeraldas	4,84	2,29	2,28	1,98	0,00	4,06	1,62	2,62	2,19
Galápagos	0,00	0,78	0,00	2,19	0,00	0,00	0,00	2,43	1,92
Guayas	1,57	2,96	2,20	2,47	1,43	3,30	1,83	1,08	2,28
Imbabura	3,85	2,70	2,25	2,98	1,35	3,22	2,10	2,25	2,61
Loja	2,23	3,35	3,20	2,75	4,36	2,96	1,48	1,46	2,94
Los Ríos	1,33	1,99	1,03	2,54	0,00	-0,28	1,33	1,36	1,76
Manabí	3,39	1,77	1,94	3,79	0,80	2,08	1,72	2,02	2,50
Morona Santiago	1,33	3,05	1,50	0,94	0,00	1,01	0,60	1,94	1,49
Napo	1,96	1,51	0,72	-2,39	0,00	0,00	0,32	1,67	0,70
Orellana	0,00	0,00	0,30	1,11	0,00	0,00	0,00	0,21	0,78
Pastaza	1,04	1,31	0,15	-2,52	0,00	1,06	0,56	1,47	-0,45
Pichincha	2,06	2,07	2,80	1,07	1,49	2,12	2,09	0,66	2,21
Santa Elena	0,06	0,74	-0,29	0,98	0,00	0,46	0,31	0,04	0,29
Santo Domingo	0,98	1,01	0,77	1,21	0,36	1,36	-0,17	1,07	0,94
Sucumbios	2,81	1,78	1,72	0,41	0,00	0,00	0,36	1,82	1,27
Tungurahua	0,13	0,65	0,21	-1,02	0,33	0,17	0,01	-0,99	-0,19
Zamora Chinchipe	1,45	0,00	1,97	17,89	0,00	0,00	0,74	1,54	2,77

Fuente: SENESCYT - SNIIESE 2014

Figura 5.3: Años promedio de diferencia en años entre duración oficial de la carrera y tiempo real de estudios

Fm²	Duración oficial																		
	2.00	2.50	3.00	3.50	3.60	4.00	4.10	4.50	4.80	5.00	5.30	5.40	5.50	5.60	6.00	6.50	7.00	7.50	8.00
-	0.0007	0.0008	0.0008	0.0013	0.0001	0.0674	0.0012	0.1093	0.0001	0.27588939	0.0030	0.0068	0.10188625	0.0016	0.2512	0.0707	0.0932	0.0116	0.0023
0.0004	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
0.0009	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.00000193	-	-	-	-
0.0016	-	-	-	-	-	-	-	-	-	0.00000193	0.00000193	-	-	-	-	-	-	-	-
0.0025	-	-	-	-	-	-	-	0.00000193	-	-	-	0.00000193	-	-	-	-	-	-	-
0.0081	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
0.0081	-	-	-	-	0.00000193	-	-	-	-	-	-	-	-	-	-	-	-	-	-
0.0081	-	-	-	-	-	-	-	-	-	-	-	-	-	0.00000193	-	-	-	-	-
0.0289	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.00000193	-	-	-
0.0484	-	-	0.00000193	-	-	-	-	-	0.00000193	-	-	-	-	-	-	-	-	-	-
0.0961	-	-	-	-	0.00000193	-	-	-	-	-	-	-	-	-	-	-	-	-	-
0.1024	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
0.1369	-	-	-	-	-	-	0.00000193	-	-	0.00000193	-	-	-	-	-	-	-	-	-
0.2116	0.00000193	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
0.3600	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
0.5329	-	-	-	0.00000193	-	-	-	-	-	-	-	-	0.00000193	-	-	-	-	-	-
0.6400	-	0.00000193	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1.2544	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.00000193	-	-

Figura 5.4: Probabilidad de la intersección entre Tiempo real y Tiempo oficial 95

Área	Promedio Duracion oficial
Agricultura	6,12
Ingeniería, industria y construcción	5,80
Ciencias	5,65
Salud y servicios sociales	5,64
Ciencias sociales, educacion comercial y derecho	5,54
Servicios	5,36
Humanidades y artes	5,20
Educacion	5,17

Fuente: SENESCYT - SNIIESE 2014

Figura 5.5: Duración oficial de estudios por área del conocimiento

Universidad	Tiempo duracion oficial carrera
Universidad Tecnológica Equinoccial	6,4202
Escuela Superior Politécnica de Chimborazo	5,8476
Escuela Superior Politécnica del Litoral	5,7799
Universidad de Cuenca	5,5979
Escuela Politecnica Nacional	5,4540
Universidad de las Américas	5,4421
Universidad de Guayaquil	5,3239
Universidad Central del Ecuador	5,0686
Universidad San Francisco de Quito	4,4236

Fuente: SENESCYT - SNIIESE 2014

Figura 5.6: Duración oficial de estudios por universidad

Anexo C

Análisis descriptivo de la variable tiempo de estudio hasta la titulación en carreras con duración oficial de 5 años

A continuación se realizará el análisis descriptivo del tiempo de titulación real de los estudiantes, considerando únicamente a aquellos cuya carrera tiene una duración oficial de 5 años. Se escogió esta duración oficial ya que es la más usual, así la nueva población a estudiar consta de 143.121 titulados.

Al calcular el promedio del tiempo de estudio sin considerar la duración oficial de la carrera, de cierta manera se distorsiona la información, en efecto si se calcula el tiempo promedio de una carrera con 2 años de duración y se lo compara con otra de duración oficial de 8 años erróneamente se concluiría que la primera carrera realizó una mejor gestión en lo referente a titulación de sus estudiantes.

Al segmentar la población y trabajar con un grupo homogéneo en cuanto a duración oficial de la carrera se refiere, la comparación de la variable, tiempo promedio de estudio hasta la titulación, por carrera, área, provincia o como sea requerida, es válida, ya que se estaría comparando entre valores que son comparables entre sí.

1. Tiempo promedio de estudio por provincia.

Provincia	Tiempo Promedio
Azuay	6,85
Bolívar	4,98
Cañar	5,86
Carchi	4,39
Chimborazo	4,19
Cotopaxi	6,38
El Oro	7,55
Esmeraldas	7,16
Galápagos	5,58
Guayas	6,61
Imbabura	8,15
Loja	8,35
Los Ríos	7,99
Manabí	6,95
Morona Santiago	6,40
Napo	5,34
Orellana	5,93
Pastaza	4,21
Pichincha	8,26
Santa Elena	6,30
Santo Domingo	6,25
Sucumbios	5,45
Tungurahua	5,10
Zamora Chinchipe	13,84

Fuente: SENESCYT - SNIESE

Figura 5.7: Años promedio de estudio por provincia - carreras 5 años

La provincia con menor tiempo promedio en titular a sus estudiantes es Chimborazo, mientras que la que más tardó es Zamora Chinchipe, con un promedio de 13,84 años. Este último dato evidencia la distorsión producida al no considerar la duración oficial de la carrera, nótese que el valor obtenido por esta provincia cuando se considera la totalidad de los titulados (Fig 4.2) es de 7,49 años.

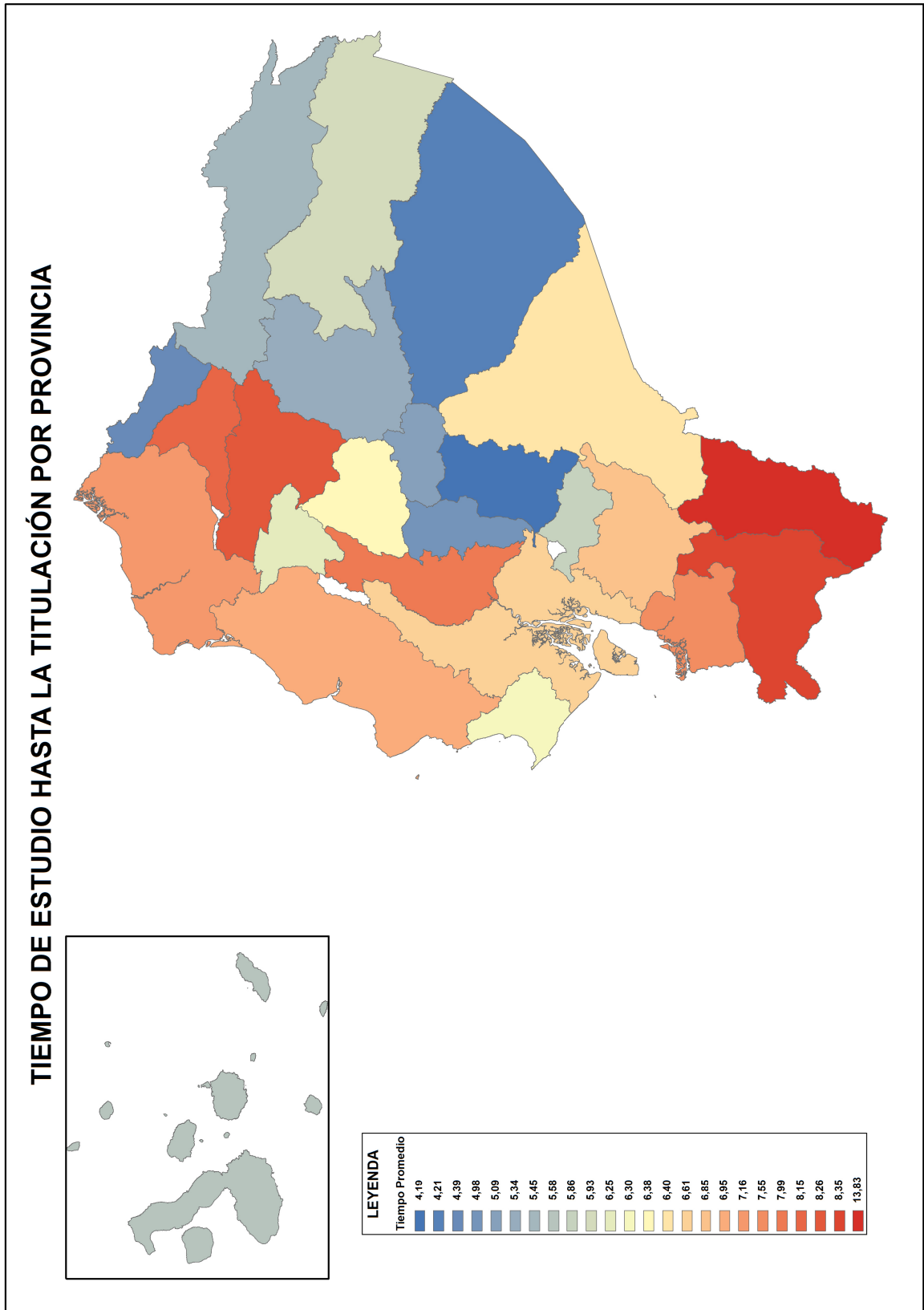


Figura 5.8: Tiempo promedio de estudio por provincia - carreras 5 años

2. Tiempo promedio de estudio por área y provincia.

Provincia	Agricultura	Ciencias	C. Sociales, Ed. Comercial y Derecho	Educación	Humanidades y Artes	Ingeniería, Industria y Construcción	Salud y Servicios sociales	Servicios
Azuay	7,26		6,78	7,36	7,75	6,49	5,93	6,73
Bolívar				4,98				
Cañar		5,89	6,00	6,48		6,55	4,57	
Carchi				3,29			4,42	
Chimborazo			4,89	3,99			2,52	5,65
Cotopaxi	7,03	6,97	6,27	2,98	7,45	6,74		7,02
El Oro			10,68	7,56	7,48		6,03	
Esmeraldas		9,42	9,33	6,93		10,53	7,11	7,93
Galápagos		5,58						
Guayas	6,73	7,65	6,33	6,33	6,65	8,16	6,39	6,42
Imbabura	9,59	9,78	8,27	4,78	6,62	8,32		8,67
Loja	6,71	8,88	8,15	9,05	11,47	8,55	6,99	6,56
Los Ríos			7,32	8,35			6,54	7,27
Manabí	8,33	7,40	6,72	6,85	6,49	7,02	7,30	7,48
Morona Santiago			6,07	9,03				
Napo			5,34					
Orellana			5,17	6,11				
Pastaza	5,50		5,05	2,58		6,06		6,14
Pichincha	8,12	7,30	8,78	6,59	6,43	7,75	8,34	5,69
Santa Elena				6,68				4,65
Santo Domingo		6,94	5,41	6,35		8,83	4,66	
Sucumbios			5,67	5,42				
Tungurahua		5,36	5,18	4,06	6,19	5,82	4,03	6,21
Zamora Chinchipe			6,13	22,89			7,36	

Fuente: SENESCYT - SNIIESE.

Figura 5.9: Años promedio de estudio por área y provincia - carreras 5 años

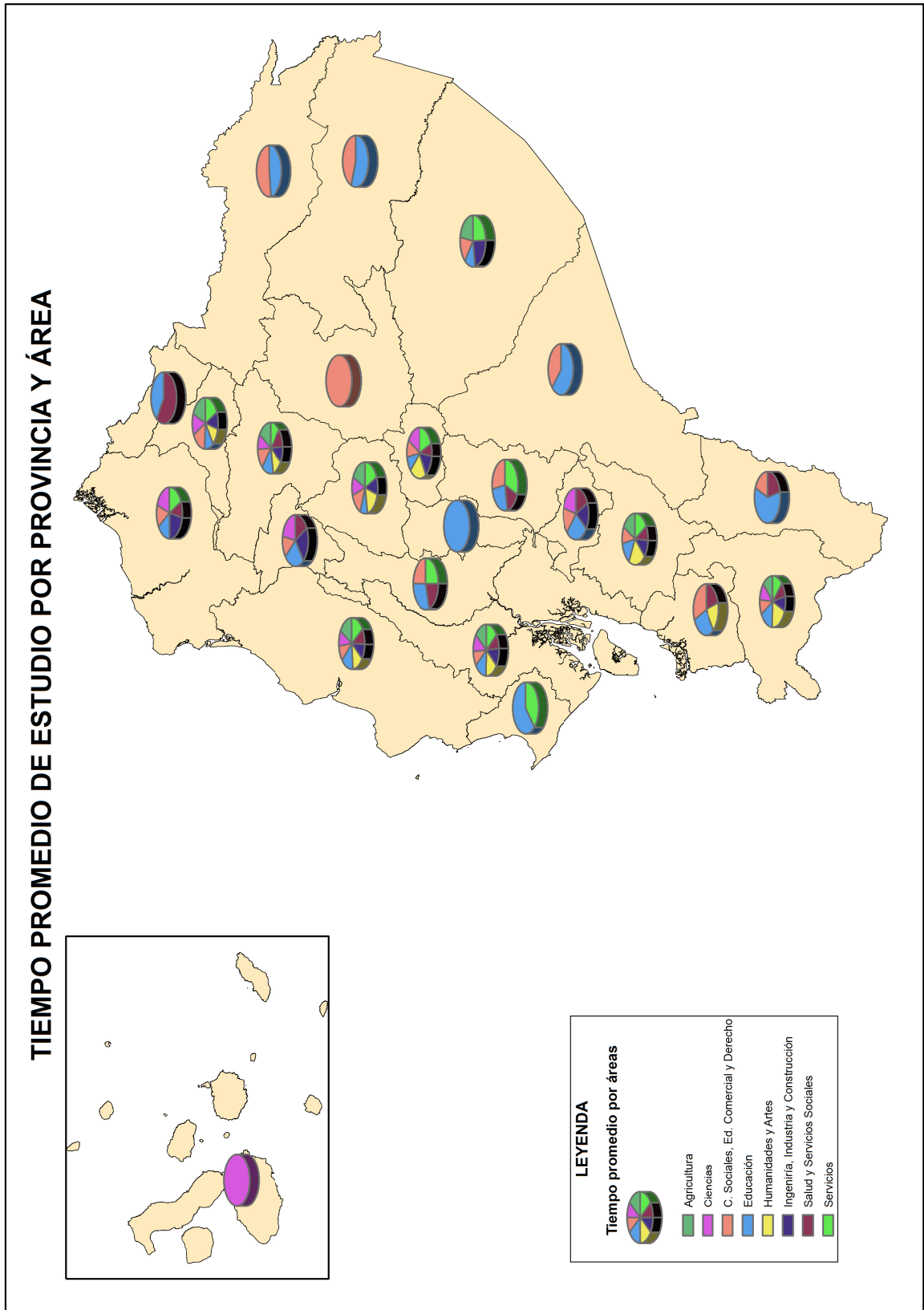


Figura 5.10: Tiempo promedio de estudio por área y provincia - carreras 5 años

3. Diferencia entre tiempo de titulación real y tiempo de titulación oficial por provincia.

Provincia	Diferencia Promedio
Azuay	1,85
Bolívar	-0,02
Cañar	0,86
Carchi	-0,61
Chimborazo	-0,81
Cotopaxi	1,38
El Oro	2,55
Esmeraldas	2,16
Galápagos	0,58
Guayas	1,61
Imbabura	3,15
Loja	3,35
Los Ríos	2,99
Manabí	1,95
Morona Santiago	1,40
Napo	0,34
Orellana	0,93
Pastaza	-0,79
Pichincha	3,26
Santa Elena	1,30
Santo Domingo	1,25
Sucumbios	0,45
Tungurahua	0,10
Zamora Chinchipe	8,84

Fuente: SENESCYT - SNIIESE.

Figura 5.11: Diferencia entre titulación real y oficial por provincia - carreras 5 años

Al comparar los valores obtenidos considerando la totalidad de titulados y a aquellos cuyas carreras tienen una duración de 5 años, las provincias de Zamora Chinchipe, Chimborazo, Galápagos y Los Ríos difieren en más de un año.

Particularmente llama la atención la provincia de Zamora Chinchipe que registra una diferencia promedio de 2,77 años en el primer caso y de 8,84 años en el segundo.

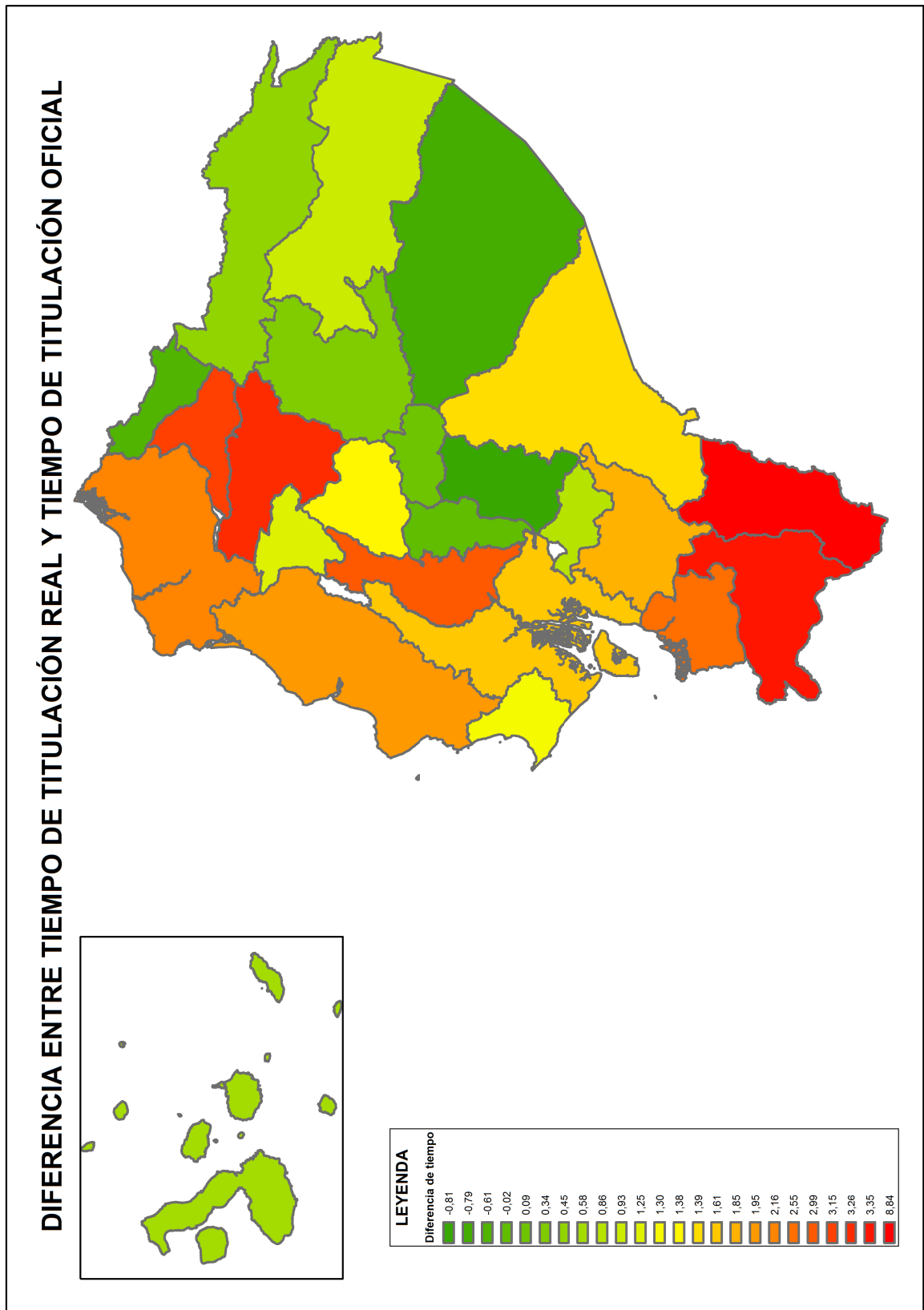


Figura 5.12: Diferencia promedio entre tiempo real y oficial por provincia - carreras 5 años

Bibliografía

- [1] HALL, P., C. HEYDE. (1980). *Martingale Limit Theory and Its Application*. New York, Academic Press Inc.
- [2] KNIGHT, K. (1999). *Mathematical Statistics*. United Kingdom, Chapman and Hall/CRC.
- [3] VALENCIA, E., Y. POVEDA, ET AL. (2012). *Desigualdad de Chebyshev bidimensional*. Scientia et Technical. 17(5): 242-246.
- [4] HOEL, P., S. PORT, ET AL. (1987). *Introduction to Stochastic Processes*. Boston, Waveland Press Inc.
- [5] RÜDT, D. (2010). *The Central Limit Theorem*. University of Toronto.
- [6] BLAIOTTA, J. AND P. DELIEUTRAZ. (2004). *Teorema Central del Límite*. Universidad de Buenos Aires.
- [7] KARLIN, S. AND H. TAYLOR. (1975). *A first course in Stochastic Processes*. United Kingdom, Academic Press Inc.
- [8] JIMENEZ, J., E. TODOROVA, ET AL. (2011). *Probabilidad de ruina en el modelo clásico de Cramer-Lundberg*. Temas de Ciencia y Tecnología. 15(45):9-18.
- [9] CABAÑA, E. (2002). *El Proceso de Wiener y el Teorema del Límite Central*. Boletín de la Asociación Matemática Venezolana. 9(2):181-192.
- [10] GÓMEZ, W. (2012). *El Proceso de Wiener y el Teorema del Límite Central*. Tesis Lic. Estadística. Nueva Segovia de Barquisimeto, Universidad Centroccidental Lisandro Alvarado.
- [11] RINCÓN, L. (2004). *¿Qué es la esperanza condicional?*. Miscelánea Matemática. (39):17-30.
- [12] MARMOLEJO, M. AND É. VALENCIA. (2009). *Martingalas discretas. Aplicaciones*. Revista Integración. 27(2):135-171.

- [13] HERNÁNDEZ, R. (2002). *Una introducción a la Teoría de Martingalas*. Tesis Lic. Estadística. Sonora, México, Universidad de Sonora.
- [14] NÚÑEZ, J. A. (2011). *Simulación de soluciones de ecuaciones diferenciales estocásticas con aplicación a las finanzas*. Tesis Ing. Departamento de Matemática. Quito, Escuela Politécnica Nacional.
- [15] ROSS, S. (2010). *Introduction to Probability Models*. 10 ed. Los Angeles, Elsevier Inc.
- [16] ECOS DE LA ECONOMÍA (2011). *Matemática Actuarial. No vida*.
- [17] ROSENTHAL, J. (2006). *A first look at rigorous probability theory*. 2 ed. London, World Scientific Publishing Co.
- [18] JACOD, J. AND P. PROTTER. (2004). *Probability Essentials*. Berlin, Springer.
- [19] SEVIM, C (2005). *On gamblers ruin problem in some models* Tesis Maestría. Ankara - Turquía, Bilkent University.
- [20] SIGMAN, K. (2009). *Stopping Times*.
- [21] HELLAND, I. (1982). *Central Limit Theorems for Martingales whith Discrete or Continuos Time*. Scand J Statist. (9): 79-94.
- [22] HANSEN, B. (2009). *Lecture Notes on Nonparametrics*. Wisconsin.
- [23] RODRÍGUEZ, L. (2012). *Construcción de kernels y funciones de densidad de probabilidad*. Escuela Superior Politécnica del Litoral.
- [24] LEDL, T. (2002). *Kernel density estimation. Theory, aspects of dimension and application in discriminant analysis*. Tesis Maestría. Viena, Universidad de Viena.
- [25] ARMSTRONG, D. (2005). *Optimal Bandwidth for the kernel*. R-Code University of Maryland.
- [26] CHERNOFF, H. (1959). *Sequential design of experiments*. The Annals of Mathematical Statistics . 30(3):755-770.
- [27] HORNA, L. (1985). *Sobre una aplicación del teorema central del límite para martingalas*. Memorias de la V conferencia de Jóvenes Científicos de la Universidad Estatal de Moscú.(1):48 - 51
- [28] MALYUTOV, M., V. SPOKOĬNYĬ, ET AL. (1993). *On asymptotic properties of estimates under sequential design*.

- [29] PULFORD, G.W. (2003). *Markov chain analysis of the sequential probability ratio test for automatic track maintenance*. Sixth International Conference of Information Fusion. (2): 1258 - 1265.
- [30] LAI, T.L. (2011). *Sequential analysis: some classical problems and new challenges*. Statistica Sinica. 11: 303 - 408.
- [31] UNESCO (2011). *Clasificación Internacional Normalizada de la Educación CINE 2011*. Instituto de Estadística de la UNESCO: 76-78.
- [32] RAMIREZ, G., M. VÁSQUEZ, ET AL. (2011). *Teoría de la robabilidad: Nociones fundamentales*. Universidad Central de Venezuela.
- [33] MONTES, F. (2007). *Introducción a la probabilidad*. Universidad de Valencia.
- [34] GRAFE, J. AND R. ESCUDER (1984). *Tipos extremos de distribuciones de probabilidad y funciones de distribución asociadas*. Estadística Española. 104(1): 35-47.
- [35] GONZÁLEZ, L. *Variable Aleatoria Continua*. Universidad de Sevilla.
- [36] FERNÁNDEZ, A. *Función Característica*. Tesis Ing. Departamento de Matemática. Universidad Nacional de Tucumán.
- [37] BICKEL, P., K. DOKSUM. (1977). *Mathematical Statistics*. San Francisco, Holden-Day Inc.
- [38] SHIRYAEV, A.N. (1995). *Probability*. New York, Springer.
- [39] MARTÍNEZ, J. AND J. VILLALÓN. (2003). *Introducción al Cálculo Estocástico. Aplicado a la Modelación Económico-financiera-actuarial*. Coruña, Netbiblo, S.L.