



## Proyecto Final: Simulación de algunos teoremas de probabilidad y estadística

Samantha Arzate **155804**

Mariana Martínez **166297**

Jorge Méndez **165696**

Tonantzin Real **163491**

18 de diciembre de 2020

Profesor Jorge De la Vega Góngora

# Índice

<b>1. Introducción</b>	<b>1</b>
<b>2. Datos</b>	<b>1</b>
<b>3. Métodos</b>	<b>1</b>
<b>4. Resultados</b>	<b>2</b>
4.1. Teorema del Límite Central . . . . .	2
4.1.1. ¿Qué pasaría si la varianza no fuera finita? . . . . .	6
4.1.2. ¿Y si las variables a aproximar fueran dependientes entre sí? . . . . .	7
4.1.3. ¿Y si no se distribuyeran idénticamente? . . . . .	9
4.2. Intervalos de Confianza . . . . .	10
4.3. Teorema de la convergencia dominada de Lebesgue . . . . .	14
4.4. Teorema de Bayes . . . . .	18
<b>5. Conclusiones</b>	<b>23</b>
<b>6. Referencias</b>	<b>24</b>
<b>7. Anexos</b>	<b>26</b>

## 1. Introducción

Debido a que nuestro cerebro no fue diseñado para manejar grandes cantidades de números o información al mismo tiempo, como sí puede hacerlo una computadora hoy en día, resulta relevante el papel que juega la simulación a la hora de ilustrar o ejemplificar conceptos más complejos. Este trabajo pretende facilitar la comprensión de algunos resultados estadísticos conocidos y estudiados anteriormente, haciendo uso de métodos vistos este semestre.

## 2. Datos

Todas las simulaciones del **Teorema del Límite Central** y de **intervalos de confianza** fueron realizadas con las funciones de R que generan valores de las distribuciones presentadas en este trabajo. Todos estos valores fueron generados con la semilla 123. La semilla usada para el cálculo de las integrales de las  $f_{i,n}$  fue 18. Usamos como  $n$  la secuencia del 10 al 1000 con brincos de tamaño 10. Por último, en el problema captura y recaptura, que se resuelve con un enfoque bayesiano, se utiliza la semilla 165696 para generar las muestras aleatorias de tamaño 100,000 de la distribución a priori.

## 3. Métodos

Por un lado, para las simulaciones del **Teorema del Límite Central** se programó una función que generó  $N$  muestras de tamaño  $n$  de las distribuciones uniforme, exponencial, Poisson y binomial y otra función que plasmó en una sola imagen las densidades y la media de todas las muestras para que se facilitara la presentación de resultados.

Además, para analizar los supuestos del TLC, se construyeron, a través de R, tres escenarios diferentes en los que se violara el supuesto a revisar y se respetaran los otros dos. Posteriormente, se intentó forzar una aproximación normal sobre las medias muestrales para cada caso y se comparó con los datos reales obtenidos.

Por otro lado, para las simulaciones de los **intervalos de confianza**, se programó una función que contó los intervalos que contenían al parámetro poblacional y otra función que graficó esta situación. Para la cobertura de los intervalos de confianza, se programó otra función en la que, para cada valor de  $\theta \in [0, 1]$ , se mostrara el porcentaje de cobertura del intervalo de confianza correspondiente.

En el caso del **Teorema de la convergencia dominada de Lebesgue (TCDL)** primero se dan unas definiciones previas necesarias para entender el teorema. Luego se dan 3 ejemplos de dicho teorema. Para poder calcular las integrales de  $f_i$  y de  $f_{i,n}$  (para las  $n$  mencionadas en la sección anterior), se usó el Método de Monte Carlo Crudo. Para cada  $n$  se calculó el intervalo de confianza de la forma  $\left(2\hat{\theta} - \hat{\theta}_{1-\alpha/2}, 2\hat{\theta} - \hat{\theta}_{\alpha/2}\right)$  con  $\alpha = 0.1$ . El  $\hat{\theta}_{1-\alpha/2}$  para cada  $n$  se calculó con los cuantiles de la muestra de las integrales de  $f_{i,n}$  de 1 a  $n$  para todas las  $n$  consideradas. Esto permite visualizar que no

solo la integral converge puntualmente, sino que los intervalos de confianza sí incluyen la integral límite. En un ejemplo se muestra que el TCDL no se cumple si no se cumplen todos los supuestos.

Por último, en la sección dedicada al **teorema de Bayes**, se comienza con el enunciamiento del teorema. Procedemos a explicar la relevancia de este resultado para el desarrollo de la inferencia bayesiana y presentamos los pasos a seguir para aplicar esta metodología de análisis de datos. Después, presentamos un ejemplo teórico de la aplicación del teorema y su importancia para actualizar la probabilidad de que algo es verdad dadas las evidencias. Al final, como ejemplo práctico, presentamos el método de Computación Bayesiana Aproximada que es un enfoque que estima la función de verosimilitud basado en simulaciones.

## 4. Resultados

### 4.1. Teorema del Límite Central

Uno de los resultados más relevantes y más usados de la teoría de Probabilidad y Estadística es el **Teorema del Límite Central**, este resultado nos ha facilitado conocer la distribución aproximada de la media de una variable aleatoria cuando  $n \rightarrow \infty$ ; aplicable, por ejemplo, en procesos estocásticos con los Procesos de Poisson Compuestos o como motivación para las simulaciones de Monte Carlo.

**Teorema 1** (Teorema del Límite Central (TLC)). *Sea  $X_1, X_2, \dots, X_n$  un conjunto de variables aleatorias, independientes e idénticamente distribuidas con media  $\mu$  y varianza  $0 < \sigma^2 < \infty$ . Sea*

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

*Entonces,*

$$\lim_{n \rightarrow \infty} \Pr \left( \frac{S_n - n\mu}{\sigma\sqrt{n}} \leq z \right) = \Phi(z).$$

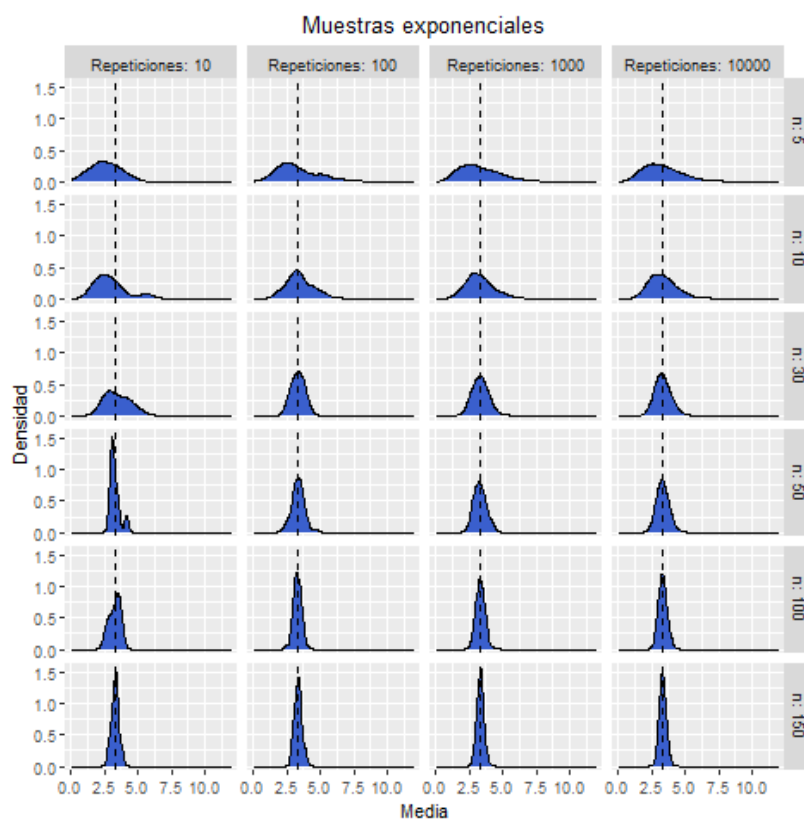
*Es decir, si  $n$  es suficientemente grande, la variable aleatoria*

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

*tiene aproximadamente una distribución normal con  $\mu_{\bar{X}} = \mu$  y  $\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}$ .*

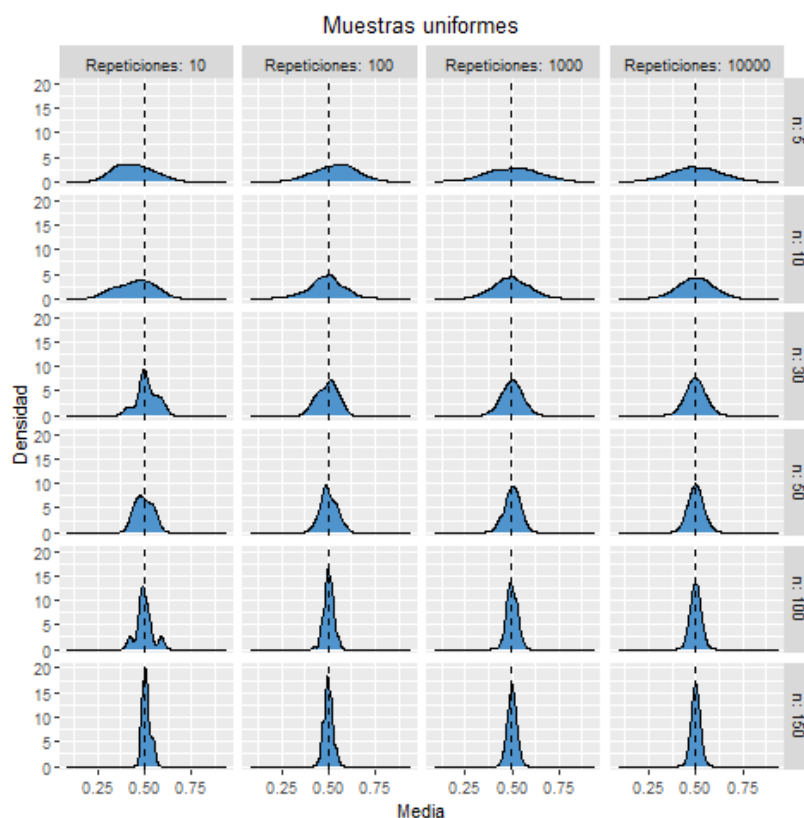
Para comprender mejor el teorema, a continuación presentamos la simulación de cuatro distribuciones (dos continuas y dos discretas), cada una con diferentes tamaños de muestra  $n$  y diferente número de repeticiones por simulación.

## Variables aleatorias continuas



**Figura 1:** Variables aleatorias exponenciales

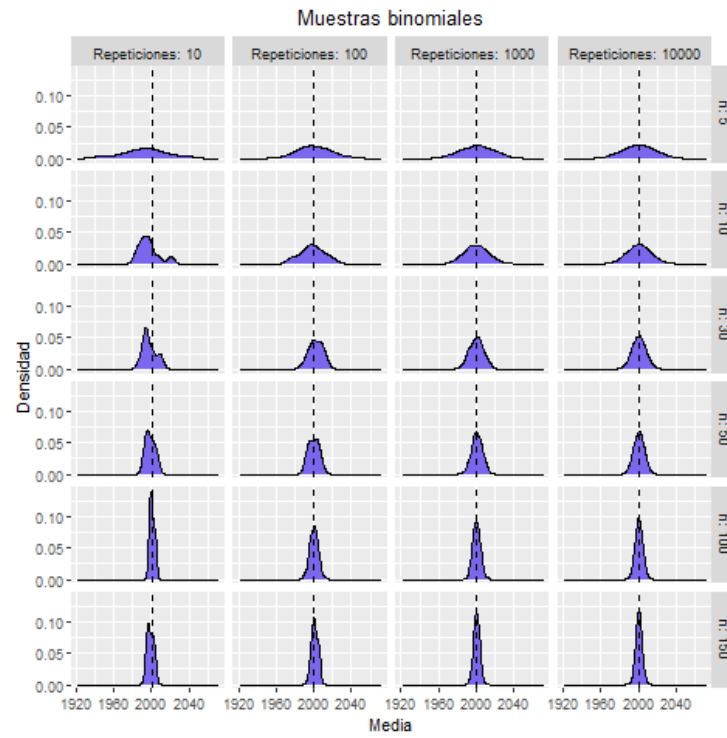
Notemos que a pesar de que las densidades provienen de dos diferentes variables aleatorias continuas, conforme la  $n$  aumenta, la densidad se asemeja a la de una campana gaussiana. Tanto en la Figura 1 como en la Figura 2 podemos notar que la mejor aproximación a una distribución normal sucede cuando las repeticiones son 10000 y a partir de un tamaño de muestra  $n = 30$  pues las densidades siguen más claramente una normal. De igual forma, podemos notar que entre menos repeticiones y menor tamaño de muestra, las densidades no se parecen a una normal. Por último podemos ver que la media en la mayor parte de las gráficas corta a la mitad a las densidades; es decir, hay simetría.



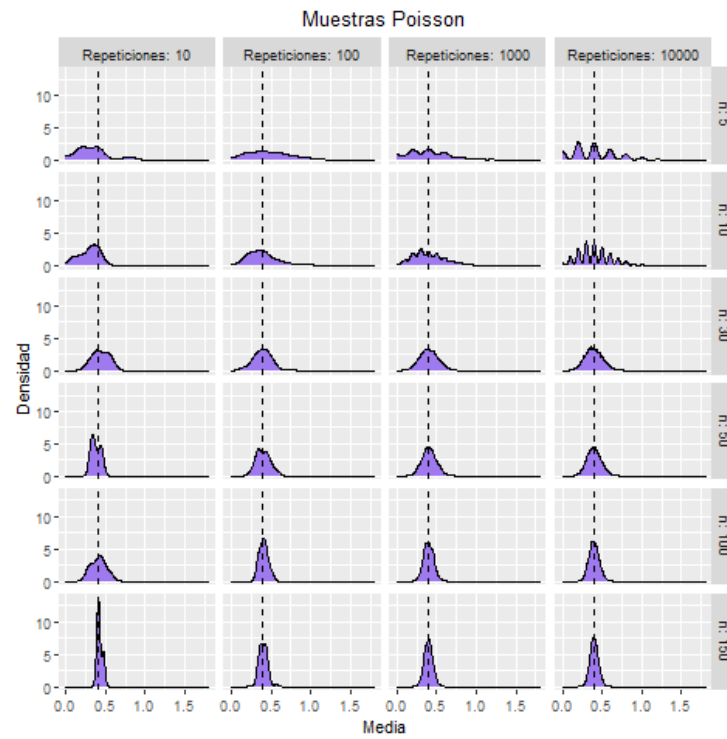
**Figura 2:** Variables aleatorias Uniformes (0,1)

### Variables aleatorias discretas

Lo que se puede decir con estas distribuciones es muy similar a lo que sucede con las distribuciones continuas: a mayores repeticiones y tamaño de muestra, mejor aproximación a una densidad normal. En la Figura 4 llama la atención la densidad que se forma con las repeticiones 1000 y 10000 y  $n = 5$  y  $n = 10$ ; sin embargo, en ambos casos, conforme el tamaño de la muestra crece, la densidad deja de tener varias ondas y, como bien menciona el TLC, se asemeja a una distribución normal. Situación que desde un inicio es más clara para la Figura 3.



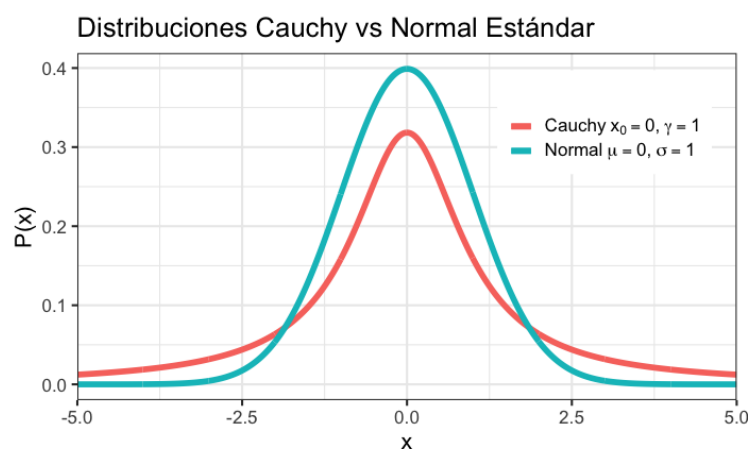
**Figura 3:** Variables aleatorias Binomial



**Figura 4:** Variables aleatorias Poisson

### 4.1.1. ¿Qué pasaría si la varianza no fuera finita?

Como vimos anteriormente, el Teorema del Límite Central (en su versión original) requiere que los datos a aproximar con una distribución normal provengan de una distribución con varianza finita. La distribución Cauchy, la cual es un caso especial de la distribución  $t$ , es un ejemplo de distribución con varianza no finita, de hecho, la distribución Cauchy ni siquiera tiene media finita. A continuación podemos ver la gráfica de la densidad Cauchy (naranja) junto a la gráfica de la densidad de una distribución Normal estándar (verde) para ayudarnos a compararlas. Notemos que la distribución Cauchy tiene un pico más corto y estrecho que la distribución normal, pero tiene colas más anchas.



**Figura 5:** Aunque la distribución Cauchy es más baja que la distribución Normal, sus colas son más anchas

Para ilustrar el problema que existiría al tratar de aproximar variables Cauchy con el Teorema del Límite Central, generamos 1000 muestras de tamaño 100 a partir de una distribución  $\text{Cauchy}(0,1)$  y calculamos la media de cada muestra obtenida. Observar estas 1000 medias muestrales debería darnos una idea de la distribución de las mismas. A continuación, las estadísticas descriptivas obtenidas a partir de nuestras 1000 muestras:

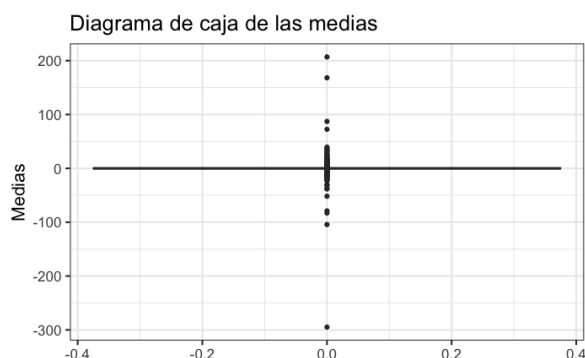
Media	Desv.Est.	Mínimo	Q1	Mediana	Q3	Máximo
3.857e-03	1.502e+01	-2.948e+02	-9.416e-01	3.549e-02	9.561e-01	2.069e+02

**Tabla 1:** Estadísticas descriptivas sobre las medias muestrales obtenidas, fijando la semilla en 123

Notemos que la desviación estándar es mucho mayor que el rango intercuartílico, y los valores mínimos y máximos parecen ser valores atípicos muy extremos.

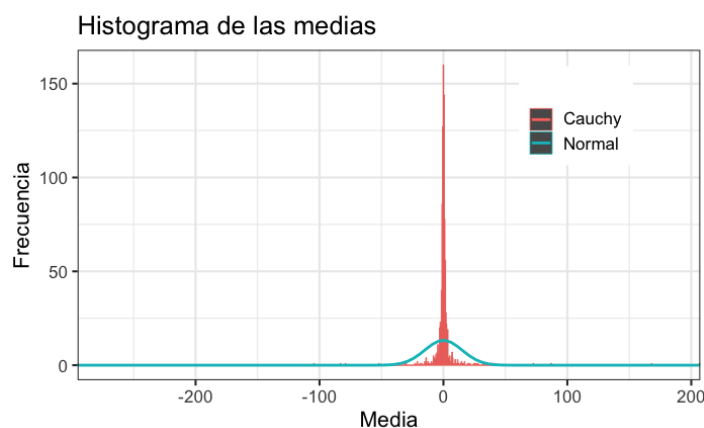
A continuación, podemos ver la gráfica de caja de las medias de estas 1000 muestras. Cabe resaltar que es bastante inusual, con una caja muy delgada y varios valores atípicos extremos.





**Figura 6:** Diagrama de caja para las medias de las muestras obtenidas

Aquí hay un histograma de las 1000 medias muestrales que ilustra la distribución muestral de las medias para muestras de tamaño 100 de la distribución de Cauchy. Además, para facilitar la comparación, sobrepusimos una distribución normal con la misma media y desviación estándar de las 1000 medias simuladas. Así, podemos concluir que, efectivamente, esta distribución muestral está lejos de ser normal!

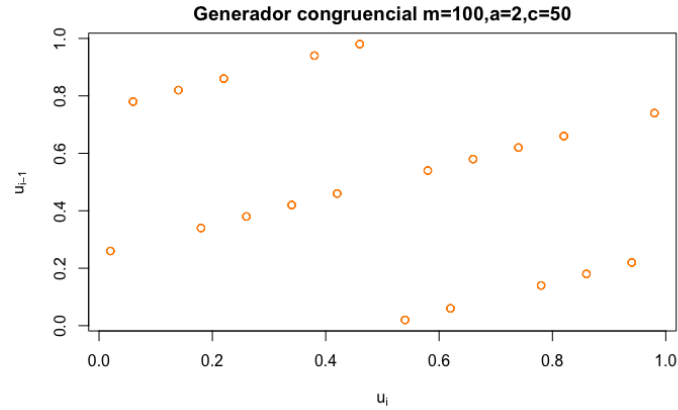


**Figura 7:** Hist. de las medias muestrales obtenidas vs frec. normales esperadas

#### 4.1.2. ¿Y si las variables a aproximar fueran dependientes entre sí?

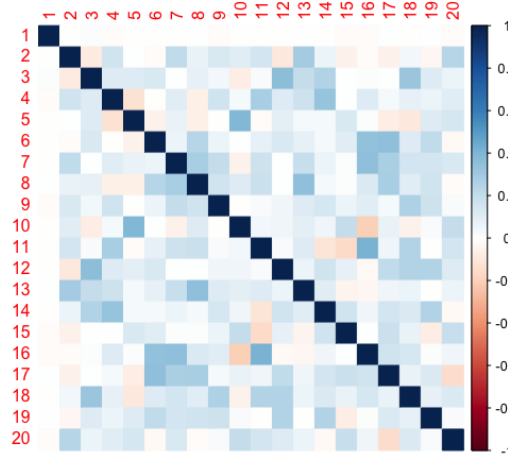
Para ilustrar la falla en la aproximación de la distribución muestral a través del TLC, generaremos otras 1000 muestras tamaño 100 de una distribución Uniforme(0,1), pero simuladas de tal manera que dichas 100 variables aleatorias estén correlacionadas.

Así, con la intención de introducir dependencia a las variables generadas, decidimos utilizar generadores lineales congruenciales (GLC) a la hora de simularlas. Pues, como vimos en clase, una de las deficiencias en la calidad de los GLC es que las muestras generadas pueden mostrar cierto nivel de dependencia serial, en particular autocorrelación.



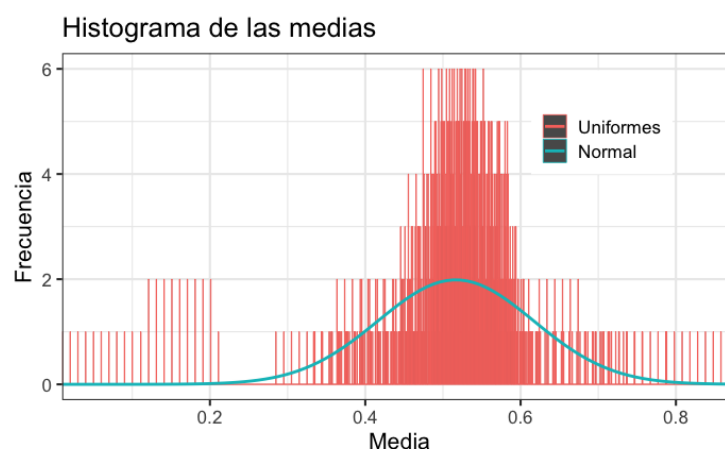
**Figura 8:** Ejemplo de simulación de variables aleatorias  $U(0,1)$  a través de GLM

Una vez generadas nuestras muestras, corroboramos que efectivamente hayan indicios de dependencia lineal entre las variables mediante el siguiente diagrama de correlaciones. (Notar que para facilitar su visualización, nos limitamos a presentar en este archivo el diagrama de correlación para las primeras 20 variables simuladas).



**Figura 9:** Las variables simuladas tienen correlaciones distintas a cero

Finalmente, tras calcular las medias de las muestras obtenidas, intentamos ajustar una distribución normal a los datos obtenidos, como lo haríamos según el TLC. Sin embargo, como era de esperarse, el ajuste no fue nada bueno y podemos verlo gráficamente en la siguiente figura.

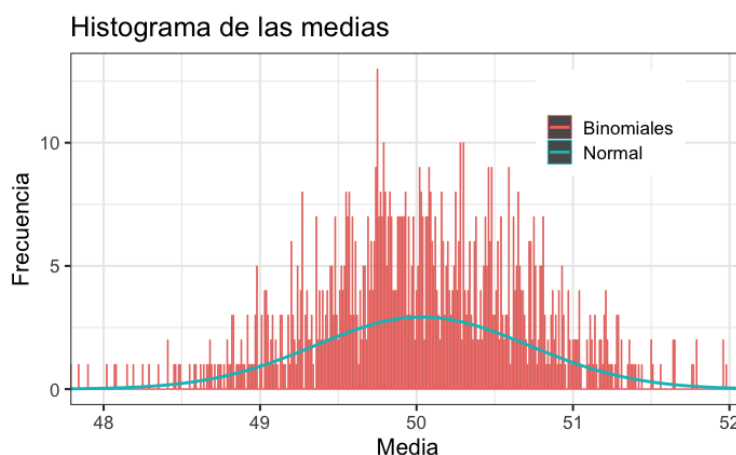


**Figura 10:** Comparación entre el histograma de las medias muestrales obtenidas y las frecuencias normales esperadas.

#### 4.1.3. ¿Y si no se distribuyeran idénticamente?

El último de los supuestos del TLC que nos queda por analizar es el supuesto de que todas las variables de la sucesión, deben ser idénticamente distribuidas. Así, con la intención de ver qué podría pasar si este supuesto no se cumple, generamos, una vez más, 1000 muestras de tamaño 100, pero esta vez a partir de distribuciones Binomiales con  $n = j$  y  $p = 1/j$  para cada  $j = 1, \dots, 100$ .

Luego, tras calcular las medias de las muestras obtenidas y tratar de ajustar una distribución normal a su histograma, obtuvimos el siguiente resultado:



**Figura 11:** Comparación entre el histograma de las medias de la muestra de tamaño 1000, contra las frecuencias normales esperadas. (Semilla 123)

A pesar de que el histograma obtenido parece tener una forma acampanada, el ajuste normal con la media y la varianza recomendadas por el TLC claramente no son suficientes para otorgar un buen ajuste.

## 4.2. Intervalos de Confianza

Si bien el concepto de **intervalos de confianza** (IC) se ve formalmente hasta estadística matemática, a partir de ese momento es recurrente en varias materias. En esta sección presentaremos diversos gráficos de los intervalos de confianza para la media  $\mu$  de la distribución normal y exponencial.

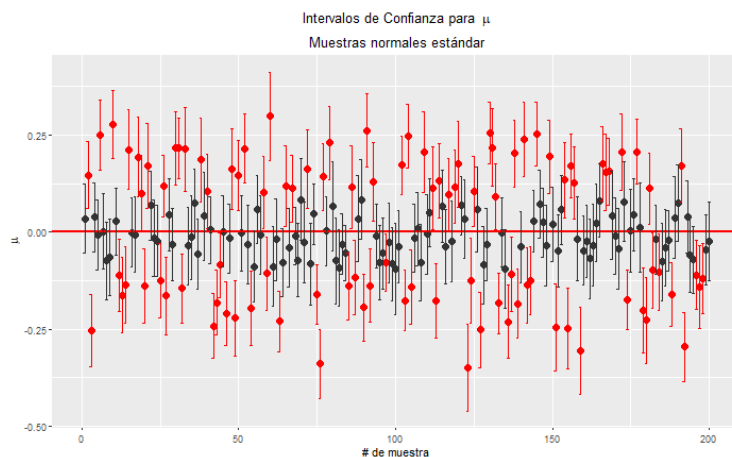
La idea que nos enseñan acerca de los intervalos de confianza para la media poblacional es que, como solo conocemos la media muestral, un IC para  $\mu$  del  $100(1 - \alpha) \%$  significa que la media real  $\mu$  estará contenida en el  $100(1 - \alpha) \%$  de los intervalos de la media de las muestras que se tomen de la población completa.

Recordemos que para una  $X_1, X_2, \dots, X_n$  muestra aleatoria  $N(\mu, \sigma^2)$  con  $\sigma^2$  conocida, el intervalo de confianza del  $100(1 - \alpha) \%$  para  $\mu$  está dado por:

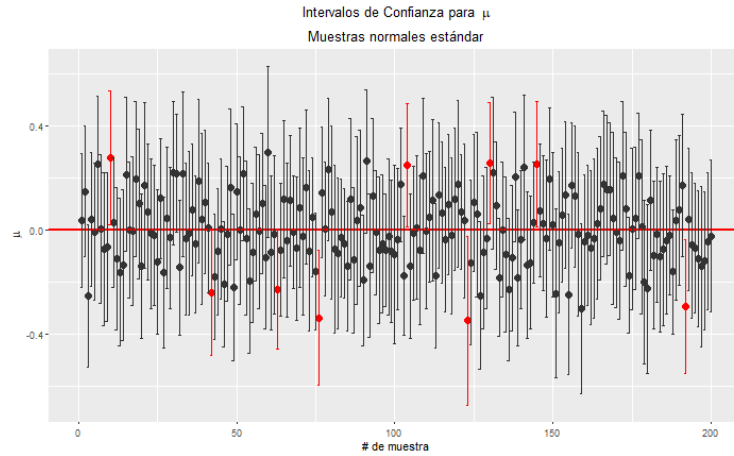
$$\bar{x} \pm z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

donde  $z_{1-\frac{\alpha}{2}}$  es el percentil  $1 - \frac{\alpha}{2}$  de una  $N(0, 1)$

Para entender mejor a qué se refiere, visualicemos primero las siguientes figuras de variables aleatorias normales estándar:



**Figura 12:** Intervalo de confianza del 50 %



**Figura 13:** Intervalo de confianza del 95 %

En ambas gráficas, se muestran 200 IC (líneas negras y rojas) para las medias muestrales correspondientes (puntos negros y rojos) donde la línea horizontal que pasa por 0 se refiere a la media teórica pues las variables fueron normales estándar. Notemos que en la Figura 12, prácticamente la mitad de los intervalos presentados son rojos; es decir, dichos intervalos no contienen a  $\mu = 0$ . De las 10000 muestras que se generaron, el 50.18 % de los IC contuvieron a  $\mu$ , lo cual es aproximadamente el mismo porcentaje de los 200 IC mostrados en la esta figura. En cuanto a la Figura 13 podemos ver que únicamente 9 de los 200 IC presentados no contienen a  $\mu = 0$ , lo que significa que el 95.5 % de los IC sí contienen al parámetro. En este caso, se obtuvo que de las 10000 muestras que se generaron, el 94.86 % de los IC contuvieron a  $\mu$ .

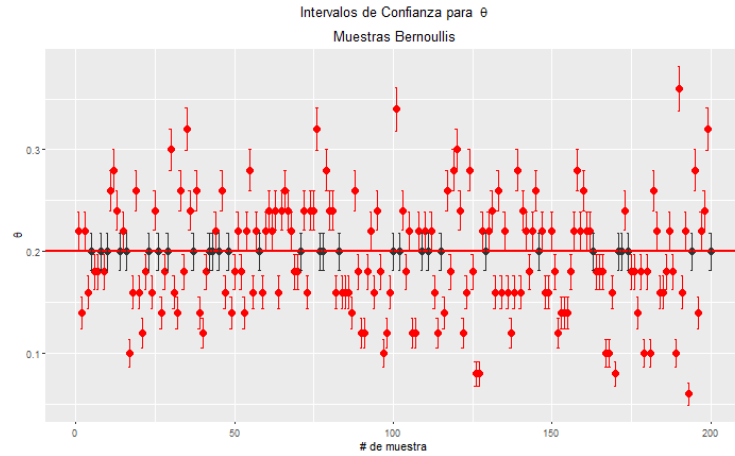
Las figuras anteriores nos permitieron ver que los IC del  $1 - \alpha$  definido para  $\mu$  de variables aleatorias normales estándar, fueron acorde a lo que nos enseñaron en estadística matemática. A continuación mostraremos qué pasa en el caso de variables aleatorias binomiales.

Para este caso tenemos que si  $X_1, X_2, \dots, X_n$  es una muestra aleatoria  $Bin(n, \theta)$  cuando  $n$  es grande, el intervalo de confianza del  $100(1 - \alpha) \%$  para  $\theta$  está dado por:

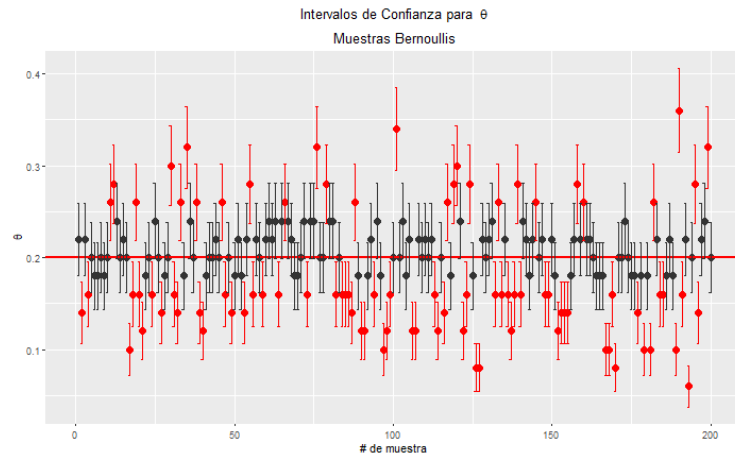
$$\bar{x} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\bar{x}(1-\bar{x})}{n}}$$

donde  $z_{1-\frac{\alpha}{2}}$  es el percentil  $1 - \frac{\alpha}{2}$  de una  $N(0, 1)$

Observemos que la Figura 14 tiene muchos más IC que no contienen a  $\theta = 0.2$  que la Figura 15, lo cual tiene sentido puesto que la primera figura es un IC del 25 % mientras que el segundo es del 50 %. De las 10000 muestras generadas, para el IC del 25 %, únicamente el 13.35 % de los intervalos contuvieron a la media teórica  $\theta = 0.2$  mientras que para el IC del 50 %, el 50.38 % lo contuvieron. Pero, ¿cómo es posible que un IC supuestamente del 25 % únicamente contenga al parámetro  $\theta$  en el 13.35 % de los casos! ¿Será que sucede algo similar si elegimos otro intervalo? Veamos el caso del IC del 95 %:



**Figura 14:** Intervalo de confianza del 25 %

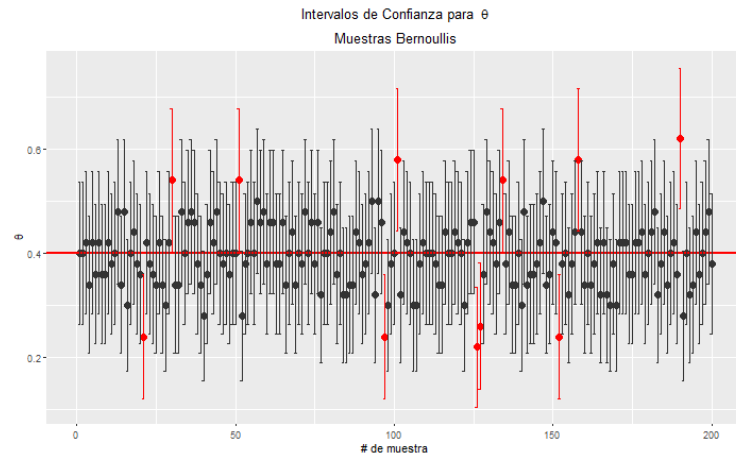


**Figura 15:** Intervalo de confianza del 50 %

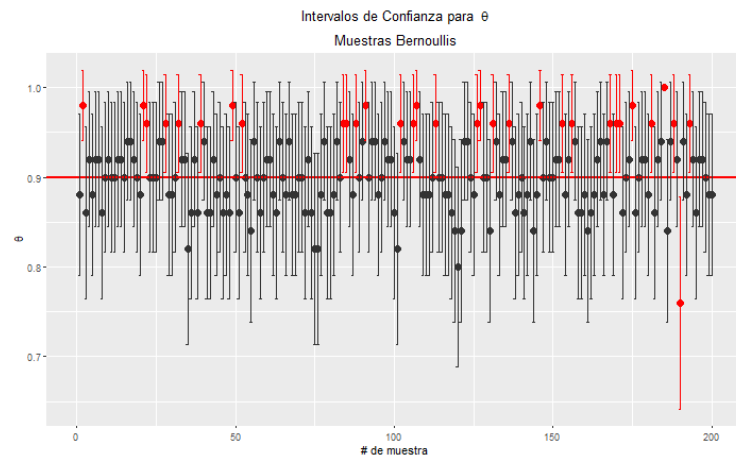
**Figura 16:** Variables aleatorias Binomial(50,0.2)

Es evidente que la Figura 18 tiene muchos más IC que no contienen al parámetro real  $\theta = 0.9$  que la Figura 17 pues en ese caso solo 11 de 200 IC no contienen a  $\theta = 0.4$ ; es decir, el 94.5 % de los intervalos sí contienen a  $\theta$ . De las 10000 muestras que se generaron, el 94.5 % de los IC referentes a las variables Bin(50,0.4) sí incluyeron a  $\theta = 0.4$ , lo cual era de esperarse puesto que el IC buscado era del 95 %. Sin embargo, para las variables Bin(50,0.9) únicamente el 88 % de los intervalos contuvieron a  $\theta = 0.9$ . Pero, ¿por qué pasa esto si nada más cambiamos a  $\theta$ ?

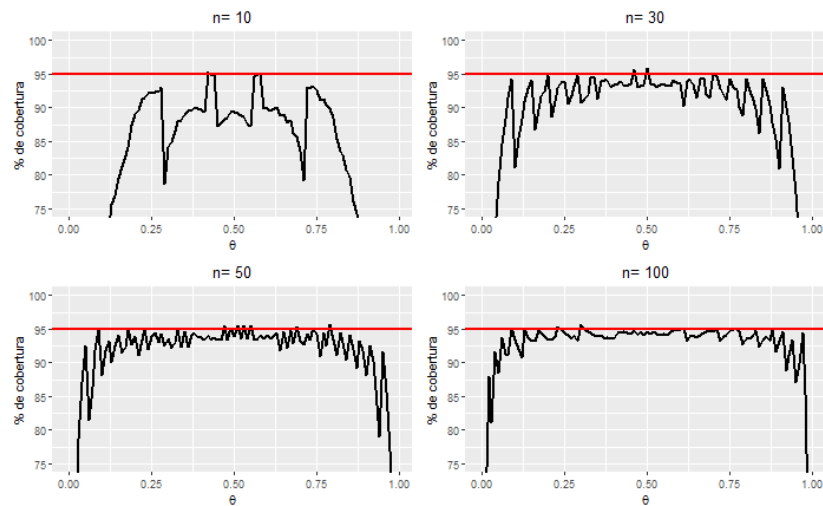
Para entender por qué sucede lo que hemos mencionado, mostraremos la cobertura de los IC para el caso binomial para diferentes tamaños de muestra  $n$  y valores de  $\theta$ .



**Figura 17:** Intervalo de confianza del 95 % para  $\text{Bin}(50,0.4)$



**Figura 18:** Intervalo de confianza del 95 % para  $\text{Bin}(50,0.9)$



**Figura 19:** Cobertura de intervalos de confianza para  $\theta$

En la Figura 19 hay cuatro gráficos que muestran la cobertura de intervalos de confianza del 95 % para todos los valores de  $\theta$  de la distribución Binomial y con diferentes tamaños de muestra  $n$ . Notemos que conforme  $n$  aumenta, la cobertura de los IC aumenta para más valores de  $\theta$ . Cuando  $n = 10$ , únicamente dos valores (pareciera  $\theta = 0.45$  y  $\theta = 0.55$ ) alcanzan el porcentaje de cobertura del 95 % mientras que para todos los demás valores de  $\theta$  la cobertura es mucho menor que la deseada. En el caso de  $n = 30$ , hay más valores de  $\theta$  para los cuales se alcanza e inclusive se pasa del 95 %. Cuando  $n = 50$ , hay varios picos que alcanzan la cobertura del 95 %; sin embargo, podemos ver que para valores de  $\theta$  cercanos a 0 y a 1, el porcentaje de cobertura alcanza 88 % e incluso valores menores al 75 %. Finalmente, cuando  $n = 100$  el rango de valores de  $\theta$  para los cuales se alcanza una cobertura del 95 % es mucho mayor que para el resto de los tamaños de muestra  $n$  presentados. Sobre todo podemos ver que cuando  $\theta \approx 1$ , el porcentaje de cobertura aumentó significativamente con respecto a todas las demás  $n$ .

### 4.3. Teorema de la convergencia dominada de Lebesgue

El teorema que vamos a discutir en esta sección es uno de los teoremas estrella de la teoría de la medida. Primero vamos a hablar un poco sobre qué es teoría de la medida, qué hace, luego vamos a dar unas nociones básicas de esta rama tan abstracta de las matemáticas. Estas nociones son necesarias para poder entender lo que el teorema de la convergencia dominada de Lebesgue (TCDL) dice. No vamos a incluir demostraciones pues no es el punto de este trabajo, pero si el curioso lector quiere saber más sobre esto le podemos recomendar que lea a Grabinsky (2016) en [4] o a Tao (2011) en [9]. Si el lector quiere saber más sobre el impacto que tiene la teoría de la medida le podemos recomendar el libro de Ash y Doleans-Dade (1999) en [7]. Todas estas referencias las puede encontrar en este trabajo en la sección de bibliografía. Para poder entender mejor nuestros dos teoremas estrella y como introducción a la teoría de la medida en general, vamos a empezar dando unas definiciones básicas de esta bella rama.

**Definición 1.** Una clase no vacía del conjunto potencia de  $\mathcal{X}$ ,  $\mathcal{S} \subseteq \mathcal{P}(\mathcal{X})$  se llama  $\sigma$ -álgebra si cumple con los siguientes puntos:

- $E, F \in \mathcal{S} \Rightarrow E \setminus F \in \mathcal{S}$
- Si  $E_1, E_2, \dots$  es una sucesión de elementos en  $\mathcal{S}$ , entonces  $\bigcup_{k \in \mathbb{N}} E_k \in \mathcal{S}$
- $\mathcal{X} \in \mathcal{S}$

Tener esta definición ayuda mucho pues, tal vez nos topemos con una clase de conjuntos de  $\mathcal{X}$ , pero que no necesariamente sea una  $\sigma$ -álgebra pero que sí querríamos definir una medida sobre ella. En tal caso trabajaríamos con la  $\sigma$ -álgebra generada.

**Definición 2.** Supongamos que  $\mathcal{E}$  es una clase de conjuntos de  $\mathcal{X}$  (es decir, un conjunto cuyos elementos son subconjuntos de elementos de  $\mathcal{X}$ ). En este caso  $\mathcal{E}$  puede o no formar



una  $\sigma$ -álgebra. Entonces siempre va a existir una única  $\sigma$ -álgebra denotada como  $S(\mathcal{E})$  con las siguientes propiedades:

- $\mathcal{E} \subseteq S(\mathcal{E})$
- $S(\mathcal{E})$  es la menor (con menor elementos)  $\sigma$ -álgebra que contiene a  $\mathcal{E}$

A esta  $\sigma$ -álgebra le vamos a llamar la  $\sigma$ -álgebra generada por  $\mathcal{E}$ .

A continuación definimos a los Borelianos, una de las  $\sigma$ -álgebras más importantes.

**Definición 3.** La  $\sigma$ -álgebra de Borel en  $\mathbb{R}$  se define como:

$$\mathcal{B}_{\mathbb{R}} = S(\mathcal{A}),$$

donde  $\mathcal{A}$  son los conjuntos abiertos de  $\mathbb{R}$ .

**Definición 4.** Una vez que tengamos a  $\mathcal{X}$  y a una  $\sigma$ -álgebra definida con esto, entonces la pareja  $(\mathcal{X}, \mathcal{S})$  se llama espacio medible.

Es decir, ya tenemos una pareja para poder “medir”.

**Definición 5.** Sean  $(\mathcal{X}, \mathcal{S})$  y  $(\hat{\mathcal{X}}, \hat{\mathcal{S}})$  dos espacios medibles. Decimos que una función  $f : \mathcal{X} \rightarrow \hat{\mathcal{X}}$  es  $\mathcal{S} - \hat{\mathcal{S}}$  medible si  $f^{-1}(\hat{E}) \in \mathcal{S}$  para cualquier  $\hat{E} \in \hat{\mathcal{S}}$ .

Para simplificar notación, diremos que una función  $f$  es  $\mathcal{S}$  medible si  $(\hat{\mathcal{X}}, \hat{\mathcal{S}})$  es  $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ . Diremos que  $f$  es Borel-medible si también  $(\mathcal{X}, \mathcal{S})$  es  $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ .

Lo que vamos a definir a continuación muy probablemente impresione a nuestro querido lector pues a raíz de esta definición vamos a poder escribir  $[-\infty, \infty]$  sin ningún problema.

**Definición 6.** Definimos a la recta real extendida al agregarle dos símbolos,  $-\infty$  y  $\infty$ , es decir  $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty\} \cup \{\infty\}$ .

Con estos dos símbolos las operaciones válidas y propiedades en  $\overline{\mathbb{R}}$  son: no se puede dividir entre  $-\infty$ ,  $\infty$  o  $0$ ,  $-\infty < x < \infty$  para cualquier  $x \in \mathbb{R}$  y respetamos el orden usual de  $\mathbb{R}$ ,  $(\pm\infty) + x = x + (\pm\infty) = \pm\infty$ ,  $\infty + \infty = \infty$ ,  $(\pm\infty)(\pm\infty) = \infty$  y  $(\pm\infty)(\mp\infty) = -\infty$ . Con la recta real extendida podemos hacer definiciones de los Borelianos, funciones medibles, etc. como las hicimos para la recta real. Dadas todas estas definiciones básicas ya podemos definir lo que es una medida.

**Definición 7.** Sea  $(\mathcal{X}, \mathcal{S})$  un espacio medible fijo. Una medida,  $\mu$ , es una función conjuntista, es decir  $\mu : \mathcal{S} \rightarrow \overline{\mathbb{R}}$  tal que:

- $\mu(E) \geq 0$  para todo  $E \in \mathcal{S}$
- $\mu(\emptyset) = 0$

- La medida de la unión de elementos disjuntos es la suma de la medida de estos elementos, es decir:

$$\mu \left( \bigsqcup_{k \in \mathbb{N}} E_k \right) = \sum_{k \in \mathbb{N}} \mu(E_k)$$

Todos estos puntos son lo que nos esperaríamos de una medida (hablando coloquialmente). ¿Qué tiene de distinto esto? La respuesta es que nunca especificamos quién es el espacio medible  $(\mathcal{X}, \mathcal{S})$ , puede ser un conjunto abstracto, un conjunto de funciones, un conjunto de cosas concretas, etc. En este trabajo nos vamos a centrar en la pregunta ¿cuándo podemos “meter” un límite a una integral? Para esto vamos a por fin enunciar el teorema de la convergencia dominada de Lebesgue (TCDL).

**Teorema 2** (Teorema de la convergencia dominada de Lebesgue (TCDL)). *Sea  $(\mathcal{X}, \mathcal{S}, \mu)$  un espacio de medida y  $(f_n)_{n=1}^{\infty}$  una sucesión en la cerradura de  $(\mathcal{X}, \mathcal{S})$  supongamos que existe una función  $g$  positiva y que  $\int |g| d\mu < \infty$ . Supongamos que  $|f_n| \leq g$  casi donde sea con respecto a  $\mu$  y para todo  $n \in \mathbb{N}$ , entonces si  $f(x) = \lim_{n \rightarrow \infty} f_n(x)$ :*

1.  $f \in \mathcal{L}_1(\mu)$ , es decir,  $\int |f| d\mu < \infty$
2. para todo  $E \in \mathcal{S}$  tenemos que  $\int_E f d\mu = \lim_{n \rightarrow \infty} \int_E f_n d\mu$ .

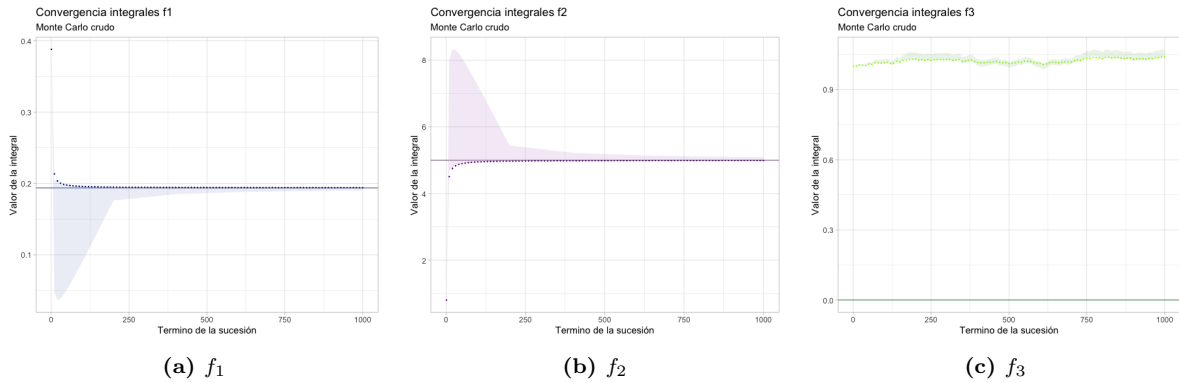
Vamos a ejemplificar este teorema. El querido lector se podrá dar cuenta que las funciones de probabilidad acumulada son un caso particular de medidas, se le invita a revisar que cumplen los 3 puntos de la definición de medida. Por ende vamos a usarlas para ejemplificar el TCDL. Vamos a usar integración de Monte Carlo crudo y luego con variadas antitéticas. Dentro de estas funciones hay una que no cumple con los supuestos del TCDL y por ende no podemos “meter” el límite a la integral. Usando simulación vamos a graficar esto para que el querido lector pueda ver que sí son necesarios todos los supuestos del teorema. Notamos que la función que no cumple con los supuestos es  $f_{3,n}$  pues no hay función alguna que la domine por el factor  $n$  si  $0 < x \leq 1/n$ . Además vamos a hacer uso del teorema de la modificación de la integral pues todas las funciones son Riemann-integrables. Es decir, podemos “modificar” la integral de la siguiente manera usando la derivada de la medida.

$$\int_E f(x) d\mu = \int_E f(x) \mu'(x) dx.$$

Sucesión de funciones	Función límite	Medida usada
$f_{1,n} = \frac{nx+x}{n}$	$f_1 = x$	Distribución exponencial con $\lambda = 2$
$f_{2,n} = \frac{nx^3 - 10nx^2 + 25nx - 10x + 25}{1+nx}$	$f_2 = (x - 5)^2$	Distribución gamma con $\alpha = 5$ y $\beta = 1$
$f_{3,n} = \begin{cases} n & \text{si } 0 < x \leq 1/n \\ 0 & \text{otro caso} \end{cases}$	$f_3 = \begin{cases} \infty & \text{si } x = 0 \\ 0 & \text{otro caso} \end{cases}$	Distribución uniforme en $[0, 1]$

**Figura 20:** Ejemplos simulados para el TCDL

Usando la técnica de Monte Carlo Crudo obtenemos las siguientes gráficas. Podemos ver la convergencia en cada uno de los casos mencionados arriba. Pusimos intervalos de confianza del estilo  $(2\hat{\theta} - \hat{\theta}_{1-\alpha/2}, 2\hat{\theta} - \hat{\theta}_{\alpha/2})$  con  $\alpha = 0.1$  para mostrar que, efectivamente el límite teórico se encuentra dentro de estas bandas además que la estimación puntual parece que sí converge.

**Figura 21:** Ejemplos simulados TCDL

Con puntitos representamos los valores que las sucesiones de integrales (las integrales de cada  $f_n$ ) van tomando para cada  $n$  y con líneas el valor aproximado de la integral de la función límite. Como podemos ver la convergencia de la sucesión verde (relacionada con  $f_{3,n}$ ) es distinta al valor de la integral de su función límite  $f_3$ . Esto se debe a que esta sucesión de funciones no cumple con los supuestos necesarios para el TCDL, es decir para “meter” el límite a la integral.

Con la gráfica 21c ya podemos tener más claro que, efectivamente, no podemos “meter” límites a integrales, pues ambos son procesos límites y no es tan trivial intercambiarlos. De hecho, ni las bandas de los intervalos de confianza incluyen a la integral del límite.

Las funciones que “dominan” a  $f_{1,n}$  y  $f_{2,n}$  son:

$$\begin{aligned} g_1 &= x + 1 \\ g_2 &= (x - 5)^2 \end{aligned}$$

El curioso lector puede confirmar que, efectivamente,  $\int |g_i| d\mu < \infty$  para las distintas medidas usadas para  $i = 1, 2$ . También podemos apreciar en esta segunda gráfica que, independientemente de cuál técnica de integración por simulación usemos, obtenemos resultados similares.

#### 4.4. Teorema de Bayes

Bayes (1763) analiza el siguiente experimento mental: se lanza una bola sobre una mesa y se marca la posición de la bola. Después, se lanzan más bolas y en cada lanzamiento registramos si la bola cayó a la izquierda o a la derecha, adelante o atrás, de la primera bola lanzada. Thomas Bayes calculó la probabilidad condicional de la posición de la primera bola dada la secuencia de lanzamientos de las demás bolas y notó que entre más pelotas eran lanzadas, más actualizaba su idea de dónde estaba la primera bola. Aunque no estaba completamente seguro de la localización, con cada nueva pieza de evidencia, podía estar cada vez estaba más seguro de la posición real. Esta es la idea detrás de todo que dio origen a lo que conocemos como Teorema de Bayes.

**Teorema 3** (Teorema de Bayes). *Sea  $\{A_1, A_2, \dots, A_n\}$  un conjunto de sucesos mutuamente excluyentes y exhaustivos, y tales que la probabilidad de cada uno de ellos es distinta de 0. Sea  $B$  un evento cualquiera del que se conocen las probabilidades condicionales,  $P(B|A_i)$ . Entonces, la probabilidad  $P(A_i|B)$  viene dada por la expresión:*

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)}$$

donde

$P(A_i)$  son las probabilidades a priori

$P(B|A_i)$  es la probabilidad de  $B$  en la hipótesis  $A_i$

$P(A_i|B)$  son las probabilidades a posteriori.

Con base en la definición de probabilidad condicionada se obtiene la Fórmula de Bayes, también conocida como Regla de Bayes:

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{k=1}^n P(B|A_k)P(A_k)}$$

La regla de Bayes proporciona un modo natural de actualización de las creencias cuando aparece nueva información. Este proceso de aprendizaje inductivo por medio de este teorema es la base de la Inferencia Bayesiana. La diferencia fundamental entre la estadística clásica (frecuentista) y la bayesiana es el concepto de probabilidad. Para la estadística clásica es un concepto objetivo, que se encuentra en la naturaleza, mientras

que para la estadística bayesiana responde a la siguiente pregunta: ¿qué nos dicen los datos  $X$  acerca del parámetro  $\theta$ ? Ignora toda evidencia externa. Por el contrario, en el caso bayesiano, además de la muestra también juega un papel fundamental la información previa o externa que se posee en relación a los fenómenos que se tratan de modelar. En este caso, la pregunta que se responde es: ¿cómo cambian nuestros juicios originales acerca del valor de la cantidad desconocida  $\theta$  a la luz de los datos  $X$ ?

El objetivo es encontrar la distribución condicional de todas aquellas cantidades de interés cuyo valor desconocemos dado el valor conocido de las variables observadas.

### Enfoque bayesiano

El problema que nos encontramos es el siguiente: tenemos datos  $X$  y queremos conocer el valor de ciertas cantidades  $\theta$  que nos interesan, como parámetros del modelo, mediciones, etc. Como estadísticos, postulamos un modelo de probabilidad:

$$p(x|\theta)$$

Desde el punto de vista bayesiano, además,  $\theta$  debe tener una distribución de probabilidad  $p(\theta)$ , que refleja nuestra incertidumbre inicial acerca del parámetro y se conoce como distribución a priori. Al incluir esta distribución, el análisis Bayesiano usa la probabilidad para representar la incertidumbre en todas las partes de un modelo estadístico.  $X$ , también es conocido, por lo que condicionamos en su valor observado  $x$ . Por lo tanto, nuestro conocimiento acerca del valor de  $\theta$  queda descrito a través de su distribución final,  $p(\theta|x)$ , conocida como distribución a posteriori. El Teorema de Bayes nos dice como encontrarla y los pasos a seguir son los siguientes:

1. Especificación del modelo muestral,  $p(x|\theta)$
2. Especificación de una distribución inicial,  $p(\theta)$
3. Cálculo de la distribución final,  $p(\theta|x)$ , vía el Teorema de Bayes.
4. Resumen de la información contenida en  $p(\theta|x)$  para hacer inferencias sobre las cantidades de interés.

### Ejemplo teórico

Consideremos el primer ejemplo. Supongamos que sales positivo a un test para detectar el COVID-19. El doctor que te realizó la prueba te dijo que de cada 100 personas que tienen la enfermedad, el test puede detectar 99 casos. Ahora, ¿cuál es la probabilidad de que actualmente tengas la enfermedad, dado que el test fue positivo?. El Teorema de Bayes puede ser de gran ayuda. Definamos los siguientes eventos:

$H = \text{Tener COVID} - 19$

$\neg H = \text{No tener COVID} - 19$

$E = \text{Salir positivo en la prueba}$

$\neg E = \text{Salir negativo en la prueba}$

La distribución a priori  $P(H)$  es difícil de determinar, pero en este caso un punto razonable para empezar es la frecuencia de la enfermedad en la población, supongamos

que es 0.001. Por los demás datos del problema,  $P(E|H) = 0.99$ ,  $P(E|\neg H) = .01$ ,  $P(\neg H) = 0.999$ . Entonces tenemos lo siguiente:

$$P(H|E) = \frac{P(E|H)P(H)}{P(E|H)P(H) + P(E|\neg H)P(\neg H)} = \frac{.99(.001)}{.99(.001) + .999(.01)} = 9\%$$

Se obtiene una probabilidad de 9 % de realmente tener la enfermedad después de salir positivo y es un valor muy bajo. Parece magia, pero es sentido común aplicado a matemáticas. La formula de Bayes no está hecha para aplicarse una sola vez, fue creada para ser usada múltiples veces y cada vez ganando evidencia y actualizando la probabilidad de que algo es verdad. El primer ejemplo, cuando sales positivo a una prueba, qué pasaría si fueras al doctor y vuelves a aplicarte la prueba pero ahora por otro laboratorio independiente y ese test también vuelve a ser positivo. Ahora, ¿cuál es la probabilidad de que actualmente tengas la enfermedad? En lugar de usar  $P(H) = 0.001$  como distribución a priori, usamos la distribución posterior obtenida en el ejemplo anterior,  $P(H) = 0.09$ . Calculamos la nueva probabilidad posterior:

$$P(H|E) = \frac{.99(.09)}{.99(.09) + .91(.01)} = 90.73\%$$

La nueva probabilidad basado en dos pruebas positivas, es del 91 %. Hay 91 % de probabilidad de tener la enfermedad, dos diferentes resultados de dos diferentes laboratorios incrementan las probabilidades pero aún no es tan alta como el nivel de precisión del test, que es del 99 %.

El Teorema de Bayes es de gran importancia porque nos permite determinar la probabilidad de las causas a partir de los efectos que han podido ser observados. Esta idea también puede combinarse con métodos de simulación para analizar nuestros datos y resolver problemas, el método de Computación Bayesiana Aproximada es un ejemplo de ello.

### Computación Bayesiana Aproximada (CBA)

Utilizaremos el método conocido como Computación Bayesiana Aproximada, que, aunque es ineficiente computacionalmente, es fácil de entender. El método CBA es un enfoque basado en simulación que no requiere una formulación explícita de la función de verosimilitud  $p(x|\theta)$ . En cambio, este método aproxima la función de verosimilitud mediante simulaciones, cuyos resultados se comparan con los datos observados. CBA utiliza un mecanismo de aceptar-rechazar para realizar el cálculo posterior.

Necesitamos tres cosas para definir nuestro proceso de estimación:

- 1 *Datos*. Es el valor observado, nuestro conjunto de observaciones  $x_1, \dots, x_n$ .
- 2 *Un modelo generador (MG)*. Un modelo generador es cualquier tipo de programa computacional, expresión matemática o conjunto de reglas, que recibe como argumento un conjunto de parámetros fijos y nos devuelve datos simulados. Su función principal es aproximar la función de verosimilitud mediante simulaciones.

3. *Distribución a priori.* ¿Qué información se tiene sobre el modelo, antes de observar los datos?.

Para entender mejor la aplicación de este método resolveremos un ejercicio conocido.

### Problema de captura y recaptura

Un estadístico está interesado en el número  $N$  de peces que hay en un estanque. Él captura 250 peces, los marca y los regresa al estanque. Unos cuantos días después regresa y atrapa peces hasta que obtiene 50 peces marcados, en ese punto también tiene 124 peces no marcados (la muestra total es de 174 peces).

a) ¿Cuál es la estimación de  $N$ ?, ¿de qué manera se puede resolver este problema utilizando Inferencia Bayesiana?

### Solución con CBA

Regresamos a la estructura de un *CBA*. En este problema en concreto los *datos* son la observación obtenida,  $x$  son los *peces no marcados* que se extraen después de obtener 50 peces marcados. El *modelo generador* corresponde al proceso de *marcar y recapturar* que se define más adelante. El parámetro a estimar es  $\theta = \text{número de peces en el lago}$ . En este caso, lo que necesitamos estimar es el número de peces en el lago y sabemos que el número de peces no marcados es igual a 124.

#### *Marcar y recapturar*

Definiremos una función que realice el procedimiento de marcar y recapturar a los peces. Será nuestro *MG* que simularemos y nos ayudará a estimar el número de peces en el lago  $\theta$ .

1. Capturar  $\theta$  peces.
  2. Marcar 250 y los regresamos al lago.
  3. Después de un tiempo, capturar peces hasta que se obtienen 50 peces marcados.
  4. Contar, en los resultados del paso anterior, cuántos peces no están marcados.
- No contamos el número de peces marcados porque este siempre será igual a 50. La observación que nos interesa son los peces no marcados.

#### *Estimación del modelo*

Definimos el siguiente procedimiento para ajustar el modelo:

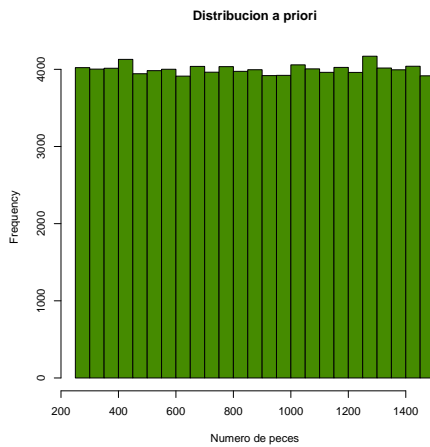
1. Obtener una muestra aleatoria de tamaño  $n$  de la distribución a priori del parámetro  $\theta_1, \dots, \theta_n$ .
2. Aplicar el *MG* a cada observación obtenida de la muestra y así obtener un conjunto de observaciones  $x_1, \dots, x_n$ .
3. Nos quedaremos que las  $\theta_i$  cuyos valores generados sean iguales a los observados. En este caso en particular, nos quedaremos con las  $\theta_i$  que generan 124 peces no marcados.
4. La distribución de aquellos parámetros que cumplen la condición representa la probabilidad de que la observación haya sido producida por cierto valor del parámetro.

A partir de esta distribución podemos hacer estimaciones del valor desconocido.

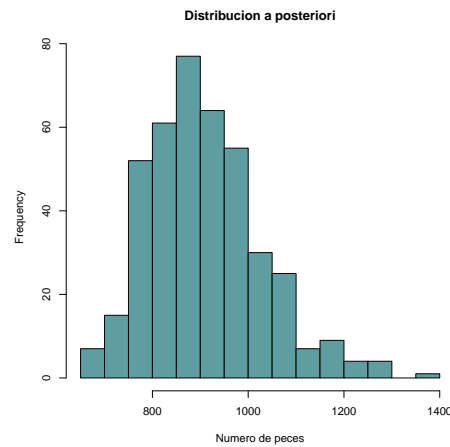
Ahora, presentaremos los resultados utilizando dos distribuciones a priori distintas.

#### *Distribución a priori no informativa*

Supongamos que la distribución a priori de  $\theta \sim Unif\{250, \dots, 1500\}$ . Cuando la distribución a priori es la distribución uniforme la llamamos *distribución a priori no informativa* porque no contiene mucha información sobre el parámetro. Le estamos asignando el mismo peso de probabilidad a todos los valores. A continuación programamos el *MG*, extraemos la muestra de la distribución a priori y graficamos su histograma.



**Figura 22:** Distribución a priori



**Figura 23:** Distribución posterior

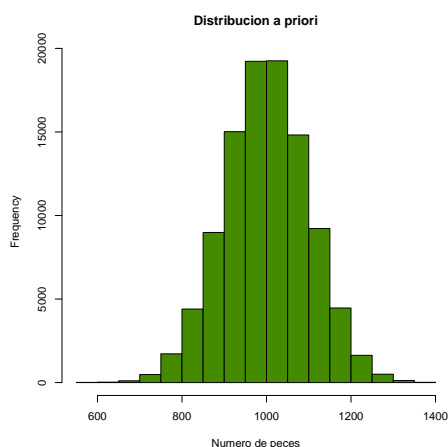
Aunque la distribución es uniforme discreta al principio, cuando observamos la distribución posterior del parámetro es distinto al inicial. No parece que la distribución posterior sea simétrica, está sesgada a la izquierda. La mayoría de las observaciones se encuentran entre 800 y 950. Utilizamos los datos observados después de aplicar el modelo generador de datos para actualizar nuestro conocimiento sobre el parámetro y obtener una mejor aproximación de la distribución del mismo. Por ejemplo, los valores menores a 600 o mayores 1400 nunca simulaban datos iguales a los que nosotros observamos, por eso es que la probabilidad posterior es muy pequeña. La distribución posterior es lo que el modelo sabe sobre el parámetro después de observar los datos. La distribución posterior contiene información del modelo y de los datos. A pesar de obtener como resultado final la distribución del parámetro, nos interesa conocer un valor puntual. Generalmente tomamos el valor que mayor probabilidad tiene de generar los datos (*MLE*) o podemos tomar el valor esperado de la distribución. El número de peces estimados, utilizando el método de *MLE*, es el siguiente: 869 peces

#### *Distribución a priori informativa*

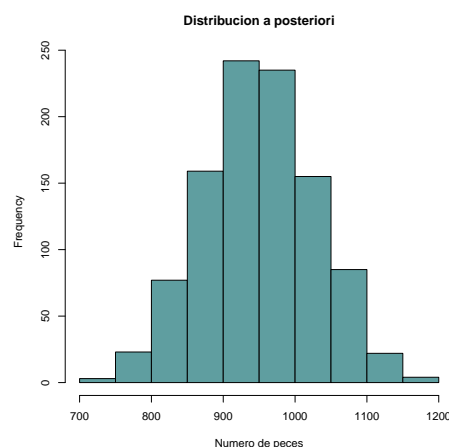
Una ventaja de utilizar el enfoque bayesiano, es que puedes incluir fuentes de información además de los datos, por ejemplo, la opinión de un experto. Supongamos que



el pescador más experimentado del lugar te comenta lo siguiente: "Siempre ha habido muchos peces en el lago. Alrededor de 1000, diría yo". Esta información se incluye en la distribución a priori del parámetro. Cambiamos la distribución de la uniforme a otra distribución más informativa. Elegimos como distribución a priori una distribución normal con media  $\mu = 1,000$  y una desviación estándar de 100.



**Figura 24:** Distribución a priori



**Figura 25:** Distribución posterior

Parece ser que después de usar la información otorgada por el pescador experto y la información proveniente de los datos, es más probable que el número de peces esté entre 900 y 1,000. Además, la distribución a posterior ahora toma una forma simétrica y parecida a la normal, por influencia de la distribución a priori que elegimos en esta ocasión. El número de peces estimados, utilizando el método de *MLE*, es el: 955. Incluir la opinión del experto dentro del análisis hace que las estimaciones cambien y mejoren (suponiendo que el experto no nos engaña).

## 5. Conclusiones

Tras simular algunos escenarios en los que los supuestos del TLC no se satisfacen y analizar los resultados obtenidos, fuimos capaces de reconocer la importancia de verificar el cumplimiento de los mismos antes de intentar aplicar el teorema, lo cual puede extenderse para cualquier otro resultado estadístico o matemático en general. Además, una implicación interesante para este curso en particular es que dado que las simulaciones de Monte Carlo están motivadas por el TLC, debemos tener cuidado al intentar generar simulaciones de Monte Carlo cuando trabajamos con distribuciones de cola ancha, como la Cauchy o la distribución *t* de Student con dos o menos grados de libertad.

En cuanto a los intervalos de confianza, es importante investigar acerca de la cobertura de los mismos pues vemos que, al menos en la distribución binomial, a pesar de que se busque un intervalo de confianza del  $100(1 - \alpha)\%$ , dependiendo del tamaño de

muestra  $n$  y valor del parámetro  $\theta$  que se elija, realmente no se estará obteniendo un intervalo del  $\alpha$  deseada.

Cuando las personas hablan del Teorema de Bayes, llegan a discutir lo contraintuitivo que es, pero en realidad es lo contrario: somos buenos entendiendo la idea. Cuando se introduce este resultado a la comunidad científica, hace una analogía con un hombre que ha estado toda su vida en una cueva y de pronto sale y ve el sol salir por primera vez. Se pregunta si el sol siempre hace eso o si solo es un capricho. Después cada día que el sol sale, el hombre tiene un poco más de información y está más confiado de que esa es la forma en que el mundo funciona. Por lo tanto, la regla de Bayes proporciona un modo natural de actualización de las creencias cuando aparece nueva información.

Usualmente teoría de la medida es considerada como una rama difícil y abstracta de las matemáticas. No es común hacer visualizaciones de las diferentes convergencias de las que se habla en estas clases. A pesar de esto, sí se pueden hacer visualizaciones de ejemplos usando simulación. En varias ramas de matemática moderna (topología, optimización numérica, etc.) los investigadores usan mucho la simulación para dar un contexto y poder derivar propiedades importantes en sus respectivos campos de estudio. Por esto decimos que la simulación no solo es importante en el contexto práctico, sino también en el académico.

## 6. Referencias

- [1] Conchi Ausín. *Introducción a la Inferencia Bayesiana*. URL: <http://halweb.uc3m.es/esp/Personal/personas/causin/esp/2012-2013/SMB/Tema6.pdf>. (accesado: 29.11.2020).
- [2] L.D. Brown, T. Cai y A. Dasgupta. «Interval Estimation for a Binomial Proportion». En: *Statistical Science* (2001). URL: <https://projecteuclid.org/euclid.ss/1009213286>.
- [3] C. Dalitz. «Construction of Confidence Intervals». En: *Technical Report No. 2017-01* (2017). URL: <https://arxiv.org/pdf/1807.03582.pdf>.
- [4] Guillermo Grabinsky. *Teoría de la medida*. Las prensas de ciencias, 2000.
- [5] Clara Grazian y Yanan Fan. «A review of Approximate Bayesian Computation methods via density estimation: inference for simulator-models». En: *School of Mathematics and Statistics, UNSW Sydney, Kensington 2052, Australia* (2019). URL: <https://arxiv.org/pdf/1909.02736.pdf>.
- [6] Eduardo Gutiérrez Peña. *Estadística bayesiana: Teoría y Conceptos Básicos*. URL: <http://www.dpye.iimas.unam.mx/soriano/BAYES/DOCUMENTOS/NOTAS/INFERENCIA%20BAYESIANA.pdf>. (accesado: 29.11.2020).
- [7] Catherine A. Doléans-Dade Robert B. Ash. *Probability and Measure Theory*. Academic Press, 1999.

- [8] Brandon Rohrer. *How Bayes Theorem works*. URL: <https://www.youtube.com/watch?v=5NMxi0GL39M>. (accesado: 27.11.2020).
- [9] Terence Tao. *An Introduction to Measure Theory*. American Mathematical Society, 2011.
- [10] R. Wicklin. *Coverage probability of confidence intervals: A simulation approach*. URL: <https://blogs.sas.com/content/iml/2016/09/08/coverage-probability-confidence-intervals.html>. (accesado: 07.11.2020).

## **7. Anexos**

### ANEXO 1: Propuesta de trabajo

## 1. Motivación del tema

Como proyecto final decidimos aplicar la simulación a ciertos conceptos de probabilidad y estadística. Si bien los conceptos con los que vamos a trabajar ya fueron estudiados anteriormente, consideramos que con las herramientas obtenidas en la clase de simulación, podemos comprobar estos resultados de una manera distinta. A pesar de que en semestres anteriores estos resultados ya los demostramos, creemos que mediante la simulación de estos procesos y de ejemplos concretos, reforzaremos su aprendizaje y facilitaremos su comprensión.

## 2. Alcance del proyecto

El propósito de nuestro proyecto es comprobar e ilustrar algunos resultados vistos en cursos pasados y que, mediante la simulación de sus respectivos modelos, logremos un mejor entendimiento y comprensión de los conceptos estudiados. Consideramos que al nosotros generar los modelos necesarios para cada uno de los conceptos presentados, con nuestro proyecto alcanzaremos a ejemplificar los resultados de estos teoremas y demás nociones mostradas.

## 3. Resultados esperados

Esperamos reforzar el conocimiento de los cursos anteriores de teoría de la medida y la rama de estadística y probabilidad mediante la generación de diferentes modelos que simulen los resultados de teoremas y conceptos vistos en esos cursos. A partir de la generación de estos modelos, buscamos analizar la necesidad de las hipótesis presentadas así como la posibilidad de modificar algunos de los supuestos dados. De igual forma, pretendemos mostrar resultados que dependan de diferentes valores de parámetros y evaluar su efecto en los conceptos.

## 4. Actividades a realizar

Los resultados que queremos presentar son los del Teorema del Límite Central, teorema de Bayes y teorema de la convergencia dominada de Arzelà. Para cada uno de estos teoremas queremos enunciarlos, hacer hincapié en la importancia de sus supuestos, utilizar diferentes tamaños y distribuciones de muestra para el modelado que produzcan resultados visuales y, por último, mostrar algunas aplicaciones de los mismos. Con el fin de resaltar la importancia de los supuestos queremos exhibir, en cada caso, qué sucede cuando estos no se cumplen.

## 5. Metodología

- Ilustrar el **teorema de la convergencia dominada de Arzelà**, el cual puede ser visto como un corolario (para el caso de la integral de Riemann) del teorema de la convergencia dominada de Lebesgue. Dicho teorema es uno de los más importantes dentro de la teoría de la medida. El propósito es dar unos ejemplos donde se cumpla el teorema. Dado que el teorema involucra integrales, éstas se van a calcular con los métodos vistos en clase (en particular los métodos de Monte Carlo junto con métodos de reducción de varianza).
- Realizar una breve introducción del análisis bayesiano de datos. Se dará un ejemplo teórico simple para entender la alta relevancia del **teorema de Bayes**, puesto que tiene vinculación íntima con la comprensión de la probabilidad de aspectos causales dados los efectos observados. A través de simulación y el método de análisis basado en este teorema, se pretende resolver el problema de “captura y recaptura”, en el que un estadístico está interesado en el número de peces que hay en un estanque.
- Mostrar cómo, dadas 4 muestras aleatorias de diferentes distribuciones (tentativamente exponencial, uniforme, binomial y Poisson) con diferentes tamaños de muestra, el **Teorema del Límite Central** se cumple. Tras enunciar este teorema y comparar los resultados dependiendo de la muestra que se utilice, se expondrán tres situaciones de la vida real donde este teorema está presente.
- Finalmente, y como complemento de los resultados obtenidos en los puntos anteriores, planeamos simular qué pasaría si alguno de los **supuestos** establecidos en sus respectivos teoremas **no se satisface**. Pues en estos casos, la simulación parecería ser nuestra mejor aliada para tratar de reconocer el comportamiento aberrante frente al que nos podamos encontrar.