DEPARTAMENTO DE COMPUTACION

Facultad de Ciencias Exactas y Naturales - UBA

Trabajo Práctico I

Ley de los Grandes Números y Teorema Central del Límite

Probabilidad y Estadística

Segundo Cuatrimestre de 2017

Integrante	LU	Correo electrónico
Parral, Guillermo Eduardo	280/16	guillermoeparral@gmail.com
Regert, Fernando	282/15	fernandostds9@gmail.com
Sassone, Federico Sebastián	602/13	fede.sassone@hotmail.com

Facultad de Ciencias Exactas y Naturales

Universidad de Buenos Aires

Ciudad Universitaria - (Pabellón I/Planta Baja)

Intendente Güiraldes 2160 - C1428EGA

Ciudad Autónoma de Buenos Aires - Rep. Argentina

Tel/Fax: (54 11) 4576-3359

http://www.fcen.uba.ar

Índice

1. Introducción

La idea del siguiente trabajo es realizar un conjunto de experimentos con variables aleatorias variando la cantidad de muestras. Estos se harán en dos tandas, una para variables con una distribución exponencial y otra para variables con distribución binomial. Se busca luego poder graficar esos resultados, lograr analizarlos y encontrar en los mismos justificaciones a partir de la aplicación de la Ley de los grandes números y el Teorema central del límite.

Debajo, las definiciones de los mismos.

1.1. Ley de los grandes números

Si X_1, X_2, X_3, ... es una sucesión infinita de variables aleatorias independientes e idénticamente distribuidas que cumplen $E(|X_i|) < \infty$ y tienen el valor esperado μ, entonces

$$P(\lim_{n\to\infty} \overline{X_n} = \mu) = 1$$

es decir, el promedio de las variables aleatorias converge a μ casi seguramente. En otras palabras, el valor esperado de una variable aleatoria como es el promedio a largo plazo al hacer un muestreo repetitivo.

1.2. Teorema central del límite

Sea X_1, X_2, ..., X_n un conjunto de variables aleatorias, independientes e idénticamente distribuidas con media μ y varianza $0 < \sigma^2 < \infty$. Sea

$$S_n = X_1 + ... + X_n$$

Entonces

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

Es decir, que sin importar la distribución de X_i y con solo la existencia de media y varianza, se garantiza una distribución normal cuando n es suficientemente grande.

A continuación, los distintos ejercicios para ambas distribuciones.

2. Distribución Exponencial(λ)

2.1. Ejercicio 1

Se generaron 3000 observaciones de variables aleatorias e independientes con distribución $E(\lambda)$ y obtuvimos $\overline{X_n}$ para cada una de ellas, a continuación los gráficos (n) vs. $\overline{X_n}$ fijando semilla de R dentro y fuera del loop.

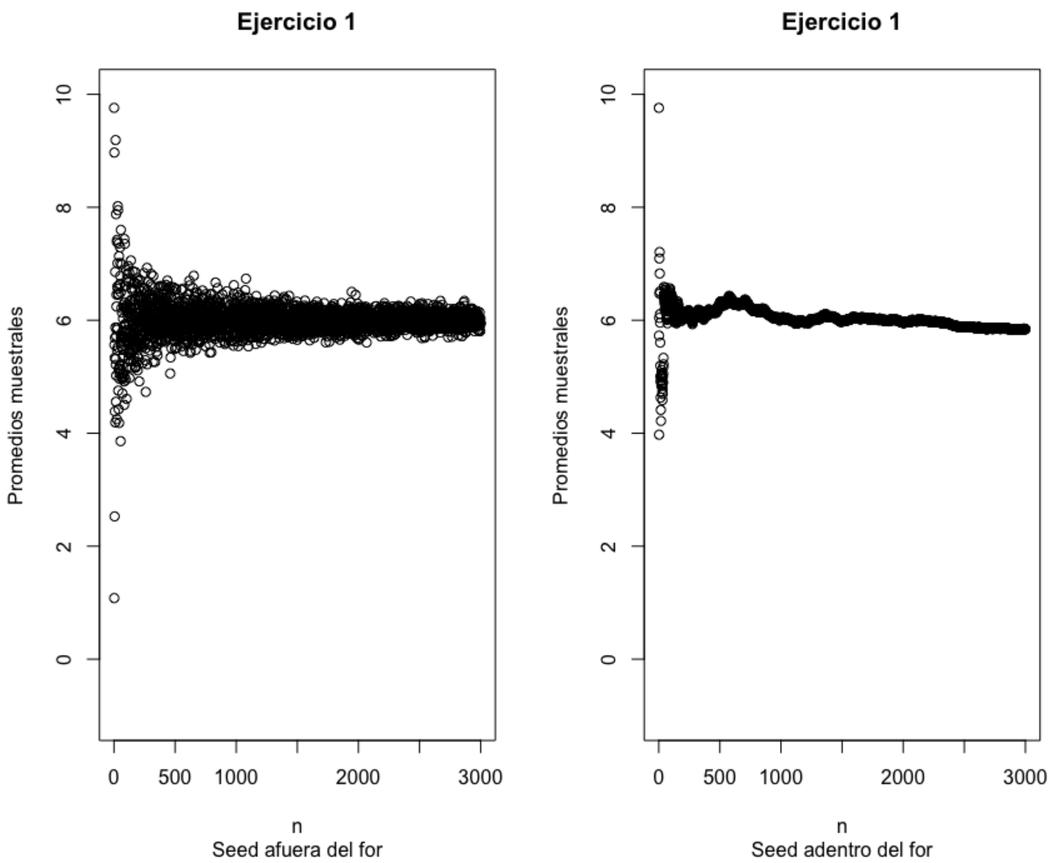

Figura 1: Comparación de promedios muestrales incrementando número de observaciones para una distribución exponencial. Semilla fuera del for vs dentro.

Recordemos el valor esperado de una Exponencial: Sea $X \sim E(\lambda)$ luego, $E[X] = \frac{1}{\lambda}$.

Nuestro valor de λ fue de $\frac{1}{6}$.

En los gráficos podemos apreciar que cuando la cantidad de muestras se acerca a 3000, los valores de los promedios muestrales se acercan a 6. Tiene sentido ya que la Ley de los grandes números enuncia justamente esto, para un n lo suficientemente grande, el promedio de las muestra se acerca a el valor esperado de la distribución.

Con respecto a los cambios en el gráfico según donde se setee la semilla: Al estar fuera del for, se genera más aleatoriedad a la muestra; mientras que al estar dentro del mismo, cada muestra se calcula con el mismo punto de partida agregando elementos. Esto puede apreciarse en una distribución mucho más uniforme de los resultados en el gráfico de la derecha. Es decir cuando se aplica la semilla dentro del for los promedios muestrales distan menos de la esperanza en general.

2.2. Ejercicio 2

Empezando por 2 observaciones, valor que iremos aumentando, se realizaron observaciones de variables aleatorias independientes con distribución E(λ), se les calculo el promedio y se repitio este proceso 1000 veces , para cada cantidad de observaciones se realizaron Histogramas, Boxplots y QQ-plots donde podremos observar la densidad de los valores que toman los promedios obtenidos. A continuación, los resultados.

2.2.1. Histogramas

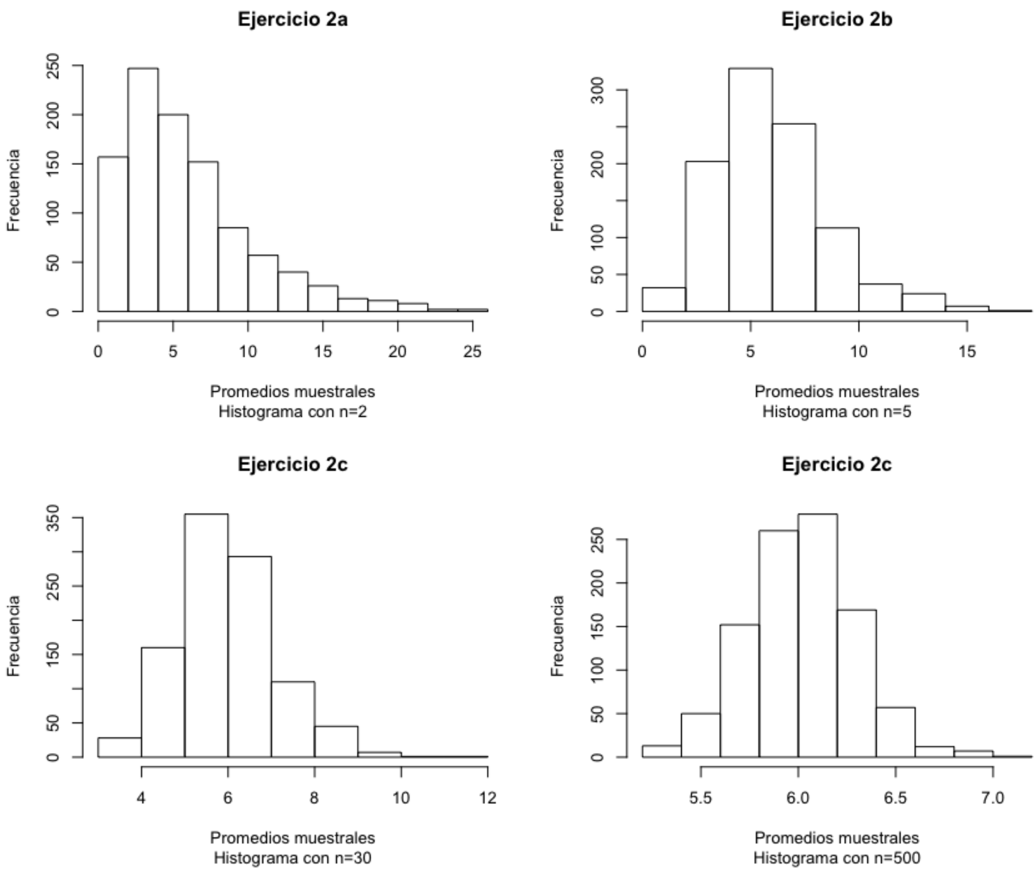

Figura 2: Comparación de Histogramas de distribuciones exponenciales incrementando el número de muestras.

Al aumentar el número de muestras podemos ver que la forma del histograma se asemeja cada vez más a la curva de una normal. De hecho, se puede observar que el punto más alto del histograma se alcanza en 6, aproximadamente, lo cual es similar al caso de una variable aleatoria con distribución normal y media 6.

2.2.2. BoxPlots

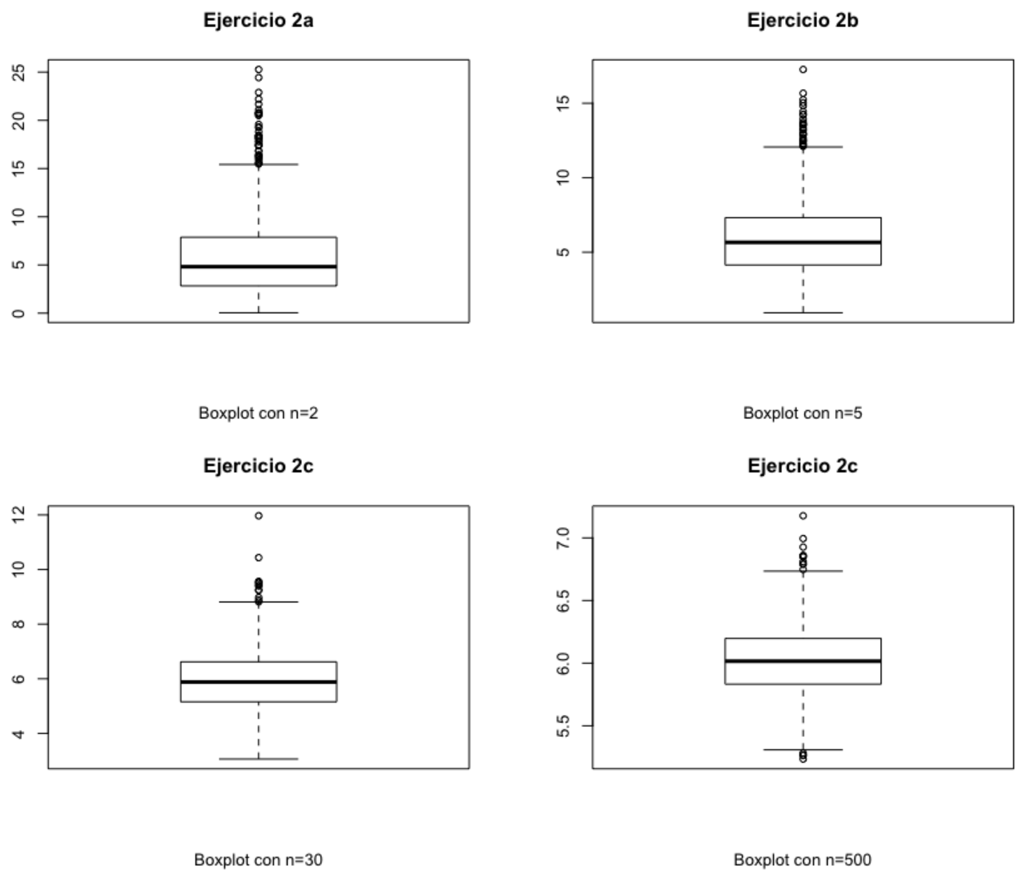

Figura 3: Comparación de Boxplots de distribuciones exponenciales incrementando el número de muestras.

Al aumentar el número de muestras, los diagramas de caja se vuelven más simétricos respecto a su mediana y su forma se asemeja cada vez mas al diagrama de cajas de una Distribución Normal. Esto se puede ver en los tamaños de los cuartiles. Al ser cada vez más simétricos el segundo y el tercero y contener la mayor cantidad de muestras, se asemejan a la campana de una Normal. El primer y cuarto cuaril representan las colas de la Normal.

2.2.3. QQ-Plots

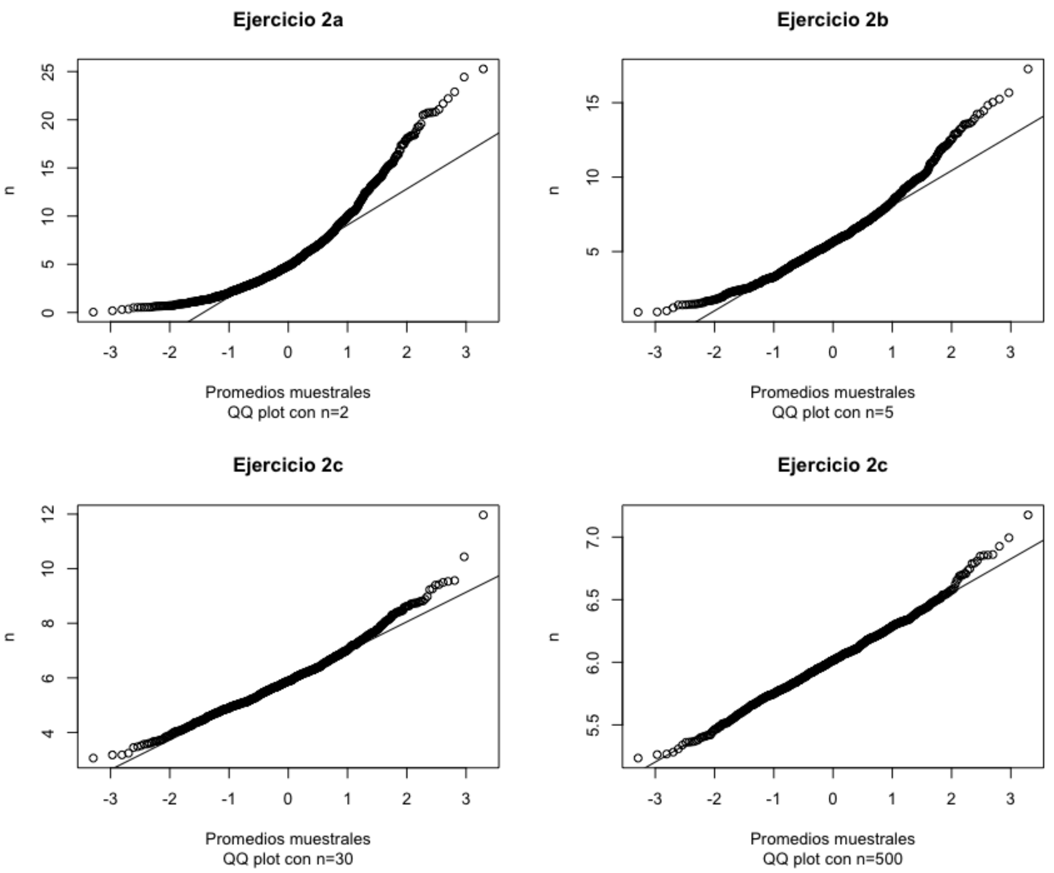

Figura 4: Comparacion de QQ-plots de distribuciones exponenciales incrementando el número de muestras.

Nuevamente, al aumentar el número de observaciones, la representación se acerca más a la de una Distribución Normal porque, como se ve en el gráfico, la forma que toman los promedios muestrales se asemejan a la recta que representa a la distribución normal. Se puede ver que para un menor número de muestras, el QQ-plot tiene una cola más liviana a izquierda y más pesada a derecha, lo cual tiene sentido si se mira la curva de una exponencial: la mayor cantidad de "peso" recae en el lado izquierdo mientras que en la normal recae en el centro. En los gráficos con menos cantidad de observaciones el gráfico se asemeja a una Distribucion Exponencial.

2.2.4. Comparación de Boxplots en paralelo

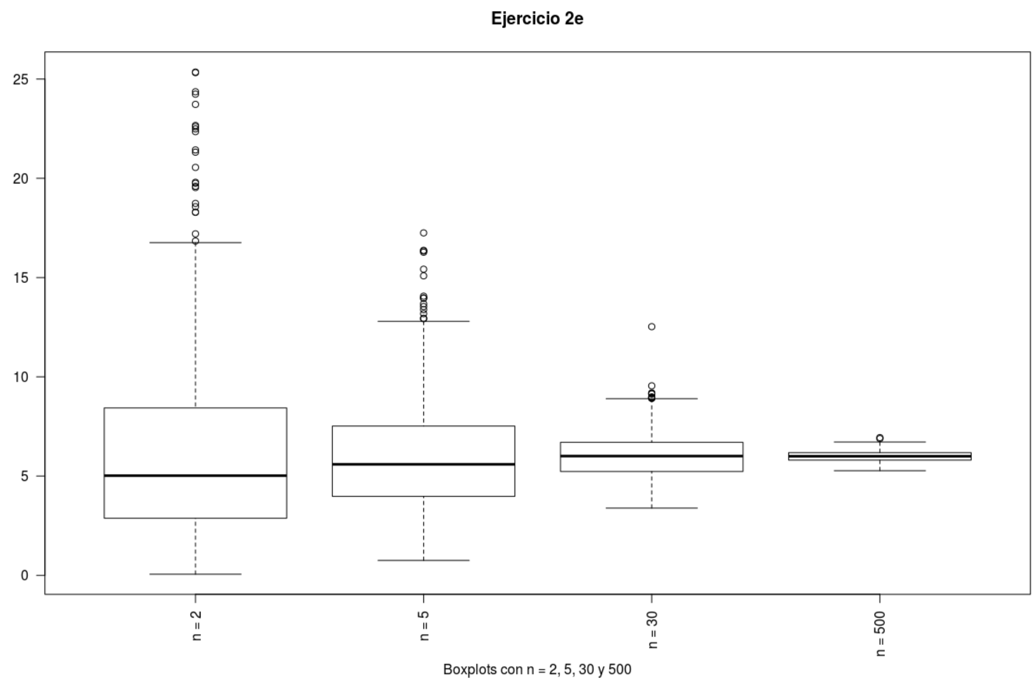

Figura 5: Comparacion de BoxPlots en paralelo de distribuciones exponenciales incrementando el número de muestras.

Viendo los gráficos de Boxplots en paralelo podemos apreciar que al aumentar el número de observaciones, los promedios muestrales se encuentran muchísimo más cerca del 6 y los outliers se alejan muy poco del mismo. El segundo y tercer cuartil se vuelven cada vez más simétricos y se asemejan a la representación de una distribución Normal. Todo esto se condice, nuevamente, con lo planteado en el Ejercicio 1. La ley de los grandes números nos indica que la mayor cantidad de promedios se concentran en en valor esperado de la distribución.

2.2.5. Conclusión

Nuevamente, veamos la definición de la Ley de los Grandes Números:

$$P(\lim_{n\to\infty} \overline{X_n} = \mu) = 1$$

Es decir, el promedio muestral tiende a la esperanza de la variable aleatoria cuando N tiende a infinito. Esto se puede apreciar en todos los experimentos ya que el promedio muestral tiende a 6 (que es la esperanza teorica de la distribución $E(\lambda)$ dada en la consigna).

2.3. Ejercicio 3

Repasemos nuevamente la esperanza y varianza de nuestra distribución:

$X \sim E(\lambda)$, luego $E[X] = \frac{1}{\lambda}$ y $Var[X] = (\frac{1}{\lambda^2})$

Nuestro parámetro $\lambda = 1/6$. Luego $E[X] = 6$ y $Var[X] = 36$.

Los valores del promedio son los mismos que los obtenidos en el Ejercicio 2.

En particular, para los siguientes experimentos utilizaremos la definición del teorema central del limite que nos dice que aplicando a los promedios la transformacion:

$$\frac{\overline{x_n} - E(X_1)}{\sqrt{\frac{Var(X_1)}{n}}}$$

La distribución de la muestra se aproxima a la distribucion normal estándar si el n es suficientemente grande.

Entonces, para los siguientes experimentos, aplicaremos la transformación a los valores de los promedios e incrementaremos el numero de muestras. Con los datos obtenidos graficamos histogramas, diagrama de caja y QQ-Plots comparando n con la variable resultante una vez que se aplica la transformación.

2.3.1. Boxlots en paralelo

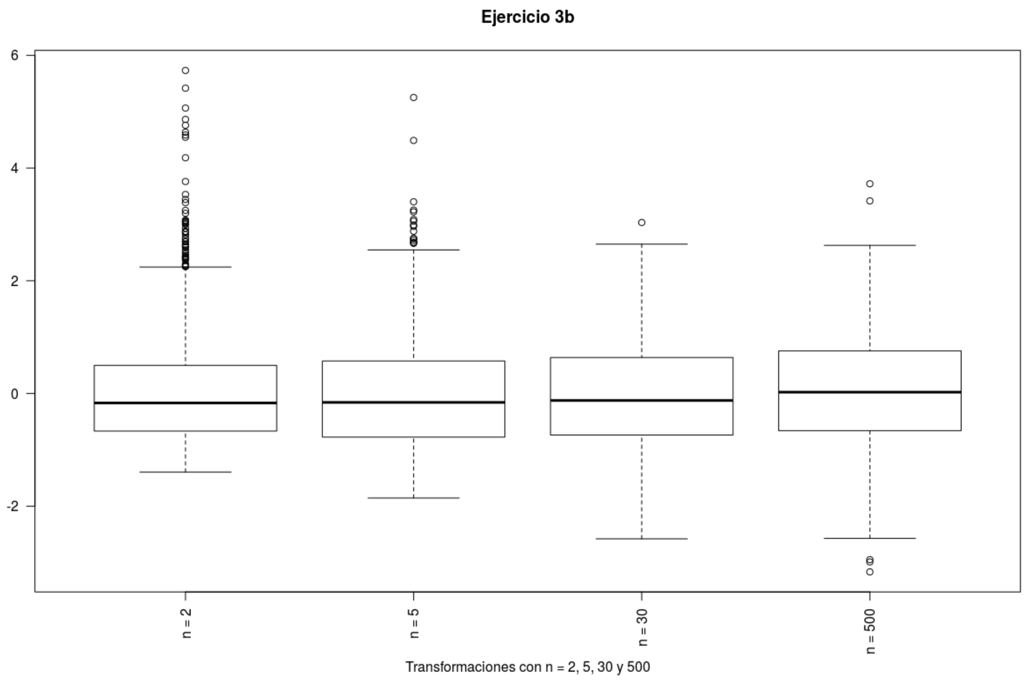

Figura 6: Boxplots en paralelo de las transformaciones aplicadas a los conjuntos de datos del Ejercicio 2

Como podemos ver en el gráfico anterior, a medida que la muestra se incrementa, el boxplot representa más y más al de una Normal centrada en 0. Se puede ver que los cuartiles dos y tres se van volviendo más simétricos y los outliers disminuyen. También los cuartiles 1 y 4, que representan a ambas colas en una distribución normal, ganan simetría.

2.3.2. QQ-Plots

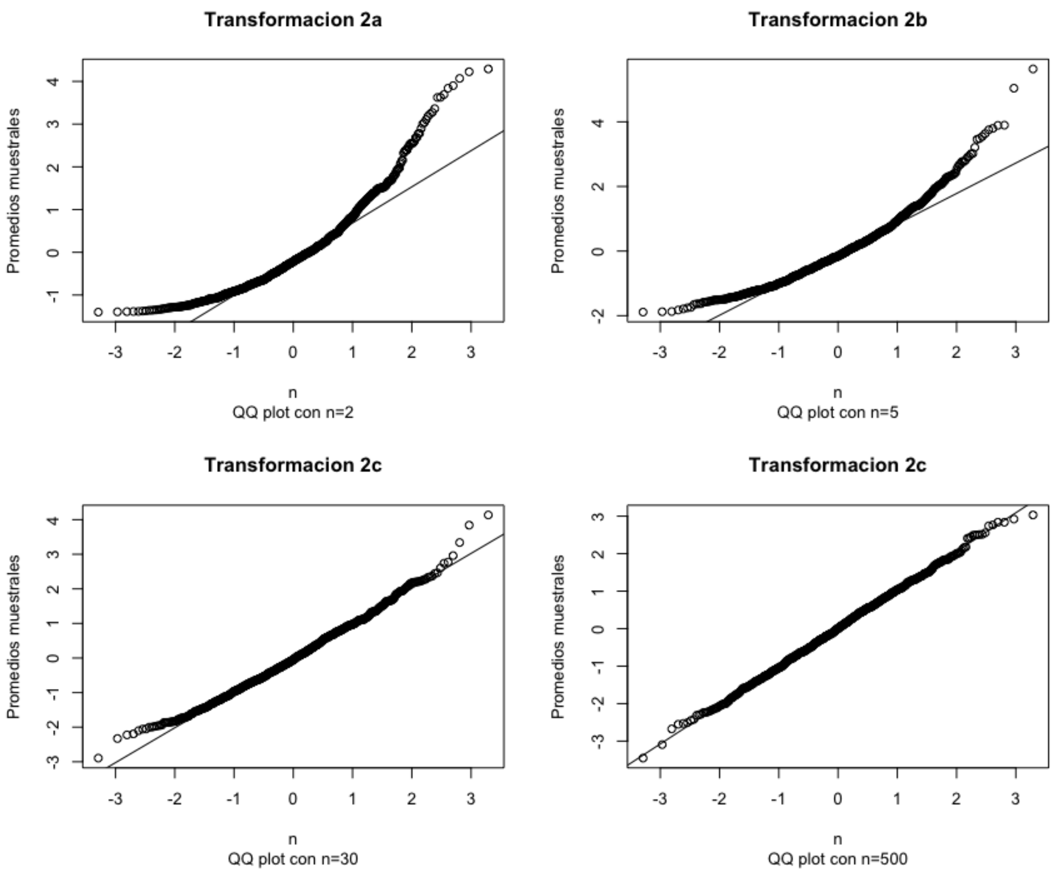

Figura 7: QQ-Plots de las transformaciones aplicadas a los conjuntos de datos del Ejercicio 2

Nuevamente, podemos observar como al aplicar la transformacion y aumentar el número de muestras, la distribución se asemeja más a una normal (representada por la línea diagonal en todos los gráficos).

2.3.3. Histogramas

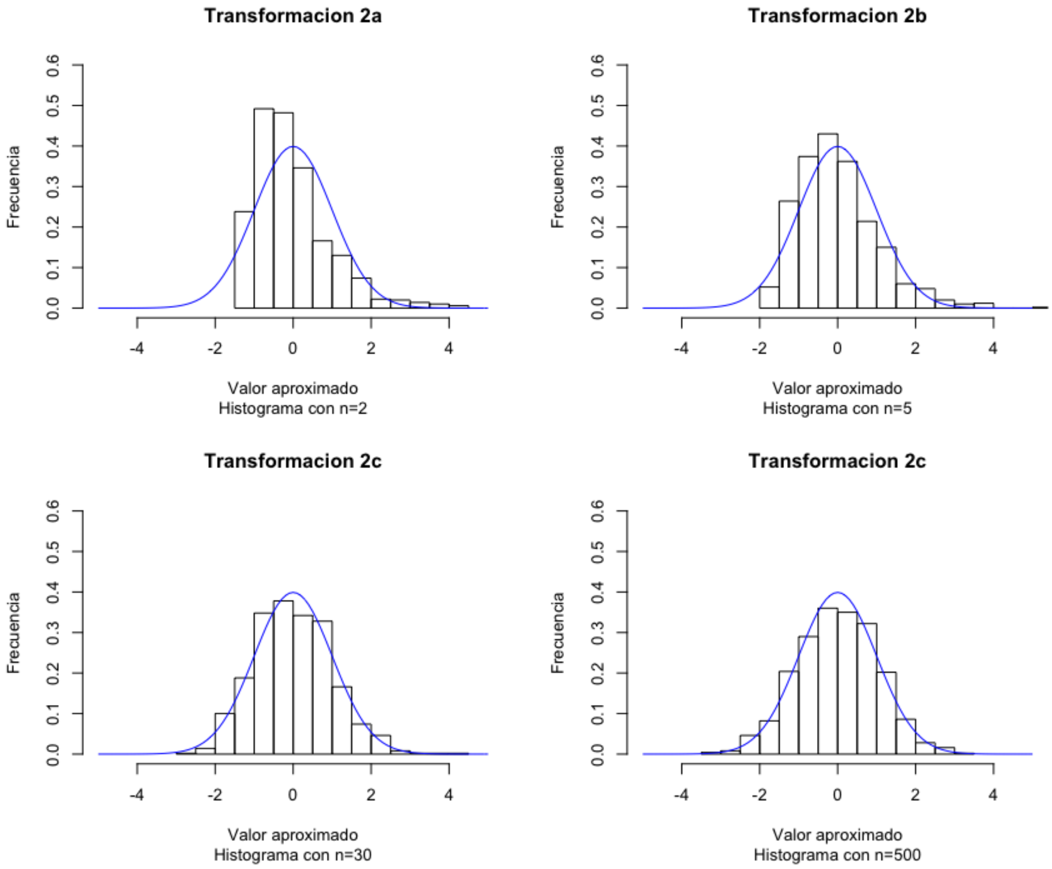

Figura 8: Comparación de Histogramas aumentando los tamaños de las muestras.

La línea azul en los gráficos representa la funcion densidad de una distribucion normal estándar. Se puede observar como a medida que aumenta el número de muestras los histogramas se asemejan cada vez mas a dicha función ademas de centrarse en el 0 como ocurriria en una distribucion normal estandar.

2.3.4. Conclusión

Se les aplicó la transformación a los conjuntos de datos del Ejercicio 2 y se comprobó mediante la comparación de los distintos gráficos que las distribuciones se asemejan a una Normal cuando la cantidad de muestras aumenta. La justificación a este comportamiento está en el Teorema central del límite. Independientemente de la distribución de nuestras variables, se garantiza una distribución normal con n lo suficientemente grande.

3. Distribución Binomial(n, p)

3.1. Ejercicio 1

Se generaron 3000 observaciones de variables aleatorias e independientes con distribución $Bi(n, p)$ y obtuvimos $\overline{X_n}$ para cada una de ellas, a continuación los gráficos (n) vs. $\overline{X_n}$ fijando semilla de R dentro y fuera del loop.

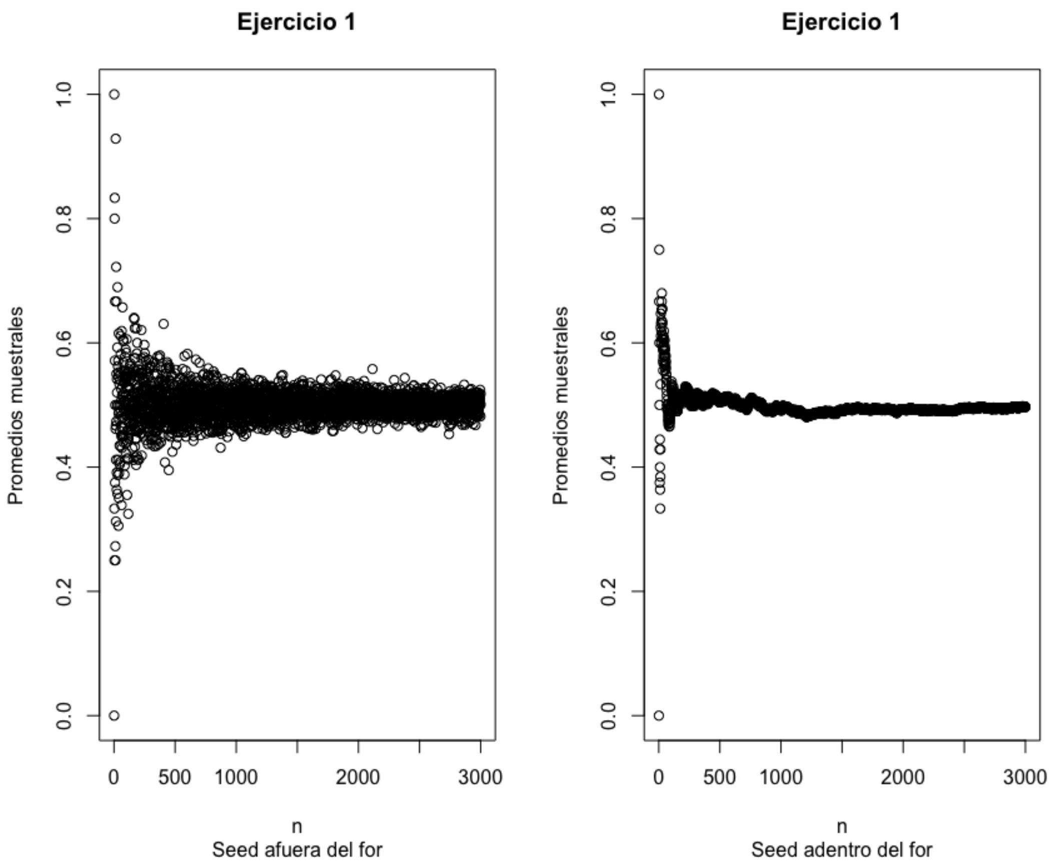

Figura 9: Comparación de promedios muestrales incrementando número de observaciones para una distribución binomial. Semilla fuera del for vs dentro.

Recordemos el valor esperado de una binomial: $X \sim Bi(n, p)$, luego $E(X) = n * p$.

En nuestro caso, $n = 5$ y $p = 1$.

Luego, por la definición de la ley de los grandes números, cuando $n \to \infty$, el promedio de las variables tiende a la Esperanza de la distribucion, es decir, $5 * 0, 1 = 0{,}5$ en nuestro caso.

Esto se ve claramente en los gráficos, donde para un número pequeño de muestras el promedio varía bastante mientras que al crecer las muestras el mismo converge a $0{,}5$.

Respecto a los cambios en los gráficos según dónde se setee la semilla de R, la justificación es la misma que en el Ejercicio 1 de la distribución exponencial: Al estar fuera del for, se genera más aleatoriedad a la muestra; mientras que al estar dentro del mismo, cada muestra se calcula con el mismo punto de partida agregando elementos. Esto puede apreciarse en una distribución mucho más uniforme de los resultados en el gráfico de la derecha. Es decir, cuando se aplica la semilla dentro del for los promedios en general distan menos del valor de la esperanza.

3.2. Ejercicio 2

Empezando por 2 observaciones, valor que iremos aumentando, se realizaron observaciones de variables aleatorias independientes con distribución Binomial(n, p), se les calculo el promedio y se repitio este proceso 1000 veces , para cada cantidad de observaciones se realizaron Histogramas, Boxplots y QQ-plots donde podremos observar la densidad de los valores que toman los promedios obtenidos. A continuación, los resultados.

3.2.1. Histogramas

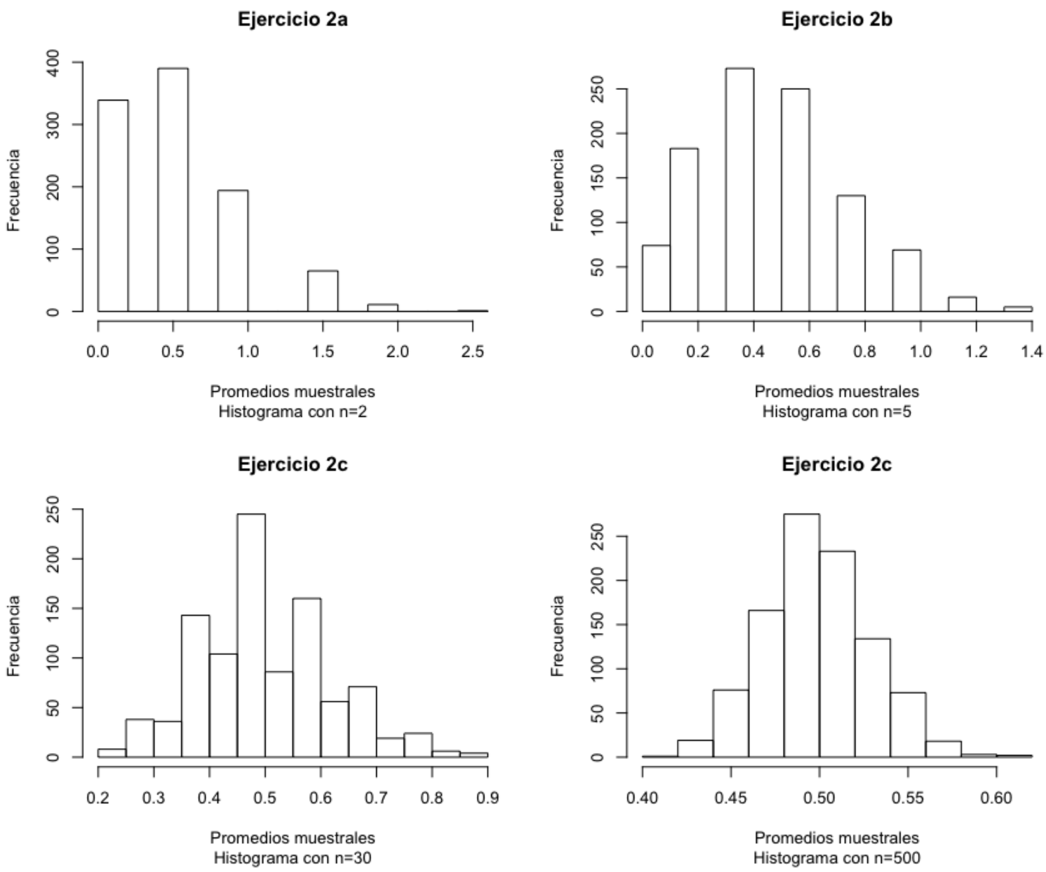

Figura 10: Histogramas de distribuciones Binomiales aumentando el número de muestras.

A medida que aumentamos la cantidad de muestras, los resultados son menos dispersos. En 2a hay outliers hasta en 2.5 al igual que en 2b, llegando a 1.4. A partir de las 30 muestras la distancia se reduce y la mayor cantidad de resultados se acerca al valor esperado. En 500 muestras todos los resultados se encuentan a $|0{,}1|$ de distancia del valor esperado: 0.5 $= n * p$. La justificación proviene de la Ley de los grandes números: para un n lo suficientemente grande, el promedio de las muestra se acerca a el valor esperado de la distribución.

3.2.2. QQ-Plots

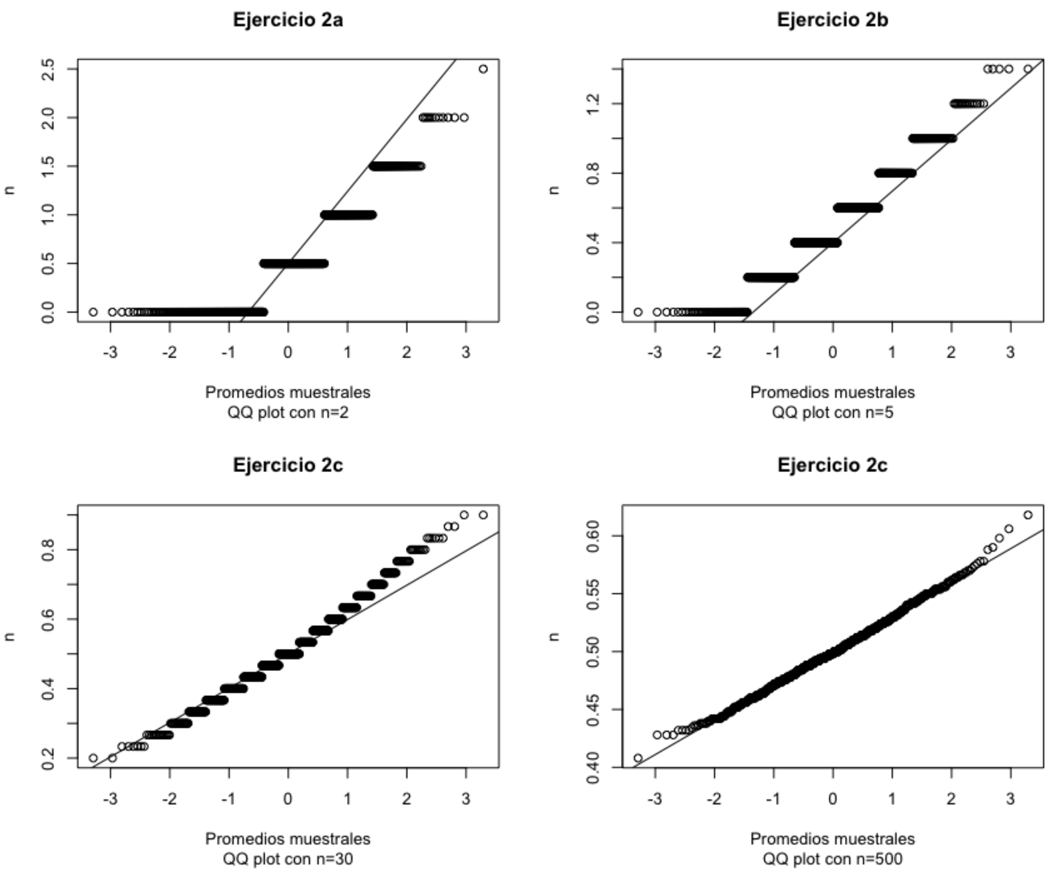

Figura 11: QQ-Plots de distribuciones Binomiales aumentando el número de muestras.

Podemos ver como los resultados comienzan poco centrados en el promedio esperado y al llegar a las 500 repeticiones, lo hacen. Adicionalmente, la forma del QQ-plot se asemeja a la de una distribución normal, (establecida como guía). Llama la atención los centrados que están los resultados en ciertos valores de n para muestras más pequeñas, mientras que al aumentar la cantidad de las mismas, los resultados se ven mejor distribuídos. Creemos que esto se debe a que la distribución binomial es discreta y la normal continua, luego los valores empiezan a dispersarse más a medida que aumenta n, asemejándose más y más a una distribución normal.

3.2.3. BoxPlots

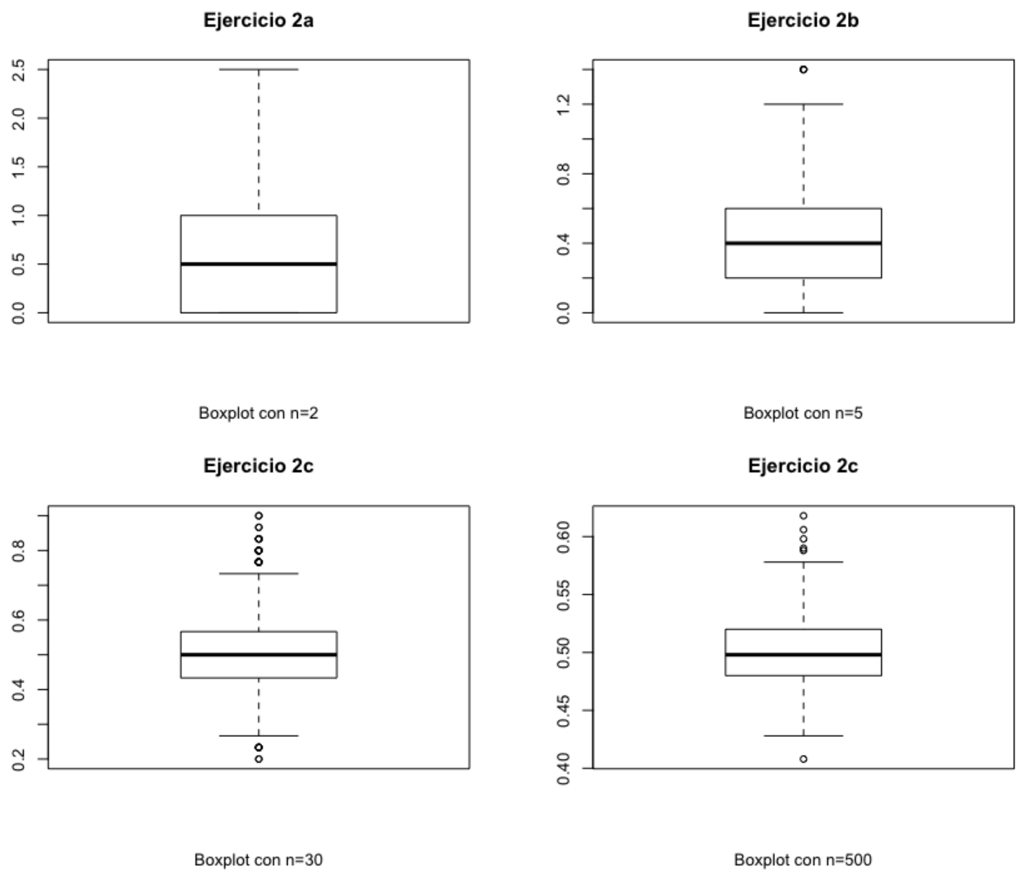

Figura 12: QQ-Plots de distribuciones Binomiales aumentando el número de muestras.

Otra forma gráfica de ver la conclusión anterior. En 2a el cuarto cuartil ocupa más de la mitad del gráfico. A medida que aumenta la cantidad de muestras, el boxplot se vuelve más simétrico, hasta que finalmente en el caso de $n = 500$ está centrado en el valor esperado.

3.2.4. BoxPlots paralelos

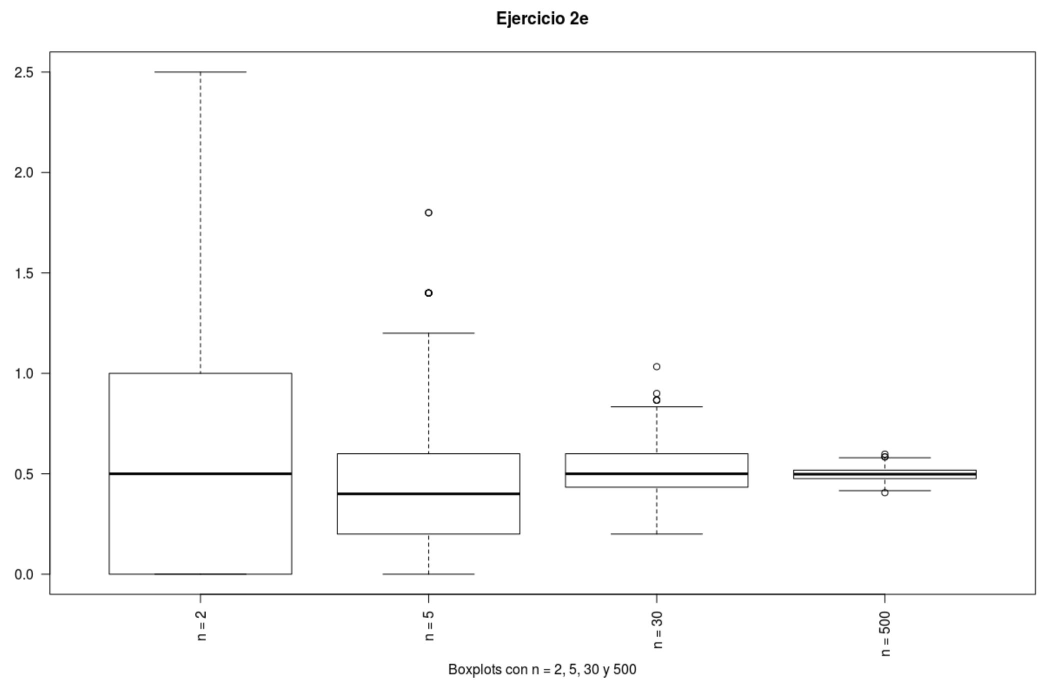

Figura 13: QQ-Plots de distribuciones Binomiales aumentando el número de muestras.

Al ver los boxplots en paralelo se hace todavía más evidente como los datos se concentran en el promedio esperado según la ley de los grandes números y como la forma del gráfico se acerca cada vez más a la de una distribución normal, teniendo sus valores centrados y poco distribuídos.

3.2.5. Conclusión

Recordemos el valor esperado de nuestra distribución:

Como $X \sim Bi(n, p)$, luego $E[X] = np$, como nuestros valores asignados eran $n = 5$, $p = 0{,}1$, nuestro valor esperado era $0{,}5$.

Se pueden apreciar las definiciones de la ley de los grandes números y el teorema central del límite en los experimentos previos ya que en todos ellos, el promedio muestral tendió a la esperanza de la variable aleatoria cuando $n \to \infty$. Además, las formas de los gráficos se asemejó cada vez más a la de una distribución normal.

3.3. Ejercicio 3

Repasemos nuevamente la esperanza y varianza de nuestra distribución:

$X \sim Bi(n,p)$, luego $E[X] = np$ y $Var[X] = np(1-p)$

Nuestro parámetro $n = 5$ y $p = 0{,}1$. Luego $E[X] = 0{,}5$ y $Var[X] = 0{,}45$.

Los valores del promedio son los mismos que los obtenidos en el Ejercicio 2.

En particular, para los siguientes experimentos utilizaremos la definición del teorema central del limite que nos dice que aplicando a los promedios la transformacion:

$$\frac{\overline{x_n} - E(X_1)}{\sqrt{\frac{Var(X_1)}{n}}}$$

La distribución de la muestra se aproxima a la distribucion normal estándar si el n es suficientemente grande.

Entonces, para los siguientes experimentos, aplicaremos la transformación a los valores de los promedios e incrementaremos el numero de muestras. Con los datos obtenidos graficamos histogramas, diagrama de caja y QQ-Plots comparando n con la variable resultante una vez que se le aplica la transformación.

3.3.1. Boxplots en paralelo

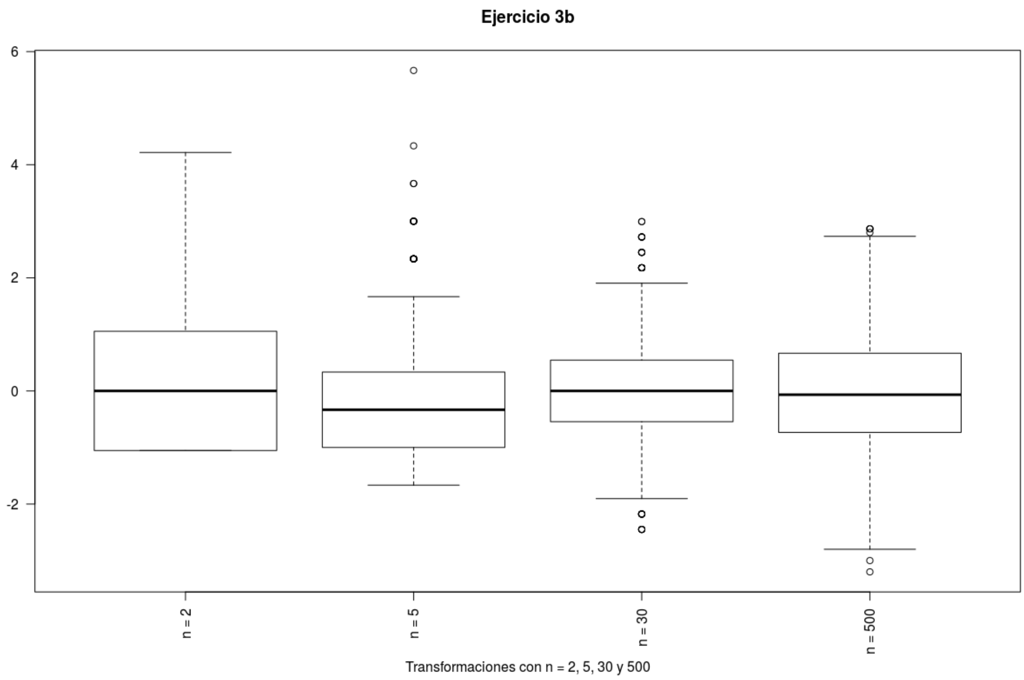

Figura 14: Boxplots en paralelo de transformaciones de distribuciones Binomiales aumentando el número de muestras.

La transformación envía correctamente los valores a un muestreo centrado en 0. A medida que n incrementa, el boxplot representa más y más al de una Normal centrada en 0. Se puede ver que el primer cuartil incrementa a medida que crece la muestra y los outliers empiezan a distribuirse a la izquierda del cero. Finalmente, llegando a una muestra de 500 los cuartiles 2 y 3 se vuelven simétricos así también como los 1 y 4, que representan a ambas colas en una distribución normal.

3.3.2. QQ-Plots

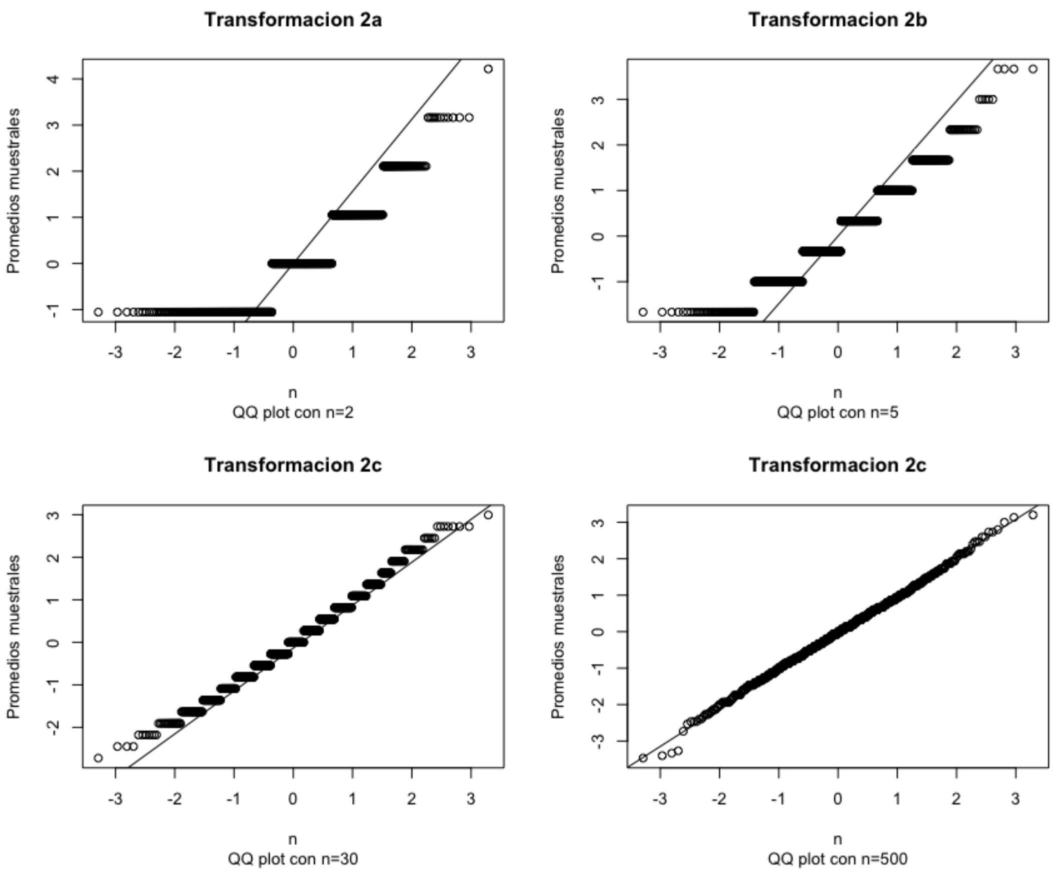

Figura 15: QQ-Plots de transformaciones de distribuciones Binomiales aumentando el número de muestras.

Al igual que en el Ejercicio 2, las muestras se van pareciendo a la guía de la normal a medida que incrementa n. También el comportamiento deja de parecerse al de una distribución discreta para acercarse al de una continua. Nótese adicionalmente que, al igual que en los Boxplots anteriores, los valores comienzan dispersos pero finalmente se centran cerca de la mediana.

3.3.3. Histogramas

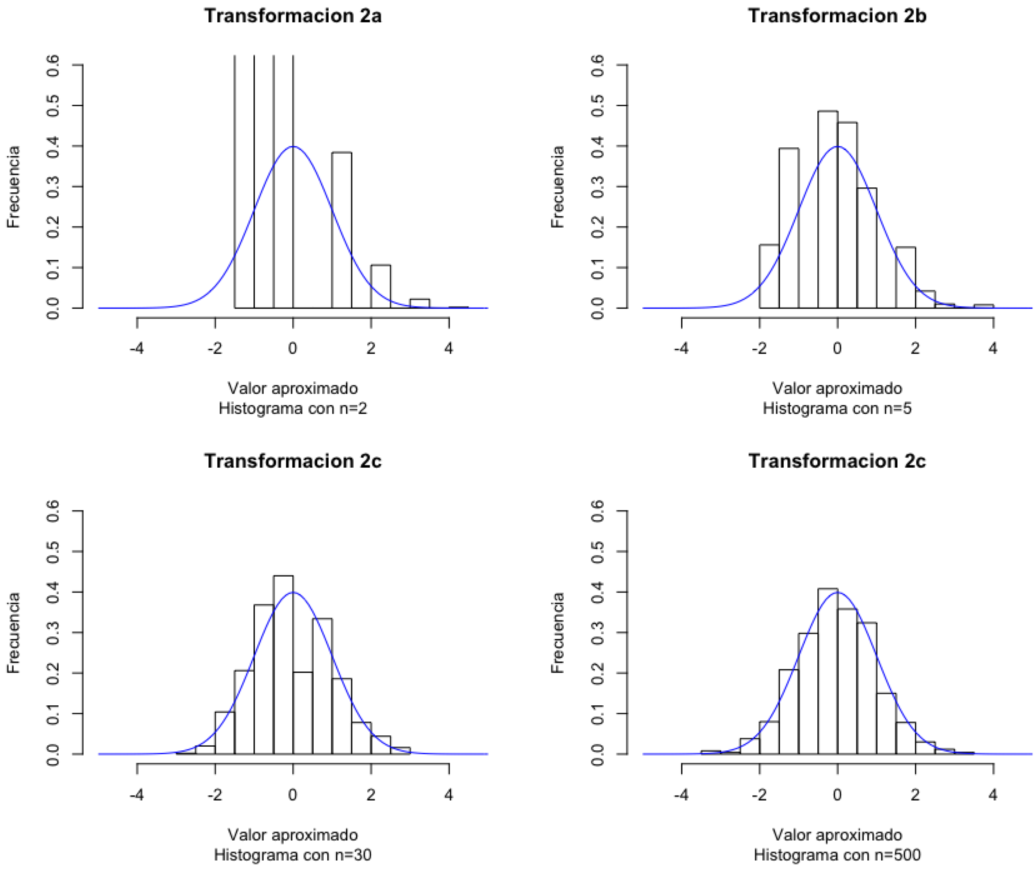

Figura 16: QQ-Plots de transformaciones de distribuciones Binomiales aumentando el número de muestras.

Nuevamente, la linea azul representa la función de densidad de una Distribución normal. Podemos ver que con $n = 2$ los valores de la transformación no se acercan a los esperados, pero al incrementar n los gráficos se amoldan a la distribución normal y los valores se centran en 0.

3.3.4. Conclusión

Las transformaciones aplicadas a los conjuntos de datos del Ejercicio 2 centraron los resultados en 0. Tal como era esperado, al aumentar el número de muestras, las distribuciones se comportaron como una normal. Al llegar al mismo resultado mediante dos distribuciones distintas, solo nos queda concluir que independientemente de la distribución de nuestras variables, se garantiza una distribución normal con n lo suficientemente grande, tal como lo dice el Teorema central del límite.