

Simulación de algunos teoremas de probabilidad y estadística

Samantha Arzate Mariana Martínez Jorge Méndez Tonantzin Real
Diciembre 2020

Conceptos de cursos anteriores:

- Aplicaciones
- Visualizaciones
- Simulación

Teorema del Límite Central

Teorema

Teorema (Teorema del Límite Central (TLC))

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

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

Entonces,

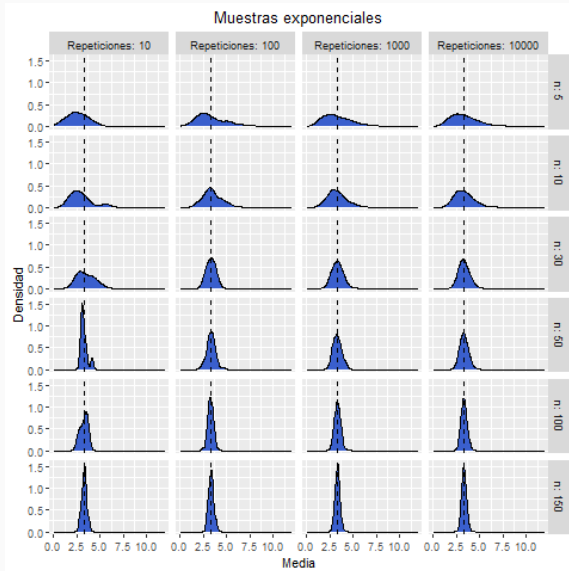
$$\lim_{n \rightarrow \infty} \Pr \left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq z \right) = \Phi(z).$$

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

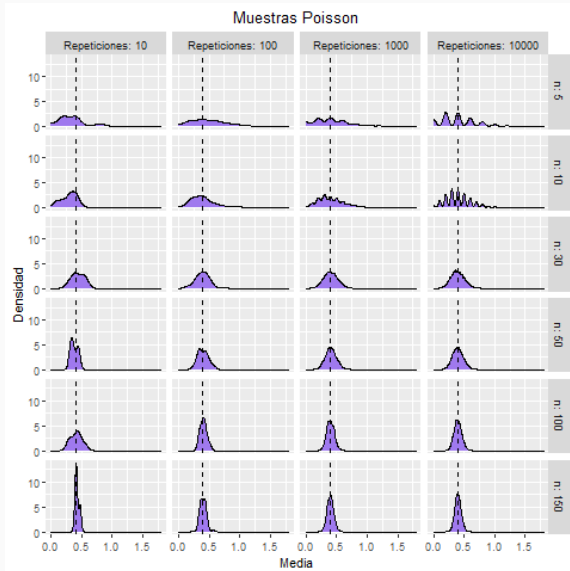
$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

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

Distribución exponencial



Distribución Poisson



Importancia del cumplimiento de los supuestos del TLC

Supuestos requeridos por el TLC

Como vimos anteriormente, el Teorema Central del Límite puede ser aplicado sobre un conjunto X_1, X_2, \dots, X_n de variables aleatorias

- Con media μ y varianza $0 < \sigma^2 < \infty$
- Independientes
- Idénticamente distribuidas

¿Qué comportamiento aberrante podríamos observar si alguno de estos supuestos no se satisface?

¿Qué pasaría si la varianza no fuera finita?

La distribución **Cauchy** es un ejemplo de distribución con **varianza no finita**, de hecho, la distribución Cauchy ni siquiera tiene media finita.

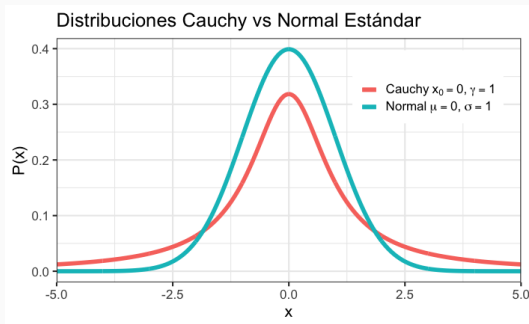


Figura 1: Aunque la densidad de una Cauchy es más baja y estrecha que la densidad de una Normal, sus colas son más anchas

Tras generar 1000 muestras de tamaño 100 a partir de una distribución Cauchy(0,1), obtuvimos los siguientes resultados:

Media	Desv.Est.	Mínimo	Q1	Mediana	Q3	Máximo
0.004	15.019	-294.770	-0.942	0.035	0.956	206.885

Tabla 1: Estadísticas descriptivas de las medias muestrales obtenidas

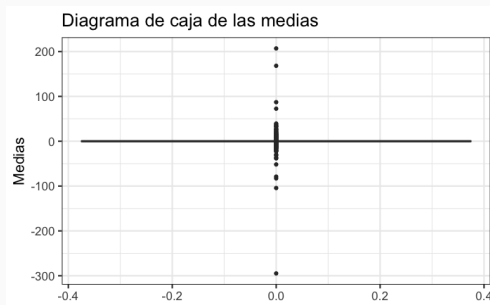


Figura 2: Diagrama de caja para las medias muestrales obtenidas

Como podemos concluir a partir de la siguiente gráfica, ¿esta distribución muestral está lejos de ser normal!

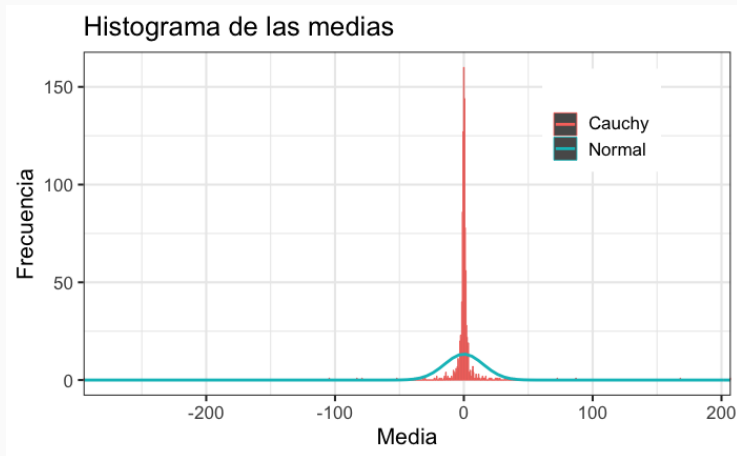


Figura 3: Comparación entre el histograma de las medias muestrales obtenidas y las frecuencias normales esperadas

¿Y si las variables X_1, X_2, \dots, X_n no fueran independientes?

Con la intención de introducir **dependencia** entre las variables uniformes(0,1) generadas, decidimos utilizar **generadores lineales congruenciales (GLC)** a la hora de simularlas.

¿Por qué GLC? Porque, como vimos al principio del curso, 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**.

A continuación un diagrama de correlación para las 20 primeras de 100 variables simuladas a través de GLC.

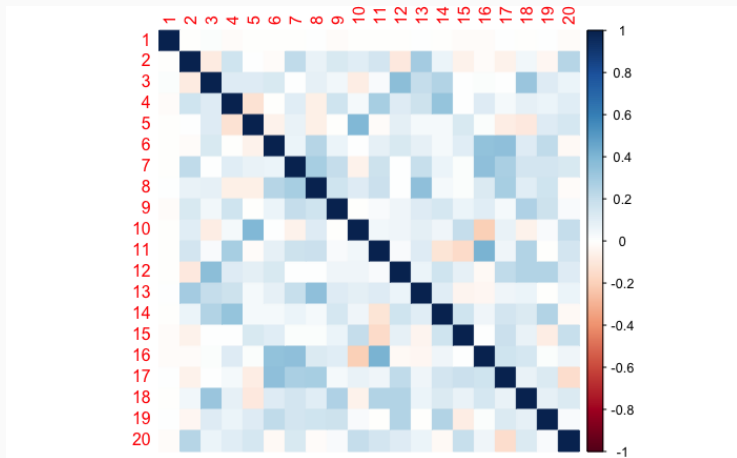


Figura 4: Las variables simuladas tienen correlaciones distintas a cero

Tras calcular las medias de 1000 muestras de tamaño 100 obtenidas con $a = 2i$, $c = 10i$, $z_0 = i$, para $i = 1, \dots, 1000$, 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.

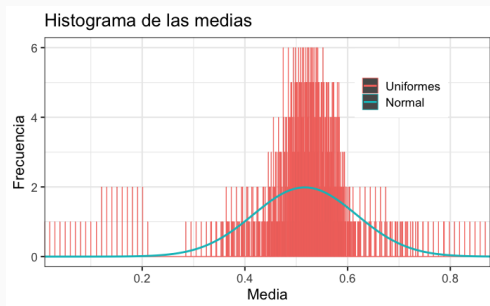


Figura 5: Comparación entre el histograma de las medias muestrales obtenidas y las frecuencias normales esperadas.

¿Y si las variables no estuvieran idénticamente distribuidas?

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 variable X_j , con **$j = 1, \dots, 100$** .

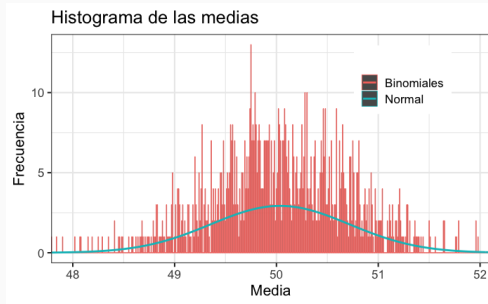


Figura 6: Comparación entre el histograma de las medias de la muestra de tamaño 1000, contra las frecuencias normales esperadas. (Semilla 123)

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.

Intervalos de Confianza

Distribución Normal

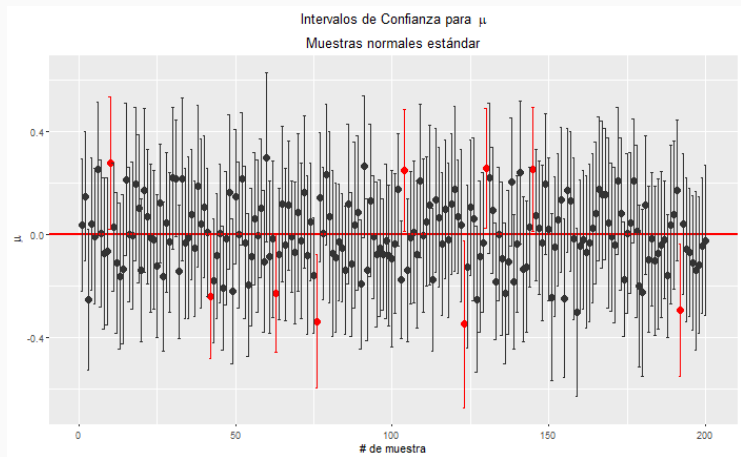


Figura 7: Intervalos de confianza del 95 %

Distribución Bernoulli

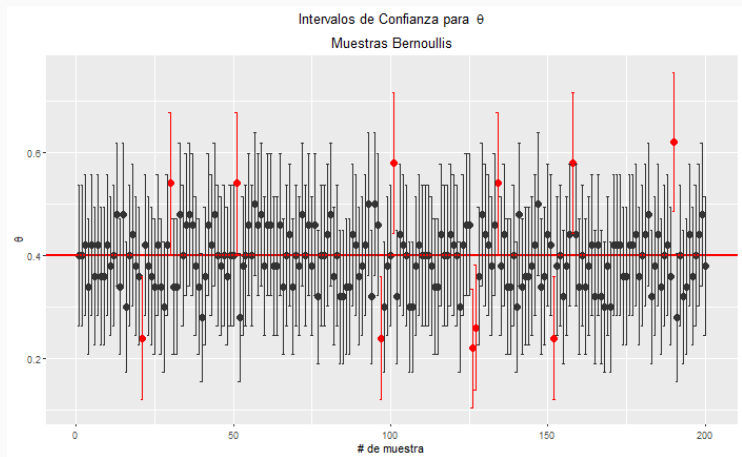


Figura 8: Intervalos de confianza del 95 % para $\theta = 0.4$

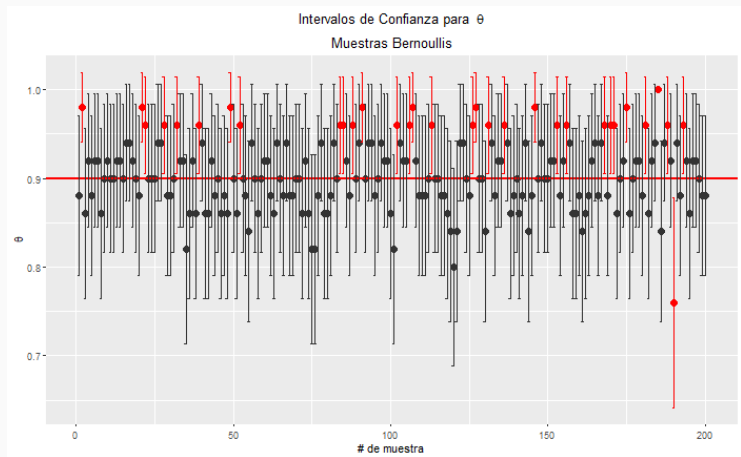


Figura 9: Intervalos de confianza del 95 % para $\theta = 0.9$

Cobertura de los intervalos de confianza

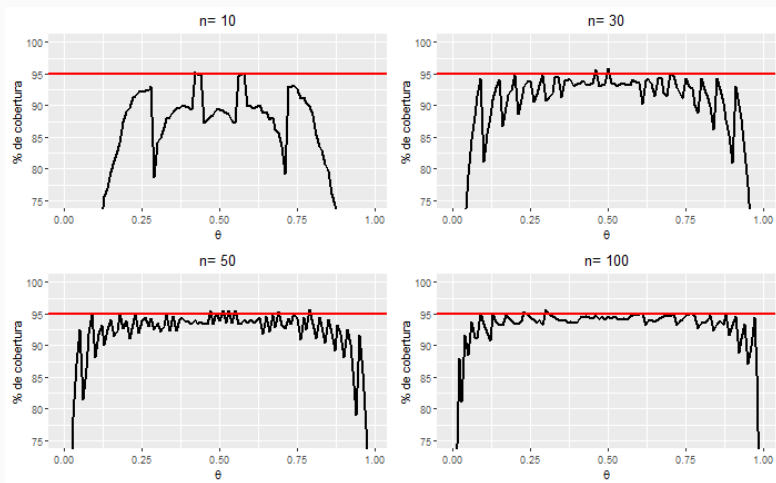


Figura 10: Cobertura de intervalos de confianza para θ

Teorema de la convergencia dominada de Lebesgue

¿Teoría de la medida?

¿Qué es esto?

¿Para qué sirve?

¿Qué tiene que ver esto con “proba”?

¿Por qué es interesante?

¿Una medida de masa? ¿De volumen? ¿De longitud?

Teoría de la medida como la conocemos en matemáticas se empezó a desarrollar a inicios del siglo XX por matemáticos como Émile Borel, Henri Lebesgue, Johann Radon y Maurice Fréchet.

Es una forma de poner en un contexto matemático y generalizar nociones como “área”, “volumen”, “masa”, etc. Los generaliza y se va más allá, se mete con teoría de aproximación, análisis, probabilidad y teoría de la integración.

(Casi) todo empieza con una σ -álgebra

Definición

Una clase no vacía del conjunto potencia de \mathcal{X} , $\mathcal{S} \subseteq \mathcal{P}(\mathcal{X})$ se llama σ -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
$$\bigcup_{k \in \mathbb{N}} E_k \in \mathcal{S}$$
- $\mathcal{X} \in \mathcal{S}$

También podemos generarlas a partir de un conjunto de conjuntos cualquiera.

Definición

La σ -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

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

Definición

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 $\hat{E} \in \hat{\mathcal{S}}$.

Definición

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

Por fin, la definición de medida

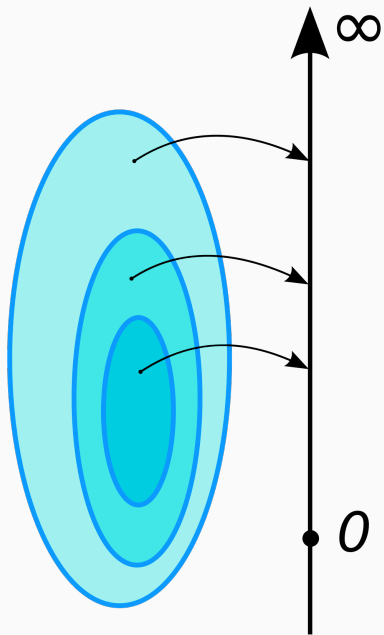
Definición

Sea $(\mathcal{X}, \mathcal{S})$ un espacio medible fijo. Una medida, μ , 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:

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

Esto es un buen ejemplo visual de una medida de un conjunto cualquiera.



Teorema (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 μ y para todo $n \in \mathbb{N}$, entonces si $f(x) = \lim_{n \rightarrow \infty} f_n(x)$:

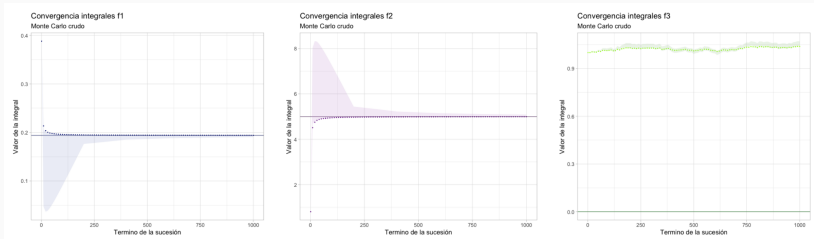
1. $f \in \mathcal{L}_1(\mu)$, es decir, $\int |f| d\mu < \infty$
2. para todo $E \in \mathcal{S}$ tenemos que $\int_E f d\mu = \lim_{n \rightarrow \infty} \int_E f_n d\mu$.

Ejemplos, por favor (y simulación)

Como las funciones de probabilidad acumulada son medidas, las vamos a usar para nuestros ejemplos. Tenemos las siguientes funciones a examinar (y aproximar sus integrales vía Monte Carlo).

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 $[0, 1]$

Simulación brillando



Teorema de Bayes

La posición de una pelota

Consideremos el siguiente experimento:

- Se lanza una pelota sobre una mesa y se marca la posición de la pelota. Se desconoce la posición exacta.
- Se lanza una segunda pelota y registramos si cayó a la izquierda o derecha, adelante o atrás, de la primera bola.
- Repetimos el paso anterior muchas veces.

Thomas Bayes calculó la probabilidad condicional de la posición de la primera bola dada la secuencia de lanzamientos de las demás bolas y nota que entre más pelotas eran lanzadas, más actualizaba su idea de dónde estaba la primera bola.

Teorema (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 cero. 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:

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

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.

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).

Obtener una estimación de N

Usaremos el método de Computación Bayesiana Aproximada:

1. **Distribución a priori $P(N_i)$** Obtener una muestra aleatoria de tamaño 100,000 de la distribución a priori del parámetro $N_1, \dots, N_{100,000}$.
2. **Simulación del MG $P(B|N_i)$** Aplicar el MG a cada observación obtenida de la muestra y así obtener un conjunto de observaciones $B_1, \dots, B_{100,000}$.
3. $\sum_{k=1}^n P(B|N_k)P(N_k)$. Nos quedaremos que las N_i cuyos valores generados B_i sean iguales a los observados. En este caso en particular, nos quedaremos con las N_i que generan 124 peces no marcados.
4. **$P(N_i|B)$** . La distribución de aquellos parámetros que cumplen la condición representa la probabilidad de que la observación haya sido producida por cierto valor del parámetro.

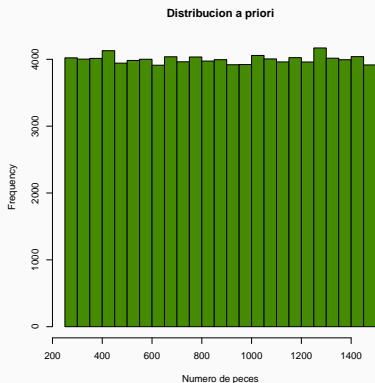


Figura 12: Distribución a priori no informativa

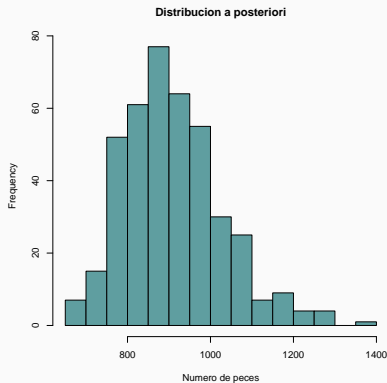


Figura 13: Distribución a posteriori

- Importancia de la verificación de supuestos
- Simular para poder ver comportamiento aberrante esperado
- Efecto de tamaño de muestra y número de repeticiones
- Efecto de elección de parámetros
- Simplificación del entendimiento de temas abstractos
- Actualización nuestras creencias a la luz de nueva evidencia