Análisis estadístico de datos simulados

Dr. Misael Erikson Maguiña Palma

Convergencia

Promedio Teórico Fp.

Datos Simulados de la V. A.

Supongamos que estamos interesados en aproximar la media teórica $\boxed{\mu} = \boxed{E(X)}$ a partir de una secuencia i.i.d. $X_1, X_2, \cdots, X_n$ mediante la media muestral $\bar{X}_n$. Una justificación teórica de la validez de la aproximación obtenida mediante simulación es *la ley (débil) de los grandes números*:

- Si $X_1, X_2, \cdots$ es una secuencia de v.a.'s independientes con:

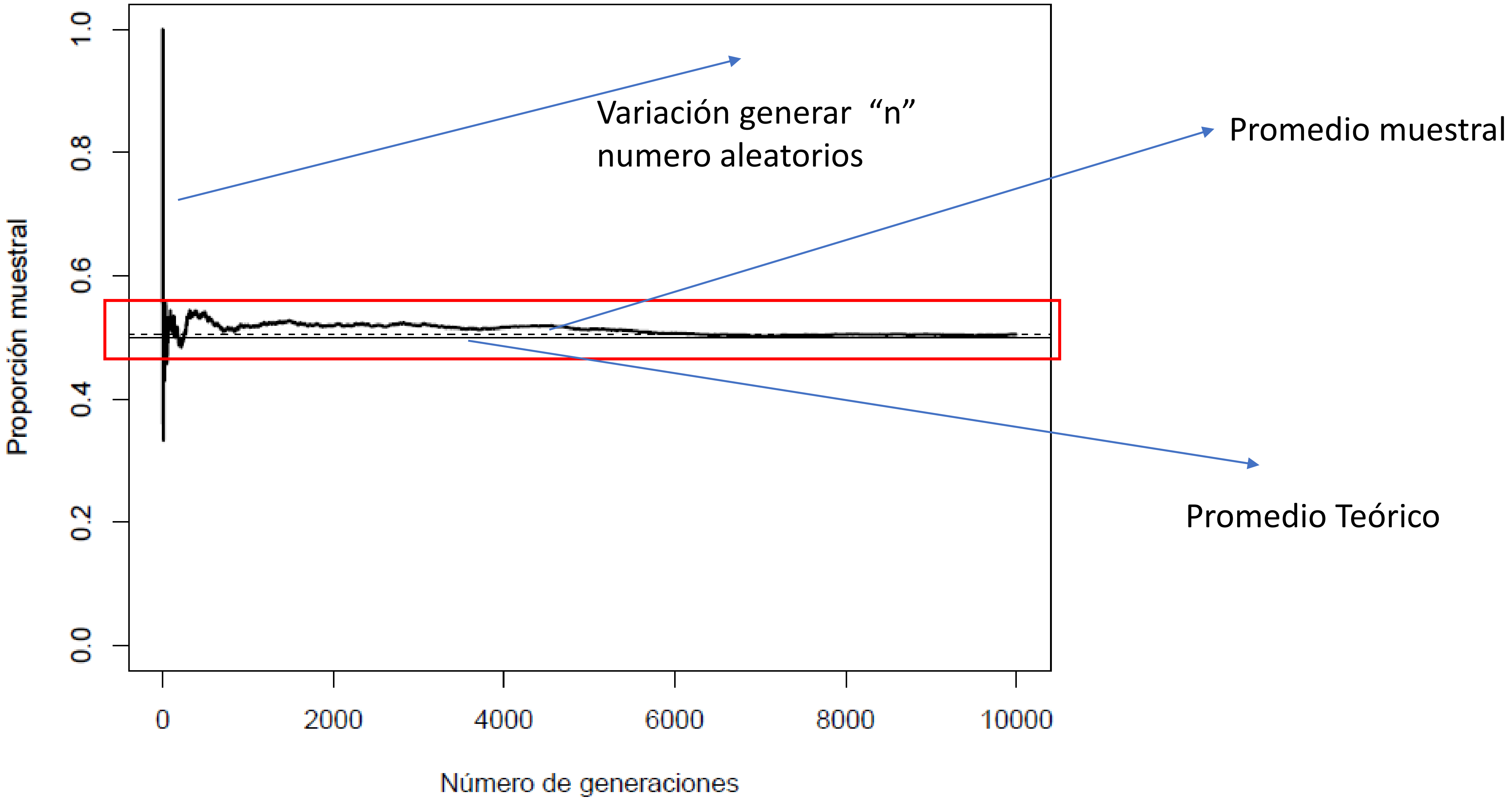
$$E(X_i) = \mu \text{ y } Var(X_i) = \sigma^2 < \infty,$$

entonces $\overline{X}_n = (X_1 + \cdots + X_n)/n$ converge en probabilidad a μ. i.e. para cualquier $\varepsilon > 0$:

$$\lim_{n\to\infty} P\left(\left|\overline{X}_n - \mu\right| < \varepsilon\right) = 1.$$

- *La ley fuerte* establece la convergencia casi segura.

Aproximación de la proporción en función del número de generaciones.



Teorema central del límite

Si $X_1, X_2, \cdots$ es una secuencia de v.a.'s independientes con $E(X_i) = \mu$ y $Var(X_i) = \sigma^2 < \infty$, entonces:

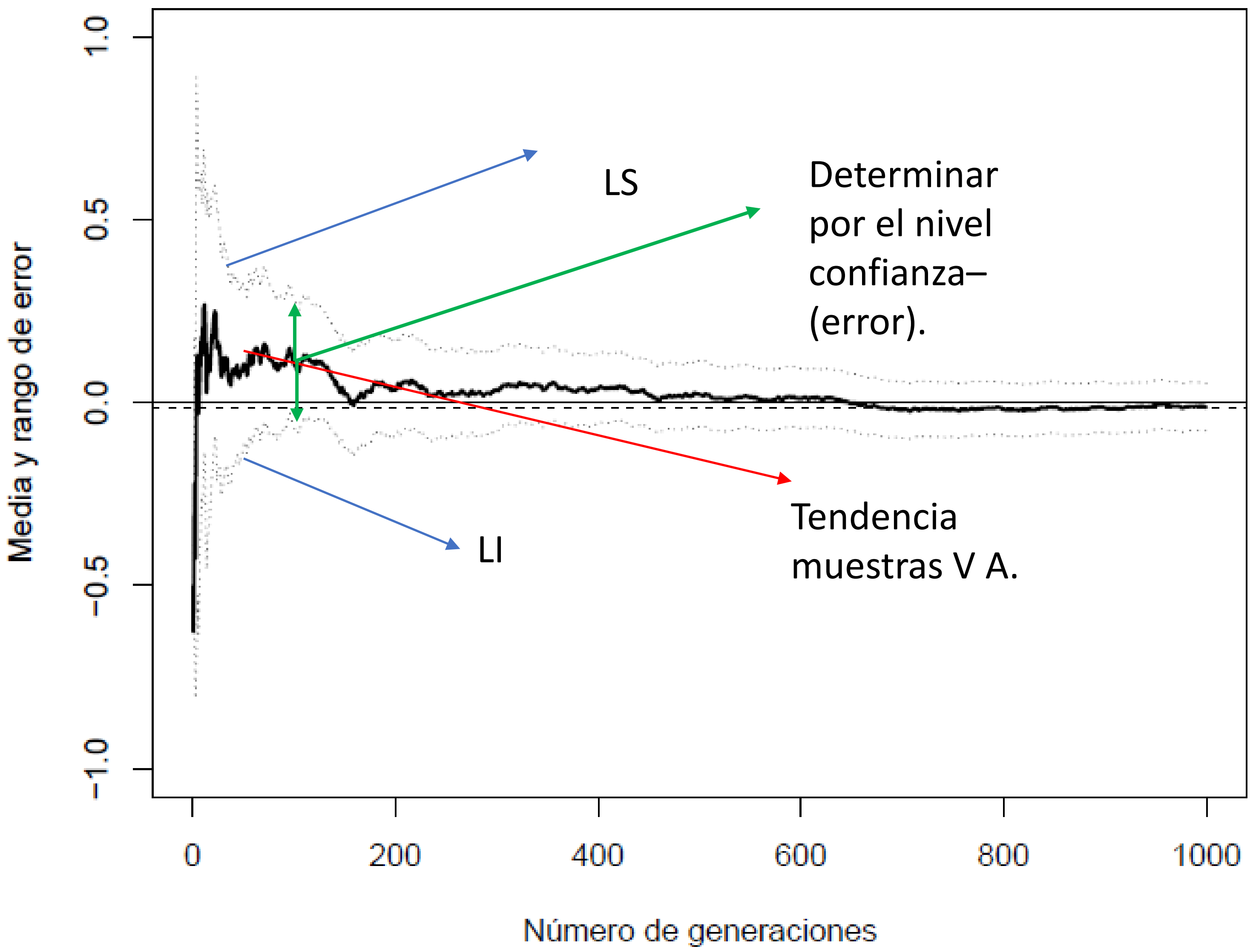
$$Z_n = \frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{d} N(0,1)$$

i.e. $\lim\limits_{n \to \infty} F_{Z_n}(z) = \Phi(z)$. Por tanto, un intervalo de confianza asintótico para μ es:

$$IC_{1-\alpha}(\mu) = \left(\overline{X}_n - \boxed{z_{1-\alpha/2} \frac{\widehat{S}_n}{\sqrt{n}}}, \overline{X}_n + z_{1-\alpha/2} \frac{\widehat{S}_n}{\sqrt{n}} \right).$$

Podemos considerar que $z_{1-\alpha/2} \frac{\widehat{S}_n}{\sqrt{n}}$ es la precisión obtenida (con nivel de confianza $1 - \alpha$).

Gráfico de convergencia incluyendo el error de la aproximación.



Reto (tiempo media hora)

Continúe generando variables aleatorias normales unitarias hasta generar n de ellas, donde $n \geq 30$ es tal que $S/\sqrt{n} < 0.1$, y S es la desviación estándar muestral de los n datos.

(a) ¿Cuántas normales cree que se generen?

(b) ¿Cuántas normales generó?

(c) ¿Cuál es la media muestral de todas las normales generadas?

(d) ¿Cuál es la varianza muestral?

(e) Comente los resultados de (c) y (d) ¿Son sorprendentes?

Tiempo de entrega 11:30 AM

Reto (tiempo media hora)

Se puede mostrar que si sumamos números aleatorios hasta que su suma sea mayor que 1, entonces la cantidad esperada de sumandos es igual a e. Es decir, si

$$N = \min\left\{n\colon \sum_{i=1}^{n} U_i > 1\right\}$$

entonces $E[N] = e$.

(a) Sírvase de lo anterior para estimar e mediante 1000 ejecuciones de simulación.

(b) Estime la varianza del estimador en (a) y dé una estimación de e mediante un intervalo de confianza de 95 por ciento.

Intervalo de confianza para la media

A partir del enunciado del Ejercicio 8.1, se deduce que el intervalo de confianza (de nivel $1 - \alpha$) para la media μ de una población normal con varianza conocida es:

$$IC_{1-\alpha}(\mu) = \left(\overline{X} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}, \ \overline{X} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \right).$$

La idea es que el $100(1 - \alpha)\%$ de los intervalos así construidos contentrán el verdadero valor del parámetro.

Ejemplo

1) Crear un conjunto de datos muestras con 500 muestras de tamaño $n = 10$ de una $N(1, 2)$. Añadir al conjunto de datos las estimaciones de la media y desviación típica obtenidas con cada una de las muestras.
2) Repetir el anterior considerando muestras de una Poisson(1) (tener en cuenta que $X \sim Poisson(\lambda)$ $\mu X = \sigma X = \lambda$). ¿Qué ocurre con la distribución de Poisson?

Contrastes de hipótesis

Analizar el comportamiento del contraste de Kolmogorov-Smirnov para contrastar normalidad empleando repetidamente este test, considerando 1000 pruebas con muestras de tamaño 30 de una $N(0, 1)$. Comparar gráficamente el ajuste de la distribución del p-valor a la de referencia (estudiar el tamaño del contraste).