

Clase 13: Ley de los grandes números

Matías Carrasco

27 de agosto de 2019

Resumen Veremos la LGN para sucesiones iid de v.a. con varianza finita. Motivaremos su contenido con varios ejemplos de sumas de v.a.

Índice

Distribución de una suma	1
Desigualdades de Markov y Chebyshev	8
Ley de los grandes números (LGN)	12

Distribución de una suma

En esta sección vamos a estudiar la distribución de la suma $Z = X + Y$ de dos variables independientes X e Y . Esto ya lo hemos hecho cuando estuvimos estudiando la densidad y la fpp de una variable $Z = g(X, Y)$. Ahora veremos en más detalle qué ocurre con el caso especial de $g(x, y) = x + y$, siendo nuestro principal objetivo entender la distribución de una suma de un gran número de variables.

Caso discreto

En el caso discreto la distribución de la suma queda determinada por la fórmula

$$P(X + Y = z) = \sum_x P(X = x, Y = z - x).$$

Como además X e Y son independientes, la distribución conjunta $P(X = x, Y = z - x)$ se factoriza como producto

$$P(X = x, Y = z - x) = P(X = x)P(Y = z - x).$$

Reescribiendo esta ecuación en términos de las f.p.p. de cada variable, obtenemos la fórmula de *convolución* discreta

$$p_Z(z) = \sum_x p_X(x)p_Y(z - x). \quad (1)$$

Esta fórmula describe como obtener la f.p.p. de Z en función de las de X e Y . La suma debe efectuarse en todos los valores de x en el recorrido de X .

■ **Ejemplo 1** Sean X_1 y X_2 los resultados de lanzar dos veces un dado. Vamos a determinar la f.p.p. de la suma $S_2 = X_1 + X_2$. Los valores

posibles para la suma son $\{2, \dots, 12\}$. Sea s un número cualquiera de estos, la fórmula de convolución (1) nos dice que

$$p_{S_2}(s) = \sum_{x=1}^6 p_{X_1}(x)p_{X_2}(s-x)$$

Pero $p_{X_2}(s-x)$ no es cero solamente cuando $s-x$ es también un número en $\{1, \dots, 6\}$. Es decir, se deben cumplir las dos condiciones siguientes:

$$x \in \{1, \dots, 6\} \quad \text{y} \quad s-x \in \{1, \dots, 6\}.$$

Estas dos condiciones son equivalentes a

$$x \in \{1, \dots, 6\} \cap \{s-6, \dots, s-1\}.$$

Llamemos $N(s)$ a la cantidad de números x que hay en esa intersección.¹ Para los x en la intersección, ambos factores $p_{X_1}(x)$ y $p_{X_2}(s-x)$ son iguales a $1/6$, por lo que la fórmula da

$$p_{S_2}(s) = \sum_{x \text{ en la intersección}} p_{X_1}(x)p_{X_2}(s-x) = N(s)/36.$$

Nos resta calcular $N(s)$. Los valores de $N(s)$ los podemos obtener viendo la Tabla 1.

Resulta entonces que $N(s)$ aumenta linealmente de a uno hasta $s = 7$, y luego decrece también linealmente de a uno. Entonces, la fpp de S_2 es

Valor de s	2	3	4	5	6	7	8	9	10	11	12
$p_{S_2}(s)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

En lugar de hacerlo de forma analítica con la fórmula de convolución (1), haremos ahora un razonamiento visual para ilustrar lo que ésta fórmula esconde.

El dibujo de la Figura 1 muestra la distribución conjunta de los dos resultados X_1 y X_2 . Cada punto verde representa un caso equiprobable, en la horizontal ponemos los de X_1 y en la vertical los de X_2 . Los puntos rojos en el interior de las celdas representan casos equiprobables del par (X_1, X_2) . Como las variables son independientes, la cantidad de puntos rojos en la celda (i, j) se obtiene multiplicando la cantidad de verdes en i por la de verde en j . En este caso esto es siempre $1 \times 1 = 1$, por eso ponemos un solo punto rojo en cada celda.

Para obtener la distribución de la suma S_2 , debemos agrupar las celdas que suman un mismo valor. Estas son las diagonales que cruzan el cuadro desde la punta superior izquierda a la inferior derecha. Si deslizamos los puntos rojos a lo largo de estas diagonales, para que la fila de puntos comience en una base paralela a la diagonal opuesta, obtenemos una hermosa imagen de la distribución. Cada punto rojo

¹ $N(s)$ depende s . Por ejemplo, si $s = 2$ hay un solo x que verifica ambas condiciones y es $x = 1$.

		x					
		1	2	3	4	5	6
s	2	X					
	3	X	X				
	4	X	X	X			
	5	X	X	X	X		
	6	X	X	X	X	X	
	7	X	X	X	X	X	X
	8		X	X	X	X	X
	9			X	X	X	X
	10				X	X	X
	11					X	X
	12						X

Tabla 1: Cada cruz X representa un punto de la intersección.

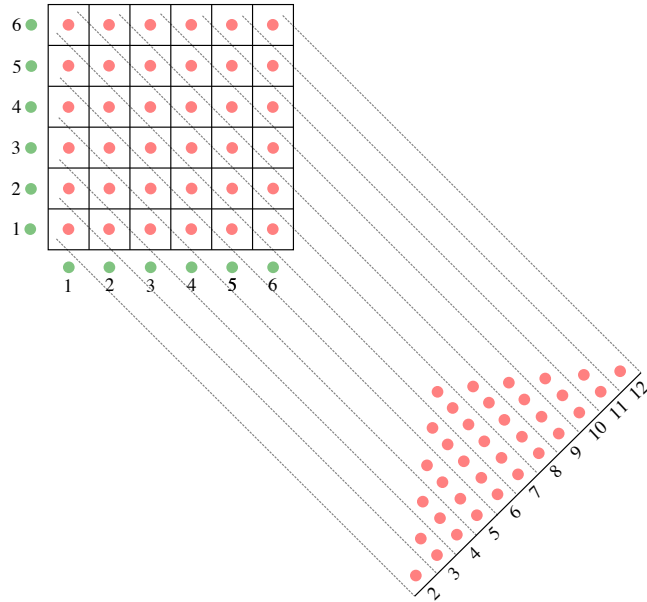


Figura 1: Ilustración de la fórmula de convolución para hallar la fpp de S_2 .

es un caso equiprobable de S_2 . Observar la forma triangular que se obtiene. ■

■ *Ejemplo 2* Supongamos ahora que lanzamos una vez más el dado del ejemplo anterior. Llamemos X_3 al resultado y $S_3 = X_1 + X_2 + X_3$ la suma de los tres resultados. ¿Cuál es la distribución de S_3 ?

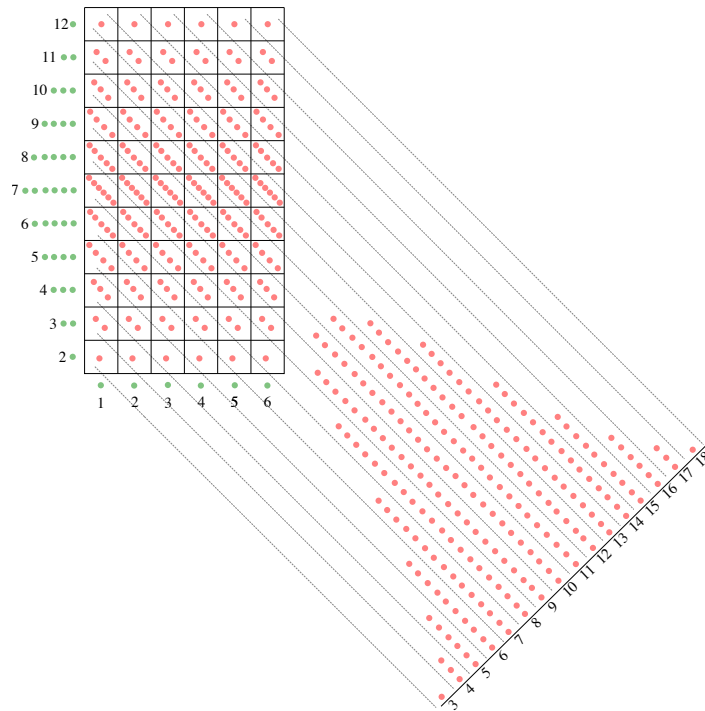


Figura 2: Ilustración de la fórmula de convolución para hallar la fpp de S_3 .

En el dibujo de la Figura 2 se muestra el mismo mecanismo que nos ayudo a encontrar la distribución de S_2 . En este caso, como $S_3 = X_2 + S_2$, los puntos verdes en la horizontal representan casos equiprobables de X_3 , y los de la vertical casos equiprobables de S_2 . Observar ahora que las celdas interiores no tienen la misma cantidad de puntos rojos, pues los puntos verdes de S_2 no están uniformemente repartidos. De todos modos, el mecanismo es el mismo: dejamos deslizar los puntos sobre las diagonales, y el resultado es la distribución de S_3 . ■

Dos cosas interesantes se observan de los ejemplos con los dados. La primera es que la forma de las distribuciones de $S_1 = X_1$, S_2 , y S_3 son cada vez más acampanadas. Lo segundo es que las distribuciones se van concentrando al rededor de un valor central. En este segundo punto nos centraremos hoy.

Llamemos $\bar{X}_1 = S_1$, $\bar{X}_2 = S_2/2$ y $\bar{X}_3 = S_3/3$ los promedios de los lanzamientos del dado de los ejemplos anteriores. El valor esperado de cada uno de ellos es igual 3.5, el valor esperado de un solo lanzamiento:

$$E(\bar{X}_1) = E(S_1) = E(X_1) = 3.5$$

$$E(\bar{X}_2) = E(S_2) / 2 = (E(X_1) + E(X_2)) / 2 = 3.5$$

$$E(\bar{X}_3) = E(S_3) / 3 = (E(X_1) + E(X_2) + E(X_3)) / 3 = 3.5$$

Calculemos la probabilidad de que el promedio esté entre 3 y 4 en cada caso. Basta contar los puntos rojos correspondientes. Obtenemos así la siguiente tabla:

i	Rango \bar{X}_i	Rango S_i	Probabilidad
1	3-4	3-4	$2/6 \approx 33\%$
2	3-4	6-8	$16/36 \approx 44\%$
3	3-4	9-12	$104/216 \approx 48\%$

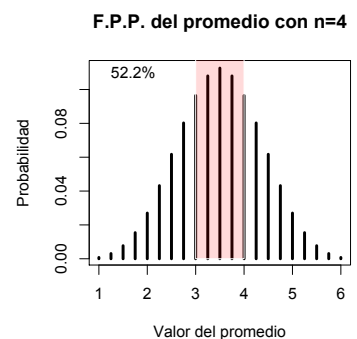
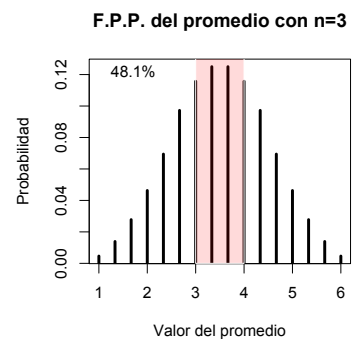
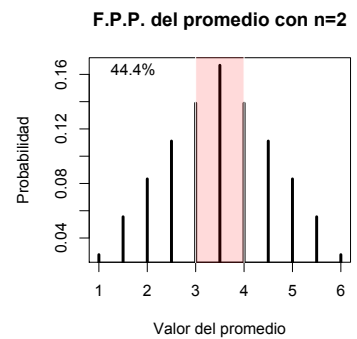
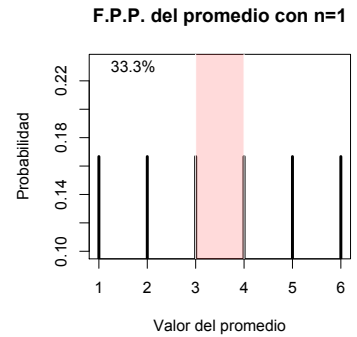
La probabilidad va aumentando. ¿Será que al promediar más y más dados, la probabilidad aumente más y más?

He aquí las probabilidades para el promedio de 4 y 5 dados:

i	Rango \bar{X}_i	Rango S_i	Probabilidad
4	3-4	12-16	$676/1296 \approx 52\%$
5	3-4	15-20	$4332/7776 \approx 56\%$

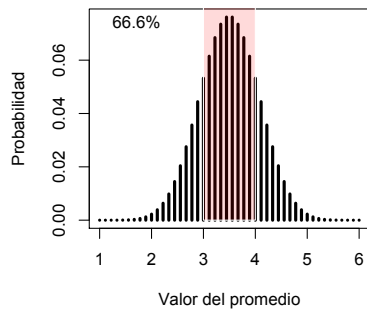
De hecho, Abraham De Moivre (1667-1754), un matemático francés que vivió exiliado casi toda su vida en Inglaterra por cuestiones religiosas, calculó la fórmula exacta para la distribución de la suma de n dados. Su fórmula calcula la probabilidad de obtener una suma igual a s en n lanzamientos, y es

$$p_n(s) = \frac{1}{6^n} \sum_{k=0}^M (-1)^k \binom{n}{k} \binom{s-6k-1}{n-1}, \quad M := \left\lfloor \frac{s-n}{6} \right\rfloor$$

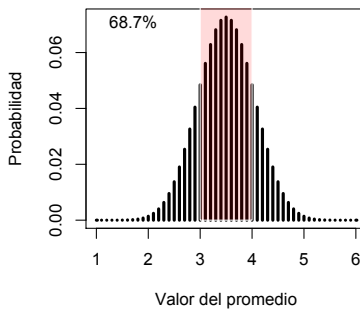


Esta maravillosa fórmula no sirve de mucho sin una computadora. Pero teniendo una, podemos graficar y calcular probabilidades en segundos. Las páginas siguientes contienen las gráficas para algunos valores de n . En cada gráfico, arriba a la izquierda se muestra la probabilidad de que el promedio caiga entre 3 y 4.

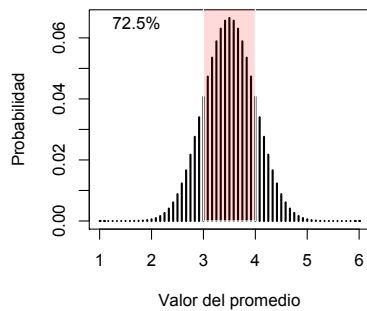
F.P.P. del promedio con n=9



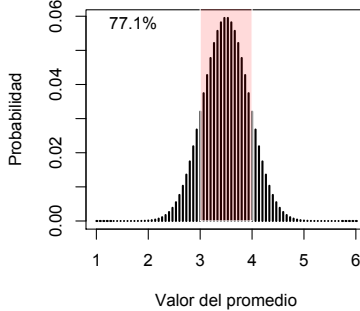
F.P.P. del promedio con n=10



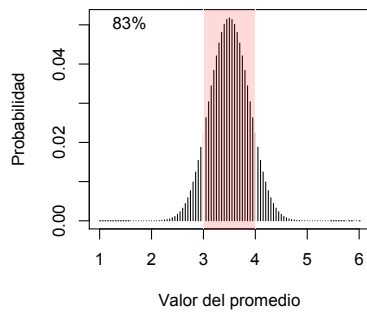
F.P.P. del promedio con n=12



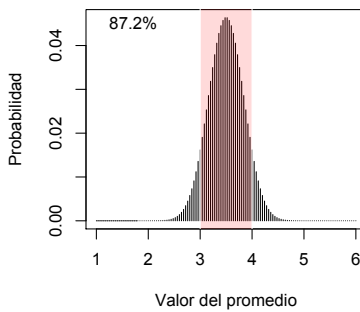
F.P.P. del promedio con n=15



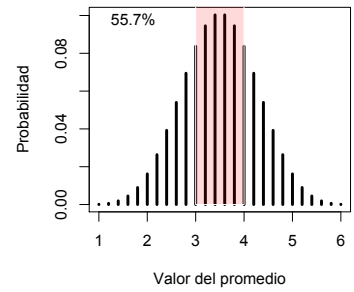
F.P.P. del promedio con n=20



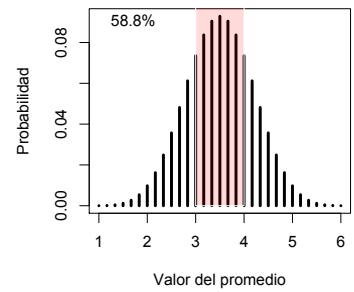
F.P.P. del promedio con n=25



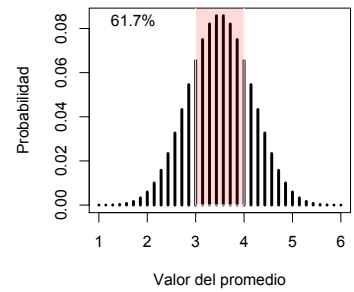
F.P.P. del promedio con n=5



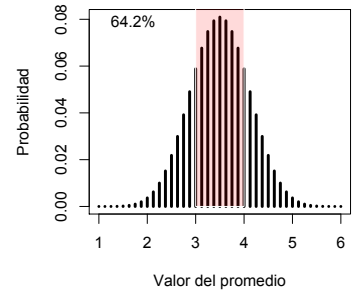
F.P.P. del promedio con n=6



F.P.P. del promedio con n=7



F.P.P. del promedio con n=8



Podemos completar la tabla con los cálculos hechos en computadora:

i	Rango \bar{X}_i	Rango S_i	Probabilidad
6	3-4	18-24	$\approx 59\%$
7	3-4	21-28	$\approx 62\%$
8	3-4	24-32	$\approx 64\%$
9	3-4	27-36	$\approx 67\%$
10	3-4	30-40	$\approx 69\%$
12	3-4	36-48	$\approx 73\%$
15	3-4	45-60	$\approx 77\%$
20	3-4	60-80	$\approx 83\%$
25	3-4	75-100	$\approx 87\%$

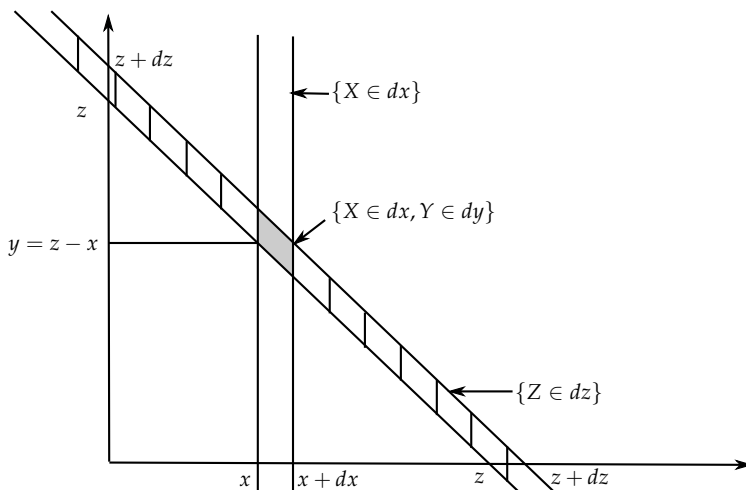
Estas probabilidades parecen crecer a 100% a medida que lanzamos más y más dados. También la forma de la distribución de \bar{X}_n se hace más y más acampanada.

Lo mismo ocurre si cambiamos el rango, por ejemplo si queremos que el promedio caiga entre 3.25 y 3.75. A medida que lanzamos más dados es más probable que el promedio caiga en ese rango. Cuando el rango es más chico, debemos lanzar más dados para alcanzar un mismo nivel de probabilidad. Este fenómeno es exactamente lo que establece la Ley de los grandes números.

Caso continuo

El mismo fenómeno ocurre también cuando promediamos variables continuas. La distribución de la suma $Z = X + Y$ de dos variables continuas independientes también tiene una fórmula de convolución para la densidad, análoga a la fórmula (1).

Llamemos $p(x, y)$ a la densidad conjunta de X e Y . El evento $\{Z \in dz\}$ es el conjunto rayado en el diagrama siguiente



Este evento puede descomponerse en fetas verticales correspondientes a los valores de X , como está sugerido en el diagrama. El paralelo-

gramo sombreado en gris contenido en el evento $\{Z \in dz\}$ y cercano al punto $(x, z - x)$, representa la intersección de los eventos $\{X \in dx\}$ y $\{Z \in dz\}$ y tiene área $dx dz$. La densidad de probabilidad cerca de este pequeño paralelogramo es $p(x, z - x)$, por lo que

$$P(X \in dx, Z \in dz) = p(x, z - x) dx dz.$$

Para obtener la densidad marginal de Z basta integrar en todos los valores de x , y luego dividir por dz , es decir

$$p(z) = \frac{P(Z \in dz)}{dz} = \int_{-\infty}^{+\infty} p(x, z - x) dx.$$

Como X e Y son independientes, la densidad conjunta se factoriza

$$p(x, y) = p_X(x) p_Y(y).$$

De aquí resulta la fórmula de convolución de densidades para la densidad de Z :

$$p_Z(z) = \int_{-\infty}^{+\infty} p_X(x) p_Y(z - x) dx. \tag{2}$$

Notar la similitud con la ecuación (1).

■ **Ejemplo 3** Supongamos que X e Y son independientes, cada una con distribución uniforme en $(0, 1)$. Queremos hallar la densidad de $Z = X + Y$, y como en el caso discreto, haremos argumentos visuales para mostrar lo que la fórmula de convolución (1) esconde.

Aquí (X, Y) tiene distribución uniforme en el cuadrado de lado 1. Para $0 < z < 1$, el evento $\{X + Y \in dz\}$ está representado en el diagrama de la Figura 3 por un conjunto de área $z dz + \frac{1}{2}(dz)^2$. Para calcular esta área hemos dividido al conjunto en un paralelogramo de altura z perpendicular a los lados de ancho dz , y la mitad de un cuadrado de lado dz . Ignorando el término $(dz)^2$ por despreciable en comparación con dz , resulta

$$P(Z \in dz) = z dz,$$

ya que el área total del cuadrado es 1.

Para $1 \leq z < 2$ se puede ver de forma similar que

$$P(Z \in dz) = (2 - z) dz.$$

Entonces, la densidad de Z viene dada por

$$\begin{cases} z & \text{si } 0 < z < 1; \\ 2 - z & \text{si } 1 \leq z < 2; \\ 0 & \text{en otro caso.} \end{cases}$$

Observar que es de forma triangular como en el caso discreto (Fig 4).

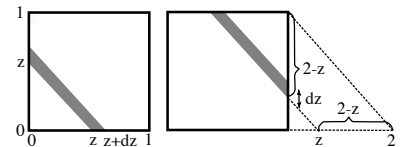


Figura 3: El evento $\{X + Y \in dz\}$.

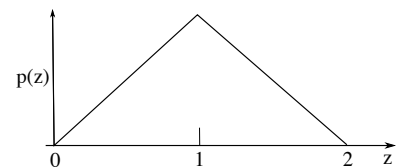


Figura 4: La densidad de Z .

■ **Ejemplo 4** Consideremos ahora $T = X + Y + W$ en donde $X, Y,$ y W son independientes con distribución uniforme en $[0, 1]$. Para calcular la densidad de T , escribimos $T = Z + W$ en donde $Z = X + Y$ tiene la distribución triangular que calculamos en el ejemplo anterior.

Aplicando la fórmula de convolución (2) vemos que

$$p_T(t) = \int_{-\infty}^{+\infty} p_Z(z)p_W(t-z)dz = \int_{t-1}^t p_Z(z)dz = P(t-1 \leq Z \leq t),$$

ya que $p_W(w) = 1$ si $w \in [0, 1]$ y 0 si no.

Basta entonces calcular esta probabilidad. Hay tres casos para discutir:

Caso 1: $0 < t < 1$. Entonces $t - 1 < 0$, por lo que

$$p_T(t) = P(t-1 \leq Z \leq t) = \frac{1}{2}t^2$$

pues es el área del triángulo de la Figura 5.

Caso 2: $1 \leq t < 2$. Entonces $0 \leq t - 1 < 1$. El área relevante es 1 menos dos triángulos, por lo que

$$p_T(t) = P(t-1 \leq Z \leq t) = 1 - \frac{(2-t)^2}{2} - \frac{(t-1)^2}{2} = -t^2 + 3t - 3/2.$$

Ver la Figura 6.

Caso 3: $2 < t < 3$. Entonces $1 < t - 1 < 2$. El área relevante es ahora un triángulo

$$p_T(t) = P(t-1 \leq Z \leq t) = \frac{(3-t)^2}{2}.$$

Ver la Figura 7.

En resumen, la densidad de T es una función partida, definida en los intervalos $(0, 1), (1, 2)$ y $(2, 3)$ por las funciones cuadráticas que hemos calculado en cada caso. Observar la forma simétrica y acampanada de la densidad de T (Fig 8).

Al igual que en el caso discreto, la convolución de dos densidades se puede calcular numéricamente con una computadora. En las figuras que siguen (en la página siguiente) se muestran las densidades de promediar n variables uniformes independientes, de $n = 1$ a $n = 6$. Notar que el valor de la densidad en el promedio $\mu = 0.5$ es cada vez mayor. Notar a su vez la forma simétrica y acampanada de las distribuciones.

Desigualdades de Markov y Chebyshev

Las desigualdades de Markov y Chebyshev establecen cotas superiores para las colas de una distribución de una variable en función de

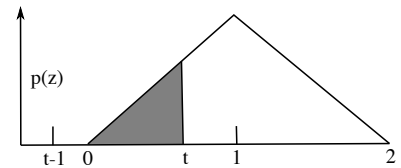


Figura 5: Caso 1: $0 < t < 1$

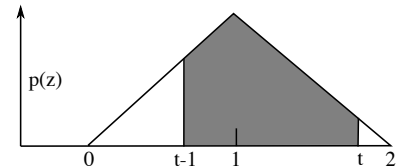


Figura 6: Caso 2: $1 \leq t < 2$

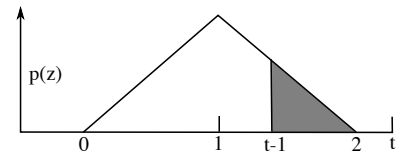


Figura 7: Caso 3: $2 < t < 3$

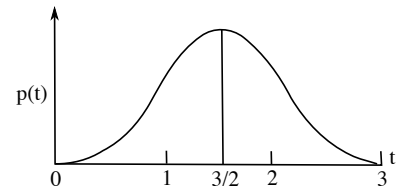
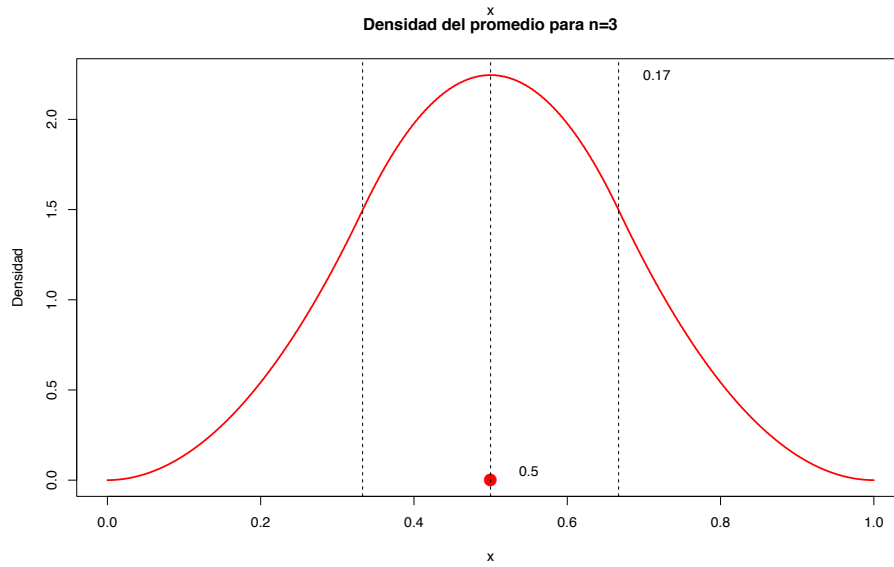
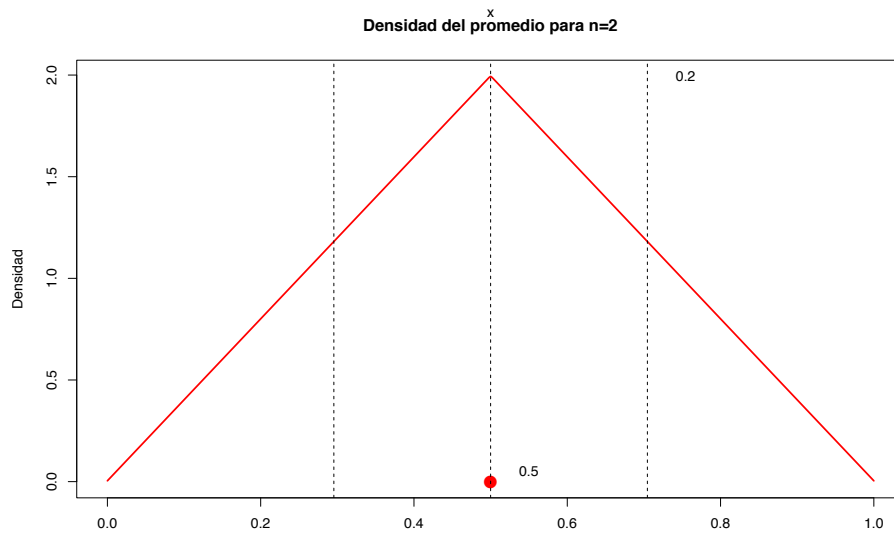
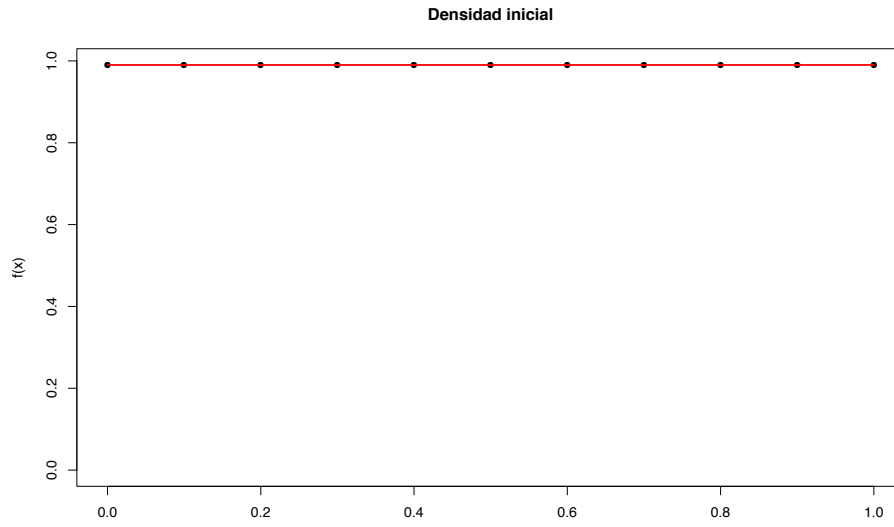
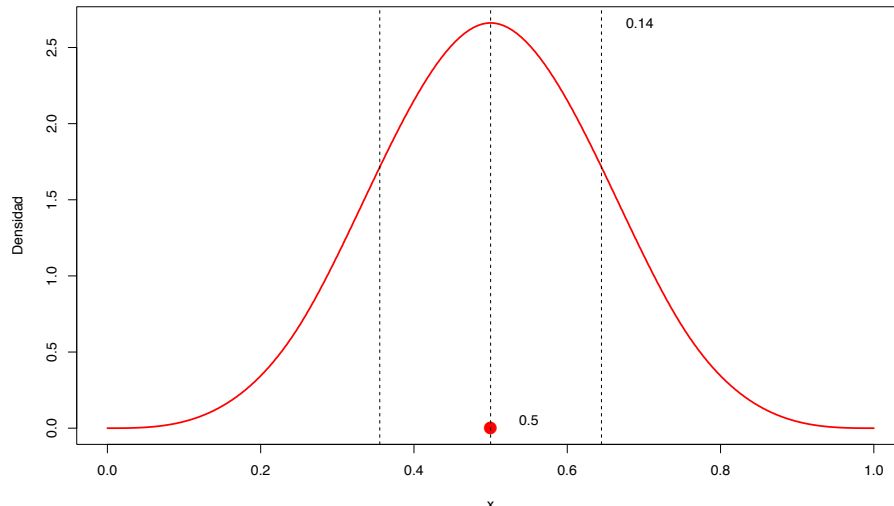


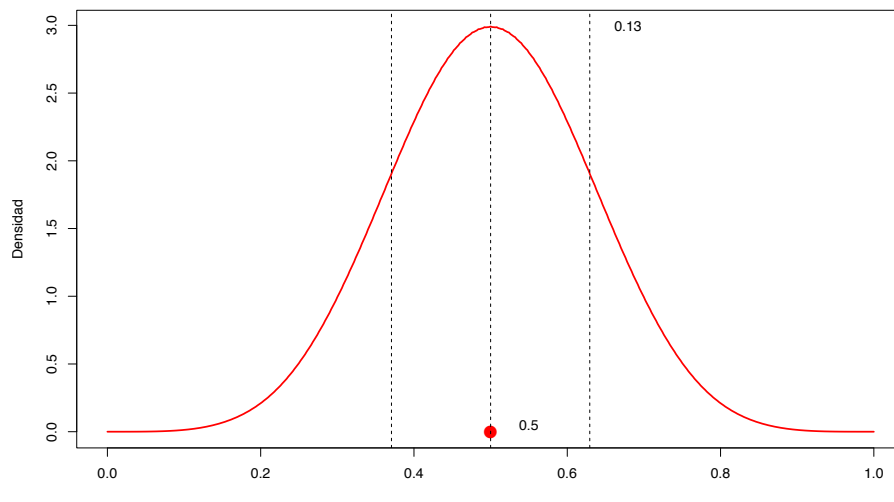
Figura 8: La densidad de T .



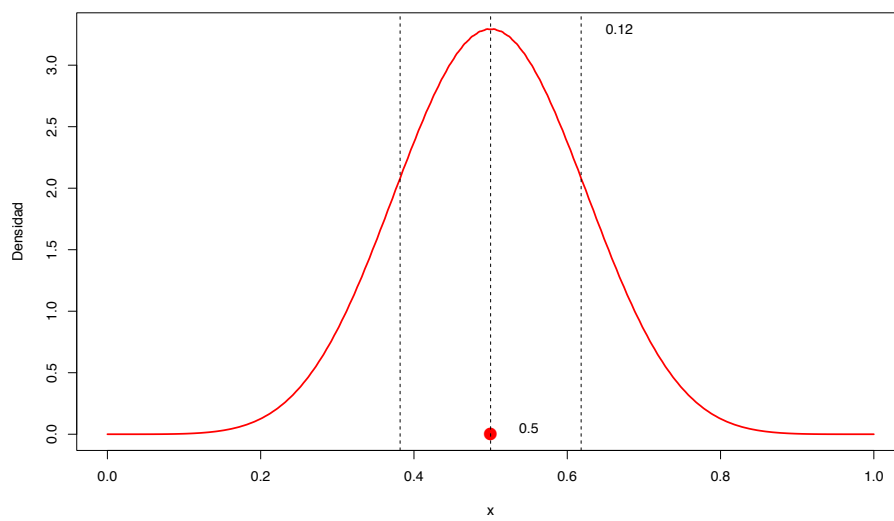
Densidad del promedio para n=4



Densidad del promedio para n=5



Densidad del promedio para n=6



sus momentos. Aunque son muy simples, la demostración de la LGN se basa en ellas.

La más fundamental es la desigualdad de Markov, de la cual deriva la de Chebyshev.

Desigualdad de Markov

Sea X una variable aleatoria positiva. Entonces, para todo $t > 0$ vale que

$$P(X \geq t) \leq \frac{E(X)}{t}.$$

Demostración. Fijemos $t > 0$, y consideremos la variable aleatoria Bernoulli Y que vale 1 si $X \geq t$ y 0 si no. La probabilidad de éxito de Y es

$$p = P(Y = 1) = P(X \geq t).$$

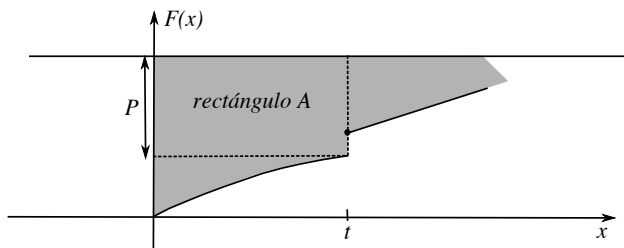
Notar que $tY \leq X$, pues cuando $Y = 0$ la desigualdad es trivial, y cuando $Y = 1$ es porque $X \geq t$. Entonces $E(tY) \leq E(X)$. Por otro lado, la esperanza de Y es p , de donde $tP(X \geq t) \leq E(X)$. Dividiendo por t se deduce la desigualdad de Markov. \square

Para aquellos más afines a las demostraciones visuales he aquí una alternativa. Recordar que la esperanza de X se puede calcular con la función de distribución mediante la fórmula

$$E(X) = \int_0^{\infty} (1 - F(x)) dx,$$

que no es otra cosa que el área gris por encima del gráfico de $F(x)$ en el diagrama.

En el gráfico hemos indicado un rectángulo A cuyos lados son t y $P = P(X \geq t)$. La clave es que $P(X \geq t)$ es el límite por izquierda de $F(x)$ cuando x tiende a t , y por lo tanto está contenido en la zona gris. Esto sin importante la eventualidad de una discontinuidad de F en t .



Claramente

$$E(X) = \text{Área gris} \geq \text{Área de } A = tP(X \geq t),$$

de donde deducimos igual que antes la desigualdad de Markov.

Desigualdad de Chebyshev

Sea X una variable de esperanza $\mu = E(X)$ y varianza $\sigma^2 = \text{Var}(X)$. Entonces, para todo $\epsilon > 0$ vale que

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

Demostración. Consideremos la variable $Y = (X - \mu)^2$. Claramente $Y \geq 0$ y su esperanza es

$$E(Y) = E((X - \mu)^2) = \text{Var}(X) = \sigma^2.$$

Si aplicamos la desigualdad de Markov a Y con ϵ^2 , obtenemos

$$P(Y \geq \epsilon^2) \leq \frac{E(Y)}{\epsilon^2} = \frac{\sigma^2}{\epsilon^2}.$$

El evento $\{Y \geq \epsilon^2\}$ es el mismo que $\{|X - \mu| \geq \epsilon\}$, pues por definición $\sqrt{Y} = |X - \mu|$. Entonces la probabilidad a la izquierda en la desigualdad anterior es igual a $P(|X - \mu| \geq \epsilon)$. \square

Ley de los grandes números (LGN)

La primera versión de la LGN fue probada por Bernoulli en 1713. El Teorema de Bernoulli da una base rigurosa a la interpretación de las probabilidades que usamos al principio del curso. Recordar que interpretamos la probabilidad p de un cierto evento como la proporción del tiempo que el evento ocurre cuando el experimento se repite muchas veces de forma independiente y en las mismas condiciones.

La LGN es una generalización de esta interpretación a promedios generales. En palabras se puede enunciar así:

El promedio de muchas realizaciones independientes está (con alta probabilidad) cerca de la esperanza de la distribución subyacente.

Veamos cómo se relaciona esta afirmación sobre promedios con la interpretación de una probabilidad. En el contexto de los ensayos de Bernoulli, las realizaciones corresponden a los ensayos. La frecuencia relativa de éxitos en n ensayos (la proporción del tiempo que el evento ocurre) se puede escribir como un promedio

$$\bar{X}_n = \frac{X_1 + \cdots + X_n}{n} = \frac{1}{n} \sum_{i=1}^n X_i,$$

en donde X_i es la variable Bernoulli que indica la ocurrencia de éxito en el i -ésimo ensayo. Es usual también escribir $S_n = X_1 + \cdots + X_n$ para la suma.

Cada X_i es Bernoulli de parámetro p , en donde p es la probabilidad de éxito. En particular, todas tienen la misma distribución. Más aún, las X_i son independientes. En la jerga probabilística se dice que X_1, \dots, X_n son independientes e idénticamente distribuidas, y esto se abrevia con la sigla i.i.d..

Usando esta notación el Teorema de Bernoulli se enuncia de la siguiente manera:

$$\text{Para todo } \epsilon > 0: P(|\bar{X}_n - p| < \epsilon) \rightarrow 1 \text{ cuando } n \rightarrow \infty.$$

En general, si X_1, \dots, X_n son variables i.i.d. con esperanza μ y varianza σ , entonces las esperanzas y varianzas del promedio y la suma son

Variable	Esperanza	Varianza	Desvío
S_n	$n\mu$	$n\sigma^2$	$\sqrt{n}\sigma$
\bar{X}_n	μ	σ^2/n	σ/\sqrt{n}

El cálculo es el siguiente:

$$E(S_n) = E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i) = n\mu$$

$$\text{Var}(S_n) = \text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i) = n\sigma^2$$

Para el cálculo de la varianza hemos usado que las variables son independientes. Como $\bar{X}_n = S_n/n$, se tiene

$$E(\bar{X}_n) = E(S_n/n) = E(S_n)/n = \mu$$

$$\text{Var}(\bar{X}_n) = \text{Var}(S_n/n) = \text{Var}(S_n)/n^2 = \sigma^2/n$$

Recordar que las constantes salen al cuadrado de la varianza.

Dos cosas son sorprendentes de los cálculos anteriores:

- La primera es que el desvío de la suma no crece linealmente con n , si no que lo hace como \sqrt{n} .
- El desvío del promedio tiende a cero a medida que n tiende a ∞ .

El segundo punto establece que al promediar muchas veces una variable, su variabilidad decrece. Intuitivamente los excesos por encima de la media compensan el déficit por debajo de ésta, lo que hace que el promedio sea más estable.

Sin embargo, el segundo punto, formalmente igual de difícil que el primero, le llevo muchísimo tiempo a los matemáticos reconocerlo. En cierto sentido nos dice que las mediciones no aportan todas la misma información, si no que a medida que obtenemos más datos la

información aportada por cada medición es menor. Es bastante contraintuitivo a primera vista, aunque luego de una reflexión pausada no parece tan descabellado.

■ *Ejemplo 5* En Inglaterra, en el siglo XII no existía una autoridad central fuerte, lo que representaba un problema para la política monetaria. Había un rey, pero su autoridad era contrarrestada por la de varios poderosos barones.

Al mismo tiempo, existía la necesidad comercial de una moneda confiable. La principal fuente de acuñación británica era la Casa de la Moneda de Londres, que funcionaba de forma independiente de la corona. El rey y los barones llevaban oro y plata a la Casa y recibían monedas a cambio.

El rey estipulaba el peso y la finura de las monedas, y monitoreaba regularmente el cumplimiento de los estándares estipulados a través de pomposos juicios.

Los juicios de la Casa de Moneda datan de finales del 1200 y se siguen haciendo hoy en día. Cada día de producción, una selección (razonablemente aleatoria) de monedas era separada y puesta en una caja llamada Pyx para pruebas posteriores.

Cada tres meses, la Pyx se abría en presencia de jueces que representaban a partes interesadas en la precisión de las monedas. Las mismas eran analizadas para determinar la finura y el peso. Nos concentraremos en lo que sucedía con el peso que es más interesante desde el punto de vista estadístico.

Todos sabían que inevitablemente habría alguna variación en el peso de monedas diferentes. Por ende, el contrato con el rey y los barones especificaba tanto un peso objetivo (llamémoslo M_0) y una tolerancia permitida que se llamaba el “remedio” (denotada por R). Si el peso estaba por debajo de $M_0 - R$, la Casa debía devolver el dinero faltante al rey. Incluso, al maestro de la Casa se le podía cortar la mano, o algo peor. No estaba bien visto robarle al rey.

Las monedas demasiado pesadas también eran un problema, ya que podían ser eliminadas de circulación y convertidas en lingotes por los empresarios alertas. Pero en estos casos no había ganancia para la Casa de la Moneda, y el enfoque principal de la prueba era entonces en las monedas livianas.

Las monedas se pesaban en lotes, pues la precisión de aquellos tiempos no permitía pesar una sola moneda con exactitud.

Si, por ejemplo, se pesan 100 monedas de oro en un lote, claramente el objetivo debe ser $100M_0$. Pero, ¿cuál sería el remedio? Es decir, ¿cuánto se debe tolerar?

Es una pregunta muy simple, y por su respuesta casi que se puede reconocer a una persona bien instruida en estadística.

La elección que hicieron (hasta empezado el siglo XIX) fue que el remedio en ese caso es simplemente $100R$; solo si el lote pesaba menos de $100M_0 - 100R$, la Casa no pasaría la prueba.

Durante mucho tiempo se pensó que los errores crecen linealmente con el número de mediciones. Pero el cálculo que hicimos más arriba nos dice que esto está equivocado; esto es demasiado generoso para la casa de la moneda.

Es un estándar tan bajo que un maestro alerta podría apuntar casi tan bajo, como por ejemplo acuñando a un objetivo de $M_0 - 0.5R$, o incluso $M_0 - 0.8R$, y prácticamente no correr riesgo de fallar en la prueba.

A modo de ejemplo concreto, pensemos en una moneda como la de \$10 que pesa cerca de $M_0 = 10g$. Supongamos que nuestra tolerancia para el peso de una moneda es del 5% del peso de la moneda. En este caso $R = 0.5g$.

Si hiciéramos como hacían en el siglo XII, la tolerancia para los lotes sería de $100R = 50g$. Notar que el peso promedio es de $100M_0 = 1000g$, pero si los pesos de las monedas varían de forma independiente, (es decir que las monedas individuales no están estadísticamente relacionadas entre sí) un remedio apropiado para un lote de 100 sería $10R$ (5g), no $100R$. Con pesos estadísticamente independientes, la variación aumenta como *la raíz cuadrada* del número de monedas.

Por supuesto, en los primeros años del juicio del Pyx, incluso los mejores matemáticos no estaban al tanto de lo que ahora se conoce como la regla de la \sqrt{n} .

Sin embargo, hubo un maestro de la Casa de la Moneda que fue un matemático mejor que el promedio: Isaac Newton. De 1696 a 1727 fue maestro de la Casa de la Moneda. Y a su muerte en 1727, Newton tenía una considerable fortuna. ¿Habría visto Newton el error en los procedimientos en el juicio del Pyx, y lo habrá explotado de forma astuta para su beneficio? ■

La regla de la \sqrt{n} es la clave detrás de la demostración de la LGN.

Ley de los grande números

Sean X_1, X_2, \dots variables i.i.d. con esperanza $\mu = E(X_i)$ y varianza $\sigma^2 = Var(X_i)$. Entonces, para todo $\epsilon > 0$

$$P(|\bar{X}_n - \mu| < \epsilon) \rightarrow 1$$

cuando n tiende a infinito.

Notación: Escribiremos $\bar{X}_n \xrightarrow{P} \mu$ para indicar que \bar{X}_n está, con probabilidad muy alta, tan cerca de μ como queramos.

Demostración. Tomando complementos es equivalente probar que

$$P(|\bar{X}_n - \mu| \geq \epsilon) \rightarrow 0$$

cuando $n \rightarrow \infty$.

Como la varianza de \bar{X}_n es σ^2/n , por la desigualdad de Chebyshev tenemos

$$P(|\bar{X}_n - \mu| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}.$$

El lado de la derecha de esta ecuación claramente tiende a cero cuando n tiende a infinito. \square