

Clase 7: El Teorema Central del Límite

Matías Carrasco

6 de octubre de 2019

Índice

1. Motivación	1
2. Teorema central del límite	3
3. Demostración informal para el caso binomial	8
4. Normalidad asintótica de estimadores	10

1. Motivación

Comencemos con un ejemplo de motivación. Al realizar una encuesta política, los resultados a menudo se informan como un número con un margen de error. Por ejemplo, $52\% \pm 3\%$ favorecen al candidato A. La “regla a ojo” es que si se sondea a n personas, entonces el margen de error es $\pm 1/\sqrt{n}$. Ahora veremos exactamente qué significa esta regla, y esto nos llevará a enunciar un resultado muy importante que se conoce con el nombre de Teorema Central del Límite (TCL).

Supongamos que hay 2 candidatos A y B. Supongamos además que la fracción de la población que prefiere A es p . Es decir, si le pregunta a una persona al azar a quién prefiere, entonces la probabilidad de que contesten A es p .

Para ejecutar la encuesta, un encuestador selecciona n personas al azar y pregunta “¿Apoya usted al candidato A o al candidato B?”. Por lo tanto, podemos ver la encuesta como una secuencia de n ensayos de Bernoulli X_1, X_2, \dots, X_n , donde X_i es 1 si la persona i prefiere A y 0 si prefiere B. La fracción de personas encuestadas que prefiere A es el promedio:

$$\hat{p} = \frac{\text{nro. personas que votan A}}{\text{Total de encuestados}} = \frac{S_n}{n} = \frac{X_1 + \dots + X_n}{n} = \bar{X}$$

Sabemos que cada $X_i \sim \text{Ber}(p)$, entonces $\mathbf{E}(X_i) = p$ y $\mathbf{var}(X_i) = p(1-p)$. Por lo tanto, la suma S_n tiene distribución binomial de parámetros n y p , de donde

$$\mathbf{E}(\hat{p}) = p \text{ y } \mathbf{var}(\hat{p}) = \frac{p(1-p)}{n}.$$

A modo de ejemplo, supongamos que encuestamos a $n = 100$ personas y que el verdadero valor de p , es decir la verdadera proporción de personas que apoyan al candidato A en la

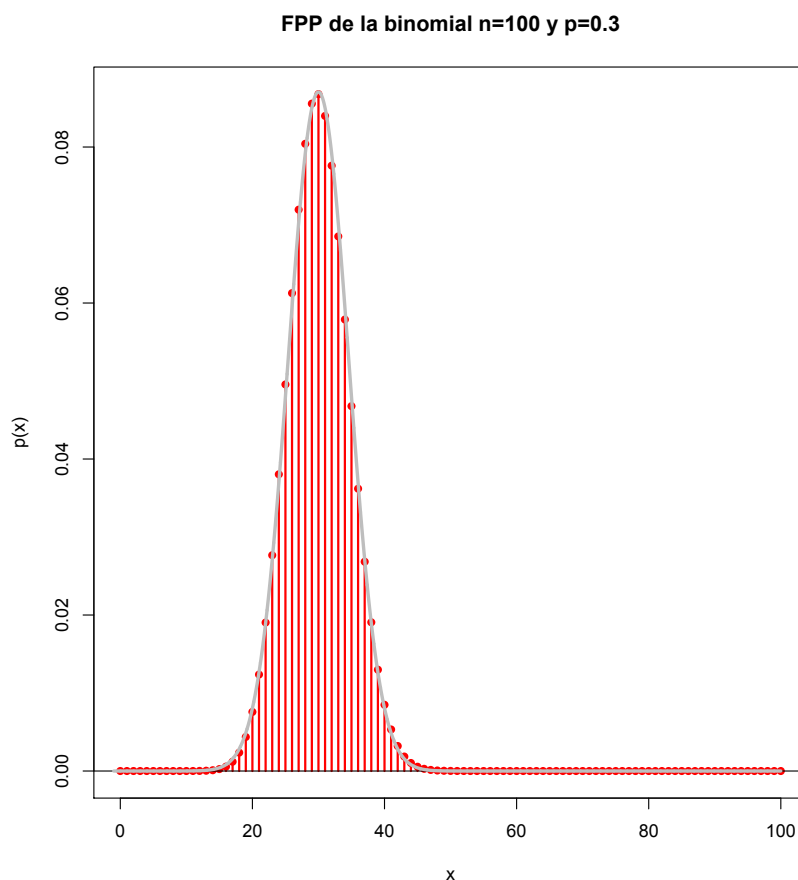


Figura 1: Gráfico de la fpp de una binomial de parámetros $n = 100$ y $p = 0.3$. En gris está superpuesta la densidad de la distribución normal con la misma esperanza y varianza.

población, es igual a 30%. En la Figura 1 se ve el gráfico de la distribución de S_{100} , igual al total de personas encuestadas que votan por A.

Si agregamos al gráfico la curva de la densidad normal que tiene la misma esperanza y varianza que S_{100} , vemos que esta pasa prácticamente por encima de los valores de la fpp. Esto sugiere que, al menos para valores grandes de n , la normal es una buena aproximación

$$\hat{p} \stackrel{d}{\approx} N\left(p, \frac{p(1-p)}{n}\right)$$

a la distribución de $\hat{p} = S_n/n$. Pongamos $\sigma^2 = p(1-p)/n$ para simplificar la notación.

En una distribución normal, el 95% de la probabilidad está dentro de 2 desviaciones estándar de la media. Asumiendo que la normal aproxima bien la distribución de \hat{p} , esto significa que en el 95% de las encuestas de n personas, la media de la muestra \hat{p} estará dentro de $2\sigma/\sqrt{n}$ de la verdadera media p .

El paso final es notar que para cualquier valor de p tenemos $\sigma \leq 1/2\sqrt{n}$.¹ Esto significa que podemos decir de manera conservadora que en el 95% de las encuestas de n personas, la media de la muestra \hat{p} está a menos de $1/\sqrt{n}$ de la media verdadera. El estadístico

¹Es un ejercicio sencillo de cálculo ver que $1/4$ es el valor máximo de $p \mapsto p(1-p)$.

(frecuentista) llama al intervalo $\hat{p} \pm 1/\sqrt{n}$ con el nombre de *intervalo de confianza* (en este caso conservador) al nivel 95% para p . Para abreviar a veces escribimos IdC en lugar de intervalo de confianza.

2. Teorema central del límite

El truco de estandarizar la normal puede imitarse con cualquier otra variable. Si X es una variable aleatoria con esperanza μ y varianza σ^2 , la estandarización de X es la variable

$$Z = \frac{X - \mu}{\sigma}.$$

Observar que Z tiene esperanza 0 (se dice que está centrada) y varianza 1 (se dice que está normalizada). No hay que engañarse, la estandarización de X no tiene porqué tener distribución normal estándar, esto solo pasa si originalmente X es normal.

Sean X_1, \dots, X_n variables i.i.d. con esperanza $\mu = \mathbf{E}(X_i)$ y varianza $\sigma^2 = \mathbf{var}(X_i)$. Como antes, escribimos

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

$$\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$$

pero ahora vamos a introducir un miembro más a la lista. Como S_n y \bar{X}_n son múltiplos una de la otra, ambas tienen la misma estandarización

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}.$$

El teorema central del límite nos permite aproximar una suma o promedio de variables aleatorias i.i.d. por una variable aleatoria normal. Esto es extremadamente útil porque generalmente es fácil hacer cálculos con la distribución normal.

Enunciado informal del TCL

Para n grande,

$$\bar{X}_n \stackrel{d}{\approx} N(\mu, \sigma^2/n), \quad S_n \stackrel{d}{\approx} N(n\mu, n\sigma^2), \quad Z_n \stackrel{d}{\approx} N(0, 1)$$

La notación $X \stackrel{d}{\approx} Y$ quiere decir que la distribución de X es aproximadamente igual a la de Y . Pero a no engañarse, esto no quiere decir que X se parezca a Y , simplemente que la función de distribución F_X se parece a F_Y .

Enunciado preciso del TCL

Sea X_1, X_2, \dots una sucesión i.i.d. de variables aleatorias con esperanza μ y varianza σ^2 .
Sea

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$$

la estandarización de la suma o promedio. Entonces para todo $z \in \mathbb{R}$,

$$\lim_{n \rightarrow \infty} F_{Z_n}(z) = \Phi(z).$$

La prueba del TCL no es muy difícil, y las herramientas utilizadas están al alcance de este curso, pero sí un poco larga y técnica. Por eso la dejamos como lectura opcional en unas notas que pueden encontrar en la página del curso.

Aplicaciones del TCL

Ejemplo 1

Se lanza una moneda justa 100 veces. Estimar la probabilidad de que salga cara en más de 55 de los lanzamientos.

Sea X_i el resultado del i -ésimo lanzamiento, por lo que $X_i = 1$ si sale cara y $X_i = 0$ si sale cruz. La cantidad total de caras es

$$S_{100} = X_1 + \cdots + X_{100}.$$

Sabemos que $\mathbf{E}(X_i) = 1/2$ y $\mathbf{var}(X_i) = 1/4$, por lo que para $n = 100$, tenemos

$$\mathbf{E}(S_{100}) = 50, \quad \mathbf{var}(S_{100}) = 25, \quad \sigma = 5.$$

El TCL dice que la distribución de la estandarización de S_{100} es aproximadamente igual a la distribución $N(0, 1)$. Esto es

$$\mathbf{P}(S_{100} > 55) = \mathbf{P}\left(\frac{S_{100} - 50}{5} > \frac{55 - 50}{5}\right) = \mathbf{P}(Z_{100} > 1) \approx 1 - \Phi(1).$$

Como $\Phi(1) = 0.8413$, resulta $\mathbf{P}(S_{100} > 55) \approx 0.1587$. ■

Ejemplo 2

Estimar la probabilidad de más de 220 caras en 400 lanzamientos de una moneda justa.

Este ejemplo es casi el calco del anterior. Usando ahora $\mu = 200$ y $\sigma = 10$, obtenemos

$$\mathbf{P}(S_{400} > 220) = \mathbf{P}\left(Z_{400} > \frac{220 - 200}{10}\right) = \mathbf{P}(Z_{400} > 2) \approx 1 - \Phi(2).$$

Como $\Phi(2) = 0.9772$, resulta $\mathbf{P}(S_{400} > 220) \approx 0.0228$.

Notar que aunque $55/100 = 220/400$, la probabilidad de más de 55 caras en 100 lanzamientos es mayor que la probabilidad de más de 220 caras en 400 lanzamientos. Esto se debe al valor mayor de n en el segundo caso. ■

Ejemplo 3

Estimar la probabilidad de obtener entre 40 y 60 caras en 100 lanzamientos de una moneda justa.

Como en el primer ejemplo, $\mu = 50$ y $\sigma = 5$, por lo que

$$\mathbf{P}(40 \leq S_{100} \leq 60) = \mathbf{P}(-2 \leq Z_{100} \leq 2) \approx \Phi(2) - \Phi(-2).$$

Como $\Phi(2) - \Phi(-2) = 2\Phi(2) - 1 = 0.9544$, resulta $\mathbf{P}(40 \leq S \leq 60) \approx 0.9544$.

Si usamos la binomial para calcular esta probabilidad el resultado es 0.9648, por lo que la aproximación normal está por debajo del valor exacto en cerca del 1%. ■

¿Por qué usar el TCL?

Dado que las probabilidades en los ejemplos anteriores se pueden calcular exactamente usando la distribución binomial, es posible que se pregunten cuál es el punto de encontrar una respuesta aproximada utilizando la TCL.

De hecho, solo pudimos calcular estas probabilidades exactamente porque los X_i eran Bernoulli y, por lo tanto, la suma S_n era binomial. En general, la distribución de S_n no será conocida, por lo que no podrá calcularse las probabilidades exactamente. También puede suceder que el cálculo exacto sea posible en teoría pero demasiado costoso computacionalmente, incluso para una computadora. El poder de la TCL es que se aplica cuando X_i tiene casi cualquier distribución, aunque algunas distribuciones pueden requerir un n más grande para que la aproximación sea buena. Veamos algunos ejemplos.

Ejemplo 4

Un dado desparejo tiene dos caras opuestas que son menos probables que las otras cuatro. Así el 1 y el 6 tienen probabilidad $1/10$ y los otros cuatro resultados tienen probabilidad $1/5$.

Estimar la probabilidad de que en 100 lanzamientos la suma esté entre 335 y 365.

Llamemos X_i al resultado del i -ésimo lanzamiento. La f.p.p de cada X_i es

Valor de X_i	1	2	3	4	5	6
f.p.p.	1/10	2/10	2/10	2/10	2/10	1/10

por lo que

$$\mathbf{E}(X_i) = 1 \cdot \frac{1}{10} + 2 \cdot \frac{2}{10} + 3 \cdot \frac{2}{10} + 4 \cdot \frac{2}{10} + 5 \cdot \frac{2}{10} + 6 \cdot \frac{1}{10} = \frac{35}{10} = 3.5$$

La varianza la podemos calcular a partir de la siguiente tabla

Valor de X_i	1	2	3	4	5	6
f.p.p.	1/10	2/10	2/10	2/10	2/10	1/10
$(X_i - 3.5)^2$	6.25	2.25	0.25	0.25	2.25	6.25

de donde resulta $\mathbf{var}(X_i) = 2.25$.

Entonces, para $S_{100} = X_1 + \dots + X_{100}$ tenemos

$$\mathbf{E}(S_{100}) = 350, \quad \mathbf{var}(S_{100}) = 225 = (15)^2.$$

Aplicando el TCL, podemos aproximar la probabilidad por

$$\begin{aligned} \mathbf{P}(335 \leq S_{100} \leq 365) &= \mathbf{P}\left(\frac{335 - 350}{15} \leq \frac{S_{100} - 350}{15} \leq \frac{365 - 350}{15}\right) \\ &= \mathbf{P}(-1 \leq Z_{100} \leq 1) \approx \Phi(1) - \Phi(-1) \end{aligned}$$

Como $\Phi(1) - \Phi(-1) = 2\Phi(1) - 1 = 0.6826$, resulta

$$\mathbf{P}(335 \leq S_{100} \leq 365) \approx 0.6826.$$

Notar que para este dado no uniforme, es casi imposible calcular la distribución de S_{100} . Solamente con una computadora ese cálculo es razonable.

¿Cuál sería la probabilidad si el dado fuera justo?

En este caso cambia la varianza de S_{100} , ya que el promedio sigue siendo el mismo, y vale $\mathbf{var}(S_{100}) = 100 \cdot 35/12 = 291.666\dots$. Así que el cálculo da

$$\begin{aligned} \mathbf{P}(335 \leq S_{100} \leq 365) &= \mathbf{P}\left(\frac{335 - 350}{17.08} \leq \frac{S_{100} - 350}{17.08} \leq \frac{365 - 350}{17.08}\right) \\ &= \mathbf{P}(-0.88 \leq Z_{100} \leq 0.88) \approx \Phi(0.88) - \Phi(-0.88) \end{aligned}$$

Como $\Phi(0.88) - \Phi(-0.88) = 2\Phi(0.88) - 1 = 0.6212$, resulta

$$\mathbf{P}(335 \leq S_{100} \leq 365) \approx 0.6212.$$

Notar que es una probabilidad menor la anterior, lo cual es coherente pues la varianza ahora es mayor. ■

Una lotería binomial

En un país muy corrupto y mafioso la lotería se juega de manera muy extraña. Los tickets son tiras de 0's y 1's de longitud $n = 1000$, y se sortea una de las tiras. Los posibles tickets de lotería son:

0	0	0	0	0	...	0	0	0	0	0
1	0	0	0	0	...	0	0	0	0	0
0	1	0	0	0	...	0	0	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0	0	0	0	0	...	0	0	0	1	0
0	0	0	0	0	...	0	0	0	0	1
1	1	0	0	0	...	0	0	0	0	0
1	0	1	0	0	...	0	0	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0	0	0	0	0	...	0	0	1	0	1
0	0	0	0	0	...	0	0	0	1	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0	1	1	1	1	...	1	1	1	1	1
1	1	1	1	1	...	1	1	1	1	1

Para jugar los habitantes deben comprar tickets, todos los que quieran, y todos los tickets salen lo mismo.

Los tickets no son equiprobables, si no que se sortea cada dígito de la tira de forma tal que la probabilidad de que un dígito sea 1 es $p = 0.1$ y de que sea un 0 es $q = 1 - p = 0.9$.

Si fueras a comprar un solo ticket, ¿cuál comprarías? La probabilidad de cada tickets depen-

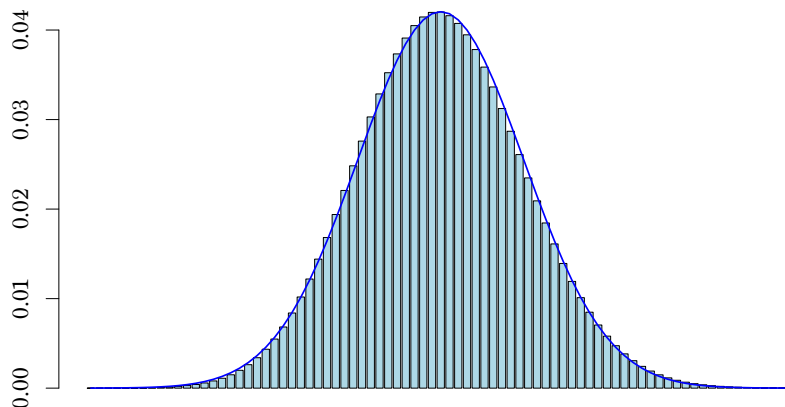


Figura 2: Aproximación normal a la binomial con $n = 1000$ y $p = 0.1$.

de solamente de la cantidad de 0's y 1's que contenga. Así, si el ticket tiene k 1's y $1000 - k$ 0's, su probabilidad es

$$p^k q^{n-k}.$$

Como $p < q$, es claro que conviene cambiar cada p por un q si queremos que la probabilidad sea máxima. Entonces, el ticket más probable es

$$00000 \dots 00000$$

que contiene solamente 0's y su probabilidad es q^n .

Sin embargo, nosotros sabemos de la LGN que con probabilidad muy alta, la cantidad de 1's en el ticket ganador es cercana al 10%. Nuestro jefe, el capo de la mafia, nos obliga a comprar una cantidad suficiente de tickets de modo que la probabilidad de ganar la lotería sea de al menos 99%.

Pero nosotros sabemos que si gastamos mucho dinero eso no le va a gustar al jefe. ¿Cuál es la cantidad mínima de plata que podemos gastar?

Consideremos la variable aleatoria S que cuenta el número de 1's en el ticket ganador. Como ya sabemos, S tiene distribución binomial de parámetros n y p . Para asegurarnos de tener el ticket ganador, podemos comprar todos los tickets que tengan cero 1's, un 1's, dos 1's, hasta que la probabilidad sea al menos 0.99.

Es decir, buscamos el valor de r_{\max} tal que

$$\mathbf{P}(S \leq r_{\max}) \geq 0.99.$$

Obviamente queremos que r_{\max} sea lo más chico posible para que el jefe no tenga que gastar demasiado dinero.

Por el TCL, la estandarización Z de S es aproximadamente normal estándar, por lo que

$$\mathbf{P}(X \leq r) = \mathbf{P}\left(Z \leq \frac{r - \mu}{\sigma}\right) \approx \Phi\left(\frac{r - \mu}{\sigma}\right),$$

en donde $\mu = \mathbf{E}(S)$ y $\sigma^2 = \mathbf{var}(S)$. Así que buscamos el menor valor de r tal que

$$\Phi\left(\frac{r-\mu}{\sigma}\right) \geq 0.99,$$

con $\mu = 100$ y $\sigma^2 = 90$. Nuevamente, de las tablas deducimos que

$$\frac{r-\mu}{\sigma} \geq 2.33,$$

o lo que es equivalente: $r \geq 2.33\sigma + \mu = 122.1$. Es decir, para asegurarnos una probabilidad de 0.99 de ganar la lotería, debemos comprar los tickets con a lo sumo $r_{\max} = 123$ cantidad de 1's.

3. Demostración informal para el caso binomial

A De Moivre ya lo cruzamos cuando estudiamos la distribución de la suma de n dados. Pero eso no fue todo lo que hizo en probabilidad. Entusiasmado con el Teorema de Bernoulli (la LGN) se puso a experimentar con monedas, y probó por primera vez una versión del TCL. Justamente, el TCL para el caso binomial se conoce como Teorema de De Moivre. Varios años después Laplace probó el teorema en su generalidad.

Cuando las variables X_i tienen distribución de Bernoulli de parámetro p , la suma S_n tiene distribución binomial de parámetros n y p . Así que en este caso tenemos una fórmula explícita para la fpp:

$$p(k) = \binom{n}{k} p^k q^{n-k}$$

en donde hemos puesto $q = 1 - p$.

La idea de De Moivre es estudiar los cocientes

$$\frac{p(k+1)}{p(k)} = \frac{p}{q} \cdot \frac{n-k}{k+1}, \quad k = 0, \dots, n-1$$

una idea que de hecho ya había usado Bernoulli para probar su LGN. Nos interesa entender estos cocientes para valores de k cercanos a np , ya que muy lejos de este valor la fpp se anula rápidamente. Pero cerca de np dichos cocientes son cercanos a 1, por lo que es mejor usar una escala logarítmica para ver lo que sucede. Es decir, definimos

$$L_k = \ln\left(\frac{p(k+1)}{p(k)}\right) = \ln\left(\frac{p}{q}\right) + \ln\left(\frac{n-k}{k+1}\right).$$

La Figura 3 muestra los valores de L_k para $n = 100$ y $p = 1/4$. Notar que en el rango de valores de k cercanos a np , en donde la fpp binomial no es prácticamente nula, los valores de L_k se agrupan muy cerca de la recta

$$y = -\frac{1}{npq}(x - np).$$

Esta recta es casi la derivada de la función

$$L(x) = \ln\left(\frac{p}{q}\right) + \ln\left(\frac{n-x}{x+1}\right)$$

en $x = np$. Digo casi porque para n grande

$$\begin{cases} L(np) = \ln\left(\frac{np}{np+1}\right) \approx 0 \\ L'(np) = -\frac{n+1}{nq(np+1)} \approx -\frac{1}{npq} \end{cases}$$

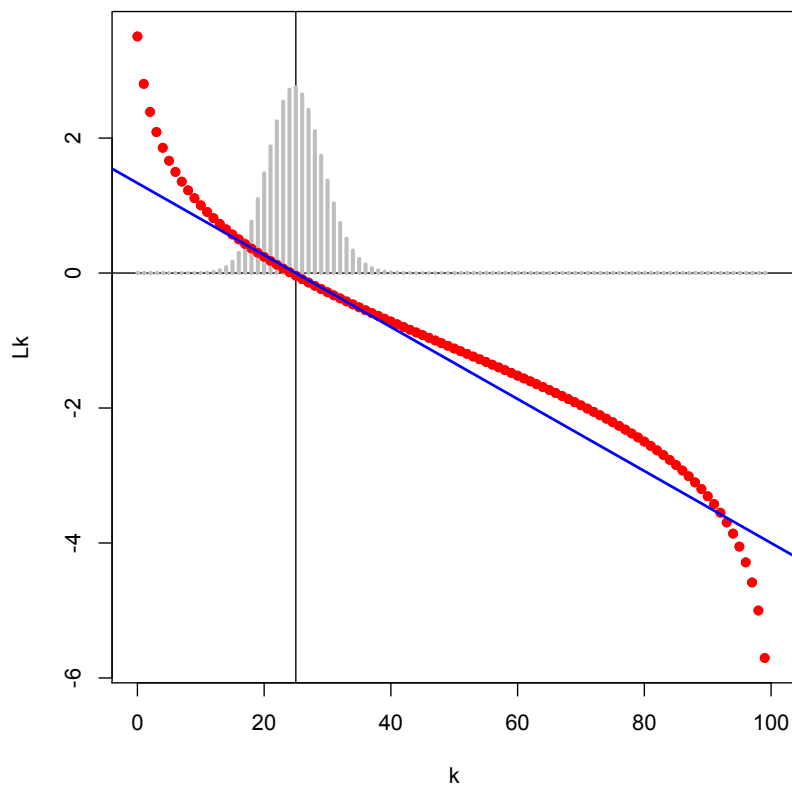


Figura 3: Los puntos en rojo representan el gráfico de L_k en función de k . Las barras en gris representan (en otra escala) la fpp binomial de parámetros $n = 100$ y $p = 1/4$. En azul se trazó la recta tangente a L_k en np .

Entonces, para el rango de valores de k cercanos a np podemos aproximar

$$L_k \approx -\frac{1}{npq}(k - np).$$

Supongamos que $k \geq np$, el argumento es similar para $k \leq np$. Notar que

$$\frac{p(k)}{p(np)} = \frac{p(k)}{p(k-1)} \cdot \frac{p(k-1)}{p(k-2)} \cdots \frac{p(np+1)}{p(np)} \approx \exp\left\{\sum_{j=np}^k L_j\right\}.$$

Pero la suma es aproximadamente

$$\sum_{j=np}^k L_j \approx \int_0^k L(x) dx = \left[-\frac{1}{2npq}(x - np)^2\right]_{np}^k = -\frac{1}{2npq}(k - np)^2.$$

En conclusión hemos probado que

$$p(k) \propto \exp \left\{ -\frac{1}{2npq} (k - np)^2 \right\} = \exp \left\{ -\frac{1}{2\sigma^2} (k - \mu)^2 \right\}$$

Esta es la idea principal detrás del TCL para el caso binomial. Si bien no es una demostración formal, porque omite varios detalles en las aproximaciones, permite entender porqué la aproximación normal es buena.

4. Normalidad asintótica de estimadores

Un estimador T_n se dice asintóticamente normal si su distribución se puede aproximar, cuando n es grande, por una distribución normal de parámetros convenientes.

Estimador asintóticamente normal

Un estimador T_n se dice asintóticamente normal si la distribución del estimador estandarizado

$$\frac{T_n - \mathbf{E}(T_n)}{\sqrt{\mathbf{var}(T_n)}}$$

se puede aproximar por una variable normal $N(0, 1)$ cuando n es grande. Esto quiere decir que para todo $t \in \mathbb{R}$, vale que

$$\lim_{n \rightarrow \infty} \mathbf{P} \left(\frac{T_n - \mathbf{E}(T_n)}{\sqrt{\mathbf{var}(T_n)}} \leq t \right) = \Phi(t)$$

en donde Φ es la función de distribución de la distribución normal estándar.

Para indicar que T_n es asintóticamente normal escribimos

$$\frac{T_n - \mathbf{E}(T_n)}{\sqrt{\mathbf{var}(T_n)}} \xrightarrow{d} N(0, 1), \text{ o incluso } \frac{T_n - \mathbf{E}(T_n)}{\sqrt{\mathbf{var}(T_n)}} \stackrel{d}{\approx} N(0, 1),$$

en donde la d nos recuerda que la convergencia es de la función de distribución.

El TCL afirma que el promedio \bar{X}_n es un estimador asintóticamente normal de la esperanza $\mu = \mathbf{E}(X)$. Este es un resultado muy importante, que nos permitirá calcular aproximaciones del error que cometemos al estimar un promedio, de la misma manera que hicimos con el margen de error para las encuestas. Recordar que como vimos antes el valor esperado del promedio es $\mathbf{E}(\bar{X}_n) = \mu$ y su varianza es $\mathbf{var}(\bar{X}_n) = \sigma^2/n$.

El promedio es un estimador asintóticamente normal (TCL)

El promedio \bar{X}_n es un estimador asintóticamente normal de la esperanza μ .

Muchas veces es necesario disponer de una versión más general del resultado anterior. Por ejemplo, el resultado sigue siendo válido si cambiamos el desvío σ (desconocido en general) por el desvío muestral S_n . Esto se conoce como el teorema de Slutsky:

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \xrightarrow{d} N(0, 1).$$

La prueba la omitimos, pero se basa en que $S_n \xrightarrow{P} \sigma$.

Teorema de Slutsky

Si T_n es un estimador consistente de σ , entonces

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{T_n} \xrightarrow{d} N(0, 1).$$

El método delta

Como hemos visto en el ejemplo de los componentes eléctricos de la clase pasada, nos interesan estimadores que se escriben como función de otros. En aquel ejemplo los tres estimadores de λ eran funciones del promedio. Sabemos que el promedio es un estimador insesgado de la esperanza, conocemos su varianza y es asintóticamente normal por el TCL. ¿Qué podemos decir entonces de una función del promedio?

El método delta sirve para mostrar que un estimador es asintóticamente insesgado, para calcular su varianza asintótica y para mostrar que es asintóticamente normal.

Sea X_1, \dots, X_n un muestreo aleatorio de X con distribución $p(x; \theta)$ que depende de un parámetro θ . Sea T_n un estimador consistente de θ y consideremos $g(T_n)$ estimador de $g(\theta)$ para una función $g: \mathbb{R} \rightarrow \mathbb{R}$ con derivada continua.

Como $T_n \xrightarrow{P} \theta$, esperamos que el comportamiento de $g(T_n)$ sea cercano al de su aproximación lineal cerca de θ , al menos para valores grandes de n . De la definición de derivada podemos escribir

$$g(T_n) = g(\theta) + g'(\theta)(T_n - \theta) + (\text{resto}).$$

Vamos a olvidarnos del resto, suponiendo que es despreciable. Si tomamos esperanza, vemos que

$$\mathbf{E}(g(T_n)) = g(\theta) + g'(\theta)\mathbf{E}(T_n - \theta),$$

que podemos reescribir como

$$\text{Sesgo}(g(T_n)) = g'(\theta)\text{Sesgo}(T_n) \quad (\text{para } n \text{ grande}).$$

Lo mismo podemos hacer para calcular el error cuadrático medio:

$$\text{ECM}(g(T_n)) = g'(\theta)^2 \text{ECM}(T_n) \quad (\text{para } n \text{ grande}).$$

Más aún, si T_n es asintóticamente insesgado, entonces $g(T_n)$ es asintóticamente insesgado y su varianza asintótica es

$$g'(\theta)^2 \mathbf{var}(T_n).$$

Notar que en este caso, usando la aproximación lineal tenemos

$$\frac{g(T_n) - g(\theta)}{\sqrt{\mathbf{var}(g(T_n))}} = \frac{g'(\theta)(T_n - \theta)}{|g'(\theta)|\sqrt{\mathbf{var}(T_n)}} = \pm \frac{T_n - \theta}{\sqrt{\mathbf{var}(T_n)}}.$$

Esto implica que si T_n es asintóticamente normal, también lo es $g(T_n)$.

El método delta

Sea T_n un estimador consistente de θ , y $g: \mathbb{R} \rightarrow \mathbb{R}$ una función con derivada continua. Entonces el estimador $g(T_n)$ de $g(\theta)$ satisface:

1. Sesgo($g(T_n)$) $\sim g'(\theta)$ Sesgo(T_n) cuando $n \rightarrow \infty$.
2. ECM($g(T_n)$) $\sim g'(\theta)^2$ ECM(T_n) cuando $n \rightarrow \infty$.

Más aún, si T_n es asintóticamente insesgado, entonces

3. $g(T_n)$ es asintóticamente insesgado.
4. La varianza asintótica de $g(T_n)$ es $\mathbf{var}(g(T_n)) \sim g'(\theta)^2 \mathbf{var}(T_n)$.

Y si T_n asintóticamente normal, entonces

5. $g(T_n)$ es asintóticamente normal.

Volvamos entonces al ejemplo de los componentes eléctricos. En el primer caso, $\hat{\lambda}_1$ es una función del promedio \bar{X}_n dada por $g(x) = 1/x$. Como \bar{X}_n es insesgado y asintóticamente normal, el método delta nos permite concluir que $\hat{\lambda}_1$ es asintóticamente insesgado y asintóticamente normal. Más aún, su varianza asintótica es

$$\mathbf{var}(\hat{\lambda}_1) \sim g'(\lambda)^2 \mathbf{var}(\bar{X}_n) = \lambda^4 \frac{\mathbf{var}(X)}{n} = \lambda^4 \frac{1/\lambda^2}{n} = \frac{\lambda^2}{n}.$$

Esto lo podemos verificar directamente ya que en este caso tenemos una fórmula exacta para la varianza.

IdC asintóticos

Disponer de un estimador T_n asintóticamente normal de un parámetro θ permite calcular el margen de error en la estimación, al menos cuando n es grande. Razonando de la misma forma que hicimos con el ejemplo de la encuesta, vemos que el intervalo

$$T_n \pm 2\mathbf{var}(T_n)^{1/2}$$

contiene al parámetro θ con probabilidad aproximada de 95%. Es decir, es un intervalo de confianza (aproximado o asintótico) al nivel 95%. Más adelante en el curso veremos en detalle como construir estos intervalos, donde por ejemplo en lugar de 2 usaremos los cuantiles de la distribución normal.

Más aún, por el Teorema de Slutsky, si la varianza de T_n es desconocida, como lo es en general pues depende usualmente del parámetro, podemos sustituirla por un estimador consistente. Por ejemplo, en el caso de los estimadores, vimos que la varianza asintótica de $\hat{\lambda}_1$ es λ^2/n . Como no conocemos λ , pues es lo que queremos estimar, no conocemos tampoco la varianza asintótica de $\hat{\lambda}_1$. Pero podemos estimarla por $\hat{\lambda}_1^2/n$, de donde el IdC asintótico queda

$$\hat{\lambda}_1 \pm 2\hat{\lambda}_1/\sqrt{n}.$$

En esta fórmula todos los términos los podemos calcular a partir de la muestra.