Aproximaciones de la distribución binomial. y distribuciones relacionadas

Pablo L. De Nápoli

Departamento de Matemática
Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires

Probabilidades y Estadística para Matemática
Segundo cuatrimestre de 2021

Bernoulli, siempre Bernoulli

Una vez más volvemos a los ensayos de Bernoulli, donde considerábamos un experimento aleatorio con dos resultados que convencionalmente se llaman

- éxito (1) con probabilidad p.
- fracaso(0) con probabilidad $q = 1 - p$.

Introdujimos las variables aleatorias de Bernoulli $X_i : \Omega^\infty \to \mathbb{R}$ dadas por

$$X_i = \begin{cases} 1 & \text{si la i-ésima realización del experimento es un éxito} \\ 0 & \text{si la i-ésima realización del experimento es un fracaso} \end{cases}$$

Las X_i son variables aleatorias discretas. También lo es el número de éxitos en n ensayos.

$$S_n = X_1 + X_2 + \ldots + X_n$$

Como ya vimos S_n tiene distribución binomial; $S_n \sim Bi(n, p)$

$$P\{S_n = k\} = b(k, n, p) = \binom{n}{k} p^k q^{n-k}, \quad 0 \leq k \leq n$$

Hoy vamos a ver dos formas para aproximar la distribución de S_n cuando n es grande.

Parte I

La aproximación de Poisson a la distribución binomial

La aproximación de Poisson

La aproximación de Poisson es una aproximación de la distribución binomial para el caso en que k es pequeño comparado con n y p es también pequeño pero $\lambda = np$ es moderado.
Empecemos desarrollando el combinatorio que aparece en la distribución binomial:

$$b(k,n,p) = \binom{n}{k} p^k q^{n-k} = \frac{n(n-1)(n-2)\ldots(n-k+1)}{k!} p^k (1-p)^{n-k}$$

Notamos que en el desarrollo del combinatorio, hay k factores en el numerador. Multiplicando y dividiendo por n^k queda:

$$b(k,n,p) = \left(1-\frac{1}{n}\right)\cdot\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\cdot\frac{(np)^k}{k!}(1-p)^{n-k}$$

Pongamos $\lambda = np$, entonces

$$b(k,n,p) = \left(1-\frac{1}{n}\right)\cdot\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\cdot\frac{\lambda^k}{k!}\left(1-\frac{\lambda}{n}\right)^{n-k}$$

La aproximación de Poisson (2)

Como

$$\lim_{n\to+\infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}$$

deducimos que si k es pequeño en comparación con n, entonces

$$b(k, n, p) \approx \frac{\lambda^k}{k!} e^{-\lambda}$$

Como formalización de esta idea, obtenemos el siguiente teorema:

Teorema (Teorema de Poisson)

Si k está fijo, y $n \to +\infty$ de modo que $\lambda = np$ permanece fijo, entonces:

$$lim_{n\to+\infty} b(k, n, p) = \frac{\lambda^k}{k!} e^{-\lambda}$$

La distribución de Poisson

Lo que obtuvimos en el límite, es otra distribución de probabilidades que se utiliza con frecuencia y se conoce como distribución de Poisson:

Definición

Sea $X : \Omega \to \mathbb{N}_0$ *una variable aleatoria entera. Diremos que* X *tiene distribución de Poisson de parámetro* $\lambda > 0$*, si*

$$P\{X = k\} = \frac{\lambda^k}{k!} e^{-\lambda}$$

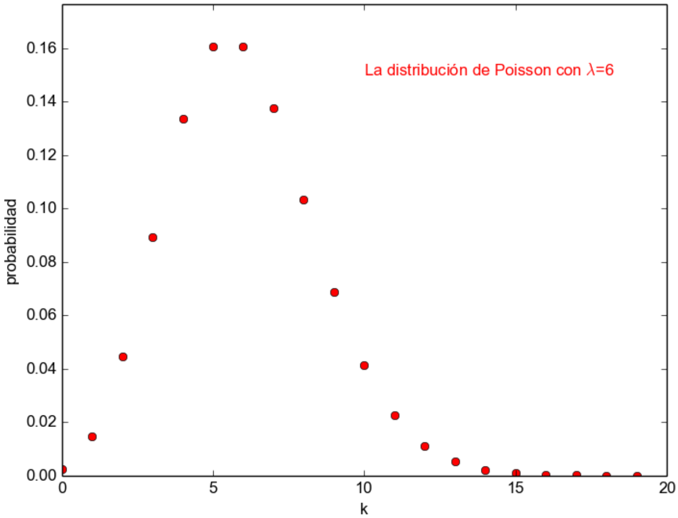
Notación: $X \sim \mathcal{P}(\lambda)$.

Hay que verificar que efectivamente tenemos una distribución de probabilidades, es decir que:

$$\sum_{k=0}^{\infty} P\{X = k\} = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} = 1$$

pero esto es inmediato, considerando el desarrollo en serie de e^{λ}.

La distribución de Poisson: gráfico



Función generatriz, esperanza y varianza

Vamos a calcular ahora la esperanza y la varianza de la distribución de Poisson: para ello utilizaremos el método de las funciones generatrices, que desarrollamos anteriormente: Si X tiene distribución de Poisson de parámetro λ, la función generatriz de su distribución de probabilidades es:

$$g_X(z) = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k z^k}{k!} = e^{-\lambda} e^{\lambda z} = e^{\lambda(z-1)}$$

Tenemos que

$$g'_X(z) = \lambda e^{\lambda(z-1)}$$

$$g''_X(z) = \lambda^2 e^{\lambda(z-1)}$$

En consecuencia deducimos que:

$$E(X) = g'_X(1) = \lambda$$

$$\text{Var}(X) = g''(1) + g'(1) - g'(1)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

Suma de variables independientes con distribución de Poisson

Otra consecuencia es la siguiente:

Proposición

Si $X \sim \mathcal{P}(\lambda_1)$, $Y \sim \mathcal{P}(\lambda_2)$ y son independientes, entonces $X + Y \sim \mathcal{P}(\lambda_1 + \lambda_2)$.

Demostración.

Como X e Y son idendepndientes

$$g_{X+Y}(z) = g_X(z) \cdot g_Y(z) = e^{\lambda_1(z-1)} e^{\lambda_2(z-1)} = e^{(\lambda_1+\lambda_2)(z-1)}$$

En consecuencia, $X + Y \sim \mathcal{P}(\lambda_1 + \lambda_2)$, ya que la distribución de probabilidades de $X + Y$ está determinada por su función generatriz. $\square$

Parte II

El teorema de De Moivre-Laplace

La Fórmula de Stirling

En muchas cuestiones del cálculo de probabilidades, resulta necesario disponer de una aproximación de $n!$ para n grande. Este es el contenido de la Fórmula de Stirling:

Teorema (Fórmula de Stirling)

$$n! \sim \sqrt{2\pi}\ n^{n+1/2} e^{-n}$$

Con más presición, se tienen las desigualdades:

$$\sqrt{2\pi}\ n^{n+1/2} e^{n} < n! < \sqrt{2\pi} e^{-n} \left(1 + \frac{1}{4n}\right)$$

Idea de la prueba:

$$\log(n!) = \sum_{k=1}^{n} logk \approx \int_2^n \log\ x\ dx$$

vean el apéndice A en mi apunte.

Teorema local de De Moivre-Laplace

Comenzamos aproximando la distribución binomial, utilizando la fórmula de Stirling

Teorema (Teorema local de De Moivre-Laplace)

$$b(k,n,p) = \frac{1}{\sqrt{2\pi npq}}\, e^{-x_k^2/2}\,(1+\beta_{n,k})$$

donde

$$x_k = \frac{k-np}{\sqrt{npq}}$$

y para $M \geq 0$,

$$\max_{|x_k|\leq M} |\beta_{n,k}| \to 0 \text{ cuando } n \to \infty \tag{1}$$

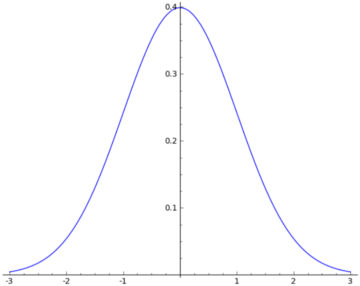
Ver la demostración en mi apunte.

Recordamos una definición: la distribución normal

Decimos que N tiene **distribución normal**, y lo notaremos $N \sim N(\mu, \sigma^2)$, si su función de densidad de probabilidad viene dada por:

$$f_N(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-(x-\mu)^2/(2\sigma^2)}$$

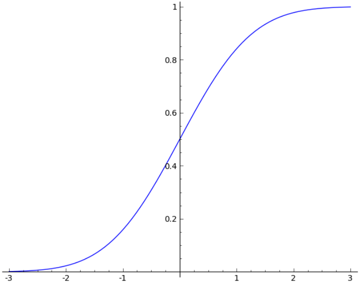
donde μ, σ son dos parámetros reales con $\sigma > 0$. El caso $\mu = 0$, $\sigma = 1$, es decir $N(0,1)$, se conoce como **distribución normal estándar**.



Recordamos una definición: la distribución normal

Si $N \sim N(0,1)$, la función de distribución acumulada de N será la función:

$$F_N(x) = P\{N \leq x\} = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-(t-\mu)^2/(2\sigma^2)}\, dt \qquad (2)$$



Un adelanto: Esperanza y varianza de variables continuas

Definición

Si X es una variable aleatoria continua con densidad de probabilidad $f(x)$ su esperanza y varianza se definen por

$$\mu_X = E[X] = \int_{\mathbb{R}} x\, f(x)\, dx \text{ si es absolutamente convergente}$$

$$\sigma_X^2 = Var[X] = \int_{\mathbb{R}} (x - \mu_X)^2\, f(x)\, dx$$

Estas fórmulas son análogas de las que vimos en la clase pasada para variables discretas

$$E[X] = \sum_i p_i x_i \text{ siempre que } \sum_i p_i |x_i| < +\infty$$

$$\mathrm{Var}(X) = \sum_i (x_i - \mu_X)^2 \cdot p_i$$

Signficado de los parámetros de la distribución normal

Proposición

Supongamos que N *se distribuye según la densidad normal* $N(\mu, \sigma^2)$. *Entonces* $E[N] = \mu$ *y* $Var(N) = \sigma^2$.

Demostración

Entonces, haciendo el cambio de variable $y = \frac{x-\mu}{\sigma}$, econtramos que

$$E[N] = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x\, e^{-(x-\mu)^2/(2\sigma^2)}\, dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\mu + \sigma y)\, e^{-y^2/2}\, dy$$

$$= \mu \left[\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2}\, dy \right] + \sigma \left[\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y\, e^{-y^2/2}\, dy \right] = \mu$$

[La segunda integral se anula, pues la densidad normal estándar es una función par]. Similarmente,

$$\mathrm{Var}(N) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-\mu)^2\, e^{-(x-\mu)^2/(2\sigma^2)}\, dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sigma^2 y^2\, e^{-y^2/2}\, dy$$

Para calcular esta integral, observamos que:

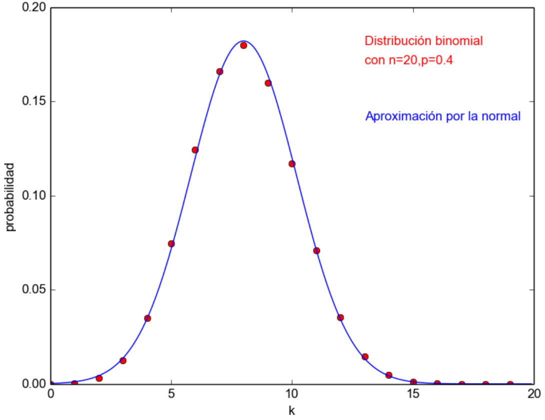
$$\left(e^{-y^2/2} \right)' = (-y) e^{-y^2/2}$$

e integramos por partes, deducimos que:

$$\mathrm{Var}(N) = \sigma^2 \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2}\, dy = \sigma^2$$

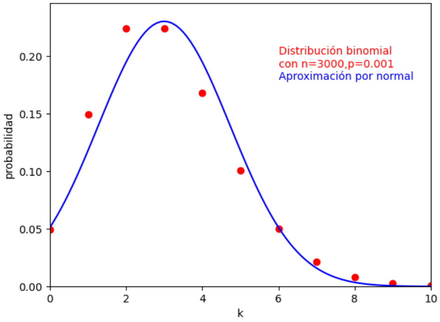
Ilustración gráfica

En el teorema Teorema local de De Moivre-Laplace estamos aproximando la distribución puntual de S_n por la función de densidad de la distribución normal (con parámetros adecuados)



Una Observación importante

La cota del error en el teorema significa la aproximación dada por el es buena en el centro de la distribución binomial, pero no en las colas de la misma. Por ejemplo, si n es grande y p es muy pequeño con np moderado. En esta situación es mejor la aproximación por la distribución de Poisson. Por simetría, tampoco es buena si p está muy cerca de 1.



¡Hagámoslo en la computadora!

Programita en Python, usando SciPy

```
from numpy import sqrt
from scipy.stats import binom, poisson, norm

def aproximar_normal(k, n, p):
        exacto = binom(n, p).pmf(k)
        mu = n * p
        q = 1 - p
        sigma = sqrt(n * p * q)
        aprox_poisson = poisson(mu).pmf(k)
        aprox_normal = norm(mu, sigma).pdf(k)

        err_rel_poisson = (aprox_poisson - exacto) / exacto
        err_rel_normal = (aprox_normal - exacto) / exacto
```

Pueden bajar mis programitas de
https://pdenapo@bitbucket.org/pdenapo/programitas-proba.git.

Ejemplo 1

Enunciado

Tiramos una moneda equilibrada 10 veces.¿Cuál es la probabilidad de obtener exactamente 3 caras?

Salida del programa

```
$python3 aproximaciones_normal.py  3 10 0.5
k= 3
n= 10
p= 0.5
Valor exacto =    0.11718750000000014
Aprox. de Poisson=   0.1403738958142805
Error rel. Poisson=   0.19785724428185886
Aprox. normal=   0.11337165224497914
Error relativo Normal= -0.03256190084284584
```

Vemos que la aproximación normal resulta buena (se equivoca en un 3,2 % mientras que la de Poisson tiene un error relativo de casi un 20 %)

Ejemplo 2

Enunciado

Una vacuna tiene una probabilidad de producir una reacción alérgica de $0,001$. Se la aplica a 2000 individuos. ¿Cuál es la probabilidad de que exactamente 3 indivuos tengan dicha reacción alérgica?.

Salida del programa

```
$ python3 aproximaciones_normal.py  3 2000 0.001
k= 3
n= 2000
p= 0.001
Valor exacto =   0.18053732803244238
Aprox. de Poisson=  0.18044704431548356
Error rel. Poisson=  -0.0005000833785609052
Aprox. normal=  0.21975057549276572
Error relativo Normal=  0.2172029900280609
```

Vemos que ahora la aproximación de Poisson es muy buena, mientras que la aproximación normal tiene un error del 21 %.

Probabilidad de que la cantidad de éxitos en un cierto intervalo

En muchas situaciones no estamos interesados en las probabilidades individuales $P\{S_n = k\}$, sino en la probabilidad de que S_n caiga en un cierto intervalo $I = (a, b]$

$$P\{a < S_n \leq b\} = \sum_{a<k\leq b} b(k, n, p) = F_{S_n}(b) - F_{S_n}(a)$$

Esto podríamos aproximarlo reemplazando la función de distribución de S_n por la de la distribución de Poisson o la de la normal (según en que régimen estemos).

Normalización de una variable aleatoria

Sea X una variable aleatoria con segundo momento finito. Entonces la variable reescaldada (o "normalizada")

$$X^* = \frac{X - E(X)}{\sqrt{\mathrm{Var}(X)}}$$

satisface que $E(X^*) = 0$ y $\mathrm{Var}(X^*) = 1$.

Sea S_n el número de éxitos en n ensayos de Bernoulli con probabilidad $p \in (0,1)$. Sabemos que S_n tiene distribución binomial que y que $E[S_n] = np$, $\mathrm{Var}(S_n) = npq$. Consideramos entonces la variable normalizada:

$$S_n^* = \frac{S_n - np}{\sqrt{npq}}$$

Nuestro objetivo es estudiar el límite de la distribución de S_n^* cuando $n \to +\infty$ distribución de S_n^* cuando $n \to +\infty$:

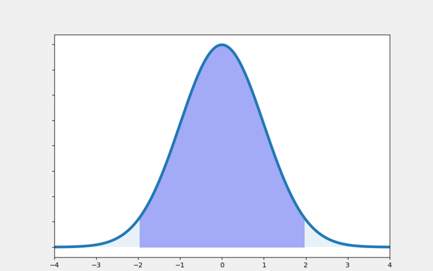
El siguiente teorema afirma que la distribución límite de la variable normalizada S_n^* está dada por la integral definida de $f_N(x)$:

Teorema (De Moivre-Laplace)

$$P\{a < S_n^* \leq b\} \to \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2}\, dx = F_N(b) - F_N(a)$$

uniformemente en a y en b cuando $n \to +\infty$.

donde F_N era la función de distribución de una variable aleatoria con distribución normal estándar. Ver la demostración en mi apunte.



Idea de la Demostración

La idea básica de la demostración es la siguiente:

$$P_n(a,b) = P\{a < S_n^* \leq b\} = \sum_{a<x_k\leq b} b(k,n,p)$$

ya que si S_n^* toma el valor x_k, entonces S_n toma el valor k.
Los puntos x_k están cada vez más próximos a medida que $n \to +\infty$, ya que

$$x_{k+1} - x_k = \frac{1}{\sqrt{npq}}$$

y por el teorema anterior $b(k,n,p) \approx f_N(x_k)(x_{k+1} - x_k)$ entonces,

$$P_n(a,b) = P\{a < S_n^* \leq b\} \approx \sum_{a<x_k\leq b} f_N(x_k)(x_{k+1} - x_k)$$

y esta es una suma de Riemann para la integral $\int_a^b f_N(x)\,dx$. Por lo tanto, conforme $n \to +\infty$, es razonable que podamos aproximar $P_n(a,b)$ por dicha integral. La demostración consiste en una formalización de esta idea.

¡Hagámoslo en la computadora!

Programita en Python, usando SciPy

```
def aproximar_normal_acumulada(k, n, p):
        exacto = binom(n, p).cdf(k)
        mu = n * p

        q=1-p
        sigma = sqrt(n * p * q)
        aprox_poisson = poisson(mu).cdf(k)
        aprox_normal = norm(mu, sigma).cdf(k)

        err_rel_poisson = (aprox_poisson - exacto) / exacto
        err_rel_normal = (aprox_normal - exacto) / exacto
```

Ejemplo 3

Enunciado

Tiramos una moneda equilibrada 30 veces.¿Cuál es la probabilidad de obtener a lo sumo 3 caras?

Salida del programa

```
python3 aproximaciones_normal_acumulada.py  3 10 0.5
k= 3
n= 30
p= 0.5
Valor exacto =              4.215165972709657e-06
Aprox. de Poisson=          0.00021137850346676174
Error rel. Poisson=         49.14713651497813
Aprox. normal=              5.885669548807499e-06
Error relativo Normal=      0.39630790030884216
```

Ejemplo 2

Enunciado

Una vacuna tiene una probabilidad de producir una reacción alérgica de $0,001$. Se la aplica a 2000 individuos. ¿Cuál es la probabilidad de que a lo sumo 3 indivuos tengan dicha reacción alérgica?.

Salida del programa

```
$ python3 aproximaciones_normal_acumulada.py  3 2000 0.001
k= 3
n= 2000
p= 0.001
Valor exacto =   0.8572137667929646
Aprox. de Poisson=  0.857123460498547
Error rel. Poisson=  -0.00010534862821369182
Aprox. normal=  0.7603598554307263
Error relativo Normal=  -0.1129868827522349
```

Vemos que ahora la aproximación de Poisson es muy buena, mientras que la aproximación normal tiene un error del $11\,\%$.

Parte III

Una Aplicación a la Estadística

Una aplicación a la estadística

Consideremos por ejemplo, una encuesta electoral para una elección donde participan dos candidatos A y B, y supongamos que cada persona puede votar por uno de ellos (y para simplificar que no hay votos en blanco). Podemos modelizar esto utilizando la distribución binomial, para ello imaginemos un experimento aleatorio donde se elige una persona al azar y se le pregunta por quien vota. Y llamemos p a la probabilidad de que vote por A ("éxito") y $q = 1 - p$ a la probabilidad de que vote por B. Alternativamente, podemos pensar que tenemos una elección en la que participan varios candidatos y que nos interesa medir la intención de voto de un determinado candidato A. En este caso, consideramos el experimento aleatorio que consiste en elegir una persona al azar, preguntarle por quien vota, y hay dos resultados posibles que nos interesan: si vota por A (con probabilidad p) o si no vota por A con probabilidad q=1-p.

Una aplicación a la estadística(2)

Nuestro objetivo es estimar la probabilidad desconocida p. Como resulta extraordinariamente costoso y complicado preguntarle a cada votante del padrón electoral por quién piensa votar, lo que suele hacerse es elegir una muestra, digamos formada por n personas. Entonces, conforme a la ley de los grandes números, si llamamos S_n a la cantidad de personas de la muestra que votan por el candidato A, podemos aproximar la probabilidad desconocida p por la frecuencia:

$$f_n = \frac{S_n}{n}$$

observada en la muestra (Estamos suponiendo que las elecciones de las distintas personas pueden considerarse independientes unas de otras, de modo que la elección de n personas encuestadas, puede considerarse como realizar n ensayos de Bernoulli, y la distribución de S_n sea dada por la distribución binomial.)

Una aplicación a la estadística (3)

Otro ejemplo análogo se da en el control de calidad en un proceso industrial. Por ejemplo, imaginemos que tenemos un lote de 10.000 lamparitas y queremos saber cuantas están falladas. Llamemos p a la probabilidad de que una lamparita elegida al azar funcione, y $q = 1 - p$ a la probabilidad de que esté fallada. Nuevamente, sería extraordinariamente costoso probar una por una las lamparitas, por lo que se hace es elegir una muestra, y aproximar p por la frecuencia f_n observada en la muestra.

Una aplicación a la estadística (4)

Una pregunta fundamental es entonces: ¿Cómo elegir el tamaño de la muestra?. Para ello, elegimos un margen de error ε, y un nivel de confianza $1-\alpha$ donde ε y α son números pequeños, y nos proponemos elegir el tamaño de la muestra de modo que podamos asegurar que la probabilidad de que f_n diste de p como mucho en ε es por lo menos $1-\alpha$, o sea:

$$P\{|f_n - p| \leq \varepsilon\} \geq 1 - \alpha$$

Por ejemplo: supongamos que queremos que muestra encuesta (o control de calidad) se equivoque como mucho en un $2\,\%$ en el $95\,\%$ de las veces que realizamos la encuesta. Entonces, elegimos $\varepsilon = 0,02$ y $\alpha = 0,05$.

Elegimos entonces x_α de modo que:

$$F_N(-x_\alpha) = \frac{\alpha}{2} \text{ o sea } x_\alpha = -F_N^{-1}\left(\frac{\alpha}{2}\right)$$

donde F_N es la función de distribución normal estándar . Por la simetría de la curva normal,

$$F_N(x_\alpha) = 1 - \frac{\alpha}{2}$$

Llamando S_n^* a la variable normalizada dada por (24), por el teorema de De Moivre Laplace:

$$P\left\{-x_\alpha \leq S_n^* \leq x_\alpha\right\} \approx \frac{1}{\sqrt{2\pi}} \int_{-x_\alpha}^{x_\alpha} e^{-x^2/2}\, dx = F_N(x_\alpha) - F_N(-x_\alpha) = 1 - \alpha$$

si n es suficientemente grande.

Una aplicación a la estadística (5)

Esta relación dice que con probabilidad $1-\alpha$ podemos asegurar que p está en el intervalo:

$$I_\alpha = \left[\frac{S_n}{n} - \ x_\alpha \sqrt{\frac{pq}{n}}, \frac{S_n}{n} + x_\alpha \ \sqrt{\frac{pq}{n}} \right]$$

I_α se llama un intervalo de confianza (asintótico) para p de nivel de confianza $1-\alpha$. En realidad en esta forma, esta relación no resulta todavía muy útil ya que no conocemos p y entonces tampoco conocemos el ancho del intervalo I_α. Pero podemos observar que:

$$pq = p(1-p) \leq \frac{1}{4} \ \forall \ p \in [0,1]$$

En consecuencia, podemos asegurar que

$$I_\alpha \subset \left[\frac{S_n}{n} - \ x_\alpha \frac{1}{2\sqrt{n}}, \frac{S_n}{n} + x_\alpha \ \frac{1}{2\sqrt{n}} \right]$$

y que (si n es grande):

$$P \left\{ \left| \frac{S_n}{n} - p \right| \leq x_\alpha \ \frac{1}{2\sqrt{n}} \right\} \geq 1-\alpha$$

En consecuencia, si queremos que se cumpla nuestro objetivo de que

$$P\{|f_n - p| \leq \varepsilon\} \geq 1 - \alpha$$

debemos elegir n para que:

$$x_\alpha \, \frac{1}{2\sqrt{n}} \leq \varepsilon$$

o sea:

$$n \geq n_0 = \left(\frac{x_\alpha}{2\varepsilon}\right)^2$$

Esta relación nos dice cuál es el tamaño (mínimo) de la muestra que necesitamos para poder garantizar un determinado margen de error con un determinado nivel de confianza. Por ejemplo, si $\alpha = 0,05$ y $\varepsilon = 0{,}02$, obtenemos que: $x_\alpha = 1,96$ y $n \geq 2401$.
Observación: Notamos que cuando $\alpha \to 0$, $x_\alpha \to +\infty$ por lo que $n_0 \to +\infty$.

Parte IV

La Distribución Multinomial

La Distribución Multinomial

Es una generalización de la distribución binomial donde consideramos experimentos con muchos varios posibles, en lugar de un experimento con sólo dos resultados. Consideramos un experimento con N resultados posibles, y supongamos que la probabilidad de que ocurra el i-ésimo resultado en una realización del experimento es p_i, de modo que:

$$\sum_{i=1}^{N} p_i = 1$$

Supongamos que repetimos el experimento n veces en condiciones independientes, y llamemos X_i a la cantidad de veces que ocurre el i-ésimo resultado, de modo que:

$$X_1 + X_2 + \ldots + X_N = n$$

Entonces, la distribución de probabilidades conjunta de las X_i viene dada por:

$$P\{X_1 = k_1, X_2 = k_2, \ldots, X_N = k_N\} = \frac{n!}{k_1! k_2 \ldots k_N!} \, p_1^{k_1} p_2^{k_2} \ldots p_N^{k_N}$$

si $k_1 + k_2 + \ldots + k_N = N$ (y cero en caso contrario). Notamos que $X = (X_1, X_2, \ldots, X_N)$ es un vector aleatorio N-dimensional.

La Distribución Multinomial (2)

Notación: $X \sim \mathcal{M}(n, p_1, p_2, \ldots, p_N)$

Esta distribución recibe este nombre, debido a su relación con el desarrollo multinomial:

$$(x_1 + x_2 + \ldots + x_N)^n = \sum_{\substack{k_N : k_1 + k_2 + \ldots + k_N = n \\ 0 \leq k_i \leq n}} \frac{n!}{k_1! k_2 \ldots k_N!} \, x_1^{k_1} x_2^{k_2} \ldots x_N^{k_N}$$

(Tomando $x_i = p_i$ se ve que las probabilidades suman 1, por lo que se trata efectivamente de una distribución de probabilidades).

Distribuciones Marginales

Una propiedad interesante de la distribución multinomial es que las distribuciones de cada una de las X_i por separado (distribuciones marginales) son binomiales:

Proposición

Si $X \sim \mathcal{M}(n, p_1, p_2, \ldots, p_N)$, *entonces*

$$X_i \sim Bi(n, p_i) \quad 0 \leq i \leq N$$

Demostración

Por simetría, basta verlo para la distribución de X_1. Si $0 \leq k_1 \leq n$,

$$
\begin{aligned}
P\{X_1 = k_1\} &= \sum_{\substack{k_N : k_2 + \ldots + k_N = n - k_1 \\ 0 \leq k_i \leq n}} P\{X_1 = k_1, X_2 = k_2, \ldots, X_N = k_N\} \\
&= \sum_{\substack{k_N : k_2 + \ldots + k_N = n - k_1 \\ 0 \leq k_i \leq n}} \frac{n!}{k_1! k_2 \ldots k_N!} \, p_1^{k_1} p_2^{k_2} \ldots p_N^{k_N} \\
&= \frac{n!}{k_1!(n-k_1)!} \, p_1^{k_1} \sum_{\substack{k_N : k_2 + \ldots + k_N = n - k_1 \\ 0 \leq k_i \leq n}} \frac{(n-k_1)!}{k_2! \ldots k_N!} \, p_2^{k_2} \ldots p_N^{k_N} \\
&= \frac{n!}{k_1!(n-k_1)!} \, p_1^{k_1} (p_2 + p_3 + \ldots + p_N)^{n-k_1} \\
&= \frac{n!}{k_1!(n-k_1)!} \, p_1^{k_1} (1 - p_1)^{n-k_1}
\end{aligned}
$$

luego

$$X_1 \sim \text{Bi}(n, p_1)$$