Como aparece la campana de Gauss

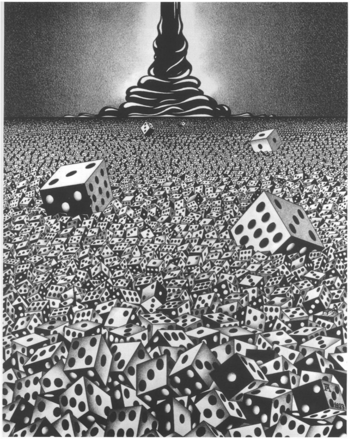

Figura: Otro dibujo de Anatoli Fomenko

Probabilidad - Clase 10
Teorema integral de De Moivre-Laplace (II)

Ernesto Mordecki

Facultad de Ciencias, Universidad de la República. Montevideo, Uruguay

Curso de la Licenciatura en Matemática - 2020

Contenidos

Previos

- Consideremos $(\Omega_n, \mathcal{A}_n, \mathbf{P})$ una serie de n experimentos de Bernoulli
- Sea $\mu(\omega)$ la cantidad de éxitos en los n experimentos.
- Queremos obtener una expresión aproximada, para la probabilidad de un suceso de la forma

$$\{\omega \in \Omega_n \colon \alpha < \mu(\omega) \leq \beta\}$$

  donde $\alpha < \beta$ son arbitrarios.
- Designamos

$$\mathbf{P}(\alpha < \mu \leq \beta) = \mathbf{P}(\{\omega \in \Omega_n \colon \alpha < \mu(\omega) \leq \beta\}).$$

Teorema (Teorema límite integral de De Moivre–Laplace)

- *Consideremos n experimentos independientes, con probabilidad de éxito p $(0 < p < 1)$, $q = 1 - p$.*
- *Sea μ la cantidad de éxitos que ocurren en esta serie.*
- *Sean a, b dos números que verifican $-\infty \leq a < b \leq \infty$ (incluimos $a = -\infty$ y $b = \infty$).*
- *Entonces, si $n \to \infty$,*

$$\mathbf{P}\left(a < \frac{\mu - np}{\sqrt{npq}} \leq b\right) - \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx \to 0, \tag{1}$$

uniformemente, para todos los valores de a, b considerados.

Una manera alternativa de escribir la convergencia que tiene lugar en (1), es, para $n \to \infty$, tenemos

$$\sup_{-\infty \leq a < b \leq \infty} \left| \mathbf{P}\left(a < \frac{\mu - np}{\sqrt{npq}} \leq b\right) - \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx \right| \to 0,$$

La demostración de este teorema se basa en la aplicación del Teorema límite **local** de De Moivre–Laplace y en el siguiente resultado. Sea para eso

$$[x] \text{ es la parte entera de } x,$$

$$\text{es decir el mayor entero que no supera a } x.$$

Lema

Es válida la desigualdad

$$P_n(m) \leq P_n\big([(n+1)p]\big),$$

para todo m $(0 \leq m \leq n)$.

Demostración del lema. Según la fórmula binomial tenemos

$$\frac{P_n(m+1)}{P_n(m)} = \frac{\binom{n}{m+1}p^{m+1}q^{n-m-1}}{\binom{n}{m}p^m q^{n-m}}$$
$$= \frac{p}{q} \times \frac{n!m!(n-m)!}{(m+1)!(n-m-1)!n!} = \frac{p}{q} \times \frac{n-m}{m+1}.$$

Entonces

$$P_n(m+1) > P_n(m) \text{ si y sólo si } (n-m)p > (m+1)q$$

$$np - mp > (m+1)(1-p) = m + 1 - mp - p$$
$$(n+1)p - 1 > m.$$

Dando vuelta:

$P_n(m+1) < P_n(m)$(crece) cuando y solo cuando$m < (n+1)p-1$

Conclusión:

- Si $(n+1)p$ no es natural, entonces $P_n([(n+1)p])$ es el máximo (porque $m_0 = [(n+1)p]$ es el mayor entero en la zona donde crece.

- Si $(n+1)p$ es natural, entonces el máximo es $P_n(m_0-1) = P_n(m_0)$ (hay dos valores) porque las probabilidades en nuestra cuenta son iguales.

Eso concluye la demostración de $P_n(m) \leq P_n([(n+1)p])$.

Demostración del teorema integral.
Introduzcamos las notaciones

- $P_n(a,b) = \mathbf{P}\left(a < \frac{\mu - np}{\sqrt{npq}} \leq b\right)$,
- $x_{n,m} = \frac{m-np}{\sqrt{npq}}$ (como la clase pasada).

Demostramos la convergencia primero con a y b finitos.

Recordemos que $x_{n,m} = \frac{m-np}{\sqrt{npq}}$. Calculamos

$$P_n(a,b) = \sum_m P_n(m), \tag{2}$$

donde la suma se efectúa en los valores de m, para los cuales $a < x_{n,m} \leq b$. Tenemos $x_{n,m+1} - x_{n,m} = 1/\sqrt{npq}$, y además

$$x_{n,0} = -\frac{np}{\sqrt{npq}} \to -\infty \ (n \to \infty),$$
$$x_{n,n} = \frac{n-np}{\sqrt{npq}} \to +\infty \ (n \to \infty).$$

Tenemos que llegar a una integral que se aproxima por una suma de Riemann. Sea

$$\Pi_n(x) = \begin{cases} 0, & \text{si } x \le x_{n,0}, \text{ ó } x > x_{n,n} + \frac{1}{\sqrt{npq}} \\ \sqrt{npq}P_n(m), & \text{si } x_{n,m} < x \le x_{n,m+1} \; (m = 0, 1, \dots, n). \end{cases}$$

y designamos mediante $\underline{m}$ y $\overline{m}$, a los (únicos) naturales, que verifican las condiciones

$$a < x_{n,\underline{m}} \le a + \frac{1}{\sqrt{npq}}, \qquad b < x_{n,\overline{m}} \le b + \frac{1}{\sqrt{npq}}.$$

$$\sqrt{npq} \quad \text{es “}1/\Delta x\text{”}$$

De esa forma

$$\int_{x_{n,m}}^{x_{n,m+1}} \Pi_n(x)dx = \sqrt{npq}P_n(m)(x_{n,m+1} - x_{n,m}) = P_n(m),$$

para $m = 0, 1, \ldots, n$, y la igualdad (2) puede ser escrita, como

$$P_n(a, b) = \sum_m \int_{x_{n,m}}^{x_{n,m+1}} \Pi_n(x)dx = \int_{x_{n,\underline{m}}}^{x_{n,\overline{m}}} \Pi_n(x)dx. \qquad (3)$$

En síntesis, hay que probar que la suma converge a la integral en $[a, b]$ de

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \qquad \text{Densidad normal}$$

. Esto se hace en dos etapas:

(a) Aproximación del intervalo: se observa que la diferencia de integral de la función $\Pi_n(x)$ en los intervalos $[a, b]$ y $[x_{n,\underline{m}}, x_{n,\overline{m}}]$ es arbitrariamente pequeña, para valores grandes de n (mediante el Lema)

(b) Aproximación de la integral: se aproximan las integrales en el intervalo $[a, b]$ de las funciones $\Pi_n(x)$ y $\varphi(x)$ (aquí se utiliza el teorema local).

Comenzamos entonces dividiendo en tres el intervalo de integración en (3), obteniendo

$$P_n(a,b) = \int_a^b \Pi_n(x)dx - \int_a^{x_{n,\underline{m}}} \Pi_n(x)dx + \int_b^{x_{n,\overline{m}}} \Pi_n(x)dx,$$

de donde resulta, que

$$\left| P_n(a,b) - \int_a^b \Pi_n(x)dx \right| \leq \int_a^{x_{n,\underline{m}}} \Pi_n(x)dx + \int_b^{x_{n,\overline{m}}} \Pi_n(x)dx,$$

Los intervalos de integración son acotados por

$$\Delta x = \frac{1}{\sqrt{npq}}$$

Acotamos el integrando

$$\Pi_n(x) \leq \sqrt{npq} P_n(m_0), \quad \forall x, \text{ (Lema)}$$

donde $m_0 = [(n+1)p]$. Para aproximar por el teorema local, $x_{n,m_0} = (m_0 - np)/\sqrt{npq}$. Tenemos

$$(n+1)p - 1 \leq m_0 \leq (n+1)p$$

$$(n+1)p - 1 - np \leq m_0 - np \leq (n+1)p - np$$

$$\frac{(n+1)p - 1 - np}{\sqrt{npq}} \leq x_{n,m_0} \leq \frac{(n+1)p - np}{\sqrt{npq}}$$

$$\frac{p-1}{\sqrt{npq}} \leq x_{n,m_0} \leq \frac{p}{\sqrt{npq}}$$

Entonces, si n es suficientemente grande

$$-1 \leq x_{n,m_0} \leq 1.$$

Aplicamos el teorema local con $C = 1$ y tenemos

$$\frac{\sqrt{npq}P_n(m_0)}{\frac{1}{\sqrt{2\pi}}e^{-x_{n,m_0}^2/2}} \to 1 \quad \text{cuando} \quad n \to \infty.$$

De aquí, como $\varphi(x) \leq 1/\sqrt{2\pi}$, obtenemos

$$\Pi_n(x) \leq \sqrt{npq}P_n(m_0) < \frac{2}{\sqrt{2\pi}}$$

para todo n suficientemente grande.

Teniendo en cuenta esta acotación, y la fórmula (15), obtenemos

$$\begin{aligned}\left|P_n(a,b) - \int_a^b \Pi_n(x)dx\right| &\leq \sqrt{npq}P_n(m_0)(x_{n,\underline{m}} - a + x_{n,\overline{m}} - b) \\ &\leq \sqrt{npq}P_n(m_0)\frac{2}{\sqrt{npq}} < \frac{4}{\sqrt{2\pi npq}},\end{aligned} \tag{4}$$

para todo n suficientemente grande (esto concluye la etapa (a) Aproximación del intervalo).

Demostremos ahora, que

$$\Pi_n(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}(1 + r_n(x)),$$

donde el resto $r_n(x)$ converge uniformemente a cero, es decir

$$\sup_{a \leq x \leq b} |r_n(x)| \to 0 \quad (n \to \infty). \tag{5}$$

Sea x fijo en el intervalo $[a, b]$. Si $x_{n,m} < x \leq x_{n,m+1}$, por el teorema local, tenemos

$$\Pi_n(x) = \sqrt{npq} P_n(m) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x_{n,m}^2}{2}} (1 + \gamma_n),$$

donde $\gamma_n \to 0$ ($n \to \infty$) uniformemente con respecto a m.

Sumando y restando en el exponente, tenemos

$$\Pi_n(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} e^{-(x^2 - x_{n,m}^2)/2} (1 + \gamma_n).$$

Es claro, que

$$\frac{|x^2 - x_{n,m}^2|}{2} = \frac{|(x - x_{n,m})(x + x_{n,m})|}{2} \leq \frac{1}{2\sqrt{npq}} 2 \max(|a|, |b|) \to 0,$$

si $n \to \infty$. Poniendo $r_n(x) = e^{-(x^2 - x_{n,m}^2)/2}(1 + \gamma_n) - 1$, la fórmula anterior es

$$\Pi_n(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} (1 + r_n(x)),$$

y se verifica la convergencia uniforme del resto.

Para concluir la etapa (b) tenemos

$$
\begin{aligned}
&\left|P_n(a,b) - \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx\right| \\
&\leq \left|P_n(a,b) - \int_a^b \Pi_n(x)dx\right| + \left|\int_a^b \Pi_n(x)dx - \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx\right| \\
&\qquad\leq \frac{4}{\sqrt{2\pi npq}} + \frac{1}{\sqrt{2\pi}} \sup_{a \leq x \leq b} |r_n(x)| \int_a^b e^{-\frac{x^2}{2}} dx \to 0,
\end{aligned}
$$

si $n \to \infty$.

Faltan los extremos a, b infinitos.

Termina el repaso

Demostremos ahora el teorema 1 sin supuestos adicionales sobre a y b. Tiene lugar la igualdad

$$\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\,dx = \sqrt{2\pi}. \tag{6}$$

Sea ε un número positivo arbitrario. Sea c una constante positiva que verifica la condición

$$\frac{1}{\sqrt{2\pi}} \int_{\{|x|>c\}} e^{-\frac{x^2}{2}}\,dx < \varepsilon. \tag{7}$$

Según se vió en la primera parte de la demostración, tiene lugar la acotación

$$\left| P_n(-c,c) - \frac{1}{\sqrt{2\pi}} \int_{-c}^{c} e^{-\frac{x^2}{2}}\,dx \right| < \varepsilon. \tag{8}$$

si n es suficientemente grande.

Además,

$$\mathbf{P}\left(\frac{\mu - np}{\sqrt{npq}} \leq -c\right) + \mathbf{P}\left(\frac{\mu - np}{\sqrt{npq}} > c\right) = 1 - P_n(-c, c)$$
$$= \frac{1}{\sqrt{2\pi}}\int_{-c}^{c} e^{-\frac{x^2}{2}} dx - P_n(-c, c) + \frac{1}{\sqrt{2\pi}}\int_{\{|x|>c\}} e^{-\frac{x^2}{2}} dx < 2\varepsilon. \tag{9}$$

para todo n suficientemente grande, en vista de la elección de c (fórmula 7) y la acotación (8).

Consideremos el caso en el que $a \leq -c < b \leq c$ (los otros dos casos a considerar son análogos). Tenemos

$$
\begin{aligned}
&\left|P_n(a,b) - \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx\right| \\
&= \left|P_n(a,-c) + P_n(-c,b) - \frac{1}{\sqrt{2\pi}} \int_a^{-c} e^{-\frac{x^2}{2}} dx - \frac{1}{\sqrt{2\pi}} \int_{-c}^b e^{-\frac{x^2}{2}} dx\right| \\
&\leq \left|P_n(a,-c)\right| + \left|\frac{1}{\sqrt{2\pi}} \int_a^{-c} e^{-\frac{x^2}{2}} dx\right| + \left|P_n(-c,b) - \frac{1}{\sqrt{2\pi}} \int_{-c}^b e^{-\frac{x^2}{2}} dx\right| \\
&\qquad < 2\varepsilon + \varepsilon + \varepsilon = 4\varepsilon,
\end{aligned}
$$

para todo n suficientemente grande, teniendo en cuenta las acotaciones (9), (7) y (8). Esto concluye la demostración del teorema.

Aplicación

Del teorema 1 se obtiene, que para n suficientemente grande, tiene lugar la identidad aproximada

$$\mathbf{P}\left(a < \frac{\mu - np}{\sqrt{npq}} \leq b\right) \approx \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx = \Phi(b) - \Phi(a),$$

si introducimos la función

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt, \tag{10}$$

definida para todo x real, la *función de distribución normal*

En R

- `pnorm(x)` da $\Phi(x)$
- `dnorm(x)` da $\varphi(x)$
- `rnorm(n)` **simula** n resultados de un experimento normal estándar, es decir, n valores, cada uno de los cuales tiene probabilidades que se calculan con las áreas de $\varphi(x)$.
- `qnorm(p)` da la función inversa de $\Phi(x)$:

```
> pnorm(2)
[1] 0.9772499
> qnorm(0.9772499)
[1] 2.000001
>
```

El teorema nos permite calcular $\mathbf{P}(\alpha \leq \mu \leq \beta)$, para α y β dados. Es claro, que

$$\mathbf{P}(\alpha \leq \mu \leq \beta) = \mathbf{P}\left(\frac{\alpha - np}{\sqrt{npq}} \leq \frac{\mu - np}{\sqrt{npq}} \leq \frac{\beta - np}{\sqrt{npq}}\right),$$

de donde

$$\mathbf{P}(\alpha \leq \mu \leq \beta) \approx \Phi\Big(\frac{\beta - np}{\sqrt{npq}}\Big) - \Phi\Big(\frac{\alpha - np}{\sqrt{npq}}\Big).$$

Ejemplo

Calcular $\mathbf{P}(\mu \leq 230)$.

- Tenemos $p = 0{,}02$, $q = 1 - p = 0{,}98$,
- $\alpha = 0$, $\beta = 230$, $n = 10000$,
- $$\mathbf{P}(\mu \leq 230) \approx \Phi\left(\frac{230 - 200}{14}\right) - \Phi\left(\frac{-200}{14}\right)$$

```
> pnorm((230-200)/14)- pnorm(-200/14)
[1] 0.9839377
>
```

Resta ver como hacer con las normales con parámetros (a, σ).

Sheldon usa la fórmula de Bayes

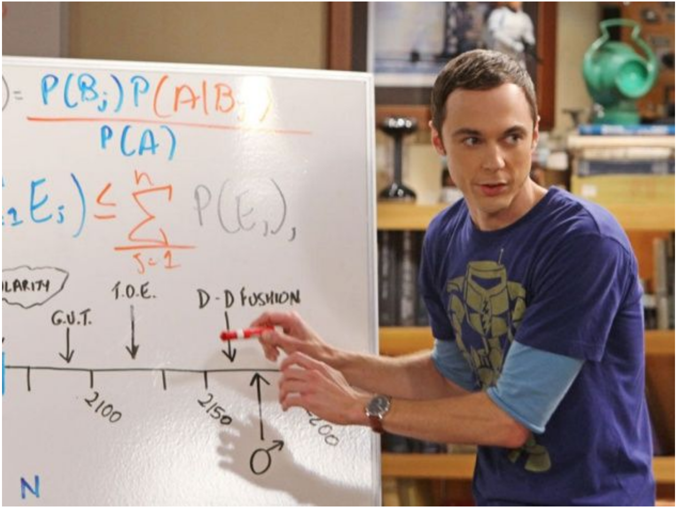

Figura: ¡Que pasen lindo!