

# La Ley de Grandes Números

Pablo Lessa

9 de octubre de 2014

## 1. Ley de Grandes Números

Te hago una pregunta personal: Si estás jugando a la ruleta apostando en cada turno por negro o rojo y ves que sale 6 veces seguidas el negro, ¿Cuál es tu siguiente apuesta? ¿Negro o rojo?

### 1.1. La historia de Joseph Jagger

Resulta que entre 1830 y 1892 en Inglaterra vivió un tipo llamado Joseph Jagger. No fundó una banda ni se hizo mundialmente famoso por su estilo de baile. Sino que comenzó su vida trabajando en una fábrica textil, presumiblemente (aunque no hice una investigación bibliográfica profunda) arreglando o manteniendo las maquinas (telares mecánicos, supongo).

A los 40 años de edad, ya un tipo maduro, se le ocurrió que las ruletas podrían tener defectos mecánicos que hicieran que los resultados de las tiradas no fueran totalmente equiprobables. Entonces, dando el siguiente paso, le pagó a unos amigos para que anotaran los resultados de todas las tiradas de un día de cada una de las mesas de ruleta del famoso casino de Monte Carlo.

Luego de analizar los datos descubrió que había una mesa de ruleta en la cual los números 7,8,9,17,18,19, 28, y 29, salieron más frecuentemente de lo esperado durante el día.

Entonces al día siguiente fue al casino de Monte Carlo y apostó a estos números repetidas veces. Durante los siguientes tres días ganó 14 mil libras que equivalen más o menos a un millón de dólares del día de hoy. Además se acumuló un grupo de apostadores que empezaron a copiar sus apuestas, también ganando dinero.

Esto obviamente no le gustó al casino. Así que en la noche cambiaron todas las mesas de lugar. Al día siguiente Joseph empezó a perder guita, pero después de un rato se acordó que ‘su mesa’ tenía una marca distintiva, la buscó por el salón, la encontró y empezó a ganar de nuevo.

Esa noche el casino desarmó todas las mesas, aceito todas las piezas mecánicas, y las volvió a armar. Al día siguiente Joseph (y sus seguidores) perdieron un poco de dinero. A lo cual Joseph decidió retirarse con lo ganado (unas 64 mil

libras en total) y nunca más volvió a apostar. Compró un montón de propiedades y vivió como un hombre rico hasta que su muerte una veintena de años más tarde.

## 1.2. La moraleja

Es un hecho empírico que para algunos tipos de ‘experimentos’ (como tirar un dado, o hacer girar una ruleta) los resultados individuales parecen impredecibles pero al repetir el experimento se presentan regularidades estadísticas que sí son predecibles.

Este tipo de idea no sólo es útil para los juegos de azar y otras frivolidades. La mecánica cuántica y teorías físicas relacionadas son de este tipo (no dan predicciones sobre experimentos individuales sino que predicen propiedades estadísticas de largas tiras de repeticiones de dichos experimentos).

## 1.3. La receta de Kolmogorov

Para los matemáticos la sección anterior plantea el siguiente problema: ¿Cómo formalizar esta observación empírica para convertirla en matemática (con definiciones precisas y argumentos claros)?

Luego de una larga historia se llegó a una receta fundacional para la teoría de probabilidades que consta en tres pasos:

1. Construir un espacio  $\Omega$  donde los puntos representan todos los resultados posibles de la secuencia de experimentos a realizar.
2. Construir una medida de probabilidad  $\mu$  en  $\Omega$  que modeliza el comportamiento del azar.
3. Intentar demostrar teoremas sobre el comportamiento de los puntos de un subconjunto de  $\Omega$  cuya medida para  $\mu$  sea 1.

## 1.4. El caso de la moneda: Paso 1

Vamos a seguir la receta de Kolmogorov para modelar una sucesión (infinita) de tiradas de moneda. Se supone que si la moneda sale de un lado anotamos 1 como resultado del experimento y 0 si sale el otro lado.

En este caso el candidato natural para espacio  $\Omega$  es el espacio de sucesiones de ceros y unos i.e.

$$\Omega = \{0, 1\}^{\mathbb{N}} = \{\omega = (\omega_1, \omega_2, \dots) : \omega_i \in \{0, 1\} \forall i\}.$$

En este espacio consideramos una distancia  $d$  definida como

$$d(\omega, \omega') = \sum_{\{n: \omega_n \neq \omega'_n\}} 2^{-n}.$$

**Ejercicio 1.** *Demostrar que  $(\Omega, d)$  es un espacio métrico separable y completo.*

**Ejercicio 2.** *Mostrar que  $(\Omega, d)$  es homeomorfo al habitual conjunto Cantor ternario en  $[0, 1]$ , i.e.  $C = \bigcap C_n$  donde  $C_0 = [0, 1]$ ,  $C_1 = [0, 1/3] \cup [2/3, 1]$ ,  $C_2 = [0, 1/9] \cup [2/9, 1/3] \cup [2/3, 7/9] \cup [8/9, 1], \dots$*

## 1.5. El caso de la moneda: Paso 2

### 1.5.1. $\sigma$ -álgebras

Una probabilidad es una función  $\mu$  que asigna a algunos subconjuntos de  $\Omega$  un número entre 0 y 1 y que cumple ciertas reglas razonables. Por motivos técnicos en los cuales prefiero no entrar no se puede definir una tal  $\mu$  sobre todos los subconjuntos de  $\omega$ . Esto complica un poco la teoría (estamos obligados a introducir las temidas  $\sigma$ -álgebras) pero no debería impedir a nadie entenderla.

**Definición 1** ( $\sigma$ -álgebra). *Una  $\sigma$ -álgebra  $\mathcal{F}$  es un subconjunto de  $2^\Omega$  (i.e. el conjunto de todos los subconjuntos de  $\Omega$ ) que es no vacío, y cerrado bajo complementos y uniones numerables (i.e. si  $A \in \mathcal{F}$  entonces  $\Omega \setminus A$  también, y si  $A_1, \dots, A_n, \dots \in \mathcal{F}$  entonces  $\bigcup A_n$  también).*

**Ejercicio 3.** *Mostrar que si  $A \subset 2^\Omega$  existe una mínima  $\sigma$ -álgebra que contiene a  $A$ . Es decir, la intersección  $\sigma(A)$  de todas las  $\sigma$ -álgebras que contienen a  $A$  es una  $\sigma$ -álgebra. Notar que  $2^\Omega$  es una  $\sigma$ -álgebra o sea que dicha intersección no es vacía.*

Dada una sucesión finita  $a_1, \dots, a_n$  de ceros y unos, definimos el cilindro  $[a_1, \dots, a_n]$  como el subconjunto de  $\Omega$  formado por aquellas sucesiones que empiezan con  $(a_1, \dots, a_n)$ . Notemos que para cada  $n$  las uniones finitas de cilindros de largo  $n$  forman una  $\sigma$ -álgebra finita (i.e. con finitos conjuntos), denotamos esta  $\sigma$ -álgebra por  $\mathcal{F}_n$ .

**Definición 2** (Borelianos). *La  $\sigma$ -álgebra más importante en cualquier espacio métrico completo y separable es la llamada  $\sigma$ -álgebra de Borel que es aquella generada por los abiertos. Denotamos por  $B(\Omega)$  la  $\sigma$ -álgebra de Borel de  $\Omega$ .*

### 1.5.2. Probabilidades

Ahora estamos listos para definir probabilidades. Una probabilidad en  $\Omega$  es una función  $\mu : B(\Omega) \rightarrow [0, 1]$  que cumple las siguientes dos propiedades

1.  $\mu(\Omega) = 1$
2.  $\mu(\bigcup_n A_n) = \sum_n \mu(A_n)$  para toda sucesión de Borelianos dos a dos disjuntos  $A_1, A_2, \dots$

A veces se consideran probabilidades con dominio en otras  $\sigma$ -álgebras pero la única razón para hacer esto es para extenderlas a todo  $B(\Omega)$ . El punto técnico (que obliga a introducir las  $\sigma$ -álgebras en primer lugar) es que no existe ninguna probabilidad 'no trivial' en la  $\sigma$ -álgebra  $2^\Omega$ .

Por otro lado, es trivial construir probabilidades en  $\sigma$ -álgebras finitas como las  $\mathcal{F}_n$ . El teorema básico que permite construir probabilidades es el siguiente.

**Teorema 1** (Teorema de Extensión de Tulcea). *Sea  $\mathcal{F}_n, n = 1, 2, \dots$  una sucesión creciente de sub- $\sigma$ -álgebras de la  $\sigma$ -álgebra de Borel  $B(\Omega)$  de algún espacio métrico separable y completo. Además supongamos que la unión de los  $\mathcal{F}_n$  genera  $B(\Omega)$  y que para toda sucesión decreciente de Borelianos no vacíos  $A_n \in \mathcal{F}_n$  la intersección  $\bigcap A_n$  es no vacía. Dada una sucesión  $\mu_n, n = 1, 2, \dots$  de probabilidades tales que  $\mu_n$  está definida en  $\mathcal{F}_n$ , y la restricción de  $\mu_{n+1}$  a  $\mathcal{F}_n$  coincide con  $\mu_n$  para todo  $n$ , existe una única probabilidad Boreliana  $\mu$  que coincide con cada  $\mu_n$  en  $\mathcal{F}_n$ .*

**Ejercicio 4.** *Verificar que la sucesión de  $\sigma$ -álgebras  $\mathcal{F}_n$  generadas por los cilindros de largo  $n$  que hemos definido en  $\Omega = \{0, 1\}^{\mathbb{N}}$  satisface las hipótesis del teorema de Tulcea.*

Con los anteriores prerrequisitos técnicos podemos finalmente construir las probabilidades que modelan el experimento de tirar una misma moneda repetidas veces.

**Ejercicio 5.** *Demostrar que para cada  $p \in [0, 1]$  existe una única probabilidad en  $\Omega$  tal que para todo  $n$  y toda sucesión de ceros y unos  $a_0, \dots, a_n$  se cumple*

$$\mu_p([a_0, \dots, a_n]) = p^k(1-p)^{n-k}$$

donde  $k$  es el número de unos en la sucesión.

### 1.5.3. Integrales y Esperanza

Dada una probabilidad en  $\Omega$  y una función  $f : \Omega \rightarrow \mathbb{R}$  medible (i.e. preimagen de todo Boreliano de  $\mathbb{R}$  es Boreliano en  $\Omega$ ) denotamos por  $\mathbb{E}_\mu(f)$  o por  $\int f(\omega)d\mu(\omega)$  la integral de Lebesgue (también llamada esperanza o valor esperado de  $f$  respecto a  $\mu$ ) de  $f$  respecto a  $\mu$  si esta es finita (i.e. si la parte positiva y negativa de  $f$  son integrables).

Recordemos que si  $f$  toma finitos valores  $x_1, \dots, x_n$  la integral o esperanza se define como

$$\mathbb{E}_\mu(f) = \sum_{k=1}^n \mu(\{\omega : f(\omega) = x_k\})x_k.$$

Si  $f$  es no negativa se define  $\mathbb{E}_\mu(f)$  como el supremo de las esperanzas de funciones medibles no negativas que toman finitos valores y son menores o iguales a  $f$  en todo punto (este supremo podría ser  $+\infty$ ).

En general si  $\mathbb{E}_\mu(|f|) < +\infty$  se define  $\mathbb{E}_\mu(f) = \mathbb{E}_\mu(f^+) - \mathbb{E}_\mu(f^-)$  donde  $f^+$  coincide con  $f$  donde  $f$  es positiva y vale 0 en el resto, y  $f^-$  coincide con  $-f$  si  $f$  es negativa y vale 0 en el resto.

**Ejercicio 6.** *Demostrar que si  $f : \Omega \rightarrow \mathbb{R}$  es continua entonces*

$$\mathbb{E}_\mu(f) = \lim_{n \rightarrow +\infty} \sum_{a_1, \dots, a_n \in \{0, 1\}} \mu([a_0, \dots, a_n])f(a_0, \dots, a_n, 0, 0, \dots)$$

para toda probabilidad  $\mu$ .

**Ejercicio 7** (Fubini).  *Demostrar que si  $f : \Omega \rightarrow \mathbb{R}$  sólo depende de las primeras  $n$  coordenadas y  $g : \Omega \rightarrow \mathbb{R}$  depende de las coordenadas  $n + 1, n + 2, \dots, n + m$  entonces*

$$\mathbb{E}_{\mu_p}(fg) = \mathbb{E}_{\mu_p}(f) \mathbb{E}_{\mu_p}(g)$$

para todo  $p \in [0, 1]$ .

Un teorema importante para construcción de medidas es el siguiente

**Teorema 2** (Teorema de Riesz).  *Sea  $X$  un espacio métrico completo y separable y  $E : C_c(X) \rightarrow \mathbb{R}$  una función lineal que asocia a cada función continua de  $X$  en  $\mathbb{R}$  un número real, y además asocia un número no negativo a cada función no negativa y cumple  $E(1) = 1$ . Entonces existe una probabilidad  $\mu$  en  $X$  tal que  $E(f) = \mathbb{E}_\mu(f)$  para toda función continua de soporte compacto  $f : X \rightarrow \mathbb{R}$ .*

Sorprendentemente el teorema anterior es más fácil de demostrar en  $\{0, 1\}^{\mathbb{N}}$  que en un espacio métrico completo y separable general. La demostración más elegante del teorema pasa por reducir el caso general al caso  $\{0, 1\}^{\mathbb{N}}$ .

### 1.6. El caso de la moneda: Paso 3

Estamos en condiciones de enunciar uno de los dos teoremas básicos de la teoría de probabilidades (el otro, el teorema central del límite, no será discutido en estas notas).

**Teorema 3** (Ley de Grandes Números).  *Para todo  $p \in [0, 1]$  se cumple*

$$\mu_p \left( \left\{ \omega \in \Omega : \lim_{n \rightarrow +\infty} \frac{\omega_1 + \dots + \omega_n}{n} = p \right\} \right) = 1.$$

Un punto importante para verificar es que la teoría que estamos desarrollando (donde las probabilidades sólo están definidas para Borelianos) es suficientemente rica como para hacer que el enunciado del anterior teorema tenga sentido.

**Ejercicio 8.**  *Demostrar que el conjunto  $\{\omega \in \Omega : \lim_{n \rightarrow +\infty} \frac{\omega_1 + \dots + \omega_n}{n} = 1/2\}$  es Boreliano.*

Un punto clave de la demostración de la ley de grandes números es el siguiente lema que puede interpretarse diciendo que no se pueden dar infinitas capas a una pared con sólo un litro de pintura.

**Ejercicio 9** (Borel-Cantelli).  *Sea  $\mu$  una probabilidad en  $\Omega$  y  $A_n$  una familia de Borelianos tales que*

$$\sum_{n=1}^{\infty} \mu(A_n) < +\infty.$$

Entonces se cumple

$$\mu(\{\omega \in \Omega : \omega \in A_n \text{ para infinitos } n\}) = 0.$$

Sin más preámbulos:

*Demostración de la ley de grandes números.* Dado  $\epsilon > 0$  definamos para cada  $n$  el conjunto  $A_n = \{\omega \in \Omega : \frac{\omega_1 + \dots + \omega_n}{n} - p > \epsilon\}$ . La idea es mostrar que la serie de  $\mu_p(A_n)$  converge (y luego lo mismo para la sucesión de Borelianos  $B_n$  definidos en forma similar a los  $A_n$  pero cambiando  $>$  por  $<$ , pero la prueba es la misma). Usando Borel-Cantelli se concluye que el conjunto de los  $\omega$  tales que eventualmente  $\frac{\omega_1 + \dots + \omega_n}{n} - p \leq \epsilon$  mide 1. Cómo esto vale para todo  $\epsilon > 0$  se concluye el teorema (intersectando los conjuntos de medida 1 obtenidos para numerables valores de  $\epsilon > 0$ ).

Para estimar  $\mu_p(A_n)$  usamos propiedades básicas de la integral (Fubini y un que  $\mathbb{E}_\mu(f) \leq \mathbb{E}_\mu(g)$  si  $f \leq g$ ). Las funciones involucradas en las integrales son las proyecciones sobre las coordenadas, i.e.  $X_n : \Omega \rightarrow \mathbb{R}$  definidas por  $X_n(\omega) = \omega_n$  donde  $\omega = (\omega_1, \dots, \omega_n, \dots)$ .

El primer paso es la siguiente observación válida para cualquier  $t > 0$  (uso la notación  $1_A$  para la función que toma el valor 1 en  $A$  y 0 en  $\Omega \setminus A$ , i.e. la indicatriz de  $A$ )

$$\mu_p(A_n) = \mathbb{E}_\mu(1_{A_n}) \leq \mathbb{E}_\mu(\exp(t(X_1 + \dots + X_n - n(p + \epsilon))))$$

donde la última desigualdad es porque  $1_{A_n}$  toma el valor 1 en  $A_n$  pero  $\exp(t(X_1 + \dots + X_n - n(p + \epsilon))) \geq 1$  en este conjunto (por definición de  $A_n$ ).

Luego se usa Fubini y en el lado derecho ( $X_n$  sólo depende de la coordenada  $n$ ) y se obtiene (poniendo  $q = 1 - p$ ):

$$\mu_p(A_n) \leq e^{-t(p+\epsilon)n} \mathbb{E}_\mu(e^{tX_1})^n = e^{-t(p+\epsilon)n} (pe^t + q)^n.$$

Ahora el truco reside en analizar la función  $t \mapsto pe^t + q$  y notar que si  $t > 0$  es suficientemente chico se tiene  $pe^t + q \leq e^{(p+\epsilon/2)t}$  (comparando derivadas en 0).

Es decir para algún  $t > 0$  suficientemente chico se cumple

$$\mu_p(A_n) \leq e^{-ten/2}$$

lo cual es suficiente para mostrar

$$\sum_n \mu_p(A_n) < +\infty$$

concluyendo la demostración.  $\square$

## 1.7. Equidistribución de dígitos decimales

En el intervalo  $[0, 1]$  existe una única probabilidad Boreliana  $\mu$   $\mathbb{E}_\mu$  coincide con la integral habitual (de Riemann) en las funciones continuas. Esta probabilidad, que cumple  $\mu([a, b]) = b - a$  para todo  $[a, b] \subset [0, 1]$ , habitualmente es llamada la distribución uniforme en  $[0, 1]$  o la medida de Lebesgue.

Cada punto  $x \in [0, 1]$  tiene asociado una expansión decimal, e.g.  $1/3 = 0,333333\dots$ . Además los únicos puntos que tienen más de una expansión decimal son racionales y por lo tanto forman un conjunto de  $\mu$ -medida nula.

Usando la ley de grandes números (la versión para 10 símbolos en lugar de 2) se obtiene el siguiente corolario.

**Corolario 1.** *Sea  $\mu$  la distribución uniforme en  $[0, 1]$ , entonces existe un conjunto Boreliano  $A \subset [0, 1]$  probabilidad 1 para  $\mu$  tal que si  $x \in A$  entonces cada dígito aparece con frecuencia  $1/10$  en la expansión decimal de  $x$ .*