Clase 3: Los axiomas de Kolmogorov

Matías Carrasco

30 de junio de 2019

Índice

Resumen En esta clase veremos la definición general de probabilidad basada en los axiomas de Kolmogorov. Deduciremos varias de las propiedades básicas de la probabilidad de estos axiomas. Luego aplicaremos los axiomas al caso de experimentos con espacios muestrales discretos. En particular, veremos algunos ejemplos discretos de espacios no equiprobables.

La definición axiomática de probabilidad

Todas las religiones tienen su mesías, y el de los probabilístas es el ruso Andréi Kolmogorov (1903-1987). Fue él quien introdujo en 1933 la definición axiomática de probabilidad, el equivalente de los diez mandamientos de Moisés. Solo que en este caso son cuatro:

Axiomas de Kolmogorov

Sea Ω un espacio muestral cualquiera. Una probabilidad es una función

$$P : \text{Eventos} \to [0,1]$$

que a cada evento A asigna un número real $P(A)$, y que para ser digna de ese nombre debe cumplir:

- *Ax.1*: $P(A) \geq 0$ para cualquier evento A.
- *Ax.2*: $P(\Omega) = 1$.
- *Ax.3*: $P(A \cup B) = P(A) + P(B)$ si A y B son incompatibles.
- *Ax.4*: Si $\{A_k\}$ es una sucesión creciente de eventos, entonces

$$P\left(\lim_{k\to\infty} A_k\right) = \lim_{k\to+\infty} P(A_k)\,.$$

Antes de ver algunos ejemplos, vamos a probar varias propiedades básicas que cumplen las probabilidades. Estas propiedades son consecuencia de las reglas básicas (axiomas), de modo que son ciertas para cualquier función P que cumpla con dichas reglas.

Complemento

Si A es un evento, la probabilidad de su complemento es

$$P(A^c) = 1 - P(A)\,.$$

Demostración. Esto se sigue de que $A \cup A^c = \Omega$ y de que esta unión es disjunta. Si aplicamos las reglas 1 y 2, en ese orden, obtenemos

$$1 = P(\Omega) = P(A) + P(A^c),$$

y de aquí despejamos la probabilidad de A^c. □

Regla de la resta

Si A es un subconjunto de B (esto lo escribimos $A \subset B$), entonces

$$P(B \setminus A) = P(B) - P(A).$$

Demostración. Recordar que el evento $B \setminus A$ consiste de aquellos elementos ω que están en B pero no en A. Para demostrarla notar que podemos escribir $B = A \cup (B \setminus A)$, y que esta unión es disjunta. Luego, por la regla 2, $P(B) = P(A) + P(B \setminus A)$. □

Monotonía

Si A es un subconjunto de B, entonces $P(A) \leq P(B)$.

Demostración. Esto se sigue inmediatamente de la regla de la resta ya que $P(B \setminus A) \geq 0$. □

Notar que la propiedad de monotonía junto con los axiomas 1 y 2 implican que

$$0 \leq P(A) \leq 1$$

para todo evento A.

División en casos

Si podemos dividir el espacio muestral Ω en subconjuntos $C_1, \ldots, C_n$ que son disjuntos dos a dos, entonces la probabilidad de cualquier evento A se descompone como

$$P(A) = \sum_{i=1}^{n} P(A \cap C_i).$$

Demostración. La demostración consiste en aplicar el axioma 3 a los conjuntos $A \cap C_i$ cuya unión es A. □

Principio de inclusión-exclusión

Si A y B son dos eventos cualesquiera, entonces

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Demostración. Esto se deduce del axioma 3 y de la regla de la resta aplicadas a la unión disjunta

$$A \cup B = [A \setminus (A \cap B)] \cup [B \setminus (A \cap B)] \cup [A \cap B] .$$

□

La cota de la unión

Si A y B son dos eventos cualesquiera, entonces

$$P(A \cup B) \leq P(A) + P(B)$$

Demostración. Esto se sigue del principio de inclusión y exclusión ya que $P(A \cap B) \geq 0$. □

Si Ω es finito el axioma 4, la continuidad de la probabilidad, se cumple de forma automática. Es una regla importante solamente para espacios muestrales infinitos. Este junto al axioma 3, implican una versión mejorada de este último, valido para cualquier cantidad numerable de eventos.

El axioma 3(∞)

Si $\{A_k\}$ es una sucesión de eventos incompatibles dos a dos (dos cualesquiera de ellos A_i y A_k son incompatibles), entonces

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k) . \tag{1}$$

Demostración. La prueba es muy simple. Primero observar que el axioma 3 vale, razonando por inducción, para cualquier cantidad finita de eventos incompatibles. Definimos

$$B_n = \bigcup_{k=1}^{n} A_k,$$

que es una sucesión creciente de eventos cuya unión es $\bigcup_n B_n = \bigcup_k A_k$. Por el axioma 3 tenemos $P(B_n) = \sum_{k=1}^{n} P(A_k)$. Y por el axioma 4, deducimos

$$P\left(\bigcup_k A_k\right) = P\left(\bigcup_n B_n\right) = \lim_{n\to\infty} P(B_n) = \lim_{n\to\infty} \sum_{k=1}^{n} P(A_k) = \sum_{k=1}^{\infty} P(A_k) .$$

□

Recíprocamente, (1) implica el axioma 4. Para ver esto, sea $\{A_k\}$ una sucesión creciente de eventos. Definimos

$$B_n := \begin{cases} A_1 & \text{si } n = 1; \\ A_n \setminus A_{n-1} & \text{si } n > 1. \end{cases}$$

Más aún, la unión sigue siendo la misma: $\bigcup_n B_n = \bigcup_k A_k$; pero la ventaja ahora es que los eventos $\{B_n\}$ son dos a dos disjuntos. Así que aplicando (1) deducimos

$$\begin{aligned} P\left(\bigcup_k A_k\right) &= P\left(\bigcup_n B_n\right) = \sum_{n=1}^{\infty} P(B_n) \\ &= P(A_1) + \sum_{n=1}^{\infty} P(A_n) - P(A_{n-1}) = \lim_{n\to\infty} P(A_n). \end{aligned}$$

Es decir que (1) es equivalente a Ax.3 + Ax.4. Es por esto que lo llamamos Axioma 3(∞), pues es como el axioma 3 pero para infinitos eventos.

Forma equivalente de los axiomas de Kolmogorov

Sea Ω un espacio muestral cualquiera. Una probabilidad es una función

$$P : \text{Eventos} \to [0, 1]$$

que a cada evento A asigna un número real $P(A)$, y que para ser digna de ese nombre debe cumplir:

- *Ax.1*: $P(A) \geq 0$ para cualquier evento A.
- *Ax.2*: $P(\Omega) = 1$.
- *Ax.3(∞)*: Si $\{A_k\}$ es una sucesión de eventos incompatibles dos a dos, entonces $P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k)$.

Modelo general de probabilidades discretas

Claramente, los ejemplos de experimentos equiprobables y de probabilidades geométricas que vimos en las dos clases pasadas son casos particulares de esta definición. Pero los axiomas de Kolmogorov no restringen los resultados de un experimento a ser equiprobables.

Ejemplo 1 Un caso particular muy importante es el modelo general de probabilidades discretas. Esto quiere decir que Ω puede ser finito o infinito, pero en caso de ser infinito debe ser numerable. Así, podemos enumerar los elementos de Ω en una lista

$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_i, \ldots\}.$$

Para definir las probabilidades de los eventos de Ω, primero definimos las probabilidades $p_i \in [0,1]$ de los eventos simples ω_i.

No cualquier elección de números p_i hará que la probabilidad del espacio muestral sea 1. Para que esto suceda, debemos imponer la condición de normalización

$$\sum_{i=1}^{\infty} p_i = 1.$$

En caso de que Ω sea finito, digamos con n elementos, los p_i serán nulos para $i > n$.

¿Cómo definimos la probabilidad de un evento? Sea A un evento en Ω, definimos la probabilidad de A como el agregado de las probabilidades de sus elementos, de modo que

$$P(A) := \sum_{\omega_i \in A} p_i.$$

La condición de normalización de los p_i garantiza que

$$P(\Omega) = \sum_{\omega_i \in \Omega} p_i = \sum_{i=1}^{\infty} p_i = 1.$$

Además, se cumple el axioma 3(∞): si podemos descomponer un evento A como una unión disjunta (posiblemente infinita) de eventos A_k, $k = 1, 2, \ldots$, entonces

$$P(A) = \sum_{k=1}^{\infty} P(A_k).$$

De hecho, por definición

$$P(A) = \sum_{\omega_i \in A} p_i = \sum_{k=1}^{\infty} \sum_{\omega_i \in A_k} p_i = \sum_{k=1}^{\infty} P(A_k).$$

en donde la segunda igualdad es consecuencia de que los A_i son disjuntos.

Notar que si Ω tiene n elementos, y elegimos $p_i = 1/n$ para todo i, entonces

$$P(A) = \sum_{i=1}^{n} p_i = \sum_{i=1}^{n} \frac{1}{n} = \frac{|A|}{n} = \frac{|A|}{|\Omega|},$$

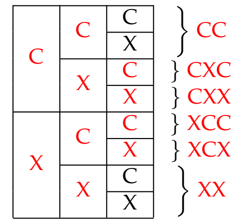
que no es otra cosa que la fórmula de casos favorable sobre casos posibles. Lo interesante del modelo general es que los $\{p_i\}$ pueden ser arbitrarios, con la sola condición de sumar 1. ■

Ejemplo 2 Consideremos el siguiente experimento: tiramos una moneda justa hasta que se repita una cara o una cruz. ¿Cómo son el espacio muestral y las probabilidades? El espacio muestral es sencillo, consiste en

$$\Omega = \{CC, XX, CXC, XCX, CXX, XCC\}$$

Pero, ¿son los resultados equiprobables?

Si tiráramos tres veces la moneda, olvidandonos por el momento de si repite o no, el espacio muestral sería el conjunto de todas las secuencias de tres caras y cruces, y consistiría entonces de $2^3 = 8$ elementos. Claramente, la probabilidad de cada secuencia sería $1/8$.



Pero en nuestro caso debemos parar si encontramos una repetición, como en las secuencias marcadas en rojo en la tabla al margen.

Notar que la secuencia que comienza con CC corresponde a dos secuencias posibles de largo tres, a saber CCC y CCX. De aquí resulta natural suponer que la probabilidad de obtener CC es $2 \times 1/8 = 1/4$. Lo mismo para XX.

Es tentador entonces definir

$$P(CC) = P(XX) = 1/4$$

$$P(CXC) = P(CXX) = P(XCC) = P(XCX) = 1/8$$

Aunque las probabilidades no sean todas iguales, la suma de todas es igual a 1. Es como si hubiéramos repartido de forma desigual una torta, las secuencias CC y XX se llevan pedazos más grandes que el resto.

¿Qué ocurre con las probabilidades de los eventos? Por ejemplo, ¿cuál es la probabilidad de que el primer lanzamiento resulte cara? Si sumamos las probabilidades de todas las secuencias de Ω que empiezan con C, obtenemos

$$P(1^{\text{er}} \text{ lanzamiento cara}) = \frac{1}{4} + \frac{1}{8} + \frac{1}{8} = \frac{1}{2},$$

que concuerda con nuestra intuición. ■

Ejemplo 3 El ejemplo anterior se puede extender a situaciones más generales, como por ejemplo, ¿qué pasa cuando lanzamos un dado hasta que se repita un resultado?

> Disponemos de n celdas distintas y comenzamos a distribuir bolas en ellas. Las bolas se distribuyen una a la vez, eligiendo una celda al azar para cada bola. El proceso termina cuando una bola cae en una celda que ya está ocupada por otra. ¿Cuántas bolas hemos distribuido al terminar el proceso?

En el caso de la moneda del ejemplo anterior, tenemos $n = 2$ celdas, una que representa cara y otra que representa cruz. En el caso del dado, tenemos $n = 6$ celdas, una por cada dígito posible que tiene el dado. En ambos casos, los lanzamientos corresponden a las bolas.

Claramente, podemos distribuir a lo sumo n bolas sin ocupar dos veces la misma celda, pero la cantidad de bolas distribuidas al terminar puede ser cualquier número entre 2 y $n + 1$.

Como hicimos en el ejemplo anterior, olvidémonos por un momento de que el proceso termina cuando una bola cae en una celda ya ocupada. Fijemos así una cierta cantidad k de bolas distinguibles a distribuir en las n celdas distintas. Para este nuevo experimento el espacio muestral es

$$\Omega_k = \{(c_1, \ldots, c_k) : c_i \in \{1, \ldots, n\}\},$$

en donde la secuencia $(c_1, \ldots, c_k)$ representa el resultado en el cual la primer bola se coloca en la celda c_1, la segunda en la celda c_2, y así sucesivamente con la k-ésima bola en la celda c_k. Como hay n posibilidades para cada bola, vemos que $|\Omega_k| = n^k$. Esto sugiere que la probabilidad de obtener una distribución dada de bolas igual a $(c_1, \ldots, c_k)$ es igual a $1/n^k$.

Volvamos a nuestro experimento original. Una secuencia $(c_1, \ldots, c_k)$ representa ahora que se coloca la primera, la segunda, ..., y la k-ésima bola en las celdas $c_1, c_2, \ldots, c_k$, y que el proceso termina en k pasos. Esto significa que las celdas $c_1, \ldots, c_{k-1}$ son todas diferentes, pero c_k es igual a una de ellas. Cualquier secuencia de este tipo representa un resultado posible, y solamente los valores $2, \ldots, n+1$ son posibles para k.

El espacio muestral es entonces

$$\Omega = \{(c_1, \ldots, c_k) : c_i \neq c_j \ \forall i, j < k \text{ y } \exists i \text{ t.q. } c_k = c_i\}.$$

De acuerdo a lo dicho más arriba, es natural suponer que la probabilidad de cada secuencia de largo k es $1/n^k$, es decir

$$P(c_1, \ldots, c_k) = \frac{1}{n^k}.$$

Debemos asegurarnos, sin embargo, que la probabilidad de Ω sea igual a uno. Esto es, que la suma de las probabilidades de todas las secuencias posibles es 1.

Para hacer esto, consideremos para cada k entre 2 y $n+1$, el evento A_k de que el proceso termine en el k-ésimo paso. Obviamente estos eventos son disjuntos dos a dos, ya que el proceso no puede terminar en dos cantidades diferentes de pasos. ¿Cuántos elementos tiene cada A_k?

Por la regla del producto, podemos elegir las celdas $c_1, \ldots, c_{k-1}$ de $(n)_{k-1}$ formas distintas. Para c_k tenemos que elegir entre las $k-1$ celdas $c_1, \ldots, c_{k-1}$. Por lo tanto $|A_k| = (k-1)(n)_k$, de donde se sigue que

$$\sum_{(c_1,\ldots,c_k)\in A_k} P(c_1, \ldots, c_k) = |A_k| \frac{1}{n^k} = \frac{(k-1)(n)_{k-1}}{n^k}.$$

Llamemos a estos números α_k. Debemos probar que

$$\sum_{k=2}^{n+1} \alpha_k = 1.$$

Es fácil ver que $\alpha_2 = 1/n$. También, un cálculo sencillo muestra que

$$1 - (\alpha_2 + \alpha_3) = 1 - \frac{1}{n} - \frac{2(n-1)}{n^2} = \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right).$$

En general, se puede mostrar por inducción que

$$1 - (\alpha_2 + \cdots + \alpha_r) = \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{r-1}{n}\right) = \frac{(n)_r}{n^r}. \qquad (2)$$

Esto muestra que

$$\alpha_2 + \cdots + \alpha_{n+1} = 1$$

porque en la fórmula anterior con $r = n+1$ el último factor es cero. Observar que (2) representa la probabilidad de que el proceso termine en más de r pasos. El lado derecho de (2) no es otra cosa que la probabilidad de que r bolas caigan en celdas distintas. ■

Ejemplo 4 Supongamos que en una clase, suficientemente grande, el profesor empieza a preguntarle a sus estudiantes el día del cumpleaños. Lo hace uno a la vez, hasta encontrar una coincidencia. ¿A cuántos estudiantes le preguntará?

Podemos pensar entonces que disponemos de $n = 365$ celdas, que representan los diferentes días del año, y que empezamos a distribuir bolas hasta que una de ellas caiga en una celda ya ocupada.

Por lo que vimos en la parte anterior, la probabilidad de preguntarle a k estudiantes (es el evento que llamamos A_k) es

$$P\,(\text{preguntar a } k \text{ estudiantes}) = \frac{(k-1)(365)_{k-1}}{365^k}.$$

Lamentablemente estas probabilidades son difíciles de calcular a mano. Se puede hacer una aproximación muy buena con herramientas elementales de cálculo, pero como una imagen vale más que mil palabras, nos contentaremos con la gráfica de la Figura 1.

Notar que el valor más probable para k es $k = 20$, con una probabilidad de 0.032. Este valor no es difícil de calcular. Si llamamos p_k a la probabilidad de preguntarle a k estudiantes, vemos que

$$\frac{p_k}{p_{k+1}} = \frac{365(k-1)}{k(365-k+1)}.$$

De aquí resulta que $p_k < p_{k+1}$ si, y solo si $k^2 - k < 365$. Resolviendo la ecuación cuadrática, esta última desigualdad es equivalente a $k < \frac{1+\sqrt{1+4\cdot 365}}{2} = 19.61$. Esto quiere decir que p_k crece hasta $k = 20$, y luego comienza a decrecer.

El caso general de n celdas es completamente análogo, y se obtiene que la probabilidad máxima se alcanza en $k \approx \sqrt{n}$. ■

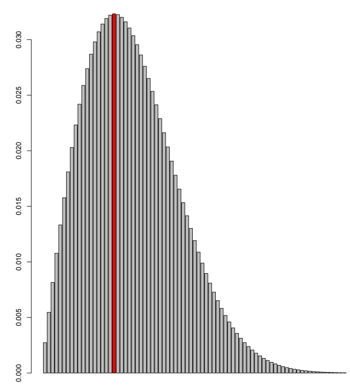


Figura 1: Esta gráfica muestra la probabilidad p_k de que el profesor se detenga luego de preguntarle a k estudiantes. Notar que el máximo se da en $k = 20$.

Ejemplo 5 Tiramos una moneda justa hasta que salga cara. ¿Cuál es la probabilidad de que tengamos que tirar un número par de veces?

Un espacio muestral razonable es

$$\Omega = \{C, XC, XXC, XXXC, XXXXC, \dots\}.$$

Usaremos el mismo truco que en los ejemplos anteriores para definir la probabilidad de cada secuencia. Olvidémonos por un momento de las reglas del juego y que aunque hayamos obtenido una cara en un determinado lanzamiento, seguimos tirando la moneda.

Al tirar una sola vez la moneda, cada uno de los resultados C o X tiene la misma probabilidad de ocurrir. Sin importar lo que hagamos después, es razonable definir la probabilidad de C como igual a $1/2$.

Esto además es coherente con lo que ocurre al tirar dos veces la moneda. Los resultados posibles son

$$CC, CX, XC, XX.$$

Estos son todos igualmente probables, con probabilidad $1/4$. Si sale C en la primer tirada habríamos ganado, pero tiramos la moneda una vez más. Puede salir C de nuevo o X. Estos son dos casos ficticios en los cuales hubiéramos parado el juego en la primer tirada. La suma de sus probabilidades es $1/4 + 1/4 = 1/2$ lo cual concuerda con nuestra definición anterior.

De hecho, este argumento funciona para cualquier cantidad de lanzamientos. Si lanzamos n veces una moneda, todas las secuencias que comienzan por C son ficticias. Como son exactamente la mitad de todas las secuencias posibles, la probabilidad de empezar con C siempre es $1/2$ sin importar cuántas veces tiremos la moneda.

En la siguiente tabla mostramos cuáles son las probabilidades asignadas a las secuencias de nuestro experimento si seguimos un razonamiento similar al anterior para cada una de ellas.

Tiradas	**Resultados posibles**	**Probabilidad**
1	C, X	$1/2$
2	CC, CX, XC, XX	$1/4$
3	CCC, CCX, CXC, CXX XCC, XCX, XXC, XXX	$1/8$
4	$CCCC, CCCN, CCXC, CCXX, CXCC, CXCX$ $CXXC, CXXX, XCCC, XCCX, XCXC, XCXX$ $XXCC, XXCX, XXXC, XXXX$	$1/16$
$\vdots$	$\vdots$	$\vdots$

En resumen, hemos definido

$$P\left(\underbrace{X \cdots X}_{k-1 \text{ veces}} C\right) = \frac{1}{2^k}, \; k = 1, 2, \dots$$

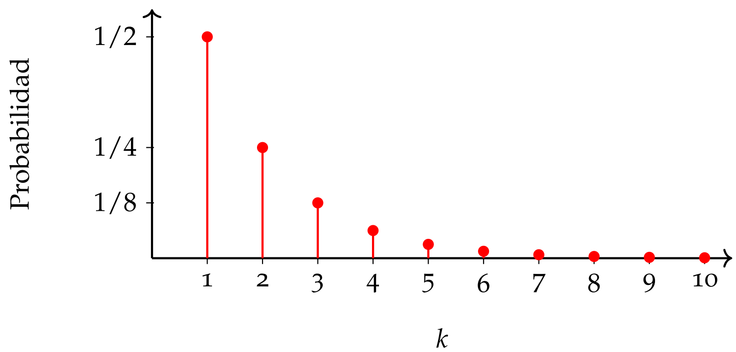
Podemos verificar que la suma de las probabilidades de todas las secuencias posibles es igual a 1. Para esto debemos recordar que la suma de una serie geométrica es

$$\sum_{j=0}^{\infty} x^j = \frac{1}{1-x}$$

si $|x| < 1$. Si la aplicamos con $x = 1/2$, obtenemos

$$P(\Omega) = \sum_{k=1}^{+\infty} P(X \cdots XC) = \sum_{k=1}^{+\infty} \frac{1}{2^k} = \frac{1}{2} \frac{1}{(1-1/2)} = 1.$$

Como estas probabilidades forman una una sucesión geométrica, en este caso de razón $1/2$, se las conoce bajo el nombre de distribución *geométrica*.



Queremos calcular la probabilidad de lanzar la moneda un número par de veces. Este evento, al que llamamos E, se puede descomponer según sean 2 lanzamientos, 4 lanzamientos, 6 lanzamientos, etc. Llamemos E_i al evento de lanzar la moneda $2i$ veces. Es decir, podemos escribir $E = \bigcup_{i=1}^{+\infty} E_i$ siendo esta una unión disjunta. Luego

$$\begin{aligned} P(E) &= \sum_{i=1}^{+\infty} P(E_i) = \sum_{i=1}^{+\infty} \frac{1}{2^{2i}} \\ &= \sum_{i=1}^{+\infty} \frac{1}{4^i} = \frac{1}{4} \frac{1}{(1-1/4)} = \frac{1}{3}. \end{aligned}$$

Es dos veces más probable tirar la moneda un número impar que un número par de veces. ■