

Probabilidades.

Luis Arenas

March 15, 2009

Chapter 1

Espacios de Probabilidad Finitos.

Al lanzar una moneda se tienen dos resultados posibles: Cara y Sello. La pregunta sobre cual es la probabilidad de obtener cara se interpreta en matemáticas como una función que asigna a cada elemento del conjunto $\Omega = \{\text{cara, sello}\}$ un número real en el intervalo $[0, 1]$ llamado su probabilidad. En otras palabras, el problema está totalmente determinado si se conocen las probabilidades $p(\text{cara})$ y $p(\text{sello})$. Por ejemplo, se dice que la moneda está equilibrada si $p(\text{cara}) = p(\text{sello})$. Uno asume, de hecho, que el conjunto Ω contiene todas las posibilidades. En particular, desestimamos la probabilidad de que la moneda quede parada de canto o un ave la devore mientras está en el aire. Esto se expresa matemáticamente mediante la ecuación

$$p(\text{cara}) + p(\text{sello}) = 1.$$

Por cierto podemos, por ejemplo, incluir la tercera probabilidad considerando un nuevo conjunto $A' = \{\text{cara, sello, canto}\}$ y asignar a sus elementos probabilidades que satisfagan la ecuación

$$p(\text{cara}) + p(\text{sello}) + p(\text{canto}) = 1,$$

y modelar el hecho de que consideramos la tercera alternativa como extremadamente improbable mediante una condición del tipo $p(\text{canto}) \ll 1$.

Mas generalmente, una función de probabilidad en el conjunto finito Ω es una función $p : \Omega \rightarrow [0, 1]$ que satisface

$$\sum_{\omega \in \Omega} p(\omega) = 1.$$

En particular, para todo subconjunto B de Ω puede definirse la probabilidad del conjunto B mediante $P(B) = \sum_{\omega \in B} p(\omega)$. La probabilidad $P(B)$ es un número entre 0 y 1. Nótese que se tienen las propiedades siguientes, cuya demostración se deja al lector:

1. $P(\emptyset) = 0$.
2. $P(\Omega) = 1$.
3. $P(B^c) = 1 - P(B)$.
4. $P(B_1 \cup B_2) = P(B_1) + P(B_2) - P(B_1 \cap B_2)$.
5. $P(B_1 \cup B_2) = P(B_1) + P(B_2)$ si B_1 y B_2 son disjuntos.
6. Si $\{B_i\}_{i \in I}$ es una familia finita de conjuntos disjuntos, entonces

$$\sum_{i \in I} P(B_i) = P\left(\bigcup_{i \in I} B_i\right).$$

7. Si $B_1 \subseteq B_2$, entonces $P(B_1) \leq P(B_2)$.

Conversamente, si $\wp(\Omega)$ es la colección de todos los subconjuntos de Ω , entonces toda función no negativa $P : \wp(\Omega) \rightarrow \mathbb{R}$ que satisface las condiciones 2 y 6 define una función de probabilidad en el conjunto Ω , ya que para todo $B \subseteq \Omega$ se tiene

$$P(B) = \sum_{\omega \in B} P(\{\omega\}),$$

y podemos definir $p(\omega) = P(\{\omega\})$. Una función $P : \wp(\Omega) \rightarrow \mathbb{R}$ que satisface 2 y 6 recibe el nombre de medida de probabilidad en el conjunto Ω . Un par (Ω, P) , donde Ω es un conjunto finito y P es una medida de probabilidad, recibe el nombre de espacio de probabilidad (finito). El conjunto Ω recibe el nombre de espacio muestral y un subconjunto $B \subseteq \Omega$ recibe el nombre de evento.

ejemplo 1.1. Sea $\Omega = \{1, 2, 3, 4, 5, 6\}$, y sea $p(i) = 1/6$ para todo i . Esta es la función probabilidad que se esperaría al arrojar un dado simétrico. El conjunto de los pares sería $B = \{2, 4, 6\}$ y el de los números compuestos sería $C = \{4, 6\}$. En este caso se tiene $P(B) = 1/2$ y $P(C) = 1/3$. En otras palabras, la probabilidad de obtener un número par es de $1/2$ y la probabilidad de obtener un número compuesto es de $1/3$. El único elemento de B que no está en C es el 2, y de hecho $p(2) = 1/6 = P(B) - P(C)$.

ejemplo 1.2. Se arroja una moneda dos veces. En este caso hay 4 resultados posibles (sin considerar caídas de canto o aves tragamonedas):

$$\Omega = \left\{ (\text{cara, cara}), (\text{cara, sello}), (\text{sello, cara}), (\text{sello, sello}) \right\}.$$

Si la moneda está equilibrada, es *razonable* asumir que estas cuatro alternativas son igualmente probables y tiene cada una la probabilidad $1/4$. En este caso la probabilidad de obtener una cara en la primera tirada es la probabilidad del evento

$$B = \left\{ (\text{cara, cara}), (\text{cara, sello}) \right\}.$$

De hecho $P(B) = 1/2$, que es la misma probabilidad que asociamos a la obtención de una cara en el lanzamiento de una moneda. Del mismo modo, la probabilidad de obtener una cara en la segunda tirada es la probabilidad del evento

$$C = \left\{ (\text{cara, cara}), (\text{sello, cara}) \right\}.$$

Como antes se tiene $P(C) = 1/2$. Nótese que en particular que

$$P\left[\{(\text{cara, cara})\}\right] = 1/4 = 1/2 \times 1/2.$$

En otras palabras, para obtener la probabilidad de obtener una cara en cada lanzamiento, multiplicamos la probabilidad de obtener una cara en el primer lanzamiento por la probabilidad de obtener una cara en el segundo lanzamiento. Intuitivamente, consideramos que si la probabilidad de obtener una cara en el segundo lanzamiento es $1/2$, la probabilidad de obtener (cara, cara) es la mitad de la probabilidad de obtener una cara en el primer lanzamiento. A continuación formalizaremos este concepto.

definición 1.3. Dos eventos B y C se dicen independientes si $P(B \cap C) = P(B)P(C)$.

ejemplo 1.4. En el caso de la moneda arrojada dos veces, el evento de obtener cara en la primera tirada es independiente del evento de obtener cara en la segunda tirada.

definición 1.5. Para Dos eventos B y C cualesquiera, tales que $P(B) \neq 0$, se define la probabilidad condicionada $P(C|B) = P(B \cap C)/P(B)$. Similarmente, para un elemento $\omega \in B$ se define $p(\omega|B) = p(\omega)/P(B)$, mientras

que $p(\omega|B) = 0$ si $\omega \in B^c$. La función $\omega \mapsto p(\omega|B)$ es una función de probabilidad, ya que es positiva y se tiene

$$\sum_{\omega \in \Omega} p(\omega|B) = \sum_{\omega \in B} p(\omega|B) = \sum_{\omega \in B} \frac{p(\omega)}{P(B)} = \frac{1}{P(B)} \sum_{\omega \in B} p(\omega) = \frac{P(B)}{P(B)} = 1.$$

La probabilidad condicionada es frecuentemente interpretada como la probabilidad *a posteriori* una vez que se ha establecido que el evento B ocurre. Se sigue de la definición que B y C son independientes si y sólo si

$$P(C|B) = \frac{P(C \cap B)}{P(B)} = \frac{P(C)P(B)}{P(B)} = P(C).$$

En otras palabras, el conocimiento de que el evento B ocurre no afecta nuestro cálculo de la probabilidad del evento C . Intuitivamente, pensamos que dos eventos son independientes si ninguno de ellos es consecuencia del otro ni existe una causa común a ambos. Por ejemplo, si yo ignoro la hora que marca mi reloj de pulsera, pero observo que el reloj de la plaza marca las 5, es mucho más probable que mi reloj marque una hora cercana a las 5 que una hora cercana a la una. Esto se debe a que existe una causa común a ambos fenómenos (ambos relojes marcan la hora). Por otro lado, en el ejemplo de las monedas, nuestra intuición nos dice que la primera vez que tiramos la moneda no afecta la segunda, por lo que ambos eventos deben ser independientes. Uno espera naturalmente que si no existe relación causal alguna entre dos fenómenos, nuestro conocimiento de uno no influya en nuestro conocimiento del otro, de modo que en particular, nuestro cálculo de las probabilidades asociadas al segundo fenómeno no deben ser influenciadas por nuestro conocimiento del primero. Es esta consecuencia de la noción intuitiva de dependencia la que se utiliza como definición en la teoría matemática, a falta de una manera mas directa de traducir a símbolos la noción intuitiva de causalidad. Es posible, sin embargo, que eventos que si estén relacionados causalmente sean matemáticamente independientes. Por ejemplo, si una caja contiene las siguientes bolas:

1. Una bola roja grande,
2. Una bola roja pequeña,
3. Una bola negra grande,

4. Una bola negra pequeña,

los eventos *bola roja* y *bola grande* son ciertamente independientes pero para cada bola grande en particular la probabilidad de extraerla es fuertemente modificada una vez que sabemos que la bola extraída fué roja. Para la bola roja grande esta probabilidad subió de $\frac{1}{4}$ a $\frac{1}{2}$, mientras que para la bola negra grande bajó de $\frac{1}{4}$ a 0. La independencia en este caso es una consecuencia del hecho de que la probabilidad no cambió *en promedio* para el conjunto de bolas grandes. Ignoraremos esta distinción en lo que sigue y nos enfocaremos en la definición matemática dada mas arriba.

ejemplo 1.6. Un ejemplo en el que no hay independencia es el siguiente: Se tiene una bolsa con dos bolas, una negra y una roja. Por turno, dos personas extraen cada una una bola y la conservan (no hay remplazo). En este caso el espacio muestral es

$$\Omega = \left\{ (\text{negra, roja}), (\text{roja, negra}) \right\},$$

de donde la probabilidad de obtener una bola roja en la segunda extracción es la misma que la de obtener una bola negra en la primera extracción, y la misma probabilidad de obtener el par (negra, roja). Las dos extracciones de este ejemplo no son independientes, salvo en el caso trivial, donde se tiene

$$p\left[(\text{negra, roja})\right] \in \{0, 1\}.$$

Nótese que si $A = \bigcup_{i \in I} A_i$ es una partición de A en conjuntos disjuntos entonces para todo subconjunto B de A se tiene

$$P(B) = \sum_{i \in I} P(B \cap A_i) = \sum_{i \in I} P(B|A_i)P(A_i). \quad (1.1)$$

En particular se tiene

$$\boxed{P(A_i|B) = \frac{P(B \cap A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{i \in I} P(B|A_i)P(A_i)}}. \quad (1.2)$$

Este resultado se conoce como Teorema de Bayes. Nos referiremos también a la relación (1.1) como *descomposición por casos*.

Una colección $\mathbb{P} = \{B_1, \dots, B_n\}$, donde $A = \bigcup_{i=1}^n B_i$ y $B_i \cap B_j = \emptyset$ si $i \neq j$, se denomina una partición de A . Nótese que $p(B_i) = P(B_i)$ es una función de probabilidad en \mathbb{P} . Diremos que es la función de probabilidad asociada a la partición.

definición 1.7. Una variable aleatoria en el espacio de probabilidad finito (Ω, P) es una función $X : \Omega \rightarrow \mathbb{R}$. Su esperanza, que será denotada $E_P(X)$, o simplemente $E(X)$ si no hay riesgo de confusión, se define por

$$E(X) = \sum_{\omega \in \Omega} X(\omega)p(\omega).$$

También se utiliza la notación $\int_{\Omega} X dP(\omega)$, en cuyo caso se denomina la integral de X con respecto a la medida de probabilidad P . Nótese que se tiene

$$\begin{aligned} E(\alpha X + \beta Y) &= \sum_{\omega \in \Omega} [\alpha X(\omega) + \beta Y(\omega)]p(\omega) = \\ &= \alpha \sum_{\omega \in \Omega} X(\omega)p(\omega) + \beta \sum_{\omega \in \Omega} Y(\omega)p(\omega) = \alpha E(X) + \beta E(Y), \end{aligned}$$

y si $X(\omega) = c$ es una constante, se tiene

$$E(X) = \sum_{\omega \in \Omega} cp(\omega) = c \sum_{\omega \in \Omega} p(\omega) = c \times 1 = c.$$

Si B es un subconjunto de Ω , la esperanza condicionada de X respecto de B es por definición

$$E(X|B) = \int_{\Omega} X(\omega)dP(\omega|B) = \sum_{\omega \in \Omega} X(\omega)p(\omega|B) = \sum_{\omega \in B} X(\omega)p(\omega|B),$$

dado que $p(\omega|B) = 0$ si $\omega \notin B$. En particular, la esperanza condicionada $E(X|B)$ depende sólo de los valores de la variable aleatoria en los puntos de B . Se sigue que si X e Y son variables aleatorias que coinciden en B se tiene

$$E(X|B) = E(Y|B).$$

Nótese que $p(\omega) = p(\omega|B)P(B)$ para todo ω en B . En particular, se tiene que si $\mathbb{P} = \{B_1, \dots, B_n\}$ es una partición de Ω , entonces

$$\begin{aligned} E(X) &= \sum_{\omega \in \Omega} X(\omega)p(\omega) = \sum_{B_i \in \mathbb{P}} \left(\sum_{\omega \in B_i} X(\omega)p(\omega|B_i) \right) P(B_i) \\ &= \sum_{B_i \in \mathbb{P}} E(X|B_i)P(B_i). \end{aligned}$$

Dada un real cualquiera r , se define $A_r(X) = \{\omega \in \Omega | X(\omega) = r\}$ y se denota

$$P(X = r) := P[A_r(X)].$$

Por cierto que $P(X = r) \neq 0$ sólo para una cantidad finita de valores r . La esperanza condicionada $E(Y|X = r)$ es por definición $E[Y|A_r(X)]$, la cual está definida siempre que $P(X = r) \neq 0$. El rango probable o esencial $R_P(X)$ (o $R(X)$ si P es claro del contexto) de X es el conjunto de los valores $r \in \mathbb{R}$ tales que $P(X = r) \neq 0$. En este caso se tiene $E(X|X = r) = r$ para todo $r \in R_P(X)$. Dejamos la demostración al lector. En particular, en términos de las probabilidades $P(X = r)$, la esperanza de X puede escribirse como

$$E(X) = \sum_{r \in R(X)} rP(X = r).$$

Por cierto, la suma precedente no cambia si se reemplaza $R(X)$ por un conjunto mayor. En lo que sigue escribiremos simplemente

$$E(X) = \sum_{r \in \mathbb{R}} rP(X = r),$$

con el entendido de que esta suma es finita.

ejemplo 1.8. Suponga que se ordena aleatoriamente los números de uno a n de manera que cada orden posible sea igualmente probable. Para cada $k = 1, \dots, n$ sea X_k una variable aleatoria que vale 1 si el número k está en su posición y 0 en caso contrario. En este caso se tiene $E(X_k) = P(X_k = 1) = 1/n$. Sea Y una variable aleatoria que cuenta la cantidad de números que se encuentran en su posición correcta. Es fácil ver que $Y = X_1 + \dots + X_n$, de donde se sigue que

$$E(Y) = \sum_{k=1}^n E(X_k) = \sum_{k=1}^n 1/n = 1.$$

definición 1.9. Dos variables aleatorias X y Y se dicen independientes si para todo par de números reales r y s los conjuntos $A_r(X)$ y $A_s(Y)$ son independientes. En este caso se tiene

$$P(X = s | Y = r) = P(X = s)$$

para cada par de números reales s y r tales que $P(Y = r) \neq 0$. Se sigue que

$$E(X|Y = r) = \sum_{s \in \mathbb{R}} sP(X = s|Y = r) = \sum_{s \in \mathbb{R}} sP(X = s) = E(X)$$

para todo real r tal que $P(Y = r) \neq 0$. En particular

$$\begin{aligned} E(XY) &= \sum_{r \in R(Y)} E(XY|Y = r)P(Y = r) = \sum_{r \in R(Y)} E(Xr|Y = r)P(Y = r) \\ &= \sum_{r \in R(Y)} rE(X|Y = r)P(Y = r) = \sum_{r \in R(Y)} rE(X)P(Y = r) = E(X)E(Y), \end{aligned}$$

para todo par de variables aleatorias independientes X e Y .

Sean ahora (Ω_1, P_1) y (Ω_2, P_2) dos espacios de probabilidad finitos, y sea $\Omega = \Omega_1 \times \Omega_2$. Entonces la función $p(\omega_1, \omega_2) = p_1(\omega_1)p_2(\omega_2)$ es una función de probabilidad en el espacio Ω , de hecho

$$\sum_{(\omega_1, \omega_2) \in \Omega} p(\omega_1, \omega_2) = \left(\sum_{\omega_1 \in \Omega_1} p_1(\omega_1) \right) \left(\sum_{\omega_2 \in \Omega_2} p_2(\omega_2) \right) = 1.$$

La medida de probabilidad P definida por p satisface

$$\begin{aligned} P(B_1 \times B_2) &= \sum_{(\omega_1, \omega_2) \in B_1 \times B_2} p(\omega_1, \omega_2) = \\ &= \left(\sum_{\omega_1 \in B_1} p_1(\omega_1) \right) \left(\sum_{\omega_2 \in B_2} p_2(\omega_2) \right) = P_1(B_1)P_2(B_2). \end{aligned}$$

Dado que

$$(B_1 \times \Omega_2) \cap (\Omega_1 \times B_2) = B_1 \times B_2,$$

se tiene que los eventos

$$\bar{B}_1 = B_1 \times \Omega_2 \text{ y } \bar{B}_2 = \Omega_1 \times B_2$$

son independientes. En particular, tomando $B_1 = A_{r_1}(X_1)$ y $B_2 = A_{r_2}(X_2)$ para variables aleatorias X_1 y X_2 y números reales r_1 y r_2 cualesquiera, se tiene el siguiente resultado:

Si cada X_i es una variable aleatoria que depende solo de la coordenada ω_i , las variables X_1 y X_2 son independientes con respecto a la medida de probabilidad producto.

Por otro lado se tiene que, si X depende sólo de ω_1 , es decir $X(\omega_1, \omega_2) = Y(\omega_1)$, entonces

$$E_P(X) = \sum_{(\omega_1, \omega_2) \in \Omega} X(\omega_1, \omega_2) p(\omega_1, \omega_2) = \sum_{\omega_1 \in \Omega_1} Y(\omega_1) p_1(\omega_1) \sum_{\omega_2 \in \Omega_2} p_2(\omega_2) = \sum_{\omega_1 \in \Omega_1} Y(\omega_1) p_1(\omega_1) = E_{P_1}(Y).$$

En general uno puede identificar una variable aleatoria que depende sólo de la primera variable con una variable aleatoria en el espacio Ω_1 . Del mismo modo, todo evento de la forma $B_1 \times \Omega_2$ se identifica con el evento B_1 en Ω_1 . Todas estas consideraciones se aplican a productos de más de dos factores.

ejemplo 1.10. La variable aleatoria $X : \{0, 1\} \rightarrow \mathbb{R}$ definida por $X(\omega) = \omega$ recibe el nombre de V.A. de Bernouilli. Supongamos que cada conjunto $\Omega_i = \{0, 1\}$ tiene dada la función de probabilidad definida por $p(1) = q$, $p(0) = 1 - q$ con p y q fijos. Esto induce una probabilidad producto en el producto cartesiano

$$\Omega = \prod_{i=1}^n \Omega_i.$$

Con respecto a esta medida, la probabilidad de obtener una secuencia dada de ceros y unos (a_1, \dots, a_n) es $q^r(1 - q)^s$, donde r es el número de unos y s el número de ceros de la secuencia. La probabilidad de obtener r unos y s ceros (si $r + s = n$) es por lo tanto $\binom{n}{r} q^r(1 - q)^s$. Esta es realmente una función de probabilidad en el conjunto $\{0, 1, \dots, n\}$, puesto que

$$\sum_{r=0}^n \binom{n}{r} q^r(1 - q)^{n-r} = (q + (1 - q))^n = 1.$$

Si X es la variable aleatoria que cuenta el número de ceros y unos en la secuencia, se tiene $X = X_1 + \dots + X_n$, donde cada X_i es una V.A. de Bernouilli en la coordenada correspondiente. Se dice que X es una V.A. Binomial o con distribución binomial. Nótese que

$$E(X) = E(X_1) + \dots + E(X_n) = np.$$

La Varianza de una variable aleatoria X se define por

$$V(X) = E\left([X - E(X)]^2\right) = E(X^2) - E(X)^2.$$

Es una medida de la dispersión de los valores de la variable aleatoria. Dejamos al lector la tarea de mostrar que $V(X) = 0$ si y sólo si X es una constante.

ejemplo 1.11. Si $X : \{0, 1\} \rightarrow \mathbb{R}$ es una V.A. de Bernoulli, se tiene

$$V(X) = (1 - p)^2 p + (0 - p)^2 q = q^2 p + p^2 q = pq(q + p) = pq.$$

También puede realizarse el cálculo como sigue:

$$V(X) = E(X^2) - E(X)^2 = E(X) - E(X)^2 = p - p^2 = pq,$$

donde se usa el hecho de que $X^2 = X$, pues esa variable aleatoria toma sólo valores 0 y 1.

La Varianza de una suma $X + Y$ satisface

$$\begin{aligned} V(X + Y) &= E\left([X + Y]^2\right) - \left(E(X) + E(Y)\right)^2 = \\ &= V(X) + V(Y) + 2\left(E(XY) - E(X)E(Y)\right). \end{aligned}$$

El término $E(XY) - E(X)E(Y)$ recibe el nombre de covarianza (o coeficiente de correlación) de las variables X e Y y se denota $\text{cov}(X, Y)$. Si la covarianza es nula, las variables X e Y se dicen incorrelacionadas, en cuyo caso $V(X + Y) = V(X) + V(Y)$. Este es el caso, en particular, si X e Y son independientes.

ejemplo 1.12. Si $X = X_1 + \dots + X_n$ donde las X_i son V.A. independientes de Bernoulli, se tiene

$$V(X) = V(X_1) + \dots + V(X_n) = npq.$$

Esta es, en particular, la varianza de una V.A. con distribución binomial

Chapter 2

Espacios de Probabilidad Numerables.

En este capítulo extenderemos las definiciones anteriores al caso en el cual el espacio $\Omega = \{\omega_1, \dots\}$ es numerable. En este caso una función de probabilidad en Ω es una función $p : \Omega \rightarrow [0, 1]$ tal que

$$\sum_{i=1}^{\infty} p(\omega_i) = 1,$$

donde la suma debe entenderse como una serie convergente, y por lo tanto absolutamente convergente ya que se trata de una serie de términos positivos. En particular, el orden de los términos es irrelevante y puede escribirse simplemente

$$\sum_{\omega \in \Omega} p(\omega) = 1.$$

Del mismo modo, para cada subconjunto B de ω se define

$$P(B) = \sum_{\omega \in B} p(\omega).$$

Esta suma es también absolutamente convergente por ser una sub-suma de la anterior. Puede también expresarse en términos de la función característica χ_B de B mediante

$$P(B) = \sum_{\omega \in \Omega} \chi_B(\omega) p(\omega).$$

Las propiedades 1-6 del capítulo anterior se extienden fácilmente a este caso. Por ejemplo se tiene

Proposición 2.1. Si B_1 y B_2 son subconjuntos de Ω , entonces

$$P(B_1 \cup B_2) = P(B_1) + P(B_2) - P(B_1 \cap B_2).$$

Demostración Sigue inmediatamente de la fórmula

$$\chi_{B_1 \cup B_2}(\omega) = \chi_{B_1}(\omega) + \chi_{B_2}(\omega) - \chi_{B_1 \cap B_2}(\omega).$$

□

Las restantes propiedades se deducen fácilmente de esta o se generalizan de manera similar. De hecho, la propiedad 5 se generaliza a familias numerables, es decir:

Proposición 2.2. Si $\{B_i\}_{i \in \mathbb{N}}$ es una familia numerable de conjuntos disjuntos, entonces $\sum_{i=1}^{\infty} P(B_i) = P\left(\bigcup_{i=1}^{\infty} B_i\right)$.

Demostración Enumeremos $\Omega = \{\omega_1, \dots\}$ como al comienzo y sea $R_M = \{\omega_M, \omega_{M+1}, \dots\}$. Nótese que $P(R_M) \rightarrow 0$ cuando $M \rightarrow \infty$ por ser la cola de una serie convergente. Sea $U = \bigcup_{i=1}^{\infty} B_i$. Para cada entero positivo N considerese el conjunto $U_N = \bigcup_{i=1}^N B_i$. Obsérvese que $P(U_N) \leq P(U)$ por la propiedad 6. Por otro lado, si N es suficientemente grande todo ω_j con $j < M$ que está contenido en algún B_i , está de hecho contenido en un B_i con $i \leq N$. Se sigue que el conjunto diferencia $U - U_N$ está contenido en R_M , de donde

$$0 \leq P(U - U_N) \leq P(R_M) \rightarrow 0.$$

Como $U = U_N \cup (U - U_N)$, se tiene $P(U_N) = P(U) - P(U - U_N) \rightarrow P(U)$. □

ejemplo 2.3. Se lanza una moneda todas las veces necesarias hasta obtener una cara por primera vez. En este caso, el conjunto Ω es el conjunto de todas las secuencias finitas

cara, sello – cara, sello – sello – cara, sello – sello – sello – cara, ...

Llamaremos ω_i al i -ésimo término de la sucesión precedente. Para una moneda equilibrada se tiene $p(\omega_i) = 2^{-i}$. En este caso, la probabilidad de obtener un número par de lanzamientos es

$$\sum_{i=1}^{\infty} p(\omega_{2i}) = \sum_{i=1}^{\infty} 2^{-2i} = 1/3.$$

Nótese que en este caso existe el evento de un número infinito de tiradas, evento que llamaremos ω_∞ . Nótese que

$$p(\omega_\infty) = 1 - \sum_{i=1}^{\infty} p(\omega_i) = 1 - \sum_{i=1}^{\infty} 2^{-i} = 0.$$

En este caso diremos que el evento ω_∞ es improbable. Esto no quiere decir que sea imposible realizar un número infinito de tiradas, sinó que la probabilidad de que eso ocurra es 0. Esta distinción será mas clara en el próximo capítulo. La variable aleatoria que cuenta el número de tiradas necesarias hasta obtener la primera cara recibe el nombre de *tiempo de espera* asociado al evento *cara*.

ejemplo 2.4. Si $\Omega = \{\omega_1, \omega_2, \dots\}$ de modo que $p(\omega_i) = \frac{\beta \alpha^i}{i!}$, se dice que Ω tiene una distribución de Poisson. Nótese que la fórmula

$$1 + \alpha + \frac{\alpha^2}{2!} + \dots = e^\alpha$$

implica que $\beta = e^{-\alpha}$. La distribución de Poisson aparece naturalmente como un caso límite de la distribución binomial. Si cada uno de n objetos se coloca aleatoriamente en una de m cajas, la probabilidad de obtener r objetos en una caja dada es $\binom{n}{r} q^r (1-q)^{n-r}$ donde q es la probabilidad que tiene un objeto dado de caer en esa caja. Si cada caja es equiprobable, se tiene $q = 1/m$. Si se toma un número cada vez mayor de cajas, entonces $q \rightarrow 0$. Supongamos que la proporción $\alpha = n/m = nq$ es constante (o tiende a una constante). Entonces se tiene

$$\binom{n}{r} q^r (1-q)^{n-r} = \frac{1}{r!} \left(1 - \frac{\alpha}{n}\right)^{n-r} \alpha^r \prod_{i=0}^{r-1} \frac{n-i}{n} \rightarrow \frac{e^{-\alpha} \alpha^r}{r!}.$$

La Distribución de Poisson se utiliza por ejemplo para estimar la probabilidad de encontrar un número dado n de peces en una pequeña porción de un estanque (que hace el papel de caja), asumiendo que la densidad de peces en el estanque es conocida. También puede utilizarse inversamente, para conocer la densidad de peces en el estanque observando la distribución del número de peces en una región pequeña.

Una variable aleatoria se define como una función arbitraria $X : \Omega \rightarrow \mathbb{R} \cup \{\infty\}$. Diremos que es finita si su imagen está contenida en \mathbb{R} . La probabilidad $P(X = r)$ se define como antes. En el caso numerable no es

necesariamente cierto que $P(X = r)$ es no nula para un número finito de valores de r (sólo puede afirmarse que el número de tales r es numerable). Sin embargo, si se obtiene el siguiente resultado.

Proposición 2.5. *Si X es una variable aleatoria finita en un espacio de probabilidad numerable, entonces $P(X > t) \rightarrow 0$ cuando $t \rightarrow \infty$.*

Demostración Sea $\epsilon > 0$. Se sigue de la definición que si $R(X)$ denota el rango de X , entonces

$$\sum_{r \in R(X)} P(X = r) = 1.$$

Enumeremos los elementos de $R(X) = \{r_1, \dots\}$. Existe un $N > 0$ tal que

$$\sum_{i=1}^N P(X = r_i) > 1 - \epsilon.$$

Ahora si $T > r_1, \dots, r_N$, se tiene que $X > T$ implica $T \neq r_1, \dots, r_n$, por lo que $P(X > T) < P(X \neq r_1, \dots, r_n) < \epsilon$. \square

definición 2.6. La esperanza $E(X)$ de una variable aleatoria X se define mediante

$$E(X) = \sum_{i=1}^{\infty} X(\omega_i) p(\omega_i),$$

si esta serie resulta ser absolutamente convergente. En tal caso diremos que la variable aleatoria X tiene esperanza definida o que es integrable. Si X es integrable el orden de los sumandos no interesa y puede escribirse

$$E(X) = \sum_{\omega \in \Omega} X(\omega) p(\omega).$$

Nótese que este es siempre el caso si X es una variable aleatoria acotada, ya que la serie de término general $p(\omega)$ es absolutamente convergente. Si es necesario recordar la medida P correspondiente a la función de probabilidad p , puede utilizarse la notación $E_P(X) = E(X)$, o bien la notación analítica

$$E(X) = \int_{\Omega} X(\omega) dP(\omega).$$

ejemplo 2.7. Nótese que resulta muy sencillo definir variables aleatorias finitas cuya esperanza no converge: Por ejemplo, si se lanza una moneda simétrica i veces para obtener una cara, se tiene que la variable aleatoria $X(\omega_i) = 2^i$ no tiene esperanza definida, ya que la serie

$$\sum_{i=1}^{\infty} X(\omega_i)p(\omega_i) = \sum_{i=1}^{\infty} 2^i 2^{-i} = \sum_{i=1}^{\infty} 1$$

no converge.

ejemplo 2.8. Si $B \subseteq \Omega$ es un evento, la función característica χ_B es una variable aleatoria. Su esperanza está dada por

$$E(\chi_B) = \sum_{i=1}^{\infty} \chi_B(\omega_i)p(\omega_i) = \sum_{\omega_i \in B} p(\omega_i) = P(B).$$

ejemplo 2.9. Si X es una variable aleatoria y $F : \mathbb{R} \cup \{\infty\} \rightarrow \mathbb{R} \cup \{\infty\}$ es una función arbitraria, entonces $F(X) = F \circ X$ es una variable aleatoria. Su esperanza está dada por

$$E[F(X)] = \sum_{i=1}^{\infty} F[X(\omega_i)]p(\omega_i),$$

si dicha serie converge. Un caso particular importante es el de la función $F(x) = |x|$. En este caso

$$E(|X|) = \sum_{i=1}^{\infty} |X(\omega_i)|p(\omega_i).$$

Nótese que en particular la condición de convergencia absoluta en la definición de variable aleatoria integrable implica que X es integrable si y sólo si $|X|$ es integrable.

Proposición 2.10. *La función $X \mapsto E(X)$ es lineal en su dominio de definición y la esperanza de la función constante c es $E(c) = c$.*

Demostración Tal como en el caso finito, podemos escribir

$$E(\alpha X + \beta Y) = \sum_{\omega \in \Omega} [\alpha X(\omega) + \beta Y(\omega)]p(\omega) =$$

$$\alpha \sum_{\omega \in \Omega} X(\omega)p(\omega) + \beta \sum_{\omega \in \Omega} Y(\omega)p(\omega) = \alpha E(X) + \beta E(Y),$$

donde la convergencia absoluta de la primera suma está garantizada por la convergencia absoluta de las restantes. Por otro lado

$$E(c) = \sum_{\omega \in \Omega} cp(\omega) = c \sum_{\omega \in \Omega} p(\omega) = c \times 1 = c.$$

□

Proposición 2.11. *La función $X \mapsto E(X)$ es monótona, en el sentido de que si $X(\omega) < Y(\omega)$ para todo $\omega \in \Omega$, se tiene $E(X) < E(Y)$.*

Demostración

$$E(X) = \sum_{\omega \in \Omega} X(\omega)p(\omega) \leq \sum_{\omega \in \Omega} Y(\omega)p(\omega) = E(Y).$$

□

ejemplo 2.12. Si $K_1 < X(\omega) < K_2$ para todo $\omega \in \Omega$, se tiene $K_1 < E(X) < K_2$.

ejemplo 2.13. Si X es una variable aleatoria integrable arbitraria se tiene $-|X| \leq X \leq |X|$, por lo que $-E(|X|) \leq E(X) \leq E(|X|)$, es decir $|E(X)| \leq E(|X|)$.

ejemplo 2.14. Si X es una variable aleatoria finita no negativa, entonces se tiene $X\chi_A \leq X\chi_B$ para todo par de subconjuntos A y B de Ω con $A \subseteq B$, por lo que $E(X\chi_A) \leq E(X\chi_B)$. En particular, tomando $B = \Omega$ se tiene $E(X\chi_A) \leq E(X)$.

ejemplo 2.15. Si X es una variable aleatoria integrable, entonces se tiene para todo subconjunto B de Ω la identidad

$$E(X\chi_B) = \sum_{\omega \in \Omega} X(\omega)\chi_B(\omega)p(\omega) = \sum_{\omega \in \Omega} X(\omega)p(\omega|B)P(B) = E(X|B)P(B).$$

Los conceptos de independencia de conjuntos y variables aleatorias, así como las definiciones de probabilidad y esperanza condicionada se traducen literalmente al caso numerable. También se traduce literalmente la fórmula

$$E(X) = \sum_{B_i \in \mathbb{P}} E(X|B_i)P(B_i),$$

si $\mathbb{P} = \{B_1, \dots, B_n\}$ es una partición finita de Ω . Esta fórmula puede demostrarse alternativamente utilizando las funciones características. De hecho, siendo \mathbb{P} una partición finita se tiene

$$\sum_{i=1}^n \chi_{B_i} = 1,$$

de donde

$$E(X) = E\left(X \sum_{i=1}^n \chi_{B_i}\right) = \sum_{i=1}^n E(X \chi_{B_i}) = \sum_{i=1}^n E(X|B_i)P(B_i).$$

Diremos que una sucesión $\{X_n\}_{n \in \mathbb{N}}$ de variables aleatorias converge a una variable aleatoria X si para todo $\omega \in \Omega$ se tiene $X_n(\omega) \rightarrow X(\omega)$.

Proposición 2.16. *Supongamos que la sucesión $\{X_n\}_{n \in \mathbb{N}}$ de variables aleatorias acotadas por una misma constante K converge a una variable aleatoria X . Entonces $E(X_n)$ converge a $E(X)$.*

Demostración Sea $\epsilon > 0$. Para cada $\omega \in \Omega$, sea $N = N(\epsilon, \omega)$ el menor entero tal que $|X_n(\omega) - X(\omega)| < \epsilon$ para todo $n > N$. La función $\omega \mapsto N(\epsilon, \omega)$ es una variable aleatoria finita (por hipótesis) que denotaremos N_ϵ . En particular, $P(N_\epsilon > n) \rightarrow 0$ cuando $n \rightarrow \infty$. Ahora bien:

$$E(|X_n - X|) = E\left(|X_n - X| \mathbb{1}_{N_\epsilon > n}\right) P(N_\epsilon > n) +$$

$$E\left(|X_n - X| \mathbb{1}_{N_\epsilon \leq n}\right) P(N_\epsilon \leq n) \leq 2KP(N_\epsilon > n) + \epsilon \rightarrow \epsilon.$$

Como $\epsilon > 0$ era arbitrario, se concluye $E(|X_n - X|) \rightarrow 0$ cuando $n \rightarrow \infty$. Esto implica

$$0 \leq \left|E(X_n) - E(X)\right| \leq E(|X_n - X|) \rightarrow 0,$$

de donde $E(X_n) \rightarrow E(X)$. □

Proposición 2.17. *Supongamos que sucesión $\{X_n\}_{n \in \mathbb{N}}$ de variables aleatorias converge uniformemente a una variable aleatoria X . Entonces $E(X_n)$ converge a $E(X)$.*

Demostración Basta tomar $Y_n = X_n - X$ y observar que $|Y_n|$ está acotado por 1 para n suficientemente grande, por lo que $E(Y_n) \rightarrow 0$. \square

Nótese que es posible dar una demostración directa del corolario observando que para n suficientemente grande se tiene $X - \epsilon < X_n < X + \epsilon$. Los detalles se dejan al lector.

Hay un segundo teorema de convergencia mas fuerte que el de convergencia acotada que utilizaremos en lo que sigue. Para ello necesitamos alguna preparación.

Proposición 2.18. *Sea Y una V.A. positiva e integrable. Existe una medida de probabilidad P_Y tal que para toda variable aleatoria X se tiene $E_Y(X) = E(XY)/E(Y)$ si algún lado de la ecuación converge.*

Demostración Para cada $\omega \in \Omega$ se define $p_Y(\omega) = Y(\omega)p(\omega)/E(Y)$. Dejamos al lector la tarea de comprobar que esta fórmula define una función de probabilidad. La esperanza de X con respecto a esta medida está dada por

$$E_Y(X) = \sum_{\omega \in \Omega} X(\omega)p_Y(\omega) = \sum_{\omega \in \Omega} X(\omega)Y(\omega)p(\omega)/E(Y),$$

de donde se sigue lo pedido. \square

Proposición 2.19. *Supongamos que la sucesión $\{X_n\}_{n \in \mathbb{N}}$ de variables aleatorias acotadas por una misma V.A. positiva e integrable Y converge a una variable aleatoria X . Entonces $E(X_n)$ converge a $E(X)$.*

Demostración Se define $Z_n = X_n/Y$ cuando $Y \neq 0$, y $Z_n = 0$ en caso contrario. Nótese que $X_n = Z_n Y$. Se define Z del mismo modo, así que $X = ZY$. La definición implica que $Z_n \rightarrow Z$ para todo punto $\omega \in \Omega$. Como X_n está acotado por Y , se tiene que $Z_n \leq 1$. Ahora se tiene

$$E(X_n) = E(Z_n Y) = E_Y(Z_n) \rightarrow E_Y(Z) = E(X).$$

\square

Proposición 2.20. *Si X es una variable aleatoria y $F : \mathbb{R} \rightarrow \mathbb{R}$ es una función, entonces toda variable aleatoria Y que es independiente con X es independiente con $F(X)$.*

Demostración Sean $r, s \in \mathbb{R}$. Basta probar que

$$P\left(F(X) = r, Y = s\right) = P\left(F(X) = r\right)P(Y = s).$$

Para eso observamos que $F(X) = r$ implica $X = t$ para algún $t \in F^{-1}(r)$. Se sigue que

$$\begin{aligned} P\left(F(X) = r, Y = s\right) &= \sum_{t \in F^{-1}(r)} P(X = t, Y = s) \\ &= \sum_{t \in F^{-1}(r)} P(X = t)P(Y = s) = P\left(F(X) = r\right)P(Y = s). \end{aligned}$$

□

definición 2.21. Una variable aleatoria $X : \Omega \rightarrow \mathbb{R}$ se dice simple si existe una partición $\mathbb{P} = \{B_1, \dots, B_n\}$ de Ω tal que X es constante en cada conjunto B_i . En este caso puede considerarse a X como una función definida en \mathbb{P} y el cálculo de su esperanza se reduce al cálculo de la esperanza de una variable aleatoria definida en un espacio finito. Utilizaremos a menudo esta observación en lo que sigue.

Proposición 2.22. Si X e Y son variables aleatorias independientes e integrables y al menos una es simple, se tiene $E(XY) = E(X)E(Y)$.

Demostración Suponiendo que Y es simple, se tiene que su rango $R(Y)$ es finito y se tiene, tal como en el caso finito:

$$\begin{aligned} E(XY) &= \sum_{r \in R(Y)} E(XY|Y = r)P(Y = r) \\ &= \sum_{r \in R(Y)} E(Xr|Y = r)P(Y = r) = \sum_{r \in R(X)} rE(X|Y = r)P(Y = r) \\ &= \sum_{r \in R(Y)} rE(X)P(Y = r) = E(X)E(Y). \end{aligned}$$

□

Proposición 2.23. Toda variable aleatoria acotada X es límite uniforme de variables aleatorias simples que son funciones en X .

Demostración Sea X una variable aleatoria tal que para todo $\omega \in \Omega$ se tiene $-K \leq X(\omega) \leq K$. Se define $X_n(\omega) = \frac{k}{n}$ si $\frac{k}{n} \leq X(\omega) < \frac{k+1}{n}$ es inmediato que para todo n y todo $\omega \in \Omega$ se tiene $|X_n(\omega) - X(\omega)| < \frac{1}{n}$. El hecho de que cada X_n es simple sigue si observamos que X_n toma sólo valores de la forma $\frac{k}{n}$ con $-K - \frac{1}{n} < \frac{k}{n} < K$ por lo que hay sólo una cantidad finita de valores posibles. Por otro lado, es inmediato que $X_n = \frac{1}{n}[nX]$ donde $[a]$ denota la función parte entera de a . \square

Proposición 2.24. *Si X e Y son variables aleatorias independientes e integrables y al menos una es acotada, se tiene $E(XY) = E(X)E(Y)$.*

Demostración Si $-K \leq X \leq K$ escribimos X como un límite uniforme de variables aleatorias X_n que son también independientes de Y . Como $|YX_n|$ está acotada por la V.A. integrable KY , se tiene por un lado $E(X_nY) \rightarrow E(XY)$ y por otro $E(X_nY) = E(X_n)E(Y) \rightarrow E(X)E(Y)$. \square

El resultado anterior es válido bajo la hipótesis de independencia e integridad de ambas variables solamente. A fin de demostrarlo, necesitaremos un teorema de descomposición de la esperanza para particiones numerables.

Proposición 2.25. *Sea X una variable aleatoria integrable. Entonces*

$$E(X|X \geq n)P(X \geq n) \rightarrow 0, \text{ cuando } n \rightarrow \infty.$$

Demostración Sea Y_n la variable aleatoria que vale 0 si $X < n$ y 1 si no. Entonces

$$\begin{aligned} E(XY_n) &= E(XY_n|X \geq n)P(X \geq n) + E(XY_n|X < n)P(X < n) \\ &= E(X|X \geq n)P(X \geq n). \end{aligned}$$

Se sigue que

$$E(X|X \geq n)P(X \geq n) = E(XY_n) = E_X(Y_n)E(X) = P_X(X \geq n),$$

y el resultado sigue de la Proposición 2.5 que el último término tiende a cero. \square

Proposición 2.26. *Sea X una variable aleatoria integrable. si $\varphi = \{B_1, B_2, \dots\}$ es una partición numerable de Ω , se tiene*

$$E(X) = \sum_{i=1}^{\infty} E(X|B_i)P(B_i),$$

y esta serie es absolutamente convergente.

Demostración Nótese que

$$\left| E(X|B_i)P(B_i) \right| \leq E(|X||B_i)P(B_i).$$

Remplazando X por $|X|$ si es necesario podemos asumir que X es positiva. Sea T una variable aleatoria tal que $T(\omega) = n$ si y sólo si $\omega \in B_n$. Para cada número natural N la partición

$$\wp_N = \{B_1, \dots, B_N, C_N\}$$

donde $C_N = \bigcup_{i=N+1}^{\infty} B_i$, satisface

$$E(X) = \sum_{i=1}^N E(X|B_i)P(B_i) + E(X|C_N)P(C_N),$$

por lo que basta probar que $E(X|C_N)P(C_N)$ converge a 0. Obsérvese que $\omega \in C_N$ si y sólo si $T(\omega) > N$ por lo que el resultado sigue de la proposición precedente. \square

En particular, si tomamos la partición formada por los conjuntos

$$A_r(X) = \{\omega \in \Omega | X(\omega) = r\},$$

se tiene:

Corolario 2.26.1. *Si X una variable aleatoria integrable, entonces*

$$E(X) = \sum_{r \in R(X)} rP(X = r),$$

donde el rango esencial $R(X)$ puede remplazarse por cualquier conjunto mayor. \square

Tal como en el caso finito, podemos escribir

$$E(X) = \sum_{r \in \mathbb{R}} rP(X = r),$$

con el entendido de que esta suma tiene un número numerable de términos no nulos y es absolutamente convergente.

Mas generalmente, con las mismas convenciones, se tiene:

Corolario 2.26.2. Si X una variable aleatoria tal que $F(X)$ es integrable, entonces

$$E[F(X)] = \sum_{r \in \mathbb{R}} F(r)P(X = r).$$

□

Proposición 2.27. Si X e Y son variables aleatorias independientes e integrables, entonces XY es también integrable y se tiene $E(XY) = E(X)E(Y)$.

Demostración Supongamos primero que se sabe de antemano que XY es integrable. Entonces podemos escribir como en el caso finito:

$$\begin{aligned} E(XY) &= \sum_{r \in R(Y)} E(XY|Y = r)P(Y = r) = \sum_{r \in R(Y)} E(Xr|Y = r)P(Y = r) \\ &= \sum_{r \in R(Y)} rE(X|Y = r)P(Y = r) = \sum_{r \in R(Y)} rE(X)P(Y = r) = E(X)E(Y), \end{aligned}$$

utilizando el resultado precedente. En el caso general, basta probar que $|XY| = |X||Y|$ es integrable, por lo que podemos suponer que X e Y son positivas. Además, si definimos $X_n = \min\{X, n\}$, se tiene que el producto X_nY es integrable ya que X_n es acotada y por lo tanto $E(X_nY) = E(X_n)E(Y) \leq E(X)E(Y)$. Como cada sub-suma finita de la suma

$$\sum_{\omega \in \Omega} X(\omega)Y(\omega)p(\omega) \tag{2.1}$$

que define $E(XY)$ es sub-suma de la suma correspondiente que define a $E(X_nY)$ para algún n , esta está acotada por $E(X)E(Y)$ por lo que (2.1) converge. □

Chapter 3

Espacios de probabilidad.

La teoría desarrollada hasta aquí permite estudiar un gran número de problemas, pero hay situaciones que no nos permite describir. Por ejemplo si queremos construir un espacio asociado al problema siguiente:

Se arroja una moneda indefinidamente. Cual es la probabilidad de que el número de caras nunca supere al de sellos durante todo el proceso?

Nos gustaría estudiar este problema definiendo un espacio de probabilidad en el cual se tuviera un punto por cada posible sucesión (infinita) de resultados cara o sello. Nótese que el conjunto de tales sucesiones no es numerable. Otro problema de este tipo es el siguiente:

Se escoge un número al azar entre 0 y 1 de modo que para cada intervalo $[a, b]$ la probabilidad de escoger un elemento de este es $b - a$.

En este capítulo desarrollaremos una teoría de probabilidades mas general, que puede aplicarse a problemas de este tipo. Por simplicidad damos aquí solamente las definiciones generales, posponiendo los resultados de existencia (por ejemplo, de los espacios antes mencionados) a un capítulo posterior (aún no escrito).

σ -Algebras

Antes de dar la definición general de espacio de probabilidad, es necesario hacer una observación. Hasta aquí la probabilidad era una función definida

en un subconjunto *arbitrario* del espacio muestral Ω . Sin embargo, es posible demostrar que no existe ninguna medida de probabilidad definida en todos los subconjuntos de $[0, 1]$ tal que la medida de un intervalo $[a, b]$ sea $b - a$. Es necesario por lo tanto restringir nuestra definición a una colección menor de subconjuntos. Esta colección debe, sin embargo, ser lo bastante amplia para incluir uniones numerables. A continuación formalizaremos este concepto:

definición 3.1. Sea Ω un conjunto arbitrario. Una σ -álgebra en Ω es una colección Σ de subconjuntos de Ω que satisface las condiciones siguientes:

1. $\emptyset \in \Sigma$.
2. Si $A \in \Sigma$ entonces su complemento A^c está en Σ .
3. Si $\{A_i\}_{i \in \mathbb{N}}$ es una familia numerable de conjuntos en Σ , entonces su unión $A = \bigcup_{i \in \mathbb{N}} A_i$ está en Σ .

Nótese que las leyes de De Morgan implican que una σ -álgebra es también cerrada bajo intersecciones numerables, puesto que si $\{A_i\}_{i \in \mathbb{N}}$ es una familia numerable de conjuntos en Σ , entonces

$$A = \bigcap_{i \in \mathbb{N}} A_i = \left(\bigcup_{i \in \mathbb{N}} A_i^c \right)^c \in \Sigma.$$

definición 3.2. Un espacio medible es un par (Ω, Σ) donde Ω es un conjunto y Σ es una σ -álgebra en Ω .

ejemplo 3.3. Si Ω es un conjunto arbitrario, la colección $\wp(\Omega)$ de todos los subconjuntos de Ω es una σ -álgebra. Si Ω es finito o numerable, esta es la única σ -álgebra que contiene a todos los conjuntos unitarios $\{\omega\}$.

ejemplo 3.4. Si Ω es un conjunto arbitrario, la colección \mathfrak{X} de todos los subconjuntos A de Ω tales que uno de los conjuntos A o A^c es numerable es una σ -álgebra.

ejemplo 3.5. Si $f : \Delta \rightarrow \Omega$ es una función arbitraria, y si A es un subconjunto de Ω , su pre-imágen $f^{-1}(A)$ se define por:

$$f^{-1}(A) = \{\delta \in \Delta \mid f(\delta) \in A\}.$$

Dejamos al lector la comprobación de las propiedades básicas de la pre-imágen:

$$f^{-1}(\emptyset) = \emptyset, \quad f^{-1}(A^c) = f^{-1}(A)^c, \quad f^{-1}\left(\bigcup_{i \in \mathbb{N}} A_i\right) = \bigcup_{i \in \mathbb{N}} f^{-1}(A_i).$$

Si Σ es una σ -álgebra en Ω , su pre-imágen es la colección

$$f^*(\Sigma) = \{f^{-1}(A) \mid A \in \Sigma\}.$$

De hecho, $f^*(\Sigma)$ es una σ -álgebra en Δ , como se deduce de las identidades precedentes.

ejemplo 3.6. Si $f : \Omega \rightarrow \Delta$ es una función arbitraria, y si Σ es una σ -álgebra en Ω , la colección

$$f_*(\Sigma) = \{B \subseteq \Delta \mid f^{-1}(B) \in \Sigma\}$$

es una σ -álgebra en Δ .

ejemplo 3.7. Toda intersección de σ -álgebras es una σ -álgebra. En particular, si T es una colección de subconjuntos de Ω , existe una menor σ -álgebra $\Sigma(T)$ que contiene a T y se llama la σ -álgebra generada por T . De hecho $\Sigma(T)$ puede caracterizarse como la intersección de todas las σ -álgebras que contienen a T . Los elementos de $\Sigma(T)$ se obtienen a partir de los elementos de T mediante sucesivos complementos y uniones numerables.

ejemplo 3.8. La σ -álgebra en \mathbb{R} generada por el conjunto de todos los intervalos del tipo $[a, b[$ se denomina la σ -álgebra de Borel de \mathbb{R} . Los elementos de esta σ -álgebra se llaman borelianos. Observese que los siguientes tipos de conjuntos son borelianos:

1. $\{r\} = \bigcap_{n=1}^{\infty} [r, r + 1/n[$.
2. $[a, b] = \bigcap_{n=1}^{\infty} [a, b + 1/n[$.
3. $]a, b[= \bigcup_{n=1}^{\infty} [a + 1/n, b[$.

Lo mismo sucede para cualquier otro tipo de intervalo.

Medidas de Probabilidad

definición 3.9. Sea (Ω, Σ) un espacio medible. Una medida en este espacio es una función $P : \Sigma \rightarrow [0, \infty]$ que satisface las condiciones siguientes:

1. $P(\emptyset) = 0$.
2. Si $\{A_i\}_{i \in \mathbb{N}}$ es una familia numerable de conjuntos disjuntos en Σ , entonces su unión $A = \bigcup_{i \in \mathbb{N}} A_i$ satisface

$$P(A) = \sum_{i=1}^{\infty} P(A_i).$$

En este caso el trío (Ω, Σ, P) se denomina un espacio de medida. Si $P(\Omega) = 1$ se dice que (Ω, Σ, P) es un espacio de probabilidad.

ejemplo 3.10. Si Ω es numerable, cada función de probabilidad p definida en Ω define una medida de probabilidad P en la σ -álgebra $\wp(\Omega)$ de todos los subconjuntos de Ω mediante $P(B) = \sum_{\omega \in B} p(\omega)$.

ejemplo 3.11. Mas generalmente, si \wp es una partición numerable de Ω , la colección de todos los subconjuntos de ω que pueden escribirse como unión de elementos de \wp forman una σ -álgebra Σ_\wp . Toda función de probabilidad en \wp define una medida de probabilidad en (Ω, Σ_\wp) .

ejemplo 3.12. Si $f : \Omega \rightarrow \Delta$ es una función arbitraria, y si P es una medida de probabilidad en (Ω, Σ) , entonces existe una medida de probabilidad en $f_*(\Sigma)$ definida por $P_f(B) = P[f^{-1}(B)]$.

ejemplo 3.13. Si P es una medida de probabilidad definida en (Ω, Σ) , y si Π es una σ -álgebra contenida en Σ , la restricción de P a Π define una medida en (Ω, Π) .

ejemplo 3.14. Si Σ es la σ -álgebra definida el ejemplo 3.4, sobre un espacio muestral no numerable Ω , entonces se define una medida de probabilidad en Σ mediante $P(A) = 0$ si A es numerable y $P(A) = 1$ en caso contrario.

definición 3.15. Una Variable Aleatoria en un espacio medible (Ω, Σ) es una función $X : \Omega \rightarrow \mathbb{R} \cup \{\infty\}$, tal que para cada intervalo $[a, b[$ se tiene

$$X^{-1}[a, b[= \left\{ \omega \in \Omega \mid X(\omega) \in [a, b[\right\} \in \Sigma.$$

En otras palabras, si (Ω, Σ, P) es un espacio de medida, la función $X : \Omega \rightarrow \mathbb{R}$ es una variable aleatoria si la probabilidad

$$P(a \leq X < b) = P(X \in [a, b]) = P_X([a, b])$$

está definida. Aquí P_X está definida en el ejemplo 3.12. Nótese que los conjuntos B para los cuales

$$X^{-1}(B) = \{\omega \in \Omega | X(\omega) \in B\} \in \Sigma$$

forman una σ -álgebra por el ejemplo 3.6, de modo que en particular $P(X \in B)$ está definida para cualquier boreliano B . Como

$$\{\omega \in \Omega | a \leq X(\omega) < b\} = \{\omega \in \Omega | X(\omega) < b\} \cap \{\omega \in \Omega | X(\omega) < a\}^c,$$

es suficiente con probar que $\{\omega \in \Omega | X(\omega) < b\} \in \Sigma$, o en otras palabras que la probabilidad $P(X < b)$ está definida. La función definida por $F(b) = P(X < b)$ recibe el nombre de función de distribución de X .

Observación 3.16. Si denotamos por \mathfrak{B} la σ -álgebra en \mathbb{R} formada por todos los borelianos, entonces para toda variable aleatoria X la función P_X definida por

$$P_X(B) = P(X \in B)$$

es una medida de probabilidad en el espacio $(\mathbb{R}, \mathfrak{B})$. Se llama la medida de Borel asociada a la variable aleatoria X .

Proposición 3.17. *Sumas de variables aleatorias son variables aleatorias.*

Demostración Nótese que si $x + y < b$, se tiene $y < b - x$. Sea ahora $\{r_n\}_n$ una sucesión de números racionales que converge a x por la derecha. Entonces $y < b - r_n$ para n suficientemente grande. Se sigue que existe algún número racional r tal que $x < r$ e $y < b - r$. Se concluye que

$$\{\omega \in \Omega | X(\omega) + Y(\omega) < b\} =$$

$$\bigcup_{r \in \mathbb{Q}} \{\omega \in \Omega | X(\omega) < r\} \cap \{\omega \in \Omega | Y(\omega) < b - r\} \in \Sigma.$$

□

Proposición 3.18. *Si X es una variable aleatoria y $G : \mathbb{R} \rightarrow \mathbb{R}$ es una función monótona, entonces $G(X) = G \circ X$ es una variable aleatoria.*

Demostración Basta ver que si G es creciente entonces para todo intervalo I en \mathbb{R} , su preimagen $G^{-1}(I)$ es un intervalo, ya que entonces

$$\{\omega \in \Omega | G[X(\omega)] \in I\} = \{\omega \in \Omega | X(\omega) \in G^{-1}(I)\} \in \Sigma.$$

Si a y b son elementos de $G^{-1}(I)$, entonces para todo c entre a y b el elemento $G(c)$ debe estar entre $G(a)$ y $G(b)$ por la monotonía de G y por lo tanto $G(c) \in I$, de donde $c \in G^{-1}(I)$. \square

Corolario 3.18.1. *Si X es una variable aleatoria, entonces αX también lo es para toda constante α . En particular combinaciones lineales de variables aleatorias son variables aleatorias.* \square

Corolario 3.18.2. *Si X es una variable aleatoria y $G : \mathbb{R} \rightarrow \mathbb{R}$ es una función continua con un número finito de máximos o mínimos, entonces $G(X) = G \circ X$ es una variable aleatoria.*

Demostración Supongamos que G es monótona en cada uno de los intervalos $] - \infty, x_0], [x_0, x_1], \dots, [x_{n-1}, x_n], [x_n, \infty[$. Basta escribir

$$G(x) = \sum_{i=0}^{n+1} G_i(x),$$

donde las funciones monótonas G_i se definen por

$$G_0(x) = \begin{cases} G(x) & \text{if } x \leq x_0 \\ G(x_0) & \text{if } x \geq x_0 \end{cases},$$

$$G_i(x) = \begin{cases} 0 & \text{if } x \leq x_{i-1} \\ G(x) - G(x_{i-1}) & \text{if } x_{i-1} \leq x \leq x_i \\ G(x_i) - G(x_{i-1}) & \text{if } x \geq x_i \end{cases},$$

para $1 \leq i \leq n$, y finalmente

$$G_{n+1}(x) = \begin{cases} 0 & \text{if } x \leq x_n \\ G(x) - G(x_n) & \text{if } x \geq x_n \end{cases}.$$

\square

Demostraremos al final de este capítulo que $G(X)$ es una variable aleatoria para toda función continua G .

Proposición 3.19. *Productos de variables aleatorias son variables aleatorias.*

Demostración Si $b, x, e y$ son positivos, entonces $xy < b$ es equivalente a $x < r$ e $y < b/r$ para algún racional positivo r , por lo que si X e Y fuesen variables aleatorias positivas el resultado sigue como en el caso de la suma. Para el caso general, escribimos $X = G_1(X) - G_2(X)$ e $Y = G_1(Y) - G_2(Y)$, donde

$$G_1(t) = 1 + \max\{0, t\}, \quad G_2(t) = 1 + \max\{0, -t\},$$

y observamos que

$$XY = G_1(X)G_1(Y) - G_1(X)G_2(Y) - G_1(X)G_2(Y) + G_1(X)G_2(Y).$$

□

definición 3.20. Una Variable Aleatoria se dice discreta si su imagen

$$X(\Omega) = \{X(\omega) | \omega \in \Omega\}$$

es numerable. La esperanza de una variable aleatoria discreta X se define por:

$$E(X) = \sum_{r \in \mathbb{R}} rP(X = r) = \sum_{r \in X(\Omega)} rP(X = r).$$

La probabilidad $P(X = r)$ está definida ya que $\{r\}$ es un boreliano.

Una partición (numerable) φ se dice mas fina que una partición φ' si todo conjunto de φ está contenido en un conjunto de φ' . Denotemos por φ_X la partición formada por los conjuntos $A_r(X)$. Entonces para cualquier partición φ mas fina que φ_X se tiene

$$E(X) = \sum_{A \in \varphi} X(a)P(A),$$

donde a es un elemento arbitrario de A . Nótese que si φ es cualquier partición numerable contenida en Σ , entonces

$$\varphi \cap \varphi_X = \{A \cap A_X | A \in \varphi, A_X \in \varphi_X\}$$

es también una partición numerable contenida en Σ y es más fina que que φ_X . La variable aleatoria X puede interpretarse como una variable aleatoria definida en el conjunto numerable $\varphi \cap \varphi_X$, de donde se obtiene, utilizando los resultados del capítulo anterior, que

$$E(X) = \sum_{\substack{A \in \varphi \\ P(A) \neq 0}} E(X|A)P(A).$$

La esperanza condicional en la fórmula precedente se define como la esperanza con respecto a la probabilidad condicional, la que también se interpreta como una probabilidad en el conjunto numerable $\wp \cap \wp_X$. Del mismo modo se demuestra que si X e Y son variables aleatorias discretas, podemos considerarlas como variables aleatorias en el conjunto numerable $\wp_X \cap \wp_Y$, de modo que cualquier combinación lineal $\alpha X + \beta Y$ es discreta y se tiene $E(\alpha X + \beta Y) = \alpha E(X) + \beta E(Y)$. Asimismo, la definición de independencia se extiende fácilmente a variables aleatorias discretas. Si X e Y son variables aleatorias discretas, integrables, e independientes, entonces la esperanza de XY está definida y se tiene

$$E(XY) = E(X)E(Y).$$

definición 3.21. Se dice que una variable aleatoria X es integrable si

$$\sum_{n \in \mathbb{Z}} nP(X \in [n, n+1[)$$

es absolutamente convergente. En otras palabras, una variable aleatoria arbitraria X es integrable si y sólo si la variable aleatoria discreta $X_1 = [X]$ tiene esperanza bien definida. En particular, si X es discreta, se tiene $[X] \leq X \leq [X] + 1$, por lo que X también tiene esperanza bien definida. De hecho si definimos $X_n = \frac{1}{n}[nX]$, entonces $0 \leq E(X - X_n) \leq \frac{1}{n}$, de modo que

$$E(X) = \lim_{n \rightarrow \infty} E(X_n). \quad (3.1)$$

Sin embargo, las variables aleatorias $X_n = \frac{1}{n}[nX]$ están definidas y son discretas para toda variable aleatoria X . Se sigue que podemos utilizar la ecuación (3.1) para definir la esperanza de una variable aleatoria arbitraria. La desigualdad $X_n \leq X_1 + 1$ prueba que cada variable aleatoria X_n tiene esperanza bien definida y las desigualdades

$$X_n \leq X \leq X_n + \frac{1}{n}, \quad X_m \leq X \leq X_m + \frac{1}{m},$$

implican $|X_n - X_m| \leq \text{Max}\{\frac{1}{n}, \frac{1}{m}\}$, por lo que se tiene

$$|E(X_n) - E(X_m)| \leq E|X_n - X_m| \leq \text{Min}\left\{\frac{1}{n}, \frac{1}{m}\right\}$$

para todo par de enteros m y n . Se sigue que la sucesión $\{E(X_n)\}$ es de Cauchy y por lo tanto converge.

Proposición 3.22. *La esperanza es lineal.*

Demostración Probaremos primero que $E(X) = -E(-X)$. Para esto observamos que si $\frac{k}{n} \leq x < \frac{k+1}{n}$ entonces $\frac{-k-1}{n} < -x \leq \frac{-k}{n}$. Se sigue que para todo $\omega \in \Omega$, $(-X)_n(\omega)$ es igual a $-X_n(\omega)$ o a $-X_n(\omega) - \frac{1}{n}$. En todo caso

$$(-X)_n \leq -X_n \leq (-X)_n + \frac{1}{n},$$

de donde

$$E(-X)_n \leq -E(X)_n \leq E(-X)_n + \frac{1}{n},$$

y el resultado sigue. Supongamos ahora que α y β son positivos. Se sigue de la definición que

$$\alpha X + \beta Y \leq (\alpha X + \beta Y)_n \leq \alpha X + \beta Y + \frac{1}{n},$$

y del mismo modo

$$\alpha X + \beta Y \leq \alpha X_n + \beta Y_n \leq \alpha \left(X + \frac{1}{n} \right) + \beta \left(Y + \frac{1}{n} \right),$$

de donde

$$\frac{-\alpha - \beta}{n} \leq (\alpha X + \beta Y)_n - (\alpha X_n + \beta Y_n) \leq \frac{1}{n},$$

y la esperanza del término central converge a $E(\alpha X + \beta Y) - \alpha E(X) - \beta E(Y)$. El caso general sigue ahora fácilmente. Por ejemplo, si α es negativo y β positivo, escribimos

$$\begin{aligned} E(\alpha X + \beta Y) &= E[(-\alpha)(-X) + \beta Y] = \\ &(-\alpha)E(-X) + \beta E(Y) = \alpha E(X) + \beta E(Y). \end{aligned}$$

□

Proposición 3.23. *Si X es una variable aleatoria integrable, y si \wp es una partición numerable arbitraria, entonces*

$$E(X) = \sum_{A \in \wp} E(X|A)P(A).$$

Demostración Basta ver que como $X_n \leq X \leq X_n + \frac{1}{n}$, se tiene

$$\begin{aligned} E(X_n) &= \sum_{A \in \varphi} E(X_n|A)P(A) \leq \sum_{A \in \varphi} E(X|A)P(A) \\ &\leq \sum_{A \in \varphi} \left(E(X_n|A) + \frac{1}{n} \right) P(A) = E(X_n) + \frac{1}{n}, \end{aligned}$$

y cada extremo de esta desigualdad converge a $E(X)$. \square

En particular, para todo evento fijo C de probabilidad no nula puede aplicarse esta fórmula a la probabilidad condicional $P(\cdot|C)$ y obtener

$$E(X|C) = \sum_{\substack{A \in \varphi \\ P(A|C) \neq 0}} E(X|A \cap C)P(A|C).$$

En particular, si C es la union de alguna subcolección de la particion φ , se tiene

$$E(X|C) = \sum_{\substack{A \in \varphi \\ P(A) \neq 0 \\ A \subseteq C}} E(X|A) \frac{P(A)}{P(C)},$$

es decir

$$E(X|C)P(C) = \sum_{\substack{A \in \varphi \\ P(A) \neq 0 \\ A \subseteq C}} E(X|A)P(A).$$

Nótese que C es la union de una subcolección de la particion φ si y sólo si está en la σ -álgebra generada por φ .

definición 3.24. Dos variables aleatorias X e Y se dicen independientes, si para todo par de intervalos $I = [a, b[$ y $J = [c, d[$ los eventos

$$A = \{\omega \in \Omega | X(\omega) \in I\}, \quad B = \{\omega \in \Omega | Y(\omega) \in J\},$$

son independientes, es decir $P(A \cap B) = P(A)P(B)$.

Lema 3.25. Si X e Y son independientes, también lo son X_n e Y_n .

Demostración Basta observar que

$$\left\{ \omega \in \Omega \mid X_n(\omega) = \frac{k}{n} \right\} = \left\{ \omega \in \Omega \mid \frac{k}{n} \leq X(\omega) < \frac{k+1}{n} \right\},$$

y del mismo modo

$$\left\{ \omega \in \Omega \mid Y_n(\omega) = \frac{k}{n} \right\} = \left\{ \omega \in \Omega \mid \frac{k}{n} \leq Y(\omega) < \frac{k+1}{n} \right\},$$

por lo que estos conjuntos son independientes. \square

Proposición 3.26. *Si X e Y son variables aleatorias integrables e independientes, entonces XY es integrable, y se tiene $E(XY) = E(X)E(Y)$.*

Demostración Para la integrabilidad, reemplazando X , Y , y XY por $|X|$, $|Y|$, y $|XY|$, de ser necesario, podemos suponer que ambas variables aleatorias son positivas y se tiene

$$(XY)_1 \leq XY \leq (X_1 + 1)(Y_1 + 1) = X_1Y_1 + X_1 + Y_1 + 1,$$

y cada una de las variables aleatorias del lado izquierdo es integrable. Como X_n e Y_n son independientes, se tiene $E(X_nY_n) = E(X_n)E(Y_n) \rightarrow E(X)E(Y)$, por otro lado

$$X_nY_n \leq XY \leq \left(X_n + \frac{1}{n} \right) \left(Y_n + \frac{1}{n} \right) = X_nY_n + \frac{1}{n}(X_n + Y_n) + \frac{1}{n^2},$$

por lo que al tomar esperanzas se tiene

$$E(X_nY_n) \leq E(XY) \leq E(X_nY_n) + \frac{1}{n}[E(X_n) + E(Y_n)] + \frac{1}{n^2},$$

y el resultado sigue. \square

Apéndice 1: Tiempos de espera

El teorema de convergencia acotada se extiende al caso general con la misma demostración utilizada en el capítulo precedente si demostramos que la función

$$N(\epsilon, \omega) = \min \left\{ m \mid n \geq m \Rightarrow |X(\omega) - X_n(\omega)| \leq \epsilon \right\}$$

es una variable aleatoria. Esta función es un ejemplo de lo que llamaremos un tiempo de espera.

Sea Y_1, Y_2, \dots una sucesión de variables aleatorias tomando valores 0 y 1. El tiempo de espera asociada a la sucesión es la función

$$N(\omega) = \min \left\{ m \mid n \geq m \Rightarrow Y(\omega) = 1 \right\}.$$

La función N se denomina en ocasiones el tiempo de espera del primer acierto de la sucesión.

Proposición 3.27. *Todo tiempo de espera es una variable aleatoria discreta.*

Demostración Sea X_1, X_2, \dots es una sucesión de variables aleatorias en un espacio medible (Ω, Σ) tomando valores 0 y 1. Basta observar que

$$N = k \iff X_k = 1 \text{ y } X_1 = \dots = X_{k-1} = 0,$$

de donde

$$\{\omega \mid N(\omega) = k\} = \{\omega \mid X_k(\omega) = 1\} \cap \bigcap_{i=1}^{k-1} \{\omega \mid X_i(\omega) = 0\}$$

y este último conjunto pertenece a la σ -álgebra Σ . □

Observese que si X_1, X_2, \dots, X_m es una colección finita de variables aleatorias discretas en un espacio medible, estas pueden considerarse variables aleatorias en un espacio numerable, por lo que cualquier función que dependa de ellas es una variable aleatoria discreta. En particular, Si X_1, X_2, \dots es una sucesión de variables aleatorias en un espacio medible (Ω, Σ) tomando valores 0 y 1, podemos definir una nueva sucesión de V.A.s Y_1, Y_2, \dots donde $Y_i = 1$ si y sólo si X_i es el n -ésimo término de la sucesión que toma el valor 1. El tiempo de espera de correspondiente recibe el nombre de tiempo de espera del n -ésimo acierto de la sucesión original.

Del mismo modo es posible definir el tiempo de espera del último acierto (recuérdese que como toda variable aleatoria, un tiempo de espera puede tomar el valor infinito). Dada una sucesión de variables aleatorias a valores 0 y 1 X_1, X_2, \dots , definimos una segunda sucesión de V.A.s Y_1, Y_2, \dots donde $Y_i = 1$ si $0 = X_{i+1} = X_{i+2} = \dots$. Cada Y_i es una variable aleatoria ya que

$$\{\omega \mid Y_i(\omega) = 1\} = \bigcap_{j=i+1}^{\infty} \{\omega \mid X_j(\omega) = 0\}.$$

Nótese que, así definido, el tiempo de espera del último acierto se anula si no hay aciertos.

Mas generalmente, Si X_1, X_2, \dots es una sucesión arbitraria de variables aleatorias en un espacio medible y si B es un boreliano (en las aplicaciones, B es generalmente un intervalo), podemos definir una nueva sucesión de V.A.s Z_1, Z_2, \dots donde $Z_i = 1$ si y sólo si $0 = X_i \in B$, y hablar así del tiempo de espera del primer ingreso (o del n -ésimo, o el último) a B . El hecho de que cada Z_i es una variable aleatoria sigue de la observación de que el conjunto $\{\omega | X_i(\omega) \in B\}$ está en la σ -álgebra.

Apendice 2: Variables aleatorias y funciones continuas

En esta sección demostraremos el resultado siguiente:

Proposición 3.28. *Si X es una variable aleatoria finita y $G : \mathbb{R} \rightarrow \mathbb{R}$ es una función continua, entonces $G(X)$ es una variable aleatoria.*

Para demostrar este resultado necesitaremos algunos lemas sobre la topología del conjunto de números reales. Recordemos que un subconjunto $U \subseteq \mathbb{R}$ se dice abierto si para todo elemento x de U existe un real $\epsilon > 0$ tal que el intervalo $]x - \epsilon, x + \epsilon[$ está contenido en U . En particular, los intervalos abiertos son conjuntos abiertos.

Lema 3.29. *Si $G : \mathbb{R} \rightarrow \mathbb{R}$ es una función continua y U es un conjunto abierto, entonces $G^{-1}(U)$ es un conjunto abierto.*

Demostración Sea $x \in G^{-1}(U)$. Por definición esto significa $G(x) \in U$. Como U es abierto, existe $\epsilon > 0$ tal que

$$]G(x) - \epsilon, G(x) + \epsilon[\subseteq U.$$

Como G es continua en x , existe $\delta > 0$ tal que $|x - y| < \delta$ implica $|G(x) - G(y)| < \epsilon$. En particular, si $y \in]x - \delta, x + \delta[$ se tiene $|x - y| < \delta$, por lo que $|G(x) - G(y)| < \epsilon$, es decir

$$G(y) \in]G(x) - \epsilon, G(x) + \epsilon[\subseteq U,$$

de donde, por definición, se tiene $y \in G^{-1}(U)$. Como y era un punto arbitrario de $]x - \delta, x + \delta[$, se tiene $]x - \delta, x + \delta[\subseteq G^{-1}(U)$. \square

Lema 3.30. *Todo subconjunto abierto de \mathbb{R} es unión numerable de intervalos.*

Demostración Basta probar que un conjunto abierto U es la unión de todos los intervalos de extremos racionales $]r, q[$ tales que $]r, q[\subseteq U$, ya que existe sólo una cantidad numerable de tales intervalos. Sin embargo, para cada elemento $x \in U$ existe un intervalo $]x - \epsilon, x + \epsilon[$ contenido en U , por lo que basta encontrar un racional r entre $x - \epsilon$ y x , así como un racional q entre x y $x + \epsilon$. \square

Corolario 3.30.1. *Todo subconjunto abierto de \mathbb{R} es un boreliano.* \square

Demostración de la proposición. Sea $G : \mathbb{R} \rightarrow \mathbb{R}$ una función continua. Entonces para todo conjunto de la forma $I =] - \infty, b[$ el conjunto $G^{-1}(I)$ es abierto y por lo tanto boreliano. Se sigue que

$$\{\omega \in \Omega | G(X)(\omega) \in I\} = \{\omega \in \Omega | X(\omega) \in G^{-1}(I)\} \in \Sigma.$$

\square

Apendice 3: El Teorema de Convergencia Dominada.

En este apéndice generalizaremos los teoremas de convergencia dominada y acotada para variables aleatorias no discretas.

Proposición 3.31. *Supongamos que la sucesión $\{X_n\}_{n \in \mathbb{N}}$ de variables aleatorias acotadas por una misma constante K converge a una variable aleatoria X . Entonces $E(X_n)$ converge a $E(X)$.*

Demostración Sea $\epsilon > 0$. Para cada $\omega \in \Omega$, sea $N = N(\epsilon, \omega)$ el menor entero tal que $|X_n(\omega) - X(\omega)| < \epsilon$ para todo $n > N$. La función $\omega \mapsto N(\epsilon, \omega)$ es un tiempo de espera finito, luego en particular una variable aleatoria finita. El resultado sigue como en el capítulo anterior una vez que probemos el siguiente lema:

Lema 3.32. *Si X es una variable aleatoria finita en un espacio de probabilidad arbitrario, entonces $P(X > t) \xrightarrow{n \rightarrow \infty} 0$.*

Demostración Observemos que $P(Y > t) \xrightarrow{n \rightarrow \infty} 0$ para toda V.A. discreta Y por los resultados del capítulo precedente. Por otro lado, para toda V.A. X se tiene

$$0 \leq P(X > t) \leq P\left(\frac{1}{n}[nX] > t - \frac{1}{n}\right),$$

de donde se sigue lo pedido. \square

Para deducir el teorema de convergencia dominada del de convergencia acotada, se define la medida P_Y , para toda V.A. positiva integrable Y mediante

$$P_Y(A) = E_Y(Y\chi_A)/E(Y).$$

Comprobaremos que P_Y define una medida. Es inmediato que $P_Y(\Omega) = 1$ y que $P_Y(A \cup B) = P_Y(A) + P_Y(B)$ si A y B son disjuntos. Consideremos ahora una unión numerable $A = \bigcup_{i=1}^{\infty} A_i$ de conjuntos A_i tales que $A_i \cap A_j$ para $i \neq j$. En este caso, $\{A^c, A_1, A_2, \dots\}$ es una partición numerable de Ω , por lo que se tiene

$$\begin{aligned} P_Y(A)E(Y) &= E(Y\chi_A) = E(Y\chi_{A^c})P(A^c) + \sum_{i=1}^{\infty} E(Y\chi_{A_i})P(A_i) \\ &= \sum_{i=1}^{\infty} E(Y|A_i)P(A_i) = \sum_{i=1}^{\infty} E(Y\chi_{A_i}) = \sum_{i=1}^{\infty} P_Y(A_i)E(Y). \end{aligned}$$

Si X es cualquier V.A. discreta finita, se tiene

$$E_Y(X)E(Y) = \sum_{r \in \mathbb{R}} r P_Y(X = r)E(Y) = \sum_{r \in \mathbb{R}} r E(Y\chi_{\{X=r\}}).$$

Y por otro lado,

$$E(XY) = \sum_{r \in \mathbb{R}} E(XY|X = r)P(X = r) = \sum_{r \in \mathbb{R}} r E(Y|X = r)P(X = r),$$

de donde sigue que $E_Y(X)E(Y) = E(XY)$. El caso general sigue ahora si recordamos que

$$\frac{1}{n}[nX] \leq X \leq \frac{1}{n}[nX] + \frac{1}{n}.$$

Dejamos los detalles al lector. El siguiente resultado sigue ahora como en el capítulo anterior.

Proposición 3.33. *Supongamos que la sucesión $\{X_n\}_{n \in \mathbb{N}}$ de variables aleatorias acotadas por una misma V.A. positiva e integrable Y converge a una variable aleatoria X . Entonces $E(X_n)$ converge a $E(X)$. \square*