PROBABILIDAD Y MEDIDA

ALGEBRAS Y σ-ALGEBRAS

Definición.- Sea Ω un conjunto y sea $\mathcal{A}$ una subfamilia de $\mathcal{P}(\Omega)$, decimos que $\mathcal{A}$ tiene estructura de álgebra si se verifica:

1) $\Omega \in \mathcal{A}$
2) Si $A \in \mathcal{A}$ entonces $A^c \in \mathcal{A}$
3) Si A y $B \in \mathcal{A}$ entonces $A \cup B \in \mathcal{A}$.

De esta definición se derivan de inmediato las siguientes propiedades

Propiedades

1) $\phi \in \mathcal{A}$
2) Si $\{A_i\}_{i=1}^{n} \subset \mathcal{A}$, entonces $\bigcup_{i=1}^{n} A_i \in \mathcal{A}$ y $\bigcap_{i=1}^{n} A_i \in \mathcal{A}$

Ejemplos de algebras

1) $\mathcal{A} = \{\phi, \Omega\}$ es un algebra (el algebra trivial)
2) $\mathcal{A} = \{$ subconjuntos de $\Omega\} = \mathcal{P}(\Omega)$ es también un algebra
3) Para $A \subset \Omega$, $\mathcal{A} = \{\phi, A, A^c, \Omega\}$ es un algebra.
4) Sea Ω un conjunto infinito, entonces $\mathcal{A} = \{A, A \subset \Omega \,/\, A$ o A^c es finito$\}$ es un algebra.

Las algebras aseguran la estabilidad para las uniones e intersecciones de sus elementos en número finito. Nada puede asegurarse acerca de estas operaciones cuando se llevan a cabo en número infinito numerable. La solución a este problema llega por vía de la generalización mediante la siguiente estructura.

Definición.- Sea Ω un conjunto y sea $\mathcal{B}$ una subfamilia de $\mathcal{P}(\Omega)$, decimos que $\mathcal{B}$ es una σ-algebra si verifica:

1) $\mathcal{B}$ es un algebra
2) Si $\{A_n\}_{n \geq 1} \subset \mathcal{B}$, entonces $\bigcup_{n \geq 1} A_n \in \mathcal{B}$.

Al igual que para las algebras también ahora $\bigcap_{n\geq 1} A_n \in \mathcal{B}$, propiedad que se deriva directamente de la definición. Como ejemplos de σ-algebras podemos citar:

Ejemplos de σ-algebras

1) $\mathcal{B} = \{\phi, \Omega\}$ es la σ-algebra trivial
2) $\mathcal{B} = \mathcal{P}(\Omega)$ es una σ-algebra.

Una estructura interesante y útil es la engendrada a partir de una familia de subconjuntos del espacio Ω inicial tomado.

Observación: Las σ-algebras se denominan también, en ocasiones, tribus.

Estructuras engendradas

Sea Ω un conjunto y sea $\mathcal{C}$ un subconjunto de $\mathcal{P}(\Omega)$ que posee o no alguna estructura concreta. Sean $\{\mathcal{A}_i\}_{i\in I}$, $\{\mathcal{B}_j\}_{j\in J}$ las familias de algebras y σ-algebras que contienen a $\mathcal{C}$. Hay que decir que dichas familias son no vacías por cuanto ambas contienen a $\mathcal{P}(\Omega)$. A partir de ellas obtenemos

$$\mathcal{A}(\mathcal{C}) = \bigcap_{i \in I} \mathcal{A}_i \quad , \quad \mathcal{B}(\mathcal{C}) = \bigcap_{j\in J} \mathcal{B}_j \, .$$

Es fácil comprobar que estas familias de conjuntos así definidas verifican:

1.a) $\mathcal{A}(\mathcal{C})$ es un algebra — 1.b) $\mathcal{B}(\mathcal{C})$ es una σ-algebra
2.a) $\mathcal{C} \subset \mathcal{A}(\mathcal{C})$ — 2.b) $\mathcal{C} \subset \mathcal{B}(\mathcal{C})$
3.a) $\mathcal{A}(\mathcal{C})$ es la mínima algebra que contiene a $\mathcal{C}$. — 3.b) $\mathcal{B}(\mathcal{C})$ es la mínima σ-algebra que contiene a $\mathcal{C}$.

$\mathcal{A}(\mathcal{C})$ y $\mathcal{B}(\mathcal{C})$ se llaman, respectivamente, algebra y σ-algebra engendrada por la familia $\mathcal{C}$.

La utilidad de esta construcción estriba en el hecho de permitir definir estructuras a partir de familias de conjuntos específicas. Sin ninguna duda el caso más interesante es cuando Ω no es un conjunto cualquiera sino que se trata de un espacio topológico. Aparece entonces la tribu de Borel, de la que nos vamos a ocupar a continuación.

La tribu de Borel, β.- Si Ω es un espacio topológico posee un sistema de abiertos y cerrados que le caracterizan. Sea Θ el sistema de abiertos y $\mathcal{F}$ el de cerrados. Definimos, mediante el proceso antes admitido, la σ-algebra engendrada por el sistema de abiertos, $\mathcal{B}(\Theta)$, dicha σ-algebra es conocida como la σ-algebra de Borel del espacio topológico Ω y se la designa mediante β. Esta σ-algebra tiene entre otras las siguientes propiedades dignas de mención:

a) $\mathcal{B}(\mathcal{F}) \equiv \beta$

b) Si S es una familia de abiertos que, siendo numerable, engendra el sistema de abiertos, es decir, si se trata de una base numerable del espacio topológico Ω, entonces $\mathcal{B}(S) = \beta$.

Un caso de tribu de Borel especialmente interesante para nosotros es aquel en que $\Omega = R$. Haciendo uso de la propiedad b) podemos afirmar que, como R posee una base numerable, la tribu de Borel de la recta real viene engendrada por una familia de intervalos cualquiera. Así por ejemplo, si tomamos $S = \{\,]x, y],\ x, y \in R,\ x<y\}$, entonces $\mathcal{B}(S) = \beta$. Si hubiéramos tomado los extremos racionales para definir los intervalos el resultado hubiera sido el mismo. Esta versatilidad en la construcción de β será útil posteriormente.

La tribu de Borel en R^2, o en general en R^k, β^2 o β^k respectivamente, se define análogamente a partir de familias de rectángulos (productos de intervalos) cuyas proyecciones son intervalos con extremos cualesquiera. Por ejemplo, para R^2, β^2 puede venir engendrada por la familia

$$S = \{]x_1, y_1] \times]x_2, y_2] \text{ , con } x_1, y_1, x_2, y_2 \in R \text{ ; } x_1 < y_1 \text{ , } x_2 < y_2 \}.$$

<u>Definición</u>.- Al par formado por un conjunto Ω y una σ-algebra de sus partes se le denomina <u>espacio medible</u> y se designa mediante $(\Omega, \mathcal{A})$.

MEDIDA

Sea $(\Omega, \mathcal{A})$ un espacio medible, definimos sobre $\mathcal{A}$ una función de conjunto, μ, que verifica

1) $\mu(\phi) = 0$

2) $\mu(A) \geq 0$, $\forall A \in \mathcal{A}$

3) si $\{A_n\}_{n \geq 1} \subset \mathcal{A}$ y $A_n \cap A_m = \phi$, $n \neq m$, entonces

$$\mu\left(\bigcup_{n \geq 1} A_n\right) = \sum_{n \geq 1} \mu(A_n).$$

Decimos entonces que μ es una medida sobre $\mathcal{A}$. A la tripleta $(\Omega, \mathcal{A}, \mu)$ la denominamos <u>espacio de medida</u>. Una consecuencia inmediata de esta definición es la <u>monotonía</u> de la medida, en efecto

si $A \subset B$, entonces $B = A + (B - A)$

aplicando 3) tenemos

$$\mu(B) = \mu(A) + \mu(B - A)$$

y como por 2) $\mu(B - A) \geq 0$, de aquí $\mu(B) \geq \mu(A)$. Es decir, la medida es <u>monótona creciente</u>.

Son muchas las propiedades de interés que posee una medida, no nos ocuparemos de ellas directamente ahora sino que las estudiaremos a medida que sean necesarias en el transcurso del desarrollo del tema. Sí tiene interés revisar ~~ahora~~ algunos de los distintos tipos de medidas con las que podemos encontrarnos.

<u>Medida σ-finita</u>.- Sea $(\Omega, \mathcal{A}, \mu)$ un espacio de medida, decimos que μ es <u>σ-finita</u> si $\forall A \in \mathcal{A}$,

$$\exists \{A_n\}_{n \geq 1} \subset \mathcal{A} \ / \ A \subset \bigcup_{n \geq 1} A_n \text{ y } \mu(A_n) < +\infty, \ \forall n.$$

<u>Medida finita</u>.- Sea $(\Omega, \mathcal{A}, \mu)$ un espacio de medida, decimos que μ es <u>finita</u> si $\mu(\Omega) < +\infty$.

Ello implica, dada la monotonía de μ, que $\mu(A) < +\infty$, $\forall A \in \mathcal{A}$.

<u>Ejemplo</u>.- Una medida de uso muy corriente en Análisis Matemático es la que se conoce con el nombre de <u>Medida de Lebesgue</u>, porque fue Lebesgue quien la introdujo a principios de siglo para desarrollar su famosa teoría de la Integración, teoría que por otra parte va directamente ligada a la teoría de la medida. En su versión <u>univariante</u>, la medida de Lebesgue se define sobre el espacio medible (R, β) como aquella que a cada intervalo le hace corresponder su longitud, cuando los extremos son finitos, o vale infinito sobre los intervalos con algún extremo infinito. No es lugar aquí para entrar en detalles

acerca de la definición y propiedades de esta medida.

En ~~su~~ versión <u>k-dimensional</u>, la medida se define sobre el espacio medible (R^k, β^k), $k \geq 2$ haciendo que sobre los hiperrectángulos de k-dimensiones, $\prod_{i=1}^{k}]x_i, y_i]$ valga el hipervolumen correspondiente (area para $k=2$, volumen para $k=3$, etc.) si los extremos de los intervalos factores son finitos, o bien valga infinito en cualquier otro caso.

Dadas las características topológicas de R y R^k, $k \geq 2$ podemos afirmar que las distintas medidas de Lebesgue son σ-finitas.

<u>Ejemplo</u>.- Otro ejemplo de medida definida sobre un espacio medible $(\Omega, \mathcal{A})$, es aquella definida como sigue:

$$\forall A \in \mathcal{A}, \quad \mu(A) = \{\text{número de elementos de } A\} = \text{card}(A).$$

Se comprueba fácilmente que es una medida a la que se conoce como <u>medida de cardinalidad</u> o <u>medida de conteo</u> (counting measure). Si Ω es finito, obviamente μ será finita. Al contrario de lo que ocurre con la medida de Lebesgue, esta medida no es σ-finita, si Ω es infinito no numerable, porque si fuera infinito numerable,

$$\Omega = \{w_1, \ldots, w_n, \ldots\} \text{ , entonces } \Omega = \bigcup_{n \geq 1} \{w_n\} \text{ y como } \mu(\{w_n\}) = 1$$

entonces μ es σ-finita pues $\forall A \in \mathcal{A}$, $A \subset \Omega$.

Vamos ahora ya a ocuparnos del objeto de nuestro estudio, a saber, las medidas de probabilidad.

MEDIDA DE PROBABILIDAD

Una <u>medida de probabilidad</u> no es más que una medida para la que $\mu(\Omega) = 1$. Es costumbre en este caso designar la medida mediante la letra P y hablar de $(\Omega, \mathcal{A}, P)$ como de un <u>espacio de probabilidad</u> y denominar a Ω <u>espacio muestral</u> y a $\mathcal{A}$ <u>σ-algebra de sucesos</u>.

<u>Ejemplo 1</u>.- Sea $\Omega = \{w_1, \ldots, w_n\}$ y sea $\mathcal{A} = \mathcal{P}(\Omega)$, definimos P de la siguiente manera

$$P(A) = \sum_{j=1}^{p} P(\{w_{i_j}\}) \text{ , con } A = \{w_{i_1}, \ldots, w_{i_p}\}$$

además $P(\{w_1\}) = \ldots = P(\{w_n\}) = 1/n$.

La función así definida es una medida de probabilidad, más concretamente se la conoce como <u>probabilidad uniforme</u> y coincide con la definición clásica debida a Laplace. Queda puesto así de manifiesto que esta nueva definición de la probabilidad es más general y contiene aquella como caso particular.

<u>Ejemplo 2</u>.- Sea $\Omega = [0,1]$ y $\mathcal{A} = \beta_{[0,1]}$ tribu de Borel sobre el espacio $[0,1]$. Definimos P mediante

$$P(A) = \text{longitud}(A) \text{ , donde } A \text{ es un subintervalo de } \Omega.$$

P define una probabilidad que se conoce como la <u>probabilidad uniforme sobre el intervalo [0,1]</u>.

Ejemplo 3.- Sea $\Omega=\{w_1,w_2\}$ y sea $\mathcal{A}=\mathcal{P}(\Omega)$, definimos P mediante

$$P(\{w_1\})=p \quad y \quad P(\{w_2\})=1-p \quad , (0\leq p\leq 1)$$

$(\Omega,\mathcal{A},P)$ constituye un espacio de probabilidad.

Propiedades de la medida de probabilidad

1) P es finitamente aditiva.

En efecto, sabemos que si $\{A_n\}_{n\geq 1}\subset\mathcal{A}$ con $A_n\cap A_m=\phi$, $n\neq m$ entonces

$$P\left(\bigcup_{n\geq1}A_n\right)=\sum_{n\geq1}P(A_n) \quad (\sigma\text{-aditividad de la medida de probabilidad}).$$

Si el número de sucesos es finito, a saber, $A_1,\ldots,A_n$, entonces podemos suponer $A_k=\phi$, $k\geq n+1$ y de aquí, aplicar la anterior condición y obtener

$$P\left(\bigcup_{i=1}^{n}A_i\right)=P\left(\bigcup_{n\geq1}A_n\right)=\sum_{n\geq1}P(A_n)=\sum_{i=1}^{n}P(A_i), \text{ por ser } P(\phi)=0.$$

2) $\forall A,B\in\mathcal{A}$, $P(A\cup B)=P(A)+P(B)-P(A\cap B)$.

$A\cup B=A\cup(B-A)$ y por la aditividad finita $P(A\cup B)=P(A\cup(B-A))=P(A)+P(B-A)$

pero $B=(B-A)\cup(A\cap B)$ y $P(B)=P(B-A)+P(A\cap B)$, es decir $P(B-A)=P(B)-P(A\cap B)$

sustituyendo

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

3) P es subaditiva, es decir, para $\{A_n\}_{n\geq1}\subset\mathcal{A}$, cualquiera $P\left(\bigcup_{n\geq1}A_n\right)\leq\sum_{n\geq1}P(A_n)$.

A partir de la sucesión $\{A_n\}_{n\geq1}$ podemos definir una nueva sucesión como sigue

$$B_n=A_n-\bigcup_{i=1}^{n-1}A_i$$

estos B_n verifican, $A_n\supset B_n$ y además $B_n\cap B_m=\phi$, $n\neq m$ con $\bigcup_{n\geq1}B_n=\bigcup_{n\geq1}A_n$, entonces

$$P\left(\bigcup_{n\geq1}A_n\right)=P\left(\bigcup_{n\geq1}B_n\right)=\sum_{n\geq1}P(B_n)\leq[\text{por monotonía}]\leq\sum_{n\geq1}P(A_n).$$

La desigualdad es también válida para un número finito de sucesos.

TEOREMA.- Sea $\{A_i\}_{i=1}^{n}\subset\mathcal{A}$, entonces

$$P\left(\bigcup_{i=1}^{n}A_i\right)=\sum_{i=1}^{n}P(A_i)-\sum_{1\leq i_1<i_2\leq n}P(A_{i_1}\cap A_{i_2})+\sum_{1\leq i_1<i_2<i_3\leq n}P(A_{i_1}\cap A_{i_2}\cap A_{i_3})-\cdots$$
$$\cdots+(-1)^{n+1}P(A_1\cap A_2\cap\ldots\cap A_n).$$

Demostración.- Se comprueba fácilmente por inducción sobre n.

Sucesiones monótonas de conjuntos (sucesos)

Sea $\{A_n\}_{n\geq1}$ una sucesión de conjuntos, se definen los límites superior e inferior de la siguiente forma

$$\overline{\lim_n}\,A_n=\bigcap_{n\geq1}\bigcup_{i\geq n}A_i \quad , \text{ límite superior}$$

$$\underline{\lim_n}\,A_n=\bigcup_{n\geq1}\bigcap_{i\geq n}A_i \quad , \text{ límite inferior}$$

cuando ambos límites coinciden, decimos que la sucesión tiene límite.

Hay una situación en la que está siempre garantizada la existencia del límite. Este el caso de las llamadas sucesiones monótonas, que pueden ser de dos tipos:

a) Una sucesión de conjuntos se dice que es monótona decreciente o contractiva si $A_n\supset A_{n+1}$, $\forall n$.

b) Una sucesión de conjuntos se dice que es monótona creciente o expansiva si $A_n\subset A_{n+1}$, $\forall n$.

Es fácil comprobar para este tipo de sucesiones que,

Si $\{A_n\}_{n\geq1}$ es contractiva $\overline{\lim_n}\,A_n=\underline{\lim_n}\,A_n=\bigcap_{n\geq1}A_n=\lim_n A_n$

Si $\{A_n\}_{n\geq1}$ es expansiva $\overline{\lim_n}\,A_n=\underline{\lim_n}\,A_n=\bigcup_{n\geq1}A_n=\lim_n A_n$

Cuando se trata de sucesiones de sucesos existe una importante propiedad de las probabilidades basadas sobre sucesiones monótonas. Es la siguiente

TEOREMA (Continuidad respecto del paso al límite).- Sea $\{A_n\}_{n\geq1}$ una sucesión monótona de sucesos, entonces

$$P\left(\lim_n A_n\right)=\lim_n P(A_n).$$

Demostración.- a) Supongamos que $\{A_n\}_{n\geq1}$ es expansiva, entonces $\lim_n A_n=\bigcup_{n\geq1}A_n$. Definimos una nueva sucesión

$$A'_n=A_n-A_{n-1} \quad , n\geq1, \quad A_0=\phi,$$

Sus elementos verifican $A'_n\cap A'_m=\phi$ y $\bigcup_{n\geq1}A'_n=\bigcup_{n\geq1}A_n$. A partir de aquí

$$P(\lim A_n)=P\left(\bigcup_{n\geq1}A_n\right)=P\left(\bigcup_{n\geq1}A'_n\right)=\sum_{n\geq1}P(A'_n)=\lim_{n\to\infty}\sum_{i=1}^{n}P(A'_i)=$$
$$=\lim_{n\to\infty}P\left(\bigcup_{i=1}^{n}A'_i\right)=\lim_{n\to\infty}P(A_n).$$

b) Si $\{A_n\}_{n\geq1}$ es contractiva la sucesión $\{A_n^c\}_{n\geq1}$ será expansiva y $\lim_n A_n^c=\bigcup_{n\geq1}A_n^c$.

Es fácil comprobar por otra parte que para un suceso A y su complementario A^c, se verifica

$$P(A)+P(A^c)=1.$$

Por el apartado anterior sabemos que

$$P(\lim_n A_n^c)=\lim_n P(A_n^c)=P\Big(\bigcup_{n\geq 1}A_n^c\Big)$$

utilizando la propiedad descrita

$$\lim_n P(A_n^c)=\lim_n\,[1-P(A_n)]=P\Big(\Big[\bigcap_{n\geq 1}A_n\Big]^c\Big)=1-P\Big(\bigcap_{n\geq 1}A_n\Big)$$

y de aquí

$$\lim_n P(A_n)=P\Big(\bigcap_{n\geq 1}A_n\Big)=P(\lim_n A_n).$$

Se deriva de aquí la llamada <u>continuidad en el ϕ</u> de la probabilidad y que damos en forma de

<u>COROLARIO</u>.- Si $\{A_n\}_{n\geq 1}$ es una sucesión monótona decreciente, tal que $\lim_n A_n=\phi$, entonces $\lim_n P(A_n)=0$.

La importancia de este corolario estriba en el hecho de que este resultado es equivalente a la aditividad completa o σ-aditividad postulada para la medida (en este caso para la probabilidad). La implicación en el sentido σ-aditividad → continuidad en ϕ está ya comprobada en el teorema anterior por cuanto en él hemos hecho uso de la citada σ-aditividad. Comprobémosla en sentido inverso.

<u>TEOREMA</u>.- Sustituyendo en la definición de medida de probabilidad la σ-aditividad por la continuidad en ϕ, tenemos que,

Si $\{A_n\}_{n\geq 1}\subset\mathcal{B}$ / $A_n\cap A_m=\phi$, entonces $P\Big(\bigcup_{n\geq 1}A_n\Big)=\sum_{n\geq 1}P(A_n)$.

Demostración.- Definamos, $A=\bigcup_{n\geq 1}A_n$, $B_n=\bigcup_{i=1}^{n}A_i$

$B'_n=A-B_n$.

La sucesión $\{B'_n\}_{n\geq 1}$ es contractiva y $\lim B'_n=\phi$, aplicando la continuidad en ϕ de la probabilidad, tendremos

$$P(\lim_n B'_n)=\lim_n P(B'_n)=0$$

pero $\lim_n P(B'_n)=\lim_n P(A-B_n)$.

Se comprueba fácilmente, que para dos sucesos cualesquiera E y F tales que $F\subset E$, se verifica que $P(E-F)=P(E)-P(F)$, aplicando esto

$$\lim_n P(B'_n)=\lim_n\,[P(A)-P(B_n)]=P(A)-\lim_n\sum_{i=1}^{n}P(A_i)=P(A)-\sum_{n\geq 1}P(A_n)=0$$

la serie es convergente

de aquí $P(A)=\sum_{n\geq 1}P(A_n)$.

NOTA.- Se supone la aditividad finita.

<u>Observaciones</u>.- Debemos señalar que en ocasiones nos encontraremos con sucesos tales que su probabilidad será 0, ello no implica en modo alguno que se trate del ϕ, de la misma manera que hay sucesos tales que su probabilidad es 1 y no por ello coinciden con el espacio muestral.

Veamos algunos ejemplos de medidas de probabilidad.

<u>Ejemplos</u>

① Sea Ω el conjunto de los naturales con el cero. Sea $\mathcal{B}$ la σ-algebra de las partes de Ω. Sobre el espacio medible $(\Omega,\mathcal{B})$ definimos una función de conjunto P, mediante

$$\forall A\in\mathcal{B},\quad P(A)=\sum_{x\in A}\frac{e^{-\lambda}\cdot\lambda^x}{x!},\quad \lambda>0.$$

La función P verifica

a) $P(\phi)=0$

b) $P(A)\geq 0$, $\forall A\in\mathcal{B}$

c) Para $\{A_n\}_{n\geq 1}\subset\mathcal{B}$, tal que $A_n\cap A_m=\phi$, $n\neq m$, tenemos

$$A=\bigcup_{n\geq 1}A_n \quad \text{y} \quad P(A)=\sum_{x\in A}\frac{e^{-\lambda}\cdot\lambda^x}{x!}=\sum_n\sum_{x\in A_n}\frac{e^{-\lambda}\cdot\lambda^x}{x!}=\sum_n P(A_n).$$

se trata pues de una medida. Además

$$P(\Omega)=\sum_{x\in\Omega}\frac{e^{-\lambda}\cdot\lambda^x}{x!}=\sum_{x\geq 0}\frac{e^{-\lambda}\cdot\lambda^x}{x!}=e^{-\lambda}\cdot\sum_{x\geq 0}\frac{\lambda^x}{x!}=e^{-\lambda}\cdot e^{\lambda}=1.$$

y es por tanto una medida de probabilidad.

② Sobre el mismo espacio medible del ejemplo anterior, definimos P, mediante

$\forall A\in\mathcal{B}$, $P(A)=1$, si A tiene un número finito de puntos

$P(A)=0$, en caso contrario.

La P así definida no es una medida de probabilidad porque ni siquiera es una medida. En efecto, no verifica la σ-aditividad, pues tomando $A_n=\{n\}$, $n\geq 1$, tendremos

$\bigcup_{n\geq 1}A_n=N$, $P(N)=0$ por contener un nº infinito de puntos

mientras que $P(A_n)=1$, $\forall n$ y por tanto $\sum_{n\geq 1}P(A_n)=+\infty$.

③ Existe una medida de probabilidad especialmente importante que da lugar a lo que genéricamente se conoce como <u>Probabilidad geométrica</u>, por cuanto la obtención de la probabilidad de cualquier suceso está directamente relacionada con las características geométricas del mismo. Veamos cuál es.

Tomemos R^n como espacio muestral y sea β^n la tribu de Borel de R^n. Sobre el espacio medible (R^n,β^n) consideremos la medida de Lebesgue en R^n, μ {Recordemos que esta medida a cada elemento de β^n le hace corresponder su hipervolumen}.

Sea ahora $\Omega\in\beta^n$ tal que $\mu(\Omega)<+\infty$. Restringimos nuestra atención a este conjunto y lo dotamos de la σ-algebra de sucesos $\beta^n_\Omega=\{A \,/\, A\subset\Omega \text{ y } A\in\beta^n\}$, definimos P de la siguiente manera

$$P(A)=\frac{\mu(A)}{\mu(\Omega)},\quad \forall A\in\beta^n_\Omega.$$

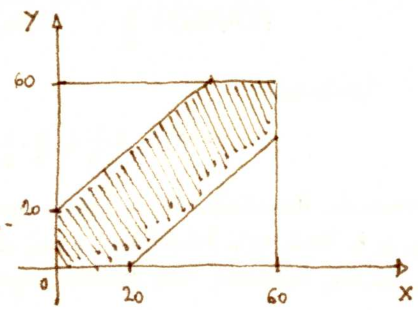
Es muy fácil comprobar que P es una medida, teniendo en cuenta que μ lo es. Es además una medida de probabilidad porque efectivamente está normalizada (vale 1 sobre Ω), ya que

$$P(\Omega) = \frac{\mu(\Omega)}{\mu(\Omega)} = 1.$$

Para $n=2$ entenderemos de inmediato el significado físico de este tipo de medida de probabilidad. En $\mathbb{R}^2$, la medida de Lebesgue es el área del conjunto de Borel considerado y la probabilidad definida supone que para un suceso A, la probabilidad es proporcional a la superficie que dicho suceso representa. Veamos un ejemplo que nos aclare estos conceptos.

Ejemplo de probabilidades geométricas.- Dos personas A y B acuerdan una cita en un lugar determinado entre las doce y la una del mediodía. El primero en llegar esperará 20 minutos al otro y después se irá. ¿Cuál es la probabilidad de que A y B se encuentren si la llegada de ambos es aleatoria dentro de la hora fijada?

En la definición de probabilidad geométrica que hemos dado hay implícita una condición de uniformidad en cuanto a la distribución de la probabilidad en todo el hipervolumen que Ω representa. Es decir que la elección de puntos en Ω es completamente aleatoria. El problema que nos ocupa explicita esta condición y podemos intentar contestarlo de acuerdo con el tipo de probabilidad admitida. Designemos por x e y los tiempos de llegada de A y B respectivamente. De acuerdo con las hipótesis que el enunciado plantea, A y B se encontrarán si



$$|x-y| \leq 20$$

(x e y se expresan en minutos)

Si representamos gráficamente la desigualdad observamos que el suceso E, se encuentran, viene representado geométricamente por la parte rayada de la figura. Mientras que todos los posibles tiempos de llegada serían la superficie del cuadrado de 60 unidades de lado y que representa el suceso Ω de la definición de probabilidad geométrica. La probabilidad buscada viene pues dada por

$$P(E) = \frac{\text{Superficie}(E)}{\text{Superficie}(\Omega)} = \frac{60^2 - 2\times 40^2/2}{60^2} = \frac{3600-1600}{3600} = \frac{2000}{3600} = \frac{5}{9}.$$

El concepto de probabilidad geométrica fue introducido por BUFFON, en el siglo 18, mediante un problema ya famoso en la historia de la Teoría de la Probabilidad, el problema de la aguja de BUFFON. Consiste en determinar la probabilidad de que al lanzar aleatoriamente una aguja de longitud $2l$, sobre un plano en el que se han pintado paralelas que distan entre sí $2a$ $(l<a)$, la aguja intersecte alguna de estas líneas. Nos ocuparemos de la solución de este problema en su momento.

Como en el siglo 18 los conceptos de medida y medida de Lebesgue no eran conocidos la solución se planteaba directamente en términos de superficies y volúmenes.

Señalar finalmente dos aspectos interesantes:

a) El concepto de probabilidad geométrica supone una generalización de la definición clásica de probabilidad. Allí en realidad utilizábamos para medir los conjuntos una medida adecuada para aquellas circunstancias, a saber, el cardinal de los conjuntos. Aquí, puesto que se trata de elementos de otra naturaleza recurrimos a la medida de Lebesgue como procedente para hacerlo.

b) Cuando introduzcamos el concepto de variable aleatoria y de independencia entre variables, veremos como este tipo de problemas puede resolverse dentro de un contexto más general, determinar la probabilidad de que las variables tomen valores en determinados conjuntos conociendo la ley que rige su probabilidad.

PROBABILIDAD CONDICIONAL

El concepto de probabilidad manejado hasta ahora es conocido también, en ocasiones, como probabilidad incondicional o absoluta por contraposición al que vamos a introducir a continuación. Antes, no obstante, de definirlo formalmente vamos a intentar dar alguna motivación intuitiva de su razón de ser.

Supongamos que tenemos un dado correcto en el que hemos pintado de rojo las caras que tienen 1, 4 y 6 puntos y de azul las restantes. Si lanzamos el dado una vez, sabemos que la probabilidad de que aparezca un 6 es $1/6$ (definición clásica), ahora bien, si después del lanzamiento nos es permitido mirar el dado, de tal forma que podamos distinguir el color de la cara, pero no su número y si observamos que dicha cara es roja, entonces podemos asegurar que la probabilidad de obtener 6 es ahora $1/3$. Esta última probabilidad viene condicionada por el hecho de conocer de antemano el color que ha aparecido en el lanzamiento y además se observa que su valor viene dado por

$$P(\text{salga 6 condicionado a que haya salido rojo}) = \frac{P(\text{salga 6 y rojo})}{P(\text{rojo})} = \frac{1/6}{1/2} = \frac{1}{3},$$

lo que utilizando una notación más simplificada y si $A=\{\text{salir } 6\}$, $B=\{\text{salir rojo}\}$, queda

$$P(A/B) = \frac{P(A\cap B)}{P(B)}$$

donde $P(A/B)$ designa la probabilidad del suceso A condicionado a que ha ocurrido B.

Daremos ahora una definición formal del concepto.

Definición.- Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad, sea A un suceso tal que $P(A)>0$. Definimos una función de conjunto sobre los elementos de $\mathcal{A}$, mediante

$$P(B/A) = \frac{P(A\cap B)}{P(A)}, \quad \forall B \in \mathcal{A}.$$

a la que denominamos probabilidad del suceso B condicionada a la ocurrencia del suceso A, o simplemente probabilidad condicional de B, dado A.

Observación.- No es una casualidad designar a la función de conjunto así definida con el nombre de probabilidad condicional. Este nombre se justifica plenamente cuando comprobamos (cosa que se hace muy fácilmente) que la función así definida es una medida de probabilidad y posee por tanto todas las propiedades inherentes al concepto.

TEOREMA (Regla de multiplicación).- Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad y sean $\{A_i\}_{i=1}^{n} \subset \mathcal{A}$, tales que $P\left(\bigcap_{i=1}^{n-1} A_i\right) > 0$. Entonces

$$P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1).P(A_2/A_1).P(A_3/A_1\cap A_2)\ldots\ldots P\left(A_n/\bigcap_{i=1}^{n-1} A_i\right).$$

La demostración es sencilla y se deja al cuidado del interesado (se supone).

El interés de la fórmula que nos proporciona reside en el hecho de que en ocasiones es mucho más sencillo obtener las probabilidades del 2º miembro de la igualdad que la probabilidad del primer miembro. Veámoslo con un ejemplo.

Ejemplo.- Una urna contiene 10 bolas idénticas, de las cuales 5 son negras, 3 rojas y 2 blancas. Se extraen cuatro bolas sin reemplazamiento. Probabilidad de que la primera sea negra, roja la segunda, blanca la tercera y la cuarta negra.

Sean $N_1 \to$ la 1ª bola es negra $\quad B_3 \to$ la 3ª bola es blanca

$R_2 \to$ la 2ª bola es roja $\quad N_4 \to$ la 4ª bola es negra.

Nos piden obtener

$$P(N_1\cap R_2\cap B_3\cap N_4)$$

haciendo uso del teorema tenemos

$$P(N_1\cap R_2\cap B_3\cap N_4) = P(N_1).P(R_2/N_1).P(B_3/R_2\cap N_1).P(N_4/B_3\cap R_2\cap N_1)$$

y

$$P(N_1) = \frac{1}{2}\ ,\ P(R_2/N_1) = \frac{1}{3}\ ,\ P(B_3/R_2\cap N_1) = \frac{1}{4}\ ,\ P(N_4/B_3\cap R_2\cap N_1) = \frac{4}{7}$$

de aquí

$$P(N_1\cap R_2\cap B_3\cap N_4) = \frac{1}{2}\cdot\frac{1}{3}\cdot\frac{1}{4}\cdot\frac{4}{7} = \frac{1}{42}$$

EL TEOREMA DE LA PROBABILIDAD TOTAL Y LA FORMULA DE BAYES

Supongamos que en el espacio de probabilidad $(\Omega, \mathcal{A}, P)$, la familia $\{A_i\}_{i\geq 1}$ constituye una partición del espacio muestral. Es decir

$$\Omega = \bigcup_{i\geq 1} A_i \quad y \quad A_i\cap A_j = \phi\ ,\ i\neq j$$

La familia además es finita o infinita numerable. Supongamos además que los sucesos que componen la partición son tales que $P(A_i) > 0$, $\forall i$. Sea ahora B un suceso cualquiera, sabemos que

$$B = B\cap\Omega = B\cap\left(\bigcup_{i\geq 1} A_i\right) = \biguplus_{i\geq 1}(B\cap A_i)$$

tomando probabilidades y por la σ-aditividad

$$P(B) = P\left(\bigcup_{i\geq 1}(B\cap A_i)\right) = \sum_{i\geq 1} P(B\cap A_i)$$

pero recordemos que $P(B/A_i) = \frac{P(A_i\cap B)}{P(A_i)} \to P(A_i\cap B) = P(B/A_i).P(A_i)$, sustituyendo

$$\boxed{P(B) = \sum_{i\geq 1} P(B/A_i).P(A_i)}$$

Fórmula que es conocida con el nombre de Fórmula de la probabilidad total. Comprobemos su utilidad.

Ejemplo.- La urna 1 contiene 2 bolas negras, la urna 2 contiene una bola negra y 2 blancas, y la urna 3 contiene tres bolas negras y tres bolas blancas. Se lanza un dado, si sale 1, 2 o 3 seleccionamos la urna 1, si sale un 4 seleccionamos la urna 2 y con 5 o 6 seleccionamos la 3. Extraemos una bola de la urna seleccionada. Probabilidad de que sea blanca.

Sean $A \to$ la bola extraída es blanca

$U_i \to$ la urna elegida es la i, $i = 1,2,3$.

$$P(A) = \sum_{i=1}^{3} P(A\cap U_i) = \sum_{i=1}^{3} P(A/U_i).P(U_i)$$

y

$$P(A/U_1) = 0\ ,\quad P(U_1) = \frac{1}{2}$$

$$P(A/U_2) = \frac{2}{3}\ ,\quad P(U_2) = \frac{1}{6}$$

$$P(A/U_3) = \frac{1}{2}\ ,\quad P(U_3) = \frac{1}{3}$$

Finalmente

$$P(A) = 0\cdot\frac{1}{2} + \frac{2}{3}\cdot\frac{1}{6} + \frac{1}{2}\cdot\frac{1}{3} = \frac{1}{9} + \frac{1}{6} = \frac{5}{18}.$$

Una consecuencia inmediata de la fórmula que acabamos de obtener es otra conocida fórmula de gran trascendencia en el desarrollo de la teoría de la Probabilidad o, más exactamente, en una nueva concepción de la misma.

Bajo las mismas condiciones antes descritas, supongamos además que B es tal que $P(B) > 0$. Entonces

$$P(A_i/B) = \frac{P(A_i\cap B)}{P(B)} = \frac{P(B/A_i).P(A_i)}{\sum_{i\geq 1} P(B/A_i).P(A_i)}\ ,\ \forall i$$

Esta fórmula es conocida como la FORMULA DE BAYES. Su importancia estriba en que aprovecha la información conocida para modificar las probabilidades de los sucesos. En efecto, si observamos la fórmula en ella suponemos conocidas las probabilidades $P(A_i)$ y $P(B/A_i)$, $\forall i$. Podemos suponer que B es un suceso que tiene lugar bajo diversas condiciones, respecto a cuya naturaleza podemos establecer una serie de hipótesis, A_i. Las probabilidades $P(A_i)$ se denominan probabilidades a priori. Cuando llevamos a cabo el experimento ocurre B y el conocimiento de este hecho nos permite obtener $P(A_i/B)$, conocidas como probabilidades a posteriori, que corrigen las $P(A_i)$ a través de la información previa adquirida (conocimiento del resultado del experimento). Veamos un ejemplo

Ejemplo.- Volvamos nuevamente a las urnas del ejemplo anterior. Calculemos $P(U_i/A)$, $i = 1,2,3$

$$P(U_1/A) = \frac{P(A/U_1).P(U_1)}{\sum_{i=1,3} P(A/U_i).P(U_i)} = 0$$

$$P(U_2/A) = \frac{\frac{2}{3}\cdot\frac{1}{6}}{\frac{5}{18}} = \frac{2}{5} \quad , \quad P(U_3/A) = \frac{\frac{1}{2}\cdot\frac{1}{3}}{\frac{5}{18}} = \frac{3}{5}$$

Obsérvese que al saber que la bola extraída es blanca las probabilidades de que la urna elegida haya sido la 2 o la 3 se han modificado y en concreto han mejorado por cuanto

Elegida urna	A priori	A posteriori
2	1/6	2/5
3	1/3	3/5

La probabilidad de elegir la urna 1 se altera y como dicha urna no tiene bolas blancas dicha probabilidad, lógicamente, vale 0 (se convierte en algo imposible de ocurrir).

La importancia de la fórmula de Bayes, insistimos, estriba en este hecho: Siempre que poseamos información a priori del experimento nos permite aprovecharla para modificar adecuadamente las probabilidades que manejamos.

INDEPENDENCIA DE SUCESOS

Supongamos una urna con 10 bolas, 6 de ellas rojas y 4 blancas. Llevamos a cabo dos extracciones con reemplazamiento y sean A_1 y A_2 los sucesos, "obtener bola roja en la primera y la segunda extracción" respectivamente. Sabemos que $P(A_1) = 6/10$, $P(A_2) = 6/10$ y además, $P(A_2/A_1) = 6/10 = P(A_2)$. Es decir, el conocimiento de lo que ocurrió en la primera extracción no es de ninguna utilidad a la hora de saber qué pasará con la segunda. Decimos, ante una situación como la presente, que ambos sucesos son independientes.

Este ejemplo nos ayudará mejor a comprender el por qué de la siguiente definición

Definición 1.- Sean A y B sucesos del espacio de probabilidad $(\Omega, \mathcal{A}, P)$, tales que $P(A)>0$. Decimos que A y B son independientes si $P(B/A) = P(B)$.

Obsérvese que si además $P(B)>0$, entonces

$$P(A/B) = \frac{P(A\cap B)}{P(B)} = \frac{P(B/A).P(A)}{P(B)} = \frac{P(B).P(A)}{P(B)} = P(A)$$

La independencia es una relación mutua entre los sucesos. Señalemos que el transcurso de la comprobación anterior, y como resultado intermedio, hemos obtenido la siguiente propiedad:

Propiedad 1.- Si A y B son independientes, entonces $P(A\cap B) = P(A).P(B)$.

Como en ocasiones muchos autores intercambian los términos de este proceso y dan como definición la propiedad, obteniendo como propiedad lo que hemos presentado como definición.

Definición 2.- Dos sucesos A y B del espacio de probabilidad $(\Omega, \mathcal{A}, P)$ decimos que son independientes si

$$P(A\cap B) = P(A).P(B)$$

De aquí y si $P(A)>0$, $\quad P(B/A) = \dfrac{P(B\cap A)}{P(A)} = \dfrac{P(B).P(A)}{P(A)} = P(B)$. Propiedad 2

Esta segunda definición tiene la ventaja de no exigir condición alguna a los sucesos A y B. Así, mientras sucesos A y B con $P(A)$ o $P(B)$ nulas no permiten una definición correcta de la probabilidad condicional, ello no impide la definición de independencia que acabamos de dar.

TEOREMA.- Si A y B son independientes también lo son A y B^c, A^c y B o A^c y B^c.

Demostración.- Inmediata.

Conviene señalar que los conceptos de independencia e incompatibilidad no deben ser confundidos. De hecho, cuando dos sucesos incompatibles A y B son tales que $P(A)>0$ y $P(B)>0$, ambos son siempre dependientes por cuanto $P(A\cap B) = P(\emptyset) = 0$, mientras que $P(A).P(B) \neq 0$.

Ejemplo.- Elegimos al azar una carta de una baraja de 52. Sea A, la carta extraída es un as y sea B, la carta extraída es una copa. Tenemos

$$P(A) = \frac{4}{52} = \frac{1}{13} \quad , \quad P(B) = \frac{13}{52} = \frac{1}{4}$$

$$P(A\cap B) = P(\text{extraer el as de copas}) = \frac{1}{52}$$

pero $P(A).P(B) = \frac{1}{13}\cdot\frac{1}{4} = \frac{1}{52} = P(A\cap B)$, luego ambos sucesos son independientes.

La definición puede generalizarse para un número finito de sucesos, pero hay entonces que introducir un matiz en la misma.

Definición.- Sean $\{A_1, \ldots, A_n\}$ una familia finita de sucesos del espacio de probabilidad $(\Omega, \mathcal{A}, P)$

a) Decimos que son mutuamente o completamente independientes si se verifican

$$P(A_{i_1}\cap A_{i_2}\cap \ldots \cap A_{i_k}) = P(A_{i_1}).\ldots.P(A_{i_k})$$

$$k = 2, \ldots, n \qquad \{i_1, \ldots, i_k\} \subset \{1, 2, \ldots, n\}.$$

b) Decimos que son independientes dos a dos si

$$P(A_i\cap A_j) = P(A_i).P(A_j) \qquad i,j = 1, \ldots, n \ , \ i\neq j.$$

Observemos que la independencia mutua exige que se cumplan

$$\binom{n}{2} + \binom{n}{3} + \ldots + \binom{n}{n} = 2^n - \binom{n}{1} - \binom{n}{0} = 2^n - n - 1$$

relaciones entre los n sucesos, $A_1, \ldots, A_n$. Además, de la definición se sigue de inmediato que esta independencia implica la independencia dos a dos. Por el contrario, el contraejemplo que damos a continuación demuestra que no es cierta la implicación contraria.

Ejemplo.- Supongamos un tetraedro cuyas caras están pintadas, una de rojo, otra de azul, otra de verde y la cuarta con los tres colores. Sean

$A \rightarrow$ la cara sobre la que se apoya el tetraedro presenta color rojo
$B \rightarrow$ id. id. azul
$C \rightarrow$ id. id. verde

Si el tetraedro está bien construido, tendremos

$$P(A) = P(B) = P(C) = \frac{1}{2}$$

$$\left.\begin{aligned} P(A\cap B) &= \tfrac{1}{4} = P(A).P(B) \\ P(A\cap C) &= \tfrac{1}{4} = P(A).P(C) \\ P(B\cap C) &= \tfrac{1}{4} = P(B).P(C) \end{aligned}\right\} \text{ hay independencia dos a dos.}$$

pero $\quad P(A\cap B\cap C) = \frac{1}{4} \neq P(A).P(B).P(C)$ y no hay independencia mutua.

Independencia entre σ-algebras.- El concepto de independencia puede hacerse extensivo a las σ-algebras en lugar de ceñirnos tan solo a los sucesos. Así, para un espacio de probabilidad $(\Omega, \mathcal{A}, P)$ supongamos que existen $\mathcal{A}_1, \ldots, \mathcal{A}_n$, subfamilias de $\mathcal{A}$ con structura de σ-algebra. Decimos que las σ-algebras $\mathcal{A}_i$, $i=1,\ldots,n$ son independientes si

$$P(A_1\cap A_2\cap \ldots \cap A_n) = P(A_1).P(A_2)\ldots P(A_n) \quad , \forall A_i\in\mathcal{A}_i, \; i=1,\ldots,n.$$

El problema de la independencia tanto de sucesos como de σ-algebras se sitúa en el contexto de espacios medibles productos y la definición sobre ellos de medidas a partir de los espacios coordenados, y que son conocidas como medidas producto. En cualquier caso este problema sobrepasa, sobradamente, los límites de este curso y no nos ocuparemos de él.

APENDICE . EL TEOREMA DE EXTENSION

Supongamos el espacio medible $(\mathbb{R}, \beta)$, la recta real provista de la tribu de Borel. Para cada intervalo I, definimos

$$P(I) = \int_I \frac{1}{\pi}\cdot\frac{1}{1+x^2}\,dx .$$

La función de conjunto así definida verifica:

a) $P(I) \geq 0$, $\forall I \in \beta$

b) $P(R) = 1$

c) Dados $\{I_j\}_{j=1}^{n}$ con $I_i\cap I_j = \phi$, $i\neq j$, entonces
$$P\left(\bigcup_{j=1}^{n} I_j\right) = \sum_{j=1}^{n} P(I_j).$$

Podemos decir que P es una función de conjunto que se asemeja mucho a una medida de probabilidad, pero no lo es por cuanto:

a) No sabemos si es σ-aditiva

b) La hemos definido sobre los intervalos de R que son una subfamilia de β, pero no sobre β.

En realidad a) es soslayable pues las propiedades de la integral nos permiten afirmar que la σ-aditividad (para intervalos) también ha de verificarse. El verdadero problema lo plantea la posible extensión de P, desde la familia de intervalos a la tribu de Borel. Este es un problema común a la construcción de muchas medidas y tiene una adecuada respuesta en la teoría de la Medida. Es muy sencillo, en ocasiones, definir una función de conjunto sobre una familia de la σ-algebra con alguna característica especial. Por ejemplo, sobre la familia de intervalos de la forma $]a,b]$, ~~con a y b finitos o infinitos pero no simultáneamente~~, podemos definir una función tal que $\mu(]a,b]) = b-a$, ~~o bien $\mu(]a,b]) = +\infty$ si a=−∞~~, si consideramos además intervalos de la forma $]-\infty,b]$, $]a,+\infty[$ y añadimos R y ϕ y definimos $\mu(]-\infty,b]) = +\infty$, $\mu(]a,+\infty[) = +\infty$, $\mu(R) = +\infty$, $\mu(\phi) = 0$. Esta función es la que dará lugar a la medida de Lebesgue antes citada. Pero, ¿cómo?

La familia de intervalos que acabamos de considerar tiene dos características a resaltar

1) La σ-algebra que engendra es la tribu de Borel.

2) Puede engendrar una structura intermedia, llamada algebra de conjuntos y que difiere de la σ-algebra en que es estable tan solo para uniones finitas, de manera sencilla. En efecto, ~~la~~ el algebra $\mathcal{Q}(\mathcal{I})$, si por $\mathcal{I}$ designamos la familia de intervalos, se caracteriza por que

$$\text{si } A\in\mathcal{Q}(\mathcal{I}) \; , \quad A = \bigcup_{i=1}^{n} I_i \quad \text{con} \quad I_i\cap I_j = \phi \; , \; i\neq j .$$

3) Es fácil comprobar que la σ-algebra que engendraríamos con $\mathcal{Q}(\mathcal{I})$, con $\mathcal{I}$ es la misma, es decir, la tribu de Borel.

La característica 2) nos permite extender de forma inmediata y natural cualquier función de conjunto definida sobre $\mathcal{I}$ a $\mathcal{Q}(\mathcal{I})$, sin más que escribiendo $\mu(A) = \sum_{i=1}^{n}\mu(I_i)$. Obsérvese que, así definida, μ verifica sobre $\mathcal{Q}(\mathcal{I})$

a) $\mu(A)\geq 0$, $\forall A\in\mathcal{Q}(\mathcal{I})$

b) $\mu(A\cup B) = \mu(A)+\mu(B) \quad$ si $A\cap B = \phi$

Llegados a este punto y con las propiedades que μ verifica podemos aplicar un conocido teorema de Teoría de la Medida, el teorema de extensión de Hahn-Caratheodory, que afirma que si sobre un algebra $\mathcal{Q}$, tenemos definida una medida μ que es σ-finita, ésta puede extenderse a la σ-algebra que $\mathcal{Q}$ engendra, de manera única.

La demostración del teorema requiere profundos conocimientos de teoría de la Medida. Era conveniente, no obstante, hacer referencia al mismo porque el proceso de construcción de medidas de probabilidad sobre el espacio medible (R,β), que hemos presentado al comienzo del apendice, es muy frecuente en teoría de la Probabilidad y sólo ahora sabemos que es lícito afirmar que P es una medida de probabilidad, porque el teorema de extensión nos asegura su definición sobre β.