#10

UNIVERSIDAD DE SONORA

Escuela de Altos Estudios

UNA DEMOSTRACION PROBABILISTA DEL TEOREMA DE RADON-NIKODYM

TESIS

Que para obtener el Título de

LICENCIADO EN MATEMATICAS

Presenta

2

Fernando Luque Vásquez

Hermosillo, Sonora 1980

UNIVERSIDAD DE SONORA

Escuela de Altos Estudios

UNA DEMOSTRACION PROBABILISTA DEL TEOREMA DE RADON-NIKODYM

TESIS

Que para obtener el Título de

LICENCIADO EN MATEMATICAS

Presenta

Fernando Luque Vásquez

Hermosillo, Sonora 1980

A la memoria de mi Madre,

A mi Padre,

A todos mis Hermanos,

A Eva...

I N D I C E

I N T R O D U C C I O N

El teorema de Radon-Nikodym juega un papel fundamental en la teoría moderna de probabilidad; un papel que le asignó Kolmogorov en su histórico libro "Foundations of the Theory of Probability" (7)*. El teorma tal y como lo usó Kolmogorov permite definir la esperanza condicional en su forma más general; sin embargo, la definición es puramente descriptiva y no es claro el enlace que existe entre la definición de esperanza condicional en el caso elemental, donde se tiene una definición constructiva, y el caso general, donde la esperanza condicional es una derivada de Radon-Nikodym. Este "punto obscuro" se debe principalmente al hecho de utilizar el teorema como un resultado ajeno a la teoría de probabilidad.

Con el fin de esclarecer el enlace entre la definición constructiva y la definición descriptiva, daremos una versión probabilista del teorema de Radon-Nikodym, obteniendo así una definición constructiva de la esperanza condicional en su forma más general.

En el Capítulo I damos una demostración del teorema de Radon-Nikodym (su versión en teoría de la medida) y desarrollamos el concepto de "condicionalidad" hasta llegar a la definición de esperanza condicional en su forma más general.

En el Capítulo II, con el fin de mostrar que el teorema de Radon-Nikodym no es necesario para definir esperanza condicional, obtenemos la definición siguiendo un camino distinto, utilizando el teorema de representación de Riesz y demostrando la equivalencia entre las definiciones obtenidas.

En el Capítulo III damos la versión probabilista del teorema de Radon-Nikodym y algunos ejemplos ilustrativos.

(*) Los números (1), (2),... se refieren a las referencias bibliográficas.

Debo agradecer a los Profesores Enrique Valle Flores, Fernando Avila Murillo y Marco Antonio Valencia, por sus consejos y sugerencias en el desarrollo de este trabajo.

CAPITULO I

En el desarrollo de este trabajo, consideremos fijo un espacio medible $(\Omega,\mathcal{A})$*, donde* Ω *es un conjunto abstracto y* $\mathcal{A}$ *es un* σ*-campo de conjuntos de* Ω*. En* $\mathcal{A}$ *definiremos una medida; es decir, una función* $\mu:\mathcal{A}\to\overline{\mathbb{R}}$ *tal que:*

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

ii) $\mu\left(\sum_{i=1}^{\infty} A_i\right)=\sum_{i=1}^{\infty}\mu A_i \quad , \quad A_i\in\mathcal{A}$

Nota1: $\sum_{i=1}^{\infty} A_i$ *significa* $\bigcup_{i=1}^{\infty} A_i$ *cuando* $A_i\cap A_j=\phi$ *si* $i\neq j$*. Se llama --- unión ajena.*

Si $\mu A<\infty \ \forall\, A\in\mathcal{A}$*, se dice que* μ *es finita. Si todo conjunto de* $\mathcal{A}$ *es unión numerable de conjuntos en* $\mathcal{A}$ *para los cuales* μ *- es finita, se dice que* μ *es* σ*-finita.*

1-1 <u>*Definición*</u>*.- Sea X una función medible cuya integral existe (posiblemente infinita). La integral indefinida* φ *sobre* $\mathcal{A}$ *está dada por:*

$$\varphi(A)=\int_A X\,d\mu=\int X I_A\,d\mu \quad .$$

<u>*Nota 2*</u>*: El indicador (o función indicatriz) de A, es la función característica de A definida por:*

$$I_A(\omega)=\begin{cases}1 & \text{si } \omega\in A\\ 0 & \text{si } \omega\notin A\end{cases} \quad .$$

1-2 <u>*Definición*</u>*.- Una función conjuntista* φ *definida en* $\mathcal{A}$ *se dice* μ*-continua si* $\mu A=0$ *implica* $\varphi(A)=0$*.*

1-3 <u>*Proposición*</u>*.- La integral indefinida de una función medible es una función* μ*-continua.*

<u>*Demostración*</u>*.- Sea* $X=I_A$ *y sea* $B\in\mathcal{A}$ *tal que* $\mu B=0$

Se tiene la afirmación para indicadores de conjuntos medibles. En el caso en que X es una función simple*, es decir, $X=\sum_{j=1}^{n} x_j A_j$, se tiene:

$$\varphi(B)=\int_B\Big(\sum_{j=1}^{n} x_j I_{A_j}\Big)d\mu=\sum_{j=1}^{n}\Big(x_j\int I_{A_jB}\,d\mu\Big)=0$$

Si X es una función medible positiva, existe una sucesión $\{X_n\}$ de funciones siples tales que $X_n \uparrow X$. Aplicando el teorema de la convergencia monótona obtenemos:

$$\varphi(B)=\int_B X\,d\mu=\int X I_B\,d\mu=\lim\int X_n I_B\,d\mu=0 \quad .$$

En el caso general, descomponemos X en su parte positiva y en su parte - negativa: $X = X^+ - X^-$ y obtenemos:

$$\varphi(B)=\int_B X\,d\mu=\int_B X^+ d\mu-\int_B X^- d\mu=0 \quad .$$

1-4 <u>Proposición</u>.- La integral indefinida es σ-aditiva; es decir,

$$\varphi\Big(\sum_{j=1}^{\infty} A_j\Big)=\sum_{j=1}^{\infty}\varphi(A_j)$$

<u>Demostración</u>:

$$\varphi\Big(\sum_{j=1}^{\infty} A_j\Big)=\int_{\sum_{j=1}^{\infty} A_j} X\,d\mu=\int\Big(X I_{\sum_{j=1}^{\infty} A_j}\Big)d\mu=\int\Big(X\sum_{j=1}^{\infty} I_{A_j}\Big)d\mu=$$

$$\int\sum_{j=1}^{\infty}(X I_{A_j})\,d\mu=\sum_{j=1}^{\infty}\int(X I_{A_j})\,d\mu=\sum_{j=1}^{\infty}\int_{A_j} X\,d\mu=\sum_{j=1}^{\infty}\varphi(A_j) \quad .$$

1-5 <u>Proposición</u>.- Si X es una función medible integrable (su integral es finita) estonces es finita c.d. y la integral indefinida es finita. Si X no es integrable, pero no obstante es finita c.d. y μ es σ-finita, entonces la integral indefinida es σ-finita.

Demostraremos la primera parte para el caso en que X es una función medible simple. La extensión se sigue de la misma manera como se hizo - en (1-3).

(*) $X:\Omega\to\mathbb{R}$ simple en este trabajo significa simple y medible. Es decir, $X(\Omega)$ es finito y $X^{-1}(B)\in\mathcal{Q}$ si $B\subset\mathbb{R}$.

Sea

$$X=\sum_{j=1}^{n} x_j I_{A_j} \quad ; \quad \int X d\mu=\sum_{j=1}^{n} x_j \mu A_j < \infty \Rightarrow x_j < \infty \text{ c.d.}$$

Para la segunda parte, descomponemos Ω en conjuntos A_n de medida finita y obtenemos que:

$$\int X d\mu = \sum_{m=-\infty}^{\infty} \sum_{n=1}^{\infty} \int_{A_n[m\leq X\leq m+1]} X d\mu$$

y cada término de la doble suma es finito.

1-6 Definición.- Una función conjuntista φ_s definida sobre $\mathcal{Q}$, se dice que es μ-singular, si se anula fuera de un conjunto μ-nulo, es decir, si existe un conjunto μ-nulo N tal que

$$\varphi_s(AN^c)=0$$

1-7 Definición.- Una medida con signo es una función conjuntista $\varphi:\mathcal{Q}\to\bar{\mathbb{R}}$ tal que

1) φ toma a lo más uno de los valores:

2) $\varphi(\phi)=0$

3) $\varphi(\sum_{n=1}^{\infty} A_n)=\sum_{n=1}^{\infty}\varphi(A_n) \quad , \quad A_n \in \mathcal{Q}$.

1-8 Proposición.- Una función conjuntista es una medida con signo si y sólo si es la diferencia de dos medidas, de las cuales al menos una es finita.

Demostración: Sea φ una medida con signo y supongamos que toma a lo más el valor $+\infty$ (queda excluído $-\infty$). Sean μ_1 y μ_2 tales que:

$$\mu_1(A)=\begin{cases}\varphi(A) & \text{si} \quad \varphi(A)\geq 0\\ 0 & \text{si} \quad \varphi(A)<0\end{cases}$$

$$\mu_2(A)=\begin{cases}-\varphi(A) & \text{si} \quad \varphi(A)\leq 0\\ 0 & \text{si} \quad \varphi(A)>0\end{cases}$$

Es claro que μ_1 y μ_2 son medidas, μ_2 es finita y

$$\varphi = \mu_1 - \mu_2 \quad .$$

Recíprocamente, sean μ_1 y μ_2 dos medidas en $\mathcal{Q}$ con $\mu_2 < \infty$. Entonces, la función conjuntista $\varphi = \mu_1 - \mu_2$ es tal que:

i) φ toma a lo más el valor $+\infty$

ii) $\varphi(\phi) = 0$ (por ser μ_1 y μ_2 medidas).

iii) $\varphi(\sum A_n) = \mu_1(\sum A_n) - \mu_2(\sum A_n) = \sum \mu_1 A_n - \sum \mu_2 A_n =$

$$\sum (\mu_1 A_n - \mu_2 A_n) = \sum \varphi(A_n) \quad .$$

1-9 <u>Teorema</u> (descomposición de Hahn).- Sea φ una medida con signo definida en $\mathcal{Q}$. Entonces existe un conjunto $D \in \mathcal{Q}$ tal que para todo $A \in \mathcal{Q}$

$$\varphi(AD) \geq 0 \quad \text{y} \quad \varphi(AD^c) \leq 0 \quad .$$

La demostración puede verse en (1) o (3).

1-10 <u>Teorema</u> (descomposición de Lebesque).- Si la medida μ y la función σ-aditiva φ definidas en $\mathcal{Q}$ son σ-finitas, entonces existe una y sólo una descomposición de φ en una función conjuntista μ-continua y σ-aditiva φ_c y una función conjuntista μ-singular y σ-aditiva φ_s

$$\varphi = \varphi_c + \varphi_s$$

Además φ_c es la integral indefinida de una función medible finita X determinada μ-c.d. A X se le llama derivada de Radon-Nikodym: $\frac{d\varphi}{d\mu}$.

<u>Demostración</u>: Puesto que Ω puede ser considerado como una unión numerable de conjuntos medibles ajenos para los cuales φ y μ son finitas, supondremos que φ y μ son medidas finitas.

Probaremos primero la unicidad: Supóngase que existen dos descomposiciones:

$$\varphi = \varphi_c + \varphi_s = \varphi'_c + \varphi'_s \quad ,$$

entonces $\varphi_c - \varphi'_c = \varphi'_s - \varphi_s = 0$ *ya que* $\varphi_c - \varphi'_c$ *es una función* μ*-conti-nua y* $\varphi'_s - \varphi_s$ *es una función* μ*-singular.*
Para ver que el integrando está definido μ*-c.d., supóngase que para cada* $A \in \mathcal{Q}$*,*

$$\varphi_c(A) = \int_A X\,d\mu = \int_A X'\,d\mu \quad .$$

Esto implica que $X = X'$ *c.d., ya que si* $\mu B = \mu[X - X' > \epsilon] > 0$ *entonces* $\int (X - X')\,d\mu > 0$ *. Así, las afirmaciones de unicidad se cumplen si probamos la existencia.*

Sea Φ *la clase de todas las funciones integrables no negativas* X *cuyas integrales indefinidas están acotadas por* φ *:*

$$\int_A X\,d\mu \leq \varphi(A) \quad , \; A \in \mathcal{Q} \ .$$

Φ *es no vacío puesto que la función medible* $X \equiv 0$ *pertenece a* Φ *. Existe una sucesión* $\{X_n\}$ *contenida en* Φ *tal que*

$$\int X_n\,d\mu \longrightarrow \sup_{X \in \Phi} \int X\,d\mu = \alpha \leq \varphi(\Omega) < \infty \ .$$

Sea $X'_n = \sup_{k \leq n} X_k$ *, de donde* $0 \leq X'_n \uparrow Y = \sup X_n$*.*
Sea

$A_k = [X_k = X'_n]$ *para n fijo y sea*

$$A'_1 = A_1 \ , \quad A'_k = A^c_1 \cdots A^c_{k-1} A_k$$

Así, $\sum_{k=1}^{n} A'_k = \bigcup_{k=1}^{n} A_k = \Omega$ *y para cada* $A \in \mathcal{Q}$ *,*

$$\int_A X'_n\,d\mu = \sum_{k=1}^{n} \int_{AA'_k} X'_n\,d\mu = \sum_{k=1}^{n} \int_{AA'_k} X_k\,d\mu \leq \sum_{k=1}^{n} \varphi(AA'_k) = \varphi(A)$$

Haciendo $n \to \infty$ *y aplicando el teorema de convergencia monótona, tenemos:*

$$\int_A Y\,d\mu \leq \varphi(A) \quad \text{y} \quad \int Y\,d\mu = \alpha$$

Por lo tanto, Y *es un elemnto maximal de* Φ *. Esta propiedad nos ayudará para demostrar que*

$$\varphi_s = \varphi - \varphi_c \geq 0 \ ,$$

donde φ_c *es la integral indefinida de* Y*, es* μ*-singular.*

Sea $D_n + D^c_n$ *una descomposición de Hahn para la función conjuntista finita y* σ*-aditiva*

$$\varphi_n = \varphi_s - \frac{1}{n}\mu \quad ;$$

entonces:

$$\varphi_n(AD_n) \leq 0 \quad y \quad \varphi_n(AD_n^c) \geq 0 \quad \forall A \in \mathcal{Q}$$

Sea $D = \bigcap_{n=1}^{\infty} D_n$ *y entonces* $D^c = \bigcup_{n=1}^{\infty} D_n^c$. *Así, para cada A y todo n,*

$$0 \leq \varphi_s(AD) \leq \frac{1}{n}\mu(AD)$$

Haciendo $n \to \infty$ *se sigue que* $\varphi_s(AD) = 0$ *y por lo tanto*

$$\varphi_s(A) = \varphi_s(AD^c).$$

Puesto que

$$\varphi_c(A) = \varphi(A) - \varphi_s(AD^c) \leq \varphi(A) - \varphi_s(AD_n^c) \ ,$$

se sigue que

$$\int_A \left(Y + \frac{1}{n} I_{D_n^c} \right) = \varphi_c(A) + \frac{1}{n}\mu(AD_n^c) \leq \varphi(A) - \varphi_n(AD_n^c) \leq \varphi(A).$$

Así que

$$\left(Y + \frac{1}{n} I_{D_n^c} \right) \in \Phi$$

Pero esto contradice el hecho demostrado que

$$\int Y d\mu = \alpha = \sup_{X \in \Phi} \int X d\mu$$

a menos que $\mu D_n^c = 0$. *Por lo tanto, todos los conjuntos* D_n^c *son* μ*-nu*<u>*los*</u> *y así lo es* D^c . *Puesto que* $\varphi_s(AD) = 0 \ \forall A$, *se concluye que* φ_s *es* μ*-singular. //*

En el caso particular de una función φ μ*-continua, el resultado anterior se reduce a:*

1-11 <u>*TEOREMA DE RADON-NIKODYM*</u>*.- Si, sobre* $\mathcal{Q}$*, la medida* μ *y la función conjuntista* σ*-aditiva* φ *son* σ*-finitas y* φ *es* μ*-continua, entonces* φ *es la integral indefinida de una función finita determinada* μ*-c.d.*

La integral indefinida de una función medible X no necesariamente finita es σ-aditiva y μ-continua, pero no necesariamente σ-finita. El siguiente teorema provee una extensión del teorema 1-11.

1-12 Teorema.- *El teorema de Radon-Nikodym resulta válido si la finitud de X y la σ-finitud de φ se suprimen sumultáneamente.*

Demostración: *Sea μ una medida finita, φ una medida μ-continua en $\mathcal{Q}$, $\mathcal{B}$ la clase de todos los conjuntos medibles tales que φ es σ-finita en $\mathcal{B}$ y sea*

$$S = \sup_{B \in \mathcal{B}} \mu B$$

Existe una suceción $\{B_n\}$, $B_n \in \mathcal{B}$ tal que

$$S = \lim \mu B_n$$

y por lo tanto,

$$B = \bigcup B_n \in \mathcal{B} \quad y \quad \mu B = S$$

Si existe $C \in \{B^c A, A \in \mathcal{Q}\}$ tal que $0 < \varphi(C) < \infty$, entonces:

$$B + C \in \mathcal{B}, \quad \mu C > 0$$

$$y$$

$$S \geq \mu(B + C) = \mu B + \mu C > S$$

Por lo tanto, mientras φ es σ-finita en $\{BA, A \in \mathcal{Q}\}$, φ puede tomar sólo los valores 0 o ∞ en $\{B^c A, A \in \mathcal{Q}\}$.
Además, es imposible tener

$$\mu C > 0 \quad y \quad \varphi(C) = 0$$

porque entonces $(B + C) \in \mathcal{B}$ y tendríamos, como antes, $S > S$. Puesto que φ es μ-continua, también es imposible tener

$$\mu C = 0 \quad y \quad \varphi(C) > 0.$$

Así, para cada $C \in \{B^c A, A \in \mathcal{Q}\}$ nos quedan dos posibilidades:

$$\mu C > 0 \quad y \quad \varphi(C) = \infty$$

$$o$$

$$\mu C = 0 \quad y \quad \varphi(C) = 0$$

En otras palabras, φ en $\{B^cA,\ A\in\mathcal{Q}\}$ es la integral indefinida de una función $X = \infty$ determinada μ-c.d. y por otro lado, por el teorema --- 1-11, φ en $\{BA,\ A\in\mathcal{Q}\}$ es la integral indefinida de una función X en - B determinada μ-c.d. Estos valores de X en B y B^c la determinan en Ω y para cada $A\in\mathcal{Q}$,

$$\int_A X\,d\mu = \int_{AB} X\,d\mu + \int_{AB^c} X\,d\mu = \varphi(AB) + \varphi(AB^c) = \varphi(A) .$$

CONDICIONALIDAD

Sea $(\Omega, \mathcal{Q}, P)$ un espacio de probabilidad, $\mathcal{B}\subset\mathcal{Q}$ un σ-campo.

Definiremos probabilidad y esperanza condicionales empezando por el caso más simple: Probabilidad del evento A dado el evento B, esperanza -- condicional de una variable aleatoria X dado el evento B; luego el caso "dado el σ-campo $\mathcal{B}$", cuando $\mathcal{B}$ es el σ-campo generado por una partición numerable de Ω y por último el caso general ($\mathcal{B}$ cualquier sub-σ-campo de $\mathcal{Q}$).

1-13 <u>Definición</u>.- La probabilidad condicional de un evento A dado un evento B no nulo, denotada por P_BA, se define por la relación

$$P_BA = \frac{P(AB)}{P(B)} \quad , \quad PB \neq 0$$

P_B es una medida de probabilidad en $\mathcal{Q}$ y podemos hablar del espacio de - probabilidad $(\Omega, \mathcal{Q}, P_B)$.

1-14 <u>Definición</u>.- La esperanza de una variable aleatoria X en el espacio de probabilidad $(\Omega, \mathcal{Q}, P_B)$, se denomina la esperanza condicional de X dado B y se denota por

$$E_B X = \int X\,dP_B$$

Pero como $P_B = 0$ en $\{AB^c, A \in \mathcal{Q}\}$, tenemos:

$$E_B X = \int_B X \, dP_B$$

y como $P_B = \frac{1}{PB} P$ en $\{AB, A \in \mathcal{Q}\}$, llegamos a

$$E_B X = \frac{1}{PB} \int_B X \, dP \, .$$

En particular,

$$PB \cdot E_B I_A = \int_B I_A \, dP = P(AB)$$

y

$$P_B A = E_B I_A$$

La esperanza condicional (y probabilidad condicional) adquiere su significado completo cuando se interpreta como valores de funciones de la siguiente manera: El número $E_B X$ se asigna ya no a B, sino a cada elemento de B y similarmente, $E_{B^c} X$ se asigna a cada elemento de B^c.

Así, considerando el σ-campo $\mathcal{B} = \{\emptyset, B, B^c, \Omega\}$, queda definida - la variable aleatoria:

$$E^{\mathcal{B}} X = (E_B X) I_B + (E_{B^c} X) I_{B^c}$$

Mas general, sea $\{B_j\}$ una partición numerable de Ω y sea $\mathcal{B}$ el σ-campo generado por esta partición. Consideremos la función elemental

$$E^{\mathcal{B}} X = \sum_{j \in J} (E_{B_j} X) I_{B_j} \quad , J \text{ numerable} , X \in \mathfrak{E}$$

donde $\mathfrak{E}$ es la familia de las variables aleatorias cuya integral existe.

Si algunos de los B_j son nulos, los correspondientes valores $E_{B_j} X$ - están indeterminados. Así, $E^{\mathcal{B}} X$ no está determinada en el evento nulo el cual es la suma de los B_j nulos.

Considerando esta posibilidad y la definición de $E_{B_j} X$, llegamos a:

1-15 <u>*Definición*</u> *(Constructiva).- La función elemental* $E^{\mathcal{B}}X$ *definida --* *P - c.d. por*

$$E^{\mathcal{B}}X = \sum_{j\in J} \left(\frac{1}{PB_j}\int_{B_j} X\,dP\right) I_{B_j}\ , \quad X \in \mathfrak{X}$$

es la esperanza condicional de X *dado* $\mathcal{B}$.

Particularizando a indicadores, definimos la función $\mathcal{B}$*-medible -* $P^{\mathcal{B}}A = E^{\mathcal{B}}I_A$, $A \in \mathcal{Q}$ *la cual será la probabilidad condicional de* A *dado* $\mathcal{B}$

En las definiciones anteriores decimos, "dado el σ *-campo* $\mathcal{B}$ *" y no "dada la partición* $\{B_j\}$ *", ya que* $E^{\mathcal{B}}X$ *determina la esperanza condi-- cional de* X *dado un evento no nulo arbitrario* $B \in \mathcal{B}$. *Obsérvese que si* B *es no nulo, existe* $K \subset J$ *tal que*

$$B = \sum_{k\in K} B_k$$
$$y$$
$$PB\,E_B X = \int_{\Sigma B_k} X\,dP = \Sigma \int_{B_k} X\,dP = \sum_{k\in K} PB_k\, E_{B_k} X = \int_B (E^{\mathcal{B}}X)\,dP\,.$$

Así, si $P_{\mathcal{B}}B = PB \ \ \forall\ B \in \mathcal{B}$ *y* $PB > 0$ *, la expresión de la derecha vendrá a ser* $\int_B (E^{\mathcal{B}}X)\,dP_{\mathcal{B}}$ *; mientras que la expresión de la iz -- quierda sabemos que es* $\int_B X\,dP$.

Hemos justificado la terminología, pero además hemos obtenido una propiedad de la esperanza condicional que permite caracterizarla, como veremos enseguida. Nótese que $E^{\mathcal{B}}X$ *es una variable aleatoria* $\mathcal{B}$ *-medible.*

1-16 <u>*Proposición*</u>*.- Si* Y *es una variable aleatoria* $\mathcal{B}$*-medible tal que*

$$\int_B Y\,dP = \int_B X\,dP \quad \forall\ B \in \mathcal{B}\ ,$$

entonces

$$P[Y = E^{\mathcal{B}}X] = 1$$

<u>*Demostración:*</u> *Si el evento* $B = [Y < E^{\mathcal{B}}X] \in \mathcal{B}$ *fuera de probabilidad positiva se tendría:*

Por lo tanto, $P[Y < E^{\mathcal{B}}X] = 0$. Similarmente, $P[Y > E^{\mathcal{B}}X] = 0$.//

Esto nos lleva a "redefinir" (ahora descriptivamente) la esperanza condicional respecto a un σ-campo numerable (generado por una partición de Ω a lo mas numerable).

1-17 <u>Definición</u> (Descriptiva).- La esperanza condicional $E^{\mathcal{B}}X$ de una variable aleatoria $X \in E$ es la única variable aleatoria $\mathcal{B}$-medible que satisface la relación

$$\int_B X\, dP = \int_B (E^{\mathcal{B}}X)\, dP_{\mathcal{B}} \quad \forall\ B \in \mathcal{B} ,$$

donde $\mathcal{B}$ es generado por una partición a lo más numerable de Ω . La unicidad es en el sentido de que si existe otra variable aleatoria $\mathcal{B}$-medible Y que satisfaga esta relación, entonces,

$$P[Y \neq E^{\mathcal{B}}X] = 0 .$$

El siguiente paso es extender la definición de esperanza condicional al caso en que el sub-σ-campo $\mathcal{B}$ no necesariamente es el σ-campo generado por una partición numerable de Ω . La definición que clásicamente se extiende es la definición descriptiva (ver (1)). Una manera de lograr tal extensión es utilizando el teorema de Radon-Nikodym. Otra posible forma de extender la definición es utilizando el teorema de representación de Riesz para funcionales lineales acotadas en $L^2(\Omega, \mathcal{Q}, P)$.

Para finalizar este capítulo extenderemos la definición de esperanza condicional, justificando tal extensión por medio del teorema de Radon-Nikodym.

1-18 <u>Definición</u>.- La espernaza condicional ($E^{\mathcal{B}}X$) de $X \in E$ dado $\mathcal{B}$ es una función $\mathcal{B}$-medible definida P -c.d. por la relación

$$\int_B (E^{\mathcal{B}} X)\, dP_{\mathcal{B}} = \int_B X\, dP \quad , \quad B \in \mathcal{B} .$$

Para justificar la definición, obsérvese que la integral indefinida φ *de* X, *es* σ*-aditiva y* P *- continua y por lo tanto, su restricción* $\varphi_{\mathcal{B}}$ *a* $\mathcal{B}$ *es* σ*-aditiva y* P *-continua. Aplicando el teorema de Radon-Nikodym extendido (1-12), la función* $\mathcal{B}$*- medible* $E^{\mathcal{B}}X$ *existe y está definida* P *-c.d.*

En el caso que $\varphi_{\mathcal{B}}$ *es* σ*- finita, aplicamos el teorema de Radon-Nikodym (1-11) y en este caso,* $E^{\mathcal{B}}X$ *es finita excepto en un evento nulo arbitrario perteneciente a* $\mathcal{B}$ *.*

La restricción de $E^{\mathcal{B}}$ *a la familia* $I_{\mathcal{Q}}$ *de indicadores de eventos es la probabilidad condicional dado* $\mathcal{B}$ *y se denota por* $P^{\mathcal{B}}$*. En otras palabras,* $P^{\mathcal{B}}$ *es una función en* $\mathcal{Q}$ *cuyos valores son funciones* $\mathcal{B}$*- medibles* $P^{\mathcal{B}}A$ *definidas* P *-c.d. por* $P^{\mathcal{B}}A = E^{\mathcal{B}}I_A$ *o directamente por*

$$\int_B (P^{\mathcal{B}}A)\, dP_{\mathcal{B}} = \int_B (E^{\mathcal{B}} I_A)\, dP_{\mathcal{B}} = \int_B I_A\, dP = P(AB) .$$

CAPITULO II

En este capítulo definiremos esperanza condicional a través del teorema de representación de Riesz para funcionales lineales acotadas en -- $L^2(\Omega,\mathcal{Q},P)$. Demostramos primeramente el teorema, definimos esperanza condicional de una variable aleatoria en $L^2(\Omega,\mathcal{Q},P)$ y extendemos la definición al caso general.

2-1 Definición.- Sea $(\Omega,\mathcal{Q},\mu)$ un espacio con medida. Definiremos $L^p(\Omega,\mathcal{Q},\mu)$ como el espacio de las funciones $\mathcal{Q}$-medibles X para las cuales.

$$\int |X|^p d\mu < \infty \quad , \ 1 \leq p < \infty$$

Si $(\Omega, \mathcal{Q}, P)$ está fijo, escribiremos simplemente L^p.

Considerando iguales a dos conjuntos que difieren en un conjunto - μ-nulo damos la siguiente definición.

2-2 Definición.- Definimos en $L^p(\Omega, \mathcal{Q}, \mu)$ una norma por medio de la - relación

$$\|X\|_p = \left[\int |X|^p d\mu\right]^{\frac{1}{p}}$$

2-3 Definiciones.- Supóngase $Y: \Omega \longrightarrow [0,\infty]$ $\mathcal{Q}$-medible. Sea S el conjunto de números reales α tales que

$$\mu[Y^{-1}(\alpha,\infty]] = 0$$

y sea $\beta = \inf S$. Puesto que

$$Y^{-1}(\beta,\infty] = \bigcup_{n=1}^{\infty} Y^{-1}(\beta + \tfrac{1}{n}, \infty]$$

y puesto que la unión numerable de conjuntos de medida cero, es de medida cero, $\beta \in S$. A β se le denomina el supremo esencial de Y y se denota $\beta = \sup \text{ess}\, Y$.

Si X es una función $\mathcal{Q}$-medible, se define $\| \ \|_\infty$ por larelación:

$$\|X\|_\infty = \sup \text{ess} |X|$$

$\mathcal{L}^{\infty}(\Omega, \mathcal{Q}, \mu)$ *denotará el conjunto de las funciones* X $\mathcal{Q}$*- medibles tales que*

$$\|X\|_{\infty} < \infty$$

A las funciones $X \in \mathcal{L}^{\infty}(\Omega, \mathcal{Q}, \mu)$ *se les denomina* *<u>esencialmente acotadas</u>.*

2-4 *<u>Proposición</u>.- Si* μ *es una medida finita, entonces:*

$$\mathcal{L}^{p} \supset \mathcal{L}^{q} \qquad 1 \leq p \leq q \leq \infty$$

<u>Demostración</u>: Si $1 \leq p \leq q < \infty$ *entonces:*

$$|X|^{p} \leq 1 + |X|^{q}$$

Integarando ambos lados de la desigualdad y por la monotonía de la integral se tiene:

$$\int |X|^{p} d\mu \leq \mu\Omega + \int |X|^{q} d\mu$$

de donde se deduce que si $X \in \mathcal{L}^{q}$ *entonces* $X \in \mathcal{L}^{p}$

Si $X \in \mathcal{L}^{\infty}$ *entonces,*

$$\infty > \mu\Omega \,(\sup \mathrm{ess}\, |X|)^{p} \geq \int |X|^{p} d\mu$$

2-5 *<u>Definición</u>.- Sea E un espacio lineal. Un producto interior o producto escalar denotado* (X, Y)*, es una función definida en el producto* $E \times E$ *y con valores en campo de escalares* $\mathbb{R}$ *tal que:*

i) $(\alpha X + \beta Y, Z) = \alpha(X, Z) + \beta(Y, Z) \quad \forall \alpha, \beta \in \mathbb{R} \quad y \quad X, Y \in E$

ii) $(X, Y) = (Y, X)$

iii) $X \neq 0 \Rightarrow (X, X) > 0$.

Nótese que si E es un espacio con producto interior, E se convierte en un espacio normado definiendo

$$\|X\| = (X, X)^{\frac{1}{2}} \qquad \forall X \in E$$

2-6 *<u>Definición</u>.- Un espacio de Hilbert es un espacio lineal con producto interior que con la norma inducida es de Banach.*

<u>Observación</u>.- La norma definida en L^{p} *, induce una métrica para el espa-*

cio L^p definida por

$$d(X,Y)=\|X-Y\|_p \quad \forall\, X,Y\in L^p \quad ; \ 1\leq p\leq\infty$$

2-7 <u>Definición</u>.- Decimos que $\{X_n\}$, $X_n \in L^p$ converge en L^p a X, si $\|X_n - X\|_p \longrightarrow 0$ cuando $n\to\infty$ y escribimos $X_n \xrightarrow{L^p} X$.

2-8 <u>Definición</u>.- Decimos que $\{X_n\}$, $X_n \in L^p$, converge en la p-ésima media a X $(X_n \xrightarrow{p} X)$ si

$$\int |X_n - X|^p \longrightarrow 0 .$$

En lo sucesivo sólo se considerarán espacios de probabilidad ---- $(\Omega, \mathcal{Q}, P)$ y escribiremos EX en lugar de $\int X dP$.

<u>Nota</u>: La convergaencia en L^p es equivalente a la convergenacia en la - p-ésima media.

2-9 <u>Teorema</u>.- El espacio $L^p(\Omega, \mathcal{Q}, P)$ es completo.

<u>Demostración</u>: Sea X_n una sucesión de Cauchy en $L^p(\Omega, \mathcal{Q}, P)$;

$$|X_m - X_n| \xrightarrow{p} 0$$

Por la desigualdad de Markov,

$$P[|X_m - X_n| \leq \epsilon] \leq \frac{1}{\epsilon^p} E|X_m - X_n|^p \to 0 \qquad m,n\to\infty ,$$

de donde $|X_m - X_n| \xrightarrow{P} 0$ y entonces

$$X_n \xrightarrow{P} X$$

para alguna $X \in L^p$.

Existe una subsucesión $X_{n'}$ tal que:

$$X_{n'} \xrightarrow{c.s.} X .$$

Puesto que

$$\int |X_m - X_{n'}|^p \to 0$$

cuando $m, n' \longrightarrow \infty$, se sigue por el lema de Fatou (Royden pág. 226).

$$\int |X_m - X|^p \leq \liminf_{n'} \int |X_m - X_{n'}|^p \longrightarrow 0 \quad \text{cuando} \quad m \to \infty$$

por lo tanto, $X_m \xrightarrow{p} X$.

2-10 <u>Teorema</u>.- El espacio $L^2(\Omega, \mathcal{Q}, P)$ es un espacio de Hilbert con el producto interior dado por

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

<u>Demostración</u>: Las propiedades del producto interior, se obtienen directamente de las propiedades de la integral. Por otro lado, la norma inducida por el producto interior es la norma definida para L^2 , así que por el teorema anterior, L^2 es de Banach.

2-11 <u>Definición</u>.- Una funcional lineal en L^p es una funcion $F: L^p \to \mathbb{R}$ tal que

$$F(\alpha X + \beta Y) = \alpha F(X) + \beta F(Y) \quad \forall\ X, Y \in L^p, \ \alpha, \beta \in \mathbb{R}$$

F es <u>continua</u> si $F(X_n) \longrightarrow F(X)$ cuando $X_n \xrightarrow{L_p} X$

F es normada o acotada si existe $M < \infty$, independiente de X, tal que

$$|F(X)| \leq M \|X\|_p .$$

La norma de F es el número dado por la relación

$$\|F\| = \sup_{X \neq 0} \frac{|F(X)|}{\|X\|} = \sup_{\|X\|_p = 1} |F(X)|$$

2-12 <u>Definición</u>.- En un espacio lineal H con producto interior, dos vectores X, Y son ortogonales si $(X, Y) = 0$.

Para todo subconjunto M de H, el conjunto $M^{\perp}$ de vectores Y que son ortogonales a todos los vectores $X \in M$ es el suplemento ortogonal de M.

2-13 <u>Teorema</u>.- Sea H un espacio de Hilbert, M un subespacio vectorial -- completo (de Hilbert). Para cada $X \in H$, existe un punto y sólo uno ---- $Y = P_M(X)$ tal que

$$\|X - Y\| = d(X, M).$$

El punto $Y = P_M(X)$ es el único punto $Z \in M$ tal que $X - Z$ es ortogonal a M. La aplicación $X \longrightarrow P_M(X)$ de H sobre M, es lineal, continua y de norma 1 si $M \neq \{0\}$; su núcleo $M' = P_M^{-1}(0)$ es el subespacio ortogonal a M.

Demostración: Sea $\alpha = d(X,M)$; por definición, existe una sucesión $\{Y_n\}$ de puntos de M tal que

$$\lim_{n\to\infty} \|X - Y_n\| = \alpha \ ;$$

demostraremos que $\{Y_n\}$ es una sucesión de Cauchy.

Aplicamos la propiedad del paralelogramo (válida para todo espacio con producto interior):

$$\|\mu + v\|^2 + \|\mu - v\|^2 = 2(\|\mu\|^2 + \|v\|^2)$$

para el caso $\mu = X - Y_n$, $v = X - Y_m$ y obtenemos:

$$\|Y_m - Y_n\|^2 = 2(\|X - Y_m\|^2 + \|X - Y_n\|^2) - 4\|X - \tfrac{1}{2}(Y_m + Y_n)\|^2$$

Pero $\frac{1}{2}(Y_m - Y_n) \in M$ de donde $\|X - \frac{1}{2}(Y_m - Y_n)\|^2 \geq \alpha^2$
Luego, si N es tal que para $n \geq N$, $\|X - Y_n\|^2 \leq \alpha^2 + \epsilon$
se tiene,

$$\|Y_m - Y_n\|^2 \leq 4\epsilon \quad \text{para } m \geq N ,\ n \geq N.$$

Por lo tanto, $\{Y_m\}$ es de Cauchy y como M es completo, $\{Y_m\}$ converge a $Y \in M$ para el cual

$$\|X - Y\| = d(X,M) .$$

Ahora supóngase que $Y' \in M$ es tal que $\|X - Y'\| = d(X,M)$
Aplicando la propiedad del paralelogramo al caso

$$\mu = X - Y' \quad , \quad v = X - Y$$

obtenemos:

$$\|Y - Y'\|^2 = 4\alpha^2 - 4\|X - \tfrac{1}{2}(Y + Y')\|^2$$

y puesto que

$$\frac{1}{2}(Y+Y') \in M, \quad \|Y-Y'\|^2 \leq 0,$$

es decir $Y = Y'$.

Sea Ahora, $Z \neq 0$ un punto cualquiera de M. Tenemos:

$$\|X-(Y+\lambda Z)\|^2 > \alpha^2 \quad \forall \lambda \neq 0 .$$

Aplicando las propiedades del producto interior obtenemos

$$-2\lambda(X-Y, Z) + \lambda^2 \|Z\|^2 > 0 \quad \forall \lambda \neq 0$$

Si $(X - Y, Z) \neq 0$, llegaríamos a una contradicción escogiendo convenientemente a λ . En consecuencia,

$$(X - Y, Z) \doteq 0$$

de donde $X - Y$ es ortogonal a M. Sea $Y' \in M$ tal que $X - Y'$ es ortogonal a M; entonces para $Z \neq 0$ en M se tiene por el teorema de Pitágoras:

$$\|X-(Y'+Z)\|^2 = \|X-Y'\|^2 + \|Z\|^2$$

y esto demuestra que $Y = Y'$ en virtud de la caracterización de Y.
Esta última caracterización de $Y = P_M(X)$ demuestra que P_M es lineal, pues si $X - Y$ y $X' - Y'$ son ortogonales a M, entonces $\lambda X - \lambda Y$ es ortogonal a M y también lo es

$$(X+X')-(Y+Y') = (X-Y) + (X'-Y') .$$

En virtud del teorema de Pitágoras, se tiene

$$\|X\|^2 = \|P_M(X)\|^2 + \|X - P_M(X)\|^2,$$

lo que prueba que $P_M(X) \leq \|X\|$ y por lo tanto P_M es continua y tiene norma ≤ 1 (ver Dieudonné pag), pero por ser $P_M(X) = X$ para $X \in M$,- se tiene $\|P_M\| = 1$ si M no se reduce a 0.

La definición de P_M implica que $M' = P_M^{-1}(0)$ está formado por los- vectores X ortogonales a M. Además, por ser

$$X = P_M(X) + (X - P_M(X)) \quad y \quad X - P_M(X) \in M'$$

se tiene : $H = M + M'$.

Finalmente, si $X \in H$ es ortogonal a M' se tiene en particular

$$(X, X - P_M(X)) = 0 \ ,$$

pero también se tiene

$$(P_M(X), X - P_M(X)) = 0$$

por lo tanto,

$$\|X - P_M(X)\|^2 = 0 \ , \text{ es decir, } X = P_M(X) \in M \quad .//$$

Nota.- A la aplicación lineal P_M se le denomina la proyección ortogonal de H sobre M.

2-14 Corolario.- Si $M \neq H$, existe $Y \in H$, $Y \neq 0$ tal que $Y \perp M$.

Demostración: Sea $X \in (H - M)$ y sea

$$Y = X - P_M(X) \neq 0 \ . \ //$$

2-15 Teorema.- Sea F una funcional lineal continua definida en un espacio de Hilbert H. Entonces existe $Y \in H$ única tal que

$$F(X) = (X, Y).$$

Demostración: Si $F(X) = 0 \ \forall X \in H$, tomamos $Y = 0$. De otra manera, definimos

$$M = \{X \mid F(X) = 0\}$$

La linealidad de F demuestra que M es un subespacio y la continuidad de F muestra que M es completo.

Puesto que $F(X) \neq 0$ para algún $X \in H$, el corolario 2-14 asegura que M - contiene puntos distintos de cero. Por lo tanto, existe $Z \in M^{\perp}$ con $\|Z\| = 1$

Sea $\mu = F(X)Z - F(Z)X$.

Puesto que

$$F(\mu) = F(X)F(Z) - F(Z)F(X) = 0$$

tenemos que $\mu \in M$ y en consecuencia

$$0 = (\mu, Z) = F(X)(Z,Z) - F(Z)(X,Z)$$

$$\Longrightarrow \quad F(X) = F(X)(Z,Z) = F(Z)(X,Z).$$

Hacemos $Y = \alpha Z$ con $\alpha = F(Z)$.

Para probar la unicidad: Si $(X,Y) = (X,Y')$ $\forall$ $X \in H$, sea

$$Z = Y - Y';$$

Entonces $(X,Z) = 0$ $\forall$ $X \in H$; en particular, $(Z,Z) = 0$ y por lo tanto - $Z = 0$.

2-16 <u>Proposición</u>.- Una funcional lineal en un espacio normado es continua si y sólo si es acotada.

<u>Demostración</u>: Sea F una funcional lineal. Si F es acotada, entonces es - continua puesto que

$$|F(X_n) - F(X)| = |F(X_n - X)| \leq M\|X_n - X\| \to 0 \quad \text{cuando } \|X_n - X\| \to 0.$$

Si F no es acotada, entonces no es continua puesto que para toda n existe un punto x_n tal que

$$|F(X_n)| > n\|X_n\|$$

y haciendo

$$Y_n = \frac{X_n}{n\|X_n\|}$$

tenemos $|F(Y_n)| > 1$ mientras que

$$\|Y_n\| = \frac{1}{n} \to 0 \quad .//$$

En base al teorema 2-10 y al teorema 2-15, tenemos el siguiente resultado, que es el teorema de representación de Riesz para funcionales - lineales acotadas en L^2.

2-17 <u>Teorema</u> Sea F una funcional lineal acotada en $L^2(\Omega, \mathcal{Q}, P)$. Existe un elemento único $Y \in L^2(\Omega, \mathcal{Q}, P)$ tal que

$$F(X) = \int (XY)\,dP \quad \forall\ X \in L^2(\Omega, \mathcal{Q}, P).$$

Vamos ahora a aplicar (2-17) para definir la esperanza condicional. Sean $X \in L^2(\Omega, \mathcal{Q}, P)$, $\mathcal{B}$ un sub-σ-campo de $\mathcal{Q}$ y $F: L^2(\Omega, \mathcal{B}, P) \to \mathbb{R}$

definida por

$$F(Z)=\int(XZ)dP .$$

Por la linealidad de la integral, resulta que F es lineal y por la desigualdad de Schwarz,

$$|F(Z)|\leq \|X\|\,\|Z\|$$

resulta que F es acotada. Aplicando el teorema 2-11 se sigue que existe $Y\in L^2(\Omega,\mathcal{B},P)$ tal que

$$F(Z)=\int(ZY)dP \quad \forall\ Z\in L^2(\Omega,\mathcal{B},P),$$

es decir,

$$\int(X\,Z)dP=\int(YZ)dP .$$

En particular, si $Z=I_B$, $B\in\mathcal{B}$, obtenemos:

$$\int_B XdP=\int_B YdP \quad \forall\ B\in\mathcal{B} .$$

Sea X positiva (no necesariamente integrable) y sea

$$X_n=\min(X,n)$$

Es claro que $X_n\in L^2(\Omega,\mathcal{B},P)$ y $X_n\uparrow X$. Para cada X_n, existe $Y_n\in L^2(\Omega,\mathcal{B},P)$ tal que:

$$\int_B X_n dP=\int_B Y_n dP \quad \forall\ B\in\mathcal{B} \quad \text{y} \quad P(0\leq Y_n\leq Y_{n+1})=1$$

Se sigue que existe una variable aleatoria $Y\in L^1(\Omega,\mathcal{B},P)$ tal que -- $P(Y_n\uparrow Y)=1$ y, aplicando el teorema de la convergencia monótona,

$$\int_B XdP=\int_B YdP \quad \forall\ B\in\mathcal{B} .$$

Ahora, en el caso en que $X=X^+-X^-$ es casi integrable, por lo - menos una de las variables aleatorias positivas X^+ o X^- es integrable- y, por lo tanto, al menos una de las variables aleatorias positivas y $\mathcal{B}$-medibles Y_+ o Y_- tales que

$$\int_B X^+dP=\int_B Y_+dP \quad , \quad \int_B X^-dP=\int_B Y_-dP \quad \forall\ B\in\mathcal{B}$$

es integrable. En consecuencia, $Y = Y_+ - Y_-$ está definida casi dondequiera, es $\mathcal{B}$-medible y ademas,

$$\int_B X dP = \int_B Y dP \quad \forall \ B \in \mathcal{B}$$

La siguiente proposición nos demuestra que la función Y obtenida anteriormente, es la esperanza condicional con respecto a $\mathcal{B}$ de la variable Xde la definición 1-17.

2-18 <u>Proposición</u>.- Una variable aleatoria $\mathcal{B}$-medible Y es la esperanza condicional con respecto a $\mathcal{B}$ de la variable aleatoria X si y sólo si

$$\int (ZX) dP = \int (ZY) dP$$

<u>Demostración</u>: Si $Y = E^{\mathcal{B}}X$, la definición 1-17 nos dice que

$$\int (ZX) dP = \int (ZY) dP$$

siempre que $Z = I_B$, $B \in \mathcal{B}$. Por linealidad, se obtiene la igualdad para Z variable aleatoria simple positiva y medible; por el teorema de la -- convergencia monótona, se tiene el resultado para el caso en que Z es una variable aleatoria positiva y $\mathcal{B}$-medible.

Observese que, en $L^2(\Omega, \mathcal{A}, P)$, la esperanza condicional no es más - que la proyección de $L^2(\Omega, \mathcal{A}, P)$ en su subespacio de Hilbert $L^2(\Omega, \mathcal{B}, P)$ Esta propiedad permite definir la esperanza condicional en $L^2(\Omega, \mathcal{A}, P)$ - (ver por ejemplo (8)).

Lo anterior se justifica ya que si $Y = P_{L^2(\Omega,\mathcal{B},P)}X$ entonces

a) $Y \in L^2(\Omega, \mathcal{B}, P)$

b) $\int (X - Y) Z dP = 0$ o $\int (XZ) dP = \int (YZ) dP \quad \forall \quad Z \in L^2(\Omega, \mathcal{B}, P)$.

ESPERANZA CONDICIONAL DADA UNA FUNCION

2-19 <u>Definición</u>.- Sea Y una función de $(\Omega, \mathcal{A}, P)$ al espacio medible -- $(\Omega', \mathcal{A}')$. Definimos los σ-campos inducidos por Y en Ω' y Ω, respectivamente, de la siguiente forma: $\mathcal{B}'_Y \subset \mathcal{A}'$ es el σ-campo de todos los conjuntos de $\mathcal{A}'$ cuya imagen inversa bajo Y son eventos $(\in \mathcal{A})$ y $\mathcal{B}_Y$ es el σ-campo de esos eventos. P_Y y $P_{Y'}$ son las probabilidades inducidas - por Y en $\mathcal{B}_Y$ y $\mathcal{B}'_Y$, respectivamente definidas por

$P_Y B = PB$, $B \in \mathcal{B}_Y$; $P'_Y B' = PB$ si $B' \in \mathcal{B}'_Y$ y $B = Y^{-1}(B')$.

Si $\mathcal{B} = \mathcal{B}_Y$, escribiremos $E^Y X$ en lugar de $E^{\mathcal{B}_Y} X$ y a $E^Y X$ la llamaremos la esperanza condicional de X dada Y . Para justificar la terminología, demostraremos que efectivamente, $E^Y X$ es una función de la función Y. Para ello usaremos el siguiente resultado.

2-20 <u>Proposición</u>.- Para toda función numérica medible g definida en $(\Omega', \mathcal{B}'_Y, P'_Y)$,

$$\int_{B'} g\,dP'_Y = \int_B g(Y)\,dP_Y \quad , \quad B' \in \mathcal{B}'_Y \quad , \quad B = Y^{-1}(B') ,$$

en el sentido que si una de las integrales existe, existe la otra y son iguales.

<u>Demostración</u>: Sea $g = I_{A'}$, $A' \in \mathcal{B}'_Y$. Tomamos $A = Y^{-1}(A')$ y obtenemos $g(Y) = I_A$ de donde

$$\int_{B'} I_{A'}\,dP'_Y = P'_Y(A'B') = P_Y(AB) = \int_B I_A\,dP_Y \quad .$$

Siendo válida para indicadores, la igualdad es válida para funciones g-simples y, por el teorema de la convergencia monótona, para g medible no-negativa. La afirmación se sigue descomponiendo la función g medible en su parte positiva y en su parte negativa.

2-21 <u>Proposición</u>.- La esperanza condicional de $X \in \mathbb{E}$ dada Y es una función de la función Y.

<u>Demostración</u>: Si φ es la integral indefinida de X y φ' en $\mathcal{B}'_Y$ está definida por

$$\varphi'(B') = \varphi(B) \ , \ B' \in \mathcal{B}'_Y \quad , \quad B = Y^{-1}(B')$$

entonces φ' es σ-aditiva y P'_Y-continua. Aplicamos el teorema de Radon-Nikodym extendido (1-12) y definimos una función medible g en $(\Omega', \mathcal{B}'_Y)$ por

$$\int_{B'} g\,dP'_Y = \varphi'(B') = \int_B X\,dP$$

Puesto que

$$\int_B X\,dP = \int_B (E^Y X)\,dP \ ,$$

se sigue aplicando 2-20 que

$$\int_B g(Y)\,dP_Y = \int_{B'} g\,dP'_Y = \int_B (E^Y X)\,dP \quad .$$

Por lo tanto, las integrales indefinidas de las funciones $\mathcal{B}_Y$-medibles $g(Y)$ y $E^Y X$ son las mismas, y así, la afirmación está probada. //

CAPITULO III

En los capítulos anteriores, definimos la esperanza condicional en su forma más general auxiliándonos de dos resultados "ajenos" a la teoría de probabilidad: El teorema de Radon-Nikodym y el teorema de representación de Riesz.

En ambos casos, la definición obtenida es puramente descriptiva: Para toda $X \in \mathcal{E}$ existe una función $\mathcal{B}$-medible $E^{\mathcal{B}}X$ tal que

$$\int_B (E^{\mathcal{B}}X)\,dP = \int_B X\,dP \qquad \forall\ B \in \mathcal{B} \qquad \text{-----------} \quad 3\text{-}1$$

Esta definición se obtiene como una generalización de la definición de esperanza condicional en el caso elemental:

$$E^{\mathcal{B}}X = \sum_j (E_{B_j}X)\, I_{B_j} \qquad \text{------------} \quad 3\text{-}2$$

donde $\{B_j\}$ es una partición a lo más numerable de Ω que genera a $\mathcal{B}$.

Consideremos la definición de la esperanza condicional como se obtuvo en el capítulo I (la esperanza condicional es una derivada de Radon-Nikodym). No es fácil "ver" la relación que existe entre la definición en el caso elemental (3-2) y la definición en el caso mas general (3-1)

Para esclarecer el enlace que existe entre (3-1) y (3-2) enunciaremos y demostraremos el teorema de Radon-Nikodym como un teorema de probabilidad, obteniendo de esta manera una definición constructiva de la esperanza condicionalen su forma más general.

Versión Probabilista del Teorema de Radon-Nikodym.

Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad, $\mathcal{B}$ un sub-σ-campo de $\mathcal{A}$ y Y una función $\mathcal{A}$-medible de Ω a $[-\infty, \infty]$ cuya integral existe. Considérese el espacio de probabilidad $(\Omega, \mathcal{B}, P_{\mathcal{B}})$ donde $P_{\mathcal{B}}$ es la restricción de P a $\mathcal{B}$ y la función conjuntista φ definida por

$$\varphi(B) = \int_B Y dP \qquad , \; B \in \mathcal{B}$$

Recuérdese que φ es σ-aditiva y $P_{\mathcal{B}}$-continua.

TEOREMA.- Existe una sucesión creciente $\{\mathcal{B}_n \,;\, n=1,2,\ldots\}$ de sub-σ-campos de $\mathcal{B}$, cada uno generado por una partición numerable $\{B_{nk};\ k=0,1,2,\ldots\}$ de Ω tal que la sucesión $\{E^{\mathcal{B}_n}Y\}$ tiene un límite puntual F, satisfaciendo para todo $B \in \mathcal{B}$,

$$\int_B (E^{\mathcal{B}_n}Y) dP_{\mathcal{B}} \longrightarrow \int_B F dP_{\mathcal{B}} \quad \text{y} \quad \int_B (E^{\mathcal{B}_n}Y) dP_{\mathcal{B}} \longrightarrow \int_B Y dP \equiv \varphi(B) \,.$$

Así, $E^{\mathcal{B}}Y$ existe y es el límite de esperanzas condicionales definidas constructivamente.

Demostración: Probaremos el teorema para Y no negativas, puesto que en el caso general podemos tomar $E^{\mathcal{B}}Y = E^{\mathcal{B}}Y^+ - E^{\mathcal{B}}Y^-$.

Sea la familia de medidas con signo

$$\mathcal{F} = \{\varphi - \alpha P_{\mathcal{B}} \,;\, \alpha \geq 0\}$$

Por el teorema de descomposición de Hahn, existe un conjunto $A_\alpha \in \mathcal{B}$ tal que para cualquier $B \in \mathcal{B}$,

$$B \subset A_\alpha \Longrightarrow (\varphi - \alpha P_{\mathcal{B}})(B) \geq 0$$
$$B \subset A_\alpha^c \Longrightarrow (\varphi - \alpha P_{\mathcal{B}})(B) \leq 0$$

o equivalentemente,

$$B \subset A_\alpha \Longrightarrow \int_B Y dP \geq \alpha P(B)$$
$$B \subset A_\alpha^c \Longrightarrow \int_B Y dP \leq \alpha P(B) \,.$$

Por supuesto, los A_α's no son únicos (ver teorema 1-9). Sin embargo, para una colección numerable de α's (usaremos el conjunto $\{\frac{k}{2^n}\,;\ k,n=0,1,2,\ldots\}$) podemos escoger los A_α de tal manera que sean crecientes. Empezamos con cualquier sucesión $\{A^*_\alpha\}$ y definimos

$$A_u = A^*_u - \bigcup_{v<w} (A^*_w - A^*_v)$$

Esto es válido porque $P(A^*_w - A^*_v) = 0$ si $w > v$. Además, puesto que $Y \geq 0$, podemos tomar $A_0 = \Omega$. Sea A_∞ el límite (intersección) de todas nuestras A_u ($P(A_\infty) = 0$ si Y es integrable) y nótese que A_∞ es también el límite de cada subsucesión

$$\left\{ A_{\frac{k}{2^n}} \; ; \; k=0,1,\ldots \right\} .$$

Así, para cada n, la sucesión

$$\mathcal{B}_n = \left\{ B_{nk} = A_{\frac{k}{2^n}} - A_{\frac{k+1}{2^n}} \; ; \; k=0,1,\ldots; \; A_\infty \right\}$$

es una partición $\mathcal{B}$-medible de Ω.

Ahora, las esperanzas condicionales de Y dado el σ-campo $\mathcal{B}_n$, donde $\mathcal{B}_n$ es (abusando de la notación) el σ-campo generado por la partición $\mathcal{B}_n$, están dadas por:

$$E^{\mathcal{B}_n} Y = \sum_k (E_{B_{nk}} Y) I_{B_{nk}} + \infty I_{A_\infty}$$

donde

$$E_{B_{nk}} Y = \int_{B_{nk}} \frac{Y dP}{P(B_{nk})} .$$

El hecho clave es que para cualquier $B \in \mathcal{B}$,

$$k2^{-n} \leq \int_{BB_{nk}} \frac{Y dP}{P(BB_{nk})} \leq (k+1)2^{-n}$$

puesto que

$$B_{nk} \subset A_{\frac{k}{2^n}} \qquad y \qquad B_{nk} \subset A^c_{\frac{k+1}{2^n}} .$$

Siendo las $\mathcal{B}_n$ encajadas, concluimos que $\{E^{\mathcal{B}_n} Y\}$ converge puntualmente a una función $\mathcal{B}$-medible F satisfaciendo para cada n y k,

$$k2^{-n} \leq F \leq (k+1)2^{-n} \quad \text{en } B_{nk} .$$

Si $P(BA_\infty) > 0$,

$$\int_B (E^{\mathcal{B}_n} Y)\, dP_{\mathcal{B}} = \infty = \int_B F\, dP = \int_B Y\, dP \;,$$

mientras que si $P(BA_\infty) = 0$, *se verifica que*

$$\left| \int_B (E^{\mathcal{B}_n} Y - F)\, dP_{\mathcal{B}} \right| \leq \sum_k \int_{BB_{nk}} \left| E^{\mathcal{B}_n} Y - F \right| dP_{\mathcal{B}} =$$

$$\sum_k \int_{BB_{nk}} \left| E_{BB_{nk}} Y - F \right| dP_{\mathcal{B}} \leq \sum_k 2^{-n}\, P(BB_{nk}) \leq 2^{-n}$$

y

$$\left| \int_B (E^{\mathcal{B}_n} Y - Y)\, dP \right| = \left| \sum_k \left[\int_{B_{nk}} \frac{Y dP}{P(B_{nk})} \right] P(BB_{nk}) - \sum_k \int_{BB_{nk}} Y dP \right| \leq$$

$$\sum 2^{-n}\, P(BB_{nk}) \leq 2^{-n} \;.//$$

Como un caso particular, consideremos $\mathcal{B} = \sigma(X)$ *donde* X *es una variable aleatoria y* $\sigma(X) = \{X^{-1}(A) : A \text{ es un Boreliano}\}$.
Sea $B_{nk} = X^{-1}\left[\frac{k}{n}, \frac{k+1}{n}\right]$ *y sea* $\mathcal{B}_n$ *el* σ*-campo generado por la partición de* Ω $\{B_{nk} ; k=0,1,\dots\}$
Definamos

$$X_n = \sum_k \frac{k}{n}\, I_{\left[\frac{k}{n} \leq X \leq \frac{k+1}{n}\right]}$$

La esperanza condicional de una variable aleatoria Y *dado* X_n, $E(Y|X_n)$, *es el caso elemental ya que* $\sigma(X_n)$ *está generado por una partición numerable de* Ω *y así, obtenemos la definición constructiva:*

$$E(Y|X) = \lim_{n\to\infty} E(Y|X_n)$$

Consideremos el caso especial de probabilidad condicional: Reemplazamos Y por $I_{[Y\in B]}$ donde B es algún conjunto de Borel y $E(Y|X)$ se reemplaza por $P(Y\in B|X)$. Para el caso particular $P(Y\in B|X=x)$ (El valor común de $P(Y\in B|X)$ en el evento $X^{-1}(\{x\})$), tenemos las tres definiciones siguientes:

$$P(Y\in B|X=x)=\lim_{n\to\infty} P(Y\in B|X_n=x_n) \quad ,\qquad \text{------------(3-3)}$$

donde las X_n son funciones elementales que convergen a X c.d. y x_n es el valor de X_n cuando $X=x$. Por ejemplo,

$$X_n=\sum_k \left(\frac{k}{n}\right) I_{\left[\frac{k}{n}\leq X<\frac{k+1}{n}\right]}$$

$$x_n=\sum_k \left(\frac{k}{n}\right) I_{\left[\frac{k}{n}\leq x<\frac{k+1}{n}\right]}$$

Otra posible definición es:

$$P(Y\in B|X=x)=\lim_{h\to 0} P(Y\in B|\, x\leq X\leq x+h) \qquad \text{-----------(3-4)}$$

La tercera definición se aplica al caso en que la distribución conjunta de X y Y es absolutamente continua con respecto a la medida de Lebesgue, es decir, para alguna densidad F,

$$P(X\in A,\ Y\in B) = \iint_{A\times B} F(x,y)\,dx\,dy$$

para cada rectángulo de Borel $A\times B$. La definición es:

$$P(Y\in B|X=x)=\frac{\int_B F(x,y)\,dy}{\int_{-\infty}^{\infty} F(x,y)\,dy} \qquad \text{------------(3-5)}$$

Para justificar la tercera definición:

$$P(Y\in B\,|\,X=x)=\lim_{h\to 0} P(Y\in B\,|\,x\le X\le x+h)$$

$$=\lim_{h\to 0}\frac{\int_x^{x+h}\left(\int_B F(u,y)dy\right)du}{\int_x^{x+h}\left(\int_{-\infty}^{\infty} F(u,y)dy\right)du}=$$

$$=\lim_{h\to 0}\frac{\frac{1}{h}\int_x^{x+h}\left(\int_B F(u,y)dy\right)du}{\frac{1}{h}\int_x^{x+h}\left(\int_{-\infty}^{\infty} F(u,y)dy\right)du}\,.$$

Si $\int_B F(u,y)dy$ *es continua en u cuando* $u=x$ *y* $F_X(u)=\int_{-\infty}^{\infty} F(u,y)dy$ *es continua en* $u=x$ *obtenemos*

$$P(Y\in B\,|\,X=x)=\frac{\int_B F(x,y)dy}{F_X(x)}$$

EJEMPLOS ILUSTRATIVOS

1.- <u>El problema de la mejor elección</u>. Se tiene una baraja de n cartas-distintas (por ejemplo n=52), las cuales están arbitrariamente graduadas. Una persona, extrae las cartas una por una e informa solo una cosa: La carta extraída es o no es "la mejor hasta ahora". Oyendo esto, se debe decidir inmediatamente si se acepta o no la carta. Se puede escoger a lo más una carta y se gana si y sólo si la carta escogida es la mejor de las n cartas.

Para n grande, puede parecer que las oportunidades de ganar son - muy escasas, pero esto no es así como lo demuestra la siguiente solu-- ción.

Se escoge un entero k entre 1 y n y se usa la siguiente estrategia: Dejar pasar las k primeras cartas no importa cuales sean. Después tomar la primera carta, la cual es la mejor hasta ahora. Sea B el evento de -

"ganar el juego" y para $i=1,2,\ldots,n$, sea G_i el evento que la mejor de las n cartas es la i-ésima. Nótese que en G_i para $i>k$ se gana si y sólo si la mejor de las primeras $i-1$ cartas está entre las primeras k. Entonces se tiene que

$$P_k(B)=\sum_i P(B|G_i)P(G_i)=\sum_{k+1\le i\le n} P(B|G_i)P(G_i)=\sum_{k+1\le i\le n}\left(\frac{k}{i-1}\right)\left(\frac{1}{n}\right).$$

Para $k=\frac{n}{2}$, esta probabilidad es al menos $\frac{1}{4}$, no importa que tan grande sea n. El valor óptimo de $k=k(n)$ satisface

$$\sum_{k<i<n} i^{-1}\le 1<\sum_{k\le i<n} i^{-1},$$

de lo cual concluimos que

$$\frac{k(n)}{n}\longrightarrow e^{-1}\quad,\quad P_{k(n)}(B)\longrightarrow e^{-1}\approx .36\ .$$

2.- Problema de Diagnostico. Supóngase que una persona puede tener una y sólo una de las enfermedades en el conjunto $\Delta=\{\delta_0,\delta_1,\ldots\}$ (δ_0 para 'salud perfecta'). Sea σ un síntoma (o colección de síntomas). Una persona escogida al azar tendrá la enfermedad δ_i con probabilidad $Q(\delta_i)$ igual a la frecuencia relativa de la enfermedad en la población. Las probabilidades condicionales, $R(S|\delta_i)$, $S\in\mathcal{L}$, donde $\mathcal{L}$ es un σ-campo de subconjuntos de Σ (el conjunto de síntomas) se conocen por medio de estudios clínicos. Lo único que un médico observa es algún síntoma σ, con lo cual él trata de diagnosticar al paciente su condición; es decir, trata de evaluar $P(\delta_i|\sigma)$.

Si Σ es numerable, esta probabilidad está bien definida:

$$P(\delta|\sigma)=\frac{P(\delta\sigma)}{P(\sigma)}$$

donde

$$P(\delta\sigma)=Q(\delta)R(\sigma|\delta)\quad,\quad P(\sigma)=\sum_i Q(\delta_i)R(\sigma|\delta_i).$$

Si Σ es no numerable, definimos en el espacio medible

$$(\Omega,\mathcal{Q})=(\Delta\times\Sigma,\ 2^{\Delta}\times\mathcal{L})$$

la medida de probabilidad

$$P(\delta\times S)=Q(\delta)R(S|\delta)$$

identificando $P(\delta\sigma)$ con $P(\delta\times\{\sigma\})$ y $P(\sigma)$ está dada por:

$$P(\sigma)=\sum_i P(\{\delta_i\}\times\{\sigma\})$$

3.- Problema X-Y. Escogemos aleatoriamente un número Y en (0,1), es decir, $P(Y \leq y)=y$ si $0<y<1$. Después escogemos aleatoriamente un número X en (Y,1). Solamente se observa X. Dado su valor, ¿cual es la distribución condicional de Y?

En este caso usaremos la definición (3-5).

Sea la distribución conjunta de X y Y:

$$P(X \in A, Y \in B)=\iint_{A\times B} F(x,y)\,dx\,dy$$

con la densidad

$$F(x,y)=\begin{cases} \frac{1}{1-y} & 0<y<x<1 \\ 0 & \text{en cualquier otro caso} \end{cases}$$

entonces, para todo $x \in (0,1)$,

$$P(Y \in B \mid X=x)=\int_B g(y|x)\,dy$$

donde

$$g(y|x)=\frac{F(x,y)}{\int_0^1 F(x,z)\,dz}$$

$$=\frac{-1}{(1-y)\,Ln(1-x)} \qquad 0<y<x<1$$

$$=0 \qquad \text{en cualquier otro caso.}$$

4.- Paradoja de Borel Supóngase que un punto es escogido aleatoriamente en la superficie de la tierra la cual consideramos esférica. Encontrar:

a) La distribución condicional de la longitud del punto, dado que cae en el ecuador.

b) La distribución condicional de la latitud del punto, dado que cae en el meridiano de Greenwich. (Kolmogorov (7) presentó una versión de este problema bajo el título: "Explicación de una paradoja de Borel").

Puesto que el punto de la esfera, se escoge de acuerdo a una distribución uniforme simetrica, ~~si~~ la parte a) tiene respuesta, la parte b) -- tendrá la misma respuesta.

Ahora tenemos la paradoja: La longitud ψ y la latitud θ son variables aleatorias perfectamente definidas. La distribución de ψ es de hecho uniforme, mientras que la distribución de θ no lo es (su densidad es proporcional a $\cos\theta$, como lo veremos después). Además, θ y ψ son independientes, así que las distribuciones condicionales de cada - una dada la otra, son simplemente sus distribuciones (no condicionales). De esta manera tendremos respuestas a la parte a) y a la parte b), ¡pero no son las mismas!. Lo anterior nos indica que la respuesta depende de la descomposición: Círculos meridianos con polos comunes en el círculo dado nos lleva a la distribución uniforme, mientras que círculos formados por la intersección de la superficie esférica con los planos paralelos al plano determinado por el círculo dado (ecuador) lleva a la distribución uniforme.

La distribución de ψ está dada por:

$$P(\psi_1 \leq \psi \leq \psi_2) = \frac{\psi_2 - \psi_1}{\pi}$$

mientras que la distribución de θ es:

$$P(\theta_1 \leq \theta \leq \theta_2) = \frac{\int_{\theta_1}^{\theta_2} 2\pi R^2 \cos\theta \, d\theta}{4\pi R^2} = \frac{1}{2}\int_{\theta_1}^{\theta_2} \cos\theta \, d\theta$$

Para demostrar la independencia:

$$P(\psi_1 \leq \psi \leq \psi_2 \, , \, \theta_1 \leq \theta \leq \theta_2) = \frac{2\int_{\theta_1}^{\theta_2} (\psi_2 - \psi_1) R^2 \cos\theta \, d\theta}{4\pi R^2} =$$

$$= \frac{2(\psi_2 - \psi_1) R^2 \int_{\theta_1}^{\theta_2} \cos\theta \, d\theta}{4\pi R^2} = \left(\frac{\psi_2 - \psi_1}{\pi}\right) \frac{1}{2}\int_{\theta_1}^{\theta_2} \cos\theta \, d\theta$$

$$= P(\psi_2 \leq \psi \leq \psi_2) P(\theta_1 \leq \theta \leq \theta_2) \, .$$

REFERENCIAS

1. M. Loeve, *Probability Theory Vol. I y II*, Springer-Verlag New York (1978).

2. S.M. Samuels, *The Radon-Nikodym as a Theorem in Probability*, American Mathematical Monthly (1978) 155-165.

3. H.L. Royden, *Real Analysis*, Macmillan Publishing Co. Inc. - New York (1968).

4. J. Neveu, *Mathematical Foundations of the Calculus of Probability*, Holden-Day Inc. San Francisco (1965) .

5. Rudin, *Real and Complex Analysis*, McGraw-Hill (1974).

6. J. Dieudonné, *Foundations of Modern Analysis*, Academic Press (1962).

7. A.N. Kolmogorov, *Foundations of the Theory of Probability*, Chelsea, New York (1956).

8. J. Neveu, *Discrete Parameter Martingales*, North Holland -- Publishing Company, New York (1975).