UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE CIENCIAS

ESPERANZA CONDICIONAL

T E S I S

QUE PARA OBTENER EL TITULO DE:

A C T U A R I A

P R E S E N T A:

MARICELA ESPERANZA ALONSO QUIROZ

DIRECTOR DE TESIS:
DR. MIGUEL ANGEL GARCIA ALVAREZ

2002

UNIVERSIDAD NACIONAL
AUTÓNOMA DE
MÉXICO

M. EN C. ELENA DE OTEYZA DE OTEYZA
Jefa de la División de Estudios Profesionales de la
Facultad de Ciencias
Presente

Comunicamos a usted que hemos revisado el trabajo escrito: ESPERANZA CONDICIONAL

realizado por ALONSO QUIROZ MARICELA ESPERANZA

con número de cuenta 8627370-1, quién cubrió los créditos de la carrera de ACTUARIA

Dicho trabajo cuenta con nuestro voto aprobatorio.

Atentamente

Director de Tesis Propietario	DR. MIGUEL ANGEL GARCIA ALVAREZ	
Propietario	M. en C. JULIO CESAR CEDILLO SANCHEZ	
Propietario	ACT. MARISA MIRANDA TIRADO	Marisa Miranda Tirado
Suplente	ACT. JAIME VAZQUEZ ALAMILLA	
Suplente	M. en C. VINICIO ANTONIO GOMEZ GUTIERREZ	V.A. Gómez G.

Consejo Departamental de MATEMATICAS

M. en C. JOSE ANTONIO FLORES DIAZ

FACULTAD DE CIENCIAS
CONSEJO DEPARTAMENTAL
DE
MATEMATICAS

2

Este trabajo está dedicado:

A Judith Quiroz y Calixto Alonso.
Por su gran esfuerzo por sacarnos adelante.

A Vini
Por su apoyo incondicional, su ejemplo, su confianza, su lucha, ...

A Migue.
Porque confía en mi.

A mis hermanos:
Pedro, Mari, Héctor, Blanca, Jaci, Pita, Conchita, Juan, Raúl, Humberto, Sergio, Lucía.

A mis amigos:
Aida, Moni, Héctor J.S, Yerard, Martha.

A Jacel, Israel, Moni, Cachito, Quimihi, Cesar, Danielito, Arely, Bety, Raulito, Iñaqui, Kenet, Laila, Jorge, Diego. Con la confianza de que algún día irán a la universidad.

A Miguel Angel García
Por su paciencia durante tantos años.

A todos los que han luchado por cambiar este mundo, por uno mejor, donde la educación sea un derecho humano elemental.
En especial : **Al CGH**, por encabezar la huelga heróica de 1999.
A la Revolución cubana y a Fidel.
A Javier y Mario

ESPERANZA CONDICIONAL

CONTENIDO

Introducción

El siguiente trabajo pretende ser de utilidad en un curso de Probabilidad II, para estudiantes de licenciatura en la Facultad de Ciencias.

Aunque es un material autocontenido, es decir, que no presupone conceptos previos de probabilidad, si se requiere que el lector tenga conocimientos de Cálculo Diferencial e Integral y cierta madurez en Matemáticas.

El tema central en este trabajo es la parte de Esperanza Condicional, el cual se desarrollará mas exhaustivamente.

La teoría de la Probabilidad es una rama de las Matemáticas como la Geometría o la Topología y se le puede dar un tratamento axiomático como a estas. A lo largo de este trabajo se usara la formalidad axiomática sin perder de vista el tratamiento intuitivo.

En el capítulo 1 se darán las definiciones y se expondrán los resultados que se requieren en los capítulos posteriores. En la primera parte del capítulo se exponen los conceptos y resultados básicos del Cálculo de Probabilidades. En la segunda parte se expone todo lo relativo a las variables aleatorias.

En la primera parte, comenzamos definiendo lo que es un espacio de probabilidad, el cual está compuesto de tres elementos: el espacio muestral, la familia de eventos y la medida de probabilidad. Después se introduce el concepto de probabilidad condicional y de independencia de eventos para concluir con la llamada Regla de la Probabilidad Total.

En la segunda parte se introduce el concepto de variable aleatoria y se define lo que es una variable aleatoria discreta y una variable aleatoria absolutamente continua. Más adelante se exponen los ejemplos básicos de distribuciones discretas y absolutamente continuas. Finalmente se introduce el concepto de Esperanza Matemática.

En el capitulo 2 se estudiara lo relativo a las distribuciones y esperanzas condicionales de variables aleatorias discretas.

Sobre distribuciones condicionales se motiva la definición, se dan ejemplos y se concluye con una proposición acerca de la distribución condicional de dos variables aleatorias independientes dada una de ellas.

En la parte de esperanzas condicionales, se empieza con la esperanza condicional de una variable aleatoria dado un evento de probabilidad positiva, basándonos en la misma idea de la probabilidad condicional de un evento dado otro evento de probabilidad positiva. A partir de donde definimos la esperanza condicional de una variable aleatoria dado que otra variable aleatoria tomo un valor especifico y con esto se define la esperanza condicional de una variable aleatoria dada otra variable aleatoria. Además se expone **la regla general de la probabilidad total** para variables aleatorias discretas y se utiliza para calcular esperanzas y varianzas no condicionales. Mas adelante se motiva una proposición que nos permitira dar **la definicion general de esperanza condicional de una variable aleatoria dada otra** sin importar como sean estas. En esta segunda parte se prueban las propiedades de las esperanza condicional de una variable aleatoria dada otra. Aunque la mayoria de las propiedades serán probadas unicamente en el caso discreto, hay algunas que se prueban en general para cualquier par de variables aleatorias.Se estudia también la esperanza condicional de una función de dos variables aleatorias discretas dada una de ellas. Finalmente se da la definición de varianza condicional de una variable aleatoria dada otra y se termina con una proposición que nos permite calcular varianzas no condicionales usando varianza condicional.

En el capítulo 3 se estudian las distribuciones y esperanzas condicionales de variables aleatorias absolutamente continuas.

En la parte de las distribuciones condicionlaes se empieza definiendo la distribución condicional de una variable aleatoria dado que otra variable aleatoria tomo un valor específico, esta condición es un evento de probabilidad cero sin embargo la definición que se da permite esto y lo único que se le pide a la variable que condiciona es que su función de densidad, que existe por ser absolutamente continua, sea mayor que cero en el valor específico.Después de esto se define la distribución condicional de una variable aleatoria dado cualquier evento.

En la sección de esperanzas condicionales, se da la definición en general de esperanza de una variable aleatoria dada otra variable aleatoria , se llega a la regla de la probabilidad total para variables aleatorias absolutamente continuas de manera general, se prueban sus propiedades. Después de esto se expone lo relativo a esperanzas condicionales de funciones de dos variables aleatorias dada una de ellas y finalmente se trata la varianza condicional para variables aleatorias absolutamente continuas.

El capítulo 4 consta de dos partes una que se dedica a las distribuciones y esperanzas condicionales en el caso mixto, es decir, en el caso en que alguna de ellas es absolutamente continua y la otra es discreta. En esta parte se define una función de densidad conjunta en el caso mixto y usándola se define la esperanza condicional en este mismo caso. La segunda parte es un esbozo de la necesidad de introducir la Teoría de la Medida como una herramienta indispensable en el desarrollo posterior de la Esperanza Condicional. Se demuestran resultados como el **teorema de Radon Nikodym** (1930) y se utiliza éste para demostrar la existencia de la esperanza condicional en el caso general.

Capítulo I

La Teoría de la Probabilidad como una rama de la Matemática tiene su origen en 1654, con las soluciones de problemas a juegos de azar , que dieron Pascal Fermat y Huygens. Sin embargo antes de estas soluciones con las que se puede decir, comienza el Cálculo de Probabilidades, ya existía evidencia de formulación y solución de problemas de probabilidad en 1477 y en 1487 en el libro de Paccioli llamado "Proporcionalitá".

No es sino hasta principios del siglo pasado que se da la fundamentación matemática de la Teoría de la Probabilidad. Es A.Kolmogorov en 1993 quien estableció la formulación moderna (axiomática) de la Teoría de la Probabilidad en su libro " Foundations of the Theory of Probability"

Espacios de Probabilidad

Lo que se busca en esta sección es modelar matemáticamente a un experimento aleatorio, es decir un experimento en el cual no se puede saber de antemano el resultado.

Espacio Muestral

Consideremos un experimento aleatorio, es decir un experimento en el cual de antemano no se sabe el resultado de éste. Al conjunto de todos los resultados posibles de dicho experimento se le llama **espacio muestral** y se le denota por Ω

Ejemplo 1.1. Consideremos el experimento de elegir al azar una persona y medir su estatura en metros. Los posibles resultados del experimento, es decir todas las estaturas posibles para una persona, estan en el conjunto Ω siguiente:

$\Omega = \{x \in \mathbb{R} \mid x \in (0,2.5)\}$

Ejemplo1.2 Un prisionero está en una celda que contiene tres puertas. La primera puerta da a un tunel que lo regresa a su celda después de dos días de caminar. La segunda da a un tunel que lo regresa a su celda después de cuatro días de caminar. La tercera lo deja libre después de un dia de caminar. Considerando que en cada ocasión el prisionero elige al azar cualquiera las puertas 1,2 y 3, que el experimento empieza cuando el prisionero elige la primera puerta y termina cuando queda libre y que el resultado del experimento es el tiempo (en días) que tarda el prisionero en salir, el espacio muestral de este experimento está dado por el siguiente conjunto Ω:

$\Omega = \{1,3,5,7,\ldots\}$

Eventos

Un evento es una característica relativa a un experimento aleatorio. Un evento ocurre si el resultado del experimento tiene la característica del evento. Por ejemplo: Si el experimento es lanzar dos dados , un posible evento es "que la suma salga 8". En este caso Ω es el conjunto: $\{(1,1),(1,2),\ldots,(1,6),(2,1),(2,2),\ldots,(2,6)\ldots(6,1)(6,2),\ldots,(6,6)\}$ Donde cada una de las entradas de los elementos de Ω representa a cada uno de los dados. El evento "que la suma salga 8" lo podemos ver como el conjunto:$\{(2,6),(3,5),(4,4),(5,3),(6,2)\}$. Si el resultado del experimento es cualquiera de los elementos de este conjunto entonces ocurre el evento "que la suma salga 8".

En general si Ω es un conjunto finito o infinito numerable, podemos definir a un evento como cualquier subconjunto de Ω.

Si un evento se caracteriza por cierta propiedad, entonces hay otro evento cuya característica es que no tiene tal propiedad. Dados dos eventos, hay otro evento que tiene la caracteristica de que cumple al menos una de las propiedades de los dos eventos dados.

Es decir: Si A y B son eventos entonces se cumple que:

i) A^c tambien es un evento.

ii) $A \cup B$ también es un evento.

Diremos que un evento es elemental si y sólo si es uno de los posibles resultados del experimento.

Si Ω es infinito no numerable, no se puede afirmar que todo subconjunto de Ω sea un evento pues nos enfretariamos aquí con el problema de la asignación de probabilidades a todo subconjunto de Ω. Éste es un problema histórico llamado "el problema de la medida" por Lebesgue en 1902.

Sin embargo lo que si podemos decir es que hay una familia $\mathcal{A}$ de subconjuntosde Ω , formado de todos los posibles eventos.

En general vamos a decir que un conjunto E es un evento si $E \in \mathcal{A}$, en donde $\mathcal{A}$ es una σ-álgebra que se define mas adelante.

Definición 1.3 Sea Ω un conjunto. Una familia $\mathcal{A}$ de subconjuntos de Ω es un álgebra si y sólo si $\mathcal{A}$ satisface las siguientes condiciones: i) $\Omega \in \mathcal{A}$.

ii) Si $A \in \mathcal{A}$ entonces $A^c \in \mathcal{A}$.

iii) Si $A_1, A_2, \ldots, A_n \in \mathcal{A}$ entonces $\bigcup_{k=1}^{n} A_k \in \mathcal{A}$.

Si queremos que nuestro conjunto de eventos sea lo mas amplio posible diremos que un evento es un elemento de una σ –álgebra , que se define a continuación:

Definición 1.4 Sea Ω un conjunto. Una familia $\mathcal{A}$ de subconjuntos de Ω es un σ –álgebra si y solo si $\mathcal{A}$ satisface las siguientes condiciones: i) $\Omega \in \mathcal{A}$.

ii) Si $A \in \mathcal{A}$ entonces $A^c \in \mathcal{A}$.

iii) Si$A_1, A_2, \ldots \in \mathcal{A}$ entonces $\bigcup_{k=1}^{\infty} A_k \in \mathcal{A}$.

En adelantenos llamaremos a los elementos de la σ –álgebra de subconjuntos de Ω eventos. Hay que notar que $\emptyset \in \mathcal{A}$, es decir que $\emptyset$ es un evento.

En general una σ –álgebra es cerrada no solo bajo complementación y uniones numerables sino también bajo intersecciones.

Si A y B son dos eventos cualesquiera, es decir dos elementos de $\mathcal{A}$ entonces diremos que A y B son mutuamente excluyentes si $A \cap B = \emptyset$. En general si $A_1, A_2, \ldots, A_n \in \mathcal{A}$ y además son ajenos dos a dos como conjuntos, diremos que $A_1, A_2, \ldots, A_n$ son eventos mutuamente excluyentes.

Medida de Probabilidad

De la sección anterior se tiene un experimento, el conjunto Ω de todos los posibles resultados del experimento, y el conjunto $\mathcal{A}$ de todos los posíbles eventos a los cuales les queremos asignar una probabilidad. A la pareja $(\Omega, \mathcal{A})$ se le llama **espacio medible**.

La probabilidad de un evento se puede interpretar como una medida de que tan creíble es la ocurrencia de ese evento. O bien como una medida de la incertidumbre asociada a ese evento.

Como ya se dijo anteriormente no necesariamente le vamos a poder asignar probabilidad a cualquier subconjunto de Ω pero a todo elemento de $\mathcal{A}$ donde $\mathcal{A}$ es la σ –álgebra generada por los intervalos abiertos si se le puede asignar una probabilidad.

Si pensamos a la probabilidad como una medida en el sentido natural del término, entonces la función $P : \mathcal{A} \to \mathbb{R}$ que a cada elemento de la σ –álgebra le asigna su probabilidad tiene que ser una función mayor o igual a cero. Además si Ω es el conjunto de todos los posibles resultados del experimento aleatorio, entonces Ω siempre ocurre y por ende $P(\Omega)$ tiene que ser la mayor probabilidad que se le asigne a un evento.

Si A y B no se intersectan con $P(A)$ y $P(B)$ sus respectivas probabilidades, entonces suena lógico asignarle al evento $A \cup B$ la probabilidad $P(A) + P(B)$ ya que si vemos a la probabilidad de un evento como el número de veces que ocurre el evento, entre el número de veces que se repite el experimento y tomando en cuenta que A y B son ajenos, el número de veces que ocurre $A \cup B$ es la suma de las veces que ocurre A más las veces que ocurre B. De todo esto surge la siguiente definición:

Definición 1.5 Sea Ω un conjunto y $\mathcal{A}$ una una σ -álgebra sobre Ω. Se dice que una función $P : A \to \mathbb{R}$ es **una medida de probabilidad** si se satisfacen las siguientes propiedades:

1) $P(A) \geq 0$ para cualquier $A \in \mathcal{A}$

2) $P(\Omega) = 1$

3) Si $A_1, A_2, \ldots$ son eventos mutuamente excluyentes, entonces

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

Además a la terna $(\Omega, \mathcal{A}, P)$ se le llama **Espacio de Probabilidad**

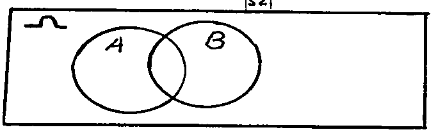
Definición 1.6 Se dice que la terna es un espacio de probabilidad si Ω es un espacio de probabilidad, $\mathcal{A}$ es una σ -álgebra sobre Ω y P es una medida de probabilidad.

Probabilidad Condicional.

Uno de los conceptos más importantes en probabilidad es el de **probabilidad condicional**

Supongamos que todos los posibles resultados de nuestro experimento aleatoria son los puntos del rectángulo de la figura de abajo. Y supongamos también que que se tienen dos eventos A y B también en la figura de abajo. La probabilidad de que ocurra A, $P(A)$ la podemos calcular como el área de A entre el area de Ω denotadas por $|A|$ y $|\Omega|$ respectivamente. De la misma forma

$$P(B) = \frac{|B|}{|\Omega|}$$



grafica

Supongamos ahora que sabemos que el resultado del experimento es un punto de B. Si queremos obtener la probabilidad de A basta con obtener el cociente $\frac{|A\cap B|}{|B|}$ pues si el resultado del experimento está en B, se puede ver a B como un nuevo espacio muestral, y el hecho de calcularle la probabilidad a A ahora se reduce a calcular el área de la intersección (pues el resultado del experimento no puede ser un punto de A que no pertenece a B, pues B ya ocurrió).

Resumiendo. La probabilidad de A dado que el resultado del experimento es un punto de B denotada por $P(A \mid B)$, es $P(A \mid B) = \frac{|A \cap B|}{|B|}$. Dividiendo ambos lados entre $|\Omega|$ tenemos que:

$$P(A \mid B) = \frac{\frac{|A \cap B|}{|\Omega|}}{\frac{|B|}{|\Omega|}} = \frac{P(A \cap B)}{P(B)}$$

Esto nos da la motivación para definir la probabilidad condicional de A dada la ocurrencia de B. para cualquier A y para cualquier B tal que $P(B) > 0$

Definición 1.7 Sean A y B dos eventos, $P(B) > 0$. Se define la probabilidad condicional de A dada la ocurrencia de B como $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$.

Independencia de eventos

Algunas veces en un experimento aleatorio, se tiene la característica de que la ocurrencia de un evento, no afecta la ocurrencia de otro. Por ejemplo, si el experimento aleatorio es lanzar dos veces una moneda. El resultado del segundo lanzamiento es independiente del resultado en el primer lanzamiento.

Deseamos reflejar esta característica de independencia al modelo matemático del experimento. Es decir queremos definir probabilísticamente el concepto de independencia de eventos.

Si A y B son dos eventos dados y ocurre que $P(A \mid B) = P(A)$ parece razonable decir que "la ocurrencia del evento B no afecta a la ocurrencia del evento A", al menos desde el punto de vista de la Probabilidad.

Asi que

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

parece ser una buena definición de independencia de eventos , sin embargo necesariamente se tendría que pedir que $P(B) > 0$

La siguiente definición de independencia se obtiene de la segunda parte de la igualdad anterior, y nos da la posibilidad de definir independencia de eventos en general sin la restricción de que B sea de probabilidad positiva.

Definición 1.8 Sean A y B eventos . Se dice que A y B son independientes si

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

Regla de la probabilidad total

La regla de la probabilidad total es un resultado práctico y útil en la Teoría de la Probabilidad.

Nos permite calcular la probabilidad de un evento B en partes, o dicho de otra forma calculándola como una suma de sus partes.

Proposición 1.9 (Regla de la probabilidad total) Si tenemos una colección de eventos $B_1, B_2, .., B_N$ todos de probabilidad positiva y mutuamente excluyentes tales que $\cup_{i=1}^{N} B_i = \Omega$. Y si B es cualquier evento entonces

$$P(B) = P(B \mid B_1)P(B_1) + P(B \mid B_2)P(B_2) + \ldots + P(B \mid B_N)P(B_N)$$

Demostración. Debido a que $B_1, B_2, .., B_N$ son mutuamente excluyentes, tenemos que los eventos $B \cap B_1, B \cap B_2, .., B \cap B_N$ son mutuamente excluyentes.

Por otro lado tenemos que

$$B = B \cap \Omega = B \cap (\cup_{i=1}^{N} B_i) = (\cup_{i=1}^{N} B \cap B_i)$$

asi que:

$$P(B) = P(\cup_{i=1}^{N} B \cap B_i) = P(B \cap B_1) + P(B \cap B_2) + .. + P(B \cap B_N)$$

$$= P(B \mid B_1)P(B_1) + P(B \mid B_2)P(B_2) + \ldots + P(B \mid B_N)P(B_N)$$

■

Variables aleatorias y funciones de distribución

Una variable aleatoria es una característica numérica de nuestro experimento. Asi pues si los resultados de nuestro experimento son solamente "ganar" o "perder". Y nos interesa calcular el número de veces que se gana , podríamos asignarle al resultado " ganar" el número 1 y a "perder" el 0. Esa asignación del 1 y el 0 a los resultados del experimento es una variable aleatoria.

Algunos ejemplos de variables aleatorias son:

-La suma que se obtiene al lanzar dos dados.

-El número de partículas cósmicas que caen en México durante un día.

-El número de soles en 50 volados.

Nos interesa la probabilidad de que la variable aleatoria en cuestión tome un valor determinado, o un conjunto de valores.

Si X es una variable aleatoria, entonces toma un determinado valor real en un subconjunto, para cada posible resultado $\omega \in \Omega$, es decir X se puede representar mediante una función $X : \Omega \to \mathbb{R}$. Denotaremos por $[X \in B]$ al conjunto $\{\omega \in \Omega \mid X(\omega) \in B\}$, $B \subseteq \mathbb{R}$, es decir es la imagen inversa del conjunto B bajo la función X. De la misma manera, denotaremos por $[X \leq x]$ al conjunto $\{\omega \in \Omega \mid X(\omega) \leq x\}$

Definición 1.10 Se dice que una función $X : \Omega \to \mathbb{R}$ es una variable aleatoria si $[X \leq x] \in \mathcal{A}$ para todo $x \in \mathbb{R}$

Uno de los conceptos más importantes en la Teoría de la Probabilidad es el de **función de distribución,** el cual se define a continuación:

Definición 1.11 Sea X una variable aleatoria. La función de distribución de la variable aleatoria X, denotada por F_X, es la función $F_X : \mathbb{R} \to \mathbb{R}$ definida por

$$F_X(x) = P(X \leq x)$$

.

$F_X(x)$ nos da la información de qué tanta probabilidad se acumula hasta el valor x. Es por eso que también se le llama la función acumulativa de X. Dos variables aleatorias pueden ser diferentes, y sin embargo tener la misma función de distribución.

Clasificación de variables aleatorias y funciones de densidad

Las variables aleatorias se pueden clasificar en cuatro grupos; las discretas, las absolutamente continuas, las continuas que no son absolutamente continuas, y las que no están en ninguno de los grupos anteriores.

Variables aleatorias discretas

En palabras, las variables aleatorias discretas son aquellas que solo toman un número a lo más numerable de valores. La siguiente definición formaliza esto:

Definición 1.12 Se dice que una variable aleatoria X es discreta si existe un conjunto finito o infinito numerable de números reales C tal que $P(X = x) > 0$ para cualquier $x \in C$ y además

$$\sum_{x \in C} P(X = x) = 1$$

A C se le llama el conjunto de posibles valores de X

Dada X una variable aleatoria discreta, se define a la función de densidad de X de la siguiente forma:

Definición 1.13 Sea X una variable aleatoria discreta, la función de densidad de X en el punto x se define como.

$f_X(x) = P(X = x)$ para cualquier $x \in \mathbb{R}$

Es claro que $f_X(x)$ tiene la siguientes propiedades

$i)$ $f_X(x) \geq 0$ para cualquier $x \in \mathbb{R}$

$ii)$ $\sum_{x \in C} f_X(x) = 1$

A toda función que satisfaga estas dos propiedades se le llama función de densidad discreta.

Densidad Bernoulli

Supongamos que un experimento aleatorio admite unicamente dos posibles resultados. Uno que ocurre con probabilidad p, al cual llamaremos éxito y, por lo tanto, el otro, al cual llamaremos fracaso, con probabilidad $(1-p)$

$\Omega = \{E, F\}$ donde E representa al éxito y F al fracaso, en tal caso:

La σ -álgebra sobre Ω es $\mathcal{A} = \{\emptyset, E, F, \{E, F\}\}$

La función de probabilidad asociada a este experimento es:

$$P(\emptyset) = 0, P(E) = p, P(F) = (1-p), P(\{E, F\}) = 1$$

Una posible variable aleatoria asociada al experimento es:

$$X(\omega) = \begin{cases} 1 & \omega = E \\ 0 & \omega = F \end{cases}$$

La función de densidad de esta variable aleatoria es

$$f_X(x) = \begin{cases} p & x = 1 \\ 1-p & x = 0 \\ 0 & e.o.c. \end{cases} = \begin{cases} p^x(1-p)^{1-x} & x = 0,1 \\ 0 & e.o.c. \end{cases}$$

Al experimento anterior se le conoce con el nombre de **ensayo de Bernoulli** y se dice que la variable aleatoria X tiene una función de densidad Bernoulli o bien que se distribuye como una Bernoulli.

Densidad Binomial

Considerense n ensayos de Bernoulli con probabilidad de éxito p, donde el resultado de cada uno de ellos es independiente del resultado de los demás. Si X es el número de éxitos en esos n ensayos entonces

$$f_X(x) = \begin{cases} \binom{n}{x} p^x(1-p)^{n-x} & x \in \{0,1,2,\ldots,n\} \\ 0 & e.o.c. \end{cases}$$

Toda variable aleatoria cuya función de densidad es como la anterior es llamada binomial con parámetros n,p,
y se dice que tiene densidad binomial o que se distribuye binomial con los mismos parámetros.

Densidad geométrica

Considérense n ensayos de Bernoulli con probabilidad de éxito p, donde el resultado de cada uno de ellos es independiente del resultado de los demás. Si X es el número de fracasos antes del primer éxito, entonces la función de densidad de X es

$$f_X(x) = \begin{cases} p(1-p)^x & x = 0,1,\ldots \\ 0 & e.o.c. \end{cases}$$

Toda variable aleatoria cuya función de densidad es como la anterior es llamada geométrica con parámetro p,
y se dice que tiene densidad geométrica con el mismo parámetro.

Densidad uniforme discreta

A toda variable cuya densidad está dada como sigue:

$$f_X(x) = \begin{cases} \frac{1}{n} & x \in \{x_1,x_2,\ldots,x_n\} \\ 0 & e.o.c. \end{cases}$$

donde $x_1,x_2,\ldots,x_n$ son números reales fijos, se dice que tiene una distribución uniforme. Esta función de densidad es un modelo para cuando el experimento aleatorio consiste de elegir al azar un elemento del conjunto $\{x_1,x_2,\ldots,x_n\}$.

Densidad Poisson

A toda variable cuya densidad está dada como sigue:

$$f_X(x) = \begin{cases} \frac{\lambda^x}{x!}e^{-\lambda} & x \in \{0,1,2,\dots\} \\ 0 & e.o.c. \end{cases}$$

, en donde $\lambda > 0$, se dice que tiene una distribución Poisson con parámetro λ.

La gráfica de unafunción de distribución de una variable aleatoria discreta, tiene un número al menos finito de discontinuidades de salto. De hecho estas discontinuidades las tiene en el conjunto de valores de X.

Variables aleatorias continuas y absolutamente continuas

Las variables aleatorias continuas son aquellas cuya función de distribución no tiene saltos, es decir que su función de distribución es continua y por lo tanto no hay un solo punto x tal que $P(X = x) > 0$.

Definición 1.14. Se dice que una variable aleatoria X es continua si su función de distribución es continua.

La siguiente proposición se puede usar también como definición de variable aleatoria continua.

Proposición 1.15. Si X es una variable aleatoria continua entonces

$$P(X = x) = 0 \qquad \text{para toda } x \in \mathbb{R}$$

Demostración: Sea X una variable aleatoria continua. Supongamos que existe $x \in \mathbb{R}$ tal que $P(X = x) > 0$, es decir

$$P(X \leq x) - P(X < x) > 0$$

entonces

$$P(X \leq x) > P(X < x)$$

es decir

$$F_X(x) > \lim_{n\to\infty} P(X \leq x_n) \quad \text{para toda } \{x_n\} \text{ sucesión monótona creciente}$$

De lo anterior se concluye que $F_X(x)$ es discontinua en x. Esto es una contradicción con la hipótesis y por lo tanto $P(X = x) = 0$ para toda $x \in \mathbb{R}$

Hay un tipo muy especial de **variables aleatorias** continuas que son llamadas **absolutamente continuas,** y son las que se puede escribir su función de distribución como:

$$F_X(x) = \int_{-\infty}^{x} f(y)dy \text{ para alguna } f \text{ integrable}$$

Es claro que si existe una f que cumpla con lo anterior, entonces existe toda una clase de ellas pues modificando f en un número finito de puntos sigue teniendo la misma propiedad. A cualquier elemento de esa clase de funciones se la llama **función de densidad de** X.

Si $f_X(x)$ es una función de densidad de una variable aleatoria absolutamente continua X, entonces

i) $f_X(x) \geq 0$
ii) $\int_{-\infty}^{\infty} f_X(u)du = 1$

Como ejemplo de distribuciones absolutamente tenemos todos los siguientes:

Densidad Uniforme continua

A toda variable aleatoria cuya función de densidad esté dada por

$$f_X(x) = \begin{cases} \frac{1}{b-a} & x \in (a,b) \\ 0 & e.o.c. \end{cases}$$

se dice que tiene una distribución uniforme continua. Esta variable nos modela elecciones al azar en el intervalo (a,b).

Una ditribución que generaliza la distribución uniforme continua en $(0,1)$ es la distribución beta con parámetros α,β.

Distribución beta

Dada una variable aleatoria X tiene distribución beta con parámetros α,β si su función de densidad está dada por

$$f_X(x) = \begin{cases} \frac{1}{B(\beta,\alpha)}x^{\alpha-1}(1-x)^{\beta-1} & x \in (0,1) \\ 0 & e.o.c. \end{cases}$$

donde $B(\beta,\alpha) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx$

Evidentemente el dividir entre $B(\beta,\alpha)$ es con el fin de que $f_X(x)$ cumpla con la propiedad ii)

Distribución gama y distribución exponencial.

Si α y λ son números reales positivos , se dice que una variable aleatoria X tiene distribución gamma de parámetros α y λ, si su función de densidad está dada por

$$f_X(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x} & x > 0 \\ 0 & e.o.c. \end{cases}$$

donde $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx$

La función Γ es una función que para n entero positivo cumple que $\Gamma(n) = (n-1)!$.En el caso de que α sea un entero positivo, la distribución gama con parámetros α y λ, modela el tiempo que transcurre hasta que ocurren n eventos. Cuando X tiene una distribución gama con parametros $1,\lambda$, se dice que X tiene una distribución exponencial con parámetro λ.

Distribución normal

Dada una variable aleatoria X tiene una distribución normal de parámetros μ y σ^2 si su función de densidad es

$$f_X(x) = \frac{1}{\sqrt[2]{2\pi}\,\sigma}e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

para cualquier $x \in \mathbb{R}$

Un caso particular de la distribución normal, es la distribución **normal estándar** que es una distribución normal con parámetros $(0,1)$

La distribución normal se obtiene como el límite de una distribución binomial con parámetros n y p cuando n tiende a ∞. Sin embargo no sólo es una buena aproximación para la distribución binomial con parámetros n y p, haciendo $\mu = np$ y $\sigma^2 = np(1-p)$, sino tambien lo es para otras como la Poisson de parámetro λ. Es por algunas de estas razones que es tan importante.

Distribuciones conjuntas

Dada una familia de n variables aleatorias $X_1, X_2, \ldots, X_n$, aún cuando tengamos las funciones de distribución de cada una de éstas, $F_{X_1}, F_{X_2}, \ldots, F_{X_n}$, no sabemos mucho de la relación que hay entre las variables aleatorias, sin embargo una función que sí nos da la relación entre ellas es la **función de distribución** conjunta de $X_1, X_2, \ldots, X_n$, que se define a continuación:

Definición 1.16. Sean , n variables aleatorias. La función $F_{X_1, X_2, \ldots, X_n} : \mathbb{R}^n \to [0,1]$ definida por $F_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n)$ es llamada la función de distribución conjunta de $X_1, X_2, \ldots, X_n$.

La función de distribución conjunta de n variables aleatorias siempre existe. Si tenemos la función de distribución de n variables aleatorias podemos tener la función de distribución de cada una de ellas. Este resultado se formaliza a continuación, sólo para el caso de dos variables aleatorias.

Proposición 1.17. Sean X y Y dos variables aleatorias y sea $F_{X,Y}$ su función de distribución conjunta, entonces

$$\lim_{x \to \infty} F_{X,Y}(x,y) = F_Y(y)$$

Demostración: Sean $y \in \mathbb{R}$ y $\{x_n\}$ una sucesión monótona creciente de números reales tales que $\lim_{x \to \infty} x_n = \infty$, entonces la sucesión de eventos $[X \leq x_n, Y \leq y]$ es monótona no decreciente y

$$[Y \leq y] = \cup_{n=1}^{\infty} [X \leq x_n, Y \leq y]$$

por lo tanto:

$$\begin{aligned}
F_Y(y) &= P(Y \leq y) = P(\cup_{n=1}^{\infty} [X \leq x_n, Y \leq y]) \\
&= P(\cup_{n=1}^{\infty} [x_{n-1} < X \leq x_n, Y \leq y]) \text{ donde } x_0 = -\infty \\
&= \sum_{n=1}^{\infty} P(x_{n-1} < X \leq x_n, Y \leq y) \\
&= \lim_{l \to \infty} \sum_{n=1}^{l} P(x_{n-1} < X \leq x_n, Y \leq y) \\
&= \lim_{l \to \infty} P(\cup_{n=1}^{l} [x_{n-1} < X \leq x_n, Y \leq y]) \\
&= \lim_{n \to \infty} P(X \leq x_n, Y \leq y) \\
&= \lim_{n \to \infty} F_{X,Y}(x_n, y) \qquad \blacksquare
\end{aligned}$$

Densidades conjuntas

Cuando se tiene una familia de n variables aleatorias $X_1,X_2,...,X_n$, la distribución conjunta es determinada por una función de densidad en el caso de que $X_1,X_2,...,X_n$ sean todas ellas discretas o bien absolutamente continuas. Pues la función de densidad conjunta de n variables aleatorias es similar a la de una variable aleatoria.

Definición 1.18. Si se tienen $X_1,X_2,...,X_n$ variables aleatorias discretas, se define a la función de densidad de $X_1,X_2,...,X_n$ como $f_{X_1,X_2,...,X_n}(x_1,x_2,...,x_n) = P(X_1 = x_1, X_2 = x_2,...,X_n = x_n)$

Definición 1.19. Se dice que la función de distribución conjunta, $F_{X_1,X_2,...,X_n}$, de las variables aleatorias $X_1,X_2,...,X_n$ es absolutamente continua si existe una función $f_{X_1,X_2,...,X_n} : \mathbb{R}^n \to \mathbb{R}$ integrable tal que

$$F_{X_1,X_2,...,X_n}(x_1,x_2,...,x_n) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f_{X_1,X_2,...,X_n}(y_1,y_2,...,y_n)dy_1\cdots dy_n$$

para cualquier $(x_1,x_2,...,x_n) \in \mathbb{R}^n$

Dado $(x_1,x_2,...,x_n) \in \mathbb{R}^n$, definamos $A = \{(y_1,y_2,...,y_n) \mid y_1 \le x_1, y_2 \le x_2, ..., y_n \le x_n\}$. La propiedad que caracteriza a una función de densidad conjunta de $X_1,X_2,...,X_n$ se puede escribir entonces de la siguiente manera:

$$P[(X_1,X_2,...,X_n) \in A] = \int \cdots \int_A f_{X_1,X_2,...,X_n}(y_1,y_2,...,y_n)dy_1\cdots dy_n$$

Se puede demostrar que esta misma relación se cumple para cualquier subconjunto $A \subset \mathbb{R}^n$ para el cual la integral $\int\cdots\int_A f_{X_1,X_2,...,X_n}(y_1,y_2,...,y_n)dy_1\cdots dy_n$ esté bien definida.

Independencia de variables aleatorias

Definición 1.20. Se dice que n variables aleatorias $X_1,X_2,...,X_n$ son independientes si para cualquier $(x_1,x_2,...,x_n) \in \mathbb{R}^n$ se tiene que

$$F_{X_1,X_2,...,X_n}(x_1,x_2,...,x_n) = F_{X_1}(x_1)F_{X_2}(x_2)\cdots F_{X_n}(x_n)$$

En el caso de que las variables aleatorias $X_1,X_2,...,X_n$ sean todas ellas absolutamente continuas, o bien todas ellas discretas la condicion de independencia se puede establecer en terminos de sus funciones de densidad

$$f_{X_1,X_2,...,X_n}(x_1,x_2,...,x_n) = f_{X_1}(x_1)f_{X_2}(x_2)\cdots f_{X_n}(x_n) \text{ para cualquier } (x_1,x_2,...,x_n) \in \mathbb{R}^n$$

Un resultado muy útil de la independencia de variables aleatorias es que ésta se "hereda" a funciones de variables aleatorias independiente.

Proposición 1.21. Sean $X_1,X_2,...,X_n$ n variables aleatorias independientes, sean $g_1,g_2,...g_n$

n funciones tales que $g_1(X_1), g_2(X_2), \ldots, g_n(X_n)$ sean variables aleatorias, entonces $g_1(X_1), g_2(X_2), \ldots, g_n(X_n)$ son independientes.

Esperanza de una variable aleatoria

La esperanza de una variable aleatoria es uno de los conceptos mas importantes en la probabilidad, este concepto surge al igual que el de probabilidad a mediados del siglo XVII, cuando se inicia la probabilidad como disciplina matemática .

La idea de la esperanza nace de tratar de resolver problemas como el siguiente que dan lugar a preguntas como ¿Cuánto debe pagar una persona por participar en un juego para que este sea justo?.

Ejemplo 1.22. Supongamos que un juego consiste en lanzar dos dados, y que 36 personas participan en el, entre ellas se encuentra Christian Huygens. Cada persona le apuesta a un posible resultado una cantidad fija de antemano, x , y la que gane se queda con los $36x$. Evidentemente este es un juego justo pues todos los participantes tienen la misma probabilidad de ganar los $36x$ pesos. Supongamos que Huygens pacta con cada un a de las otras 35 personas, de que si alguno de los dos gana el juego, este le dará al otro una cantidad. Dicho de otra forma si la persona i gana el juego esta le dara a Huygens x_i pesos , mientras que si Huygens gana hará lo propio.

$x_i, i \in \{1, \ldots 36\}$ es lo que gana Huygens si gana la persona i, considerando que el es la persona 36.

Esta forma de jugar de Huygens hace que: gane quien gane él obtenga algo. De hecho si lo que el obtiene es x_{36} se tiene que:

$$x_{36} = 36x - x_1 - x_2 - \ldots - x_{35}$$

es decir

$$x = \frac{x_1 + x_2 + \ldots + x_{35} + x_{36}}{36}$$

Es decir: lo que cada jugador debe poner para que el juego sea justo es $\frac{x_1 - x_2 - \ldots - x_{35} - x_{36}}{36}$.

La esperanza de las variables aleatorias discretas se puede ver como un promedio de los posibles valores de la variable, ponderado con la probabilidad de cada valor

Definición 1.23. Sea X una variable aleatoria discreta . Se dice que X tiene esperanza finita si la serie $\sum_x |x| f_X(x) < \infty$ y en ese caso se define a la esperanza de X como

$$E(X) = \sum_x x f_X(x)$$

El caso de la esperanza de una variable absolutamente continua, es una generalización del caso discreto.

Definición 1.24. Sea X una variable aleatoria absolutamente continua. Se dice que X tiene esperanza finita si la integral $\int_{-\infty}^{\infty} |x| f_X(x) dx < \infty$ y en tal caso se define a la esperanza de X como:

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx$$

Esperanza de funciones de una variable aleatoria.

Si tenemos una variable aleatoria X y queremos calcular la esperanza de $g(X)$ donde $g : \mathbb{R} \to \mathbb{R}$ es claro que tendriamos que pedir por principio de cuentas que $g(X)$ sea una variable aleatoria. Sin embargo tenemos un problema; y es que solo hemos definido esperanzas en el caso discreto y en el caso absolutamente continuo, y estas definiciones se han basado en la existencia de una función de densidad, y en general no sucede que una función de una variable aleatoria absolutamente continua sea o discreta o absolutamente continua.

Los siguientes resultados van encaminados a poder definir la esperanza de cualquier variable aleatoria.

La siguiente proposición trata de establecer condiciones para la existencia de la esperanza de $g(X)$ en terminos de su función de distribución, cuando X es una variable aleatoria discreta, para después tratar de generalizar.

En el caso de que X sea una variable aleatoria discreta, $g(X)$ también lo es y para que exista la esperanza de $g(X)$ tendria que suceder que $\sum_{g(x)} |g(x)| f_{g(X)}[g(x)] < \infty$ pero

$$\sum_{g(x)} |g(x)| f_{g(X)}[g(x)] = \sum_{\{x \in \mathbb{R} : g(x) > 0\}} g(x) f_X(x) + \sum_{\{x \in \mathbb{R} : g(x) < 0\}} g(x) f_X(x)$$

Proposición 1.25. Sean X una variable aleatoria discreta con función de densidad f_X y sea g cualquier función tal que $g(X)$ es una variable aleatoria, entonces:

$$a) \quad \sum_{\{x \in \mathbb{R} : g(x) > 0\}} g(x) f_X(x) = \int_0^{\infty} [1 - F_{g(X)}(z)] dz$$

$$b) \quad \sum_{\{x \in \mathbb{R} : g(x) < 0\}} g(x) f_X(x) = -\int_0^{\infty} F_{g(X)}(-z) dz$$

Demostración:

a) El conjunto $\{x \in \mathbb{R} : g(x) > 0\}$ y tales que $f_X(x) > 0$ es a lo más numerable.La demostración se hará en el caso finito. Para el caso infinito numerable se requieren argumentos que utilizan herramientas que sobrepasan el nivel de este trabajo. Sea $\{x_1, x_2, x_3, \ldots x_n\}$ el conjunto de valores de X

Se tiene que: $0 = g(x_0) < g(x_1) < g(x_2) < g(x_3) < \ldots < g(x_n)$

Se tiene que

$$\sum_{\{x:g(x)>0\}} g(x)f_X(x) = \sum_{\{x_1,x_2,\dots\}} g(x)f_X(x)$$
$$= g(x_1)f_X(x_1) + g(x_2)f_X(x_2) + g(x_3)f_X(x_3) + \dots + g(x_n)f_X(x_n)$$
$$= g(x_1)[F_{g(X)}(g(x_1)) - F_{g(X)}(0)] + g(x_2)[F_{g(X)}(g(x_2)) - F_{g(X)}(g(x_1))]$$
$$+ g(x_3)[F_{g(X)}(g(x_3)) - F_{g(X)}(g(x_2))] + \dots + g(x_n)[F_{g(X)}(g(x_n)) - F_{g(X)}(g(x_n))]$$
$$= g(x_1)[F_{g(X)}(g(x_1)) - 1 + 1 - F_{g(X)}(0)] + g(x_2)[F_{g(X)}(g(x_2)) - 1 + 1 - F_{g(X)}(g(x_1))]$$
$$+ \dots + g(x_n)[F_{g(X)}(g(x_n)) - 1 + 1 - F_{g(X)}(g(x_{n-1}))]$$
$$= g(x_1)[1 - F_{g(X)}(0)] + [g(x_2) - g(x_1)][1 - F_{g(X)}(g(x_1))]$$
$$+ [g(x_3) - g(x_2)][1 - F_{g(X)}(g(x_2))] + \dots + [g(x_n) - g(x_{n-1})][1 - F_{g(X)}(g(x_{n-1}))]$$
$$= \int_0^{g(x_1)}[1 - F_{g(X)}(z)]dz + \int_{g(x_1)}^{g(x_2)}[1 - F_{g(X)}(z)]dz + \int_{g(x_2)}^{g(x_3)}[1 - F_{g(X)}(z)]dz$$
$$+ \dots + \int_{g(x_{n-1})}^{g(x_n)}[1 - F_{g(X)}(z)]dz$$
$$= \int_0^{\infty}[1 - F_{g(X)}(z)]dz \qquad \blacksquare$$

La demostración de b) se hace de la misma manera.

Como consecuencia del resultado anterior se tiene la siguiente proposición.

Proposición 1.26. Sea X es una variable aleatoria discreta. X tiene esperanza finita si y solo si $\int_0^{\infty}[1 - F_X(z)]dz < \infty$ y $\int_0^{\infty} F_X(-z)dz < \infty$ y en ese caso se tiene que

$$E(X) = \int_0^{\infty}[1 - F_X(z)]dz + \int_0^{\infty} F_X(-z)dz$$

Este resultado nos sugiere una forma en que se puede definir la esperanza de una variable aleatoria, que no sea ni discreta ni absolutamente continua. Pues el cálculo de la esperanza en el resultado anterior solo depende de la función de distribución de la variable en cuestión.

La siguiente proposición es una generalización al caso continuo de la proposición que permitirá definir la esperanza de una variable aleatoria en general.

En el caso de que X sea una variable aleatoria absolutamente continua, $g(X)$ no necesariamente lo es. Sin embargo

$$\int_{-\infty}^{\infty}|g(x)|f_X(x)dx = \int_{\{x\in\mathfrak{R}:g(x)>0\}} g(x)f_X(x)dx - \int_{\{x\in\mathfrak{R}:g(x)<0\}} g(x)f_X(x)dx$$

Proposición 1.27. Sea X una variable aleatoria absolutamente continua, con función de densidad f_x. Sea g cualquier función boreliana Riemman integrable en todo intervalo finito, entonces

a) $\int_{\{x\in\mathfrak{R}:g(x)>0\}} g(x)f_X(x)dx = \int_0^{\infty}[1 - F_{g(X)}(z)]dz$

b) $\int_{\{x\in\mathfrak{R}:g(x)<0\}} g(x)f_X(x)dx = -\int_0^{\infty} F_{g(X)}(-z)dz$

Demostración:

a)Por ser X absolutamente continua

$$
\begin{aligned}
\int_0^{\infty}[1-F_{g(X)}(z)]dz &= \int_0^{\infty} P(g(X)>z)dz \\
&= \int_0^{\infty}\int_{\{x\in\Re:g(x)>z\}} f_X(x)dxdz \\
&= \int_0^{\infty}\int_{-\infty}^{\infty} I_{\{u\in\Re:g(u)>z\}}(x)f_X(x)dxdz \\
&= \int_0^{\infty}\int_{-\infty}^{\infty} I_{\{(u,w)\in\Re^2:g(u)>w>0\}}(x,z)f_X(x)dxdz \\
&= \int_0^{\infty}\int_{-\infty}^{\infty} I_{\{(u,w)\in\Re^2:g(u)>w>0\}}(x,z)f_X(x)dzdx \\
&= \int_{-\infty}^{\infty}\int_0^{\infty} I_{\{u\in\Re:g(u)>0\}}(x)I_{(0,g(x))}(z)f_X(x)dxdz \\
&= \int_{\{u\in\Re:g(u)>0\}}\int_0^{g(x)} f_X(x)dxdz = \int_{\{x\in\Re:g(x)>0\}} g(x)f_X(x)dx
\end{aligned}
$$

La demostración de b) recoge la misma idea que la de a). ■

Como consecuencia del resultado anterior se tiene la siguiente proposición.

Proposición 1.28. Sea X es una variable aleatoria absolutamente continua. X tiene esperanza finita si y solo si $\int_0^{\infty}[1-F_{g(X)}(z)]dz < \infty$ y $\int_0^{\infty} F_{g(X)}(-z)dz < \infty$ y en ese caso se tiene que

$$E(X) = \int_0^{\infty}[1-F_X(z)]dz + \int_0^{\infty} F_X(-z)dz$$

Las proposiciones anteriores nos permiten dar ahora una definición de esperanza que se pueda aplicar a cualquier variable aleatoria.

Definición general de esperanza

A continuación se da la definición general de esperanza.

Definición 1.29. Sea X cualquier variable aleatoria con función de distribución F_X, se dice que X tiene esperanza finita si $\int_0^{\infty}[1-F_X(z)]dz < \infty$ y $\int_0^{\infty} F_X(-z)dz < \infty$ y en ese caso se define a la esperanza de X como $E(X) = \int_0^{\infty}[1-F_X(z)]dz - \int_0^{\infty} F_X(-z)dz$

Los siguientes resultados son muy utiles.

Proposición 1.30. Sea X es una variable aleatoria discreta entonces $g(X)$ tiene esperanza finita si $\sum_x |g(x)|f_X(x) < \infty$

y en este caso se tiene que

$$E(g(X)) = \sum_x g(x)f_x(x)$$

Demostración.

$$\sum_x |g(x)|f_X(x) = \sum_{\{x\in\Re:g(x)>0\}} g(x)f_X(x) + \sum_{\{x\in\Re:g(x)<0\}} g(x)f_X(x) = \int_0^{\infty}[1-F_{g(X)}(z)]dz + \int_0^{\infty} F_{g(X)}(-z)dz < \infty$$

Por otro lado

$$
\begin{aligned}
E(g(X)) &= \int_0^\infty [1 - F_{g(X)}(z)]dz - \int_0^\infty F_{g(X)}(-z)dz \\
&= \sum_{\{x\in\Re:g(x)>0\}} g(x)f_X(x) - \left[- \sum_{\{x\in\Re:g(x)<0\}} g(x)f_X(x) \right] \\
&= \sum_x g(x)f_x(x) \quad \blacksquare
\end{aligned}
$$

Proposición 1.31. Sea X es una variable aleatoria absolutamente continua. $g(X)$ tiene esperanza finita si

$\int_{-\infty}^{\infty} |g(x)| f_X(x)dx < \infty$

y en este caso se tiene que

$$E(g(X)) = \int_{-\infty}^{\infty} g(x)f_x(x)dx$$

Demostración:

$$
\begin{aligned}
\int_{-\infty}^{\infty} |g(x)| f_X(x)dx &= \int_{\{x\in\Re:g(x)>0\}} g(x)f_X(x)dx + \int_{\{x\in\Re:g(x)<0\}} g(x)f_X(x)dx \\
&= \int_0^\infty [1 - F_{g(X)}(z)]dz + \int_0^\infty F_{g(X)}(-z)dz < \infty
\end{aligned}
$$

$$
\begin{aligned}
E(g(X)) &= \int_0^\infty [1 - F_{g(X)}(z)]dz - \int_0^\infty F_{g(X)}(-z)dz \\
&= \int_{\{x\in\Re:g(x)>0\}} g(x)f_X(x)dx - \left[- \int_{\{x\in\Re:g(x)<0\}} g(x)f_X(x)dx \right] = \int_{-\infty}^{\infty} g(x)f_x(x)dx \quad \blacksquare
\end{aligned}
$$

Propiedades de la esperanza de una variable aleatoria

Las siguientes propiedades de la esperanza de una variable aleatoria se usarán en los capítulos siguientes.

Proposición 1.32. Sean X y Y variables aleatorias y sea c cualquier constante entonces se tiene que

1) Si X tiene esperanza finita, entonces $E(cX) = cE(X)$
2) $E(c) = c$
3) Si X y Y tienen esperanza finita entonces $E(X + Y) = E(X) + E(Y)$
4) Si X y Y tienen esperanza finita y $P(X \leq Y) = 1$ entonces $E(X) \leq E(Y)$

Esperanza de una función de n variables aleatorias

Los resultados de la sección anterior se pueden generalizar, para dar lugar a dos resultados muy usados, que nos permiten calcular la esperanza de una función de n variables aleatorias absolutamente continuas todas ellas, o bien todas ellas discretas, no importando si esta nueva variable aleatoria es del tipo absolutamente continuo o discreto. No se hará aquí la generalización, sólo se escribiran los resultados.

Proposición 1.33. Sean $X_1, X_2, \ldots X_n$ variables aleatorias discretas con función de densidad conjunta $f_{X_1, X_2, \ldots X_n}$. y $g : \mathbb{R}^n \to \mathbb{R}$ cualquier función tal que $g(X_1, X_2, \ldots X_n)$ sea una variable aleatoria, entonces $g(X_1, X_2, \ldots X_n)$ tiene esperanza finita si y solo si

$$\sum_{x_1, x_2, \ldots, x_n} |g(x_1, x_2, \ldots, x_n)| f_{X_1, X_2, \ldots X_n}(x_1, x_2, \ldots, x_n) < \infty$$

y en ese caso se tiene que :

$$E[g(X)] = \sum_{x_1, x_2, \ldots, x_n} g(x_1, x_2, \ldots, x_n) f_{X_1, X_2, \ldots X_n}(x_1, x_2, \ldots, x_n)$$

Proposición 1.34. Sean $X_1, X_2, \ldots X_n$ variables aleatorias absolutamente continuas, con función de densidad conjunta $f_{X_1, X_2, \ldots X_n}$. Sea $g : \mathbb{R}^n \to \mathbb{R}$ cualquier función tal que $g(X_1, X_2, \ldots X_n)$ sea una variable aleatoria, entonces $g(X_1, X_2, \ldots X_n)$ tiene esperanza finita si y sólo si

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} |g(x_1, x_2, \ldots, x_n)| f_{X_1, X_2, \ldots X_n}(x_1, x_2, \ldots x_n) dx_1 dx_2 \cdots dx_n < \infty$$

y en ese caso se tiene que :

$$E[g(X)] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, x_2, \ldots, x_n) f_{X_1, X_2, \ldots X_n}(x_1, x_2, \ldots x_n) dx_1 dx_2 \cdots dx_n$$

Capítulo II

Distribuciones y esperanzas condicionales en el caso discreto

Distribuciones condicionales en el caso discreto.

Sean X y Y dos variables aleatorias discretas. La probabilidad de que $X = x$ dado que $Y = y$ puede calcularse de acuerdo con la definición de probabilidad condicional de un evento A dada la ocurrencia de un evento B de probabilidad positiva.

Tendríamos

$$P(X = x \mid Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}$$
$$= \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

siempre que $f_Y(y) > 0$

aquí se considera que el evento A es el evento $X = x$. i.e. A es el conjunto de todos los eventos elementales ω_i tales que $X(\omega_i) = x$. De la misma forma $B = \{\omega_i \mid Y(\omega_i) = y\}$

Resulta entonces que se puede definir una nueva función. la función de densidad condicional de X dado que Y=y con $y \in \mathbb{R}$ tal que $f_Y(y) > 0$ de la siguiente forma:

$$f_{X|Y}(x \mid y) = \begin{cases} \frac{f_{X,Y}(x,y)}{f_Y(y)} & f_Y(y) > 0 \end{cases}$$

Cabe notar que estamos viendo a esta función como una función de x.

Para verificar que $f_{X|Y}(x \mid y)$ es una función de densidad vista como función de x, basta con ver que:

$$i)\ f_{X|Y}(x,y) = \frac{P(X = x, Y = y)}{P(Y = y)} \geq 0$$

$$ii) \sum_x f_{X|Y}(x \mid y) = \sum_x \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{P(\Omega, Y = y)}{P(Y = y)} = \frac{P(Y = y)}{P(Y = y)} = 1$$

Por lo tanto $f_{X|Y}(x \mid y)$ es una función de densidad y como tal, cumple todas las propiedades de estas.

Debido a que $f_{X|Y}(x \mid y)$ es una densidad, se tiene una función de distribución de la siguiente manera:

$$F_{X|Y}(x \mid y) = P(X \leq x \mid Y = y) = \sum_{u \leq x} P(u \leq x \mid Y = y).$$

Proposición 2.1. Si X y Y son variables aleatorias independientes se tiene que

$$a) F_{X|Y}(x,y) = F_X(x)$$

y que

$$b) f_{X|Y}(x,y) = f_X(x).$$

Es decir que la ocurrencia del evento $Y = y$ no afecta a la distribución de X.

Demostración a)

$$F_{X|Y}(x \mid y) = \frac{P(X \leq x, Y = y)}{P(Y = y)} = \frac{P(X \leq x)P(Y = y)}{P(Y = y)} = P(X \leq x) = F_X(x)$$

La prueba de b) es similar a la de a) ■

Ejemplo 2.2. Si X y Y son variables aleatorias independientes ambas con distribución Poisson de parámetros λ_1 y λ_2 respectivamente, calcular e identificar la función de densidad condicional de X dado que $X + Y = n$.

Se requiere calcular $f_{X|X+Y}(x \mid n)$ es decir

$$P(X = x \mid X + Y = n) = \frac{P(X = x, X + Y = n)}{P(X + Y = n)}$$

Calculando el numerador de esta expresión y usando la hipótesis de independencia de X y Y tenemos que:

$$\begin{aligned} P(X = x, X + Y = n) &= P(X = x, Y = n - x) \\ &= P(X = x)P(Y = n - x) \\ &= \frac{\lambda_1^x}{x!}e^{-\lambda_1}\frac{\lambda_2^{n-x}}{(n-x)!}e^{-\lambda_2} \end{aligned}$$

Por otro lado el denominador lo obtenemos de la siguiente manera utilizando la independencia de Y y X.

$$\begin{aligned} P(X + Y = n) &= P(Y = n - X) \\ &= P(Y = n - X, \Omega) = P(Y = n - X, x \in \{0,1,2,\ldots\}) \\ &= P(Y = n; x = 0) + P(Y = n - 1, x = 1) + P(Y = n - 2, x = 2) + \ldots \\ &= \sum_{x=0}^{\infty} P(Y = n - x, X = x) = \sum_{x=0}^{\infty} P(Y = n - x)P(X = x) \\ &= \sum_{x=0}^{n} \frac{\lambda_1^x}{x!}e^{-\lambda_1}\frac{\lambda_2^{n-x}}{(n-x)!}e^{-\lambda_2} = e^{-(\lambda_1 - \lambda_2)}\sum_{x=0}^{n}\frac{\lambda_2^{(n-x)}}{(n-x)!}\frac{\lambda_1^x}{x!} \\ &= \frac{e^{-(\lambda_1 - \lambda_2)}}{n!}\sum_{x=0}^{n}\frac{\lambda_2^{(n-x)}n!}{(n-x)!}\frac{\lambda_1^x}{x!} = \frac{e^{-(\lambda_1 - \lambda_2)}}{n!}(\lambda_1 + \lambda_2)^n \end{aligned}$$

Podiamos haber obtenido este resultado usando el hecho de que la suma de variables aleatorias Poisson independientes es también una variable aleatoria Poisson con parametro $\lambda_1 + \lambda_2$

De todo lo anterior se concluye que:

$$f_{X|X+Y}(x \mid n) = \begin{cases} \dfrac{\frac{\lambda_1^x}{x!}e^{-\lambda_1}\frac{\lambda_2^{n-x}}{(n-x)!}e^{-\lambda_2}}{\frac{e^{-(\lambda_1+\lambda_2)}}{n!}(\lambda_1+\lambda_2)^n} & x \in \{0,1,2\ldots,n\} \\ 0 & e.o.c. \end{cases}$$

$$= \begin{cases} \dfrac{n!\lambda_1^x\lambda_2^{n-x}}{x!(n-x)!(\lambda_1+\lambda_2)^n} & x \in \{0,1,\ldots,n\} \\ 0 & e.o.c. \end{cases}$$

$$= \begin{cases} \binom{n}{x}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^x\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-x} & x \in \{0,1,\ldots,n\} \\ 0 & e.o.c. \end{cases}$$

Es decir dado que $X+Y=n$ la distribución condicional de X es una distribución binomial con parámetros n y $\frac{\lambda_1}{\lambda_1+\lambda_2}$

Ejemplo 2.3.Calcular la función de densidad condicional de X dado que $X+Y=n$, si X y Y son variables aleatorias identicamente distribuídas e independientes, ambas geométricas con parámetro p.

Se requiere calcular $f_{X|X+Y}(x \mid n)$ es decir

$$P(X=x \mid X+Y=n) = \frac{P(X=x, X+Y=n)}{P(X+Y=n)},$$

Calculando el numerador de esta expresión y utilizando la hipótesis de independencia tenemos lo siguiente:

$$\begin{aligned} P(X=x, X+Y=n) &= P(X=x, Y=n-x) \\ &= P(X=x)P(Y=n-x) \\ &= p(1-p)^x p(1-p)^{n-x} = p^2(1-p)^n. \end{aligned}$$

Calculando el numerador y usando independencia tenemos que:

$$\begin{aligned} P(X+Y=n) &= P(Y=n-X) = P(Y=n-X,\Omega) = P(Y=n-X, x\in\{0,1,2,\ldots\}) \\ &= P(Y=n, x=0) + P(Y=n-1, x=1) + P(Y=n-2, x=2) + \ldots \\ &= \sum_{x=0}^{\infty} P(Y=n-x, X=x) \\ &= \sum_{x=0}^{\infty} P(Y=n-x)P(X=x) \\ &= \sum_{x=0}^{n} p(1-p)^{n-x}p(1-p)^x \\ &= \sum_{x=0}^{n} p^2(1-p)^n = (n+1)p^2(1-p)^n \end{aligned}$$

De todo lo anterior se concluye que:

$$f_{X|X+Y}(x \mid n) = \begin{cases} \frac{p^2(1-p)^n}{(n+1)p^2(1-p)^n} & x \in \{0,1,2\ldots,n\} \\ 0 & e.o.c. \end{cases}$$

$$= \begin{cases} \frac{1}{n+1} & x \in \{0,1,\ldots,n\} \\ 0 & e.o.c. \end{cases}$$

Es decir dado que $X+Y=n$ la distribución condicional de X es una distribución uniforme en el conjunto $\{0,1,\ldots,n\}$.

Ejemplo 2.4. Se escoge un número al azar del conjunto de números $\{1,2,3,4,5\}$, sea X dicho número. Después se elige otro número del conjunto $\{1,2,\ldots,X\}$ sea este segundo número Y. Encontrar la densidad condicional de X dado que $Y=y$ para $y \in \{1,2,3,4,5\}$ ¿Son X y Y independientes?

La función de densidad conjunta de X y Y queda como sigue:

$$f_{x,y}(x,y) = P(X=x, Y=y) = P(Y=y \mid X=x)P(X=x)$$

$$f_{x,y}(x,y) = \begin{cases} \frac{1}{5x} & y \leq x \\ 0 & y > x \end{cases}$$

Encontrando a $f_y(y)$ a través de la densidad conjunta de X y Y tenemos que:

$$f_y(y) = \sum_{x=y}^{x=5} f_{x,y}(x,y) = \sum_{x=y}^{x=5} \frac{1}{5x}$$

es decir

$$f_y(y) = \begin{cases} \frac{1}{5}\left(1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5}\right) & y=1 \\ \frac{1}{5}\left(\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5}\right) & y=2 \\ \frac{1}{5}\left(\frac{1}{3}+\frac{1}{4}+\frac{1}{5}\right) & y=3 \\ \frac{1}{5}\left(\frac{1}{4}+\frac{1}{5}\right) & y=4 \\ \frac{1}{5}\left(\frac{1}{5}\right) & y=5 \\ 0 & e.o.c \end{cases}$$

De lo cual se concluye que:

$$f_{X|X+Y}(x \mid 1) = \begin{cases} \frac{1}{x\left(1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5}\right)} & x \geq 1 \\ 0 & e.o.c. \end{cases}$$

$$f_{X|X-Y}(x \mid 2) = \begin{cases} \frac{1}{x(\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5})} & x \geq 2 \\ 0 & e.o.c. \end{cases}$$

$$f_{X|X-Y}(x \mid 3) = \begin{cases} \frac{1}{x(\frac{1}{3}+\frac{1}{4}+\frac{1}{5})} & x \in \{3,4,5\} \\ 0 & e.o.c. \end{cases}$$

$$f_{X|X-Y}(x \mid 4) = \begin{cases} \frac{1}{x(\frac{1}{4}+\frac{1}{5})} & x \in \{3,4\} \\ 0 & e.o.c. \end{cases}$$

$$f_{X|X-Y}(x \mid 5) = \begin{cases} \frac{1}{x(\frac{1}{5})} & x = 5 \\ 0 & e.o.c. \end{cases}$$

Lo anterior se puede resumir en el siguiente cuadrode $f_{X|X-Y}(x \mid n)$:

$x \setminus n$	1	2	3	4	5
1	.438	0	0	0	0
2	.219	.389	0	0	0
3	.146	.259	.426	0	0
4	.1095	.194	.319	.555	0
5	.0876	.1558	.255	.444	1

Es claro que X y Y no son independientes pues $f_{X|X-Y}(x \mid n) \neq f_x(x) = \frac{1}{5}$

Ejemplo 2.5. Se lanzan dos dados .Sean X y Y el valor mas grande y el mas pequeño de los que se obtuvieron. Calcular la densidad condicional de Y dado que $X = x$ para $x \in \{1,2,3,4,5,6\}$

La densidad conjunta de X y Y queda como sigue:

$$f_{X,Y}(x,y) = P(X = x, Y = y) = \begin{cases} \frac{1}{36} & x = y \\ \frac{2}{36} & x > y \\ 0 & \text{e.o.c} \end{cases}$$

Dicha función se obtiene por lo siguiente:

i) Cuando $x = y$ existe una sola pareja, dentro de las 36 posibles, para la cual los valores mayor y menor son iguales, a saber la pareja (x,x).

ii) En el caso $x > y$ fijos, existen dos parejas que cumplen que el valor mayor es x y el valor menor sea y, tales parejas son $(x,y),(y,x)$.

iii) Por último, no exísten parejas tales que el valor mayor sea x y el valor menor y satisfaciendo la condición $x < y$.

La función de densidad de X queda como sigue:

$$f_X(x) = P(X = x) = \begin{cases} \frac{2x-1}{36} & x \in \{1,2,3,4,5,6\} \\ 0 & e.o.c. \end{cases}$$

esto es debido que hay $2x - 1$ parejas donde x es el mayor de los números. Por ejemplo si $x = 5$ las parejas son :$\{(5,5),(5,4),(5,3),(5,2),(5,1),(4,5),(3,5),(2,5),(1,5)\}$

Ejemplo 2.6. Sean X y Y dos variables aleatorias con función de densidad conjunta dada por:

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{n^2(n+1)}(x+y) & x,y \in \{1,2,\ldots,n\} \\ 0 & e.o.c. \end{cases}$$

para y$\in$ $\{1,2,\ldots,n\}$ encuentre la función de densidad condicional de X dado que $Y = y$.
Calculando $f_Y(y)$ tenemos que:
$f_Y(y) = \sum_{x\in\mathbb{R}} f_{X,Y}(x,y)$ y para $y \in \{1,2,\ldots,n\}$

$$f_Y(y) = \sum_{x=1}^{n} \frac{(x+y)}{n^2(n+1)} = \frac{ny}{n^2(n+1)} + \frac{n(n+1)}{2n^2(n+1)} = \frac{1}{n}\left(\frac{n+1+2y}{2(n+1)}\right)$$

Por lo tanto tenemos que:

$$f_Y(y) = \begin{cases} \frac{1}{n}\left(\frac{n+1+2y}{2(n+1)}\right) & y \in \{1,2,\ldots,n\} \\ 0 & e.o.c. \end{cases}$$

De donde podemos calcular

$$f_{X|Y}(x \mid y) = \frac{P(X = x, Y = y)}{P(Y = y)} \text{ si } x,y \in \{1,2,\ldots,n\}$$

$$f_{X|Y}(x \mid y) = \frac{\frac{1}{n^2(n+1)}(x+y)}{\frac{1}{n}\left(\frac{n+1+2y}{2(n+1)}\right)}$$

asi que:

$$f_y(y) = \begin{cases} \frac{2}{n}\left(\frac{x+y}{n+1+2y}\right) & x \in \{1,2,\ldots,n\} \\ 0 & e.o.c. \end{cases}$$

Ejemplo 2.7. Se eligen al azar y sin reemplazo dos tarjetas de una urna que contiene N tarjetas numeradas de 1 hasta N, en donde N es mayor que 1 . Sean X y Y el menor y el mayor de los números respectivamente, de las tarjetas seleccionadas. Encuentre la función de densidad condicional de a) X dado que $Y = y$ para $Y \in \{2, \ldots, n\}$ y b) Y dado que $X = x$ para $Y \in \{1, 2, \ldots, n-1\}$

Sean Z y W los números en las tarjetas. Es claro que $X = min\{Z, W\}$ y $Y = max\{Z, W\}$

Las funciones de densidad de X y de Y:

$$f_Y(y) = P(max\{Z, W\} = y) = P(Z = y, W < y) = \begin{cases} \frac{y-1}{\binom{N}{2}} & \text{si } y \in \{2, \ldots, n\} \\ 0 & e.o.c. \end{cases}$$

$$f_X(x) = P(min\{Z, W\} = x) = P(Z = x, W > x) = \begin{cases} \frac{N-x}{\binom{N}{2}} & \text{si } x \in \{1, 2, \ldots, n-1\} \\ 0 & e.o.c. \end{cases}$$

Necesitamos tambien calcular la función de densidad conjunta de X y Y.

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{\binom{N}{2}} & \text{si } x \in \{1, 2, \ldots, n-1\}, y \in \{1, 2, \ldots, n-1\}, x < y \\ 0 & e.o.c. \end{cases}$$

Por lo tanto las soluciones son las siguientes:

a)

$$f_{X|Y}(x \mid y) = \begin{cases} \dfrac{\frac{1}{\binom{N}{2}}}{\frac{y-1}{\binom{N}{2}}} & \text{si } x \in \{1, 2, \ldots, n-1\} \\ 0 & e.o.c. \end{cases}$$

$$f_{X|Y}(x \mid y) = \begin{cases} \frac{1}{y-1} & \text{si } x \in \{1, 2, \ldots, n-1\} \\ 0 & e.o.c. \end{cases}$$

Es decir la función de densidad condicional de X dado que $Y = y$ es una densidad uniforme en el conjunto $\{1, 2, \ldots, y-1\}$.

b)

$$f_{X|Y}(x \mid y) = \begin{cases} \frac{1}{N-x} & \text{si } y \in \{x+1, \ldots, n\} \\ 0 & e.o.c. \end{cases}$$

Es decir la función de densidad condicional de Y dado que $X = x$ es una densidad uniforme en el conjunto $\{x+1, \ldots, N\}$

Ejemplo 2.8. Consideremos una sucesión de ensayos de Bernoulli independientes en cada uno de los cuales la probabilidad de éxito es igual a p, y para cada $k \in \mathbb{N}$, sea X_k el número de ensayo en el cual ocurre el k-ésimo éxito. Encuentre la función de densidad condicional de X_1 dado que $X_2 = n$ para $n \in \mathbb{N}$.

$f_{X|Y}(x \mid y)$ en palabras se puede ver como la probabilidad de que el primer éxito ocurra en el ensayo x dado que el segundo éxito ocurre en el ensayo n.

$$
\begin{aligned}
f_{X_1|X_2}(x \mid n) &= P(X_1 = x \mid X_2 = n) \\
&= \frac{P(X_1 = x, X_2 = n)}{P(X_2 = n)} \\
&= \begin{cases} \frac{(1-p)^{n-2}p^2}{(n-1)(1-p)^{n-2}p^2} & x \in \{1,\dots,n-1\} \\ 0 & e.o.c. \end{cases}
\end{aligned}
$$

Por lo tanto

$$
f_{X_1|X_2}(x \mid n) = \begin{cases} \frac{1}{n-1} & x \in \{1,\dots,n-1\} \\ 0 & e.o.c. \end{cases}
$$

Es decir dado que $X_2 = x$, X_1 tiene una densidad uniforme en $\{1,\dots,n-1\}$

Ejemplo 2.9. (Generalización del ejercicio anterior). Consideremos las condiciones del ejemplo anterior. Encuentre la función de densidad de X_k dado que $X_j = n$ para k<j.

$$
f_{X_k|X_j}(x \mid n) = P(X_k = x \mid X_j = n) = \frac{P(X_k = x, X_j = n)}{P(X_j = n)}
$$

$$
= \begin{cases} \dfrac{\binom{x-1}{k-1}\binom{n-x-1}{j-k-1}(1-p)^{n-j}p^j}{\binom{n-1}{j-1}(1-p)^{n-j}p^j} & \text{si } x \in \{k, k+1,\dots,n-j+k\} \\ 0 & e.o.c. \end{cases}
$$

Por lo tanto

$$
f_{X_k|X_j}(x \mid n) = \begin{cases} \dfrac{\binom{x-1}{k-1}\binom{n-x-1}{j-k-1}}{\binom{n-1}{j-1}} & \text{si } x \in \{k, k+1,\dots,n-j+k\} \\ 0 & e.o.c. \end{cases}
$$

Esto es debido aque debe haber $k-1$ éxitos en los primeros $x-1$ lugares y debe haber $j-1-k$ éxitos entre el lugar x y el lugar n sin contar ni al lugar x ni al n.

Hay que notar que la función no condicionada de X_n es una densidad binomial negativa de parámetros j y p.

Hasta aquí hemos hecho ejemplos aplicando unicamente la definición de densidad condicional. El siguiente ejemplo es importante pues sirve para motivacion de un resultado importante .

Ejemplo 2.10. Sean X y Y variables aleatorias discretas e independientes. Encontrar la densidad condicional de a) Z =max(X, Y) dado que $Y = y$ y b) W = min(X, Y) dado que

$X = x$. a)

$$
\begin{aligned}
f_{Z|Y}(z \mid y) &= \frac{P(Z = z, Y = y)}{P(Y = y)} = \frac{P(\max\{X, Y\} = z, Y = y)}{P(Y = y)} \\
&= \frac{P(\{\{X = z, Y = z\} \cup \{X < z, Y = z\} \cup \{X = z, Y < z\}\}, Y = y)}{P(Y = y)} \\
&= \frac{P(\{X = z, Y = z\}, Y = y)}{P(Y = y)} + \frac{P(\{X < z, Y = z\}, Y = y)}{P(Y = y)} + \frac{P(\{X = z, Y < z\}, Y = y)}{P(Y = y)} \\
&= \begin{cases} \frac{P(X = z, Y = y)}{P(Y = y)} & z = y \\ 0 & e.o.c. \end{cases} + \begin{cases} \frac{P(X < z, Y = y)}{P(Y = y)} & z = y \\ 0 & e.o.c. \end{cases} + \begin{cases} \frac{P(X = z, Y = z)}{P(Y = y)} & z > y \\ 0 & e.o.c. \end{cases} \\
&= \begin{cases} P(X \leq z) & z = y \\ P(X = z) & z > y \\ 0 & e.o.c \end{cases}
\end{aligned}
$$

b)

$$
\begin{aligned}
f_{W|X}(w \mid x) &= \frac{P(W = w, X = x)}{P(X = x)} = \frac{P(\min\{X, Y\} = w, X = x)}{P(W = x)} \\
&= \frac{P(\{X = w, Y \geq w\} \cup \{X > w, Y = w\}, X = x)}{P(X = x)} \\
&= \frac{P(\{X = z, Y \geq w\}, X = x)}{P(X = x)} + \frac{P(\{X > w, Y = w\}, X = x)}{P(X = x)} \\
&= \begin{cases} \frac{P(X = x, Y \geq w)}{P(X = x)} & w = x \\ 0 & e.o.c. \end{cases} + \begin{cases} \frac{P(X = x, Y = w)}{P(X = x)} & w < x \\ 0 & e.o.c. \end{cases} \\
&= \begin{cases} P(Y \geq w) & w = x \\ P(Y = w) & w < x \\ 0 & e.o.c \end{cases}
\end{aligned}
$$

El último paso sólo es válido utilizando la hipótesis de independencia de X y Y.

Poniendo atención a los resultados obtenidos nos damos cuenta que: a) La función de densidad condicional del máximo entre X y Y. dado que $Y = y$ es igual que la función de densidad de la variable aleatoria $\max(X, y)$.b).La función de densidad condicional del mínimo entre X y Y. dado que $X = x$ es igual que la función de densidad de la variable aleatoria $\min(x, Y)$. Este resultado en general es válido. La forma general sería como sigue:

Proposición 2.11. Sean X y Y. dos variables aleatorias discretas y sea $g : \mathbb{R}^2 \to \mathbb{R}$ cualquier función. Si $y \in \mathbb{R}$ es tal que $P(Y = y) > 0$ entonces

$$f_{g(X,Y)|Y}(z \mid y) = f_{g(X,y)}(z)$$

Demostración :

$$
\begin{aligned}
f_{g(X,Y)|Y}(z \mid y) &= P(g(X,Y) = z \mid Y = y) \\
&= \frac{P(g(X,Y) = z, Y = y)}{P(Y = y)} \\
&= \frac{P(g(X,y) = z, Y = y)}{P(Y = y)} \\
&= \frac{P(g(X,y) = z)P(Y = y)}{P(Y = y)} \\
&= P(g(X,y) = z) = f_{g(X,y)}(z) \qquad \blacksquare
\end{aligned}
$$

Hay que notar que en la penúltima igualdad se utiliza el hecho de que $g(X,y)$ es solamente una función de X, pues y está fija.

Ejemplo 2.12. Sean X y Y dos variables aleatorias independientes ambas con distribución uniforme en el conjunto $\{1,2,...,n\}$. Para x y $y \in \{1,2,...,n\}$. Encontrar la densidad condicional de a) $Z = max(X,Y)$ dado que $Y = y$ y b) $W = \min(X,Y)$ dado que $X = x$.

a)

$$
f_{Z|Y}(z \mid y) = f_{\max\{X,y\}}(z) = \begin{cases} P(X \le z) & z = y \\ P(X = z) & z > y \\ 0 & e.o.c \end{cases}
$$

$$
= \begin{cases} \frac{y}{n} & z = y \\ \frac{1}{n} & z \in \{y+1, y+2, ..., n\} \\ 0 & e.o.c \end{cases}
$$

b)

$$
f_{W|X}(w \mid x) = f_{\min\{x,Y\}}(w) = \begin{cases} P(Y \ge w) & w = x \\ P(Y = w) & w < x \\ 0 & e.o.c \end{cases}
$$

$$
= \begin{cases} \frac{n-x+1}{n} & w = x \\ \frac{1}{n} & w \in \{1,2,...,x-1\} \\ 0 & e.o.c \end{cases}
$$

Esperanza Condicional de una variable aleatoria discreta dado un evento

Consideremos el problema de lanzar dos dados. Sea X la suma de ambos. si sabemos que el segundo dado cayó en 6 ¿Cuál es la esperanza de X?

En este problema se nos pide encontrar la esperanza de X condicionada con el hecho de que el segundo dado cayo en 6, es decir se nos pide la esperanza condicional de X dado que ocurrió el evento B, el segundo dado cayó en 6. Podiamos denotar a esta esperanza de la siguiente forma $E(X \mid B)$.

Es claro que la condición de que B ha ocurrido altera la probabilidad de que X tome algún valor. De aquí que se ocurra calcular

$$\begin{aligned} E(X \mid B) &= 7P(X=7 \mid B) + 8P(X=8 \mid B) + 9P(X=9 \mid B) \\ &+ 10P(X=10 \mid B) + 11P(X=11 \mid B) + 12P(X=12 \mid B) \\ &= \frac{7+8+9+10+11+12}{6} = 9.5 \end{aligned}$$

Es decir para calcular $E(X \mid B)$ hacemos los calculos normales solo que usando la función de densidad condicional de X dado que ocurrio el evento B.

Usando el hecho de que:

$$\Omega = \left\{ \begin{array}{c} (1,1),(1,2),\ldots,(1,6) \\ \ldots \\ (6,1),(6,2),\ldots,(6,6) \end{array} \right\}$$

y

$$B = \{(1,6),(2,6),\ldots,(6,6)\}$$

el evento $\{X = x\} = \{\omega \in \Omega \mid X(\omega) = x\}$ y el evento $\{X = x \cap B\}$ se expresan en la siguiente tabla:

x	$X = x$	$\{X = x \cap B\}$
2	(1,1)	Ø
3	(1,2),(2,3)	Ø
4	(1,3),(2,2),(3,1)	Ø
5	(1,4),(2,3),(3,2),(4,1)	Ø
6	(1,5),(2,4),(3,3),(4,2),(5,1)	Ø
7	(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)	(1,6)
8	(2,6),(3,5),(4,4),(5,3),(6,2)	(2,6)
9	(3,6),(4,5),(5,4),(6,3)	(3,6)
10	(4,6),(5,5),(6,4)	(4,6)
11	(5,6),(6,5)	(5,6)
12	(6,6)	(6,6)

Si $I_B = \begin{cases} 1 & \text{si } B \text{ ocurre} \\ 0 & \text{si } B \text{ no ocurre} \end{cases}$ el evento $\{XI_B = x\}$ ocurre solamente si el evento $\{X = x \cap B\}$ también ocurre. es decir si $X = x$ y el resultado del experimento es un elemento de B y por lo tanto $I_B = 1$.

De aqui que:

$$E(X \mid B) = \sum_{x \in \mathbb{R}} xP(X = x \mid B) = \sum_{x \in \mathbb{R}} x\frac{P(X = x, B)}{P(B)} = \frac{1}{P(B)} \sum_{x \in \mathbb{R}} xP(\{X = x\} \cap B)$$
$$= \frac{1}{P(B)} \sum_{x \in \mathbb{R}} xP(XI_B = x) = \frac{1}{P(B)} E(XI_B).$$

Cabe notar que la esperanza condicional de una variable dado un evento, es la esperanza de la variable restringiendo el espacio muestral al evento, al igual que en probabilidad condicional.

En la última serie de igualdades no ha importado quien es B, de donde podemos definir a la esperanza condicional de una variable aleatoria discreta de esperanza finita X, dado un evento B de probabilidad positiva como:

$$E(X \mid B) = \sum_{x \in \mathbb{R}} xP(X = x \mid B)$$

y como anteriormente se vio esto es equivalente a definirla asi:

$$E(X \mid B) = \frac{E(XI_B)}{P(B)}$$

Esta última igualdad se puede generalizar pues, $\frac{E(XI_B)}{P(B)}$ se puede calcular independientemente de quienes son X y B basta con que X tenga esperanza finita y B sea de probabilidad positiva.

Definición 2.13 . Sea X una variable aleatoria de esperanza finita . Sea B un evento de probabilidad positiva. Se define a la esperanza condicional de X dado el evento B como:

$$E(X \mid B) = \frac{E(XI_B)}{P(B)}$$

Ejemplo 2.14. Un dado honesto es lanzado sucesivamente. Sean X y Y respectivamente el número de lanzamientos necesarios para obtener un 6 y un 5. Encontrar a)$E(X)$ b) $E(X \mid Y = 1)$ c)$E(X \mid Y = 5)$.

a)

$$f_X(x) = P(X = x) = \begin{cases} (\frac{5}{6})^{x-1}\frac{1}{6} & x \in \{1,2,...\} \\ 0 & e.o.c. \end{cases}$$

por lo tanto

$$E(X) = \sum_{x=1}^{\infty} x\left(\frac{5}{6}\right)^{x-1}\frac{1}{6} = \frac{1}{5}\sum_{x=1}^{\infty} x\left(\frac{5}{6}\right)^{x}$$

Para calcular esta suma definamos $S = r + 2r^2 + 3r^3 + ... + nr^n$
De aqui que:

$$rS = r^2 + 2r^3 + +... + nr^{n-1}$$

y sea G la famosa progresión geométrica $G = r + r^2 + ... + r^n$
es claro que $S = rS + G - nr^{n-1}$, despejando S de la igualdad anterior tenemos que:

$$S = \frac{G + nr^{n-1}}{1-r}$$

usando el sabido hecho de que $G = \frac{r - r^{n+1}}{1-r}$ sustituyendo G en S tenemos que:

$$S = \frac{\frac{r-r^{n+1}}{1-r} - nr^{n-1}}{(1-r)}$$

y debido a que $|r| < 1$

$$\lim_{n\to\infty} S = \frac{r}{(1-r)^2}$$

Por lo tanto

$$\sum_{x=1}^{\infty} x\left(\frac{5}{6}\right)^{x} = 30$$

y entonces $$E(X) = \frac{30}{5} = 6.$$

Este resultado es intuitivo, pues en 6 lanzamientos del dado se espera que salgan algo asi como una vez cada número de los del conjunto$\{1,2,3,4,5,6\}$ en particular el 6.

b) De acuerdo con el resultado del inciso anterior tenemos que:

$$P(X=x \mid Y=1)=\begin{cases} (\frac{5}{6})^{x-2}\frac{1}{6} & x\in\{2,3,\ldots\} \\ 0 & e.o.c \end{cases}$$

y por lo tanto

$$E(X \mid Y=1)=\sum_{x=2}^{\infty}x\left(\frac{5}{6}\right)^{x-2}\frac{1}{6}=\left(\frac{6}{5}\right)^{2}\frac{1}{6}\sum_{x=2}^{\infty}x\left(\frac{5}{6}\right)^{x}=\frac{6}{5^{2}}\left[30-\frac{5}{6}\right]=7$$

Este resultado tambien es intuitivo pues, debido a la independencia de X y Y intrínseca al experimento, al no caer 6 el primer lanzamiento, se le suma 1 al valor esperado de X.

c) $$P(X=x \mid Y=5)=\begin{cases} (\frac{5}{6})^{x-1}\frac{1}{6} & x\in\{1,2,3,4\} \\ (\frac{5}{6})^{x-2}\frac{1}{6} & x\in\{6,7,\ldots\} \\ 0 & e.o.c. \end{cases}$$

De acuerdo con los resultados en a) tenemos que:

$$E(X \mid Y=5)=\sum_{x=1}^{4}x\left(\frac{5}{6}\right)^{x-1}\frac{1}{6}+\sum_{x=6}^{\infty}x\left(\frac{5}{6}\right)^{x-2}\frac{1}{6}=\frac{8401}{1296}=6.4823$$

Este ejemplo es interesante , pues uno podria pensar que si el quinto lanzamiento es 5 hay que sumarle un 5 a la esperanza de X para obtener lo que se nos pide, sin embargo el hecho de que el quinto lanzamiento haya sido 5 no quiere decir que antes no pueda haber un 6.

Ejemplo 2.15. Una urna contiene 4 bolas blancas y 6 negras. se sacan 3 bolas y sin reemplazarlas se sacan otras 5. Sean X y Y respectivamente, el número de bolas blancas en la primera y segunda extracción. Calcular $E(X \mid Y=y)$ para $y\in\{1,2,3,4\}$.

$$\begin{aligned} E(X \mid Y=y) &= \sum_{x\in R}xP(X=x \mid Y=y) \\ &= \sum_{x\in R}x\frac{P(X=x,Y=y)P(X=x)}{P(Y=y)P(X=x)} \\ &= \sum_{x\in R}x\frac{P(Y=Y \mid X=x)P(X=x)}{P(Y=y)} \end{aligned}$$

$$P(X=x)=\begin{cases} \frac{\binom{4}{x}\binom{6}{3-x}}{\binom{10}{x}} & x\in\{0,1,2,3\} \\ 0 & e.o.c \end{cases}$$

es decir X es una hipergeométrica(10,4)

$$P(Y = y \mid X = x) = \begin{cases} \dfrac{\binom{4-x}{y}\binom{6-(3-x)}{5-y}}{\binom{7}{5}} & y \in \{1,2,3,4\} \\ 0 & e.o.c \end{cases}$$

$$= \begin{cases} \dfrac{\binom{4-x}{y}\binom{3-x}{5-y}}{\binom{7}{5}} & y \in \{1,2,3,4\} \\ 0 & e.o.c \end{cases}$$

Es decir Y tambien es una hipergeométrica

Ejemplo 2.16. Sean X y Y dos variables aleatorias independientes, ambas con distribución uniforme en el conjunto $\{1,2,3,\ldots,2n\}$ $n > 1$. Encontrar $a)$ $E(X \mid X > 2Y)$ y $b)$ $E(X \mid X + Y > 4)$

$$a) \qquad E(X \mid X > 2Y) = \frac{1}{P(X > 2Y)} E(XI_{\{x>2y\}})$$

$$P(X > 2Y) = \sum_{y=1}^{2n} P(Y = y, X > 2y) = \sum_{y=1}^{2n} P(Y = y) P(X > 2y)$$
$$= \sum_{y=1}^{n-1} \sum_{x=2y-1}^{2n} \frac{1}{4n^2} = \frac{1}{4} - \frac{1}{4n}$$

Por definición $E(XI_{\{x>2y\}}) = \sum_{x=1}^{2n} xP(X = x, x > 2Y)$ y como X y Y son independientes lo anterior sepuede escribir como

$$\sum_{y=1}^{n-1} \sum_{x=2y-1}^{2n} xP(X = x)P(Y = y)$$

Las parejas de x y y que cumplen que $P(X = x, x > 2Y) > 0$ son las que cumplen las siguientes condiciones: que $x,y \in \{1,2,\ldots,2n\}$ y $x > 2y$.

$$E(XI_{\{x>2y\}}) = \sum_{y=1}^{n-1} \sum_{x=2y-1}^{2n} x \frac{1}{4n^2}$$
$$= \sum_{y=1}^{n-1} \sum_{x=2y-1}^{2n} x \frac{1}{4n^2} = \frac{1}{3}n - \frac{1}{8} - \frac{5}{24n}$$

Por lo tanto $E(X \mid X > 2Y) = \frac{\frac{1}{3}n - \frac{1}{8} - \frac{5}{24n}}{\frac{1}{4} - \frac{1}{4n}} = \frac{4}{3}n + \frac{5}{6}$

$b)$ $$E(X \mid X+Y>4) = \frac{1}{P(X+Y>4)} E(XI_{\{x+y>4\}})$$

$$P(X+Y>4) = \sum_{x=1}^{2n} P(Y>4-X, X=x) = \sum_{x=1}^{2n} P(Y>4-X)P(X=x)$$
$$= \sum_{x=1}^{4} \frac{2N-(4-x)}{2N}\frac{1}{2N} + \sum_{x=5}^{2n}\frac{2n}{4n^2}$$
$$= \frac{4n-3}{2n^2} + \frac{(2n-4)2n}{4n^2} = \frac{2n^2-3}{2n^2}$$

Por definición $E(XI_{\{X+Y>4\}}) = \sum_{x=1}^{2n} xP(X=x, x+Y>4)$ y como X y Y son independientes lo anterior se puede escribir como:

$$\sum_{x=1}^{2n} xP(X=x)P(Y>4-x)$$

Al igual que en el inciso anterior tenemos que encontrar las parejas de x y de y que cumplen que $P(Y>4-x)>0$ y que $P(X=x)>0$ para obtener que:

$$E(XI_{\{X+Y>4\}}) = \sum_{x=1}^{4} x\frac{1}{2n}\frac{2n-(4-x)}{2n} + \sum_{x=5}^{2n} x\frac{1}{2n}$$
$$= \frac{1}{2}\frac{2n^2+n-10}{n} + \frac{5}{2n^2}(2n-1) = \frac{1}{2}\frac{-5+2n^3+n^2}{n^2}$$

Por lo tanto

$$E(X \mid X+Y>4) = \frac{\frac{2n^2-3}{2n^2}}{\frac{1}{2}\frac{-5+2n^3+n^2}{n^2}}$$
$$= \frac{-3+2n^2}{-5+2n^3+n^2}$$

Esperanza Condicional de una variable aleatoria discreta dada otra variable aleatoria discreta

Hasta aqui hemos calculado esperanzas condicionales de una variable aleatoria discreta dado un evento de probabilidad positiva. En particular hemos encontrado esperanzas condicionales de una variable aleatoria discreta dado un evento que está en términos de otra variable aleatoria

Si X y Y dos variables aleatorias discretas y calculamos la esperanza condicional de X dado el evento $Y=y$, entonces $E(X \mid Y=y)$ se puede ver como una función de y, pues para cada y tal que $P(Y=y)>0$ hay un valor de $E(X \mid Y=y)$.

Es claro que si $P(Y=y)>0$ entonces $E(X \mid Y=y)$ se puede calcular de tres formas, estas son

1) $E(X \mid Y=y) = \sum_{x\in\mathbb{R}} xf_{X|Y}(x \mid y)$,

2) $E(X \mid Y=y) = \sum_{x \in \Re} x \frac{f_{X,Y}(x,y)}{f_Y(y)}$,
3) $E(X \mid Y=y) = \frac{1}{f_Y(y)} E(X I_{Y=y})$

Por lo anterior se puede definir la esperanza condicional de una variable aleatoria discreta X dada otra variable aleatoria discreta Y de la siguiente forma:

Definición 2.17. Sean X y Y dos variables aleatorias discretas tales que $E(X) < \infty$. La esperanza condicional de la variable aleatoria discreta X dada la variable aleatoria discreta Y denotada por $E(X \mid Y)$ se define como :

$$E(X \mid Y) = h(Y)$$

donde

$$h(y) = \left\{ \begin{array}{cc} \sum_{x \in \Re} x f_{X|Y}(x \mid y) & \text{si } f_Y(y) > 0 \\ 0 & e.o.c. \end{array} \right\}$$

Hay que notar que $h(y)$ no es otra cosa que $E(X \mid Y=y)$.

El procedimiento para calcular $E(X \mid Y)$ es calcular $E(X \mid Y=y)$ usando cualquiera de las tres formas y después sustituir Y en el lugar de y.

Ejemplo 2.18. Sean X y Y dos variables aleatorias discretas con función de densidad conjunta dada por

Para $y = 1, 2, \ldots, N$, se tiene,

$$f_{X|Y}(x \mid y) = \left\{ \begin{array}{cc} \frac{2(x+y)}{N(N+1+2y)} & x \in \{1, 2, \ldots, N\} \\ 0 & e.o.c. \end{array} \right\}$$

(ver ejemplo 2.6)

Así que,

$$\begin{aligned} E[X \mid Y=y] &= \sum_{x=1}^{N} x f_{X|Y}(x \mid y) = \sum_{x=1}^{N} \frac{x 2(x+y)}{N(N+1+2y)} \\ &= \frac{2}{N(N+1+2y)} \left[\sum_{x=1}^{N} x^2 + y \sum_{x=1}^{N} x \right] \\ &= \frac{2}{N(2y+N+1)} \left[\frac{1}{6} N(N+1)(2N+1) + \frac{1}{2} N(N+1) y \right] \\ &= \frac{1}{2y+N+1} \left[\frac{1}{3} (N+1)(2N+1) + (N+1) y \right] \end{aligned}$$

Por lo tanto,

$$E[X \mid Y] = \frac{1}{2Y+N+1} \left[\frac{1}{3} (N+1)(2N+1) + (N+1) Y \right]$$

Notese que $E(X \mid Y)$ es una variable aleatoria, que es una función de la variable aleatoria Y. Y como tal cabe preguntarse si $E(X \mid Y)$ tiene esperanza finita para que en caso de que tenga calcularla. La respuesta es sí, $E(X \mid Y)$ tiene esperanza finita y por lo tanto tenemos la siguiente proposición

Proposición 2.19. Sea X una variable aleatoria discreta de esperanza finita y sea Y cualquier variable aleatoria discreta entonces

$$E(E(X \mid Y)) < \infty$$

Demostración: Recordemos que por definicion de esperanza , para que esta exista, la serie

$$\sum_y |E(X \mid Y)| P(Y = y)$$

debe ser convergente. Así que

$$\sum_y |E(X/Y)| P(Y = y) = \sum_y \left| \sum_x x P(X = x/Y = y) \right| P(Y = y)$$

y usando el hecho de que:

$$\left| \sum_x x P(X = x/Y = y) \right| \leq \sum_x |x| P(X = x/Y = y)$$

entonces

$$\begin{aligned}\sum_y |E(X \mid Y)| P(Y = y) &\leq \sum_y \sum_x |x| P(X = x \mid Y = y) P(Y = y)\\ &= \sum_y \sum_x |x| \frac{P(X = x, Y = y)}{P(Y = y)} P(Y = y) = \sum_y \sum_x |x| P(X = x, Y = y)\\ &= \sum_x \sum_y |x| P(X = x, Y = y) = \sum_x |x| P(X = x) = E(|X|) < \infty \qquad \blacksquare\end{aligned}$$

Ahora que sabemos que $E(E(X/Y)) < \infty$ calculémosla:

$$\begin{aligned}E(E(X \mid Y)) &= \sum_y E(X \mid Y) P(Y = y) = \sum_y \left[\sum_x x P(X = x \mid Y = y) \right] P(Y = y)\\ &= \sum_y \sum_x x \frac{P(X = x, Y = y)}{P(Y = y)} P(Y = y) = \sum_y \sum_x x P(X = x, Y = y)\\ &= \sum_x \sum_y x P(X = x, Y = y) = \sum_x x P(X = x) = E(X)\end{aligned}$$

Con esto queda demostrado el utilísimo resultado siguiente:

Proposición 2.20.(regla general de la probabilidad total para variables aleatorias

discretas).

Sea X una variable aleatoria discreta de esperanza finita y sea Y cualquier variable aleatoria discreta entonces

$$E(E(X \mid Y)) = E(X).$$

Este resultado se parece a la conocida regla de la probabilidad total:

$$P(X=x) = \sum_y P(X=x \mid Y=y)P(Y=y)$$

De hecho si B es un evento cualquiera

$$X = \begin{cases} 1 & B \text{ ocurre} \\ 0 & \text{si } B \text{ no ocurre} \end{cases}$$

tenemos que:

$$E(X) = \sum_y E(X \mid Y=y)P(Y=y) = \sum_y P(X=1 \mid Y=y)P(Y=y) = P(X=1)$$

Es decir: la regla de la probabilidad total no es otra cosa que un caso particular de la última proposición de dónde el nombre de **regla general de la probabilidad total**, que, como se verá mas adelante no sólo es válida para variables aleatorias discretas.

Ejemplo2.21. Un prisionero está en una celda que contiene tres puertas. La primera puerta da a un túnel que lo regresa a su celda después de dos días de caminar. La segunda da a un túnel que lo regresa a su celda después de cuatro días de caminar. La tercera lo deja libre después de un día de caminar. Si se asume que en cada ocasión el prisionero elige las puertas 1,2 y 3 con las respectivas probabilidades .5, .3 y .2 ¿Cuál es el número esperado de días que le llevará al prisionero recobrar su libertad?

Sea X: El número de días que le toma al prisionero recobrar su libertad.

Sea Y : La puerta que escoge inicialmente. Tenemos que:

$E(X \mid Y=3) = 1$

$E(X \mid Y=2) = 4 + E(X)$

$E(X \mid Y=1) = 2 + E(X)$. De aquí que:

$$E(X) = \sum_{y=1}^{3} E(X \mid Y=y)P(Y=y) = [2+E(X)].5 + [4+E(X)].3 + .2$$
$$= 1 + .5E(X) + 1.2 + .3E(X) + .2$$

Despejando $E(X)$ tenemos que:

$$E(X) = 12$$

Es decir en promedio le llevará al prisionero 12 días el ser libre.

Este problema sería muy dificil de resolver sin usar la gran herramienta que es la regla general de la probabilidad total. Tan sólo tómense en cuenta cuántos valores puede tomar X e intentese calcular $f_X(x)$.

Ejemplo 2.22. (Esperanza y varianza de un número aleatorio de variables aleatorias.) El número de personas que entran a un elevador en la planta baja de un edificio es una variable aleatoria Poisson con media 10. Si hay N pisos aparte de la planta baja y si cada persona se puede bajar del elevador en cualquiera de esos N pisos con igual probabilidad. Calcular a) el número esperado de paradas que hace el elevador hasta quedar vacío y b) La varianza del número de paradas que hace el elevador

Sea X : El número de paradas que hace el elevador hasta quedarse vacio.

Sea Y : el número de personas que entran al elevador en la planta baja

$$Sea\ X_i = \begin{cases} 1 & \text{Si al menos una persona baja del elevador en el piso } i \\ 0 & e.o.c. \end{cases}$$

$$X = \sum_{i=1}^{N} X_i$$

a) Queremos encontrar $E(X)$, donde $X = \sum_{i=1}^{N} X_i$

De aquí que queremos encontrar $E\left[\sum_i X_i\right]$

Tenemos que

$$E(X_i \mid Y = y) = P(X_i = 1 \mid Y = y) = 1 - P(X_i = 0 \mid Y = y)$$

El evento $X_i = 0$ dado que $Y = y$ en palabras es "Que nadie baje en el piso i dado que subieron y personas en la planta baja" Que es lo mismo que decir "Que toda la gente se baje en el piso que sea menos en el i-esimo dado que subieron y personas en la planta baja".

De aqui que:

$$E(X_i \mid Y = y) = 1 - \left(\frac{N-1}{N}\right)^y$$

pero

$$E(X_i) = E(E(X_i \mid Y)) = E\left[1 - \left(\frac{N-1}{N}\right)^y\right] = 1 - E\left[\left(\frac{N-1}{N}\right)^y\right]$$
$$= 1 - \sum_{y=0}^{\infty}\left(\frac{N-1}{N}\right)^y \frac{10^y}{y!}e^{-10} = 1 - e^{-10}\sum_{y=0}^{\infty}\frac{\left(\frac{10(N-1)}{N}\right)^y}{y!} = 1 - e^{-10}e^{\frac{10(N-1)}{N}} = 1 - e^{-\frac{10}{N}}$$

De aquí que la esperanza requerida es :

$$E(X) = E\left[\sum_i X_i\right] = \sum_i E(X_i) = NE(X_i) = N[1 - e^{-\frac{10}{N}}]$$

b) Para calcular la varianza del número de paradas que hace el elevador, tomemos en cuenta el inciso anterior.

y consideremos que

$$var(X) = var(\sum_{i=1}^{N} X_i) = \sum_{i=1}^{N} var(X_i) + 2\sum_{i<j} cov(X_i, X_j)$$

Tenemos que:

$$\begin{aligned} var(X_i) &= E(X_i^2) - [E(X_i)]^2 \\ &= E(E(X_i^2 \mid Y)) - [E(X_i)]^2 = E(1 - \left(\frac{N-1}{N}\right)^y) - (1 - e^{-\frac{10}{N}})^2 \\ &= 1 - e^{-\frac{10}{N}} - (1 - e^{-\frac{10}{N}})^2 = (1 - e^{-\frac{10}{N}})e^{-\frac{10}{N}} \end{aligned}$$

El hecho de que $E(X_i^2 \mid Y)$ sea igual a $1 - (\frac{N-1}{N})^y$ se justifica tomando en cuenta que los valores de x son exactamente los mismos quelos de x^2 y estos son solo 0 y 1.

Por otro lado tenemos que si $i,j \in \{1,2,...,N\}$ y además $i < j$

$$cov(X_i, X_j) = E(X_iX_j) - E(X_i)E(X_j)$$

Calculemos primero $E(X_iX_j)$ tomando en cuenta que

$$E(X_iX_j) = E(E(X_iX_j \mid Y))$$

pero

$$\begin{aligned} E(X_iX_j \mid Y = y) &= P(X_iX_j = 1 \mid Y = y) = 1 - P(X_iX_j = 0 \mid Y = y) \\ &= 1 - [P(X_i = 0, X_j = 0 \mid Y = y) + P(X_i = 1, X_j = 0 \mid Y = y) + P(X_i = 0, X_j = 1 \mid Y = y)] \\ &= 1 - [P(X_i = 0, X_j = 0 \mid Y = y)] \\ &\quad - [P(X_j = 0 \mid Y = y) - P(X_i = 0, X_j = 0 \mid Y = y)] \\ &\quad - [P(X_i = 0 \mid Y = y) - P(X_i = 0, X_j = 0 \mid Y = y)] \\ &= 1 - \left(\frac{N-2}{N}\right)^y - \left[\left(\frac{N-1}{N}\right)^y - \left(\frac{N-2}{N}\right)^y\right] - \left[\left(\frac{N-1}{N}\right)^y - \left(\frac{N-2}{N}\right)^y\right] \\ &= 1 + \left(\frac{N-2}{N}\right)^y - 2\left(\frac{N-1}{N}\right)^y \end{aligned}$$

Tomando esperanza tenemos que

$$\begin{aligned} E(X_iX_j) &= E(1 + \left(\frac{N-2}{N}\right)^y - 2\left(\frac{N-1}{N}\right)^y) \\ &= 1 + E\left[\left(\frac{N-2}{N}\right)^y\right] - 2E\left[\left(\frac{N-1}{N}\right)^y\right] \\ &= 1 + e^{-\frac{20}{N}} - 2[e^{-\frac{10}{N}}] \end{aligned}$$

Así que

$$cov(X_i, X_j) = 1 + e^{-\frac{20}{N}} - 2[e^{-\frac{10}{N}}] - [1 - e^{-\frac{10}{N}}]^2$$

Así que sustituyendo en $var(\sum_{i=1}^{N} X_i) = \sum_{i=1}^{N} var(X_i) + 2\sum_{i<j} cov(X_i, X_j)$

$$var(X) = N(1 - e^{-\frac{10}{N}})e^{-\frac{10}{N}} + 2\binom{N}{2}\left[1 + e^{-\frac{20}{N}} - 2[e^{-\frac{10}{N}}] - [1 - e^{-\frac{10}{N}}]^2\right]$$

Ejemplo 2.23. Sea X una variable aleatoria discreta de esperanza finita y sea Y cualquier variable aleatoria discreta acotada, calcular $E(YE(X \mid Y))$

Para poder calcular esta esperanza primero debemos saber si esta existe. Sea M la cota de la variable aleatoria Y

$$\begin{aligned}
E(|YE(X \mid Y)|) &= \sum_y |yE(X \mid Y = y)|P(Y = y) \\
&= \sum_y \left| y \sum_x xP(X = x \mid Y = y) \right| P(Y = y) \\
&= \sum_y \left| \sum_x yxP(X = x \mid Y = y) \right| P(Y = y) \\
&\leq \sum_y \sum_x |yx| P(X = x \mid Y = y)P(Y = y) \\
&= \sum_y |y| \sum_x |x| P(X = x, Y = y) \\
&\leq \sum_y M \sum_x |x| P(X = x, Y = y) \\
&= M \sum_y \sum_x |x| P(X = x, Y = y) = ME(|x|) < \infty
\end{aligned}$$

Así que ahora la podemos calcular:

$$\begin{aligned}
E(YE(X \mid Y)) &= \sum_y yE(X \mid Y = y)P(Y = y) \\
&= \sum_y y\left[\sum_x xP(X = x/Y = y) \right] P(Y = y) = \sum_y \sum_x yx \frac{P(X = x, Y = y)}{P(Y = y)} P(Y = y) \\
&= \sum_y \sum_x yxP(X = x, Y = y) = E(XY)
\end{aligned}$$

De lo anterior tendriamos el siguiente resultado que ya queda demostrado.

Proposición 2.24. Sea X una variable aleatoria discreta de esperanza finita y sea Y cualquier variable aleatoria acotada discreta entonces

$$E(YE(X \mid Y)) = E(XY).$$

¿Qué resultado tendriamos si calculamos $E(Y^2E(X/Y))$? Veamos

$$E(Y^2E(X \mid Y)) = \sum_y y^2 E(X \mid Y=y)P(Y=y)$$

$$= \sum_y y^2\left[\sum_x xP(X=x/Y=y)\right]P(Y=y) = \sum_y\sum_x y^2 x\frac{P(X=x,Y=y)}{P(Y=y)}P(Y=y)$$

$$= \sum_y\sum_x y^2 xP(X=x,Y=y) = E(Y^2X).$$

De acuerdo con el ejemplo tendríamos el resultado siguiente que queda demostrado con el ejemplo anterior.

Proposición 2.25.Sea X una variable aleatoria discreta de esperanza finita y sea Y cualquier variable aleatoria discreta acotada, entonces

$$E(Y^2E(X \mid Y)) = E(Y^2X).$$

Los resultados anteriores nos dan pie para un resutado general.

Proposición que caracteriza a $E(X \mid Y)$.Sea X una variable aleatoria discreta de esperanza finita y sea Y cualquier variable aleatoria discreta entonces

$$E(f(Y)E(X \mid Y)) = E(f(Y)X).$$

para toda $f:\mathbb{R}\to\mathbb{R}$ acotada

Demostración:

Primero demostremos que existe la esperanza de $E(f(Y)E(X \mid Y))$ para después calcularla

Sea $f:\mathbb{R}\to\mathbb{R}$ acotada y sea M la cota de f.

$$E(|f(Y)E(X \mid Y)|) = \sum_y |f(y)E(X \mid Y=y)|P(Y=y)$$

$$= \sum_y \left|f(y)\sum_x xP(X=x \mid Y=y)\right|P(Y=y) \le \sum_y\sum_x |f(y)x|\frac{P(X=x,Y=y)}{P(Y=y)}P(Y$$

$$= \sum_y\sum_x |f(y)||x|P(X=x,Y=y) \le M\sum_y\sum_x |x|P(X=x,Y=y) = ME(|X|) < \infty$$

Ahora podemos calcularla

$$E(f(Y)E(X \mid Y)) = \sum_y f(y)E(X \mid Y=y)P(Y=y)$$

$$= \sum_y f(y)\left[\sum_x xP(X=x \mid Y=y)\right]P(Y=y) = \sum_y \sum_x f(y)x\frac{P(X=x,Y=y)}{P(Y=y)}P(Y=$$

$$= \sum_y \sum_x f(y)xP(X=x,Y=y) = E(f(Y)X). \qquad \blacksquare$$

Hay que notar que se pide que f sea acota para garantizar que $E(f(Y)E(X \mid Y))$ sea finita..

La proposición anterior es muy importante pues nos da las dos características principales de $E(X \mid Y)$ cuando X y Y son como en la proposición: 1) La variable aleatoria $E(X \mid Y)$ tiene esperanza finita y 2) Para toda f:$\mathbb{R} \to \mathbb{R}$ acotada $E(f(Y)E(X \mid Y)) = E(f(Y)X)$

Estas dos características nos dan la pauta para otro resultado tambien importante :

Proposición (que rompe con la unicidad de $E(X/Y)$**).** Sea $g : \mathbb{R} \to \mathbb{R}$ una función que cumple que:

1) $E(g(Y)) < \infty$

2) $E(f(Y)g(Y)) = E(f(Y)X)$ para toda $f : \mathbb{R} \to \mathbb{R}$ acotada , entonces $P(g(Y) = E(X/Y)) = 1$

Demostración: Sea $g : \mathbb{R} \to \mathbb{R}$ una función que cumple que:

1) $E(g(Y)) < \infty$ y 2) $E(f(Y)g(Y)) = E(f(Y)X)$ para toda $f : \mathbb{R} \to \mathbb{R}$ acotada

Por 2) se tiene que:

$$\sum_y f(y)g(y)P(Y=y) = \sum_y \sum_x f(y)xP(X=x,Y=y)$$

para toda $f : \mathbb{R} \to \mathbb{R}$ acotada. en particular esto es válido para

$$f(y) = \begin{cases} 1 & P(Y>0) \\ 0 & e.o.c. \end{cases}$$

asi que en este caso tenemos que:

$$\sum_y g(y)P(Y=y) = \sum_y \sum_x xP(X=x,Y=y)$$

de aqui que:

$$g(y)P(Y=y) = \sum_x xP(X=x,Y=y)$$

por lo tanto:

$$g(y) = \sum_x x \frac{P(X = x, Y = y)}{P(Y = y)} = E(X/Y = y)$$

para todas las y tales que $P(Y = y) > 0$.

En conclusión tenemos que si $P(Y = y) > 0$. entonces

$$g(y) = E(X \mid Y = y)$$

. De aqui que se concluya que:

$$P(g(y) = E(X \mid Y = y)) = 1 \qquad \blacksquare$$

A partir de este resultado, a cualquier función $g : \mathbb{R} \to \mathbb{R}$ que cumpla con 1) y con 2) de la proposición anterior se le llama una **versión de la esperanza condicional de X dada** Y. Sin importar como son X y Y.

Si Y es una variable aleatoria discreta, para encontrar otra versiónde $E(X \mid Y)$ teniendo una, basta con cambiar la regla de correspondencia en los puntos en los que $P(Y = y) = 0$.

Cuando X yY son variables aleatorias discretas, dos versiones de $E(X \mid Y)$ pueden ser completamente distintas como lo muestra el siguiente ejemplo, y sin embargo son distintas en conjuntos de probabilidad cero y por ende no afecta en terminos probabilísticos ese problema. Pueden ser muy distintas, y sin embargo vistas probabilisticamente todas las versiones de $E(X \mid Y)$ son iguales.

Ejemplo 2.26. Considérense las condiciones del ejemplo 2.18. Se tiene que

$E(X \mid Y = 3) = 1$

$E(X \mid Y = 2) = 4 + E(X) = 6 + 12 = 16$

$E(X \mid Y = 1) = 2 + E(X) = 2 + 12 = 14$

Tenemos que

$$h_1(Y) = \begin{cases} 14 & y = 1 \\ 16 & y = 2 \\ 1 & y = 3 \\ 0 & e.o.c. \end{cases}$$

y

$$h_2(Y) = \begin{cases} 14 & y = 1 \\ 16 & y = 2 \\ 1 & y = 3 \\ -5 & y > 3 \\ 0 & e.o.c \end{cases}$$

son dos versiones de $E(X/Y)$

A la versionde$E(X \mid Y)$ que cumple con la definición dada para $E(X \mid Y)$ le llamaremos la **version original**

de $E(X \mid Y)$ o simplemente le llamaremos $E(X \mid Y)$. En terminos practicos, cualquier versiónde $E(X \mid Y)$ es igual que $E(X \mid Y)$. Debido a que todas las versiones de $E(X \mid Y)$ son iguales iguales, incluyendo a la versión original, excepto en un conjunto de probabilidad cero. De aquí en adelante cuando se hable de $E(X \mid Y)$ nos estaremos refiriendo a cualquier versión de $E(X \mid Y)$.

Utilizando la proposición que rompe con la unicidad de $E(X \mid Y)$ resolvamos el siguiente problema.

Ejemplo 2.27. Sean X y Y dos variables aleatorias independientes ambas con densidad uniforme en el conjunto $\{1,2,...n\}$. Encontrar $E(XY \mid X+Y)$

Sea $Z = X + Y$.

Tenemos que

$$f_Z(z) = P(X+Y=z) = P(Y=z-X) = \sum_{x=1}^{n} P(Y=z-x, X=x)$$
$$= \begin{cases} \sum_{x=1}^{z-1} P(Y=z-x)P(X=x) & z \in \{2,3,...,n\} \\ \sum_{x=z-n}^{n} P(Y=z-x)P(X=x) & z \in \{n+1,...,2n\} \end{cases}$$
$$= \begin{cases} \sum_{x=1}^{z-1} \frac{1}{n^2} & z \in \{2,3,...,n\} \\ \sum_{x=1}^{z-1} \frac{1}{n^2} & z \in \{n+1,...,2n\} \end{cases}$$

Por lo tanto tenemos que:

$$f_Z(z) = \begin{cases} \frac{z-1}{n^2} & z \in \{2,3,...,n\} \\ \frac{2n-1-z}{n^2} & z \in \{n+1,...,2n\} \end{cases}$$

Se busca una función h tal que

$$E(f(X+Y)h(X+Y)) = E(f(X+Y)XY)$$

para cualquier función f acotada. Desarrollando la igualdad de arriba y usando $f_Z(z)$ tenemos que:

$$\sum_{z=2}^{2n} f(z)h(z)f_Z(z) = \sum_{y=1}^{n}\sum_{x=1}^{n} f(x+y)xyf_{X,Y}(x,y)$$

de aqui se sigue que:

$$\sum_{z=2}^{n-1} f(z)h(z)\frac{z-1}{n^2} + \sum_{z=n+2}^{2n} f(z)h(z)\frac{2n+1-z}{n^2} = \sum_{y=1}^{n}\sum_{x=1}^{n} f(x+y)xy\frac{1}{n^2}$$
$$= \sum_{z=2}^{n-1}\sum_{y=1}^{z-1} f(z)(z-y)y\frac{1}{n^2} + \sum_{z=n+2}^{2n}\sum_{y=z-n}^{n} f(z)(z-y)y\frac{1}{n^2}$$

por lo tanto obtenemos las siguientes igualdades:

$$1)\qquad \sum_{z=2}^{n+1} f(z)h(z)\frac{z-1}{n^2} = \sum_{z=2}^{n+1}\sum_{y=1}^{z-1} f(z)(z-y)y\frac{1}{n^2}$$

$$2)\qquad \sum_{z=n+2}^{2n} f(z)h(z)\frac{2n+1-z}{n^2} = \sum_{z=n+2}^{2n}\sum_{y=z-n}^{n} f(z)(z-y)y\frac{1}{n^2}$$

De aqui que:

$$h(z)\frac{z-1}{n^2} = \sum_{y=1}^{z-1}(z-y)y\frac{1}{n^2} \text{ y } h(z)\frac{2n+1-z}{n^2} = \sum_{y=z-n}^{n}(z-y)\frac{1}{n^2}$$

Despejando a $h(z)$ tenemos que:

$$h(z) = \begin{cases} \dfrac{\sum_{y=1}^{z-1}(z-y)y\frac{1}{n^2}}{\frac{z-1}{n^2}} & z \in \{2,3,...,n\} \\ \dfrac{\sum_{y=z-n}^{n}(z-y)\frac{1}{n^2}}{\frac{2n+1-z}{n^2}} & z \in \{n+1,...,2n\} \end{cases} := \begin{cases} \dfrac{\frac{1}{6}z\frac{z^2-1}{n^2}}{\frac{z-1}{n^2}} & z \in \{2,3,...,n\} \\ \dfrac{-\frac{1}{2}z\frac{-2n-1+z}{n^2}}{\frac{2n+1-z}{n^2}} & z \in \{n+1,...,2n\} \end{cases}$$

Asi que una versión de $E(XY \mid X+Y)$ es $h(z)$ donde $Z = X+Y$ es como sigue.

$$h(z) = \begin{cases} \frac{1}{6}(z+1)z & z \in \{2,3,...,n\} \\ \frac{1}{2}z & z \in \{n+1,...,2n\} \end{cases}$$

Nota: La igualdad

$$\sum_{y=1}^{n}\sum_{x=1}^{n} f(x+y)xy\frac{1}{n^2} = \sum_{z=2}^{n+1}\sum_{y=1}^{z-1} f(z)(z-y)y\frac{1}{n^2} + \sum_{z=n+2}^{2n}\sum_{y=z-n}^{n} f(z)(z-y)y\frac{1}{n^2}$$

sale de rearreglar lo que buscamnos, es decir de que $\sum_{y=1}^{n}\sum_{x=1}^{n} f(x+y)xy\frac{1}{n^2}$ es igual que sumar las entradas de la siguiente tabla:

$f(2)1$	$f(3)2$	$f(4)3$	...	$f(n+1)n$
$f(3)2$	$f(4)4$	$f(5)6$	...	$f(n+2)2n$
$f(4)3$	$f(5)6$	$f(6)9$	...	$f(n+3)3n$
...	...	...	...	...
$f(n-1)(n-2)$	$f(n)2(n-2)$	$f(n+1)3(n-2)$	...	$f(2n-2)(n-2)n$
$f(n)(n-1)$	$f(n+1)2(n-1)$	$f(n+2)3(n-1)$	...	$f(2n-1)(n-1)n$
$f(n+1)n$	$f(n+2)2n$	$f(n+3)3n$	...	$f(2n)n^2$

La esperanza condicional de X dada Y tiene otras propiedades importantes. Una de ellas es la siguiente, que está acorde con la definición de densidad condicional.

Proposición 2.28. Sean X y Y variables aleatorias discretas independientes y X con esperanza finita. Entonces

$$E(X \mid Y) = E(X).$$

Demostración.
$E(X \mid Y = y) = \sum_x xP(X = x \mid Y = y)$ y debido a la independencia de X y Y.
$= \sum_x xP(X = x) = E(X)$ ■

Hay que notar que este es un resultado intuitivo, pues si Y es independiente de X, la ocurrencia del evento $Y = y$ no afecta a la ocurrencia del evento $X = x$, y por ende no afecta a $E(X)$.

La siguiente propiedad de la esperanza condicional de X dada Y está basada en la regla de la probabilidad total.

Proposición 2.29. Sea X una variable aleatoria discreta de esperanza finita. Sea Y otra variable aleatoria discreta. Sea $g : \mathbb{R} \to \mathbb{R}$ cualquier función tal que $g(Y)$ es una variable aleatoria . Entonces

$$E[E(X \mid Y) \mid g(Y)] = E(X \mid g(Y))$$

Demostración.
Primero demostremos que $E[E(X \mid g(Y))] < \infty$

$$\begin{aligned} E[|E(X \mid g(Y))|] &= \sum_y |E(X \mid g(y))| f_Y(y) \\ &= \sum_y \left| \sum_x x \frac{f_{X,g(Y)}(x, g(y))}{f_{g(Y)}(g(y))} \right| f_Y(y) \\ &\leq \sum_y \sum_x |x| \frac{f_{X,g(Y)}(x, g(y))}{f_{g(Y)}(g(y))} f_Y(y) \text{ Pero } P(g(Y) = g(y)) = P(Y = y) \text{ asi que} \\ &= \sum_y \sum_x |x| f_{X,g(Y)}(x, g(y)) = E(|X|) < \infty \end{aligned}$$

Ahora si probemos la igualdad.
Si $h(g(Y))$ es una versión de $E(X \mid g(Y))$ tendría que cumplir con que: para toda $f : \mathbb{R} \to \mathbb{R}$ acotada se cumple lo siguente:

$$E[f(g(Y))h(g(Y))] = E[f(g(Y))X] \text{.........}(1)$$

Para probar lo que se desea se tendría que probar que $E[E(X \mid Y) \mid g(Y)]$ cumple con lo mismo que h es decir que cumple con la ecuación (1). eso seria suficiente para probar que $E[E(X \mid Y) \mid g(Y)]$ es una versión de $E(X \mid g(Y))$.

Tenemos que:
$E[f(g(Y))E[E(X \mid Y) \mid g(Y)]] = E[f(g(Y))E(X \mid Y)]$ usando la propiedad que caracteriza a la esperanza de $E(X \mid Y)$ dada $g(Y)$
$= E[f(g(Y))X]$ usando la propiedad que caracteriza a la esperanza de X dada Y

Asi que $E[f(g(Y))E[E(X \mid Y) \mid g(Y)]]$ es una versión de $E[f(g(Y))X]$ ■

La demostración de este resultado no usa el hecho de que X y Y sean variables aleatorias discretas salvo por el hecho de que la proposición que caracteriza a la esperanza condicional solo se ha probado para variables aleatorias discretas.. Por lo tanto la misma demostración nos serrvira para el caso de variables aleatorias absolutamente continuas, habiendo por supuesto probado la proposición que caracteriza a la esperzanza condicional en el caso absolutamente continuo.

De la misma forma que las distribuciones condicionales cumplen con las propiedades de las distribuciones comunes, asi la esperanza condicional cumple con las siguientes tres propiedades que tambien tiene la esperanza común.

Proposición 2.30. Sean X y Y variables aleatorias discretas independientes y X con esperanza finita. Entonces:

1) $E(cX \mid Y) = cE(X \mid Y)$ para cada c constante.

2) $E(c \mid Y) = c$ para toda c constante.

3) $E(X_1 + X_2 \mid Y) = E(X_1 \mid Y) + E(X_2 \mid Y)$ donde X_1 y X_2 son variables aleatorias discretas de esperanza finita.

Demostración:

$$1)E(cX \mid Y = y) = \sum_x cxP(X = x \mid Y = y)$$
$$= c\sum_x xP(X = x \mid Y) = cE(X \mid Y)$$

$$2)E(c \mid Y) = \sum_x cP(X = x \mid Y = y) = c$$

$$3)E(X_1 + X_2 \mid Y) = \sum_{x_1}\sum_{x_2} x_1 x_2 f_{X_1X_2|Y}(x_1,x_2 \mid y)$$
$$= \sum_{x_1}\sum_{x_2} x_2 f_{X_1X_2|Y}(x_1,x_2 \mid y) + \sum_{x_1}\sum_{x_2} x_1 f_{X_1X_2|Y}(x_1,x_2 \mid y)$$
$$= \sum_{x_2}\sum_{x_1} x_2 f_{X_1X_2|Y}(x_1,x_2 \mid y) + \sum_{x_1}\sum_{x_2} x_1 f_{X_1X_2|Y}(x_1,x_2 \mid y)$$
$$= \sum_{x_2} x_2 f_{X_2|Y}(x_2 \mid y) + \sum_{x_1} x_1 f_{X_1|Y}(x_1 \mid y) = E(X_1 \mid Y) + E(X_2 \mid Y) \quad \blacksquare$$

La siguiente propiedad es intuitiva pero no por eso es fácil de probar en el caso general.

proposición 2.31. Sean X, Y y Z variables aleatorias discretas tales que X y Z tienen esperanza finita, tales que $X \leq Z$ entonces

$$E(X \mid Y) \leq E(Z \mid Y).$$

Demostración:

$$E(X \mid Y = y) = \frac{E(XI_{Y=y})}{P(Y = y)}$$

y por otro lado tenemos que

$$E(Z \mid Y=y) = \frac{E(ZI_{Y=y})}{P(Y=y)}$$

así que lo que queremos demostrar se deduce a demostrar que

$$E(XI_{Y=y}) \leq E(ZI_{Y=y})$$

para cada y tal que $P(Y=y) > 0$
por hipótesis tenemos que

$$P(X \leq Z) = 1$$

entonces

$$P(XI_{Y=y} \leq ZI_{Y=y}) = 1$$

así que

$$P(0 \leq ZI_{Y=y} - XI_{Y=y}) = 1$$

es decir

$$ZI_{Y=y} - XI_{Y=y}$$

es una variable aleatoria no negativa con probabilidad 1, entonces

$$E[ZI_{Y=y} - XI_{Y=y}] \geq 0$$

y por lo tanto

$$E[ZI_{Y=y}] - E[XI_{Y=y}] \geq 0$$

entonces

$$E[ZI_{Y=y}] \geq E[XI_{Y=y}]$$

y como esto es valido para cada y tal que $P(Y=y) > 0$ entonces

$$E(X \mid Y) \leq E(Z \mid Y). \qquad \blacksquare$$

Una interpretación de esta propiedad es como sigue : Si una variable aleatoria es menor que otra su promedio es menor que el de esa otra. Asumiendo que la esperanza condicional de una variable aleatoria dada otra es como la esperanza común pero restringiendo el espacio muestral., se tiene la propiedad.

Esperanza condicional de una función de dos variables aleatorias discretas dada una de ellas.

Ejercicio 2.32. Sean X y Y variables aleatorias con función de densidad conjunta dada por

$$f_{X,Y}(x,y) = \begin{cases} \frac{2x}{n^2(n-1)} & x,y \in \{1,2,...,n\} \\ 0 & e.o.c \end{cases}$$

calcular $E(XY \mid Y)$.

Usando la definición de esperanza condicional para variables aleatorias discretas, tendríamos lo siguiente

$$E(XY \mid Y=y) = \sum_w w \frac{P(XY=w, Y=y)}{P(Y=y)}$$

si hacemos $x = \frac{w}{y}$ esto es

$$E(XY \mid Y=y) = \sum_w w \frac{P(X=\frac{w}{y}, Y=y)}{P(Y=y)} = \sum_x xy \frac{P(X=x, Y=y)}{P(Y=y)}$$

calculando

$$P(Y=y) = \sum_{x=1}^{n} \frac{2x}{n^2(n+1)} = \frac{2x}{n^2(n+1)} \sum_{X=1}^{n} x$$

$$= \frac{2}{n^2(n+1)} \frac{n(n+1)}{2} = \begin{cases} \frac{1}{n} & y \in \{1,2,...,n\} \\ 0 & e.o.c \end{cases}$$

por lo tanto tenemos que:

$$E(XY \mid Y=y) = \sum_{x=1}^{n} xy \frac{\frac{2x}{n^2(n-1)}}{\frac{1}{n}} = \sum_{x=1}^{n} 2x^2 \frac{y}{n(n+1)} = 2\frac{y}{n(n+1)} \sum_{x=1}^{n} x^2 = \frac{1}{3}(2n+1)y$$

Asi que :

$$E(XY \mid Y) = \frac{1}{3}(2n+1)Y.$$

En este ejemplo usamos la función de densidad condicional de X y Y y no la función de densidad condicional de XY dada Y, ésta no es una característica de este ejemplo, sino que se puede hacer en general. La siguiente proposición recoge el resultado.

Proposición 2.33. Sean X y Y variables aleatrorias discretas. Sea $g : \mathbb{R} \to \mathbb{R}$ tal que $g(X,Y)$ es una variable aleatoriade esperanza finita, entonces

$$h(Y) = \sum_x g(x,y) f_{X|Y}(x \mid y) \text{ es una versión de } E(g(X;Y) \mid Y)$$

Demostración.

Primero demostremos que $h(Y)$ está bien definida es decir , que la serie $\sum_x |g(x,y)| f_{X|Y}(x \mid y)$ converge para toda y tal que $P(Y=y) > 0$.

Tenemos que

$$\sum_y \sum_x |g(x,y)|f_{X,Y}(x,y) < \infty$$

pues por hipótesis $E(g(X,Y)) < \infty$ lo cual implica que

$$\sum_x |g(x,y)|f_{X,Y}(x,y) < \infty$$

para cada $y \in \mathbb{R}Y$ si $P(Y=y) > 0$. entonces

$$\sum_x |g(x,y)|f_{X|Y}(x \mid y) < \infty$$

entonces $E[h(Y)] < \infty$.

Lo segundo es probar que $h(Y)$ tiene esperanza finita para poderla proponer como una version de $E(g(X,Y) \mid Y)$.

$$E(|h(Y)|) = \sum_y |h(y)|f_Y(y) = \sum_y \left| \sum_x g(x,y) \frac{f_{X,Y}(x,y)}{f_Y(y)} \right| f(x,y)$$

$$= \sum_y \sum_x |g(x,y)|f_{X,Y}(x,y) = E(|g(x,y)|) < \infty \text{ y por lo tanto } E(g(x,y)) < \infty$$

Por último probemos que en efecto h(Y) es una versión de $E(g(X;Y)/Y)$. Sea $f : \mathbb{R} \to \mathbb{R}$ acotada , entonces

$$\begin{aligned} E[f(Y)h(Y)] &= \sum_y f(y)h(y)P(Y=y) \\ &= \sum_y f(y) \left[\sum_x g(x,y) \frac{f_{X,Y}(x,y)}{f_Y(y)} \right] P(Y=y) \\ &= \sum_y f(y) \sum_x g(x,y) f_{X,Y}(x,y) \\ &= \sum_y \sum_x f(y) g(x,y) f_{X,Y}(x,y) = E[f(Y)g(X,Y)] \qquad \blacksquare \end{aligned}$$

En palabras lo que dice este resultado es que podemos calcular la esperanza condicional de $g(X;Y)$ dada la variable aleatoria Y ponderando con $f_{X/Y}$ en lugar de hacerlo como la definición lo dice con $f_{g(X,Y)/Y}$.

Hay que notar que este resultado sirve para calcular esperanzas condicionales de una función de X. digamos $g(X)$ dada la variable aleatoria Y, teniendo solamente la densidad condicional de X dado Y y no la densidad de la función de X dada Y. Ademas se cumple los siguientes resultados

Proposición 2.34. Sean X y Y variables aleatorias discretas independientes y X con esperanza finita. Sea $g : \mathbb{R} \to \mathbb{R}$ una función tal que $g(Y)X$ es una variable aleatoria de esperanza finita, entonces $g(Y)E(X \mid Y)$ tiene esperanza finita y

$$E(g(Y)X \mid Y) = g(Y)E(X \mid Y)$$

Demostración: Probemos que $g(Y)E(X \mid Y)$ tiene esperanza finita

$$\begin{aligned} E[|g(Y)E(X \mid Y)|] &= \sum_y |g(y)E(X \mid Y = y)| f_Y(y) \\ &= \sum_y \left| g(y) \sum_x x \frac{f_{X,Y}(x,y)}{f_Y(y)} \right| f_Y(y) \\ &\leq \sum_y \sum_x |g(y)x| f_{X,Y}(x,y) = E(|g(Y)X|) < \infty \end{aligned}$$

Ahora probemos que $E(g(Y)X \mid Y) = g(Y)E(X \mid Y)$ y debido a la proposición

$$\begin{aligned} E(g(Y)X \mid Y = y) &= \sum_x g(y) x f_{X|Y}(x \mid y) \\ &= g(y) \sum_x x f_{X|Y}(x \mid y) = g(Y)E(X \mid Y = y) \text{ asi que} \end{aligned}$$

$$E(g(Y)X \mid Y) = g(Y)E(X \mid Y) \qquad \blacksquare$$

Proposición 2.35. Sean X y Y variables aleatorias discretas . Sea $g : \mathbb{R} \to \mathbb{R}$ tal que $g(X)$ es una variable aleatoria de esperanza finita, entonces

$$E[E(g(X) \mid Y)] = E[g(X)]$$

Demostración.

$$\begin{aligned} E[E(g(X) \mid Y)] &= \sum_y E(g(X) \mid Y = y) P(Y = y) \\ &= \sum_y \left[\sum_x g(x) \frac{f_{X,Y}(x,y)}{f_Y(y)} \right] P(Y = y) \\ &= \sum_y \sum_x g(x) f_{X,Y}(x,y) = \sum_x \sum_y g(x) f_{X,Y}(x,y) \\ &= \sum_x g(x) f_X(x) = E(g(X)) \qquad \blacksquare \end{aligned}$$

Ejemplo 2.36. Considere un juego donde, quien juega gana lo que apuesta con probabilidad p y pierde lo que apuesta con probabilidad $(1-p)$. Cuando $p > \frac{1}{2}$ una forma muy popular de apostar en Inglaterra conocida como "la estratégia de Kelley" es siempre apostar $2p-1$ de lo que te quedae de dinero. Calcular la cantidad esperada de dinero que tendra un jugador después de n juegos si al principio tenía x pesos y empleo la estrategia de Kelley.

Sea Y_n : La cantidad de dinero que tiene el epostador después del $n-esimo$ juego.

Sea Z_n : El número de partidas ganadas después del $n-ésimo$ juego.

Se pide encontrar $E(Y_n)$.

La variable aleatoria Z_n es necesaria debido a que para saber cual es la ganancia, al apostador le basta con saber cuantas partidas gano y cuantas perdió, sin importar el orden en que esto sucedio pues al final de cuentas la ganancia es la misma independientemente del orden de los ganados y de los perdidos.

Z_n es binomial con parámetros n y p.

Por otro lado si $z \in \{0,1,\ldots,n\}$ se tiene que

$E(Y_n/Z_n = z) = x[1+(2p-1)]^z[1-(2p-1)]^{n-z}$

Y como

$$E(Y_n) = E(E(Y_n \mid Z_n))$$

entonces

$$\begin{aligned}
E(Y_n) &= \sum_{z=0}^{n} E(Y_n \mid Z_n = z)P(Z_n = z) \\
&= \sum_{z=0}^{n} x[1+(2p-1)]^z[1-(2p-1)]^{n-z}\binom{n}{z}p^z(1-p)^{n-z} \\
&= x\sum_{z=0}^{n} 1+(2p-1)p^z[1-(2p-1)(1-p)]^{n-z}\binom{n}{z}
\end{aligned}$$

La anterior es una suma de Newton y por lo tanto

$$\begin{aligned}
E(Y_n) &= x([1+(2p-1)]+[1-(2p-1)])^n \\
&= x[1+(2p-1)^2]^n.
\end{aligned}$$

Ejemplo 2.37. Sean X y Y variables aleatorias independientes ambas con distribución uniforme en $\{1,2,\ldots n\}$. Encontrar $E(\min\{X,Y\})$ dada la variable aleatoria Y.

$$\begin{aligned}
E[\min\{X,Y\} \mid Y=y] &= \sum_x \min\{x,y\} f_{X/Y}(x,y) = \sum_x \min\{x,y\}\frac{1}{n} \\
&= \sum_{x=1}^{y} x\frac{1}{n} + \sum_{x=y+1}^{n} y\frac{1}{n} = \left[\frac{y(y+1)}{2}\right]\frac{1}{n} + \frac{n-y}{n} = \frac{y(y-1)}{2n} + 1
\end{aligned}$$

Asi que

$$E(\min\{X,Y\}) = \frac{Y(Y-1)}{2n} + 1$$

Con este ejemplo no solo se ve la aplicación del resultado anterior sino se tiene que:

$$E[\min\{X,Y\} \mid Y = y] = E[\min\{X,y\}]$$

ya que $E[\min\{X,y\}] = \sum_x \min\{x,y\} f_{XY}(x,y)$ por definicion.

Asi lo anterior da lugar a la siguiente proposición.

Proposición 2.38. Sean X y Y variables aleatorias independientes, discretas. Sea $g : \mathbb{R} \to \mathbb{R}$ tal que $E(g(X,Y)) < \infty$ entonces

$$E(g(X,Y) \mid Y = y) = E(g(X,y))$$

Demostración:

$$E(g(X,Y) \mid Y = y) = \sum_x g(x,y) \frac{f_{XY}(x,y)}{f_Y(y)}$$

$$= \sum_x g(x,y) P(X = x) = E(g(X,y)) \qquad \blacksquare$$

Varianza Condicional

Asi como la definición de esperanza condicional de la variable aleatoria X dada la variable aleatoria Y es analoga a la definición usual de esperanza de una variable aleatoria solo que todas las distribuciones condicionadas al hecho de que se conoce Y. Podemos tambien definir a la varianza condicional de X dada la variable aleatoria Y como la varianza usual solo que todas las esperanzas que entervienen en la definición estarán condicionadas a la variable aleatoria Y.

Nos quedaria la siguiente definición.

Definición 2.39. Sean X y Y variables aleatorias. Se define a la varianza condicional de X dada la variable aleatoria Y como

$$Var(X \mid Y) = E[(X - E(X \mid Y))^2 \mid Y]$$

La varianza condicional de X dada la variable aleatoria Y es una nueva variable aleatoria que se puede ver como una función de la variable aleatoria Y.

Un primer resultado que también tiene su análogo en la varianza usual es el siguiente:

Proposición 2.40.

$$var(X \mid Y) = E(X^2 \mid Y) - [E(X \mid Y)]^2$$

Demostración.

$$\begin{aligned} var(X \mid Y) &= E[(X-E(X \mid Y)^2 \mid Y] = E\left[X^2 - 2XE(X \mid Y) + [E(X \mid Y)]^2 \mid Y\right] \\ &= E\left[X^2 - 2XE(X \mid Y) + [E(X/Y)]^2 \mid Y\right] \\ &= E(X^2 \mid Y) - E(2XE(X \mid Y) \mid Y) + E\left[[E(X \mid Y)]^2 \mid Y\right] \\ &= E(X^2 \mid Y) - 2[E(X \mid Y)]^2 + [E(X \mid Y)]^2 \\ &= E(X^2 \mid Y) - [E(X \mid Y)]^2 \qquad \blacksquare \end{aligned}$$

Si calculamos ahora la esperanza de la variable aleatoria $var(X \mid Y)$ llegamos a un resultado muy importante de la varianza condicional.

Tenemos que

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

Asi que podemos enunciar la siguiente proposición cuya demostración se reduce a despejar $var(X)$de la igualdad anterior.

Proposición 2.41.

$$var(X) = E(var(X \mid Y)) + var(E(X \mid Y))$$

Este resultado nos puede ayudar a resolver problemas como el siguiente.

Ejemplo 2.42. Un minero está atrapado en una mina que tiene 3 túneles. El primer tunel lo lleva a la salidad de la mina después de caminar 3 horas. El segundo tunel lo regresa a la mina después de caminar 5 horas. El tercer tunel lo regresa a la mina después de caminar 7 horas.Si se asume que el minero escoge en cada ocasión un tunel al azar. Calcular la varianza del tiempo que le lleva al minero salir de la mina.

Sea X :el tiempo en horas que le lleva al minero salir de la mina.

Sea Y la puerta que inicialmente escoge.

Se requiere calcular $var(X)$: Es claro que sería difícil calcularla sin la útil ayuda que nos da ,el saber el tunel que elige al principio el minero, pues la densidad de X difícil de calcular.

Podemos usar la fórmula anterior $var(X) = E(var(X \mid Y)) + var(E(X \mid Y))$

Y para calcularla tenemos que:

$$E(X \mid Y) = \begin{cases} 3 & y=1 \\ 5+E(X) & y=2 \\ 7+E(X) & y=3 \\ 0 & e.o.c \end{cases}$$

De aqui que

$$E(X) = E(E(X \mid Y)) = \frac{1}{3}(3+5+E(X)+7+E(X)) = 5+\frac{2}{3}E(X).$$

Despejando $E(X)$ tenemos que $E(X) = 15$ y por ende

$$E(X/Y) = \begin{cases} 3 & y=1 \\ 20 & y=2 \\ 22 & y=3 \\ 0 & e.o.c \end{cases}$$

Ahora podemos calcular

$$\begin{aligned} var(E(X \mid Y)) &= \frac{1}{3}[(3-E(E(X \mid Y)))^2+(20-E(E(X \mid Y)))^2+(22-E(E(X \mid Y)))^2] \\ &= \frac{1}{3}[(3-E(X))^2+(20-E(X))^2+(22-E(X))^2] = \frac{218}{3} \end{aligned}$$

Por otro lado calculemos $var(X \mid Y) = E(X^2 \mid Y) - (E(X \mid Y))^2$

$$E(X^2 \mid Y) = \begin{cases} 9 & y=1 \\ E((5+X)^2) & y=2 \\ E((7+X)^2) & y=3 \\ 0 & e.o.c \end{cases}$$

Esta última igualdad es debida al hecho de que, si el minero escoge la segunda puerta caminara 5 horas y estará de nuevo en la mina, en la misma situación del inicio, y el tiempo que le llevara salir de la mina es $X+5$ y por lo tanto el cuarado del tiempo que le toma al minero salir de la mina es $(X+5)^2$. El razonamiento en los demás casos es análogo.

Tenemos entonces que:

$$E(X^2 \mid Y) = \begin{cases} 9 & y=1 \\ E(5)+10E(X)+E(X^2) & y=2 \\ E(7)+14E(X)+E(X^2) & y=3 \\ 0 & e.o.c \end{cases} = \begin{cases} 9 & y=1 \\ 5+10(15)+E(X^2) & y=2 \\ 7+14(15)+E(X^2) & y=3 \\ 0 & e.o.c \end{cases}$$

Calculando

$$E(X^2) = \sum_{y=1}^{3} E(X^2 \mid Y = y)P(Y = y) = \frac{1}{3}[9 + 5 + 10(15) + E(X^2) + 7 + 14(15) + E(X^2)]$$

Resolviendo esta ecuación para $E(X^2)$ tenemos que $E(X^2) = 443$ y por lo tanto

$$E(X^2 \mid Y) = \begin{cases} 9 & y = 1 \\ 618 & y = 2 \\ 702 & y = 3 \\ 0 & e.o.c \end{cases} \qquad var(X \mid Y) = \begin{cases} 218 & y = \{2,3\} \\ 0 & e.o.c \end{cases}$$

entonces

$$E(var(X \mid Y)) = \frac{2}{3}218$$

y en total tenemos que

$$var(X) = \frac{2}{3}(218) + \frac{1}{3}(218) = 218$$

Capítulo III

Distribuciones y Esperanzas condicionales en el caso absolutamente continuo.

Distribucion Condicional de una variable aleatoria dado el valor de otra variable aleatoria

Si tenemos dos variables aleatorias X y Y absolutamente continuas, se puede definir la función de densidad condicional de la variable aleatoria X dado que la variable aleatoria Y toma el valor y de la misma forma que se hizo en el caso discreto.

Definición 3.1. Sean X y Y variables aleatorias absolutamente continuas. Se define a la función de densidad de X y Y para todos los valores de Y tales que $f_Y(y) > 0$ de la siguiente forma:

$$f_{X|Y}(x \mid y) = \begin{cases} \frac{f_{X,Y}(x,y)}{f_Y(y)} & f_Y(y) > 0 \\ 0 & e.o.c \end{cases}$$

donde $f_{X,Y}(x,y)$ es la función de densidad conjunta de X y Y.

Nótese que $f_{X|Y}(x \mid y)$ es una función de x. Al igual que en variables aleatorias discretas, la función de densidad condicional de una variable aleatoria X dada la variable aleatoria Y, es una función de densidad, vista como función de los valores de X si ambas son absolutamente continuas.

Proposición 3.2. Sean X y Y variables aleatorias absolutamente continuas entonces. $f_{X|Y}(x \mid y)$ tal como se definió arriba es una función de densidad vista como función de x.

Demostración. Tenemos que demostrar que i) $f_{X|Y}(x \mid y) \geq 0$ y ii) $\int_{-\infty}^{\infty} f_{X|Y}(x \mid y)dx = 1$

Sea y tal que $f_Y(y) > 0$, tenemos que $f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} \geq 0$.

Por otra parte $\int_{-\infty}^{\infty} f_{X|Y}(x \mid y)dx = \int_{-\infty}^{\infty} \frac{f_{X,Y}(x,y)}{f_Y(y)}dx = \frac{1}{f_Y(y)} \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx = \frac{1}{f_Y(y)} f_Y(y) = 1$ ■

Como toda densidad $f_{X|Y}(x \mid y)$ tiene asociada a ella una función de distribución.

Definición 3.3. La función de distribución condicional de la variable aleatoria X dado que la variable aleatoria Y toma el valor de y, se define de la siguiente forma

$$F_{X|Y}(x \mid y) = P(X \leq x \mid Y = y) = \int_{-\infty}^{x} f_{X|Y}(u \mid y)du$$

La definición anterior nos permite calcular $P(X \in A \mid Y = y)$

(*para cualquier A* evento) se tendria que

$$P(X \in A \mid Y = y) = \int_A f_{X|Y}(u \mid y)du$$

Esta forma de definir a la función de densidad condicional de X dada Y y a su función de distribución recoge la idea intuitiva de que si X es independiente de Y entonces el valor de Y no afecta a la función de densidad condicional de X dado que $Y = y$. Sobre esto es el siguiente resultado.

Proposición 3.4. Sean X y Y variables aleatorias absolutamente continuas e independientes, entonces

i) $f_{X|Y}(x \mid y) = f_X(x)$

ii) $F_{X|Y}(x \mid y) = F_X(x)$

Demostración: Sea y tal que $f_Y(y) > 0$

i) $f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{f_X(x)f_Y(y)}{f_Y(y)} = f_X(x)$

ii) $F_{X|Y}(x \mid y) = P(X \leq x \mid Y = y) = \int_{-\infty}^{x} f_{X|Y}(u \mid y)du = \int_{-\infty}^{x} f_X(u)du = F_X(x)$ ■

Al igual que pasa con las distribuciones no condicionales, si tenemos la función de distribución condicional de X dado que $Y = y$ entonces $f_{X|Y}(x \mid y)$ es cualquier función que cumpla que

$$F_{X|Y}(x \mid y) = \int_{-\infty}^{x} f_{X|Y}(u \mid y)du$$

En el caso en que X y Y son absolutamente continuas, siempre existe la densidad condicional de X dado que $Y = y$. Aunque no necesariamente es única. O dicho de otra forma puede haber varias densidades iguales en todos lados excepto quizás en un conjunto de probabilidad 0.

Ejemplo 3.5. Sean X y Y variables aleatorias con función de densidad conjunta dada por

$$f_{X,Y}(x,y) = \begin{cases} \frac{12}{5}y(2-x-y) & x,y \in (0,1) \\ 0 & e.o.c. \end{cases}$$

Encontrar la densidad condicional de la X dado $Y = y$

Si $y \in (0,1)$

$$f_Y(y) = \int_0^1 \frac{12}{5}y(2-x-y)dx = \frac{12}{5}y\int_0^1(2-x-y)dx = \frac{12}{5}y\left(\frac{3}{2}-y\right)$$

de donde

$$f_Y(y) = \begin{cases} \frac{12}{5}y(\frac{3}{2} - y) & y \in (0,1) \\ 0 & e.o.c. \end{cases}$$

entonces

$$f_{X|Y}(x \mid y) = \begin{cases} \frac{\frac{12}{5}y(2-x-y)}{\frac{12}{5}y(\frac{3}{2}-y)} & x \in (0,1) \\ 0 & e.o.c. \end{cases} = \begin{cases} \frac{(2-x-y)}{(\frac{3}{2}-y)} & x \in (0,1) \\ 0 & e.o.c. \end{cases}$$

Ejemplo 3.6. Sean X y Y variables aleatorias independientes ambas con función de densidad exponencial de parámetro λ. Encontrar la densidad condicional de X dado que $X + Y = s$

Calculemos primero

$$F_{X-Y}(s) = \iint_{\{(x,y)\in\mathbb{R}^2 : x-y<s, x>0, y>0\}} f_{X,Y}(x,y)dydx$$
$$= \int_0^s \int_0^{s-x} \lambda^2 e^{-\lambda(x+y)}dydx = \int_0^s (-\lambda e^{-\lambda s}u - e^{-u\lambda} + 1)dx = 1 - e^{-\lambda s}(1 + s\lambda)$$

asi que

$$f_{X-Y}(s) = \frac{dF_{X-Y}(s)}{ds} = \begin{cases} \frac{d(1-e^{-\lambda s}(1-s\lambda))}{ds} & x > 0 \\ 0 & e.o.c. \end{cases} = \begin{cases} \lambda^2 e^{-\lambda s}s & x > 0 \\ 0 & e.o.c. \end{cases}$$

Este resultado se puede obtener evaluando s en una densidad gamma con parámetros $(\lambda, 2)$. Pues es sabido que la suma de variables aleatorias exponenciales independientes es gamma.

Por otro lado calculemos

$$F_{X,X-Y}(x;s) = P(X \leq x, X + Y \leq s) = \iint_{\{(u,y)\in\mathbb{R}^2 : u\leq x, u-y<s, u>0, y>0\}} f_{X,Y}(u,y)dydu$$
$$= \int_0^{\min\{x,s\}} \int_0^{s-u} \lambda^2 e^{-\lambda(u-y)}dydu = \int_0^{\min\{x,s\}} (-\lambda e^{-\lambda s} + \lambda e^{-u\lambda})du$$
$$= \begin{cases} 0 & x < 0 \\ \int_0^x (-\lambda e^{-\lambda s} + \lambda e^{-u\lambda})du & x < s \\ \int_0^s (-\lambda e^{-\lambda s} + \lambda e^{-u\lambda})du & x \geq s \end{cases} := \begin{cases} 0 & x < 0 \\ 1 - \lambda x e^{-\lambda s} - e^{-\lambda x} & x < s \\ 1 - \lambda e^{-\lambda s}s - e^{-\lambda s} & x \geq s \end{cases}$$

De donde

$$f_{X,X-Y}(x,s) = \frac{dF_{X,X-Y}(x,s)}{ds} = \begin{cases} 0 & x < 0 \\ \frac{d(1-\lambda x e^{-\lambda s} - e^{-\lambda x})}{dsdx} & x < s \\ \frac{d(1-\lambda e^{-\lambda s}s - e^{-\lambda s})}{dxds} & x \geq s \end{cases} = \begin{cases} \lambda^2 e^{-\lambda s} & x < s \\ 0 & e.o.c \end{cases}$$

finalmente se tiene que:

$$f_{X|X+Y}(x \mid s) = \begin{cases} \frac{\lambda^2 e^{-\lambda s}}{\lambda^2 e^{-\lambda s} s} & 0 < x < s \\ 0 & e.o.c \end{cases} = \begin{cases} \frac{1}{s} & 0 < x < s \\ 0 & e.o.c \end{cases}$$

En conclusión la función de densidad condicional de X dado que $X+Y=s$ es una densidad uniforme en $(0,s)$.

Ejemplo 3.7. Sean X y Y variables aleatorias absolutamente continuas con función de densidad conjunta dada por:

$$f_{X,Y}(x,y) = \begin{cases} c(x^2-y^2)e^{-x} & x \geq 0, -x < y < x \\ 0 & e.o.c. \end{cases}$$

Encontrar la densidad condicional de Y dado que $X = x$

Tenemos que $F_{Y|X}(y \mid x) = P(Y \leq y \mid X = x) = \int_{-\infty}^{y} f_{Y|X}(u \mid y)du$

Calculemos primero $f_X(x)$.

$$f_X(x). = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$$

entonces

$$= \int_{-x}^{x} c(x^2-y^2)e^{-x}dy = \frac{4}{3}x^3ce^{-x}$$

asi que

$$f_X(x) = \begin{cases} \frac{4}{3}x^3ce^{-x} & x \geq 0 \\ 0 & e.o.c. \end{cases}$$

si $x \geq 0$ entonces

$$f_{Y|X}(y \mid x) = \begin{cases} \frac{c(x^2-y^2)e^{-x}}{\frac{4}{3}x^3ce^{-x}} & -x < y < x \\ 0 & e.o.c. \end{cases} = \begin{cases} \frac{3}{4}\frac{x^2-y^2}{x^3} & -x < y < x \\ 0 & e.o.c. \end{cases}$$

:

Ejemplo 3.8. Sean X y Y variables aleatorias independientes amabas con distribución gamma de parámetros α, λ y β, λ respectivamente. Calcular la densidad condicional de $X+Y$ dado que $\frac{X}{X+Y} = w$

Como X y Y son independientes entonces

$$f_{X,Y}(x,y) = \begin{cases} \frac{\lambda^\alpha \lambda^\beta}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} y^{\beta-1} e^{-\lambda(x+y)} & x,y > 0 \\ 0 & e.o.c. \end{cases}$$

Para encontrar $f_{X+Y|\frac{X}{X+Y}}(z \mid w)$ necesitamos calcular $f_{X+Y,\frac{X}{X+Y}}(z,w)$ es decir la densidad conjunta de $X+Y$ y $\frac{X}{X+Y}$. Para ello se usará el teorema de cambio de variable. Tenemos todas las condiciones para usar este teorema pues la transformación $z = x+y, w = \frac{x}{x+y}$ es invertible y su inversa es $x = zw, y = z - zw$. Las derivadas parciales $\frac{dx}{dz} = w,\ \frac{dx}{dw} = z,\ \frac{dy}{dz} = 1-w,\ \frac{dy}{dw} = -z$ son continuas y el jacobiano

$$\begin{vmatrix} \frac{dx}{dz} & \frac{dx}{dw} \\ \frac{dx}{dw} & \frac{dy}{dw} \end{vmatrix} = \begin{vmatrix} w & z \\ 1-w & -z \end{vmatrix} = |-z| = z$$

es distinto de cero pues X y Y son mayores que 0.

De acuerdo a esto tenemos que

$$\begin{aligned} f_{X+Y,\frac{X}{X+Y}}(z,w) &= z f_{X,Y}(zw, z(1-w)) \\ &= \begin{cases} z\frac{\lambda^{\alpha}\lambda^{\beta}}{\Gamma(\alpha)\Gamma(\beta)}(zw)^{\alpha-1}(z(1-w))^{\beta-1}e^{-\lambda(zw+z-zw)} & zw>0, z(1-w)>0 \\ 0 & e.o.c. \end{cases} \\ &= \begin{cases} \frac{\lambda^{\alpha}\lambda^{\beta}}{\Gamma(\alpha)\Gamma(\beta)} z^{\alpha} w^{\alpha-1} z^{\beta-1}(1-w)^{\beta-1}e^{-\lambda z} & z>0, w\in(0,1) \\ 0 & e.o.c. \end{cases} \\ &= \begin{cases} \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha+\beta)} z^{\alpha+\beta-1}e^{-\lambda z}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} w^{\alpha-1}(1-w)^{\beta-1} & z>0, w\in(0,1) \\ 0 & e.o.c. \end{cases} \end{aligned}$$

La función de arriba es el producto de dos funciones de densidad una gama con parámetros $\alpha+\beta, \lambda$ y otra beta con parámetros α, β. Por lo tanto Z y W son independientes y por ende

$$f_{X+Y|\frac{X}{X+Y}}(z,w) = \begin{cases} \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha+\beta)} z^{\alpha+\beta-1}e^{-\lambda z} & z>0 \\ 0 & e.o.c. \end{cases}$$

es decir: La densidad condicional de $X+Y$ dada $\frac{X}{X+Y} = w$ es gamma con parámetros $\alpha+\beta, \lambda$.

Hay que notar que en los ejemplos anteriores se esta condicionando con un evento que tiene probabilidad 0 .

Distribucion Condicional de una variable aleatoria dado un evento.

Considérese un problema como el siguiente: Se eligen dos puntos al azar del intervalo $(0,1)$. Si llamamos x al primer punto y y al segundo y se sabe que la suma de ambos resulto ser mayor que $\frac{1}{2}$ ¿Cuál es la densidad condicional de X?

En este problema se requiere la distribución condicional de X dado un evento de probabilidad positiva, para ello tenemos la siguiente definición.

Definición 3.9. Sea A un evento de probabilidad positiva y sea X una variable aleatoria absolutamente continua entonces la densidad condicional de X dado A se define de la siguiente manera

$$F_{x|A}(x \mid A) = P(X \leq x \mid A) = \frac{P(X \leq x, A)}{P(A)}$$

En nuestro problema la distribución condicional de X dado A quedaria como sigue

$$F_{x|A}(x \mid A) = P(X \leq x \mid X+Y > \frac{1}{2}) = \frac{P(X \leq x, X+Y > \frac{1}{2})}{P(X+Y > \frac{1}{2})}$$

Tenemos que

$$P(X+Y > \frac{1}{2}) = \iint_{\{(x,y)\in\mathbb{R}^2: x+y>\frac{1}{2}\}} f_{X,Y}(x,y)\,dydx = \iint_{\{(x,y)\in\mathbb{R}^2: x+y>\frac{1}{2}; x,y\in(0,1)\}} dydx$$
$$= \int_0^{\frac{1}{2}}\int_{\frac{1}{2}-x}^{1} dydx = \int_0^{\frac{1}{2}}\left(\frac{1}{2}+x\right)dx = \frac{3}{8}$$

Por otra parte

$$P(X \leq x, X+Y > \frac{1}{2}) = \iint_{\{(u,y)\in\mathbb{R}^2: u\leq x, u+y>\frac{1}{2}; x,y\in(0,1)\}} dydu = \int_0^{\min\{\frac{1}{2},x\}}\int_{\frac{1}{2}-u}^{1} dydu$$

$$= \int_0^{\min\{\frac{1}{2},x\}}(\frac{1}{2}+u)du = \begin{cases} \int_0^{\frac{1}{2}}(\frac{1}{2}+u)du & \in (\frac{1}{2}x, 1) \\ \int_0^{x}(\frac{1}{2}+u)du & x \in (0, \frac{1}{2}] \\ 0 & x \leq 0 \\ 1 & x \geq 1 \end{cases} = \begin{cases} \frac{3}{8} & x \in (\frac{1}{2}, 1) \\ \frac{x}{2}+\frac{x^2}{2} & x \in (0, \frac{1}{2}] \\ 0 & x \leq 0 \\ 1 & x \geq 1 \end{cases}$$

Asi que $F_{x|A}(x \mid A)$ queda como sigue:

$$F_{x|A}(x \mid A) = \begin{cases} \frac{\frac{3}{8}}{\frac{3}{8}} & x \in (\frac{1}{2}, 1) \\ \frac{\frac{x}{2}-\frac{x^2}{2}}{\frac{3}{8}} & x \in (0, \frac{1}{2}] \\ 0 & x \leq 0 \\ 1 & x \geq 1 \end{cases} = \begin{cases} 0 & x \leq 0 \\ \frac{4}{3}x + \frac{4}{3}x^2 & x \in (0, \frac{1}{2}) \\ 1 & x \geq \frac{1}{2} \end{cases}$$

Y por lo tanto $f_{X|A}(x \mid A)$ que es la densidad condicional de la variable aleatoria X dado el evento A obtenida derivando $F_{x|A}(x \mid A)$ nos queda como sigue:

$$f_{X|A}(x \mid A) = \frac{dF_{x|A}(x \mid A)}{dx} = \begin{cases} \frac{d(\frac{4}{3}x - \frac{4}{3}x^2)}{dx} & x \in (0, \frac{1}{2}) \\ 0 & e.o.c. \end{cases} = \begin{cases} \frac{4}{3} + \frac{8}{3}x & x \in (0, \frac{1}{2}] \\ 0 & e.o.c. \end{cases}$$

Esperanza condicional de una variable aleatoria absolutamente continua dado un evento

La definición de esperanza condicional de una variable aleatoria dado un evento de probabilidad positiva , dada en el capitulo anterior, tiene la virtud de ser general, es decir, se puede aplicar independientemente de quién sea X, lo único que se le pide esque tenga esperanza finita y que el evento sea de probabilidad positiva.

La definición es la siguiente

Definición 3.10. Sea X una variable aleatoria de esperanza finita . Sea B un evento de probabilidad positiva. Se define a la esperanza condicional de X dado el evento B como:

$$E(X \mid B) = \frac{E(XI_B)}{P(B)}$$

Ejemplo 3.11. La función de densidad conjunta de X y Y está dada por .

$$f_{X,Y}(x,y) = \begin{cases} \frac{e^{-\frac{x}{y}} e^{-y}}{y} & 0 < x < \infty, 0 < y < \infty \\ 0 & e.o.c \end{cases}$$

Calcular a) $E(X \mid Y > X)$ y b) $E(X^2 \mid Y > X)$

a) Tenemos que

$$P(Y > X) = \int\int_{\{(x,y)\in\mathbb{R}^2: x<y, x>0, y>0\}} \frac{e^{-\frac{x}{y}} e^{-y}}{y} dxdy$$
$$= \int_0^\infty \int_0^y \frac{e^{-\frac{x}{y}} e^{-y}}{y} dxdy = \int_0^\infty (-e^{-1-y} + e^{-y})dy = 1 - e^{-1}$$

Por otro lado como $XI_{\{Y>X\}}$ es una funcion de las variables aleatorias X y Y se tiene que

$$E(XI_{\{Y>X\}}) = \int\int_{\{(x,y)\in\mathbb{R}^2: x<y, x>0, y>0\}} x\frac{e^{-\frac{x}{y}} e^{-y}}{y} dxdy$$
$$= \int_0^\infty \int_0^y x\frac{e^{-\frac{x}{y}} e^{-y}}{y} dxdy = \int_0^\infty (-2ye^{-1-y} + ye^{-y})dy = 1 - 2e^{-1}$$

Asi que

$$E(X/Y > X) = \frac{1-2e^{-1}}{1-e^{-1}} \approx .41802$$

b) Como $X^2I_{\{Y>X\}}$ es una funcion de las variables aleatorias X y Y se tiene que

$$E(X^2I_{\{Y>X\}}) = \int\int_{\{(x,y)\in\mathbb{R}^2: x<y, x>0, y>0\}} x^2\frac{e^{-\frac{x}{y}} e^{-y}}{y} dxdy$$
$$= \int_0^\infty \int_0^y x^2\frac{e^{-\frac{x}{y}} e^{-y}}{y} dxdy = 4 - 10e^{-1}$$

Asi que

$$E(X^2/Y > X) = \frac{4 - 10e^{-1}}{1 - e^{-1}} \approx .508\,14$$

Ejemplo 3.12. Sea X una variable aleatoria con distribución uniforme en $(0,1)$. Encontrar $E(X \mid X > \frac{1}{2})$

Tenemos que

$$E(XI_{\{x>\frac{1}{2}\}}) = \int_{\{x\in\mathbb{R}: x>\frac{1}{2}0, x\in(0,1)\}} x dx = \int_{\frac{1}{2}}^{1} x dx = \frac{3}{8}$$

Considerando que $P(X > \frac{1}{2})$ tenemos que

$$E\left(X \mid X > \frac{1}{2}\right) = \frac{\frac{3}{8}}{\frac{1}{2}} = \frac{3}{4}$$

Esperanza condicional de una variable aleatoria absolutamente continua dada otra variable aleatoria absolutamente continua

Una forma de definir a la esperanza condicional de una variable aleatoria absolutamente continua X dado el evento $Y = y$, es definiéndola como la esperanza usual solo que usando la densidad condicional de X dado que $Y = y$. Esta forma de hacerlo resulta natural pues generaliza al caso discreto. Con esto se puede definir a la esperanza condicional de la variable aleatoria X dada la variable aleatoria Y de la siguiente forma:

Definición 3.13. Sean X y Y variables aleatorias absolutamente continuas con función de densidad condicional de X dado que $Y = y$ dada por $f_{X|Y}(x \mid y)$. X de esperanza finita. Se define a la esperanza condicional de X dada la variable aleatoria Y como $h(Y)$ donde

$$h(y) = \begin{cases} \int_{-\infty}^{\infty} x f_{X|Y}(x \mid y) dx & \text{si } f_Y(y) > 0 \text{ y } \int_{-\infty}^{\infty} |x| f_{X,Y}(x,y) dx < \infty \\ 0 & e.o.c. \end{cases}$$

y se denota por $E(X \mid Y)$.

Se pide que X sea de esperanza finita para que $h(Y)$ esté bien definida y además tenga esperanza finita.

La siguiente definición ya se había esbozado en el capitulo anterior y fue motivada también del caso discreto.

Definición 3.14. Sean X y Y variables aleatorias absolutamente continuas. $E(X) < \infty$. Si existe una funcion $g : \mathbb{R} \to \mathbb{R}$ que cumple que $i)$ $E(g(Y)) < \infty$ y $ii)$ $E(f(Y)g(Y)) = E(f(Y)X)$ para toda $f : \mathbb{R} \to \mathbb{R}$ acotada, entonces se dice que $g(Y)$ es una versión de la esperanza condicional de X dada Y.

$h(Y)$ cumple con $i)$ y $ii)$ de la segunda definición. Sin embargo no toda función que cumpla con esto, es igual a $h(Y)$. En otras palabras no solo hay una versión de la esperanza condicional de X dado Y. Lo que si se puede asegurar es que si se tienen dos versiones de la esperanza condicional de X dado Y estas son iguales con probabilidad 1. Los siguientes resultados resumen los comentarios anteriores.

Proposición 3.15 (Proposición que rompe con la unicidad). Sea X una variable aleatoria absolutamente continua de esperanza finita y sea Y otra variable aleatoria absolutamente continua, entonces, $h(Y)$ es una version de la esperanza condicional de X dada Y.

Demostración: El hecho de que el conjunto de puntos $y \in \mathbb{R}$ tales que $\int_{-\infty}^{\infty}|x|f_{X,Y}(x,y)dx = \infty$ tiene medida cero, está implicito en la demostración. aunque esto no es obvio no se demostrará aqui.

Se quiere demostrar que $h(Y)$ es tal que

i) $E(h(Y)) < \infty$ y

ii) Si $f : \mathbb{R} \to \mathbb{R}$ es acotada, entonces $E[f(Y)h(Y)] = E(f(Y)X)$

Demostremos primero *i)*

$$\begin{aligned}\int_{-\infty}^{\infty}|h(y)|f_Y(y)dy &= \int_{-\infty}^{\infty}\left|\int_{-\infty}^{\infty} x\frac{f_{X,Y}(x,y)}{f_Y(y)}dx\right|f_Y(y)dy \\ &\leq \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|x|\frac{f_{X,Y}(x,y)}{f_Y(y)}dxf_Y(y)dy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|x|f_{X,Y}(x,y)dxdy = E(|X|) < \infty\end{aligned}$$

ii) Sea $f : \mathbb{R} \to \mathbb{R}$ cualquier función acotada

$$\begin{aligned}E[f(Y)h(Y)] &= \int_{-\infty}^{\infty} f(y)h(y)f_Y(y)dy = \int_{-\infty}^{\infty} f(y)f_Y(y)\left[\int_{-\infty}^{\infty} x\frac{f_{X,Y}(x,y)}{f_Y(y)}dx\right]dy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf_{X,Y}(x,y)f(y)dxdy = E(f(Y)X)\end{aligned}$$ ■

Ejemplo 3.16. Sean X y Y variables aleatorias con función de densidad conjunta dada por:

$$f_{X,Y}(x,y) = \begin{cases} c(x^2-y^2)e^{-x} & x > 0, -x < y < x \\ 0 & e.o.c.\end{cases}$$

encontrar $E(Y \mid X)$

De acuerdo con el ejemplo 3.7 tenemos que :

$$f_{Y|X}(y \mid x) = \begin{cases} \frac{3}{4}\frac{x^2-y^2}{x^3} & -x < y < x \\ 0 & e.o.c.\end{cases}$$

Entonces:

$$\begin{aligned}E(Y \mid X = x) &= \int_{-\infty}^{\infty} yf_{Y|X}(y,x)dy \\ &= \int_{-\infty}^{\infty} y\frac{3}{4}\frac{x^2-y^2}{x^3}dy = 0\end{aligned}$$

pues

$$\int_{-\infty}^{0} y\frac{3}{4}\,\frac{x^2-y^2}{x^3}dy = \int_{0}^{\infty} y\frac{3}{4}\,\frac{x^2-y^2}{x^3}dy$$

asi que:

$$E(Y \mid X) = 0$$

Ejemplo 3.17. Sean X y Y variables aleatorias independientes ambas con distribución exponencial de parámetro λ. Sea $Z = X + Y$. Calcular $E(X \mid Z)$

Tenemos que

$$\begin{aligned} F_{X,X+Y}(x,z) &= P(X \leq x, Z \leq z) \\ &= P(X \leq x, X+Y \leq z) = \iint_{\{(u,y)\in\mathbb{R}^2: u<x, u+y<z, u>0, y>0\}} f_{X,Y}(x,y)dydx \\ &= \int_0^x \int_0^{z-u} \lambda^2 e^{-\lambda(u+y)} dydu = \int_0^x (-\lambda e^{-z\lambda} + \lambda e^{-u\lambda})du \quad = 1 - \lambda e^{-z\lambda}x - e^{-x\lambda} \end{aligned}$$

asi que

$$f_{X,X+Y}(x,z) = \frac{d^2 F_{X,X+Y}(z)}{dxdz} = \begin{cases} \frac{d^2(1-\lambda e^{-z\lambda}x - e^{-x\lambda})}{dxdz} & 0 < x < z \\ 0 & e.o.c. \end{cases} = \begin{cases} \lambda^2 e^{-\lambda z} & 0 < x < z \\ 0 & e.o.c. \end{cases}$$

por otra parte

$$f_{X+Y}(z) = \begin{cases} \lambda^2 z e^{-\lambda z} & 0 < z \\ 0 & e.o.c. \end{cases}$$

Asi que

$$f_{X|X+Y}(x \mid z) = \begin{cases} \frac{1}{z} & 0 < x < z \\ 0 & e.o.c. \end{cases}$$

Pues, como se vio en el ejemplo 3.6, la suma de exponenciales independientes de parámetro λ es gamma $(2,\lambda)$

Asi que

$$E(X \mid Z) = \int_{-\infty}^{\infty} x f_{X|Z}(x \mid z)dx = \begin{cases} \int_0^z x\frac{1}{z}dx & z > 0 \\ 0 & e.o.c. \end{cases} = \begin{cases} \frac{1}{2}z & z > 0 \\ 0 & e.o.c. \end{cases}$$

Finalmente tenemos que

$$h(Z) = \frac{Z}{2}$$

Ejemplo 3.18. Sean X y Y variables aleatorias independientes ambas con distribución uniforme en $(0,1)$. Encontrar $E(XY \mid X+Y)$

Para que $h(X+Y)$ sea una versión de la esperanza condicional $E(XY \mid X+Y)$ se tiene que cumplir que $E[f(X+Y)h(X+Y)] = E[f(X+Y)XY]$ para toda función $f : \mathbb{R} \to \mathbb{R}$

acotada. Pues esta es la caracterización de la esperanza condicional de una variable dada otra.

Sea $f:\mathbb{R}\rightarrow\mathbb{R}$ acotada. Sea $Z = X + Y$ Desarrollando la igualdad de arriba tenemos que:

$$\int_{-\infty}^{\infty} f(z)h(z)f_z(z)dz = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x+y)xyf_{X,Y}(x,y)dydx$$

Sustituyendo a $f_z(z) = \begin{cases} z & 0 < z < 1 \\ 2-z & 1 \le z < 2 \\ 0 & e.o.c. \end{cases}$ en la igualdad de arriba y tomando en cuenta que X y Y son independientes tenemos que:

$$\int_0^1 f(z)h(z)zdz + \int_1^2 f(z)h(z)(2-z)dz = \int_0^1\int_0^1 f(x+y)xydydx$$
$$= \int_0^1\int_0^z f(z)(z-y)ydydz + \int_1^2\int_{z-1}^1 f(z)(z-y)ydydz$$

Asi que:

$$\int_0^1 f(z)h(z)zdz + \int_1^2 f(z)h(z)(2-z)dz = \int_0^1\int_0^z f(z)(z-y)ydydz + \int_1^2\int_{z-1}^1 f(z)(z-y)ydydz$$

Y por lo tanto

$$1)\ \int_0^1 f(z)h(z)zdz = \int_0^1\int_0^z f(z)(z-y)ydydz$$

$$2)\ \int_1^2 f(z)h(z)(2-z)dz = \int_1^2\int_{z-1}^1 f(z)(z-y)ydydz$$

asi que:

$$1)\quad f(z)h(z)z = \int_0^z f(z)(z-y)ydy$$

$$2)\qquad f(z)h(z)(2-z) = \int_{z-1}^1 f(z)(z-y)ydy$$

y por ende

$$1)\qquad h(z)z = \int_0^z (z-y)ydy$$

$$2)\qquad h(z)(2-z) = \int_{z-1}^1 (z-y)ydy$$

Despejando $h(z)$ tenemos que:

$$h(z)=\begin{cases}\frac{\int_0^z (z-y)y\,dy}{z} & z\in(0,1)\\ \frac{\int_{z-1}^1 (z-y)y\,dy}{(2-z)} & z\in[1,2)\\ 0 & e.o.c\end{cases}=\begin{cases}\frac{\frac{1}{6}z^3}{z} & z\in(0,1)\\ \frac{-\frac{2}{3}+z-\frac{1}{6}z^3}{(2-z)} & z\in[1,2)\\ 0 & e.o.c\end{cases}=\begin{cases}\frac{1}{6}z^2 & z\in(0,1)\\ \frac{1}{6}z^2+\frac{1}{3}z-\frac{1}{3} & z\in[1,2)\\ 0 & e.o.c\end{cases}$$

Regla general de la probabilidad total para variables aleatorias absolutamente continuas.

Un caso particular de la proposición 3.15 que establece que Si $f:\mathbb{R}\to\mathbb{R}$ es acotada, entonces $E[f(Y)h(Y)] = E(f(Y)X)$, se tiene cuando $f(y) = 1$ para toda $y\in\mathbb{R}$.

Sustituyendo esta f en la igualdad se obtiene:

$$E[E(X \mid Y)] = E(X)$$

que en el caso en que Y sea absolutamente continua, esta igualdad se traduce en:

$$E(X) = \int_{-\infty}^{\infty} E(X \mid Y=y)f_Y(y)dy$$

A esta igualdad se le llama regla general de la probabilidad total.

esta regla , aparte de útil pues a veces como en el siguiente ejemplo , lo que se tiene como dato es la esperanza condicional y lo que se quiere es la esperanza común.

Si en el caso discreto es muy intuitiva, en el caso absolutamente continuo es una generalización de éste.

Ejemplo 3.19. Se elige al azar un punto en el intervalo $(0,1)$ Sea y ese número. Sea X una variable aleatoria con distribuión uniforme en $(y,1)$. Encontrar $E(X)$.

Tenemos que

$$E(X) = E(E(X \mid Y)) = \int_{-\infty}^{\infty} E(X \mid Y=y)f_Y(y)dy$$
$$= \int_0^1\int_y^1\left(\frac{x}{1-y}\right)dxdy = \int_0^1\left(\frac{1}{2}y+\frac{1}{2}\right)dy = \frac{3}{4}$$

Ejemplo 3.20. Supongamos que se envía una señal aleatoria X desde un lugar A de tal manera que su distribución es $N(\mu,\sigma^2)$. Supongamos además que, cuando $X=x$, el valor Y, que se recibe en un lugar B, tiene una distribución $N(x,a^2)$, en donde a es una constante distinta de cero. Encuentre $E(Y)$, $Var(Y)$, $Cov(X,Y)$, $f_{X,Y}(x,y)$ y $f_Y(y)$

Para encontrar $E(Y)$ tenemos que usar la regla de la probabilidad total para variables aleatorias absolutamente continuas.

$$E(Y) = \int_{-\infty}^{\infty} E[Y \mid X=x]f_X(x)dx = \int_{-\infty}^{\infty} xf_X(x)dx = E(X) = \mu$$

Ahorta encontremos $E(Y^2)$ usando otra vez a regla de la probabilidad total

$$E(Y^2) = \int_{-\infty}^{\infty} E[Y^2 \mid X = x] f_X(x)dx = \int_{-\infty}^{\infty} (a^2 + x^2) f_X(x)dx = a^2 + \int_{-\infty}^{\infty} x^2 f_X(x)dx$$
$$= a^2 + E(X^2) = a^2 + \sigma^2 + \mu^2$$

$$Var(Y) = E(Y^2) - [E(Y)]^2 = a^2 + \sigma^2 + \mu^2 - \mu^2 = a^2 + \sigma^2$$

$$Cov(X,Y) = E[XY] - E[X]E[Y] = E[E(XY \mid X)] - \mu^2 = E[XE(Y \mid X)] - \mu^2$$
$$= E[X^2] - \mu^2 = \sigma^2 + \mu^2 - \mu^2 = \sigma^2$$

En el problema esta dada $f_{Y|X}(y \mid x)$ asi que nos sirve para calcular $f_{X,Y}(x,y)$

$$f_{X,Y}(x,y) = f_{Y|X}(y \mid x) f_X(x) = \frac{1}{\sqrt{2\pi}\, a} \exp\left\{-\frac{1}{2a^2}(y-x)^2\right\} \frac{1}{\sqrt{2\pi}\, \sigma} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$
$$= \frac{1}{2\pi a\sigma} \exp\left\{-\frac{1}{2}\left[\left(\frac{y-x}{a}\right)^2 + \left(\frac{x-\mu}{\sigma}\right)^2\right]\right\}$$
$$= \frac{1}{2\pi a\sigma} \exp\left\{-\frac{1}{2}\left[\left(\frac{(y-\mu)-(x-\mu)}{a}\right)^2 + \left(\frac{x-\mu}{\sigma}\right)^2\right]\right\}$$

Se sabe que Y tiene distribución normal, entonces,
$f_Y(y) = \frac{1}{\sqrt{2\pi}\sqrt{a^2-\sigma^2}} \exp\left\{-\frac{1}{2(a^2-\sigma^2)}(y-\mu)^2\right\}$

Ejemplo 3.21. Sea X un número que se elige al azar en el intervalo $(0,1)$ y Y un número que se elige al azar en el intervalo $(0,X)$. Encuentre la función de densidad Y.

Se tiene

$$f_{Y|X}(y \mid x) = \begin{cases} \frac{1}{x} & \text{si } 0 < y < x < 1 \\ 0 & \text{en otro caso} \end{cases}$$

Así que,

$$f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X}(y \mid x) f_X(x)dx = \int_0^1 f_{Y|X}(y \mid x)dx$$
$$= \begin{cases} \int_y^1 \frac{1}{x}dx & \text{si } 0 < y < 1 \\ 0 & \text{en otro caso} \end{cases} = \begin{cases} -\ln y & \text{si } 0 < y < 1 \\ 0 & \text{en otro caso} \end{cases}$$

Capítulo IV

Distribuciones condicionales en el caso mixto. Versión moderna de la esperanza condicional

Distribuciones condicionales en el caso mixto

Ya hemos definido esperanzas condicionales en el caso general, es decir, en el caso en que las variables aleatorias implicadas en la esperanza sean de cualquier forma es decir: continuas, discretas, absolutamente continuas o ninguna de éstas.

En el caso de las distribuciones condicionales, sólo se han definido en el caso en que las variables aleatorias implicadas son o bien todas absolutamente continuas o bien todas discretas.

En este capítulo se definirán la distribución condicional de una variable aleatoria absolutamente continua dada una discreta, y la distribución de una variable aleatoria discreta dada una absolutamente continua; esto nos permitirá definir una función de densidad de dos variables aleatorias en el caso mixto.

Las propiedades que cumple la esperanza condicional para variables aleatorias discretas y para variables aleatorias absolutamente continuas tambien se cumplen en el caso mixto y sus demostraciones no se harán pues lo importante en este capítulo es definir una función de densidad en el caso mixto, y ejemplificar la esperanza condicional en este caso.

Si tenemos una variable aleatoria discreta X y una variable absolutamente continua Y, se puede definir la densidad condicional $f_{X|Y}(x,y)$ como la siguiente probabilidad $P(X = x \mid Y = y)$. Para calcular esta probabilidad condicional, no podemos usar la función de densidad conjunta de X y Y pues ésta ni siquiera está definida en este caso.

Una forma de hacerlo es usando a la esperanza condicional de la función indicadora en $X = x$ dado que $Y = y$, pues ésta si está bien definida. Así, si

$$I_{X=x} = \begin{cases} 1 & \text{si } X = x \\ 0 & \text{si } X \neq x \end{cases}$$

se define

$$P(X = x \mid Y = y) = E(I_{X=x} \mid Y = y)$$

De acuerdo a esto se puede definir la densidad condicional de X dado que $Y = y$, de la siguiente forma:

Definición 4.1. Sea X una variable aleatoria discreta y sea Y una variable aleatoria absolutamente continua. Se define la función de densidad condicional de X dado que $Y = y$, para toda $x \in \mathbb{R}$, como:

$$f_{X|Y}(x \mid y) = h(y)$$

donde $h(Y)$ es una versión de la esperanza condicional de $I_{X=x}$ dada la variable aleatoria Y.

Obsérvese que si $P(X = x) = 0$, entonces una versión de la esperanza condicional de $I_{X=x}$, dada la variable aleatoria Y, es la función idénticamente cero; así que, en este caso, tomaremos para h a la función idénticamente cero.

Además, en cualquier caso, siendo $I_{X=x}$ una función no negativa, podemos tomar h como una función no negativa.

Utilizando la definición anterior se puede hacer la siguiente definición:

Definición 4.2. Sea X una variable aleatoria discreta y sea Y una variable aleatoria absolutamente continua. Se define la función de densidad condicional de Y dado que $X = x$, para toda $y \in \mathbb{R}$, como:

$$f_{Y|X}(y \mid x) = \begin{cases} \frac{f_{X|Y}(x|y) f_Y(y)}{f_X(x)} & f_X(x) > 0 \\ 0 & e.o.c. \end{cases}$$

Proposición 4.3. Si $f_X(x) > 0$, entonces la función $y \longrightarrow f_{Y|X}(y \mid x)$ es una función de densidad.

Demostración. Por la definición, $f_{Y|X}(y \mid x)$ es una función no negativa, así que entonces únicamente resta probar que $\int_{-\infty}^{\infty} f_{Y|X}(y \mid x)dy = 1$.

$\int_{-\infty}^{\infty} f_{Y|X}(y \mid x)dy = \int_{-\infty}^{\infty} \frac{f_{X|Y}(x|y) f_Y(y)}{f_X(x)} dy = \frac{1}{f_X(x)} \int_{-\infty}^{\infty} f_{X|Y}(x \mid y) f_Y(y)dy$

$= \frac{1}{f_X(x)} E[E(I_{X=x} \mid Y)] = \frac{1}{f_X(x)} E(I_{X=x}) = \frac{1}{f_X(x)} P[X = x] = 1$ ■

si $f_X(x) > 0$ se tiene que

$$f_{Y|X}(y \mid x) f_X(x) = \frac{f_{X|Y}(x \mid y) f_Y(y) f_X(x)}{f_X(x)} = f_{X|Y}(x \mid y) f_Y(y)$$

En el caso en que $f_X(x) = 0$, se tiene $f_{Y|X}(y \mid x) = f_{X|Y}(x \mid y) = 0$ para cualquier $y \in \mathbb{R}$, así que también en este caso se tiene $f_{Y|X}(y \mid x) f_X(x) = f_{X|Y}(x \mid y) f_Y(y)$.

De acuerdo con el resultado anterior se puede hacer la siguiente definición:

Definición 4.4. Sea X una variable aleatoria discreta y sea Y una variable aleatoria absolutamente continua se define la función de densidad conjunta de X y Y para toda $(x, y) \in \mathbb{R}^2$ como:

$$f_{X,Y}(x,y) = f_{X|Y}(x \mid y) f_Y(y)$$

El siguiente ejemplo es la versión moderna de un problema resuelto por Thomas Bayes en 1763

Ejemplo 4.5. Supongamos que se realizan $n + m$ ensayos de Bernoulli independientes con probabilidad de éxito x en cada uno de ellos. Supongamos también que esta x no se sabe de antemano sino que se elige al azar del intervalo $(0, 1)$. Si se sabe que de los $n + m$ ensayos realizados, n resultaron ser éxitos ¿Cual es la densidad condicional de la probabilidad de éxito?

Sea X: La probabilidad de éxito en los ensayos
Es claro que

$$f_X(x) = \begin{cases} 1 & \text{si } x \in (0,1) \\ 0 & e.o.c. \end{cases}$$

Sea Y: El número de éxitos en $n+m$ ensayos
Se nos pide encontrar $f_{X|Y}(x \mid n)$.

$$f_{X|Y}(x \mid n) = \frac{f_{Y|X}(y \mid x) f_X(x)}{f_Y(n)} = \frac{\binom{n+m}{n} x^n (1-x)^m}{P(Y=n)}$$

dado que la probabilidad de éxito es x, la densidad de Y es binomial con parámetros $n+m, x$.

Calculando por separado a $P(Y=y)$ tenemos que:

$$\begin{aligned} P(Y=n) &= E(E(I_{Y=n} \mid X)) = \int_{-\infty}^{\infty} E(I_{Y=y} \mid X=x) f_X(x)dx \\ &= \int_{-\infty}^{\infty} P(Y=n \mid X=x) f_X(x)dx \\ &= \int_0^1 \binom{n+m}{n} x^n (1-x)^m dx = \binom{n+m}{n} \int_0^1 x^n (1-x)^m dx \end{aligned}$$

Evaluando $\int_0^1 x^n(1-x)^m dx$ por partes tenemos que:
Sea $u = x^n$ y sea $dv = (1-x)^m dx$, entonces $du = nx^{n-1}dx$ y $v = \frac{(1-x)^{m+1}}{m+1}$, así que

$$\int_0^1 x^n(1-x)^m dx = x^n \frac{(1-x)^{m+1}}{m+1} \Big|_0^1 - \int_0^1 nx^{n-1} \frac{(1-x)^{m+1}}{m+1} dx = \int_0^1 nx^{n-1} \frac{(1-x)^{m+1}}{m+1} dx$$

y haciendo este procedimiento n veces se tiene que

$$\begin{aligned} \int_0^1 x^n(1-x)^m dx &= \frac{n}{m+1} \frac{n-1}{m+2} \frac{n-2}{m+3} \cdots \frac{2}{m+n-1} \frac{1}{m+n} \int_0^1 (1-x)^{m+n} dx \\ &= \frac{n}{m+1} \frac{n-1}{m+2} \frac{n-2}{m+3} \cdots \frac{2}{m+n-1} \frac{1}{m+n} \frac{1}{m+n+1} \\ &= \frac{n}{m+1} \frac{n-1}{m+2} \frac{n-2}{m+3} \cdots \frac{2}{m+n-1} \frac{1}{m+n} \frac{1}{m+n+1} = \frac{n!m!}{(n+m+1)!} \end{aligned}$$

Así que entonces

$$P(Y=n) = \binom{n+m}{n} \int_0^1 x^n(1-x)^m dx = \frac{(n+m)!}{n!m!} \frac{n!m!}{(n+m+1)!} = \frac{1}{m+n+1}$$

y por lo tanto

$$\begin{aligned} f_{X|Y}(x \mid n) &= \frac{\binom{n+m}{n} x^n(1-x)^m}{P(Y=n)} \\ &= \frac{\binom{n+m}{n} x^n(1-x)^m}{\frac{1}{m+n+1}} = (m+n+1)\binom{n+m}{n} x^n(1-x)^m \end{aligned}$$

En conclusión: La densidad condicional de X dado que n ensayos resultaron ser éxitos es una densidad beta con parámetros $n+1$ y $m+1$.

Este ejemplo nos da una interpretación muy concreta de la densidad beta o dicho de otra forma se puede motivar la densidad beta con este problema.

Si de antemano se supiera cuál de los n ensayos fueron los éxitos la densidad condicional de X dado Y no cambiaría pues lo único que pasaría sería que en (1) desaparecerían ($\binom{n+m}{n}$) del numerador y del denominador.

Ejemplo 4.6. Sea Y una variable aleatoria con uniforme en el intervalo (0, 1) y supongamos que, para cada valor y de Y, X es una variable aleatoria con distribución geométrica de parámetro y. Para $x \in \{0, 1, \ldots\}$, encuentre e identifique la distribución condicional de Y dado que $X = x$.

Para $x \in \{0, 1, \ldots\}$ se tiene,

$$P[X = x] = E(E(I_{X=x} \mid Y)) = \int_{-\infty}^{\infty} E(I_{X=x} \mid Y = y) f_Y(y)dy = \int_{-\infty}^{\infty} P[X = x \mid Y = y] f_Y(y)dy$$

$$= \int_0^1 y(1-y)^x dy = \frac{x!}{(x+2)!} = \frac{1}{(x+1)(x+2)}$$

Para $x \in \{0, 1, \ldots\}$ y $0 < y < 1$, se tiene,

$$f_{Y|X}(y \mid x) = \frac{f_{X|Y}(x \mid y) f_Y(y)}{f_X(x)} = (x+1)(x+2)y(1-y)^x$$

Es decir, dado que $X = x$, Y tiene distribución beta de parámetros 2 y $x+1$.

Versión Moderna de la Esperanza Condicional

Hasta aquí se ha desarrollado la esperanza condicional de una variable aleatoria, de dos variables aleatorias y de una función de dos variables aleatorias dada otra variable aleatoria. Si se quisiera desarrollar ahora la esperanza condicional de una función de n variables aleatorias condicionando con m variables aleatorias, se podría hacer de la misma forma en que se hizo lo anterior. Es claro que se necesitaría hablar de una función de densidad conjunta de esas $n+m$ variables aleatorias, o bien tratar de generalizar a este caso la definición que se dio para el caso de la esperanza condicional de una variable aleatoria dada otra, en el caso en que las variables aleatorias sean cualesquiera. Sin embargo, si quisiéramos definir la esperanza condicional de una función de n variables aleatorias condicionando con una infinidad de variables aleatorias, ¿qué haríamos?

Por problemas como éste es que **la Teoría de la Medida** se ha vuelto una herramienta indispensable en el estudio de la esperanza condicional, en especial la **integral de Lebesgue**.

Por lo anterior, a continuación se expone la herramienta necesaria de la teoría de la medida para llegar a demostrar que se puede definir en general la esperanza condicional, es decir condicionando incluso con una infinidad de variables aleatorias, o dicho de otra forma con la $\sigma - \acute{a}lgebra$ generada por éstas.

Esperanza de una variable aleatoria

Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad consistente en un espacio muestral Ω, una $\sigma - \acute{a}lgebra$ de subconjuntos de Ω, y una función de probabilidad $P : \mathcal{A} \longrightarrow [0,1]$

La esperanza de la variable aleatoria X se define como $E(X) = \int_{\Omega} XdP$

Donde $\int_{\Omega} XdP$ es la integral de Lebesgue de X con respecto a la medida de probabilidad P.

La integral de Lebesgue.

Lo que a continuación se expone, esta tratado en el caso particular de nuestro espacio de probabilidad, $(\Omega, \mathcal{A}, P)$.

Definición 4.7. Se dice que una variable aleatoria $\varphi : \Omega \to \mathbb{R}$ es **simple** si se puede representar en la forma:

$\varphi = \sum_{j=1}^{n} a_j I_{A_j}$ donde $A_i \cap A_j = \emptyset$ para toda $i \neq j$ y $a_j \in \mathbb{R}$, siendo todos los a_j distintos. A_i eventos, tales que $\bigcup_{i=1}^{n} A_i = \Omega$.

Es decir , una variable aleatoria simple toma solamente un número finito de valores, y además tiene una representación única debido a que los a_j son todos distintos y $A_i \cap A_j = \emptyset$ y tales que $\bigcup_{i=1}^{n} A_i = \Omega$

Definición 4.8. Si φ es una variable aleatoria simple se define la integral de φ con respecto a P como $\int \varphi dP = \sum_{j=1}^{n} a_j P(A_j)$

Hay que notar que esta definición no permite que las integrales sean $-\infty$ o $+\infty$.

Definición 4.9. Si X es una variable aleatoria que toma solamente valores no negativos, se define la integral de X con respecto a P como

$$\int XdP = \sup \int \varphi dP = \sup \sum_{j=1}^{n} a_j P(A_j)$$

Donde el supremo de $\int \varphi dP$ se toma sobre todas las funciones φ tales que $0 \leq \varphi(\omega) \leq X(\omega)$ para toda $\omega \in \Omega$.

Debidoa que la función $\varphi(\omega) = 0$ para toda $\omega \in \Omega$, es simple y cumple que $\varphi(\omega) \leq X(\omega)$ para toda $\omega \in \Omega$ se puede garantizar que el conjunto $\int \varphi dP$ no es vacío.

Ejemplo 4.10.

Se elige un punto al azar en el intervalo $[0,1)$. en este caso $\Omega = [0,1)$, $\mathcal{A}$ es la $\sigma - algebra$ de Borel (la generada por todos los intervalos) Sea X la variable aleatoria definida como

$$X(\omega) = \begin{cases} 1 & \text{si } \omega \text{ es racional de } (0,1) \\ 0 & \text{si } \omega \text{ es irracional de } (0,1) \end{cases}$$

X es variable aleatoria pues para cualquier $E \subseteq \mathbb{R}$ se tiene que $X^{-1}(E) \in \{\Omega, \Omega \cap \mathbb{Q}, \Omega \cap \mathbb{I}, \emptyset\} \subseteq \mathcal{A}$

La esperanza de X se puede calcular facilmente usando integral de Lebesgue. X es una función simple pues es de la forma $\sum_{j=1}^{1} I_{\Omega \cap \mathbb{Q}}$ y por lo tanto

$\int_{\Omega} XdP = \int I_{\Omega \cap \mathbb{Q}} dP = P[\Omega \cap \mathbb{Q}] = 0$

Definición 4.11. Se define a la parte positiva (negativa) de cualquier variable aleatoria X como X^{+} (X^{-}) como

$$X^{+} = \begin{cases} X & \text{si } X \geq 0 \\ 0 & e.o.c. \end{cases}$$

$$X^{-} = \begin{cases} X & \text{si } X \leq 0 \\ 0 & e.o.c. \end{cases}$$

Definición 4.12. Si X es una variable aleatoria que cumple que $\int X^{+}dP$ y $\int X^{-}dP$ son finitas, se define la integral de X con respecto a P como

$$\int XdP = \int X^{+}dP - \int X^{-}dP$$

En adelante siempre que se hable de conjunto medible, nos estaremos refiriendo a un elemento de $\mathcal{A}$ a menos que se indique otra cosa.

Dada una variable aleatoria $X : \Omega \longrightarrow \mathbb{R}$, la familia de conjuntos

$$B_\alpha = \{\omega \in \Omega \mid X(\omega) \leq \alpha\} = X^{-1}[(-\infty, \alpha]]$$

es una familia de eventos. Dicha familia es creciente en el sentido de que si $\alpha < \beta$ entonces $B_\alpha \subseteq B_\beta$. Siempre que tengamos una variable aleatoria X, tendremos una familia creciente de subconjuntos de Ω ,$\{B_\alpha\}$, con $\alpha \in \mathbb{R}$.

Una pregunta que nos podemos hacer es si sucederá lo recíproco: Dada una familia creciente de eventos de de Ω, es decir de elementos de $\mathcal{A}, \{B_\alpha\}$, con $\alpha \in \mathbb{R}$ ¿existirá una variable aleatoria $X : \Omega \longrightarrow \bar{\mathbb{R}}$ tal que $B_\alpha = X^{-1}[(-\infty, \alpha]]$ para toda $\alpha \in \mathbb{R}$? La respuesta a esta pregunta es no necesariamente, pero casi. Veamos.

Una variable aleatoria es primero que nada una función que va de Ω a $\bar{\mathbb{R}}$. Como tal, tendría que cumplir que cuando $\omega \in B_\alpha$, $X(\omega) \in (-\infty, \alpha]$, es decir que $X(\omega) \leq \alpha$.

La igualdad anterior permanece válida fijando ω y variando α.

Como $X(\omega) \leq \alpha$ para toda α tal que $\omega \in B_\alpha$ entonces $X(\omega) \leq \inf\{\alpha \mid \omega \in B_\alpha\}$

¿Qué sucedería si $X(\omega) < \inf\{\alpha \mid \omega \in B_\alpha\}$?

Pasaría que podría existir un número real β tal que $\omega \notin B_\beta$ y $X(\omega) < \beta < \inf\{\alpha \mid \omega \in B_\alpha\}$ pero si $X(\omega) < \beta$ entonces $X(\omega) \leq \beta$ y por lo tanto $\omega \in X^{-1}[(-\infty, \beta]]$ y esto implicaría que

$$B_\beta \neq X^{-1}[(-\infty, \beta]]$$

esto es una contradicción con lo que se requiere de la variable X.

Por lo anterior, si queremos una variable aleatoria X tal que $X^{-1}[(-\infty, \alpha]]$ sea un elemento de $\{B_\alpha\}$ entonces X tendría que ser tal que $X(\omega) = \inf\{\alpha \mid \omega \in B_\alpha\}$. Pero, recordemos que una variable aleatoria definida sobre el espacio de probabilidad $(\Omega, \mathcal{A}, P)$, es una función $X : \Omega \longrightarrow \bar{\mathbb{R}}$ tal que la imagen inversa de cualquier conjunto de la forma $[X \leq x]$ $x \in \mathbb{R}$ está en $\mathcal{A}$. El siguiente resultado nos demuestra que en efecto X definida como $\inf\{\alpha \mid \omega \in B_\alpha\}$ es una variable aleatoria con $\alpha \in D \subseteq \mathbb{R}$ D numerable.

Proposición de existencia. Sea D un conjunto numerable de números reales, $\{B_\alpha\}$ con $\alpha \in D$, una familia de eventos tales que si $\alpha < \beta$ entonces $B_\alpha \subseteq B_\beta$, con $\alpha, \beta \in D$, entonces existe una variable aleatoria $X : \Omega \longrightarrow \bar{\mathbb{R}}$ tal que
$\{\omega \in \Omega \mid X(\omega) < \alpha\} \subseteq B_\alpha \subseteq \{\omega \in \Omega \mid X(\omega) \leq \alpha\}$ para cualquier $\alpha \in D$.

Demostración: Para cada $\omega \in \Omega$, sea $X(\omega) = \inf\{\alpha \in D \mid \omega \in B_\alpha\}$

Se tiene que $B_\alpha \subseteq \{\omega \in \Omega \mid X(\omega) \leq \alpha\}$ para cualquier $\alpha \in D$ debido a lo siguiente:

Sea $\omega_0 \in B_{\alpha_0}, \alpha_0 \in D$ entonces
$\inf\{\alpha \in D \mid \omega_0 \in B_\alpha\} \leq \alpha_0$ eso implica que
$X(\omega_0) \leq \alpha_0$ es decir
$\omega_0 \in \{\omega \in \Omega \mid X(\omega) \leq \alpha_0\}$ pero esto es valido para cualquier $\omega_0 \in B_{\alpha_0}$
así que $B_\alpha \subseteq \{\omega \in \Omega \mid X(\omega) \leq \alpha\}$

Para ver que X es una variable aleatoria vamos a probar que la imagen inversa de cualquier intervalo de la forma $(-\infty, \alpha)$ pertenece a $\mathcal{A}$. Para esto basta con probar que $X^{-1}[(-\infty, \alpha)]$ es unión a lo mas numerable de elementos de $\mathcal{A}$. Y como por hipótesis, los B_α con $\alpha \in D$ son elementos de $\mathcal{A}$, entonces es suficiente con mostrar que $X^{-1}[(-\infty, \alpha)]$ es unión de algunos B_α.

Sea $\alpha \in \mathbb{R}$

Tenemos que demostrar que $\{\omega \in \Omega \mid X(\omega) < \alpha\}$ es un evento

caso 1) Sea $\{\omega \in \Omega \mid X(\omega) < \alpha\} = \emptyset$, el conjunto vacío es elemento de $\mathcal{A}$.

caso 2) Sea $\{\omega \in \Omega \mid X(\omega) < \alpha\} \neq \emptyset$ entonces existe $\beta \in D$ tal que $\beta < \alpha$ y $\omega_0 \in B_\beta$ debido a lo siguiente: Considerese el conjunto de $\{\beta \in D \mid \omega_0 \in B_\beta\}$ si dicho conjunto fuera vacio, $X(\omega_0) = \inf\emptyset = \infty$ entonces no se cumpliría que $X(\omega_0) < \alpha$ por lo tanto dicho conjunto no es vacío.

Por otro lado $X(\omega_0) = \inf\{\beta \in D \mid \omega_0 \in B_\beta\} < \alpha$

Como el ínfimo de cualquier conjunto es la máxima cota inferior del conjunto y α es mayor que éste, entonces α no puede ser cota inferior de $\{\beta \in D \mid \omega_0 \in B_\beta\}$ Entonces hay elementos del conjunto menores que α. Es decir, existe $\beta \in D$ tal que $\omega_0 \in B_\beta$ y además $\beta < \alpha$.

En resumen, para todo $\omega_0 \in \{\omega \in \Omega \mid X(\omega) < \alpha\}$ existe algún $\beta \in D$ tal que $\omega_0 \in B_\beta$ y $\beta < \alpha$.

De aquí se sigue que $\{\omega \in \Omega \mid X(\omega) < \alpha\} \subseteq \bigcup_{\{\beta \in D, \beta < \alpha.\}} B_\beta$

Recíprocamente si $\omega \in \bigcup_{\{\beta \in D, \beta < \alpha.\}} B_\beta$
entonces $\omega \in B_\beta$ para alguna $\beta \in D, \beta < \alpha$
entonces $X(\omega) = \inf\{\beta \in D \mid \omega \in B_\beta\} \leq \beta$ y por transitividad $X(\omega) < \alpha$ y por lo tanto $\omega \in \{\omega \in \Omega \mid X(\omega) < \alpha\}$ y en conclusión
$\{\omega \in \Omega \mid X(\omega) < \alpha\} = \bigcup_{\{\beta \in D, \beta < \alpha.\}} B_\beta$. Es decir $\{\omega \in \Omega \mid X(\omega) < \alpha\}$ es una unión numerable de eventos y por lo tanto X es una variable aleatoria.

Nos falta probar que $\{\omega \in \Omega \mid X(\omega) < \alpha\} \subseteq B_\alpha$

Tenemos que $\{\omega \in \Omega \mid X(\omega) < \alpha\} = \bigcup_{\{\beta \in D, \beta < \alpha.\}} B_\beta$ pero como $\beta < \alpha$ entonces $B_\beta < B_\alpha$ y por lo tanto
$\bigcup_{\{\beta \in D, \beta < \alpha.\}} B_\beta \subseteq B_\alpha$ de donde $\{\omega \in \Omega \mid X(\omega) < \alpha\} \subseteq B_\alpha$ ■

Recordando la pregunta que nos hicimos:¿Existirá una variable aleatoria $X : \Omega \longrightarrow \mathbb{R}$ tal que $B_\alpha = X^{-1}[(-\infty, \alpha]]$ para toda $\alpha \in \mathbb{R}$? La proposición anterior responde a esto, pero no para cualquier α de los reales, sino con $\alpha \in D$ donde D es un conjunto numerable,.

La siguiente proposición junto con el resultado anterior asegura que se cumple para cualquier $\alpha \in \mathbb{R}$. La demostración no se hará aquí, aunque la demostración de que $X(\omega) \leq Y(\omega)$ es inmediata.

Proposición 4.13. Con las hipótesis de la proposición anterior.

Si $X(\omega) = \inf\{\alpha \in \mathbb{Q} \mid \omega \in B_\alpha\}$ y $Y(\omega) = \inf\{\alpha \in \mathbb{R} \mid \omega \in B_\alpha\}$ entonces $X(\omega) = Y(\omega)$

Donde $\mathbb{Q}$ son los numeros racionales de $\mathbb{R}$

Definición 4.14. Se dice que $\mu : \mathcal{A}. \to \bar{\mathbb{R}}$ es una **medida** si se cumplen las siguientes condiciones.

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

ii) $\mu(E) \geq 0$ Para cualquier $E \in \mathcal{A}$

iii) μ es $\sigma - aditiva$

[Es decir que si $A_1, A_2, \ldots$ son mutuamente excluyentes, entonces $\mu(\bigcup_i A_i) = \sum_i \mu(A_i)$]

Ejemplo 4.15. una medida de probabilidad P es una medida.

Definición 4.16 Se dice que $\upsilon : \mathcal{A}. \to \bar{\mathbb{R}}$ es una **medida con signo** si se cumplen las siguientes condiciones.

i) $\upsilon(\emptyset) = 0$

ii) υ es $\sigma - aditiva$

[Es decir que si $A_1, A_2, \ldots$ son mutuamente excluyentes, entonces $\upsilon(\bigcup_i A_i) = \sum_i \upsilon(A_i)$]

iii) υ toma a lo mas uno de los valores $-\infty, \infty$.

Ejemplo 4.17. Sea $X : \Omega \longrightarrow \bar{\mathbb{R}}$ una variable aleatoria.

La función $\upsilon : \mathcal{A}. \to \bar{\mathbb{R}}$ que a cada elemento A de $\mathcal{A}$ le asigna la función $\upsilon(A) = \int_A X dP$, la integral de Lebesgue de X sobre A, es una medida con signo.

Ejemplo 4.18. Sean μ_1 y μ_2 dos medidas entonces $\mu_1 - \mu_2$ es una medida con signo

Definición 4.19. Sea υ una medida con signo y $A \in \mathcal{A}$.

Se dice que A es un **conjunto positivo** (resp. conjunto negativo) con respecto a υ si $\upsilon(E) \geq 0$ (resp. ≤ 0) para cualquier evento $E \in A$.

Proposición 4.20. La unión de una colección finita o infinita numerable de conjuntos positivos es un conjunto positivo.

Demostración. Sean $A = \bigcup_n A_n$ con cada A_n conjunto positivo y E un evento tal que $E \subseteq A$.

Queremos demostrar que $\upsilon(E) \geq 0$

Sea $E_n = E \cap A_n \cap A_1^c \cap A_2^c \cap \cdots \cap A_{n-1}^c$

Como $E_n \subseteq A_n$ y $E_{n+k} \subseteq A_n^c$ entonces

$$E_n \cap E_{n+k} = \emptyset$$

en palabras los E_n 's son ajenos por parejas o, dicho en probabilidad, mutuamente excluyentes.

Por otra parte

$$E = E \cap A = E \cap [\bigcup_n A_n] = \bigcup_n (E_n \cap A_n) \supseteq \bigcup_n E_n$$

y como $E_n \subseteq E$ para toda n entonces $\bigcup_n E_n \subseteq E$, de modo que $E = \bigcup_n E_n$

Pero $E_n \subseteq A_n$ y A_n es un conjunto positivo, entonces $\upsilon(E_n) \geq 0$

En conclusión:

$$\upsilon(E) = \upsilon(\cup_n E_n) = \sum_n \upsilon(E_n) \geq 0 \qquad \blacksquare$$

Proposición 4.21. Sea υ una medida con signo, sea E un conjunto medible tal que $0 < \upsilon(E) < \infty$ entonces existe un conjunto positivo $A \subseteq E$ tal que $\upsilon(A) > 0$

La idea de la demostración es quitarle al conjunto E subconjuntos de medida negativa hasta quedarse con un subconjunto positivo de medida positiva.

Antes de empezar la demostración obsérvese que si B es un subconjunto de E de medida negativa, entonces, $E = B \cup (E - B)$ y pasa lo siguiente $0 < \upsilon(B) + \upsilon(E - B) < \infty$. Esto implica que $\upsilon(B)$ y $\upsilon(E - B)$ son finitos pues si alguno de los dos fuera infinito, su suma tambien lo sería. Mas aún $\upsilon(E) - \upsilon(E - B) = \upsilon(B) < 0$.

En resumen $0 < \upsilon(E - B) < \infty$ y $-\infty < \upsilon(B) < 0$. Hechas estas observaciones comencemos la demostración:

Demostración.

Caso 1) Sea E un conjunto positivo. Tomamos $A = E$ y ya está.

Caso 2) E no es un conjunto positivo con respecto a υ. Esto quiere decir que existen conjuntos $B \subseteq E$ tales que $\upsilon(B) < 0$ es decir $\{B \subseteq E \mid \upsilon(B) < 0\} \neq \emptyset$.

Sea n_1 el menor entero para el cual existe un conjunto medible $B \subseteq E$ tal que $\upsilon(B) < -\frac{1}{n_1}$

Sea B_1 un conjunto medible con esa propiedad.

Sea $E_1 = E - B_1$ entonces $0 < \upsilon(E_1) < \infty$, si E_1 es positívo ya acabamos , sino repetimos el proceso.

Sea n_2 el menor entero para el cual existe un conjunto medible $B \subseteq E_1$ tal que $\upsilon(B) < -\frac{1}{n_2}$

Sea B_2 un conjunto medible con la propiedad anterior

Sea $E_2 = E_1 - B_2$ entonces $0 < \upsilon(E_2) < \infty$, si E_1 es positívo ya acabamos , si no, repetimos el proceso.

Hay dos posibilidades:

Posibilidad 1) En un número finito de pasos llegamos a algún $E_n \subseteq E$ tal que $0 < \upsilon(E_n) < \infty$ y E_n positivo y ya se acaba la demostración.

Posibilidad 2) Se genera una sucesión de conjuntos $\{B_k\}$ y una sucesión de enteros positivos $\{n_k\}$ tales que cada B_k es medible y $\upsilon(B_k) < \frac{-1}{n_k}$. En este caso se propone al conjunto $A = E - \cup_{k=1}^{\infty} B_k$ como el conjunto positivo buscado. Claramente $A \subseteq E$. Sea $B = \cup_{k=1}^{\infty} B_k$,

$$\upsilon(B) = \upsilon(\cup_{k=1}^{\infty} B_k) = \sum_{k=1}^{\infty} \upsilon(B_k) < 0$$

pues los B_k son ajenos entre sí y todos de medida negativa. Por la observación previa a la demostración tenemos que $-\infty < \upsilon(\cup_{k=1}^{\infty} B_k) < 0$. Surge la pregunta ¿será positivo el conjunto A ?, vamos a ver .

Sea $G \subseteq A$. Demostremos que $\upsilon(G) \geq 0$

Tenemos que $G \subseteq A = E - \cup_{k=1}^{\infty} B_k \subseteq E - B_1$ entonces

$$G \subseteq E - B_1$$

Como n_1 es el menor entero para el cual existen conjuntos medibles $B \subseteq E$ tales que

$$\upsilon(B) < -\frac{1}{n_1}$$

Para $n_1 - 1$ no hay conjuntos medibles $B \subseteq E$ tales que

$$\upsilon(B) < -\frac{1}{n_1 - 1}$$

Como $G \subseteq E$ y es medible , entonces

$$\upsilon(G) \geq -\frac{1}{n_1 - 1}$$

$$G \subseteq E - \cup_{k=1}^{\infty} B_k \subseteq E - (B_1 \cup B_2) = (E - B_1) - B_2 = E_1 - B_2$$

entonces

$$G \subseteq E_1$$

La forma en que se eligió a n_2 nos garantiza que $\upsilon(G) \geq -\frac{1}{n_k - 1}$ para toda $k \in \mathbb{N}$

Por otro lado, la sucesión $\{n_k\}$ es creciente; pues :

Supongamos que n_{k-1} es menor que n_k

n_{k-1} es un entero para el cual existen conjuntos $B \subseteq E_k$ medibles, tales que $\upsilon(B) < -\frac{1}{n_k - 1}$, pero como $E_k = E_{k-1} - B_k \subseteq E_{k-1}$, entonces los conjuntos $B \subseteq E_k$ medibles, tales que $\upsilon(B) < -\frac{1}{n_k - 1}$, también son tales que

$$B \subseteq E_{k-1}$$

entonces

n_k ya no es el menor entero con la propiedad requerida y por lo tanto $n_k < n_{k-1}$

Sin embargo ningún término de la sucesión $\{n_k\}$ se repite una infinidad de veces. De hecho si algún término se repitiera una infinidad de veces entonces la sucesión se quedaría constante a partir de cierta s es decir $n_s = n_{s-1} = n_{s-2} = \ldots$

Por lo que $\lim_{k \to \infty}\{n_k\} = \infty$ y entonces $\upsilon(G) \geq -\frac{1}{n_k - 1}$ para cualquier $k \in \mathbb{N}$

De lo anterior se sigue que

$$\upsilon(G) \geq 0$$

y por lo tanto A es positivo ■

El siguiente resultado será usado para probar el teorema de Radón Nikodym que a su vez será usado para probar la existencia de la esperanza condicional.

Teorema de la descomposición de Hahn (1920). Sea ν una medida con signo, entonces existe un conjunto positivo A y un conjunto negativo B tales que

$$A \cap B = \emptyset \qquad \text{y} \qquad \Omega = A \cup B$$

La idea de la demostración de este teorema es la siguiente: Si ν no toma el valor ∞, se trata de encontrar un conjunto positivo de medida máxima. El complemento de dicho conjunto será un conjunto negativo. Y si, ν no toma el valor $-\infty$ buscaremos un conjunto negativo de medida mínima y el complemento de dicho conjunto será un conjunto positivo.

Demostración:

Como ν es una medida con signo, no puede tomar el valor $+\infty$ en un conjunto E_1 y el valor $-\infty$ en un conjunto E_2. Supongamos que no toma el valor $+\infty$.

Sea $\lambda = \sup\{\nu(A) \mid A \text{ es un conjunto positivo con respecto a } \nu\}$

Como el conjunto vacío es positivo y $\nu(\emptyset) = 0$, entonces $\lambda \geq 0$

Por definición de supremo, para cada $n \in \mathbb{N}$, existe un conjunto positivo A_n tal que

$$\lambda - \frac{1}{n} < \nu(A_n) \leq \lambda$$

Entonces existe una sucesión de conjuntos positivos $\{A_n\}$ tal que $\lim_{n\to\infty}\{A_n\} = \lambda$

Se proponen a los siguientes conjuntos como los conjuntos de la descomposición de Hahn

$$A = \bigcup_{n=1}^{\infty} A_n \qquad \text{y} \qquad B = A^c$$

Hay que demostrar que A y B cumplen con las condiciones del teorema.

A es un conjunto positivo ya que es la unión numerable de conjuntos positivos. Como A es positivo $\lambda \geq \nu(A)$

pero $A - A_n \subseteq A$ para toda $n \in \mathbb{N}$

Entonces $\nu(A - A_n) \geq 0$ pues A es positivo, esto implica que

$\nu(A) \geq \nu(A_n)$ para toda $n \in \mathbb{N}$ y tomando el limite cuando n tiende a ∞ se tiene que $\nu(A) \geq \lambda$. En conclusión

$$\nu(A) = \lambda$$

Pero como $\nu(E) < \infty$ para cualquier E entonces $\lambda < \infty$.

Para probar que B es negativo, supongamos que no lo es. Entonces existe un conjunto $E \subseteq B$ tal que $\nu(E) > 0$. E cumple la hipótesis de la proposición anterior, por lo tanto existe un conjunto positivo $D \subseteq E$ tal que $\nu(D) > 0$. Pero sucediendo esto λ ya no sería el supremo del conjunto $\{\nu(A) \mid A \text{ es un conjunto positivo con respecto a } \nu\}$

pues

$$\lambda = \nu(A) < \nu(A) + \nu(D) = \nu(A \cup D)$$

con $A \cup D$ conjunto positivo y $A \cap D = \emptyset$ pues $D \subseteq E \subseteq B = A^c$. Por lo tanto B es un conjunto negativo

En conclusión A y B cumplen ser tales que

$$A \cap B = \emptyset \qquad \text{y} \qquad \Omega = A \cup B$$

En el caso de que ν no tome el valor $-\infty$, pero si el valor $+\infty$, aplicamos el razonamiento anterior a $-\nu$ y llegamos a la misma conclusión. ■

En el teorema de la descomposición de Hahn

1). En el caso de que $\nu(E) = \int_E XdP$ un conjunto positivo con respecto a ν sería

$$\{\omega \in \Omega \mid X(\omega) \geq 0\}$$

y un conjunto negativo seria $\{\omega \in \Omega \mid X(\omega) \leq 0\}$

2) Los conjuntos A y B de la descomposición de Hahn, no están determinados de manera única, pues podemos quitar y poner los conjuntos de medida cero, que son positivos y negativos al mismo tiempo. es decir. Si C es un conjunto tal que $\nu(C) = 0$ entonces

$$A' = A - C \qquad \text{y} \qquad B' = B \cup C$$

cumplirían con ser otra descomposición de Hahn.

Definición 4.22 Sea ν una medida con signo. Una pareja de conjuntos medibles (A, B) tales que A es positivo
y B es negativo

$$A \cap B = \emptyset \qquad \text{y} \qquad \Omega = A \cup B$$

. se le llama una **descomposición de Hahn** para ν.

Definición 4.23. Sean P y ν dos **medidas**. Se dice que ν es **abolutamente continua** con respecto a P si $\nu(E) = 0$ para cualquier conjunto medible E tal que $P(E) = 0$. Se denota $\nu << P$.

Ejemplo 4.24. Si P es la medida del espacio de probabilidad $(\Omega, \mathcal{A}, P)$ y $X : \Omega \to [0, \infty)$ una variable aleatoria no negativa, entonces $\nu : \mathcal{A} \to \mathbb{R}$ definida por $\nu(E) = \int_E XdP$ es una medida absolutamnete continua con respecto a P.

El concepto de continuidad absoluta de una medida con respecto a otra medida es importante porque es la hipótesis fundamental del teorema de Radon Nikodym

Teorema de Radom Nikodym.(1930) (Adaptación libre)

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

Sea $\nu : \mathcal{A} \to \mathbb{R}$ una medida absolutamente continua con respecto a P, entonces existe una variable aleatoria $X : \Omega \to \mathbb{R}$ no negativa tal que

$$\nu(E) = \int_E XdP$$

para cualquier evento E.

La idea de la demostración es construir una familia de eventos, que cumpla las hipótesis del lema que nos garantiza la existencia de una variable aleatoria, X, y después probar que cumple las propiedades requeridas.

Demostración:

Sea $\nu_\alpha = \nu - \alpha P$ una medida con signo para cada $\alpha \in \mathbb{Q}$.

Sea (A_α, B_α) una descomposición de Hahn para ν_α, donde A_α es positivo y B_α es negativo $A_\alpha \cap B_\alpha = \emptyset$ y $A_\alpha \cup B_\alpha = \Omega$

En el caso en que $\alpha = 0$ se puede tomar $A_\alpha = \Omega$, $B_\alpha = \emptyset$

Si $\alpha < \beta$ se tiene que

$$B_\alpha - B_\beta = B_\alpha \cap A_\beta$$

así que

$$B_\alpha - B_\beta \subseteq B_\alpha \quad \text{y} \quad B_\alpha - B_\beta \subseteq A_\beta$$

Como $B_\alpha - B_\beta \subseteq B_\alpha$ y B_α es negativo para ν_α entonces

$$\nu_\alpha(B_\alpha - B_\beta) \leq 0$$

de aquí que

$$\nu(B_\alpha - B_\beta) - \alpha P(B_\alpha - B_\beta) \leq 0 \quad \text{por definición de } \nu_\alpha$$

entonces

$$\nu(B_\alpha - B_\beta) \leq \alpha P(B_\alpha - B_\beta) \qquad (1)$$

Por otra parte, como $B_\alpha - B_\beta \subseteq A_\beta$ y A_β es positivo para ν_β entonces

$$\nu_\beta(B_\alpha - B_\beta) \geq 0$$

de aquí que

$$\nu(B_\alpha - B_\beta) - \beta P(B_\alpha - B_\beta) \geq 0 \quad \text{por definición de } \nu_\beta$$

entonces

$$\nu(B_\alpha - B_\beta) \geq \beta P(B_\alpha - B_\beta) \qquad (2)$$

De (1) y de (2) se sigue que

$$\beta P(B_\alpha - B_\beta) \leq \alpha P(B_\alpha - B_\beta)$$

Pero por hipótesis $\alpha < \beta$ entonces $P(B_\alpha - B_\beta) = 0$

Intuitivamente esto quiere decir que la probabilidad del evento B_β es mayor o igual que la probabilidad del evento B_α cuando $\alpha < \beta$ ya que:

$$
\begin{aligned}
P(B_\alpha) &= P\left(B_\alpha \cap B_\beta^c \cup B_\alpha \cap B_\beta\right) \\
&= P\left(B_\alpha \cap B_\beta^c\right) + P(B_\alpha \cap B_\beta) \\
&= P(B_\alpha - B_\beta) + P(B_\alpha \cap B_\beta) \le P(B_\beta)
\end{aligned}
$$

O sea que si tomamos una sucesión creciente de números racionales $\alpha_1 < \alpha_2 < \alpha_3 < \ldots$ entonces la sucesión de eventos $\beta_{\alpha_1}, \beta_{\alpha_2}, \beta_{\alpha_3}, \ldots$ será creciente desde el punto de vista de la probabilidad

$$P(\beta_{\alpha_1}) \le P(\beta_{\alpha_2}) \le P(\beta_{\alpha_3}) \le \ldots$$

(Aunque no necesariamente es creciente desde el punto de vista de la contención.)

Sea $F = \bigcup_{\{\alpha,\beta \in \mathbb{Q} \mid \alpha<\beta\}} (B_\alpha - B_\beta)$ entonces $P(F) = 0$ ya quer F es la unión numerable de eventos de probabilidad cero.

Por otra parte $B_a - F \le B_b - F$ si a y b son racionales con $a < b$ ya que

$$
\begin{aligned}
B_a - F &= \{\omega \in B_a \mid \omega \notin F\} \\
&= \{\omega \in B_a \mid \omega \notin \textstyle\bigcup_{\{\alpha,\beta \in \mathbb{Q} \mid \alpha<\beta\}} (B_\alpha - B_\beta)\}
\end{aligned}
$$

Si $\omega \in B_a - F$ entonces ω no puede estar en la unión de los $B_\alpha - B_\beta$, en particular ω no puede estar en $B_a - B_b$ y por lo tanto ω tiene que estar en $B_a \cap B_b, \omega \in B_b - F$.

Sean

$$B_\alpha{}' = B_\alpha - F \quad \text{y} \quad A_\alpha{}' = (B_\alpha{}')^c \quad \text{para cada } \alpha \in \mathbb{Q}$$

entonces B_α y $B_\alpha{}'$ se diferencian por un conjunto de probabilidad cero que es $B_\alpha \cap F$ y también A_α y $A_\alpha{}'$ se diferencian por un conjunto de probabilidad cero que es $B_\alpha \cap F$ también.

No solo se tiene que $P(B_\alpha \cap F) = 0$ sino que también se tiene que

$$\nu(B_\alpha \cap F) = 0$$

utilizando la hipótesis de que $\nu << P$, es decir que ν es absolutamente continua con respecto a P.

Esto implica que $\left(A_\alpha{}', B_\alpha{}'\right)$ es otra descomposición de Hahn para ν_α puesto que :

1) $A_\alpha{}'$ es positivo con respecto a ν_α. Pues si

$$
\begin{aligned}
&E \subseteq A_\alpha{}' \\
&E \subseteq A_\alpha \cup (B_\alpha \cap F)
\end{aligned}
$$

así que

$$E = E \cap A_\alpha \cup E \cap (B_\alpha \cap F)$$

$$\nu_\alpha(E) = \nu_\alpha(E \cap A_\alpha) + \nu_\alpha(E \cap (B_\alpha \cap F))$$

así que

$$\nu_\alpha(E) \geq 0 \text{ porque } \nu_\alpha(E \cap (B_\alpha \cap F)) = 0 \text{ y } \nu_\alpha(E \cap A_\alpha) \geq 0 \text{ ya que } A_\alpha \geq 0$$

2) $B_\alpha{}'$ es negativo con respecto a ν_α. Pues si

$$E \subseteq B_\alpha{}'$$
$$E \subseteq B_\alpha - F$$

así que

$$E \subseteq B_\alpha$$

$$\nu_\alpha(E) \leq 0 \text{ porque } B_\alpha \text{ es negativo con respecto a } \nu_\alpha$$

3) $A_\alpha{}' \cap B_\alpha{}' = \emptyset$ y $A_\alpha{}' \cup B_\alpha{}' = \Omega$ pues $A_\alpha{}' = (B_\alpha{}')^c$

De hecho la familia $\{(A_\alpha{}', B_\alpha{}') \mid \alpha \in \mathbb{Q}\}$ cumple que cada $(A_\alpha{}', B_\alpha{}')$ es una descomposición de Hahn para ν_α, pero tambien cumple que si $\alpha < \beta$ entonces $P(B_\alpha{}' - B_\beta{}') = 0$ y entonces $P(B_\alpha{}') \leq P(B_\beta{}')$, así que

$$B_\alpha{}' \subseteq B_\beta{}' \text{ es creciente en el sentido de la contención.}$$

Para simplificar la notación, vamos a olvidar la familia original de descomposiciones de Hahn y vamos a trabajar de aquí en adelante solamente con la nueva familia. De aquí en adelante la pareja (A_α, B_α) denotará a la pareja que anteriormente denotabamos por $(A_\alpha{}', B_\alpha{}')$. Con esta nueva notación, la familia $\{B\alpha\}_{\alpha\in\mathbb{Q}}$ es creciente.

Aplicando la proposición de existencia a esta familia, podemos concluir que existe una variable aleatoria $X : \Omega \to \overline{\mathbb{R}}$ tal que $\{\omega \in \Omega \mid X(\omega) < \alpha\} \subseteq B_\alpha \subseteq \{\omega \in \Omega \mid X(\omega) \leq \alpha\}$ para toda $\alpha \in \mathbb{Q}$

Como para $\alpha = 0$ la pareja original $(\Omega, \emptyset)$ queda intacta bajo la sustitución $B_0{}' = B_0 - F$ y $A_0{}' = (B_0{}')^c$; podemos decir que $A_0 = \Omega, B_0 = \emptyset$ esto implica que

$$\{\omega \in \Omega \mid X(\omega) < 0\} \subseteq \emptyset \subseteq \{\omega \in \Omega \mid X(\omega) \leq 0\}$$

así que

$$\{\omega \in \Omega \mid X(\omega) < 0\} = \emptyset$$

es decir $X(\omega) \geq 0$ para todo $\omega \in \Omega$, i.e. X es una variable aleatoria no negativa.

Lo que nos falta por demostrar es que $\nu(E) = \int_E XdP$ para todo $E \in \mathcal{A}$

Sean α y $\beta \in \mathbb{Q}$ con $\alpha < \beta$. Se tiene que

$$\alpha \leq X(\omega) \leq \beta \text{ para todo } \omega \in B_\beta - B_\alpha$$

Debido a que $\omega \in B_\beta$ y $B_\beta \subseteq \{\omega \in \Omega \mid X(\omega) \leq \beta\}$ entonces

$$X(\omega) \leq \beta$$

Y por otra parte $\omega \notin B_\alpha$ implica que

$$\omega \in (B_\alpha)^c \subseteq \{\omega \in \Omega \mid X(\omega) < \alpha\}^c = \{\omega \in \Omega \mid X(\omega) \geq \alpha\}$$

lo cual implica que

$$X(\omega) \geq \alpha$$

Juntando ambas desigualdades se obtiene que

$$\alpha \leq X(\omega) \leq \beta \text{ para todo } \omega \in B_\beta - B_\alpha$$

Sea ahora $E \in \mathcal{A}$ un evento cualquiera. Sea $E_{\alpha\beta} = E \cap (B_\beta - B_\alpha)$. De aquí se obtiene que

$$\int_{E_{\alpha\beta}} \alpha dP \leq \int_{E_{\alpha\beta}} X dP \leq \int_{E_{\alpha\beta}} \beta dP$$

Es decir

$$\alpha P(E_{\alpha\beta}) \leq \int_{E_{\alpha\beta}} X dP \leq \beta P(E_{\alpha\beta}) \tag{2}$$

Por otra parte, como

$$E_{\alpha\beta} \subseteq B_\beta - B_\alpha = B_\beta \cap A_\alpha$$

entonces

$$\nu_\alpha(E_{\alpha\beta}) \geq 0 \text{ porque } A_\alpha \text{ es positivo con respecto a } \nu_\alpha \text{ y}$$

$$\nu_\beta(E_{\alpha\beta}) \leq 0 \text{ porque } B_\beta \text{ es negativo con respecto a } \nu_\beta$$

Sustituyendo las definiciones de ν_α y ν_β se obtiene

$$\nu(E_{\alpha\beta}) - \alpha P(E_{\alpha\beta}) \geq 0 \text{ y}$$

$$\nu(E_{\alpha\beta}) - \beta P(E_{\alpha\beta}) \leq 0$$

entonces

$$\nu(E_{\alpha\beta}) \geq \alpha P(E_{\alpha\beta}) \text{ y}$$

$$\nu(E_{\alpha\beta}) \leq \beta P(E_{\alpha\beta})$$

De donde

$$\alpha P(E_{\alpha\beta}) \leq \nu(E_{\alpha\beta}) \leq \beta P(E_{\alpha\beta}) \tag{3}$$

Combinando las desigualdades (2) y (3) para obtener que

$$\nu(E_{\alpha\beta}) - (\beta - \alpha)P(E_{\alpha\beta}) \leq \int_{E_{\alpha\beta}} X dP \leq \nu(E_{\alpha\beta}) + (\beta - \alpha)P(E_{\alpha\beta})$$

Debido a lo siguiente:
De (2) se sigue que $\alpha P(E_{\alpha\beta}) \leq \int_{E_{\alpha\beta}} X dP$ que es lo mismo que $0 \leq \int_{E_{\alpha\beta}} X dP - \alpha P(E_{\alpha\beta})$

De (3) se sigue que $\nu(E_{\alpha\beta}) \leq \beta P(E_{\alpha\beta})$ lo que implica que $\nu(E_{\alpha\beta}) - \beta P(E_{\alpha\beta}) \leq 0$ por lo tanto

$$\nu(E_{\alpha\beta}) - \beta P(E_{\alpha\beta}) \leq \int_{E_{\alpha\beta}} XdP - \alpha P(E_{\alpha\beta})$$

así que

$$\nu(E_{\alpha\beta}) - \beta P(E_{\alpha\beta}) + \alpha P(E_{\alpha\beta}) \leq \int_{E_{\alpha\beta}} XdP$$

$$\nu(E_{\alpha\beta}) - (\beta - \alpha)P(E_{\alpha\beta}) \leq \int_{E_{\alpha\beta}} XdP$$

Por otra parte:
De (2) se sigue que también que:
$\int_{E_{\alpha\beta}} XdP \leq \beta P(E_{\alpha\beta})$ que es lo mismo que $\int_{E_{\alpha\beta}} XdP - \beta P(E_{\alpha\beta}) \leq 0$
De (3) se sigue que $\alpha P(E_{\alpha\beta}) \leq \nu(E_{\alpha\beta})$ lo que implica que $0 \leq \nu(E_{\alpha\beta}) - \alpha P(E_{\alpha\beta})$ por lo tanto

$$\int_{E_{\alpha\beta}} XdP - \beta P(E_{\alpha\beta}) \leq \nu(E_{\alpha\beta}) - \beta P(E_{\alpha\beta})$$

así que

$$\int_{E_{\alpha\beta}} XdP \leq \nu(E_{\alpha\beta}) - (\beta - \alpha)P(E_{\alpha\beta})$$

así que efectivamente:

$$\nu(E_{\alpha\beta}) - (\beta - \alpha)P(E_{\alpha\beta}) \leq \int_{E_{\alpha\beta}} XdP \leq \nu(E_{\alpha\beta}) + (\beta - \alpha)P(E_{\alpha\beta})$$

La desigualdad anterior es válida en particular para

$$\alpha = \frac{k}{n} \quad < \beta \quad \frac{k+1}{n} \quad \text{con} \in \mathbb{N} \ \text{ y } \ k \in \{0, 1, 2, \ldots\}$$

Definamos ahora para simplificar la notación a $E_k = E_{\alpha\beta}$ con α y β como acabamos de mencionar. Aplicando la desigualdad anterior se obtiene

$$\nu(E_k) - \frac{1}{n}P(E_k) \leq \int_{E_k} XdP \leq \nu(E_k) + \frac{1}{n}P(E_k) \qquad (4)$$

pues $(\beta - \alpha) = \frac{k-1}{n} - \frac{k}{n} = \frac{1}{n}$.

También se tiene que los conjuntos E_k son ajenos por parejas pues, la sucesión $B_0, B_{\frac{1}{n}}, B_{\frac{2}{n}}, \cdots$ es una sucesión creciente, es decir

$$B_0 \subseteq B_{\frac{1}{n}} \subseteq B_{\frac{2}{n}} \subseteq \cdots$$

y $E_k = E - (B_{\frac{k+1}{n}} - B_{\frac{k}{n}})$

Como los E_k son ajenos podemos sumar sobre k en la inecuación (4) obteniendo lo siguiente:

$$\nu(\cup_{k=0}^{\infty} E_k) - \frac{1}{n}P(\cup_{k=0}^{\infty} E_k) \leq \int_{E_k} XdP \leq \nu(\cup_{k=0}^{\infty} E_k) + \frac{1}{n}P(\cup_{k=0}^{\infty} E_k) \qquad (5)$$

Sea $E_{\infty} = E - \cup_{k=0}^{\infty} B_{\frac{k}{n}}$,
como

$$\{\omega \in \Omega \mid X(\omega) < \infty\} \subseteq \cup_{k=0}^{\infty} B_{\frac{k}{n}}$$

pues si $\omega_0 \in \{\omega \in \Omega \mid X(\omega) < \infty\}$ entonces existe $k_0 \in \mathbb{N}$ tal que $X(\omega_0) < \frac{k_0}{n}$
así que $\omega_0 \in \left\{\omega \in \Omega \mid X(\omega) < \frac{k_0}{n}\right\} \subseteq B_{\frac{k_0}{n}}$
entonces se tiene que $E_{\infty} \subseteq \{\omega \in \Omega \mid X(\omega) = \infty\}$ es decir $X(\omega) = \infty$ para todo $\omega \in E_{\infty}$.
Por otra parte, para cualquier $k \in \{0, 1, 2, ...\}$, se tiene que $E_{\infty} \subseteq A_{\frac{k}{n}}$
así que

$$\nu(E_{\infty}) - \frac{k}{n}P(E_{\infty}) \geq 0 \quad \text{es decir } \nu(E_{\infty}) \geq \frac{k}{n}P(E_{\infty})$$

pero esto sucede para toda k, entonces, si $P(E_{\infty}) > 0$ no queda otra que $\nu(E_{\infty}) = \infty$, pero si $P(E_{\infty}) = 0$ entonces $\nu(E_{\infty}) = 0$ porque ν es absolutamente continua con respecto a P.

Es decir se cumple la igualdad

$$\nu(E_{\infty}) = \int_{E_{\infty}} XdP \qquad (6)$$

Finalmente ; como $E = E_{\infty} \cup [\cup_{k=0}^{\infty} E_k]$ y $E_{\infty} \cap [\cup_{k=0}^{\infty} E_k] = \emptyset$ se tiene que :

$$\nu(E) = \nu(E_{\infty}) + \nu(\cup_{k=0}^{\infty} E_k) \qquad \text{y}$$

$$\int_E XdP = \int_{E_{\infty}} XdP + \int_{\cup_{k=0}^{\infty} E_k} XdP$$

Como $P(\cup_{k=0}^{\infty} E_k) \leq P(E)$ entonces combinando lo anterior y (5) y (6) se obtiene que:

$$\begin{aligned}
\int_E XdP &= \int_{E_{\infty}} XdP + \int_{\cup_{k=0}^{\infty} E_k} XdP \\
&= \nu(E_{\infty}) + \int_{\cup_{k=0}^{\infty} E_k} XdP \\
&\leq \nu(E_{\infty}) + \nu(\cup_{k=0}^{\infty} E_k) + \frac{1}{n}P(\cup_{k=0}^{\infty} E_k) \\
&= \nu(E) + \frac{1}{n}P(\cup_{k=0}^{\infty} E_k) \\
&\leq \nu(E) + \frac{1}{n}P(E)
\end{aligned}$$

en conclusión:

$$\int_E XdP \leq \nu(E) + \frac{1}{n}P(E)$$

y por otra parte como

$$\begin{aligned}\int_E XdP &= \int_{E_\infty} XdP + \int_{\bigcup_{k=0}^{\infty} E_k} XdP \\ &= \nu(E_\infty) + \int_{\bigcup_{k=0}^{\infty} E_k} XdP \\ &\geq \nu(E_\infty) + \nu(\textstyle\bigcup_{k=0}^{\infty} E_k) - \frac{1}{n}P(\bigcup_{k=0}^{\infty} E_k) \\ &= \nu(E) - \frac{1}{n}P(\textstyle\bigcup_{k=0}^{\infty} E_k) \\ &\geq \nu(E) - \frac{1}{n}P(E)\end{aligned}$$

en conclusión

$$\nu(E) - \frac{1}{n}P(E) \leq \int_E XdP \leq \nu(E) + \frac{1}{n}P(E) \text{ para toda } n \in \mathbb{N}$$

de aquí se concluye que :

$$\nu(E) = \int_E XdP \text{ para cualquier } E \in \mathcal{A} \qquad \blacksquare$$

Esperanza Condicional

Esperanza Condicional de una variable aleatoria X dado un evento A

Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad. Sea $X : \Omega \to \mathbb{R}$ una variable aleatoria de esperanza finita (Recordemos que $E(X) = \int_\Omega XdP$)

Sea $A \in \mathcal{A}$ un evento.

Como ya se vió en el primer capítulo, si ya ocurrió el evento A, ya no puede ocurrir cualquier evento elemental, esto altera la probabilidad de ocurrencia de cualquier evento B.

La función $P(\bullet \mid A) : \mathcal{A} \to [0, 1]$ definida por $P(B \mid A)$ para todo $B \in \mathcal{A}$, tiene todas las propiedades para ser una medida de probabilidad. De hecho la terna $(\Omega, \mathcal{A}, P(\bullet \mid A))$ es un espacio de probabilidad diferente a $(\Omega, \mathcal{A}, P)$ .La función $X : \Omega \to \mathbb{R}$ sigue siendo una variable aleatoria con respecto al espacio de probabilidad $(\Omega, \mathcal{A}, P(\bullet \mid A))$. Podemos preguntarnos por la esperanza de X con respecto a este nuevo espacio. Naturalmente será $\int_\Omega XdP(\bullet \mid A)$.

Definición 4.25 Sea $(\Omega, \mathcal{A}, P(\bullet \mid A))$ un espacio de probabilidad. Sea $X : \Omega \to \mathbb{R}$ una variable aleatoria , entonces

$$E(X \mid A) = \int_\Omega XdP(\bullet \mid A).$$

Observese que si $P(A) > 0$ entonces

$$\int_\Omega XdP(\bullet \mid A) = \frac{1}{P(A)} \int_\Omega XdP$$

un argumento intuitivo para convencerse de esta igualdad es el siguiente: Si $Y : \Omega \to \mathbb{R}$ es una función simple de esperanza finita, asi que $Y(\omega)=\sum_{k=1}^{n}\alpha_k I_{A_k}(\omega)$ para algunos eventos ajenos $A_1,A_2,...,A_n$ y algunos números reales $\alpha_1,\alpha_2,...,\alpha_n$

Se tiene entonces que

$$\int_\Omega YdP(\bullet \mid A)=\sum_{k=1}^{N}\alpha_k P(A_k \mid A)=\sum_{k=1}^{N}\alpha_k\frac{P(A_k\cap A)}{P(A)}$$

$$=\frac{1}{P(A)}\sum_{k=1}^{N}\alpha_k P(A_k\cap A)$$ y como Y es simple y I_A también lo es

$$I_A(\omega)Y(\omega)=\sum_{k=1}^{n}\alpha_k I_{A\cap A_k}(\omega)$$

se tiene que

$$=\frac{1}{P(A)}\int_\Omega I_A YdP=\frac{1}{P(A)}\int_A YdP$$

Como la igualdad vale para cualquier función simple y la integral de Lebesgue de X se calcula en base a funciones simples, entonces la igualdad debe ser válida en general.

Lo anterior es la esencia de la demostración del siguiente resultado

Proposición 4.26. Sea $(\Omega,\mathcal{A},P(\bullet \mid A))$ un espacio de probabilidad. Sea $X:\Omega\to\mathbb{R}$ una variable aleatoria, si $P(A)>0$ entonces

$$\int_\Omega XdP(\bullet \mid A)=\frac{1}{P(A)}\int_\Omega XdP$$

Lo que a continuación sigue pretende motivar la necesidad de demostrar la existencia de la esperanza condicional.

Sea $A\in\mathcal{A}$ y fijémonos en la $\sigma-\acute{a}lgebra$ generada por el evento A, es decir en $\{A,A^c,\emptyset,\Omega\}$.

$\{A,A^c,\emptyset,\Omega\}$ es una $sub\ \sigma-\acute{a}lgebra$ de $\mathcal{A}$, es decir, tambien es $\sigma-\acute{a}lgebra$ y está contenida en $\mathcal{A}$.

Si X es una variable aleatoria con respecto al espacio $(\Omega,\mathcal{A},P)$ y la esperanz a de X es finita, podemos hablar de $E(X \mid A)$ y de $E(X \mid A^c)$, así que podemos definir una nueva variable aleatoria $Z:\Omega\to\mathbb{R}$ donde

$$Z(\omega)=\begin{cases} E(X \mid A) & \text{si } \omega\in A \\ E(X \mid A^c) & \text{si } \omega\in A^c \end{cases}$$

esta nueva variable aleatoria recibirá el nombre de **esperanza condicional de** X **dada la sub**$\sigma-\acute{a}lgebra$ $\{A,A^c,\emptyset,\Omega\}$.

La variable aleatoria Z cumple las siguientes propiedades

Esto lo podemos generalizar al caso en que tengamos tres elementos de $\mathcal{A}$, A,B y C mutuamente excluyentes y tales que su unión sea Ω. Podríamos definir

$$Z(\omega) = \begin{cases} E(X \mid A) & \text{si } \omega \in A \\ E(X \mid B) & \text{si } \omega \in B \\ E(X \mid C) & \text{si } \omega \in C \end{cases}$$

esta nueva variable aleatoria sería la esperanza condicional de X dada la sub$\sigma - \acute{a}lgebra$ generada por los eventos A, B y C.

¿Qué quiere decir que una $\sigma - \acute{a}lgebra$ sea generada por los eventos A, B y C.?

Definición 4.27. Sea una familia F de subconjuntos de Ω. Se define a la $\sigma - \acute{a}lgebra$ **sea generada** por F como la mas pequeña $\sigma - \acute{a}lgebra$ que contiene a F

En el caso de que tengamos n eventos $A_1, A_2, \ldots, A_n$ mutuamente excluyentes y tales que $\bigcup_{i=1}^{n} A_i = \Omega$ podríamos definir $Z(\omega) = E(X \mid A_i)$ si $\omega \in A_i$ esta nueva variable aleatoria sería la esperanza condicional de X dada la sub$\sigma - \acute{a}lgebra$ generada por los eventos $A_1, A_2, \ldots, A_n$.

Sean $\mathcal{F}$ una sub$\sigma - \acute{a}lgebra$ de $\mathcal{A}$ y $X : \Omega \to \mathbb{R}$ una variable aleatoria con esperanza finita definida en el espacio $(\Omega, \mathcal{A}, P)$

Todo lo anterior nos da pie para la siguiente pregunta :¿Existirá una variable aleatoria $Z : \Omega \to \mathbb{R}$ con las mismas propiedades que en los casos anteriores? Es decir, con las siguientes propiedades:

1) Z es medible con respecto a $\mathcal{F}$
2) Z tiene esperanza finita
3) $\int_A ZdP = \int_A XdP$ Para cualquier $A \in \mathcal{F}$

El siguiente resultado contesta la pregunta planteada

Teorema (Existencia de la variable aleatoria Esperanza Condicional)

Sea $X : \Omega \to \mathbb{R}$ una variable aleatoria de esperanza finita en $(\Omega, \mathcal{A}, P)$

Sea $\mathcal{F}$ una sub$\sigma - \acute{a}lgebra$ de $\mathcal{A}$, entonces existe una variable aleatoria $Z : \Omega \to \mathbb{R}$ que es medible en $(\Omega, \mathcal{F}, P)$, de esperanza finita y tal que $\int_A ZdP = \int_A XdP$ para cualquier $A \in \mathcal{F}$

Ademas, si Y es otra variable aleatoria con las mismas propiedades que Z, entonces $Y = Z$ con probabilidad 1

Demostración: Sean X^{+} y X^{-} la parte positiva y negativa, respectivamente, de X. Sean $Q_{+} : \mathcal{F} \to \mathbf{R}$ y $Q_{-} : \mathcal{F} \to \mathbf{R}$ mediante las relaciones $Q_{+}(B) = \int_B X^{+} dP$ y $Q_{-}(B) = \int_B X^{-} dP$, respectivamente.

Q_{+} y Q_{-} son absolutamente continuas con respecto a P, de manera que, por el teorema de Radón-Nikodym, existen dos variables aleatorias no negativas $Z_{+} : \Omega \to \bar{\mathbf{R}}$ y $Z_{-} : \Omega \to \bar{\mathbf{R}}$ $\mathcal{F}$-medibles tales que, $Q_{+}(A) = \int_A Z_{+} dP$ y $Q_{-}(A) = \int_A Z_{-} dP$ para cualquier $B \in \mathcal{F}$. En particular, tanto Z_{+} como Z_{-} tienen esperanza finita. Así que $Z = Z_{+} - Z_{-}$ satisface las condiciones del teorema.

Por otra parte, si Y es otra variable aleatoria con las mismas propiedades que Z, entonces

$$\int_A YdP = \int_A ZdP$$

para cualquier evento $A \in \mathcal{F}$, así que $Y = Z$ excepto en un conjunto de probabilidad cero ■

Bibliografía

García Alvarez, M.A, Introducción a la teoría de la Probabilidad.
Volumen 1, Notas de Clase.
Facultad de Ciencias . UNAM

García Alvarez, M.A, Introducción a la teoría de la Probabilidad.
Volumen 2, Notas de Clase.
Facultad de Ciencias . UNAM

Ross, Sheldom, A First Course in Probability
Ed. Macmillan Publishing Company
1984, N.Y.

Kolmogorov, A.N, Foundations of the Theory of Probability.
Ed. Chelsea Publishing
1956, N.Y.

García Alvarez, M.A, Introducción al Cálculo Estocástico
Notas de clase.
Facultad de Ciencias. UNAM

Halmos. Measure Theory
Ed. D. Van Nostrand Company.
1956

Bartle, The Elements of Integration.
Ed.John Wiley and Sons.
1996.N.Y.