87

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

2Ej

FACULTAD DE CIENCIAS

NOTAS DE APOYO PARA EL CURSO DE PROBABILIDADES II

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

A C T U A R I O

P R E S E N T A:

LUIS GUILLERMO DE LA ROSA JIMENEZ

MÉXICO, D.F. 1996

VNIVERSDAD NACIONAL
AVFNMA DE
MEXICO

M. en C. Virginia Abrín Batule
Jefe de la División de Estudios Profesionales de la
Facultad de Ciencias
P r e s e n t e

Comunicamos a usted que hemos revisado el trabajo de Tesis:

"NOTAS DE APOYO PARA EL CURSO DE PROBABILIDADES II"

realizado por LUIS GUILLERMO DE LA ROSA JIMENEZ

con número de cuenta 8703479-8 **, pasante 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. JOSE ANTONIO FLORES DIAZ	
Propietario	DR. JOSE RODOLFO MENDOZA BLANCO	José Rodolfo Mendoza Blanco
Suplente	M. en C. BEATRIZ RODRIGUEZ FERNANDEZ	Beatriz Rodríguez Fdz.
Suplente	MAT. MARGARITA ELVIRA CHAVEZ CANO	

Consejo Departamental de Matemáticas

M. en C. ALEJANDRO BRAVO MOJICA

Agradecimientos

Dedico este trabajo a mis Padres, por su ayuda, esfuerzo y entrega.
A Mary, por su cariño, paciencia y ayuda para conmigo.

Por su contribución a la revisión de este trabajo deseo dejar una constancia de reconocimiento a las siguientes personas:

Al Dr. Miguel Angel García por su valiosa ayuda. Al Dr. José Rodolfo Mendoza, porque su interés en este trabajo fue para mi de inestimable valor. A la Dra. Begoña Fernández por sus valiosos comentarios. De igual manera quiero agradecer también a las maestras Beatriz Rodríguez y Margarita Chavez Cano así como al maestro José Antonio Flores por su interés en la revisión de esta tesis. También quiero agradecer a José Luis Batún y a Sergio Nava por el tiempo que dedicaron para refinar este manuscrito.

Contenido

Introducción

El material que se estudia en un segundo curso de probabilidad a nivel licenciatura, se encuentra en una gran variedad de textos, en algunas ocasiones de difícil acceso para la mayoría de los estudiantes. Por lo que se hizo una revisión bibliográfica donde se recopiló parte del material visto en el curso y se presenta en este trabajo como un apoyo al estudiante. Los temas que se tratan en este trabajo se exponen en siete capítulos:

En el primer capítulo se pretende hacer un rápido recordatorio de los conceptos más importantes, que se pueden ver en un curso de probabilidad introductorio, conceptos tales como: probabilidad, probabilidad condicional, espacio muestral, variable aleatoria, etc.

El segundo capítulo contiene una breve introducción a vectores aleatorios y su distribución, así como sus principales propiedades, y por último la relación que guardan con sus distribuciones marginales.

A continuación en el tercer capítulo se muestra como obtener la distribución de una función de un vector aleatorio. En la primera parte de este capítulo se trata el caso de funciones definidas de $\mathcal{R}^n$ a $\mathcal{R}$ y en la segunda parte, funciones definidas de $\mathcal{R}^n$ a $\mathcal{R}^n$. Finalmente, se ilustra la forma de encontrar la distribución de funciones de variables aleatorias mediante varios ejemplos.

El cuarto capítulo esta dedicado a motivar las principales variables aleatorias utilizadas en Estadística, como son las variables aleatorias χ^2, F, t de Student, etc., dando una breve reseña histórica de cada una de ellas, así como también sus principales aplicaciones. Adicionalmente, se incluye la teoría de las estadísticas de orden en su caso más sencillo, es decir, cuando se tienen n variables aleatorias absolutamente continuas, independientes e identicamente distribuidas.

En el quinto capítulo se desarrolla el concepto de esperanza, varianza, así como también el de momentos, tanto centrales como alrededor del origen de una variable aleatoria.

En el sexto capítulo, mediante las herramientas desarrolladas anteriormente se estudian dos de las distribuciones multivariadas más importantes, las cuales son la distribución normal multivariada y la distribución multinomial.

Finalmente, en el último capítulo se desarrolla la teoría de las distribuciones condicionales.

Es importante mencionar que los conceptos matemáticos que se utilizaron en este trabajo son los que se aprende en cursos de licenciatura de los primeros cuatro semestres de las carreras de Actuaría o Matemáticas, de aquí que en ocasiones se tenga que hacer uso de resultados sumamente fuertes sin demostración alguna, ya que para ello se necesita que el alumno tenga dominio de materias como son Análisis matemático o Teoría de la medida, que normalmente un alumno de cuarto semestre de licenciatura no tiene.

Capítulo 1

Resumen de Conceptos básicos

Uno de los primeros intentos de desarrollar una teoría de la probabilidad con rigor matemático se debe a Laplace, en su trabajo titulado, Theorie analytique des probabilités (1812) Laplace dió la definición de lo que ahora conocemos como probabilidad clásica de un evento, el cual tiene un número finito de posibles resultados y que consiste tan solo en calcular el número de maneras en que puede ocurrir un evento dividido entre el número total de posibles resultado, esto siempre y cuando cada posible resultado ocurra con la misma frecuencia. Esta definición es sumamente restringida, ya que existen experimentos, en los cuales no todos los posibles resultados tienen la misma frecuencia, es más, en la mayoría de los experimentos puede obtener un número infinito de posibles resultados. No fue si no hasta principios de este siglo cuando un matemático ruso A. N. Kolmogorov sentó las bases de lo que ahora conocemos como teoría de la probabilidad, dando una base axiomática para ésta, en su trabajo Fundamentos de la Teoría de la Probabilidad en 1933. De acuerdo con este desarrollo, los eventos aleatorios son representados por conjuntos y la probabilidad es tan solo una medida definida sobre esos conjuntos. De aquí que el desarrollo de la Teoría de la Medida dió un fundamento lógico y consistente a la Teoría de la Probabilidad, además de ayudar grandemente al desarrollo de la matemática moderna. Por lo cual es de extrema importancia mencionar las bases axiomáticas de la probabilidad dadas por Kolmogorov.

1.1 Espacios de probabilidad

Supongamos que tenemos que realizar un experimento cuyo resultado no se puede predecir con certeza. Sin embargo, supongamos que el conjunto de todos los posibles resultados es conocido. Este conjunto de posibles resultados de un experimento es conocido como el espacio muestral, que comúnmente denotaremos por la letra griega Ω.

Ejemplo 1.1 *El experimento consiste en lanzar dos dados, entonces el espacio muestral consiste en 36 puntos*

$$\Omega = \begin{cases} (1,1),(1,2),(1,3),(1,4),(1,5),(1,6) \\ (2,1),(2,2),(2,3),(2,4),(2,5),(2,6) \\ (3,1),(3,2),(3,3),(3,4),(3,5),(3,6) \\ (4,1),(4,2),(4,3),(4,4),(4,5),(4,6) \\ (5,1),(5,2),(5,3),(5,4),(5,5),(5,6) \\ (6,1),(6,2),(6,3),(6,4),(6,5),(6,6) \end{cases}$$

donde (i,j) es el evento donde aparece i en el primer dado y j en el segundo.

Ejemplo 1.2 *El experimento consiste en la medida en horas de la duración de un foco seleccionado al azar de la producción de una fábrica de focos, entonces el espacio muestral consiste de todos los números no negativos, esto es*

$$\Omega = [0, \infty)$$

Cualquier subconjunto de Ω se le conoce como un evento. Algunos ejemplos de eventos son los siguientes:

Ejemplo 1.3 *En el ejemplo 1.1, si $E = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$, entonces E es el evento donde la suma de los dos números obtenidos es igual a siete.*

Ejemplo 1.4 *En el ejemplo 1.2, si $E = (20000, 60000)$, entonces E es el evento en donde el foco dura entre 20000 y 60000 horas.*

También tenemos que para cualquier par de eventos E y F de un espacio muestral, se definen los eventos $E \cup F$ el cual es el evento donde E o F ocurra y $E \cap F$ como el evento donde E y F ocurren. $E \cup F$ se le conoce como la unión de E y F, mientras que $E \cap F$ se le conoce como la intersección de E y F, en otras palabras, $E \cup F$ consiste de todos los elementos que están en E y F, mientras que $E \cap F$ consiste en todos los elementos que están tanto en E como en F.

De manera análoga se pueden definir uniones e intersecciones de más de dos eventos, así como también uniones e intersecciones de una infinidad de conjuntos. Finalmente para cualquier evento E definiremos un nuevo evento que se le conoce como el complemento de E el cual denotaremos por E^c y el cual consiste de todos los elementos del espacio muestral que no están en E. Todas estas operaciones son las operaciones básicas que se estudian en teoría de conjuntos. Por lo que daremos por hecho que se está familiarizado con los conceptos básicos que se manejan en teoría de conjuntos.

A continuación para asignar una medida de probabilidad tenemos que definir a cuales subconjuntos de Ω le podemos asignar una medida, ya que en ocasiones, como es el caso del conjunto de números reales, no es posible definir una medida de probabilidad a cada uno de los subconjuntos de $\mathcal{R}$, por lo que tenemos que definir primeramente una familia de subconjuntos sobre el espacio muestral que cumpla ciertas propiedades y la que se encuentren todos los posibles eventos de interés que puedan ocurrir al realizar un cierto experimento cuyo espacio muestral sea Ω. Para ello daremos la definición de σ–álgebra, sobre la cual definiremos nuestra medida de probabilidad.

Definición 1.1 *Sea Ω un espacio muestral. Una colección $\mathcal{A}$ no vacía de subconjuntos de Ω es llamada σ–álgebra si cumple las siguientes propiedades:*

1. $\Omega \in \mathcal{A}$.

2. *Si A está en $\mathcal{A}$, entonces A^c está en $\mathcal{A}$.*

3. *Si A_n está en $\mathcal{A}$, para toda $n \in \mathcal{N}$, entonces $\cup_{n=1}^{\infty} A_n$ está en $\mathcal{A}$.*

Ejemplos de este tipo de familias son:

Ejemplo 1.5 *Si Ω es cualquier subconjunto, la familia de todos los subconjuntos de Ω, la cual denotaremos por 2^{Ω}, es claramente una σ–álgebra.*

Ejemplo 1.6 *Otra σ–álgebra definida sobre cualquier conjunto Ω es la σ–álgebra discreta, consistente únicamente de dos conjuntos*

$$\mathcal{A} = \{\Omega, \emptyset\}.$$

Ejemplo 1.7 *Si Ω es cualquier conjunto y $\mathcal{G}$ es una familia de subconjuntos de Ω, entonces hay una σ–álgebra más pequeña de Ω que contiene a $\mathcal{G}$, ya que se puede ver fácilmente que la* *intersección de todas las $\sigma-$álgebras de Ω que contienen a $\mathcal{G}$ es nuevamente una $\sigma-$álgebra. A esta $\sigma-$álgebra se le conoce como la $\sigma-$álgebra generada por $\mathcal{G}$.*

Ejemplo 1.8 *Si $\Omega = \mathcal{R}$, entonces a la $\sigma-$ álgebra generada por todos los intervalos abiertos (a, b) en $\mathcal{R}$ se le conoce como la $\sigma-$álgebra de Borel y la denotaremos por la letra $\mathcal{B}$. Observe que $\mathcal{B}$ es también generada por todos los intervalos cerrados $[a, b]$ en $\mathcal{R}$. Cualquier conjunto en $\mathcal{B}$ lo llamaremos conjunto de Borel.*

Con estas definiciones podemos enunciar la definición de lo que vamos a entender como una probabilidad y la cual se basa principalmente en los axiomas dados por Kolmogorov.

Definición 1.2 *Una medida de probabilidad P sobre una $\sigma-$álgebra de subconjuntos $\mathcal{A}$ de Ω es una función real que tiene por dominio $\mathcal{A}$ y satisface las siguientes propiedades:*

1. $P\{\Omega\} = 1$,

2. $P\{A\} \geq 0$ *para todo conjunto* $A \in \mathcal{A}$.

3. *Si A_n es una familia finita o infinita numerable de conjuntos mutuamente disjuntos en $\mathcal{A}$, entonces*
$$P\left\{\bigcup_n A_n\right\} = \sum_n P\{A_n\}.$$
Un espacio de probabilidad, denotado por $(\Omega, \mathcal{A}, P)$ es un espacio muestral Ω, una $\sigma-$álgebra de subconjuntos $\mathcal{A}$ y una medida de probabilidad P definida sobre $\mathcal{A}$.

Obs. 1.1 *A una función que satisface la tercera propiedad se le llama contablemente aditiva.*

Obs. 1.2 *Si Ω contiene a lo más n $(n < \infty)$ elementos y $\mathcal{A} = 2^{\Omega}$, a cada elemento individual $\{\omega_i\}$ se le conoce como un evento elemental y es suficiente asignar probabilidad a cada $\{\omega_i\}$ para calcular la probabilidad de cualquier conjunto en $\mathcal{A}$. De esta manera, si $A \in 2^{\Omega}$, entonces $P\{A\} = \sum_{\omega \in A} P\{\omega\}$.*

Obs. 1.3 *Si Ω contiene un número a lo más numerable de elementos (es decir, Ω tiene un numero infinito de elementos, pero se puede establecer una relación de orden) y $\mathcal{A} = 2^{\Omega}$, entonces no es posible que todos sus elementos tengan la misma probabilidad de ocurrencia,* *y en este caso, también es suficiente definir la probabilidad sobre cada evento elemental para poder calcular la probabilidad de cualquier conjunto A, ya que si $A \in 2^{\Omega}$, definimos $P\{A\} = \sum_{\omega \in A} P\{\omega\}$.*

Obs. 1.4 *Si Ω no es numerable y, $\mathcal{A}$ contiene a todos los eventos elementales, entonces no se puede asignar una probabilidad positiva a cada uno de estos eventos, sin violar la condición de que $P\{\Omega\} = 1$. En este caso, no es posible que la $\sigma-$álgebra en la que definamos nuestra probabilidad sea 2^{Ω}.*

Ejemplo 1.9 *Sea $\Omega = \{1,2,3,\ldots\}$ el conjunto de todos los números positivos y sea $\mathcal{A} = 2^{\Omega}$ y definamos P como sigue:*

$$P\{i\} = \frac{1}{2^i}, i = 1,2,\ldots$$

Entonces $\sum_{i=1}^{\infty} P\{i\} = 1$, por lo que P define una probabilidad sobre 2^{Ω}.

Ejemplo 1.10 *Sea $\Omega = [0,\infty)$ y $\mathcal{A} = \mathcal{B}$, la $\sigma-$álgebra de Borel restringida a Ω. Definamos P como sigue, para cada intervalo $I \subseteq \Omega$,*

$$P\{I\} = \int_I e^{-x} dx.$$

Donde esta integral es la integral de Riemann. Claramente P define una probabilidad y esta se puede extender sobre la mayoría de los subconjuntos de $\mathcal{B}$ salvo aquellos conjuntos que no son Riemann integrables[1], ya que $P\{I\} \geq 0$, $P\{\Omega\} = 1$ y P es contablemente aditiva.

A continuación enunciaremos algunas de las propiedades más simples que tienen las probabilidades y que se pueden demostrar de una manera sumamente directa ocupando todas las propiedades que se dan en la definición 1.2.

1.2 Algunas proposiciones

En las proposiciones siguientes tendremos $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad y $E, F \in \mathcal{A}$.

[1] Es importante mencionar que para que esta probabilidad este bien definida sobre todos los conjuntos en $\mathcal{B}$ es necesario utilizar la integral de Lebesgue, la cual puede ser estudiada en cualquier libro de integración avanzado. Es to es ya que el conjunto $A = \{x \in \mathcal{R} : x \text{ es racional}\}$ es un conjunto en $\mathcal{B}$ el cual no es Riemann integrable y si Lebesgue integrable.

Proposición 1.1

$$P\{\emptyset\} = 0.$$

Dem.

Es inmediata de la definición de probabilidad.

Proposición 1.2

$$P\{E^c\} = 1 - P\{E\}.$$

Dem.

$$\begin{aligned} 1 &= P\{\Omega\} \\ &= P\{E \cup E^c\} \\ &= P\{E\} + P\{E^c\} \end{aligned}$$

por lo tanto

$$P\{E^c\} = 1 - P\{E\}.$$

Proposición 1.3 *Si $E \subseteq F$, entonces $P\{E\} \leq P\{F\}$.*

Dem.

Ya que $E \subseteq F$ esto sigue que a F lo podemos expresar como sigue

$$F = E \cup (E \cap F^c),$$

donde E y $E \cap F^c$ son eventos disjuntos, de aquí que

$$\begin{aligned} P\{F\} &= P\{E \cup (F \cap E^c)\} \\ &= P\{E\} + P\{F \cap E^c\} \\ &\geq P\{E\} \end{aligned}$$

ya que $P\{F \cap EF^c\} \geq 0$.

Proposición 1.4

$$P\{E \cup F\} = P\{E\} + P\{F\} - P\{E \cap F\}.$$

Dem.

Se deja como ejercicio.

Definición 1.3 *Una sucesión $E_n, n \in \mathcal{N}$ de eventos, decimos que es creciente si*

$$E_1 \subseteq E_2 \subseteq \ldots \subseteq E_n \subseteq E_{n+1} \subseteq \ldots,$$

Análogamente, decimos que $E_n, n \in \mathcal{N}$ es una sucesión decreciente si

$$E_1 \supseteq E_2 \supseteq \ldots \supseteq E_n \supseteq E_{n+1} \supseteq \ldots,$$

En ambos casos podemos definir el límite de una sucesión de conjuntos, por lo que cuando E_n es creciente el límite se define por

$$\lim_{n\to\infty} E_n = \bigcup_{n=1}^{\infty} E_n$$

y cuando E_n es decreciente lo definimos como

$$\lim_{n\to\infty} E_n = \bigcap_{n=1}^{\infty} E_n$$

Teorema 1.1 *Si $(\Omega, \mathcal{A}, P)$ es un espacio de probabilidad, y E_n es una sucesión de conjuntos creciente o decreciente en $\mathcal{A}$, entonces*

$$\lim_{n\to\infty} P\{E_n\} = P\left\{\lim_{n\to\infty} E_n\right\}$$

Dem.

Supongamos primero que $E_{n,n\in\mathcal{N}}$ es una secuencia de conjuntos crecientes. Definamos los eventos $F_{n,n\in\mathcal{N}}$ por

$$\begin{aligned} F_1 &= E_1 \\ F_n &= E_n \cap \left(\bigcup_{j=1}^{n-1} E_j\right)^c = E_n \cap E_{n-1}^c \qquad si\ n = 2, \ldots \end{aligned}$$

ya que E_n es creciente. Además F_n es una familia de conjuntos numerable mutuamente disjunta y

$$\bigcup_{i=1}^{\infty} F_i = \bigcup_{i=1}^{\infty} E_i \qquad y \qquad \bigcup_{i=1}^{n} F_i = \bigcup_{i=1}^{n} E_i \quad para\ toda\ n \in \mathcal{N}$$

De aquí tenemos que

$$\begin{aligned}
P\left\{\bigcup_{i=1}^{\infty} E_i\right\} &= P\left\{\bigcup_{i=1}^{\infty} F_i\right\} \\
&= \sum_{i=1}^{\infty} P\{F_i\} \\
&= \lim_{n\to\infty} \sum_{i=1}^{n} P\{F_i\} \\
&= \lim_{n\to\infty} P\left\{\bigcup_{i=1}^{n} F_i\right\} \\
&= \lim_{n\to\infty} P\left\{\bigcup_{i=1}^{n} E_i\right\} \\
&= \lim_{n\to\infty} P\{E_n\}
\end{aligned}$$

por lo que queda demostrado el teorema cuando E_n es una secuencia creciente. Cuando E_n es una secuencia decreciente se tiene que

$$P\left\{\bigcap_{i=1}^{\infty} E_i\right\} = 1 - P\left\{\bigcup_{i=1}^{\infty} E_i^c\right\}$$

y ya que $E^c_{n, n\in\mathcal{N}}$ es una secuencia creciente tenemos que

$$\begin{aligned}
P\left\{\bigcap_{i=1}^{\infty} E_i\right\} &= 1 - \lim_{n\to\infty} P\{E_n^c\} \\
&= \lim_{n\to\infty} 1 - P\{E_n^c\} \\
&= \lim_{n\to\infty} P\{E_n\},
\end{aligned}$$

lo cual establece el teorema.

Es claro que como las proposiciones anteriores podemos enunciar muchas más, por lo que aquí vimos las propiedades más importantes, dejando algunas demostraciones como ejercicio.

1.3 Probabilidad condicional

En esta sección daremos la definición de probabilidad condicional, la cual ha sido de gran relevancia, ya que gracias a ella se han desarrollado herramientas como la estadística Bayesiana. Sea E y F eventos definidos sobre un espacio de probabilidad.

Definición 1.4 *El evento $E \mid F$ es la ocurrencia de E dado que ya sabemos que ocurrió F. Como sabemos que F ha ocurrido, tenemos que F se convierte ahora en nuestro nuevo* *espacio muestral y la probabilidad de que ocurra $E \mid F$ será la probabilidad de $E \cap F$ dividida entre la probabilidad de F, es decir,*

$$P\{E \mid F\} = \frac{P\{E \cap F\}}{P\{F\}}$$

es claro que esto está bien definido siempre y cuando $P\{F\} > 0$ y no podremos utilizar tal definición si $P\{F\} = 0$.

Es muy fácil probar que la probabilidad condicional cumple todas las propiedades enunciadas en la definición 1.2, lo cual dejamos como ejercicio.

Ejemplo 1.11 *Supóngase que la población de una cierta ciudad consiste en un 40% de hombres y un 60% de mujeres. Supongase también que el 50% de los hombres y el 30% de las mujeres fuma. Dado que una persona seleccionada es fumador, encontrar la probabilidad de que sea hombre.*

Sol.

Para ello, sea S el conjunto consistente de toda la población, sea H el evento donde la persona seleccionada es hombre, M el evento donde la persona seleccionada es mujer, F el evento donde la persona seleccionada fuma y N el evento donde la persona seleccionada no fuma. Entonces dada la información, tenemos que $P\{F \mid H\} = \frac{1}{2}$, $P\{F \mid M\} = \frac{3}{10}$, $P\{H\} = \frac{2}{5}$, $P\{M\} = \frac{3}{5}$. El problema consiste en calcular $P\{H \mid F\}$, por lo que

$$P\{H \mid F\} = \frac{P\{H \cap F\}}{P\{F\}}$$

Pero tenemos que

$$\begin{aligned} P\{H \cap F\} &= P\{F \mid H\} P\{H\} \\ &= \tfrac{1}{2}\tfrac{2}{5} = \tfrac{1}{5} \end{aligned}$$

Ya que F es la unión de dos conjuntos disjuntos $F \cap H$ y $F \cap M$, tenemos que

$$P\{F\} = P\{F \cap H\} + P\{F \cap M\}$$

Además tenemos que

$$\begin{aligned} P\{F \cap M\} &= P\{M\} P\{F \mid M\} \\ &= \tfrac{3}{5}\tfrac{3}{10} = \tfrac{9}{50} \end{aligned}$$

por lo que

$$\begin{aligned} P\{F\} &= \tfrac{1}{5} + \tfrac{9}{50} \\ &= \tfrac{19}{50} \end{aligned}$$

por último tenemos que

$$\begin{aligned} P\{H \mid F\} &= \frac{\frac{1}{5}}{\frac{19}{50}} \\ &= \frac{10}{19}. \end{aligned}$$

1.4 Eventos independientes

Una concepto de extrema importancia en la Teoría de la Probabilidad, es el concepto de independencia entre dos eventos, y el cual esta relacionado con el hecho de que la ocurrencia de cualquiera de los eventos no afecta la ocurrencia del otro para efecto de cálculo de probabilidades. Este hecho, lo podemos expresar en términos de probabilidades como sigue:

Definición 1.5 *Dos eventos $E, F \in \mathcal{A}$ se dice que son independientes, si*

$$P\{E \cap F\} = P\{E\} P\{F\}.$$

Por la definición 1.4, en muchas ocasiones la definición de independencia esta dada como: E y F son independientes si $P\{E \mid F\} = P\{E\}$ (lo cual también implica que $P\{F \mid E\} = P\{F\}$ si $P\{E\} > 0$), esto siempre y cuando la $P\{F\} > 0$. Esto último hace más conveniente la primera definición aunque menos intuitiva. Cuando dos eventos no son independientes los llamaremos dependientes.

Es claro que el concepto de independencia puede ser extendido para más de dos eventos, para ello existen, dos conceptos de independencia uno mucho más fuerte que el otro.

Definición 1.6 *Los eventos $E_1, E_2, \ldots, E_n$ se dice que son mutuamente independientes si para toda $0 \leq i, j \leq n$, $i \neq j$, tenemos que*

$$P\{E_i \cap E_j\} = P\{E_i\} P\{E_j\}.$$

Definición 1.7 *Los eventos $E_1, E_2, \ldots, E_n$ se dice que son completamente independientes, si cada subcolección de estos conjuntos $E_{1'}, E_{2'}, \ldots, E_{m'}$, $m' \leq n$ de esos eventos cumple*

$$P\{E_{1'} \cap E_{2'} \cap \ldots \cap E_{m'}\} = P\{E_{1'}\} P\{E_{2'}\} \ldots P\{E_{m'}\}.$$

Es claro que si $E_1, E_2, \ldots, E_n$ son completamente independientes, entonces también son mutuamente independientes, pero el recíproco no se cumple. A continuación damos un ejemplo de esto.

Ejemplo 1.12 *Tome cuatro idénticas canicas. En la primera, escribo los números A_1, A_2, A_3. En las otras tres, escriba A_1, A_2 y A_3 respectivamente. Ponga las cuatro canicas en una urna y tome una aleatoriamente. Sea E_i el evento, donde se obtiene en la canica el símbolo A_i $i = 1, 2, 3$, entonces,*

$$\begin{aligned} P\{E_1\} = P\{E_2\} = P\{E_3\} &= \frac{1}{2} \\ P\{E_1 \cap E_2\} = P\{E_2 \cap E_3\} = P\{E_1 \cap E_3\} &= \frac{1}{4} \\ P\{E_1 \cap E_2 \cap E_3\} &= \frac{1}{4}. \end{aligned}$$

Esto sigue que aunque los eventos E_1, E_2, E_3 no son completamente independientes, si son mutuamente independientes.

Ejemplo 1.13 *En este ejemplo la $P\{E_1 \cap E_2 \cap E_3\} = P\{E_1\}P\{E_2\}P\{E_3\}$, pero E_1, E_2, E_3 no son independientes a pares y de aquí no son completamente independientes. Sea $\Omega = \{1,2,3,4\}$, y sea p_i la probabilidad de obtener $\{i\}$, $i = 1, 2, 3, 4$. Sea $p_1 = \frac{\sqrt{2}}{2} - \frac{1}{4}$, $p_2 = \frac{1}{4}$, $p_3 = \frac{3}{4} - \frac{\sqrt{2}}{2}$, $p_4 = \frac{1}{4}$. Sea $E_1 = \{1,3\}$, $E_2 = \{2,3\}$, $E_3 = \{3,4\}$. Entonces*

$$\begin{aligned} P\{E_1 \cap E_2 \cap E_3\} &= P\{3\} \\ &= \tfrac{3}{4} - \tfrac{\sqrt{2}}{2} \\ &= \tfrac{1}{2}\left(1 - \tfrac{\sqrt{2}}{2}\right)\left(1 - \tfrac{\sqrt{2}}{2}\right) \\ &= (p_1 + p_2)(p_2 + p_3)(p_3 + p_4) \\ &= P\{E_1\}P\{E_2\}P\{E_3\} \end{aligned}$$

Pero $P\{E_1 \cap E_2\} = \frac{3}{4} - \frac{\sqrt{2}}{2} \neq P\{E_1\}P\{E_2\}$ y esto sigue que E_1, E_2, E_3 no son completamente independientes.

1.5 La Fórmula de Bayes

A continuación enunciaremos una de las propiedades más importantes de la probabilidad condicional y que es el fundamento de toda una corriente estadística denominada Estadística Bayesiana, la cual se basa principalmente en esta fórmula para poder realizar inferencia. Para poder enunciar la fórmula de Bayes, enunciaremos primeramente el siguiente resultado.

Teorema 1.2 *Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad. Supongase que H_i, $i = 1, 2, \ldots, n$ es una colección de eventos en $\mathcal{A}$ tales que son una partición disjunta de Ω (es decir*

$\bigcup_{i=1}^{n} H_i = \Omega$ *y* $H_i \cap H_j = \emptyset$ *para todo* $i \neq j$*) y* $P\{H_i\} > 0$ *para toda* $i = 1, 2, \ldots, n$*. Sea* E *otro evento también en* $\mathcal{A}$*, entonces*

$$P\{E\} = \sum_{i=1}^{n} P\{E \mid H_i\} P\{H_i\} \tag{1.1}$$

Dem.

Dado que $\bigcup_{i=1}^{n} H_i = \Omega$*, tenemos que*

$$\begin{aligned} P\{E\} &= P\left\{\left(\bigcup_{i=1}^{n} H_i\right) \cap E\right\} \\ &= P\left\{\bigcup_{i=1}^{n} (H_i \cap E)\right\} \\ &= \sum_{i=1}^{n} P\{(H_i \cap E)\} \\ &= \sum_{i=1}^{n} P\{E \mid H_i\} P\{H_i\}, \end{aligned}$$

ya que $(H_i \cap E) \cap (H_j \cap E) = \emptyset$ *para* $i \neq j$.

La ecuación 1.1 se le conoce comúnmente con el nombre de fórmula de probabilidad total, y enuncia el hecho de que la probabilidad de un evento siempre es igual a la suma del producto de las probabilidades condicionales de el evento dado cada uno de los conjuntos en una partición disjunta de Ω, por la probabilidad de cada uno de esos conjuntos. Esta fórmula, en ocasiones es de gran utilidad para calcular probabilidades, cuando se sabe la probabilidad para cada uno de los conjuntos en una partición disjunta de Ω y también las probabilidades condicionales del conjunto dado cada uno de los conjuntos de la partición.

Ejemplo 1.14 *Una compañía de seguros cree que la gente puede dividirse en dos clases, aquellos propensos a tener accidentes y aquellos que no lo son. Sus estadísticas muestran que las personas propensas a tener accidentes, fijando un año, tienen un accidente con probabilidad* 0.1*, mientras que las que no son propensas, tienen probabilidad* 0.05 *de sufrir un accidente. Si asumimos que el* 30% *de la gente está propensa a sufrir accidentes, ¿cuál es la probabilidad de que una póliza dada por la compañía tenga que pagar por un accidente dentro del año que cubre ésta?*

Sol.

Sea A_1 *el evento donde el dueño de la póliza sufre una accidente, sea* A *el evento, donde la póliza es firmada por una persona propensa al riesgo. De aquí la probabilidad* *deseada es $P\{A_1\}$, la cual está dada por*

$$\begin{aligned} P\{A_1\} &= P\{A_1 \mid A\} P\{A\} + P\{A_1 \mid A^c\} P\{A^c\} \\ &= (.1)(.3) + (.05)(.7) = 0.056 \end{aligned}$$

Teorema 1.3

$$P\{H_i \mid E\} = \frac{P\{H_i\} P\{E \mid H_i\}}{\sum_{i=1}^{n} P\{E \mid H_i\} P\{H_i\}} \tag{1.2}$$

<u>*Dem.*</u>

La demostración es sumamente sencilla utilizando la fórmula de probabilidad total.

Ejemplo 1.15 *Supongase que tenemos tres cartas idénticas, exceptuando que la primera carta está pintada de rojo en ambos lados, la segunda, está pintada de negro también en ambos lados y por último, la tercera carta está pintada de un lado de rojo y del otro de negro. Las tres cartas son puestas en un sombrero y revueltas, después es elegida una de ellas aleatoriamente. Si uno de los lados de la carta tomada aleatoriamente es rojo ¿cuál es la probabilidad de que el otro lado sea de colo negro?*

<u>*Sol.*</u>

Sea RR, NN, RN los eventos que denotan respectivamente el tomar la primera, segunda y tercera carta respectivamente. Sea R el evento, donde la carta tomada tiene uno de sus lados lado rojo, entonces la probabilidad que deseamos calcular está dada por

$$\begin{aligned} P\{RN \mid R\} &= \frac{P\{R|RN\}P\{RN\}}{P\{R|RR\}P\{RR\}+P\{R|RN\}P\{RN\}+P\{R|NN\}P\{NN\}} \\ &= \frac{1\frac{1}{3}}{1\frac{1}{3}+1\frac{1}{3}+0\frac{1}{3}} = \frac{1}{2} \end{aligned}$$

1.6 Variables Aleatorias

En el sección anterior, definimos lo que es una función de probabilidad sobre una $\sigma-$álgebra $\mathcal{A}$. Es claro $\mathcal{A}$ es una $\sigma-$álgebra arbitraria de Ω, por lo que comúnmente P no es fácil manejarla. Sin embargo, en la práctica nuestros eventos pueden ser representados por números reales, es decir, que comúnmente la probabilidad de cada uno de nuestros eventos en $\mathcal{A}$ es calculada con el número que representa a este evento. Para ser más preciso existe una función definida sobre $\mathcal{A}$ en los números reales, con la cual hasta el momento hemos trabajado implícitamente, ya que es muy difícil definir una regla de correspondencia, que defina a la probabilidad directamente sobre $\mathcal{A}$, de esta forma daremos a continuación una de las definiciones más importantes en la Teoría de la Probabilidad y que está íntimamente ligada al concepto de función medible que se trabaja en Teoría de la medida.

Definición 1.8 *Sea Ω un espacio muestral y $\mathcal{A}$ una $\sigma-$álgebra definida sobre Ω, entonces decimos que una función $X : \Omega \to \mathcal{R}$ la cual mapea elementos de Ω en los reales, es una variable aleatoria si, la imágenes inversas sobre todos los subconjuntos de $\mathcal{R}$ en $\mathcal{B}$ son elementos de $\mathcal{A}$, es decir*

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{A} \text{ para toda } B \in \mathcal{B}. \tag{1.3}$$

Es muy común también que estas funciones sean llamadas funciones medibles, por lo cual, cuando hablemos de éstas nos referimos a una funciones que cumplan la condición 1.3.

Sea $x \in \mathcal{R}$, y considere los intervalos semicerrados $(-\infty, x]$. Ya que $(-\infty, x] \in \mathcal{B}$, tenemos que si X es una variable aleatoria, entonces

$$X^{-1}(-\infty, x] = \{X(\omega) \leq x\},$$

es un evento en $\mathcal{A}$.

También, si B es un conjunto de Borel en $\mathcal{R}$, entonces B puede ser obtenido por un número numerable de operaciones de uniones, intersecciones, y diferencias de intervalos semicerrados. Usando la definición de variable aleatoria, junto con el hecho anteriormente mencionado podemos enunciar los siguientes teoremas.

De ahora en adelante para facilitar la notación y para que no haya confusión escribiremos $\{X \in B\}$ en lugar de

$$\{\omega \in \Omega : X(\omega) \in B\},$$

Así el conjunto

$$\{\omega \in \Omega : X(\omega) \leq x\},$$

lo escribiremos como $\{X \leq x\}$.

Teorema 1.4 *X es una variable aleatoria si y sólo si para cada $x \in \mathcal{R}$*

$$\{X \leq x\} \in \mathcal{A}$$

Note que la noción de probabilidad no entra en la definición de variable aleatoria.

Teorema 1.5 *Sea X una variable aleatoria definida sobre una $\sigma-$álgebra $\mathcal{A}$ en Ω, y a, b constantes. Entonces $aX + b$ es también una variable aleatoria sobre $\mathcal{A}$.*

Dem.

La prueba es muy simple y se deja como ejercicio.

Ejemplo 1.16 *Para cualquier conjunto $A \subseteq \Omega$, definimos*

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

$I_A(\omega)$ se le conoce como la función indicadora del conjunto A. I_A es una variable aleatoria, si y sólo si $A \in \mathcal{A}$.

Ejemplo 1.17 *Sea $\Omega = \{Sol,\ águila\}$ y sea $\mathcal{A} = 2^{\Omega}$. Defina X por $X(sol) = 1$, $X(águila) = 0$. Entonces*

$$X^{-1}(-\infty, x] = \begin{cases} \emptyset & si\ x < 0, \\ \{águila\} & si\ 0 \leq x < 1, \\ \{Sol,\ águila\} & si\ x \geq 1, \end{cases}$$

y claramente X es una variable aleatoria.

Obs. 1.5 *Si Ω es un espacio muestral discreto, es decir Ω es un conjunto con un número numerable de elementos puntuales y $\mathcal{A} = 2^{\Omega}$, entonces cada función real definida sobre $\mathcal{A}$ es una variable aleatoria.*

Ejemplo 1.18 *Sea $\Omega = [0,1]$ y $\mathcal{A} = \mathcal{B} \cap [0,1]$, la $\sigma-$álgebra de Borel sobre el intervalo $[0,1]$. Definimos X una variable aleatoria por*

$$X(\omega) = w, \qquad si\ \omega \in [0,1].$$

Claramente X es una variable aleatoria y cualquier subconjunto en $\mathcal{B}$ es un evento.

1.6.1 Distribución de probabilidad de una variable aleatoria

En esta sección veremos que cada variable aleatoria induce un espacio de probabilidad y además definiremos lo que es una función de distribución, dando sus principales propiedades. Para ello tendremos Ω un espacio muestral, $\mathcal{A}$ una $\sigma-$álgebra definida sobre Ω y X una variable aleatoria sobre Ω.

Teorema 1.6 *La variable aleatoria X definida en un espacio de probabilidad $(\Omega, \mathcal{A}, P)$ induce un espacio de probabilidad $(\mathcal{R}, \mathcal{B}, Q)$ por medio de la correspondencia*

$$Q\{B\} = P\left\{X^{-1}(B)\right\} = P\{X \in B\} \qquad para\ toda\ B \in \mathcal{B}.$$

Escribiremos $Q = P\{X^{-1}\}$ y llamaremos a Q o a $P\{X^{-1}\}$ la distribución de probabilidad de X.

Dem.

Claramente $Q\{B\} \geq 0$ para toda $B \in \mathcal{B}$, y también tenemos que

$$Q\{\mathcal{R}\} = P\{X \in \mathcal{R}\} = P\{\Omega\} = 1.$$

Sea $B_i \in \mathcal{B}, i = 1, 2, \ldots$ con $B_i \cap B_j = \emptyset$ para toda $i \neq j$. Ya que la imagen inversa de una unión disjunta de conjuntos de Borel es la unión de sus imágenes inversas, tenemos que

$$\begin{aligned} Q\{\textstyle\bigcup_{i=1}^{\infty} B_i\} &= P\{X^{-1}(\textstyle\bigcup_{i=1}^{\infty} B_i)\} \\ &= P\{\textstyle\bigcup_{i=1}^{\infty} X^{-1}(B_i)\} \\ &= \sum_{i=1}^{\infty} P\{X^{-1}(B_i)\} = \sum_{i=1}^{\infty} Q\{B_i\} \end{aligned}$$

Esto sigue que $(\mathcal{R},\mathcal{B},Q)$ es un espacio de probabilidad.

En este caso Q es una función definida sobre los conjuntos borelianos de $\mathcal{R}$, los cuales consisten de todas las uniones e intersecciones numerables y complementos, así como la combinación de estas tres operaciones entre conjuntos abiertos por lo que no es sencillo expresar muchos de estos eventos en términos de eventos elementales, por lo que resulta que esta función no sea sencillo definirla, de aquí que introduzcamos la siguiente función puntual definida sobre $\mathcal{R}$.

Definición 1.9 *Una función real F definida sobre $\mathcal{R}$ que es no decreciente, continua por la derecha y que satisface que*

$$\lim_{x\to\infty} F(x) = 1 \text{ y } \lim_{x\to-\infty} F(x) = 0$$

es llamada una función de distribución.

Obs. 1.6 *Dado que F es continua por la derecha y no decreciente, se tiene que si b es cualquier número real y b_n es una secuencia decreciente que converge a b, entonces $\lim_{n\to\infty} F(b_n) = F(b)$ (en este caso al $\lim_{n\to\infty} F(b_n)$ se le conoce como límite por la derecha de F y comúnmente se le denota por $F(b+)$), además se puede probar que el límite por la izquierda existe, el cual se define de manera análoga y el cual denotaremos por $F(b-)$. En general, tendremos que*

$$F(x-) \leq F(x) = F(x+),$$

para toda $x \in \mathcal{R}$, y x es un punto de discontinuidad para F si y sólo si $F(x+)$ y $F(x-)$ no son iguales. Este resultado se tiene en general para cualquier función no decreciente F, por lo cual tenemos que en la definición de función de distribución, el hecho de que la función sea no decreciente es de extrema importancia, ya que si ésta no fuera continua por la derecha, podemos definir una nueva función como

$$F^*(x) = F(x+),$$

la cual ya es continua por la derecha en $\mathcal{R}$. Es importante mencionar que algunos autores, hacen que la función de distribución sea continua por la izquierda y esto se logra definiendo la función de distribución como

$$F_X(x) = P\{X < x\},$$

la cual claramente es continua por la izquierda (en particular los de origen ruso tienen esta definición de función de distribución).

Definición 1.10 *Sea X una variable aleatoria definida sobre un espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Defina una función puntual $F_X(.)$ en $\mathcal{R}$ por*

$$F_X(x) = P\{X \leq x\} \qquad \text{para toda } x \in \mathcal{R} \tag{1.4}$$

La función F_X es llamada la función de distribución de X.

Teorema 1.7 *La función F_X definida en 1.4 es una función de distribución.*

Dem.

Sea $x_1 < x_2$. Entonces claramente tenemos que $(-\infty, x_1] \subseteq (-\infty, x_2)$ y por las propiedades elementales de la probabilidad, tenemos que

$$F_X(x_1) = P\{X \leq x_1\} \leq P\{X \leq x_2\} = F_X(x_2)$$

por lo cual F_X es una función no decreciente. Ya que F_X es no decreciente basta probar únicamente que para $x \in \mathcal{R}$ y cualquier secuencia de números que decrecen a x, es decir

$$x_1 > x_2 > \ldots > x_n > \ldots > x \quad y \quad \lim_{n\to\infty} x_n = x$$

se tiene que $F_X(x_n) \to F_X(x)$. Sea $A_k = \{x < X \leq x_k\}$, entonces $A_k \in \mathcal{A}$ y también

$$\lim_{k\to\infty} A_k = \bigcap_{k=1}^{\infty} A_k = \emptyset,$$

ya que ningún intervalo $(x, x_k]$ *contiene a* x. *Esto implica que* $\lim_{k\to\infty} P\{A_k\} = 0$. *Pero*

$$\begin{aligned} P\{A_k\} &= P\{X \le x_k\} - P\{X \le x\} \\ &= F_X(x_k) - F_X(x) \end{aligned}$$

de aquí que

$$\lim_{k\to\infty} F(x_k) = F(x),$$

por lo cual F *es continua por la derecha.*

Finalmente, sea x_n *una sucesión de números decrecientes a* $-\infty$. *Entonces*

$$\{X \le x_n\} \supseteq \{X \le x_{n+1}\} \quad \textit{para cada } n \in \mathcal{N}$$

y

$$\lim_{n\to\infty} \{X \le x_n\} = \bigcap_{i=1}^{\infty} \{X \le x_n\} = \emptyset,$$

por lo cual

$$F(-\infty) = \lim_{n\to\infty} P\{X \le x_n\} = P\left\{\lim_{n\to\infty} \{X \le x_n\}\right\} = 0.$$

De manera análoga podemos demostrar que

$$F(\infty) = \lim_{n\to\infty} P\{X \le x_n\} = 1$$

si es que x_n *es una sucesión que crece a* ∞, *por lo que la prueba esta completa.*

El siguiente resultado, el cual se enunciara sin prueba, establece una correspondencia entre la probabilidad inducida Q definida sobre $\mathcal{B}$ y una función de distribución F definida en $\mathcal{R}$.

Teorema 1.8 *Dada una probabilidad* Q *definida en* $\mathcal{B}$, *existe una función de distribución que satisface*

$$Q(-\infty, x] = F(x) \quad \textit{para toda } x \in \mathcal{R}, \tag{1.5}$$

e inversamente, dada una función de distribución F, *existe una única probabilidad* Q *definida sobre* $\mathcal{B}$ *que satisface 1.5.*

Teorema 1.9 *Cada función de distribución define una función de distribución sobre algún espacio de probabilidad.*

<u>*Dem.*</u>

Sea F una función de distribución, por el teorema 1.8 existe una única probabilidad Q definida en $\mathcal{B}$ que satisface

$$Q(-\infty, x] = F(x) \quad \text{para toda } x \in \mathcal{R}.$$

Sea $(\mathcal{R},\mathcal{B},Q)$ *el espacio de probabilidad en la que definimos la variable aleatoria*

$$X(\omega) = \omega, \quad \omega \in \mathcal{R}.$$

Entonces

$$Q\{X \leq x\} = Q(-\infty, x] = F(x),$$

y F es la función de distribución de la variable aleatoria X.

Obs. 1.7 *Si X es una variable aleatoria en un espacio de probabilidad, dado por* $(\Omega, \mathcal{A}, P)$*, vimos que por el teorema 1.7 que* $F_X(x) = P\{X \leq x\}$ *es una función de distribución asociada a X. El teorema 1.9 asegura que a cada F la podemos asociar con alguna variable aleatoria. De aquí para cada variable aleatoria, existe una función de distribución e inversamente. De ahora en adelante cuando hablemos de una variable aleatoria, asumiremos que ésta está definida sobre algún espacio de probabilidad.*

Ejemplo 1.19 *Sea X la variable aleatoria definida sobre* $(\Omega, \mathcal{A}, P)$ *por*

$$X(\omega) = c \quad \text{para toda } \omega \in \Omega.$$

Entonces

$$P\{X = c\} = 1,$$

y

$$F_X(x) = Q(-\infty, x] = P\left\{X^{-1}(-\infty, x]\right\} = \begin{cases} 0 & si\ x < c \\ 1 & si\ x \geq c \end{cases}.$$

Esta variable aleatoria se le conoce comúnmente como variable aleatoria degenerada.

Ejemplo 1.20 *Sea* $\Omega = \{sol, águila\}$ *y X definida por*

$$X(sol) = 1, X(águila) = 0.$$

Si P asigna probabilidades iguales a $\{sol\}$ *y a* $\{águila\}$*, entonces*

$$P\{X = 0\} = \frac{1}{2} = P\{X = 1\},$$

y

$$F_X(x) = Q(-\infty, x] = \begin{cases} 0 & si\ x < 0, \\ \frac{1}{2} & si\ 0 \leq x < 1, \\ 1 & si\ x \geq 1. \end{cases}$$

Ejemplo 1.21 *En el ejemplo 1.18. Para cada subintervalo I de $[0,1]$, sea $P\{I\}$ la longitud del intervalo. Entonces la función de distribución de la variable aleatoria definida por $X(\omega) = \omega$, $\omega \in [0,1]$, está dada por*

$$F_X(x) = Q(-\infty, x] = P\left\{X^{-1}(-\infty, x]\right\} = \begin{cases} 0 & si\ x < 0 \\ x & si\ x \in [0,1] \\ 1 & si\ x > 0 \end{cases},$$

ya que $P\{(-\infty, x]\} = P\{[0,x]\} = x$ para toda $x \in [0,1]$.

1.6.2 Variables aleatorias discretas y continuas

En este trabajo nos restringiremos exclusivamente a dos tipos de variables aleatorias, las cuales son las variables aleatorias absolutamente continuas y las variables discretas las cuales definiremos a continuación, no significando esto que sean las únicas que existen, ya que en muchos casos las variables aleatorias no son de ninguno de estos dos tipos, pero tanto las variables aleatorias discretas como las absolutamente continuas, son las de mayor uso en la estadística y en las aplicaciones prácticas. Las variables aleatorias discretas son aquellas que asumen a lo más un número numerable de valores, y las variables aleatorias continuas son aquellas cuya función de distribución es una función absolutamente continua, lo cual definiremos con más claridad a continuación.

Definición 1.11 *Una variable aleatoria X definida sobre $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad, se dice que es discreta, si existe un conjunto numerable $D \subseteq \mathcal{R}$ tal que $P\{X \in D\} = 1$. Los puntos de D los cuales tienen una probabilidad positiva son llamados puntos de salto de la función de distribución de X y sus probabilidades las llamaremos saltos de la función de distribución de X.*

Obs. 1.8 *Note que en cualquier caso $D \in B$ ya que si $x \in \mathcal{R}$, entonces*

$$\{x\} = \bigcap_{n=1}^{\infty} \left(x - \frac{1}{n}, x + \frac{1}{n}\right)$$

y

$$F_X(x) = Q(-\infty, x] = \begin{cases} 0 & si\ x < 0, \\ \frac{1}{2} & si\ 0 \leq x < 1, \\ 1 & si\ x \geq 1. \end{cases}$$

Ejemplo 1.21 *En el ejemplo 1.18. Para cada subintervalo I de $[0,1]$, sea $P\{I\}$ la longitud del intervalo. Entonces la función de distribución de la variable aleatoria definida por $X(\omega) = \omega$, $\omega \in [0,1]$, está dada por*

$$F_X(x) = Q(-\infty, x] = P\left\{X^{-1}(-\infty, x]\right\} = \begin{cases} 0 & si\ x < 0 \\ x & si\ x \in [0,1] \\ 1 & si\ x > 0 \end{cases},$$

ya que $P\{(-\infty, x]\} = P\{[0,x]\} = x$ para toda $x \in [0,1]$.

1.6.2 Variables aleatorias discretas y continuas

En este trabajo nos restringiremos exclusivamente a dos tipos de variables aleatorias, las cuales son las variables aleatorias absolutamente continuas y las variables discretas las cuales definiremos a continuación, no significando esto que sean las únicas que existen, ya que en muchos casos las variables aleatorias no son de ninguno de estos dos tipos, pero tanto las variables aleatorias discretas como las absolutamente continuas, son las de mayor uso en la estadística y en las aplicaciones prácticas. Las variables aleatorias discretas son aquellas que asumen a lo más un número numerable de valores, y las variables aleatorias continuas son aquellas cuya función de distribución es una función absolutamente continua, lo cual definiremos con más claridad a continuación.

Definición 1.11 *Una variable aleatoria X definida sobre $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad, se dice que es discreta, si existe un conjunto numerable $D \subseteq \mathcal{R}$ tal que $P\{X \in D\} = 1$. Los puntos de D los cuales tienen una probabilidad positiva son llamados puntos de salto de la función de distribución de X y sus probabilidades las llamaremos saltos de la función de distribución de X.*

Obs. 1.8 *Note que en cualquier caso $D \in \mathcal{B}$ ya que si $x \in \mathcal{R}$, entonces*

$$\{x\} = \bigcap_{n=1}^{\infty} \left(x - \frac{1}{n}, x + \frac{1}{n}\right)$$

Por lo que tenemos que $\{x\} \in \mathcal{B}$, de aquí, dado que D es un conjunto a lo más numerable tenemos que

$$D = \bigcup_{x \in D} \{x\},$$

y dado que B puede ser representado como una unión numerable de conjuntos en $\mathcal{B}$, entonces $D \in \mathcal{B}$. Es claro que si X es una variable aleatoria discreta, y toma únicamente los valores x_i, $i = 1, 2, \ldots$ con probabilidad p_i, entonces tendremos que $\sum_{i=1}^{\infty} p_i = 1$.

Definición 1.12 *Si X es una variable aleatoria discreta, entonces definimos la función real f_X como sigue*

$$f_x(x) = P\{X = x\},$$

y se le conoce como función de densidad de probabilidades de X.

Obs. 1.9 *Es importante mencionar que en el caso discreto el conjunto*

$$T = \{x \in \mathcal{R} : f_X(x) > 0\}$$

es a lo más numerable y además tenemos que $T \subseteq D$, por lo que podemos establecer un orden en el conjunto T. Comúnmente denotaremos por x_i, $i = 1, 2, \ldots$ a todos los posibles puntos que están en T, es decir

$$\bigcup_{i=1}^{\infty} \{x_i\} = T.$$

Además la función de distribución de X está dada por

$$F_X(x) = P\{X \leq x\} = \sum_{x_i \leq x} f_X(x_i).$$

También tenemos que si I_A es la función indicadora del conjunto A, podemos escribir nuestra variable aleatoria como

$$X(\omega) = \sum_{i=1}^{\infty} x_i I_{\{X = x_i\}}(\omega).$$

Ejemplo 1.22 *Sea X una variable aleatoria con función de densidad de probabilidades, dada por*

$$f_X(x) = \begin{cases} \frac{6}{\pi^2}\frac{1}{x^2} & si\ x = 1, 2, \ldots \\ 0 & e.o.c. \end{cases},$$

por lo que la función de distribución de X, esta dada por

$$F_X(x) = \begin{cases} 0 & si\ x < 1 \\ \frac{6}{\pi^2}\sum\limits_{k=1}^{[x]}\frac{1}{k^2} & si\ x \geq 1 \end{cases},$$

donde $[x]$ es la menor parte entera de x.

Podemos observar que si X es una variable aleatoria discreta, entonces la función de distribución de X, F_X será una función de saltos, esto es que si x_n, $n \in \mathcal{N}$ son los posibles valores que toma X, entonces $F_x(x)$ es constante en los intervalos $[x_{i-1}, x_i)$ y da un salto de tamaño p_i (donde $p_i = P\{X = x_i\}$) en x_i.

Ejemplo 1.23 *Sea X una variable aleatoria discreta, cuya función de densidad de probabilidades está dada por*

$$f_X(x) = \begin{cases} \frac{1}{4} & si\ x = 1 \\ \frac{1}{2} & si\ x = 2 \\ \frac{1}{8} & si\ x = 3, 4 \\ 0 & e.o.c. \end{cases}.$$

Por lo que la función de distribución de X, F_X está dada por

$$F_X(x) = \begin{cases} 0 & si\quad x < 1 \\ \frac{1}{4} & si\ 1 \leq x < 2 \\ \frac{3}{4} & si\ 2 \leq x < 3 \\ \frac{7}{8} & si\ 3 \leq x < 4 \\ 1 & si\ 4 \leq x \end{cases}.$$

Es claro que F_X es una función de saltos (ver fig.1), mientras que la gráfica de f_X consiste únicamente de unos cuantos puntos (ver fig.1).

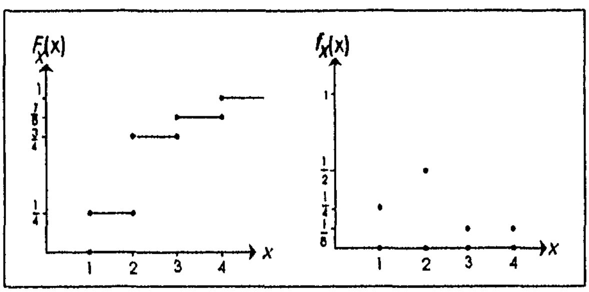


Figura 1. F_X y f_X

A continuación consideraremos las variables aleatorias con una función de densidad que no tenga saltos. Estas variables aleatorias son llamadas variables aleatorias continuas, y nos restringiremos a estudiar un subconjunto de estas variables, las cuales son las variables absolutamente continuas.

Definición 1.13 *Sea X una variable aleatoria definida sobre $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad, con función de distribución F_X. Entonces decimos que X es una variable continua si F_X es una función continua. Si F_X es una función absolutamente continua, es decir, existe una función f_X tal que para cada número real x tenemos que*

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt.$$

entonces decimos que X es una variable absolutamente continua. La función f_X es llamada la función de densidad de X.

Obs. 1.10 *Es importante mencionar que cuando tengamos una variable aleatoria discreta, f_X la llamaremos función de densidad de probabilidades, mientras que si X es una variable absolutamente continua, entonces a f_X la llamaremos simplemente función de densidad de X. Claramente estas dos funciones surgen de distinta manera, ya que en el primer caso, f_X representa la probabilidad de que $X = x$, mientras que si X es una variable absolutamente continua, esta asume cada valor real con probabilidad cero, es decir $P\{X = x\} = 0$ para toda X, por lo cual $f_X(x)$ no representa probabilidad alguna, lo cual se demostrará a continuación.*

Teorema 1.10 *Sea X una variable aleatoria arbitraria con función de distribución F_X y $a \in \mathcal{R}$. Entonces*

$$P\{X = a\} = F_X(a) - F_X(a-).$$

<u>Dem.</u>

Claramente, tenemos que si $a < b$ son números reales, entonces

$$P\{a < X \leq b\} = F_X(b) - F_X(a), \tag{1.6}$$

ya que $\{X \leq b\} = \{X \leq a\} \cup \{a < X \leq b\}$ por lo que

$$P\{X \leq b\} = P\{X \leq a\} + P\{a < X \leq b\},$$

por lo que tenemos la ecuación 1.6. Para calcular $P\{X = a\}$, calcularemos primero la probabilidad de $\{X < a\}$, por lo que

$$\begin{aligned} P\{X < a\} &= P\{\lim_{n\to\infty}\{X \leq a - x_n\}\} \\ &= \lim_{n\to\infty} P\{X \leq a - x_n\} \\ &= \lim_{n\to\infty} F_X(a - x_n), \end{aligned}$$

donde x_n es una secuencia de números no negativos que convergen a cero, por lo que claramente tenemos que

$$P\{X < a\} = F_X(a-),$$

por la definición de límite por la izquierda. Ahora, dado que

$$P\{X \leq a\} = P\{X < a\} + P\{X = a\},$$

tenemos que

$$P\{X = a\} = F_X(a) - F_X(a-),$$

lo cual establece el teorema.

Utilizando este resultado, es muy fácil ver que si X es una variable aleatoria continua, entonces para toda $a \in \mathcal{R}$, se tiene que

$$P\{X = a\} = 0$$

dado que $F_X(a-) = F_X(a)$

Obs. 1.11 *Note que $f_X(x) \geq 0$ para todo $x \in \mathcal{R}$ y satisface que*

$$\lim_{x\to\infty} F_X(x) = 1 = \int_{-\infty}^{\infty} f_X(t)dt.$$

Sea a y b dos números reales cualesquiera, tales que $a < b$, entonces

$$\begin{aligned} P\{a < X \leq b\} &= F_X(a) - F_X(b) \\ &= \int_a^b f_X(t)dt. \end{aligned}$$

Sea B un boreliano en $\mathcal{R}$. Ya que B puede ser obtenido con un número numerable de uniones, intersecciones, diferencias de intervalos, los siguientes resultados se sostienen.

Teorema 1.11 *Sea X una variable absolutamente continua con función de densidad f_X. Entonces para cada $B \in \mathcal{B}$*

$$P\{B\} = \int_B f(t)dt.$$

Donde utilizamos la integral de Lebesgue. Dado que aquí no estudiaremos la integral de Lebesgue y que la mayoría de las probabilidades de los eventos que son de interés se pueden calcular utilizando la integral de Riemann (ya que esta coincide con la de Lebesgue si es que ésta está bien definida) podemos reemplazar la integral de Lebesgue por la integral de Riemann en la mayoría de los casos. Además si F diferenciable en x y f_X es continua en X tenemos que

$$F'_X(x) = \frac{dF(x)}{dx} = f_X(x).$$

Teorema 1.12 *Cada función real no negativa en $\mathcal{R}$ que satisface*

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

es una función de densidad de una variable X absolutamente continua.

Dem.

En vista del teorema 1.9, es suficiente mostrar que la función

$$F(x) = \int_{-\infty}^{x} f(t)dt \quad \text{para toda } x \in \mathcal{R}$$

define una función de distribución. Es claro que

$$\lim_{x\to\infty} F(x) = 1 \text{ y } \lim_{x\to-\infty} F(x) = 0,$$

se cumple y también tenemos que si $x_2 > x_1$,

$$\begin{aligned} F(x_2) &= \textstyle\int_{-\infty}^{x_1} f_X(t)dt + \int_{x_1}^{x_2} f_X(t)dt \\ &\geq \textstyle\int_{-\infty}^{x_1} f_X(t)dt = F(x_1). \end{aligned}$$

Finalmente F es continua, por lo que es continua por la derecha.

Ejemplo 1.24 *Sea X una variable aleatoria, que tiene una función de densidad triangular, es decir, su función de densidad f_X (ver fig.2) está dada por*

$$f_X(x) = \begin{cases} x & si\ 0 < x \leq 1, \\ 2-x & si\ 1 \leq x \leq 2, \\ 0 & e.o.c. \end{cases}.$$

La cual fácilmente se puede checar que es una función de densidad. Entonces X tiene función de distribución F_X (ver fig.3) dada por

$$F_X(x) = \begin{cases} 0 & si\ x \leq 0 \\ \int_0^x t dt = \frac{x^2}{2} & si\ 0 < x \leq 1 \\ \int_0^1 t dt + \int_1^x (2-t) = 2x - \frac{x^2}{2} - 1 & si\ 1 < x \leq 2 \\ 1 & si\ x > 2 \end{cases}.$$

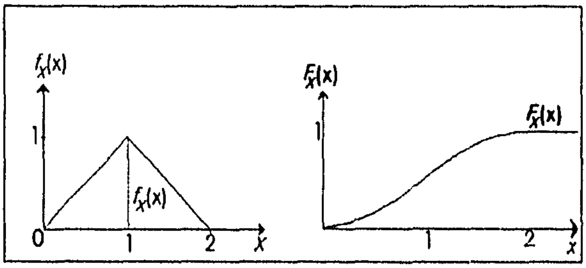


Figura 2. F_X y f_X

Ejemplo 1.25 *Sea $k > 0$ una contante y sea X una variable aleatoria, cuya función de densidad está dada por*

$$f_X(x) = \begin{cases} k(x-1) & si\ 0 < x < 1 \\ 0 & c.o.c. \end{cases}.$$

Entonces, dado que $\int_0^1 f_X(x)dx = k/6$ tenemos que $k = 6$.

Obs. 1.12 *Es importante recalcar nuevamente que las variables aleatorias discretas, y las continuas, son una pequeña parte de toda la familia de variables aleatorias, pero éstas son de gran importancia, ya que son las que surgen más frecuentemente en la práctica, además de esto, si X es una variable aleatoria que no es discreta, ni continua, entonces la función de distribución de X, F_X se puede descomponer como la suma de dos funciones de distribución, una de las cuales es continua y la otra discreta (ver Kai Lai Chung [4] pg. 7,8). Por último, antes de ver los ejemplos de las variables aleatorias más importantes en la práctica, se dará un ejemplo de una variable, la cual no es ni discreta, ni continua.*

Ejemplo 1.26 *La variable aleatoria X tiene por función de distribución F_X dada por*

$$F_X(x) = \begin{cases} 0 & si\ x < 0 \\ \frac{x}{2} & si\ 0 \leq x < 1 \\ \frac{2}{3} & si\ 1 \leq x < 2 \\ \frac{11}{12} & si\ 2 \leq x < 3 \\ 1 & si\ 3 \leq x \end{cases}.$$

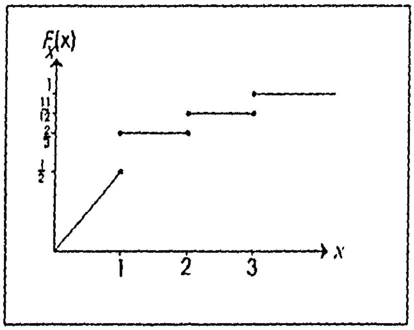


Figura 3. Claramente vemos que F_X no es una función continua, pero tampoco es una función de saltos.

Podemos ver en la figura 3 que F_X no es una función continua, ni tampoco es una función de saltos, por lo cual tenemos que X no es una variable continua ni discreta.

En la siguiente sección daremos una breve reseña de las más importantes variables aleatorias, tanto en el caso discreto, como en el continuo.

1.6.3 Las Variables aleatorias más conocidas

Variables discretas

Variable aleatoria Bernoulli Supongase que tenemos una prueba o un experimento, cuyo resultado puede ser clasificado únicamente de dos formas: "éxito" o "fracaso". Si nosotros, a la variable aleatoria X le damos el valor de 0 si obtenemos un fracaso y le damos el valor 1 si obtenemos un éxito, entonces fácilmente podemos ver que la función de densidad de probabilidades de X está dada por

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

donde p, $0 \leq p \leq 1$ es la probabilidad de obtener un éxito en nuestro experimento. Una variable aleatoria X con tal función de distribución, se le conoce como una variable aleatoria Bernoulli. De ahora en adelante un experimento en el cual surga una variable aleatoria Bernoulli lo llamaremos un ensayo Bernoulli.

Variable aleatoria Binomial Supongase que se realizan n ensayos Bernoulli independientes, donde en cada uno de estos ensayos se obtiene un éxito con probabilidad p ($0 \leq p \leq 1$). Si la variable aleatoria X denota el número de éxitos que se obtuvieron en los n ensayos Bernoulli, entonces, a la variable aleatoria X se le conoce con el nombre de variable aleatoria binomial con parámetros (n, p). En este caso diremos que X se distribuye $Bin(n, p)$. Se puede probar que la función de densidad de probabilidades de una variable aleatoria X que se distribuye Binomial con parámetros (n, p) está dada por

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

donde $\binom{n}{x}$ son la combinaciones de n en x, es decir el número de maneras de tomar x objetos n objetos y esta dado por $\binom{n}{x} = \frac{n!}{(n-x)!x!}$. Note que por el teorema del binomio, tenemos que

$$\begin{aligned} \sum_{i=1}^{n} f_X(1) &= \sum_{i=1}^{n} \binom{n}{i} p^i (1-p)^{n-i} \\ &= [p + (1-p)]^n \\ &= 1. \end{aligned}$$

y claramente, $f_X(x) \geq 0$ para toda $x \in \mathcal{R}$. Es claro que la función de distribución de X está dada por

$$F_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ \sum_{i=1}^{[x]} \binom{n}{i} p^i (1-p)^{n-i} & \text{si } 0 \leq x \leq n \\ 1 & \text{si } x > n \end{cases}.$$

Variable uniforme discreta Supongase que se tiene una urna, dentro de la cual se tienen n bolas exactamente iguales, numeradas del 1 al n, y se toma aleatoriamente una bola de la urna (es decir la probabilidad de tomar cualquiera de ellas es exactamente la misma). Si X denota el número de la bola que se obtuvo entonces decimos que X es una variable aleatoria discreta con parámetro n. Es claro que la función de densidad de probabilidades, está dada por

$$f_X(x) = \begin{cases} \frac{1}{n} & \text{si } x = 1, 2, \ldots, n \\ 0 & \text{e.o.c.} \end{cases},$$

y la función de distribución de X está dada por

$$F_X(x) = \begin{cases} 0 & \text{si } x < 1 \\ \frac{1}{n} & \text{si } x \in [1, 2) \\ \frac{2}{n} & \text{si } x \in [2, 3) \\ \vdots & \\ \frac{n-1}{n} & \text{si } x \in [n-1, n) \\ 1 & \text{si } x \geq 1 \end{cases}.$$

Variable aleatoria geométrica Supongase que se tiene un experimento, consistente en realizar ensayos Bernoulli independientes hasta obtener un éxito, donde cada uno de ellos tiene una probabilidad p $(0 \leq p \leq 1)$ de ser éxito. Si X es la variable aleatoria que denota el número de ensayos requeridos para obtener el primer éxito, entonces diremos que X es una variable aleatoria geométrica con parámetro p. También diremos que X se distribuye geométricamente con parámetro p. Es claro que la función de densidad de probabilidades, está dada por

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

ya que si A_i es el evento, donde se obtiene un fracaso en el i−ésimo ensayo, entonces

$$\begin{aligned} P\{X = j\} &= P\left\{A_1 \cap A_2 \cap \ldots \cap A_{j-1} \cap A_j^c\right\} \\ &= P\{A_1\} P\{A_2\} \cdots P\{A_{j-1}\} P\left\{A_j^c\right\} \\ &= (1-p)^{j-1}p, \end{aligned}$$

ya que los ensayos son independientes.

Además es muy fácil ver que

$$\begin{aligned} \textstyle\sum_{i=1}^{\infty} f_X(i) &= \textstyle\sum_{i=1}^{\infty}(1-p)^{i-1}p \\ &= p\textstyle\sum_{i=1}^{\infty}(1-p)^{i-1} \\ &= p\frac{1}{1-(1-p)} \\ &= 1. \end{aligned}$$

Variable aleatoria Poisson Una variable aleatoria X, la cual toma uno de los valores $0, 1, 2, \ldots$ se dice que es una variable aleatoria Poisson con parámetro λ ($\lambda > 0$), si su función de densidad de probabilidades, está dada por

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

La variable aleatoria Poisson, surge de diversas maneras en una gran cantidad de aplicaciones prácticas, y también como una aproximación a la variable aleatoria binomial cuando la n es suficientemente grande y np es pequeña, de aquí la gran importancia que esta variable aleatoria tiene. Es fácil comprobar que f_X es una función de densidad de probabilidades, ya que

$$\sum_{i=1}^{\infty} e^{-\lambda}\frac{\lambda^i}{i!} = e^{-\lambda}\sum_{i=1}^{\infty}\frac{\lambda^i}{i!} = e^{-\lambda}e^{\lambda} = 1,$$

y por supuesto $f_X(x) \geq 0$ para todo $x \in \mathcal{R}$. Algunos experimentos, donde comúnmente surge esta variable aleatoria son los siguientes:

1. El número de errores en una página de un libro.
2. El número de gente en un pueblo que vive más de cien años de edad.
3. El número de clientes que entran en un cierto banco durante el día.
4. El número de $\alpha-$partículas que se descargan en un material radioactivo en un periodo de tiempo fijo.

Variables absolutamente continuas

Variable aleatoria Uniforme Una variable aleatoria continua X se dice que es una variable aleatoria uniforme en el intervalo (a, b) ($a < b$), si su función de densidad está dada por

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

Es claro que f_X es una función de distribución, ya que $f_X(x) \geq 0$ para toda $x \in \mathcal{R}$, y

$$\int_{-\infty}^{\infty} f_X(t)dt = \int_a^b \frac{1}{b-a}dt = 1.$$

En otras palabras la probabilidad de que X este dentro de un subintervalo de (a,b) es la longitud de este, dividido entre la longitud de (a,b). También tenemos que la función de distribución de X está dada por

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt = \begin{cases} 0 & \text{si } x \leq a \\ \int_0^x \frac{1}{b-a}dt & \text{si } a < x < b \\ 1 & \text{si } b \leq x \end{cases},$$

por lo que

$$F_X(x) = \begin{cases} 0 & \text{si } x \leq a \\ \frac{x-a}{b-a} & \text{si } a < x < b \\ 1 & \text{si } b \leq x \end{cases}.$$

La variable aleatoria uniforme en el intervalo $(0,1)$ es de gran importancia, ya que se puede demostrar teóricamente que si X es una variable aleatoria arbitraria entonces $Y = F_X(X)$ tiene una distribución uniforme en el intervalo $(0,1)$.

Variable aleatoria Exponencial Una variable aleatoria continua cuya función de densidad para $\lambda > 0$ está dada por

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{e.o.c.} \end{cases}$$

se le conoce como variable aleatoria exponencial y es un caso particular de la variable aleatoria Gamma la cual será la siguiente variable aleatoria continua que estudiemos. Es claro que f_X es una función de densidad, ya que

$$\int_{-\infty}^{\infty} f_X(x)dx = \int_0^{\infty} \lambda e^{-\lambda x}dx = -e^{-\lambda x}\,|_0^{\infty} = 1.$$

También tenemos que la función de distribución de X para $x \geq 0$ está dada por

$$F_X(x) = \int_0^x \lambda e^{-\lambda t}dt = 1 - e^{-\lambda x}.$$

Una de las aplicaciones más comúnes de la variable aleatoria Poisson es cuando X representa la duración de vida útil de algún objeto, por ejemplo X puede representar la vida en horas de un foco, o la duración de una llamada telefónica. La variable aleatoria Exponencial posee la propiedad de que es desmemoriada, es decir,

$$P\{X > s + t \mid X > t\} = P\{X > s\},$$

la cual no tiene ninguna otra variable aleatoria.

Variable aleatoria Gamma Una variable aleatoria continua cuya densidad está dada por

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

para algún $\lambda > 0$ y $\alpha > 0$ se le conoce como variable aleatoria Gamma con parámetros α, λ. $\Gamma(\alpha)$ es llamada función gamma y se define por

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

Se puede demostrar que

$$\Gamma(\alpha + 1) = \alpha\Gamma(\alpha),$$

por lo cual se puede probar por inducción que

$$\Gamma(n) = (n-1)!$$

ya que $\Gamma(1) = 1$.

Variable aleatoria normal Decimos que X es una variable aleatoria normal (o que X se distribuye normalmente) con parámetros μ y σ^2 si la función de densidad de X está dada por

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{si } -\infty < x < \infty.$$

La gráfica de esta función de densidad es simétrica alrededor de μ. A la gráfica de la función de densidad de una variable aleatoria normal se le conoce como curva gaussiana (ver fig. 4), en honor a Gauss que fue uno de los primeros matemáticos que la utilizaron.

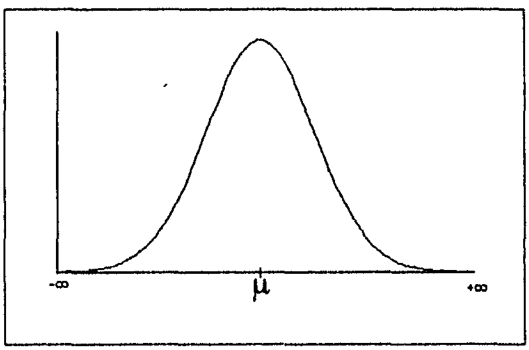


Figura 4. Función de densidad normal.

La variable aleatoria Normal es una de las más importantes variables aleatoria (si no es que la más importante), ya que las sumas de variables aleatorias independientes identicamente distribuidas, se aproximan en la mayoría de las ocasiones a una variable aleatoria normal con parámetros $\mu = 0$ y $\sigma^2 = 1$.

Si X es una variable aleatoria normal con parámetros μ y σ^2, entonces si definimos una nueva variable aleatoria por $Y = aX + b$ con a y b constantes ($a \neq 0$), entonces Y se distribuye normal con parámetros $a\mu + b$ y $a^2\sigma^2$. Para probar esto, calcularemos primero la función de distribución de Y suponiendo que $a > 0$

$$\begin{aligned}
F_X(x) &= P\{Y \leq x\} \\
&= P\{aX + b \leq x\} \\
&= P\left\{X \leq \tfrac{x-b}{a}\right\} \\
&= F_X\left(\tfrac{x-b}{a}\right) \\
&= \int_{-\infty}^{\frac{x-b}{a}} \tfrac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt \\
&= \int_{-\infty}^{x} \tfrac{1}{\sqrt{2\pi}a\sigma} \exp\left\{-\tfrac{(v-(a\mu+b))^2}{2a^2\sigma^2}\right\} dv,
\end{aligned}$$

ya que se hizo el cambio de variable $t = (v - b)/a$. De aquí podemos concluir que Y se distribuye normalmente con parámetros $a\mu + b$ y $a^2\sigma^2$.

Utilizando este resultado, tenemos que si X se distribuye normal con parámetros μ y σ^2, entonces $Z = (X - \mu)/\sigma$ se distribuye normal con parámetros $\mu = 0$ y $\sigma^2 = 1$. En este caso diremos que Z se distribuye normal estándar o también que es una distribución normal unitaria y a en este caso a la función de distribución de Z la denotaremos por Φ.

Hasta aquí no hemos dado una fórmula explícita para la función de distribución de una variable aleatoria normal, y el porque de ésto es que no hay una forma explícita de calcular la integral

$$\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt,$$

por lo que existen tablas donde esta integral se calcula numéricamente, únicamente para el caso donde nuestra variable se distribuye normal estándar, dado que cualquier variable aleatoria normal se puede estandarizar que es lo que hemos hecho en el párrafo anterior. Los valores para $\Phi(z)$ pueden ser obtenidos de la tabla de la distribución normal dada al final de este trabajo.

1.6.4 La función de densidad de una variable aleatoria $Y = g(X)$

En la práctica, es muy común encontrar el problema de que dada una variable aleatoria absolutamente continua X, sea necesario encontrar la función de densidad de una función de ella, es decir, la distribución de $g(X)$. Para ello, es necesario que $\{g(X) \leq y\}$ sea un evento y en tal caso, diremos que g es una función medible. Para esto, ilustraremos el problema con algunos ejemplos, y después enunciaremos un teorema que nos ayudará en algunos casos a encontrar la función de densidad de una función de X, ya que en esta sección consideraremos solamente variables aleatorias continuas.

Ejemplo 1.27 *Sea X una variable aleatoria distribuida uniformemente en el intervalo $[0,1]$. Calcular la función de densidad de $Y = X^n$.*

Sol.

Para ello calcularemos primeramente la función de distribución de Y para $0 \leq y \leq 1$. Esto es

$$\begin{aligned} F_Y(y) &= P\{Y \leq y\} \\ &= P\{X^n \leq y\} \\ &= P\left\{X \leq y^{\frac{1}{n}}\right\} \\ &= F_X(y^{\frac{1}{n}}) \\ &= y^{\frac{1}{n}}, \end{aligned}$$

por lo que derivando tenemos que

$$f_Y(y) = \begin{cases} \frac{1}{n} y^{\frac{1}{n}-1} & si\ 0 \leq y \leq 1 \\ 0 & e.o.c. \end{cases}.$$

En el último ejemplo, es claro que $g(x) = x^n$ es una función estrictamente creciente para $n \geq 1$, además de ser derivable, por lo que es invertible. En general, podemos encontrar muy fácilmente las función de densidad de $Y = g(X)$, cuando g es una función derivable estrictamente creciente o decreciente, lo cual es enunciado en el siguiente teorema que daremos sin demostración, y la cual se basa principalmente en la regla de la cadena.

Teorema 1.13 *Sea g una función derivable, estrictamente creciente o decreciente en un intervalo I, $g(I)$ de g y g^{-1} la función inversa de g. Sea X una variable absolutamente continua que tiene una función de densidad f_X tal que $f_X(x) = 0$ para todo $x \notin I$. Entonces*

$Y = g(X)$ tiene una función de densidad f_Y dada por

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| & si\ y \in g(I) \\ 0 & c.o.c. \end{cases}.$$

Ejemplo 1.28 *Sea X una variable aleatoria con función de densidad f_X y sean a y b constantes tales que $b \neq 0$. Entonces por el teorema 1.13 tenemos que la variable aleatoria $Y = a + bX$ tiene función de densidad dada por*

$$f_Y(y) = \frac{1}{|b|} f_X\left(\frac{y-a}{b}\right) \quad si\ -\infty < y < \infty,$$

ya que $g(x) = a + bx$, por lo que $g^{-1}(y) = \frac{y-a}{b}$ y $\frac{d}{dy} g^{-1}(y) = \frac{1}{b}$. Lo cual podemos comprobar muy fácilmente, si suponemos primero que $b > 0$, por lo cual tenemos que

$$\begin{aligned} F_Y(y) &= P\{Y \leq y\} \\ &= P\{a + bX \leq y\} \\ &= P\left\{X \leq \tfrac{y-a}{b}\right\} \\ &= F_X\left(\tfrac{y-a}{b}\right), \end{aligned}$$

por lo que derivando, tenemos que

$$f_Y(y) = \frac{1}{b} f_X\left(\frac{y-a}{b}\right)$$

donde $1/|b| = 1/b$, análogamente, tenemos que si $b < 0$, entonces

$$\begin{aligned} F_Y(y) &= P\{Y \leq y\} \\ &= P\{a + bX \leq y\} \\ &= P\left\{X \geq \tfrac{y-a}{b}\right\} \\ &= 1 - F_X\left(\tfrac{y-a}{b}\right), \end{aligned}$$

por lo que derivando nuevamente, tenemos que

$$f_Y(y) = -\frac{1}{b} f_X\left(\frac{y-a}{b}\right),$$

donde ahora $1/|b| = -1/b$.

1.7 Ejercicios

1. Considere un comité de cinco personas denotados por A, B, C, D y E. Supongase que se deben seleccionar un director y un secretario para el comité. Asumiendo que

un miembro del comité no puede ocupar los dos cargos, exprese el espacio muestral, asociado a estas selecciones. ¿Cuál es el evento donde A es elegido para alguno de los dos puestos?

2. En cada uno de los siguientes experimentos. ¿Cuál es el espacio muestral?

   (a) El experimento, donde 4 productos son seleccionados aleatoriamente de la producción de una fábrica.

   (b) Un libro es abierto en cualquier página y el número de errores es contado.

   (c) Dos cartas son tomadas (i) con reemplazo, (ii) sin reemplazo de una baraja ordinaria.

3. Sea E, F, G tres eventos. Encontrar expresiones en términos de operaciones de conjuntos, para los eventos donde

   (a) E únicamente ocurra,

   (b) E y F ocurran, pero G no ocurra,

   (c) al menos uno de los tres eventos ocurran,

   (d) los tres eventos ocurran,

   (e) ningún evento ocurra,

   (f) a lo más uno de los eventos ocurra,

   (g) a lo más dos de los eventos ocurran.

4. Si $A \subseteq B$, probar que
$$P\{B - A\} = P\{B\} - P\{A\},$$
donde $A - B = A \cap B^c$.

5. Si $P\{E\} = .9$ y $P\{F\} = .8$, mostrar que $P\{E \cap F\} \geq .7$. En general, mostrar que
$$P\{E \cap F\} \geq P\{E\} + P\{F\} - 1,$$
la cual se le conoce como la desigualdad de Bonferroni.

6. Probar que si A y B son mutuamente excluyentes y si $P\{A\} > 0$ y $P\{B\} > 0$ entonces A y B no pueden ser independientes.

7. Probar el teorema 1.4.

8. Probar la desigualdad de Boole

$$P\left\{\bigcup_{i=1}^{n} E_i\right\} \leq \sum_{i=1}^{n} P\{E_i\}.$$

9. Probar que

$$P\{E \cap F^c\} = 1 - P\{E\} - P\{F\} + P\{E \cap F\}.$$

10. Probar que

$$\begin{aligned} &P\{E_1 \cup E_2 \cup \ldots \cup E_n\} \\ &= \textstyle\sum_{i=1}^{n} P\{E_i\} - \sum_{i_1 < i_2} P\{E_{i_1} \cap E_{i_2}\} + \ldots \\ &+(-1)^{r+1} \sum_{i_1 < i_2 < \ldots < i_r} P\{E_{i_1} \cap E_{i_2} \cap \ldots \cap E_{i_r}\} \\ &+ \ldots + (-1)^{n+1} P\{E_1 \cap E_2 \cap \ldots \cap E_n\}. \end{aligned}$$

11. Probar que si F es un evento fijo tal que $P\{F\} > 0$, entonces la probabilidad condicional de $E \mid F$ es una probabilidad de acuerdo a la definición 1.2.

12. Probar o dar un contraejemplo de los siguientes enunciados:

    (a) Si E es independiente de F y E es independiente de G, entonces E es independiente de $F \cup G$.

    (b) Si E es independiente de F, E es independiente de G y $F \cap G = \emptyset$ entonces E es independiente de $F \cup G$.

    (c) Si E es independiente de F, F es independiente de G y E es independiente de $F \cap G$, entonces G es independiente de $E \cap F$.

13. Sean $E_1, E_2, \ldots, E_n$ eventos aleatorios, entonces

$$P\{E_1 \cap \ldots \cap E_n\} = P\{E_1\} P\{E_2 \mid E_1\} P\{E_3 \mid E_1 \cap E_2\} \ldots P\left\{E_n \mid \bigcap_{i=i}^{n-1} E_i\right\}.$$

14. Probar que si $E_1, E_2, \ldots, E_n$ son eventos completamente independientes, entonces

$$P\{E_1 \cup E_2 \cup \ldots \cup E_n\} = 1 - [1 - P\{E_1\}][1 - P\{E_2\}] \ldots [1 - P\{E_n\}].$$

15. Sea A y B dos eventos tales que $P\{A\} = p_1 > 0$, $P\{B\} = p_2 > 0$, y $p_1 + p_2 > 0$. Mostrar que $P\{B \mid A\} \geq 1 - [(1 - p_2)/p_1]$.

16. Dos dígitos son tomados aleatoriamente sin reemplazo del conjunto de números

$$\{1, 2, 3, 4, 5, 6, 7, 8\}.$$

   (a) Encontrar la probabilidad de que ambos dígitos sean mayores que 5.

   (b) Mostrar que la probabilidad de que la suma de los dígitos sea igual a 5 es la misma que la probabilidad de que la suma de los dígitos sea 13.

17. La probabilidad de que una familia tenga exactamente k niños es $p_k = \alpha p^k$, $k = 1, 2, \ldots, 0 < p < 1$. Supongase que la probabilidad de que cualquiera de los niños tenga ojos verdes es $0 < b < 1$, independientemente de los otros. ¿Cuál es la probabilidad de que una familia elegida aleatoriamente tenga exactamente r $(r \geq 0)$ niños con ojos verdes?

18. En el problema anterior, si tenemos que p_k está dado por:

$$\begin{aligned} p_k &= \alpha p^k \quad \text{si } k = 1, 2, \ldots \\ p_0 &= 1 - \tfrac{\alpha p}{1-p}. \end{aligned}$$

   Supongase que en la distribución del sexo en los k niños es igualmente probable un niño que una niña. Encontrar la probabilidad que una famllia tenga exactamente r niños $(r \geq 0)$. Encontrar la probabilidad condicional de que una familia tenga al menos dos niños dado que tiene al menos un niño.

19. Supongase que una fábrica tiene dos máquinas A y B que hacen el 60% y 40% de la producción respectivamente. En la salida de producción en cada una de las máquinas, la máquina A produce 3% de productos defectuosos, mientras que la máquina B produce 5% de productos defectuosos. Encontrar la probablidad de que dado un producto defectuoso, éste haya sido producido por la máquina B.

20. Un estudiante realiza una examen de selección múltiple, en el cual en cada pregunta se le ofrecen cinco posibles respuestas. Obviamente si el estudiante conoce la respuesta, el elegirá la respuesta correcta. En otro caso el seleccionara una de las cinco respuestas aleatoriamente. suponga que el estudiante conoce las respuestas del 70% de todo el examen.

   (a) ¿Cuál es la probabilidad de que en una pregunta tomada al azar el estudiante dé la respuesta adecuada?

(b) Si el estudiante contesta correctamente una pregunta, ¿cuál es la probabilidad de que el estudiante conozca la respuesta?

21. Suponga que en una prueba para detectar el VIH 90% de los pacientes con dicho virus reaccionaron positivamente, mientras que el 5% de la gente que no presentaba el virus reaccionó positivamente a la prueba. Si en un hospital el 1% de todos los pacientes tiene dicho virus. Calcular la probabilidad de que un paciente elegido aleatoriamente en el hospital reaccione positivamente ante esta prueba.

22. Una urna contiene cinco bolas blancas y cuatro negras. Cuatro bolas son transferidas a una segunda urna. Una bola es tomada de la segunda urna y resultó ser negra. Si otra bola es tomada de la segunda urna. Encontrar la probabilidad de que esa bola sea blanca.

23. Una tipo de probabilidad que no tratamos en este capítulo, pero que nos ayuda a resolver una gran cantidad de problemas, es la *probabilidad geométrica*. Sea Ω un subconjunto en $\mathcal{B}$ con $0 < medida(\Omega) < \infty$ (en este caso la medida de que hablamos puede ser la longitud, el área o volumen de una región en $\mathcal{R}$, $\mathcal{R}^2$ o $\mathcal{R}^3$ o más generalmentegeneralmente una medida definida sobre los conjuntos borelianos de $\mathcal{R}^n$[2]). Si se toma aleatoriamente un punto x de Ω (esto es, que la probabilidad de tomar este punto de distintos subconjuntos de Ω con medidas iguales, es la misma) y $B \in \mathcal{B} \cap \Omega$ es decir $\mathcal{B}$ restringido al conjunto Ω. Entonces la probabilidad de que $x \in B$

$$P\{x \in B\} = \frac{medida(B)}{medida(\Omega)}.$$

Demostrar que esta probabilidad cumple la definición 1.2.

24. Si Mary y Luis se quedan de ver en un restaurante entre la 1 y las dos de la tarde y ambos llegan aleatoriamente en esa hora y cada uno de ellos espera al otro únicamente por 15 minutos. Calcular la probabilidad de que se encuentren.

25. Este problema fue resuelto primeramente por Bertrand, por lo que lleva el nombre de paradoja de Bertrand, y muestra el cuidado que debe de tenerse al utilizarse el término aleatorio.

[2] La medida que comúnmente se utiliza para medir áreas o volumenes en $\mathcal{R}^n$ es la medida de Lebesgue. La cual se estudia en cursos de integración avanzada.

Una cuerda de un círculo unitario es elegida aleatoriamente. Calcular la probabilidad de que la longitud de la cuerda exceda a la longitud de los lados del triángulo inscrito en ella.

Para resolver este problema Bertrand propuso tres soluciones distintas, las cuales llevan a resultados distintos, por lo cual se consideró como una paradoja.

(a) Primeramente el consideró que el círculo es simétrico, por lo que puede fijarse una dirección y tomar aleatoriamente un punto sobre el diámetro con la dirección elegida de la circunferencia para tomar la cuerda perpendicular al diámetro que pasa sobre el punto elegido.

(b) La segunda solución la hizo fijando un punto en la circunferencia, después, eligió un ángulo entre 0 y π el cual será la dirección de la cuerda elegida.

(c) Por último, note que para fijar la posición de la cuerda basta dar su punto medio.

Resuelva el problema, considerando cada uno de estos planteamientos. Después de ello opine el porque de los distintos resultados.

26. Un camino de longitud r es dividido por dos puntos elegidos aleatoriamente. ¿Cuál es la probabilidad de que los nuevos tres segmentos formen un triángulo equilátero?

27. Un punto es tomado aleatoriamente dentro de una esfera de radio R. ¿Cuál es la probabilidad de que la distancia del centro de la esfera al punto más cercano sea mayor o igual que r $(0 \leq r \leq R)$?

28. Probar el teorema 1.5.

29. Probar que si F es una función de distribución, entonces si $b \in \mathcal{R}$ y b_n es una sucesión creciente a b, entonces $\lim_{n\to\infty} F(b_n)$ existe (en este caso $\lim_{n\to\infty} F(b_n)$ es tan solo el límite por la izquierda de F).

30. Probar que si F es una función de distribución, entonces el conjunto de discontinuidades de F es a lo más numerable.

31. Sea X el número de soles al lanzar una moneda. ¿Cuál es Ω? ¿Cuáles son valores que asignan a X los elementos de Ω? ¿Qué representan los eventos $\{X \leq 2.75\}$,

$\{.5 \leq X \leq 1.72\}$?. Además dé la función de densidad de probabilidades para X, asumiendo que la moneda es justa (es decir que la probabilidad de un sol es igual a la de un águila).

32. Un dado es lanzado dos veces. Sea X la suma de sus caras y Y el valor absoluto de su diferencia. ¿Cuál es Ω? ¿Cuáles son los valores asignados a X y Y por los elementos de Ω? ¿Cuál es la función de densidad de probabilidades para X, asumiendo que el dado no está cargado?

33. Si es que sabe programar en algún lenguaje de cómputo, desarrolle un programa en el lenguaje de su preferencia, el cual pueda calcular la función de distribución de una variable aleatoria binomial con parámetros n y p, donde el usuario dé el valor que desea calcular de la función de distribución, así como los valores de n y p.

    Es importante mencionar que si usted calcula 300! de manera directa en casi cualquier computadora, ésta marcará un error, por lo que se recomienda calcular las combinaciones de n en i utilizando la función logaritmo, es decir,

$$\begin{aligned} \ln \tbinom{n}{i} &= \ln n! - \ln(n-i)! - \ln i! \\ &= \textstyle\sum_{j=1}^{n} \ln j - \sum_{j=1}^{n-i} \ln j - \sum_{j=1}^{i} \ln j, \end{aligned}$$

    y claramente

$$\binom{n}{i} = \exp\left(\ln \binom{n}{i}\right),$$

    evitando así los problemas númericos. El objetivo de este ejercicio es mostrar, la dificultad que existe para calcular la función de distribución binomial cuando n es muy grande. Es claro que hace dos siglos, cuando se trabajo por primera vez con esta función de densidad no se podían realizar programas en computadora como el arriba sugerido, por lo que se motivaron diversas aproximaciones a esta variable aleatoria, y lo que dio origen a lo que hoy conocemos como teorema central del límite.

34. ¿Son las siguientes funciones funciones de distribución?:

    (a) $F(x) = \begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } 0 \leq x < \frac{1}{2} \\ 1 & \text{si } x \geq \frac{1}{2} \end{cases}$.

    (b) $F(x) = \frac{1}{\pi} \arctan x$ si $-\infty < x < \infty$.

(c) $F(x) = \begin{cases} 0 & \text{si } x \leq 1 \\ 1 - \frac{1}{x} & \text{si } x > 1 \end{cases}$.

(d) $F(x) = \begin{cases} 1 - e^{-x} & \text{si } x \geq 0 \\ 0 & \text{e.o.c.} \end{cases}$.

35. Sea X con función de distribución dada por F_X.

   (a) Si F_X es la función de distribución definida en el problema 34a. Encontrar $P\left\{X > \frac{1}{4}\right\}$, $P\left\{\frac{1}{3} < X \leq \frac{3}{8}\right\}$.

   (b) Si F_X es la función de distribución definida en el problema 34d. Encontrar $P\{-\infty < X < 2\}$.

36. Encontrar las funciones de densidad de las funciones de distribución definidas en los problemas 34a, 34c y 34d.

37. Sea

$$f_\theta(x) = \begin{cases} \theta^2 x e^{-\theta x} & \text{si } x > 0 \\ 0 & \text{e.o.c.} \end{cases},$$

donde $\theta > 0$. ¿Es f_θ una función de densidad? Encontrar su función de distribución correspondiente?.

38. Sea

$$f_\theta(x) = \begin{cases} \frac{x+1}{\theta(\theta+1)} e^{-\frac{x}{\theta}} & \text{si } x > 0 \\ 0 & \text{e.o.c.} \end{cases},$$

donde $\theta > 0$. En tal caso encontrar su función de distribución.

39. ¿Para qué valores de k las siguientes funciones definen funciones de densidad de probabilidad para alguna variable aleatoria?

   (a) $f(x) = \begin{cases} k\frac{\lambda^x}{x!} & \text{si } x = 0, 1, 2, \ldots \text{ y } \lambda > 0 \\ 0 & \text{e.o.c.} \end{cases}$.

   (b) $f(x) = \begin{cases} \frac{k}{n} & \text{si } x = 1, 2, \ldots, n \\ 0 & \text{e.o.c.} \end{cases}$.

40. Sea la función

$$f(x) = \begin{cases} \frac{\theta\alpha^\theta}{x^{\theta+1}} & \text{si } x > \alpha \\ 0 & \text{e.o.c.} \end{cases},$$

donde α y θ son constantes positivas. Mostrar que f es una función de densidad. Si X es una variable aleatoria con tal función de densidad, entonces se dice que X es una variable aleatoria Pareto con parámetros θ y α.

41. Sea la función

$$f(x) = \frac{1}{2\beta} e^{-\frac{|x-\alpha|}{\beta}} \quad \text{si } -\infty < x < \infty,$$

donde α y β son constantes y $\beta > 0$. Mostrar que una función de densidad. Si X es una variable aleatoria con tal función de densidad, entonces se dice que es una variable aleatoria doble exponencial con parámetros α y β.

42. Supongase que tenemos un experimento, donde se elige un ángulo θ entre $-\pi$ y π radianes uniformemente y desde el punto $(0,1)$ se tira una cuerda hacia abajo, la cual forma con el eje Y un ángulo θ. Si X denota la distancia del origen a la intersección de esta cuerda con el eje X. Calcular la función de distribución de X, a partir de está obtener su función de densidad. Esta variable aleatoria, es conocida con el nombre de variable aleatoria Cauchy y su forma más general se puede obtener, suponiendo que la cuerda es tirada hacia abajo, desde el punto (α, β), donde $\alpha \in \mathcal{R}$ y $\beta > 0$. Calcular la función de distribución y de densidad para X en este último caso.

43. La función de distribución

$$F(x) = \begin{cases} 0 & \text{si } x \leq v \\ 1 - \exp\left\{-\left(\frac{x-v}{\alpha}\right)^{\beta}\right\} & \text{si } x > v \end{cases},$$

donde $\alpha, \beta > 0$ se le conoce como la función de distribución Weibull con parámetros v, α y β. Demostrar que esta función es una función de distribución y encontrar su función de densidad.

44. Si X es normal con parámetros $\mu = 10$ y $\sigma^2 = 36$, calcular

    (a) $P\{X > 5\}$;

    (b) $P\{4 < X < 16\}$;

    (c) $P\{X < 8\}$;

    (d) $P\{X < 20\}$;

    (e) $P\{X > 16\}$

(f) $P\{|X-5|>8\}$.

45. La mediana de una variable aleatoria continua X se define como el valor m tal que $F_X(m)=\frac{1}{2}$, es decir, el valor donde X es igualmente probable que resulte mayor o menor que este. Encontrar la mediana de X, si X se distribuye

    (a) uniformemente sobre el intervalo (a,b);

    (b) normal con parámetros μ y σ^2;

    (c) exponencial con parámetro λ.

46. Si X es una variable aleatoria exponencial con parámetro λ, y $c>0$, mostrar que cX es una variable aleatoria exponencial con parámetro λ/c.

47. Si X es una variable aleatoria exponencial con parámetro λ. Verificar que X también se distribuye como una variable aleatoria gamma con parámetros $\alpha=1$ y λ.

48. Si X es una variable aleatoria normal estándar. Demostrar que X^2 se distribuye como una variable aleatoria gamma con parámetros $\alpha=\frac{1}{2}$ y $\lambda=\frac{1}{2}$.

49. Sea
$$Y=\left(\frac{X-v}{\alpha}\right)^{\beta}.$$
Si X es una variable aleatoria Weibull con parámetros v, α, β $\alpha,\beta>0$. Mostrar que Y es una variable aleatoria exponencial con parámetro $\lambda=1$ y viceversa.

50. Mostrar que $\Gamma(x+1)=x\Gamma(x)$. A partir de aquí demostrar que $\Gamma(n+1)=n!$ para todo $n=0,1,\ldots$

51. Calcular $\Gamma(n+\frac{1}{2})$ para $n=0,1,2,\ldots$

52. Si f_X es la función de distribución de una variable aleatoria gamma con parámetros α y λ verificar que
$$\int_{-\infty}^{\infty} f_X(x)dx=1.$$

53. Si X es una variable aleatoria uniforme distribuida en el intervalo (a,b). ¿Cuál será la relación lineal con U la cual se distribuye uniformemente en $(0,1)$.

54. Sea X una variable aleatoria con función de distribución f_X. Encontrar la función de de la variable aleatoria $Y=a+bX$ $(b\neq 0)$.

55. Encontrar la función de densidad de $Y = e^X$ cuando X se distribuye normalmente con parámetros μ y σ^2. Esta variable aleatoria se dice que tiene una distribución lognormal con parámetros μ y σ^2.

56. Probar el teorema 1.13.

Capítulo 2

Vectores Aleatorios

En un experimento, en muchas ocasiones, el resultado de éste, puede estar expresado en varias cantidades, y deseamos comúnmente calcular la probabilidad de obtener alguna combinación de ellas. Todas esas cantidades pueden ser representadas comúnmente por variables aleatorias, de aquí la necesidad de trabajar no solamente con variables aleatorias individuales, ya que una de esas variables, puede determinar completamente el comportamiento de las otras variables o viceversa. En este capítulo, veremos la forma de calcular probabilidades, de manera conjunta de todas estas variables aleatorias. Para ello daremos la definición de lo que entenderemos por un vector aleatorio, para posteriormente definir una función de distribución conjunta de un vector aleatorio junto con sus principales propiedades. Posteriormente, estudiaremos los vectores aleatorios de tipo discreto y para finalizar estudiaremos los vectores aleatorios de tipo absolutamente continuo, junto con sus principales propiedades.

2.1 Vectores aleatorios

Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad fijo, pero en cualquier caso arbitrario.

Definición 2.1 *El vector* $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ *de variables aleatorias en* $(\Omega, \mathcal{A}, P)$ *definidas por*

$$\mathbf{X}(\omega) = (X_1(\omega), X_2(\omega), \ldots, X_n(\omega)), \quad \omega \in \Omega,$$

es llamado un vector aleatorio de dimensión n si la imagen inversa de cada intervalo en $\mathcal{R}^n$ de la forma

$$I = \{(x_1, x_2, \ldots, x_n) : -\infty < x_i \leq a_i, a_i \in \mathcal{R}, i = 1, 2, \ldots, n\}$$

está también en $\mathcal{A}$, en otras palabras, si

$$\mathbf{X}^{-1}(I) = \{\omega : X_1(\omega) \leq a_1, \ldots, X_n(\omega) \leq a_n\} \in \mathcal{A}$$

Es claro que esta definición es tan solo una extensión de la definición de variable aleatoria para una sola dimensión. Es importante hacer notar, que en este caso, $\mathbf{X}$ es tan solo una función medible definida de Ω en $\mathcal{R}^n$.

De ahora en adelante fijaremos mayor atención para el caso de dimensión $n = 2$, ya que todo lo que hagamos en este caso es muy similar para cuando tenemos que la dimensión es mayor que 2.

Definición 2.2 *La función $F_{X,Y}(\cdot, \cdot)$ definida por*

$$F_{X,Y}(x, y) = P\{X \leq x, Y \leq y\}, \qquad \text{para todo } (x, y) \in \mathcal{R}^2,$$

es conocida como la función de distribución conjunta del vector aleatorio (X, Y). En este caso X y Y son también variables aleatorias, las cuales tienen por su parte funciones de distribución F_X y F_Y y las cuales las llamaremos las funciones de distribución marginales de X y Y respectivamente.

Más adelante mostraremos que existe obviamente una estrecha relación entre $F_{X,Y}$ y F_X y F_Y. Además si para todo $A, B \in \mathcal{B}$, los eventos $\{X \in A\}$ y $\{Y \in B\}$ son eventos independientes, claramente la función de distribución conjunta es tan sólo el producto de las funciones de distribución marginales de las variables que conforman a nuestro vector aleatorio. Antes de ello daremos las propiedades más importantes de las funciones de distribución conjuntas.

Dado que $F_{X,Y}$ define a una probabilidad, se puede mostrar que

i. $F_{X,Y}$ es una función no decreciente y continua por la derecha con respecto a cada coordenada, y ii.

$$\begin{aligned}
\lim_{\substack{x\to+\infty \\ y\to+\infty}} F_{X,Y}(x,y) &= 1, \\
\lim_{y\to-\infty} F_{X,Y}(x,y) &= 0 \qquad \text{para todo } y \in \mathcal{R} \\
\lim_{y\to-\infty} F_{X,Y}(x,y) &= 0 \qquad \text{para todo } x \in \mathcal{R},
\end{aligned}$$

Por otra parte, es fácil comprobar que si (X, Y) es un vector aleatorio y $a_1 < b_1$ y $a_2 < b_2$ entonces

$$\begin{aligned}
P\{a_1 < X \leq b_1, a_2 < Y \leq b_2)\} &= F_{X,Y}(b_1, b_2) + F_{X,Y}(a_1, a_2) \\
&\quad -F_{X,Y}(b_1, a_2) - F_{X,Y}(a_1, b_2)
\end{aligned}$$

(Ver Harris [2], 73).

Por este hecho tenemos que si una función F satisface (i) y (ii) esto no son suficiente, para que F sea la función de distribución de algún vector aleatorio, lo cual puede ser mostrado con el siguiente ejemplo.

Ejemplo 2.1 *Sea F una función de dos variables definida por*

$$F(x,y) = \begin{cases} 0 & si\ x < 0\ o\ x+y < 1\ o\ y < 0, \\ 1 & e.o.c. \end{cases}$$

Es claro que F satisface (i) y (ii). Sin embargo F no es un función de distribución, ya que si suponemos que X y Y son variables aleatorias con dicha función de distribución, entonces

$$\begin{aligned}
P\left\{\tfrac{1}{3} < X \leq 1, \tfrac{1}{2} < Y \leq 1\right\} &= F(1,1) + F(\tfrac{1}{3}, \tfrac{1}{3}) - F(1, \tfrac{1}{3}) - F(\tfrac{1}{3}, 1) \\
&= 1 + 0 - 1 - 1 = -1 < 0.
\end{aligned}$$

Por lo cual, tenemos que F no puede ser una función de distribución conjunta, ya que con ésta se puede calcular probabilidades y obviamente las probabilidades nunca pueden ser negativas. De aquí que sea necesaria otra condición para poder decir que una función es una función de distribución de un vector aleatorio bivariado. Por lo que enunciaremos el siguiente teorema, el cual no será demostrado en estas notas.

Teorema 2.1 *Una función F de dos variables es una función de distribución conjunta de algún vector aleatorio bivariado si y sólo sí esta satisface las siguientes condiciones*

1. *F es una función no decreciente y continua por la derecha con respecto a cada uno de sus argumentos,*

2.
$$\begin{aligned}\lim_{x\to-\infty} F(x,y) &= \lim_{y\to-\infty} F(x,y) = 0 \quad \textit{para toda } x,y \in \mathcal{R},\\ \lim_{\substack{x\to+\infty\\ y\to+\infty}} F(x,y) &= 1\end{aligned}$$

y

3. *para cada* (x_1,y_1), (x_2,y_2) *con* $x_1 < y_1$ *y* $x_2 < y_2$ *la desigualdad*
$$F(x_2,y_2) - F(x_2,y_1) + F(x_1,y_1) - F(x_1,y_2) \geq 0$$

se sostiene.

Es claro que la generalización de la definición de una función de distribución, al caso n−dimensional es la siguiente.

Definición 2.3 *Una función* $F_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n)$ *es la función de distribución conjunta del vector aleatorio* $\mathbf{X} = (X_1,X_2,\ldots,X_n)$ *si*
$$F_{\mathbf{X}}(x_1,x_2,\ldots,x_n) = P\{X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n\}.$$
En este caso las variables aleatorias X_1, X_2, $\ldots$, X_n *son también variables aleatorias con funciones de distribución* F_{X_1}, F_{X_2}, $\ldots$, F_{X_n} *las cuales y para* $(i = 1, \ldots, n)$ F_{X_i} *la llamaremos función de distribución marginal de* X_i.

Análogamente al resultado dado en el teorema 2.1 se tiene el siguiente resultado.

Teorema 2.2 *Una función F de n variables es una función de distribución conjunta de algún vector aleatorio* $\mathbf{X} = (X_1,X_2,\ldots,X_n)$ *si y sólo sí esta satisface las siguientes condiciones:*

1. *F es una función no decreciente y continua por la derecha con respecto a cada uno de sus argumentos,*

2.
$$\begin{aligned}F(-\infty,x_2,\ldots,x_n) &= F(x_1,-\infty,x_3,\ldots,x_n) = \ldots\\ &= F(x_1,\ldots,x_{n-1},-\infty) = 0\\ F(+\infty,+\infty,\ldots,+\infty) &= 1\end{aligned}$$

y

3. *para cada* $(x_1, x_2, \ldots, x_n) \in \mathcal{R}^n$ *y para todo* $\varepsilon_i > 0$ $(i = 1, 2, \ldots, n)$ *la desigualdad*

$$\begin{aligned}
&F(x_1+\varepsilon_1, x_2+\varepsilon_2, \ldots, x_n+\varepsilon_n) \\
&-\sum_{i=1}^{n} F(x_1+\varepsilon_1, \ldots, x_{i-1}+\varepsilon_{i-1}, x_i, x_{i+1}+\varepsilon_{i+1}, \ldots, x_n+\varepsilon_n) \\
&+\sum_{\substack{i,j=1\\ i<j}}^{n} F(x_1+\varepsilon_1, \ldots, x_{i-1}+\varepsilon_{i-1}, x_i, x_{i+1}+\varepsilon_{i+1}, \ldots, x_{j-1}, x_j+\varepsilon_{j-1}, \\
&x_j, x_{j+1}+\varepsilon_{j+1}, \ldots, x_n+\varepsilon_n) \\
&+\ldots \\
&+(-1)^n F(x_1, x_2, \ldots, x_n) \geq 0
\end{aligned}$$

se sostiene.

Aquí nos restringiremos únicamente al caso de vectores aleatorios bivariados, y a los casos, donde estos vectores son absolutamente continuos o discretos en cada entrada. Para ello definiremos primeramente los vectores aleatorios discretos.

Definición 2.4 *Un vector aleatorio bivariado* $\mathbf{X} = (X_1, X_2)$*, es un vector aleatoria discreto si el conjunto de valores* (x_i, y_j) *que puede tomar el vector aleatorio es a lo más numerable, es decir, existe un conjunto* $A \subset \mathcal{R}^2$ *a lo más numerable el cual cumple que*

$$P\{\mathbf{X} \in A\} = 1.$$

En este caso la función $f_{\mathbf{X}}$ *la llamaremos función de densidad de probabilidades conjunta de* $\mathbf{X}$ *y la definiremos por*

$$f_{\mathbf{X}}(x_1, x_2) = P\{X_1 = x_1, X_2 = x_2\}.$$

Es claro que

$$\sum_{(x_1,x_2)\in A} f_{\mathbf{X}}(x_1, x_2) = 1.$$

Además tenemos también que

$$F_{\mathbf{X}}(x_1, x_2) = \sum_{(x_1',x_2')\in B} f_{\mathbf{X}}(x_1', x_2'),$$

donde $B = \{(x_1', x_2') : x_1' \leq x_1, x_2' \leq x_2\}$.

Ejemplo 2.2 *Supóngase que tres bolas son tomadas aleatoriamente de una urna que contiene tres bolas rojas, cuatro blancas, y cinco bolas azules. Si* X *y* Y *denotan respectivamente el número de bolas rojas y blancas que fueron elegidas, entonces es fácil probar que la función de densidad conjunta de probabilidades* $f_{X,Y}$ *de* X *y* Y *está dada por la siguiente tabla*

$i\backslash j$	0	1	2	3	$f_X(i)$
0	$\frac{10}{220}$	$\frac{40}{220}$	$\frac{30}{220}$	$\frac{4}{220}$	$\frac{84}{220}$
1	$\frac{30}{220}$	$\frac{60}{220}$	$\frac{18}{220}$	0	$\frac{108}{220}$
2	$\frac{15}{220}$	$\frac{12}{220}$	0	0	$\frac{27}{220}$
3	$\frac{1}{220}$	0	0	0	$\frac{1}{220}$
$f_Y(j)$	$\frac{56}{220}$	$\frac{112}{220}$	$\frac{48}{220}$	$\frac{4}{220}$	

Es claro que la última columna y el último renglón de esta tabla es tan sólo la suma del renglón y la columna correspondientes. Esto debido a que $f_X(i)$ y $f_Y(j)$ en el caso de que nuestro vector aleatorio sea discreto tenemos que

$$\begin{aligned} f_X(x) &= P\{X = i\} \\ &= \sum_{(x,j)\in A} f_{X,Y}(x,j) \end{aligned}$$

y

$$\begin{aligned} f_Y(y) &= P\{Y = j\} \\ &= \sum_{(i,y)\in A} f_{X,Y}(i,j), \end{aligned}$$

donde A es el conjunto a lo más numerable del cual hablamos en la definición 2.4, ya que la probabilidad de que X tome el valor x es la unión disjunta de los eventos donde $X = x$ y Y toma todos sus posibles valores, la cual es una unión numerable, donde estos dos hechos se pueden comprobar muy fácilmente. En el caso más general es muy fácil extender estas definiciones y dado que nuestro interés principal será los vectores aleatorios absolutamente continuos, a continuación daremos la definición de estos únicamente en el caso bivariado.

Definición 2.5 *Una vector aleatoria bivariado se dice que es absolutamente continuo si existe una función no negativa $f_{X,Y}(\cdot,\cdot)$ tal que para cada $(x,y) \in \mathcal{R}^2$ tenemos que*

$$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u,v)dvdu,$$

donde $F_{X,Y}$ es la función de distribución conjunta de (X,Y). La función $f_{X,Y}$ es llamada función de densidad conjunta de (X,Y).

Claramente, tenemos que

$$\int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f_{X,Y}(u,v)dv \right] du = 1$$

Además si f es una función continua en (x,y) tenemos que

$$\frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y} = f_{X,Y}(x,y). \tag{2.1}$$

Otra propiedad de la función de densidad conjunta de un vector aleatorio bivariado es

$$P\{(X,Y)\in B\} = \int\int_B f_{X,Y}(u,v)dudv, \tag{2.2}$$

para todo B conjunto boreliano de $\mathcal{R}^2$ donde utilizamos la integral de Riemann si es que ésta está bien definida[1]. Por lo que se desprende que

$$P\{a<X<b, c<Y<d\} = \int_a^b \int_c^d f_{X,Y}(u,v)dudv.$$

En el caso de que (X,Y) sea un vector absolutamente continuo, tenemos que X y Y a su vez son variables aleatorias absolutamente continuas con funciones de distribución dadas por

$$F_X(x) = \int_{-\infty}^{x}\int_{-\infty}^{\infty} f_{X,Y}(u,v)dvdu \tag{2.3}$$

y

$$F_Y(y) = \int_{-\infty}^{y}\int_{-\infty}^{\infty} f_{X,Y}(u,v)dudv \tag{2.4}$$

ya que

$$\begin{aligned} P\{X\in A\} &= P\{X\in A, Y\in(-\infty,\infty)\} \\ &= \int_A \int_{-\infty}^{\infty} f_{X,Y}(u,v)dvdu \end{aligned}$$

para todo $A \in \mathcal{R}$ medible. De 2.3 y 2.4 tenemos también por el teorema fundamental del cálculo que las funciones de densidad marginales, tanto de X como de Y están dadas por

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

y

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

Ejemplo 2.3 *La función de densidad conjunta de X y Y está dada por*

$$f_{X,Y}(x,y) = \begin{cases} \frac{6}{7}\left(x^2+\frac{xy}{2}\right) & si\ 0<x<1\ y\ 0<y<2 \\ 0 & e.o.c. \end{cases}$$

(a) Verificar que esta función es realmente una función de densidad para el vector aleatorio (X,Y).

(b) Calcular la función de densidad marginal de X y de Y.

[1]En otro caso podemos utilizar la integral de Lebesgue, pero en la mayoria de los casos prácticos no es necesario.

(c) *Encontrar* $P\{X>Y\}$.

Sol.

Para a) tenemos primero que $f_{X,Y}\geq 0$ *para todo* $(x,y)\in\mathcal{R}^2$, *además tenemos que*

$$\begin{aligned}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f_{X,Y}(x,y)dxdy &= \int_0^2\int_0^1\tfrac{6}{7}\left(x^2+\tfrac{xy}{2}\right)dxdy\\ &= \tfrac{6}{7}\left[\int_0^2\tfrac{x^3}{3}\Big|_0^1 dy+\int_0^2 y\tfrac{x^2}{4}\Big|_0^1 dy\right]\\ &= \tfrac{12}{21}+\tfrac{3}{14}\tfrac{y^2}{2}\Big|_0^2\\ &= \tfrac{12}{21}+\tfrac{3}{7}\\ &= 1,\end{aligned}$$

lo cual comprueba que $f_{X,Y}$ *es la función de densidad de* X *y* Y.

Para b) tenemos que si $x\in(0,1)$ *entonces*

$$\begin{aligned}f_X(x) &= \int_{-\infty}^{\infty}f_{X,Y}(x,y)dy\\ &= \tfrac{6}{7}\int_0^2\left(x^2+\tfrac{xy}{2}\right)dy\\ &= \tfrac{12}{7}x^2+\tfrac{6}{7}x\\ &= \tfrac{6}{7}(2x^2+x),\end{aligned}$$

por lo tanto

$$f_X(x)=\begin{cases}\tfrac{6}{7}(2x^2+x) & si\ 0<x<1\\ 0 & e.o.c.\end{cases},$$

y si $y\in(0,2)$ *entonces tenemos que*

$$\begin{aligned}f_Y(y) &= \int_{-\infty}^{\infty}f_{X,Y}(x,y)dx\\ &= \tfrac{6}{7}\int_0^1\left(x^2+\tfrac{xy}{2}\right)dx\\ &= \tfrac{2}{7}+\tfrac{3y}{14},\end{aligned}$$

por lo cual

$$f_Y(y)=\begin{cases}\tfrac{2}{7}+\tfrac{3y}{14} & si\ 0<y<2\\ 0 & e.o.c.\end{cases}.$$

Para c) tenemos que

$$\begin{aligned}P\{X>Y\} &= \iint_{\{(x,y):x>y\}}f_{X,Y}(x,y)dxdy\\ &= \iint_{\{(x,y):x>y\}\cap(0,1)\times(0,2)}\tfrac{6}{7}\left(x^2+\tfrac{xy}{2}\right)dxdy\\ &= \tfrac{6}{7}\int_0^1\int_0^x\left(x^2+\tfrac{xy}{2}\right)dydx\\ &= \tfrac{6}{7}\int_0^1 x^3+\tfrac{x^3}{4}dx\\ &= \tfrac{15}{56}.\end{aligned}$$

Cuando tenemos la función de densidad conjunta de n variables aleatorias absolutamente continuas, entonces se puede calcular también la función de distribución conjunta de un subconjunto de ellas, es decir, si $X_{i_1}, X_{i_2}, \ldots, X_{i_k}$ con $1 \le k \le n$. Para ello podemos ver que análogamente a la ecuación 2.2 tenemos que

$$P\{\mathbf{X} \in B\} = \int \underset{B}{\ldots} \int f_{\mathbf{X}}(x_1, x_2, \ldots, x_n) dx_1 dx_2 \ldots dx_n,$$

donde $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ y B es un conjunto que representa algún evento del espacio muestral donde estamos trabajando. Así tenemos que si $I = \{i_1, i_2, \ldots, i_k\}$ es un subconjunto de $\{1, 2, \ldots, n\}$ entonces

$$\begin{aligned} F_{\mathbf{X}'}(x_{i_1}, x_{i_2}, \ldots, x_{i_k}) &= P\{X_{i_1} \le x_{i_1}, X_{i_2} \le x_{i_2}, \ldots, X_{i_k} \le x_{i_k}) \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[\int_{-\infty}^{x_{i_1}} \cdots \int_{-\infty}^{x_{i_k}} f_{\mathbf{X}}(t_1, t_2, \ldots, t_n) \prod_{j=1}^{k} dt_{i_j} \right] \prod_{i \in I^c} dt_i, \end{aligned}$$

donde $\mathbf{X}' = (X_{i_1}, X_{i_2}, \ldots, X_{i_k})$. De aquí tenemos que la función de densidad conjunta de $X_{i_1}, X_{i_2}, \ldots, X_{i_k}$ es

$$f_{\mathbf{X}'}(x_{i_1}, x_{i_2}, \ldots, x_{i_k}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\mathbf{X}}(x_1, x_2, \ldots, x_n) \prod_{i \in I^c} dt_i,$$

por lo que a partir de la función de densidad conjunta de $\mathbf{X}$ podemos obtener distribuciones conjuntas de cualquier subconjunto de las n variables aleatorias que conforman nuestro vector aleatorio, así como también las funciones de distribución marginales para cada una de estas variables.

Es claro que si (X, Y) es un vector aleatorio bivariado con función de distribución conjunta $F_{X,Y}$, entonces esta función determina de manera única a las funciones de distribución marginales F_X y F_Y, pero si tenemos que F_X y F_Y son funciones de distribución para X y Y, entonces podemos mostrar que (X, Y) puede tener varias funciones de distribución conjuntas.

Ejemplo 2.4 *Si X y Y son variables aleatorias con funciones de distribución*

$$F_X(x) = \begin{cases} 1 & si\ x \ge 1 \\ \frac{4}{7}x^3 + \frac{3}{7}x^2 & si\ 0 < x < 1 \\ 0 & si\ x \le 0 \end{cases},$$

y

$$F_Y(y) = \begin{cases} 1 & si\ y \geq 2 \\ \frac{2}{7}y + \frac{3}{14}y^2 & si\ 0 < y < 2 \\ 0 & si\ y \leq 0 \end{cases},$$

se puede comprobar muy fácilmente que

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

es una función de distribución conjunta para X y Y, la cual cumple la definición enunciada anteriormente. Pero también tenemos que la función de densidad conjunta vista en el ejemplo 2.3

$$f'_{X,Y}(x,y) = \begin{cases} \frac{6}{7}\left(x^2 + \frac{xy}{2}\right) & si\ 0 < x < 1\ y\ 0 < y < 2 \\ 0 & e.o.c. \end{cases},$$

determina otra función de distribución conjunta para X y Y que claramente no es el producto de las funciones de distribución, con la cual podemos obtener las mismas funciones de distribución marginales para X y Y.

En el ejemplo anterior podemos ver que la función de distribución conjunta establece la relación que existe entre las dos variables aleatorias y si de alguna manera el comportamiento de una de las variables no afecta el comportamiento de la otra y viceversa, entonces podemos ver fácilmente que la función de distribución conjunta de X y Y es tan sólo el producto de las funciones de distribución marginales. Este hecho se puede expresar de manera más exacta definiendo lo que entenderemos por independencia de dos variables aleatorias.

2.2 Variables Aleatorias independientes

Para ilustrar el concepto de independencia de dos variables aleatorias es conveniente ilustar primero éste con un ejemplo.

Ejemplo 2.5 *Sea* $\mathbf{X} = (X_1, X_2)$ *un vector aleatorio el cual se puede obtener tomando un punto aleatoriamente de manera uniforme del cuadrado unitario*

$$U = \{(x,y) : 0 \leq x \leq 1, 0 \leq y \leq 1\}$$

de esta manera podemos ver que

$$P\{\mathbf{X} \in A\} = \frac{\text{area } A \cap U}{\text{area } U} = \text{area } A$$

por lo cual tenemos que

$$F_{\mathbf{X}}(x_1, x_2) = \text{area } B_{x_1,x_2},$$

donde

$$B_{x_1,x_2} = \{(t_1, t_2) \in U : t_1 \leq x_1, t_2 \leq x_2\},$$

donde claramente

$$\text{area } B_{x_1,x_2} = 1$$

si $x_1 > 1$ *y* $x_2 > 1$,

$$\text{area } B_{x_1,x_2} = x_1 x_2$$

si $(x_1, x_2) \in U$,

$$\text{area } B_{x_1,x_2} = x_1$$

si $0 \leq x_1 \leq 1$ *y* $x_2 > 1$,

$$\text{area } B_{x_1,x_2} = x_2$$

si $0 \leq x_2 \leq 1$ *y* $x_1 > 1$ *y*

$$\text{area } B_{x_1,x_2} = 0$$

en otro caso. De aquí podemos ver que aplicando 2.1 tenemos que

$$f_{X_1,X_2}(x_1, x_2) = \begin{cases} 1 & si\ 0 \leq x_1, x_2 \leq 1 \\ 0 & e.o.c. \end{cases}.$$

Además, podemos ver que X_1 *y* X_2 *son variables aleatorias uniformes en el intervalo* $[0, 1]$ *y con esto vemos que*

$$f_{X_1,X_2}(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2).$$

También vemos que el tomar un punto uniformemente en U *se pueda hacer, tomando dos puntos sobre el intervalo* $[0, 1]$ *los cuales determinen la abscisa y la ordenada del punto, y claramente tenemos que estas elecciones son independientes.*

A continuación enunciaremos el concepto de independencia de dos variables aleatorias, el cual está relacionado con la definición de independencia de dos eventos.

Definición 2.6 *X y Y son variables aleatorias independientes si y sólo sí*

$$F_{X,Y}(x,y) = F_X(x)F_Y(y) \text{ para toda } (x,y) \in \mathcal{R}^2.$$

Con esta definición podemos ver que en el ejemplo 2.5 las variables X_1 y X_2 son independientes. A continuación veremos una implicación de esta definición, la cual está enunciada en el siguiente lema.

Lema 2.3 *Si X y Y son variables aleatorias independientes y $a < c$ y $b < d$ números reales, entonces*

$$P\{a < X \leq c, b < Y \leq d\} = P\{a < X \leq c\}\, P\{b < Y \leq d\}.$$

Dem.

Se deja como ejercicio.

Ahora enunciaremos teoremas que nos darán condiciones necesarias y suficientes para la independencia de variables aleatorias en el caso discreto y en el caso absolutamente continuo.

Teorema 2.4 *Una condición necesaria y suficiente para que las variables aleatorias (discretas o absolutamente continuas) X y Y sean independientes es que*

$$f_{X,Y}(x,y) = f_X(x)f_Y(y)$$

para todo $(x,y) \in \mathcal{R}^2$.

Dem.

Nos reduciremos únicamente a la prueba cuando X y Y son variables aleatorias absolutamente continuas, siendo de manera análoga en el caso discreto.

Si $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ para todo $(x,y) \in \mathcal{R}^2$, entonces

$$\begin{aligned} F_{X,Y}(x,y) &= \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(u,v)dvdu \\ &= \int_{-\infty}^{x}\int_{-\infty}^{y} f_X(u)f_Y(v)dvdu \\ &= \int_{-\infty}^{x} f_X(u)du \int_{-\infty}^{y} f_Y(v)du \\ &= F_X(x)F_Y(y), \end{aligned}$$

con lo que concluye nuestra prueba.

Tenemos también un resultado más general que el lema 2.3 y el cual está enunciado en el siguiente teorema.

Teorema 2.5 *X y Y son independientes si y sólo sí*

$$P\{X \in A_1, Y \in A_2\} = P\{X \in A_1\} P\{Y \in A_2\}$$

para todos los conjuntos de Borel A_1 y A_2 en $\mathcal{R}$.

Dem.

La demostración es directa utilizando el teorema 2.4 y el teorema de Fubini para integrales.

También podemos ver que si definimos nuevas variables aleatorias $U = f(X)$ y $V = g(Y)$ con X y Y variables aleatorias independientes, entonces podemos enunciar el siguiente teorema.

Teorema 2.6 *Si X y Y son variables aleatorias independientes y f y g funciones Borel medibles. Entonces $U = f(X)$ y $V = g(Y)$ son también independientes.*

Dem.

Tenemos que

$$P\{f(X) \leq x, g(Y) \leq y\} = P\left\{X \in f^{-1}(-\infty, x], Y \in g^{-1}(-\infty, y]\right\},$$

y por el teorema 2.5 y dado que $f^{-1}(-\infty, x]$ y $g^{-1}(-\infty, y]$ son conjuntos de Borel tenemos que

$$\begin{aligned} P\{X \in f^{-1}(-\infty, x], Y \in g^{-1}(-\infty, y]\} &= P\{X \in f^{-1}(-\infty, x]\} P\{Y \in g^{-1}(-\infty, y]\} \\ &= P\{f(X) \leq x\} P\{g(Y) \leq y\}, \end{aligned}$$

por lo tanto U y V son variables aleatorias independientes.

Esta última condición para la independencia de dos variables aleatorias es una condición necesaria más no suficiente, ya que podemos dar un ejemplo de lo contrario.

Ejemplo 2.6 *Sean X y Y variables aleatorias con función de densidad conjunta*

$$f_{X,Y}(x,y) = \begin{cases} \frac{1+xy}{4} & si\ |x| < 1\ y\ |y| < 1 \\ 0 & e.o.c. \end{cases}$$

Es fácil comprobar que X y Y no son variables aleatorias independientes, ya que por el teorema 2.4 y dado que

$$f_X(x) = \frac{1}{2} \quad si\ |x| < 1 \quad y \quad f_Y(y) = \frac{1}{2} \quad si\ |y| < 1,$$

son las funciones de densidad marginales de X y Y respectivamente. Sin embargo tenemos que las variables aleatorias X^2 y Y^2 son variables aleatorias independientes, ya que para u y v números reales positivos, tenemos que

$$\begin{aligned} P\{X^2 \leq u, Y^2 \leq v\} &= P\{|X| \leq \sqrt{u}, |Y| \leq \sqrt{v}\} \\ &= \int_{-\sqrt{v}}^{\sqrt{v}} \int_{-\sqrt{u}}^{\sqrt{u}} f_{X,Y}(x,y)dxdy \\ &= \tfrac{1}{4} \int_{-\sqrt{v}}^{\sqrt{v}} \int_{-\sqrt{u}}^{\sqrt{u}} [(1+xy)\,dx]\,dy \\ &= \sqrt{u}\sqrt{v} \\ &= P\{X^2 \leq u\} P\{Y^2 \leq v\} \end{aligned}$$

para todo $0 \leq u, v \leq 1$.

Es claro que el concepto de independencia entre dos variables aleatorias, se puede generalizar al caso de más de dos, como se hizo con los conceptos de independencia de dos eventos. Por lo que a continuación se dará una definición más general de este concepto.

Definición 2.7 *Una colección de variables aleatorias $X_1, X_2, \ldots, X_n$, se dice que son mutuamente o completamente independientes si y sólo sí*

$$F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i) \qquad \text{para todo } (x_1, x_2, \ldots, x_n) \in \mathcal{R}^n,$$

donde $F_{X_1,X_2,\ldots,X_n}$ es la función de distribución conjunta de $(X_1, X_2, \ldots, X_n)$ y F_{X_i} ($i = 1, 2, \ldots, n$) es la función de distribución marginal de X_i. X_1, X_2, ..., X_n se dice que son independientes por pares si y sólo si cada X_i y X_j $i \neq j$ ($i, j = 1, 2, \ldots, n$) son independientes.

Obviamente cuando tenemos independencia completa, ésta es mucho más fuerte que tener solamente independencia a pares, lo cual es claramente enunciado por el siguiente teorema.

Teorema 2.7 *Si $X_1, X_2, \ldots, X_n$ son completamente independientes, entonces cada subcolección X_{i_1}, X_{i_2}, ..., X_{i_k} de $X_1, X_2, \ldots, X_n$ son también independientes.*

Dem.

Se deja como ejercicio.

Es claro que por el último teorema independencia completa implica independencia a pares, no así de manera contraria, ya que se puede dar un ejemplo donde se cumpla la independencia a pares y no se cumpla la independencia completa (ver ejercicio 15). Para finalizar este capítulo daremos la definición de una secuencia de variables independientes idénticamente distribuidas y que es uno de los supuestos más fuertes en algunos de los teoremas más importantes en la Teoría de la Probabilidad. Para ello definiremos primero lo que entenderemos por una secuencia de variables aleatorias independientes y lo que significa que dos variables aleatorias sean idénticamente distribuidas.

Definición 2.8 *Una secuencia* $\{X_n\}_{n\in\mathcal{N}}$ *de variables aleatorias se dice que son independientes si para cada* $n = 2, 3, \ldots$ *las variables aleatorias* $X_1, X_2, \ldots, X_n$ *son independientes.*

Definición 2.9 *Decimos que* X *y* Y *son variables aleatorias idénticamente distribuidas si* X *y* Y *tienen la misma función de distribución, esto es*

$$F_X(x) = F_Y(x) \qquad \text{para toda } x \in \mathcal{R},$$

donde F_X *y* F_Y *son las funciones de distribución de* X *y de* Y*, respectivamente.*

Definición 2.10 *Decimos que* $\{X_n\}_{n\in\mathcal{N}}$ *es una secuencia de variables aleatorias independientes idénticamente distribuidas (iid) si* $\{X_n\}_{n\in\mathcal{N}}$ *es una secuencia de variables aleatorias independientes y la distribución de* X_n $(n = 1, 2, \ldots)$ *es la misma que la de* X_1*.*

Es importante mencionar que si dos variables aleatorias X y Y están idénticamente distribuidas, eso no significa que $X = Y$ con probabilidad 1. Si $P\{X = Y\} = 1$ entonces diremos que X y Y son variables aleatorias equivalentes. Es claro que el concepto de variables aleatorias independientes, puede ser extendido hacia los vectores aleatorios como sigue.

Definición 2.11 *Dos vectores aleatorios* $(X_1, X_2, \ldots, X_n)$ *y* $(Y_1, Y_2, \ldots, Y_m)$ *se dice que son independientes si*

$$F_{X_1,\ldots,X_n,Y_1,\ldots,Y_m}(x_1,\ldots,x_n,y_1,\ldots,y_m) = F_{X_1,\ldots,X_n}(x_1,\ldots,x_n)F_{Y_1,\ldots,Y_m}(y_1,\ldots,y_m)$$

para todo $(x_1,\ldots,x_n,y_1,\ldots,y_m) \in \mathcal{R}^{n+m}$.

Es claro también que la independencia de $(X_1, X_2, \ldots, X_n)$ y $(Y_1, Y_2, \ldots, Y_m)$ no implica la independencia de las variables $X_1, X_2, \ldots, X_n$ o de las variables $Y_1, Y_2, \ldots, Y_m$ (ver ejercicio 17).

2.3 Ejercicios

1. Demostrar que si F es una función de distribución de un vector aleatorio bivariado, entonces
$$\begin{aligned}\lim_{\substack{x\to+\infty\\y\to+\infty}} F_{X,Y}(x,y) &= 1,\\ \lim_{y\to-\infty} F_{X,Y}(x,y) &= 0 \qquad \text{para todo } y\in\mathcal{R},\\ \lim_{y\to-\infty} F_{X,Y}(x,y) &= 0 \qquad \text{para todo } x\in\mathcal{R}.\end{aligned}$$

2. Verificar lo que se hizo en el ejemplo 2.4.

3. Demostrar el lema 2.3.

4. Demostrar el teorema 2.4 para el caso discreto.

5. Sea
$$F(x,y)=\begin{cases}1 & \text{si } x+2y\geq 1\\ 0 & \text{e.o.c.}\end{cases}.$$
¿Es F una función de distribución en el plano?

6. Sea T el triángulo formado por los puntos $(0,0)$, $(0,\sqrt{2})$, y $(\sqrt{2},\sqrt{2})$. Sea $F(x,y)$ es área de la intersección de T con la región $\{(x_1,x_2) : x_1\leq x, x_2\leq y\}$. Mostrar que F define una función de distribución en el plano y encontrar las funciones de distribución marginales.

7. Sea (X,Y) un vector aleatorio, el cual tiene una función de densidad conjunta definida por
$$f_{X,Y}(x,y)=\frac{1}{2}$$
para todo (x,y) en el cuadrado cuyas esquinas son los puntos $(1,0)$, $(0,1)$, $(-1,0)$ y $(0,-1)$ y 0 en otro caso. Encontrar las funciones de densidad marginales para X y Y.

8. En el ejemplo 2.5, encontrar la probabilidad de que la distancia de (X,Y) al centro del cuadrado sea mayor que $\frac{1}{4}$.

9. La función de densidad conjunta de X y Y está dada por
$$f(x,y)=\begin{cases}e^{-x-y} & \text{si } 0\leq x,y<\infty\\ 0 & \text{e.o.c.}\end{cases}.$$
Calcular $P\{X>Y\}$ y $P\{X<a\}$.

10. Sea

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

la función de densidad conjunta de (X,Y,Z). Calcular

$$P\{X<Y<Z\} \text{ y } P\{X=Y<Z\}.$$

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

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

¿Son X y Y independientes? ¿Que pasa si $f_{X,Y}$ estuviera dada por

$$f_{X,Y}(x,y)=\begin{cases} 2 & \text{si } 0<x<y, 0<y<1 \\ 0 & \text{e.o.c.} \end{cases}?$$

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

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

Encontrar el valor de c y las funciones de densidad marginales de X y Y.

13. El vector aleatorio bivariado de Cauchy (X,Y) tiene una función de densidad conjunta dada por

$$f_{X,Y}(x,y)=\frac{c}{2\pi}(c^2+x^2+y^2),$$

donde c es una constante positiva. Encontrar las funciones de densidad marginal de X y de Y.

14. El vector aleatoria bivariado gamma (X,Y) tiene una función de densidad conjunta dada por

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

donde α y β son constantes positivas. Encontrar las funciones de densidad marginales para X y Y.

15. Si X_1 y X_2 son variables aleatorias independientes con función de densidad de probabilidades dada por

$$f_X(x) = \begin{cases} \frac{1}{2} & \text{si } x = \pm 1 \\ 0 & \text{e.o.c.} \end{cases}.$$

Si definimos $X_3 = X_1 X_2$. Mostrar que X_1, X_2 y X_3 son independientes a pares pero no completamente.

16. Sea (X_1, X_2, X_3) un vector aleatorio con función de densidad conjunta

$$f_{X_1,X_2,X_3}(x_1, x_2, x_3) = \begin{cases} \frac{1}{4} & \text{si } (x_1, x_2, x_3) \in A \\ 0 & \text{e.o.c.} \end{cases},$$

donde $A = \{(1,0,0), (0,1,0), (0,0,1), (1,1,1)\}$. ¿Son X_1, X_2 y X_3 independientes a pares? ¿Son $X_1 + X_2$ y X_3 independientes?

17. Sea (X, Y) un vector aleatorio con función de densidad conjunta

$$f_{X,Y}(x, y) = \begin{cases} \frac{1}{4}\left[1 + xy(x^2 - y^2)\right] & \text{si } |x| < 1 \text{ y } |y| < 1 \\ 0 & \text{e.o.c.} \end{cases},$$

y (U, V) un vector aleatorio con función de densidad conjunta

$$f_{U,V}(u, v) = \frac{1}{2\pi\sqrt{(1-\rho^2)}} \exp\left\{-\frac{1}{2(1-\rho^2)}(u^2 - 2\rho uv + v^2\right\},$$

donde $|\rho| < 1$ es una constante. Sea la función de densidad conjunta de (X, Y, U, V) dada por

$$f_{X,Y,U,V}(x, y, u, v) = f_{X,Y}(x, y) f_{U,V}(u, v).$$

Mostrar que tanto X y Y como U y V no son independientes, pero si lo es (X, Y) de (U, V).

18. Probar el teorema 2.7.

19. Sea $X_1, X_2, \ldots, X_n$ variables aleatorias con función de densidad conjunta $f_{X_1,X_2,\ldots,X_n}$ y con funciones de densidad marginal f_{X_i} para cada X_i ($i = 1, 2, \ldots, n$). Probar que $X_1, X_2, \ldots, X_n$ son mutuamente independientes si y sólo sí

$$f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i) \qquad \text{para todo } (x_1, x_2, \ldots, x_n) \in \mathcal{R}^n.$$

Capítulo 3

Distribuciones de funciones de vectores aleatorios

En este capítulo se verá la manera de obtener la distribución de variables aleatorias que surgen como funciones de vectores aleatorios. En la primera parte de este capítulo veremos la manera de obtener la distribución de una variable aleatoria la cual surge como una función de un vector aleatorio. En la segunda parte, veremos la manera de como obtener la distribución de un vector aleatorio el cual surge a su vez de otro vector aleatorio, del cual conocemos su distribución.

3.1 La función de densidad de una variable aleatoria $\mathbf{Y} = g(X_1, X_2, \ldots, X_n)$

Un problema fundamental en la teoría de la probabilidad es el de encontrar la distribución de una variable aleatoria $\mathbf{Y}$ la cual surge como una función $\mathbf{Y} = g(X_1, X_2, \ldots, X_n)$ de n variables aleatorias.

Primeramente tomaremos n variables aleatorias absolutamente continuas las cuales tienen una función de densidad conjunta $f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$. La función de densidad conjunta de estas variables aleatorias posee la siguiente propiedad:

$$P\{(X_1, \ldots, X_n) \in A\} = \int \cdots \int_A f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) dx_1 dx_2 \ldots dx_n. \tag{3.1}$$

La función real $g(x_1, \ldots, x_n)$ determina una nueva variable aleatoria Y y estamos interesados en conocer la distribución de esta variable aleatoria, para ello tendremos que calcular:

$$\begin{aligned} F_Y(y) &= P\{Y \leq y\} \\ &= P\{g(X_1, \ldots, X_n) \leq y\}. \end{aligned}$$

Sea $A = \{(x_1, \ldots, x_n) : g(x_1, \ldots, x_n) \leq y\}$, por la primera observación acerca de la función de densidad conjunta tenemos que:

$$\begin{aligned} F_Y(y) &= P\{g(X_1, X_2, \ldots, X_n) \leq y\} \\ &= \int \cdots \int_A f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) dx_1 \ldots dx_n. \end{aligned} \tag{3.2}$$

Ahora obtendremos la función de distribución de la suma de dos variables aleatorias X y Y absolutamente continuas, por lo que tenemos que si $g(x, y) = x+y$ podemos enunciar el siguiente teorema:

Teorema 3.1 *Si X y Y son dos variables aleatorias absolutamente continuas y $Z = X+Y$ entonces su función de densidad de Z está dada como sigue:*

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} f_{X,Y}(x, z-x)dx \\ &= \int_{-\infty}^{\infty} f_{X,Y}(z-x, x)dx. \end{aligned} \tag{3.3}$$

Dem.

Sea $Z = g(X, Y)$. Entonces:

$$\begin{aligned} F_Z(z) &= P\{Z \leq z\} \\ &= P\{X+Y \leq z\} \\ &= \iint_{x+y \leq z} f_{X,Y}(x, y)dxdy. \end{aligned}$$

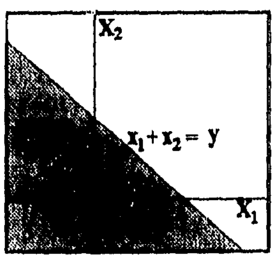


Figura 1.

Ahora bien podemos escribir la integral anterior como sigue (ver fig. 1):

$$\begin{aligned} F_Z(z) &= \int_{-\infty}^{\infty}\int_{-\infty}^{z-x} f_{X,Y}(x,y)dydx \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{z} f_{X,Y}(x,u-x)dudx, \end{aligned}$$

fijando x y haciendo $u = x+y$ en la integral con respecto a y, se tiene que $du = dy$, y además los límites de integración con respecto a y cambian y se tiene que si $y \in (-\infty, z-x)$ entonces $u \in (-\infty, z)$. Por lo anterior, y utilizando el teorema fundamental del cálculo se tiene que:

$$\begin{aligned} f_Z(z) &= F_Z'(z) \\ &= \int_{-\infty}^{\infty} f_{X,Y}(x,z-x)dx. \end{aligned}$$

Por otro lado, podemos demostrar también de forma análoga lo siguiente:

$$f_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(z-x,x)dx.$$

Por consiguiente, tenemos

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} f_{X,Y}(x,z-x)dx \\ &= \int_{-\infty}^{\infty} f_{X,Y}(z-x,x)dx. \end{aligned}$$

Es claro que si X y Y son dos variables aleatorias independientes tendremos que:

$$\begin{aligned} f_{X+Y}(z) &= \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)dx \qquad (3.4) \\ &= \int_{-\infty}^{\infty} f_X(z-x)f_Y(x)dx \end{aligned}$$

Esta última función de densidad recibe el nombre de convolución de las funciones $f_X(.)$ y $f_Y(.)$, y se puede escribir de la siguiente manera: $f_{X+Y}(.) = f_X(.) * f_Y(.)$.

En el caso de tener variables aleatorias discretas, tenemos un resultado análogo a las ecuaciones 3.3 y 3.4, el cual se da en el siguiente teorema:

Teorema 3.2 *Si X y Y son variables aleatorias discretas con función de densidad $f_{X,Y}$, entonces la variable aleatoria $Z = X+Y$ tiene una función de densidad de la siguiente forma:*

$$\begin{aligned} f_{X+Y}(z) &= \sum_{x:f_{X,Y}(x,z-x)>0} f_{X,Y}(x,z-x) \qquad (3.5) \\ &= \sum_{x:f_{X,Y}(z-x,x)>0} f_{X,Y}(z-x,x). \end{aligned}$$

Dem.

Como $f_{X+Y}(z) = P\{X+Y=z\}$ y puesto que el evento

$$\{X+Y=z\} = \bigcup_{x:f_{X,Y}(x,z-x)>0} \{X=x, Y=z-x\}$$

para toda $z \in \mathcal{R}$ se tiene que estos eventos claramente son mutuamente excluyentes, y

$$\begin{aligned}
f_{X+Y}(z) &= P\{X+Y=z\} \\
&= P\left\{\bigcup_{x:f_{X,Y}(x,z-x)>0} \{X=x, Y=z-x\}\right\} \\
&= \sum_{x:f_{X,Y}(x,z-x)>0} P\{X=x, Y=z-x\} \\
&= \sum_{x:f_{X,Y}(x,z-x)>0} f_{X,Y}(x, z-x).
\end{aligned}$$

En forma análoga se puede demostrar la segunda igualdad.

De forma similar a la suma de variables aleatorias independientes absolutamente continuas , tenemos que

$$\begin{aligned}
f_{X+Y}(z) &= \sum_{x:f_X(x)\ f_Y(z-x)>0} f_X(x) f_Y(z-x) \qquad (3.6)\\
&= \sum_{x:f_X(z-x)\ f_Y(x)>0} f_X(z-x) f_Y(x).
\end{aligned}$$

En el caso del producto de dos variables aleatorias tenemos el siguiente resultado:

Teorema 3.3 *Sean X y Y dos variables absolutamente continuas con función de densidad conjunta $f_{X,Y}(x,y)$, entonces:*

$$\begin{aligned}
f_{XY}(z) &= \int_{-\infty}^{\infty} \frac{1}{|x|} f_{X,Y}(x, \frac{z}{x}) dx \qquad (3.7)\\
&= \int_{-\infty}^{\infty} \frac{1}{|x|} f_{X,Y}(\frac{z}{x}, x) dx.
\end{aligned}$$

Dem.

Utilizando 3.2 tenemos que

$$\begin{aligned}
F_{XY}(z) &= P\{XY \leq z\} \\
&= \iint_{xy\leq z} f_{X,Y}(x,y) dx dy,
\end{aligned}$$

si $z \geq 0$, (ver figura 2) podemos expresar esta integral como sigue:

$$F_{XY}(z) = \int_{-\infty}^{0}\int_{\frac{z}{x}}^{\infty} f_{X,Y}(x,y)dydx + \int_{0}^{\infty}\int_{-\infty}^{\frac{z}{x}} f_{X,Y}(x,y)dydx.$$

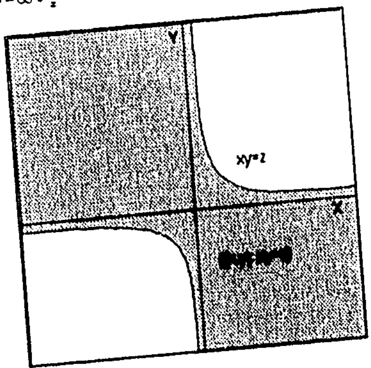


Figura 2. $z \geq 0$

Fijando x y haciendo el cambio de variable $u = xy$ en la integral con respecto a y, se tiene $y = \frac{u}{x}$ y $dy = \frac{du}{x}$, por lo que los límites de integración cambian en la integral con respecto a y, esto es que si $y \in (-\infty, \frac{z}{x})$, entonces $u \in (z, -\infty)$, a partir de aquí tenemos que

$$\begin{aligned}
F_{XY}(z) &= \int_{-\infty}^{0}\int_{z}^{-\infty} f_{X,Y}(x,\frac{u}{x})\frac{du}{x}dx + \int_{0}^{\infty}\int_{-\infty}^{z} f_{X,Y}(x,\frac{u}{x})\frac{du}{x}dx \\
&= \int_{-\infty}^{0}\int_{-\infty}^{z} f_{X,Y}(x,\frac{u}{x})(-\frac{1}{x})dudx + \int_{0}^{\infty}\int_{-\infty}^{z} f_{X,Y}(x,\frac{u}{x})\frac{1}{x}dudx \\
&= \int_{-\infty}^{0}\int_{-\infty}^{z} f_{X,Y}(x,\frac{u}{x})\left|\frac{1}{x}\right|dudx + \int_{0}^{\infty}\int_{-\infty}^{z} f_{X,Y}(x,\frac{u}{x})\left|\frac{1}{x}\right|dudx \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{z} f_{X,Y}(x,\frac{u}{x})\left|\frac{1}{x}\right|dudx.
\end{aligned}$$

Por último, utilizando el teorema fundamental del cálculo podemos ver

$$\begin{aligned}
f_{XY}(z) &= F'_{XY}(z) \\
&= \int_{-\infty}^{\infty} f_{X,Y}(x,\frac{z}{x})\left|\frac{1}{x}\right|dx.
\end{aligned}$$

De la misma manera se puede demostrar también el resultado para $z < 0$. La otra parte se demuestra de manera similar.

Con el mismo método se pueden demostrar fórmulas para la diferencia y el cociente de dos variables aleatorias absolutamente continuas, las cuales son

$$\begin{aligned} f_{X-Y}(z) &= \int_{-\infty}^{\infty} f_{X,Y}(z+x,x)dx \\ &= \int_{-\infty}^{\infty} f_{X,Y}(x,x-z)dx, \end{aligned} \tag{3.8}$$

$$f_{\frac{X}{Y}}(z) = \int_{-\infty}^{\infty} |x|\, f_{X,Y}(zx,x)dx. \tag{3.9}$$

Es claro que cuando X y Y son variables aleatorias independientes se tiene que las formulas anteriores serán:

$$\begin{aligned} f_{XY}(z) &= \int_{-\infty}^{\infty} f_X(x) f_Y(\frac{z}{x}) \left|\frac{1}{x}\right| dx \\ &= \int_{-\infty}^{\infty} f_X(\frac{z}{x}) f_Y(x) \left|\frac{1}{x}\right| dx, \end{aligned} \tag{3.10}$$

$$\begin{aligned} f_{X-Y}(z) &= \int_{-\infty}^{\infty} f_X(z+x) f_Y(x)dx \\ &= \int_{-\infty}^{\infty} f_X(x) f_Y(x-z)dx, \end{aligned} \tag{3.11}$$

$$f_{\frac{X}{Y}}(z) = \int_{-\infty}^{\infty} |x|\, f_X(zx) f_Y(x)dx. \tag{3.12}$$

A continuación se darán ejemplos de distribuciones que se obtienen por medio de una función de variables aleatorias.

Ejemplo 3.1 *Sean X y Y variables aleatorias independientes discretas con distribuciones binomial con parámetros (n,p) y (m,p) respectivamente. Encontrar la función de distribución de $Z = X + Y$.*

Sol.

Aplicando 3.6 tenemos que

$$\begin{aligned} f_Z(z) &= \sum_{x:\, f_X(x) f_Y(z-x)>0} f_X(x) f_Y(z-x) \\ &= \sum_{i=0}^{n} f_X(i) f_Y(z-i) \\ &= \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} \binom{m}{z-i} p^{z-i} (1-p)^{m-z+i} \end{aligned}$$

$$\begin{aligned}
&= \sum_{i=0}^{n} p^i p^{z-i}(1-p)^{n-i}(1-p)^{m-z+i}\binom{n}{i}\binom{m}{z-i} \\
&= \sum_{i=0}^{n} p^z(1-p)^{n+m-z}\binom{n}{i}\binom{m}{z-i} \\
&= p^z(1-p)^{n+m-z}\sum_{i=0}^{n}\binom{n}{i}\binom{m}{z-i} \\
&= \binom{m+n}{z}p^z(1-p)^{n+m-z},
\end{aligned}$$

ya que $\sum_{i=0}^{n}\binom{n}{i}\binom{m}{z-i} = \binom{m+n}{z}$, donde $\binom{i}{j} = 0$ si $i < j$, de aquí se tiene que

$$f_Z(z) = \begin{cases} \binom{m+n}{z}p^z(1-p)^{n+m-z} & \text{si } z = 0, 1, \ldots, n \\ 0 & \text{e.o.c.} \end{cases},$$

por lo que $Z = X + Y$ también se distribuye binomial con parámetros $(n+m, p)$.

Ejemplo 3.2 *Sean X y Y variables aleatorias independientes con una distribución Poisson con parámetros λ_1 y λ_2 respectivamente. Encontrar la distribución de $Z = X + Y$.*

Sol.

Aplicando nuevamente 3.6 tenemos que

$$\begin{aligned}
f_Z(z) &= \sum_{x: f_X(x)f_Y(z-x)>0} f_X(x)f_Y(z-x) \\
&= \sum_{x: f_X(x)f_Y(z-x)>0} f_X(x)f_Y(z-x) \\
&= \sum_{i=0}^{\infty} f_X(i)f_Y(z-i) \\
&= \sum_{i=0}^{z} f_X(i)f_Y(z-i) \\
&= \sum_{i=0}^{z} e^{-\lambda_1}\frac{\lambda_1{}^i}{i!}e^{-\lambda_2}\frac{\lambda_2{}^{z-i}}{(z-i)!} \\
&= e^{-\lambda_1-\lambda_2}\sum_{i=0}^{z}\frac{\lambda_1{}^i\lambda_2{}^{z-i}}{i!(z-i)!} \\
&= \frac{e^{-\lambda_1-\lambda_2}}{z!}\sum_{i=0}^{z}\frac{z!}{i!(z-i)!}\lambda_1{}^i\lambda_2{}^{z-i},
\end{aligned}$$

es claro que la suma es el desarrollo del binomio $(\lambda_1 + \lambda_2)^z$ por lo cual tenemos que

$$f_Z(z) = \begin{cases} \frac{e^{-\lambda_1-\lambda_2}}{z!}(\lambda_1+\lambda_2)^z & \text{si } z = 0, 1, 2, \ldots \\ 0 & \text{e.o.c.} \end{cases},$$

por lo cual $Z = X + Y$ tiene distribución Poisson con parámetros $\lambda_1 + \lambda_2$.

Ejemplo 3.3 *Sean X y Y variables aleatorias independientes cada una con distribución uniforme en (a, b). Encontrar la función de densidad de la variable aleatoria $Z = X + Y$.*

Sol.

Aplicando la formula 3.4 tenemos que

$$
\begin{aligned}
f_{X+Y}(z) &= \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)dx \\
&= \int_a^b f_X(x) f_Y(z-x)dx,
\end{aligned}
$$

ya que $f_X(x) = 0$ si $x \notin (a, b)$, de la misma forma

$$f_Y(z-x) = 0 \text{ si } z - x \notin (a, b).$$

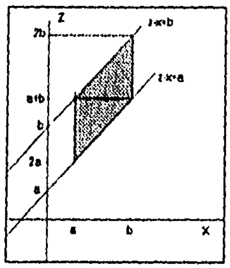


Figura 3.

Observando la fig. 3 tenemos que si $z \in (2a, a+b)$ entonces $x \in (a, z-a)$, y si $z \in [a+b, 2b)$ entonces $x \in [z-b, b)$, por lo que

$$
\begin{aligned}
f_Z(z) &= \begin{cases} \int_a^{z-a} f_X(x) f_Y(z-x)dx & \text{si } z \in (2a, a+b) \\ \int_{z-b}^{b} f_X(x) f_Y(z-x)dx & \text{si } z \in [2a, a+b) \\ 0 & \text{e.o.c.} \end{cases} \\
&= \begin{cases} \int_a^{z-a} \frac{1}{(b-a)^2} dx & \text{si } z \in (2a, a+b) \\ \int_{z-b}^{b} \frac{1}{(b-a)^2} dx & \text{si } z \in [2a, a+b) \\ 0 & \text{e.o.c.} \end{cases},
\end{aligned}
$$

por último tenemos que

$$f_Z(z) = \begin{cases} \frac{z-2a}{(b-a)^2} & \text{si } z \in (2a, a+b) \\ \frac{2b-z}{(b-a)^2} & \text{si } z \in [2a, a+b) \\ 0 & \text{e.o.c.} \end{cases}.$$

La distribución de Z se le conoce como la distribución de Simpson.

Otra manera de calcular la función de densidad de la variable aleatoria Z, es primero calcular la función de distribución, como sigue

$$\begin{aligned} F_Z(z) &= P\{Z \leq z\} \\ &= P\{X + Y \leq z\} \\ &= \int\int_{x+y \leq z} f_X(x) f_Y(y) dx dy. \end{aligned}$$

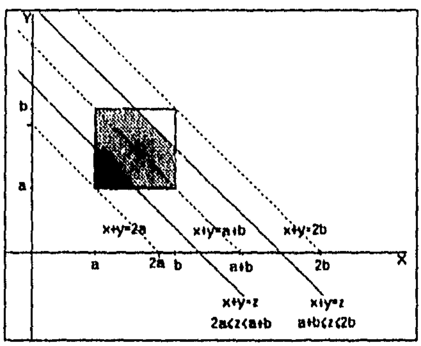


Figura 4.

La última integral se puede calcular fácilmente razonando geométricamente dado que $f_X(x)f_Y(y)$ es distinto de 0 únicamente cuando $(x, y) \in [a, b] \times [a, b]$ y observando la figura 4 se tiene:

$$F_Z(z) = \begin{cases} 0 & \text{si } z \leq 2a \\ \frac{1}{(b-a)^2} \frac{(z-2a)^2}{2} & \text{si } z \in (2a, a+b) \\ \frac{1}{(b-a)^2} \frac{(2b-z)^2}{2} & \text{si } z \in [2a, a+b) \\ 1 & \text{si } z \geq 2b \end{cases}.$$

Si derivamos y simplificamos esta función de distribución, obtendremos la misma función de densidad anterior.

Ejemplo 3.4 *Sean X y Y variables aleatorias exponenciales con parámetros λ_1 y λ_2 respectivamente. Encontrar la distribución de $Z = \frac{X}{Y}$. También calcular $P\{X < Y\}$.*

Sol.

Aplicando 3.12 se tiene

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} |x| f_X(zx) f_Y(x) dx \\ &= \int_0^{\infty} x f_X(zx) f_Y(x) dx, \end{aligned}$$

ya que si $zx \leq 0$ entonces $f_X(zx) = 0$, además si $x < 0$ entonces $f_X(x) = 0$, de aquí se tiene que para $z \geq 0$

$$\begin{aligned} f_Z(z) &= \int_0^{\infty} x\lambda_1 e^{-\lambda_1(zx)}\lambda_2 e^{-\lambda_2 x} dx \\ &= \lambda_1\lambda_2 \int_0^{\infty} x e^{-(\lambda_1.z+\lambda_2)x} dx \end{aligned}$$

si $u = (\lambda_1.z + \lambda_2)x$, $du = (\lambda_1.z + \lambda_2)\ du$ se tiene:

$$\begin{aligned} f_Z(z) &= \lambda_1\lambda_2 \int_0^{\infty} \frac{u}{(\lambda_1.z+\lambda_2)} e^{-u} \frac{1}{(\lambda_1.z+\lambda_2)} du \\ &= \frac{\lambda_1\lambda_2}{(\lambda_1.z+\lambda_2)^2} \int_0^{\infty} u e^{-u} du \end{aligned}$$

si $w = u$, $dw = du$, $dv = e^{-u}du$, $v = -e^{-u}$

$$\begin{aligned} f_Z(z) &= \frac{\lambda_1\lambda_2}{(\lambda_1.z+\lambda_2)^2} \left[-ue^{-u}|_0^{\infty} + \int_0^{\infty} e^{-u} du \right] \\ &= \frac{\lambda_1\lambda_2}{(\lambda_1.z+\lambda_2)^2} \left[-e^{-u}|_0^{\infty} \right] \end{aligned}$$

por lo que

$$f_Z(x) = \begin{cases} \frac{\lambda_1\lambda_2}{(\lambda_1.z+\lambda_2)^2} & \text{si } z > 0 \\ 0 & \text{e.o.c.} \end{cases}.$$

Ahora para calcular $P\{X < Y\}$ se tiene:

$$\begin{aligned} P\{X < Y\} &= P\left\{ \frac{X}{Y} < 1 \right\} \\ &= \int_0^1 \frac{\lambda_1\lambda_2}{(\lambda_1.z+\lambda_2)^2} dz \end{aligned}$$

si $u = \lambda_1.z + \lambda_2$, $du = \lambda_1 dz$ entonces

$$\begin{aligned} P\{X < Y\} &= \lambda_1\lambda_2 \int_{\lambda_1}^{\lambda_1+\lambda_2} u^{-2}\frac{1}{\lambda_1}du \\ &= \lambda_2 \frac{u^{-1}}{-1}\Big|_{\lambda_1}^{\lambda_1+\lambda_2} \\ &= \lambda_2\left[\frac{-1}{\lambda_1+\lambda_2}+\frac{1}{\lambda_2}\right] \\ &= \frac{\lambda_1}{\lambda_1+\lambda_2}, \end{aligned}$$

por lo cual se tiene que

$$P\{X < Y\} = \frac{\lambda_1}{\lambda_1+\lambda_2}.$$

3.2 La función de densidad conjunta de un vector aleatorio $\mathbf{U} = g(X_1, X_2, \ldots, X_n)$

En la sección anterior calculamos la densidad de una sola variable aleatoria que surge como una función de n variables aleatorias. En esta sección consideraremos que tenemos n variables aleatorias de este tipo, y lo que deseamos es calcular la densidad conjunta de ellas, en otras palabras si $\mathbf{X} = (X_1, \ldots, X_n)$ es un vector aleatorio con n componentes, consideraremos una función $\mathbf{Y} = G(X)$, donde $G : \mathcal{R}^n \to \mathcal{R}^n$, en otras palabras $\mathbf{Y} = (Y_1, \ldots, Y_n)$ donde cada componente $Y_i = G_i(X_1, \ldots, X_n), i = 1, 2, \ldots, n$. Es claro que si $\mathbf{X}$ es un vector de variables aleatorias absolutamente continuas se tiene que $\mathbf{Y} = (Y_1 = G_1(X_1, \ldots, X_n), \ldots, Y_n = G_n(X_1, \ldots, X_n))$.

Para poder obtener la función de densidad conjunta de Y se tiene que observar una propiedad de la función de densidad conjunta de X, la cual se da a continuación.

Si T es cualquier conjunto en el rango de Y se tiene que

$$\begin{aligned} P\{\mathbf{Y} \in T\} &= P\left\{\mathbf{X} \in G^{-1}(T)\right\} \\ &= \int \cdots \int_{G^{-1}(T)} f_{\mathbf{X}}(x_1, \ldots, x_n)dx_1 \ldots dx_n, \end{aligned}$$

donde $G^{-1}(T) = \{x \in \mathcal{R}^n : G(x) \in T\}$ y $f_{\mathbf{X}}(x_1, \ldots, x_n)$ es la función de densidad conjunta de las variables aleatorias $X_1, \ldots, X_n$.

Si definimos el conjunto

$$T_{\mathbf{y}} = \{\mathbf{y} \in \mathcal{R}^n : Y_1 \le y_1, Y_2 \le y_2, \ldots, Y_n \le y_n\},$$

podemos enunciar el siguiente teorema:

Teorema 3.4 *Si* $\mathbf{X}$ *es un vector aleatorio de variables absolutamente continuas con función de densidad* $f_{\mathbf{X}}(x_1,\ldots,x_n)$ *y* $\mathbf{Y}=G(\mathbf{X})$ *una función cuyo rango es un subconjunto de* $\mathcal{R}^n$ *entonces*

$$F_{\mathbf{Y}}(y_1,\ldots,y_n)=\int\limits_{G^{-1}(T_y)}\cdots\int f_{\mathbf{X}}(x_1,..,x_n)dx_1\ldots dx_n.$$

Dem.

Es inmediata observando 3.1.

Un resultado análogo se puede obtener para variables aleatorias discretas.

Teorema 3.5 *Si* $\mathbf{X}$ *es un vector aleatorio discreto, entonces la función de densidad de* $\mathbf{Y}$, $f_{\mathbf{Y}}(y_1,\ldots,y_n)$, *es*

$$\begin{aligned} f_{\mathbf{Y}}(y_1,\ldots,y_n) &= P\{\mathbf{Y}=\mathbf{y}\} \\ &= P\{\mathbf{X}=G^{-1}(\mathbf{y})\}. \end{aligned}$$

Si tenemos que la función G es una función continua, y que cada imágen de $y\in\mathcal{R}^n$ puede tener a lo más un número infinito numerable de inversas. Podemos introducir la familia de transformaciones inversas siguiente:

$$\left.\begin{array}{l} X_1=G_{1j}^{-1}(Y_1,\ldots,Y_n) \\ X_2=G_{2j}^{-1}(Y_1,\ldots,Y_n) \\ \vdots \\ X_n=G_{nj}^{-1}(Y_1,\ldots,Y_n) \end{array}\right\} j=1,2,..,$$

las cuales son funciones bien definidas de $\mathcal{R}^n$ en $\mathcal{R}$. De acuerdo a lo anterior nosotros podemos replantear los teoremas 3.4 y 3.5 en términos de transformaciones inversas.

Corolario 3.6 *Si* $\mathbf{X}$ *es un vector aleatorio absolutamente continuo y* $G(\mathbf{x})$ *es una función continua que cumple las suposiciones anteriores entonces*

$$F_{\mathbf{Y}}(y_1,\ldots,y_n)=\sum_j\int\limits_{G_j^{-1}(T_y)}\cdots\int f_{\mathbf{X}}(x_1,\ldots,x_n)dx_1\ldots dx_n,$$

donde $G_j^{-1}=(G_{1j}^{-1},G_{2j}^{-1},\ldots,G_{nj}^{-1})$ *la cuales son funciones de* $\mathcal{R}^n$ *en* $\mathcal{R}^n$.

Corolario 3.7 *Si X es un vector aleatorio discreto y G es una función continua, entonces*

$$\begin{aligned} P\{Y_1 = y_1, \dots, Y_n = y_n\} &= f_{\mathbf{Y}}(\mathbf{y}) \\ &= \sum_j P\left\{X_1 = G_{1j}^{-1}(y_1,\dots,y_n),\dots,X_n = G_{nj}^{-1}(y_1,\dots,y_n)\right\}. \end{aligned}$$

A continuación enunciaremos un teorema que es análogo al teorema de cambio de variable en una sola variable.

Teorema 3.8 *Sea $\mathbf{X}$ un vector de variables aleatorias absolutamente continuas con función de densidad $f_{\mathbf{X}}(\mathbf{x})$ y supongase que para $x_i = G_{ij}^{-1}(y_1,\dots,y_n)$, $\frac{\partial G_{ij}^{-1}}{\partial y_k}$ es continua para $i,k = 1,2,\dots,n$ y para $j = 1,2,\dots$, además que cada jacobiano J_i de las transformaciones inversas*

$$J_j = \begin{vmatrix} \frac{\partial G_{1j}^{-1}}{\partial y_1} & \frac{\partial G_{1j}^{-1}}{\partial y_2} & \cdots & \frac{\partial G_{1j}^{-1}}{\partial y_n} \\ \frac{\partial G_{2j}^{-1}}{\partial y_1} & \frac{\partial G_{2j}^{-1}}{\partial y_2} & \cdots & \frac{\partial G_{2j}^{-1}}{\partial y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial G_{nj}^{-1}}{\partial y_1} & \frac{\partial G_{nj}^{-1}}{\partial y_2} & \cdots & \frac{\partial G_{nj}^{-1}}{\partial y_n} \end{vmatrix}$$

no es cero. Entonces si para cada $\mathbf{y}$ en el rango de G hay a lo más un número numerable de inversas, se tiene que

$$f_{\mathbf{Y}}(\mathbf{y}) = \sum_j f_{\mathbf{X}}\left(G_{1j}^{-1}(y_1,\dots,y_n),\dots,G_{nj}^{-1}(y_1,\dots,y_n)\right)|J_j|.$$

<u>Dem.</u>

Para facilitar la prueba haremos la convención de mantener el rango de la suma fijo, para lograr ésto y puesto que para cada $y \in \mathcal{R}^n$ el número de inversas no es necesariamente constante, tomaremos $f_{\mathbf{X}}(G_j^{-1}(\mathbf{y})) = 0$ si este punto $y \in \mathcal{R}^n$ no está en el dominio de esa inversa en particular.

Aplicando 3.2 se tiene que para $B \subset \mathcal{R}^n$

$$P\{\mathbf{Y} \in B\} = \int_B \cdots \int f_{\mathbf{Y}}(y_1,\dots,y_n)dx_1\dots dx_n,$$

por el teorema 3.4 se tiene que

$$\begin{aligned} P\{\mathbf{Y} \in B\} &= \int_{G^{-1}(B)} \cdots \int f_{\mathbf{X}}(x_1,\dots,x_n)dx_1\dots dx_n \\ &= \sum_j \int_{G_j^{-1}(B)} \cdots \int f_{\mathbf{X}}(x_1,\dots,x_n)dx_1\dots dx_n. \end{aligned}$$

Ahora como cada $G_j^{-1}(B)$ del conjunto B es una transformación uno a uno, de aquí aplicando el teorema de cambio de variable para integrales (discutido en la mayoría de los libros de cálculo avanzado) resulta que

$$\begin{aligned} F_{\mathbf{Y}}(\mathbf{y}) &= \sum_j \int_B \cdots \int f_{\mathbf{X}}\left(G_{1j}^{-1}(y_1,\ldots,y_n),\ldots,G_{nj}^{-1}(y_1,\ldots,y_n)\right)|J_j|\,dx_1\ldots dx_n \\ &= \int_B \cdots \int \sum_j f_{\mathbf{X}}\left(G_{1j}^{-1}(y_1,\ldots,y_n),\ldots,G_{nj}^{-1}(y_1,\ldots,y_n)\right)|J_j|\,dx_1\ldots dx_n. \end{aligned}$$

Puesto que el conjunto B es un conjunto arbitrario en el rango de G (sin embargo debe ser un conjunto medible) se tiene que lo que está dentro de esta integral debe ser la función de densidad conjunta de $\mathbf{Y}$ por lo cual se tiene que

$$f_{\mathbf{Y}}(\mathbf{y}) = \sum_j f_{\mathbf{X}}\left(G_{1j}^{-1}(y_1,\ldots,y_n),\ldots,G_{nj}^{-1}(y_1,\ldots,y_n)\right)|J_j|.$$

Lo cual establece el teorema.

A continuación enunciaremos dos corolarios los cuales son de gran utilidad en la práctica.

Corolario 3.9 *Sea $\mathbf{X}$ un vector aleatorio de variables absolutamente continuas y sea $G(\mathbf{X})$ una transformación uno a uno de $\mathcal{R}^n$ en $\mathcal{R}^n$ y supongase que para $x_i = G_i^{-1}(y_1, y_2, \ldots, y_n)$, la $\frac{\partial G_i^{-1}}{\partial y_k}$ es continua para todo $i, k = 1, 2, \ldots, n$ y el jacobiano de la transformación J de la transformación inversa*

$$J = \begin{vmatrix} \frac{\partial G_1^{-1}}{\partial y_1} & \frac{\partial G_1^{-1}}{\partial y_2} & \cdots & \frac{\partial G_1^{-1}}{\partial y_n} \\ \frac{\partial G_2^{-1}}{\partial y_1} & \frac{\partial G_2}{\partial y_2} & \cdots & \frac{\partial G_2^{-1}}{\partial y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial G_n^{-1}}{\partial y_1} & \frac{\partial G_n^{-1}}{\partial y_2} & \cdots & \frac{\partial G_n^{-1}}{\partial y_n} \end{vmatrix}$$

no es cero . Entonces para cada $\mathbf{y}$ en el rango de G se tiene que

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}\left(G_1^{-1}(y_1,\ldots,y_n),\ldots,G_n^{-1}(y_1,\ldots,y_n)\right)|J|.$$

Ejemplo 3.5 *Encontrar la distribución de una transformación lineal de un vector aleatorio absolutamente continuo $\mathbf{Y}$ definido como sigue:*

$$y_i = \sum_{j=1}^{n} a_{ij}x_j, i = 1, 2, \ldots, n$$

cuando el $\det(a_{ij}) \neq 0$.

Sol.

Como $\det(a_{ij}) \neq 0$, podemos escribir $\mathbf{X}$ en términos de $\mathbf{Y}$ por lo que obtenemos

$$X_i = \sum_{j=1}^{n} c_{ij} y_j, i = 1, \ldots, n.$$

Podemos obtener el jacobiano de la transformación inversa el cual es

$$\begin{aligned} J &= \det(c_{ij}) \\ &= \frac{1}{\det(a_{ij})}, \end{aligned}$$

por lo que aplicando el corolario 3.9 se tiene:

$$f_{\mathbf{Y}}(y_1, y_2 \ldots, y_n) = f_{\mathbf{X}} \left(\sum_{j=1}^{n} c_{1j} y_j, \ldots, \sum_{j=1}^{n} c_{nj} y_j \right) \left| \det(a_{ij})^{-1} \right|.$$

Ejemplo 3.6 *Supongase que X_1 y X_2 son variables aleatorias independientes cada una con una distribución normal estándar.*

1. *Encontrar la distribución conjunta de $\frac{X_1+X_2}{\sqrt{2}}$ y $\frac{X_2-X_1}{\sqrt{2}}$.*

2. *Argumentar que $2X_1X_2$ y ${X_2}^2 - {X_1}^2$ tienen la misma distribución.*

Sol.

Sea

$$\left. \begin{aligned} Y_1 &= \tfrac{X_1+X_2}{\sqrt{2}} \\ Y_2 &= \tfrac{X_2-X_1}{\sqrt{2}} \end{aligned} \right\} G(X_1, X_2) = (Y_1, Y_2),$$

calculando la transformación inversa se tiene que:

$$\left. \begin{aligned} X_1 &= \tfrac{Y_1-Y_2}{\sqrt{2}} \\ X_1 &= \tfrac{Y_1+Y_2}{\sqrt{2}} \end{aligned} \right\} G^{-1}(Y_1, Y_2) = (X_1, X_2),$$

por lo que el jacobiano de la función inversa es

$$\begin{aligned} J &= \begin{vmatrix} \frac{\partial \frac{y_1-y_2}{\sqrt{2}}}{\partial y_1} & \frac{\partial \frac{y_1-y_2}{\sqrt{2}}}{\partial y_2} \\ \frac{\partial \frac{y_1+y_2}{\sqrt{2}}}{\partial y_1} & \frac{\partial \frac{y_1+y_2}{\sqrt{2}}}{\partial y_2} \end{vmatrix} \\ &= \begin{vmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{vmatrix} \\ &= 1 \end{aligned}$$

Aplicando el corolario 3.9 se tiene que

$$\begin{aligned}
f_Y(y_1,y_2) &= f_X\left(\frac{y_1-y_2}{\sqrt{2}},\frac{y_1+y_2}{\sqrt{2}}\right)\\
&= \frac{e^{-\frac{\left(\frac{y_1-y_2}{\sqrt{2}}\right)^2}{2}}}{\sqrt{2\pi}}\cdot\frac{e^{-\frac{\left(\frac{y_1+y_2}{\sqrt{2}}\right)^2}{2}}}{\sqrt{2\pi}}\\
&= \frac{e^{-\frac{y_1{}^2}{2}-\frac{y_2{}^2}{2}}}{2\pi}\\
&= \frac{e^{-\frac{y_1{}^2}{2}}}{\sqrt{2\pi}}\cdot\frac{e^{-\frac{y_2{}^2}{2}}}{\sqrt{2\pi}},
\end{aligned}$$

ya que $f_Y(y_1,y_2)=f_{Y_1}(y_1)\cdot f_{Y_2}(y_2)$ se tiene que Y_1 y Y_2 son variables aleatorias independientes con distribución normal estándar.

Además para el segundo punto se tiene que

$$\begin{aligned}
X_2{}^2-X_1{}^2 &= (X_2-X_1)(X_2+X_1)\\
&= \frac{2(X_2-X_1)(X_2+X_1)}{2}\\
&= 2\frac{(X_2-X_1)}{\sqrt{2}}\cdot\frac{(X_2+X_1)}{\sqrt{2}},
\end{aligned}$$

y como $\frac{(X_2-X_1)}{\sqrt{2}}$ y $\frac{(X_2+X_1)}{\sqrt{2}}$ son dos variables aleatorias independientes con distribución normal estándar, se tiene que $X_2{}^2-X_1{}^2$ tiene la misma distribución que $2X_1X_2$.

Ejemplo 3.7 *Si X_1 y X_2 son variables aleatorias exponenciales con parámetro λ, encontrar la distribución conjunta de $Y_1=\frac{X_1}{X_2}$ y $Y_2=X_1+X_2$ y las distribuciones marginales de Y_1 y Y_2.*

<u>Sol.</u>

$$\left.\begin{array}{l} y_1=\frac{x_1}{x_2}\\ y_2=x_1+x_2\end{array}\right\} G(x_1,x_2) = (y_1,y_2),$$

$$\left.\begin{array}{l} x_1=\frac{y_1y_2}{1+y_1}\\ x_2=\frac{y_2}{1+y_1}\end{array}\right\} G^{-1}(y_1,y_2) = (x_1,x_2),$$

el jacobiano de la transformación inversa es:

$$J = \begin{vmatrix} \frac{\partial\frac{y_1y_2}{1+y_1}}{\partial y_1} & \frac{\partial\frac{y_1y_2}{1+y_1}}{\partial y_1}\\ \frac{\partial\frac{y_2}{1+y_1}}{\partial y_1} & \frac{\partial\frac{y_2}{1+y_1}}{\partial y_2}\end{vmatrix}$$

$$
\begin{aligned}
&= \begin{vmatrix} \frac{y_2}{(1+y_1)^2} & \frac{y_1}{1+y_1} \\ \frac{-y_2}{(1+y_1)^2} & \frac{1}{1+y_1} \end{vmatrix} \\
&= \frac{y_2}{(1+y_1)^2}.
\end{aligned}
$$

Utilizando el corolario 3.9 se tiene que

$$
\begin{aligned}
f_{Y_1,Y_2}(y_1,y_2) &= f_{X_1,X_2}\left(\tfrac{y_1 y_2}{1+y_1}, \tfrac{y_2}{1+y_1}\right) \cdot \tfrac{y_2}{(1+y_1)^2} \\
&= \lambda^2 \cdot e^{-\lambda \frac{y_1 y_2}{1+y_1}} \cdot e^{-\lambda \frac{y_2}{1+y_1}} \tfrac{y_2}{(1+y_1)^2} \text{ si } \tfrac{y_1 y_2}{1+y_1} > 0 \text{ y } \tfrac{y_2}{1+y_1} > 0 \\
&= \lambda^2 \tfrac{e^{-\lambda \cdot y_2} y_2}{(1+y_1)^2} \text{ si } y_1 > 0 \text{ y } y_2 > 0 \\
&= \begin{cases} \lambda^2 \tfrac{e^{-\lambda \cdot y_2} y_2}{(1+y_1)^2} \text{ si } y_1 > 0 \text{ y } y_2 > 0 \\ 0 \text{ e.o.c.} \end{cases}.
\end{aligned}
$$

Para calcular las funciones de densidad marginales se tiene que

$$
\begin{aligned}
f_{Y_1}(y_1) &= \int_0^\infty \lambda^2 \frac{e^{-\lambda \cdot y_2} y_2}{(1+y_1)^2} dy_2 \\
&= \frac{\lambda^2}{(1+y_1)^2} \int_0^\infty y_2 e^{-\lambda y_2} dy_2 \\
&= \frac{1}{(1+y_1)^2} \text{ si } y_1 > 0,
\end{aligned}
$$

de aquí Y_1 y Y_2 son variables aleatorias independientes puesto que:

$$
f_{Y_1,Y_2}(y_1,y_2) = \begin{cases} \lambda^2 e^{-\lambda \cdot y_2} y_2 \frac{1}{(1+y_1)^2} & \text{si } y_1 > 0 \text{ y } y_2 > 0 \\ 0 & \text{e.o.c.} \end{cases},
$$

por lo que resulta que

$$
f_{Y_2}(y_2) = \begin{cases} \lambda^2 y_2 e^{-\lambda \cdot y_2} & \text{si } y_2 > 0 \\ 0 & \text{e.o.c.} \end{cases}.
$$

Ejemplo 3.8 *Supongase X_1 y X_2 son variables aleatorias independientes cada una con distribución uniforme sobre el intervalo $(0,1)$. Encontrar la función de distribución conjunta de $Y_1 = X_1 + X_2$ y $Y_2 = X_2 - X_1$.*

<u>Sol.</u>

Como

$$
\left.\begin{array}{l} Y_1 = X_1 + X_2 \\ Y_2 = X_2 - X_1 \end{array}\right\} G(X_1, X_2),
$$

$$
\left.\begin{array}{l} X_1 = \frac{Y_1 - Y_2}{2} \\ X_2 = \frac{Y_1 + Y_2}{2} \end{array}\right\} G^{-1}(Y_1, Y_2).
$$

El jacobiano de la transformación inversa es

$$J = \begin{vmatrix} \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{vmatrix} = \frac{1}{2}.$$

Utilizando el corolario 3.9 tenemos que

$$f_{Y_1,Y_2}(y_1,y_2) = f_{X_1,X_2}\left(\frac{y_1-y_2}{2}, \frac{y_1+y_2}{2}\right)\frac{1}{2}.$$

Es claro que si $\frac{y_1-y_2}{2}$ y $\frac{y_1+y_2}{2}$ no pertenecen al intervalo $(0,1)$ (ver fig. 5) entonces

$$f_{X_1,X_2}\left(\frac{y_1-y_2}{2}, \frac{y_1+y_2}{2}\right) = 0,$$

por lo que se tiene:

$$f_{Y_1,Y_2}(y_1,y_2) = \begin{cases} \frac{1}{2} & \text{si } 0 < \frac{y_1-y_2}{2} < 1 \\ 0 & \text{e.o.c.} \end{cases}.$$

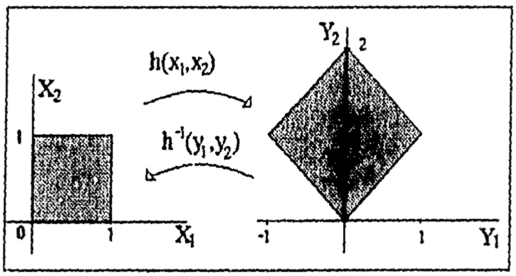


Figura 5.

Ejemplo 3.9 *Sea X_1 y X_2 variables aleatorias independientes cada una con una distribución normal estándar. Encontrar la función de densidad conjunta de $Y_1 = {X_1}^2$ y $Y_2 = {X_1}^2 + {X_2}^2$. ¿Son Y_1 y Y_2 independientes?*

Sol.

$$\left.\begin{aligned} y_1 &= {x_1}^2 \\ y_2 &= {x_1}^2 + {x_2}^2 \end{aligned}\right\} G(x_1,x_2).$$

Es claro que para cada punto (y_1,y_2) en el rango de esta transformación existen cuatro puntos (x_{1i},x_{2i}) $i = 1,2,3,4$ tales que $G(x_{1i},x_{2i}) = (y_1,y_2)$ para cada $i = 1,2,3,4$ por lo que podemos definir cuatro funciones inversas como en el teorema 3.8, de aquí se tiene que

$$\left.\begin{array}{l} x_{11} = \sqrt{y_1} \\ x_{21} = \sqrt{y_2 - y_1} \end{array}\right\} G_1^{-1}(y_1, y_2),$$

$$\left.\begin{array}{l} x_{12} = -\sqrt{y_1} \\ x_{22} = \sqrt{y_2 - y_1} \end{array}\right\} G_2^{-1}(y_1, y_2),$$

$$\left.\begin{array}{l} x_{13} = -\sqrt{y_1} \\ x_{23} = -\sqrt{y_2 - y_1} \end{array}\right\} G_3^{-1}(y_1, y_2),$$

$$\left.\begin{array}{l} x_{11} = \sqrt{y_1} \\ x_{23} = -\sqrt{y_2 - y_1} \end{array}\right\} G_4^{-1}(y_1, y_2).$$

Las cuales son transformaciones uno a uno de un conjunto

$$A = \{(y_1, y_2) : y_2 > y_1 > 0\}$$

en conjuntos

$$\begin{aligned} H_1 &= \{(x_1, x_2) : x_1 > 0, x_2 > 0\}, \\ H_2 &= \{(x_1, x_2) : x_1 < 0, x_2 > 0\}, \\ H_3 &= \{(x_1, x_2) : x_1 < 0, x_3 < 0\}, \\ H_4 &= \{(x_1, x_2) : x_1 > 0, x_2 < 0\} \end{aligned}$$

(ver fig.6).

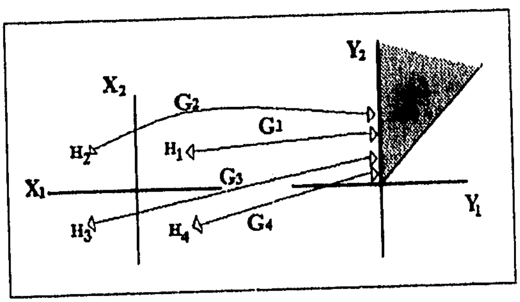


Figura 6.

Ahora calcularemos el jacobiano de cada una de esta transformación inversa:

$$J_1 = \begin{vmatrix} \frac{1}{2\sqrt{y_1}} & 0 \\ -\frac{1}{2\sqrt{y_2 - y_1}} & \frac{1}{2\sqrt{y_2 - y_1}} \end{vmatrix}$$

$$= \frac{1}{4\sqrt{y_1}\sqrt{y_2 - y_1}}.$$

Fácilmente se puede ver que

$$\begin{aligned} |J_1| &= |J_2| = |J_3| = |J_4| \\ &= \frac{1}{4\sqrt{y_1}\sqrt{y_2 - y_1}} \text{ si } y_2 > y_1 > 0. \end{aligned}$$

Por lo que utilizando el teorema 3.8 se tiene que

$$\begin{aligned} f_{Y_1,Y_2} & (y_1, y_2) \\ &= \sum_{i=1}^{4} f_{X_1,X_2}(x_{1i}, x_{2i}) \frac{1}{4\sqrt{y_1}\sqrt{y_2-y_1}} \\ &= \frac{1}{4\sqrt{y_1}\sqrt{y_2-y_1}} \left[f_{X_1,X_2}(\sqrt{y_1}, \sqrt{y_2 - y_1}) + \ldots + f_{X_1,X_2}(\sqrt{y_1}, -\sqrt{y_2 - y_1}) \right] \\ &= \frac{1}{4\sqrt{y_1}\sqrt{y_2-y_1}} \left[\frac{4e^{-\frac{y_1}{2} - \frac{(y_2-y_1)}{2}}}{2\pi} \right] \\ &= \frac{e^{-\frac{y_2}{2}}1}{2\pi\sqrt{y_1}\sqrt{y_2-y_1}} \text{ si } y_2 > y_1 > 0, \end{aligned}$$

por lo que:

$$f_{Y_1,Y_2}(y_1, y_2) = \begin{cases} \frac{e^{-\frac{y_2}{2}}1}{2\pi\sqrt{y_1}\sqrt{y_2-y_1}} & \text{si } y_2 > y_1 > 0 \\ 0 & \text{e.o.c.} \end{cases}.$$

Es claro que y_1 y y_2 no son independientes.

3.3 Ejercicios.

1. Demostrar las fórmulas 3.9 y 3.7.

2. Sea X y Y variables aleatorias discretas independientes cada una con distribución uniforme en $\{1, 2, .., N\}$, encontrar:

   (a) $P\{X > Y\}$.

   (b) $P\{X = Y\}$.

   (c) Calcular la función de densidad de $|Y - X|$.

3. Sea X y Y variables aleatorias discretas independientes cada una con densidades geométricas con parámetros p_1 y p_2, encontrar:

   (a) $P\{X > Y\}$.

(b) $P\{X = Y\}$.

(c) Calcular la función de densidad de $X + Y$.

4. Asuma que las variables aleatorias discretas mutuamente independientes X_i $(i = 1, 2, .., n)$ cada una con una distribución binomial negativa con parámetros r_i $(i = 1, 2, .., n)$ donde cada r_i es un entero positivo, es decir cada X_i tiene una función de densidad como sigue:

$$f_{X_i}(x) = \begin{cases} \binom{r_i + x - 1}{x} p^{r_i}(1-p)^x & \text{si } x = 0, 1, 2, .. \\ 0 & \text{e.o.c.} \end{cases}.$$

Mostrar que la función de densidad de probabilidad de $Y = \sum_{i=1}^{n} X_i$ tiene una distribución binomial negativa con parámetros $r = \sum_{l=1}^{n} r_i$.

5. Probar que si dos variables aleatorias discretas independientes X_1 y X_2 distribuidas identicamente, las cuales toman valores del 1 al 6 y las cuales tienen la propiedad de que su suma satisface

$$P\{X_1 + X_2 = k\} = P\{X_1 + X_2 = 14 - k\} = \frac{k-1}{36}$$

para $k = 2, 3, 4, 5, 6$ y

$$P\{X_1 + X_2 = 7\} = \frac{1}{6}$$

entonces

$$P\{X_1 = k\} = P\{X_2 = k\} = \frac{1}{6}$$

para $k = 1, 2, 3, 4, 5, 6$.

6. Dos puntos son seleccionados aleatoriamente en una línea de longitud L, uno es tomado a la derecha del punto medio y el otro a la izquierda (en otras palabras los puntos son variables aleatorias independientes X y Y las cuales tienen distribuciones uniformes, la primera en el intervalo $(0, \frac{L}{2})$ y la segunda en $(\frac{L}{2}, L)$). Encontrar la probabilidad que la distancia entre los dos puntos sea mayor que $\frac{L}{3}$.

7. En el problema 6, encontrar la probabilidad de que los 3 segmentos de línea formados de 0 a X, de X a Y y de Y a L puedan formar los tres lados de un triángulo (recordar la desigualdad del triángulo).

8. Sean X y Y variables aleatorias que están distribuidas uniformemente en el interior de un triángulo con vértices en $(0,0)$, $(2,0)$ y $(1,2)$. Encontrar $P\{X \leq 1, Y \leq 1\}$.

9. Supóngase que un punto es tomado uniformemente en un cuadrado de área 1 teniendo por vértices los puntos $(0,0)$, $(0,1)$, $(1,0)$, $(1,1)$. Sean X y Y las coordenadas del punto tomado.

   (a) Encontrar la distribución marginal de X y Y.

   (b) ¿Son X y Y independientes?

   (c) Encontrar la probabilidad de que la distancia de (X,Y) al centro del cuadrado sea mayor que $\frac{1}{4}$.

10. Sea
$$f_{X,Y}(x,y) = \begin{cases} \frac{3x}{2} & \text{si } 0 < x < 1, |y| < x \\ 0 & \text{e.o.c.} \end{cases}.$$
Encontrar la función de densidad de $X - Y$.

11. Sea
$$f_{X,Y,Z}(x,y,z) = \begin{cases} e^{-(x+y+z)} & \text{si } x > 0, y > 0, z > 0 \\ 0 & \text{e.o.c.} \end{cases},$$
encontrar la la función de densidad de su promedio, es decir la densidad de $W = (X+Y+Z)/3$.

12. Sea X y Y variables aleatorias independientes cada una con una distribuida uniformemente en $(0,1)$. Encontrar:

   (a) $P\left\{|X-Y| \leq \frac{1}{2}\right\}$.

   (b) $P\left\{\left|\frac{X}{Y} - 1\right| \leq \frac{1}{2}\right\}$.

13. Sea X y Y variables aleatorias independientes cada una con densidad normal con parámetros $\mu = 0$ y σ^2. Encontrar $P\{X^2 + Y^2 \leq 1\}$.

14. Sean X y Y variables aleatorias distribuidas uniformemente en la región
$$A = \left\{(x_1, x_2) : {x_1}^2 + {x_2}^2 \leq r^2\right\}.$$
Si $R = \sqrt{{X_1}^2 + {X_2}^2}$. Encontrar la función de densidad de R.

15. Sea X y Y variables aleatorias independientes cada una con una distribución normal estándar. Encontrar la función de densidad de $Z = \frac{X}{Y}$.

16. Si X tiene una distribución uniforme sobre el intervalo $(0,1)$ y Y tiene una distribución exponencial con parámetro $\lambda = 1$. Encontrar la distribución de

    (a) $Z = X + Y$.

    (b) $Z = \frac{X}{Y}$. Asuma que estas dos variables son independientes.

17. Sea X y Y variables aleatorias independientes ambas con distribución exponencial con parámetro λ. Probar que $Z = \frac{X}{X+Y}$ se distribuye uniformemente en el intervalo $(0,1)$.

18. Las variables aleatorias X y Y son independientes cada una con funciones de densidad dadas por

$$f_X(x) = \begin{cases} \frac{1}{\pi\sqrt{1-x^2}} & \text{si } |x| < 1 \\ 0 & \text{e.o.c.} \end{cases},$$

y

$$f_Y(y) = \begin{cases} ye^{-\frac{y^2}{2}} & \text{si } y > 0 \\ 0 & \text{e.o.c.} \end{cases}.$$

Probar que la variable $Z = XY$ tiene una distribución normal.

19. Si X, Y y Z son variables aleatorias independientes con una distribución uniforme sobre el intervalo $(-1,1)$. ¿Cuál es la función de densidad de $U = \frac{XY}{Z}$?

20. Sea

$$f_{X,Y}(x,y) = \begin{cases} \frac{2}{\pi}(1 - x^2 - y^2) & \text{si } x^2 + y^2 \leq 1 \\ 0 & \text{e.o.c.} \end{cases}.$$

Si $U = \sqrt{X^2 + Y^2}$ y $V = \tan^{-1}(\frac{X}{Y})$. Encontrar la función de densidad conjunta de U y V.

21. Si X y Y son variables aleatorias independientes ambas con distribución uniforme sobre el intervalo $(0,1)$, calcular la densidad conjunta de

    (a) $U = X + Y, V = \frac{X}{Y}$,

    (b) $U = X, V = \frac{X}{Y}$,

(c) $U = X + Y, V = \frac{X}{X+Y}$.

22. Si X y Y tienen una función de densidad conjunta dada por:

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

   (a) Calcule la función de densidad conjunta de $U = XY$ y $V = \frac{X}{Y}$.

   (b) Encuentre las funciones de densidad marginales. ¿Son X y Y independientes?

23. Si X, Y y Z son variables aleatorias independientes cada una con distribución exponencial con $\lambda = 1$. Calcule la función de densidad conjunta de $U = X+Y$, $V = X+Z$, $W = Y+Z$.

24. Si X y Y son variables aleatorias independientes con parámetro λ, encontrar la función de densidad conjunta de $U = X + Y$, $V = e^X$.

25. Si X y Y son variables aleatorias independientes con distribución normal estándar. Encontrar la función de densidad conjunta de $U = aX + bY$ y $V = cX + dY$ donde $ad - cb \neq 0$.

26. Si X y Y tienen una función de densidad dada por

$$f_{X,Y}(x,y) = \begin{cases} 4xye^{-(x^2+y^2)} & \text{si } x > 0, y > 0 \\ 0 & \text{e.o.c.} \end{cases}.$$

   Encontrar la función de densidad conjunta de $U = X$, $V = \sqrt{X^2 + Y^2}$, a partir de ésta encontrar la densidad marginal de V.

27. Sean X y Y variables aleatorias independientes cada una con una distribución normal estándar. Encontrar la función de densidad conjunta de $U = X^2 + Y^2$ y $V = \frac{X}{Y}$.

28. Sean X y Y variables aleatorias independientes distribuidas uniformemente sobre el intervalo $(0,1)$. Encontrar la función de densidad conjunta de $R = \sqrt{X^2 + Y^2}$ y $\theta = \tan^{-1}(\frac{Y}{X})$.

Capítulo 4

Distribuciones muestrales

En el capítulo anterior desarrollamos la teoría para poder encontrar la distribución de una variables aleatoria que surge por medio de una función de variables aleatorias, en este capítulo daremos algunos ejemplos de variables aleatorias de gran utilidad en estadística, utilizando las herramientas que desarrollamos anteriormente.

4.1 Distribución Gamma

Una de las distribuciones más importantes en la teoría de la probabilidad y que nos será de gran utilidad en este capítulo es la distribución gamma, esta tiene una función de densidad dada por:

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

donde $\Gamma(\alpha)$ es una función definida como sigue:

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx, \; \alpha > 0.$$

Su importancia radica en que esta es el caso más general de algunas variables aleatorias de suma importancia como son, la variable $\chi^2_{(n)}$ que veremos más adelante y también la variable aleatoria exponencial.

Para esto primero encontraremos la distribución de la suma de n variables aleatorias cada una con una distribución gamma con parámetros $(\alpha_1, \lambda), (\alpha_2, \lambda), \ldots, (\alpha_n, \lambda)$ respectivamente.

Teorema 4.1 *Sean $X_1, X_2, \ldots, X_n$ variables aleatorias independientes que tienen cada una distribución gamma con parámetros (α_1, λ), $(\alpha_2, \lambda), \ldots, (\alpha_n, \lambda)$ respectivamente. Entonces $Y = X_1 + X_2 + \ldots + X_n$ tiene una distribución gamma con parámetros $(\alpha_1 + \alpha_2 + \ldots + \alpha_n, \lambda)$.*

Dem.

Primero demostraremos que $Y = X_1 + X_2$ tienen una distribución gamma con parámetros $(\alpha_1 + \alpha_2, \lambda)$. Utilizando el 3.4 tenemos

$$f_{X_1+X_2}(z) = \int_{-\infty}^{\infty} f_{X_1}(x) \cdot f_{X_2}(z-x)dx$$
$$= \int_0^z f_{X_1}(x) f_{X_2}(z-x)dx,$$

ya que $f_{X_1}(x) = 0$ si $x \leq 0$ y $f_{X_2}(z-x) = 0$ si $x \geq z$ o $x \leq 0$,

$$f_{X_1+X_2}(z) = \int_0^z \frac{\lambda^{\alpha_1}}{\Gamma(\alpha_1)} x^{\alpha_1-1} e^{-\lambda x} \frac{\lambda^{\alpha_2}}{\Gamma(\alpha_2)} (z-x)^{\alpha_2-1} e^{-\lambda(z-x)} dx$$
$$= \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} e^{-\lambda z} \int_0^z x^{\alpha_1-1}(z-x)^{\alpha_2-1} dx.$$

Si hacemos $x = zu$ entonces $dx = zdu$ y $0 \leq u \leq 1$ entonces

$$f_{X_1+X_2}(z) = \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} e^{-\lambda z} \int_0^1 (zu)^{\alpha_1-1}(z-zu)^{\alpha_2-1} z du$$
$$= \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} e^{-\lambda z} z^{\alpha_1-\alpha_2-1} \int_0^1 u^{\alpha_1-1}(1-u)^{\alpha_2-1} du,$$

por lo que tenemos:

$$f_{X_1+X_2}(z) = \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} e^{-\lambda z} z^{\alpha_1-\alpha_2-1} \int_0^1 u^{\alpha_1-1}(1-u)^{\alpha_2-1} du \quad \text{si } z > 0,$$

si $c = \frac{\int_0^1 u^{\alpha_1-1}(1-u)^{\alpha_2-1}du}{\Gamma(\alpha_1)\Gamma(\alpha_2)}$ podemos escribir $f_{X_1+X_2}(z)$ como sigue

$$f_{X_1+X_2}(z) = c\lambda^{\alpha_1+\alpha_2} e^{-\lambda z} z^{\alpha_1+\alpha_2-1},$$

y como $\int_{-\infty}^{\infty} f_{X_1+X_2}(z)dz = 1$ tenemos que

$$c\int_0^{\infty} e^{-\lambda z} z^{\alpha_1+\alpha_2-1} dz = c\Gamma(\alpha_1+\alpha_2)$$
$$= 1,$$

por lo que $c = \frac{1}{\Gamma(\alpha_1+\alpha_2)}$ de aquí podemos concluir que

$$\int_0^1 u^{\alpha_1-1}(1-u)^{\alpha_2-1} du = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1+\alpha_2)}.$$

Esta última integral se conoce como la función beta, y se denotará como sigue:

$$B(\alpha_1,\alpha_2)=\int_0^1 u^{\alpha_1-1}(1-u)^{\alpha_2-1}du=\frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1+\alpha_2)}$$

que es importante para definir la variable aleatoria con el mismo nombre. Por lo que se tiene que:

$$f_{X_1+X_2}(z)=\begin{cases}\frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1+\alpha_2)}e^{-\lambda z}z^{\alpha_1-\alpha_2-1} & si\ z>0\\ 0 & e.o.c.\end{cases}.$$

Por lo que tenemos que $Y=X_1+X_2+\ldots+X_n$ *tiene una distribución gamma con parámetros* $(\alpha_1+\alpha_2+\ldots+\alpha_n,\lambda)$.

4.2 Distribución χ^2

Definición 4.1 *A la distribución Gamma con parámetros* $\alpha=\frac{n}{2}$ *y* $\lambda=\frac{1}{2}$ *se le conoce como distribución* χ^2 *(ji-cuadrada) con* n *grados de libertad. Podemos decir también que* Z *tiene una distribución* $\chi^2_{(n)}$.

La distribución χ^2 tiene una estrecha relación con la distribución normal como se muestra en el siguiente teorema:

Teorema 4.2 *Sean* $X_1,X_2,\ldots,X_n$ *variables aleatorias independientes idénticamente distribuidas con una distribución normal estándar, si*

$$Z={X_1}^2+{X_2}^2+\ldots+{X_n}^2,$$

entonces Z *tiene una distribución* χ^2 *con* n *grados de libertad.*

Dem.

Si $n=1$ y X es una variable aleatoria normal estándar tenemos:

$$\begin{aligned}F_{X^2}(z) &= P\left\{X^2\le z\right\}\\ &= P\{|X|\le\sqrt{z}\}\\ &= P\{-\sqrt{z}\le X\le\sqrt{z}\}\\ &= P\{X\le\sqrt{z}\}-P\{X\le-\sqrt{z}\}\\ &= \int_{-\infty}^{\sqrt{z}}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx-\int_{-\infty}^{-\sqrt{z}}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx \quad \text{si } z>0\end{aligned}$$

Utilizando las propiedades de la función de distribución, el teorema fundamental del cálculo y el hecho de que $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ tenemos:

$$\begin{aligned} f_{X^2}(z) &= \frac{1}{2\sqrt{2\pi z}}e^{-\frac{z}{2}} + \frac{1}{2\sqrt{2\pi z}}e^{-\frac{z}{2}} \\ &= \frac{1}{\sqrt{2\pi z}}e^{-\frac{z}{2}} \\ &= \frac{z^{\frac{1}{2}-1}e^{-\frac{1}{2}z}(\frac{1}{2})^{\frac{1}{2}}}{\Gamma(\frac{1}{2})} \quad \text{si } z > 0 \end{aligned}$$

Por lo que se tiene que $f_{X^2}(z)$ tiene una distribución gamma con parámetros $\alpha = \frac{1}{2}$ y $\lambda = \frac{1}{2}$.

De aquí, por el 4.1 podemos concluir que Z tiene una distribución gamma con parámetros $\alpha = \frac{n}{2}$ y $\lambda = \frac{1}{2}$.

Definición 4.2 *Si $Z = \sqrt{\chi^2_{(n)}}$ entonces decimos que Z tiene una distribución χ (ji) con n grados de libertad, podemos decir que Z tiene una distribución $\chi_{(n)}$.*

Teorema 4.3 *Si Z es una variable aleatoria χ, entonces Z tiene una función de densidad dada como sigue:*

$$f_Z(z) = \begin{cases} \frac{2}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} z^{n-1}e^{-\frac{w^2}{2}} & si \quad z > 0 \\ 0 & e.o.c. \end{cases}.$$

Dem.

Calculemos primero la función de distribución de Z

$$\begin{aligned} F_Z(z) &= P\left\{\sqrt{Z} \leq z\right\} \\ &= P\left\{Z \leq z^2\right\} \\ &= \int_0^{z^2} \frac{u^{\frac{n}{2}-1}e^{-\frac{1}{2}u}(\frac{1}{2})^{\frac{n}{2}}}{\Gamma(\frac{n}{2})}du, \end{aligned}$$

por lo que se tiene que

$$\begin{aligned} f_Z(z) &= F'_Z(z) \\ &= \frac{(z^2)^{\frac{n}{2}-1}e^{-\frac{z^2}{2}}(\frac{1}{2})^{\frac{n}{2}}2z}{\Gamma(\frac{n}{2})} \\ &= \frac{2}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}z^{n-1}e^{-\frac{z^2}{2}}, \end{aligned}$$

por lo cual tenemos que

$$f_Z(z) = \begin{cases} \frac{2}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} z^{n-1} e^{-\frac{z^2}{2}} & \text{si } z > 0 \\ 0 & \text{e.o.c.} \end{cases}.$$

La distribución χ^2 fue obtenida por primera vez por Bienagmé en 1838, como la distribución límite de una variable aleatoria discreta definida como $\sum_{j=1}^{k} \frac{(N_i - np_i)^2}{np_i}$ cuando el vector aleatoria $(N_1, N_2, \ldots, N_n)$ tiene una distribución multinomial con parámetros $n, p_1, p_2, \ldots, p_n$.

La variable aleatoria χ^2 aparece nuevamente en 1900 como una aproximación para las estadísticas ji-cuadradas, usadas para construir tablas de contingencia, las cuales también son distribuciones discretas.

Por último la familia de distribuciones gamma puede representar una gran cantidad de situaciones físicas.

Las dos variables obtenidas a continuación son de gran utilidad en aplicaciones de la estadística matemática.

4.3 Distribución F

Definición 4.3 *Si X y Y son variables aleatorias independientes cada una con distribución $\chi^2_{(n)}$ y $\chi^2_{(m)}$ respectivamente y $Z = \frac{X/n}{Y/m}$ entonces la variable aleatoria Z se dice que tiene una distribución F con (n, m) grados de libertad.*

Teorema 4.4 *Si Z tiene una distribución F con (n, m) grados de libertad entonces su función de densidad es como sigue:*

$$f_Z(z) = \begin{cases} n^{\frac{n}{2}} m^{\frac{m}{2}} \frac{\Gamma(\frac{n+m}{2})}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})} \frac{z^{\frac{n}{2}-1}}{(nz+m)^{\frac{n+m}{2}}} & si \quad z > 0 \\ 0 & e.o.c. \end{cases}$$

Dem.

Calcularemos primeramente la función de densidad de $\frac{X}{n}$ y $\frac{Y}{m}$ respectivamente, por lo cual se tiene:

$$\begin{aligned} F_{\frac{X}{n}}(u) &= P\left\{\frac{X}{n} \leq u\right\} \\ &= P\left\{X \leq nu\right\} \\ &= \int_0^{nu} f_X(x)dx, \end{aligned}$$

por consiguiente, utilizando el teorema fundamental del cálculo se tiene que:

$$f_{\frac{X}{n}}(u) = nf(nu) = \begin{cases} \frac{n}{2}\frac{e^{-\frac{nu}{2}}(\frac{nu}{2})^{\frac{n}{2}-1}}{\Gamma(\frac{n}{2})} & si\ u > 0 \\ 0 & e.o.c. \end{cases}.$$

Análogamente se tiene que

$$f_{\frac{Y}{n}}(v) = \begin{cases} \frac{n}{2}\frac{e^{-\frac{nv}{2}}(\frac{nv}{2})^{\frac{n}{2}-1}}{\Gamma(\frac{n}{2})} & si\ v > 0 \\ 0 & e.o.c. \end{cases},$$

de aquí:

$$f_{\frac{X}{n},\frac{Y}{m}}(u,v) = \begin{cases} \frac{n}{2}\frac{e^{-\frac{nu}{2}}(\frac{nu}{2})^{\frac{n}{2}-1}}{\Gamma(\frac{n}{2})} \cdot \frac{m}{2}\frac{e^{-\frac{mu}{2}}(\frac{mv}{2})^{\frac{m}{2}-1}}{\Gamma(\frac{m}{2})} & si\ u,v > 0 \\ 0 & e.o.c. \end{cases}$$

Utilizando 3.7 se tiene:

$$f_Z(z) = \int_{-\infty}^{\infty} |x|\, f_{X,Y}(zx,x)dx$$

$$= \int_0^{\infty} x \cdot \frac{n}{2}\frac{e^{-\frac{nzx}{2}}\,(\frac{nzx}{2})^{\frac{n}{2}-1}}{\Gamma(\frac{n}{2})} \cdot \frac{m}{2}\frac{e^{-\frac{mx}{2}}\,(\frac{mx}{2})^{\frac{m}{2}-1}}{\Gamma(\frac{m}{2})}dx$$

$$= \frac{n^{\frac{n}{2}}m^{\frac{m}{2}}z^{\frac{n}{2}-1}}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})2^{\frac{n+m}{2}}}\int_0^{\infty} x^{\frac{n+m}{2}-1}e^{-(\frac{nz+m}{2})x}dx,$$

si hacemos el cambio de variable $u = \frac{nz+m}{2}x$, $du = \frac{nz+m}{2}dx$ *tenemos que*

$$f_Z(z) = \frac{n^{\frac{n}{2}}m^{\frac{m}{2}}}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})} \cdot z^{\frac{n}{2}-1}(nz+m)^{-\frac{n+m}{2}}\int_0^{\infty} u^{-\frac{n+m}{2}-1}e^{-u}du$$

$$= \frac{n^{\frac{n}{2}}m^{\frac{m}{2}}}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})}\Gamma(\frac{n+m}{2}) \cdot z^{\frac{n}{2}-1} \cdot (nz+m)^{-\frac{n+m}{2}}.$$

Por lo cual se establece el teorema.

Es claro que si Z_1 tiene una distribución F con (n,m) grados de libertad y Z_2 también tiene una distribución F pero con (m,n) grados de libertad entonces Z_1 y ${Z_2}^{-1}$ tienen la misma distribución, de forma análoga Z_2 y ${Z_1}^{-1}$ tienen también la misma distribución.

La importancia de la distribución F en la estadística matemática se debe principalmente a que es aplicable en pruebas estándar asociadas con el análisis de varianza, además de que existe una relación entre las distribuciones F y las distribuciones binomiales, la cual es de gran utilidad para evitar los problemas del cálculo de la función de distribución binomial.

4.4 Distribución t de Student

La otra distribución, la cual también surge como un cociente de variables aleatorias, de suma importancia en la estadística matemática es la distribución t de Students.

Definición 4.4 *Si X y Y son variables aleatorias independientes, donde X tiene una distribución normal estándar y Y tiene una distribución $\chi^2_{(n)}$. Diremos que una variables aleatoria tiene una distribución t de Student (o simplemente distribución t) con n grados de libertad, si tiene la misma función de distribución que $T = \frac{X}{\sqrt{Y/n}}$.*

Teorema 4.5 *Si T tiene una distribución t de Student con n grados de libertad entonces su función de densidad está dada como sigue*

$$f_T(t) = \frac{1}{\sqrt{n\pi}} \cdot \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \cdot \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}.$$

Dem.

Para calcular la función de densidad de T primeramente encontraremos la función de densidad de $Z = \sqrt{\frac{Y}{n}}$. Por el teorema 4.3 establecimos la distribución de $\sqrt{Y}$ por lo que:

$$\begin{aligned} F_Z(z) &= P\{Z \le z\} \\ &= P\left\{\frac{\sqrt{Y}}{\sqrt{n}} \le z\right\} \\ &= P\left\{\sqrt{Y} \le \sqrt{n}z\right\} \\ &= \int_0^{\sqrt{n}z} f_{\sqrt{Y}}(u)du, \end{aligned}$$

de aquí, utilizando el teorema fundamental del cálculo tenemos que

$$\begin{aligned} f_Z(z) &= \sqrt{n} f_{\sqrt{Y}}(\sqrt{n}z) \\ &= \begin{cases} 2(\frac{n}{2})^{\frac{n}{2}} \frac{z^{n-1}e^{-\frac{nz^2}{2}}}{\Gamma(\frac{n}{2})} & si\ z > 0 \\ 0 & e.o.c. \end{cases} \end{aligned}$$

Ya que X y Z son variables aleatorias independientes, se tiene que

$$f_{X,Z}(x,z) = \begin{cases} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} \cdot 2(\frac{n}{2})^{\frac{n}{2}} \frac{z^{n-1}e^{-\frac{nz^2}{2}}}{\Gamma(\frac{n}{2})} & si\ z > 0 \\ 0 & e.o.c. \end{cases}$$

Utilizando 3.7 se tiene que

$$f_T(t) = \int_{-\infty}^{\infty} |x| f_{X,Y}(tx,x)dx$$

$$= \int_0^{\infty} x \frac{e^{-\frac{t^2x^2}{2}}}{\sqrt{2\pi}} \cdot 2(\frac{n}{2})^{\frac{n}{2}} \frac{x^{n-1}e^{-\frac{nx^2}{2}}}{\Gamma(\frac{n}{2})} dx$$

$$= \int_0^{\infty} \frac{x^n 2(\frac{n}{2})^{\frac{n}{2}}}{\sqrt{2\pi}\Gamma(\frac{n}{2})} e^{-\frac{(t^2+n)x^2}{2}} dx$$

$$= \frac{2(\frac{n}{2})^{\frac{n}{2}}}{\sqrt{2\pi}\Gamma(\frac{n}{2})} \int_0^{\infty} x^n e^{-\frac{(t^2+n)x^2}{2}} dx,$$

si hacemos el cambio de variable $x = \sqrt{\frac{2u}{t^2+n}}$, $dx = \sqrt{\frac{2}{t^2+n}}\frac{1}{2\sqrt{u}}du$ *tenemos que*

$$f_T(t) = \frac{2(\frac{n}{2})^{\frac{n}{2}}}{\sqrt{2\pi}\Gamma(\frac{n}{2})} \int_0^{\infty} \left(\frac{2u}{t^2+n}\right)^{\frac{n}{2}} e^{-u} \sqrt{\frac{2}{t^2+n}} \frac{1}{2\sqrt{u}} du$$

$$= \frac{n^{\frac{n}{2}}}{\sqrt{\pi}\Gamma(\frac{n}{2})}(t^2+n)^{-\frac{n+1}{2}} \int_0^{\infty} u^{\frac{n+1}{2}-1}e^{-u}du$$

$$= \frac{n^{\frac{n+1}{2}}}{\sqrt{n\pi}\Gamma(\frac{n}{2})}(t^2+n)^{-\frac{n+1}{2}}\Gamma(\frac{n+1}{2})$$

$$= \frac{1}{\sqrt{n\pi}} \cdot \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \cdot \left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}.$$

Por lo que se establece el teorema.

Otra manera de escribir esta función de distribución es:

$$f_T(t) = \frac{1}{\sqrt{n}B(\frac{1}{2},\frac{n}{2})} \cdot \left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}$$

ya que $\frac{\Gamma(\frac{1}{2})\Gamma(\frac{n}{2})}{\Gamma(\frac{n+1}{2})} = B(\frac{1}{2},\frac{n}{2})$ y como $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ tenemos $\frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi}\Gamma(\frac{n}{2})} = \frac{1}{B(\frac{1}{2},\frac{n}{2})}$.

Podemos observar que si la distribución t de Student tiene un grado de libertad entonces, esta también se comporta como una distribución Cauchy estándar.

La distribución t de Student fue determinada por primera vez por W. S. Gosset quien escribió bajo el seudónimo de "Student". La mayor aplicación de la distribución t de Student es en la construcción de pruebas e intervalos de confianza relativos a valores esperados de distribuciones normales.

4.5 Distribución Beta

Esta distribución surge de manera natural como una función de variables aleatorias, esto es:

Teorema 4.6 *Si X_1 y X_2 son dos variables aleatorias independientes ambas con distribución χ^2 con n_1 y n_2 grados de libertad, entonces $V = \frac{X_1}{X_1+X_2}$ es una variable aleatoria beta con parámetros $a = \frac{n_1}{2}$, $b = \frac{n_2}{2}$.*

<u>*Dem.*</u>

Sea

$$\left.\begin{array}{l} y_1 = \frac{x_1}{x_1+x_2} \\ y_2 = x_1 + x_2 \end{array}\right\} = G(x_1, x_2).$$

La cual es claramente una transformación uno a uno, por lo que su transformación inversa esta dada por:

$$\left.\begin{array}{l} x_1 = y_1 y_2 \\ y_2 = y_2(1 + y_1) \end{array}\right\} = G^{-1}(y_1, y_2).$$

El jacobiano de la transformación inversa es:

$$J = \begin{vmatrix} y_2 & y_1 \\ -y_2 & 1 - y_1 \end{vmatrix}$$
$$= y_2.$$

Utilizando 3.9 se tiene:

$$f_{Y_1,Y_2}(y_1, y_2) = |y_2| \cdot f_{X_1,X2}(y_1 y_2, y_2(1 - y_1))$$
$$= |y_2| \frac{(y_1 y_2)^{\frac{n_1}{2}-1} e^{-\frac{y_1 y_2}{2}}}{2^{\frac{n_1}{2}} \Gamma(\frac{n_1}{2})} \cdot \frac{[y_2(1 - y_1)]^{\frac{n_2}{2}-1} e^{-\frac{y_2(1-y_1)}{2}}}{2^{\frac{n_2}{2}} \Gamma(\frac{n_2}{2})}$$

si $y_1 y_2 > 0$, $y_2(1 - y_1) > 0$ *ya que si* $y_1 y_2 > 0$, $y_2(1 - y_1) > 0$ *entonces se tiene que* $0 < y_1 < 1$ *y* $y_2 > 0$. *Por lo cual tenemos que:*

$$f_{Y_1,Y_2}(y_1, y_2) = \frac{y_2^{\frac{n_1+n_2}{2}-1}\, y_1^{\frac{n_1}{2}-1} (1 - y_1)^{\frac{n_2}{2}-1} e^{-\frac{y_2}{2}}}{2^{\frac{n_1+n_2}{2}} \Gamma(\frac{n_1}{2}) \Gamma(\frac{n_2}{2})}$$
$$= \frac{y_2^{\frac{n_1+n_2}{2}-1} e^{-\frac{y_2}{2}}}{2^{\frac{n_1+n_2}{2}} \Gamma(\frac{n_1+n_2}{2})} \cdot \frac{\Gamma(\frac{n_1+n_2}{2})}{\Gamma(\frac{n_1}{2}) \Gamma(\frac{n_2}{2})}\, y_1^{\frac{n_1}{2}-1} (1 - y_1)^{\frac{n_2}{2}-1}$$
$$= \frac{y_2^{\frac{n_1+n_2}{2}-1} e^{-\frac{y_2}{2}}}{2^{\frac{n_1+n_2}{2}} \Gamma(\frac{n_1+n_2}{2})} \cdot B(\frac{n_1}{2}, \frac{n_2}{2})\, y_1^{\frac{n_1}{2}-1} (1 - y_1)^{\frac{n_2}{2}-1}$$
$$\text{si } 0 < y_1 < 1 \text{ y } y_2 > 0.$$

De aquí se tiene que

$$f_{Y_1,Y_2}(y_1,y_2) = \begin{cases} \dfrac{y_2^{\frac{n_1+n_2}{2}-1} e^{-\frac{y_2}{2}} B(\frac{n_1}{2},\frac{n_2}{2}) y_1^{\frac{n_1}{2}-1}(1-y_1)^{\frac{n_2}{2}-1}}{2^{\frac{n_1+n_2}{2}}\Gamma(\frac{n_1+n_2}{2})} & si\ 0 < y_1 < 1\ y\ y_2 > 0 \\ 0 & e.o.c. \end{cases} .$$

Por lo que se puede observar que Y_1 y Y_2 son dos variables aleatorias independientes, la primera con una distribución χ^2 con $n_1 + n_2$ grados de libertad como se vio anteriormente y la segunda es una variable aleatoria beta con parámetros $a = \frac{n_1}{2}$ y $b = \frac{n_2}{2}$, por lo que tenemos que:

$$f_{Y_1}(y_1) = \begin{cases} B(\frac{n_1}{2},\frac{n_2}{2})\, y_1^{\frac{n_1}{2}-1}(1-y_1)^{\frac{n_2}{2}-1} & si\ 0 < y_1 < 1 \\ 0 & e.o.c. \end{cases} ,$$

por lo que se establece el teorema.

También se puede demostrar:

Teorema 4.7 *Si $X_1, X_2, \ldots, X_m$ son variables independientes cada una con distribución χ^2, con n_i grados de libertad para $i = 1, 2, \ldots, m$ entonces*

$$\begin{aligned} Y_1 &= \frac{X_1}{X_1 + X_2} \\ Y_2 &= \frac{X_1 + X_2}{X_1 + X_2 + X_2} \\ &\vdots \\ Y_{m-1} &= \frac{X_1 + X_2 + \ldots + X_{m-1}}{X_1 + X_2 + \ldots + X_{m-1} + X_m} \end{aligned}$$

son variables aleatorias mutuamente independientes cada una con distribución beta con parámetros $a = \sum_{i=1}^{j} n_i, b = \frac{n_j}{2} + 1$ para cada Y_j, $j = 1, 2, .., m$.

Otras manera en la cual surge la variable aleatoria beta, es en la distribución de las estadísticas de orden de n variables aleatorias independientes idénticamente distribuidas, cada una con un una distribución uniforme sobre el intervalo $(0,1)$, por lo que tenemos que la k-ésima estadística de orden (que definiremos más adelante) tiene una función de densidad dada como sigue:

$$f_{X_{(k)}}(x) = \begin{cases} \frac{1}{B(k,n-k+1)} x^{k-1}(1-x)^{n-k} & \text{si } 0 \le x \le 1 \\ 0 & \text{e.o.c.} \end{cases} .$$

Además cuando la variable aleatoria beta tiene parámetros $a = b = \frac{1}{2}$, se le conoce como la distribución arco seno ya que su función de distribución se da como sigue:

$$F_X(x) = \begin{cases} \frac{2}{\pi} \operatorname{arcsen} \sqrt{x} & \text{si } 0 < x < 1 \\ 0 & \text{e.o.c.} \end{cases} .$$

Esta variable surge de manera interesante en la teoría de procesos estocásticos cuando se estudian las caminatas aleatorias. Además de que tiene una gran diversidad de usos en la estadística matemática.

4.6 Estadísticas de orden

Una herramienta de gran utilidad en la estadística matemática es el cálculo de las estadísticas de orden de n variables aleatorias, para lo cual daremos a continuación su definición:

Definición 4.5 *Si $X_1, X_2, \ldots, X_n$ son variables aleatorias y $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ son las mismas variables aleatorias arregladas en orden de magnitud ascendente entonces $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ se les llama las estadísticas de orden correspondientes a X_1, X_2, $\ldots$, X_n.*

Ejemplo 4.1 *Considere una máquina que tiene n componentes cuyos tiempos de falla X_1, X_2, $\ldots$, X_n son variables aleatorias, podemos ver fácilmente que $X_{(k)}$ es el tiempo para que k de los componentes fallen, además, si la máquina se descompone cuando k de sus componentes fallan $X_{(k)}$ es el tiempo de falla de la máquina.*

Es claro que

$$\begin{aligned} X_{(1)} &= \min\{X_1, X_2, \ldots, X_n\} \\ \text{y } X_{(n)} &= \max\{X_1, X_2, \ldots, X_n\} \end{aligned}$$

A continuación desarrollaremos la teoría referente a las estadísticas de orden, suponiendo que las n variables aleatorias son independientes y absolutamente continuas, y que cada una tiene una función de distribución $F(\cdot)$ y una función de densidad $f(\cdot)$.

Teorema 4.8 *Sean $X_1, X_2, \ldots, X_n$ variables aleatorias independientes absolutamente continuas idénticamente distribuidas cada una con una función de distribución $F(\cdot)$ y una* *función de densidad $f(\cdot)$, y sean $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ las estadísticas de orden de X_1, $X_2, \ldots, X_n$, entonces la función de distribución marginal para $X_{(k)}$ es dada por*

$$F_{X_{(k)}}(x) = \sum_{i=k}^{n} \binom{n}{i} F^i(x) \cdot (1 - F(x))^{n-i}.$$

Dem.

Sea Z_x = número de $X_i \leq x$, es claro Z_x es una variable aleatoria binomial con parámetros n y $p = F(x)$, por lo que se tiene:

$$\begin{aligned} F_{X_{(k)}}(x) &= P\left\{X_{(k)} \leq x\right\} \\ &= P\{Z_x \geq k\} \\ &= \sum_{i=k}^{n} \binom{n}{i} F^i(x) \cdot (1 - F(x))^{n-i}, \end{aligned}$$

por lo que se establece el teorema.

Corolario 4.9 *Sean $X_1, X_2, \ldots, X_n$ variables aleatorias independientes absolutamente continuas idénticamente distribuidas cada una con una función de distribución $F(\cdot)$ y una función de densidad $f(\cdot)$ y sean $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ las estadísticas de orden de X_1, X_2, $\ldots$, X_n entonces*

$$F_{X_{(n)}}(x) = F^n(x) \quad y \quad F_{X_{(1)}}(x) = 1 - [1 - F(x)]^n$$

Teorema 4.10 *Sean $X_1, X_2, \ldots, X_n$ variables aleatorias independientes absolutamente continuas idénticamente distribuidas cada una con una función de distribución $F(\cdot)$, y una función de densidad $f(\cdot)$, y sean $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ las estadísticas de orden de X_1, X_2, $\ldots$, X_n entonces para $k = 1, 2, \ldots, n$*

$$f_{X_{(k)}}(x) = \frac{n!}{(k-1)!(n-k)!} F^{k-1}(x) \cdot [1 - F(x)]^{n-k} f(x).$$

Dem.

Tenemos que $f_{X_{(k)}}(x) = F'_{X_{(k)}}(x)$ de aquí tenemos que

$$\begin{aligned} f_{X_{(k)}}(x) &= \frac{d}{dx} \sum_{i=k}^{n} \binom{n}{i} F^i(x) \cdot (1 - F(x))^{n-i} \\ &= \sum_{i=k}^{n} \binom{n}{i} \frac{d}{dx} F^i(x) \cdot (1 - F(x))^{n-i} \end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=k}^{n} \frac{n!}{(n-i)!(i)!} i F^{i-1}(x)\,[1-F(x)]^{n-i} f(x) \\
&\quad - \sum_{i=k}^{n} \frac{n!(n-i)}{(n-i)!i!} F^{i}(x)\,[1-F(x)]^{n-i-1} f(x) \\
&= \sum_{i=k}^{n} \binom{n}{i} \Big[i\,[1-F(x)]^{n-i} F^{i-1}(x) f(x) \\
&\quad - (n-i) F^{i}(x)\,[1-F(x)]^{n-i-1} f(x) \Big],
\end{aligned}
$$

es claro que en la segunda suma si $i = n$ esta término de la suma es 0 por lo que

$$
\begin{aligned}
f_{X_{(k)}}(x) &= \sum_{i=k}^{n} \frac{n!}{(n-i)!(i-1)!} F^{i-1}(x)\,[1-F(x)]^{n-i} f(x) \\
&\quad - \sum_{i=k}^{n-1} \frac{n!}{(n-i-1)!i!} F^{i}(x)\,[1-F(x)]^{n-i-1} f(x),
\end{aligned}
$$

si en la segunda suma hacemos un cambio en el índice $i = j - 1$ entonces tenemos que

$$
\begin{aligned}
f_{X_{(k)}}(x) &= \sum_{i=k}^{n} \frac{n!}{(n-i)!(i-1)!} F^{i-1}(x)\,[1-F(x)]^{n-i} f(x) \\
&\quad - \sum_{j=k+1}^{n} \frac{n!}{(n-j)!(j-1)!} F^{j-1}(x)\,[1-F(x)]^{n-j} f(x) \\
&= \frac{n!}{(n-k)!(k-1)!} F^{k-1}(x)\,[1-F(x)]^{n-k} f(x),
\end{aligned}
$$

por lo que se establece el teorema.

Corolario 4.11 *Sean $X_1, X_2, \ldots, X_n$ variables aleatorias independientes absolutamente continuas idénticamente distribuidas cada una con una función de distribución $F(\cdot)$ y una función de densidad $f(\cdot)$ y sean $X_{(1)}$, $X_{(2)}$, $\ldots$, $X_{(n)}$ las estadísticas de orden de X_1, X_2, $\ldots$, X_n entonces*

$$f_{X_{(1)}}(x) = n\,[1-F(x)]^{n-1} f(x),$$

y

$$f_{X_{(n)}}(x) = nF^{n-1}(x) \cdot f(x).$$

Es claro que si se desea trabajar de una manera más general con las estadísticas de orden, lo más conveniente sería conocer su función de densidad conjunta por lo que enunciaremos el siguiente teorema:

Teorema 4.12 *Sean $X_1, X_2, \ldots, X_n$ variables aleatorias independientes absolutamente continuas idénticamente distribuidas cada una con una función de distribución $F(\cdot)$, y una función de densidad $f(\cdot)$, y sean $X_{(1)}$, $X_{(2)}$, ..., $X_{(n)}$ las estadísticas de orden de X_1, X_2, ..., X_n entonces*

$$f_{X_{(1)},X_{(2)},\ldots,X_{(n)}}(x_1, x_2, \ldots, x_n) = \begin{cases} n!f(x_1)f(x_2)\ldots f(x_n) & si\ x_1 < x_2 < \ldots < x_n \\ 0 & e.o.c. \end{cases}.$$

Dem. Dado que

$$f_{X_{(1)},X_{(2)},\ldots,X_{(n)}}(x_1, x_2, \ldots, x_n) = \frac{\partial^n}{\partial x_1 \partial x_2 \ldots \partial x_n} F_{X_{(1)},X_{(2)},\ldots,X_{(n)}}(x_1, x_2, \ldots, x_n)$$

y recordando la definición de derivada parcial de orden múltiple tenemos:

$$\begin{aligned}
f & \ X_{(1)},X_{(2)},\ldots,X_{(n)}(x_1, x_2, \ldots, x_n) \\
= & \lim_{\substack{\Delta x_i \to 0 \\ i=1,2,\ldots,n}} \frac{1}{\prod_{i=1}^{n} \Delta x_i} P\left\{x_1 < X_{(1)} \le x_1 + \Delta x_1, \ldots, x_n < X_{(n)} \le x_n + \Delta x_n\right\} \\
= & \lim_{\substack{\Delta x_i \to 0 \\ i=1,2,\ldots,n}} \frac{1}{\prod_{i=1}^{n} \Delta x_i} P\{una\ X_i \in (x_1, x_1 + \Delta x_1], \ldots, una\ X_i \in (x_n, x_n + \Delta x_n]\} \\
= & \lim_{\substack{\Delta x_i \to 0 \\ i=1,2,\ldots,n}} \frac{n!}{\prod_{i=1}^{n} \Delta x_i} [F(x_1 + \Delta x_1) - F(x_1)] \ldots [F(x_n + \Delta x_n) - F(x_n)] \\
= & \ n! \lim_{\substack{\Delta x_1 \to 0 \\ i=1,2,\ldots,n}} [F(x_1 + \Delta x_1) - F(x_1)] \ldots \lim_{\Delta x_n \to 0} [F(x_n + \Delta x_n) - F(x_n)] \\
= & \ n!f(x_1) \ldots f(x_n) \ si\ x_1 < \ldots < x_n,
\end{aligned}$$

de aquí se tiene que

$$f_{X_{(1)},X_{(2)},\ldots,X_{(n)}}(x_1, x_2, \ldots, x_n) = \begin{cases} n!f(x_1)f(x_2)\ldots f(x_n) & si\ x_1 < x_2 < \ldots < x_n \\ 0 & e.o.c. \end{cases},$$

por lo que se establece el teorema.

Ejemplo 4.2 *En un camino cuya longitud es de un kilómetro se encuentran 3 personas distribuidas aleatoriamente en forma uniforme. Encontrar la probabilidad de que la distancia entre cada una de ellas sea mayor que a, donde $a \le \frac{1}{2}$ km.*

Sol.

Ya que las tres personas se encuentran distribuidas aleatoriamente en forma uniforme en un camino de longitud 1, definimos:

X_1 =la posición en el camino de la primera persona.

X_2 =la posición en el camino de la segunda persona.

X_3 =la posición en el camino de la tercera persona.

Es claro que X_1, X_2, X_3 tienen una distribución uniforme sobre el intervalo $(0,1)$, por lo que utilizando el teorema 4.12 se tiene que si

$A = \{$la distancia entre cada una de las personas sea mayor que $a\}$

entonces:

$$\begin{aligned}
P\{A\} &= P\left\{X_{(2)} - X_{(1)} \leq a, X_{(3)} - X_{(2)} \leq a\right\} \\
&= \iiint_{\{(x_1,x_2):x_2-x_1\leq a, x_3-x_2\leq a, x_1<x_2<x_3\}} 3! \cdot f(x_1)f(x_2)f(x_3)dx_1dx_2dx_3 \\
&= 6 \iiint_{\{(x_1,x_2):x_2-x_1\leq a, x_3-x_2\leq a, x_1<x_2<x_3\}} dx_1dx_2dx_3 \\
&= 6\int_0^{1-2a}\int_{x_1}^{1-a}\int_{x_2+a}^{1} dx_3dx_2dx_1 \\
&= 6\int_0^{1-2a}\int_{x_1}^{1-a} 1 - x_2 - a dx_2dx_1,
\end{aligned}$$

haciendo $u_2 = 1 - x_2 - a$, $du_2 = -dx_2$ se tiene que

$$\begin{aligned}
P\{A\} &= 6\int_0^{1-2a}\left[-\int_{1-2a-x_1}^{0} u_2du_2\right]dx_1 \\
&= 6\int_0^{1-2a}\left[\int_0^{1-2a-x_1} u_2du_2\right]dx_1 \\
&= 6\int_0^{1-2a}\frac{(1-2a-x_1)^2}{2}dx_1,
\end{aligned}$$

haciendo $u_1 = 1 - 2a - x_1, du_1 = -dx_1$ se tiene

$$\begin{aligned}
P\{A\} &= -3\int_{1-2a}^{0} u_1{}^2 du_1 \\
&= 3\int_0^{1-2a} u_1{}^2 du_1 \\
&= (1-2a)^3.
\end{aligned}$$

Definición 4.6 *Si $n = 2m + 1$ es impar (es decir m es un entero) entonces $X_{(m+1)}$ se le conoce como la mediana de $X_1, X_2, \ldots, X_n$. En general el $100p\%$ percentil es definido únicamente si $(n+1)p$ es un número entero y es representado por la $(n+1)p$ estadística de orden,(i.e. por $X_{((n+1)p)}$). La mediana corresponde a $p = \frac{1}{2}$; tenemos también que el primer cuartil y el tercer cuartil para $p = \frac{1}{4}, \frac{3}{4}$ respectivamente.*

Ejemplo 4.3 *Si $X_1, X_2, \ldots, X_5$ son variables aleatorias independientes idénticamente distribuidas cada una con una distribución uniforme sobre el intervalo $(0, a)$. Encontrar la probabilidad de que la mediana de la muestre este entre $\frac{1}{4}a$ y $\frac{3}{4}a$.*

Sol.

Del teorema 4.10 tenemos que:

$$\begin{aligned} P\left\{\frac{1}{4}a < x < \frac{3}{4}a\right\} &= \frac{30}{a^5}\int_{\frac{1}{4}a}^{\frac{3}{4}a} x^2(a-x)^2 dx \\ &= \frac{30}{a^5}\int_{\frac{1}{4}a}^{\frac{3}{4}a} (a^2x^2 - 2ax^3 + x^4)dx \\ &= \frac{30}{a^5}\left(\frac{a^2x^3}{3} - \frac{ax^4}{2} + \frac{x^5}{5}\right)\Big|_{\frac{1}{4}a}^{\frac{3}{4}a} \\ &= \frac{203}{256}. \end{aligned}$$

Una variable aleatoria relacionada con las estadísticas de orden, es el rango de $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ y se define como $R = X_{(n)} - X_{(1)}$.

Teorema 4.13 *Sean $X_1, X_2, \ldots, X_n$ variables aleatorias independientes absolutamente continuas idénticamente distribuidas cada una con una función de distribución $F(\cdot)$, y una función de densidad $f(\cdot)$, y sean $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ las estadísticas de orden de $X_1, X_2, \ldots, X_n$ entonces $R = X_{(n)} - X_{(1)}$ es el rango de $X_1, X_2, \ldots, X_n$ entonces su función de densidad es:*

$$f_R(r) = \begin{cases} n(n-1)\int_{-\infty}^{\infty} f(x)f(r+x)\left[F(r+x) - F(x)\right]^{n-2} dx & si\ r > 0 \\ 0 & e.o.c. \end{cases}.$$

Dem.

Asumiremos que $n \geq 2$ ya que si $n = 1$ se tiene que $R = 0$. Para calcular la función de densidad de R, primero calcularemos la densidad conjunta de $X_{(1)}$ y $X_{(n)}$ y para lo cual podemos hacer la siguiente observación:

$$\begin{aligned} P\left\{X_{(1)} > x, X_{(n)} \leq y\right\} &= P\left\{x < X_1 \leq y, x < X_2 \leq y, \ldots, x < X_n \leq y\right\} \\ &= \begin{cases} [F(y) - F(x)]^n & si\ x \leq y \\ 0 & e.o.c. \end{cases}. \end{aligned}$$

Dado que

$$P\left\{X_{(n)} \leq y\right\} = P\left\{X_{(1)} > x, X_{(n)} \leq y\right\} + P\left\{X_{(1)} \leq x, X_{(n)} \leq y\right\},$$

se tiene que

$$
\begin{aligned}
F_{X_{(1)},X_{(n)}}(x,y) &= P\left\{X_{(1)} \leq x, X_{(n)} \leq y\right\} \\
&= P\left\{X_{(n)} \leq y\right\} - P\left\{X_{(1)} > x, X_{(n)} \leq y\right\} \\
&= \begin{cases} F^n(y) - [F(y) - F(x)]^n & si\ x \leq y \\ F^n(y) & si\ x > y \end{cases}.
\end{aligned}
$$

Por lo tanto

$$
F_{X_{(1)},X_{(n)}}(x,y) = \begin{cases} F^n(y) - [F(y) - F(x)]^n & si\ x \leq y \\ F^n(y) & si\ x > y \end{cases}.
$$

Por consiguiente

$$
\begin{aligned}
f_{X_{(1)},X_{(n)}}(x,y) &= \frac{\partial^2}{\partial x \partial y} F_{X_{(1)},X_{(n)}}(x,y) \\
&= \begin{cases} n(n-1)\left[F(y) - F(x)\right]^{n-2} f(x) f(y) & si\ x \leq y \\ 0 & e.o.c. \end{cases}.
\end{aligned}
$$

De aquí podemos utilizar 3.11 para calcular la función de densidad de R por lo cual

$$
\begin{aligned}
f_R(r) &= P\left\{X_{(n)} - X_{(1)} \leq r\right\} \\
&= \int_{-\infty}^{\infty} f_{X_{(1)},X_{(n)}}(x, r+x)dx \\
&= n(n-1)\int_{-\infty}^{\infty} [F(r+x) - F(x)]^{n-2} f(x) f(r+x)dx \ si\ r > 0.
\end{aligned}
$$

Por lo tanto

$$
f_R(r) = \begin{cases} n(n-1)\int_{-\infty}^{\infty} [F(r+x) - F(x)]^{n-2} f(x) f(r+x)dx & si\ r > 0 \\ 0 & e.o.c. \end{cases},
$$

lo cual establece el teorema.

Ejemplo 4.4 *Si $X_1, X_2, \ldots, X_n$ son variables aleatorias independientes idénticamente distribuidas con distribución uniforme sobre el intervalo $(0,1)$. Calcule la función de densidad de $R = X_{(n)} - X_{(1)}$.*

Sol.

Utilizando el teorema 4.13 tenemos:

$$
\begin{aligned}
f_R(r) &= n(n-1)\int_{-\infty}^{\infty} [F(r+x) - F(x)]^{n-2} f(x) f(r+x)dx \\
&= n(n-1)\int_0^{1-r} [r + x - x]^{n-2} dx \\
&= n(n-1)\int_0^{1-r} [r + x - x]^{n-2} dx \ si\ 0 < r < 1.
\end{aligned}
$$

Por lo que

$$f_R(r) = \begin{cases} n(n-1)r^{n-2}(1-r) & si\ 0 < r < 1 \\ 0 & e.o.c. \end{cases}.$$

4.7 Ejercicios

1. Si X tiene una distribución gamma con parámetros (n, λ).

   (a) ¿Cuál es la distribución de cX si $x > 0$?

   (b) Mostrar que $U = \frac{1}{2\lambda} \cdot \chi^2_{(2n)}$ tiene una la misma distribución que X.

2. Encuentre la función densidad de la variable aleatoria $U = \frac{\chi^2_{(n)}}{n}$ donde $\chi^2_{(n)}$ es una variable aleatoria ji-cuadrada con n grados de libertad.

3. Encuentre la función densidad de la variable aleatoria $U = \frac{\chi}{\sqrt{n}}$ donde χ es una variable aleatoria ji con n grados de libertad.

4. Si X tiene una distribución F con (n, m) grados de libertad, muestre que

$$Y = \frac{m\frac{X}{n}}{1 + m\frac{X}{n}}$$

   tiene una distribución beta.

5. Probar que si X tiene una distribución F con (m, n) grados de libertad, y si Y tiene una distribución F con (n, m) grados de libertad, entonces para cada $x > 0$ se tiene que:

$$P\{X \leq x\} = 1 - P\left\{Y \leq \frac{1}{x}\right\}.$$

6. Si X tiene una distribución t de Student con n grados de libertad entonces X^2 tiene una distribución F con $(1, n)$ grados de libertad.

7. Si Z tiene una distribución Cauchy, i.e., Z tiene una función de densidad

$$f_Z(z) = \frac{1}{\pi(1 + z^2)}.$$

   Probar que Z^2 tiene una distribución F con $(1, 1)$ grados de libertad.

8. Probar que si X tiene una distribución t de Student con n grados de libertad, entonces la función de distribución de X es simétrica, i.e., $F_X(x) = 1 - F_X(-x)$.

9. Probar que si X tiene una distribución t de Student con n grados de libertad. Probar que

$$P\{|X| \le x\} = 1 - 2(1 - P\{X \le x\})$$

10. Demostrar el teorema 4.7.

    Si $X_1, X_2, \ldots, X_n$ son variables aleatorias, definimos las variables aleatorias siguientes:

    $\overline{X}_k = \frac{1}{k}\sum_{i=1}^{k} X_i \ (k = 1, 2, \ldots, n).$

    $S_k^2 = \frac{1}{k-1}\sum_{i=1}^{k}(X_i - \overline{X_k})^2 \ (k = 2, 3, \ldots, n).$

    $\overline{X}_{n-k} = \frac{1}{n-k}\sum_{i=k+1}^{n} X_i \ (k = 1, 2, \ldots, n-1).$

    $S_{n-k}^2 = \frac{1}{n-k-1}\sum_{i=k+1}^{n}(X_i - \overline{X}_{n-k})^2 \ (k = 1, 2, \ldots, n-1).$

    $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$

    $S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}.$

    $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$

11. Si $X_1, X_2, \ldots, X_n$ variables aleatorias independientes idénticamente distribuidas, cada una con una distribución normal estándar.

    (a) ¿Cuáles son las distribuciones de $\overline{X}_k$ y $\overline{X}_{n-k}$?

    (b) ¿Cuál es la distribución de $Y = k\overline{X}_k + (n-k)\overline{X}_{n-k}$?

    (c) ¿Cuál es la distribución de $Y = \frac{1}{2}\left\{\overline{X}_k + \overline{X}_{n-k}\right\}$?

12. Si $X_1, X_2, \ldots, X_n$ son variables aleatorias independientes idénticamente distribuidas, cada una con una distribución normal con parámetros μ y σ^2. Calcular las distribuciones de las siguientes variables aleatorias.

    (a) De $\frac{X_i - \mu}{\sigma^2}$ para $i = 1, 2, \ldots, n$.

    (b) De $\overline{X}_k$ y $\overline{X}_{n-k}$.

    (c) De S y S^2.

(d) De $\frac{1}{2}\left\{\overline{X}_k + \overline{X}_{n-k}\right\}$.

(e) De $\frac{(k-1)S_k^2+(n-k-1)S_{n-k}^2}{\sigma^2}$.

(f) De $\frac{S_k^2}{S_{n-k}^2}$

(g) De $(\overline{X}-\mu)/(S/\sqrt{n})$.

13. Sean $X_1, X_2, \ldots, X_n$ variables aleatorias independientes idénticamente distribuidas con función de distribución $F(\cdot)$, y función de densidad $f(\cdot)$. La variable aleatoria

$$M = \left[X_{(1)} + X_{(n)}\right]/2,$$

la cual es el promedio entre la variable aleatoria más grande y la más pequeña, es llamada rango medio. Mostrar que su función de distribución es

$$F_M(m) = n\int_{-\infty}^{\infty}\left[F(2m-x) - F(x)\right]^{n-1} f(x)dx.$$

14. Sean $X_1, X_2, \ldots, X_n$ variables aleatorias independientes idénticamente distribuidas cada una con una distribución uniforme sobre el intervalo $(0, 1)$. Encontrar:

(a) La función de densidad de $X_{(1)}$.

(b) La función de densidad de $X_{(n)}$.

(c) La función de densidad conjunta de $X_{(1)}$ y $X_{(n)}$.

(d) La función de densidad del rango de $X_1, X_2, \ldots, X_n$.

(e) La función de densidad del rango medio de $X_1, X_2, \ldots, X_n$.

(f) La función de densidad de de la mediana suponiendo que n es impar.

(g) La función de densidad conjunta de $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$.

15. Sean $X_1, X_2, \ldots, X_n$ variables aleatorias independientes idénticamente distribuidas cada una con una distribución uniforme sobre el intervalo (a, b). Encontrar los incisos a), b),...,g) definidos en el problema anterior.

16. Sean $X_1, X_2, \ldots, X_n$ variables aleatorias independientes idénticamente distribuidas con función de distribución $F(\cdot)$ y función de densidad $f(\cdot)$. Encontrar la función de densidad conjunta de $X_{(i)}$ y $X_{(j)}$, para $i, j = 1, 2, \ldots, n$.

17. Sean $X_1, X_2, \ldots, X_n$ variables aleatorias independientes idénticamente distribuidas cada una con una distribución uniforme sobre el intervalo $(0,1)$. Demostrar que para $1 \leq k \leq n+1$ se tiene que

$$P\left\{X_{(k)} - X_{(k-1)} > t\right\} = (1-t)^n \text{ si } 0 \leq t \leq 1,$$

donde $X_{(0)} \equiv 0$ y $X_{(n+1)} \equiv$t.

18. Sean $X_1, X_2, \ldots, X_n$ variables aleatorias independientes idénticamente distribuidas con función de distribución $F(\cdot)$, y función de densidad $f(\cdot)$. Si X es una variable aleatoria que es independiente de X_i para $i = 1, 2, \ldots, n$, que también tiene una función de distribución $F(\cdot)$ y función de densidad $f(\cdot)$. Determinar:

    (a) $P\left\{X > X_{(n)}\right\}$.
    (b) $P\left\{X > X_{(n)}\right\}$.
    (c) $P\left\{X_{(i)} < X < X_{(j)}\right\}$, si $1 \leq i < j \leq n$.

19. Si X_1, X_2, X_3, X_4, X_5 son variables aleatorias independientes idénticamente distribuidas cada una con una distribución exponencial con parámetro λ, calcular

$$P\left\{X_{(1)} \leq a\right\} \text{ y } P\left\{X_{(n)} \leq a\right\}$$

si $a > 0$.

20. Si tres camiones se descomponen en puntos aleatorias distribuidos en un camino de longitud L, encontrar la probabilidad de que al menos dos camiones se encuentren a una distancia mayor que d, donde $d \leq \frac{L}{2}$.

21. Si X_1, X_2, X_3, X_4, X_5 son variables aleatorias independientes idénticamente distribuidas cada una con una distribución uniforme sobre el intervalo $(0,1)$, calcular la probabilidad de que la mediana de X_1, X_2, X_3, X_4, X_5 se encuentre en el intervalo $(0, \frac{1}{2})$ o en el intervalo $(\frac{3}{4}, 1)$.

Capítulo 5

Esperanza, Varianza y Momentos

En este capítulo veremos algunas características numéricas de las variables aleatorias, como son la esperanza, la varianza, y más generalmente los momentos. Sabemos que una variable aleatoria está completamente determinada por su función de distribución, ya que si conocemos ésta, podemos determinar los valores que puede tomar esa variable aleatoria y sus probabilidades. Sin embargo, en un gran número de casos, necesitamos únicamente tener una idea general del comportamiento de una variable aleatoria. Por lo cual resulta de extrema importancia para la teoría de la probabilidad conocer la esperanza, la varianza y los momentos de una variable aleatoria, los cuales son características numéricas que nos ayudan a identificar más fácilmente la distribución de una cierta variable aleatoria.

5.1 Esperanza

Consideremos un juego de azar en el cual es posible obtener n resultados para los cuales conocemos que la probabilidad de obtener el primer resultado es p_1, para el segundo es p_2, para el i-ésimo p_i y por último para el n-ésimo es p_n. Al resultado de este juego le podemos asignar una variable aleatoria X, donde $X = i$ si se obtiene el i-ésimo posible resultado, $i = 1, \ldots, n$. Como X es una variable aleatoria tenemos que

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

Supongamos que realizamos un número muy grande de juegos, y para cada uno de ellos alguien nos da i pesos si obtenemos el i-ésimo resultado. Es válido suponer que si nos pagan por obtener un cierto resultado, también soliciten un pago por realizar cada juego, de tal forma que al terminar de jugar, tanto el que paga, como el jugador desearían no haber perdido dinero (juego justo). De esta manera nos interesaría saber la cantidad exacta que tendría que pagar un jugador por tener derecho a realizar un juego. Por el concepto de probabilidad frecuentista sabemos que si jugamos un número muy grande de juegos la proporción aproximada de veces que se obtiene el i-ésimo resultado es p_i, por lo que al final en promedio pagaremos

$$1p_1 + 2p_2 + \ldots + np_n = \sum_{i=1}^{n} ip_i,$$

ésto es siempre que el número de experimentos realizados fue muy grande. Esta constante es la que comúnmente se le llama esperanza y se define de manera similar para cualquier variable aleatoria discreta, es decir:

Definición 5.1 *Sea X una variable aleatoria discreta que toma valores en $x_1, x_2, \ldots$ y sea $f_X(\cdot)$ su función de densidad, si la serie*

$$\sum_{i=1}^{\infty} |x_i| P\{X = x_i\}$$

converge, entonces

$$E[X] = \sum_{i=1}^{\infty} x_i P\{X = x_i\}$$

se le llama esperanza de la variable aleatoria X.

Ejemplo 5.1 *Un juego muy popular de dados en los bares británicos, el cual es jugado por la casa lanzando 3 dados y en el cual un jugador puede apostar a cualquiera de los resultados del 1 al 6. Si uno de los tres dados muestra el valor apostado el jugador no pierde ni gana; si dos dados muestran el valor escogido el pago es de dos a uno; si los tres dados muestran su selección el pago es de tres a uno; por último si ninguno de los dados muestra la selección del jugador entonces el jugador pierde. Calcule la ganancia esperada si el jugador apuesta una unidad en este juego.*

Sol. Sea W la variable aleatoria que denota la ganancia del jugador, es claro que W puede tomar los valores $-1, 0, 1, 2$ por lo cual

$$E[W] = \sum_{w: P\{W=w\}>0} wP\{W = w\}$$

donde únicamente falta calcular la función de distribución de W. Sean $X_i, i = 1,2,3$ variables aleatorias que denotan el resultado del i-ésimo experimento. Si el jugador apuesta al resultado j-ésimo $j = 1, \ldots, 6$ entonces

$$\begin{aligned}
P\{W=-1\} &= P\{X_1 \neq j, X_2 \neq j, X_3 \neq j\} \\
&= \left(\tfrac{5}{6}\right)^3, \\
P\{W=0\} &= P\{X_1 = j, X_2 \neq j, X_3 \neq j\} + \\
&\quad P\{X_1 \neq j, X_2 = j, X_3 \neq j\} + \\
&\quad P\{X_1 \neq j, X_2 \neq j, X_3 = j\} \\
&= 3\left(\tfrac{1}{6}\right)\left(\tfrac{5}{6}\right)^2, \\
P\{W=1\} &= P\{X_1 = j, X_2 = j, X_3 \neq j\} + \\
&\quad P\{X_1 = j, X_2 \neq j, X_3 = j\} + \\
&\quad P\{X_1 \neq j, X_2 = j, X_3 = j\} \\
&= 3\left(\tfrac{1}{6}\right)^2\left(\tfrac{5}{6}\right), \\
P\{W=2\} &= P\{X_1 = j, X_2 = j, X_3 = j\} \\
&= \left(\tfrac{1}{6}\right)^3,
\end{aligned}$$

ya que todas las X_i son variables aleatorias independientes, por lo que el valor esperado de W sería

$$\begin{aligned}
E[W] &= \sum_{w:P\{W=w\}>0} wP\{W=w\} \\
&= -1P\{W=-1\} + 0P\{W=0\} + \\
&\quad 1P\{W=1\} + 2P\{W=2\} \\
&= -\left(\tfrac{5}{6}\right)^3 + 3\left(\tfrac{1}{6}\right)^2\left(\tfrac{5}{6}\right)2\left(\tfrac{1}{6}\right)^3 \\
&= -\tfrac{108}{108} \\
&= -1.
\end{aligned}$$

Este resultado indica que a la larga el jugador pierde en este juego, por consiguiente este juego no es un juego justo y no conviene jugarlo.

Ejemplo 5.2 *Un contratista ha encontrado a través de la experiencia que la oferta más baja para un trabajo (excluyendo su propia oferta) es una variable aleatoria que se distribuye uniformemente en el intervalo* $(\frac{3}{4}c, 2c)$*, donde c es el costo estimado del contratista (sin ganancias o pérdidas) del trabajo. Si la ganancia es definida como cero si el contratista no obtiene el trabajo (su oferta es mayor que la oferta más baja) y como la diferencia entre su oferta y su costo estimado c si se obtiene el trabajo. ¿Qué podría ofrecer el (en términos de c) para maximizar su ganancia esperada?*

Sol.

Sea G_k la variable aleatoria que denota la ganancia esperada, considerando que el contratista siempre realiza una oferta k, donde k se encuentra en el intervalo $(c, 2c)$, ya que si k es menor que c el contratista siempre tendría pérdida. Sea O la variable aleatoria que denota la oferta más baja, por consiguiente es claro que

$$G_k = \begin{cases} k - c & si\ k \leq O \\ 0 & si\ k > O \end{cases},$$

calculando la función de densidad tenemos

$$P\{G_k = x\} = \begin{cases} P\{k > O\} & si\ x = 0 \\ P\{k \leq O\} & si\ x = k - c \\ 0 & e.o.c. \end{cases},$$

de donde utilizando la definición anterior tenemos que

$$\begin{aligned} E[G_k] &= \sum_{x=0,k-c} xP\{G_k = x\} \\ &= 0P\{k > O\} + (k-c)P\{k \leq O\} \\ &= (k-c)\int_k^{2c} \frac{dx}{2c - \frac{3}{4}c} \\ &= \frac{4(k-c)}{5c}(2c - k) \\ &= \frac{4}{5}c(3ck - k^2 - 2c^2). \end{aligned}$$

Dado que $E[G_k]$ es una función que depende de k únicamente si el valor de c es fijo, podemos encontrar sus valores máximos (si es que los hay) derivando esta función con respecto a k, por lo que tenemos que si

$$h(k) = G_k,$$

tenemos que

$$h'(k) = \frac{4}{5c}(3c - 2k),$$

igualando con cero esta ecuación tenemos que

$$k = \frac{3}{2}c,$$

derivando nuevamente se puede comprobar que $k = \frac{3}{2}c$ es un máximo para esta función. Por lo que se le recomendaría al contratista que siempre obtuviera una ganancia igual a la mitad de sus costos.

Para las variables aleatorias absolutamente continuas resulta de forma natural, dada la generalización del concepto de suma por medio de la integral, definir la esperanza como sigue:

Definición 5.2 *Si $f(x)$ es la función de densidad de una variable aleatoria absolutamente continua X y*

$$\int_{-\infty}^{\infty} |x| f(x)dx$$

converge entonces la esperanza de X se denotará por $E[X]$ y se define como

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx.$$

Ejemplo 5.3 *La vida en horas de una lámpara fluorescente es una variable aleatoria que tiene una función de densidad de probabilidad que depende de una constante positiva α dada por*

$$f(x) = \begin{cases} \alpha^2 xe^{-\alpha x} & si\ x > 0 \\ 0 & e.o.c. \end{cases}.$$

Calcular la duración esperada de tal lámpara.

Sol.

Dada la definición anterior tenemos que

$$\begin{aligned} E[X] &= \textstyle\int_{-\infty}^{\infty} xf(x)dx \\ &= \textstyle\int_0^{\infty} xf(x)dx \\ &= \textstyle\int_0^{\infty} \alpha^2 x^2 e^{-\alpha x}dx, \end{aligned}$$

integrando por partes y haciendo $u = \alpha^2 x^2$ y $dv = e^{-\alpha x}dx$ tenemos que

$$\begin{aligned} E[X] &= -2\alpha^2 xe^{-\alpha x} |_0^{\infty} + 2\textstyle\int_0^{\infty} \alpha xe^{-\alpha x}dx \\ &= 2\textstyle\int_0^{\infty} \alpha xe^{-\alpha x}dx, \end{aligned}$$

nuevamente si hacemos $u = 2\alpha x$ y $dv = e^{-\alpha x}dx$ entonces tenemos

$$\begin{aligned} E[X] &= -2\alpha xe^{-\alpha x} |_0^{\infty} + 2\textstyle\int_0^{\infty} e^{-\alpha x}dx \\ &= -2\tfrac{e^{\alpha x}}{\alpha} |_0^{\infty} \\ &= \tfrac{2}{\alpha}. \end{aligned}$$

Por consiguiente la duración esperada de una lámpara sera $\frac{2}{\alpha}$ horas.

Ejemplo 5.4 *Muestre que si X es una variable aleatoria Cauchy, es decir, X tiene una función de densidad*

$$f(x) = \frac{1}{\pi(1+x^2)},$$

entonces no tiene definida una esperanza.

Sol.

En la definición de esperanza se dice que si

$$\int_{-\infty}^{\infty} |x|\, f(x)dx$$

no converge entonces tenemos que $E[X]$ no existe, por lo que tenemos que

$$\int_{-\infty}^{\infty} \frac{|x|}{\pi(1+x^2)}dx = 2\int_{0}^{\infty} \frac{x}{\pi(1+x)^2}dx,$$

y basta ver que

$$\int_{0}^{\infty} \frac{x}{\pi(1+x)^2}dx$$

no existe. Ya que

$$\int_{0}^{\infty} \frac{x}{\pi(1+x)^2}dx = \lim_{h\to\infty}\int_{0}^{h} \frac{x}{\pi(1+x)^2}dx$$

haciendo el cambio de variable $x = \tan u$ podemos tratar de resolver resolver esta integral, y de aquí podemos observar que este límite diverge, ésto es

$$\begin{aligned} \lim_{h\to\infty} \textstyle\int_0^h \frac{x}{\pi(1+x)^2}dx &= -\tfrac{1}{\pi}\lim_{h\to\infty} \ln \tfrac{1}{\sqrt{1+x^2}} \Big|_0^h \\ &= -\tfrac{1}{\pi}\lim_{h\to\infty} \ln \tfrac{1}{\sqrt{1+h^2}}, \end{aligned}$$

por lo que tenemos que $E[X]$ no existe.

A continuación enunciaremos un teorema de gran utilidad para el cálculo de la esperanza de una variable aleatoria, la cual es conocida como la ley del estadístico inconsciente, la cual nos ayuda a obtener la esperanza de cualquier función de una variable aleatoria sabiendo únicamente la función de distribución de la variable aleatoria.

Teorema 5.1 (Ley del Estadístico Inconsciente)

Si X es una variable aleatoria y g es una función real tal que $E[g(X)]$ existe tenemos que:

1. *Si X es una variable aleatoria discreta con función de densidad f_X, entonces*

$$E[g(X)] = \sum_{x:f(x)>0} g(x)f_X(x).$$

2. *Si X es una variable aleatoria absolutamente continua con función de densidad f_X, entonces*

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

Este teorema puede generalizarse para cualquier variable aleatoria, pero esto no lo haremos ya que para ello se requieren herramientas de probabilidad avanzada.

La demostración de la ley del estadístico inconsciente en el caso discreto es directa, así que la haremos únicamente para el caso de variables absolutamente continuas. Esta demostración se basa principalmente en el siguiente lema que es válido para cualquier variable aleatoria con esperanza finita, el cual demostraremos únicamente para variables absolutamente continuas

Lema 5.2 *Para cualquier variable aleatoria Y con esperanza finita, se tiene*

$$E[Y] = \int_0^{\infty} P\{Y > y\} dy - \int_{-\infty}^{0} P\{Y < -y\} dy.$$

Dem.

Si Y tiene una función de densidad f_Y tenemos que

$$F_Y(y) = \int_{-\infty}^{y} f_Y(u) du,$$

por lo que tenemos que

$$\int_0^{\infty} P\{Y > y\} dy = \int_0^{\infty} \int_y^{\infty} f_Y(u) du dy,$$

intercambiando el orden de integración se tiene que

$$\begin{aligned} \int_0^{\infty} P\{Y > y\} dy &= \int_0^{\infty} \int_0^{y} f_Y(u) du dy \\ &= \int_0^{\infty} \int_0^{u} dy f_Y(u) du \\ &= \int_0^{\infty} u f_Y(u) du. \end{aligned}$$

De manera análoga tenemos que

$$\begin{aligned} \int_{-\infty}^{0} P\{Y < -y\} dy &= \int_{-\infty}^{0} \int_{-\infty}^{-y} f_Y(u) du dy \\ &= \int_{-\infty}^{0} \int_0^{-u} f_Y(u) dy du \\ &= \int_{-\infty}^{0} f_Y(u) \int_0^{-u} dy du \\ &= \int_{-\infty}^{0} f_Y(u)(-u) du \\ &= -\int_{-\infty}^{0} u f_Y(u) du. \end{aligned}$$

De la definición de esperanza para una variable absolutamente continua se tiene que

$$\begin{aligned}
E[Y] &= \textstyle\int_{-\infty}^{\infty} u f_Y(u) du \\
&= \textstyle\int_0^{\infty} u f_Y(u) du + \int_0^{\infty} u f_Y(u) du \\
&= \textstyle\int_0^{\infty} P\{Y > y\}\, dy - \int_0^{\infty} P\{Y < -y\}\, dy,
\end{aligned}$$

lo que concluye con la demostración de este lema.

Este lema es de gran importancia en la demostración del teorema 5.1, por lo que la demostración del teorema 5.1 es como sigue:

Dem. **(Ley del Estadístico inconsciente)**

Sea $g(x)$ cualquier función de valor real, entonces, dado que $g(X)$ es una variable aleatoria por el lema 5.2 se tiene que

$$\begin{aligned}
E[g(X)] &= \textstyle\int_0^{\infty} P\{g(X) > x\}\, dx - \int_0^{\infty} P\{g(X) < -x\}\, dx \\
&= \int_0^{\infty} \int_{y:g(y)>x} f_X(y) dy dx - \int_0^{\infty} \int_{y:g(y)<-x} f_X(y) dy dx,
\end{aligned}$$

de aquí utilizando el teorema de Fubini tenemos que

$$\begin{aligned}
E[g(X)] &= \int_{y:g(y)>x} f_X(y) \textstyle\int_0^{g(y)} dx dy - \displaystyle\int_{y:g(y)<x} f_X(y) \textstyle\int_0^{-g(y)} dx dy \\
&= \int_{y:g(y)>x} g(y) f_X(y) dy + \int_{y:g(y)<x} g(y) f_X(y) dy \\
&= \textstyle\int_{-\infty}^{\infty} g(y) f_X(y) dy,
\end{aligned}$$

por lo que el teorema queda demostrado.

La ley del estadístico inconsciente (L.E.I.) nos ayuda a calcular de una manera muy sencilla esperanzas, que de otra forma serían muy difíciles de calcular, y es de gran ayuda para calcular otras características numéricas de las variables aleatorias, como son la varianza y los momentos.

Ejemplo 5.5 *Sea X una variable aleatoria continua con esperanza finita y función de densidad f. Mostrar que $E[|X - a|]$ es mínima cuando a es la mediana (la mediana de X cuando esta es continua es el valor a tal que $P\{X \leq a\} = .5$).*

Sol.

Tenemos que la función $g(a) = E[|X-a|]$. Por la L.E.I. tenemos que

$$\begin{aligned}
g(a) &= E[|X-a|] \\
&= \int_{-\infty}^{\infty} |x-a| f(x)dx \\
&= \int_a^{\infty} (x-a) f(x)dx \\
&\quad -\int_{-\infty}^{a} (x-a) f(x)dx \\
&= \int_a^{\infty} x f(x)dx - a\int_a^{\infty} f(x)dx \\
&\quad -\int_{-\infty}^{a} x f(x)dx + a\int_{-\infty}^{a} f(x)dx \\
&= \int_a^{\infty} x f(x)dx - \int_{-\infty}^{a} x f(x)dx \\
&\quad -aP\{X > a\} + aP\{X \leq a\} \\
&= \int_a^{\infty} x f(x)dx + \int_{-\infty}^{a} x f(x)dx + aF(a) \\
&\quad -a(1-F(a)) - 2\int_{-\infty}^{a} x f(x)dx \\
&= E[X] - 2\int_{-\infty}^{a} x f(x)dx \\
&\quad -a + 2aF(a),
\end{aligned}$$

de aquí, utilizando el teorema fundamental del cálculo y dado que $E[X]$ es una constante tenemos que

$$\begin{aligned}
g'(a) &= -2af(a) - 1 + 2af(a) + 2F(a) \\
&= 2F(a) - 1,
\end{aligned}$$

igualando con cero tenemos que

$$F(a) = .5,$$

por lo que tenemos lo que queríamos mostrar

Ejemplo 5.6 *Sea Z una variable aleatoria con distribución normal estándar, y para una a fija, sea*

$$X = \begin{cases} Z & si\ Z > a \\ 0 & e.o.c. \end{cases},$$

Mostrar que $E[X] = \frac{1}{\sqrt{2\pi}} e^{-\frac{a^2}{2}}$.

Sol.

Por la L.E.I. tenemos que

$$E[X] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} g(u) e^{-\frac{u^2}{2}} du,$$

donde

$$g(u) = \begin{cases} u & si\ u > a \\ 0 & e.o.c. \end{cases},$$

por lo que

$$
\begin{aligned}
E[X] &= \frac{1}{\sqrt{2\pi}} \int_a^{\infty} u e^{-\frac{u^2}{2}} du \\
&= -\frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \Big|_a^{\infty} \\
&= \frac{1}{\sqrt{2\pi}} e^{-\frac{a^2}{2}},
\end{aligned}
$$

por lo tanto

$$E[X] = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

5.2 Varianza

La esperanza de una variable aleatoria es de gran utilidad, ya que nos da un promedio de los posibles valores de nuestra variable, pero si nosotros queremos saber que tanto están dispersos los valores que se obtienen con ésta, nosotros no podemos contestar estas preguntas conociendo únicamente la esperanza de la variable, de esta forma desearíamos tener una medida de esta variabilidad, la cual mediremos con la varianza, ya que ésta nos da una idea de ésto.

Definición 5.3 *Si X es una variable aleatoria con media μ, entonces la varianza de X está definida por*

$$\operatorname{Var}(X) = E\left[(X-\mu)^2\right].$$

Claramente tenemos que la varianza es finita si y sólo sí

$$\int (x-\mu)^2 f_X(x) dx < \infty$$

en el caso de tener una variable aleatoria absolutamente continua y

$$\sum_x (x-\mu)^2 f_X(x) < \infty$$

en el caso de una variable aleatoria discreta.

La razón de medir la variabilidad de esta forma y no de otra, por ejemplo el de obtener la esperanza de la distancia entre X y su media, es que ésta última en ocasiones es muy difícil de calcular ya que interviene un valor absoluto y en ocasiones resulta dificil de calcular, por lo que comúnmente se opta por la definición de varianza que hemos dado, aunque en ocasiones también podemos utilizar a la esperanza de la distancia entre X y su media también como una medida de variabilidad de X.

Ejemplo 5.7 *Sea X una variable aleatoria normal con parámetros μ y σ^2. Calcular la varianza y la esperanza de X.*

Sol.

Recordemos que

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x f_x(x) dx \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx, \end{aligned}$$

haciendo $x = \sigma u + \mu$ tenemos que $u = \frac{x-\mu}{\sigma}$ y $dx = \sigma du$, por lo cual

$$\begin{aligned} E[X] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma u + \mu) e^{-\frac{u^2}{2}} du \\ &= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u e^{\frac{u^2}{2}} du \\ &\quad + \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{u^2}{2}} du \\ &= \mu. \end{aligned}$$

Por la L. E. I. podemos ver que

$$\begin{aligned} \mathrm{Var}(X) &= E\left[(X-\mu)^2\right] \\ &= \int_{-\infty}^{\infty} (x-\mu)^2 f_x(x) dx \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (x-\mu)^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx, \end{aligned}$$

haciendo $x = \sigma u + \mu$ tenemos que $u = \frac{x-\mu}{\sigma}$ y $dx = \sigma du$, por lo que

$$\mathrm{Var}(X) = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u^2 e^{-\frac{u^2}{2}} du,$$

calculando esta integral por partes tenemos que

$$\begin{aligned} \mathrm{Var}(X) &= \frac{\sigma^2}{\sqrt{2\pi}} - u e^{\frac{u^{24}}{2}} \Big|_{-\infty}^{\infty} \\ &\quad + \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{u^2}{2}} du \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{u^2}{2}} du \\ &= \sigma^2. \end{aligned}$$

Dado que en ocaciones el cálculo de la varianza de una variable aleatoria resulta muy complicado utilizando la definición anterior, debemos encontrar una manera más sencilla de calcular la varianza de una variable aleatoria. Así que a continuación estudiaremos las propiedades más importantes de la esperanza, así como también las propiedades de la varianza.

Proposición 5.1 *Si a y b son constantes, entonces*

$$E[aX+b]=aE[X]+b.$$

<u>*Dem.*</u>

Haremos esta demostración únicamente en el caso en el que X sea una variable aleatoria absolutamente continua. La demostración en el caso continuo es completamente análoga por lo cual la omitiremos.

Por la L.E.I. tenemos que

$$\begin{aligned} E[aX+b] &= \int_{-\infty}^{\infty}(ax+b)f(x)dx \\ &= \int_{-\infty}^{\infty}axf(x)dx \\ &\quad +\int_{-\infty}^{\infty}bf(x)dx \\ &= a\int_{-\infty}^{\infty}xf(x)dx \\ &\quad +b\int_{-\infty}^{\infty}f(x)dx \\ &= aE[X]+b \end{aligned}$$

De esta proposición se puede esperar que la esperanza sea un operador lineal. Lo cual lo demostraremos más adelante, para ello enunciaremos el siguiente teorema:

Proposición 5.2 *Si $a\in\mathcal{R}$, entonces $E[a]=a$.*

<u>*Dem.*</u>

Es claro que a es una variable aleatoria que toma el valor a con probabilidad 1, entonces su función de densidad esta definida como sigue:

$$f_a(x)=\begin{cases} 1 & si\ x=a \\ 0 & e.o.c. \end{cases}$$

de aquí tenemos que

$$\begin{aligned} E[a] &= \sum_x xf_a(x) \\ &= af_a(a) \\ &= a \end{aligned}$$

Podemos establecer un enunciado análogo a la L.E.I., que dice que si X y Y son dos variables aleatorias absolutamente continuas que tienen una función de densidad conjunta $f_{X,Y}(\cdot,\cdot)$ y $g(x,y)$ es una función de $\mathcal{R}^2$ en $\mathcal{R}$,entonces

$$E[g(X,Y)]=\int_{\mathcal{R}^2}g(x,y)f_{X,Y}(x,y)dxdy.$$

También si X y Y son variables aleatorias discretas con función de densidad conjunta $f_{X,Y}(\cdot,\cdot)$ tenemos que

$$E[g(X,Y)] = \sum_x \sum_y g(x,y) f_{X,Y}(x,y).$$

De esta manera podemos enunciar la siguiente proposición:

Proposición 5.3 *Si X y Y son dos variables aleatorias absolutamente continuas con función de densidad conjunta $f_{X,Y}(\cdot,\cdot)$, entonces tenemos que*

$$E[X+Y] = E[X] + E[Y].$$

<u>*Dem.*</u>

$$\begin{aligned} E[X+Y] &= \int_{\mathcal{R}^2} (x+y) f_{X,Y}(x,y)dxdy \\ &= \int_{\mathcal{R}^2} x f_{X,Y}(x,y)dxdy \\ &\quad + \int_{\mathcal{R}^2} y f_{X,Y}(x,y)dxdy \\ &= E[X] + E[Y]. \end{aligned}$$

De aquí por las proposiciones 5.1 y 5.3 tenemos que la esperanza es un operador lineal. Por esta propiedad de la esperanza podemos encontrar una manera más sencilla de calcular la varianza, por lo que enunciaremos la siguiente proposición:

Proposición 5.4 $\text{Var}(X) = E[X^2] - E^2[X]$.

<u>*Dem.*</u>

$$\begin{aligned} \text{Var}(X) &= E[(X-\mu)^2] \\ &= E[X^2 - 2\mu X + \mu^2] \\ &= E[X^2] \\ &\quad -2\mu E[X] + E[\mu^2] \\ &= E[X^2] - 2\mu^2 + \mu^2 \\ &= E[X^2] - \mu^2 \\ &= E[X^2] - E^2[X]. \end{aligned}$$

Ejemplo 5.8 *Sea X una variable aleatoria gamma con parámetros (λ, α). Calcular la esperanza y la varianza de X.*

<u>*Sol.*</u>

Es claro que

$$\begin{aligned} E[X] &= \int_0^\infty x \frac{\lambda^\alpha e^{-\lambda x} x^{\alpha-1}}{\Gamma(\alpha)} dx \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty e^{-\lambda x} x^\alpha dx, \end{aligned}$$

haciendo $u = \lambda x$ *tenemos que* $du = \lambda dx$, *por lo que*

$$\begin{aligned} E[X] &= \frac{\lambda^\alpha}{\Gamma(\alpha)\lambda} \int_0^\infty e^{-u} \left(\frac{u}{\lambda}\right)^\alpha du \\ &= \frac{1}{\Gamma(\alpha)\lambda} \int_0^\infty e^{-u} u^\alpha du \\ &= \frac{1}{\Gamma(\alpha)\lambda} \int_0^\infty e^{-u} u^{(\alpha+1)-1} du \\ &= \frac{1}{\Gamma(\alpha)\lambda} \Gamma(\alpha+1) \\ &= \frac{\alpha}{\lambda}. \end{aligned}$$

También tenemos que

$$\begin{aligned} \mathrm{Var}(X) &= E\left[(X-\mu)^2\right] \\ &= \int_0^\infty (x-\mu)^2 \frac{\lambda^\alpha e^{-\lambda x} x^{\alpha-1}}{\Gamma(\alpha)} dx, \end{aligned}$$

la cual es una integral en extremo complicada, pero dado que

$$\mathrm{Var}(X) = E\left[X^2\right] - E^2[X]$$

necesitamos calcular únicamente la $E\left[X^2\right]$ *para obtener la* $\mathrm{Var}(X)$, *por lo cual*

$$\begin{aligned} E\left[X^2\right] &= \int_0^\infty x^2 \frac{\lambda^\alpha e^{-\lambda x} x^{\alpha-1}}{\Gamma(\alpha)} dx \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty e^{-\lambda x} x^{\alpha+1} dx, \end{aligned}$$

y haciendo $u = \lambda x$ *tenemos que* $du = \lambda dx$, *por lo que*

$$\begin{aligned} E\left[X^2\right] &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{1}{\lambda} \int_0^\infty e^{-u} \left(\frac{u}{\lambda}\right)^{\alpha+1} du \\ &= \frac{1}{\Gamma(\alpha)\lambda^2} \int_0^\infty e^{-u} u^{\alpha+1} du \\ &= \frac{1}{\Gamma(\alpha)\lambda^2} \int_0^\infty e^{-u} u^{(\alpha+2)-1} du \\ &= \frac{1}{\Gamma(\alpha)\lambda^2} \Gamma(\alpha+2) \\ &= \frac{(\alpha+1)\alpha}{\lambda^2}, \end{aligned}$$

por lo tanto

$$\begin{aligned} \mathrm{Var}(X) &= \frac{(\alpha+1)\alpha}{\lambda^2} - \frac{\alpha^2}{\lambda^2} \\ &= \frac{\alpha^2+\alpha-\alpha^2}{\lambda^2} \\ &= \frac{\alpha}{\lambda^2}. \end{aligned}$$

Es importante mencionar que la esperanza nos es de gran utilidad cuando queremos establecer la relación existente entre dos variables aleatorias, ya que si X y Y son variables aleatorias independientes, entonces podemos enunciar la siguiente proposición:

Proposición 5.5 *Si X y Y son variables aleatorias independientes, entonces para cualesquiera h y g funciones de $\mathcal{R}$ en $\mathcal{R}$*

$$E[g(X)h(Y)] = E[g(X)]\,E[h(Y)].$$

<u>Dem.</u>

Si suponemos que X y Y son dos variables aleatorias absolutamente continuas con función de densidad $f_{X,Y}(\cdot,\cdot)$, entonces

$$\begin{aligned}
E[g(X)h(Y)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f_{X,Y}(x,y)dxdy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f_X(x)f_Y(y)dxdy \\
&= \int_{-\infty}^{\infty} g(x)f_X(x)dx \int_{-\infty}^{\infty} h(y)f_Y(y)dy \\
&= E[g(X)]\,E[h(Y)].
\end{aligned}$$

Definición 5.4 *Sean X y Y variables aleatorias independientes, entonces la covarianza entre X y Y se denotará por $\mathrm{Cov}(X,Y)$ y está definida como sigue*

$$\mathrm{Cov}(X,Y) = E[(X - E[X])(Y - E[Y])].$$

Se puede ver de manera muy fácil que

$$\mathrm{Cov}(X,Y) = E[XY] - E[X]\,E[Y].$$

Por esta igualdad y por la proposición 5.5 podemos enunciar el siguiente corolario:

Corolario 5.3 *Si X y Y son variables aleatorias independientes, entonces*

$$\mathrm{Cov}(X,Y) = 0.$$

El recíproco no es cierto, ya que podemos dar un ejemplo sencillo o un poco más elaborada, en donde no se cumpla el recíproco.

Ejemplo 5.9 *Si X es una variable aleatoria beta con parámetros a y b. Calcular $E[X^n]$. En particular si tenemos que $a = b = \frac{1}{2}$ y $Y = 4X(1-X)$. Calcular la covarianza entre X y Y. ¿Son X y Y independientes?*

<u>Sol.</u>

Primeramente calculemos la $E[X^n]$, por lo que tenemos que

$$\begin{aligned} E[X^n] &= \tfrac{1}{B(a,b)}\int_0^1 x^n x^{a-1}(1-x)^{b-1}dx \\ &= \tfrac{1}{B(a,b)}\int_0^1 x^{n+a-1}(1-x)^{b-1}dx \\ &= \tfrac{1}{B(a,b)}B(n+a,b) \\ &= \tfrac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\tfrac{\Gamma(n+a)\Gamma(b)}{\Gamma(n+a+b)} \\ &= \tfrac{\Gamma(a+b)\Gamma(n+a)}{\Gamma(a)\Gamma(n+a+b)}. \end{aligned}$$

Se puede demostrar que Y se distribuye beta con parámetros $a = b = \frac{1}{2}$. Por lo cual tenemos que

$$\mathrm{Cov}(X,Y) = E[XY] - E[X]E[Y]$$

de aquí tenemos que basta calcular la esperanza de X y la esperanza de XY, por lo cual tenemos que

$$\begin{aligned} E[X] &= \tfrac{\Gamma(1)\Gamma(\frac{3}{2})}{\Gamma(\frac{1}{2})\Gamma(2)} \\ &= \tfrac{\frac{1}{2}\Gamma(\frac{1}{2})}{\Gamma(\frac{1}{2})} \\ &= \tfrac{1}{2}. \end{aligned}$$

Además tenemos que

$$\begin{aligned} E[XY] &= E[X(4X(1-X))] \\ &= E[4X^2 - 4X^3] \\ &= 4E[X^2] - 4E[X^3]. \end{aligned}$$

De aquí dado que

$$\begin{aligned} E[X^2] &= \tfrac{\Gamma(1)\Gamma(\frac{5}{2})}{\Gamma(\frac{1}{2})\Gamma(3)} \\ &= \tfrac{\frac{3}{2}\frac{1}{2}\Gamma(\frac{1}{2})}{\Gamma(\frac{1}{2})\Gamma(3)} \\ &= \tfrac{3}{8}. \end{aligned}$$

De manera similar tenemos que

$$\begin{aligned} E[X^3] &= \tfrac{\Gamma(1)\Gamma(\frac{7}{2})}{\Gamma(\frac{1}{2})\Gamma(4)} \\ &= \tfrac{\frac{5}{2}\frac{3}{2}\frac{1}{2}}{3!} \\ &= \tfrac{5}{16}. \end{aligned}$$

Por lo cual

$$\begin{aligned} E[XY] &= 4E[X^2] - 4E[X^3] \\ &= \tfrac{12}{8} - \tfrac{20}{16} \\ &= \tfrac{1}{4}. \end{aligned}$$

Por lo cual tenemos que

$$\begin{aligned}\text{Cov}(X,Y) &= E[XY] - E[X]E[Y] \\ &= \tfrac{1}{4} - \tfrac{1}{2}\tfrac{1}{2} \\ &= 0.\end{aligned}$$

Además tenemos que X y Y no son variables aleatorias independientes ya que podemos ver que dado un valor de X, el valor de Y se encuentra completamente determinado.

En general, se puede mostrar que si la $\text{Cov}(X,Y)$ es positiva, entonces si X crece, Y también lo hará. En cambio si la $\text{Cov}(X,Y)$ es negativa, entonces tenemos que si X crece, Y tenderá a decrecer.

La covarianza nos ayuda también a calcular varianzas de sumas de variables aleatorias.

Proposición 5.6 *Si X y Y son variables aleatorias,entonces*

$$\text{Var}(X \pm Y) = \text{Var}(X) + \text{Var}(Y) \pm 2\,\text{Cov}(X,Y).$$

<u>*Dem.*</u>

$$\begin{aligned}\text{Var}(X \pm Y) &= E\left[(X \pm Y - E[X \pm Y])^2\right] \\ &= E\left[(X - E[X] \pm (Y - E[Y]))^2\right] \\ &= E[(X - E[X])^2 \\ &\quad \pm 2\,(X - E[X])\,(Y - E[Y]) \\ &\quad + (Y - E[Y])^2] \\ &= E[(X - E[X])^2] \\ &\quad \pm 2E\,[(X - E[X])\,(Y - E[Y])] \\ &\quad + E\left[(Y - E[Y])^2\right] \\ &= \text{Var}(X) + \text{Var}(Y) \pm 2\,\text{Cov}(X,Y).\end{aligned}$$

Podemos demostrar en general que si $X_1, X_2, \ldots, X_n$ son variables aleatorias, entonces

$$\text{Var}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \text{Var}(X_i) + 2\sum_{i<j} \text{Cov}(X_i, X_j).$$

De esta fórmula se puede ver que la varianza no es un operador lineal como lo es la esperanza, también se puede ver por el corolario 5.3 que si $X_1, X_2, \ldots, X_n$ son variables aleatorias independientes, entonces

$$\operatorname{Var}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \operatorname{Var}(X_i).$$

Por último, podemos enunciar la siguiente proposición

Proposición 5.7 *Si X es una variable aleatoria y $a, b \in \mathcal{R}$, entonces*

$$\operatorname{Var}(aX + b) = a^2 \operatorname{Var}(X).$$

Dem.

Si X es una variable aleatoria absolutamente continua, tenemos que

$$\begin{aligned}
\operatorname{Var}(aX + b) &= E\left[(aX + b)^2\right] - E^2[aX + b] \\
&= E\left[a^2X^2 + 2abX + b^2\right] \\
&\quad - (aE[X] + b)^2 \\
&= a^2E\left[X^2\right] + 2abE[X] + b^2 \\
&\quad - (a^2E^2[X] + 2abE[X] + b^2) \\
&= a^2\left(E\left[X^2\right] - E^2[X]\right) \\
&= a^2 \operatorname{Var}(X).
\end{aligned}$$

Ejemplo 5.10 *Si X es una variable aleatoria con esperanza μ y varianza σ^2 ($\sigma^2 > 0$), entonces a la variable $Z = \frac{X-\mu}{\sigma}$ se le conoce como la normalización de la variable aleatoria X. Probar que $E[Z] = 0$ y $\operatorname{Var}(Z) = 1$.*

Sol.

Dadas todas las propiedades anteriormente vistas de la esperanza y la varianza, tenemos que

$$\begin{aligned}
E[Z] &= E\left[\tfrac{X-\mu}{\sigma}\right] \\
&= E\left[\tfrac{X}{\sigma} - \tfrac{\mu}{\sigma}\right] \\
&= E\left[\tfrac{X}{\sigma}\right] - E\left[\tfrac{\mu}{\sigma}\right] \\
&= \tfrac{1}{\sigma}E[X] - \tfrac{1}{\sigma}E[\mu] \\
&= \tfrac{1}{\sigma}E[\mu] - \tfrac{1}{\sigma}E[X] \\
&= 0.
\end{aligned}$$

Además, tenemos que

$$\begin{aligned}
\mathrm{Var}(Z) &= \mathrm{Var}\left(\frac{X-\mu}{\sigma}\right) \\
&= \frac{1}{\sigma^2}\mathrm{Var}(X-\mu) \\
&= \frac{1}{\sigma^2}\left(E\left[(X-\mu)^2\right] - E^2[X-\mu]\right) \\
&= \frac{1}{\sigma^2}\mathrm{Var}(X) \\
&= \frac{\sigma^2}{\sigma^2} \\
&= 1,
\end{aligned}$$

ya que $E^2[X-\mu] = 0$

La siguiente proposición es una caracterización de las variables aleatorias que toman un solo valor real con probabilidad 1. A estas variables aleatorias también se les conoce como variables aleatorias degeneradas.

Proposición 5.8 *X es una variable aleatoria degenerada en $a \in \mathcal{R}$, (i.e. $P\{X=a\}=1$) si y sólo sí*

$$\mathrm{Var}(X) = 0.$$

Dem.

Supongamos que X es una variable aleatoria degenerada en a. Por lo que se puede ver fácilmente que

$$E[X] = a,$$

ya que X es una variable aleatoria discreta, de esta manera, tenemos que

$$\begin{aligned}
\mathrm{Var}(X) &= E\left[(X-a)^2\right] \\
&= \textstyle\sum_x (x-a)^2 f_X(x) \\
&= 0.
\end{aligned}$$

dado que X toma únicamente el valor a.

Ahora si la $\mathrm{Var}(X) = 0$, *el problema es más complicado pero se resuelve de una manera muy sencilla utilizando teoría de la medida, por lo que aquí no lo demostraremos.*

Una medida más exacta del la relación existente entre X y Y la podemos obtener conociendo el coeficiente de correlación existente entre X y Y, el cual se define como sigue:

Definición 5.5 *Si X y Y son variables aleatorias con varianzas positivas σ_X^2 y σ_Y^2 respectivamente, entonces denotaremos como $\rho_{X,Y}$ al coeficiente de correlación entre X y Y, el cual se define como*

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}.$$

El coeficiente de correlación nos da una medida del grado de relación lineal que existe entre X y Y, ya que podemos enunciar el siguiente teorema:

Teorema 5.4 *Para cada par de variables aleatorias X y Y para las cuales $\rho_{X,Y}$ exista tenemos que*

$$|\rho_{X,Y}| \leq 1,$$

y además tenemos que

$$|\rho_{X,Y}| = 1$$

si y sólo sí existen números reales $a \neq 0$ y b tales que

$$P\{Y = aX + b\} = 1,$$

también podemos ver que $a > 0$ si $\rho_{X,Y} = 1$ y $a < 0$ si $\rho_{X,Y} = -1$.

Dem.

Primero probemos que

$$|\rho_{X,Y}| \leq 1,$$

para ello vemos que

$$\begin{aligned}
0 &\leq \mathrm{Var}(\tfrac{X}{\sigma_X} + \tfrac{Y}{\sigma_Y}) \\
&= \mathrm{Var}(\tfrac{X}{\sigma_X}) + \mathrm{Var}(\tfrac{Y}{\sigma_Y}) \\
&\quad +2\,\mathrm{Cov}(\tfrac{X}{\sigma_X}, \tfrac{Y}{\sigma_Y}) \\
&= \tfrac{1}{\sigma_X^2}\mathrm{Var}(X) + \tfrac{1}{\sigma_Y^2}\mathrm{Var}(Y) \\
&\quad +2\left(E\left[\tfrac{X}{\sigma_X}\tfrac{Y}{\sigma_Y}\right] - E\left[\tfrac{X}{\sigma_X}\right]E\left[\tfrac{Y}{\sigma_Y}\right]\right) \\
&= 2 + 2\tfrac{1}{\sigma_X\sigma_Y}\left(E[XY] - E[X]E[Y]\right) \\
&= 2(1 + \rho_{X,Y}),
\end{aligned}$$

de aquí tenemos que

$$\rho_{X,Y} \geq -1.$$

Por otro lado

$$\begin{aligned} 0 &\leq \operatorname{Var}(\tfrac{X}{\sigma_X} - \tfrac{Y}{\sigma_Y}) \\ &= \operatorname{Var}(\tfrac{X}{\sigma_X}) + \operatorname{Var}(\tfrac{Y}{\sigma_Y}) \\ &\quad -2\operatorname{Cov}(\tfrac{X}{\sigma_X}, \tfrac{Y}{\sigma_Y}) \\ &= 2(1-\rho_{X,Y}), \end{aligned}$$

por lo cual también se tiene que

$$\rho_{X,Y} \leq 1,$$

por lo que podemos concluir que

$$|\rho_{X,Y}| \leq 1.$$

Para la segunda parte de esta demostración, primero supondremos que

$$|\rho_{X,Y}| = 1,$$

si $\rho_{X,Y} = 1$, *tenemos que*

$$\operatorname{Var}(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}) = 0,$$

de aquí por la proposición 5.8 podemos ver que

$$P\left\{\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y} = k\right\} = 1$$

para alguna $k \in \mathcal{R}$, *por lo cual*

$$P\left\{Y = \frac{\sigma_Y}{\sigma_X}X - k\sigma_Y\right\} = 1$$

por lo que tenemos que $a = \frac{\sigma_Y}{\sigma_X} > 0$, *de manera análoga podemos ver que si* $\rho_{X,Y} = -1$, *entonces*

$$P\left\{Y = -\frac{\sigma_Y}{\sigma_X}X + k\sigma_Y\right\} = 1,$$

por lo cual tenemos que $a = -\frac{\sigma_Y}{\sigma_X}$.

Si suponemos que existen números reales $a \neq 0$ *y* b *tales que*

$$P\{Y = aX + b\} = 1,$$

entonces tenemos que

$$\operatorname{Var}(Y) = a^2 \operatorname{Var}(X),$$

de donde $\sigma_X = |a|\sigma_Y$ *además tenemos que*

$$\begin{aligned}
\mathrm{Cov}(X,Y) &= = E[XY] - E[X]E[Y] \\
&= E[X(aX+b)] \\
&\quad -E[X]E[(aX+b)] \\
&= aE[X^2] + bE[X] \\
&\quad -aE^2[X] - bE[X] \\
&= a\left(E[X^2] - E^2[X]\right) \\
&= a\,\mathrm{Var}(X),
\end{aligned}$$

por lo cual tenemos que

$$\begin{aligned}
\rho_{X,Y} &= \tfrac{a\sigma_X^2}{|a|\sigma_X^2} \\
&= \begin{cases} 1 & si\ a > 0 \\ -1 & si\ a < 0 \end{cases}.
\end{aligned}$$

Cuando $\rho_{X,Y} = 0$ se dice que X y Y no están correlacionadas (que es distinto a decir que X y Y son independientes), si $|\rho_{X,Y}| = 1$, entonces se dice que X y Y están perfectamente correlacionadas.

5.3 Momentos

En el ejemplo 5.9 pudimos ver que fue de gran ayuda el conocer la esperanza de X^k. En esta sección veremos que esta esperanza es conocida con el nombre de momento de $k-$*ésimo* orden alrededor del origen.

Definición 5.6 *Si X es una variable aleatoria, entonces a $E\left[(X-a)^k\right]$ se le conoce como el $k-$ésimo momento de X alrededor de a. Además, si $a = 0$ tenemos $E\left[X^k\right]$ que se le llama el $k-$ésimo momento alrededor del origen. También si $a = E[X]$, entonces $E\left[(X-E[X])^k\right]$ se le conoce como el $k-$ésimo momento central. A la $E\left[|X-a|^k\right]$ se le conoce por el $k-$ésimo momento absoluto de X de orden k alrededor de a.*

Es claro que el primer momento central es cero y el segundo momento la ya bien conocida por nosotros varianza. También podemos establecer una fácil relación entre los momentos centrales y los momentos alrededor del origen, ya que utilizando el hecho de que

$$(a-b)^n = \sum_{i=0}^{n} \binom{n}{i} (-1)^i a^{n-i} b^i,$$

y que la esperanza es un operador lineal tenemos que para todo $k \in \mathcal{N}$

$$\begin{aligned} E\left[(X - E[X])^k\right] &= E\left[\sum_{i=0}^{k} \binom{n}{i}(-1)^i X^{k-i} E^i[X]\right] \\ &= \sum_{i=0}^{k} \binom{n}{i}(-1)^i E^i[X]\, E\left[X^{k-i}\right]. \end{aligned}$$

Si denotamos por μ_k al $k-$*ésimo* momento alrededor del origen y como $\mu_{(k)}$ al $k-$*ésimo* momento central, entonces tenemos que

$$\begin{aligned} \mu_{(k)} &= \sum_{i=0}^{k} \binom{k}{i}(-1)^i (\mu_1)^i \mu_{k-i} \\ &= \sum_{i=0}^{k-2} \binom{k}{i}(-1)^i (\mu_1)^i \mu_{k-i} \\ &\quad + k(-1)^{k-1} (\mu_1)^{k-1} \mu_1 \\ &\quad + (-1)^k (\mu_1)^k \\ &= \sum_{i=0}^{k-2} \binom{k}{i}(-1)^i (\mu_1)^i \mu_{k-i} \\ &\quad + (-1)^{k-1} (\mu_1)^k (k-1), \end{aligned}$$

por lo que tenemos que

$$\begin{aligned} \mu_{(k)} &= \sum_{i=0}^{k-2} \binom{k}{i}(-1)^i (\mu_1)^i \mu_{k-i} \\ &\quad + (-1)^{k-1} (\mu_1)^k (k-1), \end{aligned} \tag{5.1}$$

de donde vemos que

$$\begin{aligned} \mu_{(1)} &= 0 \\ \mu_{(2)} &= \mu_2 - \mu_1^2 \\ \mu_{(3)} &= \mu_3 - 3\mu_2\mu_1 + 2\mu_1^3 \\ \mu_{(4)} &= \mu_4 - 4\mu_3\mu_1 + 6\mu_2\mu_1^2 - 3\mu_1^4. \end{aligned}$$

Así, podemos ver por 5.1 que si los primeros k momentos alrededor del origen existen, entonces el $k-$*ésimo* momento central también existe, por lo que a continuación enunciaremos un teorema que nos dice cuando los momentos alrededor del origen existen.

Teorema 5.5 *Si X es una variable aleatoria absolutamente continua, y μ_k existe ($k \in \mathcal{N}$, entonces tenemos que μ_j existe para $j = 1, 2, \ldots, k-1$.*

Dem.

Ya que el $k-$ésimo momento existe tenemos que

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

además podemos ver para todo $n \in \mathcal{N}$ que

$$\begin{aligned}\int_{-\infty}^{\infty} |x|^n f_X(x)dx &= \int_{|x|\le 1} |x|^n f_X(x)dx \\ &\quad + \int_{|x|>1} |x|^n f_X(x)dx \\ &\le \int_{|x|\le 1} f_X(x)dx \\ &\quad + \int_{|x|>1} |x|^n f_X(x)dx \\ &\le 1 + \int_{|x|>1} |x|^n f_X(x)dx,\end{aligned}$$

ya que $1 \le j < k$ tenemos que

$$\int_{|x|>1} |x|^j f_X(x)dx \le \int_{|x|>1} |x|^k f_X(x)dx,$$

y dado que

$$\begin{aligned}\int_{-\infty}^{\infty} |x|^j f_X(x)dx &\le 1 + \int_{|x|>1} |x|^j f_X(x)dx \\ &\le 1 + \int_{|x|>1} |x|^k f_X(x)dx \\ &< \infty.\end{aligned}$$

5.4 Ejercicios

1. Un hombre con un arco dispara a un blanco, si su flecha está a lo más 5 cm. del blanco, entonces recibe 10 puntos, si su flecha está entre 5 y 15 cm. del blanco, recibe 3 puntos, en caso contrario éste no recibe puntos. Encontrar el número esperado de puntos que obtendrá, si

   (a) Su disparos caen uniformemente distribuidos en un círculo de 30 cm. de radio alrededor del blanco.

   (b) La distancia tanto vertical como horizontal al blanco son variables aleatorias normales con $\mu = 0$ y $\sigma^2 = 36$.

2. Una caja contiene 10 focos, de los cuales tres son defectuosos. Si se sacan focos aleatoriamente y se prueban, hasta encontrar un foco defectuoso, encontrar la esperanza del número de focos que tenemos que probar, antes de encontrar un foco defectuoso.

3. Sea Z una variable aleatoria discreta que toma los valores 0,1,2 y 3 y tiene una función de densidad dada por

$$\begin{aligned}f(0) &= \frac{1}{4}, \quad f(2) = \frac{1}{8}, \\ f(1) &= \frac{1}{2}, \quad f(4) = \frac{1}{8}.\end{aligned}$$

Determinar la media y la varianza de Z.

4. Suponga que X es una variable aleatoria que tiene una función de densidad dada por

$$f(x) = \begin{cases} Rx^{R-1} & \text{si } 0 \leq x \leq 1 \\ 0 & \text{c.o.c.} \end{cases},$$

donde $R > 0$ es una constante. Determine la esperanza y la varianza de X.

5. Una variable aleatoria X tiene una función de distribución dada por

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 - (1-x)^A & \text{si } 0 \leq x \leq 1 \\ 1 & \text{si } x > 1 \end{cases},$$

donde $A > 0$ es una constante. Determine la esperanza y la varianza de X.

6. Una compañia de seguros suscribe una póliza en la cual se establece que una cantidad de dinero A debe ser pagada si algún evento E ocurre dentro de un año. Si la compañia estima que E ocurrirá con una probabilidad p, ¿cuál podría ser la cuota que tenga que pagar el cliente, para que la compañia obtenga una ganancia esperada del 10 por ciento sobre A?

7. La función de densidad de X esta dada por

$$f(x) = \begin{cases} a + bx & \text{si } 0 \leq x \leq 1 \\ 0 & \text{c.o.c.} \end{cases}.$$

Si $E[X] = \frac{3}{5}$, encontrar a, b.

8. Un vendedor de periódicos compra los periódicos a N\$2.00 y los vende a N\$3.00. Sin embargo el no puede regresar los periódicos que no venda. Si la demanda de su periódico es una variable aleatoria binomial con $n = 300$, $p = \frac{1}{3}$ aproximadamente, ¿cuántos periódicos podría el comprar para maximizar su ganancia esperada?

9. La función de densidad de una variable aleatoria X es

$$f_X(x) = \frac{1}{2a} e^{-\frac{|x-a|}{a}}.$$

X se dice que tiene una distribución Laplace. Encontrar la esperanza y la varianza de X.

10. Determine la esperanza y la varianza de una variable aleatoria X con función de densidad dada por
$$f(x) = \begin{cases} x & \text{si } 0 \leq x < 1 \\ 2 - x & \text{si } 1 \leq x \leq 2 \\ 0 & \text{e.o.c.} \end{cases}.$$

11. Sea X una variable aleatoria distribuida distribuida de acuerdo a una ley logaritmica normal, es decir tiene una función de densidad dada por
$$f(x) = \begin{cases} \frac{1}{\beta x \sqrt{2\pi}} e^{-\frac{1}{2\beta^2}(\log x - \alpha)^2} & \text{si } x > 0 \\ 0 & \text{e.o.c.} \end{cases},$$
con $\beta > 0$ y α un cualquier número real. Encontrar la esperanza y la varianza de X.

12. La magnitud en la velocidad de una molécula se puede modelar por medio de una variable aleatoria X cuya función de densidad está dada por
$$f_X(x) = \begin{cases} \frac{4x^2}{\alpha^3 \sqrt{\pi}} e^{-\frac{x^2}{\alpha^2}} & x > 0 \\ 0 & \text{e.o.c.} \end{cases},$$
donde α es una constante positiva. Se dice que X tiene una distribución de Maxwell. Encontrar la velocidad promedio de la molécula, y la varianza de ésta.

13. Sea X una variable aleatoria distribuida normal con $\mu = 0$. Encontrar $E[\cos X]$. Use ésto para encontrar $E[\cos Y]$ donde Y se distribuye normal con media μ y varianza σ^2.

14. Probar que si $P\{X \leq 0\} = 1$ entonces $E[X] \leq 0$.

15. Probar que si $P\{X \geq 0\} = 1$ entonces $E[X] \geq 0$.

16. Probar que si X y Y son dos variables aleatorias tales que $P\{X \leq Y\} = 1$ entonces $E[X] \leq E[Y]$.

17. Probar que si $a < b$ son dos números reales y $P\{a \leq X \leq b\} = 1$, entonces la esperanza de X existe y $a \leq E[X] \leq b$, también probar que la varianza existe y $\text{Var}[X] \leq \frac{(b-a)^2}{4}$.

18. La entropía $H(X)$ de una variable aleatoria X se define, si es que $f_X(x)$ existe ($f_X(x)$ es la función de densidad de la variable aleatoria X) como
$$H(X) = -E[\log f_X(X)],$$

ésto es la esperanza del valor negativo del logaritmo de la función de densidad evaluada en X (en algunos contextos se acostumbra tomar logaritmo base 2 en la definición de entropía). Encontrar la entropía de las siguientes variables:

(a) Si $P\{X=0\}=p$; $P\{X=1\}=1-p$.

(b) Si $P\{X=a\}=p$; $P\{X=b\}=1-p$.

(c) Si X tiene una distribución uniforme discreta.

(d) Si X tiene una distribución normal.

(e) Si X tiene una distribución exponencial.

(f) Si

$$f_X(x)=\frac{\alpha}{2}e^{-\alpha|x|}\,(\alpha>0).$$

19. Mostrar que

$$E[X]=\int_0^{\infty}\{1-F_X(x)\}\,dx-\int_{-\infty}^{0}F_X(x).$$

Dar una explicación intuitiva acerca de esta igualdad.

20. Si X tiene una distribución normal con parámetros μ y σ^2. Mostrar que $E[X]=\mu$ y $\operatorname{Var}X=\sigma^2$. Si $\mu=0$ encontrar una fórmula para el $k-$*ésimo* momento alrededor del origen. A partir de aquí encontrar el $k-$*ésimo* momento si X tiene parámetros μ y σ^2.

21. Si X tiene una distribución gamma con parámetros λ y α. Encontrar una fórmula para el $k-$*ésimo* momento alrededor del origen y una expresión para la varianza de X.

22. Si X se distribuye beta con parámetros a y b. ¿Para qué valores de n está definido el $n-$*ésimo* momento de X? Para esa n dar una expresión para éste y encontrar la varianza de X.

23. Sea X una variable aleatoria distribuida uniformemente en el intervalo $(0,1)$. Utilizando la ley del estadístico inconsciente, calcule $E[X^n]$ y después verifique el resultado utilizando la definición de esperanza.

24. Una variable aleatoria X se distribuye uniformemente sobre el intervalo $(0,b)$. Encontrar $E\left[|X-a|^k\right]$, donde $E[X]=a$.

25. Sea X una variable aleatoria con función de densidad dada por

$$f(x) = \begin{cases} \frac{ka^k}{(x+a)^{k+1}} & \text{si } x \geq 0 \\ 0 & e.o.c. \end{cases},$$

donde $a, k > 0$. Mostrar que $E\left[|X|^{\alpha}\right] < \infty$ para $\alpha < k$.

26. Sea X una variable aleatoria con función de densidad dada por

$$f(x) = \frac{c}{(1+x^2)^m},$$

donde $m \geq 1$ y $c = \Gamma(m)/\left[\Gamma(1/2)\Gamma(m-1/2)\right]$. Mostrar que $E\left[X^{2r}\right]$ existe si y sólo si $2r < 2m - 1$. ¿Cuál es $E\left[X^{2r}\right]$ si $2r < 2m - 1$?

27. La distribución Pareto con parámetros α y β ($\alpha, \beta > 0$) es definida por una función de densidad dada por

$$f(x) = \begin{cases} \frac{\beta\alpha^{\beta}}{x^{\beta+1}} & \text{si } x \geq \alpha \\ 0 & \text{e.o.c.} \end{cases}.$$

Mostrar que el momento de orden n existe si y sólo sí $n < \beta$. Sea $\beta > 2$. Encontrar la media y la varianza de esta distribución.

28. Para una variable aleatoria X con función de densidad dada por

$$f(x) = \begin{cases} \frac{1}{2}x & \text{si } 0 \leq x < 1 \\ \frac{1}{2} & \text{si } 1 \leq x < 2 \\ \frac{1}{2}(3-x) & \text{si } 2 \leq x < 3 \\ 0 & \text{e.o.c.} \end{cases},$$

mostrar todos sus momentos existen. Encontrar la media y la varianza de X.

29. Para una variable aleatoria X Poisson con función de densidad dada por

$$\begin{aligned} f(x) &= P\{X = x\} \\ &= \begin{cases} e^{-\lambda}\frac{\lambda^x}{x!} & \text{si } x = 0, 1, \ldots \\ 0 & \text{e.o.c.} \end{cases}. \end{aligned}$$

Mostrar que $E[X] = \lambda$, $E\left[X^2\right] = \lambda + \lambda^2$, $E\left[X^3\right] = \lambda + 3\lambda^2 + \lambda^3$.

30. Para cada una de las siguientes funciones de densidad, encontrar $E[X]$, $E[Y]$, σ_X^2, σ_Y^2, σ_X, σ_Y.

(a) $f(x,y) = \begin{cases} e^{-x-y} & x,y > 0 \\ 0 & \text{e.o.c.} \end{cases}$.

(b) $f(x,y) = \begin{cases} 2e^{-x-y} & 0 < x < y < \infty \\ 0 & \text{e.o.c.} \end{cases}$.

(c) $f(x,y) = \begin{cases} \frac{6-x-y}{8} & 0 < x < 2; 2 < y < 4 \\ 0 & \text{e.o.c.} \end{cases}$.

(d) $f(x,y) = \begin{cases} 4xy & 0 < x, y < 1 \\ 0 & \text{e.o.c.} \end{cases}$.

(e) $f(x,y) = \begin{cases} \frac{1}{8}(x^2 - y^2)e^{-x} & 0 < x < \infty; |y| < x \\ 0 & \text{e.o.c.} \end{cases}$.

(f) $f(x,y) = \begin{cases} 4x(1-y) & 0 < x, y < 1 \\ 0 & \text{e.o.c.} \end{cases}$.

(g) $f(x,y) = \begin{cases} x + y & 0 < x, y < 1 \\ 0 & \text{e.o.c.} \end{cases}$.

(h) $f(x,y) = \begin{cases} 3x & 0 < y < x < 1 \\ 0 & \text{e.o.c.} \end{cases}$.

(i) $f(x,y) = \begin{cases} 2 & 0 < x < y < 1 \\ 0 & \text{e.o.c.} \end{cases}$.

31. Si $X_1, X_2, \ldots, X_5$ son variables aleatorias con $E[X_i] = 3$, $\sigma^2_{X_i} = 5, \sigma_{X_i X_j} = 2$ $(i \neq j)$. Entonces para $Y = \sum_{i=1}^{5} X_i$, encontrar $E[X]$, σ^2_Y, σ_{YX_i}.

32. Sea $U = \sum_{i=1}^{n} \alpha_i X_i$, $V = \sum_{j=1}^{n} \beta_j X_j$. Encontrar ρ_{UV} en términos de α_i, β_j y σ_{ij} $(1 \leq i, j \leq n)$. Use ésto para encontrar

   (a) $U = X_1 + X_2$, $V = X_1 - X_2$,

   (b) U, V como en 32a pero $\sigma_1^2 = \sigma_2^2$, $\sigma_{12} = 0$,

   (c) $U = X_1 + X_3, V = X_2 + X_3$,

   (d) U, V como en 32c pero $\sigma_1^2 = \sigma_2^2 = \sigma_3^2$, $\sigma_{12} = \sigma_{23} = \sigma_{13} = 0$,

   (e) $U = \sum_{i=1}^{m} X_i$, $V = \sum_{i=k+1}^{n} X_i (k \leq m < n)$
   $\sigma_{ij} = 0$, $\sigma_j^2 = \sigma^2 (j = 1, 2, \ldots, n)$,

(f) $U = a_1X_1 + a_2X_2$, $V = b_1X_1 + b_2X_2$; elija b_1, b_2 tales que $\rho_{UV} = 0$.

33. Sean X y Y variables aleatorias independientes con esperanza y varianza finita. Mostrar que para $U = XY$ y $V = X/Y$,

$$\sigma_U^2 = \sigma_X^2\sigma_Y^2 + \mu_X^2\sigma_Y^2 + \mu_Y^2\sigma_X^2,$$
$$\sigma_V^2 = \frac{\mu_Y^2\sigma_X^2 - \mu_X^2\sigma_Y^2}{\mu_Y^2(\sigma_Y^2 + \mu_Y^2)}, \text{si } E\left[Y^2\right] \neq 0, E\left[Y \neq 0\right].$$

34. Análogamente al ejercicio 18, la entropía para un vector aleatoria $(X_1, X_2, \ldots, X_n)$, si es que existe su función de densidad conjunta $f(x_1, x_2, \ldots, x_n)$, se define por

$$H(X_1, X_2, \ldots, X_n) = -E\left[\log f(X_1, X_2, \ldots, X_n)\right].$$

Mostrar que si $X_1, X_2, \ldots, X_n$ son variables aleatorias mutuamente independientes, entonces

$$H(X_1, X_2, \ldots, X_n) = \sum_{i=1}^{n} H(X_i),$$

ésto es que la entropía de variables aleatorias independientes es la suma de sus respectivas entropias.

Capítulo 6

Distribuciones multivariadas particulares

6.1 Distribución normal multivariada

Esta distribución es de gran utilidad cuando tenemos muchas variables aleatorias normales las cuales no son independientes. Es claro que cuando éstas son independientes, su función de densidad conjunta es tan solo el producto de sus funciones de densidad marginales, lo que no sucede cuando no son independientes. Para poder estudiar esta distribución conviene que el lector tenga un conocimiento básico de la teoría de matrices, ya que la utilizaremos para dar una definición adecuada de la distribución normal multivariada. Para lo cual enunciaremos la siguiente definición:

Definición 0.1 *Las n variables aleatorias X_1, $X_2, \ldots$, X_n se dice que tienen una distribución normal multivariada si existen n variables aleatorias independientes Z_1, Z_2 $, \ldots,$ Z_n cada una con una distribución normal estándar, n constantes μ_1, $\mu_2, \ldots$, μ_n y una matriz $A = (a_{ij})$ no-singular de $n \times n$, tales que*

$$\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \ddots & a_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{pmatrix} + \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix}.$$

Esta última ecuación matricial la podemos escribir como sigue:

$$\mathbf{X} = A\mathbf{Z} + \mu,$$

donde $\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}$, $A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a22 & \ddots & a_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$, $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix}$ *y* $\mathbf{Z} = \begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{pmatrix}$.

De aquí tenemos que $\mathbf{X}$ se obtiene de $\mathbf{Z}$ aplicando una transformación de $\mathcal{R}^n$ en $\mathcal{R}^n$ tal que

$$\begin{aligned} \mathrm{x} &= G(\mathrm{z}) \\ &= A\mathrm{z} + \mu, \end{aligned}$$

la cual es una transformación uno a uno ya que A es una matriz no-singular, por lo cual existe una transformación inversa

$$\begin{aligned} G^{-1}(\mathrm{x}) &= A^{-1}(\mathrm{x} - \mu) \\ &= \mathrm{z}, \end{aligned}$$

la cual también es una transformación uno a uno de $\mathcal{R}^n$ en $\mathcal{R}^n$. Es fácil ver que la función de densidad conjunta de $\mathbf{Z}$ es

$$f_{\mathbf{Z}^t}(\mathrm{z}^t) = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}\mathrm{z}^t\mathrm{z}}.$$

Si $A^{-1} = (b_{ij})$, se puede ver fácilmente que $\frac{\partial z_i}{\partial x_j} = b_{ij}$ para $i, j = 1, 2, \ldots, n$, por lo que el jacobiano de la transformación inversa es $J = |A^{-1}|$. Por consiguiente recordando el corolario 3.9 tenemos:

$$\begin{aligned} f_{\mathbf{X}^t}(\mathrm{x}^t) &= f_{\mathbf{Z}^t}(\mathrm{z}^t)\left|\left|A^{-1}\right|\right| \\ &= \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}\mathrm{z}^t\mathrm{z}} \left|\left|A^{-1}\right|\right| \\ &= \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}(\mathrm{x}-\mu)^t(A^{-1})^t A^{-1}(\mathrm{x}-\mu)} \left|\left|A^{-1}\right|\right|. \end{aligned}$$

Ahora si $C = AA^t$ se tiene que

$$\begin{aligned} C^{-1} &= (AA^t)^{-1} \\ &= (A^t)^{-1}A^{-1}, \end{aligned}$$

y además tenemos que

$$\begin{aligned} \left|C^{-1}\right| &= \left|(A^t)^{-1}A^{-1}\right| \\ &= \left|(A^t)^{-1}\right|\left|A^{-1}\right| \\ &= \left|A^{-1}\right|^2, \end{aligned}$$

de donde

$$\left\|A^{-1}\right\| = \sqrt{|C^{-1}|},$$

por lo que tenemos que

$$f_{\mathbf{X}^t}(\mathbf{x}^t) = \frac{\sqrt{|C^{-1}|}}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}(\mathbf{x}-\mu)^t C^{-1}(\mathbf{x}-\mu)},$$

la cual es la función de densidad conjunta de una distribución normal multivariada.

A continuación estudiaremos las matrices C ya que ésta establece las relaciones de dependencia entre las variables aleatorias $X_1, X_2, \ldots, X_n$. Para ésto primeramente enunciaremos el siguiente lema:

Lema 6.1 *La matriz C^{-1} es una matriz simétrica y definida positiva.*

Dem.

Esta es simétrica ya que

$$\begin{aligned} (C^{-1})^t &= ((AA^t)^{-1})^t \\ &= ((AA^t)^t)^{-1} \\ &= (A^tA)^{-1} \\ &= C^{-1}, \end{aligned}$$

y también es definida positiva ya que para todo vector $\mathbf{x} \neq \overline{0}$, $\mathbf{x} \in \mathcal{R}^n$ se tiene que

$$\begin{aligned} \mathbf{x}^t C^{-1}\mathbf{x} &= \mathbf{x}^t(AA^t)^{-1}\mathbf{x} \\ &= \mathbf{x}^t(A^t)^{-1}A^{-1}\mathbf{x} \\ &= \mathbf{x}^t(A^{-1})^t A^{-1}\mathbf{x} \\ &= (A^{-1}\mathbf{x})^t(A^{-1}\mathbf{x}) \\ &> 0, \end{aligned}$$

puesto que A^{-1} es no-singular.

Definición 6.2 *Si $\mathbf{U} = (U_{ij})$ es una matriz de variables aleatorias entonces $E[\mathbf{U}]$ denotará la matriz de esperanzas $(E(U_{ij}))$. También si $G(x) = (g_{ij}(x))$ es una matriz de funciones definidas sobre algún intervalo $[a, b]$, entonces la integral $\int_a^b G(x)dx$ denotará la matriz de integrales $\left(\int_a^b g_{ij}(x)dx\right)$. Por último $d\mathbf{x}^t$ denotará a $dx_1 dx_2 \ldots dx_n$.*

De estas definiciones, se puede ver fácilmente que $E[\mathbf{X}] = \boldsymbol{\mu}$, i.e., $E[X_i] = \mu_i$ para $i = 1, 2, \ldots, n$. Sea C_0 que denota la matriz de covarianzas de X, i.e.,

$$\begin{aligned} C_0 &= (Cov(X_i, X_j)) \\ &= (E[(X_i - \mu_i)(X_j - \mu_j)]) \\ &= (E[X_i X_j] - \mu_i \mu_j), \end{aligned}$$

la cual es una matriz de números reales. Podemos ver también que

$$C_0 = E\left[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^t\right].$$

De aquí tenemos que

$$\begin{aligned} C_0 &= E\left[(\mathbf{X} - \mu)(\mathbf{X} - \mu)^t\right] \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (\mathbf{x} - \mu)(\mathbf{x} - \mu)^t f_{\mathbf{X}^t}(\mathbf{x}^t) d\mathbf{x}^t \\ &= \frac{\sqrt{|C^{-1}|}}{(2\pi)^{\frac{n}{2}}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^t e^{-\frac{1}{2}(\mathbf{x}-\mu)^t C^{-1}(\mathbf{x}-\mu)} d\mathbf{x}^t. \end{aligned}$$

Ahora consideraremos la transformación

$$\mathbf{x} = A\mathbf{z} + \boldsymbol{\mu},$$

para la cual se tiene que su jacobiano es

$$\left|\frac{\partial \mathbf{x}^r}{\partial \mathbf{z}^r}\right| = \|A\| = \sqrt{|C|},$$

ya que es una transformación lineal, tenemos que

$$\left\|A^{-1}\right\| = \sqrt{|C^{-1}|},$$

por lo cual

$$\begin{aligned} C_0 &= \frac{\sqrt{|C^{-1}|}}{(2\pi)^{\frac{n}{2}}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \sqrt{|C|} A\mathbf{z} (A\mathbf{z})^t e^{-\frac{1}{2}(A\mathbf{z})^t C^{-1} A\mathbf{z}} d\mathbf{z}^t \\ &= \frac{\sqrt{|C^{-1}||C|}}{(2\pi)^{\frac{n}{2}}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} A\mathbf{z}\mathbf{z}^t A^t e^{-\frac{1}{2}\mathbf{z}^t A^t (A^t)^{-1} A^{-1} A\mathbf{z}} d\mathbf{z}^t \\ &= A\left[\frac{1}{(2\pi)^{\frac{n}{2}}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathbf{z}\mathbf{z}^t e^{-\frac{1}{2}\mathbf{z}^t \mathbf{z}} d\mathbf{z}^t\right] A^t, \end{aligned}$$

se tiene además que

$$\frac{1}{(2\pi)^{\frac{n}{2}}}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\mathbf{z}\mathbf{z}^t e^{-\frac{1}{2}\mathbf{z}^t\mathbf{z}}d\mathbf{z}^t$$

es la matriz de covarianzas de las variables aleatorias $Z_1, Z_2, \ldots, Z_n$, i.e., $C_1 = (Cov(z_i, z_j))$, se tiene que ésta es la matriz identidad ya que $Cov(Z_i, Z_j) = 0$ puesto que Z_i es independiente de Z_j si $i \neq j$ y $Cov(z_i, z_i) = 1$ dado que

$$\begin{aligned} Cov(Z_i, Z_i) &= \frac{1}{(2\pi)^{\frac{1}{2}}}\int_{-\infty}^{\infty} z^2 e^{-\frac{1}{2}z^2}dz \\ &= 1, \end{aligned}$$

de aquí tenemos:

$$\begin{aligned} C_0 &= A\left[\frac{1}{(2\pi)^{\frac{n}{2}}}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\mathbf{z}\mathbf{z}^t e^{-\frac{1}{2}\mathbf{z}^t\mathbf{z}}d\mathbf{z}^t\right]A^t \\ &= AIA^t \\ &= AA^t \\ &= C. \end{aligned}$$

De este resultado podemos enunciar el siguiente teorema.

Teorema 6.2 *Si $X_1, X_2, \ldots, X_n$ tienen una distribución normal multivariada, entonces su función de densidad conjunta está dada por*

$$f_{\mathbf{x}^t}(\mathbf{x}^t) = \frac{\sqrt{|C^{-1}|}}{(2\pi)^{\frac{n}{2}}}e^{-\frac{1}{2}(\mathbf{x}-\mu)^t C^{-1}(\mathbf{x}-\mu)},$$

donde $\mu = E[\mathbf{X}]$ y C es la matriz de covarianzas de $X_1, X_2, \ldots, X_n$.

Hasta aquí obtuvimos la función de densidad correspondiente a una distribución normal multivariada. Es importante observar que en la mayoría de los libros de texto la definición de la distribución normal multivariada es enunciada casi como el teorema 6.2 solo que $\frac{\sqrt{|C^{-1}|}}{(2\pi)^{\frac{n}{2}}}$ es definida únicamente como una constante K positiva, la matiz C^{-1} es considerada como una matriz A definida positiva y μ como un vector de números reales. A continuación enunciaremos un teorema que determina completamente a la función de densidad conjunta de una distribución normal multivariada , y el cual es el recíproco del teorema 6.2:

Teorema 6.3 *Si* $\mathbf{X}^t = (X_1, X_2, \ldots, X_n)$ *es un vector de variables aleatorias con una función de densidad conjunta dada por:*

$$f_{\mathbf{X}^t}(\mathbf{x}^t) = Ke^{-\frac{1}{2}(\mathbf{x}-\mu)^t A(\mathbf{x}-\mu)},$$

donde μ *es un vector de constantes,* $K > 0$ *una constante y* A *es una matriz simétrica definida positiva, entonces* $\mathbf{X}^t$ *tiene una distribución normal multivariada.*

Dem.

Recordando la teoría de matrices se tiene que si A *es una matriz simétrica definida positiva entonces existe una matriz* $A^{\frac{1}{2}}$ *de* $n \times n$ *tal que* $A^{\frac{1}{2}} A^{\frac{1}{2}} = A$. *Sea*

$$\mathbf{y} = A^{\frac{1}{2}}(\mathbf{x} - \mu)$$

una transformación de $\mathcal{R}^n$ *en* $\mathcal{R}^n$ *la cual se puede ver fácilmente que es uno a uno y que*

$$\mathbf{x} = (A^{\frac{1}{2}})^{-1}\mathbf{y} + \mu$$

es su transformación inversa. Sean $Y_1, Y_2, \ldots, Y_n$ *definidas como*

$$\begin{aligned} \mathbf{Y} &= (Y_1, Y_2, \ldots, Y_n) \\ &= A^{\frac{1}{2}}(\mathbf{X} - \mu). \end{aligned}$$

De aquí tenemos que

$$\mathbf{X} = (A^{\frac{1}{2}})^{-1}\mathbf{Y} + \mu,$$

donde $(A^{\frac{1}{2}})^{-1}$ es una matriz no singular y μ es un vector de números reales, por lo que recordando la definición 6.1 para mostrar que $\mathbf{X}$ tiene una distribución normal multivariada únicamente necesitamos demostrar que $Y_1, Y_2, \ldots, Y_n$ son variables aleatorias independientes cada una con una distribución normal estándar. Para lo cual calcularemos la función de densidad conjunta de Y como sigue:

$$\begin{aligned} f_{\mathbf{Y}^t}(\mathbf{y}^t) &= \frac{K}{\sqrt{|A|}} e^{-\frac{1}{2}\left[\left(A^{\frac{1}{2}}\right)^{-1}\mathbf{y}\right]^t A \left[\left(A^{\frac{1}{2}}\right)^{-1}\mathbf{y}\right]} \\ &= \frac{K}{\sqrt{|A|}} e^{-\frac{1}{2}\mathbf{y}^t\left[\left(A^{\frac{1}{2}}\right)^{-1}\right]^t A^{\frac{1}{2}} A^{\frac{1}{2}} \left(A^{\frac{1}{2}}\right)^{-1}\mathbf{y}} \\ &= \frac{K}{\sqrt{|A|}} e^{-\frac{1}{2}\mathbf{y}^t\left(A^{\frac{1}{2}}\right)^{-1} A^{\frac{1}{2}} A^{\frac{1}{2}} \left(A^{\frac{1}{2}}\right)^{-1}\mathbf{y}} \\ &= \frac{K}{\sqrt{|A|}} e^{-\frac{1}{2}\mathbf{y}^t\mathbf{y}}. \end{aligned}$$

Aprovechando el hecho que f_{Y^t} es una función de densidad conjunta tenemos que:

$$\frac{K}{\sqrt{|A|}}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}e^{-\frac{1}{2}\mathbf{y}^t\mathbf{y}}d\mathbf{y}^t = 1 = \frac{1}{(2\pi)^{\frac{n}{2}}}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}e^{-\frac{1}{2}\mathbf{y}^t\mathbf{y}}d\mathbf{y}^t,$$

por lo cual es claro que $K = \frac{\sqrt{|A|}}{(2\pi)^{\frac{n}{2}}}$ de aquí tenemos que $Y_1, Y_2, \ldots, Y_n$ son variables aleatorias independientes cada una con una distribución normal estándar, lo que establece el teorema.

A continuación enunciaremos varios teoremas relacionados con la distribución normal multivariada los cuales no demostraremos y que son de gran relevancia en la estadística matemática (ver [13], el cual dedica un capítulo completo para dar un pequeño resumen sobre la teoría de matrices y también un capítulo completo a la distribución normal multivariada).

Teorema 6.4 *Si las variables aleatorias $X_1, X_2, \ldots, X_n$ tienen una distribución normal multivariada, entonces cualquier subconjunto de ellas también tiene una distribución normal multivariada.*

Teorema 6.5 (Teorema de Cochran) *Si $\mathbf{X}^t = (X_1, X_2, \ldots, X_n)$ son n variables aleatorias independientes, cada una con una distribución normal estándar, si*

$$Q_1(\mathbf{x}^t), Q_2(\mathbf{x}^t), \ldots, Q_k(\mathbf{x}^t),$$

son k formas cuadráticas definidas sobre $\mathcal{R}^n$ tales que

$$x^t x = \sum_{j=1}^{k} Q_j(x^t)$$

para toda $x^t \in \mathcal{R}^n$, y $r_1 + r_2 + \ldots + r_k = n$ donde $r_j =$ rango $[Q_j(x^t)]$, entonces

$$Q_1(\mathbf{x}^t), Q_2(\mathbf{x}^t), \ldots, Q_k(\mathbf{x}^t)$$

son k variables aleatorias, y $Q_j(x^t)$ tiene una distribución χ^2 con r_j grados de libertad.

Este teorema es de suma importante para la estadística, en la parte conocida como análisis de varianza.

En las aplicaciones es muy importante hacer énfasis en el caso de la distribución normal bivariada y trivariada que a continuación definiremos, ya que éstas son usadas en una gran cantidad de aplicaciones. Una de sus más viejas aplicaciones está en el control del fuego de artillería, ya que el cálculo aproximado (en una artillería terrestre) de las desviaciones de un blanco en un plano se pueden modelar con una distribución normal bivariada, de manera análoga la distribución normal trivariada es útil para los blancos aéreos.

La distribución normal bivariada la podemos definir como sigue:

Definición 6.3 *Si X_1, X_2 tienen una función de densidad conjunta dada como sigue:*

$$\begin{aligned} f_{X_1,X_2}(x_1,x_2) = \; & \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2\sqrt{1-\rho^2}}\left\{\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2\right.\right. \\ & \left.\left. -2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)+\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right\}\right]. \end{aligned} \tag{6.1}$$

Donde $E[X_i] = \mu_i$, $Var(X_i) = \sigma_i^2$, ($i = 1, 2$) y la correlación entre X_1 y X_2 es ρ, entonces se dice que X_1 y X_2 tienen una distribución normal bivariada.

Se puede probar fácilmente que la distribución normal bivariada es también una distribución normal multivariada lo que significa que cumple con la definición .

La distribución normal bivariada también es llamada la distribución Gaussiana, distribución Laplace-Gauss o distribución Bravais.

En muchos casos es suficiente estudiar la distribución estandarizada obtenida cuando $\mu_1 = \mu_2 = 0$; $\sigma_1 = \sigma_2 = 1$ la cual tiene una función de densidad conjunta como sigue:

$$f_{X_1,X_2}(x_1,x_2) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2\sqrt{1-\rho^2}}\left({x_1}^2 - 2\rho x_1 x_2 + {x_2}^2\right)\right].$$

Si tenemos que en la distribución normal bivariada $\sigma_1 = \sigma_2$ y $\rho = 0$ entonces ésta es llamada una función de densidad circular normal, si $\sigma_1 \neq \sigma_2$ y $\rho = 0$ entonces es común que se le llame densidad elíptica normal.

De una manera similar se puede definir la distribución normal trivariada, la cual es también una distribución normal multivariada y para la cual únicamente definiremos la distribución normal trivariada o trinormal estandarizada.

Definición 6.4 *Si X_1, X_2 y X_3 tienen una función de densidad conjunta dada como sigue:*

$$f_{X_1,X_2,X_3}(x_1,x_2,x_3) = \frac{1}{(2\pi)^{\frac{3}{2}}\Delta^{\frac{1}{2}}} \exp\left[-\frac{1}{2}\sum_{i=1}^{3}\sum_{j=1}^{3} A_{ij}x_i x_j\right].$$

Donde $\Delta = 1 - {\rho_{23}}^2 - {\rho_{13}}^2 - {\rho_{12}}^2 + 2\rho_{23}\rho_{13}\rho_{12}$, $A_{11} = \frac{1-{\rho_{23}}^2}{\Delta}$, $A_{22} = \frac{1-{\rho_{13}}^2}{\Delta}$, $A_{33} = \frac{1-{\rho_{23}}^2}{\Delta}$, $A_{12} = A_{21} = \frac{\rho_{13}\rho_{23}-\rho_{12}}{\Delta}$, $A_{13} = A_{31} = \frac{\rho_{12}\rho_{23}-\rho_{13}}{\Delta}$, $A_{23} = A_{32} = \frac{\rho_{12}\rho_{13}-\rho_{23}}{\Delta}$, entonces se dice que X_1 , X_2 y X_3 tienen una distribución estandarizada normal trivariada o trinormal.

De modo semejante a la distribución bivariada, tenemos que si en la distribución trivariada todas las ρ son cero y todas las σ son iguales entonces a la distribución normal trivariada se le suele llamar distribución normal esférica; si todas las ρ son cero pero no todas las σ son iguales entonces la distribución trivariada se le suele llamar distribución normal elipsoidal.

La distribución normal multivariada se emplea comúnmente como una aproximación a distribuciones conjuntas para las cuales sus distribuciones marginales no se distribuyen exactamente normales, también es muy usada en la estructura teórica para la construcción de pruebas estadísticas y procedimientos de estimación.

La distribución normal multivariada ha sido estudiada de una manera más extensa que otros distribuciones multivariadas, en particular la distribución normal bivariada fue estudiada a principios del siglo XIX y de forma más extensa en el último cuarto de ese siglo por Schols y Galton, los cuales introdujeron la idea de correlación y regresión e impulsaron el conocimiento de las posibles formas de la distribución multivariada, más tarde Pearson aplicó la distribución normal bivariada en el estudio de datos biomédicos siendo él mismo el que realizó los primeros cálculos para los valores de probabilidad en la distribución normal bivariada.

A continuación estudiaremos otra distribución multivariada, en este caso discreta, que es de gran importancia por tener una amplia variedad de aplicaciones, la cual es la distribución multinomial.

6.2 Distribución multinomial

Esta distribución la cual es una generalización de la variable aleatoria binomial, puede ser definida como sigue:

Definición 6.5 *Considerese una serie de n pruebas independientes en las cuales solamente uno de los eventos E_1, $E_2, \ldots$, E_k sucede y en los cuales la probabilidad de que suceda E_i es p_i $(i = 1, 2, \ldots, k)$ para cada prueba, entonces la distribución conjunta de las variables aleatorias X_1, $X_2, \ldots$, X_k que representan el número de ocurrencias de los eventos E_1, $E_2, \ldots$, E_k en n pruebas, se le llama distribución multinomial con parámetros n, p_1, $p_2, \ldots$, p_n.*

Es importante hacer notar que $\sum_{j=1}^{k} p_j = 1$ y $\sum_{j=i}^{k} X_j = n$ lo cual surge de manera natural de la definición. A continuación calcularemos la función de densidad conjunta de una distribución multinomial con parámetros n, $p_1, p_2, \ldots, p_n$.

Teorema 6.6 *Si $X_1, X_2, \ldots, X_k$ tienen una distribución multinomial con parámetros n, $r_1, r_2, \ldots, r_n$ entonces su función de densidad conjunta esta dada*[1] *como sigue*

$$f_{X_1,X_2,\ldots,X_k}(x_1, x_2, \ldots, x_k) = \begin{cases} n! \prod_{j=1}^{k} \frac{p_j^{x_j}}{x_j!} & si\ x_1, x_2, \ldots, x_k \in \mathcal{N}\ y\ \sum_{j=i}^{k} x_j = n \\ 0 & e.o.c. \end{cases}.$$

Dem.

Para justificar este teorema tenemos que calcular

$$f_{X_1,X_2,\ldots,X_k}(x_1, x_2, \ldots, x_k) = P\{X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k\},$$

esto significa que queremos que en las n pruebas, exactamente x_1 de éstas se obtenga el evento E_1, x_2 de éstas se obtenga el evento $E_2, \ldots$, x_k de éstas se obtenga el evento E_k, donde $\sum_{j=i}^{k} X_j = n$. De aquí podemos ver fácilmente que la probabilidad de que en las n pruebas exactamente, las primeras x_1 de éstas se obtenga el evento E_1, las siguientes x_2 de éstas se obtenga el evento $E_2, \ldots$, y las últimas x_k de éstas se obtenga el evento E_k, donde $\sum_{j=i}^{k} X_j = n$ es $\prod_{j=1}^{k} p_j^{x_j}$, ya que las n pruebas son independientes. Por éste último hecho también es claro que la probabilidad de cualquier ordenamiento de ellas que resulten en que exactamente x_1 de éstas se obtenga el evento E_1, x_2 de éstas se obtenga el evento $E_2, \ldots, x_k$ de éstas se obtenga el evento E_k, donde $\sum_{j=i}^{k} X_j = n$, es $\prod_{j=1}^{k} p_j^{x_j}$, por lo cual recordando la teoría combinatoria tenemos que el número de ordenamientos de estas n

[1] En algunos libros de texto cuando se define a la distribución multinomial, en cada prueba se consideran $k+1$ eventos distintos con probabilidad r_i $(i = 1.2, \ldots, k+1)$, y se expresa a la función de densidad en términos únicamente de k variables aleatorias como sigue:

$$f_{X_1,X_2,\ldots,X_k}(x_1, x_2, \ldots, x_k) = \begin{cases} n! \prod_{j=1}^{k+1} \frac{p_j^{x_j}}{x_j!} & \text{si } x_1, x_2, \ldots, x_k \in \mathcal{N} \text{ y } \sum_{j=i}^{k+1} x_j = n \\ 0 & \text{e.o.c.} \end{cases},$$

Donde $x_{k+1} = n - \sum_{j=1}^{k} x_j$ y $p_{k+1} = 1 - \sum_{j=1}^{k} r_j$ están completamente determinadas por las primeras k variables aleatorias ya que $X_{k+1} = n - \sum_{j=i}^{k} X_j$.

pruebas es $\frac{n!}{\prod_{j=1}^{k} n_j!}$*, por lo tanto tenemos que*

$$f_{X_1,X_2,\ldots,X_k}(x_1,x_2,\ldots,x_k) = \begin{cases} n!\prod_{j=1}^{k}\frac{p_j^{x_j}}{x_j!} & si\ x_1,x_2,\ldots,x_k \in \mathcal{N}\ y \sum_{j=i}^{k} x_j = n \\ 0 & e.o.c. \end{cases}.$$

Por lo que se establece el teorema.

Note que si X_1 y X_2 tienen una distribución multinomial con parámetros n, p_1, p_2 entonces cada una de ellas tiene una distribución binomial con parámetros (n, p_1) y (n, p_2) respectivamente donde $p_1 + p_2 = 1$.

Ejemplo 6.1 *Supongase que un dado es lanzado 10 veces. Calcular la probabilidad de que en los resultados del dado se obtengan:*

1. *Exactamente una vez 1, dos veces 2, tres veces 3, tres veces cuatro y una vez 6.*

2. *Exactamente cuatro veces 3 y dos veces 6.*

3. *Sol.*

En la primera parte la probabilidad de obtener ese evento es exactamente

$$P\left\{X_1 = 1, X_2 = 2, X_3 = 3, \dot{X}_4 = 3, X_5 = 0, X_6 = 1\right\},$$

donde $X_1, X_2, \ldots, X_6$ tienen una distribución multinomial con parámetros $n = 10, p_1 = p_2 = \ldots = p_6 = \frac{1}{6}$ por lo que

$$\begin{aligned} P\{X_1 = 1, X_2 = 2, X_3 = 3, X_4 = 3, \quad & X_5 = 0, X_6 = 1\} \\ &= \frac{10!}{1!2!3!3!0!1!} \times \frac{1}{6}\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^3\left(\frac{1}{6}\right)^3\left(\frac{1}{6}\right)^0\frac{1}{6} \\ &= .00083352 \ . \end{aligned}$$

Para la segunda parte podemos definir una nueva distribución multinomial X_3, X_6, U donde X_3 es el número de veces que sale 3, X_6 es el número de veces que se sale 6 y U es el número de veces que no sale ni 3 ni 2, por lo que:

$$\begin{aligned} P\left\{X_3 = 4, \dot{X}_6 = 2, U = 4\right\} &= \frac{10!}{4!2!4!} \times \left(\frac{1}{6}\right)^4\left(\frac{1}{6}\right)^2\left(\frac{4}{6}\right)^4 \\ &= 0.0133363, \end{aligned}$$

ya que la probabilidad de que en un lanzamiento no salga ni el 4 ni el 2 es $\frac{4}{6}$.

Es fácil demostrar que la distribución conjunta de cualquier subconjunto $X_{a_1}, X_{a_2}, \ldots, X_{a_s}$ tiene también una distribución multinomial (con una variable $X_{s+1} = n - \sum_{j=i}^{s} X_{a_j}$). De hecho se tiene que su función de densidad conjunta está dada como sigue:

$$f_{X_{a_1},\ldots,X_{a_s}}(x_{a_1},\ldots,x_{a_s}) = \frac{n!}{\left(n - \sum_{j=i}^{s} x_{a_j}\right)! \prod_{j=1}^{s} x_{a_j}!} \left(1 - \sum_{j=i}^{s} p_{a_j}\right)^{n - \sum_{j=i}^{s} x_{a_j}} \prod_{j=1}^{s} p_{a_j}^{x_{a_j}}$$

si $x_{a_1}, x_{a_2}, \ldots, x_{a_s} \in \mathcal{N}$ y $\sum_{j=i}^{s} x_{a_j} \leq n$.

Esta distribución es empleada en diversos campos del análisis estadístico. También es usada en las mismas circunstancias que la distribución binomial cuando hay más de dos posibles resultados en un experimento.

Ejemplo 6.2 *Un importante campo de aplicación para la distribución multinomial es en la teoría cinética de la física clásica, en particular en la estadística termodinámica de "Maxwell-Boltzmann". Supongase que se desea determinar el estado de equilibrio de un sistema físico, compuesto por un número muy grande de N partículas de la misma naturaleza las cuales ocupan k celdas en un espacio sincronizado, ésto es que a cada partícula le es asignada una celda en un espacio de seis dimensiones (tres para la posición y tres para la velocidad). Las celdas se forman por una fina subdivisión del espacio, cada asignación de las N partículas entre las n celdas constituye un microestado. Un macroestado, es descrito como la n-éada $(x_1, x_2, \ldots, x_n)$ cuya j-ésima componente x_j es el número de partículas en el j-ésimo microestado. El estado de equilibrio del sistema de partículas se define como el macroestado $(x_1, x_2, \ldots, x_n)$ con la mayor probabilidad de ocurrencia, de aquí que para resolver este problema es necesario encontrar la distribución de las variables aleatorias $X_1, X_2, \ldots, X_n$, donde X_j describe el número de partículas que se encuentran en el j-ésimo microestado. Es fácil ver que $X_1, X_2, \ldots, X_n$ tiene una distribución multivariada con parámetros N, $p_1, p_2, \ldots, p_n$ donde p_i es la probabilidad de que una partícula esté en un microestado.*

La distribución multinomial también puede ser aplicada cuando se tienen datos obtenidos en una muestra aleatoria que pueden ser acomodados en un número finito de grupos, pero en la práctica la distribución multinomial tiene muchos problemas para poder ser aplicada en la distribución de poblaciones dado que existen problemas con respecto al cálculo numérico, de aquí que se hayan desarrollado diversas aproximaciones para calcular ésta (de hecho muchas de estas aproximaciones hacen uso de distribuciones continuas). Otra situación en la cual puede ser aplicada la distribución multinomial es en el cálculo de tablas de contingencia cuando en éstas se representan la incidencia conjunta de dos, tres o k factores.

6.3 Ejercicios

1. Probar que si $\mathbf{X} = X_1, X_2, \ldots, X_n$ tienen una distribución normal multivariada, entonces $E[\mathbf{X}] = \mu$, el cual se vio en la definición 6.1.

2. Probar que si $Z_1, Z_2, \ldots, Z_n$ son variables aleatorias independientes idénticamente cada una con una distribución normal estándar, entonces $Cov(Z_i, Z_i) = 1$ si $i = 1, 2, \ldots, n$.

3. Probar que si $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ tiene una distribución normal multivariada y D es una matriz no-singular de $n \times n$, entonces el vector de variables aleatorias $\mathbf{U} = D\mathbf{X}$ tiene también una distribución normal multivariada.

4. Probar que si $X_1, X_2, \ldots, X_n$ tienen una distribución normal multivariada y $\{k_1, k_2, \ldots, k_n\}$ es una permutación de los enteros $\{1, 2, \ldots, n\}$ entonces $X_{k_1}, X_{k_2}, \ldots, X_{k_n}$ también tiene una distribución normal multivariada.

5. Probar que si $X_1, X_2, \ldots, X_n$ tienen una distribución normal multivariada y $c_1, c_2, \ldots, c_n$ son constantes no todas cero, entonces la variable aleatoria

$$W = c_1X_1 + c_2X_2 + \ldots + c_nX_n,$$

tiene una distribución normal.

6. Probar que si $X_1, X_2, \ldots, X_n$ tienen una distribución normal multivariada con vector de esperanzas μ y matriz de covarianzas C entonces $X_1, X_2, \ldots, X_n$ son independientes si y sólo si C es una matriz diagonal.

7. Los artículos que produce una máquina se pueden clasificar de cuatro distintos tipos A, B, C y D. Se sabe que estos artículos se producen en las siguientes proporciones:

Tipo A	Tipo B	Tipo C	Tipo D
0.3	0.4	0.2	0.1

¿Cuál es la probabilidad de que haya exactamente un artículo de cada tipo en una muestra de cuatro artículos seleccionados al azar de un lote producido por la máquina?

8. Un vendedor ambulante vende cepillos de tres tamaños, a los que llama grande, extra-grande y gigante. Estima que al entrevistar a una persona las probabilidades son: 0.4 de no vender, 0.3 de vender un cepillo grande y 0.1 de vender un cepillo gigante. Encuentre la probabilidad de que en cuatro entrevistas venda i) cero cepillos, ii) cuatro cepillos grandes, iii) por lo menos un cepillo de cada clase.

9. Se lanza nueve veces un dado legal. ¿Cuál es el número más probable de veces en las que aparecerá i) un 6, ii) un número par?

Capítulo 7

Distribuciones condicionales

En el capítulo 1 definimos el concepto de probabilidad condicional, en este capítulo utilizaremos esa definición para definir las distribuciones condicionales que son basicamente una extensión de ese concepto, además encontraremos la relación de las distribuciones condicionales con el concepto de independencia.

Primeramente daremos estas definiciones para dos variables aleatorias y después extenderemos este concepto para k variables aleatorias.

Ya que existen algunas dificultades, que veremos más adelante, para definir estos conceptos cuando tenemos variables aleatorias absolutamente continuas, primero daremos estas definiciones en el caso de variables aleatorias discretas, ya que surge de manera más natural que en el caso en que tenemos variables aleatorias absolutamente continuas.

7.1 Funciones de densidad y distribución condicionales para variables aleatorias discretas

Definición 7.1 *Sean X y Y variables aleatorias discretas con una función de densidad conjunta $f_{X,Y}(\cdot,\cdot)$. La función de densidad discreta de Y dado que $X=x$, es denotada por $f_{Y|X}(\cdot \mid x)$, y se define como*

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

Podemos ver que la variable aleatoria Y dado que $X = x$ cuando $f_X(x) = 0$ no puede tomar ningún valor, por lo que a la función de densidad la definiremos como cero ya que el evento $\{X = x, Y = y\}$ tiene probabilidad cero, puesto que para que suceda este evento X debe tomar el valor de x lo cual sucede con probabilidad cero dado que $f_X(x) = 0$.

Es claro que si X y Y son variables aleatorias discretas las cuales toman probabilidades positivas únicamente en los valores $x_1, x_2, \ldots$ y $y_1, y_2, \ldots$ respectivamente (i.e., $f_X(x_i) > 0$ y $f_Y(y_i) > 0$ para $i = 1, 2, \ldots$) entonces

$$\begin{aligned} f_{Y|X}(y_j \mid x_i) &= \tfrac{f_{X,Y}(x_i, y_j)}{f_X(x_i)} \\ &= \tfrac{P\{X=x_i, Y=y_j\}}{P\{X=x_i\}} \\ &= P\{Y = y_j \mid X = x_i\} \text{ para } i, j = 1, 2, \ldots \end{aligned}$$

Teorema 7.1 *Si X y Y son variables discretas entonces $f_{Y|X}(\cdot \mid x)$ es una función de densidad cuando x es un valor fijo y $f_X(x) > 0$.*

Dem.

Es claro que $f_{Y|X}(\cdot \mid x)$ es una función mayor o igual que 0 ya que $f_{X,Y}(\cdot, \cdot) \geq 0$ y $f_X(x) > 0$. Falta probar que

$$\sum_{y_i \in \mathcal{R}: f_{X,Y}(x, y_i) > 0} f_{Y|X}(y_i \mid x) = 1,$$

por lo que

$$\begin{aligned} \sum_{y_i \in \mathcal{R}: f_{X,Y}(x, y_i) > 0} f_{Y|X}(y_i \mid x) &= \sum_{y_i \in \mathcal{R}: f_{X,Y}(x, y_i) > 0} \tfrac{f_{X,Y}(x, y_j)}{f_X(x)} \\ &= \tfrac{1}{f_X(x)} \sum_{y_i \in \mathcal{R}: f_{X,Y}(x, y_i) > 0} f_{X,Y}(x, y_j) \\ &= \tfrac{1}{f_X(x)} f_X(x) \\ &= 1, \end{aligned}$$

por lo que se establece el teorema.

Definición 7.2 *Si X y Y son variables aleatorias discretas con función de densidad conjunta $f_{X,Y}(\cdot, \cdot)$, la función de distribución condicional de Y dado que $X = x$ denotada por $F_{Y|X}(\cdot \mid x)$ se define como*

$$F_{Y|X}(y \mid x) = \begin{cases} P\{Y \leq y \mid X = x\} & si\ f_X(x) > 0 \\ 0 & si\ f_X(x) = 0 \end{cases}.$$

Es claro que

$$\begin{aligned} F_{Y|X}(y \mid x) &= P\{Y \leq y \mid X = x\} \\ &= \sum_{y_i \in \mathcal{R}: y_i \leq y} f_{Y|X}(y_i \mid x) \end{aligned}$$

Ejemplo 7.1 *A tiene dos centavos y B tiene un centavo. Ellos acuerdan jugar volados hasta que alguno de los dos no le queden centavos. Encontrar la función de densidad de la duración del juego si A gana.*

Sol.

Primeramente encontraremos la densidad de la duración del juego. Si X denota el número de tiros que dura el juego entonces

$$\begin{aligned} P\{X=1\} &= P\{\text{ Gane A}\} \\ &= \frac{1}{2}, \\ P\{X=2\} &= P\{\text{Gane B}\}\,P\{\text{Gane B}\} \\ &= \frac{1}{4}, \\ P\{X=3\} &= P\{\text{Gane B}\}\,P\{\text{Gane A}\}\,P\{\text{Gane A}\} \\ &= \frac{1}{8}, \\ &\vdots \\ P\{X=n\} &= \frac{1}{2^n}. \end{aligned}$$

Por lo cual

$$f_X(x) = \begin{cases} \frac{1}{2^x} & \text{si } x = 1, 2, \ldots \\ 0 & \text{e.o.c.} \end{cases}.$$

Ahora la función de densidad de X dado que A gana es

$$\begin{aligned} f_{X|A\text{ gana}}(x) &= P\{X = x \mid A \text{ gana}\} \\ &= \frac{P\{X = x, A \text{ gana}\}}{P\{A \text{ gana}\}}. \end{aligned}$$

Para calcular $P\{A \text{ gana}\}$ tenemos

$$\begin{aligned} P\{A \text{ gana}\} &= \sum_{i=1}^{\infty} \frac{1}{2^{2i-1}} \\ &= 2\sum_{i=1}^{\infty} \frac{1}{(2^2)^i} \\ &= 2\sum_{i=1}^{\infty} \frac{1}{4^i} \end{aligned}$$

$$\begin{aligned} &= 2\left[\frac{\frac{1}{4}}{1-\frac{1}{4}}\right] \\ &= \frac{2}{3}, \end{aligned}$$

y

$$P\{X = x, A \text{ gana}\} = \begin{cases} \frac{1}{2^x} & \text{si } x = 1, 3, 5, \ldots \\ 0 & \text{e.o.c.} \end{cases}.$$

Por lo que

$$f_{X|A\text{ gana}}(x) = \begin{cases} \frac{3}{2^{x+1}} & \text{si } x = 1, 3, 5, \ldots \\ 0 & \text{e.o.c.} \end{cases}.$$

En ocasiones si se tienen la función de densidad condicional de Y dado que $X = x$ y la función de densidad marginal de X y deseamos calcular la función de densidad marginal de Y, podemos recordar la fórmula de probabilidad total con la cual podemos obtener el siguiente resultado:

Corolario 7.2 *Si X y Y son variables aleatorias discretas y se tiene que $f_{Y|X}(\cdot \mid x)$ es la función de densidad conjunta de Y dado que $X = x$ y $f_X(\cdot)$ entonces la función de densidad marginal de Y se puede calcular como*

$$f_Y(y) = \sum_{x_i : f_X(x_i) > 0} f_{Y|X}(y \mid x_i) f_X(x_i).$$

Dem.

Sea $R(X) = rango(X)$ i.e.,

$$\begin{aligned} R(X) &= \{x \in \mathcal{R} : \exists \omega \in \Omega \text{ t. q. } X(\omega) = x\} \\ &= \{x_1, x_2, \ldots\}, \end{aligned}$$

y

$$A_i = \{\omega \in \Omega : X(\omega) = x_i\},$$

la cual es una partición del espacio muestral Ω de X, la cual cumple las siguientes propiedades:

(a) $\bigcup_{i=1}^{\infty} A_i = \Omega$,

(b) $A_i \cap A_j = \emptyset$ si $i \neq j$.

Es claro que

$$f_X(x_i) = P\{A_i\},$$

y

$$f_{Y|X}(y \mid x_i) = P\{Y = y \mid A_i\},$$

por lo que recordando la fórmula de la probabilidad total tenemos que:

$$\begin{aligned} f_Y(y) &= P\{Y = y\} \\ &= \sum_{i:P\{A_i\}>0} P\{Y = y \mid A_i\} P\{A_i\} \\ &= \sum_{x_i:f_X(x_i)>0} f_{Y|X}(y \mid x_i) f_X(x_i), \end{aligned}$$

por lo que el teorema queda establecido.

Ejemplo 7.2 *Sea X_1 el número que obtenemos al lanzar un dado. Entonces X_1 monedas son lanzadas y el número de soles es denotado por X_2. Encontrar la función de densidad marginal de X_2.*

Sol.

Utilizando la fórmula de probabilidad total tenemos que

$$f_{X_2}(x_2) = \sum_{x_1:f_{X_1}(x_1)>0} f_{X_2|X_1}(x_2 \mid x_1) f_{X_1}(x_1).$$

Es fácil calcular $f_{X_2|X_1}(x_2 \mid x_1)$ por lo cual

$X_1 \setminus X_2$	0	1	2	3	4	5	6
1	$\frac{1}{2}$	$\frac{1}{2}$	0	0	0	0	0
2	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	0	0	0	0
3	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	0	0	0
4	$\frac{1}{16}$	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{1}{16}$	0	0
5	$\frac{1}{32}$	$\frac{5}{32}$	$\frac{5}{16}$	$\frac{5}{16}$	$\frac{5}{32}$	$\frac{1}{32}$	0
6	$\frac{1}{64}$	$\frac{3}{32}$	$\frac{15}{64}$	$\frac{5}{16}$	$\frac{15}{64}$	$\frac{3}{32}$	$\frac{1}{64}$

También tenemos que X_1 tiene una distribución uniforme discreta que puede tomar los valores $1, 2, \ldots, 6$ y que X_2 puede tomar únicamente los valores de $0, 1, \ldots 6$, de aquí tenemos que

$$\begin{aligned} f_{X_2}(0) &= \sum_{i=1}^{6} f_{X_2|X_1}(0 \mid i) f_{X_1}(i) \\ &= \frac{1}{6}\left(\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \frac{1}{32} + \frac{1}{64}\right) \\ &= \frac{1}{6}\frac{63}{64} \\ &= \frac{21}{128}, \end{aligned}$$

$$\begin{aligned}
f_{X_2}(1) &= \sum_{i=1}^{6} f_{X_2|X_1}(1 \mid i) f_{X_1}(i) \\
&= \frac{1}{6}\left(\frac{1}{2}+\frac{1}{2}+\frac{3}{8}+\frac{1}{4}+\frac{5}{32}+\frac{3}{32}\right) \\
&= \frac{1}{6}\frac{15}{8} \\
&= \frac{5}{16},
\end{aligned}$$

$$\begin{aligned}
f_{X_2}(2) &= \sum_{i=2}^{6} f_{X_2|X_1}(2 \mid i) f_{X_1}(i) \\
&= \frac{1}{6}\left(\frac{1}{4}+\frac{3}{8}+\frac{3}{8}+\frac{5}{16}+\frac{15}{64}\right) \\
&= \frac{1}{6}\frac{99}{64} \\
&= \frac{33}{128},
\end{aligned}$$

$$\begin{aligned}
f_{X_2}(3) &= \sum_{i=3}^{6} f_{X_2|X_1}(3 \mid i) f_{X_1}(i) \\
&= \frac{1}{6}\left(\frac{1}{8}+\frac{1}{4}+\frac{5}{16}+\frac{5}{16}\right) \\
&= \frac{1}{6}\frac{8}{8} \\
&= \frac{1}{6},
\end{aligned}$$

$$\begin{aligned}
f_{X_2}(4) &= \sum_{i=4}^{6} f_{X_2|X_1}(4 \mid i) f_{X_1}(i) \\
&= \frac{1}{6}\left(\frac{1}{16}+\frac{5}{32}+\frac{15}{64}\right) \\
&= \frac{1}{6}\frac{29}{64} \\
&= \frac{29}{384},
\end{aligned}$$

$$\begin{aligned}
f_{X_2}(5) &= \sum_{i=5}^{6} f_{X_2|X_1}(5 \mid i) f_{X_1}(i) \\
&= \frac{1}{6}\left(\frac{1}{32}+\frac{3}{32}\right)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{6}\frac{4}{32} \\
&= \frac{1}{48},
\end{aligned}
$$

$$
\begin{aligned}
f_{X_2}(6) &= f_{X_2|X_1}(6 \mid 6) f_{X_1}(6) \\
&= \frac{1}{6}\left(\frac{1}{64}\right) \\
&= \frac{1}{384}.
\end{aligned}
$$

Por lo cual tenemos que:

$$
f_{X_2}(x_2) = \begin{cases} \frac{21}{128} & \text{si } x_2 = 0 \\ \frac{5}{16} & \text{si } x_2 = 1 \\ \frac{33}{128} & \text{si } x_2 = 2 \\ \frac{1}{6} & \text{si } x_2 = 3 \\ \frac{29}{384} & \text{si } x_2 = 4 \\ \frac{1}{48} & \text{si } x_2 = 5 \\ \frac{1}{384} & \text{si } x_2 = 6 \\ 0 & \text{e.o.c.} \end{cases}.
$$

Es fácil extender el concepto de distribución condicional para cuando tenemos más de dos variables, para ello enunciaremos la siguiente definición:

Definición 7.3 *Sean $X_1, X_2, \ldots, X_n$ variables aleatorias discretas con función de densidad conjunta $f_{X_1,\ldots,X_n}(\cdot,\ldots,\cdot)$ y sean*

$$
H = \{k_1, k_2, \ldots, k_j\} \;\; y \;\; H^c = \{k_{j+1}, k_{j+2}, \ldots, k_n\}, j \leq n-1
$$

subconjuntos no vacíos de los enteros $\{1, 2, \ldots, n\}$. La función de densidad conjunta de $X_{k_1}, X_{k_2}, \ldots, X_{k_j}$ dado que

$$
X_{k_{j+1}} = x_{k_{j+1}}, X_{k_{j+2}} = x_{k_{j+2}}, \ldots, X_{k_n} = x_{k_n},
$$

se denota como sigue

$$
f_{X_{k_1},\ldots,X_{k_j}|X_{k_{j+1}},\ldots,X_{k_n}}(\cdot,\ldots,\cdot \mid x_{k_{j+1}},\ldots,x_{k_n}),
$$

y esta definida por

$$
\begin{aligned}
f_{X_{k_1},\ldots,X_{k_j}|X_{k_{j+1}},\ldots,X_{k_n}} & (x_{k_1},\ldots,x_{k_j} \mid x_{k_{j+1}},\ldots,x_{k_n}) \\
&= \frac{f_{X_1,\ldots,X_{k_n}}(x_1,\ldots,x_n)}{f_{X_{k_{j+1}},\ldots,X_{k_n}}(x_{k_{j+1}},\ldots,x_{k_n})}
\end{aligned}
$$

si $f_{X_{k_{j+1}},\ldots,X_{k_n}}(x_{k_{j+1}},\ldots,x_{k_n}) > 0$, *y como*

$$f_{X_{k_1},\ldots,X_{k_j}|X_{k_{j+1}},\ldots,X_{k_n}}(x_{k_1},\ldots,x_{k_j} \mid x_{k_{j+1}},\ldots,x_{k_n}) = 0$$

si $f_{X_{k_{j+1}},\ldots,X_{k_n}}(x_{k_{j+1}},\ldots,x_{k_n}) = 0$, *donde*

$$f_{X_{k_{j+1}},\ldots,X_{k_n}}(x_{k_{j+1}},\ldots,x_{k_n})$$

es la función de densidad conjunta de $X_{k_{j+1}},\ldots,X_{k_n}$.

Ejemplo 7.3 *Si* $X_1,\ldots,X_r$ *tienen una distribución conjunta multinomial con parámetros* n, p_1, p_2, $\ldots$, p_r, *es claro que* X_i *depende de alguna manera del resultado que se obtenga en* X_j *si* $i \neq j$ *ya que* $\sum_{j=i}^{r} X_j = n$, *por lo que es interesante calcular la función de densidad condicional la cual es*

$$f_{X_{k_1},\ldots,X_{k_j}|X_{k_{j+1}},\ldots,X_{k_r}}(x_{k_1},\ldots,x_{k_j} \mid x_{k_{j+1}} + \ldots + x_{k_r} = s),$$

donde $s = 0, 1, \ldots, n$.

Sol.

Para poder encontrar esta probabilidad definamos, una nueva variables aleatoria U, la cual está dada por

$$U = X_{k_{j+1}} + \ldots + X_{k_r},$$

se puede ver fácilmente que X_{k_1}, X_{k_2}, $\ldots$, X_{k_j}, U tienen una distribución multinomial con parámetros n, p_{k_1}, p_{k_2}, $\ldots$, p_{k_j}, p_α donde

$$p_\alpha = p_{k_{j+1}} + p_{k_{j+2}} + \ldots + p_{k_r},$$

de aquí podemos escribir la probabilidad que deseamos de la siguiente manera:

$$\begin{aligned} f_{X_{k_1},\ldots,X_{k_j}|U}(x_{k_1},\ldots,x_{k_j} &\mid U = s) \\ &= \frac{P\{X_{k_1}=x_{k_1},\ldots,X_{k_j}=x_{k_j},U=s\}}{P\{U=s\}}. \end{aligned}$$

Sólo queda por calcular $P\{U = s\}$. Sea

$$W = X_{k_1} + X_{k_2} + \ldots + X_{k_j},$$

es fácil ver que U y W tienen también una distribución multinomial con parámetros n, p_α, p_β donde

$$p_\beta = p_{k_1} + p_{k_2} + \ldots + p_{k_j},$$

por lo que

$$
\begin{aligned}
P\{U=s\} &= P\{U=s, W=n-s\} \\
&= \tfrac{n!}{s!(n-s)!}\, p_\alpha{}^{s}\, p_\beta{}^{n-s} \\
&\quad \text{para } s=0,1,\ldots,n.
\end{aligned}
$$

Por lo cual

$$
\begin{aligned}
& f_{X_{k_1},\ldots,X_{k_j}|X_{k_{j+1}},\ldots,X_{k_s}} \quad (x_{k_1},\ldots,x_{k_j} \mid x_{k_{j+1}}+\ldots+x_{k_s}=s) \\
&= \frac{P\{X_{k_1}=x_{k_1},X_{k_2}=x_{k_2},\ldots,X_{k_j}=x_{k_j},U=s\}}{P\{U=s\}} \\
&= \frac{n!\left(\prod\limits_{i=1}^{j}\frac{p_{k_i}{}^{x_{k_i}}}{x_{k_i}!}\right)\frac{p_\alpha{}^{s}}{s!}}{\frac{n!}{s!(n-s)!}\, p_\alpha{}^{s}\, p_\beta{}^{n-s}} \\
&= \frac{(n-s)!}{p_\beta{}^{n-s}} \prod_{i=1}^{j} \frac{p_{k_i}{}^{x_{k_i}}}{x_{k_i}!} \\
& \text{cuando} \quad x_{k_i}\in\mathcal{N},\ \sum_{i=1}^{j} x_i = n-s \text{ y } s=0,1,2,\ldots,n.
\end{aligned}
$$

De esta forma la función de densidad conjunta de la probabilidad condicional de

$$X_{k_1}, X_{k_2}, \ldots, X_{k_j}$$

dado que

$$X_{k_{j+1}} = x_{k_{j+1}}, X_{k_{j+2}} = x_{k_{j+2}}, \ldots, X_{k_s} = x_{k_s},$$

está dada como sigue

$$
\begin{aligned}
& f_{X_{k_1},\ldots,X_{k_j}|X_{k_{j+1}},\ldots,X_{k_s}} \quad (x_{k_1},\ldots,x_{k_j} \mid x_{k_{j+1}}+\ldots+x_{k_s}=s) \\
&= \begin{cases} \dfrac{(n-s)!}{p_\beta{}^{n-s}} \displaystyle\prod_{i=1}^{j} \frac{p_{k_i}{}^{x_{k_i}}}{x_{k_i}!} & \begin{array}{l} \text{si } x_{k_i}\in\mathcal{N}, \\ \sum_{i=1}^{j} x_i = n-s \\ \text{y } s=0,1,2,\ldots,n \end{array} \\ 0 & \text{e.o.c.} \end{cases},
\end{aligned}
$$

donde $p_\beta = p_{k_1} + p_{k_2} + \ldots + p_{k_j}$.

7.2 Funciones de densidad y distribución condicionales para variables aleatorias absolutamente continuas.

Cuando tenemos variables que no son discretas, la manera en la cual se definen las distribuciones condicionales, tiene algunos problemas puesto que si se tiene una variable aleatoria X que no es discreta entonces existe $x \in \mathcal{R}$ el cual puede ser tomado por la variable aleatoria tal que $P\{X = x\} = 0$, de tal forma que la distribución condicional de Y dado que $X = x$ no puede ser definida con el concepto de probabilidad condicional dado en el capítulo 1, puesto que utilizando esa definición tenemos que :

$$\begin{aligned} P\{Y \leq y \mid X = x\} &= \frac{P\{X = x, Y \leq y\}}{P\{X = x\}} \\ &= \frac{0}{0}. \end{aligned}$$

Ya que el evento $\{Y \leq y\}$ dado que ha sucedido el evento $\{X = x\}$ claramente tiene una probilidad positiva tenemos que buscar una definición alternativa cuando X es una variable aleatoria que no es discreta. Definiremos este concepto únicamete cuando tenemos una variable aleatoria absolutamente continua.

Definición 7.4 *Sean X, Y variables aleatorias absolutamente continuas entonces la probabilidad condicional de $X \in A$ dado que $Y = y$ se define como:*

$$P\{X \in A \mid Y = y\} = \lim_{\epsilon \to 0} \frac{P\{X \in A, y < Y \leq y + \epsilon\}}{P\{y < Y \leq y + \epsilon\}}$$

si $P\{y < Y \leq y + \epsilon\} > 0$ para todo $\epsilon > 0$ y como

$$P\{X \in A \mid Y = y\} = 0$$

si $P\{y < Y \leq y + \epsilon\} = 0$ para alguna $\epsilon > 0$.

Esta definición a primera vista podría parecer conveniente cuando Y no es discreta ni absolutamente continua, pero no es así, dado que este límite podría no existir para valores de $y \in \mathcal{R}$ que puede tomar la variable Y (ver Parzen [9]).

Cuando tenemos variables aleatorias absolutamente continuas la definición de la función de distribución de X dado que $Y = y$ se puede dar como sigue:

Definición 7.5 *Sean X, Y variables aleatorias absolutamente continuas entonces la función de distribución condicional de X dado que $Y = y$ se define como*

$$\begin{aligned} F_{X|Y}(x \mid y) &= P\{X \leq x \mid Y = y\} \\ &= \lim_{\epsilon \to 0} \frac{P\{X \leq x, y < Y \leq y + \epsilon\}}{P\{y < Y \leq y + \epsilon\}} \end{aligned}$$

si $P\{y < Y \leq y + \epsilon\} > 0$ *para todo* $\epsilon > 0$ *y*

$$F_{X|Y}(x \mid y) = 0$$

si $P\{y < Y \leq y + \epsilon\} = 0$ *algún* $\epsilon > 0$.

Esta última definición resulta poco útil puesto que está expresada en términos de un límite que a primera vista es desconocido por nosotros, para evitar esta dificultad enunciaremos el siguiente teorema:

Teorema 7.3 *Si X, Y son variables aleatorias absolutamente continuas con funciones de distribución conjunta y marginales derivables entonces la función de distribución conjunta de X dado que $Y = y$ está dada como:*

$$F_{X|Y}(x \mid y) = \begin{cases} \frac{\int_{-\infty}^{x} f_{X,Y}(t,y)dt}{f_Y(y)} & si\ f_Y(y) > 0 \\ 0 & si\ f_Y(y) = 0 \end{cases}.$$

<u>Dem.</u>

Es claro que

$$F_{X|Y}(x \mid y) = \lim_{\epsilon \to 0} \frac{P\{X \leq x, y < Y \leq y + \epsilon\}}{P\{y < Y \leq y + \epsilon\}},$$

por lo que tenemos que

$$\begin{aligned} F_{X|Y}(x \mid y) &= \lim_{\epsilon \to 0} \frac{\frac{1}{\epsilon} P\{X \leq x, y < Y \leq y + \epsilon\}}{\frac{1}{\epsilon} P\{y < Y \leq y + \epsilon\}} \\ &= \lim_{\epsilon \to 0} \frac{\frac{1}{\epsilon}\{F_{X,Y}(x, y+\epsilon) - F_{X,Y}(x,y)\}}{\frac{1}{\epsilon}\{F_Y(y+\epsilon) - F_Y(y)\}} \\ &= \frac{\lim_{\epsilon \to 0} \frac{1}{\epsilon}\{F_{X,Y}(x,y+\epsilon) - F_{X,Y}(x,y)\}}{\lim_{\epsilon \to 0} \frac{1}{\epsilon}\{F_Y(y+\epsilon) - F_Y(y)\}} \\ &= \frac{\frac{\partial F_{X,Y}(x,y)}{\partial y}}{\frac{dF_Y(y)}{dy}} \\ &= \frac{\int_{-\infty}^{x} f_{X,Y}(t,y)dt}{f_Y(y)} \ \text{si}\ f_Y(y) > 0. \end{aligned}$$

Donde $F_{X,Y}(\cdot,\cdot)$ y $f_{X,Y}(\cdot,\cdot)$ es la función de distribución y de densidad conjunta de X y Y respectivamente, $F_Y(\cdot)$ es la función de distribución de Y y $f_Y(\cdot)$ es la función de densidad de Y.

A partir de este teorema podemos calcular fácilmente la función de densidad de X dado que $Y = y$, por lo cual enunciaremos el siguiente corolario:

Corolario 7.4 *Si X y Y son dos variables absolutamente continuas, entonces la función de densidad de X dado que $Y = y$ está dada como sigue*

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

Del último corolario tenemos que

$$P\{X \in A \mid Y = y\} = \int_A f_{X|Y}(x \mid y)dx, \tag{7.1}$$

cuando $f_Y(y) > 0$, por lo que la función de densidad condicional de X dado que $Y = y$ está definida de una manera consistente cuando se fija el valor de y y $f_Y(y) > 0$. Considerando este resultado podemos enunciar el siguiente teorema:

Teorema 7.5 *Si X y Y son dos variables absolutamente continuas, con función de densidad conjunta continua en todo $\mathcal{R}$ entonces la función de densidad conjunta de X y Y está dada como sigue:*

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

Dem.

Esto se desprende inmediatamente del corolario 7.4 si $f_Y(y) > 0$, si $f_Y(y) = 0$ tenemos que demostrar entonces que $f_{X,Y}(x, y) = 0$. Por consiguiente podemos observar que

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dx = 0,$$

y dado que $f_{X,Y}(x, y) \geq 0$ para todo $(x, y) \in \mathcal{R}^2$ tenemos que $f_{X,Y}(x, y) = 0$ para toda $x \in \mathcal{R}$ dado que $f_{X,Y}$ es una función continua[1].

Ejemplo 7.4 *Supongase que el número de accidentes de un automovilista durante un año es una variable aleatoria Y teniendo una distribución Poisson con parámetro λ, donde λ depende del automovilista. Si tomamos un automovilista aleatoriamente de alguna población, podemos dejar que λ varie y defina una nueva variable aleatoria absolutamente continua X la cual tiene una función de densidad f_X. Encontrar la función de densidad marginal de Y dado que X es una variable aleatoria gamma con parámetros (α, β). A partir de aquí encontrar la función de densidad conjunta de X dado que $Y = y$.*

[1] Si $f_{X,Y}$ no fuera continua, entoces puede ser que existan $x \in \mathcal{R}$ tales que $f_{X,Y}(x, y) > 0$, pero en este caso el conjunto de tales x's seria muy pequeño, en otras palabras este conjunto tiene medida de Lebesgue cero.

Sol.

La densidad condicional de Y dado que $X = \lambda$, es una densidad Poisson con parámetro λ y está dada como:

$$f_{Y|X}(y \mid \lambda) = \begin{cases} \frac{\lambda^y e^{-\lambda}}{y!} & \text{para } y = 0, 1, \ldots \\ 0 & \text{e.o.c.} \end{cases}.$$

De aquí despejando la función de densidad conjunta $f_{X,Y}(\cdot,\cdot)$ de

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

tenemos que

$$f_{X,Y}(\lambda, y) = \begin{cases} \frac{f_X(\lambda)\lambda^y e^{-\lambda}}{y!} & \text{para } y = 0, 1, \ldots \\ 0 & \text{e.o.c.} \end{cases}.$$

Donde $f_X(\lambda)$ es la función de densidad de una variable aleatoria gamma con parámetros α y $\lambda = \beta$, por lo que se tiene que:

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

y

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

De aquí tenemos que

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f_{X,Y}(\lambda, y) d\lambda \\ &= \int_0^{\infty} \frac{\beta^\alpha \lambda^{\alpha-1} e^{-\lambda\beta}}{\Gamma(\alpha)} \frac{\lambda^y e^{-\lambda}}{y!} d\lambda \\ &= \frac{\beta^\alpha}{y!\Gamma(\alpha)} \int_0^{\infty} \lambda^{\alpha+y-1} e^{-\lambda(\beta+1)} d\lambda \\ &= \frac{\beta^\alpha}{y!\Gamma(\alpha)} \frac{1}{(\beta+1)^{\alpha+y}} \Gamma(\alpha+y) \\ &= \frac{\Gamma(\alpha+y)\beta^\alpha}{y!\Gamma(\alpha)(\beta+1)^{\alpha+y}} \text{ si } y = 0, 1, \ldots \end{aligned}$$

La función de densidad conjunta de X dado que $Y = y$ se puede calcular fácilmente ya que tenemos la función de densidad conjunta de X y Y y también la función de densidad marginal de Y, por lo que tenemos que

$$\begin{aligned} f_{X|Y}(\lambda \mid y) &= \frac{f_{X,Y}(\lambda,y)}{f_Y(y)} \\ &= \frac{\frac{\beta^\alpha \lambda^{\alpha-1} e^{-\lambda\beta}}{\Gamma(\alpha)} \frac{\lambda^y e^{-\lambda}}{y!}}{\frac{\Gamma(\alpha+y)\beta^\alpha}{y!\Gamma(\alpha)(\beta+1)^{\alpha+y}}} \\ &= \frac{\lambda^{\alpha+y-1} e^{-\lambda(\beta+1)}(\beta+1)^{\alpha+y}}{\Gamma(\alpha+y)\beta^\alpha} \text{ si } \lambda > 0 \text{ y } y = 0, 1, \ldots \end{aligned}$$

Ejemplo 7.5 *La función de densidad conjunta de X y Y es dada por*

$$f_{X,Y}(x,y) = \begin{cases} xe^{-x(y+1)} & si\ x > 0\ \ y\ y > 0 \\ 0 & e.\ o.c. \end{cases}.$$

Encontrar la distribución condicional de X dado que $Y = y$.

Encontrar la distribución condicional de Y dado que $X = x$.

Sol.

Para resolver ambos incisos se necesita primeramente obtener las distribuciones marginales de X y Y, las cuales se dan a continuación

$$\begin{aligned} f_X(x) &= \int_0^\infty xe^{-x(y+1)}dy \\ &= e^{-x} \text{ si } x > 0, \\ f_Y(y) &= \int_0^\infty xe^{-x(y+1)}dx \\ &= \frac{1}{(y+1)^2} \text{ si } y > 0. \end{aligned}$$

De esta manera para el primer inciso se tiene

$$\begin{aligned} f_{X|Y}(x \mid y) &= \frac{f_{X,Y}(x,y)}{f_Y(y)} \\ &= \frac{xe^{-x(y+1)}}{\frac{1}{(y+1)^2}} \\ &= (y+1)^2 xe^{-x(y+1)} \text{ si } x > 0 \text{ y } y > 0, \end{aligned}$$

y para el segundo tenemos

$$\begin{aligned} f_{Y|X}(y \mid x) &= \frac{f_{X,Y}(x,y)}{f_X(x)} \\ &= \frac{xe^{-x(y+1)}}{e^{-x}} \\ &= xe^{-xy} \text{ si } x > 0 \text{ y } y > 0. \end{aligned}$$

El teorema que enunciaremos a continuación es de gran utilidad y es análogo a un teorema enunciado en la sección dedicada a las distribuciones condicionales de variables aleatorias discretas:

Teorema 7.6 *Si X y Y son variables aleatorias absolutamente continuas y se tiene que $f_{Y|X}(\cdot \mid x)$ es la función de densidad de Y dado que $X = x$ y $f_X(\cdot)$ es la función de densidad marginal de X entonces se tiene que la función de densidad marginal de Y se puede calcular como*

$$f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X}(y \mid x) f_X(x) dx. \tag{7.2}$$

Dem.

Primeramente tenemos que

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx,$$

y por otro lado tenemos que

$$f_{Y|X}(y \mid x) = \begin{cases} \frac{f_{X,Y}(x,y)}{f_X(x)} & si\ f_X(x) > 0 \\ 0 & si\ f_X(x) = 0 \end{cases}, \tag{7.3}$$

podemos ver también que si $f_X(x) = 0$ entonces tenemos que

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy = 0,$$

de donde tenemos que si $f_X(x) = 0$ entonces

$$\int_{-\infty}^{y_0} f_{X,Y}(x,y)dy = 0 \tag{7.4}$$

para todo $y_0 \in \mathcal{R}$. A partir de esto tenemos que para todo $y_0 \in \mathcal{R}$ se cumple que

$$\begin{aligned}
P\{Y \leq y_0\} &= \textstyle\int_{-\infty}^{y_0} f_Y(y)dy \\
&= \textstyle\int_{-\infty}^{y_0} \int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy \\
&= \textstyle\int_{-\infty}^{\infty} \int_{-\infty}^{y_0} f_{X,Y}(x,y)dydx \\
&= \int_{\{x: f_X(x) > 0\}} \textstyle\int_{-\infty}^{y_0} f_{X,Y}(x,y)dydx \text{ por 7.4} \\
&= \int_{\{x: f_X(x) > 0\}} \textstyle\int_{-\infty}^{y_0} f_{Y|X}(y \mid x) f_X(x)dydx \text{ por 7.3} \\
&= \textstyle\int_{-\infty}^{\infty} \int_{-\infty}^{y_0} f_{Y|X}(y \mid x) f_X(x)dxdy \\
&= \textstyle\int_{-\infty}^{y_0} \int_{-\infty}^{\infty} f_{Y|X}(y \mid x) f_X(x)dydx,
\end{aligned}$$

por lo tanto tenemos que

$$f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X}(y \mid x) f_X(x)dx.$$

Se puede dar un resultado más general al anterior teorema y el cual únicamente enunciaremos:

Teorema 7.7 *Si X y Y son variables aleatorias absolutamente continuas entonces*

$$P\{X \in A, Y = y\} = \int_A f_{Y|X}(y \mid x) f_X(x) dx. \tag{7.5}$$

Ejemplo 7.6 *Considere el decaimiento de una partícula de un cierto tipo en una cámara de nubes. Asuma que la variable aleatoria X es el tiempo en el cual decae la partícula, obedece a una distribución exponencial con parámetro θ. Sin embargo se tiene que el parámetro θ depende del tipo de partícula tomada. Considere que si en la cámara de nubes es tomada aleatoriamente una partícula, entonces el parámetro θ obedece a una distribución gamma con parámetros (α, λ). Encontrar la función de densidad marginal de X.*

<u>Sol.</u>

Es claro que se tiene lo siguiente:

$$f_\theta(\theta) = \frac{\lambda^\alpha \theta^{\alpha-1} e^{-\lambda\theta}}{\Gamma(\alpha)} \text{ si } \theta > 0,$$

y

$$f_{X|\theta}(x \mid \theta) = \theta e^{-\theta x} \text{ si } x > 0 \text{ y } \theta > 0.$$

Por consiguiente tenemos, si $x > 0$

$$\begin{aligned} f_X(x) &= \int_0^\infty f_{X|\theta}(x \mid \theta) f_\theta(\theta) d\theta \\ &= \int_0^\infty \frac{\lambda^\alpha \theta^\alpha e^{-\theta(\lambda+x)}}{\Gamma(\alpha)} d\theta \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty \theta^\alpha e^{-\theta(\lambda+x)} d\theta. \end{aligned}$$

Si hacemos $u = \theta(\lambda + x)$ entonces $du = (\lambda + x)d\theta$, por lo cual

$$\begin{aligned} f_X(x) &= \frac{\lambda^\alpha}{\Gamma(\alpha)(\lambda+x)^{\alpha+1}} \int_0^\infty u^\alpha e^{-u} du \\ &= \frac{\lambda^\alpha \Gamma(\alpha+1)}{\Gamma(\alpha)(\lambda+x)^{\alpha+1}} \\ &= \frac{\lambda^\alpha \alpha}{(\lambda+x)^{\alpha+1}} \text{ si } x > 0. \end{aligned}$$

La extensión de la definición de distribuciones condicionales cuando se tienen n variables aleatorias es directa pero demasiado tediosa, por lo cual se enunciará el siguiente teorema, cuya demostración se dejará como ejercicio:

Teorema 7.8 *Sean $X_1, X_2, \ldots, X_n$ variables aleatorias absolutamente continuas con función de densidad conjunta $f_{X_1,\ldots,X_n}(\cdot,\ldots,\cdot)$ y sean*

$$\begin{aligned} H &= \{k_1, k_2, \ldots, k_j\} \\ y\ H^c &= \{k_{j+1}, k_{j+2}, \ldots, k_n\} \quad j \leq n-1 \end{aligned}$$

subconjuntos no vacíos de los enteros $\{1, 2, \ldots, n\}$. La función de densidad conjunta de $X_{k_1}, X_{k_2}, \ldots, X_{k_j}$ dado que

$$X_{k_{j+1}} = x_{k_{j+1}}, X_{k_{j+2}} = x_{k_{j+2}}, \ldots, X_{k_n} = x_{k_n},$$

se denota como sigue

$$f_{X_{k_1},\ldots,X_{k_j}|X_{k_{j+1}},\ldots,X_{k_n}}(\cdot,\ldots,\cdot \mid x_{k_{j+1}},\ldots,x_{k_n}),$$

y está definida por

$$\begin{aligned} f_{X_{k_1},\ldots,X_{k_j}|X_{k_{j+1}},\ldots,X_{k_n}} & (x_{k_1},\ldots,x_{k_j} \mid x_{k_{j+1}},\ldots,x_{k_n}) \\ = & \frac{f_{X_1,\ldots,X_{k_n}}(x_1,\ldots,x_n)}{f_{X_{k_{j+1}},\ldots,X_{k_n}}(x_{k_{j+1}},\ldots,x_{k_n})} \end{aligned} \tag{7.6}$$

si $f_{X_{k_{j+1}},\ldots,X_{k_n}}(x_{k_{j+1}},\ldots,x_{k_n}) > 0$, y

$$f_{X_{k_1},\ldots,X_{k_j}|X_{k_{j+1}},\ldots,X_{k_n}}(x_{k_1},\ldots,x_{k_j} \mid x_{k_{j+1}},\ldots,x_{k_n}) = 0$$

si

$$f_{X_{k_{j+1}},\ldots,X_{k_n}}(x_{k_{j+1}},\ldots,x_{k_n}) = 0,$$

donde $f_{X_{k_{j+1}},\ldots,X_{k_n}}(x_{k_{j+1}},\ldots,x_{k_n})$ es la función de densidad conjunta de $X_{k_{j+1}}, \ldots, X_{k_n}$.

Podemos también enunciar el siguiente teorema el cual no demostraremos:

Teorema 7.9 *Si $X_1, \ldots, X_n$ y X_{n+1} son variables aleatorias absolutamente continuas y se tiene que $f_{X_1,\ldots X_n|X_{n+1}}(\cdot,\ldots,\cdot \mid x_{n+1})$ es la función de densidad conjunta de $X_1, \ldots, X_n$ dado que $X_{n+1} = x_{n+1}$ y $f_{X_{n+1}}(\cdot)$ es la función de densidad marginal de X_{n+1} entonces la función de densidad conjunta de $X_1, X_2, \ldots, X_n$ se puede calcular como*

$$\begin{aligned} f_{X_1,\ldots,X_n} & (x_1,\ldots,x_n) \\ = \int_{-\infty}^{\infty} & f_{X_1,\ldots,X_n|X_{n+1}}(x_1,\ldots,x_n \mid x) f_{X_{n+1}}(x)dx. \end{aligned} \tag{7.7}$$

Ejemplo 7.7 *Sea W una variable aleatoria gamma con parámetros (α, λ) y supongase que dando la condición de que $W = w$, las variables aleatorias $X_1, \ldots, X_n$ son n variables aleatorias independientes identicamente distribuidas cada una con una distribución exponencial con parámetro w. Demostrar que la distribución condicional de W dado que $X_1 = x_1$, $X_2 = x_2, \ldots, X_n = x_n$ tiene una distribución gamma con parámetros $(\alpha + n, \lambda + \sum_{i=1}^{n} x_i)$.*

Sol.

Dado que cuando condicionamos que $W = w$ las variables $X_1, X_2, \ldots, X_n$ son n variables aleatorias independientes identicamente distribuidas con una distribución exponencial con parámetro w, se tiene que

$$f_{X_1,\ldots,X_n|W}(x_1, \ldots, x_n \mid w) = \begin{cases} w^n e^{w \sum_{i=1}^{\infty} x_i} & \text{si } x_i > 0 \\ & i = 1, 2, \ldots, n \\ 0 & \text{e.o.c.} \end{cases}.$$

También podemos obtener la función de densidad conjunta de $f_{X_1,\ldots,X_n,W}(\cdot, \ldots, \cdot)$ ya que

$$f_{X_1,\ldots,X_n,W}(x_1, \ldots, x_n, w) = f_{X_1,\ldots,X_n|W}(x_1, \ldots, x_n \mid w) f_W(w),$$

de aquí tenemos

$$\begin{aligned} f_{X_1,X_2,\ldots,X_n,W}(x_1, \ldots, x_n, w) &= w^n e^{w \sum_{i=1}^{\infty} x_i} \frac{\lambda^\alpha w^{\alpha-1} e^{-\lambda w}}{\Gamma(\alpha)} \\ &= \frac{\lambda^\alpha w^{\alpha+n-1} e^{-w(\lambda + \sum_{i=1}^{n} x_i)}}{\Gamma(\alpha)} \quad \begin{array}{l} \text{si } x_i > 0, w > 0 \\ i = 1, 2, \ldots, n \end{array}. \end{aligned}$$

También se tiene que

$$f_{W|X_1,\ldots,X_n}(w \mid x_1, \ldots, x_n) = \frac{f_{X_1,\ldots,X_n,W}(x_1, \ldots, x_n, w)}{f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)}.$$

Por lo que necesitamos conocer $f_{X_1,\ldots,X_n}(\cdot, \ldots, \cdot)$, por lo que utilizando el último teorema tenemos que

$$\begin{aligned} f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) &= \int_0^\infty f_{X_1,\ldots,X_n|W}(x_1, \ldots, x_n \mid w) f_W(w) dw \\ &= \int_0^\infty \frac{\lambda^\alpha w^{\alpha+n-1} e^{-w(\lambda + \sum_{i=1}^{n} x_i)}}{\Gamma(\alpha)} dw \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty w^{\alpha+n-1} e^{-w(\lambda + \sum_{i=1}^{n} x_i)} dw \text{ si } x_i > 0. \end{aligned}$$

Haciendo $u = w(\lambda + \sum_{i=i}^{n} x_i)$ tenemos $du = (\lambda + \sum_{i=i}^{n} x_i)\,dw$ de aquí tenemos

$$\begin{aligned} f_{X_1,X_2,\ldots,X_n}(x_1,\ldots,x_n) &= \frac{\lambda^{\alpha}}{\Gamma(\alpha)\,(\lambda + \sum_{i=i}^{n} x_i)^{\alpha+n}} \int_0^{\infty} u^{\alpha+n-1} e^{-u} du \\ &= \frac{\lambda^{\alpha}\Gamma(\alpha+n)}{\Gamma(\alpha)\,(\lambda + \sum_{i=i}^{n} x_i)^{\alpha+n}}. \end{aligned}$$

Finalmente vemos que

$$\begin{aligned} f_{W|X_1,\ldots,X_n}(w \mid x_1,\ldots,x_n) &= \frac{\frac{\lambda^{\alpha} w^{\alpha+n-1} e^{-w\left(\lambda+\sum_{i=i}^{n} x_i\right)}}{\Gamma(\alpha)}}{\frac{\lambda^{\alpha}\Gamma(\alpha+n)}{\Gamma(\alpha)\left(\lambda+\sum_{i=i}^{n} x_i\right)^{\alpha+n}}} \\ &= \frac{w^{\alpha+n-1}\left(\lambda+\sum_{i=i}^{n} x_i\right)^{\alpha+n} e^{-w\left(\lambda+\sum_{i=i}^{n} x_i\right)}}{\Gamma(\alpha+n)} \end{aligned}$$

si $x_i > 0, w > 0, i = 1, 2, \ldots, n.$

Por lo que tenemos que la función de densidad condicional de W dado que $X_1 = x_1$, $X_2 = x_2, \ldots, X_n = x_n$ tiene una distribución gamma con parámetros $(\alpha + n, \lambda + \sum_{i=1}^{n} x_i)$.

7.3 Esperanza condicional

Un concepto de gran utilidad es el concepto de esperanza condicional y el cual surge de manera natural, cuando deseamos calcular la esperanza de una distribución condicional, ésto es:

Definición 7.6 *Si X y Y son variables aleatorias discretas con función de densidad condicional $f_{X|Y}(\cdot \mid y)$. La esperanza condicional de X dado que $Y = y$ se denotará por*

$$E[X \mid Y = y]$$

y si

$$\sum_{x: f_{X|Y}(x|y)>0} |x|\, f_{X|Y}(x \mid y) < \infty$$

entonces está definida como

$$E[X \mid Y = y] = \sum_{x: f_{X|Y}(x|y)>0} x f_{X|Y}(x \mid y) \tag{7.8}$$

si $f_Y(y) > 0$ y

$$E[X \mid Y = y] = 0$$

si $f_Y(y) = 0$.

Si X y Y son variables aleatorias absolutamente continuas con función de densidad condicional $f_{X|Y}(\cdot \mid y)$, la esperanza condicional de X dado que $Y = y$ se denotará por $E[X \mid Y = y]$ y si

$$\int_{-\infty}^{\infty} |x| \, f_{X|Y}(x \mid y) dx < \infty$$

se define como

$$E[X \mid Y = y] = \begin{cases} \int_{-\infty}^{\infty} x f_{X|Y}(x \mid y) dx & si \ f_Y(y) > 0 \\ 0 & si \ f_Y(y) = 0 \end{cases}. \tag{7.9}$$

Ejemplo 7.8 *Una urna contiene 4 bolas blancas y 6 bolas negras. Si X y Y denotan el número de bolas blancas obtenidas al sacar muestras aleatorias sin reemplazo de tamaño 3 y 5 respectivamente. Encontrar la esperanza condicional de X dado que $Y = y$, para $y = 0, 1, 2, 3, 4$.*

Sol.

Podemos calcular $f_{X|Y}(\cdot \mid y)$ para $y = 0, 1, 2, 3, 4$ por lo que tenemos que $f_{X|Y}$ está dado en la siguiente tabla:

$X \setminus Y$	0	1	2	3	4
0	0	0	$\frac{1}{10}$	$\frac{2}{5}$	1
1	0	$\frac{3}{10}$	$\frac{3}{5}$	$\frac{3}{5}$	0
2	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{10}$	0	0
3	$\frac{2}{5}$	$\frac{1}{10}$	0	0	0

De donde tenemos que $E[X \mid Y = i]$, $i = 0, 1, 2, 3, 4$ es

$$\begin{aligned} E[X \mid Y = 0] &= \sum_{i=2}^{3} i f_{X|Y}(i \mid 0) \\ &= 2\left(\frac{3}{5}\right) + 3\left(\frac{2}{5}\right) \\ &= \frac{12}{5}, \\ E[X \mid Y = 1] &= \sum_{i=1}^{3} i f_{X|Y}(i \mid 1) \\ &= \frac{3}{10} + 2\left(\frac{3}{5}\right) + 3\left(\frac{1}{10}\right) \\ &= \frac{9}{5}, \\ E[X \mid Y = 2] &= \sum_{i=0}^{2} i f_{X|Y}(i \mid 2) \end{aligned}$$

$$
\begin{aligned}
&= \frac{3}{5} + 2\left(\frac{3}{10}\right) \\
&= \frac{6}{5}. \\
E[X \mid Y = 3] &= \sum_{i=0}^{1} i f_{X|Y}(i \mid 3) \\
&= 1\left(\frac{3}{5}\right) \\
&= \frac{3}{5} \\
E[X \mid Y = 4] &= \sum_{i=0}^{0} i f_{X|Y}(i \mid 4) \\
&= 0
\end{aligned}
$$

Ejemplo 7.9 *Si X y Y tienen una función de densidad conjunta dada como sigue:*

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

Calcular la esperanza condicional de Y dado $X = x$, si $x > 0$.

Sol.

Calculando la función de densidad conjunta de Y dado $X = x$ tenemos que si $x > 0$

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

De aquí se tiene que si $x > 0$

$$
\begin{aligned}
E[Y \mid X = x] &= \int_{-\infty}^{\infty} y f_{Y|X}(y \mid x) dy \\
&= \int_{-x}^{x} y \frac{3(x^2-y^2)}{4x^3} dy \\
&= \frac{3}{4x^3} \int_{-x}^{x} x^2 y - y^3 dy \\
&= \frac{3}{4x^3} \left[\frac{y^2}{2} \Big|_{-x}^{x} - \frac{y^4}{4} \Big|_{-x}^{x} \right] \\
&= 0.
\end{aligned}
$$

Como las distribuciones condicionales satisfacen todas las propiedades que satisfacen las distribuciones ordinarias tenemos que la esperanza condicional satisface también las propiedades de la esperanza, por consiguiente si $g(x)$ es una función real tenemos las siguientes fórmulas:

$$E\left[g(X) \mid Y=y\right] = \begin{cases} \displaystyle\sum_{x:f_{X|Y}(x|y)>0} g(x) f_{X|Y}(x \mid y) & \text{si } X \text{ y } Y \text{ son discretas.} \\ \int_{-\infty}^{\infty} g(x) f_{X|Y}(x \mid y)dx & \text{si } X \text{ y } Y \text{ son absolutamente continuas.} \end{cases} \tag{7.10}$$

y

$$E\left[\sum_{i=1}^{n} X_i \mid Y=y\right] = \sum_{i=1}^{n} E\left[X_i \mid Y=y\right]. \tag{7.11}$$

Análogamente tenemos también que

$$E\left[g(X,Y) | Y=y\right] \begin{cases} \displaystyle\sum_{x:f_{X|Y}(x|y)>0} g(x,y) f_{X|Y}(x \mid y) & \text{si } X \text{ y } Y \text{ son discretas.} \\ \int_{-\infty}^{\infty} g(x,y) f_{X|Y}(x \mid y)dx & \text{si } X \text{ y } Y \text{ son absolutamente continuas.} \end{cases} \tag{7.12}$$

Ejemplo 7.10 *La función de densidad conjunta de X y Y es dada por*

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

Calcular $E\left[X^2 \mid Y=y\right]$.

Sol.

Para calcular $E\left[X^2 \mid Y=y\right]$ se necesita primeramente tener la función de densidad condicional de X dado que $Y=y$, por lo que si calculamos $f_{X|Y}(\cdot \mid y)$ cuando $y>0$ tenemos que

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

De aquí tenemos que si $y>0$

$$\begin{aligned} E\left[X^2 \mid Y=y\right] &= \int_{-\infty}^{\infty} x^2 f_{X|Y}(x \mid y)dx \\ &= \int_0^{\infty} \frac{x^2 e^{-\frac{x}{y}}}{y} dx \\ &= \frac{1}{y}\int_0^{\infty} x^2 e^{-\frac{x}{y}} dx. \end{aligned}$$

Si hacemos $u=\frac{x}{y}$ tenemos que $du=\frac{1}{y}dx$ de donde

$$\begin{aligned} E\left[X^2 \mid Y=y\right] &= \frac{1}{y}\int_0^{\infty} y^2 u^2 e^{-u} y du \\ &= y^2 \int_0^{\infty} u^2 e^{-u} du, \end{aligned}$$

haciendo esta integral por partes, tenemos que

$$E\left[X^2 \mid Y = y\right] = 2y^2,$$

dado que

$$\int_0^\infty u^2 e^{-u} du = 2.$$

Si definimos $h(y) = E\{X \mid Y = y\}$, entonces $E\{X \mid Y\} = h(Y)$. Por lo que tenemos que por si misma $E\{X \mid Y\}$ es una variable aleatoria. Una propiedad de extrema importancia de la esperanza condicional es la siguiente:

Teorema 7.10 *Si X y Y son variables aleatorias discretas o continuas ambas con esperanza finita, entonces*

$$E\{X\} = E\{E\{X \mid Y\}\}, \tag{7.13}$$

es decir que si X y Y son variables discretas se tiene

$$E\{X\} = \sum_{y: f_Y(y)>0} E\{X \mid Y = y\} f_Y(y), \tag{7.14}$$

y si X y Y son variables absolutamente continuas se tiene

$$E\{X\} = \int_{-\infty}^{\infty} E\{X \mid Y = y\} f_Y(y) dy. \tag{7.15}$$

<u>*Dem.*</u>

Si X y Y son variables aleatorias discretas entonces tenemos que

$$\begin{aligned}
E\{X\} &= \sum_{x: f_X(x)>0} x f_X(x) \\
&= \sum_{x: f_X(x)>0} x \sum_{y: f_Y(y)>0} f_{X,Y}(x,y) \\
&= \sum_{x: f_X(x)>0} \sum_{y: f_Y(y)>0} x f_{X,Y}(x,y) \\
&= \sum_{x: f_X(x)>0} \sum_{y: f_Y(y)>0} x \frac{f_{X,Y}(x,y)}{f_Y(y)} f_Y(y) \\
&= \sum_{x: f_X(x)>0} \sum_{y: f_Y(y)>0} x f_{X|Y}(x \mid y) f_Y(y) \\
&= \sum_{y: f_Y(y)>0} f_Y(y) \sum_{x: f_X(x)>0} x f_{X|Y}(x \mid y) \\
&= \sum_{y: f_Y(y)>0} E\{X \mid Y = y\} f_Y(y) \\
&= E\{E\{X \mid Y\}\}.
\end{aligned}$$

Si X y Y son variables aleatorias absolutamente continuas entonces tenemos que

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x f_X(x)dx \\ &= \int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} f_{X,Y}(x,y)dydx \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X,Y}(x,y)dxdy, \end{aligned}$$

por otra parte tenemos que si $f_Y(y) > 0$ tenemos que

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

o equivalentemente

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

además $f_{X|Y}(x \mid y)f_Y(y) = 0$ cuando $f_Y(y) = 0$.

También podemos ver que si $f_Y(y) = 0$ entonces

$$\begin{aligned} \int_{-\infty}^{\infty} x f_{X,Y}(x,y)dx &= \int_{-\infty}^{\infty} x f_{X|Y}(x \mid y) f_Y(y)dx \\ &= 0, \end{aligned}$$

para ello sea $y \in \mathcal{R}$ tal que $f_Y(y) = 0$, es claro entonces que

$$\int_{-\infty}^{\infty} f_{X,Y}(x,y)dx = 0,$$

y como $f_{X,Y}(x,y) \geq 0$ para todo $x, y \in \mathcal{R}$, tenemos que

$$\int_{-n}^{n} f_{X,Y}(x,y)dx = 0$$

para toda $n \in \mathcal{N}$, de aquí podemos ver que

$$\begin{aligned} \int_{-n}^{n} |x| f_{X,Y}(x,y)dx &\leq n \int_{-n}^{n} f_{X,Y}(x,y)dx \\ &= 0, \end{aligned}$$

de modo que

$$\int_{-n}^{n} x f_{X,Y}(x,y)dx = 0$$

para toda $n \in \mathcal{N}$ de donde se sigue que

$$\begin{aligned} \int_{-\infty}^{\infty} x f_{X,Y}(x,y)dx &= \lim_{n\to\infty} \int_{-n}^{n} x f_{X,Y}(x,y)dx \\ &= 0. \end{aligned}$$

Con este resultado vemos que

$$\begin{aligned}
E[X] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X,Y}(x,y)dxdy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X|Y}(x \mid y) f_Y(y)dydx \\
&= \int_{-\infty}^{\infty} f_Y(y)\int_{-\infty}^{\infty} x f_{X|Y}(x \mid y)dxdy \\
&= \int_{-\infty}^{\infty} E[X \mid Y = y] f_Y(y)dy \\
&= E[E[X \mid Y]]
\end{aligned}$$

como se quería demostrar.

Las últimas dos ecuaciones con las cuales obtenemos $E[X]$ podían ser interpretadas como el promedio de la esperanza condicional de X dado que $Y = y$ para todos los valores que toma Y. Este resultado es de gran utilidad y nos ayuda a calcular la $E[X]$ fácilmente. Los siguientes ejemplos ilustran la manera en que se puede utilizar este resultado:

Ejemplo 7.11 *Un niño que se encuentra perdido en un bosque, el puede tomar tres caminos diferentes, si toma el primer camino regresará a su punto de partida en 6 horas, si toma el segundo camino regresará también en 5 horas, si toma el tercer camino éste lo sacará del bosque en 9 horas. ¿Cuál es el tiempo esperado que el niño tardará en salir del bosque? Supongase que el niño no puede distinguir entre los tres caminos cuando regresa a su punto de partida.*

Sol.

Sea X la variable que denota el tiempo en que sale el niño del bosque. Sea Y la variable aleatoria que denota el camino que toma el niño, la cual es fácil ver que se distribuye de manera uniforme y toma los valores de 1,2,3 por lo que tenemos que

$$E[X] = \sum_{i=1}^{3} E[X \mid Y = i] f_Y(y).$$

Es claro que

$$\begin{aligned}
E[X \mid Y = 1] &= 6 + E[X] \\
E[X \mid Y = 2] &= 5 + E[X] \\
E[X \mid Y = 3] &= 9,
\end{aligned}$$

por lo que

$$E[X] = (6 + E[X])\frac{1}{3} + (5 + E[X])\frac{1}{3} + (9)\frac{1}{3}.$$

De aquí tenemos que despejando de la última ecuación $E[X]$ tenemos

$$E[X] = 20.$$

Ejemplo 7.12 *El número diario de automóviles que recibe un estacionamiento es una variable aleatoria con media 50 y el número de horas que permanece cada automóvil en el estacionamiento es una variable aleatoria independiente del número de automóviles diario con la siguiente distribución*

$$f(x) = \begin{cases} \left(\frac{1}{2}\right)^x & si\ x = 1, 2, \ldots \\ 0 & c.o.c. \end{cases}.$$

Si cada hora cuesta N\$6.00. ¿Cuál es el monto de dinero esperado que recauda diariamente el estacionamiento?

Sol.

Si N es el número de automóviles que recibe diariamente el estacionamiento y X_i es el número de horas que permaneció el automóvil i-ésimo en el estacionamiento. Entonces el monto de dinero recaudado diarlamente por el estacionamiento es $\$6E\left[\sum_{i=1}^{N} X_i\right]$. Por lo que tenemos que

$$E\left[\sum_{i=1}^{N} X_i\right] = E\left[E\left[\sum_{i=1}^{N} X_i \mid N\right]\right].$$

De aquí tenemos que por 7.12 y por la independencia entre X_i y N tenemos que

$$\begin{aligned} E\left[\sum_{i=1}^{N} X_i \mid N = n\right] &= \sum_{j=0}^{\infty}\left(\sum_{i=i}^{n} j\right) f_{X_j|N}(j|n) \\ &= \sum_{j=i}^{n}\sum_{i=0}^{\infty} i f_{X_i}(i) \\ &= nE[X_i]. \end{aligned}$$

Como X_i tiene una distribución geométrica con parámetro $p = \frac{1}{2}$ tenemos que

$$\begin{aligned} E[X_i] &= \frac{1}{p} \\ &= 2. \end{aligned}$$

Por lo cual

$$E\left[\sum_{i=1}^{N} X_i \mid N = n\right] = 2n,$$

y de aquí

$$E\left[\sum_{i=1}^{N} X_i \mid N\right] = 2N.$$

Por último tenemos que

$$\begin{aligned} E\left[\sum_{i=1}^{N} X_i\right] &= E\left[E\left[\sum_{i=1}^{N} X_i \mid N\right]\right] \\ &= E[2N] \\ &= 2(500) \\ &= 1000. \end{aligned}$$

Por lo que se tiene que el monto de dinero esperado que recauda el estacionamiento es de N$6,000.00.

Si condicionamos adecuadamente una variable aleatoria no solo podemos calcular esperanzas de variables aleatorias, sino también podemos calcular probabilidades. Si queremos calcular la probabilidad de que ocurra un evento A podemos definir la variable aleatoria X como el indicador del evento A, ésto es

$$X = \begin{cases} 1 \text{ si } A \text{ ocurre} \\ 0 \text{ si } A \text{ no ocurre} \end{cases}.$$

Es claro, por esta definición que

$$E[X] = P\{A\},$$

y

$$E[X \mid Y = y] = P\{A \mid Y = y\}$$

para cualquier variable aleatoria Y. Por consiguiente podemos ver que:

$$P\{A\} = \begin{cases} \sum\limits_{y: f_Y(y) > 0} P\{A \mid Y = y\} f_Y(y) & \text{si } Y \text{ es discreta} \\ \int_{-\infty}^{\infty} P\{A \mid Y = y\} f_Y(y) dy & \text{si } Y \text{ es absolutamente continua} \end{cases}.$$

Ejemplo 7.13 *Supongase que X_1 y X_2 son variables aleatorias independientes cada una función de distribución $F_{X_i}(\cdot)$ y con una función de densidad $f_{X_i}(\cdot)$. Calcular la distribución de la resta de X_1 y X_2.*

Sol.

Para calcular la distribución de $X_1 + X_2$ primeramente calcularemos su función de distribución, por lo que tenemos que

$$\begin{aligned}
F_{X_1+X_2}(z) &= P\{X_1 - X_2 \leq z\} \\
&= \int_{-\infty}^{\infty} P\{X_1 - X_2 \leq z \mid X_2 = x_2\} f_{X_2}(x_2) dx_2 \\
&= \int_{-\infty}^{\infty} P\{X_1 - x_2 \leq z \mid X_2 = x_2\} f_{X_2}(x_2) dx_2 \\
&= \int_{-\infty}^{\infty} P\{X_1 \leq z + x_2\} f_{X_2}(x_2) dx_2 \text{ por indep. de } X_1 \text{ y } X_2 \\
&= \int_{-\infty}^{\infty} F_{X_1}(z + x_2) f_{X_2}(x_2) dx_2.
\end{aligned}$$

Ejemplo 7.14 *El número de accidentes que una cierta persona tiene en un cierto año es una variable aleatoria Poisson con parámetro λ. Sin embargo supongase que el parámetro λ cambia de acuerdo a la persona, siendo igual a 2 en el 60% de las personas y a 3 en la proporción restante. Si una persona es tomada aleatoriamente. Calcular la probabilidad de que tenga cero accidentes y 3 accidentes exactamente.*

Sol.

Sea X la variable aleatoria que denota el número de accidentes que tiene una persona tomada aleatoriamente de la población. Si Y es una variable aleatoria que toma el valor de uno si la persona que fue tomada tiene $\lambda = 2$ y Y.es cero si $\lambda = 3$. Entonces tenemos que

$$\begin{aligned}
P\{X = 0\} &= \sum_{i=0}^{1} P\{X = 0 \mid Y = i\} f_Y(i) \\
&= e^{-2}\frac{2^0}{0!}(.4) + e^{-3}\frac{3^0}{0!}(.6) \\
&= 0.0840063,
\end{aligned}$$

$$\begin{aligned}
P\{X = 3\} &= \sum_{i=0}^{1} P\{X = 3 \mid Y = i\} f_Y(i) \\
&= e^{-2}\frac{2^3}{3!}(.4) + e^{-3}\frac{3^3}{3!}(.6) \\
&= 0.2066039.
\end{aligned}$$

7.4 Ejercicios

1. Demostrar el teorema 7.8.

2. El número de huevos puestos por cierto insecto es una variable aleatoria Poisson X con parámetro λ. Cada huevo tiene la misma probabilidad p de empollar, además cada uno de ellos empolla de manera independiente de los otros. Calcular la función de densidad del número de progenitores del insecto.

3. Tres monedas son lanzadas. Sea X el número de soles obtenidos en los dos primeros lanzamientos y Y el número de águilas obtenidas en los dos últimos lanzamientos.

   (a) Calcular la función de densidad conjunta de X y Y.

   (b) Encontrar la función de densidad condicional de Y dado que $X = 1$.

   (c) Encontrar la función de densidad condicional de X dado que $Y = 2$.

4. Considere una muestra de tamaño dos sin reemplazo, obtenida de una urna que contiene tres bolas, numeradas del 1 al 3. Sea X el número obtenido en la primera bola tomada y Y el mayor número en las dos bolas tomadas.

   (a) Encontrar la función de densidad conjunta de X y Y.

   (b) Encontrar $P\{X = 1 \mid Y = 3\}$

5. Sea Y una variable aleatoria con una distribución Poisson con parámetro λ. Asuma que la distribución condicional de X dado que $Y = y$ tiene una distribución binomial con parámetros $n = y$ y p. Encontrar la distribución marginal de X (asuma que $X = 0$ si $y = 0$).

6. Seis cartas son tomadas sin reemplazo de una baraja Inglesa. Sea X la variable aleatoria que denota el número de ases obtenidos y Y el número de reyes.

   (a) Encontrar la densidad conjunta de X y Y.

   (b) Encontrar la densidad condicional de X dado que $Y = y$.

7. Tres monedas son lanzadas n veces. Sea X la variable que denota el número de veces en que no aparece sol, Y el número de veces que aparece un sol y Z el número de veces que aparecen dos soles.

(a) Encontrar la densidad conjunta de X, Y y Z.

(b) Encontrar la densidad condicional de X y Y dado Z.

8. Supongase que W, la cantidad de humedad en el aire en un día dado, es una variable aleatoria gamma con parámetros (α, λ). Supongase también que dado que $W = w$, el número de accidentes durante el día es una variable aleatoria N con una distribución Poisson con parámetro λ. Mostrar que la distribución condicional de W dado que $N = n$ tiene una distribución gamma con parámetros $(\alpha + n, \lambda + 1)$.

9. El dueño de una tienda de televisores a observado que el 45% de los clientes que entran a su tienda compran una televisión a color, 15% de ellos compran una televisión blanco y negro, mientras el 45% restante entra solo a curiosear. Si 5 clientes entran a su tienda en un día dado, ¿Cuál es la probabilidad de que venda exactamente 2 televisores a color y 1 blanco y negro?

10. El número de gente que entra en una farmacia de 9 a 10 de la mañana es una variable aleatoria X que se distribuye Poisson con parámetro $\lambda = 10$. Calcule la probabilidad de que entre las 9 y las 10 de la mañana entren a lo más 3 hombres a la farmacia dado que 10 mujeres entraron en ésta a esa hora.

11. La función de densidad discreta de X está dada por

$$f_X(x) = \begin{cases} \frac{x}{3} & \text{si } x = 1, 2, 3 \\ 0 & \text{e.o.c.} \end{cases},$$

y $f_{Y|X}(y \mid x)$ tiene una distribución binomial con parámetros $p = \frac{1}{2}$ y $n = x$, i.e.,

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

Encontrar la distribución conjunta de X y Y.

12. Una urna contiene 4 bolas blancas y 6 bolas negras. Si X y Y denotan el número de bolas blancas obtenidas al sacar muestras aleatorias sin reemplazo de tamaño 3 y 5 respectivamente. Encontrar la función de densidad condicional de X dado que $Y = y$, para $y = 0, 1, 2, 3, 4$.

13. Sean X y Y variables aleatorias independientes con una distribución Poisson con parámetro λ Encontrar la función de densidad condicional de X dado Z, donde $Z = X + Y$.

14. La función de densidad conjunta de dos variables aleatorias discretas X y Y está definida como sigue

$$f_{X,Y}(x, y)$$

$X \backslash Y$	1	2
1	$\frac{1}{8}$	$\frac{1}{4}$
2	$\frac{1}{8}$	$\frac{1}{2}$

   (a) Calcule la función de densidad condicional de X dado que $Y = i, i = 1, 2$.

   (b) ¿Son X y Y independientes?

15. Si X_1 y X_2 tienen una distribución normal bivariada como se definió en el capítulo anterior. Mostrar que la densidad condicional de X_1 dado que $X_2 = x$, es una densidad normal con parámetros $\mu = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y - \mu_2)$ y $\sigma = {\sigma_1}^2(1 - \rho^2)$.

16. La función de densidad conjunta de X y Y es dada por

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

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

17. Una fábrica de chips para computadoras, produce diariamente $100,000$ chips, de los cuales un $100p\%$ de ellos son defectuosos. La proporción p no es constante y varía de un día a otro. El dueño de la fábrica ha encontrado que p varía de acuerdo a una distribución beta con parámetros $a = 1$ y $b = 99$. Si X_1 denota el número de productos defectuosos en un cierto día y X_2 la proporción de éstos en ese día. En muchas ocasiones la distribución de X_2 se le conoce como la distribución de Polya y surge en muchos otros contextos.

   (a) Encontrar la función de densidad marginal X_1.

   (b) Encontrar la función de densidad condicional de X_1 dado X_2.

   (c) Encontrar $P\{X_1 = 80\}$.

18. Dado que X_1 es una variable aleatoria distribuida uniformemente en el intervalo $(0,1)$. Sea X_2 una variable aleatoria distribuida uniformemente sobre el intervalo $(X_1, 1)$. Encontrar la función de densidad marginal de X_2. Extender este resultado si se tienen k variables $X_1, X_2, \ldots, X_k$, donde X_j esáa distribuida uniformemente sobre el intervalo $(X_{j-1}, 1)$ para $j = 2, 3, \ldots, k$.

19. Sea

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

Encontrar lo siguiente:

(a) $P\{X < Y \mid X > 2Y\}$.

(b) $P\{X > 2 \mid X + Y > 1\}$.

(c) $P\left\{X^2 + Y^2 \leq 1 \mid Y = \frac{1}{2}\right\}$.

20. Sea

$$f_{X,Y}(x,y) = \begin{cases} 3x & \text{si } 0 < y < x < 1 \\ 0 & \text{e.o.c.} \end{cases} .$$

Encontrar lo siguiente:

(a) Encontrar la función de densidad condicional de Y dado X.

(b) Encontrar la función de densidad condicional de X dado Y.

(c) $P\left\{X > \frac{1}{2} \mid Y < 2X\right\}$.

(d) $P\left\{X > \frac{1}{2} \mid X + Y < 1\right\}$.

(e) $P\left\{X^2 + Y^2 < \frac{1}{2} \mid Y = \frac{1}{2}\right\}$.

21. Sea

$$f_{X,Y}(x,y) = \begin{cases} \frac{6}{7}(x^2 + \frac{xy}{2}) & \text{si } 0 < x < 1, 0 < y < 2 \\ 0 & \text{e.o.c.} \end{cases} .$$

(a) Encontrar la función de densidad condicional de X dado Y.

(b) Encontrar la función de densidad de Y dado X.

(c) Encontrar $P\left\{Y > \frac{1}{2} \mid X < \frac{1}{2}\right\}$.

22. Si X_1 y X_2 tienen una distribución bivariada, representan las salidas diarias (en miles de unidades) de un cierto producto en cierta tienda en dos días consecutivos, tiene parámetros $\mu_1 = \mu_2 = 3$, $\sigma_1 = \sigma_2 = 1$ y $\rho = 0.8$. Encontrar K tal que:

    (a) $P\{X_2 > K\} = 0.05$.

    (b) $P\{X_2 > K \mid X_1 = 2\} = 0.05$.

    (c) $P\{X_2 > K \mid X_1 = 1\} = 0.05$.

    Supongase que la tienda en un día dado desea tener suficientes unidades del producto tal que con una probabilidad de 0.95 pueda satisfacer toda la demanda del producto durante el día.

    (d) ¿Qué tan grande debe ser el inventario por la mañana si ayer salieron 2000 unidades?

    (e) ¿Qué tan grande debe ser el inventario por la mañana si se desconoce las salidas del día anterior?

23. Sean X y Y variables aleatorias independientes distribuidas identicamente. Sea $Z = Y - X$. Sea $A = \{|Z| \leq 1\}$. Encontrar i)$P\{A \mid X = 1\}$,ii) la función de densidad condicional de $f_{Z|X}(\cdot \mid x)$, iii)$P\{Z \leq 0 \mid A\}$.

    (a) Si X y Y se distribuyen uniformes sobre el intervalo $(0, 2)$.

    (b) Si X y Y se distribuyen normales con parámetros $\mu = 0$ y $\sigma = 2$.

    (c) Si X y Y se distribuyen exponenciales con parámetro $\lambda = 1$.

24. Sean X y Y variables aleatorias independientes identicamente distribuidas. Sea $U = X + Y$ y $V = X - Y$. Sea $A = \{|V| \leq 1\}$. Encontrar i)$P\{A \mid U = 1\}$,ii)$f_{V|U}(\cdot \mid u)$, y iii)$P\{U \geq 0 \mid A\}$.

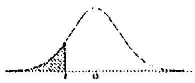
    (a) Si X y Y se distribuyen uniformes sobre el intervalo $(0, 2)$.

    (b) Si X y Y se distribuyen normales con parámetros $\mu = 0$ y $\sigma = 2$.

    (c) Si X y Y se distribuyen exponenciales con parámetro $\lambda = 1$.

Tabla1. Valores de la función de distribución normal estándar[1].

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx = P\{Z \leq z\}$$



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.5	.0002	.0002	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003
-3.4	.0003	.0003	.0004	.0004	.0004	.0004	.0004	.0004	.0005	.0005
-3.3	.0005	.0005	.0005	.0005	.0006	.0006	.0006	.0006	.0006	.0007
-3.2	.0007	.0007	.0007	.0008	.0008	.0008	.0008	.0009	.0009	.0009
-3.1	.0010	.0010	.0010	.0011	.0011	.0011	.0012	.0012	.0013	.0013
-3.0	.0013	.0014	.0014	.0015	.0015	.0016	.0016	.0017	.0018	.0018
-2.9	.0019	.0019	.0020	.0021	.0021	.0022	.0023	.0023	.0024	.0025
-2.8	.0026	.0026	.0027	.0028	.0029	.0030	.0031	.0032	.0033	.0034
-2.7	.0035	.0036	.0037	.0038	.0039	.0040	.0041	.0043	.0044	.0045
-2.6	.0047	.0048	.0049	.0051	.0052	.0054	.0055	.0057	.0059	.0060
-2.5	.0062	.0064	.0066	.0068	.0069	.0071	.0073	.0075	.0078	.0080
-2.4	.0082	.0084	.0087	.0089	.0091	.0094	.0096	.0099	.0102	.0104
-2.3	.0107	.0110	.0113	.0116	.0119	.0122	.0125	.0129	.0132	.0136
-2.2	.0139	.0143	.0146	.0150	.0154	.0158	.0162	.0166	.0170	.0174
-2.1	.0179	.0183	.0188	.0192	.0197	.0202	.0207	.0212	.0217	.0222
-2.0	.0228	.0233	.0239	.0244	.0250	.0256	.0262	.0268	.0274	.0281
-1.9	.0287	.0294	.0301	.0307	.0314	.0322	.0329	.0336	.0344	.0351
-1.8	.0359	.0367	.0375	.0384	.0392	.0401	.0409	.0418	.0427	.0436
-1.7	.0446	.0455	.0465	.0475	.0485	.0495	.0505	.0516	.0526	.0537
-1.6	.0548	.0559	.0571	.0582	.0594	.0606	.0618	.0630	.0643	.0655
-1.5	.0668	.0681	.0694	.0708	.0721	.0735	.0749	.0764	.0778	.0793
-1.4	.0808	.0823	.0838	.0853	.0869	.0885	.0901	.0918	.0934	.0951
-1.3	.0968	.0985	.1003	.1020	.1038	.1056	.1075	.1093	.1112	.1131
-1.2	.1151	.1170	.1190	.1210	.1230	.1251	.1271	.1292	.1314	.1335
-1.1	.1357	.1379	.1401	.1423	.1446	.1469	.1492	.1515	.1539	.1562
-1.0	.1587	.1611	.1635	.1660	.1685	.1711	.1736	.1762	.1788	.1814
-.9	.1841	.1867	.1894	.1922	.1949	.1977	.2005	.2033	.2061	.2090
-.8	.2119	.2148	.2177	.2206	.2236	.2266	.2296	.2327	.2358	.2389
-.7	.2420	.2451	.2483	.2514	.2546	.2578	.2611	.2643	.2676	.2709
-.6	.2743	.2776	.2810	.2843	.2877	.2912	.2946	.2981	.3015	.305
-.5	.3085	.3121	.3156	.3192	.3228	.3264	.3300	.3336	.3372	.3409
-.4	.3446	.3483	.3520	.3557	.3594	.3632	.3669	.3707	.3745	.3783
-.3	.3821	.3859	.3897	.3936	.3974	.4013	.4052	.4090	.4129	.4168
-.2	.4207	.4247	.4286	.4325	.4364	.4404	.4443	.4483	.4522	.4562
-.1	.4602	.4641	.4681	.4721	.4761	.4801	.4840	.4880	.4920	.4960

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.999	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998

[1]Si X es una variable aleatoria normal con parámetros μ y σ^2, entonces la función de distribución de X puede ser calculada conociendo únicamente los valores de la función de distribución de Z, la cual es una variable aleatoria normal estándar, ya que $Z = (X - \mu)/\sigma$. Esto es que

$$F_X(x) = P\{X \leq x\} = \Phi\left(\frac{x - \mu}{\sigma}\right).$$

Bibliografía

[1] B. V. Gndenko, "*The Theory of Probability,*" Chelsea Publishing Company, New York, 1968.

[2] Bernard Harris, "*Theory of Probability,*" Addison-Wesley, Reading, Mass., 1966.

[3] Paul G. Hoel, Sidney J. Port, Charles J. Stone , "*Introduction to Probability Theory,*" Houghton Mifflin Company, Boston, 1971.

[4] Kai Lai Chung, "*A Course in Probability Theory,*" Academic Press. Inc., San Diego Ca., 1968.

[5] Johnson N., Kotz S., "*Distributions in Statistics, Continuos Univariate Distributions-I,*" John Wiley and Sons, New York, 1970.

[6] Johnson N., Kotz S., "*Distributions in Statistics, Continuos Univariate Distributions-II,*" John Wiley and Sons, New York, 1970.

[7] Johnson N., Kotz S., "*Distributions in statistics, Discrete Distributions,*" John Wiley and Sons, New York, 1969.

[8] A. M. Mood, F. A. Graybill, D. C. Boes, "*Introduction to the Theory of Statistics,*" McGraw-Hill Book Company, Singapore, 1974.

[9] Emanuel Parzen, "*Modern Probability Theory and Its Applications,*" John Wiley and Sons, New York, 1992.

[10] V. K. Rohatgi, "*An Introduction to Probability Theory and Mathematical Statistics,*" John Wiley and Sons, New York, 1976.

[11] S. Ross, "*Introduction to Probability Models,*" Academic Press, Inc., Orlando, Florida, 1985.

[12] S. Ross, "*A First Course in Probabity,*" Macmillan Publishing Company, New York, 1989.

[13] Howard G. Tucker, "*An introduction to Probability and Mathematical Statistics,*" Academic Press, New York, 1965.