Resumen Probabilidad

Docente: José Domingo Restrepo Álvarez

Índice

Nota al lector

Estas notas se crean con el objetivo de dar una cuenta de lo que estamos realizando en el curso de Probabilidad a cargo del profesor José Domingo Restrepo Álvarez, dando los teoremas y definiciones respectivos, para llevar a cabo un buen curso de Probabilidad, y dar así una introducción fuerte a Inferencia Estadística, los temas se resumen en seguir los primeros seis capítulos del libro *Introduction to Probability and Mathematical Statistics*, daremos uno que otro ejemplo para ejemplificar mucho mejor la teoría (esto cuándo lo consideremos necesario cabe aclarar), y ser claros en que estamos desarrollando dentro de este curso.

De igual manera diremos cuándo consideremos pertinente revisar los ejemplos que realiza el autor del libro guía, dado que hay muchos que están muy bien construídos y permiten guiarse al momento de cada uno probar y retarse al hacerlos. Se dejarán por entre medio una que otra tarea de tal modo que la persona que lea estas notas pueda hacerse ciertas preguntas interesantes a lo largo del curso que podrían ayudar a desarrollar más la idea necesaria dentro del curso.

Además daremos demostraciones de ciertos teoremas al nivel del curso[1] que consideremos necesarios, y/o interesantes para desarrollar efectivamente el trayecto de este curso, algunos de estos teoremas son resultados importantes de la Probabilidad, tales como la desigualdad de Chebyshev, la de Markov, y hallar ciertas características de las Variables Aleatorias. En ocasiones tomaremos ejemplos del libro guía para ir desarrollando el concepto de forma mucho más clara, buscando desarrollar ciertos objetivos y habilidades que para más adelante podrán servirnos.

No queda más que desear éxitos en el camino de la Probabilidad, y que sea un grato trayecto y deseando que puedan tener un gran docente como lo fue el profesor José Domingo Restrepo Álvarez con nosotros en este curso.

Jose David Palacio Arias Estudiante de Matemáticas de la Universidad de Antioquia.

[1]Está estipulado ver este curso en 3er nivel, por lo que conocimientos en Teoría de la Medida, Análisis Real, y demás fundamentaciones de la Probabilidad, no se han desarrollado fuertemente, por lo avanzados con respecto al curso que son.

1. Parcial 1

Lo normal sería comenzar definiendo que es una probabilidad, y los efectos que esta tiene y trae en el munod cotidiano, pero podemos comenzar cuestionándonos ¿porqué? Existen muchos conceptos que deberemos definir, y mostrar, dado que no salen de manera inmediata, entraremos mostrando que es un Experimento Aleatorio, un Experimento Determinista, un Espacio Muestral, lo que es una Medida de Probabilidad y un Espacio de Probabilidad, conceptos necesarios para definir éxitosamente lo que es la probabilidad de una forma más bien formal, el único problema de este curso es la formalidad, no es mucha dado que se necesita un curso más adelante llamado *Teoría de la Medida*.

1.1. Combinatoria

Muchas veces ocurre que varias personas llegan al curso sin saber mucho de Combinatoria, herramienta que es fundamental para el Cálculo de Probabilidades, por lo que dejaremos consignado aquí varios desarrollos de la misma:

1.1.1. Principio Multiplicativo

Pensemos que en ocasiones puedo realizar cierta acción y obtengo un número n_1 de resultados, y que realizo otra acción y obtengo un número n_2 de resultados, puedo realizar ambas actividades de una forma $n_1 \cdot n_2$, por ejemplo:

Si tiro un dado de 10 caras, y saco una canica de una bolsa de 7 canicas (distinguibles todas), obtendré que las posibles salidas que tengo serán de $10 \cdot 7 = 70$, además de eso tendremos que el principio multiplicativo se cumple para cualquier cantidad de acciones que realicemos[2] en nuestro experimento aleatorio.

1.1.2. Variación

Cuando nos pidan formar a partir de un conjunto de n elementos, grupos de cierto tamaño r, lo consideraremos una *Variación*, existen distintas formas de realizar una variación, ya sea con repetición o sin repetición, en el caso de que deseemos una Variación con Repetición tomaremos la siguiente fórmula:

$$VR_n^r = n^r \tag{1}$$

Dónde aquí podremos elegir sin problema de repetir los elementos, un ejemplo es:

Si tengo 8 digitos (del 1 al 8), ¿Cuántos números de 5 dígitos puedo formar?

Como no me están restringiendo los dígitos, significa que yo puedo repetirlos sin problema alguno, por lo que queda que:

$VR_8^5 = 8^5 = 32768$

Si se restringen los elementos del conjunto a sin repetición, tendremos que la variación será:

$$V_n^r = \frac{n!}{(n-r)!} \tag{2}$$

Así, sí tomamos en cuenta el ejemplo anterior, tendremos que la cantidad de números de 5 dígitos sin repetición es:

[2] Enfaticemos que los procesos que se pueden realizar son netamente finitos, dado que si fueran infinitos, el producto sería infinito, cosa que no deseamos.

$V_8^5 = \frac{8!}{3!} = 6720$ Es claro que la cantidad de números iba a ser menor que la que no tiene repeticiones, dado que este proceso excluye varios números.

1.1.3. Permutación

El concepto de Permutación entra en vigor cuando estamos analizando el elegir conjuntos de tantos elementos dónde el orden nos es muy importante, aquí podemos aclarar que una permutación es un caso específico de una variación, pero de aquí en adelante daremos un tratamiento permutativo y combinatorio, dado que es con estos con los que definiremos la probabilidad con mayor habilidad y claridad. Podemso tener la siguiente fórmula para las Permutaciones:

$$nPm = \frac{n!}{(n-m)!} \tag{3}$$

Veamos ahora un caso específico de las permutaciones.

1.1.4. Permutaciones Circulares y Permutaciones con Repeticiones

Cabe el caso dónde nos pidan organizar objetos, personas, elementos, entre otros en un entorno dónde no conocemos el orden, es decir no sabemos que lugar va primero, y que lugar va después, por ejemplo en una mesa circular de 8 puestos, no podemos decir de forma clara cuál es la primera silla, y cuál es la última, por lo que entramos en conflicto con la definición de orden y permutación, así definiremos que la cantidad de posibilidades en un entorno de este tipo será:

$$PC_p = (p-1)!$$

Por lo que a las 8 personas podemos organizarlas de 7! formas distintas. Definamos ahora que pasa con las Permutaciones sí se llega a dar el caso de repeticiones.

Cabe el caso de que entre los objetos que estemos permutando, haya repetición en estos, es decir, que no podamos distinguir uno de otro, por lo que aquí sí estamos permutando estos objetos nos quedará finalmente que la cantidad de veces que se repita un elemento, irá en relación con el total, más claramente:

$$P_m^{m_1,m_2,\cdots,m_k} = \frac{m!}{m_1!m_2!\cdots m_k!}$$

Dónde $m_1, m_2, \cdots, m_k$ son la cantidad de veces que se repite los objetos $1, 2, \cdots, k$, demos un ejemplo que lo clarifique.

¿De cuántas formas podemos organizar la palabra TITIRIBI?

Si identificamos, la letra T se repite 2 veces, la letra I se repite 4 veces, la letra R, y B 1 vez, por lo que la permutación con repetición nos resulta en:

$$P_8^{4,2} = \frac{8!}{4!2!} = 840$$

Así hay 840 formas de organizar la palabra TITIRIBI.

1.1.5. Combinación

Las combinaciones entrarán cuando dentro de nuestra muestra no se nos pida tener en cuenta el orden, es decir que elegir un grupo en cualquier orden siempre será el mismo, daré un ejemplo antes de ir a mostrar la fórmula:

Si tengo que realizar un grupo de personas, ¿Cómo puedo elegir a Pablo, Sara, y Raúl? Dado que el tener a Raúl, Sara, Pablo; Sara Raúl Pablo; Pablo, Raúl, Sara. Y demás combinaciones son lo mismo en esencia, al final solo tendremos una, a esto se refiere la combinación, solo le interesa la cantidad de grupos formados, más no la cantidad teniendo en cuenta distintos órdenes de las mismas personas, así podemos definir que:

$$nCm = \frac{n!}{m!(n-m)!} \tag{4}$$

Así hemos definido la cantidad de grupos formados sin que nos importe el orden.

Luego podemos establecer la siguiente relación:

$$nCm = \frac{nPm}{m!} \Leftrightarrow m!(nCm) = nPm \tag{5}$$

1.1.6. Objetos Indistinguibles

Como lo mencionamos anteriormente, estos objetos surgen cuándo no sabemos la distinción entre uno y otro, es decir, cuándo al manejarlos solo conocemos que hay esa cantidad de objetos, por lo que hay que tener mucho más cuidado al momento de tratar con estos, porque la intuición fácilmente puede hacernos un engaño y darnos a ver una respuesta que creemos correcta, pero que no necesariamente lo es.

1.2. Espacio Muestral y Experimentos Aleatorios

Cuándo tenemos un experimento del que de antemano no conocemos su respuesta será lo que conoceremos como experimento aleatorio, y aquellos experimentos de los que sí podamos conocer de antemano su solución los llamaremos experimentos deterministas. Los experimentos tanto aleatorios como no aleatorios usualmente tienen un rango de definición, es decir, las posibles respuestas se encuentran albergadas dentro de un conjunto al cuál llamaremos *Espacio Muestral*. Nuestro espacio muestral lo denotaremos por Ω.

Cuando tiramos un dado de 6 caras, no sabemos que número va a salir (esto significa que es un experimento aleatorio), pero sí sabemos que ese número está entre 1 y 6, es decir las posibles respuestas son los naturales de 1 a 6, así $\Omega = \{1,2,3,4,5,6\}$.

Si consideraramos el experimento de lanzar un dado de 6 caras 3 veces, y que la suma de los resultados sea mayor a 8, el espacio muestral sería:

$$\Omega = \{(a,b,c) : a,b,c \in \{1,\cdots,6\}; a+b+c > 8\}$$

1.2.1. Espacio Muestral Discreto

Cuándo se da el caso que la cantidad de elementos de nuestro espacio muestral es un número finito (natural naturalmente) o el espacio muestral tiene la misma cantidad de elementos de los naturales, lo llamaremos *espacio muestral discreto*. Todos los ejemplos que hemos dado hasta el momento han sido de espacios muestrales discretos.

[3] Aquí entramos a un espacio muestral con una cantidad de elementos igual a la de los reales.

1.2.2. Espacio Muestral Continuo

Veamos como sería un experimento aleatorio con un espacio muestral continuo[3]:

Supongamos que una lámpara de calor es sometida a prueba, donde X es la cantidad de luz producida (en lumens), y Y la cantidad de energía calorífica (en joules), se miden ambas cantidades, un espacio muestral adecuado sería el del producto cartesiano de los reales no negativos, es decir:

$$\Omega = \{(x,y) \in \mathbb{R} \times \mathbb{R} : x,y \in [0,\infty)\}$$

Esto dado que tanto el valor de x, como el de y pueden ser cualquier número real no negativo.

1.3. Eventos

Una pregunta muy importante es el preguntarse que cosa es un evento. Definir un evento como un subconjunto de Ω es algo más bien trillado, dado que no todo subconjunto definido en un espacio muestral poseerá complemento[4] y no podremos determinar la probabilidad de que no ocurra cierto evento. He aquí el problema con nuestra definición, por lo que idearemos la idea de $\sigma-$álgebra, para poder dar idea de que es formalmente un evento.

La estructura algebraica "$\sigma-$álgebra"tendrá las siguientes propiedades:

Sea $\mathscr{A}$ una familia de subconjuntos de Ω, sí se cumple que:

- $\Omega \in \mathscr{A}$.
- $A \in \mathscr{A} \Rightarrow A^c \in \mathscr{A}$.
- $\forall n \in \mathbb{N}(A_n \in \mathscr{A}) \Rightarrow \bigcup_{n\in\mathbb{N}} A_n \in \mathscr{A}$.

Decimos que $\mathscr{A}$ es una $\sigma-$álgebra.

Para ser exactos dentro de nuestra definición, tendremos que un evento es aquel elemento de la $\sigma-$álgebra, así tendremos que se forma una dupla con $(\Omega,\mathscr{A})$ al cuál lo llamaremos **Espacio Medible**, y a la tripla $(\Omega,\mathscr{A},\mathbb{P})$ la llamaremos **Espacio de Probabilidad**. De igual forma, el docente nos brinda una definición un poco más "amigable"para entender que significa ser un evento:

El evento es un conjunto que cumple una propiedad dada, y qué provienen del espacio muestral.

Lo cuál es cierto, dado que podemos definir lo que es ser un evento con cierta probabilidad bajo la $\sigma-$álgebra que definamos[5], y no tener problemas bajo los complementos como puede llegar a ocurrir con simplemente Ω.

Más adelante veremos un ejemplo cuándo definamos la función de probabilidad mediante los axiomas de Kolmogorov.

1.4. Axiomas de Kolmogorov

La Probabilidad dentro del programa de Hilbert no tenia una estructura axiomática, fue Kolmogorov el que la dotó de la misma, con eso ya

[4]Me explico, puede darse la ocasión de que el espacio muestral esté definido de esa forma, sin que aparezca el complemento.

[5]Podemos definir muchos tipos de $\sigma-$álgebra, por ejemplo tenemos que $\mathscr{A} = \{\phi,\Omega\}$ es mi $\sigma-$álgebra trivial, mientras que sí consideramos $\mathscr{A} = \mathscr{P}(\Omega)$, esta la llamaremos nuestra $\sigma-$álgebra total, dado que es la mayor $\sigma-$álgebra que contenga los eventos de Ω.

se consideraría una rama formal de la matemática. Los axiomas de Kolmogorov dicen:

Sea Ω un espacio muestral, y $\mathscr{A}$ una σ-álgebra, luego una función $P : \Omega \to \mathbb{R}$ es una función de Probabilidad sobre $(\Omega, \mathscr{A})$, si se cumplen los siguientes axiomas:

1. La probabilidad de cualquier suceso S es no negativa:

   $\forall S \in \mathscr{A}(P(S) \geq 0)$.

2. La probabilidad del suceso seguro es igual a 1:

   $P(\Omega) = 1$.

3. Si $\{S_n\}_{n\in\mathbb{N}} \subseteq \mathscr{A}$ son sucesos mutuamente excluyentes, tenemos que:

$$P\left(\bigcup_{n=1}^{\infty} S_n\right) = \sum_{n=1}^{\infty} P(S_n)$$

Habiendo definido esto, tenemos que la terna $(\Omega, \mathscr{A}, P)$ se le denomina espacio de probabilidad.

Así podemos demostrar varias propiedades de la probabilidad:

- $P(\phi) = 0$.
- Para cualquier suceso S se da $P(S) \leq 1$.
- $P(\Omega/S) = 1 - P(S)$.
- Si $A \subseteq B$ entonces $P(A) \leq P(B)$.
- $P(A/B) = P(A) - P(A \cap B)$.
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Habiendo definido esto, podemos dar el ejemplo que mencionamos antes:

Sea $\Omega = \{1, 2, 3\}, \mathscr{A} = \{\phi, \Omega, \{1\}, \{2, 3\}\}$, definiremos $\mathbb{P}$ como:

$$\mathbb{P}(A) = \begin{cases} 1 & si \quad 3 \in A \\ \\ 0 & si \quad 3 \notin A \end{cases}$$

Vemos que sin importar el evento, todas las posibles probabilidades son mayores o iguales a 0 ✓

Vemos que $1 \in \Omega$, por lo que $\mathbb{P}(\Omega) = 1$✓

Luego podemos probar que cuando los eventos son disjuntos la probabilidad de la unión es la suma de sus probabilidades, y no hay problemas al unirlas y generar el conjunto Ω, es decir, se respeta la suma de probabilidades.

Estas son de las propiedades básicas dentro de la probabilidad, definamos ahora lo que es la probabilidad clásica.

1.5. Probabilidad Clásica

Cuándo los eventos de un espacio muestral tienen la misma posibilidad de ocurrir, la denominaremos probabilidad clásica, un ejemplo base de esto es al lanzar un dado todas las caras tienen exactamente la misma probabilidad de salir, que en uno de 6 caras sería $1/6$ para cada cara. Al lanzar una moneda, la parte de cara, y la parte de sello tiene la misma probabilidad $1/2$ cada una.

1.6. Probabilidad Condicional

Definición: Sean A,B eventos con $\mathbb{P}(B) > 0$. La probabilidad condicional del evento A, dado el evento B, está definida por

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}; \mathbb{P}(B) > 0 \tag{6}$$

La probabilidad condicional cumple las tres propiedades de la defínciión de probabilidad:

1. $0 \leq \mathbb{P}(A|B) \leq 1$
2. $\mathbb{P}(\Omega|B) = \mathbb{P}(B|B) = 1$
3. Si A_1,A_2 son eventos mutuamente excluyentes disjuntos entonces

$$\mathbb{P}(A_1 \cup A_2|B) = \mathbb{P}(A_1|B) + \mathbb{P}(A_2|B)$$

Teorema: Sean A,B eventos, se tiene que:

$$\mathbb{P}(A\cap B) = \mathbb{P}(A)\cdot\mathbb{P}(B|A) = \mathbb{P}(B)\cdot\mathbb{P}(A|B) \tag{7}$$

El cuál lo conoceremos como *Multiplicación de Probabilidades*

Observación:

- El teorema se puede extender a más eventos:

$$\mathbb{P}(A_1\cap\cdots\cap A_n) = \mathbb{P}(A_1)\cdot\mathbb{P}(A_2|A_1)$$
$$\cdot\mathbb{P}(A_3|A_2\cap A_1)\cdots\mathbb{P}(A_n|A_1\cap\cdots\cap A_{n-1})$$

- Esta formula es útil en problemas que involucran el muestreo sin reemplazamiento.

Teorema de la Probabilidad Total. *Sí los eventos $B_1,\cdots,B_n$ constituyen una partición del espacio muestral Ω de tal forma que $\mathbb{P}(B_i) \neq 0$ para $i = 1,\cdots,k$. Entonces para cualquier evento $A \subset S$ tenemos:*

$$\mathbb{P}(A) = \sum_{i=1}^{k}\mathbb{P}(B_i\cap A) = \sum_{i=1}^{k}\mathbb{P}(B_i)\mathbb{P}(A|B_i) \tag{8}$$

Demostración. Sea ξ el experimento dado, Ω el espacio muestral asociado a ξ. Sean $B_1,B_2,\cdots B_n$ una partición de Ω, esto significa: $B_i\cap B_j = \phi\forall i\neq j$, y $\bigcup_{i=1}^{k} B_i = \Omega$ y tal que $\mathbb{P}(B_i) \neq 0\forall i = 1,\cdots,k$.

Sea A un evento de Ω con la ayuda del diagrama de Venn

$A = (B_1\cap A)\cup\cdots\cup(B_k\cap A)$, las parcelas así dadas son disjuntas dos a dos y la unión es el evento A, así tomando probabilidad en ambos lados tenemos:

$$\mathbb{P}(A) = \mathbb{P}[(B_1\cap A)\cup\cdots\cup(B_k\cap A)]$$
$$\mathbb{P}(B_1\cap A)+\cdots+\mathbb{P}(B_k\cap A)$$
$$\sum_{i=1}^{k}\mathbb{P}(B_i\cap A) = \sum_{i=1}^{k}\mathbb{P}(B_i)\mathbb{P}(A|B_i)$$

□

Teorema de Bayes. *Sea Ω el espacio muestral, sí los eventos $B_1,\cdots B_k$ constituyen una partición de Ω dónde $\mathbb{P}(B_i)\neq 0\forall i = 1,\cdots,k$ entonces cualquier evento A en Ω es tal que $\mathbb{P}(A) \neq 0$*

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(B_j)\mathbb{P}(A|B_j)}{\mathbb{P}(A)} = \frac{\mathbb{P}(B_j)\mathbb{P}(A|B_j)}{\sum_{i=1}^{n}\mathbb{P}(B_i)\mathbb{P}(A|B_i)} \tag{9}$$

La demostración de este teorema es inmediata al teorema de probabilidad total.

Veamos ahora lo que son los eventos independientes:

1.7. Eventos Independientes

Definición: Dos eventos A, B son llamados independientes si

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$$

De lo contrario se dice que A, B son eventos dependientes.

Teorema: Si A, B son eventos tales que $\mathbb{P}(A) > 0$ y $\mathbb{P}(B) > 0$ entonces A, B son independientes sii cualquiera de las siguientes afirmaciones se cumple:

$$\mathbb{P}(A|B) = \mathbb{P}(A) \qquad \mathbb{P}(B|A) = \mathbb{P}(B)$$

Teorema: Dos eventos A, B se dicen independientes sii los siguientes pares de eventos también lo son:

1. A y B^c.
2. A^c y B.
3. A^c y B^c.

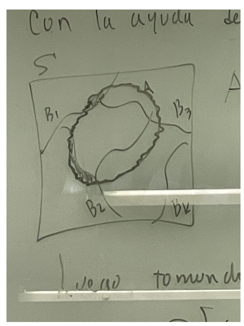

Figura 1: Diagramas de Venn del Teorema de Probabilidad Total

1.8. Variables Aleatorias

Una variable aleatoria es una función que determinaremos por X, dónde llevará elementos del espacio muestral en el conjunto de los reales. Es decir:

$$X : \Omega \to \mathbb{R} \tag{10}$$

$$\omega \mapsto X(\omega) = x \in \mathbb{R}$$

Definimos auxiliarmente que que el rango de esta función es un subconjunto de $\mathbb{R}$. Es decir podemos definir la función biyectiva como:

$$X : \Omega \to \mathbb{D} \subseteq \mathbb{R}$$

Ejemplo:

Considere el lanzamiento de dos monedas legales, defina su espacio muestral, y la variable aleatoria X, es contar el número de caras obtenidas en las dos monedas, así $\Omega = \{(c,c),(s,s),(s,c),(c,s)\}$, dónde c significa cara, y s significa sello.

Así podemos determinar que:

$$X(\{(c,c)\} = 2$$

Dado que salieron 2 caras.

$$X(\{(c,s)\}) = 1 = X(\{(s,c)\})$$

Dado que salió 1 cara.

$$X(\{(s,s)\}) = 0$$

Ahora nos podemos preguntar por:

$$\mathbb{P}[X(\{(c,c)\} = 2] = \mathbb{P}[X = 2] = \frac{1}{4}$$

ASí también podemos preguntarnos por:

$$\mathbb{P}[X = 1] = \frac{1}{2}$$

$$\mathbb{P}[X = 0] = \frac{1}{2}$$

Y podemos evidenciar que la suma de todas las probabilidades es 1, basta con sumar $1/4 + 1/2 + 1/4 = 1$. Cumpliendo de este modo con los axiomas de Kolmogorov[6].

Ejemplo: Considerese un dado de cuatro caras equilibrado. Un jugador arroja el dado dos veces y la puntuación es el máximo de dos números de puntos qie sañen. Aquí, en cada cara se encuentra un número de puntos, digamos $\{1,2,3,4\}$. Definimos una variable aleatoria, en particular

$$e = (i,j); i,j = 1,2,3,4, X(e) = \text{máx}(i,j)$$

$$S = \{(1,1),(1,2),(1,3),\cdots,(4,4)\}$$

Luego podemos generar que mi evento B_i sea: Que la segunda componente, el máximo sea i. Así tendremos 4 opciones:

$$B_4 : (1,4),(2,4),(3,4),(4,4)$$

$$B_3 : (1,3),(2,3),(3,3),(4,3)$$

$$B_2 : (1,2),(2,2),(3,2),(4,2)$$

$$B_1 : (1,1),(2,1),(3,1),(4,1)$$

Así podemos obtener que si X cuenta que el número sea el máximo entre las componentes tendrmeos:

$\mathbb{P}(X = 2) = \frac{3}{16}, \mathbb{P}(X = 3) = \frac{5}{16}, \mathbb{P}(X = 1) = \frac{1}{16}, \mathbb{P}(X = 4) = \frac{7}{16}$.

1.9. Variables Aleatorias Discretas

Definición: Si el conjunto de todos los posibles valores de la V.A.[7] X es un conjunto contable $x_1, x_2, \cdots$ entonces X es llamada V.A. Discreta, y la función

$$\mathbb{P}[X = x], \text{para } x = x_1, x_2, \cdots$$

Que asigna probabilidades a cada valor posible x es llamada función de masa de probabilidad.

[6] La clave se encuentra en definir de manera exitosa la variable aleatoria X, para efectividad en el curso.

[7] Variable Aleatoria.

1.9.1. Varios Teoremas de Probabilidad

Teorema. *Una función $f(x)$ es un f.m.p*[8] *si y solo si satisface las siguientes dos propiedades para un conjunto de núemros reales infinito contable $x_1, x_2, \cdots$.*

1. $f(x_i) \geq 0$, *para todo* x_i.
2. $\sum_{x_i} f(x_i) = 1$.

Demostración. 1 es inmediata de los axiomas.

Para 2: Para eventos $[X = x_1], [X = x_2], \cdots$, constituyen una partición del espacio muestral. Luego $\sum_{x_i} f(x_i) = \sum_{x_i} \mathbb{P}[X = x_i]$, y todo esto llega a ser el espacio muestral, dado que es una partición del mismo y su probabilidad es 1, por lo que establece la igualdad. □

Función de Distribución Acumulada. *Sea X un V.A. la función F_x definida sobre $\mathbb{R}$ por medio de:*

$$F_x(x) := \mathbb{P}[X \leq x] \tag{11}$$

Se llama función de distribución (ó función de distribución acumulada) de la V.A. X.

Ejemplo:

Suponga que se lanza una moneda corriente 3 veces consecutivas, y sea X la V.A. X : El número de caras obtenidas. Encuentre la función de distribución acumulada de la V.A.

Sabemos que $|S| = 8$, así podemos calcular las probabilidades:

- $$\mathbb{P}[X = 3] = \frac{1}{8}$$

- $$\mathbb{P}[X = 2] = \mathbb{P}[X = 1] = \frac{3}{8}$$

- $$\mathbb{P}[X = 0] = \frac{1}{8}$$

Así podemos determinar que la función de distribución acumulada es:

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

La relación entre $F(x), f(x)$ para una función de distribución discreta está dada por el siguiente resultado.

Teorema. *Sea X una V.A. discreta con f.m.p $f(x)$ y función de distribución acumulada (F.D.A) $F(x)$. Si los posibles valores que puede tomar la V.A. X son indexados en forma creciente $x_1 < x_2 < x_3$ entonces $f(x_1) = F(x_1)$ y para cualquier $i > 1$ $f(x_i) = F(x_i) - F(x_{i-1})$.*

Además si $x < x_1$ entonces $F(x) = 0$, y para cualquier otro real x.

$$F(x) = \sum_{x_i \leq x} f(x_i) \tag{12}$$

Dónde la suma se toma sobre todos los i tales que $x_i < x$.

Teorema. *La función $F(x)$ de alguna V.A. X satisface las siguientes propiedades.*

[8] Función de masa de probabilidad.

1. *Para todo $a, b \in \mathbb{R}$, con $a < b$ entonces*

$$F_X(a) \leq F_X(b)$$

Esto nos dice que $F(x)$ es no decreciente.

2.
$$\lim_{x \to -\infty} F(x) = 0$$

3.
$$\lim_{x \to \infty} F(x) = 1$$

4.
$$\lim_{x \downarrow x_0} F_i(x) = F(x_0)$$

5.
$$\lim_{h \to 0^+} F(x+h) = F(x)$$

Demostración.

2 Tenemos que para cualquier $n \in \mathbb{N}$ se satisafce que:

$$\{X \leq -n\} \supset \{X \leq -(n+1)\} \supset \{X \leq -(n+2)\}$$

Y además:

$$\bigcap_{n=1}^{\infty} \{X \leq -n\} = \phi$$

Por tanto tendremos que:

$$\lim_{x \to -\infty} F_X(x) = \lim_{x \to -\infty} F(-n) = \lim_{x \to -\infty} \mathbb{P}(X \leq -n)$$

$$= \lim_{x \to -\infty} \mathbb{P}(\phi) = \mathbb{P}(\phi) = 0.$$

$\square$

Teorema. *Sea X una V.A. real, F_X es una F.D.A, y $a, b \in \mathbb{R}$ con $a < b$, entonces:*

1. $\mathbb{P}[a < X \leq b] = \mathbb{P}_X[X \leq b] - \mathbb{P}[X \leq a] = F_X(b) - F_X(a)$.
2. $\mathbb{P}[X = x] = F_X(x) - F_X(x^-)$.

Demostración. Note que:

$$(-\infty < X < b] = (-\infty < X < a] \cup (a < X < b]$$

Así sacando probabilidades, y aplicando definición de acumulada obtenemos que:

$$F_X(b) = F_x(a) + \mathbb{P}(a < X \leq b)$$

Obteniendo así lo que deseamos. $\square$

Teorema. *Con las hipótesis del teorema inmediatamente anterior se cumplen las siguientes propiedades:*

1.
$$F(x^-) := \lim_{h \to 0^-} F(x-h) = \mathbb{P}(X < x)$$

2.
$$\mathbb{P}[a \leq X \leq b] = F(b) - F(a^-)$$

3.
$$\mathbb{P}[a \leq X < b] = F(b^-) - F(a^-)$$

4.
$$\mathbb{P}[a < X \leq b] = F(b) - F(a)$$

5.
$$\mathbb{P}[a < X < b] = F(b^-) - F(a)$$

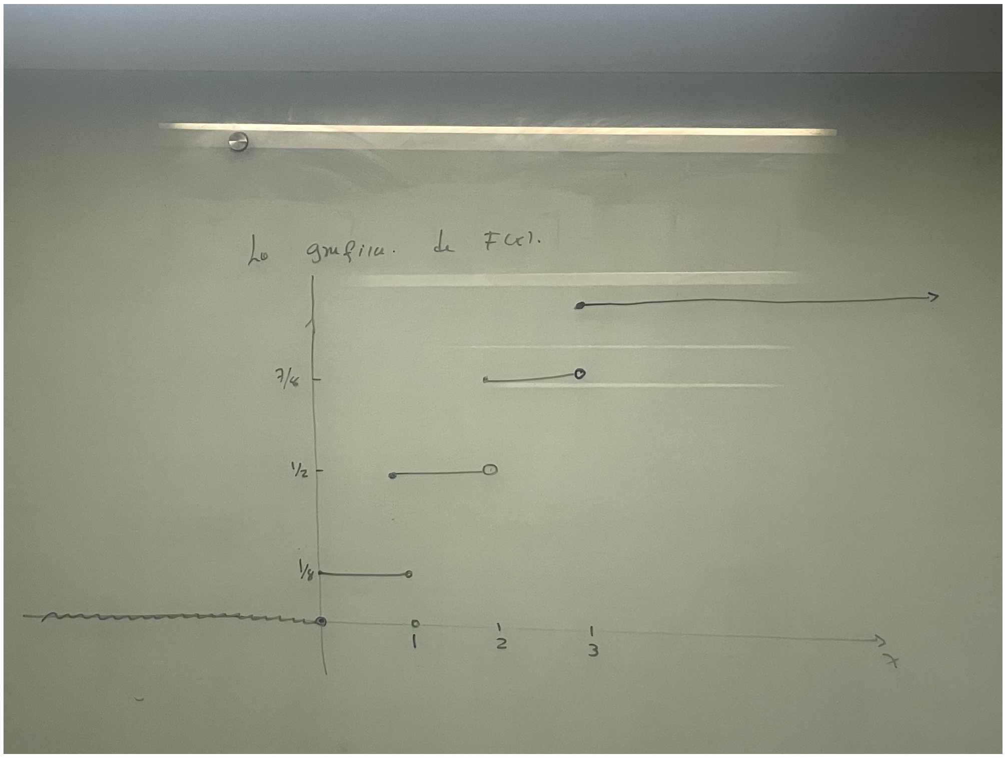

Figura 2: Ejemplo de Función Acumulada

6. *Si $\mathbb{P}(a < X < b) = 0$ entonces $F_X(x)$ es constante en el intervalo (a, b).*

Las funciones de distribución escalonadas tienen discontinuidades que son llamadas saltos.

Definición. *Sea X una V.A. y F su F.D.A. Se dice que F presenta un salto en un punto $a \in \mathbb{R}$ si*

$$F(a) - F(a^-) \neq 0$$

La diferencia $F(a) - F(a^-)$ se llama longitud de salto y se tiene que es igual a $\mathbb{P}(X = a)$.

2. Parcial 2

2.1. Variable Aleatoria Continua, y Propiedades

Definición Se dice que una v.a X es de tipo discreto si su espacio muestral tiene cardinal finito, o infinito numerable.

Definición .Se dice que una v.a X es de tipo continuo si su función de distribución acumulada ($F(x)$) es una función continua para todo $x \in \mathbb{R}$.

Observaciones:

1. Si X es una v.a. discreta, con $f(x)$ como f.m.p tenemos que se cumplen las siguientes condiciones:

   - $f(x_i) \geq 0, \forall x \in \mathbb{R}$.
   - $$\sum_{x \in S} f(x) = 1$$

2. Si la v.a. X es discreta con F.D.A $F_X(x)$, entonces el número de saltos de $F_X(x)$ es a lo sumo numerable.

3. El soporte de la v.a. X discreta es el espacio de puntos de X que tiene probabilidad positiva.

4. Si la v.a. X es de tipo continuo entonces el número de saltos de la función $F(x)$ es nulo.

5. La v.a. X es absolutamente continua si y solo si una función real no negativa e integrable f_α tal que para todo $x \in \mathbb{R}$ satisface que:

   $$F_X(x) = \int_{-\infty}^{x} f_\alpha(t)dt \tag{13}$$

   La función f_α recibe el nombre de función densidad de probabilidad (f.d.p) de la v.a. X, y es Riemann-Integrable.

6. El soporte de la v.a. X continua consiste en todos los puntos x tales que $f_\alpha(x) \geq 0$.

   Dada $f_\alpha(x)$ es continua sobre el soporte de la v.a. X u $F_X(\infty) = 1$, la f.d.p. satisface:

   - $$F_X(x) \geq 0$$
     .
   - $$\int_{-\infty}^{\infty} f_\alpha(x)dx = 1 \tag{14}$$
   - El soporte de X tiene a lo sumo un número finito de discontinuidades en todo intervalo finito que sea subconjunto del conjunto soporte.

7. Si $a, b \in \mathbb{R}$ entonces:

   $$\mathbb{P}[a < X < b] = F(b) - F(a)$$

   Luego obtenemos que lo que tenemos es igual a:

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

   Y todo esto es igual a:

   $$\mathbb{P}[a \leq X \leq b]$$

La cuál al tiempo es equivalente a:

$$\mathbb{P}[X \leq b] - \mathbb{P}[X \leq a]$$

8. Del Teorema Fundamental del Cálculo tenemos:

$$\frac{d}{dx}F(x) = f(x) \Rightarrow F'(x) = f(x) \quad (15)$$

Esto nos permite concluir que a partir de alguna de las dos funciones podemos obtener la otra, si tenemos la de distribución, derivando obtenemos la de densidad, y si tenemos la de densidad, intengrando entre menos infinito y x, obtenemos la función de distribución.

2.2. Valor Esperado de una Variable Aleatoria

Definición. *Sea X una v.a. definida sobre el espacio de probabilidad* $(\Omega, \mathscr{A}, \mathbb{P})$.

1. *Si X es una v.a. de tipo discreto con f.m.p.* $P(x) = \mathbb{P}[X = x]$, *y con valores para X con* $x_1, x_2, \cdots$ *y si:*

$$\sum_{k=2}^{\infty} |x_k|\mathbb{P}[X = x] < \infty \quad (16)$$

entonces la esperanza de X es:

$$\mathbb{E}(x) = \sum_{k=1}^{\infty} x_k \mathbb{P}[X = x] \text{ si la suma existe} \quad (17)$$

2. *Si X es una v.a. de tipo continuo y f.d.p.* $f(x)$, *y si:*

$$\int_{-\infty}^{\infty} |x| f(x) dx < \infty \quad (18)$$

Entonces la esperanza de X es:

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x) dx \quad (19)$$

1. Observación: Por notación se tiene que la esperanza matemática de una v.a. X se denota por:

$$\mu = \mathbb{E}(X) = \mu_X \quad (20)$$

2. Si X es una v.a. que tome solo un númeor finito de valores, entonces $\mathbb{E}(X)$ existe.

La esperanza podemos interpretarla como el promedio de todos los datos que poseemos en un experimento. Haciendo una analogía con Física, es buscar el punto de equilibrio de un objeto cualquiera.

Ejemplo: Un experimento consiste en clasificar 3 objetos como defectuosos (D) o no defectuosos (N). Suponga que su v.a. X es el número de objetos defectuosos.

- Escriba el espacio muestral más adecuado para el experimento.

$$\Omega = \{DDD, DDN, DND, NDD, DNN, DND$$

$$, NND, NNN\}$$

- Suponga que todos los posibles resultados son igual de posibles. Determine el valor esperado de la v.a. X.

$$\mathbb{E}(X)=\sum_{x} xf(x)$$

$$\mathbb{E}(X)=0\cdot\frac{1}{8}+1\cdot\frac{3}{8}+2\cdot\frac{3}{8}+3\cdot\frac{1}{8}=\frac{12}{8}=1{,}5$$

Luego, podemos interpretar que de los 3 objetos que tomamos, 1.5 están defectuosas [9], pero esto físicamente no trae sentido, por lo que diremos que en promedio de los 3 objetos que tomamos, 2 son defectuosos.

Los resultados nos tiene que dar en las mismas dimensiones que trae el conteo.

Teorema. *Sea X una v.a. y $g:\mathbb{R}\to\mathbb{R}$, una función tal que $Y=g(X)$ es una v.a., entonces:*

$$\mathbb{E}(Y)=\sum_{x\in\Omega} g(x)p(x) \text{ si la v.a. } X \text{ es discreta}$$

$$\mathbb{E}(Y)=\int_{-\infty}^{\infty} gx)f(x)dx \text{ si } X \text{ es v.a. continua}$$

dónde $p(x),f(x)$ son la función masa de probabilidad, y función de densidad de probabilidad de X respectivamente.

Demos algunas propiedades del valor absoluto:

Teorema:

Sea X una v.a. real:

1. Si $\mathbb{P}(X\geq 0)=1$, y $\mathbb{E}(X)$ existe entonces $\mathbb{E}(X)\geq 0$.
2. $\mathbb{E}(k)=k, k\in\mathbb{R}$.
3. Si X es acotada, esto es, si existe una constante real $M>0$, tal que $\mathbb{P}(|X|\leq M)=1$, entonces $\mathbb{E}(M)$ existe.
4. Si k_1,k_2 son constantes y $g_1(x),g_2(x)$ funciones tales que $g_1(X),g_2(X)$ son v.a. cuyos valores esperados existen entonces:

$$\mathbb{E}[k_1g_1(X)+k_2g_2(X)]=k_1\mathbb{E}[g_1(X)]+k_2\mathbb{E}[g_2(X)]$$

5. Si g_1,g_2 son funciones tales que $g_1(X),g_2(X)$ son v.a. cuyos valores esperados existen y $g_1(X)\leq g_2(X)$, entonces:

$$\mathbb{E}(g_1(X))\leq\mathbb{E}(g_2(X))$$

   En particular se tiene que: $|\mathbb{E}(X)|\leq\mathbb{E}(|X|)$.

6. Sea X una v.a. discreta con f.m.p. $p(x)$ Si el soporte de X es $\{a_1,a_2,\cdots\}$ se sigue que:

$$\mathbb{E}(X)=a_1\mathbb{P}(X=x_1)+a_2\mathbb{P}(X=x_2)+\cdots$$

De igual modo tenemos que se cumple que:

$$\mathbb{E}(XY)=\mathbb{E}(X)\cdot\mathbb{E}(Y) \tag{21}$$

Si y solo si son variablea aleatorias independientes. A la esperanza en general la llamaremos μ, es decir $\mathbb{E}(X)=\mu$, además tenemos que no se cumple la siguiente propiedad:

$$\mathbb{E}\left(\frac{1}{X}\right)\neq\frac{1}{\mathbb{E}(X)}$$

[9] Pero esto es un significado matemático, por lo que hay que traer a colación el significado físico.

Definiremos de igual modo la varianza de una variable aleatoria con base en su esperanza, la cuál es:

$$\text{Var }(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2$$

En contraste con la esperanza, la varianza no es un operador lineal, pero podemos determinar las siguientes propiedades:

1. Var $(X) \geq 0$.
2. Var $(a) = 0, \forall a \in \mathbb{R}$.
3. Var $(aX) = a^2$Var (X).

De este modo, al haber determinado estas propiedades, podemos pasar a un tema de aproximaciones de las mismas.

2.2.1. Aproximación de la Media y la Varianza

Sea X una v.a. y $H(X)$ una función de la v.a. la cuál se puede extender en series de Taylor. Una expresión para la aproximación de la media y la varianza de la v.a. X se puede conseguir por medio de esa extensión en series.

Sea $H(x)$ una función real, y $H(X)$ una v.a. que sea de clase $C^\infty(I)$, es decir ser **infinitamente derivable**, dónde I es un abierto, y $\mu \in I(\mu = \mathbb{E}(X)$, la función $H(x)$ tiene una aproximación de Taylor alrededor de μ:

$$H(x) = H(\mu) + H'(\mu)(x-\mu) + \frac{H''(\mu)}{2!}(x-\mu)^2 \tag{22}$$

El cuál conocemos como el 2do polinomio de Taylor, así la esperanza de la v.a. $H(X)$ sería:

$$\begin{aligned}\mathbb{E}(X) &= \mathbb{E}(H(\mu)) + \mathbb{E}(H'(\mu)(X-\mu)) \\ &+ \frac{1}{2!}\mathbb{E}(H''(\mu)(X-\mu)^2) \\ &= H(\mu) + \frac{H''(\mu)}{2}\sigma_X^2\end{aligned}$$

Pero en el caso de la varianza, existe un problema, dado que al obtener la segunda derivada llegamos a algo llamado la asimetría, por lo que no usaremos la cuadratura de la función, sino la linealización, que es usar netamente la primera derivada, así nos queda:

$$\begin{aligned}\text{Var }(H(X)) &= \text{Var }(H(\mu) + H'(\mu)(X-\mu)) \\ &= (H'(\mu))^2\text{Var }(X) = (H'(\mu))^2\sigma_X^2\end{aligned} \tag{23}$$

Con esto dicho, hemos terminado el tema de las variables aleatorias, introduciremos ahora unos resultados de Probabilidad, para llevar a cabo más problemas de probabilidad, y empezar a dar resultados de aproximar procesos y como resolver problemas de forma mucho más fácil y simple. Es el tema de las distribuciones de probabilidad, a lo largo del estudio del tema, se han dado cuenta que hay experimentos que cumplen cierta característica peculiar y se han catalogado, las distribuciones podemos verlas como continuas o discretas, y usaremos lo visto anteriormente para llevarlas a cabo de forma efectiva.

Como hemos expresado anteriormente clasificaremos inicialmente las distribuciones entre discretas y continuas, daremos varios conceptos que ayudan a entender el tema probabilístico de distribuciones de una mejor manera.

2.3. Distribuciones de Probabilidad: Discretas

Lo mejor es siempre empezar con la distribución que parece ser la más normal[10] en el sentido intuitivo, y es la **Distribución Uniforme Discreta**[11].

2.3.1. Distribución Uniforme

La cuál viene siendo un tipo de probabilidad dónde todos los casos tienen la misma probabilidad de ocurrir, es decir la probabilidad clásica que hemos venido trabajando. Denotaremos a N como el tamaño de nuestra población, más formalmente sería:

$$\mathbb{P}(x) = \begin{cases} \dfrac{1}{N} & si \quad x = x_1, x_2, \cdots \\ \\ 0 & e.o.p \end{cases} \tag{24}$$

Cuándo X tiene una distribución uniforme discreta, la denotaremos como: $X \sim UNIF(N)$, dado que solo depende del tamaño de la población.

Teorema: Sea $X \sim UNIF(N)$, tenemos que:

- $\mathbb{E}(X) = \dfrac{1}{N} \sum_{i=1}^{N} x_i$.
- Var $(X) = \dfrac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2$.

2.3.2. Procesos de Bernoulli

Definiremos aquí lo que se llaman **Procesos de Bernoulli**, para entender lo que realizaremos a continuación Un proceso con las siguientes características se denominará **proceso de Bernoulli**:

1. El experimento consiste en realizar n intentos repetidos.
2. Los resultados de cada uno de los intentos puede clasificarse como éxito ó fracaso.
3. La probabilidad de éxito, representada por p permanece constante en cada uno de los intentos.
4. Los eventos repetidos son independientes.

2.3.3. Distribución Bernoulli

La distribución Bernoulli consiste en realizar un experimento una sola vez, y observar si es cierto el evento o no, siendo p la probabilidad de éxito, y $q = 1 - p$ la probabilidad de fracaso.

Definición: Una v.a. X tiene distribución Bernoulli, si y solo si su f.m.p estpa dada por:

$$\mathbb{P}(x) = \begin{cases} p^x (1-p)^{1-x} & si \quad x = 0, 1 \\ \\ 0 & e.o.p \end{cases}$$

Si se cumplen todas estas condiciones, se dice que $X \sim Ber(p)$.

[10] Aclaremos que no es la distribución normal :v.

[11] Se aclara lo discreto, dado que en las continuas existe una de este tipo también.

Tenemos los siguientes teoremas acerca de la Bernoulli:

Teorema: Sea $X \sim Ber(p)$, luego se da que:

1. $\mathbb{E}(X) = p$.
2. $\text{Var}\,(X) = p - p^2 = pq = p(1-p)$.

2.3.4. Distribución Binomial

Sea X una v.a. que denota el número de éxitos en n ensayos de Bernoulli. Entonces X tiene una f.m.p Binomial de la siguiente forma:

$$\mathbb{P}(n,p) = Bin(n,p) = \begin{cases} \binom{n}{x} & si \quad x = 0,1,\cdots,n \\ \\ 0 & e.o.p \end{cases} \tag{25}$$

Para demostrar que es una función de probabilidad, se sugiere usar el Binomio de Newton.

Al cumplirse estas condiciones, decimos que $X \sim Bin(x,n,p)$.

Teorema: La media y la varianza de la v.a. $X \sim Bin(n,p)$ es:

1. $\mathbb{E}(X) = np$.
2. $\text{Var}\,(X) = np(1-p)$.

2.3.5. Experimento Hipergeométrico

Un experimento hipergeométrico es aquel que posee las siguientes características:

1. Una muestra aleatoria de tamaño n se selecciona sin reemplazo de un total de N resultados ó artículos posibles.
2. k resultados ó artículos del total N, pueden clasificarse como N éxitos y $N-k$ fracasos.

De este modo podemos empezar a definir lo que sería nuestra distribución hipergeométrica.

2.3.6. Distribución Hipergeométrica

Definición: La distribución de probabilidad de la v.a. X es hipergeométrica sí su f.m.p es:

$$Hip(N,n,k) = \begin{cases} \dfrac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} & si \quad x \in R \\ \\ 0 & e.o.p \end{cases} \tag{26}$$

Dónde $R := [\text{máx}\{0, n-(N-k), \text{mín}\{k,n\}]$

Teorema: La media y la varianza de la v.a X hipergeométrica es:

- $\mathbb{E}(X) = n\dfrac{k}{N}$.
- $\text{Var}\,(X) = \dfrac{N-n}{N-1} \cdot n \cdot \dfrac{k}{N} \cdot \left(1 - \dfrac{k}{N}\right)$.

2.3.7. Distribución Binomial Negativa

Definición: Si repetimos intentos/ensayos independientes pueden dar como resultado un éxito con probabilidad p, y un fracaso con probabilidad $1-p$, entonces la distribución de probabilidad de la v.a. X definida como el número de ensayos en el cuál ocurre el k-ésimo éxito es:

$$BN(k,p) = \begin{cases} \binom{x-1}{k-1} p^k q^{x-k} & si \quad x = k, k+1, \cdots \\ \\ 0 & e.o.p \end{cases} \tag{27}$$

Teorema: Sea $X \sim BN(k,p)$ luego su esperanza y varianza será:

1. $\mathbb{E}(X) = \dfrac{k}{p}$.

2. Var $(X) = \dfrac{k(1-p)}{p^2}$.

Entre esta distribución, y la que viene existe una relación, tal y como existe entre la Binomial y la Bernoulli.

2.3.8. Distribución Geométrica

Dadas las hipótesis de la Binomial Negativa, y definiendo la v.a. X como el número de intentos en el cuál ocurre el 1^{er} éxito:

$$G(p) = \begin{cases} p(1-p)^{x-1} & si \quad x = 1,2,\cdots \\ \\ 0 & e.o.p \end{cases} \tag{28}$$

Decimos que $X \sim Geo(p)$.

Teorema: La esperanza y la varianza de una v.a. que distribuye Geométricamente es:

1. $\mathbb{E}(X) = \dfrac{1}{p}$.

2. $\text{Var}(X) = \dfrac{1-p}{p^2}$.

Llegó el momento de la distribución más bonita en palabras de nuestro profesor, dado que las utilidades, y usos son variados, y se diferencia del resto por el efecto que trae.

2.3.9. Distribución Poisson

La distribución de una v.a. X Poisson que representa el número de resultados que ocurren en un intervalo de tiempo dado, ó región específica indicada por la letra t, es:

$$Poi(\lambda t) = \begin{cases} \dfrac{e^{-\lambda t}(\lambda t)^x}{x!} & si \quad x = 0,1,2,\cdots \\ \\ 0 & e.o.p \end{cases} \tag{29}$$

Dónde λ es el número promedio de resultados por unidad de tiempo o región.

Teorema: La Media y la Varianza de la Distribución Poisson es λt para ambas.

Con esto damos terminadas las distribuciones discretas, sigue ver ahora como será la relación con las continuas.

2.4. Distribuciones de Probabilidad: Continuas

Dentro de las distribuciones continuas, resaltaremos en este punto, una de las más importantes dentro del estudio de fenómenos probabilísticos, la distribución normal, que al verla aparentemente no tiene nada de normal pero tiene una de las propiedades más bonitas.

2.4.1. Distribución Normal

La distribución Normal es también llamada Distribución Gaussiana. La distribución normal depende de 2 parámetros, μ, σ^2, la cuáles son la media y la varianza. Si X es una v.a. con distribución normal, lo denotaremos como: $X \sim N(\mu, \sigma^2)$. En el caso de que $\mu = 0$, y de que $\sigma^2 = 1$ decimos que nuestra distribución es normal estándar, y la variable aleatoria Z se denotaría: $Z \sim N(0,1)$.

Definición: La función de densidad de probabilidad de una v.a. X Normal con media μ y varianza σ^2 es dada por la ecuación:

$$f(\mu,\sigma^2) = \begin{cases} \dfrac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} & si \quad x \in \mathbb{R} \\ 0 & e.o.p \end{cases} \tag{30}$$

Dada la fórmula, pasaremos a dar varias propiedades sobre la curva normal.

Propiedades de la Curva Normal:

1. El punto más alto se da en la media.

2. La curva es simétrica alrededor de su eje vertical.

3. La curva tiene sus puntos de inflexión en $x = \mu \pm \sigma$, y es cóncava hacia abajo en el intervalo $[\mu - \sigma, \mu + \sigma]$, y es cóncava hacia arriba en otra parte.

4. La curva normal se acerca a un eje horizontal en forma asintótica en cualquiera de las dos direcciones alejándose de la media.

5. El área total bajo la curva y sobre el eje horizontal es igual a 1.

En este caso encontrar probabilidades es encontrar **áreas bajo la curva**. Así al hallar la probabilidad entre dos puntos, será equivalente a que encuentres el área entre los dos puntos, es decir, recurrimos a la teoría de integración.

Muchas veces resolver estas integrales es muy difícil, por lo que recurrimos a las tablas de probabilidades, en este caso de la distribución normal, para poder usar su tabla debemos normalizarla, es decir, crear una nueva v.a. tal que esa v.a. que llamaremos Z, sea del siguiente modo: $Z \sim N(0,1)$, para lograr esta variable aleatoria usaremos la siguiente fórmula:

$$Z = \frac{X - \mu}{\sigma}$$

Es decir que con esto definido podemos establecer que la probabilidad $\mathbb{P}(x_1 \leq X \leq x_2) = \mathbb{P}\left(\dfrac{x_1 - \mu}{\sigma} \leq \dfrac{X - \mu}{\sigma} \leq \dfrac{x_2 - \mu}{\sigma}\right) = \mathbb{P}(z_1 \leq Z \leq z_2)$, dónde $z_i = \dfrac{x_i - \mu}{\sigma}$.

Daremos unas observaciones usando la simetría y las propiedades de la distribución normal:

1. La función de densidad que llamaremos $\phi(z)$ definida por:

   $$\phi(z) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt$$

   Tiene la propiedad de que $\phi(-z) = 1 - \phi(z)$, esto es demasiado útil dado que en ocasiones en la tabla nos pedirán probabilidades negativas[12] y con este concepto queda todo arreglado.

2. Si $X \sim N(\mu, \sigma^2)$ con $\sigma > 0$ entonces $Z = \dfrac{X - \mu}{\sigma} \sim N(0,1)$.

[12] Esto debido a la transformación realizada para normalizarla.

3. Si Z es normal estándar entonces $X = \sigma Z + \mu \sim N(\mu, \sigma^2)$.

Teorema: Sea $X \sim N(\mu, \sigma^2)$, luego:

1. $\mathbb{E}(X) = \mu$.
2. $\mathrm{Var}(X) = \sigma^2$

Consideremos ahora un caso de problemas que parecen continuos, pero no necesariamente lo son, y como usar la normal para resolverlos.

2.4.2. Corrección por Continuidad

Cuándo estamos en las variables aleatorias continuas, muchas veces nos interesaría preguntarnos por una probabilidad puntual, cosa que es imposible (inicialmente) dado que las continuas trabajan solo con intervalos, y la probabilidad puntual da 0[13]. Es en este caso que buscaremos un **factor de corrección** para poder llevar a cabo lo que es la probabilidad de forma correcta, este factor de correción, nos permitirá acercarnos al punto en una vecindad dónde la probabilidad tome un sentido real, en general nuestro factor de correción lo tomaremos como 0.5 para todos los casos, y con esto, generamos un intervalo, dónde sí queremos hallar la probabilidad de algún punto, sea calcular realmente $\mathbb{P}(p - 0{,}5 \leq X \leq p + 0{,}5)$, luego si queremos hallar las probabilidades con nuestra tabla de probabilidades, debemos normalizar, y con esto se usa la fórmula de: $z_1 = \dfrac{x_1 - 0{,}5 - \mu}{\sigma}, z_2 = \dfrac{x_2 + 0{,}5 - \mu}{\sigma}$, y con eso ya es usar la tabla común y corriente.

[13] Esto en virtud de que la integral de un punto hasta el mismo, es 0 siempre.

2.4.3. Aproximación de la Distribución Binomial a la Normal

El siguiente resultado que permite utilizar áreas bajo la curva normal para aproximar propiedades binomiales cuando n es demasiado grande.

Teorema: Sea X una v.a. con media $\mu = np$ y varianza $\sigma^2 = np(1-p)$ entonces la forma límite de la distribución de

$$Z = \frac{X - np}{\sqrt{np(1-p)}}$$

Conforme $n \to \infty$, es la distribución normal con media 0, y varianza 1.

Con esto damos por terminado las propiedades de la función normal.

Veamos ahora más distribuciones continuas.

2.4.4. Distribución Uniforme Continua

Definición: La v.a. continua X está distribuida uniforme sobre el intervalo $[a,b], a < b; a, b \in \mathbb{R}$ si su f.d.p es:

$$f(x) = \begin{cases} \dfrac{1}{b-a} & si \quad x \in [a,b] \\ 0 & e.o.p \end{cases} \tag{31}$$

Incluyamos otro teorema:

Teorema: La media y la varianza de la v.a $X \sim UNIF_c([a,b])$ son:

1. $\mathbb{E}(X) = \dfrac{a+b}{2}$.
2. $\mathrm{Var}(X) = \dfrac{(b-a)^2}{12}$.

Y además en este caso podemos decir quién es su función acumulada, la cuál es:

$$F(x) = \begin{cases} 0 & si \quad x < a \\ \\ \dfrac{x-a}{b-a} & si \quad a \leq x \leq b \\ \\ 1 & \quad x \geq b \end{cases} \tag{32}$$

2.4.5. Distribución Exponencial

Definición: La v.a. X tiene una distribución exponencial con parámetro λ, si su f.d.p está dada por:

$$f_X(\lambda) = \begin{cases} \lambda e^{-\lambda x} & si \quad x \geq 0 \\ \\ 0 & \quad e.o.p \end{cases} \tag{33}$$

Dónde λ es un parámetro llamado parámetro de escala. Tenemos el siguiente teorema:

Teorema: La Media y la Varianza de X es:

1. $\mathbb{E}(X) = \dfrac{1}{\lambda}$.
2. $\text{Var}(X) = \dfrac{1}{\lambda^2}$.

Es decir mi factor de escala está afectando mis datos, si $\lambda \to \infty$ se reduce el espacio de mis datos, y al contrario si $\lambda \to 0$. Y mi funcion de distribución acumulada fes: $1 - e^{-\lambda t}$, y así $\mathbb{P}(X \leq t_0) = F(t_0) = 1 - e^{-\lambda t_0}$.

2.4.6. Propiedad de la No Memoria

Si $X \sim Exp(\lambda)$, entonces:

$$\mathbb{P}(X > s+t | X > s) = \mathbb{P}(X > t)$$

Para $s, t > 0$.

2.4.7. Distribución Gamma

Definición: La v.a. X se dice que tiene distribución Gamma sí se cumple que su f.f.p está dada por:

$$f_X(\kappa, \theta) = \begin{cases} \dfrac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-\frac{x}{\theta}} & si \quad x \geq 0 \\ \\ 0 & \quad e.o.p \end{cases} \tag{34}$$

Dónde $\kappa, \theta \geq 0, \Gamma(\kappa) > 0$. Recordemos que $\Gamma(\alpha)$ se refiere a la función Gamma, por lo que vale recordar quién es:

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy; \Gamma(\alpha) = (\alpha - 1)!$$

.

Teorema: Si $X \sim Gam(\lambda, \theta)$, entonces:

1. $\mathbb{E}(X) = \kappa\theta$.
2. $\text{Var}(X) = \kappa\theta^2$.

2.4.8. Distribución Beta

La distribución Beta es una extensión de la Distribución Uniforme, se conoce también como *Beta de Euler*. La v.a. X continua distribuye Beta con parámetros $\alpha, \beta > 0$ si su f.d.p es dada por:

$$f(x,\alpha,\beta)=\begin{cases}\dfrac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} & si \quad x\geq 0\\ \\ 0 & e.o.p\end{cases} \tag{35}$$

Luego se dice que $X\sim Beta(\alpha,\beta)$.

Teorema: Si $X\sim Beta(\alpha,\beta)$, tenemos que:

1. $\mathbb{E}(X)=\dfrac{\alpha}{\alpha+\beta}$.
2. $\text{Var}(X)=\dfrac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}$.

Ahora sí elegimos ciertos parámetros, obtenemos una nueva distribución:

2.4.9. Distribución χ^2 (Chi-Cuadrada)

Si elegimos $\kappa=\dfrac{\nu}{2}$, y $\theta=2$, obtenemos que:

$$f\left(\frac{\nu}{2},2\right)=\begin{cases}\dfrac{x^{\nu/2-1}e^{-x/2}}{\Gamma\left(\frac{\nu}{2}\right)2^{\nu/2}} & si \quad x\geq 0\\ \\ 0 & e.o.p\end{cases} \tag{36}$$

Así decimos que $X\sim\chi^2_\nu$.

Teorema: Tenemos que

1. $\mathbb{E}(X)=\nu$.
2. $\text{Var}(X)=2\nu$.

2.4.10. Teorema de Aproximación

Sea $X\sim Bin(n,p)$, cuándo $n\to\infty, p\to 0;\mu=np$ constante se dice que podemos aproximar la distribución binomial, con una distribución Poisson, de parámetro μ, es decir $Poi(\mu)$.

En este punto veremos qué ocurre cuándo conocemos dos datos muy importantes de las variables aleatorias, las cuáles son la media y la varianza, pero cabe el caso dónde no conozcamos la función/distribución de probabilidad, por lo que tendremos que invocar ciertas desigualdades que nos serán útiles en este proceso.

2.4.11. Desigualdad de Markov

Inicialmente la desigualdad de Markov nos permite establecer una correlación entre la esperanza y la probabilidad de que la v.a. tome ciertos valores:

Sea $U(X)$ una función no negativa de la v.a. X. Si existe $\mathbb{E}[U(X)]$, entonces para todo $c>0$ se tiene que:

$$\mathbb{P}[U(X)\geq c]\leq\frac{\mathbb{E}(U(X))}{c} \tag{37}$$

Demostración: Sea X una v.a. de tipo continuo[14], además sea $A=\{x:U(x)\geq c\}$, y $f(x)$ la f.d.p (f.m.p para discreto) de la v.a. X, entonces tenemos que:

$$\mathbb{E}(U(X))=\int_{-\infty}^{\infty}U(x)f(x)\,dx=\int_A U(x)f(x)\,dx$$
$$+\int_{A^c}U(x)f(x)\,dx\geq\int_A U(x)f(x)\,dx\geq c\int_A f(x)\,dx$$
$$=c\,\mathbb{P}(X\in A)=c\,\mathbb{P}(U(X)\geq c)$$

[14] Para el caso discreto es la misma idea, basta con cambiar de integral a sumatoria/serie.

La gran importancia de esta desigualdad es probar de forma más efectiva la desigualdad de Chebyshev, importante también por generar una cota.

2.4.12. Desigualdad de Chebyshev

Para la desigualdad de Chebyshev necesitaremos la media y la varianza dentro de la formulación, la desigualdad dice así:

Sea X una v.a. con $\mathbb{E}(X) = \mu$, y $Var(X) = \sigma^2$, entonces para cualquier $k > 0$ se tiene que:

$$\mathbb{P}[|X-\mu| \geq k\sigma] \leq \frac{1}{k^2} \tag{38}$$

Para demostrarlo basta con usar la desigualdad de Markov, y sustituir en $U(X) = (X-\mu)^2$, y $c = k^2\sigma^2$.

Además de esa forma de enunciarla, existe una alternativa que es:

$$\mathbb{P}[|X-\mu| < k\sigma] \geq 1 - \frac{1}{k^2} = \frac{k^2-1}{k^2}$$

Observación: Cuándo $k = 2$ el teorema indica que la v.a. X tiene una probabilidad de al menos 3/4 de caer dentro de 2 desviaciones estándar de la media, es decir que 3/4 o más[15] de las observaciones de **cualquier** distribución cae en el intervalo $(\mu - 2\sigma, \mu + 2\sigma)$.

Realicemos un ejemplo con datos reales de la época de pandemia:

Supongamos que el Metro de Medellín tiene un promedio de 200.7 pasajeros en miles por dia en pandemia, con una desviación estándar de 12.14 en miles ¿Con qué frecuencia los pasajeros están dentro de 2 desviaciones de la media, y cuál es dicho intervalo?

Solución: Definamos cuál será mi variable aleatoria:

Sea $X :=$ # de pasajeros por día en el Metro de Medellín en la época de pandemia, sabemos que:

- $\mathbb{E}(X) = 200{,}7$
- $Var(X) = 12, 14$

Luego sabemos que:

$$\mathbb{P}[|X - 200{,}7| < 2 * 12{,}14] \geq \frac{3}{4} =$$

$$\mathbb{P}[176{,}42 < X < 224{,}98] \geq 3/4$$

De esta forma podemos concluir que al menos el 75 % de los días estará en un radio de 2 desviaciones estándar con respecto a la media la cantidad de personas que transporta en miles el Metro de Medellín en época de Pandemia.

2.4.13. Función Generadora de Momentos

Definiremos además una función que nos auxiliará en los momentos de buscar encontrar características de ciertas distribuciones[16] que hemos visto[17], definamos de forma explícita la misma:

Sea X una v.a. entonces el valor esperado

[15] Recordemos que es una cota inferior, puede ser ese número como alguno mayor.

[16] Tales como la media, y con unos procesos extras la varianza.

[17] Cabe aclarar que no todas tienen una Generadora de manera explícita, habría que recurrir al Análisis Numérico para llegar a una aproximación del valor buscado.

$$M_X(t) = \mathbb{E}[e^{tX}] \qquad (39)$$

es llamada función generadora de momentos (F.G.M.)[18] de nuestra v.a. X. Si este valor esperado existe para todo valor t en algún intervalo de la forma $-h < t < h$, para algún $h > 0$.

Observaciones:

1. Si $t = 0$ entonces $M_X(t) = 1$.
2. No toda distribución tiene F.G.M.
3. La F.G.M. es única.
4. Si 2 v.a. tienen la misma F.G.M. entonces tienen la misma distribución.

Hagamos varios ejemplos de como hallar la F.G.M.

Sea X una v.a. continua con f.d.p $f(x) = e^{-x}, x > 0$ y 0 en otro caso. Lo que haremos será aplicar la definición, es decir:

$$M_X(t) = \mathbb{E}(e^{tX}) = \int_0^\infty e^{tx}e^{-x}\,dx = \int_0^\infty e^{-x(1-t)}\,dx$$

$$\left[-\frac{1}{1-t}e^{-x(1-t)}\right]_0^\infty = -\frac{1}{1-t}(0-1) = \frac{1}{1-t},{}_{(1-t>0)}$$

Demos un ejemplo distinto:

Sea X una v.a. discreta con f.m.p. $p(x) = (1/2)^{x+1}, x = 0,1,2,\cdots$, y 0 en otro caso, la F.G.M. sería:

$$M_X(t) = \sum_{x=0}^\infty e^{tx}(1/2)^{x+1} = \frac{1}{2}\sum_{x=0}^\infty (e^t/2)^x =$$

Luego multiplicando por $1/2$, y tomando $s = e^t/2$, tendremos que

$$\frac{1}{2}(1+s+s^2+\cdots) = \frac{1}{2}\frac{1}{1-s} = \frac{1}{2-e^t}; t < \ln 2$$

Pero de igual forma un caso para analizar muy importante es el de las distribuciones que ya conocemos, por lo que procederemos con el proceso para sacarlas.

2.4.14. F.G.M. de las Distribuciones del Curso (Discretas)

Recordemos que la distribución BErnoulli es un caso especiico de Binomial[19], por lo que procederemos a estudiar en inicio la Binomial.

F.G.M Distribución Binomial

Recordemos que sí $X \sim Bin(n,p)$, mi $p(x) = \binom{n}{x}p^xq^{n-x}$, así, aplicando la definición obtenemos que:

$$M_X(t) = \sum_{x=0}^n e^{tx}\binom{n}{x}p^xq^{n-x} = \sum_{x=0}^n \binom{n}{x}(pe^t)^xq^{n-x}$$

Para lo cuál aplicando Binomio de Newton, nos queda en que:

$$M_X(t) = (pe^t + q)^n \qquad (40)$$

Así, para la Bernoulli la F.G.M. es $pe^t + q$.

F.G.M. Distribución Geométrica

Recordemos que $p(x) = pq^{x-1}$, con los valores $x = 1,2,3,\cdots$

$$M_X(t) = \sum_{x=1}^\infty e^{tx}pq^{x-1} = pe^t\sum_{x=0}^\infty (qe^t)^x =$$

[18] En el libro guía, se encuentra como MGF *Moment Generator Function*.

[19] Tomar $n = 1$ en la Binomial.

$$pe^t \frac{1}{1-qe^t} = \frac{pe^t}{1-qe^t}$$

Con la condición de que $t < -\ln q$.

F.G.M. Distribución Binomial Negativa

Sabemos que $p(x) = \binom{x-1}{r-1} p^r q^{x-r}$, para los valores $x = r, r+1, \cdots$, el proceso de la F.G.M sería:

$$M_X(t) = \sum_{x=r}^{\infty} \binom{x-1}{r-1} p^r q^{x-r} e^{tx} =$$

Realizando el producto de e^{tr}, obtenemos que:

$$p^r e^{tr} \sum_{x=r}^{\infty} \binom{x-1}{r-1} q^{x-r} (e^t)^{x-r} =$$

$$p^r e^{tr} \sum_{x=r}^{\infty} \binom{x-1}{r-1} (qe^t)^{x-r}$$

Para mayor facilidad, y no cargar de tantas cosas el documento, analicemos la serie que está ahí por separado. Recordemos que:

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

Por lo que tomando $q = qe^t$ obtenemos que:

$$\sum_{x=r}^{\infty} \binom{x-1}{r-1} (qe^t)^{x-r} = (1-qe^t)^{-r}$$

A lo que reemplazando en la ecuación se obtiene que:

$$M_X(t) = \left(\frac{pe^t}{1-qe^t} \right)^r \tag{41}$$

F.G.M. Distribución Poisson

Si $X \sim Poi(\mu)$ tendremos que:

$$M_X(t) = \sum_{x=0}^{\infty} e^{tx} \frac{e^{-\mu} \mu^x}{x!} = e^{-\mu} \sum_{x=0}^{\infty} \frac{(e^t \mu)^x}{x!} = e^{-\mu} e^{e^t \mu}$$

De forma que podemos concluir que para cualquier $t \in \mathbb{R}$ la F.G.M. es:

$$M_X(t) = e^{\mu(e^t - 1)} \tag{42}$$

En el caso de la Distribución Hipergeométrica en el libro guía no se data una forma de encontrarla, y en el Apéndice B[20] aparece una nota que dice que resolverla no es muy fácil para el nivel y momento del curso, se necesita una habilidad mayor dentro de la combinatoria, cosa que muchos estudiantes al nivel del pregrado y de ver el curso no poseen. Para la Distribución Uniforme se deja para el lector realizarla. Veamos que pasa con las Continuas.

2.4.15. F.G.M. de las Distribuciones del Curso (Continuas)

F.G.M. Distribución Uniforme

Sea $X \sim Unif(a,b)$, sabemos que $f(x) = \dfrac{1}{b-a}, x \in (a,b)$ y 0 en otro caso. Aplicando la definición con integrales queda:

$$M_X(t) = \int_a^b \frac{e^{tx}}{b-a} dx = \frac{1}{b-a} \int_a^b e^t x\, dx = \frac{e^{bt} - e^{at}}{(b-a)t} \tag{43}$$

F.G.M. Distribución Normal

Sea $X \sim N(\mu, \sigma^2)$, luego su F.G.M. es:

$$M_X(t) = e^{\mu t + \sigma^2 t^2 / 2} \tag{44}$$

Queda de tarea demostrarlo.

[20]Lugar dónde se encuentran las distribuciones con su información más importante.

F.G.M. Distribución Exponencial
Sea $X \sim Exp(\theta)$ luego la F.G.M. es:

$$M_X(t) = \int_0^{\infty} e^{tx} \frac{1}{\theta} e^{-x/\theta}\, dx = \frac{1}{\theta}\int_0^{\infty} e^{-x(-t+1/\theta)}\, dx$$

$$= \frac{1}{\theta} \frac{1}{\frac{1-\theta t}{\theta}} \left[-e^{-x(-t+1/\theta)}\right]_0^{\infty} = \frac{1}{1-\theta t}$$

F.G.M. Distribución Gamma
Sea $X \sim \Gamma(\theta, \kappa)$ luego la función generadora de momentos es:

$$M_X(t) = \int_0^{\infty} e^{tx} \frac{x^{\kappa-1} e^{-x/\theta}}{\theta^{\kappa}\Gamma(\kappa)}\, dx =$$

$$\frac{1}{\theta^{\kappa}\Gamma(\kappa)} \int_0^{\infty} x^{\kappa-1} e^{(t-1/\theta)x}\, dx$$

Analicemos internamente esa integral, hagamos la siguiente sustitución:

$$u = -(t-1/\theta)x \Rightarrow du = -(t-1/\theta)dx$$

Luego la integral queda:

$$\int_0^{\infty} \frac{u^{\kappa-1}}{\left(\frac{1}{\theta}-t\right)^{\kappa-1}} e^{-u} \frac{du}{\frac{1}{\theta}-t}$$

De la cuál podemos sacar la función

$$\left(\frac{1}{\theta}-t\right)^{-\kappa}$$

Y teniendo en cuenta que lo de adentro de la Integral es la Función Gamma de κ, podemos operar y nuestra integral resulta en:

$$\frac{1}{\theta^{\kappa}\Gamma(\kappa)} \left(\frac{1}{\theta}-t\right)^{-\kappa} \Gamma(\kappa) \Rightarrow M_X(t) = (1-\theta t)^{-\kappa} \tag{45}$$

F.G.M. Distribución χ^2
Sea $X \sim \chi^2(\nu)$ luego su F.G.M. es algo muy parecido a la de la Gamma, dado que hay tomar $\theta = 2, \kappa = \frac{\nu}{2}$, quedando:

$$M_X(t) = (1-2t)^{-\nu/2} \tag{46}$$

F.G.M Distribución Beta
Sea $X \sim Beta(a,b)$ para desarrollar su Generadora de Momentos, ocurre el mismo problema de la Hipergeométrica, las tácticas y métodos no están a nuestro nivel.

Con esto hemos terminado el tema de las Funciones Generadoras de Momento, teniendo esto en cuenta, daremos propiedades fundamentales de la F.G.M.

2.4.16. Propiedades de la Función Generadora de Momentos

Teorema
Sí la F.G.M. de X existe entonces:

- $\mathbb{E}(X^r) = M_X^{(r)}(0)$.
- $$M_X(t) = 1 + \sum_{r=1}^{\infty} \frac{\mathbb{E}(X^r)t^r}{r!}$$

[21] Recordemos que la diferenciación en este caso es con respecto a t, dado que esa es la variable de la F.G.M.

Haremos la prueba de cada una, para el primero haremos el continuo, y queda de trabajo hacer el discreto. Sabemos cuál es la fórmula de la F.G.M.

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x)\,dx$$

Luego tendremos que:

$$M_X^{(r)}(t) = \int_{-\infty}^{\infty} x^r e^{tx} f_X(x)\,dx^{21}$$

Ahora veamos lo que ocurre cuándo hacemos $t = 0$:

$$M_X^{(r)}(0) = \int_{-\infty}^{\infty} x^r e^0 f_X(x)\,dx = \mathbb{E}(X^r) \qquad (47)$$

Para el segundo punto tomamos la expansión en series de Maclaurin, y reemplazamos $M_X^{(r)}(0)$ por lo que acabamos de probar, es decir $\mathbb{E}(X^r)$.

Daremos otro teorema importante con respecto a esta función:

Teorema: Si $Y = aX + b$, entonces $M_Y(t) = e^{bt} M_X(at)$.

Demostración:

$$M_Y(t) = \mathbb{E}(e^{tY}) = \mathbb{E}(e^{t(aX+b)}) = \mathbb{E}(e^{t(aX)} e^{bt})$$

$$e^{bt}\mathbb{E}(e^{(aX)t}) = e^{bt} M_X(at) \qquad (48)$$

Este teorema nos es útil para calcular por ejemplo el r-ésimo momento alrededor la media μ, el cuál sería:

$$\mathbb{E}[(X-\mu)^r] = \frac{d^r}{dt^r}[e^{-\mu t} M_X(t)]_{t=0}$$

De igual modo hasta aquí llega el temario del Parcial 2 del curso, toma el tema de Variable Aleatoria Continua, Distribuciones de Probabilidad, Discretas y Continuas, y las Desigualdades importantes, y la Función Generadora de Momentos, ¡muchos éxitos en el Parcial!

3. Parcial 3

Dentro de este capítulo definiremos nuevas formas de calcular probabilidades, dado que tendremos la opción de usar varias distribuciones al tiempo, usando lo que llamremos como *Vectores Aleatorios* las cuáles siguen una idea parecida a lo que hemos veniudo haciendo desde hace tiempo. Dentro del actuar de la probabilidad existe el caso de que deseemos ver distintos fenómenos en un mimso experimento, por ejemplo, en un examen, la nota es un experimento (podemos interesarnos con la probabilidad de sacar cierto rango de nota, o cierta nota específica), y el tiempo invertido a estudiarle al examen es otro, aquí podemos evidenciar que cabe el caso de combinar dos tipos distintos de distribuciones.

Definición: Sean $X_1, X_2, \cdots, X_n$ n variables aleatorias definidas sobre un mismo espacio de probabilidad $(\Omega, \mathscr{A}, \mathbb{P})$, y además sea $\tilde{X} : \Omega \to \mathbb{R}^n$ definidio por:

$$\tilde{X}(\omega)^{22} = (X_1(\omega), X_2(\omega), \cdots, X_n(\omega)) \tag{49}$$

se llama **vector aleatorio n-dimensional**.

En la mayoría de este curso analizaremos principalmente vecores aleatorios 2-dimensionales. Sea ξ un experimento aleatorio y Ω un espacio muestral asociado a ξ, y consideremos las avriables aleatorias X_1, X_2 que asigna a cada elemento $\omega \in \Omega$ uno y sólo un parordenado de números reales $X_1(\omega) = x_!, X_2(\omega) = x_2$, luego nuestro vector aleatorio es: $\tilde{X} = (X_1, X_2)$.

El espacio de (X_1, X_2) es un conjunto de parejas ordenadas

$$D := \{(x_1, x_2) : X_1(\omega) = x_1, X_2(\omega) = x_2, \omega \in \Omega\}$$

(Revisar la figura 3)

$X_1(\omega), X_2(\omega)$ son dos funciones dónde cada una de ellas asigna un número real a cada resultado $\omega \in \Omega$.

Definición: Un vector aleatorio (X_1, X_2) es discreto su su espacio D es finito ó infinito contable. EL vector aleatorio (X_1, X_2) es un vector aleatorio continuo si (X_1, X_2) puede tomar todos los valores en un conjunto no numerable del plano euclidiano.

Definición: Sea $\tilde{X}$ un vector aleatorio $n-$dimensionalla probabilidad está dada por:

$$\mathbb{P}_{\tilde{X}}(B) := \mathbb{P}(\tilde{X} \in B) \tag{50}$$

se llama distribución del vector aleatorio $\tilde{X}$.

Se puede definir una única $\mathbb{P}_{X_1,X_2}$ en términos de la función de disribución acumulada como:

$$F_{X_1,X_2}(x_1, x_2) = \mathbb{P}[\{X_1 \leq x_1\} \cap \{X_2 \leq x_2\}]$$

Se puede verificar que:

$$\mathbb{P}[a_1 \leq X_1 \leq b_1, a_2 \leq X_2 \leq b_2] = F(b_1, b_2)$$

$$-F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2)$$

Podemos ver la figura 4 para entender gráficamente el proceso.

Ahora, nos interesa saber y conocer como sería una probabilidad en términos de las dos variables aleatorias.

[22] Existe la posibilidad de que la notación varíe un poco, dado que en ocasiones ciertas notaciones hacerlas en el formato LaTeX es un poco difícil.

3.1. Función de Probabilidad Conjunta

De aquí en adelante simplemente estaremos escribiendo solo la función de probabilidad, dejamos por escrito que se dará: $x_i \in X_i, \forall i \in \{1, \cdots, n\}$, esto con el objetivo de no sobrecargar la notación en el momento de explicar y hacer los cálculos respectivos.

3.1.1. Función Masa de Probabilidad Conjunta

Sea (X_1, X_2) un vector aleatorio discreto, luego la **función de masa de probabilidad conjunta** de (X_1, X_2) está definida por:

$$\mathbb{P}_{X_1,X_2}(x_1,x_2) = \mathbb{P}[X_1 = x_1, X_2 = x_2] \qquad (51)$$

Y además satisface las siguientes propiedades:

- $0 \leq \mathbb{P}(x_1,x_2) \leq 1$
- $\sum_{x_1}\sum_{x_2} \mathbb{P}(x_1,x_2) = 1$

Ahora, para un evento B tenemos:

$$\mathbb{P}[(x_1,x_2) \in B] = \sum_B \mathbb{P}(x_1,x_2) \qquad (52)$$

3.1.2. Función Densidad de Probabilidad Conjunta

Si (X_1, X_2) es de tipo continuo tenemos que nuestra función de densidad de probabilidad conjunta definida de la misma forma que en el caso discreto, debe de cumplir las sigueintes propiedades:

- $$0 \leq \mathbb{P}(x_1,x_2) \leq 1$$
- $$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_1,x_2)dx_1dx_2 = 1$$

Ahora daremos varias propiedades de la Acumulada que esbozamos al inicio de este capítulo:

Teorema 4.2.2 Una función $F(x_1,x_2)$ es una CDF[23] bivariada si y solo si:

- $\lim_{x_1 \to -\infty} F(x_1,x_2) = F(-\infty, x_2) = 0, \forall x_2$
- $\lim_{x_2 \to -\infty} F(x_1,x_2) = F(x_!, -\infty) = 0, \forall x_1$
- $\lim_{\substack{x_1 \to \infty \\ x_2 \to \infty}} F(x_1,x_2) = F(\infty, \infty) = 1$
- $F(b_1,b_2) - F(a_1,b_2) - F(b_1,a_2) + F(a_1,a_2) \geq 0$
- $\lim_{h \to 0^+} F(x_1,x_2+h) = \lim_{h \to 0^+} F(x_1+h,x_2) = F(x_1,x_2), \forall x_1,x_2$

Veamos propiedades de la Acumulada en su versión continua:

$$F(x_1,x_2) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2} F(t_1,t_2)\,dt_2dt_1 \qquad (53)$$

Si recordamos que en el caso 1-dimensional teníamos una característica especial para obtener la f.d.p. y la CDF, la cuál era derivando la CDF para obtener la f.d.p. e integrando la f.d.p. para obtener la CDF. Para estos vectores aleatorios existirá una relación muy parecida:

[23] Recordemos: *Cumulative Distributive Function*.

$$f(x_1,x_2) = \frac{\partial^2}{\partial x_1 \partial x_2} F(x_1,x_2) \tag{54}$$

Recordemos que todas estas definiciones podemos extenderlas a vectores aleatorios $k-$dimensionales, solo que por efectos del curso nos quedaremos en el caso de $k = 2$.

Dado que pueden existir varias condiciones de mis experimentos, requeriremos definir regiones de integración dónde nuestros experimentos tengan sentido de ocurrir, tales como, así que definiremos:

$$\mathbb{P}[\tilde{X} \in A] = \iint_A f(x_1,x_2)\ dx_2 dx_1 \tag{55}$$

Recomiendo ver el ejemplo 4.3.1 del libro de Bain[24]

Tendremos en muchos casos eventos que tengan dos o más objetivos de análisis, y desearemos ver en ocasiones como se comportan una fracción de estos eventos, ver su comportamiento para determinar resultados, para esto definiremos lo que significa ser una función marginal de probabilidad.

3.2. Función Marginal

Definiremos ahora que es la función marginal, que nos permitirá dar análisis en cuestión a menos variables de las originales de mi vector aleatorio, y además nos deja determinar como sería la independencia de las variables aleatorias.

[24] Introduction to Probability and Mathematical Statistics.

3.2.1. Función de Probabilidad Marginal

La marginal lo que hará será suponer que ya todo el espacio de una de las Variables Aleatorias que componen a mi Vector Aleatorio ya ocurrió, para dejarlo todo en términos de la otra, lo definiremos como:

$$f_1(x_1) = \int_{-\infty}^{\infty} f(x_1,x_2)\ dx_2 \tag{56}$$

$$f_2(x_2) = \int_{-\infty}^{\infty} f(x_1,x_2)\ dx_1$$

Podemos entenderla como hacer la integral con respecto a la variable contraria, de una forma más intuitiva.

Para el caso discreto debemos de analizar es las sumatorias/series de la función, es decir:

$$f_1(x_1) = \sum_{x_2} f(x_1,x_2)$$

$$f_2(x_2) = \sum_{x_1} f(x_1,x_2)$$

Nuevamente, recomiendo leer el ejemplo 4.3.2 del libro guía.

3.2.2. Función Acumulada de Probabilidad Marginal

Similarmente definimos la marginal de la CDF como:

$$F_1(x_1) = \lim_{x_2 \to \infty} F(x_1,x_2) \tag{57}$$

$$F_2(x_2) = \lim_{x_1 \to \infty} F(x_1,x_2)$$

Veamos ahora un nuevo concepto necesario

3.3. Variables Aleatorias Independientes

Para determinar la dependencia o independencia necesitamos que todos los eventos posibles lo sean (se nos muestra esto al inicio del capítulo 4 sección 4) así que daremos la definición de que las V.A. de mi Vector Aleatorio son independientes si y solo si:

$$f(x_1,x_2) = f_1(x_1)f_2(x_2) \tag{58}$$

Dónde $f(x_1,x_2), f_1(x_1), f(x_2)$ son la conjunta y las marginales respectivamente, de igual forma se cumple la propiedad de que son independientes si y solo si:

$$F(x_1,x_2) = F_1(x_1)F(x_2) \tag{59}$$

Esto nos puede dar una idea de como es el comportamiento de estas funciones y la independencia de las v.a.

Definiremos también lo que sería el conjunto soporte:

Conjunto Soporte: Se define como: $\{(x_1,x_2) : f(x_1,x_2) > 0\}$.

Ahora la necesidad de tener el conjunto soporte es para poder dar una definitiva a cuándo son independientes dos variales aleatorias:

Teorema 4.4.2

Dos variables aleatorias X_1,X_2 con f.d.p. conjunta $f(x_1,x_2)$ son independientes si y solo si:

1. El conjunto soporte es expresado como un par cartesiano.
2. La f.d.p. puede ser factorizada en el producto de dos funciones de la forma: $f(x_1,x_2) = j(x_1)h(x_2)$

Recomiendo mirar el ejemplo 4.4.1,4.4.2, y 4.4.3.

3.4. Distribuciones Condicionales

De igual forma como logramos determinar en el 1$^{\text{er}}$ parcial, aquí podemos extender el concepto de dependencia de una variable aleatoria a otra, muchas veces querremos encontrar probabilidades con la siguiente estructura:

$$\mathbb{P}[X_1 = x_1 | X_2 = x_2] = \frac{\mathbb{P}[X_1 = x_1, X_2 = x_2]}{\mathbb{P}[X_2 = x_2]} = \frac{f(x_1,x_2)}{f_2(x_2)}$$

Así tendremos que podemos definir la **función de densidad de probabilidad condicional**:

Si X_1,X_2 son v.a. con f.d.p. $f(x_1,x_2)$ la f.m.p. condicional sería:

$$f(x_2|x_1) := \frac{f(x_1,x_2)}{f_1(x_1)}; f_1(x_1) > 0 \tag{60}$$

Para el caso discreto conservamos la idea que traemos de Condicional, para el caso continuo hay que hacer unas aclaraciones, para eso usaremos un argumento muy parecido al factor de corrección, para poder hallar esas probabildades. Considero necesario revisar los ejemplos 4.5.1, 4.5.2 y 4.5.3.

3.5. Muestras Aleatorias

Muchas veces analizaremos diversos comportamientos bajo la misma población, por ejemplo en un grupo de bombillas, analizar una específicamente, y ver como es su fallo. A estos estudios que haremos es lo que iremos llamando *Muestra Aleatoria*.

Definamos que es:

El conjunto de variables aleatorias $X_1, \cdots, X_n$ se dice que es una **muestra aleatoria** de tamaño n de una población con f.d.p. $f(x)$ si la función de probabilidad conjunta es:

$$f(x_1, x_2, \cdots, x_n) = f(x_1)f(x_2)\cdots f(x_n) \quad (61)$$

Recomiendo revisar el ejercicio 22 de los propuestos por Bain, y el resuelto 4.6.1.

Ahora veamos distribuciones especiales.

3.6. Distribuciones Especiales

Además de las distribuciones ya mencionadas, existen otras que en varias variables permiten analizar distintos fenómenos.

3.6.1. Hipergeométrica Extendida

Existirá el caso dónde mediante procesos hipergeométricos debamos analizar un problema dónde depende de más de una variable, así que podemos dar una dependencia de los $n+1$ tipos en n variables estableciendo restricciones entre los valores de las variables, me hago explicar con este ejemplo:

Un cajón tiene 1000 semillas de flores, de las cuáles 400 son rojas, 400 son blancas y 200 son rosas, si se seleccionan 10 semillas al azar sin reemplazo, definamos como X_1 el número de semillas rojas, y X_2 el número de semillas blancas, ahora su función dependerá de dos variables, pero sabemos que el número máximo de semillas es de 10, es decir que las semillas restantes que serían las rosadas estarían atadas a las blancas, y a las rojas, y al número 10 que es el total de nuestra muestra. Es decir que tenemos la siguiente información:

- $x_1, x_2 \geq 0$.
- $x_1 + x_2 \leq 0$.

Podemos plantear la probabilidad como:

$$f(x_1, x_2) = \frac{\binom{400}{x_1}\binom{400}{x_2}\binom{200}{10-x_1-x_2}}{\binom{1000}{10}}$$

Ahora:

$$f(2,5) = \frac{\binom{400}{2}\binom{400}{5}\binom{200}{10-2-5}}{\binom{1000}{10}} = 0{,}0331.$$

Ahora más generalmente tenemos:

$$f(x_1, \cdots, x_k) = \frac{\binom{M_1}{x_1}\binom{M_2}{x_2}\cdots\binom{M_{k+1}}{x_{k+1}}}{\binom{N}{n}} \quad (62)$$

Dónde se tiene que: $\forall i(0 \leq x_i \leq M_i)$, dónde además: $M_{k+1} = N - \sum_{i=1}^{k} M_i$; $x_{k+1} = n - \sum_{i=1}^{n} x_i$. Ahora diremos que el vector aleatorio distribuye de esta forma si y solo si:

$$\tilde{X} \sim HYP(n, M_1, M_2, \cdots, M_k, N)$$

3.6.2. Distribución Multinomial

Supongamos que hay $k+1$ eventos mutuamente excluyentes y exhaustivos $E_1, E_2, \cdots, E_k, E_{k+1}$ que pueden ocurrir en cualquier intento de un experimento, y sea $p_i = \mathbb{P}(E_i)$, para $i = 1, 2, \cdots, k+1$. En n intentos independientes del experimento, nosotros definiremos como $X_i =$ número de ocurrencias del evento E_i. El vector $\tilde{X} = (X_1, X_2, \cdots, X_k)$ se dice

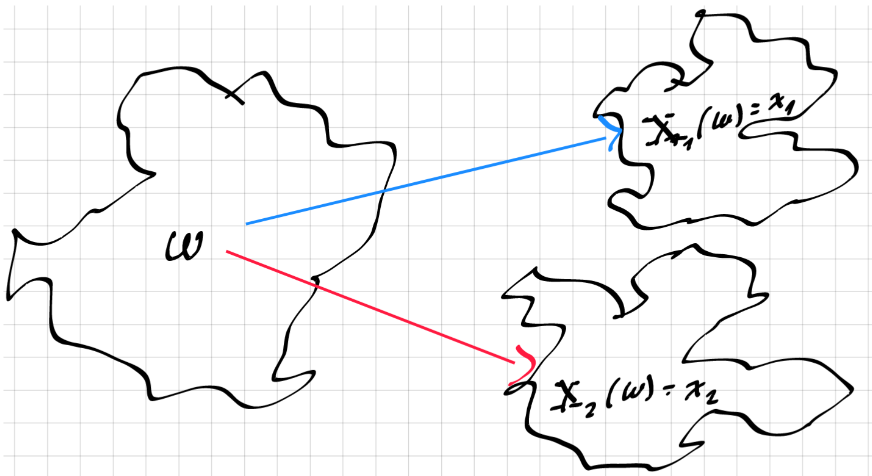

Figura 3: Vector Aleatorio definido como función.

que tiene distribución multinomial si tiene una f.m.p. conjunta de la forma:

$$f(x_1,x_2,\cdots,x_k) = \frac{n!}{x_1!x_2!\cdots x_{k+1}!} p_1^{x_1} p_2^{x_2} \cdots p_{k+1}^{x_{k+1}} \tag{63}$$

Dónde tenemos:$\forall i(0 \leq x_i \leq n)$; $x_{k+1} = n - \sum_{i=1}^{k} x_i$; $p_{k+1} = 1 - \sum_{i=1}^{k} p_i$

Luego diremos que:

$$\tilde{X} \sim MULT(n, p_1, \cdots, p_k) \tag{64}$$

Y como algo especial que tenemos con esta distribución se cumple que:

Si el vector distribuye multinomial, tendremos que:

$$X_i \sim Bin(n, p_i) \tag{65}$$

La definición de la multinomial nos recuerda un poco a la permutación con repeticiones.

Veamos ahora propiedades de las funciones de variable aleatoria que es una parte de nuestro 3er parcial.

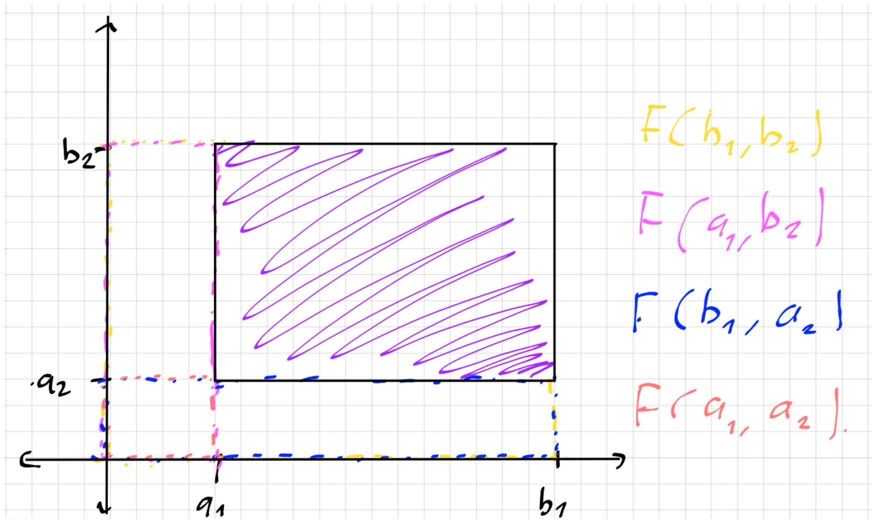

Figura 4: Función de Distribución Conjunta

Continuando con lo visto en el capítulo 4 del libro, podemos estructurar las siguientes propiedades que van teniendo nuestras variables aleatorias, empecemos con propiedades de la esperanza:

3.7. Propiedades de las Variables Aleatorias

Muchas veces necesitaremos estimar los estadísticos, y parámetros desconocidos, cosa que se nos mostrará y dará las herramientas para hacerlo.

3.7.1. Propiedades del Valor Esperado

Muchas veces tendremos dependencia de una función que querramos analizar, es decir que puede darse el caso de que:

$$\mathbb{E}(Y) = \mathbb{E}(g(\tilde{X}));\ \tilde{X} = (X_1, X_2, \cdots, X_k)$$

Dónde tendremos las definiciones usuales, es decir:

$$E[g(\tilde{X})] = \sum_{x_1} \cdots \sum_{x_k} u(x_1, \cdots, x_k) f(x_1, \cdots, x_k)$$

Esto en el caso discreto, y para el continuo, sería realizar las $k-$integrales en todo $\mathbb{R}^k$.

Tenemos un teorema de suma de v.a.

Teorema: Sea $f(x_1, x_2)$ la conjunta de X_1, X_2, luego:

$$\mathbb{E}[X_1 + X_2] = \mathbb{E}[X_1] + \mathbb{E}[X_2]$$

Este teorema se puede generalizar a:

$$\mathbb{E}\left(\sum_{i=1}^{k} a_i X_i\right) = \sum_{i=1}^{k} a_i \mathbb{E}(X_i)$$

Teorema:

Si X, Y son v.a. independientes y $g(x), h(y)$ funciones entonces:

$$\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)]\mathbb{E}[h(Y)]$$

Este teorema se puede extender a varias variables independientes es decir:

$$\mathbb{E}[u_1(X_1)\cdots u_k(X_k)] = \mathbb{E}[u_1(X_1)]\cdots\mathbb{E}[(u_k(X_k)]$$

3.7.2. Covarianza

Definiremos ahora la **Covarianza** de dos variables aleatorias X,Y como:

$$Cov(X,Y) = \mathbb{E}[(X-\mu_X)(Y-\mu_Y)] \tag{66}$$

Otra forma de denotar la covarianza es mediante el símbolo σ_{XY}.

Demos propiedades de la Covarianza:

- $Cov(aX,bY) = abCov(X,Y)$.
- $Cov(X+a,Y+b) = Cov(X,Y)$.
- $Cov(X,aX+b) = aVar(X)$.

Teorema 5.2.5: Si X,Y son variables aleatorias entonces:

$$Cov(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y); Cov(X,Y) = 0 \tag{67}$$

Implica que X,Y son independientes.

Se recomienda ver el ejercicio 5.2.1 y 5.2.2.

Ahora nosotros somos capaces de definir una forma de aproximar la media y la varianza:

Sea $H(x,y)$ una función con derivadas parciales en un rectángulo que contiene al par (μ_X,μ_Y), y un par de variables aleatorias (X,Y) con medias y varianzas: μ_X,μ_Y; σ_X^2,σ_Y^2, y covarianza σ_{XY}, así tendremos:

1.
$$\mathbb{E}[H(X,Y)] = H(\mu_X,\mu_Y) + \frac{\partial^2 H}{\partial x^2}\sigma_X^2 + \frac{\partial^2 H}{\partial y^2}\sigma_Y^2$$
.

2.
$$Var[H(X,Y)] = \left(\frac{\partial H}{\partial x}\right)^2\sigma_X^2 + \left(\frac{\partial H}{\partial y}\right)^2\sigma_Y^2 +$$
$$2\left(\frac{\partial H}{\partial x}\right)\left(\frac{\partial H}{\partial y}\right)\sigma_{XY}$$

3.8. Correlación

Nosotros tenemos que si $Cov(X,Y) \neq 0$ entonces X,Y son dependientes. Pero que $Cov(X,Y) = 0$ no implica que ambos sean independientes. Definamos algo además que nos permite ver como se relacionan:

3.8.1. Definición 5.3.1. Coeficiente de Correlación

Si X,Y son variables aleatorias con varianzas σ_X^2,σ_Y^2, y covarianza $\sigma_{XY} = Cov(X,Y)$, entonces **El coeficiente de correlación** de X,Y es:

$$\rho = \frac{\sigma_{XY}}{\sigma_X\sigma_Y} \tag{68}$$

Teorema: Si ρ es el coeficiente de correlación de X,Y entonces:

$$-1 \leq \rho \leq 1$$

, y $\rho = \pm 1$ si y solo si $Y = aX + b$ $a \neq 0$.

3.9. Esperanza Condicional

Tendremos además varias propiedades relacionadas a la esperanza y varianza de las v.a.

Definiremos la esperanza condicional de variables aleatorias con función de probabilidad conjunta:

$$\mathbb{E}(Y|x) = \sum_{y} y f(y|x) \tag{69}$$

$$\mathbb{E}(Y|x) = \int_{-\infty}^{\infty} y f(y|x)\, dy$$

Tenemos un teorema que nos dice:

$$\mathbb{E}(Y) = \mathbb{E}(\mathbb{E}(Y|X)) \tag{70}$$

Se recomienda leer el ejemplo 5.4.2 para comprender mucho mejor lo que se expresó.

Veamos otro teorema sobre la independencia de las variables:

Si X, Y son v.a. independientes tenemos:

$$\mathbb{E}(Y|x) = \mathbb{E}(Y)\ ;\ \mathbb{E}(X|y) = \mathbb{E}(X)$$

Definamos ahora Varianza Condicional:

$$Var(Y|x) = \mathbb{E}(Y^2|x) - [\mathbb{E}(Y|x)]^2 \tag{71}$$

Otros teoremas importantes son:

$$Var(Y) = \mathbb{E}_x[Var(Y|X)] + Var_x[\mathbb{E}(Y|X)]$$

$$\mathbb{E}(h(X,Y)) = \mathbb{E}_x\{E[h(X,Y)|X]\}$$

$$\mathbb{E}(g(X)Y|x) = g(x)\mathbb{E}(Y|x)$$

Si $\mathbb{E}(Y|x)$ es una función lineal de x entonces tenemos:

$$\mathbb{E}(Y|x) = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1) \tag{72}$$

Y además tenemos que:

$$\mathbb{E}_x[Var(Y|X)] = \sigma_2^2(1-\rho^2)$$

Veamos ahora una nueva distribución que nos permitirá hacer análisis con una distribución muy importante.

3.10. Distribución de la Normal Bivariada

Se dice que un par de variables aleatorias tienes distribución normal bivariada si se cumple que:

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{\Gamma\cdot -\frac{1}{2(1-\rho^2)}\right\} \tag{73}$$

Dónde Γ es:

$$\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]$$

Además tenemos un teorema importante que sería:

Teorema 5.4.7 Si $(X,Y) \sim BVN(\mu_1,\mu_2,\sigma_1,\sigma_2,\rho)$ tenemos:

$$X \sim N(\mu_1,\sigma_1);\ Y \sim N(\mu_2,\sigma_2)$$

Y ρ es el coeficiente de correlación entre X, Y.

3.10.1. Normal Bivariada Condicional

Definamos también las condicionales:

Si $(X,Y) \sim BVN(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$, entonces tenemos:

1. La condicional en $X = x$ es:

$$Y|x \sim N\left[\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1), \sigma_2^2(1-\rho^2)\right]$$

2. La condicional en $Y = y$ es:

$$X|y \sim N\left[\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y-\mu_2), \sigma_1^2(1-\rho^2)\right]$$

Podemos realizarnos la siguiente pregunta:

¿Para la Normal Bivariada si $\rho = 0$ podemos concluir que X, Y son independientes? Se realiza esta pregunta dado que permite evaluar sí se entendieron bien los conceptos de esta distribución conjunta.

Dentro del libro guía también existe el concepto de función generadora de momentos conjunta, lo cuál es la sección 5.5, pero por tiempos del curso y prioridades se deja como lectura este tema, de igual forma algo parecido se recobrará esta idea cuándo en el capítulo dle 4to parcial veamos las transformaciones y por consiguiente el método de la generadora de momentos.

4. Parcial 4

Muchas veces nos es interesante analizar funciones que dependen de otras variables aleatorias, y ver como es el comportamiento de esa nueva función $Y = f(X)$, dónde X es una variable aleatoria y f es una función.

Así de esta forma obtenemos varias tácticas para atacar este tipo de problemas:

4.1. Técnica de la Acumulada

Supongamos que X es una variable aleatoria con acumulada $F_X(x)$, y digamos que $Y = u(X)$, tendremos entonces que desearemos expresar la acumulada de Y en términos de la de X, lo cuál quedaría como:

$$F_Y(y) = P(u(X) \leq y) \tag{74}$$

Lo cuál en el caso continuo puede interpretarse de la forma:

$$F_Y(y) = \int_{-\infty}^{u^{-1}(y)} f(x)\, dx$$

De otra forma podemos que se dé el caso de que la X se mueva entre un x_1, x_2; $x_1 < x_2$ dónde ocurre que: $[u(X) \leq y] \equiv [x_1 \leq X \leq x_2]$, así nos resulta que:

$$F_Y(y) = \int_{x_1}^{x_2} f_X(x)\, dx = F_X(x_2) - F_X(x_1)$$

Y tenemos que se cumple que: $f_Y(y) = \frac{d}{dy} F_Y(y)$

Ver los ejemplos 6.2.1, 6.2.2, y 6.2.3.

De igual modo se puede aplicar este método y concepto a funciones de variables aleatorias, dónde estaremos aplicando la misma idea, solo que estableciendo restricciones que irán surgiendo conforme se desarrolle el modelo probabilístico. Recomiendo mirar el ejemplo 6.2.4. y el teorema 6.2.1.

4.2. Métodos de Transformación

También iremos desarrollando métodos de tal forma que podamos calcular de forma mucho más simple la función de probabilidad que es el alma de conocer la distribución y el comportamiento de neustro fenómeno.

Introduzcamos el primer tipo de transformaciones.

4.2.1. Transformaciones 1-1 (Inyectivas)

Muchas veces vamos a querer analizar esa dependenia de la función $Y = u(X)$, por lo que se propone analizar inicialmente la función pero no con variable aleatoria, sino solamente variable real, es decir:

Buscaremos dónde la ecuación $y = u(x)$ tiene solución **única**, para ir y hacer nuestros cálculos, los cuales partiremos dependiendo el caso:

Caso Discreto

Cuándo entremos a analizar un caso discreto de funciones de variable aleatoria tendremos que:

$$f_y(y) = f_x(w(y)),\ y \in B$$

[25] Aquí es dónde se ve la importancia de tener una funcion 1-1, para poder definir con exactitud la función a evaluar.

Dónde $w(y)$ es la función inversa[25] de $u(x) = y$, y B es el conjunto soporte de $f_y(y)$.

Demostremos esta propiedad:

Demostración: Sabemos que: $f_y(y) = \mathbb{P}[Y = y] = \mathbb{P}[u(X) = y] = \mathbb{P}[X = w(y)] = f_x(w(y))$, y sabemos que en todo este tiempo $y \in B$. $\square$

Ejemplo 6.3.1 Consideremos que $X \sim GEO(p)$, así resulta que

$$f_x(x) = pq^{x-1} \; x = 1,2,3,\cdots$$

Podemos considera la transformación: $Y = X - 1$, es decir que $u(x) = x - 1$, Por lo que: $w(y) = y + 1$, y por la propiedad demostrada nos queda que:

$$f_y(y) = f_x(y+1) = pq^y, \; y = 0,1,2,\cdots$$

, esto dado que sí x empezaba en 1, al restarle 1 (que es la transformación de Y) nos queda que la Y comienza desde el 0.

Podemos interpretar esta transformación como: *El número de fallas antes del primer acierto.*

Veamos el caso continuo que difiere un poco:

Caso Continuo

Para esta parte tendremos el mismo desarrollo de la anterior, es decir las mismas hipótesis, y tendremos ahora:

$$f_y(y) = f_x(w(y)) \left| \frac{d}{dy} w(y) \right| \tag{75}$$

La demostración de este caso sale parecida a la anterior, solo que deberemos de emepzar del método de la acumulada, y con este derivar (de ahí sale la derivada) y obtenemos la función común y corriente. Y además hay que analizar los casos dónde la función $w(y)$ sea creciente y decreciente.

Para efectos prácticos llamaremos a la derivada como el *Jacobiano*

Si tenemos que $Y = e^X$, dónde $F_x(x) = 1 - e^{-2x}$, $0 < x < \infty$ obtenemos que la transformación inversa es $x = \ln y$, por lo que resulta:

$$f_y(y) = f_x(\ln y)\frac{1}{y} = 2e^{-2\ln y}\left(\frac{1}{y}\right) = 2y^{-3}, \; 1 < y < \infty$$

Si observamos es igual e idéntico al ejemplo 1 del capítulo 6, podemos ahora definir una nueva distribución:

4.2.2. Distribución Lognormal

Decimos que $Y \sim LOGN(\mu, \sigma^2)$, si se da que: $\ln Y \sim N(\mu, \sigma^2)$, finalmente nos queda que:

$$f_y(y) = \frac{1}{y\sigma\sqrt{2\pi}} e^{-(\ln y - \mu)^2/2\sigma^2} \qquad 0 < y < \infty \tag{76}$$

Ahora, existirán las ocasiones dónde deseemos obtener las probabilidades de esta distribución, lo cuál nos resulta en:

$$F_y(y) = \mathbb{P}(Y \le y) = \mathbb{P}(\ln Y \le \ln y) = \mathbb{P}(X \le \ln y) \tag{77}$$

$$= \Phi\left(\frac{\ln y - \mu}{\sigma}\right)$$

De la Distribución Pareta con parámetros 1, 1 se puede obtener algo muy parecido a la Lognormal, la cuál en este caso se llama *Distribución Logística*.

4.2.3. Transformaciones que no son 1-1

Para este caso de las Transformaciones que no son uno a uno, usaremos la estrategia de ir definiendo subintervalos/regiones dónde la Variable Aleatoria si lo sea, y procedemos a sumar las contribuciones de cada elemento de la variable aleatoria $Y = u(X)$.

Recomiendo mirar de forma muy especial el ejemplo 6.3.7, 6.3.8, y muy importante el ejemplo 6.3.9.

4.2.4. Transformaciones Conjuntas

Podemos usar muchas veces transformaciones que dependan de muchas variables, por lo que daremos este teorema para entender que se debe de realizar:

Teorema 6.3.5. Si $\tilde{X}$ es un vector aleatorio, dónde $Y = u(\tilde{X})$, luego tenemos que:

$$f_Y(y_1, y_2, \cdots, y_k) = f_{\tilde{X}}(x_1, x_2, \cdots, x_k)$$

Dónde los x_i son las soluciones de $y = u(\tilde{x})$, e implícitamente dependen de los valores de $y_1, \cdots, y_k$.

Para el caso bidimensional recordemos que:

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix} \tag{78}$$

Es lo que llamamos **Jacobiano**.

Veamos que ocurre con el caso continuo:

Teorema 6.3.6. Si $\tilde{X}$ es un vector aleatorio, dónde $Y = u(\tilde{X})$, luego tenemos que:

$$f_Y(y_1, y_2, \cdots, y_k) = f_{\tilde{X}}(x_1, x_2, \cdots, x_k)|J|$$

Dónde los x_i son las soluciones de $y = u(\tilde{x})$, e implícitamente dependen de los valores de $y_1, \cdots, y_k$.

Se recomienda revisar y estudiar bien el ejemplo 6.3.12, y 6.3.13 especialmente.

4.3. Suma de Variables Aleatorias

Inicialmente podemos usar lo que conocemos como **Fórmula de Convolución**, sí tenemos que

$S = X_1 + X_2$, dónde son v.a. continuas con función de probabilidad conjunta $f(x_1, x_2)$, luego podemos derivar una fórmula general que sería:

$$f_S(s) = \int_{-\infty}^{\infty} f(t, s-t)\, dt$$

Sí se da la independencia podemos separar en las marginales internamente en la integral. Como tal en el curso no vimos esta fórmula, dado que esto va de la mano con la Transformada de Laplace, y eso es algo que no todos hemos visto, se deja de igual forma la recomendación de ver el ejemplo 6.4.1.

Supongamos que X_1, X_2 son variables aleatorias Gamma independientes, dónde $X_i \sim GAM(1, \alpha_i)$ luego, la conjunta sería:

$$f(x_1, x_2) = \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)} x_1^{\alpha_1 - 1} x_2^{\alpha_2 - 1} e^{-x_1 - x_2} \quad 0 < x_i < \infty$$

Definamos las siguientes transformaciones:

$$Y_1 = X_1 + X_2; Y_2 = \frac{X_1}{X_1 + X_2}$$

dónde sus transformaciones inversas serían:

$$x_1 = y_1 y_2 \; x_2 = y_1(1 - y_2)$$

así tendremos que el Jacobiano es:

$$J = \begin{vmatrix} y_2 & y_1 \\ 1-y_2 & -y_1 \end{vmatrix} = -y_1$$

Luego aplicando el método nos queda:

$$f(y_1,y_2) = \frac{(y_1y_2)^{\alpha_1-1}[y_1(1-y_2)]^{\alpha_2-1}e^{-y_1}|-y_1|}{\Gamma(\alpha_1)\Gamma(\alpha_2)}$$

Desarrollando y juntando nos queda que:

$$f_{y_1,y_2}(y_1,y_2) = \frac{y_2^{\alpha_1-1}(1-y_2)^{\alpha_2-1}y_1^{\alpha_1+\alpha_2-1}e^{-y_1}}{\Gamma(\alpha_1)\Gamma(\alpha_2)}$$

Dónde $(y_1,y_2) \in B = \{(y_1,y_2) : 0 < y_1 < \infty,\ 0 < y_2 < 1\}$, y la independencia de estas variables, se hereda de la independencia de las X_i, lo inicial es saber factorizarlas.

Podemos determinar que: $Y_1 \sim GAM(1, \alpha_1 + \alpha_2)$.

Se puede demostrar que la suma de Gammas independientes nos vuelve a arrojar otra Gamma, siempre y cuándo se conserve la primera componente, es decir en términos prácticos queda:

Si $X_i \sim GAM(\alpha, \beta_i)$ entonces la transformación $Y = \sum_{i=1}^{n} X_i \sim GAM\left(\alpha, \sum_{i=1}^{n} \beta_i\right)$

La prueba de esto se puede realizar con la definición y usando el Jacobiano, pero con el siguiente método será muy efectiva su realización.

4.3.1. Método de la Generadora de Momentos

Muchas veces tendremos los casos dónde se suman las variables aleatorias *independientes*[26], así que para poder determinar la distribución de la suma necesitaremos invocar la función generadora de momentos de la siguiente forma:

Sabemos que $Y = \sum_{i=1}^{n} X_i$, dónde cada X_i es independiente de los otros, luego sí deseamos conocer la forma de la generadora de momentos de Y será:

$$M_y(t) = \mathbb{E}[e^{yt}] = \mathbb{E}\left[e^{\left(\sum_{i=1}^{n} x_i\right)t}\right] = \mathbb{E}[e^{x_1t}]\cdots\mathbb{E}[e^{x_nt}]$$

$$= M_{x_1}(t)\cdots M_{x_n}(t)$$

Es decir en resumidas cuentas tenemos que si deseamos conocer la forma de distribución d euan suma de variables aleatorias independientes, solo nos preocuparemos por el producto de las generadoras de momentos de las ya conocidas.

Por ejemplo, si deseamos conocer la distribución de Poisson independientes:

Si $X_i \sim POI(\lambda_i)$, y tenemos que $Y = \sum_{i=1}^{n} X_i$, dónde cada X_i es independiente de las otras, luego usando el método de la generadora de momentos tendremos:

$$M_y(t) = M_{x_1}(t)\cdots M_{x_n}(t) = e^{\lambda_1(e^t-1)}\cdots e^{\lambda_n(e^t-1)}$$

$$= e^{\left(\sum_{i1}^{n} \lambda_i\right)(e^t-1)}$$

Es decir que la suma de Poisson independientes es una Poisson nuevamente pero de parámetro la suma, es decir:

[26] El hecho de ser independientes es algo muy importante

$$Y \sim POI\left(\sum_{i=1}^{n} \lambda_i\right)$$

Se deja de tarea hacer la prueba de:

- Normales Independientes.
- Exponenciales Independientes.
- Gammas Independientes.
- ¿Si hago la división de dos normales estándar qué me arroja?
- Si $X_1 \sim GAM(\theta,\kappa)$, $X_2 \sim EXP(\theta)$, ¿Cómo distribuye $Y = X_1 + X_2$?
- Demuestre que si $X \sim GAM(\theta,1)$, entonces $X \sim EXP(\theta)$.

4.4. Estadísticos de Orden

Legalmente este tema por tiempo no se dió en el curso, así que daré el preámbulo del temario y espero más adelante anexarlo de forma completa en el resumen.

En muchas ocasiones desearemos analizar el orden de ocurrencias de ciertos eventos, por ejemplo la vida de una pila, en un grupo de 4 pilas, puede que una dure 3 semanas(digamos X_1), otra 1 semana (digamos X_2), otra 2 semanas (digamos X_3) y la otra 5 semanas (digamos X_4), si los organizamos mediante ocurrencias obtendremos que nos queda en orden:

$$X_2, X_3, X_1, X_4$$

Llamaremos a este nuevo orden definido como $Y_1 = X_2, Y_2 = X_3, Y_3 = X_1, Y_4 = X_4$, a esas variables Y_i son las que conoceremos como **Estadísticos de Orden**.

En el mayor de los casos para los efectos de este curso usaremos muestras aleatorias, o a lo sumo variables independientes, dado que el ser no independiente es mucho más complejo por todas las contribuciones que pueden llegar a ocurrir.

En general podemos hallar la distribución conjunta de estos estadísticos mediante todas las posibilidades que pueden tomar, en este caso 4![27], así nos resulta que:

$$f(y_1,\cdots,y_4) = 4! f(y_1) f(y_2) f(y_3) f(y_4)$$

Dónde $y_1 < y_2 < y_3 < y_4$, y es 0 en otro caso.

De igual forma podemos calcular marginales de cada estadístico de orden, para ver la distribución de probabilidad de cada uno de estos, por ejemplo:

$$f(y_1) = \int_{y_1}^{\infty}\int_{y_2}^{\infty}\int_{y_3}^{\infty} 4! f(y_1) f(y_2) f(y_3) f(y_4)\, dy_2 dy_3 dy_4$$

Y a partir de todo esto, podemos venir y realizar distintos cálculos, como por ejemplo ¿Cuál es la probabilidad de que el primer estadístico de orden cumpla cierta propiedad? Para mayor profundidad del tema, revisar la sección 6.5 del libro de Bain.

Para los que llegaron hasta aquí, espero hayan sido gratas estas notas elaboradas durante el curso de Probabilidad, un abrazo y saludo fraternal.

Jose David Palacio Arias

[27] Recordemos que por combinatoria se toman de 4 elementos grupos de 1, y eso resulta en 4!, dado que aquí sí me interesa el orden formado.

5. Bibliografía sugerida

Encontraremos varios textos a seguir durante el curso, siendo el principal y libro guía:

- Bain, Lee; Engelhardt, Max. *Introduction to Probability and Mathematical Statistics*

De igual modo existen más libros que pueden servirnos para orientar nuestro estudio en el campo de la probabilidad:

- Blanco, Lilina. *Probabilidad*.
- Hogg, Robert; McKean, Joseph; Craig, Allen. *Introduction to Mathematical Statistics*.
- Rincón, Luis. *Introducción a la Probabilidad*.

Son varios libros con los que he tenido contacto, y puedo decir que son libros que explican bien el curso de Probabilidad, el Liliana Blanco es un poco más duro, por lo que tomar ese de uno, sería con cuidado. Pero de igual modo trae una cantidad de ejercicios interesante para uno desafiarse y prepararse para el examen, se toma como libro base y guía el Bain, dado que el profesor que con más frecuencia da el curso de Inferencia Estadística es el Doctor Daya Krishna Nagar, y es él quién más gusta y adora usar el libro mencionado, entonces para evitar problemas de notación y fórmulas se decide seguir el mismo libro del curso siguiente. Esto por experiencia del autor de estas notas es que recomienda eso.