

Resumen de Estadística I

Jose Luis Nogueira Alonso.

Facultad de informática.

Prólogo

Este documento contiene un resumen del temario de esta asignatura, es muy importante que tengas en cuenta que todo su contenido lo he enfocado como complemento para el estudio de esta. Me he basado en los temarios originales de la asignatura, pero ten en cuenta que he omitido algunas partes del temario, por tanto te recomiendo que completes este texto apoyándote en algún libro para ampliar aquellas partes en las cuales no he entrado en profundidad.

También es importante que tengas en cuenta que es un resumen enfocado principalmente como complemento para el estudio de la asignatura, por ese motivo es posible que en algunos casos haya sido poco fiel a las definiciones originales de algunos puntos, y lo haya definido con un lenguaje poco técnico, para de este modo hacer más comprensibles algunos aspectos.

Espero que te pueda ser de alguna utilidad este documento y en la medida de lo posible me comuniques cualquier posible fallo que pueda haber cometido en su confección, por otra parte si deseas ampliar o modificar alguna parte del mismo no tendré ningún problema en facilitarte los fuentes para su modificación. Si así lo deseas puedes ponerte en contacto con migo en la dirección de correo **scero@ono.com**.

Un saludo, Jose Luis Nogueira Alonso.

Índice

1. Experimentos aleatorios	3
1.1. Usos de la probabilidad en la informática	3
1.2. Modelo matemático	3
1.2.1. Espacio muestral (Ω)	3
1.2.2. Álgebra de sucesos (Λ)	3
1.2.3. Probabilidad (P)	4
1.3. Técnicas de combinatoria	4
1.3.1. Contar con los dedos de las matemáticas.	4
1.3.2. Contar palabras.	5
1.4. Contar subconjuntos de k elementos	5
1.5. Los instrumentos del probabilista	5
1.6. Método de cálculo por complementario	6
2. Modelos aleatorios dinámicos	7
2.1. Diferencia entre modelos estáticos y modelos dinámicos	7
2.2. Probabilidad condicionada	7
2.2.1. Propiedades de la probabilidad condicionada	7
2.2.2. Pasos para la resolución de un problema dinámico simple	7
2.2.3. Cálculo dinámico de la probabilidad de la intersección	7
2.3. Diagramas de árbol	7
2.3.1. Formación de los diagramas de árbol	8
2.3.2. Fórmula de la probabilidad total	8
2.3.3. Fórmula de Bayes	8
2.4. Independencia entre sucesos	8
2.4.1. Definiciones	8
2.4.2. Independencia entre varios sucesos	8
2.4.3. Propiedades de los sucesos independientes	9
2.5. Métodos recursivos para calcular probabilidades	9
2.5.1. Definición	9
2.5.2. Diagramas de estado	9
2.5.3. Ejemplo de un problema con diagramas de estados	9
2.5.4. Circuitos eléctricos	10
3. Variables aleatorias discretas	11
3.1. Conceptos de las variables aleatorias	11
3.1.1. Métodos para definir una variable aleatoria	11
3.1.2. Racha	11
3.2. Distribución de una variable aleatoria discreta	11
3.2.1. Distribución de una función $Y = g(x)$	11
3.2.2. Distribución de una variable aleatoria condicionada por un suceso	11
3.3. Esperanza matemática de una variable discreta	11
3.3.1. Esperanza de una función de una variable aleatoria	11
3.3.2. Esperanza de una función lineal	12
3.4. Momentos y varianza	12
3.5. Funciones generatrices	12
3.6. Modelos de distribuciones discretas	12
3.6.1. Distribución de Bernoulli	12
3.6.2. Distribución de Bernoulli	12
3.6.3. Distribución geométrica	13
3.6.4. Distribución de Poisson	13
4. Vectores aleatorios discretos	14
4.1. Introducción	14
4.2. Función de probabilidad conjunta	14
4.2.1. Cálculo con la función de probabilidad conjunta	14
4.3. Distribuciones marginales	14
4.4. Valor esperado de una función $f(X, Y)$	15
4.4.1. Propiedades de la esperanza matemática	15

4.5. Distribuciones condicionadas	15
4.6. Variables independientes	15
4.6.1. Conjuntos y sucesiones de variables independientes	16
4.6.2. Esperanza del producto de variables independientes	16
4.6.3. Varianza de la suma de variables independientes	16
4.7. Desigualdad de Chebishev	16
4.8. Ley de los grados números	16
4.9. Métodos de cálculo por condicionamientos	16
4.9.1. Método para calcular esperanzas por condicionamiento	16
4.9.2. Método para calcular probabilidades por condicionamiento	16
5. Variables aleatorias continuas	17
5.1. Introducción	17
5.2. Probabilidades geométricas	17
5.3. Variables aleatorias continuas	18
5.4. Esperanza matemática de una variable continua	18
5.5. Esperanza matemática de una función	19
5.6. Distribuciones con variables continuas	19
5.6.1. Distribución uniforme en un intervalo	19
5.6.2. Distribución exponencial	19
5.6.3. Distribución normal	20
5.6.4. La distribución “natural” de los números	20
5.7. Variables mixtas	21
5.8. Funciones de distribución	21
5.9. Tres modelos de funciones de distribución	21
5.9.1. Variables discretas	21
5.9.2. Variables continuas	22
5.9.3. Variables mixtas	22
5.10. Función de distribución de $Y = g(X)$	22
5.11. Distribución de una variable condicionada por un suceso	23
6. Vectores aleatorios continuos	24
6.1. Introducción	24
6.2. Función de distribución conjunta	24
6.3. Distribuciones marginales	24
6.4. Variables independientes	24
6.5. Funciones con densidad condicionada	25
6.6. Modelos dinámicos	25
6.7. Densidad de la suma de variables independientes	25
6.8. Densidad de una función $g(X, Y)$	25
6.9. Esperanza matemática de una función $g(X, Y)$	25
6.10. Covarianza entre X e Y	26
6.11. Cálculo de la densidad del vector (U, V) función de (X, Y)	26
7. Simulación estocástica	27
7.1. Introducción	27
7.2. Generadores de números aleatorios y pseudoaleatorios	27
7.3. Generadores de congruencias lineales (GCL)	27
7.4. Transformaciones generales	28
7.4.1. Algoritmo de Von Neumann	28
7.5. Simulación de variables normales	28
7.5.1. Por método <i>aceptar-rechazar</i>	28
7.5.2. Por método <i>BOX-MULLER</i>	29
7.5.3. Algoritmo de MARSAGLIA	29
7.6. Simulación de variables discretas	29
7.6.1. Algoritmo por cuantiles	29
7.6.2. Algoritmo de BERNOULLI de parámetro p	29
7.6.3. Algoritmo de simulación de una variable <i>binomial</i> de parámetro n y p	30
7.7. Simulación de permutaciones al azar	30

1. Experimentos aleatorios

1.1. Usos de la probabilidad en la informática

Los principales usos que se dan en el campo de la ingeniería informática a la probabilidad son:

- Análisis de algoritmos. Para poder comparar distintos algoritmos y medir su eficacia es necesario evaluar la velocidad del mismo ante “inputs” aleatorios.
- Generación de números aleatorios. Tarea muy importante que deben realizar los computadores y utilizados en diversas aplicaciones.
- Simulación. Para realizar modelos de simulación reales, es necesario tener en cuenta una serie de valores aleatorios dentro de un conjunto.
- Transmisión de información. Siempre que se trabaje con un modelo de transmisión de información hay que tener muy en cuenta el ruido, ya que este es inevitable.
- Análisis de patrones. Aquellos modelos de análisis de patrones de imágenes.

1.2. Modelo matemático

La rama de las matemática (conjunto de teorías) que estudia los modelos aleatorios o estocásticos se denomina *calculo de probabilidades*.

El modelo matemático con el cual trabajaremos de momento se puede resumir en la terna (Ω, Λ, P) . Cada elemento de esta terna serán estudiados con detenimiento este punto.

1.2.1. Espacio muestral (Ω)

Se conoce con este nombre al conjunto de todos los posibles resultados que puedan derivarse de un experimento aleatorio. Un ejemplo muy sencillo sería si realizamos un experimento con una moneda; su espacio muestral sería el siguiente: $\Omega = \{\odot, \otimes\}$.

El mayor problema que puede encontrarse para determinar el espacio muestral en un problema es la *simetría*¹.

Un ejemplo con simetría sería el caso de tener dos urnas y dos pelotas, de manera que se deben introducir aleatoriamente las pelotas en la urnas. En un principio se podría pensar que el espacio muestral es el siguiente:

$$\left\{ \begin{array}{|c|c|} \hline \circ & \circ \\ \hline \end{array} , \begin{array}{|c|c|} \hline \circ & \circ \\ \hline \end{array} , \begin{array}{|c|c|} \hline & \circ \\ \hline \end{array} \right\}$$

Pesar esto sería un error, pues la situación en que cada pelota está en una urna puede lograrse poniendo la primera pelota en la urna izquierda y la segunda pelota en la urna derecha o viceversa; por lo cual serían dos situaciones en lugar de una.²

Los elementos del espacio muestral se pueden representar simbólicamente, como se ha visto en los ejemplos anteriores, o bien de una manera formal con la nomenclatura: $\Omega = \{w_1, w_2, \dots, w_n\}$

1.2.2. Álgebra de sucesos (Λ)

Lo primero que debe quedar claro es que cuando se hace referencia a *suceso*, se hace referencia a un acontecimiento que puede ocurrir o no ocurrir, como resultado de un experimento.

Para trabajar con sucesos se utiliza las operaciones lógicas de conjuntos, tales como $\cup, \cap, negacion(\Lambda^c)$.

Existen diversos tipos de sucesos tales como:

- Sucesos seguros. Son aquellos que ocurrirán en todos los resultados.
- Sucesos imposibles. Son aquellos que nunca llegarán a ocurrir.
- Sucesos contrarios. Se obtienen como negación de un suceso.
- Sucesos simples. Son aquellos formados por un conjunto de un único elemento del espacio muestral.
- Sucesos compuestos. Son aquellos formados por un conjunto de mas de un elemento del espacio muestral.

La clase de todos los sucesos se designa por Λ , y *está formada por el conjunto de las partes de Ω* .

¹ Aquellos casos en los que los resultados parecen similares, pero no el mismo.

² Para facilitar la comprensión de esto es recomendable numerar las pelotas, de esta manera se verá que no es lo mismo la posición 1 - 2, que la posición 2 - 1.

1.2.3. Probabilidad (P)

Es la medida de la facilidad con que ocurren los posibles sucesos. La probabilidad se fundamenta sobre los axiomas de KOLMOGOROV, que son los siguientes:

- $P(\Omega)=1$
- Si A y B son disyuntos: $P(A \cup B) = P(A) + P(B)$ (Aditiva).

Asignación de probabilidades Es necesario asignar un valor de probabilidad a cada posible suceso de un modelo, como es evidente en muchos modelos no todos los sucesos tienen porque tener las mismas posibilidades de salir como resultado. Por estos motivos es necesario formalizar estas asignaciones de probabilidades entre los sucesos.

En un modelo *finito* para definir P se debe dar un valor a cada suceso *simple* de Λ de manera que se cumpla $P(\Lambda) = \sum_{w \in \Lambda} P(w)$.

Regla de Laplace Laplace estudia los *modelos uniformes*, estos son aquellos que son *finitos* y además *simétricos o intercambiables*³.

En la regla de Laplace se hacen las siguiente definiciones:

- Casos favorables. Son todos aquellos sucesos simples.
- Casos posibles. Son todos los posibles sucesos, tanto simples como compuestos.

Para los **casos favorables** (sucesos simples), si todos fuesen posibles se aplica la regla:

$$P(\Lambda) = \frac{\sum_{w \in \Lambda} \#(\Lambda)}{\#(\Omega)} = \frac{1}{\#(\Omega)}$$

Con $\#(\Lambda)$ se indica el número de elementos que forma el suceso, en este caso particular, al estar trabajando con sucesos simples este valor siempre es 1. Así mismo $\#(\Omega)$ indica el número de elementos de Ω (también llamado *cardinalidad*).

Para los **sucesos compuestos** se aplica la misma regla, con la diferencia de que $\#(\Lambda)$ no será igual a uno, sino al número de elementos del suceso compuesto.

Otras propiedades que se pueden aplicar en probabilidad

- $P(A^c)=1-P(A)$
- $P(A-B)=P(A)-P(A \cap B)$ // $P(B-A)=P(B)-P(A \cap B)$
- $P(A \cup B)=P(A)+P(B)-P(A \cap B)$
- $P(A^c \cup B) = P(B - A)$
- Dados n sucesos $A_i : 1 \leq i \leq n$, se cumple que $P(\bigcup_{i=1}^n A_i) \leq \sum_{i=1}^n P(A_i)$

1.3. Técnicas de combinatoria

Como ya se ha visto en la regla de Laplace en muchas ocasiones resulta de gran interés conocer el número de elementos que forma un conjunto *finito*. Esto que en casos sencillos se puede hacer directamente, pero según aumenta la complicación se necesitan técnicas para poder realizar este conteo. A continuación se exponen tres técnicas para resolver este problema.

1.3.1. Contar con los dedos de las matemáticas.

Consiste en establecer una relación biyectiva entre los elementos que forman el conjunto a contar y el conjunto patrón $\{1, 2, \dots, n\}$

Ejemplos 1. Se necesitan contar todos los números enteros que hay entre 3 y 25, ambos inclusive.

$$\begin{array}{cccccc} 3 & 4 & 5 & \dots & 24 & 25 \\ \updownarrow & \updownarrow & \updownarrow & & \updownarrow & \updownarrow \\ 1 & 2 & 3 & \dots & 22 & 23 \end{array}$$

2. Contar los números pares mayores de 1 y menores de 123.

$$\begin{array}{ccccccc} 2 \cdot 1 = 2 & 2 \cdot 2 = 4 & 2 \cdot 3 = 6 & \dots & 2 \cdot 60 = 120 & 2 \cdot 61 = 122 \\ \updownarrow & \updownarrow & \updownarrow & & \updownarrow & \updownarrow \\ 1 & 2 & 3 & \dots & 60 & 61 \end{array}$$

³Son intercambiables aquellos modelos en los cuales los posibles resultados *simples* son igual de factibles. Por ejemplo una moneda, es igual de factible el resultado *cara* o el resultado *cruz*.

1.3.2. Contar palabras.

Esta técnica consiste en interpretar los elementos del conjunto que se desean contar como *secuencias* o *palabras* formadas por símbolos de un alfabeto, de modo que cumplan ciertas restricciones.

Ejemplos 1. Se desea contar el número de palabras distintas que pueden hacerse servir en un *password* de cuatro posiciones, manera que el alfabeto a utilizar es $\{A, B, C, 0, 1, 2\}$. Además la primera letra del *password* no puede ser un número.

Como primer apartado se hace una representación gráfica del problema, en la cual se anotan las posibles restricciones.

$$\begin{array}{cccc} \odot & \odot & \odot & \odot \\ \text{alfabetico} / \neq \text{resto} & \neq \text{resto} & \neq \text{resto} & \neq \text{resto} \end{array}$$

Se comienza por la celda con más restricciones, en este caso particular la primera, y se evalúa cuantos valores puede haber en ella, en este caso 3. Como todas las demás celdas tienen las mismas restricciones se pasa a la siguiente; esta puede tener cualquier valor excepto el ya utilizado en la primera casilla, por tanto puede tener 5 valores. Repitiendo este proceso obtenemos que el resultado del problema es $3 \cdot 5 \cdot 4 \cdot 3 = 180$ palabras.

1.4. Contar subconjuntos de k elementos

Consiste en utilizar para el conteo, los llamados, números combinatorios $\binom{n}{k}$; esta expresión es leída como “ n sobre k ”.

El resultado escalar de un número combinatorio sería el resultante de la siguiente formula:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Ejemplo Se desea hallar el número de manos distintas que se pueden encontrar en el juego de poker. Para resolver este problema se sabe que una baraja tiene cuatro palos, y cada uno de estos trece cartas; así mismo se sabe que una mano es un conjunto de cinco cartas.

Aplicando esta técnica, se debe pensar que hay que coger cinco cartas de todas las que tiene la baraja; a tales efectos el resultado sería:

$$4\text{palos} \cdot 13\text{cartas} = 52\text{cartas}$$

$$\binom{52}{5}$$

1.5. Los instrumentos del probabilista

La probabilidad, a pesar de tener una base teórica, es evidentemente práctica, por ese motivo debe ser orientada hacia la resolución de problemas prácticos. Como los problemas que se pueden plantear pueden ser muy diversos, siempre se intentan reducir estos a una serie de modelos prefijados y típicos en probabilidad. Estos modelos son los siguientes:

- Monedas. Sirve para modelar pruebas repetidas en idénticas condiciones, es frecuente el uso de monedas cargadas, es decir, $P \neq 50\%$. Solo se pueden utilizar en modelos de dos posibles resultados.
- Dados. Son más generales que las monedas, pues pueden tener n caras.
- Bolas y urnas. Sirven para infinidad de modelos. Se pueden usar de los siguientes modos:
 - Extraer bolas de una urna que las contiene, dividido en:
 - Con reemplazamiento, la bola extraída se devuelve a la urna antes de la siguiente extracción.
 - Sin reemplazamiento, la bola extraída se deja fuera.
 - Repartir las bolas entre varias urnas vacías.
- Ruletas. En este modelo se considera que la ruleta se encuentra en una circunferencia de longitud l , para hacer el modelo más similar a los términos de probabilidad.

1.6. Método de cálculo por complementario

En el momento de realizarse el recuento de *casos favorables*, este se simplifica mucho si es una *conjunción* de condiciones, dicho de otra manera, si los elementos del suceso deben cumplir todas las condiciones que se impongan simultáneamente.

Por el contrario, el recuento es más difícil cuando el suceso se define como *disyunción* de condiciones.

Esto nos lleva a pensar que en muchas ocasiones es más sencillo calcular el complementario de un suceso, que el mismo.

2. Modelos aleatorios dinámicos

2.1. Diferencia entre modelos estáticos y modelos dinámicos

Se conocen como *modelos estáticos* aquellos en los cuales aparece el “Azar” una sola vez en todo el experimento. Este tipo de modelos ya han sido tratados en el tema anterior.

Son *modelos dinámicos* aquellos que hacen referencia a experimentos que a su vez están divididos en pequeños subexperimentos, ordenados en el tiempo. De manera que los resultados de los primeros subexperimentos influyen en los resultados de los siguientes.

Si pretendemos aplicar un modelo estático a un problema dinámico, encontraremos las siguientes dificultades:

- El problema no será intuitivo, pues se debería interpretar que el “Azar” solo actúa una vez. Lo cual hace que tengamos que trabajar con un espacio muestral extraño.
- Se hace difícil conseguir un modelo simétrico, así como también se hace difícil evaluar la probabilidad de los diferentes sucesos.

2.2. Probabilidad condicionada

Es un tipo de probabilidad utilizado en los modelos dinámicos, se representa gráficamente de la siguiente manera:

$P(\cdot|A)$ Esta expresión se lee como: “probabilidad condicionada por A” o “probabilidad condicionada porque a ocurrido A”.

Si tenemos dos sucesos A y B los cuales son sucesos de un experimento de probabilidad en un modelo dinámico, la probabilidad de que B sea condicionada por A , viene dada por:

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

2.2.1. Propiedades de la probabilidad condicionada

$$P(A|A) = 1$$

$$P(B^c|A) = 1 - P(B|A)$$

$$P(B_1 \cup B_2|A) = P(B_1|A) + P(B_2|A)$$

2.2.2. Pasos para la resolución de un problema dinámico simple

Estos pasos que se reflejan a continuación no son pasos formales, sino una pequeña guía sobre como resolver problemas con modelos dinámicos muy simples.

- Hay que dividir y tener claras los diferentes subexperimentos que se deben realizar.
- Hallar las probabilidades de los elementos del experimento A .
- Hallar las probabilidades de los elementos del experimento B .
- Aplicar la fórmula $P(B|A) = \frac{P(A \cap B)}{P(A)}$ para hallar la probabilidad condicionada.
- Repetir estos pasos para tantos subexperimentos como haya.

2.2.3. Cálculo dinámico de la probabilidad de la intersección

En este apartado se muestra una formalización de la idea intuitiva que se reflejaba en el apartado anterior. De esta manera se formaliza el hecho de que “la probabilidad de que ocurran A y B simultáneamente es igual a la probabilidad de que ocurra A , por la probabilidad de que ocurra B si A a ocurrido”. Esto se puede expresar con la fórmula:

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

2.3. Diagramas de árbol

Mediante estos diagramas se puede reflejar de forma gráfica los diferentes resultados que se pueden obtener en los distintos subexperimentos de un modelo dinámico. Estos diagramas son de gran utilidad en la resolución de muchos problemas, pues dan una visión muy clara sobre el problema

2.3.1. Formación de los diagramas de árbol

En el primer nivel debe haber un único vértice (representado normalmente por un punto), desde el cual parten tantas ramas como resultados pueda tener el primer subexperimentos.

En el segundo nivel se encuentran tantas etiquetas como posibles resultados pueda haber tenido el primer subexperimentos. Desde cada una de estas etiquetas parten tantas ramas como posibles resultados pueda tener el segundo subexperimentos desde el punto del árbol que corresponde.

Estos pasos se repiten hasta que estén reflejados todos los subexperimentos en el árbol.

Una vez todo el árbol está dibujado es necesario indicar en cada rama la probabilidad que esta tiene de cumplirse.

Estos árboles se deben utilizar acorde con las siguientes reglas, para la resolución de problemas:

- Para obtener la probabilidad de un resultado final⁴ se deben multiplicar las probabilidades de todas las ramas que llevan a este resultado, desde la raíz⁵.
- Si hay varias ramas del último nivel que poseen resultados iguales⁶, se puede afirmar que la probabilidad de que ese resultado ocurra, es igual a la suma de las probabilidades de que ocurra cada uno de ellos.

2.3.2. Fórmula de la probabilidad total

Expresa matemáticamente la idea del último punto del apartado anterior.

$$P(B) = \sum_{i=1}^n P(A_i)P(B|A_i)$$

2.3.3. Fórmula de Bayes

Mediante esta fórmula se puede calcular la probabilidad de un proceso anterior, conociendo la probabilidad de un proceso final. En esta fórmula se conoce $P(A)$ como *probabilidad a priori*, y $P(A/B)$ como *probabilidad a posteriori*.

$$P(A_j|B) = \frac{P(A_j)P(B|A_j)}{\sum_{i=1}^n P(A_i)P(B|A_i)}$$

2.4. Independencia entre sucesos

2.4.1. Definiciones

- Un suceso A se denomina favorable⁷ a B si se cumple $P(B|A) > P(B)$.
- Se dice que A es desfavorable a B si $P(B|A) < P(B)$.
- Si $P(B|A) = P(B)$, se dice que B es independiente de A .
- Si $P(B|A) \neq P(B)$, se dice que B es dependiente de A .
- Los sucesos A y B son independientes si $P(A \cap B) = P(A)P(B)$.

2.4.2. Independencia entre varios sucesos

Para probar que n sucesos son independientes, se debe aplicar el siguiente criterio:

$$\forall k : 2 \leq k \leq n$$

$$P(A_{i1} \cap A_{i2} \cap \dots \cap A_{ik}) = P(A_{i1}) \cdot P(A_{i2}) \cdot \dots \cdot P(A_{in})$$

Seguidamente se ilustra un ejemplo que las normas que deben cumplir tres sucesos (A, B, C) para ser independientes entre ellos:

$$P(A_1 \cap A_2) = P(A_1)P(A_2)$$

$$P(A_1 \cap A_3) = P(A_1)P(A_3)$$

$$P(A_2 \cap A_3) = P(A_2)P(A_3)$$

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3)$$

⁴Son aquellos resultados que se encuentran en una rama de último nivel.

⁵Cada resultado final únicamente puede tener un camino desde la raíz.

⁶A cada uno de estos resultados se ha llegado por un camino único.

⁷favorable no implica que A sea causa de B , sino que B es más frecuente cuando A ocurre.

2.4.3. Propiedades de los sucesos independientes

En caso de tener dos sucesos independiente (A y B), se cumplen los siguientes criterios:

$$A_1^c \cap A_2^c = (A_1 \cup A_2)^c$$

$$P([A_1 \cup A_2] \cap A_3) = P(A_1 \cup A_2)P(A_3)$$

2.5. Métodos recursivos para calcular probabilidades

2.5.1. Definición

Se llaman métodos recursivos aquellos utilizados para calcular la probabilidad de un experimento que se puede repetir infinitas veces.

Para el cálculo de la probabilidad se utiliza un algoritmo recursivo, similar a los empleados en programación.

2.5.2. Diagramas de estado

Se ha decidido comenzar la explicación de los métodos recursivo por este apartado, debido a que es un método gráfico, por el cual se pueden obtener una serie de datos útiles para el calculo con métodos recursivos.

Un diagrama de estados está formado por una serie de estados (Inicio, E_0 , E_1 , E_2 , ..., E_n , Final); y por una serie de flechas que indican las transiciones entre ellos.

Normas para construir los diagramas de estado:

- Todos los estados se representa por un círculo, dentro del cual está su nombre y, si se desea, cualquier otro dato que sirva para identificar su utilidad.
- Se colocan los estados Inicio y Final, los cuales siempre estarán en cualquier diagrama.
- Se colocan todos los estados que necesita el problema, llamando a cada uno de ellos E_i , siendo i el número de estado.
- Se deben unir todos los estados, mediante flechas, mostrando las transiciones posibles y lógicas entre ellos. Es importante tener en cuenta que todo estado tendrá tantas flechas como posibles resultados se puedan derivar de él.
- Se colocará sobre cada flecha, o transición de estado, la probabilidad de que esta ocurra.
- En caso de que de un estado únicamente parta una flecha, eso quiere decir obligatoriamente que la probabilidad de que eso ocurra es uno.

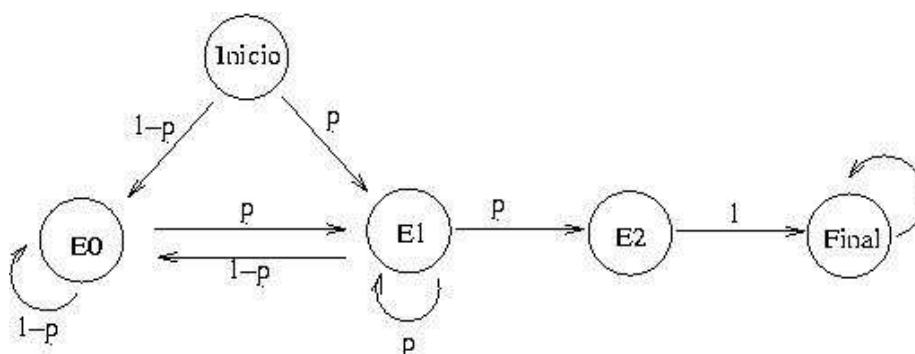
Con este sistema se representan gráficamente las probabilidades y posibles caminos que un problema puede seguir.

Una vez se ha obtenido el diagrama de estados, viene la parte más complicada, que consiste en deducir, por métodos matemáticos la *ecuación recursiva de las probabilidades* del último estado (que es el que deseamos conseguir).

2.5.3. Ejemplo de un problema con diagramas de estados

Enunciado Lanzamos repetidas veces una moneda, que tiene probabilidad de cara igual a p , hasta que aparecen dos caras consecutivas. ¿Cuál es la probabilidad de que hagamos n lanzamientos?

Para este enunciado se debería diseñar un diagrama de estados como el siguiente:



Teniendo el diagrama de estados, se debe formalizar matemáticamente la probabilidad de cada estado, respecto al lanzamiento concreto. Para hacer esto se utilizará la nomenclatura $P_k(n)$, en la cual k corresponde al estado al cual nos referimos y n a la tirada. Siguiendo esta nomenclatura y observando el diagrama, se puede afirmar que *los estados iniciales*⁸ son los siguientes:

$P_0(1) = 1 - p$... Probabilidad de estar en el estado cero, tras la primera tirada.

$P_1(1) = p$... Probabilidad de estar en el estado uno, tras la primera tirada.

$P_2(1) = 0$... Probabilidad de estar en el estado dos, tras la primera tirada (es un suceso *imposible*).

Tras haber formalizado los estados iniciales, se debe formalizar la probabilidad de estar en los diferentes estados para la tirada $(n+1)$. Para hacer esto se debe tener en cuenta, en que estado se puede estar en n y que probabilidad hay de pasar de ese estado al que vamos a estudiar. Con esta formalización se obtiene:

1. $P_0(n+1) = (1-p)P_0(n) + (1-p)P_1(n)$... Se realiza una suma pues es posible llegar a este estado de dos maneras, siguiendo en el mismo, o desde el estado E_1 .
2. $P_1(n+1) = P \cdot P_0(n)$... Solo se puede llegar a este estado desde el E_0 .
3. $P_2(n+1) = P \cdot P_1(n)$... Solo se puede llegar a este estado desde el E_1 .

Una vez obtenido este conjunto de ecuaciones, se debe obtener la *ecuación recursiva de probabilidades*. Dicha ecuación está compuesta por diferentes términos que solo pueden ser P_i , siendo i constante en toda la ecuación; y la variable P . Todos los términos deben colocarse de manera que queden igualados a cero.

Para realizar la ecuación antes descrita se elige la ecuación 1, y en ella se substituyen los valores de P_0 , por sus correspondiente. Para ello se despeja primero P_0 en la ecuación 2, con lo que se obtiene:

$$P_0(n) = \frac{1}{P} P_1(n+1)$$

En la ecuación 1, la variable P_0 , se encuentra con forma $P_0(n)$ y $P_0(n+1)$; de manera que por la última ecuación obtenida se deduce de manera inmediata:

$$P_0(n+1) = \frac{1}{P} P_1(n+2)$$

Con estas dos nuevas ecuaciones, ya se puede realizar la substitución en la ecuación 1, quedando esta con la forma:

$$\frac{1}{P} P_1(n+2) = (1-P) \frac{1}{P} P_1(n+1) + (1-P) P_1(n)$$

$$\frac{1}{P} P_1(n+2) - (1-P) \frac{1}{P} P_1(n+1) - (1-P) P_1(n) = 0$$

$$P_1(n+2) - (1-P) P_1(n+1) - P(1-P) P_1(n) = 0 \text{ (ecuación recursiva de probabilidades de } P_1 \text{)}.$$

El objetivo de este ejercicio es obtener dicha ecuación sobre P_2 por ser esta la probabilidad del último estado, para conseguir esto se actúa de manera similar a como se ha echo hasta ahora; y se substituye en la ecuación recursiva de probabilidades de P_1 , estos valores por los de P_2 .

El primer paso para esto es despejar P_1 en la ecuación 3, y deducir todos los términos para la substitución, procedemos así:

$$P_1(n) = \frac{1}{P} P_2(n+1)$$

$$P_1(n+1) = \frac{1}{P} P_2(n+2)$$

$$P_1(n+2) = \frac{1}{P} P_2(n+3)$$

Se substituye P_1 por su valor en la ecuación, y de este modo ya se consigue el resultado final:

$$\frac{1}{P} P_2(n+3) - (1-P) \frac{1}{P} P_2(n+2) - P(1-P) \frac{1}{P} P_2(n+1) = 0$$

$$P_2(n+3) - (1-P) P_2(n+2) - P(1-P) P_2(n+1) = 0$$

Gracias a esta ecuación recursiva, junto con las condiciones iniciales $P_2(1) = 0, P_2(2) = P^2$ permiten programar el cálculo de $P_2(n)$ para cualquier n .

2.5.4. Circuitos eléctricos

Esta es otra forma de mostrar combinaciones lógicas de varios elementos, las normas para el cálculo de la probabilidad de estos elementos son las siguientes:

- Paralelo

$$P(ab) = 1 - (1-x)(1-y)$$

- Serie

$$P(ab) = x \cdot y$$

⁸Entendiendo por estados iniciales, la probabilidad de estar en cada estado después de la primera tirada.

3. Variables aleatorias discretas

3.1. Conceptos de las variables aleatorias

Una variable aleatoria es una representación numérica de un resultado de fenómenos aleatorios.

Una variable aleatoria definida sobre el espacio de probabilidad (Ω, P) es la función $X : \Omega \rightarrow R$ o $F(\Omega)$.

Son variables aleatorias discretas aquellas formadas por un conjunto de valores finitos o infinitos numerables.

3.1.1. Métodos para definir una variable aleatoria

Exhaustivo Es aquel método en el que se crea una lista con todos los posibles valores de $X(\omega)$; $\omega \in \Omega$. Es un método que pierde su efectividad cuantos más elementos pueda tener la variable aleatoria $X(\omega)$, pues es más difícil su representación.

Descriptivo El método descriptivo, consiste en formalizar o definir $X(\omega)$ de manera concreta, mediante fórmulas o palabras.

3.1.2. Racha

Se crea un conjunto con los resultados obtenidos en un experimento, de manera que estos resultados se ordenan cronológicamente. Se considera *racha* todo intervalo del conjunto en que los resultados son iguales.

3.2. Distribución de una variable aleatoria discreta

La probabilidad de que ocurra el suceso i de una variable aleatoria $P(X = i)$ se debe calcular para cada valor de i , siendo cada probabilidad igual a la suma de probabilidades de que ocurran los sucesos simples que cumplen la condición $P(X = i)$.

La probabilidad de una variable aleatoria es la suma de las probabilidades de sus miembros.

$$P(X \in B) = \sum_{i \in B} P(X = i)$$

3.2.1. Distribución de una función $Y = g(x)$

Teniendo una variable aleatoria X , cualquier función $Y = g(X)$ también es una variable aleatoria.

La función $g(x)$ es una función que se aplica sobre el conjunto de resultados de la variable aleatoria X . Por tanto su distribución de probabilidades dependerá de las probabilidades de los elementos de la variable aleatoria X .

3.2.2. Distribución de una variable aleatoria condicionada por un suceso

La probabilidad condicionada ya fue tratada con anterioridad; de manera que por definición ya se sabe que conocer, el hecho de que ha ocurrido un suceso A altera la probabilidad de que una variable tome un valor entre los posibles.

$$P(X = i | A) = \frac{P(\{X = i\} \cap A)}{P(A)}; i \in X(\Omega)$$

3.3. Esperanza matemática de una variable discreta

La *esperanza matemática* ($E\{X\}$), también es conocida como *valor esperado*, este es un promedio de posibles valores y probabilidad de estos.

$$E\{X\} = \sum_{i \in X(\Omega)} i \cdot P(X = i)$$

3.3.1. Esperanza de una función de una variable aleatoria

Siguiendo la definición de esperanza:

$$E\{Y\} = E\{g(x)\} = \sum_{i \in X(\Omega)} g(i) \cdot P(X = i)$$

Como se puede observar para conocer la esperanza de una función, no es necesario conocer su distribución.

3.3.2. Esperanza de una función lineal

En este caso se supone que la función Y es una función lineal de X , de manera que $y = ax + b$.
Para este caso particular la esperanza se calcula como

$$E\{Y\} = E\{ax + b\} = \sum_{i \in X(\Omega)} (ai + b)P(X = i)$$

si se simplifica esta expresión se obtiene

$$E\{ax + b\} = aE\{X\} + b$$

3.4. Momentos y varianza

Se conoce como *momentos de X* a todas las funciones de forma $g(x) = x^m$, siendo m un valor entero. De manera que $E\{X\}$ es la media de primer orden.

La *varianza* es la desviación de un resultado respecto al valor esperado (esperanza), se representa por

$$\sigma_x^2 = E\{(X - E\{X\})^2\}$$

3.5. Funciones generatrices

Suponiendo que se tiene una variable aleatoria X , tal que $X(\Omega) \in \mathbb{N}$

$$P(X = k) = p_k$$

para $k = 0, 1, 2, \dots + \infty$ y $\sum_k p_k = 1$

En este caso la función generatriz estaría definida como

$$G(s) = E\{S^x\} = \sum_{k=0}^{\infty} s^k p_k$$

En todos los caso siempre se cumple que $G'(0) = p_1$ y $G'(1) = E\{X\}$, de igual modo $G''(1) = E\{X^2\} - E\{X\} \Rightarrow E\{X^2\}G''(1) + G'(1)$

Esto es importante, pues la *varianza* se puede definir como

$$\sigma_x^2 = E\{X^2\} - (E\{X\})^2 = G''(1) + G'(1) - [G'(1)]^2$$

3.6. Modelos de distribuciones discretas

3.6.1. Distribución de Bermoulli

Su paradigma, o modelo que intenta representar, es lanzar una moneda y contar si ha salido cara o no.
En este modelo se suponen las siguientes probabilidades

$$P(x = 0) = 1 - p; \quad P(x = 1) = p$$

Gráficamente se puede representar de la siguiente forma:

Se pueden aplicar las siguientes fórmulas

$$\begin{aligned} E\{x\} &= 0 \cdot (1 - p) + 1p = p \\ E\{x^2\} &= 0^2 \cdot (1 - p) + 1^2p = p \\ \sigma_x^2 &= E\{x\} - E\{x^2\} = p - p^2 = p(1 - p) \end{aligned}$$

3.6.2. Distribución de Bermoulli

Se modela partiendo de la idea de que se lanza n veces una moneda, la cual tiene una probabilidad p de salir cara.
La probabilidad de este modelo se puede representar como

$$P(x = k) = \binom{n}{k} p^k \cdot (1 - p)^{n-k}; \text{ para } k = 0, 1, \dots, n$$

Gráficamente se puede representar de la siguiente forma:

Se pueden aplicar las siguientes fórmulas

$$\begin{aligned} E\{x\} &= n \cdot p \\ \sigma_x^2 &= n \cdot p \cdot (1 - p) \end{aligned}$$

3.6.3. Distribución geométrica

Se modela partiendo de la idea de que se lanza una hasta que se obtenga como resultado *cara*, y en ese momento contar el número de lanzamientos necesarios que han sido necesarios para obtener ese resultado cara. Se supone que la moneda tiene una probabilidad p de obtener cara.

La probabilidad de este modelo se puede representar como

$$P(x = k) = p \cdot (1 - p)^{k-1}; \text{ si } k = 0, 1, \dots, n$$

Gráficamente se puede representar de la siguiente forma:

Se pueden aplicar las siguientes fórmulas

$$E\{x\} = \frac{1}{p}$$
$$\sigma_x^2 = \frac{(1-p)}{p^2}$$

3.6.4. Distribución de Poisson

Esta distribución se usa para contar el número de sucesos poco probables que se han producido en un intervalo de tiempo dado.

La probabilidad de este modelo se puede representar como

$$P(x = k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}; \text{ para } k = 0, 1, 2, \dots$$

la letra $\lambda > 0$ designa el parámetro de distribución.

Se pueden aplicar las siguientes fórmulas

$$E\{x\} = \lambda$$
$$\sigma_x^2 = \lambda$$

4. Vectores aleatorios discretos

4.1. Introducción

Se llaman *vectores aleatorios discretos* a aquellos con la forma (X_1, X_2, \dots, X_n) , donde cada X_i es una variable aleatoria discreta, y todas ellas están definidas sobre el mismo espacio (Ω, P) . Estos vectores se usan para medir simultáneamente varios valores.

Un ejemplo de uso de un vector sería aquel definido como (X, Y) , donde:

- X = número de resultados pares
- Y = máximo de los resultados

En este ejemplo puede verse que para cada resultado del experimento se genera un par de resultados para este vector. Formalmente se puede definir como que el resultado escoge un elemento ω perteneciente al espacio muestral Ω ; como consecuencia de esto se obtienen resultados para el vector $(X(\omega), Y(\omega))$. En caso de que se deseara representar una sucesión de un experimento ilimitado esto se podría hacer como $\{X_i\}_{i=1}^{\infty}$.

4.2. Función de probabilidad conjunta

Se llama *función de probabilidad conjunta* de X e Y a la que describe los valores posibles de (X, Y) y la probabilidad con que ocurre cada valor posible.

Para este tipo de distribuciones se parte de la definición de

$$p_{i,j} = P(X = i, Y = j); i \in X(\Omega), j \in Y(\Omega)$$

donde $p_{i,j}$ es la probabilidad de que ocurra (X, Y) .

Un conjunto o matriz de números $(p_{i,j})$, $i \in X(\Omega)$, $j \in Y(\Omega)$, define una distribución conjunta si cumple las condiciones:

- $p_{i,j} \geq 0$
- $\sum_{i \in X(\Omega)} \sum_{j \in Y(\Omega)} p_{i,j} = 1$

4.2.1. Cálculo con la función de probabilidad conjunta

Cualquier suceso definido por medio de condiciones impuestas a las variables X e Y puede ponerse con la forma

$$\{\omega; (X(\omega), Y(\omega)) \in B\}$$

donde B es el para (x, y) que cumple la condición impuesta a las variables X, Y . Esta notación se puede simplificar con

$$\{(X, Y) \in B\}$$

En este tipo de distribuciones se cumple

$$P((X, Y) \in B) = \sum_{(i,j) \in B} P(X = i, Y = j)$$

4.3. Distribuciones marginales

Con este nombre se conoce al caso particular de desear calcular $P(X = i)$ donde la variable X debe cumplir la condición impuesta, mientras que la variable Y puede coger cualquier valor del espacio muestral (Ω) . En caso de cumplirse el mismo modelo para cualquier otra variable⁹ también sería válida esta definición.

Cuando se maneja la distribución conjunta de un vector, se acostumbra a denominar *marginales* a las distribuciones *unidimensionales* de sus componentes. Por tanto $P(X = i)$ sería la distribución unidimensional de la variable X .

	Y=0	Y=1	Y=2	
X=0	1/6	1/6	1/6	3/6
X=1	0	0	1/6	1/6
X=2	1/6	0	1/6	2/6
	2/6	1/6	3/6	

⁹Generalmente se trabaja con vectores bidimensionales, por tanto otra variable sería la Y

4.4. Valor esperado de una función $f(X, Y)$

Por extensión del caso unidimensional tratado en el tema anterior se puede afirmar que

$$E\{f(X, Y)\} = \sum_{i \in X(\Omega)} \sum_{j \in Y(\Omega)} f(i, j) P(X = i, Y = j)$$

Un caso concreto muy habitual es la *esperanza de la suma de dos variables*, en estos caso se parte de la suposición de que $f(X, Y) = X + Y$, de modo que

$$E\{X + Y\} = \sum_{i \in X(\Omega)} \sum_{j \in Y(\Omega)} (i + j) P(X = i, Y = j)$$

si se reordenan los términos se puede llegar a la siguiente conclusión

$$\begin{aligned} E\{X + Y\} &= \sum_{i \in X(\Omega)} \sum_{j \in Y(\Omega)} i \cdot P(X = i, Y = j) + \sum_{i \in X(\Omega)} \sum_{j \in Y(\Omega)} j \cdot P(X = i, Y = j) = \\ &= \sum_{i \in X(\Omega)} i \sum_{j \in Y(\Omega)} P(X = i, Y = j) + \sum_{j \in Y(\Omega)} j \sum_{i \in X(\Omega)} P(X = i, Y = j) = \\ &= \sum_{i \in X(\Omega)} i \cdot P(X = i) + \sum_{j \in Y(\Omega)} j \cdot P(Y = j) = E\{X\} + E\{Y\} \end{aligned}$$

4.4.1. Propiedades de la esperanza matemática

La esperanza matemática para funciones de dos variables tiene como característica más destacable que

$$E\{aX + bY\} = a \cdot E\{X\} + b \cdot E\{Y\}$$

Esta propiedad se llama *propiedad de linealidad*, gracias a esta se pueden realizar muchos cálculos de manera simbólica. Un ejemplo de esto puede verse si se desea calcular la covarianza semejante, para la cual se puede seguir el siguiente razonamiento

$$\begin{aligned} \sigma_{x,y} &= E\{(X - E\{X\}) \cdot (Y - E\{Y\})\} \\ &= E\{XY - X \cdot E\{Y\} + E\{X\} \cdot E\{Y\}\} \\ &= E\{XY\} - 2 \cdot E\{X\}E\{Y\} + E\{X\}E\{Y\} \\ &= E\{XY\} - E\{X\} \cdot E\{Y\} \end{aligned}$$

Gracias a este desarrollo se puede afirmar la expresión $\sigma_{x,y} = E\{XY\} - E\{X\} \cdot E\{Y\}$.

4.5. Distribuciones condicionadas

Considerando la probabilidad de suceso $X = x$ condicionada por $Y = y$

$$P(X = x | Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}$$

En caso de que $X(\Omega)$ varíen mientras el valor de Y esta fijo, se tiene una distribución denominada *X condicionada por Y = y*, esta se puede representar como $X | Y = y$.

Esta condicionalidad se puede representar dinámicamente por

$$P(X = x, Y = y) = P(Y = y) \cdot P(X = x | Y = y)$$

La esperanza matemática de una distribución $X | Y = j$ se simboliza por

$$E\{X | Y = j\} = \sum_{i \in X(\Omega)} i \cdot P(X = i | Y = j)$$

4.6. Variables independientes

En estas variables se cumple la condición

$$P(X = i, Y = j) = P(X = i)P(Y = j)$$

los $\{X = i\}$ y $\{Y = j\}$ son independientes, si además, la condición anterior se verifica para toda $i \in X(\Omega)$ y para toda $j \in Y(\Omega)$, se puede decir que X e Y son independientes. Esto significa que conocer el valor de una no es significativo para conocer el valor de la otra.

Si A y B son sucesos independientes, y X e Y son *independientes*, se cumple

$$P(X \in A, Y \in B) = P(X \in A) \cdot P(Y \in B)$$

4.6.1. Conjuntos y sucesiones de variables independientes

En caso de trabajar con un conjunto finito de variables X_1, X_2, \dots, X_n , serán independientes si los valores tomados por algunas no modifican a las restantes. Si las variables de un conjunto son independientes, esto implica que estas lo son dos a dos, sin embargo, *el hecho de que las variables de un conjunto sean independientes dos a dos no implica que el conjunto sea independiente.*

4.6.2. Esperanza del producto de variables independientes

Sin demasiadas complicaciones se puede entender que para dos variables independientes se cumple lo siguiente

$$E\{XY\} = \sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} x \cdot y \cdot P(X=x)P(Y=y) = E\{X\} \cdot E\{Y\}$$

Siguiendo esta propiedad se puede generalizar para el caso de una serie de n variables independientes, por tanto

$$E\{\prod_{i=1}^k f_i(X_i)\} = \prod_{i=1}^k E\{f_i(X_i)\}$$

4.6.3. Varianza de la suma de variables independientes

De forma general se puede afirmar que $\sigma_{x_1+x_2+x_3+\dots+x_n}^2 = \sigma_{x_1}^2 + \sigma_{x_2}^2 + \dots + \sigma_{x_n}^2$

4.7. Desigualdad de Chebishev

Si se pretende estimar la probabilidad de que X tome valores más alejados de la media (μ) que un número $a > 0$, es decir, se pretende estimar la probabilidad del suceso $|X - \mu| > a$. Según esta desigualdad

$$P(|X - \mu| > a) \cdot P(X=x) \leq \frac{\sigma_x^2}{a^2}$$

4.8. Ley de los grandes números

Con esta ley se pretende formalizar aquellos casos en que se realizan una serie de pruebas repetidas independientes, que son modeladas mediante la sucesión (X_1, X_2, \dots, X_n) , este modelo tiene como media (μ) y como varianza (σ^2).

El promedia de valores observados en las n primeras pruebas

$$\bar{X}_n = \frac{X_1 + X_2 + \dots + X_n}{n}$$

$$E\{\bar{X}_n\} = \frac{1}{n}(E\{X_1\} + E\{X_2\} + \dots + E\{X_n\}) = \frac{n \cdot \mu}{n} = \mu$$

Por otra parte basándose en $\sigma_{\bar{X}_n}^2 = \frac{\sigma^2}{n}$ y combinando esta expresión con la desigualdad de CHEBYSHEV se llega a la siguiente conclusión

$$P(|\bar{X}_n - \mu| > a) \leq \frac{\sigma^2}{a^2 n} \\ \lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| > a) = 0$$

4.9. Métodos de cálculo por condicionamientos

4.9.1. Método para calcular esperanzas por condicionamiento

Este es uno de los métodos más potentes que se pueden encontrar en la estadística, se basa en la fórmula

$$E\{Y\} = \sum_{x \in X(\Omega)} E\{Y | X=x\} \cdot P(X=x)$$

4.9.2. Método para calcular probabilidades por condicionamiento

Si se considera que la función indicadora I_A es igual a la probabilidad del suceso $P(A)$, en base al método anterior se llega a la siguiente conclusión

$$E\{I_A\} = \sum_{x \in X(\Omega)} E\{I_A | X=x\} \cdot P(X=x) = \sum_{x \in X(\Omega)} E\{A | X=x\} \cdot P(X=x) = P(A)$$

De este modo se obtiene otra versión de la fórmula de probabilidad total.

5. Variables aleatorias continuas

5.1. Introducción

Este tema en un principio puede parecer muy abstracto, aunque con un estudio detallado del mismo se verá que no lo es tanto.

Las *variables aleatorias continuas* estudian modelos con valores continuos, en un intervalo acotado (a, b) o no acotado $(0, \infty)$. Para poder estudiar estos modelos se hace, en muchos casos, necesario tratar como una serie de subconjuntos los valores a tratar.

Para una introducción simple a este tipo de modelos se tomará como patrón una ruleta ideal de longitud, que tiene como perímetro longitud uno. En esta ruleta ideal existe exactamente la misma probabilidad de obtener cualquier valor entre $(0, 1]$; para que en este modelo ideal exista exactamente la misma probabilidad de que aparezca cualquiera de los infinitos puntos del intervalo es necesario considerar que $p = 0$. Pues el número de puntos es infinito y si la probabilidad tuviese un valor positivo por muy pequeño que fuese $\sum_{i=0}^{\infty} i \cdot \mu = \infty$.

Como consecuencia de la anterior deducción se ve la necesidad de trabajar con subconjuntos de uno principal, de manera que cada uno de los valores que se desean estudiar deberá pertenecer a un subconjunto (I_1, I_2, \dots, I_n) . Cada uno de estos subconjuntos debe tener una longitud $\ell(I_i)$. Con estos conceptos se puede llegar a las conclusiones siguientes

$$P(X \in I_1) + P(X \in I_2) + \dots + P(X \in I_n) = n \cdot P(X \in I_1) = 1$$

$$P(X \in I_1) = \ell(I_1) = \frac{1}{n}$$

En caso de que se desee trabajar con subintervalos $(a, b] \subset (0, 1]$ se puede realizar la siguiente aproximación

$$P(X \in (a, b]) = \ell((a, b]) = b - a$$

Para aquellos casos en que se desea trabajar con un intervalo (J) que no tiene longitud uno, se aplicaría lo siguiente

$$P(X \in I_1) = \frac{1}{n} = \frac{\ell(I_1)}{\ell(J)}$$

En cierta manera este método tiene similitud con la “regla de LAPLACE” para la elección al azar de un elemento dentro de un conjunto finito $\frac{\text{casos favorables}}{\text{casos posibles}}$.

Ejemplo Se elige un número al azar en el intervalo $(0, 1]$, ¿cuál es la probabilidad de que su cuadrado sea mayor que $\frac{1}{3}$?

Para que $X^2 > \frac{1}{3}$, debe ocurrir que $X > \frac{1}{\sqrt{3}}$, por tanto la probabilidad pedida es

$$p = P(X \in (\frac{1}{\sqrt{3}}, 1))$$

por tanto $p = 1 - \frac{1}{\sqrt{3}}$.

5.2. Probabilidades geométricas

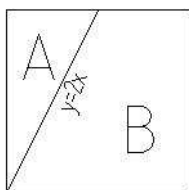
Son aquellas probabilidades descritas dentro de un modelo, generalmente bidimensional. Este tipo de problemas deben intentar modelarse dentro de un espacio geométrico $[0, 1] \times [0, 1]$; con frecuencia puede ayudar a la resolución una representación gráfica.

En general la propiedad más importante, de este tipo de problemas es aquella en que si se elige un punto al azar del conjunto A , donde A es subconjunto de R, R^2, \dots , la probabilidad de que el punto elegido pertenezca a $B \subset A$ es

$$P(X \in B) = \frac{\text{medida}(B)}{\text{medida}(A)}$$

Ejemplo Se eligen dos puntos X, Y al azar en el intervalo $(0, 1]$, ¿cuál es la probabilidad de que $Y > 2X$?

Este problema puede representarse con el siguiente gráfico:



Se ve como región inferior, B , aquella que verifica $Y < 2X$, mientras que la región superior, A , verifica $Y > 2X$. Por tanto se tiene que

$$P(Y > 2X) = \frac{\text{rea}(A)}{\text{rea total}}$$

dado que $\text{rea}(A) = \frac{1}{4}$ y el $\text{rea total} = 1$, resulta que $P(Y > 4) = \frac{1}{4}$.

5.3. Variables aleatorias continuas

Hasta este punto siempre se ha supuesto que las variables aleatorias continuas son siempre uniformes, es decir, que en cualquiera de sus infinitos puntos siempre existen los mismos valores. En algunas circunstancias se verá que esto no es correcto.

En un modelo *uniformes* se puede afirmar que $P(X \in (a, b]) = b - a$, esto no es cierto en modelos *no uniformes* y para el estudio de estos se debe recurrir a una analogía con la física.

Un modelo *no uniforme* se puede tratar como un hilo construido con un hilo con una aleación de plomo y cobre, de manera que la aleación no se distribuye por igual, sino que en un extremo del hilo hay plomo puro y en el otro cobre puro. En este caso existen infinitos valores en el gradiente intermedio del hilo, de manera que el peso del mismo varía desde un punto a otro. Para tratar este sistema con diversas masas en física se ha inventado la noción de *densidad*, la cual se puede emplear en dos contextos

- Macroscópico $\text{densidad} = \frac{\text{masa}}{\text{longitud}}$
- Microscópico, cuando se estudia en punto x $f(x) = \lim_{h \rightarrow 0} \frac{\text{masa intervalo}(x-h, x+h)}{2h}$ o bien $f(x) = \int_a^b f(t) \cdot dt$

La estadística ha tomado este modelo de cálculo y ha creado una analogía entre el concepto de probabilidad y de masa. Con la diferencia de que la probabilidad debe estar limitada al valor máximo 1, mientras que la masa no tienen dicha limitación.

Cualquier *función de densidad de probabilidad* debe cumplir:

1. $f(x) \geq 0$
2. $\int_{-\infty}^{\infty} f(t) \cdot dt = 1$

Se considera que una función es de *densidad uniforme*, en el intervalo J cuando esta es constante sobre los puntos de J y cero sobre los restantes, tal que

$$f(x) = \begin{cases} c, & \text{si } x \in J \\ 0, & \text{si } x \notin J \end{cases}$$

donde la constante c viene determinada por la relación

$$\int_J f(t) \cdot dt = \int_J c \cdot dt = c \cdot \int_J dt = 1$$

5.4. Esperanza matemática de una variable continua

Si se considera la variable continua X , que toma valores en el intervalo $I = (0, 1]$. La esperanza matemática de X , intuitivamente es el promedio de los valores que la variable toma y las probabilidades que alcanza esos valores.

Puesto que en variables continuas $P(X = x) = 0$ para cada valor posible x . Por tanto es necesario dividir en intervalo I en subintervalos iguales

$$0 = \frac{0}{n} < \frac{1}{n} < \frac{2}{n} < \dots < \frac{n-1}{n} < \frac{n}{n} = 1$$

Tomando la probabilidad

$$P\left(\frac{k}{n} < X \leq \frac{k+1}{n}\right)$$

Partiendo de todo esto se puede llegar a la conclusión de que

$$\sum_{k=0}^n \frac{k}{n} P\left(\frac{k}{n} < X \leq \frac{k+1}{n}\right) \approx \int_0^1 t \cdot f(t) \cdot dt$$

Como conclusión a todo esto se puede afirmar que la esperanza matemática de la variable X con función de densidad f es

$$E\{X\} = \int_{-\infty}^{\infty} t \cdot f(t) \cdot dt$$

siempre que sea finita.

5.5. Esperanza matemática de una función

En concepto se debe seguir el mismo procedimiento que cuando se desea calcular sobre una variable X , de manera que se tomará como intervalo $X \in (x, x + dx]$. Por este razonamiento se puede llegar a

$$\begin{aligned} E\{g(X)\} &\approx \sum g \cdot \left(\frac{k}{n}\right) \cdot \frac{1}{n} \\ E\{g(X)\} &= \int_{-\infty}^{+\infty} g(t) \cdot f(t) \cdot dt \\ \sigma_X^2 &= \int_{-\infty}^{\infty} (t - E\{X\})^2 \cdot f(t) \cdot dt \\ \sigma_X^2 &= \int_{-\infty}^{\infty} t^2 \cdot f(t) \cdot dt - (E\{X\})^2 = E\{X^2\} - (E\{X\})^2 \end{aligned}$$

5.6. Distribuciones con variables continuas

5.6.1. Distribución uniforme en un intervalo

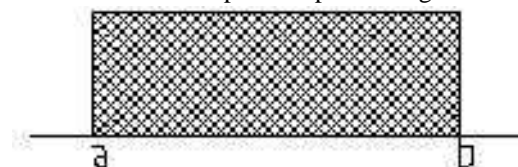
Esta distribución es extremadamente importante, pues a partir de ella se puede calcular cualquier otra distribución. Su paradigma es que en el intervalo $[a, b]$, con $a < b$. Su función de densidad es

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{si } x \in [a, b] \\ 0, & \text{si } x \notin [a, b] \end{cases}$$

La esperanza de X es de $E\{X\} = \int_a^b t \cdot \frac{dt}{b-a} = \frac{a+b}{2}$.

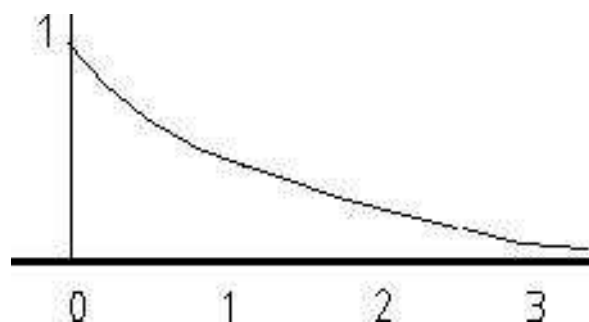
Así mismo, $E\{X^2\} = \int_a^b t^2 \cdot \frac{dt}{b-a} = \frac{b^3-a^3}{3 \cdot (b-a)} = \frac{1}{3}(b^2+ab+a^2)$ de aquí se puede calcular $\sigma_X^2 = E\{X^2\} - (E\{X\})^2 = \frac{1}{12}(b-a)^2$.

Esta distribución se puede representar gráficamente como



5.6.2. Distribución exponencial

Su paradigma es el tiempo que tarda en desintegrarse un átomo de una masa de material radiactivo. Esto dicho de otro modo implica que es una distribución descendente, de modo que cuando mayor es el valor menor es el índice de descenso que sufre.



La función de densidad exponencial es

$$f(x) = \begin{cases} \lambda \cdot e^{-\lambda \cdot x}, & \text{si } x > 0 \\ 0, & \text{si } x \leq 0 \end{cases}$$

Como se puede ver cuanto mayor es el valor $\lambda > 0$, más rápidamente decae la función y más probable es que la variable tome un valor próximo al origen.

$$P(X > x) = \int_x^{\infty} \lambda \cdot e^{-\lambda t} dt = (e^{-\lambda t})|_x^{\infty} = e^{-\lambda x}, \text{ para } x > 0$$

Por tanto se puede interpretar la igualdad

$$P(X > x + h | X > x) = P(X > h)$$

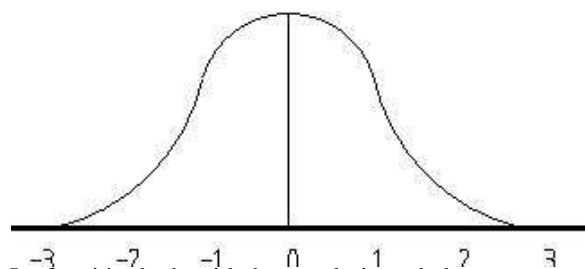
de manera que si transcurren x unidades de tiempo sin que el suceso que esperamos haya ocurrido, la probabilidad de que se retrase al menos otras h unidades de tiempo es la misma que al comienzo. Por tanto el proceso no tiene envejecimiento, y la esperanza matemática puede darse por

$$E\{X\} = \int_0^{\infty} t \cdot \lambda \cdot e^{-\lambda t} dt = \frac{1}{\lambda}$$

En consecuencia de la esperanza

$$\sigma_X^2 = E\{X^2\} - (E\{X\})^2 = \frac{1}{\lambda^2}$$

5.6.3. Distribución normal



La función de densidad normal viene dada por

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-(x-\mu)^2/2\sigma^2}, -\infty < x < \infty$$

Esta distribución depende de dos parámetros la *media de su distribución* (μ) y su *varianza* (σ). Para hacer referencia a esta distribución con sus correspondientes parámetros se emplea la expresión $N(\mu, \sigma)$.

Cálculo de distribuciones normales Si a una variable normal se le aplica una transformación lineal, la variables que resulta también es normal. De modo que si X es una variable con distribución $N(\mu, \sigma)$ y $a, b \in \mathbb{R}$, entonces $aX + b$ tiene una distribución normal de media

$$E\{aX + b\} = aE\{X\} + b = a\mu + b$$

y varianza

$$\sigma_{aX+b}^2 = \sigma_{aX}^2 = a^2 \sigma_X^2 = a^2 \cdot \sigma^2$$

Puesto que

$$\sqrt{\sigma_{aX+b}^2} = \sqrt{a^2 \sigma^2} = |a| \sigma$$

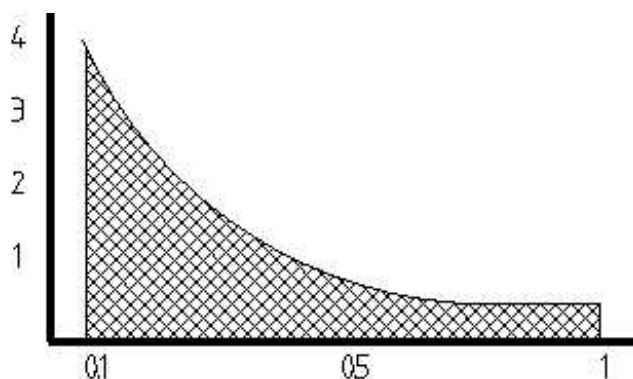
la variable $aX + b$ tiene una distribución $N(a\mu + b, |a|\sigma)$.

5.6.4. La distribución “natural” de los números

Para poder explicar esta distribución es importante recordar que de forma normalizada cualquier número se puede expresar con la forma $m(x) \cdot 10^n$, donde $m(x) \in (1/10, 1)$ es la *mantisa* y n la *característica*.

Por evidencia empírica se ha podido demostrar que los números se distribuyen de modo que el logaritmo de su mantisa es un número elegido al azar en un rango de variación $[1/10, 1]$. Esta hipótesis equivale a suponer que la función de densidad de la mantisa es

$$f(m) \begin{cases} \frac{1}{m \cdot \ln 10} & \text{si } m \in [1/10, 1] \\ 0 & \text{si } m \notin [1/10, 1] \end{cases}$$



5.7. Variables mixtas

Este tipo de variables reciben este nombre pues tienen una componente continua y una discreta. Un ejemplo *tipo* de este tipo de variables es una ruleta de longitud uno, como ya se vio al principio del tema este tipo de aparatos son claramente continuos, pues existen infinitos puntos entre $(0, 1]$, por tanto X es una variable continua. Ahora bien si se aplica la siguiente función de X

$$G = \begin{cases} X & \text{si } 0 \leq X \leq \frac{1}{2} \\ 0 & \text{si } \frac{1}{2} \leq X < 1 \end{cases}$$

por tanto en este caso $P(G = 0) = \frac{1}{2}$, por tanto en este intervalo no se comporta como una variable continua propiamente, así como tampoco se comporta como una discreta como una discreta en el intervalo $(\frac{1}{2}, 1)$.

5.8. Funciones de distribución

Son una herramienta que permite hacer una teoría general de la probabilidad, con afirmaciones válidas para todas las distribuciones posibles, sin necesidad de distinguir entre casos.

Dada una variable X , la función de distribución describe cómo está repartida la probabilidad entre los valores de esta. Toda función de distribución verifica las siguientes propiedades:

- F es no decreciente: si $x < x'$, se tiene $F(x) \leq F(x')$
- F es continua por la derecha: $\lim_{h \downarrow 0} F(x+h) = F(x)$
- Se cumple $\lim_{x \rightarrow \infty} F(x) = 1$ y $\lim_{x \rightarrow -\infty} F(x) = 0$

Si se fija x y $h \downarrow 0$, se tiene

$$\lim_{h \downarrow 0} F(x+h) = \lim_{h \downarrow 0} P(X \leq x+h) = P(X \leq x) = F(x)$$

luego F es continuo por la derecha.

Se tiene

$$\begin{aligned} \lim_{x \rightarrow \infty} F(x) &= \lim P(X \leq x) = P(X \in \mathbb{R}) = 1 \\ \lim_{x \rightarrow -\infty} F(x) &= \lim P(X \leq x) = P(X \in \emptyset) = 0 \end{aligned}$$

5.9. Tres modelos de funciones de distribución

5.9.1. Variables discretas

Suponiendo una variable discreta X , con una distribución binomial de parámetros $n = 3$, $p = \frac{1}{2}$. La función de probabilidad de X es igual a

$$P(X = 0) = \frac{1}{8}; P(X = 1) = P(X = 2) = \frac{3}{8}; P(X = 3) = \frac{1}{8}$$

Desglosando por partes se puede llegar a las siguientes conclusiones:

- Si $x < 0$, entonces $F(X) = 0$, ya que X no puede tomar valores menores o iguales que $x < 0$
- Si $0 \leq x < 1$, entonces $\{X \leq x\} = \{X = 0\}$ y $F_X(x) = P(X = 0) = 0,125$
- Si $1 \leq x < 2$, entonces $\{X \leq x\} = \{X = 0\} \cup \{X = 1\}$ y $F_X(x) = P(X = 0) + P(X = 1) = 0,5$
- Si $2 \leq x < 3$, entonces $\{X \leq x\} = \{X = 0\} \cup \{X = 1\} \cup \{X = 2\}$ y $F_X(x) = P(X = 0) + P(X = 1) + P(X = 2) = 0,875$
- Si $x \geq 3$, entonces $\{X \leq x\} = \{X = 0\} \cup \{X = 1\} \cup \{X = 2\} \cup \{X = 3\}$ y $F_X(x) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) = 1$

En resumen la función de distribución de X quedaría definida como

$$F_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ 0,125 & \text{si } 0 \leq x < 1 \\ 0,5 & \text{si } 1 \leq x < 2 \\ 0,875 & \text{si } 2 \leq x < 3 \\ 1 & \text{si } x \geq 3 \end{cases}$$

5.9.2. Variables continuas

Dada la variable aleatoria exponencial de parámetro $\lambda = 1$, para hallar su función de distribución se distingue entre:

- si $x < 0$, entonces $F_Y(x) = 0$, ya que se tiene

$$F_Y(x) = \int_{-\infty}^x 0 \cdot dt = 0$$

- si $x \geq 0$, entonces $F_Y(x) = \int_{-\infty}^x f(t) \cdot dt = \int_{-\infty}^0 0 \cdot dt + \int_0^x e^{-t} \cdot dt = (-e^{-t})|_0^x = 1 - e^{-x}$

Por tanto

$$F_Y(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ e^{-x} & \text{si } x > 0 \end{cases}$$

Dada una variable aleatoria X y una función de distribución F , en caso de que $F'(x)$ exista salvo p ara un número numerable de puntos se cumple que $\int_{-\infty}^{\infty} F'(t) \cdot dt = 1$ por lo que X es continua, y tienen una función de densidad y es igual a F' .

$$F'_Y(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ e^{-x} & \text{si } x > 0 \end{cases}$$

5.9.3. Variables mixtas

Considerando como variable mixta G , que se define como

$$G = \begin{cases} X & \text{si } 0 \leq X \leq \frac{1}{2} \\ 0 & \text{si } \frac{1}{2} < X < 1 \end{cases}$$

Para este caso particular se distingue entre:

- Si $x < 0$, entonces $P(G \leq x < 0) = 0$ y $F_G(x) = 0$
- Si $0 \leq x \leq \frac{1}{2}$, entonces se tiene

$$P(G \leq x) = P(G = 0) + P(0 < G \leq x) = \frac{1}{2} + \int_0^x dt = \frac{1}{2} + x$$

$$\text{luego } F_G(x) = x + \frac{1}{2}$$

- Si $x > \frac{1}{2}$, entonces $P(G \leq x) = 1$ y $F_G(x) = 1$

La función de distribución viene definida por

$$F_G(x) = \begin{cases} 0 & \text{si } x < 0 \\ x + \frac{1}{2} & \text{si } 0 \leq x \leq \frac{1}{2} \\ 1 & \text{si } x > \frac{1}{2} \end{cases}$$

Con lo visto en el punto anterior, se puede $F'_G(x)$ existe en todos los puntos excepto en $x = 0$ y $x = \frac{1}{2}$. Se cumple:

$$F'_G(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 & \text{si } 0 \leq x \leq \frac{1}{2} \\ 0 & \text{si } x > \frac{1}{2} \end{cases}$$

5.10. Función de distribución de $Y = g(X)$

Este método consiste en que partiendo de una variable aleatoria X , se calcula la función de distribución de otra $Y = f(X)$ que es función de X . Con este método se puede evaluar si la variable es discreta, continua o mixta. Para cada intervalo I , se cumple que

$$P(Y \in I) = P(X \in g^{-1}(I))$$

donde $g^{-1}(I) = \{x; g(x) \in I\}$.

5.11. Distribución de una variable condicionada por un suceso

En este apartado se tratará una técnica para que dada una variable aleatoria X continua, con función de densidad f , se pueda hallar la distribución que tiene dicha variable X .

Si I es un intervalo, la probabilidad del suceso $\{X \in I\}$ es igual a

$$P(X \in I | X \in A) = \frac{P(\{X \in I\} \cap \{X \in A\})}{P(X \in A)} = \int_{I \cap A} \frac{f(x)}{P(X \in A)} dx$$

Dado el cálculo anterior se puede afirmar que la *función de densidad condicionada* de X por $X \in A$ se representa por $f_{X|A}$ es igual a

$$f_{X|A}(x) = \begin{cases} \frac{f(x)}{P(X \in A)} & \text{si } x \in A \\ 0 & \text{si } x \notin A \end{cases}$$

Por tanto el *factor de proporcionalidad* viene dado por $\frac{1}{P(X \in A)}$.

6. Vectores aleatorios continuos

6.1. Introducción

El punto principal de este tema, es el estudio de la probabilidad de que el vector aleatorio (X, Y) pertenezca a cualquier rectángulo, esto se puede obtener como la integral de $f(x, y)$ sobre el rectángulo. De manera formal esta idea se puede expresar como

$$P(X \in I_1, Y \in I_2) = \int \int_{I_1 \times I_2} f(x, y) \cdot dy \cdot dx$$

por tanto también se cumple que

$$P((X, Y) \in B) = \int \int_{(x, y) \in B} f(x, y) \cdot dy \cdot dx$$

Para abordar estas cuestiones es preciso poder evaluar dos puntos básicos

1. Condiciones que debe cumplir $f(x, y)$ para ser función de densidad de algún vector unitario. Para ello deben cumplirse las siguientes condiciones:
 - $f(x, y) \geq 0$, para $-\infty < x < +\infty, -\infty < y < +\infty$
 - $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) \cdot dy \cdot dx = 1$
2. Formas para calcular la función de densidad dada una distribución. Este es el punto que más ampliamente se irá tratando en los siguientes apartados.

6.2. Función de distribución conjunta

La función de distribución conjunta del vector (X, Y) es una función de dos variables tal que

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y)$$

Esta función es igual a la probabilidad de que el vector (X, Y) tome un valor del cuadrante $(-\infty, x] \times (-\infty, y]$, esto se puede expresar mediante

$$F_{X,Y}(x, y) = P((X, Y) \in (-\infty, x] \times (-\infty, y])$$

Si es conocida la función de densidad del vector calcular la distribución se reduce a calcular la integral

$$F_{X,Y}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(t, s) \cdot ds \cdot dt$$

6.3. Distribuciones marginales

Si X e Y tienen distribución conjunta continua, cada una de estas componentes tienen una distribución unidimensional continua.

La función de densidad de X es igual a

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) \cdot dy$$

existe una función análoga para la densidad Y .

Las funciones de densidad $f_X(x)$ y $f_Y(y)$ se denominan *densidades marginales*.

6.4. Variables independientes

Reciben este nombre aquellas variables en que conocer el valor que ha tomado una de ellas no modifica la probabilidad de que la otra tome un valor prefijado.

Dada dos variables independientes X, Y que tienen una función de densidad conjunta $f(x, y)$, una condición necesaria y suficiente para que sean independientes es que se verifique

$$f(x, y) = f_X(x)f_Y(y)$$

6.5. Funciones con densidad condicionada

Dadas dos variables X, Y se denomina *distribución de Y condicionada por $X=x$* a la expresión $Y|X = x$, esto puede aplicarse sobre cualquiera de las variables, por tanto puede existir también $X|Y = y$.

Sobre las variables condicionadas puede calcularse *funciones de densidad*. Esto se haría de la siguiente manera

$$f_{Y|X=x}(y) = \begin{cases} \frac{f(x,y)}{f_X(x)} & \text{si } f_X(x) > 0 \\ 0 & \text{si } f_X(x) = 0 \end{cases}$$

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

6.6. Modelos dinámicos

En caso de tener una variable Y condicionada por $X = x$, su densidad se puede expresar como

$$f(x, y) = f_X(x)f(y|x)$$

6.7. Densidad de la suma de variables independientes

Dadas dos variables independientes X, Y con sus correspondientes funciones de densidad marginal $f_X(x)$ y $f_Y(y)$. Se puede determinar la distribución de la variable suma $S = X + Y$.

Se cumple que

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

$$F_Y(y) = \int_{-\infty}^y f_Y(t)dt$$

La función de densidad de la suma de las variables X, Y es igual a

$$f_S(s) = \int_{-\infty}^{\infty} f_Y(s-x)f_X(x)dx; -\infty < s < \infty$$

6.8. Densidad de una función $g(X, Y)$

En este apartado se va a estudiar la técnica para dadas dos variables independientes X, Y , se calcule una función de la densidad conjunta de ellas. De manera que cuando $g(X, Y) = X + Y$ y $f(x, y) = f_X(x)f_Y(y)$.

El método seguido consistirá en calcular la distribución de Z y después analizando su derivada determinar si tiene densidad. Los pasos serán los siguientes:

1. $F_Z(z) = P(g(X, Y) \leq z) = \int \int_{\{g(x,y) \leq z\}} f(x, y) \cdot dx \cdot dy$ para $-\infty < z < \infty$
2. Calcular $dF_Z(z)/dz$ y estudiar si se cumple

$$\int_{-\infty}^{\infty} \frac{dF_Z(z)}{dz} dz = 1$$

en caso afirmativo $dF_Z(z)/dz$ es la función de densidad de F_Z .

6.9. Esperanza matemática de una función $g(X, Y)$

De manera informal, se puede definir que la esperanza de una función $g(X, Y)$ es el promedio de valores de g respecto a la distribución de probabilidad (X, Y) . Por tanto se puede decir que para dos variables aleatorias X, Y con función de densidad conjunta $f(x, y)$

$$E\{g(X, Y)\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) \cdot f(x, y) \cdot dy \cdot dx$$

De manera unidimensional y conociendo la densidad marginal de $Z = g(X, Y)$

$$E\{Z\} = \int z \cdot f_Z(z) \cdot dz$$

6.10. Covarianza entre X e Y

$$g(X, Y) = (X - E\{X\})(Y - E\{Y\})$$

Se denomina *covarianza* al valor esperado de $g(X, Y)$ y se representa por $Cov(X, Y)$, de modo que

$$Cov(X, Y) = (X - E\{X\})(Y - E\{Y\})$$

Las propiedades de la covarianza se pueden reflejar con

$$\begin{aligned} Cov(X, Y) &= E\{XY\} - E\{XE\{Y\}\} - E\{YE\{X\}\} + E\{E\{X\}E\{Y\}\} = \\ &= E\{XY\} - E\{Y\}E\{X\} - E\{X\}E\{Y\} + E\{X\}E\{Y\} = E\{XY\} - E\{X\}E\{Y\} \end{aligned}$$

Se conoce como *coeficiente de correlación*

$$\rho_{X,Y} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

6.11. Cálculo de la densidad del vector (U, V) función de (X, Y)

Dado un vector (X, Y) con función de densidad conjunta $f(x, y)$. Supóngase

$$\begin{aligned} U &= g_1(X, Y) \\ V &= g_2(X, Y) \end{aligned}$$

En este apartado se verá como se puede calcular la ley de probabilidad de (U, V) a partir de (X, Y) . Existen dos métodos para ello:

Para el *primer método* no se impone ninguna condición a g_1 y g_2 . de manera que la función de distribución bidimensional viene dada por los valores (u, v)

$$F_{U,V}(u, v) = P(U \leq u, V \leq v) = \int \int_{g_1(x, y) \leq u, g_2(x, y) \leq v} f(x, y) \cdot dx \cdot dy$$

Para analizar si $F_{U,V}(u, v) = \frac{\partial^2 F_{U,V}(u, v)}{\partial u \partial v}$.

Finalmente es necesario realizar la comprobación de la condición

$$F_{U,V}(u, v) = \int_{-\infty}^u \int_{-\infty}^v f_{U,V}(s, t) \cdot dt \cdot ds$$

El *segundo método* es más automático, aunque exige que las funciones g_1 y g_2 cumplan ciertas condiciones. Para ello se supone la aplicación biyectiva $(x, y) \rightarrow (u, v)$, y $u = g_1(x, y)$ y $v = g_2(x, y)$ que tienen derivadas parciales continuas, de modo que se anulan de la siguiente manera

$$J(x, y) = \begin{vmatrix} \frac{\partial g_1}{\partial x} & \frac{\partial g_1}{\partial y} \\ \frac{\partial g_2}{\partial x} & \frac{\partial g_2}{\partial y} \end{vmatrix} \neq 0$$

entonces la función de densidad (U, V) es igual a

$$f_{U,V}(u, v) = \frac{1}{|J(x(u, v), y(u, v))|} \cdot f_{X,Y}(x(u, v), y(u, v))$$

donde $(x(u, v), y(u, v))$, es la solución del sistema de ecuaciones

$$\begin{cases} g_1(x, y) = u \\ g_2(x, y) = v \end{cases}$$

7. Simulación estocástica

7.1. Introducción

Este tipo de simulación se realiza por computador, se debe emplear en aquellos casos en que no se puede resolver un determinado problema de forma analítica¹⁰. El principal problema para poder realizar este tipo de simulación consiste en programar el computador de manera que sea capaz de sortear números según la ley de probabilidad definida por *cualquier* función de distribución.

Dada una variable uniforme X en el intervalo $[0, 1]$ y F es la función de distribución, denominada función de cuantías de F a la función Q , definida por:

$$Q(u) = \min\{x; F(x) \geq u\}$$

Para cualquier función de distribución F , debe existir una función Q para todo $u \in (0, 1)$ y, si U es una variable uniforme en $(0, 1)$, la variable $X = Q(U)$ tienen función de distribución F .

7.2. Generadores de números aleatorios y pseudoaleatorios

Existen diferentes generadores de números aleatorios, estos se pueden clasificar como:

- Mecánicos. Son los más rudimentarios, resultan lentos si se debe obtener una gran cantidad de números, pero su principal ventaja es lo impredecibles que son. El ejemplo más típico es el bombo con números.
- Electrónicos. Principalmente se basan en el ruido imperceptible de algunos elementos¹¹, como por ejemplo las resistencias. Estos métodos también son lentos para obtener grandes cantidades de números.
- Pseudoaleatorios. Son métodos basados en algoritmos de software. Uno de los ejemplos más destacados es el de VON NEUMANN que propuso formar una sucesión a partir de un número, elevando al cuadrado el anterior y extrayendo los dígitos centrales.

$$X_n = a \cdot X_{n-1} \pmod{m}$$

Este algoritmo no es aleatorio, sino determinista. Si se conoce a , m y X_0 . Por norma general, este tipo de generadores suelen ser suficientes para los programas de simulación, pues generan muchos números en poco tiempo, además de permitir sucesivas refinaciones, pues siempre se obtienen los mismos valores.

- Generados imprevisibles. Son los más difíciles de conseguir, para poder generarlos se han diseñado los *generadores no lineales* y otros dispositivos¹².

7.3. Generadores de congruencias lineales (GCL)

Este sistema produce una secuencia de números entre 0 y 1. Para implementarlo se necesita escoger los siguientes valores:

m el módulo de la congruencia

a el multiplicador

c la translación

X_0 la semilla

Elegidos estos valores se debe calcular mediante recurrencia los siguientes

$$X_n = (a \cdot X_{n-1} + c) \pmod{m}$$

El resultado devuelto X_n , se le debe aplicar una última operación para obtener U_n

$$U_n = \frac{X_n}{m}$$

Puesto que $X_n \in \{0, 1, 2, \dots, m-1\}$, los valores de $U_n \in [0, 1]$.

El principal problema de este método es que conocido X_i se pueden proveer los valores futuros, además estos son cíclicos. Para generar el ciclo lo más grande posible es conveniente elegir el valor de m lo mayor posible, y es conveniente que este valor sea una potencia de la base del sistema de numeración con que se está trabajando¹³.

La condición necesaria y suficiente para que un generador GLC tenga un periodo de longitud máxima m es que se cumpla:

¹⁰Este es el método empleado hasta ahora, en temas anteriores

¹¹Este se llama *ruido blanco*

¹²Estos suelen conectarse al puerto serie de la máquina

¹³Esta recomendación afecta a la velocidad de cálculo de nuevos números

- m y c son primos entre si
- Se cumpla que $a \equiv 1 \pmod{p}$, para cada p primo que divide a m
- Si 4 divide a m , entonces $a \equiv 1 \pmod{4}$

7.4. Transformaciones generales

Para poder obtener los valores de una variable aleatoria con una función de distribución no uniforme, es necesario aplicar una transformación al resultado presentado por el generador de números aleatorios. Para esto se emplean los conceptos vistos en el tema anterior.

7.4.1. Algoritmo de Von Neumann

Si se desea obtener la variable aleatoria X que tiene función de densidad $f(x)$. Supongamos que sabemos generar valores de otra variable Y con función de densidad $g(y)$ y se cumple que

$$\frac{f(x)}{g(x)} \leq C < \infty$$

según Von Neumann se debe aplicar el siguiente algoritmo:

1. Generar Y con densidad $g(y)$
2. Generar U con función de densidad uniforme en $(0, 1)$
3. si se cumple $CU \leq \frac{f(Y)}{g(Y)}$ devolver $X = Y$; en otro caso, volver a 2

Este algoritmo produce valores distribuidos con función de densidad $f(x)$.

Este método también se denomina método de *aceptar-rechazar*.

7.5. Simulación de variables normales

7.5.1. Por método *aceptar-rechazar*

Dada una variable normal (Z) de media 0 y varianza 1, primero se simula la variable $X = |Z|$ que tiene una función de densidad igual a:

$$f(x) = \frac{2}{\sqrt{2\pi}} \cdot e^{-x^2/2}, \text{ si } 0 < x < \infty$$

Por el método general se sortea el signo de X y se devuelve Z .

Partiendo de la variable Y con distribución exponencial

$$g(y) = e^{-x}, \text{ si } 0 < x < \infty$$

se obtiene

$$\frac{f(x)}{g(x)} = \frac{2}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}+x} = \sqrt{\frac{2e}{\pi}} \cdot e^{-(x-1)^2/2} \leq \sqrt{\frac{2e}{\pi}}$$

luego $C = \sqrt{\frac{2e}{\pi}}$.

Ahora se debe aplicar el siguiente algoritmo:

1. Generar U_1 con densidad uniforme. Hacer $Y = -\ln U_1^2$, que tienen distribución exponencial de parámetro $\lambda = 1$
2. Generar U_2 con función de densidad uniforme en $(0, 1)$
3. Si se cumple

$$U_2 \leq e^{-\frac{1}{2}(Y-1)^2}$$

hacer $X = Y$; en caso contrario volver al paso 2

4. Generar U_3 con densidad uniforme. Si $U_3 < 0,5$, devolver $Z = -X$; en otro caso, devolver $Z = X$

7.5.2. Por método *BOX-MULLER*

Partiendo del vector (X, Y) , cuyas variables son independientes y con una distribución normal de media 0 y varianza 1. La función de densidad conjunta de (X, Y) es

$$f(x, y) = \frac{1}{2\pi} \cdot e^{-\frac{1}{2}(x^2+y^2)}, \text{ si } -\infty < x < \infty, -\infty < y < \infty$$

Se calcula la función de densidad conjunta en coordenadas polares, obteniendo

$$\begin{aligned} x &= r \cdot \cos \theta \\ y &= r \cdot \sin \theta \end{aligned}$$

o bien $r^2 = x^2 + y^2$ y $\theta = \arctan(y/x)$, donde $-\infty < r < \infty$ y $0 \leq \theta < 2\pi$.

El jacobiano es

$$\frac{\partial(x, y)}{\partial(r, \theta)} = \begin{vmatrix} \cos \theta & -r \cdot \sin \theta \\ \sin \theta & r \cdot \cos \theta \end{vmatrix} = r \cdot (\cos^2 \theta + \sin^2 \theta) = r$$

y la función de densidad conjunta (R, Θ)

$$f(r, \theta) = f(x(r, \theta), y(r, \theta)) \cdot \frac{\partial(x, y)}{\partial(r, \theta)} = \frac{1}{2\pi} \cdot r \cdot e^{-\frac{1}{2}r^2}$$

Para este algoritmo primero es necesario generar U_1 con densidad uniforme, una vez obtenida esta $\Theta = 2\pi U_1$; la variable Θ tiene una distribución uniforme entre 0 y 2π .

Luego será necesario hallar U_2 con densidad uniforme y $W = -2 \ln U_2$. Para este caso W tiene una distribución exponencial de parámetro $\frac{1}{2}$. Posteriormente se aplica $R = \sqrt{W}$.

El algoritmo BOX-MULLER, que sirve para generar dos variables normales independientes tiene los siguientes pasos:

1. Generar U_1 y U_2 con densidad uniforme
2. Devolver

$$X = \sqrt{-2 \ln U_2} \cos 2\pi U_1, Y = \sqrt{-2 \ln U_2} \sin 2\pi U_1$$

El principal problema de este algoritmo es que al emplear \cos y \sin se puede incrementar el tiempo de ejecución.

7.5.3. Algoritmo de MARSAGLIA

1. Generar U_1 y U_2 con densidad uniforme entre 0 y 1
2. Hacer $V_1 = 2U_1 - 1$ y $V_2 = 2U_2 - 1$ (tanto V_1 como V_2 tienen densidad uniforme entre -1 y 1)
3. Calcular $W = V_1^2 + V_2^2$. Si $W < 1$ pasa a paso 4, si $W \geq 1$, volver a paso 1
4. Devolver

$$X = V_1 \sqrt{-\frac{2}{W} \ln W}, Y = V_2 \sqrt{-\frac{2}{W} \ln W}$$

7.6. Simulación de variables discretas

Para la simulación de variables discretas, aunque en teoría esta no puede tomar un número numerable de valores, se parte de la premisa de que las precisiones inferiores a las soportadas por un ordenador se consideran cero.

7.6.1. Algoritmo por cuantiles

1. Generar U con densidad uniforme entre 0 y 1
2. Hacer $i = 1$
3. Mientras se cumpla $U \geq \sum_{i=1}^j p_i$, hacer $i = i + 1$
4. Devolver $X = x_i$

Para que este algoritmo funcione con más eficiencia es recomendable ordenar los valores de X en orden decreciente.

7.6.2. Algoritmo de BERNOULLI de parámetro p

1. Generar U con densidad uniforme entre 0 y 1
2. Si $U \leq p$, devolver $X = 1$; en otro caso, devolver $X = 0$

7.6.3. Algoritmo de simulación de una variable *binomial* de parámetro n y p

Se parte del hecho de que una variable *binomial*, Y , de parámetros n y p , se puede descomponer como suma de n variables de BERNOULLI independientes

$$Y = X_1 + X_2 + \dots + X_n$$

Algoritmo:

1. Hacer $i = 1$
2. Generar U_i con densidad uniforme entre 0 y 1
3. Si $U_i \leq p$, hacer $X_i = 1$; en otro caso, hacer $X_i = 0$
4. Hacer $i = i + 1$. Si $i < n + 1$ pasar al paso 2
5. Devolver $Y = X_1 + X_2 + \dots + X_n$

7.7. Simulación de permutaciones al azar

Es muy frecuente que determinados experimentos requieran una simulación en la cual se elige un subconjunto de elementos de uno dado.

Un algoritmo directo sería es de seleccionar de una lista de n elementos un literal (x_1, x_2, \dots, x_n) . De manera que se selecciona un elemento de los n de la lista y se coloca en primera posición, posteriormente un elemento de $n - 1$ y se coloca en segunda, y así hasta escoger todos los deseados. Este algoritmo presenta el problema de que es necesario recordar los valores elegidos en pasos anteriores.

Uno de los algoritmos más óptimos para realizar esta operación es el siguiente, se parte de una ordenación de valores $(x_1 x_2 \dots x_{n-1} x_n)$.

1. Hacer $k = n$
2. Generar U con densidad uniforme entre 0 y 1, y hacer $I = 1 + [kU]$
3. Intercambiar x_1 con x_k
4. Hacer $k = k - 1$. Si $k > 1$ ir a paso 2
5. Devolver la ordenación en memoria