FACULTAD DE CIENCIAS

GRADO EN MATEMÁTICAS

Cálculo de Probabilidades 1

Resumen realizado por

Jose Maria Buades Rubio

Diciembre-2019

Cálculo de Probabilidades 1

1. La Experiencia del Azar

1.1 El concepto del Azar

Existe un tipo de experimentos denominados aleatorios. En ellos, se dice que el resultado del fenómeno es consecuencia del azar. De manera que se atribuyen al azar todos los efectos que no están predeterminados por sus causas.

1.2 La idea de Probabilidad

Se trata de medir el grado de verosimilitud de los diversos acontecimientos posibles, asignando una probabilidad a cada uno de ellos, que informe de la frecuencia con que hay que esperar que se presente cada uno, después de numerosas observaciones del fenómeno.

2. El Modelo Matemático de la Probabilidad

2.1 Introducción

En el estudio de cualquier fenómeno en que interviene el azar hay dos facetas fundamentales de las que ocuparse: Primero los posibles acontecimientos que pueden producirse y, en segundo lugar, la valoración de la probabilidad con que puede presentarse cada uno.
1 Un modelo matemático acerca de las situaciones aleatorias debe precisar la manera genérica de describir ambos aspectos y las reglas ineludibles a las que deben someterse tales descripciones.

2.2 Espacio muestral y sucesos

Los resultados posibles constituyen un conjunto finito que se denomina espacio muestral del fenómeno aleatorio en cuestión y se designa genéricamente por Ω.
Los subconjuntos del espacio muestral Ω se denominan sucesos.
Los subconjuntos de un único elemento se denominan sucesos simples.
Los sucesos complejos son aquellos que tienen más de un elemento, y son, por tanto, uniones de sucesos simples.

2.3 El Concepto de Probabilidad

Se pretende establecer una aplicación:

$$\mathrm{P} : \mathcal{P}(\Omega) \longrightarrow [0,1]$$

que asigne a cada suceso $A \in \mathcal{P}(\Omega)$, un número real, comprendido entre 0 y 1, que proporcione la probabilidad del suceso A.

En el caso de que Ω sea un conjunto finito, debe verificar:
(a) Si $A, B \subset \Omega$ y $A \cap B = \emptyset$, entonces:

$$\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B)$$

(b) $\mathrm{P}(\Omega) = 1$.

Los valores, $\mathrm{P}(A)$, se denominan probabilidades de los diversos sucesos A y el par (Ω, P) recibe el nombre de espacio de probabilidad finito.

2.4 Primeras Propiedades de la Probabilidad

1. Si $A \subset B \subset \Omega$, entonces $\mathrm{P}(A) \leq \mathrm{P}(B)$.

2. Si $A \subset B \subset \Omega$, entonces $\mathrm{P}(B - A) = \mathrm{P}(B) - \mathrm{P}(A)$.

3. Para cualquier $A \subset \Omega$, se cumple $\mathrm{P}(A^c) = 1 - \mathrm{P}(A)$.

4. $\mathrm{P}(\emptyset) = 0$.

5. Ya que $B = (B \cap A) \cup (B \cap A^c)$, se puede expresar
$\mathrm{P}(B) = \mathrm{P}(B \cap A) + \mathrm{P}(B - A)$
Lo cual enseña a descomponer la probabilidad del suceso B en dos casos: aquél en que también sucede A y aquél que A no ocurre.

6. Ya que $A \cup B = A \cup (B - A)$ se obtiene:
$\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)$

7. Si $A_1, A_2, A_3, ..., A_n$ son sucesos disjuntos dos a dos, es decir que se verifiquen $A_i \cap A_j = \emptyset$ para $i \neq j$, se cumple:
$\mathrm{P}(A_1 \cup A_2 \cup ... \cup A_n) = \mathrm{P}(A_1) + \mathrm{P}(A_2) + \cdots + \mathrm{P}(A_n)$

8. Si $\Omega = \{\omega_1, \omega_2, ..., \omega_n\}$ es el espacio muestral de un espacio de probabilidad finito (Ω, P) y $A = \{\omega_{i_1}, \omega_{i_2}, ..., \omega_{i_n}\}$ es un suceso, se cumple:
$\mathrm{P}(A) = \mathrm{P}(\{\omega_{i_1}\}) + \mathrm{P}(\{\omega_{i_2}\}) + \cdots + \mathrm{P}(\{\omega_{i_n}\})$
En consecuencia, basta conocer la probabilidad de cada uno de los sucesos simples para poder determinar la probabilidad de cualquier suceso. Dichos datos serán coherentes cuando sea
$\mathrm{P}(\{\omega_1\}) + \mathrm{P}(\{\omega_2\}) + \cdots + \mathrm{P}(\{\omega_n\}) = 1$

3. La Asignación de Probabilidades

3.1 Introducción

El *ajuste del modelo* trata de extraer de los mecanismos del azar empleados o de las condiciones del fenómeno aleatorio observado, un espacio muestral y, sobre todo, una asignación de probabilidades a los sucesos que resulten conformes con la realidad estudiada.

3.2 La Regla de Laplace

En un espacio muestral con n elementos, si puede observarse que todos los sucesos simples son igualmente probables, cada uno debe tener probabilidad $1/n$.

La probabilidad de un suceso A relativo a un fenómeno aleatorio es

$$\mathrm{P}(A) = \frac{\text{número de casos favorables a } A}{\text{número de casos posibles}}$$

supuesto que todos los casos posibles son igualmente probables

3.3 Métodos Combinatorios

Si se dispone de un espacio muestral con la característica de **homogeneidad** deseada podemos aplicar la regla de Laplace, y por tanto el problema de determinar la probabilidad de un suceso se convierte en un problema de combinatoria.

3.4 La concordancia entre el modelo y la realidad

No siempre se puede obtener un espacio muestral equiprobable, por ejemplo, la probabilidad de ser niño (P(niño) = 0.51) es ligeramente superior a la probabilidad de ser niña (P(niña) = 0.49).

Otro ejemplo más llamativo es el modelo de Boltzman-Maxwell, Bose-Einstein y Fermi-Dirac. Donde las partículas subatómicas se comportan de forma inidentificable, que hace que los mecanismos de azar que rigen para ellas sean diferentes de los habituales en el mundo macroscópico.

4. Las Fórmulas de Inclusión-Exclusión

4.1 Introducción

En muchas ocasiones la dificultad viene en la gran variedad de sucesos que un espacio contiene y de la posibilidad de describir, en términos simples, sucesos muy complejos.

4.2 Probabilidad de una unión de sucesos

Teorema 4.1 Si $A_1, A_2, ..., A_n$ son sucesos cualesquiera de un espacio de probabilidad, se cumple:

$$\mathrm{P}(A_1 \cup A_2 \cup ... \cup A_n) = S_1 - S_2 + S_3 - \cdots (-1)^{n-2} S_{n-1} + (-1)^{n-1} S_n$$

donde

$$S_k = \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq n} \mathrm{P}(A_{i_1} \cap A_{i_2} \cap ... \cap A_{i_k})$$

Para $n = 3$

$$\begin{aligned} \mathrm{P}(A_1 \cup A_2 \cup A_3) = &\, \mathrm{P}(A_1) + \mathrm{P}(A_2) + \mathrm{P}(A_3) \\ &- (\mathrm{P}(A_1 \cap A_2) + \mathrm{P}(A_1 \cap A_3) + \mathrm{P}(A_2 \cap A_3)) \\ &+ (\mathrm{P}(A_1 \cap A_2 \cap A_3)) \end{aligned}$$

4.3 *Probabilidad de que se realicen m sucesos

Teorema 4.2 En un espacio de probabilidad, si $A_1, A_2, ..., A_n$ son sucesos cualesquiera, la probabilidad $\mathrm{P}_{[m]}$ de que ocurran exactamente m de ellos, es

$$\mathrm{P}_{[m]} = S_m - \binom{m+1}{m} S_{m+1} + \binom{m+2}{m} S_{m+2} - \cdots + (-1)^{n-m} \binom{n}{m} S_n$$

donde

$$S_k = \sum_{1 \le i_1 < i_2 < \cdots < i_k \le n} \mathrm{P}(A_{i_1} \cap A_{i_2} \cap ... \cap A_{i_k})$$

Para $m = 0$, la probabilidad de que no ocurra ninguno de los n sucesos es

$$\mathrm{P}_{[0]} = 1 - \mathrm{P}(A_1 \cup A_2 \cup ... \cup A_n)$$

que según el resultado de la sección anterior vale

$$\mathrm{P}_{[0]} = 1 - S_1 + S_2 - S_3 + \cdots + (-1)^{n+1} S_n$$

El teorema es válido en el caso $m = 0$ si se define $S_0 = 1$.

5. Extensiones del Modelo Matemático

5.1 Introducción

Hasta ahora se ha tratado de describir un fenómeno aleatorio con un número finito de resultados posibles. En este capítulo se pretende ampliar el modelo para que pueda cubrir el caso de un número numerable de sucesos.

5.2 Espacios muestrales numerables

En ocasiones, el espacio muestral Ω que describe los posibles resultados de un fenómeno aleatorio es un conjunto numerable (en el sentido de infinito numerable). Por ejemplo, el número de lanzamientos de una moneda hasta obtener la primera cara.

En espacios muestrales numerables, no es necesario modificar la noción de suceso: Cualquier subconjunto de Ω es un suceso; simple o compuesto, según que contenga uno o más de los resultados posibles.

En casos como los anteriores, en que Ω es infinito numerable, el conjunto de sucesos, $\mathcal{P}(\Omega)$ no es numerable, sino que su cardinal es aún mayor ($2^{\mathbb{N}}$ exactamente). A diferencia con el caso de espacios muestrales finitos, es posible entonces considerar sucesiones de sucesos

$$A_1, A_2, A_3, ..., A_n, ...$$

mediante las cuales pueden formarse nuevos sucesos. En primer lugar

$$\bigcap_{n=1}^{\infty} A_n \quad \text{y} \quad \bigcup_{n=1}^{\infty} A_n$$

que describen, respectivamente, la ocurrencia de todos o alguno de los sucesos A_n. Pero también se puede considerar

$$\bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} A_n$$

que expresa que ocurren todos los A_n, a partir de algún m o, dicho más simplemente todos los A_n excepto un número finito de ellos.

Y, así mismo,

$$\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n$$

formado por aquellos resultados ω que, para cualquier m dado, pertenecen a un A_n con $n \geq m$; es decir, los que implican la ocurrencia de un número infinito de sucesos A_n.

5.3 Propiedades adicionales de la probabilidad

Sobre un espacio muestral numerable Ω, una probabilidad es una aplicación

$$\mathrm{P} : \mathcal{P}(\Omega) \longrightarrow [0, 1]$$

que verifique
(a) Si A_n es una sucesión de sucesos tales que $A_i \cap A_j = \emptyset$ para $i \neq j$, entonces

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

(b) $\mathrm{P}(\Omega) = 1$.

Si se cumplen ambas condiciones, (Ω, P) constituye un espacio de probabilidad numerable.
Por lo que se cumple la siguiente propiedad:

9. En un espacio de probabilidad numerable $\Omega = \{\omega_1, \omega_2, ..., \omega_n, ...\}$, cualquier suceso $A = \{\omega_{i_1}, \omega_{i_2}, ..., \omega_{i_k}, ...\}$ tiene probabilidad

$$\mathrm{P}(A) = \sum_{k=1}^{\infty} \mathrm{P}(\{\omega_{i_k}\})$$

Así pues, la probabilidad de cada uno de los sucesos simples es un conjunto de datos suficientes para poder determinar la probabilidad de cualquier suceso. Para que dichos datos sean correctos, tendrá que ser

$$\sum_{n=1}^{\infty} \mathrm{P}(\{\omega_n\}) = 1$$

10. Si A_n es una sucesión creciente de sucesos (es decir $A_n \subset A_{n+1}$), se cumple

$$\lim_{n\to\infty} \mathrm{P}(A_n) = \mathrm{P}\left(\bigcup_{n=1}^{\infty} A_n\right)$$

Análogamente, si A_n es una sucesión decreciente de sucesos (para la cual $A_{n+1} \subset A_n$), se cumple

$$\lim_{n\to\infty} \mathrm{P}(A_n) = \mathrm{P}\left(\bigcap_{n=1}^{\infty} A_n\right)$$

5.4 Modelos continuos

En blanco, la asignatura CP2 trata sobre modelos continuos.

6. Probabilidad Condicionada

6.1 Introducción

Es muy útil considerar situaciones en las que el experimento no ha concluido, pero ya se dispone de alguna información acerca de la actuación del azar en la realización del fenómeno que se está llevando a cabo. Es indudable que tal información puede modificar los juicios y previsiones sobre los resultados finales.

Se requiere para ello del concepto de probabilidad condicionada que se introduce a continuación.

6.2 Probabilidad condicionada

En espacios muestrales finitos, es frecuente que la probabilidad condicionada de A por B pueda calcularse directamente, enumerando los nuevos "casos posibles" - aquellos en los que sucede B - y los "casos favorables" - aquellos en los que sucede también A, para obtener la fórmula:

$$\mathrm{P}(A|B) = \frac{\text{número de casos favorables a } A \text{ y } B}{\text{número de casos en los que ocurre } B}$$

Para cualquier espacio muestra, se define:
Si A y B son sucesos de un cierto espacio de probabilidad y se cumple $\mathrm{P}(B) > 0$, la probabilidad de A condicionada por B es

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

A partir de la probabilidad condicionada se puede obtener la intersección

$$\begin{aligned}\mathrm{P}(A_1 \cap A_2 \cap ... \cap A_n) = &\mathrm{P}(A_1)\cdot \\ &\mathrm{P}(A_2|A_1)\cdot \\ &\mathrm{P}(A_3|A_1 \cap A_2)\cdot \\ &\cdots \\ &\mathrm{P}(A_n|A_1 \cap A_2 \cap ... \cap A_{n-1})\end{aligned}$$

6.3 Propiedades

Si en un espacio de probabilidades (Ω, P), se condiciona por un suceso fijo, $B \subset \Omega$, se obtiene una nueva asignación de probabilidades a los diversos sucesos de Ω, distinta de P, pero que no deja por ello de ser una probabilidad. Concretamente:

1. Si $A_1, A_2, ..., A_n, ...$ son sucesos disjuntos se cumple

$$\mathrm{P}(A_1 \cup A_2 \cup ... \cup A_n \cup ...|B) =$$
$$\mathrm{P}(A_1|B) + \mathrm{P}(A_2|B) + \cdots + \mathrm{P}(A_n|B) + \cdots$$

De forma que $\mathrm{P}(\cdot|B)$ es numerable aditiva.

2. Por otra parte,

$$\mathrm{P}(\Omega|B) = \frac{\mathrm{P}(\Omega \cap B)}{\mathrm{P}(B)} = 1$$

y más concretamente $\mathrm{P}(B|B) = 1$.
Con lo cual $(\Omega, \mathrm{P}(\cdot|B))$ ó $(B, \mathrm{P}(\cdot|B))$ son espacios de probabilidad. En consecuencia, $\mathrm{P}(\cdot|B)$ tiene todas las propiedades de cualquier probabilidad.

Fórmula de las probabilidades totales

Si $B_1, B_2, ..., B_n, ...$ es una familia se sucesos de probabilidad positiva, disjuntos dos a dos, tales que

$$B_1 \cup B_2 \cup ... \cup B_n \cup ... = \Omega$$

se verifica

$$\mathrm{P}(A) = \mathrm{P}(B_1)\mathrm{P}(A|B_1) + \mathrm{P}(B_2)\mathrm{P}(A|B_2) + \cdots + \mathrm{P}(B_n)\mathrm{P}(A|B_n) + \cdots$$

Fórmula de Bayes

Si $B_1, B_2, ..., B_n, ...$ es una familia se sucesos de probabilidad positiva, disjuntos dos a dos, tales que

$$B_1 \cup B_2 \cup ... \cup B_n \cup ... = \Omega$$

para cualquier suceso A de probabilidad positiva, de acuerdo con la definición, se tiene

$$\mathrm{P}(B_i|A) = \frac{\mathrm{P}(B_i \cap A)}{\mathrm{P}(A)} = \frac{\mathrm{P}(B_i) \cdot \mathrm{P}(A|B_i)}{\mathrm{P}(A)}$$

lo cual combinado con la fórmula de las probabilidades totales, se puede expresar

$$\mathrm{P}(B_i|A) = \frac{\mathrm{P}(B_i) \cdot \mathrm{P}(A|B_i)}{\mathrm{P}(B_1) \cdot \mathrm{P}(A|B_1) + \mathrm{P}(B_2) \cdot \mathrm{P}(A|B_2) + \cdots + \mathrm{P}(B_n) \cdot \mathrm{P}(A|B_n)}$$

6.4 El método recurrente

A veces interesa calcular la probabilidad p_n de que cierto suceso ocurra en la etapa n-ésima. Relacionando p_n con la probabilidad p_{n-1} que corresponde a la etapa anterior, de modo que la relación hallada

$$p_n = f(p_{n-1})$$

sea válida cualquiera que sea n, el problema está resuelto.

A veces, la relación recurrente no es tan simple. Puede ocurrir que p_n sea función no sólo de p_{n-1} sino también de p_{n-1}, etc. y que se cumple una relación de la forma

$$p_n = f(p_{n-1}, p_{n-2}, ..., p_{n-k})$$

Interesantes es el ejemplo 6.14, donde la probabilidad depende de los dos anteriores.

7. Independencia de Sucesos

7.1 Sucesos dependientes e independientes

Dos sucesos A y B de un espacio de probabilidad se dicen independientes si

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

Obsérvese que, si A y B son sucesos independientes, también lo son A y B^c. Lo mismo sucede por consiguiente, con A^c y B o con A^c y B^c.

7.2 Espacios producto

Sea (Ω_1, P_1) y (Ω_2, P_2) dos espacios probabilísticos, la realización conjunta de ambos experimentos se describe mediante el espacio muestral

$$\Omega = \Omega_1 \times \Omega_2$$

en el cual se considera la probabilidad definida por

$$\mathrm{P}(A \times B) = \mathrm{P}_1(A)\mathrm{P}_2(B)$$

para cualesquiera sucesos $A \subset \Omega_1$ y $B \subset \Omega_2$. El espacio de probabilidad (Ω, P) se denomina espacio de probabilidad producto de (Ω_1, P_1) y (Ω_2, P_2).

Si Ω es un conjunto finito, o numerable, de resultados de un fenómeno aleatorio, el conjunto $\Omega^{\mathbb{N}}$ de sucesiones $(a_i)_{i=1}^{\infty}$ compuestas por elementos de Ω, describe los resultados posibles de una serie ilimitada de repeticiones del fenómeno.

La probabilidad P sobre Ω permite definir una probabilidad producto, P^*, para todos los conjuntos cilíndricos $\Omega^{\mathbb{N}}$. El mecanismo para hacerlo es considerar para cada conjunto cilíndrico C, que precise el comportamiento de k términos de la sucesión, el conjunto B de Ω^k que constituye la base de C. Es decir, si

$$C = \{(a_i)_{i=1}^{\infty} | a_1 \in A_1, a_2 \in A_2, ..., a_k \in A_k\}$$

donde $A_1, A_2, ..., A_k$ son sucesos de Ω, entonces

$$B = A_1 \times A_2 \times ... \times A_k$$

La probabilidad de B en el espacio producto Ω^k y la probabilidad de C en $\Omega^{\mathbb{N}}$ coinciden. Esto es:

$$\mathrm{P}^*(C) = \mathrm{P}(A_1)\mathrm{P}(A_2)\cdots\mathrm{P}(A_k)$$

7.3 Independencia de varios sucesos

En un espacio de probabilidad, tres sucesos A_1, A_2 y A_3, son independientes si se cumplen las condiciones
1. $\mathrm{P}(A_1 \cap A_2) = \mathrm{P}(A_1)\mathrm{P}(A_2)$
2. $\mathrm{P}(A_1 \cap A_3) = \mathrm{P}(A_1)\mathrm{P}(A_3)$
3. $\mathrm{P}(A_2 \cap A_3) = \mathrm{P}(A_2)\mathrm{P}(A_3)$
4. $\mathrm{P}(A_1 \cap A_2 \cap A_3) = \mathrm{P}(A_1)\mathrm{P}(A_2)\mathrm{P}(A_3)$

Las tres primeras condiciones definen la independencia dos a dos de los tres sucesos. Sin embargo, no implican la cuarta; de forma que tres sucesos pueden ser independientes dos a dos, sin ser independientes.

Se puede extender a una familia de sucesos.
En un espacio de probabilidad, los sucesos $\{A_i | i \in I\}$ se dicen independientes si

$$\mathrm{P}(A_{i_1} \cap A_{i_2} \cap ... \cap A_{i_k}) = \mathrm{P}(A_{i_1})\mathrm{P}(A_{i_2})\cdots\mathrm{P}(A_{i_k})$$

cualquiera que sea $k \in \mathbb{N}$ y cualquiera que sean $i_1, i_2, ..., i_k \in I$.

7.4 *La independencia condicional

Sean A, B y C tres sucesos en un espacio de probabilidad (Ω, P). Si se verifica

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

se dice que A y B son independientes condicionalmente a C. Habida cuenta que

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

así que la primera condición puede escribirse de la siguiente forma:

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

o, por simetría

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

Lo que resulta sorprendente, es que puede darse las condiciones:

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

$$P(A|B \cap C^c) = P(A|C^c)$$

y no darse

$$P(A|B) = P(A)$$

O puede darse el caso contrario: dos sucesos A y B pueden ser independientes, pero no ser condicionalmente independientes a C ni a C^c.

Paradoja de Simpson

Consiste en que, simultáneamente, pueden darse las dos desigualdades

$$P(A|B \cap C) > P(A|C) \text{ y } P(A|B \cap C^c) > P(A|C^c)$$

y, sin embargo, cumplirse

$$P(A|B) < P(A)$$

8. Variables Aleatorias

8.1 Concepto de variable aleatoria

En un espacio de probabilidad discreto (Ω, P), se denomina variable aleatoria, a cualquier función

$$X : \Omega \longrightarrow \mathbb{R}$$

Como en el caso de los sucesos, cada variable aleatoria puede definirse de dos formas:

- mediante la especificación de su valor concreto sobre cada elemento $\omega \in \Omega$. Lo cual tendrá que hacerse a través de una fórmula, salvo si Ω tiene muy pocos elementos.

- mediante una descripción precisa de la correspondencia que establece entre Ω y $\mathbb{R}$.

El concepto de variable aleatoria generaliza el de suceso. Cada suceso $A \subset \Omega$ puede identificarse con la variable aleatoria

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

que suele denominarse *función indicatriz* del suceso A. Son afirmaciones equivalentes $I_a(\omega) = 1$ y $\omega \in A$; así que A e I_A especifican, de forma distinta, cuales son los elementos de A.
Cada variable aleatoria X tiene asociado gran número de sucesos. Concretamente:

- Para cada número real $a \in \mathbb{R}$, el suceso X toma el valor a se simboliza por $\{X = a\}$ y, con toda precisión, significa

  $$\{\omega \in \Omega | X(\omega) = a\}$$

  es decir, $\{X = a\} = X^{-1}(\{a\})$

- Dado un intervalo $(a, b]$ de $\mathbb{R}$, X toma un valor en $(a, b]$ o, en símbolos, $\{a < X \leq b\}$, es el suceso

$$\{X < a\} = \{\omega \in \Omega | a < X(\omega) \leq b\} = X^{-1}(a, b]$$

- En general, para cualquier subconjunto B de $\mathbb{R}$, el suceso X toma un valor en B significa

$$\{X \in B\} = \{\omega \in \Omega | X(\omega) \in B\} = X^{-1}(B)$$

8.2 Distribución de una variable aleatoria

La colección de probabilidades $\mathrm{P}\{X \in B\}$, correspondientes a cada subconjunto B de $\mathbb{R}$, se denomina la distribución de la variable aleatoria X.
Aunque la distribución de una variable aleatoria asigna una probabilidad a cada subconjunto de $\mathbb{R}$, no es necesario hacer una lista de los subconjuntos de $\mathbb{R}$ con sus probabilidades. Por el contrario, para variables aleatorias definidas en un espacio de probabilidad discreto el método más sencillo es mediante una función. En efecto, la imagen, $X(\Omega)$, de la variable aleatoria es un conjunto finito o numerable:

$$X(\Omega) = \{x_1, x_2, x_3, ..., x_n, ...\}$$

cuyos elementos, valores posibles de la variable, junto con el valor de

$$p_k = \mathrm{P}\{X = x_k\} \text{ para cada } x_k \in X(\Omega)$$

proporcionan datos suficientes para determinar la distribución completa. De hecho, para cada subconjunto B de $\mathbb{R}$, será

$$\mathrm{P}\{X \in B\} = \sum_{\{k | x_k \in B\}} p_k$$

Una variable aleatoria X es discreta si toma a lo sumo un número numerable de valores:

$$X(\Omega) = \{x_1, x_2, x_3, ..., x_n, ...\}$$

La función que a cada x_k asigna el valor

$$p_k = \mathrm{P}\{X = x_k\}$$

se denomina función de probabilidad de X.
Naturalmente, debe cumplirse

$$\sum_{k=1}^{\infty} p_k = 1$$

puesto que $\mathrm{P}\{X \in X(\Omega)\} = 1$.

Si se requiere de un espacio de probabilidad (Ω, P) en el que considerar una variable aleatoria discreta de la que sólo se conoce su distribución, la elección más simple es tomar:
$\Omega = \{x_1, x_2, ..., x_k, ...\}$ el conjunto de puntos a los que la función de probabilidad asigna un valor positivo.
$\mathrm{P}(\{x_k\}) = p_k$, para cada x_k, lo cual define una probabilidad P en Ω
$X(x_k) = x_k$ que es, desde luego, una variable aleatoria que tiene la función de probabilidad requerida.

La función

$$F(x) = \mathrm{P}\{X \leq x\} = \sum_{\{k | x_k \leq x\}} p_k$$

definida para cada $x \in \mathbb{R}$, se denomina función de distribución de la variable aleatoria X.
El valor $F(x)$ acumula las probabilidades de los valores x_k en el intervalo $(-\infty, x])$. Por eso, a veces, se denomina *función de distribución acumulativa.*

8.3 Variables aleatorias simultáneas

Si X_1, X_2 son variables aleatorias discretas definidas en el mismo espacio de probabilidad, la distribución conjunta de (X_1, X_2) es la función que asigna a cada subconjunto B de $\mathbb{R}^2$ la probabilidad

$$\mathrm{P}\{(X_1, X_2) \in B\}$$

En realidad, la distribución conjunta de (X_1, X_2) se caracteriza por la función de probabilidad conjunta que hace corresponder

$$\mathrm{P}\{X_1 = x_1, X_2 = x_2\}$$

a cada par (x_1, x_2) de valores posibles de (X_1, X_2) que siempre será un subconjunto de $X_1(\Omega_1) \times X_1(\Omega_2)$.
En este contexto:
Las funciones de probabilidad $\mathrm{P}\{X_1 = x_1\}$ y $\mathrm{P}\{X_2 = x_2\}$, se denominan funciones de probabilidad marginales de X_1 y X_2 respectivamente. Claramente:

$$\mathrm{P}\{X_1 = x_1\} = \sum_{x_2 \in X_2(\Omega)} \mathrm{P}\{X_1 = x_1, X_2 = x_2\}$$

$$\mathrm{P}\{X_2 = x_2\} = \sum_{x_1 \in X_1(\Omega)} \mathrm{P}\{X_1 = x_1, X_2 = x_2\}$$

La función de probabilidad

$$\mathrm{P}\{X_2 = x_2 | X_1 = x_1\} = \frac{\mathrm{P}\{X_1 = x_1, X_2 = x_2\}}{\mathrm{P}\{X_1 = x_1\}}$$

donde x_1 es fijo y x_2 variable, corresponde a la distribución de X_2 condicionada por $X_1 = x_1$. De igual forma $\mathrm{P}\{X_1 = x_1 | X_2 = x_2\}$ es la función de probabilidad de la distribución de X_1 condicionada por $X_2 = x_2$.

El concepto de distribución conjunta se extiende a más de dos variables aleatorias. La función de probabilidad conjunta de 3 variables aleatorias, especifica el valor de

$$\mathrm{P}\{X_1 = x_1, X_2 = x_2, X_3 = x_3\}$$

para cada terna (x_1, x_2, x_3) de valores posibles de las variables. Ello determina la distribución marginal de cada variable:

$$\mathrm{P}\{X_1 = x_1\} = \sum_{x_2 \in X_2(\Omega)} \sum_{x_3 \in X_3(\Omega)} \mathrm{P}\{X_1 = x_1, X_2 = x_2, X_3 = x_3\}$$

También, cualquier pareja de variables tiene su propia distribución conjunta que es una distribución marginal bidimensional de la distribución conjunta tridimensional. Por ejemplo

$$\mathrm{P}\{X_1 = x_1, X_2 = x_2\} = \sum_{x_3 \in X_3(\Omega)} \mathrm{P}\{X_1 = x_1, X_2 = x_2, X_3 = x_3\}$$

En cuanto a las distribuciones condicionales, sólo hacen uso de la definición de probabilidad condicionada. Por ejemplo, con 4 variables, la función de probabilidad condicionada

$$\mathrm{P}\{X_4 = x_4, X_3 = x_3 | X_2 = x_2, X_1 = x_1\} = \frac{\mathrm{P}\{X_1 = x_1, X_2 = x_2, X_3 = x_3, X_4 = x_4\}}{\mathrm{P}\{X_1 = x_1, X_2 = x_2\}}$$

8.4 Variables aleatorias independientes

Dos variables aleatorias discretas X_1 y X_2, definidas en el mismo espacio de probabilidad, se denomina independientes si se verifica

$$\mathrm{P}\{X_1 = x_1, X_2 = x_2\} = \mathrm{P}\{X_1 = x_1\}\,\mathrm{P}\{X_2 = x_2\}$$

cualesquiera que sean x_1 y x_2 entre los valores posibles de las variables.

Las variables aleatorias discretas $X_1, X_2, ..., X_r$, definidas en el mismo espacio de probabilidad, se denominan independientes si

$$\mathrm{P}\{X_1 = x_1, X_2 = x_2, ..., X_r = x_r\} = \mathrm{P}\{X_1 = x_1\}\,\mathrm{P}\{X_2 = x_2\} \cdots \mathrm{P}\{X_r = x_r\}$$

cualesquiera que sean $x_1, x_2, ..., x_r$ dentro de los conjuntos de valores posibles $X_1, X_2, ..., X_r$, respectivamente.

Lema

Si $X_1, X_2, ..., X_r$ son variables aleatorias discretas e independientes y $C_1, C_2, ..., C_r$ son subconjuntos cualesquiera de sus conjuntos de valores posibles, se verifica

$$\mathrm{P}\{X_1 \in C_1, X_2 \in C_2, ..., X_r \in C_r\} = \mathrm{P}\{X_1 \in C_1\}\,\mathrm{P}\{X_2 \in C_2\} \cdots \mathrm{P}\{X_r \in C_r\}$$

En consecuencia, los sucesos $\{X_1 \in C_1\}, \{X_2 \in C_2\}, ..., \{X_r \in C_r\}$ son independientes.

9. Esperanza Matemática

9.1 Valor esperado de una variable aleatoria

Si X es una variable aleatoria definida en un espacio de probabilidad discreto (Ω, P), el valor esperado, esperanza matemática o media de X es

$$E[X] = \sum_{\omega \in \Omega} X(\omega)\mathrm{P}(\omega)$$

en el supuesto de que

$$\sum_{\{\omega | X(\omega) > 0\}} X(\omega)\mathrm{P}(\omega) < \infty \text{ ó } \sum_{\{\omega | X(\omega) < 0\}} -X(\omega)\mathrm{P}(\omega) < \infty$$

Si las dos series anteriores divergen, se dice que $E[X]$ no existe; mientras que, si la primera diverge y la segunda converge, se toma $E[X] = +\infty$ y, si es al revés, $E[X] = -\infty$.

Si el conjunto de valores posibles de una variable aleatoria X es

$$X(\Omega) = \{x_1, x_2, ..., x_k, ...\}$$

su valor esperado, si existe, se expresa

$$E[X] = \sum_k x_k \mathrm{P}\{X = x_k\} = \sum_k x_k p_k$$

donde p_k es la función de probabilidad de X.

9.2 Propiedades de la esperanza matemática

1. Cualquier constante $c \in \mathbb{R}$ puede considerarse como un variable aleatoria; basta definir $X(\omega) = c$ para todo $\omega \in \Omega$. Su distribución se denomina distribución causal en c, puesto que asigna toda su probabilidad a dicho punto de la recta real. Por lo que

$$E[c] = c$$

2. Propiedad de Linealidad. Sean X_1 y X_2 variables aleatorias discretas definidas en el mismo espacio de probabilidad, cuyas medias son finitas. Si $c_1, c_2 \in \mathbb{R}$, la variable aleatoria $c_1X_1 + c_2X2$ cumple:

$$E[c_1X_1 + c_2X_2] = c_1E[X_1] + c_2E[X_2]$$

3. Sea X una variable aleatoria no negativa; es decir $X(\omega) \geq 0$ para todo $\omega \in \Omega$. Entonces

$$E[X] \geq 0$$

4. Si I_A es la función indicatriz de un suceso A, se tiene

$$E[I_A] = \mathrm{P}(A)$$

a) Si A y B son sucesos disjuntos, es claro que $I_{A\cup B} = I_A + I_B$, de forma que

$$\mathrm{P}(A \cup B) = E[I_{A\cup B}] = E[I_A] + E[I_B] = \mathrm{P}(A) + \mathrm{P}(B)$$

b) I_Ω es la variable aleatoria constante, igual a 1, luego se deduce $\mathrm{P}(\Omega) = 1$. Además, se observa que $I_A + I_{A^c} = 1$, prueba que $\mathrm{P}(A) + \mathrm{P}(A^c) = 1$.
c) Como $I_A \geq 0$ para cualquier suceso A, se deduce que $\mathrm{P}(A) \geq 0$. Además se obtiene que $\mathrm{P}(A) \leq \mathrm{P}(B)$ si $A \subset B$, puesto que $I_A \leq I_B$.
5. Si X es una variable aleatoria que toma sólo valores enteros no negativos, se cumple

$$E[X] = \sum_{m=1}^{\infty} \mathrm{P}\{X \geq m\}$$

o lo que es lo mismo $E[X] = \sum\limits_{m=0}^{\infty} \mathrm{P}\{X > m\}$. Por consiguiente, para una variable aleatoria X con valores enteros es

$$E[|X|] = \sum_{m=1}^{\infty} \mathrm{P}\{|X| \geq m\}$$

6. Si X e Y son variables aleatorias independientes, con esperanza finita, se verifica

$$E[XY] = E[X]E[Y]$$

9.3 Esperanza condicionada y métodos recurrentes

Sea un cierto suceso B con probabilidad $\mathrm{P}(B) > 0$. Si X es una variable aleatoria con valores posibles $x_1, x_2, ...$, el conocimiento de que el suceso B se ha producido, modifica la distribución de X puesto que altera toda la atribución de probabilidades sobre el espacio muestral. En concreto, los valores

$$\mathrm{P}\{X = x_k|B\} = \frac{\mathrm{P}(\{X = x_k\}) \cap B}{P(B)} \text{ con } k = 1, 2, 3...$$

constituyen la distribución de X condicionada por B, que da lugar a la esperanza matemática condicionada por B:

$$E[X|B] = \sum_{\omega \in \Omega} X(\omega)\mathrm{P}\{\omega|B\} = \sum_k x_k \mathrm{P}\{X = x_k|B\}$$

por lo que

$$E[X|B] = \frac{E[X \cdot I_B]}{\mathrm{P}(B)}$$

lo cual prueba que $E[X|B]$ existe, por lo menos, si $E[X]$ existe.

10. Análisis Descriptivo de las Distribuciones de Probabilidad

10.1 Introducción

Se desea asociar con cada distribución de probabilidad ciertas medidas que resuman su fisionomía de la mejor manera posible.

10.2 Momentos de una distribución

El momento de orden r respecto del origen de una variable aleatoria X, o de su distribución de probabilidad, es la esperanza matemática de X^r

$$E[X^r] = \sum_{\omega \in \Omega} X^r(\omega) \mathrm{P}\{\omega\} = \sum_k x_k^r p_k$$

que, cuando existe, suele designarse por α_r.
Cuando el momento de orden r es finito, también son finitos todos los momentos de orden inferior a r.

El momento de orden $r > 1$ respecto a la media o momento centrado de orden r de una variable aleatoria X, o de su distribución, es la esperanza matemática de $(X - E[X])^r$. Por tanto, si se le designa por μ_r, es

$$\mu_r = E[(X - E[X])^r] = \sum_k (x_k - \alpha_1)^r p_k$$

Por las propiedades de linealidad de la esperanza se obtiene:

$$\begin{aligned}
\mu_2 &= \alpha_2 - \alpha_1^2 \\
\mu_3 &= \alpha_3 - 3\alpha_1\alpha_2 + 2\alpha_1^3 \\
\mu_4 &= \alpha_4 - 4\alpha_1\alpha_3 + 6\alpha_1^2\alpha_2 - 3\alpha_1^4
\end{aligned}$$

La varianza de la distribución se designa por $V(X)$ ó $\sigma^2(X)$, o simplemente σ^2 si no da a confusión.

$$\sigma^2 = E\left[(X - E[X])^2\right] = \sum_k (x_k - \alpha_1)^2 p_k = E[X^2] - E[X]^2$$

Dese luego siempre es $\sigma^2 \geq 0$, así que se cumple $E[X^2] \geq E[X]^2$.
Si X se mide en ciertas unidades, σ^2 se mide en unidades cuadradas. Su raíz cuadrada, σ, que tiene la ventaja de tener las mismas unidades que la variable aleatoria, se denomina desviación típica de la distribución.

A veces, se utiliza el coeficiente de variación de la variable X, definido como $\sigma(X)/E[X]$.
Al ser$E[cX] = cE[X]$ se tiene

$$\sigma^2(cX) = E[(cX - E[cX])^2] = E[c^2(X - E[X])^2] = c^2\sigma^2(X)$$

También puede medirse alrededor de un punto genérico, mediante $E[(X - a)^2]$. Aplicando las propiedades de linealidad de la esperanza, resulta

$$E[(X - a)^2] = E[X^2 - 2aX + a^2] = E[X^2] - 2aE[X] + a^2$$

y alcanza en $a = E[X]$ su valor mínimo.

Desiguladad de Tchebychev

Cualquiera que sea la distribución de probabilidad de una variable aleatoria X, se verifica

$$\mathrm{P}\{|X - E[X]| > k\sigma\} \leq \frac{1}{k^2}$$

para cualquier $k > 0$, o lo que es lo mismo

$$\mathrm{P}\{|X - E[X]| > c\} \leq \frac{\sigma^2}{c^2}$$

para cualquier $c > 0$
Se puede generalizar como

$$\mathrm{P}\{|X - E[X]| > c\} \leq \frac{E[|X - E[x]|^r]}{c^r}$$

El momento central de tercer orden es de gran utilidad

$$\mu_3 = \sum_k (x_k - \alpha_1)^3 p_k$$

que se considera una medida de la asimetría de la distribución alrededor de su media.
Con el fin de obtener un número adimensional que no dependa de las unidades en que se exprese la variable, al asimetría se mide mediante el cociente

$$\gamma_3 = \mu_3/\sigma^3$$

coeficiente de asimetría de la distribución.

El momento central μ_4 sirve para evaluar el apuntamiento

$$\gamma_4 = \alpha_4/\sigma^4 - 3$$

se denomina coeficiente de apuntamiento o curtosis de la distribución. Dividir por σ^4 sirve para obtener un número adimensional; en cuanto a la razón de restarle 3 es conseguir que la distribución normal $N(0,1)$ tenga coeficiente de apuntamiento igual a cero.

10.3 Momentos de una distribución conjunta

Respecto al origen, los momentos de la distribución conjunta son el valor esperado del producto de dos potencias de las variables

$$\alpha_{r,s} = E[X_1^r X_2^s]$$

Respecto el centro $(E[X_1], E[X_2])$, se obtiene

$$\mu_{r,s} = E[(X_1 - E[X_1])^r (X_2 - E[X_2])^s]$$

Los valor esperado del producto de variables

$$\alpha_{1,1} = E[X_1 X_2]$$

y la covarianza

$$\mu_{1,1} = E[(X_1 - E[X_1])(X_2 - E[X_2])]$$

que suele representarse por $Cov(X_1, X_2) = E[X_1X_2] - E[X_1]E[X_2]$.
El coeficiente de correlación

$$\rho(X_1, X_2) = \frac{Cov(X_1, X_2)}{\sigma(X_1)\sigma(X_2)}$$

Como $Cov(c_1X_1, c_2X_2) = c_1c_2Cov(X_1, X_2)$ se tiene que $\rho(c_1X_1, c_2X_2) = \rho(X_1, X_2)$ de forma que el coeficiente de correlación no depende de las unidades de medida.
Dos variables independientes tienen covarianza nula y, por tanto, $\rho(X_1, X_2) = 0$. Las variables cuyo coeficiente de correlación es nulo se denomina incorreladas pero, recuérdese, ello no significa que sean independientes.

La matriz de covarianzas de X_1 y X_2 es la matriz

$$\Sigma(X_1, X_2) = \begin{pmatrix} \sigma^2(X_1) & Cov(X_1, X_2) \\ Cov(X_1, X_2) & \sigma^2(X_2) \end{pmatrix}$$

Se trata de una matriz simétrica semidefinida positiva.

El coeficiente de correlación $-1 < \rho(X_1, X_2) < 1$ mide muy bien la existencia de una relación lineal entre las variables.
Siendo la mejor previsión de X_2, la que se obtiene mediante la función lineal de X_1, usando la recta

$$x_2 = a^*(x_1 - E[X_1]) + E[X_2]$$

donde

$$a^* = \frac{Cov(X_1, X_2)}{\sigma^2(X_1)}$$

y x_1 en función de x_2

$$x_1 = \frac{Cov(X_1, X_2)}{\sigma^2(X_2)}(x_2 - E[X_2]) + E[X_1]$$

Si $X_1, X_2, ..., X_n$ son variables aleatorias independientes, se verifica

$$\sigma^2(X_1 + X_2 + \cdots + X_n) = \sigma^2(X_1) + \sigma^2(X_2) + \cdots + \sigma^2(X_n)$$

10.4 Otros indicadores de posición y dispersión

10.4.1 Indicadores de posición

Moda

La moda de una distribución que asigna probabilidades $p_1, p_2, ...$ a los puntos $x_1, x_2, ...$, es el valor x_k para el cual p_k es máxima.

Mediana

La mediana de una distribución es aquel valor M que deja probabilidad 1/2 tanto por debajo como por encima de él; más exactamente M ha de cumplir

$$\mathrm{P}\{X \le M\} \ge 1/2 \qquad \text{y} \qquad \mathrm{P}\{X \ge M\} \ge 1/2$$

Al igual que la media es el valor de a que hace mínimo $E[(X-a)^2]$, la mediana minimiza $E[|X-a|]$.

10.4.2 Indicadores de dispersión

En balística $M(|X - M(X)|)$ se denomina desviación probable. Que es la mediana de la desviación absoluta a la mediana.

Se denomina cuantil de orden p de la distribución al menor valor x_k de la variable tal que $\mathrm{P}\{X \le x_k\} \ge p$ con $p \in [0, 1]$.

$$c_p = \min \{x_k | F(x_k) \geq p\}$$

Por lo que $c_{1/2}$ es la mediana. Por otra parte, $c_{1/4}$ y $c_{3/4}$ se llaman respectivamente el primer y tercer cuartil. El intervalo $[c_{1/4}, c_{3/4}]$ tiene probabilidad superior o igual $1/2$.

11. Pruebas Repetidas

11.1 Introducción

Cuando se considera situaciones en las que un mismo experimento aleatorio se repite, una vez tras otra, sin que el resultado de cada realización pueda influir en el resultado de las demás. Se denomina un *esquema de pruebas repetidas* i, con más precisión, *repetidas e independientes.*

11.2 La distribución binomial y la aproximación de Poisson

Si en cada realización de un experimento aleatorio el suceso A tiene probabilidad p de ocurrir, al repetir n veces el experimento, el número total de veces que se presenta el suceso A es una variable aleatoria X_n, con función de probabilidad

$$\mathrm{P}\left\{X_n = k\right\} = \binom{n}{k} p^k q^{n-k} \qquad k = 0, 1, 2, ..., n$$

donde $q = 1 - p$. Tratándose de la distribución binomial $B(n, p)$.

$$E[x_n] = np \qquad \text{y} \qquad \sigma^2(X_n) = np(1-p)$$

El cálculo con la distribución binomial sólo es sencillo para valores pequeños de n.

Teorema de Poisson

Si n tiende a infinito y p tiende a cero, de tal manera que el producto np converge a una constante λ, se verifica

$$\lim \binom{n}{k} p^k q^{n-k} = e^{-\lambda \frac{\lambda^k}{k!}}$$

para cada valor fijo de k.
Por lo que en estas condiciones, la distribución binomial se aproxima a la

distribución de Poisson.

La aproximación es buena desde valores moderados de n aunque, cuanto más grande sea λ, mayor ha de ser n para que el error sea pequeño.

11.3 La aproximación normal a la distribución binomial

Cuando tanto n como np tienen valores elevados, la aproximación de Poisson da malos resultados.

Torema de de Moivre-Laplace

Cuando n tiende a infinito, si k tiende hacia infinito de manera que

$$x_k = \frac{k - np}{\sqrt{npq}}$$

se mantiene acotado en valor absoluto por alguna constante A, se cumple

$$\binom{n}{k} p^k q^{n-k} \sim \frac{1}{\sqrt{2\pi npq}} e^{-x_k^2/2}$$

en el sentido de que el cociente de ambos miembros converge a 1. Más exactamente se tiene

$$\binom{n}{k} p^k q^{n-k} \sim \frac{1}{\sqrt{2\pi npq}} e^{-x_k^2/2}(1 + \rho_n)$$

donde $\sqrt{n}|\rho_n|$ está acotado, a partir de un n en adelante, por una constante que sólo depende de A (y de p).

Corolario

Si X_n es una variable aleatoria con distribución binomial $B(n, p)$ y

$$Z_n = \frac{X_n - npq}{\sqrt{npq}}$$

para cualesquiera $a < b \in \mathbb{R}$, se cumple

$$\mathrm{P}\{a < Z_n \leq b\} \longrightarrow \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx$$

cuando n tiende a infinito. En las mismas condiciones se tiene también

$$\mathrm{P}\{Z_n \leq y\} \longrightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^y e^{-x^2/2} dx$$

para cualquier $y \in \mathbb{R}$.

Una variable X_n, con distribución binomial $B(n,p)$, se puede expresar

$$X_n = \sum_{i=1}^{n} I_i$$

donde $I_1, I_2, ..., I_i, ...$ es una sucesión de variables de Bernoulli. Además la variable tipificada Z_n es

$$Z_n = \frac{X_n - E[X_n]}{\sigma(X_n)}$$

y la distribución Z_n se aproxima a la $\mathcal{N}(0,1)$ cuando n crece.

11.4 La ley débil de los grandes números

Ley de los grandes números de Bernoulli

Si X_n tiene distribución binomial $B(n,p)$, para cualquier $\epsilon > 0$ se cumple

$$\lim_{n\to\infty} \mathrm{P}\left\{\left|\frac{X_n}{n} - p\right| < \epsilon\right\} = 1$$

Ley de los grandes números de Tchebychev

Si $I_1, I_2, ..., I_n$ son variables aleatorias independientes, con la misma distribución de media μ y varianza σ^2, para cualquier $\epsilon > 0$ se cumple

$$\lim_{n\to\infty} \mathrm{P}\left\{\left|\frac{X_n}{n} - \mu\right| < \epsilon\right\} = 1$$

Por supuesto el teorema de Bernoulli es el caso particular de este, en que cada sumando I_i sólo puede tomar los valores 1 y 0 con probabilidades p y $1-p$ respectivamente, con lo cual $\mu = p$.

I. Series más usadas

Sumatorios

$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2} \qquad \sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6} \qquad \sum_{i=1}^{n} i^3 = \frac{n^2(n+1)^2}{4}$$

Series geométricas

$$\sum_{i=0}^{n} c^i = \frac{c^{n+1}-1}{c-1}, c \neq 1 \qquad \sum_{i=0}^{\infty} c^i = \frac{1}{1-c}, |c| < 1 \qquad \sum_{i=1}^{\infty} c^i = \frac{c}{1-c}, |c| < 1$$

$$\sum_{i=0}^{n} i \cdot c^i = \frac{nc^{n+2} - (n+1)c^{n+1} + c}{(c-1)^2}, c \neq 1 \qquad \sum_{i=0}^{\infty} i \cdot c^i = \frac{c}{(1-c)^2}, |c| < 1$$

Números combinatorios

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

$$\sum_{k=0}^{n} \binom{n}{k}^2 = \binom{2n}{k} \qquad \sum_{k=0}^{n} \binom{n}{k} = 2^n \qquad \sum_{k=0}^{n} \binom{r+k}{k} = \binom{r+n+1}{n}$$

$$\sum_{k=0}^{n} \binom{k}{m} = \binom{n+1}{m+1}$$

$$\sum_{k=0}^{m-1} \binom{m}{k}\binom{n-1}{m-k-1} = \binom{n+m-1}{n} \qquad \sum_{k=0}^{n} \binom{r}{k}\binom{s}{n-k} = \binom{r+s}{n}$$

$$\binom{n}{k} = \sum_{j=1}^{n-k+1} \binom{n-j}{k-1} = \sum_{j=0}^{k} \binom{n-j-1}{k-j}$$

$$\binom{n}{k} = (-1)^k \binom{k-n-1}{k} \qquad \sum_{k=1}^{n} \frac{\binom{n-1}{k-1}}{\binom{n+m}{k}} = \frac{n+m+1}{(m+1)(m+2)}$$

Fórmula Stirling

$$n! \sim \sqrt{2n\pi}n^n e^{-n}$$

en el sentido de que el cociente de ambos tiende hacia 1 cuando n crece hacia infinito. Con más precisión

$$1 < \frac{n!}{\sqrt{2n\pi}n^n e^{-n}} < e^{1/(8n)}$$

Variables Aleatorias

Binomial. Número de aciertos en n pruebas de Bernoulli.

$$X = B(n,p)\ P\{X = x\} = \binom{n}{x} p^x (1-p)^{n-x}, x = 0, 1, ..., n$$

$$E[X] = np \qquad \sigma^2(X) = np(1-p)$$

Geométrica. Número de intentos hasta el primer acierto.

$$X = Geom(p)\ P\{X = x\} = (1-p)^x p, x = 0, 1, 2, ...$$

$$E[X] = \frac{1-p}{p} \qquad \sigma^2(X) = \frac{1-p}{p^2}$$

Binomial Negativa. Número de fallos hasta el n acierto

$$X = BN(n,p)\ P\{X = x\} = \binom{n+x-1}{n-1} (1-p)^x p^n, x = 0, 1, 2, ...$$

$$E[X] = n\frac{1-p}{p} \qquad \sigma^2(X) = n\frac{1-p}{p^2}$$

Poisson

$$X = Poisson(\lambda)\ P_X(x) = \frac{e^{-\lambda}\lambda^x}{x!}, x = 0, 1, 2, ...$$

$$E[X] = \lambda = \sigma^2(X)$$

Normal

$$X = N(\mu, \sigma)\ f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, -\infty < x < \infty$$

$$E[X] = \mu \qquad \sigma^2(X) = \sigma^2$$

Número e

$$e = 1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24} + \frac{1}{120} + \cdots$$

$$e^x = \lim_{n \to \infty} \left(1 + \frac{x}{n}\right)^n$$

$$\left(1 + \frac{1}{n}\right)^n < e < \left(1 + \frac{1}{n}\right)^{n+1}$$

Sumatorio de Poisson

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{\lambda}$$

Función Gamma

$$\Gamma(p) = \int_{0^+}^{+\infty} x^{p-1} e^{-x} \, dx$$

$$\Gamma(p+1) = p \cdot \Gamma(p)$$

$\Gamma(0) = (-1)! = \infty$ $\Gamma(1) = 0!$ $\Gamma(2) = 1!$ $\Gamma(3) = 2!$

$\Gamma(1/2) = \sqrt{\pi}$ $\Gamma(3/2) = 1/2\sqrt{\pi}$ $\Gamma(5/2) = 3/4\sqrt{\pi}$ $\Gamma(7/2) = 15/8\sqrt{\pi}$

Función Beta

$$\beta(p,q) = \int_{0^+}^{1^-} x^{p-1}(1-x)^{q-1} \, dx$$

$$\beta(p,q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}, \; (p > 0, q > 0)$$

Teorema de Taylor

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + ... \frac{f^{(n)}(a)}{n!}(x-a)^n + E_n(x)$$

$$e^x = 1 + x + \frac{1}{2}x^2 + \frac{1}{6} + \cdots = \sum_{i=0}^{\infty} \frac{x^i}{i!}$$

$$ln(1+x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \frac{1}{4}x^4 + \cdots = \sum_{i=1}^{\infty} (-1)^{i+1} \frac{x^i}{i}$$

$$ln\frac{1}{1-x} = x + \frac{1}{2}x^2 + \frac{1}{3}x^3 + \frac{1}{4}x^4 + \cdots = \sum_{i=1}^{\infty} \frac{x^i}{i}$$

$$(1+x)^n = 1 + nx + \frac{n(n+1)}{2}x^2 + \cdots = \sum_{i=0}^{\infty} \binom{n}{i} x^i$$