Guillermo Ayala y Francisco Montes Universidad de Valencia

Probabilidad básica

Guillermo.Ayala@uv.es, Francisco.Montes@uv.es

Índice general

Prólogo

Nos movemos en ambientes con incertidumbre. Tomamos decisiones sin saber muy bien qué va a pasar y, por lo tanto, si era correcta la decisión **previamente** tomada. Convivimos con la incertidumbre. Y no lo podemos evitar. Quizás podemos intentar cuantificarla con objeto de proceder con sensatez.[1]

[1] Si tenemos claro qué es eso de la sensatez.

Este material son unas notas para un primer curso en Probabilidad orientado a estudiantes de Matemáticas. Se proponen pruebas de los resultados a un nivel introductorio. En ocasiones, son pruebas parciales o utilizando resultados que no vemos. Es un primer curso de Probabilidad. Contienen, además, una amplia colección de problemas. Son unas notas de clase que intentan estar conectadas con otros textos u otro tipo de material. Indicamos las referencias de donde sacamos problemas para que el estudiante pueda buscar más ejercicios en esos textos.

Otras referencias bibliográficas con un nivel matemático similar al que aquí utilizamos son [1, 9, 10, 11, 6, 12].

Utilizamos el lenguaje R para trabajar con los modelos probabilísticos. **¿Para qué un lenguaje de programación en un curso de Probabilidad?** Nos evita cálculos estúpidos o usar las antediluvianas tablas de percentiles propias del siglo.[2] También es útil utilizar la apps **Probability distributions**[3] para trabajar con las distribuciones de probabilidad.

[2] Usarlas hoy es un ejercicio de nostalgia masoca.

[3] Disponible gratuitamente para Android e iOS.

Capítulo 1

Espacio de probabilidad

1.1 Causalidad y aleatoriedad

A cualquiera que preguntemos cuanto tiempo tardaríamos en recorrer los 350 kilómetros que separan Valencia de Barcelona, si nos desplazamos con velocidad constante de 100 kms/hora, nos contestará sin dudar que 3 horas y media. Su actitud será muy distinta si, previamente a su lanzamiento, le preguntamos por la cara que nos mostrará un dado. Se trata de dos fenómenos de naturaleza bien distinta,

- *el primero* pertenece a los que podemos denominar **deterministas**, aquellos en los que la relación causa-efecto aparece perfectamente determinada. En nuestro caso concreto, la conocida ecuación $e = v \cdot t$, describe dicha relación,

- *el segundo* pertenece a la categoría de los que denominamos **aleatorios**, que se caracterizan porque aun repitiendo en las mismas condiciones el experimento que lo produce, el resultado variará de una repetición a otra dentro de un conjunto de posibles resultados.

La Probabilidad pretende emular el trabajo que los físicos y, en general, los científicos experimentales han llevado a cabo. Para entender esta afirmación observemos que la ecuación anterior, $e = v \cdot t$, es un resultado experimental que debemos ver como *un modelo matemático* que, haciendo abstracción del móvil concreto y del medio en el que se desplaza, describe la relación existente entre el espacio, el tiempo y la velocidad. La Probabilidad utiliza *modelos aleatorios o estocásticos* mediante los cuales podremos estudiar el comportamiento de los fenómenos aleatorios. [1]

[1] Estas son unas notas de Probabilidad. Hay dos denominaciones a este tipo de contenido: Una es la de *Cálculo de Probabilidades* y la otra la de *Teoría de la Probabilidad*. Se suele entender que un curso de Cálculo de Probabilidades no se ocupa de la fundamentación y se centra más en el uso de los resultados probabilísticos. Un curso de Teoría de la Probabilidad presenta una fundamentación rigurosa de la Probabilidad utilizando Teoría de la Medida. Estas notas serían más un curso de Cálculo de Probabilidades. Hablaremos de Probabilidad en este manual.

1.2 Experimento, resultado, espacio muestral y suceso

Nuestro interlocutor sí que será capaz de responder que el dado mostrará una de sus caras. Al igual que sabemos que la extracción al azar de una carta de una baraja española pertenecerá a uno de los cuatro palos: oros, copas, espadas o bastos. Es decir, el *experimento* asociado a nuestro fenómeno aleatorio da lugar a un *resultado*, ω, de entre un conjunto de posibles resultados. Este conjunto de posibles resultados recibe el nombre de espacio muestral y lo denotaremos por Ω. A los subconjuntos de resultados los llamaremos **sucesos aleatorios** o, simplemente, **sucesos**. Cuando el resultado del experimento pertenece al suceso A ($\omega \in A$) decimos que **ha ocurrido** o **se ha realizado** A.

Veamos ejemplos sencillos de experimentos aleatorios indicando los espacios muestrales y algún suceso de interés.

[4] ¿Cómo empezar un curso de Probabilidad sin hablar de lanzar una moneda? Mejor no decir qué moneda. Dejémoslo en que la moneda es una peseta. Si se nos pierde tampoco perdemos tanto.

Ejemplo 1.1 (Lanzamiento de una moneda) [4] *Cogemos la moneda y la lanzamos. Al resultado consistente en que vemos la cara le llamamos cara y cruz a la situación en que vemos una cruz. Por tanto, el espacio muestral es*

$$\Omega = \{\omega_1, \omega_2\},$$

donde entendemos que ω_1 significa que ha salido cara y ω_2 significa que ha salido cruz. Son elementos de un conjunto. Podemos denotarlos como nos de la gana siempre que sea claro. Por ejemplo, representar la cara por C y la cruz con $+$. Si lo hacemos el espacio muestral es

$$\Omega = \{C, +\}.$$

Y los sucesos aleatorios de interés serían

$$A = \{C\}, \ , B = \{+\}.$$

Con palabras, el suceso A significa que nos sale cara y el suceso B que nos sale cruz.

Ejemplo 1.2 (Lanzamiento de un dado) [2] *Ahora el espacio muestral podríamos indicarlo con*

$$\Omega = \{1, \ldots, 6\},$$

donde el número 1 significa que nos sale el número 1. Y así los demás. Sucesos que pueden interesar (sobre todo si estamos apostando y por lo tanto ganamos y perdemos dinero) podrían ser

$$A = \{2, 4, 6\},$$

que indica que obtenemos un número par. O

$$A = \{5, 6\},$$

que el número obtenido es mayor o igual a 5. O cada uno de los resultados elementales consistentes en obtener un número determinado.

[2] Del dado tampoco nos libramos.

Ejemplo 1.3 (Lanzamiento de dos monedas) *Al lanzar dos monedas el espacio muestral viene definido por* $\Omega = \{CC, C+, +C, ++\}$. *Dos ejemplos de sucesos en este espacio pueden ser:*

A =*{Ha salido una cara}={C+,+C},*

B =*{Ha salido más de una cruz}={++}.*

Ejemplo 1.4 (Un número al azar en el intervalo unitario) *Parece un experimento raro.*[5] *En este caso los posibles resultados son cada número real en el intervalo unitario* $[0, 1]$. *Posibles sucesos serían:* $A = [0.27, 0.5]$, *que el número aleatorio sea mayor o igual a* 0.27 *y menor o igual que* 0.5.

[5] De hecho, es uno de los experimentos más importantes.

Ejemplo 1.5 (Elegir un punto al azar en el círculo unidad) *Su espacio muestral es el disco unitario centrado en el origen:* $\Omega = \{x \in \mathbb{R}^2 : \parallel x \parallel \leq 1\}$. *Ejemplos de sucesos pueden ser:*

$A = \{\omega : \parallel \omega \parallel < 0.5\}$,

$B = \{\omega : \ 0,3 < \parallel \omega \parallel < 0.75\}$.

Ej. 1 — En cada uno de los siguientes experimentos indicar el espacio muestral correspondiente.

1. Lanzamiento de una moneda.
2. Lanzamiento simultáneo de dos monedas.
3. Lanzamiento simultáneo de n monedas.
4. Lanzamiento de un dado.
5. Lanzamiento simultáneo de dos dados.
6. Elección al azar de una persona en el censo de votantes de la Comunidad Valenciana.
7. Elección al azar de un árbol en un bosque.
8. Selección aleatoria (y sin reemplazamiento) de 7 números entre los números que van del 1 al 50.
9. Damos una mano en una baraja de póquer.
10. Lanzamos un dardo sobre una diana.
11. Partimos una rosquilleta en dos trozos de un modo aleatorio.

Sucesos, conjuntos y σ-álgebra de sucesos

Puesto que los sucesos no son más que subconjuntos de Ω, podemos operar con ellos de acuerdo con las reglas de la teoría de conjuntos. Todas las operaciones entre conjuntos serán aplicables a los sucesos y el resultado de las mismas dará lugar a nuevos sucesos cuyo significado debemos conocer. Existen, por otra parte, sucesos cuya peculiaridad e importancia nos lleva a asignarles nombre propio. De estos y de aquellas nos ocupamos a continuación:

Suceso cierto o seguro: cuando llevamos a cabo cualquier experimento aletorio es seguro que el resultado pertenecerá al espacio muestral, por lo que Ω, en tanto que suceso, ocurre siempre y recibe el nombre de *suceso cierto* o *seguro*.

Suceso imposible: en el extremo opuesto aparece aquel suceso que no contiene ningún resultado que designamos mediante $\emptyset$ y que, lógicamente, no ocurre nunca, razón por la cual se le denomina *suceso imposible*.

Sucesos complementarios: la ocurrencia de un suceso, A, supone la no ocurrencia del suceso que contiene a los resultados que no están en A, es decir, A^c. Ambos sucesos reciben el nombre de *complementarios*.

Unión de sucesos: la unión de dos sucesos, $A \cup B$, da lugar a un nuevo suceso que no es más que el conjunto resultante de dicha unión. En consecuencia, $A \cup B$ *ocurre* cuando el resultado del experimento pertenece a A, a B o ambos a la vez.

Intersección de sucesos: la intersección de dos sucesos, $A \cap B$, es un nuevo suceso cuya *realización* tiene lugar si el resultado pertenece a ambos a la vez, lo que supone que ambos ocurren simultáneamente.

Sucesos incompatibles: Existen sucesos que al no compartir ningún resultado su intersección es el suceso imposible, $A \cap B = \emptyset$. Se les denomina, por ello, *sucesos incompatibles*. Un suceso A y su complementario A^c, son un buen ejemplo de sucesos incompatibles.

¿Nos va a interesar la probabilidad de todos los posibles sucesos? O simplemente, ¿podremos dar la probabilidad de cualquier suceso? Conviene una pequeña reflexión: la noción de suceso es un concepto que surge con naturalidad en el contexto de la experimentación aleatoria. De la misma forma, es necesario que los sucesos poseean una mínima *estructura* que garantice la estabilidad de las operaciones *naturales* que con ellos realicemos, entendiendo por naturales la *complementación*, la *unión* y la *intersección*. Esta dos últimas merecen comentario aparte para precisar que no se trata de uniones e intersecciones en número cualquiera, puesto que más allá de la numerabilidad nos movemos con dificultad. Bastará pues que se nos garantice que uniones e intersecciones numerables de sucesos son estables y dan lugar a otro suceso. Existe una estructura algebraica que verifica las condiciones de estabilidad que acabamos de enumerar.

Definición 1.1 (σ-álgebra de conjuntos) *Una familia de conjuntos $\mathbb{A}$ definida sobre Ω decimos que es una σ-álgebra si:*

1. $\Omega \in \mathbb{A}$.
2. $A \in \mathbb{A} \Rightarrow A^c \in \mathbb{A}$.
3. $\{A_n\}_{n\geq 1} \subset \mathbb{A} \Rightarrow \bigcup_{n\geq 1} A_n \in \mathbb{A}$.

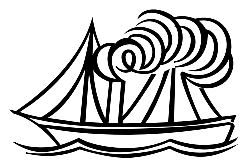
La familia de las partes de Ω[3] , $\mathbb{P}(\Omega)$, cumple con la definición y es por tanto una σ-álgebra de sucesos, de hecho la más grande de las existentes. En muchas ocasiones excesivamente grande para nuestras necesidades, que vienen determinadas por el colección inicial de sucesos de interés o importantes.

[3] Recordar que dado un conjunto A, se define el conjunto partes de A,$\mathbb{P}(A)$, como otro conjunto formado por todos los subconjuntos del conjunto dado.

Ejemplo 1.6 *Si suponemos que nuestro experimento consiste en elegir al azar un número en el intervalo [0,1], nuestro interés se centrará en conocer si la elección pertenece a cualquiera de los posibles subintervalos de [0,1]. La σ-álgebra de sucesos generada a partir de ellos, que es la menor que los contiene, se la conoce con el nombre de* σ-álgebra de Borel en [0,1], $\beta_{[0,1]}$, *y es estrictamente menor que* $\mathbb{P}([0,1])$. *No todos los subconjuntos del intervalo unitario son conjuntos de Borel.*

En resumen, el espacio muestral vendrá acompañado de la correspondiente σ-álgebra de sucesos, la más conveniente al experimento. La pareja que ambos constituyen, $(\Omega, \mathbb{A})$, recibe el nombre de *espacio medible*.

Señalemos por último que en ocasiones no es posible economizar esfuerzos y $\mathbb{A}$ coincide con $\mathbb{P}(\Omega)$. Por ejemplo, cuando el espacio muestral es numerable.



1.3 Probabilidad

Ya sabemos que la naturaleza aleatoria del experimento impide *predecir* de antemano el resultado que obtendremos al llevarlo a cabo. Queremos conocer si cada suceso de la σ-álgebra se realiza o no. En otras palabras, si dado un suceso A el resultado ω del experimento está en A o no. Esto no es posible. No podemos predecir en cada realización del experimento si el resultado va a estar o no en cada suceso. En Probabilidad la pregunta se formula del siguiente modo: ¿qué posibilidad hay de que tenga lugar cada uno de los sucesos? La respuesta exige un tercer elemento que nos proporcione esa información: Una función de conjunto P, es decir, una función definida sobre la σ-álgebra de conjuntos de sucesos, que a cada uno de ellos le asocie un valor numérico que exprese la mayor o menor probabilidad o posibilidad o certidumbre de producirse cuando se realiza el experimento. Esta función de conjunto se conoce como medida de probabilidad o simplemente probabilidad.

Empezamos con un poco de historia. El concepto de probabilidad aparece ligado en sus orígenes a los juegos de azar, razón por la cual se tiene constancia del mismo desde tiempos remotos. A lo largo de la historia se han hecho muchos y muy diversos intentos para formalizarlo, dando lugar a otras tantas *definiciones* de probabilidad que adolecían todas ellas de haber sido confeccionadas *ad hoc*, careciendo por tanto de la generalidad suficiente que permitiera utilizarlas en cualquier contexto. No por ello el interés de estas definiciones es menor. Veamos un par de ejemplos.

El **método frecuencialista** se puede aplicar en experimentos susceptibles de ser repetidos en las mismas condiciones una infinidad de veces. En este caso, la probabilidad de un suceso A, $P(A)$, se define como el límite[4] al que tiende la *frecuencia relativa de ocurrencias* del

[4] Debemos advertir que no se trata aquí de un límite puntual en el sentido habitual del Análisis. Más adelante se introducirá el tipo de convergencia al que nos estamos refiriendo.

suceso A. Por ejemplo, el típico problema de lanzar una moneda. Si la moneda está bien construida todos tenemos la intuición de que si lanzamos un gran número de veces la moneda y nos fijamos en la frecuencia de veces que sale cara entonces en el límite la frecuencia de caras será de $1/2$ que sería la probabilidad del suceso consistente en salir cara.

En el **método clásico o fórmula de Laplace** si el experimento tiene un espacio muestral **finito** con n resultados posibles, $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$, la probabilidad de un suceso A ($\subset \Omega$) se define como

$$P(A) = \frac{|A|}{|\Omega|}.$$

[5] Esta definición de probabilidad se suele expresar diciendo que la probabilidad del suceso es el cociente del número de casos favorables (el cardinal de A) dividido por el número de casos posibles (el número total de resultados o cardinal del espacio muestral Ω) o simplemente *casos favorables partido por los casos posibles*. Es la conocida fórmula de Laplace , que la propuso a finales del siglo XVIII.

Las anteriores definiciones son aplicables cuando las condiciones exigidas al experimento son satisfechas y dejan un gran número de fenómenos aleatorios fuera de su alcance. Estos problemas se soslayan con la definición axiomática propuesta por A.N. Kolmogorov en 1933 [3].

Definición 1.2 (medida de probabilidad) *Una función de conjunto, P, definida sobre la σ-álgebra $\mathbb{A}$ es una medida de probabilidad si:*

1. $P(A) \geq 0$ *para todo* $A \in \mathbb{A}$.
2. $P(\Omega) = 1$.
3. *P es numerablemente aditiva, es decir, si $\{A_n\}_{n\geq 1}$ es una sucesión de sucesos disjuntos de $\mathbb{A}$, entonces*

$$P\left(\bigcup_{n\geq 1} A_n\right) = \sum_{n\geq 1} P(A_n).$$

A la terna $(\Omega, \mathbb{A}, P)$ la denominaremos **espacio de probabilidad**.

Los axiomas propuestos por Kolmogorov no son más que la generalización de las propiedades que posee la frecuencia relativa. La definición se apoya en las aproximaciones previas existentes y al mismo tiempo las incluye como situaciones particulares que son.

Ejemplo 1.7 (Probabilidades discretas) *Supongamos que observamos el número de microorganismos en un cultivo; el número de defectos en un monitor de ordenador o en una pieza de cerámica o, en general, en una unidad de producto. O el número de personas a favor de una propuesta mediante una encuesta electoral. En todos los ejemplos indicados el conjunto de posibles resultados es finito o, si es infinito, en cualquier caso es numerable. Veamos la situación con generalidad.*

1. *Tenemos un espacio muestral, $\Omega = \{\omega_n\}_{n\geq 1}$, numerable.*

[5] Denotamos por $|A|$ el cardinal del conjunto A.

2. *Tomamos como σ-álgebra de conjuntos la familia formada por todos los posibles subconjuntos de* Ω.

3. *Sea* p *una función no negativa definida sobre* Ω *verificando:* $\sum_{n\geq 1} p(\omega_n) = 1$.

4. *Definimos*
$$P(A) = \sum_{\omega_n \in A} p(\omega_n).$$

Se verifica que la función de conjunto que acabamos de definir es una medida de probabilidad. Un espacio de probabilidad así recibe el nombre de espacio de probabilidad discreto.[6] *Hay un caso particular de especial interés, el llamado* espacio de probabilidad discreto uniforme, *en el que* Ω *es finito,* $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$. *Además la función* p *viene dada por*
$$p(\omega_n) = \frac{1}{N}, \ \forall \omega_n \in \Omega.$$
Entonces, para $A = \{\omega_{n_1}, \omega_{n_2}, \ldots, \omega_{n_m}\}$ *se tiene*
$$P(A) = \frac{m}{n},$$
que es la fórmula de Laplace. El nombre de uniforme *se justifica porque la masa unitaria de probabilidad está uniformemente repartida sobre todos los posibles resultados ya que es constante en cada punto.*

Los ejemplos más inmediatos de espacios de probabilidad discretos uniformes pueden ser los siguientes:

1. *Lanzamiento de una moneda correcta: dos resultados (cara y cruz) equiprobables.*

2. *Lanzamiento de un dado correcto: seis resultados equiprobables.*

3. *La lotería nacional donde los resultados son los distintos números que se juegan y todos ellos son equiprobables.*[6]

4. *En un estudio de opinión si elegimos al azar a un individuo de la población. El espacio muestral es el conjunto de individuos que componen la población en estudio (habitualmente los habitantes de una nación o región). Se elige* al azar, *esto es, uniformemente a una persona de la población.*

[6] Al menos para nosotros lo son. Sin embargo, no lo es para muchas personas. Cuando ocurrió el atendado en las Ramblas de Barcelona el 17 de agosto de 2017 se agotaron inmediatamente los décimos de la Lotería Nacional correspondientes a la fecha 17817. Podemos consultarlo en http://www.elmundo.es/cataluna/2017/09/16/59bc352a468aebb46c8b4596.html Hay que destacar (y alabar) la enorme preocupación que los compradores mostraban por las personas fallecidas.

No todos los experimentos producen una cantidad numerable de resultados.

Ejemplo 1.8 (Probabilidades geométricas) *Un ejemplo conocido puede ser lanzar un dardo en un diana. Si el lanzador no es un experto*[7] *lo más que logrará será que el dardo de en la diana pero sin poder precisar dónde. De alguna manera es un* lanzamiento uniforme *en la diana. Con un poco más de formalismo este experimento se puede ver como elegir un punto al azar en un círculo de radio 1, el espacio muestral resultante estará formado por todos los puntos del círculo*

[6] Si el espacio medible, $(\Omega, \mathbb{A})$, es arbitrario pero la probabilidad P verifica que existe un subconjunto numerable de Ω, $D = \{\omega_k, \ k \geq 1\}$, y una sucesión de valores no negativos, $\{p_k\}_{k\geq 1}$, tal que $P(A) = \sum_{\omega_k \in A\cap D} p_k$, se dice que la *probabilidad es discreta*. Observemos que debe verificarse $P(D) = \sum_{\omega_k \in D} p_k = 1$.

[7] Como sabemos los expertos lanzadores de dardos suelen ser también expertos bebedores de cerveza antes o después del lanzamiento.

$\Omega = \{(\omega_1, \omega_2);\ \omega_1^2 + \omega_2^2 \leq 1\}$. *La elección al azar implica que la masa de probabilidad se distribuye uniformemente en todo el círculo, lo que significa que cualquiera de sus puntos puede ser igualmente elegido. Las características del espacio no permiten afirmar, como hacíamos en el caso finito, que cada punto tiene la misma probabilidad. Ahora la uniformidad se describe afirmando que la probabilidad de cualquier suceso A es directamente proporcional a su área, $P(A) = k|A|$. Pero $P(\Omega) = 1 = k|\Omega| = k\pi$, de donde $k = 1/\pi$, y de aquí,*

$$P(A) = \frac{|A|}{|\Omega|} = \frac{|A|}{\pi},\ \forall A \subset \Omega.$$

Por ejemplo, si A ={puntos que distan del centro menos de 1/2}, A será el círculo de radio 1/2 y

$$P(A) = \frac{\pi/4}{\pi} = \frac{1}{4}.$$

Ej. 2 — (*[9, pág 3]*)Lanzamos un dado correcto y anotamos el valor en la cara superior. Elegimos un segundo dado y repetimos la operación. Se pide:

1. ¿Cuál es la probabilidad que de la suma de los dos números sea igual a 5?
2. ¿Qué probabilidad tenemos de que un dado muestre el número 2 y el otro el 4?
3. ¿Qué probabilidad tenemos de que el segundo dado muestre un número mayor al primero?
4. ¿Qué probabilidad tenemos de que el segundo dado muestre un número menor al primero?

Ej. 3 — (*[9, pág. 4]*)Repetimos el problema anterior pero vamos a suponer que los dados tienen n caras (con $n \geq 4$) y que están bien construidos.[8]

1.4 Propiedades de la probabilidad

De la definición de probabilidad se deducen algunas propiedades muy útiles.

Proposición 1.1 *1. La probabilidad del vacío es cero.*

2. Aditividad finita: Si $A_1, \ldots, A_n$ son elementos disjuntos de $\mathbb{A}$ entonces

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i).$$

3. $P(A^c) = 1 - P(A)$.

4. Monotonía: Si $A, B \in \mathbb{A},\ A \subset B,$ entonces $P(A) \leq P(B)$.

5. Probabilidad de una unión cualquiera de sucesos: Si $A_1, \ldots, A_n \in \mathbb{A}$, entonces

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \ldots + (-1)^{n+1} P(A_1 \cap \ldots \cap A_n). \quad (1.1)$$

[8] Todas las caras equiprobables.

6. *Continuidad de la probabilidad: Sea $\{A_n\}_{n\geq 1}$ una sucesión monótona creciente de sucesos y sea A su límite.*[9] *Se tiene que*

$$P(A) = \lim_{n\to+\infty} P(A_n), \tag{1.2}$$

o que es continua desde abajo. *Análogamente si $A_n \downarrow A$*[10] *entonces*

$$P(A) = \lim_{n\to+\infty} P(A_n). \tag{1.3}$$

o que es continua desde arriba.

7. *Subaditividad: Dados los sucesos $A_1, \ldots, A_n$ tenemos*

$$P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i). \tag{1.4}$$

Si tenemos una sucesión de sucesos $\{A_n\}_{n\geq 1}$ entonces

$$P\left(\bigcup_{n\geq 1} A_n\right) \leq \sum_{n\geq 1} P(A_n). \tag{1.5}$$

La desigualdad anterior (bien en el caso finito o numerable) es conocida como la desigualdad de Boole.

Demostración 1.1 1. *La probabilidad del vacío es cero: $\Omega = \Omega \cup \emptyset \cup \emptyset \cup \ldots$ y por la aditividad numerable, $P(\Omega) = P(\Omega) + \sum_{k\geq 1} P(\emptyset)$, de modo que $P(\emptyset) = 0$.*

2. *Aditividad finita: Si $A_1, \ldots, A_n$ son elementos disjuntos de $\mathbb{A}$, aplicando la σ-aditividad, la propiedad anterior y haciendo $A_i = \emptyset,\ i > n$ tendremos*

$$P\left(\bigcup_{i=1}^{n} A_i\right) = P\left(\bigcup_{i\geq 1} A_i\right) = \sum_{i=1}^{n} P(A_i).$$

3. *Se deduce de aquí fácilmente que $\forall A \in \mathbb{A},\ P(A^c) = 1 - P(A)$.*

4. *Monotonía: Si $A, B \in \mathbb{A},\ A \subset B$, entonces de $P(B) = P(A) + P(B \setminus A)$ se deduce que $P(A) \leq P(B)$.*

5. *Probabilidad de una unión cualquiera de sucesos (fórmula de inclusión-exclusión): Para su obtención observemos que si $n = 2$ es cierta, pues*

$$\begin{aligned} P(A_1 \cup A_2) &= P(A_1) + P(A_2 \setminus A_1) = \\ &P(A_1) + P(A_2 \setminus A_1) + P(A_1 \cap A_2) - P(A_1 \cap A_2) = \\ &P(A_1) + P(A_2) - P(A_1 \cap A_2). \end{aligned} \tag{1.6}$$

El resto se sigue por inducción.

---

[9] Se dice que la sucesión de conjuntos A_n crece al conjunto límite A y lo denotamos por $A_n \uparrow A$ si $A_n \subset A_{n+1},\ \forall n$ y $\bigcup_{n\geq 1} A_n = A$.

[10] Se dice que los conjuntos A_n decrecen al conjunto A y lo denotamos por $A_n \downarrow A$ si $A_{n+1} \subset A_n,\ \forall n$ y $\bigcap_{n\geq 1} A_n = A$.

6. *Continuidad de la probabilidad: Sea* $\{A_n\}_{n\geq 1}$ *una sucesión monótona creciente de sucesos y sea* A *su límite. Si, a partir de la sucesión inicial, definimos* $B_n = A_n \setminus \bigcup_{i=1}^{n-1} A_i = A_n \setminus A_{n-1}$, *para* $n > 1$, *y* $B_1 = A_1$, *se tiene*

$$P(A) = P\left(\bigcup_{n\geq 1} A_n\right) = P\left(\bigcup_{n\geq 1} B_n\right) = \sum_{n\geq 1} P(B_n) =$$

$$\lim_{n\to+\infty} \sum_{j=1}^{n} P(B_j) = \lim_{n\to+\infty} P\left(\bigcup_{j=1}^{n} B_j\right) = \lim_{n\to+\infty} P(A_n). \quad (1.7)$$

Esta propiedad se conoce como continuidad desde abajo. *Si la sucesión es decreciente y* $A_n \downarrow A$, *la sucesión de complementarios será creciente y aplicando la continuidad desde abajo y que* $P(B^c) = 1 - P(B)$, *tendremos que* $P(A_n) \downarrow P(A)$, *que se conoce como* continuidad desde arriba.

7. *Subaditividad: Dados los sucesos* $A_1, \ldots, A_n$, *la relación existente entre la probabilidad de la unión de los* A_i *y la probabilidad de cada uno de ellos es la siguiente:*

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

En efecto, sean $B_1 = A_1$ *y* $B_i = A_i \setminus \bigcup_{j=1}^{i-1} A_j$ *para* $i = 2, \ldots, n$. *Los* B_i *son disjuntos y* $\bigcup_{i=1}^{n} B_i = \bigcup_{i=1}^{n} A_i$. *Por la aditividad finita y la monotonía de* P *se tiene*

$$P\left(\bigcup_{i=1}^{n} A_i\right) = P\left(\bigcup_{i=1}^{n} B_i\right) = \sum_{i=1}^{n} P(B_i) \leq \sum_{i=1}^{n} P(A_i).$$

Si se trata de una sucesión de sucesos, $\{A_n\}_{n\geq 1}$, *se comprueba tomando límites en ambos miembros de la desigualdad anterior que*

$$P\left(\bigcup_{n\geq 1} A_n\right) \leq \sum_{n\geq 1} P(A_n).$$

Ejercicios

* **Ej. 4** — Supongamos dos sucesos A y B. Demostrar que

$$P(A \cap B) \geq P(A) + P(B) - 1.$$

Ej. 5 — (*[9, pág. 32]*)Utilizando la desigualdad de Boole y el hecho de que $(\cup_{i=1}^{n} A_i)^c = \cap_{i=1}^{n} A_i^c$ demostrar que

$$P(B_1 \cap \ldots \cap B_n) \geq \sum_{i=1}^{n} P(B_i) - (n-1).$$

* **Ej. 6** — [[5, Cap. 2]] Supongamos unos sucesos A_n con $n \geq 1$ y tales que $P(A_n) = 1$. Demostrar que $P(\cap_{n\geq 1} A_n) = 1$.

1.5 Probabilidad condicionada

Si compramos un número para una rifa que se celebra anualmente durante las fiestas de verano en nuestro pueblo y que está compuesta por 100 boletos numerados del 1 al 100, sabemos que nuestra probabilidad ganar el premio, suceso que designaremos por A, vale

$$P(A) = \frac{1}{100}$$

Supongamos que a la mañana siguiente de celebrase el sorteo alguien nos informa que el boleto premiado termina en 5. Con esta información, ¿continuaremos pensando que nuestra probabilidad de ganar vale 10^{-2}? Desde luego sería absurdo continuar pensándolo si nuestro número termina en 7, porque evidentemente la *nueva* probabilidad valdría $P'(A) = 0$, pero aunque terminara en 5 también nuestra probabilidad de ganar habría cambiado, porque los números que terminan en 5 entre los 100 son 10 y entonces

$$P'(A) = \frac{1}{10},$$

10 veces mayor que la inicial.

Supongamos que nuestro número es el 35 y repasemos los elementos que han intervenido en la nueva situación. De una parte, un suceso original $A =${*ganar el premio con el número 35*}, de otra, un suceso $B =${*el boleto premiado termina en 5*} de cuya ocurrencia se nos informa *a priori*. Observemos que $A \cap B =${*el número 35*} y que la nueva probabilidad encontrada verifica,

$$P'(A) = \frac{1}{10} = \frac{1/100}{10/100} = \frac{P(A \cap B)}{P(B)},$$

poniendo en evidencia algo que cabía esperar, que la *nueva* probabilidad a depende de $P(B)$. Estas propiedades observadas justifican la definición que damos a continuación.

Definición 1.3 (Probabilidad condicionada) *Sea* $(\Omega, \mathbb{A}, P)$ *un espacio de probabilidad y sean* A *y* B *dos sucesos, con* $P(B) > 0$*, se define la* probabilidad de A condicionada a B *mediante la expresión,*

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

A la anterior expresión se la denomina *probabilidad* porque verifica los tres axiomas que definen el concepto de probabilidad, como fácilmente puede comprobarse. De entre los resultados y propiedades que se derivan de este nuevo concepto, tres son especialmente relevantes: el teorema de factorización, el teorema de la probabilidad total y el teorema de Bayes.

Ejercicios

* **Ej. 7** — Consideremos un espacio de probabilidad $(\Omega, \mathbb{A}, P)$ B tal que $P(B) > 0$. Se pide demostrar que la función de conjunto

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

considerada sobre la σ-álgebra de conjuntos $\mathbb{A}$ es una medida de probabilidad, esto es, que verifica los axiomas de una medida de probabilidad.

* **Ej. 8** — Diremos que un conjunto B repele al conjunto A si $P(A|B) < P(A)$, y lo atrae si $P(A|B) > P(A)$. Demostrar que si B atrae a A, entonces A atrae a B y B^c repele a A.

Teorema de factorización

De la definición de probabilidad condicionada se deduce que la probabilidad de la intersección de dos sucesos puede expresarse de la forma $P(A \cap B) = P(A|B)P(B)$. El teorema de factorización extiende este resultado para cualquier intersección finita de sucesos.

Consideremos los sucesos $A_1, A_2, \ldots, A_n$, tales que $P(\cap_{i=1}^n A_i) > 0$, por inducción se comprueba fácilmente que

$$P\left(\bigcap_{i=1}^n A_i\right) = P\left(A_n | \cap_{i=1}^{n-1} A_i\right) P\left(A_{n-1} | \cap_{i=1}^{n-2} A_i\right) \ldots P(A_2|A_1)P(A_1).$$

Ejemplo 1.9 *En una urna que contiene 5 bolas blancas y 4 negras, llevamos a cabo 3 extracciones consecutivas sin reemplazamiento. ¿Cuál es la probabilidad de que las dos primeras sean blancas y la tercera negra?*

Cada extracción altera la composición de la urna y el total de bolas que contiene. De acuerdo con ello tendremos (la notación es obvia)

$$P(B_1 \cap B_2 \cap N_3) == P(N_3|B_1 \cap B_2)P(B_2|B_1)P(B_1) = \frac{4}{7} \cdot \frac{4}{8} \cdot \frac{5}{9} \quad (1.11)$$

* **Ej. 9** — Demostrar el teorema de factorización.

Teorema de la probabilidad total

Supongamos que $A_1, A_2, \ldots, A_n$ constituyen un partición del Ω, esto es, $\cup_{i=1}^n A_i = \Omega$ y $A_i \cap A_j = \emptyset$. Supongamos también que tienen probabilidad no nula, $P(A_i) > 0$, $\forall i$. Entonces cualquier suceso B puede expresarse como $B = \cup_{i=1}^n B \cap A_i$ donde los conjuntos $B \cap A_i$ son disjuntos. Por tratarse de una unión disjunta podemos escribir

$$P(B) = \sum_{i=1}^n P(B \cap A_i) = \sum_{i=1}^n P(B|A_i)P(A_i). \quad (1.12)$$

Este resultado se conoce con el nombre de teorema de la probabilidad total.

* **Ej. 10** — Tenemos tres urnas. La primera tiene una bola blanca y una negra, la segunda una blanca y dos negras y la tercera tiene una bola blanca y tres negras. El experimento consiste en elegir una urna al azar. Una vez elegida la urna, elegimos una bola al azar dentro de la urna. Se pide calcular la probabilidad de obtener una bola blanca.

* **Ej. 11** — [[5, sección 1.4]] Tenemos n urnas. La r-ésima urna tiene r-1 bolas rojas y n-r bolas verdes. Elegimos una urna al azar y sacamos dos bolas al azar sin reemplazamiento. Determinar la probabilidad de:

1. La segunda bola es verde;

2. la segunda bola es verde, dado que la primera es verde.

Aplicación 1.1 (Encuesta sobre cuestiones *delicadas*) *Es bien conocida la reticencia de la gente a contestar cualquier encuesta, reticencia que se convierte en clara desconfianza y rechazo si el cuestionario aborda lo que podríamos denominar* temas delicados*: situación económica, creencias religiosas, afinidades políticas, costumbres sexuales, consumo de estupefacientes, ... El rechazo y la desconfianza están casi siempre basados en la creencia de una insuficiente garantía de anonimato. Es comprensible, por tanto, el afán de los especialistas en convencer a los encuestados de que el anonimato es absoluto. El teorema de la probabilidad total puede ayudar a ello.*

Supongamos que un sociólogo está interesado en conocer el consumo de drogas entre los estudiantes de un Instituto de Bachillerato. Elige 100 estudiantes al azar y para garantizar la confidencialidad de las respuestas, que sin duda redundará en un resultado más fiable, diseña una estrategia consistente en que cada estudiante extrae al azar un bola de un saco o urna que contiene 100 bolas numeradas del 1 al 100, conservándola sin que nadie la vea,

- *si el número de la bola elegida está entre el* 1 *y el* 70*, contesta a la pregunta* ¿has consumido drogas alguna vez?*,*
- *si el número de la bola elegida está entre el* 71 *y el* 100*, contesta a la pregunta* ¿es par la última cifra de tu DNI?.

En ambos casos la respuesta se escribe sobre un trozo de papel sin indicar, lógicamente, a cuál de las dos preguntas se está contestando.

Realizado el proceso, las respuestas afirmativas han sido 25 y para estimar la proporción de los que alguna vez han consumido droga aplicamos (1.12),

$$\begin{aligned} P(\text{sí}) \quad &= \quad P(\text{sí}|\textit{pregunta delicada})P(\textit{pregunta delicada})+ \\ &\quad\; P(\text{sí}|\textit{pregunta intrascendente})P(\textit{pregunta intrascendente}) \end{aligned}$$

Sustituyendo,

$$0.25 = P(\text{sí}|\textit{pregunta delicada}) \times 0.7 + 0.5 \times 0.3,$$

y despejando,

$$P(\text{sí}|\textit{pregunta delicada}) = \frac{0.25 - 0.15}{0.7} \approx 0.14$$

Es obvio que $P(\textit{pregunta intrascendente})$ *ha de ser conocida muy aproximadamente, como en el caso de la terminaciones del DNI, que por mitades deben de ser pares o impares.*

Teorema de Bayes

Puede tener interés, y de hecho así ocurre en muchas ocasiones, conocer la probabilidad asociada a cada elemento de la partición dado que ha ocurrido B, es decir, $P(A_i|B)$.

Teorema 1.1 (Teorema de Bayes) *Supongamos que los sucesos* $\{A_i\}_{i=1}^n$ *forman una partición del espacio muestral* Ω[11] *y consideramos un suceso arbitrario* B *tal que* $P(B) > 0$ *entonces se verifica*

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

[11] Estamos asumiendo $A_i \cap A_j = \emptyset$ si $i \neq j$ y que $\cup_{i=1}^n A_i = \Omega$.

Este resultado, conocido como el teorema de Bayes, permite conocer el cambio que experimenta la probabilidad de A_i como consecuencia de haber ocurrido B. En el lenguaje habitual de la Probabilidad a $P(A_i)$ se la denomina probabilidad *a priori* y a $P(A_i|B)$ probabilidad *a posteriori*, siendo la ocurrencia de B la que establece la frontera entre el antes y el después. ¿Cuál es, a efectos prácticos, el interés de este resultado? Veámoslo con un ejemplo.

Ejemplo 1.10 *Tres urnas contienen bolas blancas y negras. La composición de cada una de ellas es la siguiente:* $U_1 = \{3B, 1N\}$, $U_2 = \{2B, 2N\}$, $U_3 = \{1B, 3N\}$. *Se elige al azar una de las urnas, se extrae de ella una bola al azar y resulta ser blanca. ¿Cuál es la urna con mayor probabilidad de haber sido elegida?*

Mediante U_1, U_2 *y* U_3, *representaremos también la urna elegida. Estos sucesos constituyen una partición de* Ω *y se verifica, puesto que la elección de la urna es al azar,*

$$P(U_1) = P(U_2) = P(U_3) = \frac{1}{3}.$$

Si B=\{la bola extraída es blanca\}, *tendremos*

$$P(B|U_1) = \frac{3}{4}, \quad P(B|U_2) = \frac{2}{4}, \quad P(B|U_3) = \frac{1}{4}.$$

Lo que nos piden es obtener $P(U_i|B)$ *para conocer cuál de las urnas ha originado, más probablemente, la extracción de la bola blanca. Aplicando el teorema de Bayes a la primera de las urnas,*

$$P(U_1|B) = \frac{\frac{1}{3}\cdot\frac{3}{4}}{\frac{1}{3}\cdot\frac{3}{4} + \frac{1}{3}\cdot\frac{2}{4} + \frac{1}{3}\cdot\frac{1}{4}} = \frac{3}{6},$$

y para las otras dos, $P(U_2|B) = 2/6$ *y* $P(U_3|B) = 1/6$. *Luego la primera de las urnas es la que con mayor probabilidad dió lugar a una extracción de bola blanca.*

[12]

1.6 Independencia

A veces conocer la ocurrencia de un suceso no modifica la probabilidad de ocurrencia de otro. Y es muy interesante. Veamos el concepto de independencia de dos sucesos y luego vamos a extender este concepto al caso de colecciones de sucesos.

Dos sucesos independientes

Veamos un ejemplo simple.

[12] El teorema de Bayes es uno de estos resultados nos llevan a pensar que *la cosa no era para tanto*. Se tiene ante él la sensación que produce lo trivial, hasta el punto de atrevernos a pensar que lo hubiéramos podido deducir nosotros mismos de haberlo necesitado, aunque afortunadamente el Reverendo Thomas Bayes se ocupó de ello en un trabajo titulado *An Essay towards solving a Problem in the Doctrine of Chances*, publicado en 1763. Conviene precisar que Bayes no planteó el teorema en su forma actual, que es debida a Laplace.

Ejemplo 1.11 (Un tetraedro coloreado) *Tenemos un tetraedro con una cara roja, una cara negra, una cara blanca y la cuarta cara pintada con los tres colores. Admitimos que el tetraedro está bien construido, de manera que al lanzarlo sobre una mesa tenemos la misma probabilidad de que se apoye sobre una cualquiera de las cuatro caras, a saber, $p = \frac{1}{4}$. El experimento consiste en lanzar el tetraedro y ver en que posición ha caido. Si*

$$R = \{\textit{el tetraedro se apoya en una cara con color } \text{rojo}\}$$

$$N = \{\textit{el tetraedro se apoya en una cara con color } \text{negro}\}$$

$$B = \{\textit{el tetraedro se apoya en una cara con color } \text{blanco}\},$$

En este ejemplo podemos comprobar con facilidad que $P(R|N) = P(R)$ y esto es cierto para cada par de sucesos que tomemos entre los tres considerados.

¿Qué ocurre con un par de sucesos donde no se modifica la probabilidad de uno condicionando al otro? Es claro que se deduce la siguiente igualdad

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

Esto nos permite dar la siguiente definición.

Definición 1.4 (Sucesos independientes) *Sean A y B dos sucesos. Decimos que A y B son sucesos independientes si $P(A \cap B) = P(A)P(B)$.*

De esta definición se obtiene como propiedad,

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

y su simétrica $P(B|A) = P(B)$.

En ocasiones se define la independencia de dos sucesos a partir de este resultado, obteniéndose entonces como propiedad la que nosotros hemos dado como definición. Existe equivalencia entre ambas definiciones, aunque a fuerza de ser rigurosos, hay que matizar que definir el concepto a partir de la probabilidad condicional exige añadir la condición de que el suceso condicionante tenga probabilidad distinta de cero. Hay además otra ventaja a favor de la definición basada en la factorización de $P(A \cap B)$, pone de inmediato en evidencia la *simetría* del concepto.

* **Ej. 12** — Lanzamos un dado y consideramos los sucesos A, obtener un número par, y B, obtener un número impar. ¿Son A y B independientes?

* **Ej. 13** — Consideramos el experimento propuesto en el ejemplo 1.11. Se pide

1. Comprobar que todos los pares de sucesos que podemos formar con R,N y B son independientes. De hecho, se dice que son independientes dos a dos.

2. ¿Son independientes los sucesos R y $N \cap B$?

Un conjunto de sucesos independientes

Cuando trabajamos con más de dos sucesos simultáneamente la cosa ya no es tan simple. El ejemplo 1.11 tenía un cierto grado de sorpresa. Saber algo sobre si un color ha aparecido nos debiera decir algo sobre si alguno de los otros ha salido. Y no ha sido así. Tengo tres sucesos: R,N y B. Supongamos que conozco si se han producido dos de ellos. ¿Sigo sin modificar mi probabilidad de que ocurra el otro? Por ejemplo, sé que ha ocurrido N y B, esto es, sé que se ha producido $N \cap B$: ¿cuál es la probabilidad condicionada de R? Tenemos que $P(B|N \cap B) = 1$. Por lo tanto hay algún tipo de dependencia entre *todos* estos sucesos. En Probabilidad se define independencia y por negación la dependencia.

El concepto de independencia puede extenderse a una familia finita de sucesos de la siguiente forma.

Definición 1.5 (Independencia mutua) *Se dice que los sucesos de la familia* $\{A_1, \ldots, A_n\}$ *son* mútuamente independientes *cuando*

$$P(A_{k_1} \cap \ldots \cap A_{k_m}) = \prod_{i=1}^{m} P(A_{k_i}) \tag{1.13}$$

siendo $\{k_1, \ldots, k_m\} \subset \{1, \ldots, n\}$ *y los* k_i *distintos.*

Conviene señalar que la independencia mutua de los n sucesos supone que han de verificarse $\binom{n}{n} + \binom{n}{n-1} + \ldots \binom{n}{2} = 2^n - n - 1$ ecuaciones del tipo dado en (1.13).

Si solamente se verificasen aquellas igualdades que implican a dos elementos diríamos que los sucesos son *independientes dos a dos*, que es un tipo de independencia menos restrictivo que el anterior. De hecho, en el ejemplo 1.11 tenemos independencia dos a dos pero no independencia mutua.

El tipo de independencia habitualmente exigida es la mutua, a la que nos referiremos simplemente como *independencia*.

Definición 1.6 *Una* colección de sucesos es infinita *diremos que son* **mutuamente independientes** *cuando cualquier subcolección finita lo sea.*

* **Ej. 14** — Probar que si tenemos tres sucesos *mutuamente independientes* entonces $P(A|B \cap C) = P(A)$.

Independencia de clases de sucesos

Si A y B son sucesos independientes, ni A ni B nos proporcionan información sobre la ocurrencia del otro. Parece lógico que tampoco nos digan mucho sobre los complementarios respectivos. La pregunta es ¿son A y B^c independientes? La respuesta afirmativa la deducimos a continuación.

$$\begin{aligned} P(A \cap B^c) = P(A) - P(A \cap B) = \\ P(A) - P(A)P(B) = P(A)(1 - P(B)) = P(A)P(B^c). \end{aligned} \tag{1.14}$$

Del mismo modo se comprueba que A^c es independiente tanto de B como de B^c. Resulta así que las conjuntos de sucesos $\{A, A^c\}$ y $\{B, B^c\}$ son *independientes* en el sentido que al tomar un suceso de cada una de ellos, los sucesos son independientes. De forma más general podemos hablar de clases independientes de sucesos.

Definición 1.7 (Clases independientes de sucesos) *Las clases de sucesos* $\mathbb{A}_1, \ldots, \mathbb{A}_n \subset \mathbb{A}$ *se dicen* independientes*, si al tomar* A_i *en cada* $\mathbb{A}_i$*,* $i = 1, \ldots, n$*, los sucesos de la familia* $\{A_1, \ldots, A_n\}$ *son independientes.*

Notemos que en la definición no se exige que los elementos de cada clase $\mathbb{A}_i$ sean independientes entre sí. De hecho A y A^c sólo lo son si $P(A) = 0$ ó $P(A) = 1$.

Para una *colección infinita de clases de sucesos* la anterior definición se extiende con facilidad. Diremos que $\{\mathbb{A}_n\}_{n\geq 1} \subset \mathbb{A}$ son *independientes* si cualquier subcolección finita lo es.

1.7 Probabilidades geométricas

En lo anterior hemos visto el experimento consistente en obtener un punto aleatorio uniforme en un círculo. Este tipo de experimentos en donde cada resultado corresponde con un punto y el espacio muestral es un región del plano son frecuentes e importantes. Estas probabilidades reciben el nombre de probabilidades geométricas. Si $\lambda_2(A)$ denota el área de A entonces la probabilidad de A, de que salga A, de que el punto aleatorio esté en A viene dada por

$$P(A) = \frac{\lambda_2(A)}{\lambda_2(\Omega)}. \tag{1.15}$$

Por las propiedades del área en $\mathbb{R}^2$ se sigue que lo que acabamos de definir es una medida de probabilidad.

La definición que acabamos para subconjuntos de $\mathbb{R}^2$ se puede extender de un modo simple a cualquier $\Omega \subset \mathbb{R}^k$ que tenga volumen finito.

1.8 Odds

Consideremos un suceso A con probabilidad $P(A)$ los odds del suceso es el cociente

$$odds(A) = \frac{P(A)}{1 - P(A)}. \tag{1.16}$$

1.9 Distribución empírica

Supongamos unos valores numéricos dados, $x_1, \ldots, x_n$,[13] y consideremos como espacio muestral $\Omega = \{x_1, \ldots, x_n\}$, como sucesos todos

[13] En Estadística será una muestra observada de datos. Por ejemplo, tensiones arteriales de un grupo de pacientes o número de empleados en una muestra de empresas.

los posibles subconjuntos de Ω y como medida de probabilidad, P, la uniforme es decir

$$P(\{x_i\}) = \frac{1}{n}, \tag{1.17}$$

y, para cualquier suceso A,

$$P(A) = \frac{|A|}{n}, \tag{1.18}$$

siendo $|A|$ el cardinal de A. A esta distribución se la conoce como **distribución empírica** sobre $\{x_1, \ldots, x_n\}$ y tiene una gran importancia en Estadística.

1.10 Ejercicios

Los asteriscos que preceden a cada ejercicio en este capítulo y en los posteriores indican su dificultad: *, baja; **, media y ***, alta. Los problemas de una sección muy probablemente utilizando los conceptos y resultados de las secciones previas.

Fórmula de Laplace

En esta sección nos encontramos en la situación de resultados equiprobables por lo que básicamente hemos de contar los casos favorables a los sucesos de interés y el total de resultados del experimento.

* **Ej. 15** — Se lanzan al aire dos monedas bien construidas. De las siguientes afirmaciones cuál, si alguna, te parece la solución correcta a la pregunta ¿cuál es la probabilidad de que aparezcan dos caras? Razona la respuesta.

1. Puesto que o bien aparecen dos caras o bien no aparecen dos caras la probabilidad es $\frac{1}{2}$.
2. El número de caras obtenido puede ser 0, 1 o 2. La probabilidad de 2 caras es $\frac{1}{3}$.
3. Aunque sean monedas iguales, vamos a considerar que podemos etiquetarlas como moneda 1 y moneda 2. Teniendo en cuenta ese orden, los posibles resultados son CC, C+, +C y ++. La probabilidad de dos caras, CC, es $\frac{1}{4}$.

* **Ej. 16** — Supongamos que tres corredores del equipo A y tres del equipo B participan en una carrera. Si los seis tienen las mismas aptitudes y no hay empates, ¿cuál es la probabilidad de que los tres corredores del equipo A lleguen en primero, segundo y tercer lugar y los tres corredores del equipo B lleguen en cuarto, quinto y sexto lugar?

*** **Ej. 17** — Una urna contiene 100 bolas, de las cuales r son rojas. Supongamos que las bolas son seleccionadas al azar de una en una y sin reemplazo. Determinar:

1. La probabilidad de que la primera bola extraída sea roja.
2. La probabilidad de que la quincuagésima bola extraída sea roja.
3. La probabilidad de que la última bola extraída sea roja.

* **Ej. 18** — La baraja española consta de 40 cartas distribuidas en cuatro palos: oros, copas, espadas y bastos. Cada uno de estos palos está compuesto por 10 cartas: as, dos, tres, cuatro, cinco, seis, siete, sota, caballo y rey. Jugamos a un juego en el que se reparten 6 cartas a cada jugador. Se pide hallar la probabilidad de cada uno de estos sucesos:

1. Tener exactamente dos reyes.
2. Tener el as de oros.
3. Tener 4 cartas iguales (nos referimos a tener 4 con mismo "número": pueden ser 4 seises, 4 doses, o también 4 reyes, 4 ases, etc).
4. Tener 6 cartas del mismo palo.

** **Ej. 19** — Juego de dados tradicional chino que se juega durante la celebración del año nuevo. En este juego se lanzan 6 dados. Según parece un lanzamiento con exactamente dos pares gana a un lanzamiento con exactamente un par. ¿Cuál es la probabilidad de cada uno de estos sucesos? En otras palabras, encuentra la probabilidad de obtener exactamente par en un lanzamiento de 6 dados y la probabilidad de obtener exactamente dos pares en un lanzamiento de 6 dados. [7]

[7] Sacar un par no significa que la puntuación del dado sea un número par sino sacar una pareja, por ejemplo un par de seises son 2 seises.

* **Ej. 20** — Un cajón contiene 10 pares de calcetines. Elegimos al azar 8 calcetines:

(a) ¿Cuál es la probabilidad de que no haya ningún par entre los 8 elegidos?

(b) ¿Cuál es la probabilidad de que haya exactamente un par entre los 8 elegidos?

** **Ej. 21** — **Poker.** La baraja francesa consta de 52 cartas distribuidas en cuatro palos o colores: tréboles, diamantes, corazones y picas. Cada uno de estos palos está compuesto por 13 cartas: uno o as, dos, tres, cuatro, cinco, seis, siete, ocho, nueve, diez y las tres figuras, que se llaman valet (V, equivalente al Jack inglés), Dame (D, equivalente a la Queen ingles) y Roi(R, equivalente al King inglés). en una mano de poker se reparten 5 cartas a cada jugador. Se pide hallar la probabilidad de cada uno de estos sucesos:

1. Tener escalera de color (5 cartas consecutivas del mismo palo)
2. Tener poker (4 cartas iguales x x x x y)
3. Tener un full (un trío y una pareja x x x y y)
4. Tener 5 cartas del mismo palo
5. Tener una escalera (5 cartas consecutivas)
6. Tener un trío (x x x y z)
7. Tener dobles parejas (x x y y z)
8. Tener una pareja (x x y z w)

** **Ej. 22** — **Problema de los cumpleaños** En una reunión hay n personas. ¿Cuál es la probabilidad de que dos de ellas tengan el mismo cumpleaños? Para resolver el problema vamos a suponer que una persona dada tiene la misma probabilidad de nacer cualquier día del año.

** **Ej. 23** — Tres grupos de amigos eligen al azar entre tres bares para ir a cenar, sin restricción en el número de grupos por bar.

1. Listar los posibles resultados y, considerando que son equiprobables, calcular la probabilidad de los sucesos: $A = \{\text{Primer bar vacío}\}$, $B = \{\text{Los dos primeros bares vacíos}\}$ y $C = \{\text{Cada bar no tiene más de un grupo}\}$.
2. Hallar las probabilidades de A, B y C si se distribuyen los tres grupos entre n bares, $n \geq 2$.

** **Ej. 24** — Tenemos k personas que se sientan aleatoriamente en n asientos con $n > k$.

1. Si los asientos están situados en fila, ¿cuál es la probabilidad de que ocupen k asientos contiguos en la fila?
2. Si los asientos están situados en círculo, ¿cuál es la probabilidad de que ocupen k asientos contiguos del círculo?
3. Si k personas se sientan aleatoriamente en una fila de 2k asientos, ¿cuál es la probabilidad de que no haya dos personas sentadas en asientos contiguos?

** **Ej. 25** — Hemos cuadriculado una cierta zona en seis columnas y cuatro filas. Denotamos por C(i,j) la celda que ocupa la fila i y la columna j. Considerad el siguiente juego. Tenemos una ficha colocada en el rectángulo marcado como C(4,1). La ficha puede moverse hacia la derecha o hacia arriba. Se pregunta:

1. ¿Cuántos posibles caminos hay para moverse desde C(4,1) hasta C(1,6)?
2. Si el jugador pasa de camino por la celda C(2,5) recibe un premio. Supongamos que cada camino tiene la misma probabilidad de ser elegido, ¿cuál es la probabilidad de que el jugador pase por C(2,5) cuando va de C(4,1) a C(1,6)?

** **Ej. 26** — Tenemos una urna con 20 bolas rojas y 10 bolas azules. Vamos extrayendo (sin reemplazamiento) bolas aleatoriamente de la urna. ¿Cuál es la probabilidad de que extraer todas las bolas rojas antes que de que todas las bolas azules hayan sido extraídas? Supongamos que añadimos a la urna 8 bolas negras de modo que ahora tenemos 20 rojas, 10 azules y 8 negras. Repetimos las extracciones sucesivas sin reemplazamiento y se pide responder la misma pregunta. ¿Cuál es la probabilidad de extraer todas las bolas rojas antes que de que todas las bolas azules hayan sido extraídas?

** **Ej. 27** — En un ascensor suben 6 personas que pueden bajar en 8 pisos. Suponiendo que cada persona elige el piso independientemente de las demás. ¿Qué probabilidad tenemos de que al menos dos bajen en el mismo piso? Responder bajo el supuesto que las personas se distinguen.

** **Ej. 28** — Una urna contiene 5 bolas numeradas de 1 a 5. Extraemos dos al azar. ¿Cuál es la probabilidad de que difieran en al menos 2 unidades si las extracciones se llevan con o sin reemplazamiento? Dar ambas soluciones si fueran n las bolas, numeradas de 1 a n.

Utilizando las propiedades de la probabilidad

** **Ej. 29** — **El problema de las coincidencias**Supongamos que 4 invitados llegan a una casa y dejan el sombrero en el vestíbulo. Si a la salida los recuperan de modo aleatorio, calcular la probabilidad de que ninguno de ellos reciba su propio sombrero. Resolver el problema suponiendo que, en lugar de 4 invitados, tenemos un número arbitrario n de invitados. [14]

*** **Ej. 30** — Se tiene una baraja francesa sin comodines con 52 cartas enumeradas de la siguiente forma: A, 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K. Después de barajar, colocamos el mazo boca abajo. Vamos mostrando y retirando las cartas del mazo mientras cantamos su enumeración de forma ordenada empezando con el A y acabando con la K. Una vez cantada la K, cantamos otra vez el A y volvemos a empezar. El objetivo del juego es mostrar toda la baraja sin que coincida el numero cantado con el numero mostrado. Si en algún momento dado coinciden, el juego se acaba y se considera fracaso. ¿Que probabilidad hay de conseguirlo? [8]

[8] Propuesto por Jorge Camarero y Eduardo García (22/09/2015).

*** **Ej. 31** — Sea $(\Omega, \mathbb{A}, \mathbb{P})$ un espacio de probabilidad. Si A y B son dos sucesos y $A \triangle B$ es la diferencia simétrica entre ambos, definimos las funciones $d_1(A,B)$ y $d_2(A,B)$ mediante

$$d_1(A,B) = P(A \triangle B),$$

y

$$d_2(A,B) = \begin{cases} \frac{P(A \triangle B)}{P(A \cup B)} & \text{si } P(A \cup B) \neq 0, \\ 0 & \text{si } P(A \cup B) = 0. \end{cases}$$

Probar que d_1 y d_2 son distancias.

**** **Ej. 32** — **Desigualdades de Bonferroni** (*[9, pág. 32]*)En este problema nos ocupamos de las desigualdades de Bonferroni. En la desigualdad de Boole hemos visto como la probabilidad de la unión de una colección finita arbitraria de sucesos es menor o igual a la suma de las probabilidades. Denotemos

$$S_1 = \sum_{i=1}^{n} P(A_i), S_2 = \sum_{1 \leq i < j \leq n} P(A_i \cap A_j), \ldots, S_k = \sum_{1 \leq i_1 < \ldots < i_k \leq n} P(A_{i_1} \cap A_{i_k}), \ldots \tag{1.19}$$

La fórmula de la inclusión-exclusión la podemos escribir como

$$P(\cup_{i=1}^{n} A_i) = \sum_{k=1}^{n} (-1)^{k+1} S_k.$$

Se trata de probar las desigualdades de Bonferroni que afirma que si k es impar entonces

$$P(\cup_{i=1}^{n} A_i) \leq \sum_{j=1}^{k} (-1)^{j+1} S_j,$$

[14] Este es un problema clásico que aparece ya citado en el libro de Montmort (1713) *Essay d'analyse sur les jeux de hazard* en relación con un juego de cartas conocido como *le Jeu du Treize* (El Juego del Trece), que consistía en descubrir al azar 13 cartas numeradas del 1 al 13 ganando el jugador cuando el orden de aparición de la carta y su número coincidían. El problema se presenta en múltiples versiones, sombreros elegidos al azar, sobres con direcciones en los que se introducen al azar cartas encabezadas con esas direcciones, permutaciones aleatorias, . . .

mientras que si k es par entonces

$$P(\cup_{i=1}^n A_i) \geq \sum_{j=1}^{k} (-1)^{j+1} S_j.$$

** **Ej. 33** — Demostrar que la axiomática de Kolmogorov (dada en la definición 1.2 es equivalente a la siguiente axiomática:

1. $P(A) \geq 0$ para todo $A \in \mathbb{A}$.
2. $P(\Omega) = 1$.
3. Si $\{A_i\}_{i=1}^n$ son disjuntos dos a dos entonces $P(\cup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$.
4. $\{A_n\}_{n\geq 1}$ es una sucesión de sucesos decreciente a un suceso A ($A_n \downarrow A$) entonces $\lim_{n\to+\infty} P(A_n) = P(A)$.

** **Ej. 34** — Supongamos un espacio de probabilidad $(\Omega, \mathbb{A}, P)$. Demostrar que las siguientes dos propiedades son equivalentes:

1. $A_n \downarrow A$ entonces $\lim_{n\to+\infty} P(A_n) = P(A)$ para cualquier $A \in \mathbb{A}$.
2. $A_n \downarrow \emptyset$ entonces $\lim_{n\to+\infty} P(A_n) = 0$.

** **Ej. 35** — Demostrar que la axiomática de Kolmogorov (dada en la definición 1.2 es equivalente a la siguiente axiomática:

1. $P(A) \geq 0$ para todo $A \in \mathbb{A}$.
2. $P(\Omega) = 1$.
3. Si $\{A_i\}_{i=1}^n$ son disjuntos dos a dos entonces $P(\cup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$.
4. $A_n \downarrow \emptyset$ entonces $\lim_{n\to+\infty} P(A_n) = 0$.

Probabilidad condicional y teorema de Bayes

* **Ej. 36** — Proporcionamos a A un trozo de papel para que escriba un signo $+$ o un signo $-$, sabiendo que escribe el primero con probabilidad 1/3. El papel pasa a B, quien lo deja como está o cambia el signo antes de pasarlo a C. A continuación C, que puede o no haber cambiado el signo, lo pasa a D, quien finalmente nos lo devuelve tras haber introducido o no algún nuevo cambio. Si comprobamos que el papel tiene escrito un signo $+$ y sabemos que la probabilidad de que B, C y D cambiaran el signo es 2/3, obtener la probabilidad de que A escribiera originalmente un signo $+$.

* **Ej. 37** — Un test para diagnosticar cierta enfermedad tiene una *sensibilidad* del 95% y una *especificidad* del 99%. Si la enfermedad en cuestión tiene una *prevalencia* del 0.5%, ¿cuál es el *valor predictivo* del test?

* **Ej. 38** — Una urna contiene 2 bolas blancas, 5 negras y 3 rojas. Extraemos consecutivamente dos bolas sin reemplazamiento. El color de la primera no se comprueba, la segunda es negra.

1. Calcular la probabilidad de que la primera haya sido blanca.
2. Calcular esta misma probabilidad en el caso de que hiciéramos una tercera extracción y la bola fuera blanca.

* **Ej. 39** — Se busca un paraguas en un inmueble tiene siete pisos. La probabilidad de que el paraguas esté en cada uno de los pisos del inmueble es $p/7$ $(0 < p < 1)$. Se han explorado en vano los seis primeros pisos. ¿Cuál es la probabilidad de que el paraguas se encuentre en el séptimo piso?

* **Ej. 40** — ([10])Tenemos en una urna 5 bolas blancas y 10 bolas negras. Lanzamos un dado correcto y extraemos sin reemplazamiento de la urna ese número de bolas. ¿Qué probabilidad tenemos de que el resultado del lanzamiento sea el número 3 si todas las bolas son blancas? Responder la pregunta anterior también el caso en que las extracciones son con reemplazamiento.

* **Ej. 41** — Tres prisioneros A, B y C son informados por su carcelero de que se ha elegido al azar a uno de ellos para ser ejecutado y que los otros dos van a ser liberados. El prisionero A le pide al carcelero que le diga en privado cuál de sus compañeros va a ser liberado, asegurándole que no pasa nada porque le dé esa información puesto que él sabe que al menos uno de los otros dos quedará libre. El carcelero no quiere contestar la pregunta porque dice que si A supiera cuál de sus dos compañeros va a ser liberado entones su propia probabilidad de ser ejecutado subiría de $\frac{1}{3}$ a $\frac{1}{2}$ porque entonces él sería uno de los dos que podría ser ejecutado. ¿Qué piensas del razonamiento del carcelero?

** **Ej. 42** — Una bola marcada puede estar en una cualquiera de las dos urnas que tenemos disponibles, con probabilidades p y $1-p$, respectivamente. La probabilidad de extraer la bola de la urna en la que está alojada es r $(r \neq 1)$. ¿Cuál es la mejor forma de utilizar n extracciones con reemplazamiento, de cualquiera de las dos urnas, para que la probabilidad de extraer la bola sea máxima?

* **Ej. 43** — Disponemos de 3 cajas de 20 piezas cada una. El número de piezas que reúnen las condiciones de calidad exigidas son, respectivamente, 20, 15 y 10. De una de las cajas elegida al azar se extrae una pieza que resulta ser buena. Se devuelve a la caja y se extrae una segunda pieza que también resulta buena. ¿Cuál es la probabilidad de que la caja elegida haya sido la tercera?

** **Ej. 44** — Supongamos que una caja contiene 5 monedas, cada una de ellas con distinta probabilidad de obtener una cara la lanzarla. Denotamos por p_i dicha probabilidad para la moneda i, $i = 1, 2, \ldots, 5$ y supongamos que $p_1 = 0$, $p_2 = 1/4$, $p_3 = 1/2$, $p_4 = 3/4$ y $p_5 = 1$.

1. Se selecciona al azar una moneda. Si la primera cara se obtiene al cuarto lanzamiento, ¿cuál es la probabilidad de que la $i-$ésima moneda haya sido la seleccionada?
2. Si se lanza de nuevo la misma moneda, ¿cuál es la probabilidad de obtener una nueva cara?
3. Si se hubiera obtenido una cruz en el primer lanzamiento, ¿cuál es la probabilidad de obtener una cara en el segundo lanzamiento?

** **Ej. 45** — Un taxi se ve involucrado en un accidente nocturno. En la ciudad hay dos compañias de taxis, los taxis *negros* y los taxis *blancos*. Se sabe que el 85% de los taxis de la ciudad son negros y el

15% restante son blancos. Un testigo del accidente afirma que el taxi involucrado era blanco y la fiabilidad de su testimonio es del 80%, es decir, es capaz de identificar correctamente el color del taxi el 80% de las veces.

1. Sin ningún cálculo previo, ¿piensas que es más probable que el taxi accidentado fuera el negro o el blanco?
2. Calcula la probabilidad de que el taxi accidentado fuera el blanco y compara ambas respuestas.
3. Supongamos que para $0 \leq p \leq 1$ el $100p\%$ de los taxis son blancos y que la fiabilidad del testigo continúa siendo del 80%. Estudia la sensibilidad a los datos de la respuesta anterior viendo como varía ésta en función de p. ¿A partir de qué valor de p la probabilidad de que haya sido el taxi blanco el accidentado supera 0.5?
4. El análisis anterior puede completarse permitiendo que la fiabilidad del testigo sea variable, $100q\%$, con $0 \leq q \leq 1$. Determina la región dentro del cuadrado

$$\{(p, q) \ : \ 0 \leq p \leq 1, \ 0 \leq q \leq 1\}$$

   en la que la probabilidad de que haya sido el taxi blanco el accidentado supera 0.5.

En todo cuanto precede nos referimos a *la probabilidad de que haya sido el taxi blanco* se sobrentiende que *dado que el testigo afirma que era blanco.*

* **Ej. 46** — Sean A y B dos sucesos. Obtener la probabilidad de $A \cap B$ si ha ocurrido A o si ha ocurrido $A \cup B$. Comentar el resultado.

*** **Ej. 47** — **Problema de Monty Hall**[15]
En un concurso de TV hay tres puertas, una de ellas esconde un coche y las otras dos sendas cabras. El concursante elige una de las puertas y obtiene como premio aquello que la puerta oculta, pero la puerta permanece cerrada y el presentador, que conoce lo que hay detrás de cada puerta, abre una de las otras dos puertas y aparece una cabra (lógicamente el presentador nunca abre la puerta que oculta el coche). El presentador se dirige entonces al concursante y le permite cambiar su elección, ¿qué le conviene hacer al concursante?
Este concurso tuvo gran éxito a principios de los 90 en los USA y una conocida columnista de la revista *Parade Magazine*, Marylin vos Savant publicó que cambiando su elección el concursante doblaba su probabilidad de ganar el coche, pues ésta pasaba del 1/3 inicial a 2/3. Su afirmación era correcta. Compruébalo.
Consideremos una generalización del problema. Una persona, B, invita a otra, A, a jugar a un juego que consiste llevar a cabo extracciones de una urna que contiene a bolas blancas y b bolas negras ($a+b \geq 3$). Se le pide a A que elija entre una de las dos siguientes estrategias (A llevará a cabo sus extracciones con los ojos vendados).

1. Extrae una bola al azar y si es blanca gana, en caso contrario pierde.
2. Extrae una bola, se deshace de ella sin ni siquiera conocer su color y a continuación B retira una bola negra de la urna. A lleva a

[15] En http://en.wikipedia.org/wiki/Monty_Hall_problem se puede encontrar una interesante exposición de este problema.

cabo una segunda extracción y si la bola extraída es blanca gana, en caso contrario pierde.

¿Qué estrategia le conviene elegir?

** **Ej. 48** — La probabilidad de que un virus informático haya infectado nuestro ordenador es $0,1$. Si el ordenador está infectado, un sistema antivirus detecta la infección con probabilidad $x = 0,95$, mientras que en caso de no infección el sistema detecta falsas infecciones con probabilidad $y = 0,03$. Interesa que el sistema antivirus tenga un elevado *valor predictivo*={*probabilidad de que el ordenador esté infectado cuando el antivirus detecta una infección*}. Calcularlo a partir de los datos anteriores. Si queremos aumentarlo, ¿donde hemos de dirigir nuestros esfuerzos, a aumentar x o a rebajar y?

Independencia

* **Ej. 49** — La probabilidad de cometer un error al medir cierta magnitud es p. ¿Cuál es el menor número de medidas que debemos efectuar para que la probabilidad de cometer algún error sea mayor que α.

* **Ej. 50** — Sean A y B sucesos con probabilidad (estrictamente) positiva. Indica si cada una de las siguientes afirmaciones es *cierta*, *falsa* o *puede ser cierta*.

1. Si A y B son incompatibles entonces son independientes.
2. Si A y B son independientes entonces son incompatibles.
3. $P(A) = P(B) = 0.6$ y A y B son incompatibles.
4. $P(A) = P(B) = 0.6$ y A y B son independientes.

** **Ej. 51** — Dos sucesos A_1 y A_2 son condicionalmente independientes respecto de B si verifican la siguiente condición:

$$P(A_1 \cap A_2|B) = P(A_1|B)P(A_2|B). \tag{1.20}$$

Demostrar que la definición dada en 1.20 es equivalente a que se verifique la siguiente igualdad.

$$P(A_2|A_1 \cap B) = P(A_2|B). \tag{1.21}$$

* **Ej. 52** — **Ingrid Bergman**La actriz sueca nació el día 29 de agosto de 1915 y murió el 29 de agosto de 1982. Nació y murió el mismo día del año.

1. ¿Cuál es la probabilidad de que una persona nazca y muera el día 29 de agosto de un año no bisiesto?
2. ¿Cuál es la probabilidad de que una persona nazca y muera un día cualquiera de un año no bisiesto?

En las preguntas anteriores asumimos que la probabilidad de nacer o morir es la misma para cada día del año.

** **Ej. 53** — Tres individuos A, B y C juegan a la ruleta rusa con un revólver cuyo cargador consta de 6 cilindros y solo uno de ellos contiene una bala. En cada ocasión se gira el cargador para seleccionar aleatoriamente uno de sus cilindros, que puede o no contener la bala.

Los individuos se disparan según el orden A, B, C, A, B, C, ... Determinar para cada uno de los tres individuos la probabilidad de que sea él quien acierte con la bala.

* **Ej. 54** — Repetimos indefinidamente una prueba en la que la probabilidad de éxito es siempre la misma, p, siendo los resultados de las pruebas independientes unos de otros (se trata de una sucesión de pruebas de Bernoulli). Obtener la probabilidad de que **a** éxitos ocurran antes que **b** fracasos.

** **Ej. 55** — Dos jugadores A y B juegan a un juego en el que cada uno de ellos puede efectuar n lanzamientos de dos dados, siendo A quien comienza. Las reglas del juego son las siguientes:

- Si A obtiene una suma 6 con los dados antes de que B haya obtenido una suma 7, A gana el juego.
- Si es B quien obtiene el 7 antes de que A haya obtenido el 6, es B quien gana.
- El juego termina en empate cuando ambos han agotado sus n lanzamientos.

Encontrar las expresiones de $p_A(n)$, $p_B(n)$ y $p_E(n)$ que denotan, respectivamente, que el ganador es A, el ganador es B o el juego termina en empate. Calcular sus límites cuando $n \to \infty$.

* **Ej. 56** — Dos jugadores A y B juegan a tirar alternativamente una moneda (empieza A). Gana el primero que saca una cara, y en ese momento se acaba el juego. Calcula la probabilidad que tiene cada jugador de ganar.

** **Ej. 57** — Tres jugadores A, B y C juegan a tirar de forma consecutiva un dado hasta que a alguno de ellos le salga un 6 (primero tira el dado A, luego B, luego C y así sucesivamente volvería a tocarle el turno a A). Gana el primero que saca un 6, y entonces acabaría el juego. Calcula la probabilidad que tiene cada jugador de ganar

** **Ej. 58** — Dos jugadores A y B juegan a tirar simultáneamente un dado cada uno. Gana el primero que saca un 6. Calcula:

- La probabilidad de que A y B empaten.
- La probabilidad de que gane A.
- La probabilidad de que gane B.

* **Ej. 59** — Sean A y B dos sucesos incompatibles con probabilidad distinta de cero. ¿Cuál es la probabilidad de que A ocurra antes que B si el experimento se repite indefinidamente?

** **Ej. 60** — **El juego de craps**Un jugador lanza dos dados, si la suma del primer lanzamiento es 7 u 11 gana, si la suma es 2, 3 o 12 pierde y si la suma es cualquier otro número continua lanzando hasta que aparezca una suma 7 o la suma que inicialmente obtuvo. Si aparece la suma 7 antes que la suma inicial pierde, en caso contrario gana. Calcular la probabilidad de que gane el juego.

* **Ej. 61** — **El segundo problema de Huygens**El holandés Christian Huygens publicó en 1657 uno de primeros libros sobre Pro-

babilidad que se conocen, *De Ratiociniis in Ludo Aleae* (Del Razonamiento en los Juegos de Azar), en el que planteaba una serie de problemas. El que se conoce como *segundo problema de Huygens* lo enunciamos a continuación
Tres jugadores A, B y C participan en el siguiente juego. Una urna contiene a bolas blancas y b negras. Los jugadores, en el orden $ABCABC\ldots$, extraen una bola con reemplazamiento hasta que uno de ellos obtiene una bola blanca y gana. Encontrar la probabilidad de ganar para cada jugador.

* **Ej. 62** — **La paradoja del caballero De Meré** En un juego consistente en lanzar repetidamente un par de dados, encontrar el menor número n de lanzamientos para que la probabilidad de obtener al menos un *doble seis* sea mayor que 0.5. [16]

** **Ej. 63** — A y B juegan a un juego en el que A tiene una probabilidad p de ganar una partida. El vencedor es aquel que gana dos partidas consecutivas. Encontrar el valor de p si se sabe que cuando A pierde la primera partida, las probabilidades de ganar el juego para A y para B son iguales.

** **Ej. 64** — Utilizando argumentos probabilísticos, probar la igualdad $(A > a)$

$$1 + \frac{A-a}{A-1} + \frac{(A-a)(A-a-1)}{(A-1)(A-2)} + \cdots \frac{(A-a)\ldots 2\cdot 1}{(A-1)\ldots(a+1)a} = \frac{A}{a}$$

Sugerencia.- Una urna con A bolas de las cuales a son blancas, extracciones sucesivas sin reemplazamiento, primera bola blanca, ...

Probabilidades geométricas

* **Ej. 65** — Elegimos aleatoriamente un punto de un segmento de longitud L, ¿cuál es la probabilidad de que la longitud del segmento limitado entre el origen del primero y el punto elegido sea menor que 1/3?

* **Ej. 66** — Elegimos un punto al azar de un círculo de radio R centímetros: ¿cuál es la probabilidad de que diste del origen más de r centímetros?

* **Ej. 67** — Se elige un punto al azar del cuadrado de vértices (0, 0), (0, 1), (1, 0) y (1, 1), ¿cuál es la probabilidad de la suma de sus coordenadas sea menor que 3/4? ¿Cuál es la probabilidad de la suma de sus coordenadas sea menor que a?

* **Ej. 68** — Se eligen al azar dos números en el intervalo $(0, 2)$. Calcular la probabilidad de que el producto de ambos sea inferior a 3.

* **Ej. 69** — Se eligen al azar dos números reales en (0,1). Se pide:

[16] El origen de la paradoja está en la pregunta que Antoine Gombauld, caballero De Meré, planteó a Pascal Observaba De Meré una discrepancia entre la realidad, deducida de su larga experiencia como jugador, y una antigua regla muy extendida entre los jugadores que afirmaba que $n = 24$. Esta errónea regla tenía su origen en la creencia de un comportamiento lineal de las probabilidades. Se sabía que si los lanzamientos eran de un solo dado y se perseguía la obtención de *un seis*, $n = 4$, pues $p_{3.1} = 0.4213$ y $p_{4.1} = 0.5177$. Se razonaba a continuación mediante una sencilla regla de tres: 4 es a 6 como 24 a 36.

1. ¿Cuál es la probabilidad de que el menor sea mayor que 1/2?
2. ¿Cuál es la probabilidad de que el menor de los dos sea mayor que 1/4 y el mayor de los dos menor que 3/4?
3. Los dos números elegidos se utilizan para señalar dos puntos por donde cortar un bastón de 1 m de longitud ¿Cuál es la probabilidad de que los tres trozos del bastón permitan construir un triángulo?

** **Ej. 70** — Elegimos al azar dos puntos B y C en el intervalo $[0, l]$ de manera que $B < C$. ¿Cuál es la probabilidad de que el intervalo $[0, B]$ sea más grande que el intervalo $[B, C]$?

Algunos más

[17]

* **Ej. 71** — La Facultad de Matemáticas ha puesto en marcha un nuevo sistema de adjudicación de becas. Los estudiantes formarán cola y comenzarán a matricularse y personal de la Secretaría hará una lista con sus fechas de nacimiento. El primer estudiante cuya fecha de nacimiento coincida con la de otro previamente matriculado recibirá como premio el dinero de todas la becas, en detrimento del resto de estudiantes. Suponiendo que la duración de cualquier año es de 365 días y que la probabilidad de nacer en un día cualquiera es siempre la misma, ¿qué lugar de la cola conviene ocupar para tener la mayor probabilidad de ganar la beca?

* **Ej. 72** — [El problema del encuentro] Dos amigos quedan en verse a la puerta de Correos (exactamente la puerta que enfrenta con la plaza San Francisco no la puerta que da a la parte de atrás de la iglesia de la Caridad) alrededor de las 12:30. Julián puede llegar en cualquier instante comprendido entre las 12:15 y las 12:45 y Enrique puede llegar en cualquier instante comprendido entre las 12:00 y las 13:00 (la llegada de uno es independiente de la del otro). ¿Cuál es la probabilidad de que el primero en llegar no espere más de 5 minutos? ¿Cuál es la probabilidad de que Julián llegue primero?[18]

* **Ej. 73** — **La paradoja de Bertrand** Elegimos una cuerda al azar en el círculo unidad. ¿Cuál es la probabilidad de que su longitud supere la del lado del triángulo equilátero inscrito en el círculo?

*** **Ej. 74** — **La paradoja de la urna vacía**Disponemos de una urna infinitamente grande y de una colección infinita de bolas numeradas. Procedemos a depositar las bolas en la urna de tres formas distintas.

1. A las 5 de la tarde menos 1 minuto introducimos las 10 primeras extrayendo la que lleva el número 10 (supongamos que la introducción y la sucesiva extracción consumen un tiempo 0). A las 5 menos $\frac{1}{2}$ minuto depositamos las 10 bolas siguientes y extraemos la que lleva el número 20. A las 5 menos $\frac{1}{4}$ de minuto las 10 siguientes extrayendo a continuación la que lleva el número 30. Y así sucesivamente.

[17] Con algunos de los problemas de esta sección hay que tener cuidado. En ocasiones simplemente se han añadido al final.

[18] El problema del encuentro se puede encontrar resuelto en [4, pág. 33].

2.El segundo procedimiento es análogo al anterior, pero las bolas que se extraen en cada ocasión son las numeradas 1, 2, 3, ...

3.En el tercer procedimiento las bolas se introducen como en los dos anteriores pero en cada decena la extracción se efectua al azar.

¿Cuantas bolas habrá en la urna a las 5 de la tarde según el procedimiento empleado?

*** **Ej. 75** — **Un camino aleatorio culé**El día 27 de julio de 1997 se celebraron elecciones a la presidencia del Barça. Había sólo dos candidatos, el señor Fernández y el señor Núñez, siendo este último el ganador. Un socio con veleidades probabilísticas se hizo la siguiente pregunta: ¿habrá ido el señor Núñez por delante del señor Fernández a lo largo de todo el escrutinio?

**** **Ej. 76** — En un puesto de palomitas hay n personas (con n par) esperando para comprar. El precio del paquete de palomitas es de 50 céntimos. Supongamos que la mitad de los clientes de la cola lleva una moneda de 50 céntimos y la otra mitad lleva una moneda de un euro. El propietario del puesto de palomitas no tiene ningún cambio cuando se forma la cola. ¿Qué probabilidad hay de que, en algún momento, el propietario no tenga cambio para devolver a un cliente de la cola y por lo tanto el cliente deba de esperar por el cambio?[19]

* **Ej. 77** — **El problema de los puntos o del reparto de la apuesta** Dos jugadores A y B juegan a un juego consistente en un número indeterminado de partidas. La probabilidad de ganar en cada partida es p para A y $1-p$ para B. Aquel de los dos que consigue antes vencer en r partidas gana el juego y la apuesta que ambos hicieron. Si el juego se interrumpe antes de finalizar, ¿cómo se debe repartir la apuesta?

** **Ej. 78** — Un aparato de diagnóstico automático emite un diagnóstico basado en el resultado de n análisis de un mismo paciente. Cada análisis, independientemente de los restantes, puede dar un resultado erróneo con probabilidad p. La probabilidad de un buen diagnóstico, condicionada al número de análisis correctos, es una función creciente de dicho número, $g(m)$. Durante una mañana la máquina ha diagnosticado a k pacientes. Encontrar la probabilidad del suceso $A =\{$*al menos un paciente está mal diagnosticado*$\}$. Particularizar el resultado para $g(m) = m/n$.

* **Ej. 79** — (*[10, pág. 116]*)Demostrar que si $P(A|B) = 1$ entonces $P(B^c|A^c) = 1$.

** **Ej. 80** — (*[10, pág. 116]*)Tenemos n tipos distintos de cupones. Se seleccionan al azar y se obtienen independientemente unos de otros de modo que un cupón del tipo i es seleccionado con probabilidad p_i de modo que $\sum_{i=1}^{n} p_i = 1$. Se pide:

[19]Este problema lo podemos encontrar en [4, pág. 31]. Allí se enuncia con billetes de cinco y 10 rublos. Cuando le formularon el problema al que escribe el problema era entrar a Viveros (parque de la ciudad de Valencia con entrada gratuita por cierto) y era con 50 pesetas (diez duros) y 100 pesetas (veinte duros). En fin, siempre el mismo bonito problema.

1. Si seleccionamos n cupones, ¿cuál es la probabilidad de que seleccionemos un cupón de cada tipo?

2. Supongamos $p_1 = p_2 = \ldots = p_n = 1/n$. Denotamos por E_i el suceso consistente en que no hemos elegido ningún cupón del tipo i entre los n seleccionados. Aplicar la fórmula de inclusión-exclusión para calcular $P(\cup_{i=1}^n E_i)$ y probar la siguiente igualdad

$$n! = \sum_{k=0}^{n} (-1)^k \binom{n}{k} (n-k)^n. \tag{1.22}$$

* **Ej. 81** — Consideramos dos cajas, una contiene una canica negra y una blanca; la otra contiene dos negras y una blanca. Una caja es seleccionada al azar, ¿Cuál es la probabilidad de de que la canica sea negra?, ¿cuál es la probabilidad de que la primera caja fuera seleccionada dado que la canica extraida sea blanca?

* **Ej. 82** — [[5, sección 1.4]] Un hombre tiene cinco monedas. Dos de estas monedas tienen dos caras, una de ellas tiene dos cruces y las dos últimas monedas son normales. Cierra los ojos, elige una moneda al azar y la lanza. ¿Qué probabilidad tenemos de que la cara inferior sea una cara?
Abre los ojos y ve que la parte superior de la moneda es cara: ¿qué probabilidad tenemos de que la cara inferior sea también una cara?
Cierra los ojos otra vez y lanza la moneda. ¿Qué probabilidad tiene de que la cara inferior sea cara?
Abre los ojos y ve que la moneda muestra una cara: ¿qué probabilidad tenemos de que la cara inferior sea cara?
Descartamos la moneda que estamos usando y volvermos a elegir otra al azar. La lanzamos. ¿Con qué probabilidad muestra cara en la parte superior?

* **Ej. 83** — Una urna contiene b bolas negras y r bolas rojas. Extraemos una bola al azar, y cuando la devolvemos a la urna, ponemos c bolas del mismo color que la bola que habíamos extraído. Ahora, extraemos otra bola. Demostrar que la probabilidad de que la primera bola fuera negra, dado que la segunda es roja es $\frac{b}{b+r+c}$.

* **Ej. 84** — Supongamos que cada bebe nacido de una pareja tiene la misma probabilidad de ser chico que de ser chica independientemente de la distribución por sexo de los otros niños en la familia. Para una pareja que tiene cinco hijos, calcula las probabilidades de los siguientes sucesos:

1. Todos los niños tengan el mismo sexo.

2. Los tres mayores son chicos y los demás chicas.

3. Los dos mayores son chicas.

4. Hay al menos una chica.

** **Ej. 85** — ¿Cómo podemos poner en dos urnas 20 bolas, de las cuales 10 son blancas y 10 negras, de tal forma que la probabilidad de extraer al azar una bola blanca de una urna seleccionada al azar sea máxima?

* **Ej. 86** — Sean $\{A_1, A_2$ y $A_3\}$ tres sucesos arbitrarios con probabilidades $\mathrm{P}(A_1)$, $\mathrm{P}(A_2)$ y $\mathrm{P}(A_3)$ respectivamente. Calcular la probabilidad de que ocurra exactamente uno de ellos.

* **Ej. 87** — Sean $\{A_1, A_2, \ldots, A_n\}$ sucesos arbitrarios con probabilidades $P(A_1, \ldots, P(A_2)$ respectivamente. Calcular la probabilidad de que ocurra exactamente uno de ellos.

* **Ej. 88** — (*[7, Sec. 2.1, ej. 1]*)Realizamos tres pruebas independientes, la probabilidad de obtener el suceso A es p en cada prueba.

1. ¿Cuál es la probabilidad de obtener al menos una vez el suceso A?
2. ¿Para qué valor de p esta probabilidad es exactamente igual a 0.5?

* **Ej. 89** — (*[7]*)Piensa en las siguientes probabilidades condicionadas y decide los valores que debieran de tener. Después calcula su valor utilizando la definición de probabilidad condicionada.

(a) $P(B \mid B)$.

(b) $P(B \mid B^c)$.

(c) $P(B \mid \Omega)$.

(d) $P(B \mid \emptyset)$.

(e) $P(B \mid A)$ cuando A es un subconjunto de B.

(f) $P(B \mid A)$ cuando B es un subconjunto de A.

* **Ej. 90** — (*[7]*)Se extraen tres cartas de una baraja francesa (52 cartas) de una en una y sin reemplazamiento. Determinar la probabilidad de

1. Todas sean rojas, esto es, que sean corazones o diamantes.
2. Las dos primeras sean negras y la tercera roja.
3. La primera y la tercera sean negras y la segunda roja.
4. La segunda y la tercera sean negras y la primera roja.
5. Exactamente una sea roja.
6. Exactamente una sea negra.
7. La primera es un as y la segunda también.
8. La primera no es un as y la segunda sí.
9. La segunda es un as.

* **Ej. 91** — Se estima que el 27% de los estudiantes de la Universidad de Valencia están en primer curso. Si el 25% de los estudiantes de primero, no nacidos en la ciudad y que viven en pisos de estudiantes están en Derecho, y la probabilidad de que un estudiante de la universidad elegido al azar sea de primer curso de Derecho, no haya nacido en la ciudad y viva en un piso de estudiantes es 0.02, encontrar la probabilidad de que un estudiante de primero, no nacido en la ciudad viva en un piso de estudiante.

* **Ej. 92** — En un examen de respuesta múltiple con k respuestas posibles el estudiante conoce la respuesta con probabilidad p y marca una respuesta al azar con probabilidad $1-p$. Encontrar la probabilidad de que haya contestado al azar si la respuesta es correcta.

** **Ej. 93** — Tenemos un dado correcto y seis personas numeradas de 1 a 6. Una de ellas coge el dado y lo lanza. Si el resultado es su propio número entonces gana, es decir, si la persona con el número 3 lanza el dado y sale el número 3 entonces gana el juego. Si sale otro número, por ejemplo el 4, entonces el dado lo lanza la persona 4. Si le sale el cuatro gana y de lo contrario se lo pasa al que tiene el número que ha salido. Supongamos que empieza la persona 1 a lanzar.

1. ¿Cuál es su probabilidad de que gane el jugador 1 en exactamente n lanzamientos del dado?
2. ¿Y la probabilidad de que gane la persona 1?
3. ¿Y la probabilidad de que gane la persona 3?

* **Ej. 94** — Uno de los vulgarismos más habituales consiste en eliminar la *d* en la sílaba final de las palabras terminadas en *ado*. Este defecto está especialmente extendido en el sur de España. Alguien en un hotel ha escrito una nota con la palabra *pesado*, no sabemos si en su forma correcta o incorrecta. Lo que sí sabemos es que al elegir al azar una letra de la palabra escrita ha resultado ser una vocal. Si en el hotel un 40% de sus huéspedes provienen del Sur y el 60% restante de la zona Central, ¿cuál es la probabilidad de que la nota la haya escrito alguien del Sur?

** **Ej. 95** — En una urna tenemos n bolas rojas y m bolas azules. Vamos eligiendo una bola cada vez y la extraemos de la urna sin restituirla. Las extracciones son realizadas hasta que tenemos r bolas rojas. Determinar la probabilidad de tener que extraer un total de k bolas para conseguirlo.

1.11 Soluciones abreviadas

Solución (Ej. 4) —

$$A\cup B \subseteq \Omega \to P(A\cup B) \leq P(\Omega) = 1 \to 1 \geq P(A)+P(B)-P(A\cap B) \to P(A\cap B) \geq P(A)+P(B)-1. \tag{1.8}$$

Solución (Ej. 5) —

$$P(\cap_{i=1}^n B_i) = 1-P(\cup_{i=1}^n B_i^c) \geq 1-\sum_{i=1}^n P(B_i^c) = 1-(n-\sum_{i=1}^n P(B_i)) = \sum_{i=1}^n P(B_i)-(n-1). \tag{1.9}$$

Solución (Ej. 6) — Vamos a probar que $P((\cap_{n\geq 1} A_n)^c) = 0$. Tenemos que

$$0 \leq P((\cap_{n\geq 1} A_n)^c) = P(\cap_{n\geq 1} A_n^c) \leq \sum_{n\geq 1} P(A_n^c) = 0. \tag{1.10}$$

Solución (Ej. 8) — Se aplica la definición de probabilidad condicionada y es inmediato.

Solución (Ej. 11) — Si llamamos U_r al suceso que indica que hemos elegido la urna número r, y V_i al suceso que indica que la iésima bola extraída es verde,

$$P(V_2) = \sum_{r=1}^n P(V_2 \mid U_r)P(U_r) = \frac{1}{n}\sum_{r=1}^n \frac{n-r}{n-1} = \frac{1}{2}$$

y

$$P(V_2 \mid V_1) = \sum_{r=1}^n \frac{1}{n}\frac{(n-r)(n-r-1)}{2(n-1)(n-2)} = \frac{2}{3}$$

Solución (Ej. 15) — El tercer apartado es correcto.

Solución (Ej. 16) — 0.05

Solución (Ej. 17) — La solución es la misma $\frac{r}{100}$.

Solución (Ej. 18) — (a)0.092078

(b)0.15

(c)0.001641

(d)0.000219

Solución (Ej. 19) — $P(A)=\frac{50}{216}$, $P(B)=\frac{25}{72}$.

Solución (Ej. 20) — (a)0.09145035
(b)0.4267683

Solución (Ej. 21) — (a)$P(\text{escalera color})=\frac{40}{\binom{52}{5}}$
(b)$P(\text{poker})=\frac{\binom{4}{4}13\binom{48}{1}}{\binom{52}{5}}$
(c)$P(\text{full})=\frac{13\cdot12\binom{4}{3}\binom{4}{2}}{\binom{52}{5}}$
(d)$P(\text{5 mismo palo})=\frac{4\binom{13}{5}}{\binom{52}{5}}$
(e)$P(\text{escalera})=\frac{10\cdot4^5-10\cdot4}{\binom{52}{5}}$
(f)$P(\text{trío})=\frac{13\binom{4}{3}\binom{12}{2}4^2}{\binom{52}{5}}$
(g)$P(\text{dobles parejas})=\frac{\binom{13}{2}11\binom{4}{2}^2 4}{\binom{52}{5}}$
(h)$P(\text{pareja})=\frac{13\binom{4}{2}\binom{12}{3}4^3}{\binom{52}{5}}$

Solución (Ej. 22) — $p=1-\frac{365\times364\times\cdots\times(365-n+1)}{365^n}$

Solución (Ej. 23) — (a)$P(A)=\frac{8}{27}$, $P(B)=\frac{1}{27}$, $P(C)=\frac{2}{9}$.
(b)$P(A)=\frac{(n-1)^3}{n^3}$, $P(B)=\frac{(n-2)^3}{n^3}$, $P(C)=\frac{n(n-1)(n-2)}{n^3}$.

Solución (Ej. 24) — (a)$(n-k+1)/\left(\begin{array}{c} n \\ k \end{array}\right)$
(b)$n/\left(\begin{array}{c} n \\ k \end{array}\right)$
(c)$2/\left(\begin{array}{c} 2k \\ k \end{array}\right)$

Solución (Ej. 25) — (a)$\binom{8}{3}=56$.
(b)$\frac{30}{56}$.

Solución (Ej. 26) — Es $\frac{1}{3}$ para ambas preguntas.

Solución (Ej. 27) — $1-\frac{8\times7\times,\ldots,\times3}{8^6}$.

Solución (Ej. 28) —

$$P(A_{con}) = \frac{(n-1)(n-2)}{n^2}, \qquad P(A_{sin}) = \frac{\frac{(n-1)(n-2)}{2}}{\binom{n}{2}} = \frac{n-2}{n},$$

que para $n = 5$ valen

$$P(A_{con}) = \frac{12}{25}, \qquad P(A_{sin}) = \frac{3}{5}.$$

Solución (Ej. 36) — $\frac{13}{41}$.

Solución (Ej. 37) — $\frac{95}{294}$.

Solución (Ej. 38) — 1. $\frac{2}{9}$.

2. $\frac{1}{8}$.

Solución (Ej. 39) — $\frac{p}{7-6p}$.

Solución (Ej. 40) —

$$\left[\binom{5}{3}\Big/\binom{15}{3}\right]\Big/\left[\sum_{i=1}^{5}\binom{5}{i}\Big/\binom{15}{i}\right]$$

Solución (Ej. 41) — La probabilidad de ser indultado el prisionero es la misma con y sin la información del carcelero e igual a $\frac{2}{3}$.

Solución (Ej. 42) —

$$m = \frac{n}{2} + \frac{\log[(1-p)/p]}{2\log(1-r)}.$$

Obsérvese que si $p = 1/2$, $m = n/2$ y las extracciones se reparten por igual en ambas urnas.

Solución (Ej. 43) — $\frac{4}{29}$.

Solución (Ej. 44) — 1. Designamos por A el suceso *"la primera cara aparece en el cuarto lanzamiento ´´* y por M_i "la moneda elegida es la i´´, nos piden calcular $P(M_i|A)$. Dadas las p_i del enunciado ya sabemos que $P(M_i|A) = 0$ para $i = 1$ e $i = 5$. Para el resto: $P(M_2|A) = \frac{27}{46}$, $P(M_2|A) = \frac{16}{46}$, $P(M_3|A) = \frac{3}{46}$.

2. $P(Cara) = \frac{39}{92}$.

3. $P(C_2|C_1^c) = \frac{1}{4}$.

Solución (Ej. 45) — 1.Que el taxi implicado es blanco.

2.0.41.

3.$0.2 \leq p$.

4.p y q deben estar en el recinto definido por $0 \leq p \leq 1$, $0 \leq q \leq 1$ y $p + q > 1$.

Solución (Ej. 46) — 1.$\frac{P(A \cap B)}{P(A)}$.

2.$\frac{P(A \cap B)}{P(A \cup B)}$.

Solución (Ej. 47) — Le conviene cambiar. Su probabilidad de ganar premio pasa de 1/3 a 2/3.

Solución (Ej. 49) —

$$n < \frac{\ln(1-\alpha)}{\ln(1-p)}.$$

Solución (Ej. 50) — Las tres primeras falsas y la cuarta puede ser cierta. Un ejemplo: dos extracciones con reemplazamiento de una urna con 6 rojas y 4 negras. Los sucesos serían salir roja o salir negra.

Solución (Ej. 52) — 1. $\frac{1}{365^2}$. 2. $\frac{1}{365}$.

Solución (Ej. 53) — $P(A) = \frac{6^2}{6^3-5^3}$, $P(B) = \frac{5 \times 6}{6^3-5^3}$, $P(C) = \frac{5^2}{6^3-5^3}$.

Solución (Ej. 54) —

$$p^a \sum_{j=0}^{b-1} \binom{a+j-1}{a-1} (1-p)^j.$$

Solución (Ej. 55) — $p_A(n) = 0.4918 \times (1 - 0.7176^n)$; $p_B(n) = 0.5082 \times (1 - 0.7176^{n+1})$; $p_E(n) = 0.8565 \times 0.7176^n$. Al pasar al límite $p_A(n) \to 0.4918$, $p_B(n) \to 0.5082$ y $p_E(n) \to 0$.

Solución (Ej. 56) —

$$P(\{Ganar A\}) = \frac{2}{3}; P(\{Ganar B\}) = \frac{1}{3}$$

Solución (Ej. 57) — $P(\{GanarA\}) = \frac{36}{91}$; $P(\{GanarB\}) = \frac{30}{91}$; $P(\{GanarC\}) = \frac{25}{91}$.

Solución (Ej. 58) — $P(\{Empatar\}) = \frac{1}{11}$; $P(\{GanarA\}) = \frac{5}{11}$; $P(\{GanarB\}) = \frac{5}{11}$.

Solución (Ej. 59) —

$$\frac{P(A)}{P(A)+P(B)}.$$

Solución (Ej. 61) — Si p_A, p_B y p_C son las probabilidades de que ganen A, B y C respectivamente entonces: $p_A = \frac{\alpha}{1-\beta^3}$; $p_B = \frac{\alpha\beta}{1-\beta^3}$ y $p_C = \frac{\alpha\beta^2}{1-\beta^3}$.

Solución (Ej. 62) — $n = 25$.

Solución (Ej. 67) — a. $P = \frac{(3/4)^2}{2}$.
b. Distinguimos tres casos: **Caso 1:** Si $a < 0$, $P = 0$; **Caso 2:** Si $0 \leq a \leq 1$, $P = \frac{a^2}{2}$; **Caso 3:** Si $1 < a \leq 2$, , $P = 1 - \frac{(2-a)^2}{2}$; **Caso 4:** Si $a > 2$, , $P = 1$.

Solución (Ej. 68) — Sea $C = \{(a,b) \in \mathbb{R}^2 : 0 < a,b < 2\}$ y sea $S = \{(a,b) \in \mathbb{R}^2 : 0 < a,b < 2 \wedge ab < 3\}$. Entonces, $Area(C) = 4$; $Area(S) = 3(1 + log(2) - log(3/2))$ y la probabilidad que buscamos es el cociente de ambas áreas.

Solución (Ej. 69) — En los tres apartados la probabilidad pedida vale $\frac{1}{4}$.

Solución (Ej. 70) — $\frac{1}{4}$.

Solución (Ej. 86) —

$$P(A_1)+P(A_2)+P(A_3)-2P(A_1\cap A_2)- \\ 2P(A_1\cap A_3)-2P(A_2\cap A_3)++3P(A_1\cap A_2\cap A_3).$$

Solución (Ej. 87) —

$$\sum_{i=1}^{n} P(A_i) - 2\sum_{i<j} P(A_i \cap A_j) + \cdots + (-1)^{n+1} n P(A_1 \cap \cdots \cap A_n)$$

Solución (Ej. 88) — (a)$1-(1-p)^3$

(b)0.2063.

Solución (Ej. 89) — (a)1

(b)0

(c)$P(B)$

(d)Sin sentido.

(e)1

(f)$\frac{P(B)}{P(A)}$

Solución (Ej. 90) — (a)0.1176

(b)0.1275

(c)0.1275

(d)0.1275

(e)0.3824

(f)0.3824

(g)0.0045

(h)0.0724

(i)0.0769

Solución (Ej. 91) — 0.4938

Solución (Ej. 92) — $\frac{1-p}{(1-p)+kp}$

Solución (Ej. 93) — Denotamos A_i con $i = 1, \ldots, 6$ el suceso consistente en que gana el jugador i. A_{1n} indica el suceso que gana la persona 1 en lanzamiento n.

1. Depende de n y viene dada por
   - $P(A_{11}) = 1/6$.
   - $P(A_{12}) = 0$.
   - $P(A_{1n}) = (5/6)^{n-2}(1/6)^2$ para $n \geq 3$.
2. $\frac{11}{36}$.
3. $\frac{5}{36}$.

Solución (Ej. 94) — $\frac{4}{9} = 0.4444$.

Solución (Ej. 95) —

$$p_k = \frac{\binom{n}{r-1}\binom{m}{(k-1)-(r-1)}}{\binom{m+n-1}{k-1}} \times \frac{n-(r-1)}{m+n-(k-1)}, \quad r \leq k \leq m+r.$$

1.12 Material complementario

El teorema de Bayes en términos de apuestas (*odds*): el valor de la evidencia

Cuando apostamos en una carrera de caballos es lógico que lo hagamos a aquel caballo que creemos ganador, es decir, aquél que tiene mayor probabilidad de ganar. Pero el mundo de las apuestas tiene un lenguaje propio y no se habla en él de probabilidad de ganar, utilizando en su lugar expresiones del tipo: las apuestas están "5 a 2 a favor" de un determinado caballo o "6 a 1 en contra" de que el Valencia gane la Liga.

¿Qué significa que las apuestas están "3 a 2 en contra" de que *Lucero del Alba* gane el Grand National? La expresión resume el hecho de que 2 de cada 5 apostantes lo hacen por dicho caballo como vencedor. Si habláramos de "5 a 2 a favor" estaríamos afirmando que 5 de cada 7 apostantes lo consideran ganador. Si queremos expresar estas afirmaciones en términos de probabilidad y denotamos por G el suceso *Lucero del Alba gana*, $P(G)$ no es más que la proporción de apostantes que piensan que ganará, es decir, $P(G) = 2/5$ o $P(G^c) = 3/5$ en el primer caso y $P(G) = 5/7$ o $P(G^c) = 2/7$ en el segundo.

Podemos establecer una sencilla relación entre ambas formas de expresar la misma idea. Si por O (del inglés *odds*) denotamos las apuestas en contra expresadas en forma de fracción, podemos escribir

$$O = \frac{P(G^c)}{P(G)},$$

que no es más que el cociente entre la probabilidad de no ganar y la de hacerlo. A su vez, como $P(G^c) = 1 - P(G)$, fácilmente se obtiene la expresión de la probabilidad de ganar en términos de las apuestas

$$P(G) = \frac{1}{O+1}. \tag{1.23}$$

Volvamos de nuevo al teorema de Bayes. Dados dos sucesos A y B escribíamos

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

Si reemplazamos A por su complementario, A^c, tenemos

$$P(A^c|B) = \frac{P(B|A^c)P(A^c)}{P(B)}.$$

Al dividir ambas expresiones obtenemos

$$\frac{P(A|B)}{P(A^c|B)} = \frac{P(B|A)}{P(B|A^c)} \times \frac{P(A)}{P(A^c)}, \tag{1.24}$$

expresión que se conoce como el *teorema de Bayes en forma de apuestas (odds)*. Comentemos el significado de los tres cocientes que aparecen en (1.24).

La izquierda de la igualdad representa las apuestas a favor de A, dado el suceso B. El segundo factor de la derecha son esas mismas apuestas obtenidas sin la información que supone conocer que ha ocurrido B. Por último, el primer factor de la parte derecha de la igualdad es el cociente entre las probabilidades de un mismo suceso, B, según que A haya ocurrido o no. Es lo que se denomina *razón de verosimilitud*.

Para ver el interés que (1.24) tiene, vamos a utilizarla en un contexto forense. Se trata de obtener el valor de una evidencia (Ev) en la discusión sobre la culpabilidad (C) o inocencia (C^c) de un sospechoso. La expresión (1.24) nos permite adaptar las apuestas *a priori* (antes de la presentación de la evidencia Ev) a favor de su culpabilidad y convertirlas en apuestas *a posteriori*, llevando a cabo dicha conversión mediante el factor

$$V = \frac{P(Ev|C)}{P(Ev|C^c)}, \tag{1.25}$$

al que se conoce como *valor de la evidencia*. Es importante destacar el hecho de que para su cálculo necesitamos dos probabilidades: las de Ev tanto si el sospechoso es culpable[20] como si es inocente.

El ejemplo que sigue ilustra el papel que este concepto puede jugar durante un juicio en la valoración de las pruebas y la consecuente ayuda que para juez o jurado supone.

Aplicación 1.2 (Harvey contra el Estado (Alaska, 1999)) *En 1993 Kimberly Esquivel, una adolescente de 14 años que vivía con su madre y su padrastro, quedó embarazada y se sometió a una operación de aborto. Poco después del aborto acusó a su padrastro, Patrick Harvey, de ser el padre*[21]*. Se llevó a cabo un análisis del DNA de los dos implicados y de una muestra del tejido del feto que el cirujano había conservado, obteniéndose el resultado que recoge la tabla.*

	locus DQ-alpha	**locus D1S80**
P. Harvey	*1.1,1.3*	*18,24*
K. Esquivel	*4.0,4.0*	*24,25*
Feto	*1.1,4.0*	*18,24,25*

De acuerdo con estos resultados el laboratorio emitió durante el juicio un informe en el que se afirmaba:

- *"... da un índice de paternidad de 6,90. Esto significa que las apuestas genéticas en favor de la paternidad son 6,90 veces más probables a favor de que Harvey sea el padre biológico de que lo sea un varón aleatoriamente elegido entre la población caucásica norteamericana".*
- *"... usando un valor neutral del 50% para las apuestas no genéticas en favor de la paternidad, obtenemos una probabilidad de paternidad del 87,34%".*

¿Cómo se obtuvieron estas cifras? Si denotamos mediante $H=\{$Harvey es el padre biológico$\}$ *y* $H^c=\{$Harvey NO es el padre biológico$\}$*, de acuerdo con las leyes de la genética y teniendo en cuenta que las frecuencias en la población de los alelos 1.1 y 18 son 13,7% y 26,5%, respectivamente, se obtiene*

$$P(\mathit{1.1\ y\ 18}|H) = 0.5 \times 0.5 = 0.25,$$

[20]Se entiende aquí *culpable* en el sentido de haber realizado verdaderamente la acción punible, no el hecho de serlo declarado por un juez o jurado

[21]El ejemplo está sacado de las notas del curso *Probability and Statistics for Law*, impartido por D. H. Kaye en la Universitat Pompeu Fabra de Barcelona en marzo de 2000

y

$$P(\textit{1.1 y 18}|H^c) = 0.137 \times 0.265 = 0.0365,$$

donde {1.1 y 18}={el feto posee los alelos 1.1 y 18}.

Lo que el informe denomina índice de paternidad no es más que el valor de la evidencia del fenotipo encontrado, es decir,

$$PI = \frac{P(\textit{1.1 y 18}|H)}{P(\textit{1.1 y 18}|H^c)} = \frac{0.25}{0.0365} = 6.90.$$

El valor neutral al que se refiere el informe supone asignar una probabilidad a priori 0,5 a H, lo que se traduce en que las apuestas a priori *a favor de la paternidad de Harvey son de 1 a 1. Aplicando (1.24) para obtener las apuestas* a posteriori

$$\frac{P(H|\textit{1.1 y 18})}{P(H^c|\textit{1.1 y 18})} = PI \times \frac{P(H)}{P(H^c)} = 6.90 \times 1 = 6.90.$$

La probabilidad de paternidad de Harvey, teniendo en cuenta la evidencia que los fenotipos aportan, puede calcularse mediante (1.23),

$$P(H|\textit{1.1 y 18}) = \frac{1}{\frac{1}{6.90}+1} = \frac{6.90}{7.90} = 0.873,$$

valor aportado en el informe en forma de porcentaje.

Comentario acerca del informe del laboratorio.- *El informe del laboratorio es incorrecto porque contiene dos errores que merecen ser comentados.*

- *El primero se refiere a la confusión entre el* índice de paternidad *y las* apuestas a favor de la paternidad*. Como ya hemos dicho el índice no es más que el valor de la evidencia, el cociente entre la probabilidad de que fuera Harvey quien aportara sus alelos y la probabilidad de que una extracción al azar de la población de genes aportara los alelos. Esta confusión es otra manera de presentarse la* falacia del fiscal.
- *La anterior objeción tiene una salvedad, ambos conceptos coinciden cuando las apuestas* a priori *a favor de la paternidad de Harvey son de 1 a 1, como ocurre en este caso. Pero para conseguirlo se ha asignado el valor 0,5 a $P(H)$, que el propio informe califica como* neutral *cuando* arbitrario *sería un calificativo más apropiado (asignar una probabilidad de 0,5 equivale, como ya dijimos anteriormente, a decidir la paternidad a cara o cruz). Un experto no necesita escoger un valor particular para las apuestas a priori. En su lugar debe dar una tabla de resultados como la que sigue, cuya valoración dejará en manos del juez o del jurado.*

P(H)	**P(H|1.1 y 18)**
0,10	*0,433*
0,30	*0,633*
0,50	*0,873*
0,70	*0,941*
0,90	*0,984*

1.12.1 Dos problemas importantes

Ejercicio 96 Problema de la aguja de Buffon

[22] ¿Cuál es la probabilidad de que una aguja que se lanza al azar sobre un entramado de líneas paralelas equidistantes, con distancia entre ellas mayor que la longitud de la aguja, corte a alguna de las paralelas?

Solución 1.1 (Problema 96) *Denotamos la longitud de la aguja con l y la separación entre líneas con d. La posición de la aguja en relación con las líneas paralelas viene perfectamente definida por la coordenada x de su punto medio y por el ángulo θ que forma con la horizontal. El espacio muestral sería $\Omega = [0, d] \times [-\pi/, \pi/2]$. La aguja tocará a la línea izquierda si*

$$x \leq \frac{l}{2}\cos(\theta),$$

y tocará a la línea derecha si

$$d - x \leq \frac{l}{2}\cos(\theta).$$

Denotemos este suceso por A. El área de una de ellas sería

$$\int_{-\pi/2}^{\pi/2} \frac{l}{2}\cos(\theta)d\theta = l. \tag{1.26}$$

El área total de las dos regiones, esto es del suceso A es pues $2l$. La probabilidad de tocar será pues (aplicando probabilidades geométricas) igual a

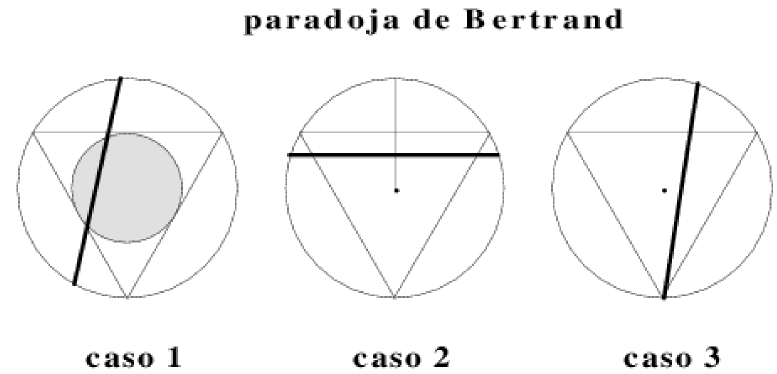
$$P(A) = \frac{2l}{\pi d}.$$

¿Y si en lugar de un retículo de líneas paralelas consideramos otros retículos posibles? Por ejemplo podemos considerar líneas perpendiculares formando cuadrados o retículos hexagonales. Entonces no se trata de tocar o no tocar a una de las líneas sino que nos aparecen las probabilidades de no tocar, de tocar a una línea o a dos o a tres dependiendo de la disposición del sistema test. Son muy interesantes aunque de un nivel que supera este texto las referencias [8] y [14].

[22] Un problema famoso que está en el origen de las probabilidades geométricas es el problema de la aguja de Buffon De hecho, el origen de la probabilidad geométrica se debe a George Louis Leclerc, Conde de Buffon, eminente naturalista francés, quién en el año 1777 publicó una obra titulada *Essai d'Arithmétique Morale* con el objeto de reivindicar a la Geometría como ciencia capaz de aportar soluciones a los problemas de azar, capacidad hasta entonces exclusivamente atribuida a la Aritmética. En ella plantea y resuelve el que ahora conocemos como problema de la aguja de Buffon. Con el tiempo se convirtió en una referencia clásica y pasó a ser conocido en la literatura probabilística como *el problema de la aguja de Buffon*. En la solución aportada por Buffon se ponía en evidencia que el problema requería *medir*, a diferencia de lo que ocurría con las soluciones a los problemas relacionados con juegos de azar discretos en los que bastaba *contar*.

Ejercicio 97 La paradoja de Bertrand

Elegimos una cuerda al azar en el círculo unidad. ¿Cuál es la probabilidad de que su longitud supere la del lado del triángulo equilátero inscrito en el círculo? [23]



Solución 1.2 (Problema 97) *La paradoja estriba en que la respuesta parece no ser única. En efecto, el valor de la probabilidad que se nos pide depende del significado que demos a* la elección al azar. *La longitud de una cuerda en un círculo puede calcularse a partir de,*

1. *la distancia al centro de su punto medio,*
2. *la posición de su punto medio sobre un radio cualquiera,*
3. *la posición de uno de sus extremos sobre la circunferencia, supuesto fijo el otro extremo.*

Cada una de estas interpretaciones supone una elección al azar sobre espacios muestrales distintos. Así,

Caso 1.- *El espacio muestral es todo el círculo unidad, C_1 y sólo las cuerdas cuyos puntos medios caen en el círculo inscrito en el triángulo equilátero, $C_{1/2}$, tienen longitud mayor que $\sqrt{3}$. Es sabido que este círculo tiene radio 1/2 y recurriendo a la probabilidad geométricas, si $A=\{$*la cuerda tiene longitud mayor que $\sqrt{3}\}$

$$P(A) = \frac{\textit{área}(C_{1/2})}{\textit{área}(C_1)} = \frac{\pi(1/2)^2}{\pi} = \frac{1}{4}.$$

Caso 2.- *El espacio muestral es ahora el segmento (radio) [0,1] y sólo las cuerdas cuyos puntos medios están en [0,1/2] cumplen la condición. Tendremos*

$$P(A) = \frac{\textit{longitud}([0,1/2])}{\textit{longitud}([0,1])} = \frac{1}{2}.$$

[23] Un ejemplo clásico de probabilidad geométrica es el que se conoce como *paradoja de Bertrand*. Como veremos a continuación la paradoja reside en el hecho de existir, aparentemente, varias soluciones al problema que se plantea. Fue propuesto por Joseph Louis François Bertrand (1822-1900), profesor de l'École Polytechnique de Paris y del Collège de France. Aunque es más conocido por la conjetura de teoría de números que lleva su nombre y que fue demostrada por Chebyshev en 1850 (para todo $n > 3$, existe siempre un número primo entre n y $2n-2$), Bertrand trabajó también en geometría diferencial y en teoría de la probabilidad.

Caso 3.- *El espacio muestral es ahora la circunferencia del círculo unidad. Si fijamos un extremo de la cuerda y elegimos al azar la posición del otro, sólo aquellas cuerdas que tengan este último extremo sobre el tercio de la circunferencia opuesto al extremo fijo cumplen la condición. Tendremos*

$$P(A) = \frac{2\pi/3}{2\pi} = \frac{1}{3}.$$

Capítulo 2

Variable aleatoria

2.1 Introducción

Supongamos el experimento consistente en elegir al azar (o si preferimos aleatoriamente) a una individuo de la Comunidad Valenciana. Obviamente el espacio muestral está formado por los distintos individuos. Si numeramos a todos los individuos podemos denotar $\Omega = \{\omega_i\}_{i=i}^N$ donde N es el número total de personas de la Comunidad. Elección al azar supone que cada individuo tiene la misma probabilidad de ser elegido. Por lo tanto el suceso formado por un solo resultado $\{\omega_i\}$ que nos indica que el individuo elegido ha sido el i-ésimo tendrá la siguiente probabilidad

$$P(\{\omega_i\}) = \frac{1}{N}.$$

El resultado mismo no suele tener un interés especial para nosotros. Seleccionamos aleatoriamente estas personas porque tenemos interés en sus ingresos, en el número de personas que conviven con él o ella, si está casada o es soltera, su glucosa o tensión arterial. Son ejemplos de valores que nos pueden interesar del individuo seleccionado dependiendo de que estemos realizando un estudio de tipo socio económico o bien un estudio de salud. Por ello nuestro interés no está en el individuo ω que obtenemos cuando seleccionamos a una persona al azar sino que nuestro interés es un valor asociado a ω que denotamos por $X(\omega)$. Lo que es aleatorio es ω porque lo seleccionamos aleatoriamente. Una vez tenemos ω el valor $X(\omega) = x$ viene dado. Otros ejemplos:

1. Elegimos al azar a una persona, una vez elegida su edad no es aleatoria, es la que tiene en el momento de observarla.
2. Elegimos al azar una muestra de agua en una planta de tratamiento de aguas residuales. Luego determinamos la demanda química de oxígeno.

A la función que asocia a un resultado un valor numérico se le llama **variable aleatoria**. Lo podemos denotar como

$$\begin{aligned} X : \Omega &\rightarrow \mathbb{R} \\ \omega &\rightarrow X(\omega) = x. \end{aligned}$$

Otros ejemplos habituales de variables aleatorias son:

- Número de llamadas que llegan a una centralita telefónica en un intervalo de tiempo.
- Suma de las caras que muestran dos dados al ser lanzados.

En resumen, nuestro interés al examinar el resultado de un experimento aleatorio no es tanto el espacio de probabilidad resultante, como la característica numérica asociada, lo que supone cambiar nuestro objetivo de Ω a $\mathbb{R}$. Hay dos razones que justifican este cambio:

1. El espacio de probabilidad es un espacio abstracto, mientras que $\mathbb{R}$ es un espacio bien conocido en el que resulta mucho más cómodo trabajar.
2. Fijar nuestra atención en la característica numérica asociada a cada resultado implica un proceso de abstracción que, al extraer los rasgos esenciales del espacio muestral, permite construir un modelo probabilístico aplicable a todos los espacios muestrales que comparten dichos rasgos.

Puesto que se trata de características numéricas ligadas a un experimento aleatorio, son ellas mismas *cantidades aleatorias*. Esto supone que para su estudio y conocimiento no bastará con saber que valores toman, habrá que conocer además la probabilidad con que lo hacen. De todo ello nos vamos a ocupar a continuación.

2.2 Variable aleatoria

La única forma que conocemos de trasladar información de un espacio a otro es mediante una aplicación. En nuestro caso la aplicación habrá de trasladar el concepto de suceso, lo que exige una mínima infraestructura en el espacio receptor de la información semejante a la σ-álgebra que contiene a los sucesos. Como nos vamos a ocupar ahora del caso unidimensional, una sola característica numérica asociada a los puntos del espacio muestral, nuestro espacio imagen es $\mathbb{R}$. En $\mathbb{R}$, los intervalos son el lugar habitual de trabajo y por tanto lo más conveniente será exigir a esta infraestructura que los contenga. Existe en $\mathbb{R}$ la llamada σ-álgebra de Borel, β, que tiene la propiedad de ser la menor de las que contienen a los intervalos, lo que la hace la más adecuada para convertir a $\mathbb{R}$ en espacio medible: $(\mathbb{R}, \beta)$. Definamos formalmente qué es una variable aleatoria.

Definición 2.1 (variable aleatoria) *Consideramos los espacios medibles $(\Omega, \mathbb{A})$ y $(\mathbb{R}, \beta)$.*

Una variable aleatoria es un aplicación, $X : \Omega \longrightarrow \mathbb{R}$, que verifica

$$X^{-1}(B) \in \mathbb{A}, \quad \forall B \in \beta. \tag{2.1}$$

[9] Cuando hacemos intervenir una variable aleatoria en nuestro proceso es porque ya estamos en presencia de un espacio de probabilidad, $(\Omega, \mathbb{A}, P)$. La variable aleatoria traslada la información probabilística relevante de Ω a $\mathbb{R}$ mediante una *probabilidad inducida* que se conoce como *ley de probabilidad de X* o *distribución de probabilidad de X*.

[9] En el contexto más general de la Teoría de la Medida, una aplicación que verifica (2.1) se dice que es una *aplicación medible*. De acuerdo con ello, una variable aleatoria no es más que una aplicación medible entre $(\Omega, \mathbb{A})$ y $(\mathbb{R}, \beta)$.

Fácilmente se comprueba el siguiente resultado.

Lema 2.1 *La familia de sucesos*

$$\sigma(X) = \{X^{-1}(B) : \ B \in \beta\}$$

es una σ-álgebra que verifica $\sigma(X) \subset \mathbb{A}$.

Definición 2.2 (σ-álgebra inducida) *Dada una variable aleatoria X se denomina* σ-álgebra inducida por X *a*

$$\sigma(X) = \{X^{-1}(B) : \ B \in \beta\}.$$

X induce sobre $(\mathbb{R}, \beta)$ una probabilidad, P_X, de la siguiente forma,

$$P_X(A) = P(X^{-1}(A)), \ \forall A \in \beta.$$

Es fácil comprobar que P_X es una probabilidad sobre la σ-álgebra de Borel, de manera que $(\mathbb{R}, \beta, P_X)$ es un espacio de probabilidad al que podemos aplicar todo cuanto se dijo en el capítulo anterior. La probabilidad P_X también recibe el nombre de *distribución* de la variable aleatoria X.

Observemos que P_X hereda las características que tenía P, pero lo hace a través de X. ¿Qué quiere esto decir? Un ejemplo nos ayudará.

Ejemplo 2.1 *Sobre el espacio de probabilidad resultante de lanzar dos dados, definimos las variables aleatorias, X como la suma de las caras e Y como el valor absoluto de la diferencia de las caras. El espacio de probabilidad sobre el que ambas están definidas es el mismo, P_X y P_Y son distintas porque viene inducidas por variables distintas. En efecto,*

$$P_X(\{0\}) = P(X^{-1}(\{0\}) = P(\emptyset) = 0,$$

sin embargo,

$$P_Y(\{0\}) = P(Y^{-1}(\{0\}) = \\ P(\{1,1\}, \{2,2\}, \{3,3\}, \{4,4\}, \{5,5\}, \{6,6\}) = \frac{1}{6}. \quad (2.2)$$

La distribución de probabilidad de X, P_X, nos proporciona cuanta información necesitamos para conocer el comportamiento probabilístico de X. No es fácil de manejar una distribución de probabilidad. Notemos que *debemos* de conocer la probabilidad de cada posible conjunto de Borel, esto es, debemos conocer $P_X(B)$ con $B \in \beta$. Se trata de un objeto matemático complejo. Esta es la razón por la que recurrimos a distintas funciones para describir (y caracterizar) la aleatoriedad de X.

Ejercicios

* **Ej. 1** — Sea X una variable aleatoria. Demostrar que $\sigma(X) = \{X^{-1}(B) : \ B \in \beta\}$ es una σ-álgebra.

* **Ej. 2** — Si tenemos un espacio de probabilidad, $(\Omega, \mathbb{A}, P)$ y una variable aleatoria X demostrar que

$$P_X(A) = P(X^{-1}(A)), \ \forall A \in \beta.$$

es una medida de probabilidad, es decir, verifica la axiomática de Kolmogorov dada en la definición 1.2.

2.3 Función de distribución de una variable aleatoria

Empezamos a estudiar funciones que nos permiten *manejar* una distribución de probabilidad de un modo simple.

Definición 2.3 (Función de distribución de una variable aleatoria) *Dada una variable aleatoria X definimos su función de distribución como*

$$F_X(x) = P_X((-\infty, x]) = P(X^{-1}\{(-\infty, x]\}) = P(X \leq x), \qquad (2.3)$$

con $x \in \mathbb{R}$.

Esta función tiene las siguientes propiedades importantes.

Proposición 2.1 *Dada una variable aleatoria X, su función de distribución asociada verifica que es: no negativa, monótona creciente, continua por la derecha, $\lim_{x\to+\infty} F_X(x) = 1$ y $\lim_{x\to-\infty} F_X(x) = 0$. Además es continua en un punto x si y solo si $P(X = x) = 0$.*

Demostración 2.1 No negatividad.- *Consecuencia inmediata de su definición.*

Monótona creciente.- *De la monotonía de la probabilidad se deduce fácilmente que $F_X(x_1) \leq F_X(x_2)$ si $x_1 \leq x_2$.*

Continuidad por la derecha.- *Consideremos una sucesión decreciente de números reales $x_n \downarrow x$. La correspondiente sucesión de intervalos verifica $\cap_n(-\infty, x_n] = (-\infty, x]$, y por la continuidad desde arriba de la probabilidad respecto del paso al límite tendremos $\lim_{x_n\downarrow x} F_X(x_n) = F_X(x)$.*

Observemos por otra parte que $(-\infty, x] = \{x\} \cup \lim_{n\to+\infty}(-\infty, x - \frac{1}{n}]$, lo que al tomar probabilidades conduce a

$$F_X(x) = P(X = x) + \lim_{n\to+\infty} F_X\left(x - \frac{1}{n}\right) = P(X = x) + F(x-), \qquad (2.4)$$

A partir de 2.4 se sigue que $F_X(x)$ es continua en x sí y solo sí $P(X = x) = 0$.

Valores límites.- *Si $x_n \uparrow +\infty$ o $x_n \downarrow -\infty$ entonces $(-\infty, x_n] \uparrow \mathbb{R}$ y $(-\infty, x_n] \downarrow \emptyset$ y por tanto*

$$F(+\infty) = \lim_{x_n\uparrow+\infty} F(x_n) = 1, \quad F(-\infty) = \lim_{x_n\downarrow-\infty} F(x_n) = 0.$$

Hemos propuesto una forma de describir la distribución de probabilidad de una variable aleatoria, su función de distribución asociada. Teníamos una variable aleatoria ya definida y describimos su distribución. Vamos a considerar la situación inversa. Nos dan una función y queremos saber si hay alguna variable aleatoria (de la que no sabremos cómo se ha definido) pero que tenga a esta función como su función de distribución asociada. Este problema va más allá de los objetivos del texto pero sí que es importante cuáles son las condiciones necesarias y suficientes para que una función dada sea la función de distribución asociada a alguna variable aleatoria. Las propiedades son las que hemos visto en la proposición 2.1. A una función que permite este camino inverso, recupero toda la distribución de probabilidad a partir de dicha función reciben el nombre de funciones de distribución de probabilidad. Se justifica la siguiente definición.

Definición 2.4 (Función de distribución de probabilidad) *Se dice la función F es una función de distribución de probabilidad si es no negativa, monótona creciente, continua por la derecha y verifica:* $\lim_{x \to +\infty} F(x) = 1$, $\lim_{x \to -\infty} F(x) = 0$.

Hemos visto previamente que dada una variable aleatoria, la función F_X definida en 2.3 es una *función de distribución de probabilidad* (en adelante simplemente función de distribución). En ocasiones se la denomina *función de distribución acumulada* porque tal y como ha sido definida nos informa de la probabilidad acumulada por la variable X hasta el punto x. Nos permite obtener probabilidades del tipo $P(a < X \leq b)$ a partir de la expresión

$$P(a < X \leq b) = F_X(b) - F_X(a).$$

La función de distribución debe su enorme importancia cuando trabajamos con una variable aleatoria al hecho de que caracteriza a la medida de probabilidad. Hemos definido F_X utilizando P_X de modo que la probabilidad determina la función de distribución. Y el recíproco: ¿es cierto? Dada una función de distribución F: ¿tenemos una única medida de probabilidad asociada? La respuesta es afirmativa pero la prueba no es simple y excede el nivel de estas notas.

2.4 Percentiles

Un concepto muy relacionado con la función de distribución de una variable aleatoria y de un gran interés práctico es el concepto de percentil.

Definición 2.5 (Percentil de orden p) *Tenemos la variable aleatoria X y consideramos un valor p ($0 \leq p \leq 1$). Se define percentil de orden p de la variable aleatoria X como un valor κ_p (no necesariamente único) que verifica las dos desigualdades siguientes*

$$P(X \leq \kappa_p) \geq p \quad y \quad P(X \geq \kappa_p) \geq 1 - p.$$

Sin duda, el percentil de mayor uso es la mediana o percentil de orden 0.5. Los percentiles de orden 0.25, 0.50 y 0.75 son los cuartiles. Se dice que son el primero, segundo y tercer cuartiles. También a los cuartiles de orden 0.25 y 0.75 se les suele denominar el cuartil inferior y el cuartil superior.

2.5 Variable aleatoria discreta

Es una variable aleatoria que toma un número finito de valores o bien toma un número infinito numerable de posibles valores. Un ejemplo muy simple y habitual es la edad de un individuo seleccionado al azar. La edad suele cuantificarse con valores enteros. Decimos que una persona tiene 24 años o que tiene 25 años. Por lo tanto es un valor entero. Si cuantificamos la edad en años entonces los valores posibles

Tabla 2.1: Función de probabilidad de una variable discreta. En la primera fila el valor y en la segunda la probabilidad de tomar este valor.

x	0.000	1.000	2.000	3.000	4.000	5.000
$P(X = x)$	0.107	0.268	0.302	0.201	0.088	0.026
x	6.000	7.000	8.000	9.000	10.000	
$P(X = x)$	0.006	0.001	0.000	0.000	0.000	

de la variable aleatoria son $0, 1, 2, \ldots$. En cualquier caso un número finito de posibles valores.

Otro ejemplo puede ser el número de organismos que observamos en una placa al microscopio, el número de huevos en una buitrera, el número de píxeles defectuosos en un monitor de ordenador, el número de árboles dentro de un quadrat en un muestreo espacial. Todos son ejemplos de variable aleatoria discreta. Como vemos habitualmente son valores que resultan de contar. Es lo más frecuente pero no siempre es así.

De un modo más formal:

Definición 2.6 (Variable aleatoria discreta) *Una variable discreta es aquella que verifica que existe un conjunto numerable D tal que la variable toma valores en este conjunto o, de otro modo,*

$$P(X \in D) = P_X(D) = 1.$$

[10]

[10] El conjunto D recibe el nombre de **soporte** de la distribución P_X. Básicamente una variable discreta es aquella cuyo soporte es un conjunto discreto.

Notemos que si $x \in D^c$ entonces $P(X = x) = 0$. Es por ello, razonable definir la siguiente función.

Definición 2.7 (Función de probabilidad) *Definimos la función de probabilidad o cuantía de una variable aleatoria discreta como*

$$f_X(x) = \begin{cases} P(X = x), & \textit{si } x \in D \\ 0, & \textit{en el resto,} \end{cases}$$

[11]

[11] En lo que sigue denotaremos indistintamente $f_X(x)$ o $P(X = x)$ indistintamente cuando hablemos de la función de probabilidad de una variable aleatoria **discreta**.

Ejemplo 2.2 *Por ejemplo, supongamos $\mathbb{D} = \{0, 1, \ldots, 10\}$ y la función de probabilidad aparece en la tabla 2.1.*

A partir de la función de probabilidad $P(X = x_i)$ podemos calcular cualquier otra probabilidad. Por ejemplo:

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2),$$

o bien,

$$P(X \geq 7) = P(X = 7) + P(X = 8) + P(X = 9) + P(X = 10).$$

También

$$P(4 \leq X \leq 7) = P(X = 4) + P(X = 5) + P(X = 6) + P(X = 7).$$

$$P(4 < X \leq 7) = P(X = 5) + P(X = 6) + P(X = 7).$$

$$P(4 < X < 7) = P(X = 5) + P(X = 6).$$

De un modo genérico podemos escribir que

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

siendo B cualquier subconjunto de la recta real (para ser precisos cualquier subconjunto de Borel de la recta).

Notemos que una función de probabilidad ha de verificar las dos propiedades básicas siguientes.

Proposición 2.2 *Si f_X es la función de probabilidad asociada a una variable aleatoria discreta entonces se verifican las dos propiedades siguientes:*

1. *f_X es no negativa.*
2. $\sum_{x_i \in D} f_X(x_i) = 1$.

Demostración 2.2

1. *Al tratarse de una probabilidad, $f_X(x) \geq 0, \ \forall x \in \mathbb{R}$,*
2. *Como $P(X \in D) = 1$ se tiene por la σ-aditividad de la medida de probabilidad que*

$$\sum_{x_i \in D} f_X(x_i) = 1.$$

De hecho en lo que sigue cuando demos la función de probabilidad de una variable discreta simplemente indicaremos los puntos en los que no se anula y el valor de la función en esos puntos. En el resto asumimos que es nula.

Si X es una variable discreta entonces es fácil comprobar que la F_X asociada viene dada por

$$F_X(x) = \sum_{x_i \leq x} P(X = x_i) = \sum_{x_i \leq x} f_X(x_i). \tag{2.5}$$

De acuerdo con esto, si $x_{(i)}$ y $x_{(i+1)}$ son dos puntos consecutivos del soporte tendremos que $\forall x \in [x_{(i)}, x_{(i+1)}[, \ F_X(x) = F_X(x_{(i)})$. Como además $P_X(x) = 0, \ \forall x \in D^c$, la función será también continua. Por otra parte $P(X = x_i) > 0$, para $x_i \in D$, con lo que los únicos puntos de discontinuidad serán lo del soporte, discontinuidad de salto finito cuyo valor es $F_X(x_{(i)}) - F_X(x_{(i-1)}) = P(X = x_i)$. Se trata por tanto de una función escalonada, cuyos saltos se producen en los puntos de D.

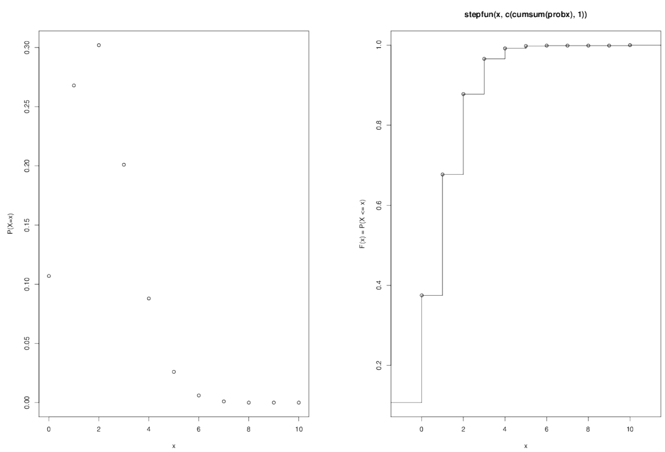
La relación entre f_X y F_X viene recogida en las dos expresiones que siguen, cuya obtención es evidente a partir de (2.4) y (2.5). La primera de ellas permite obtener F_X a partir de f_X,

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

La segunda proporciona f_X en función de F_X,

$$f_X(x) = F_X(x) - F_X(x-).$$

En resumen dada la función de probabilidad tenemos la función de distribución y viceversa, dada la función de distribución tenemos la función de probabilidad.



(a) (b)

Figura 2.1: Función de probabilidad (a) y función de distribución (b) de la distribución discreta en tabla 2.1.

Ejemplo 2.3 *Consideramos la variable aleatoria discreta tal que su función de probabilidad aparece en la tabla 2.1. Vamos a determinar la función de distribución. Sus funciones de probabilidad y de distribución las tenemos en la figura 2.1.*

2.6 Ejercicios

* **Ej. 3** — Sea X una variable aleatoria discreta con función de probabilidad definida como $f(x) = \frac{x}{10}$ para $x \in \{1, 2, 3, 4\}$(y 0 en otro caso). Calcula las siguientes probabilidades:

1. $P(1 < X < 3), P(1 \leq X < 3), P(1.5 < X < 3)$, $P(1.5 < X \leq 3), P(1.5 \leq X \leq 1.99)$.
2. $P(X \leq 1), P(X < 1), P(0 \leq X \leq 2 | 1 \leq X \leq 3)$
3. $P(X \in N)$ donde N es el conjunto de los números enteros positivos.
4. $P(X \in Q)$ donde Q es el conjunto de los números racionales.
5. Calcula y dibuja la función de distribución $F(x)$ de la variable X. Repite los dos primeros apartados del problema utilizando $F(x)$. Representa en la gráfica dibujada todas las probabilidades del primer apartado del problema.

* **Ej. 4** — Calcula la función de probabilidad de una variable aleatoria discreta X para la que $P(X > x) = \frac{1}{2^x}$ si $x \in \{1, 2, 3, \ldots\}$ y $P(X = x) = 0$ en otro caso.

* **Ej. 5** — Considera un juego que consiste en lanzar repetidamente una moneda (bien equilibrada) hasta conseguir que el resultado sea cara. Cada vez que obtienes el resultado cruz colocas una ficha de color blanco en un sombrero, pero si el resultado es cara, además de acabarse el juego, añades una ficha negra en el sombrero. De esa forma, cuando finaliza el juego, el número de fichas en el sombrero es

igual al número de lanzamientos realizados. A continuación se elige al azar una ficha del sombrero. Si es blanca ganas 100 euros, pero si es la negra pierdes 100 euros. Comprueba que la probabilidad de que ganes es $0.3069 (= 1 - \ln 2)$. (Recuerda que, desarrollando en serie de potencias, se tiene que $\ln(\frac{1}{1-x}) = \sum_{n=1}^{\infty} \frac{1}{n} x^n$ si $|x| < 1$).

* **Ej. 6** — De un lote que contiene 25 artículos, de los cuales 5 son defectuosos, se eligen 4 al azar. Sea X la variable aleatoria que cuenta el núumero de artículos defectuosos elegidos. Obtén la distribución de probabilidad de la variable aleatoria X si los artículos se escogen

1.con reemplazamiento

2.sin reemplazamiento.

* **Ej. 7** — La concha del caracol de tierra Limocolaria mantersiana puede ser rayada o pálida. El 60 por ciento de estos caracoles de tierra tiene la concha rayada. Se seleccionan 10 caracoles al azar de esta población. Calcula la probabilidad de que la proporción de caracoles con concha rayada presentes en la muestra:

1.sea del 50 por cien,

2.esté entre el 50 y el 60 por cien, ambos inclusive,

3.no sea inferior al 70 por cien.

4.sea del 60 por cien.

* **Ej. 8** — (*[10, Pág. 172]*)Elegimos dos bolas uniformemente de una urna que contiene 8 bolas blancas, 4 negras y 2 naranjas. Supongamos que ganamos dos euros por cada bola negra seleccionada y perdemos un euro por cada bola blanca que seleccionamos. Sea X la variable aleatoria que nos da la ganancia. Determinar los valores que puede tomar la variable X y cuáles son las probabilidades de estos valores.

* **Ej. 9** — (*[10, Pág. 172]*)Lanzamos dos dados correctos. Sea X la variable aleatoria igual al producto de los dos dados. Determinar $P(X = i)$ para $i = 1 \ldots, 36$.

2.6.1 Ejemplos importantes de distribuciones discretas

En lo que sigue se estudian los ejemplos de variables discretas de mayor importancia. Son modelos fácilmente reconocibles porque corresponden con experimentos que nos encontramos en la vida real. Uno de ellos será el modelo binomial. Una cuestión previa es decidir si hablamos de *variable binomial* o de *distribución binomial*. En todos los ejemplos que vamos a contemplar lo que damos es la función de probabilidad, es decir, damos los puntos con probabilidad no nula y la probabilidad de que se produzcan. Lo importante es la probabilidad de los sucesos y no la definición precisa de la variable. La variable aleatoria puede representar número de caras en n lanzamientos de una moneda correcta o puede corresponder al número de píxeles en mal estado en un monitor. Lo importante, en Probabilidad, no es la definición de la variable sino la probabilidad que induce, su distribución. Dada la distribución variables que se distribuyen igual son, desde el punto de vista probabilístico, equivalentes. Por ello, preferimos (y es lo más habitual) hablar de distribución binomial. Una variable con distribución binomial puede denominarse variable binomial. Nosotros preferimos decir que es una variable con distribución binomial.[12]

[12] Es más largo, pero somos un poco de letras.

2.7 Distribución Bernoulli

[13] Distribución Bernoulli

[13] Empezamos con algún ejemplo que motiva esta distribución.

1. Lanzamos una moneda y nos planteamos si sale cara (y lo llamamos éxito) o si no sale cara (y lo llamamos fracaso). Nos fijamos en la ocurrencia o no de un suceso: salir cara. Tenemos un espacio muestral $\Omega = \{cara, cruz\}$ y nos planteamos si se produce $A = \{cara\}$.

2. Elegimos al azar a una persona en la población de la Comunidad Valenciana y nos fijamos si es diabético (y lo consideramos éxito) o no lo es (fracaso). En este caso el espacio muestral es el conjunto de individuos de la Comunidad y el suceso que nos interesa sería el conjunto de individuos que son diabéticos en dicha Comunidad.

Tenemos, en los ejemplos anteriores, distintos experimentos donde nos fijamos en un suceso A y nos planteamos si ocurre o no ocurre este suceso. Una situación de este tipo recibe el nombre de prueba de Bernoulli. Dada una prueba de Bernoulli consideramos la variable aleatoria que nos indica si A se ha producido, esto es, la variable

$$X(\omega) = 1_A(\omega),$$

donde $1_A(\omega) = 1$ si $\omega \in A$ y cero en otro caso. Del experimento que realizamos solamente nos importa saber si un determinado suceso ha tenido lugar. De hecho, como éxito es que el resultado ω pertenezca al suceso A entonces la probabilidad de éxito será $p = P(A)$.

El rango de la variable es $\{1, 0\}$ donde 1 indica el éxito y 0 el fracaso. Una variable de este tipo recibe el nombre de variable Bernoulli. Es el ejemplo más simple de variable aleatoria. Su función de probabilidad sería: $f_X(1) = P(X = 1) = p$ y $f_X(0) = P(X = 0) = 1 - p$ que, de un modo conjunto, podemos representar como

$$P(X = x) = p^x(1-p)^{1-x} \text{ para } x = 0, 1. \tag{2.6}$$

Si una variable sigue una distribución de probabilidad se dice que es una variable Bernoulli o que se distribuye Bernoulli y se denota

$$X \sim Bi(1, p).$$

Ejemplo 2.4 (Lanzamiento de moneda correcta) *La función de probabilidad es $f_X(1) = P(X = 1) = 1/2$ y $f_X(0) = P(X = 0) = 1/2$.*

Ejemplo 2.5 (Selección aleatoria de un individuo) *Seleccionamos a un individuo al azar en la población de la Comunidad Valenciana. La función de probabilidad es $f_X(1) = P(X = 1) = p$ y $f_X(0) = P(X = 0) = 1 - p$. El valor de p nos indica la* proporción real *de diabéticos en la población. Habitualmente el valor de p no es conocido.*

* **Ej. 10** — Supongamos $X \sim Bi(1, p)$. Demostrar que $f_X(1) = P(X = 1) = p$ y $f_X(0) = P(X = 0) = 1 - p$ es una función de probabilidad.

2.8 Distribución binomial

[14]

[14] Distribución binomial

Repetimos n veces una prueba de Bernoulli (con probabilidad de éxito p) independientemente una de otra (lanzamos n veces una moneda por ejemplo) y observamos la variable aleatoria que nos da el número total de éxitos. Una variable de este tipo recibe el nombre de *variable binomial con n pruebas y una probabilidad de éxito p* y se suele denotar como

$$X \sim Bi(n, p).$$

Observemos que los valores que puede tomar esta variable son $0, 1, 2, \ldots, n$, esto es, desde cero éxitos hasta n éxitos. ¿Qué probabilidad tenemos de observar un número k éxitos?

Fenómenos aleatorios aparentemente tan diferentes como el número de hijos varones de un matrimonio con n hijos o el número de caras obtenidas al lanzar n veces una moneda correcta, son bien descritos mediante un variable binomial.

Definición 2.8 (Variable binomial) *Decimos que X es una* variable binomial *(o que sigue una distribución binomial) de parámetros n y p y lo denotamos como*

$$X \sim Bi(n, p)$$

si su función de probabilidad viene dada por

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \ \ si \ \ x \in \{0, 1, \ldots, n\}, \tag{2.7}$$

y cero en el resto.

Lema 2.2 *Si realizamos n pruebas independientes de Bernoulli con la misma probabilidad de éxito p y denotamos por X la variable que nos cuenta el número de éxitos entonces se verifica que $P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$ para $x = 0, \ldots, n$.*

Demostración 2.3 *Supongamos que denotamos por 1 un éxito y por cero fracaso. Vamos a denotar por $A_{(u_1,\ldots,u_n)}$ el suceso consistente en obtener en la prueba i el resultado u_i donde u_i igual uno indica éxito en la prueba i y cero indica fracaso. Si denotamos por $A_{(u_1,\ldots,u_n)}$ el suceso correspondiente tendremos que*

$$P(A_{(u_1,\ldots,u_n)}) = p^{\sum u_i} (1-p)^{n - \sum u_i}$$

puesto que las distintas pruebas de Bernoulli son independientes. Si X es el número total de éxitos entonces

$$\{X = x\} = \bigcup_{\sum_{i=1}^n u_i = x; u_i \in \{0,1\}} A_{(u_1,\ldots,u_n)}.$$

Notemos que los $A_{(u_1,\ldots,u_n)}$ son disjuntos cuando cambia el valor de uno de los u_i (en una prueba o tenemos un éxito o un fracaso) por lo que

$$P(\{X = x\}) = \sum_{\sum_{i=1}^n u_i = x; u_i \in \{0,1\}} P(A_{(u_1,\ldots,u_n)}).$$

Y fácilmente vemos que el número de sumandos en el sumatorio anterior es igual a $\binom{n}{x}$ pues solamente hemos de decidir en qué pruebas se producen los x éxitos ya los fracasos vienen dados como el resto de pruebas.

Una distribución binomial con una sola prueba recibe el nombre de *distribución Bernoulli*.

Veamos una propiedad interesante de la distribución binomial. Tenemos que

$$P(X = k+1) = \binom{n}{k+1} p^{k+1}(1-p)^{n-k-1} = \frac{n!}{(k+1)!(n-k-1)!} p^{k+1}(1-p)^{n-k-1} = \frac{(n-k)p}{(k+1)(1-p)} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} = \frac{(n-k)p}{(k+1)(1-p)} P(X=k). \tag{2.8}$$

Y en particular tendremos que

$$\frac{P(X=k)}{P(X=k+1)} = \frac{k+1}{n-k}\left(\frac{1-p}{p}\right). \tag{2.9}$$

El cociente de la ecuación 2.9 es menor que 1 cuando se verifica que $k < (n+1)p - 1$. Por tanto para estos valores de k la probabilidad $P(X = k+1)$ es mayor que $P(X = k)$, esto es, la probabilidad crece. Sin embargo cuando $k \geq (n+1)p - 1$ la desigualdad es la contraria y decrecemos.[15] En consecuencia el máximo lo alcanzamos para k igual a $[(n+1)p]$[1] Si ocurre que $(n+1)p$ es un número entero entonces se tendrá que

$$\frac{P(X = [(n+1)p - 1])}{P(X = [(n+1)p])} = 1, \tag{2.10}$$

y en ambos puntos se alcanza el máximo. Por cierto que lo que acabamos de determinar es la moda de la distribución binomial.[16]

[15] En palabras de un político español diríamos que hay un crecimiento negativo.

[16] La moda de una distribución discreta es el valor que maximiza la función de probabilidad.

2.8.1 Ejercicios

* **Ej. 11** — (*[11, Pág. 29]*)Lanzamos cuatro monedas correctas. Supongamos que los resultados son independientes: ¿qué probabilidad tenemos de observar dos caras y dos cruces?

* **Ej. 12** — (*[11, Pág. 30]*)Sabemos que la probabilidad de que una determinada máquina produzca una unidad defectuosa es 0.1. Además el que una unidad sea defectuosa es independiente de que las demás unidades lo sean. Supongamos una muestra de tres unidades fabricadas: ¿qué probabilidad tenemos de tener al menos una unidad de producto defectuosa?

* **Ej. 13** — (*[11, Pág. 30]*)Supongamos que el motor de un avión falla en vuelo con una probabilidad $1-p$. Supongamos que los distintos motores de un avión fallan independientemente de los demás. El avión completará el vuelo si al menos permanecen operativos la mitad de los motores que lleva. ¿Qué valores de p hacen que sea preferible volar en un avión de cuatro motores antes que en un motor de dos motores?

* **Ej. 14** — (*[11, Pág. 31]*)Supongamos que un cierto rasgo de una persona (por ejemplo, color de pelo o si es diestro o zurdo) depende de un par de genes. Suponemos que uno de ellos, d, es dominante y el otro, r, es recesivo. Si una persona es dd es dominante puro, si es

[1] Denotamos por $[x]$ la parte entera de x o mayor entero menor o igual que x.

rr es recesivo puro y si es rd es híbrido. Los dominantes puros y los híbridos tienen la misma apariencia. Un descenciente recibe un gen de cada padre. Supongamos una pareja donde los dos padres son híbridos. Tienen cuatro hijos: ¿qué probabilidad tenemos de que exactamente tres de los descencientes tenga la apariencia que proporciona el gen dominante?

* **Ej. 15** — Supongamos que realizamos n pruebas de Bernoulli independientes con una probabilidad de éxito p (la misma para todas las pruebas). Demostrar que la probabilidad de obtener exactamente x éxitos viene dada por la expresión $\binom{n}{x} p^x (1-p)^{n-x}$ donde $x = 0, \ldots, n$.

* **Ej. 16** — Demostrar que la función definida en 2.7 es una función de probabilidad.

2.9 Distribución Poisson

[17] Esta distribución aparece ligada a experimentos en los que nos interesa la ocurrencia de un determinado suceso a lo largo de un intervalo finito de tiempo[2], verificándose las siguientes condiciones:

[17] Distribución Poisson

1. La probabilidad de que el suceso ocurra en un intervalo pequeño de tiempo es proporcional a la longitud del intervalo, siendo λ el factor de proporcionalidad.

2. La probabilidad de que el suceso ocurra en dos o más ocasiones en un intervalo pequeño de tiempo es prácticamente nula.

Fenómenos como el número de partículas que llegan a un contador Geiger procedentes de una fuente radiactiva, el número de llamadas que llegan a una centralita telefónica durante un intervalo de tiempo, las bombas caídas sobre la región de Londres durante la Segunda Guerra mundial y las bacterias que crecen en la superficie de un cultivo, entre otros, pueden ser descritos mediante una variable aleatoria Poisson.

La distribución de *Poisson* de parámetro λ es una de las distribuciones de probabilidad discretas más conocida. Denotaremos que la variable X sigue una distribución de Poisson con

$$X \sim Po(\lambda).$$

El soporte de la distribución es $D_X = \{0, 1, 2, 3, \ldots\}$ y su función de probabilidad viene dada por

$$f_X(x) = P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!},\ si\ x = 0, 1, \ldots \tag{2.11}$$

y cero en el resto. [3]

[2] En un planteamiento más general, el intervalo finito de tiempo puede ser sustituido por un subconjunto acotado de $\mathbb{R}^k$

[3] Es fácil comprobar que la función que acabamos de considerar en 2.11 es de probabilidad. Es no negativa y

$$\sum_{x=0}^{+\infty} e^{-\lambda}\frac{\lambda^x}{x!} = e^{-\lambda}\sum_{x=0}^{+\infty}\frac{\lambda^x}{x!} = e^{-\lambda}e^{\lambda} = 1.$$

La función de distribución tiene por expresión

$$F_X(x) = \sum_{n \leq x} \frac{e^{-\lambda}\lambda^n}{n!}.$$

Diremos que X es una *variable Poisson de parámetro* λ o que X sigue una distribución de Poisson y lo denotaremos

$$X \sim Po(\lambda).$$

2.9.1 Ejercicios

* **Ej. 17** — Supongamos $X \sim Po(\lambda)$. Demostrar que la función definida en 2.11 es una función de probabilidad.

* **Ej. 18** — (*[11, Pág. 33]*)Supongamos que el número de errores tipográficos en una página de un libro sigue una distribución de Poisson con parámetro $\lambda = 1$. Determinar la probabilidad de que tengamos al menos un error en un página.

* **Ej. 19** — (*[11, Pág. 33]*)El número de accidentes que ocurren en una autopista sigue una distribución de Poisson con parámetro $\lambda = 3$: ¿qué probabilidad de que no tener ningún accidente en un día determinado?

* **Ej. 20** — (*[11, Pág. 33]*)En un experimento estamos contando el número de particulas α en un intervalo de un segundo con un gramo de material radioactivo. Por información anterior sabemos que dicho número aleatorio viene bien descrito mediante una distribución de Poisson con parámetro $\lambda = 3.2$. Obtener una buena aproximación para la probabilidad de no observar más de dos partículas α.

2.9.2 Aproximando una distribución binomial

¿Por qué es tan importante la distribución de Poisson en las aplicaciones? Posiblemente[4] la razón sea que aproxima a la distribución binomial cuando tenemos un gran número de pruebas y una pequeña probabilidad de éxito en cada una de ellas. Esto es hacemos muchas pruebas pero es muy raro que tengamos un éxito en una de ellas. Por ello, a esta distribución se le suele llamar la *distribución de los sucesos raros*. Realmente lo que estamos viendo es La distribución de Poisson como límite de distribuciones binomiales.

Proposición 2.3 *Sea* $\{X_n\}_{n \geq 1}$ *una sucesión de variables tales que* $X_n \sim Bi(n, p_n)$ *y* $\lim_{n \to +\infty} np_n = \lambda$ *entonces*

$$\lim_{n \to +\infty} P(X_n = x) = \frac{e^{-\lambda}\lambda^x}{x!}.$$

Demostración 2.4 *Consideremos la sucesión de variables aleatorias* $X_n \sim Bi(n, p_n)$ *en la que a medida que* n *aumenta,* p_n *disminuye de forma tal que* $np_n \approx \lambda$*. Más concretamente,* $np_n \to \lambda$*. Tendremos para la función de probabilidad que*

$$P(X_n = x) = \binom{n}{x} p_n^x (1 - p_n)^{n-x} = \frac{n!}{x!(n-x)!} p_n^x (1 - p_n)^{n-x},$$

[4]Probablemente queda mejor en estas notas

y para n suficientemente grande,

$$\begin{aligned} P(X_n = x) &\approx \frac{n!}{x!(n-x)!}\left(\frac{\lambda}{n}\right)^x \left(1-\frac{\lambda}{n}\right)^{n-x} \\ &= \frac{\lambda^x}{x!}\frac{n(n-1)\cdots(n-x+1)}{n^x}\left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-x}. \end{aligned}$$

Al pasar al límite,

$$\frac{n(n-1)\cdots(n-x+1)}{n^x} \to 1, \quad \left(1-\frac{\lambda}{n}\right)^n \to e^{-\lambda}, \quad \left(1-\frac{\lambda}{n}\right)^{-x} \to 1,$$

y tendremos

$$\lim_{n\to+\infty} f_{X_n}(x) = \frac{e^{-\lambda}\lambda^x}{x!}.$$

La utilidad de este resultado reside en permitir la aproximación de la función de cuantía de una binomial con n pruebas y probabilidad de éxito p, $Bi(n,p)$, mediante la función de cuantía de una Poisson con párametro $\lambda = np$ cuando n es grande y p pequeño.

2.10 Distribución hipergeométrica

Es una distribución de probabilidad discreta que aparece de un modo natural. Consideremos el siguiente experimento. Tenemos una urna con N bolas, de las cuales r son rojas y el resto, $N-r$, son blancas. Extraemos n de ellas *sin reemplazamiento* y consideramos la variable aleatoria X que nos da el número de bolas rojas que extraemos. Notemos que el número de bolas rojas puede tomar valores entre 0 y el mínimo entre el número de bolas que extraemos y el número de bolas rojas que tenemos en la urna. Por tanto el soporte de la distribución de X es $D_X = \{0,1,2,\ldots,\min\{n,r\}\}$. Para obtener la función de probabilidad simplemente hemos de aplicar la fórmula de Laplace. Tenemos

$$P(X = x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}, \text{ si } \ x = 0,\ldots,\min\{n,r\},$$

y cero en el resto. La variable X se dice que sigue una *distribución hipergeométrica* y lo denotaremos con

$$X \sim H(n,N,r).$$

La función que acabamos de definir se comprueba con facilidad que es una función de probabilidad. Es no negativa. Para comprobar que la suma de los valores es uno utilizamos la siguiente igualdad.

$$\sum_{i=0}^{n}\binom{a}{i}\binom{b}{n-i} = \binom{a+b}{n}.$$

[18]

[18] Para probar esta igualdad un argumento sencillo es ver que estamos contando el número de grupos de tamaño n de un total de a+b elementos mediante dos procedimientos distintos.

Nota 2.1 (Distribución binomial e hipergeométrica) *La diferencia entre los modelos binomial e hipergeométrico estriba en el tipo de extracción. Cuando ésta se lleva a cabo con reemplazamiento las sucesivas extracciones son independientes y la probabilidad de* éxito *se mantiene constante e igual a r/N, el modelo es binomial. No ocurre así si las extracciones son sin reemplazamiento. No obstante, si n es muy pequeño respecto a N y r, la composición de la urna variará poco de extracción a extracción y existirá lo que podríamos denominar una* quasi-independencia *y la distribución hipergeométrica deberá comportarse como una binomial. En efecto,*

$$
\begin{aligned}
f_X(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} = {}& \\
& \frac{r!}{x!(r-x)!}\frac{(N-r)!}{(n-x)!(N-r-n+x)!}\frac{n!(N-n)!}{N!} = \\
& \binom{n}{x}\frac{r}{N}\frac{r-1}{N-1}\cdots\frac{r-x+1}{N-x+1} \\
& \frac{N-r}{N-x}\frac{N-r-1}{N-x-1}\cdots\frac{N-r-n+x+1}{N-n+1} = \\
& \approx \binom{n}{x}p^x(1-p)^{n-x}, \qquad (2.12)
\end{aligned}
$$

con $p = \frac{r}{N}$.

Aplicación 2.1 (Estimación del tamaño de una población animal) *Y vamos a estimar dicho tamaño a partir de datos de recaptura*[19] *Queremos estimar la población de peces en un lago, para ello hemos capturado 1000 peces a los que, marcados mediante una mancha roja, hemos arrojado nuevamente al lago. Transcurrido un cierto tiempo, el necesario para que se mezclen con los restantes peces del lago, llevamos a cabo una nueva captura de otros 1000 peces entre los que hay 100 marcados. ¿Qué podemos decir acerca del total de peces en el lago?*

[19] El ejemplo está sacado de [2] un texto clásico de la Probabilidad (hay dos volúmenes) cuya lectura y consulta es recomendable y, lo mejor, un absoluto placer.

El problema que planteamos en un problema típico de estimación estadística *y vamos a dar una solución que, aunque particular para la situación descrita, está basada en una metodología de aplicación general en los problemas de estimación. Observemos en primer lugar que el número de peces marcados en la segunda captura (recaptura) es una variable aleatoria hipergeométrica, $X \sim H(1000, N, 1000)$, siempre bajo el supuesto de que ambas capturas constituyen sendas muestras aleatorias de la población total de peces del lago (en la práctica semejante suposición excluye situaciones en las que las capturas se efectúan en el mismo lugar y en un corto periodo de tiempo). Suponemos también que el número de peces en el lago, N, no cambia entre las dos capturas.*

Generalizemos el problema admitiendo tamaños arbitrarios para *ambas muestras:*

$$
\begin{aligned}
N &= \textit{población de peces en el lago (desconocida)}\\
r &= \textit{número de peces en la } 1^{\underline{a}} \textit{ captura}\\
n &= \textit{número de peces en la } 2^{\underline{a}} \textit{ captura}\\
x &= \textit{número de peces en con mancha roja en la } 2^{\underline{a}} \textit{ captura}\\
p_x(N) &= \textit{probabilidad de } x \textit{ peces con mancha roja en la } 2^{\underline{a}} \textit{ captura}
\end{aligned}
$$

Con esta formulación sabemos que

$$p_x(N) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}.$$

En la práctica, r, n y x son conocidos por observación, como en el ejemplo que planteamos, mientras que N es desconocido pero fijo y en modo alguno depende del azar. Al menos una cosa conocemos de N y es que $N \geq r + n - x$, que es el total de peces capturados entre ambas capturas. En nuestro ejemplo, $N \geq 1000 + 1000 - 100 = 1900$. ¿Qué ocurre si aceptamos $N = 1900$? Aunque se trata de un valor teóricamente posible, si calculamos $p_{100}(1900)$,

$$p_{100}(1900) = \frac{\binom{1000}{100}\binom{900}{900}}{\binom{1900}{1000}} \approx 10^{-430},$$

(podemos valernos de la fórmula de Stirling, $n! \approx \sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n}$, para aproximar las factoriales), habremos de aceptar que ha ocurrido un suceso, $X = 100$, con una probabilidad extraordinariamente pequeña. Resulta difícil de admitir una hipótesis que exige casi un milagro para que el suceso observado tenga lugar. Otro tanto nos ocurre si suponemos que N es muy grande, por ejemplo $N = 10^6$. También ahora $p_{100}(10^6)$ es muy pequeña.

Una respuesta adecuada puede ser la de buscar el valor de N que maximiza $p_x(N)$. Dicho valor, que designamos mediante $\hat{N}$, recibe el nombre de estimación máximo-verosímil *de N. Para encontrarlo, observemos que*

$$\frac{p_x(N)}{p_x(N-1)} = \frac{(N-r)(N-n)}{(N-r-n+x)N} = \frac{N^2 - Nr - Nn + rn}{N^2 - Nr - Nn + Nx},$$

de donde se deduce

$$
\begin{aligned}
p_x(N) &> p_x(N-1), \qquad \textit{si } Nx < rn,\\
p_x(N) &< p_x(N-1), \qquad \textit{si } Nx > rn.
\end{aligned}
$$

Así pues, a medida que aumenta N la función $p_x(N)$ crece primero para decrecer después, alcanzando su máximo en $N = [rn/x]$, la parte entera de rn/x. En nuestro ejemplo, $\hat{N} = 10000$.

2.11 Distribución geométrica

[20] ¿Cuál es el número de lanzamientos de una moneda (correcta) [20] Distribución geométrica

que hemos de realizar antes de obtener la primera cara? No prefijamos el número de lanzamientos de la moneda. Solamente prefijamos que queremos alcanzar el primer éxito, la primera cara, y *contamos* el número de lanzamientos que hemos realizado.

En general, consideremos una sucesión de pruebas Bernoulli (éxito o fracaso) independientes con la misma probabilidad de éxito, p. La variable aleatoria que nos interesa es el número de pruebas que hemos de realizar hasta obtener el primer éxito (donde no vamos a contar el éxito). La denotamos por X. ¿Cuál es su función de probabilidad? Supongamos que denotamos por A_i el súceso consistente en obtener éxito en la prueba i-ésima. Es claro que $\{X = 0\} = A_1$ y

$$P(X = 0) = P(A_1) = p.$$

En general si $x > 0$ tenemos

$$P(X = x) = P\left(A_1^c \cap \ldots \cap A_x^c \cap A_{x+1}\right) = P(A_x) \prod_{i=1}^{x} P(A_i^c) = p(1-p)^x,$$

donde la segunda igualdad se sigue de la independencia de los sucesos A_i. En consecuencia, si X sigue una distribución geométrica con probabilidad de éxito p entonces su función de probabilidad viene dada por

$$P(X = x) = p(1-p)^x \text{ si } x = 0, 1, \ldots,$$

y cero en el resto. [21]

[21] Cuando se define la distribución geométrica a veces se cuenta el éxito final y a veces no. Esto lo único que nos modifica es el soporte de la distribución. Es importante fijarse es si se cuenta o no.

2.12 Distribución binomial negativa

Consideremos pruebas Bernoulli independientes con la misma probabilidad de éxito, p, en cada una de ellas. Fijamos el número de éxitos que queremos obtener, por ejemplo, r éxitos. Nos fijamos en el número de pruebas adicionales que hemos de realizar hasta obtener el éxito r-ésimo. Notemos que como son pruebas adicionales, el valor mínimo será 0 indicando que las r primeras pruebas han sido un éxito. Si X es la variable que nos da este número de pruebas adicionales entonces

$$P(X = x) = \binom{r+x-1}{x} p^r (1-p)^x, \text{ si } x \geq 0.$$

Esta distribución de probabilidad recibe el nombre de *distribución binomial negativa* y la variable se dice que tiene una distribución binomial negativa.

El nombre de binomial negativa se justifica a partir de la expresión alternativa que admite la función de cuantía,

$$f_X(x) = \begin{cases} \binom{-r}{x} p^r (-(1-p))^x, & si\ x \geq 0 \\ 0, & en\ el\ resto, \end{cases}$$

obtenida al tener en cuenta que

$$\begin{aligned}
\binom{-r}{x} &= \frac{(-r)(-r-1)\cdots(-r-x+1)}{x!} \\
&= \frac{(-1)^x r(r+1)\cdots(r+x-1)}{x!} \\
&= \frac{(-1)^x (r-1)! r(r+1)\cdots(r+x-1)}{(r-1)!x!} \\
&= (-1)^x \binom{r+x-1}{x}.
\end{aligned}$$

Veamos que suma la unidad. Recordemos el desarrollo en serie de potencias de la función $f(x) = (1-x)^{-n}$,

$$\frac{1}{(1-x)^n} = \sum_{i\geq 0} \binom{n+i-1}{i} x^i, \quad |x| < 1.$$

En nuestro caso,

$$\begin{aligned}
f_X(x) = \sum_{x\geq 0} \binom{r+x-1}{x} p^r (1-p)^x = \\
p^r \sum_{x\geq 0} \binom{r+x-1}{x} (1-p)^x = p^r \frac{1}{(1-(1-p))^r} = 1. \quad (2.13)
\end{aligned}$$

Ejemplo 2.6 (El problema de las cajas de cerillas de Banach) *En un acto académico celebrado en honor de Banach, H. Steinhaus contó una anécdota acerca del hábito de fumar que aquél tenía. La anécdota se refería a la costumbre de Banach de llevar una caja de cerillas en cada uno de los bolsillos de su chaqueta, de manera que cuando necesitaba una cerilla elegía al azar uno de los bolsillos. El interés de la anécdota residía en calcular las probabilidades asociadas al número de cerillas que habría en una caja cuando, por primera vez, encontrara vacía la otra.*

Si cada caja contiene N cerillas, en el momento de encontrar una vacía la otra puede contener $0, 1, 2, \ldots, N$ cerillas. Designemos por $A_r =${el bolsillo no vacío contiene r cerillas}. Supongamos que la caja vacía es la del bolsillo izquierdo, para que ello ocurra $N-r$ fracasos *(elecciones del bolsillo derecho) deben haber precedido al $N+1$-ésimo* éxito *(elección del bolsillo derecho). En términos de una variable aleatoria $X \sim BN(N+1, 1/2)$ se trata de obtener $P(X = N-r)$. El mismo argumento puede aplicarse si la caja vacía es la del bolsillo derecho. Así pues,*

$$\begin{aligned}
p_r = P(A_r) = 2P(X = N-r) = \\
2\binom{2N-r}{N-r} \left(\frac{1}{2}\right)^{N+1} \left(\frac{1}{2}\right)^{N-r} = \binom{2N-r}{N-r} 2^{-2N+r}. \quad (2.14)
\end{aligned}$$

Por ejemplo, para $N = 50$ y $r = 4$, $p_r = 0.074790$; para $r = 29$, $p_r = 0.000232$.

Observación 2.1 *Es también habitual presentar la variable binomial negativa como el número total de pruebas necesarias para alcanzar* *el r-ésimo éxito. En este caso, el soporte de la variable es $D_X = \{r, r+1, r+2, \ldots\}$ y su función de cuantía tiene la expresión*

$$f_X(x) = \begin{cases} \binom{x-1}{x-r} p^r (1-p)^{x-r}, & si\ x \geq r \\ \\ 0, & en\ el\ resto. \end{cases}$$

** **Ej. 21** — Consideremos el problema de las cajas de cerillas propuesto en ejemplo 2.6. Se propone encontrar la solución considerando cada una de las variantes siguientes:

1. Que el número de cerillas de cada una de las cajas no es el mismo: una de ellas supondremos que tiene N cerillas y la otra M cerillas.
2. Que el fumador elige su bolsillo derecho (con M cerillas) con el doble de probabilidad que su bolsillo izquierdo (con N cerillas).

2.13 Variable aleatoria continua

Hasta el momento hemos tratado de variables discretas. Aquellas variables que con probabilidad uno toman valores en un conjunto numerable. En esta sección vamos a ver otro tipo de variables aleatorias, las variables continuas. La definición de estas variables no es tan simple. Una definición sencilla y limitada podría ser la siguiente.

Definición 2.9 (Variable continua) *Diremos que la variable aleatoria es continua o que su distribución es continua cuando existe una función no negativa, integrable y cuya integral sobre toda la recta es igual a uno y verificando que*

$$P(X \in (a,b]) = P(a < X \leq b) = \int_a^b f(x)\, dx. \tag{2.15}$$

En la igualdad anterior podemos sustituir el intervalo semiabierto $(a,b]$ por los intervalos $[a,b], [a,b), (a,b)$.

En definitiva una variable continua es aquella que la probabilidad de que la variable esté en un intervalo viene dada por la integral (de Riemann) sobre el intervalo.[5] La función f recibe el nombre de *función de densidad de probabilidad*. Como consecuencia de esta definición, entre la función de distribución y la función de densidad de probabilidad se establecen las siguientes relaciones

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

y si x es un punto de continuidad de f_X entonces

$$f_X(x) = F_X^{'}(x).$$

[5] Esta no es una definición suficiente. El tono que pretendemos tenga estas notas nos limita a esta definición. Realmente una definición rigurosa es decir que la variable X es continua cuando su distribución de probabilidad P_X es absolutamente continua respecto a la medida de Lebesgue en la recta real, esto es, si para cualquier conjunto B tal que su longitud es nula entonces $P_X(B) = 0$. Asumiendo esta definición se puede demostrar (teorema de Radon-Nikodym) que existe una función f no negativa e integrable tal que $P_X(B) = \int_B f(x)dx$ donde la integral es la integral de Lebesgue en $\mathbb{R}$. Sin embargo, este tipo de resultados queda fuera de nuestro alcance en estas notas.

6

De hecho, se puede probar que la función de distribución de una variable aleatoria continua es absolutamente continua.

Definición 2.10 (Continuidad absoluta para F_X) *Decimos que F_X es* absolutamente continua *si para cada ϵ existe un δ tal que para cualquier familia finita de intervalos disjuntos, $[a_i, b_i], i = 1, \ldots, k$*

$$\sum_{i=1}^{k} |F_X(b_i) - F_X(a_i)| < \epsilon \quad si \quad \sum_{i=1}^{k} (b_i - a_i) < \delta.$$

Observemos que con esta definición, la continuidad uniforme es un caso particular cuando la familia de intervalos contiene un único elemento. Ello supone que F_X es continua.

2.13.1 Significado de la función de densidad de probabilidad

La continuidad de F_X implica, recordemos (2.4), que en las variables aleatorias continuas $P(X = x) = 0$, $\forall x \in \mathbb{R}$. Este es un resultado que siempre sorprende y para cuya comprensión es conveniente interpretar físicamente la función de densidad de probabilidad.

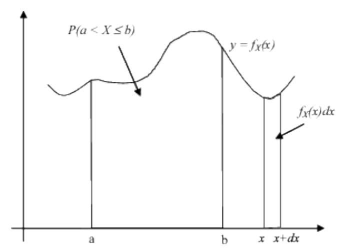


Figura 2.2: Función de densidad de probabilidad.

La función de densidad de probabilidad es sencillamente eso, una densidad de probabilidad que nos indica la cantidad de probabilidad por elemento infinitesimal de longitud. Es decir, $f_X(x)\,dx \approx P(X \in]x, x + dx])$. Ello explica que, para elementos con longitud cero, sea nula la correspondiente probabilidad. En este contexto, la probabilidad obtenida a través de la integral de Riemann pertinente se asimila a *un área*, la encerrada por f_X entre los límites de integración.

En las secciones que siguen vemos los ejemplos más importantes de distribuciones continuas.

2.13.2 Ejercicios

* **Ej. 22** — Sea X una variable aleatoria continua con función de densidad $f(x) = \frac{4}{3x^2}$ para $1 \leq x \leq 4$ (y 0 en otro caso). Calcula las siguientes probabilidades:

1. $P(1 < X < 3), P(1 \leq X < 3), P(1.5 < X < 3)$, $P(1.5 < X \leq 3), P(1.5 \leq X \leq 1.99)$.
2. $P(X \leq 1), P(X < 1), P(0 \leq X \leq 2|1 \leq X \leq 3)$
3. $P(X \in N)$ donde N es el conjunto de los números enteros positivos.
4. $P(X \in Q)$ donde Q es el conjunto de los números racionales.
5. Calcula y dibuja la función de distribución $F(x)$ de la variable X. Repite los dos primeros apartados del problema utilizando $F(x)$. Representa en la gráfica dibujada todas las probabilidades del primer apartado del problema.

[6] Esta última igualdad merece matizarse. En efecto, puesto que el origen de la densidad está en la continuidad absoluta de P_X respecto de la medida de Lebesgue, cualquier función que difiera de f_X en un conjunto con medida de Lebesgue nula será también una densidad. En otras palabras, la densidad de probabilidad no es única. Por ello sería más riguroso decir que $F_X'(x)$ es una de las posibles densidades.

* **Ej. 23** — ([13])Supongamos X una variable aleatoria con función de densidad $f(x) = cx$, para $0 \leq x \leq a$. Determinar c y la función de distribución de X.

* **Ej. 24** — ([13])Sean $f_1(x)$ y $f_2(x)$ funciones de densidad de probabilidad y supongamos λ un número tal que $0 \leq \lambda \leq 1$. Demostrar que $\lambda f_1(x) + (1-\lambda)f_2(x)$ es una función de densidad de probabilidad. ¿Es $f_1(x)f_2(x)$ una función de densidad de probabilidad?

* **Ej. 25** — ([13])Sean $F_1(x)$ y $F_2(x)$ funciones de distribución de probabilidad y supongamos λ un número tal que $0 \leq \lambda \leq 1$. Demostrar que $\lambda F_1(x) + (1-\lambda)F_2(x)$ es una función de distribución de probabilidad. ¿Es $F_1(x)F_2(x)$ una función de distribución de probabilidad?

2.14 Distribución uniforme

[22] Distribución uniforme

[22] La variable X diremos que tiene una *distribución uniforme en el intervalo* [a,b], $X \sim U(a, b)$, si su función de densidad de probabilidad es de la forma

$$f_X(x) = \begin{cases} \dfrac{1}{b-a}, & si\ x \in [a, b] \\ \\ 0, & en\ el\ resto. \end{cases}$$

La función de distribución que obtendremos integrando f_X vale

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

Surge esta variable cuando elegimos al azar un punto en el intervalo [a,b] y describimos con X su abscisa.

2.14.1 Ejercicios

* **Ej. 26** — Supongamos X uniforme en el intervalo $[a, b]$, determinar la función de distribución de la variable aleatoria X.

* **Ej. 27** — Supongamos $X \sim U(0, 5)$. Calcular las siguientes probabilidades:

1.$P(X < 3)$.

2.$P(2 < X < 3)$.

3.$P(X > 43)$.

2.15 Distribución normal

[23] Distribución normal

[23] Diremos que X es una *variable aleatoria normal* de parámetros

μ y σ^2, $X \sim N(\mu, \sigma^2)$, [7] si tiene por densidad,

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < +\infty. \tag{2.16}$$

Teorema 2.1 *La función definida en 2.16 es una función de densidad de probabilidad.*

Demostración 2.5 *En tanto que densidad, f_X debe ser no negativa e integrar uno sobre toda la recta real. La primera se deriva de inmediato de su definición. Para comprobar la segunda,*

$$I^2 = \left(\int_{-\infty}^{+\infty} f_X(x)dx\right)^2 = \left(\int_{-\infty}^{+\infty} f_X(x)dx\right)\left(\int_{-\infty}^{+\infty} f_X(y)dy\right) =$$
$$\left(\frac{1}{2\pi}\frac{1}{\sigma}\int_{-\infty}^{+\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx\right)\left(\frac{1}{\sigma}\int_{-\infty}^{+\infty} e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy\right) =$$
$$\left(\frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-\frac{z^2}{2}} \sigma dz\right)\left(\int_{-\infty}^{+\infty} e^{-\frac{v^2}{2}} \sigma dv\right) =$$
$$\frac{1}{2\pi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-\frac{z^2+v^2}{2}} dzdv =$$
$$\frac{1}{2\pi}\int_{0}^{2\pi}\left(\int_{0}^{+\infty} e^{-\frac{r^2}{2}} rdr\right)d\theta = 1, \tag{2.17}$$

donde hemos realizado los cambios $z = (x-\mu)/\sigma$ y $v = (v-\mu)/\sigma$ y posteriormente un cambio a polares $z = r\sin\theta$ y $v = r\cos\theta$.

La gráfica de f_X tiene forma de campana y es conocida como la campana de Gauss. Este matemático la introdujo cuando estudiaba los errores en el cálculo de las órbitas de los planetas. En honor suyo, la distribución normal es también conocida como distribución gaussiana. Fuera del ámbito de la Estadística y la Probabilidad quizás sea el nombre de gaussiana el más utilizado. Sin embargo, en Estadística lo habitual es hablar de la distribución normal. ¿Por qué? Pues porque cuando se estudian muchos valores aleatorios del mundo real lo **normal** es que se comporten según una distribución normal. Durante el siglo XIX se impuso esta denominación porque *normalmente* los histogramas se parecían a una campana de Gauss.

Del significado de los parámetros μ y σ^2 nos ocuparemos más adelante. Notemos que $\mu \in \mathbb{R}$ y $\sigma > 0$. Además, el eje de simetría de f_X es la recta $x = \mu$ y el vértice de la campana (máximo de f_x) está en el punto de coordenadas $(\mu, 1/\sigma\sqrt{2\pi})$.

La figura ilustra el papel que los parámetros juegan en la forma y posición de la gráfica de la función de densidad de una $N(\mu, \sigma^2)$. A medida que σ disminuye se produce un mayor apuntamiento en la campana porque el máximo aumenta y porque, recordemos, el área encerrada bajo la curva es siempre la unidad.

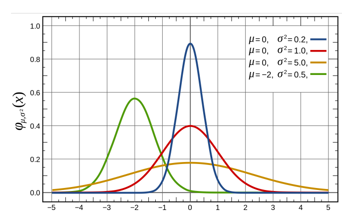


Figura 2.3: Diferentes densidades normales modificando los parámetros.

[7] Más adelante llamaremos a estos parámetros media μ y varianza σ^2 pero para entender el porqué de estos nombres hemos de estudiar previamente qué es la media y qué es la varianza.

Nota 2.2 (Calculando probabilidades con la distribución normal) *Una característica de la densidad de la normal es que su primitiva no tiene una expresión analítica, por lo que su función de distribución no tiene expresión explícita y sus valores están tabulados o se calculan por integración numérica. Esto representa un serio inconveniente si recordamos que* $P(a < X \leq b) = F_X(b) - F_X(a)$, *puesto que nos obliga a disponer de una tabla distinta para cada par de valores de los parámetros* μ *y* σ.

En realidad, no es necesario y, además, sería imposible dada la variabilidad de ambos parámetros. Podemos recurrir a la que se conoce como variable aleatoria normal tipificada, $Z \sim N(0,1)$, *cuya densidad es*

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \quad -\infty < z < +\infty.$$

En efecto, si para $X \sim N(\mu, \sigma^2)$ *queremos calcular*

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

efectuando el cambio $z = (t-\mu)/\sigma$ *tendremos*

$$F_X(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-\mu}{\sigma}} e^{-\frac{z^2}{2}} dz = \Phi\left(\frac{x-\mu}{\sigma}\right),$$

donde $\Phi(z)$ *es la función de distribución de la* $N(0,1)$. *Y, finalmente, si* $X \sim N(\mu, \sigma^2)$, *entonces*

$$P(a \leq X \leq b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right). \tag{2.18}$$

Además, puesto que es una distribución continua,

$$P(a < X \leq b) = P(a \leq X < b) = P(a < X < b) = P(a \leq X \leq b).$$

Hay que señalar que el mayor interés de la distribución normal estriba en el hecho de servir de modelo probabilístico para una gran parte de los fenómenos aleatorios que surgen en el campo de las ciencias experimentales y naturales.

El lema que sigue nos asegura que cualquier transformación lineal de un variable normal es otra normal.

Lema 2.3 *Sea* $X \sim N(\mu, \sigma^2)$ *y definamos* $Y = aX + b$, *entonces* $Y \sim N(a\mu + b, a^2\sigma^2)$.

Demostración 2.6 *Supongamos* $a > 0$, *la función de distribución de* Y *viene dada por*

$$F_Y(y) = P(Y \leq y) = P(aX+b \leq y) = P\left(X \leq \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right).$$

Su función de densidad es

$$f_Y(y) = F_Y'(y) = \frac{1}{a} f_X\left(\frac{y-b}{a}\right) = \frac{1}{a\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{y-(a\mu+b)}{a\sigma}\right)^2\right\}.$$

Si $a < 0$ entonces

$$F_Y(y) = P(Y \leq y) = P(aX+b \leq y) = P\left(X \geq \frac{y-b}{a}\right) = 1-F_X\left(\frac{y-b}{a}\right),$$

y la densidad será

$$f_Y(y) = F'_Y(y) = -\frac{1}{a} f_X\left(\frac{y-b}{a}\right) = \frac{1}{|a|\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{y-(a\mu+b)}{a\sigma}\right)^2\right\}.$$

En ambos casos se deduce que $Y \sim N(a\mu + b, a^2\sigma^2)$.

Nota 2.3 (De cómo calculaban los antiguos las probabilidades con la normal) *¿Qué problema nos planteamos? Suponemos que una cierta cantidad sigue una distribución normal con unos parámetros (media y varianza) dados. Nos planteamos cómo calcular la probabilidad de que la variable esté en un cierto intervalo. Por ejemplo, sabemos que el valor aleatorio que observamos sigue una distribución normal con media 56 y desviación típica 9. ¿Qué probabilidad tenemos de que la variable aleatoria tome un valor entre 60 y 63? Nos planteamos el valor de la siguiente probabilidad: $P(60 \leq X \leq 63)$. Para calcular este área se aplicaban las siguientes igualdades donde $Z = (X - 56)/3$,*

$$P(60 \leq X \leq 63) = P\left(\frac{60-56}{3} \leq \frac{X-56}{3} \leq \frac{63-56}{3}\right) =$$
$$P\left(\frac{60-56}{3} \leq Z \leq \frac{63-56}{3}\right) = P\left(Z \leq \frac{63-56}{3}\right) - P\left(Z \leq \frac{60-56}{3}\right). \tag{2.19}$$

Pero la variable Z es una normal estándar por lo que

$$P\left(Z \leq \frac{63-56}{3}\right) - P\left(Z \leq \frac{60-56}{3}\right) = \Phi\left(\frac{63-56}{3}\right) - \Phi\left(\frac{60-56}{3}\right).$$

De un modo genérico lo que acabamos de indicar es que si $X \sim N(56, 9)$ entonces

$$P(60 \leq X \leq 63) = \Phi\left(\frac{63-56}{3}\right) - \Phi\left(\frac{60-56}{3}\right),$$

siendo Φ la función de ditribución de una normal estándar.

Si suponemos que $X \sim N(\mu, \sigma^2)$ y tomamos dos números a y b tales que $a \leq b$ entonces

$$P(a \leq X \leq b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right), \tag{2.20}$$

Por tanto, si X es una variable aleatoria con distribución normal con media μ y desviación típica σ entonces la probabilidad de que la variable esté entre $\mu - \sigma$ y $\mu + \sigma$ es

$$P(\mu - \sigma \leq X \leq \mu + \sigma) = 0.6826895. \tag{2.21}$$

De un modo análogo si consideramos el intervalo $[\mu - 2\sigma, \mu + 2\sigma]$ entonces

$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = P(-2 \leq Z \leq 2) \tag{2.22}$$

$P(\mu - \sigma \leq X \leq \mu + \sigma)$	0.6826895
$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma)$	0.9544997
$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma)$	0.9973002

Tabla 2.2: Probabilidad de que la variable diste de la media en un número dado de desviaciones típicas

Y finalmente si consideramos el intervalo $[\mu - 2\sigma, \mu + 2\sigma]$ se tiene

$$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = P(-3 \leq Z \leq 3) \tag{2.23}$$

que es igual a 0.9973.

En la tabla 2.2 tenemos las probabilidades que hemos calculado. De un modo sencillo podemos decir: **la variable dista de la media en una desviación estándar con una probabilidad de 0.68, en dos desviaciones con una probabilidad de 0.95 y en tres desviaciones estándar con una probabilidad de 0.99**.

2.15.1 Ejercicios

* **Ej. 28** — Hace años se descubrió que los faisanes de Montana poseían una contaminación por mercurio apreciable, y se pensó que esto podía deberse a que habían sido alimentados con semillas de plantas tratadas con metilo de mercurio. Sea X la variable aleatoria que describe el nivel de mercurio (en ppm) de un faisán de Montana. Suponiendo que la distribución de X es normal de parámetros $\mu = 0.25$ ppm y $\sigma = 0.08$ ppm, calcula el percentil de orden 0.25 de su distribución y las probabilidades $P(X < 0.3), P(X > 0.17), P(0.2 < X < 0.4)$. [24]

[24] Para calcular la función de distribución de la normal se puede (y debe) utilizar la función *p*norm y para los percentiles la función *q*norm de R.

* **Ej. 29** — Un fabricante de bolas para rodamientos somete su producto al siguiente proceso de control de calidad. Las bolas son aceptadas si no pasan a través de un agujero de diámetro d_1, pero sí lo hacen a través de otro de diámetro d_2, $d_2 > d_1$. Se sabe que el diámetro D de las bolas es aleatorio con una distribución $N(\mu, \sigma^2)$, $\mu = (d_1 + d_2)/2$ y $\sigma = (d_2 - d_1)/4$. ¿Cuál es la probabilidad de rechazar una bola?

2.16 Distribución exponencial

[25] Distribución exponencial

[25] Diremos que la variable aleatoria X tiene una *distribución exponencial de parámetro* λ, $X \sim Exp(\lambda)$ siendo $\lambda > 0$, si su función de densidad es de la forma

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

La función de distribución de X vendrá dada por

$$F_X(x) = \begin{cases} 0, & si\ x \leq 0 \\ \int_0^x \lambda e^{-\lambda t}\, dt = 1 - e^{-\lambda x}, & si\ x > 0. \end{cases}$$

La distribución exponencial surge en problemas relacionados con tiempos de espera y está relacionada con la distribución de Poisson de igual parámetro. En efecto, si consideramos un proceso de desintegración radiactiva con λ desintegraciones por unidad de tiempo, el número de desintegraciones que se producen en el intervalo [0,t] es $N_t \sim Po(\lambda t)$, y el tiempo que transcurre ente dos desintegraciones consecutivas es $X \sim Exp(\lambda)$.

Nota 2.4 (La exponencial, una distribución desmemoriada) *La variable aleatoria exponencial tiene una curiosa e interesante propiedad conocida como* falta de memoria*. Consiste en la siguiente igualdad,*

$$P(X > x + t|X > t) = P(X > x), \quad \forall x, t \geq 0.$$

En efecto,

$$P(X > x+t|X > t) = \frac{P(\{X > x+t\} \cap \{X > t\})}{P(X > t)} = \frac{P(X > x+t)}{P(X > t)} = \frac{e^{-\lambda(x+t)}}{e^{-\lambda t}} = e^{-\lambda x} = P(X > x). \quad (2.24)$$

* **Ej. 30** — El tiempo de espera en minutos en la caja de un supermercado es una variable aleatoria de parámetro λ, $T \sim Exp(\lambda)$.

1.Cuál és la probabilidad de que un cliente espere más de 10 minutos?

2.Cuál debe ser el valor de λ para que solamente un 1% de los clientes espere más de 10 minutos?

* **Ej. 31** — Sea X una variable aleatoria con densidad exponencial de parámetro $\lambda = 2$. Halla las siguientes probabilidades:

1.$P(2X + 1 > 3)$

2.$P(X^2 \leq 7)$

3.$P(X \leq 2 \mid X > 0.5)$

4.$P(X > 0.5 \mid X \leq 2)$

2.17 Distribución gamma

[26] Diremos que la variable aleatoria X tiene una *distribución gamma de parámetros* α *y* β, $X \sim Ga(\alpha, \beta)$, donde $\alpha, \beta > 0$ si su función de densidad es de la forma

$$f_X(x) = \begin{cases} 0, & si\ x \leq 0 \\ \\ \dfrac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}, & si\ x > 0 \end{cases} \quad (2.25)$$

donde $\Gamma(\alpha)$ es el valor de la función Gamma en α.[27] La función definida en 2.25 es una función de densidad de probabilidad. Obviamente es no negativa. Para comprobar que su integral sobre toda la recta es uno bastará hacer el cambio $y = x/\beta$ en la correspondiente integral

$$\int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} dx = \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty y^{\alpha-1} e^{-y} \beta^\alpha dy = \frac{1}{\Gamma(\alpha)} \Gamma(\alpha) = 1.$$

Obsérvese que la distribución exponencial de parámetro λ es un caso particular de la distribución gamma. En concreto, una exponencial de parámetro λ es una distribución gamma con parámetros 1 y $1/\lambda$.

[26] Distribución gamma

[27] La función Gamma se define como $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy$ donde $\alpha > 0$.

Observación 2.2 *Necesitaremos más adelante los valores de la función gamma cuando $\alpha = n$ o $\alpha = n + \frac{1}{2}$, n natural. Es fácil comprobar, mediante sucesivas integraciones por partes, que*

$$\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1) = (\alpha - 1)(\alpha - 2)\Gamma(\alpha - 2),$$

lo que para $\alpha = n$ da lugar a

$$\Gamma(n) = (n-1)(n-2)\ldots 2\Gamma(1).$$

Pero

$$\Gamma(1) = \int_0^\infty e^{-x}dx = 1 \quad y \quad \Gamma(n) = (n-1)!.$$

Para el caso en que $\alpha = n + \frac{1}{2}$ deberemos calcular $\Gamma(\frac{1}{2})$,

$$\Gamma\left(\frac{1}{2}\right) = \int_0^\infty e^{-x}x^{-1/2}dx = \left[x = \frac{t^2}{2}\right] = \sqrt{2}\int_0^\infty e^{-t^2/2}dt = \frac{\sqrt{2}\sqrt{2\pi}}{2} = \sqrt{\pi}. \quad (2.26)$$

La última integral en (2.26), dividida por $\sqrt{2\pi}$, es la mitad del área que cubre la función de densidad de probabilidad de la normal estándar $N(0,1)$.

2.17.1 Distribución Ji-cuadrado

[28] Distribución Ji-cuadrado

[28] Una *variable aleatoria Ji-cuadrado de parámetro* r, $X \sim \chi^2_r$, es una variable aleatoria gamma con $\alpha = r/2$ y $\beta = 2$ siendo r entero no negativo. Por tanto, es una familia de densidades de probabilidad que dependen de un solo parámetro r. Su función de densidad tiene la expresión

$$f_X(x) = \begin{cases} 0, & si\ x \leq 0 \\ \\ \dfrac{1}{\Gamma(r/2)2^{r/2}}x^{r/2-1}e^{-x/2}, & si\ x > 0. \end{cases}$$

El parámetro r es también conocido como el *número de grados de libertad* de la distribución.

2.18 Distribución beta

[29] Distribución beta

[29] Diremos que la variable aleatoria X tiene una *distribución beta de parámetros α y β*, $X \sim Be(\alpha, \beta)$, si su función de densidad es de la forma

$$f_X(x) = \begin{cases} \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}, & si\ 0 < x < 1, \\ \\ 0, & en\ el\ resto. \end{cases} \quad (2.27)$$

Proposición 2.4 *La función definida en 2.27 es una función de densidad de probabilidad.*

Demostración 2.7 *Obviamente es no negativa. Para comprobar que la integral vale uno observemos que*

$$\Gamma(\alpha)\Gamma(\beta) = \left(\int_0^\infty x^{\alpha-1}e^{-x}dx\right)\left(\int_0^\infty y^{\beta-1}e^{-y}dy\right) = \int_0^\infty\int_0^\infty x^{\alpha-1}y^{\beta-1}e^{-(x+y)}dxdy. \quad (2.28)$$

Haciendo el cambio $u = x/(x+y)$ tendremos

$$\begin{aligned}\Gamma(\alpha)\Gamma(\beta) &= \int_0^\infty\int_0^1 \frac{u^{\alpha-1}}{(1-u)^{\alpha-1}}y^{\alpha-1}y^{\beta-1}e^{-y/(1-u)}dudy = \\ &= \int_0^\infty\int_0^1 \frac{u^{\alpha-1}}{(1-u)^{\alpha-1}}y^{\alpha+\beta-1}e^{-y/(1-u)}dudy = \\ &= \int_0^\infty\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}v^{\alpha+\beta-1}e^{-v}dudv = \\ &= \int_0^\infty v^{\alpha+\beta-1}e^{-v}dv\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}du = \\ &= \Gamma(\alpha+\beta)\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}du, \quad (2.29)\end{aligned}$$

donde el segundo paso se lleva a cabo mediante el cambio $v = y/(1-u)$. En definitiva,

$$\int_0^1 f_X(x)dx = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha)\Gamma(\beta)} = 1.$$

Observación 2.3 *La función de densidad de $Be(1,1)$, teniendo en cuenta que $\Gamma(1) = \Gamma(2) = 1$, tiene por expresión,*

$$f_X(x) = \begin{cases} 1, & si\ 0 < x < 1, \\ 0, & \text{en el resto,} \end{cases}$$

que corresponde a la densidad de una variable aleatoria uniforme en el intervalo unitario [0,1].

2.19 Transformación una variable aleatoria

Con mucha frecuencia tenemos interés en el estudio de una variable y en otras cantidades que calculamos (determinísticamente) a partir de esta variable. Por ejemplo, estamos interesados en el comportamiento de la variable que nos da el peso de una persona en un estudio de salud. Supongamos que la variable aleatoria que nos da el peso de esa persona elegida al azar es X y que el peso viene dado en kilogramos. ¿Qué ocurre si en un momento dado queremos trabajar con gramos? Pues quiere decir que nuestra variable de interés pasa a ser $Y = 1000X$. Un simple cambio de unidad transforma nuestra variable de interés.

Por ejemplo, si la variable original está en grados Farenheit y queremos transformar a grados centígrados entonces la transformación que estamos considerando es $Y = (X - 32)/1.8$.

x	-2	-1	1	2
$P(X = x)$	1/4	1/4	1/4	1/4

Tabla 2.3: Función de probabilidad de la variable aleatoria X.

y	1	2
$P(Y = y)$	1/2	1/2

Tabla 2.4: Función de probabilidad de la variable aleatoria Y.

Tenemos pues una variable aleatoria X pero nuestro interés realmente se centra en una transformación $g(X)$ siendo g un función real de variable real. Con algo de notación tenemos

$$Y : \Omega \xrightarrow{X} \mathbb{R} \xrightarrow{g} \mathbb{R},$$

y el valor aleatorio que nos interesa es $Y(\omega) = g(X(\omega))$. Realmente casi cualquier g nos sirve pero para ser rigurosos hemos de pedir que g sea una función medible.[8] Si pedimos esta condición entonces $Y = g(X)$ es una variable aleatoria ya que $Y^{-1}(B) = X^{-1}(g^{-1}(B)) \in \mathbb{A}$.

Tiene sentido hablar de la distribución de probabilidad asociada a Y, P_Y. Lo inmediato es preguntarse por la relación entre las distribuciones de probabilidad de ambas variables. De hecho el problema es *determinar la distribución de la variable aleatoria Y, P_Y, conocidas la distribución de X y la función g*. La relación entre ellas es la siguiente:

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

donde B un conjunto de Borel arbitrario. En particular, si $B = (-\infty, y]$, tendremos la función de distribución de Y dada por

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \in g^{-1}((-\infty, y])). \quad (2.30)$$

La variable X que estamos transformando puede ser continua o discreta.[30] Lo primero a plantearse es: Si transformamos una variable discreta: ¿cómo es Y? Y si la variable X es continua: ¿cómo es Y?[31]

[30] O ni continua ni discreta aunque en este manual nos ocupamos de estos tipos de modelos fundamentalmente.

[31] Como diría José Luis Perales: ¿Y cómo es él?

Si la variable X es discreta entonces la Y ha de ser necesariamente discreta y la función de probabilidad de Y es

$$P(Y = y) = P(X \in g^{-1}(y)) = \sum_{x \in g^{-1}(y)} P(X = x). \quad (2.31)$$

Ejemplo 2.7 *Supongamos X uniforme discreta sobre $\{-2, -1, 1, 2\}$. Su función de probabilidad la tenemos en la tabla 2.3.*

Consideramos el módulo de X, $Y = |X|$. La función de probabilidad de Y la tenemos en la tabla 2.4.

Ejemplo 2.8 *Si X es una variable discreta con soporte D, definamos Y mediante*

$$Y = signo(X) = \begin{cases} \frac{X}{|X|}, & si\ X \neq 0 \\ 0, & si\ X = 0. \end{cases}$$

[8] Decir que g es una función medible supone que la anti-imagen de un conjunto de Borel es un conjunto de Borel: $g^{-1}(B) \in \beta$ siendo $B \in \beta$.

Con esta definición, $D_Y = \{-1, 0, 1\}$, y su función de probabilidad viene dada por

$$f_Y(y) = \begin{cases} \sum_{x<0} f_X(x), & si\ y = -1 \\ \\ f_X(0), & si\ y = 0 \\ \\ \sum_{x>0} f_X(x), & si\ y = 1 \end{cases}$$

Si la variable original X es continua entonces la variable que obtenemos transformándola, Y, puede ser bien discreta bien continua (o ni continua ni discreta). Veamos ejemplos de cómo conocer la distribución de la variable Y en cada caso.

Si lo que obtenemos es una variable Y discreta hay que intentar determinar su función de probabilidad. Suele ser el camino más simple.

Ejemplo 2.9 *Si X es una variable continua, definamos Y como*

$$Y = signo(X) = \begin{cases} \frac{X}{|X|}, & si\ X \neq 0 \\ \\ 0, & si\ X = 0. \end{cases}$$

Con esta definición, $D_Y = \{-1, 0, 1\}$, y su función de probabilidad viene dada por

$$f_Y(y) = \begin{cases} \int_{-\infty}^{0} f_X(x)dx, & si\ y = -1 \\ \\ 0, & si\ y = 0 \\ \\ \int_{0}^{+\infty} f_X(x)dx, & si\ y = 1 \end{cases}$$

Si X e $Y = g(X)$ son ambas continuas entonces una primera opción puede ser intentar calcular la función de distribución de Y.

Ejemplo 2.10 *A partir de $X \sim U(0,1)$ definimos $Y = a + (b - a)X$ con $a < b$. Vamos a obtener y reconocer la distribución de Y. Para ello veamos de obtener la función de distribución de la variable transformada.*

$$F_Y(y) = P(Y \leq y) = P\left(X \leq \frac{y-a}{b-a}\right) = \frac{y-a}{b-a}, \quad a \leq y \leq b.$$

Así pues,

$$F_Y(y) = \begin{cases} 0, & y < 0; \\ \frac{y-a}{b-a}, & a \leq y \leq b; \\ 1, & y > b. \end{cases}$$

La densidad es $f_Y(y) = \frac{1}{b-a}$ en $a \leq y \leq b$ y 0 fuera. Se trata pues de una $U(a,b)$.

Ejemplo 2.11 *Sea $X \sim U(-1,1)$, uniforme continua entre -1 y +1. Definimos $Y = X^2$. Para obtener F_Y, sea $y \in [0,1]$,*

$$F_Y(y) = P(Y \leq y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y}) = \sqrt{y}. \quad (2.32)$$

Entonces,

$$F_Y(y) = \begin{cases} 0, & si\ y < 0; \\ \\ \sqrt{y}, & si\ 0 \leq y \leq 1; \\ \\ 1, & si\ y > 1. \end{cases}$$

Seguimos con el caso en que X e Y son continuas. Tenemos un par de resultados muy generales que nos dan directamente la función de densidad de Y a partir de la densidad de la variable X. Son los teoremas 2.2 y 2.3.

Teorema 2.2 *Sea X una variable aleatoria continua (con $P(X \in D) = 1$) y sea g monótona, diferenciable con $g'(x) \neq 0$, $\forall x$. Entonces $Y = g(X)$ es una variable aleatoria con función de densidad,*

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y)) \left| \frac{dg^{-1}(y)}{dy} \right|, & si\ y \in g(D) \\ 0, & en\ el\ resto. \end{cases}$$

Demostración 2.8 *Supongamos que g es monótona creciente. Tendremos, para $y \in g(D)$,*

$$F_Y(y) = P(Y \leq y) = P(X \leq g^{-1}(y)) = F_X(g^{-1}(y)).$$

Derivando respecto de y obtendremos una función de densidad para Y,

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{dF_X(g^{-1}(y))}{dg^{-1}(y)} \frac{dg^{-1}(y)}{dy} = f_X(g^{-1}(y)) \left| \frac{dg^{-1}(y)}{dy} \right|.$$

En caso de monotonía decreciente para g,

$$F_Y(y) = P(Y \leq y) = P(X \geq g^{-1}(y)) = 1 - F_X(g^{-1}(y)).$$

El resto se obtiene análogamente.

Ejemplo 2.12 *Consideremos la variable aleatoria X cuya densidad viene dada por*

$$f_X(x) = \begin{cases} 0, & si\ x < 0, \\ \frac{1}{2}, & si\ 0 \leq x \leq 1, \\ \frac{1}{2x^2}, & si\ x > 1, \end{cases}$$

Definimos una nueva variable mediante la transformación $Y = 1/X$. La transformación cumple con las condiciones del teorema, $x = g^{-1}(y) = 1/y$ y $\frac{dg^{-1}(y)}{dy} = -\frac{1}{y^2}$, por tanto la densidad de Y vendrá dada por

$$f_Y(y) = \begin{cases} 0, & si\ y < 0, \\ \frac{1}{2}\frac{1}{y^2}, & si\ 1 \leq y < \infty, \\ \frac{1}{2(1/y)^2}\frac{1}{y^2}, & si\ 0 < y < 1, \end{cases}$$

que adecuadamente ordenado da lugar a la misma densidad que poseía X.

El teorema anterior admite la siguiente generalización.

Teorema 2.3 *Sea D el dominio de g y supongamos que admite una partición finita $D = \cup_{i=1}^n D_i$, de manera que $g_i = g_{|D_i}$, restricción de g a D_i, es estrictamente monótona, diferenciable y con derivada no nula. La densidad de $Y = g(X)$ tiene la expresión,*

$$f_Y(y) = \sum_{i:y \in g_i(D_i)} f_X(g_i^{-1}(y)) \left| \frac{dg_i^{-1}(y)}{dy} \right|. \tag{2.33}$$

Demostración 2.9 *Si $D_i = (a_i, b_i)$, $i = 1, \ldots, n$, fijemos y en el rango de g. Tendremos*

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P\left(g(X) \leq y, X \in \cup_{i=1}^{n} D_i\right) \\ &= \sum_{i=1}^{n} P\left(g(X) \leq y, X \in D_i\right) = \sum_{i=1}^{n} P\left(g_i(X) \leq y\right) \quad (2.34) \end{aligned}$$

Si $y \in g_i(D_i)$ y g_i es creciente,

$$P\left(g_i(X) \leq y\right) = P\left(a_i \leq X \leq g_i^{-1}(y)\right) = F_X(g_i^{-1}(y)) - F_X(a_i).$$

Si g_i es decreciente,

$$P\left(g_i(X) \leq y\right) = P\left(g_i^{-1}(y) \leq X \leq b_i\right) = F_X(b_i) - F_X(g_i^{-1}(y)).$$

Si $y \notin g_i(D_i)$, como $I_i = g_i(D_i)$ es un intervalo, es fácil comprobar (véase la figura 2.4) que

$$\begin{aligned} P\left(g_i(X) \leq y\right) = 0 \quad & si \quad \inf I_i > y \\ P\left(g_i(X) \leq y\right) = 1 \quad & si \quad \sup I_i < y. \end{aligned}$$

En cualquiera de los dos casos, al ser constante, su contribución a f_Y es nula. En definitiva, si sustituimos en (2.34) y derivamos respecto de las $g_i^{-1}(y)$ obtendremos (2.33).

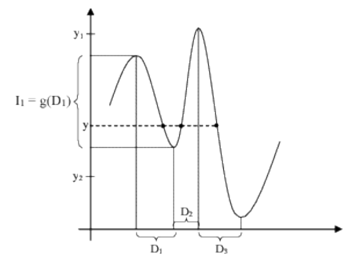


Figura 2.4: Cambio de variable

Ejemplo 2.13 *Si queremos obtener la densidad de una variable aleatoria definida mediante la transformación $Y = X^2$ a partir de $X \sim N(0,1)$, observamos en la figura 2.5 que $D = D_1 \cup D_2$, de manera que la restricción de g sobre cada D_i es una biyección que cumple las condiciones del teorema. Tenemos además que $g_1^{-1}(y) = \sqrt{y}$ y $g_2^{-1}(y) = -\sqrt{y}$. Aplicando (2.33) se obtiene*

$$f_Y(y) = \begin{cases} 0, & si\ y < 0, \\ \dfrac{1}{\sqrt{2\pi}} y^{-\frac{1}{2}} e^{\frac{-y}{2}}, & si\ y \geq 0. \end{cases}$$

Se trata de la densidad de una ji-cuadrado con un grado de libertad, χ_1^2.

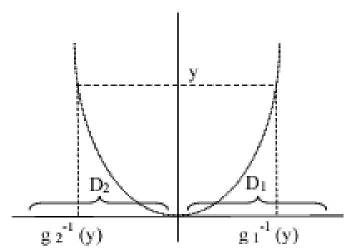


Figura 2.5: Transformación $Y = X^2$ con $X \sim N(0,1)$.

Hay dos transformaciones especialmente interesantes porque permiten obtener variables aleatorias con distribuciones preestablecidas. La primera conduce siempre a una $U(0,1)$ y la otra, conocida como *transformación integral de probabilidad*, proporciona la distribución que deseemos. Necesitamos previamente la siguiente definición.

Definición 2.11 (Inversa de una función de distribución) *Sea F una función de distribución de probabilidad en $\mathbb{R}$.* [32] *La* inversa *de F es la función definida mediante*

$$F^{-1}(x) = \inf\{t : \; F(t) \geq x\}.$$

Observemos que F^{-1} existe siempre, aun cuando F no sea continua ni estrictamente creciente. Como contrapartida, F^{-1} no es una inversa puntual de F, pero goza de algunas propiedades interesantes de fácil comprobación.

Proposición 2.5 *Sea F^{-1} la* inversa *de F. Entonces,*

[32] Esto es monótona creciente, continua por la derecha y sus límites cuando tiende a $-\infty$ y $+\infty$ son cero y uno respectivamente.

a) *para cada* x *y* t, $F^{-1}(x) \leq t \iff x \leq F(t)$,

b) F^{-1} *es creciente y continua por la izquierda, y*

c) *si* F *es continua, entonces* $F(F^{-1}(x)) = x, \ \forall x \in [0,1]$.

Podemos ya definir las dos transformaciones antes mencionadas.

Proposición 2.6 (Transformada integral de probabilidad) *Sea* $U \sim U(0,1)$, F *una función de distribución de probabilidad y definimos* $X = F^{-1}(U)$. *Entonces,* $F_X = F$.

Demostración 2.10 *Como* F^{-1} *es monótona,* X *es una variable aleatoria. Por a) en la proposición anterior,* $\forall t \in \mathbb{R}$,

$$F_X(t) = P(X \leq t) = P(F^{-1}(U) \leq t) = P(U \leq F(t)) = F(t).$$

Este resultado es la base de muchos procedimientos de simulación aleatoria porque permite obtener valores de cualquier variable aleatoria a partir de valores de una distribución uniforme. El término *generador de números aleatorios* indica un procedimiento para generar valores uniformes en el intervalo $[0,1]$. Una vez hemos generado este valor uniforme podemos generar un valor con función de distribución F considerando $X = F^{-1}(U)$.

Proposición 2.7 (Obtención de una $U(0,1)$**)** *Si* F_X *es continua,* $U = F_X(X) \sim U(0,1)$.

Demostración 2.11 *Hagamos* $F = F_X$. *Para* $x \in [0,1]$, *por la proposición 2.5 a),* $P(U \geq x) = P(F(X) \geq x) = P(X \geq F^{-1}(x))$. *La continuidad de* F *y la proposición 2.5 c) hacen el resto,*

$$P(U \geq x) = P(X \geq F^{-1}(x)) = 1 - F(F^{-1}(x)) = 1 - x.$$

2.20 Ejercicios

* **Ej. 32** — *([10, pág. 137])*Un sistema de comunicaciones consiste de n componentes, cada una de las cuales funciona independientemente y con probabilidad p. El sistema global transmite correctamente si como mínimo la mitad de sus componentes funciona. Se pide:

1. Consideremos dos sistemas, uno de cinco componentes y otro de 3 componentes. ¿Qué valores de p hacen más probable que funcione correctamente el sistema de 5 en lugar del sistema de 3?

2. En general, ¿cuándo funciona mejor un sistema con (2k+1) componentes que un sistema con (2k-1).

[9]

* **Ej. 33** — Accidentalmente, dos baterías agotadas se colocaron en un juego de cinco baterías. Para eliminar las dos baterías agotadas, las baterías se prueban una por una en orden aleatorio. La variable aleatoria X denota el número de baterías que se deben probar para encontrar las dos baterías agotadas. ¿Cuál es la función de probabilidad de X?

[9] El problema está resuelto en la referencia indicada.

* **Ej. 34** — En un proceso de fabricación de hilados se producen roturas del hilo de manera aleatoria a lo largo del tiempo. Es importante conocer cuando y cómo pueden producirse dichas roturas. Supongamos que un trabajador controla 800 husos y que la probabilidad de rotura del hilo en cada bobina, durante un cierto intervalo de tiempo τ, es $p = 0.005$. Encontrar el número de roturas más probable y la probabilidad de que se produzcan a lo sumo 10 roturas.

* **Ej. 35** — Samuel Pepy, contemporáneo de Isaac Newton, sabía que al lanzar $6n$ dados el número esperado de seises era n. A partir de este resultado deducía que los sucesos A_n={*al menos* n *seises al lanzar* 6n *dados*}, $n = 1, 2, 3$, tenían todos igual probabilidad. Isaac Newton hubo de sacarlo de su error.[10]

* **Ej. 36** — **Simulando una moneda correcta**Para simular una moneda correcta a partir de una sesgada en la que la probabilidad de obtener una cara vale $\alpha \neq 1/2$ podemos proceder como sigue. Lanzamos dos veces la moneda e interpretamos C+ como *cara* y +C como *cruz*. Si no aparece ninguno de estos dos resultados, repetimos los pares de lanzamientos hasta que aparezca alguno de ellos. Encontrar la distribución de probabilidad del número de pares de lanzamientos necesarios para poder tomar una decisión (*cara* o *cruz*) y comprobar que, en efecto, se trata de la simulación de una moneda correcta.

* **Ej. 37** — ([*13*]) 1.Determinar c cuando la variable X_1 tiene una densidad beta $f_1(x) = c\Big(x(1-x)\Big)^{1/2}$.

2.Determinar c cuando la variable X_2 tiene una densidad arco seno, $f_2(x) = c\Big(x(1-x)\Big)^{-1/2}$.

3.Determinar la función de distribución de X_2.

* **Ej. 38** — El tiempo que tardan en ser atendidos los clientes del servicio de caja de cierta sucursal bancaria es una variable aleatoria $T \sim Exp(\lambda)$, con $\lambda = 0.2$. Durante una mañana han llegado 10 clientes, ¿cuál es la probabilidad de que a lo sumo 3 de ellos hayan tardado más de 6 minutos en ser atendidos? (Suponemos que los clientes son atendidos independientemente unos de otros.)

* **Ej. 39** — ([*13*])Un motor de un avión falla con probabilidad q. Asumiendo que los fallos ocurren independientemente, determinar la probabilidad de que:

1.Al menos dos de los motores, de un total de cuatro, no fallen.

2.Al menos dos de los motores, de un total de tres, no fallen.

Comparar las probabilidades anteriores para todos los valores de q.

* **Ej. 40** — **Números desafortunados** ([*13*])Supongamos una lotería que se celebra cada semana. Sea p la probabilidad de que un número dado d sea extraído una semana cualquiera. Después de n extracciones sucesivas, sea $p(k)$ la probabilidad de que el número d

[10]El problema, que es de fácil solución y puede incluso parecer ingenuo a algún lector, se recoge aquí por su interés histórico y también porque el autor de esta colección ha tenido ocasión de comprobar que los émulos actuales de Samuel Pepy son todavía numerosos

apareciera por última vez hace k semanas. Determinar el valor de $p(k)$. [11]

* **Ej. 41** — ([13])Lanzamos un dado repetidamente hasta que obtenemos un seis. Sea A_n el suceso consistente en que el primer seis aparece en el n-ésimo lanzamiento y sea E el suceso consistente en que el número requerido de lanzamientos para obtener el primer seis es un número par. Se pide:

1. Determinar $P(E)$.
2. Determinar $p_n = P(A_n|E)$.
3. Constituyen los valores p_n una distribución de probabilidad geométrica.

* **Ej. 42** — **Muerte súbita** ([13])[12] Tenemos dos jugadores A y B que realizan una serie de pruebas independientes de modo que en cada una de ellas se puede dar una de las siguientes situaciaciones:

1. A gana con probabilidad p;
2. B gana con probabilidad q;
3. nadie gana con probabilidad 1-p-q.

El juego se detiene cuando gana uno de los dos contendientes A o B. Este el formato del juego inglés craps (o pase inglés). El procedimiento indicado también se utiliza para resolver juegos en que se llega a situaciones de empate al final del tiempo normal de juego. Son los playoffs por muerte súbita. Se pide:

1. ¿Cuál es la probabilidad a_n de que A gane en la prueba n-ésima?
2. ¿Cuál es la probabilidad α de que A gane?
3. ¿Cuál es la probabilidad $\lambda(n)$ de que el juego finalice en la prueba n?
4. Sea D_n el suceso consistente en que la duración del juego es de n pruebas y sea A_w el suceso consistente en que el jugador A es el ganador. Demostrar que D_n y A_w son independientes?

* **Ej. 43** — Sea X una variable aleatoria con distribución $Bi(n,p)$, $X \sim Bi(n,p)$. Sea Y otra variable con distribución $Bi(n,1-p)$, $Y \sim Bi(n,1-p)$. Demostrar que

$$P(X = k) = P(Y = n - k).$$

Interpreta el resultado.

** **Ej. 44** — ([13])Sea $p(k)$ la función de probabilidad de una distribución binomial $Bi(n,p)$. Demostrar que

$$(p(k))^2 \geq p(k+1)p(k-1)$$

para todo k donde

$$p(k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

[11] El problema está resuelto en la referencia indicada.

[12] Los tres primeros apartados están resueltos en la referencia indicada y el cuarto propuesto.

* **Ej. 45** — Un modelo sencillo para describir las variaciones del precio de una acción bursátil supone que un día cualquiera el precio sube 1 unidad con probabilidad p y baja 1 unidad con probabilidad $1-p$, considerando que los cambios diarios son independientes.

1. ¿Cuál es la probabilidad de que la acción conserve el precio original después de 2 días?
2. ¿Y de que haya subido 1 unidad después de 3 días?
3. Si después de 3 días la acción ha subido 1 unidad, cuál es la probabilidad de que el primer día de esos 3 la acción hubiera subido?

* **Ej. 46** — **Muestreo de aceptación** ([13])Una caja de componentes (lo que en control de calidad se le llama un lote) se recibe en una fábrica. Queremos evaluar la calidad del lote del siguiente modo. En cada lote de 100 componentes vamos a tomar al azar 10 y los evaluamos. Si tenemos como mucho una unidad defectuosa entonces aceptamos el lote. ¿Con qué probabilidad aceptamos un lote con 100 unidades de las cuales 7 son defectuosas?

* **Ej. 47** — [**La (gran) suerte de un político español**] Un (muy) famoso político español [13] tenía una gran suerte jugando a la lotería. Vamos a intentar cuantificar su suerte. Le ha tocado el primer premio en cinco sorteos distintos. Vamos a suponer que en cada uno de los sorteos hay N posibles números.

1. Nos fijamos solamente en los cinco sorteos en los que ganó el primer premio. Si suponemos que solamente compró un boleto. ¿Qué probabilidad tenía de ganar en los cinco sorteos?
2. El no jugó solamente en esos cinco sorteos. Supongamos que ha jugado en n sorteos. ¿Cuál es la probabilidad de que ganara el primer premio en cinco de ellos?
3. También es cierto que no compraría solamente un número. Supongamos que compró b números distintos en n sorteos. ¿Cuál es la probabilidad de que ganara el primer premio en cinco de ellos?
4. Supongamos $N = 10^5$ y $b = 30$, ¿cuántas veces ha de jugar para que la probabilidad de ganar cinco o más veces sea mayor que 0.1?

* **Ej. 48** — Sea X una variable aleatoria discreta con función de probabilidad definida como $f(x) = \frac{x}{10}$ para $x \in \{1, 2, 3, 4\}$(y 0 en otro caso). Calcula las siguientes probabilidades:

1. $P(1 < X < 3), P(1 \leq X < 3), P(1.5 < X < 3)$, $P(1.5 < X \leq 3), P(1.5 \leq X \leq 1.99)$.
2. $P(X \leq 1), P(X < 1), P(0 \leq X \leq 2 | 1 \leq X \leq 3)$
3. $P(X \in N)$ donde N es el conjunto de los números enteros positivos.
4. $P(X \in Q)$ donde Q es el conjunto de los números racionales.
5. Calcula y dibuja la función de distribución $F(x)$ de la variable X. Repite los dos primeros apartados del problema utilizando $F(x)$. Representa en la gráfica dibujada todas las probabilidades del primer apartado del problema.

[13] Al menos lo era a finales del siglo XX y principios del XXI.

* **Ej. 49** — Sea X una variable aleatoria continua con función de densidad $f(x) = \frac{4}{3x^2}$ para $1 \leq x \leq 4$ y 0 en otro caso. Calcula las siguientes probabilidades:

1. $P(1 < X < 3), P(1 \leq X < 3), P(1.5 < X < 3)$, $P(1.5 < X \leq 3), P(1.5 \leq X \leq 1.99)$.
2. $P(X \leq 1), P(X < 1), P(0 \leq X \leq 2|1 \leq X \leq 3)$
3. $P(X \in N)$ donde N es el conjunto de los números enteros positivos.
4. $P(X \in Q)$ donde Q es el conjunto de los números racionales.
5. Calcula y dibuja la función de distribución $F(x)$ de la variable X. Repite los dos primeros apartados del problema utilizando $F(x)$. Representa en la gráfica dibujada todas las probabilidades del primer apartado del problema.

* **Ej. 50** — Calcula la función de probabilidad de una variable aleatoria discreta X para la que $P(X > x) = \frac{1}{2^x}$ si $x \in \{1, 2, 3, \ldots\}$ y $P(X = x) = 0$ en otro caso.

* **Ej. 51** — De un lote que contiene 25 artículos, de los cuales 5 son defectuosos, se eligen 4 al azar. Sea X la variable aleatoria que cuenta el número de artículos defectuosos elegidos. Obtén la distribución de probabilidad de la variable aleatoria X si los artículos se escogen

1. con reemplazamiento,
2. sin reemplazamiento.

** **Ej. 52** — La concha del caracol de tierra Limocolaria mantersiana puede ser rayada o pálida. El 60 por ciento de estos caracoles de tierra tiene la concha rayada. Se seleccionan 10 caracoles al azar de esta población. Calcula la probabilidad de que la proporción de caracoles con concha rayada presentes en la muestra:

1. sea del 50 por cien,
2. esté entre el 50 y el 60 por cien, ambos inclusive,
3. no sea inferior al 70 por cien,
4. sea del 60 por cien.

* **Ej. 53** — Si compramos un recibo de lotería en 50 sorteos distintos y tal que la probabilidad de ganar en cada sorteo es $\frac{1}{100}$, ¿cuál es la probabilidad de cada uno de los siguientes sucesos?

1. Ganar al menos una vez.
2. Ganar exactamente una vez.
3. Ganar al menos dos veces.

* **Ej. 54** — El soporte de cierta variable continua es $[1, 5]$ y su función de densidad es proporcional a x^2.

1. ¿Cuál es esta función de densidad?
2. ¿Qué probabilidad asigna al intervalo $(1, 3]$?

* **Ej. 55** — Encuentra una función de densidad de la variable que tiene por función de distribución:

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 - (1 + 3x)e^{-3x} & \text{si } x \geq 0 \end{cases}$$

* **Ej. 56** — Si una densidad para X es $f(x) = \frac{3}{26}(x+1)^2$ para $0 < x < 2$, y 0 en el resto, encontrar la función de distribución de X.

* **Ej. 57** — Sea X una v.a. cuya función de densidad es la exponencial $f(x) = 2e^{-2x}$ per a $x > 0$. Calcular las siguientes probabilidades,

1.$P(2X + 1 > 3)$,

2.$P(X^2 \leq 7)$,

3.$P(X \leq 2 \mid X > 0.5)$,

4.$P(X > 0.5 \mid X \leq 2)$.

* **Ej. 58** — Determinar los valores de C para que las siguientes funciones sean funciones de probabilidad con soporte de la variable en los enteros positivos.

(a)$f(x) = C2^{-x}$.

(b)$f(x) = C2^{-x}/x$.

(c)$f(x) = Cx^{-2}$.

(d)$f(x) = C2^{x}/x!$.

* **Ej. 59** — Sea X una variable aleatoria cuya función de probabilidad es, sucesivamente, cada una de las consideradas en el problema 58. Calcular

(a)$P(X > 1)$.

(b)El valor más probable de X.

(c)La probabilidad de que X sea par.

** **Ej. 60** — Lanzamos n monedas, todas ellas con probabilidad de cara p. Cada moneda que muestra una cara se lanza de nuevo. Obtener la función de probabilidad del número de caras en el segundo lanzamiento.

* **Ej. 61** — Demostrar que para una variable binomial o Poisson, su función de probabilidad verifica,

$$f(k-1)f(k+1) \leq [f(k)]^2.$$

Comprobar que la misma desigualdad se cumple para $f(k) = 90/(\pi k)^4$, $k \geq 1$.

* **Ej. 62** — Determinar los valores de C para que las siguientes funciones sean funciones de densidad de probabilidad.

(a)$f(x) = C\{x(1-x)\}^{1/2}$, $0 < x < 1$.

(b)$f(x) = C\exp(-x - e^{-x})$, $x \in R$.

(c)$f(x) = C(1 + x^2)^{-m}$, $x \in R$.

* **Ej. 63** — Encontrar la densidad de $Y = aX$, $a > 0$, en función de la densidad de X. Demostrar que las variables aleatorias continuas X y $-X$ tienen la misma función de densidad si y solo si $f_X(x) = f_X(-x), \forall x \in R$.

* **Ej. 64** — Si f y g son funciones de densidad de sendas variables X e Y, comprobar que $\alpha f + (1-\alpha)g$ es también una densidad para $0 \leq \alpha \leq 1$.

2.20.1 Función de una variable aleatoria

* **Ej. 65** — Sea X una variable aleatoria con distribución exponencial de parámetro λ. Sea Y una variable aleatoria discreta definida como $Y = k$ si $k \leq X < k+1$ para $k \in \{0, 1, 2, \ldots\}$. Calcula la función de probabilidad y la función de distribución de la variable Y.

* **Ej. 66** — [9, pág. 309] Supongamos que la variable X tiene una distribución exponencial(λ). Determinar la distribución de cX cuando $c > 0$.

* **Ej. 67** — Si X es una variable aleatoria con distribución exponencial de parámetro λ, calcula la función de distribución de $Y = \frac{1}{X+1}$

* **Ej. 68** — [9, pág. 309] Demostrar que una variable aleatoria T tiene una distribución gamma con parámetros (r, λ) si y solamente si $T = T_1/\lambda$ siendo T_1 una gamma con parámetros $(r, 1)$.

* **Ej. 69** — Sea $X \sim U(0, 1)$. Calcula la función de densidad de Y, definida como $Y = X^n$.

* **Ej. 70** — Si X es un variable aleatoria con distribución uniforme en $(0, 1)$, identifica la distribución de la variable aleatoria $Y = -\frac{1}{\lambda} lnX$ siendo $\lambda > 0$.

* **Ej. 71** — [9, pág. 309] Supongamos que U tiene una distribución uniforme en $(0, 1)$. Determinar la densidad de U^2.

* **Ej. 72** — [9, pág. 309] Supongamos que X tiene una distribución uniforme en $(-1, 1)$. Determinar la densidad de $Y = X^2$.

* **Ej. 73** — [9, pág. 309] Supongamos que X tiene una distribución uniforme en $[-1, 2]$. Determinar la densidad de $Y = X^2$.

** **Ej. 74** — [Distribución Weibull] [9, pág. 310] Se pide

1. Demostrar que si T sigue una distribución Weibull(λ, α) con densidad
$$f(t) = \lambda \alpha t^{\alpha-1} e^{-\lambda t^{\alpha}} \;\; t > 0$$
donde $\lambda > 0$ y $\alpha > 0$ entonces T^{α} sigue una distribución exponencial(λ). Observar el caso particular en que $\alpha = 1$.
2. Demostrar que si U es uniforme en $(0, 1)$ entonces $T = (-\lambda^{-1} \log U)^{\frac{1}{\alpha}}$ tiene una distribución Weibull(λ, α).

* **Ej. 75** — [9, pág. 310] Supongamos Z una variable aleatoria normal estándar. Determinar la expresión de las densidades de las siguientes variables

1.$|Z|$,
2.Z^2,
3.$\frac{1}{Z}$,
4.$\frac{1}{Z^2}$.

* **Ej. 76** — Si X es una variable aleatoria con distribución uniforme en $(0,1)$, calcula:

1.La función de densidad de $Y = tan\{\pi(X - \frac{1}{2})\}$
2.La función de distribución de $Z = ln(\frac{X}{1-X})$

* **Ej. 77** — Sea X una variable aleatoria cuya densidad viene dada por

$$f_X(x) = \begin{cases} 0, & \text{si } x < 0, \\ \frac{1}{2}, & \text{si } 0 \leq x \leq 1, \\ \frac{1}{2x^2}, & \text{si } x > 1. \end{cases}$$

Encontrar la densidad de $Y = 1/X$.

* **Ej. 78** — Si X es una variable aleatoria continua con función de densidad f_X, calcular la función de densidad de $Y = X^2$.

* **Ej. 79** — Sea X una variable aleatoria continua no negativa con función de densidad f. Sea $Y = X^n$, calcula la función de densidad de Y, f_Y.

* **Ej. 80** — Sea X un variable aleatoria continua, cuya función de densidad es f_X. Hallar la función de densidad de $Y = |X|$.

* **Ej. 81** — La variable aleatoria X tiene como función de densidad,

$$f(x) = \begin{cases} 3(1-x)^2, & 0 < x < 1; \\ 0, & \text{en el resto.} \end{cases}$$

Obtener la función de densidad de $Y = (1 - X)^3$.

* **Ej. 82** — Sea X una variable aleatoria cuya función de densidad es f_X. Encuentra la función de densidad de la variable aleatoria Y, definida por $Y = aX + b$.

* **Ej. 83** — Si X es una variable aleatoria exponencial con parametro $\lambda = 1$, halla la función de densidad de la variable aleatoria Y definida por $Y = logX$.

* **Ej. 84** — Se traza aleatoriamente una linea recta a través del punto $(0,l)$. Encontrar la densidad de probabilidad de la abcisa en el origen, X, de dicha recta.

* **Ej. 85** — Sea λ el número de células por unidad de área en una preparación celular. Puede demostrarse que la distancia, D, de una célula a su vecino más próximo tiene por densidad de probabilidad,

$$f_D(d) = \begin{cases} 2\pi\lambda d e^{-\lambda\pi d^2}, & d > 0 \\ 0, & \text{en el resto.} \end{cases}$$

El área del mayor disco libre (sin contactos) centrado en una célula es $A = \pi D^2$. Encontrar la densidad de probabilidad de A.

* **Ej. 86** — La función de densidad de la variable aleatoria continua X viene dada por $f_X(x) = c(x + \sqrt{x})$ para $0 < x < 1$ y $f(x) = 0$ en el resto. ¿Cuánto vale c? Obtener la densidad de probabilidad de *1/X*.

* **Ej. 87** — ([10])Determinar la función de densidad de la variable $Y = e^X$ cuando X es una variable con distribución normal con media μ y varianza σ^2. La variable Y se dice que tiene una distribución lognormal ya que $\log Y$ tiene una distribución normal.

** **Ej. 88** — Sea X una variable aleatoria con densidad

$$f_X(x) = \begin{cases} \theta e^{-\theta x}, & x > 0,\ \theta > 0; \\ 0, & \text{en el resto.} \end{cases}$$

Obtener la distribución de probabilidad de $Y = \sin X$.

* **Ej. 89** — Un fabricante de cerveza artesanal *Xequebó* suministra una vez a la semana a sus clientes. El consumo en miles de litros de cerveza a la cervecería *GlupGlup* es una variable aleatoria con densidad de probabilidad

$$f_X(x) = \begin{cases} c(5-x)^4, & 0 \le x \le 5; \\ 0, & \text{en el resto.} \end{cases}$$

1. ¿Qué ha de valer c para que $f_X(x)$ sea una función de densidad?
2. ¿Cuál ha de ser la capacidad del depósito de cerveza para que una semana cualquiera la probabilidad de vaciarse sea 0.01?

** **Ej. 90** — Supongamos $Y \sim U(0, 5)$. ¿Qué probabilidad tenemos de que las raíces de la ecuación $4x^2 + 4Yx + Y + 2 = 0$ sean ambas números reales?

2.21 Soluciones a los ejercicios

Solución (Ej. 3) — 1.0.2, 0.3, 0.2, 0.2, 0.

2.0.1, 0, 0.5, 0.5, 1.

3.1.

4.1.

Solución (Ej. 4) —

$$f(x) = \left(\frac{1}{2}\right)^x$$

para $x = 1, 2, \dots$

Solución (Ej. 23) —

$$\int_0^a cx dx = c\frac{a^2}{2} = 1,$$

de donde $c = \frac{2}{a^2}$.

Solución (Ej. 29) — La probabilidad de aceptación de una bola viene dada por $p_A = P(d_1 < D < d_2)$, siendo la de rechazo $p_R = 1 - p_A$.

$$\begin{aligned} p_A &= P\left(\tfrac{d_1-(d_1+d_2)/2}{(d_2-d_1)/4} < Z < \tfrac{d_2-(d_1+d_2)/2}{(d_2-d_1)/4}\right) \\ &= P\left(\tfrac{(d_1-d_2)/2}{(d_2-d_1)/4} < Z < \tfrac{(d_2-d_1)/2}{(d_2-d_1)/4}\right) \\ &= P(-2 < Z < 2) = 2\Phi(2) - 1, \end{aligned}$$

de donde

$$p_R = 1 - p_A = 1 - (2\Phi(2) - 1) = 2 - 2\Phi(2) = 0.04550$$

Solución (Ej. 30) — 1.$e^{-10\lambda}$.

2.Es 0.4605

Solución (Ej. 33) — El soporte de X es $D_X = \{2, 3, 4\}$. $P(X = 2) = \frac{1}{10}$, $P(X = 3) = \frac{3}{10}$, $P(X = 4) = \frac{6}{10}$.

Solución (Ej. 45) — 1.$2p(1-p)$.

2.$3p^2(1-p)$.

3.$\frac{2}{3}$.

Solución (Ej. 49) — (a)$\frac{8}{9}, \frac{8}{9}, \frac{4}{9}, \frac{4}{9}, 0.21887$.

(b)0, 0, $\frac{3}{4}$.

(c)0.

(d)0.

Solución (Ej. 50) — (a)$\frac{3}{124}x^2$ si $1 < x < 5$.

(b)$\frac{26}{124}$.

Solución (Ej. 51) — Si es con con reemplazamiento tenemos una distribución binomial, $X \sim Bi(4, 1/5)$. Cuando lo hacemos sin reemplazamiento tenemos una distribución hipergeométrica, $X \sim H(4, 25, 5)$.

Solución (Ej. 52) — Como la proporción es $X = N/10$, donde $N \sim Bi(10, 0.4)$, los resultados se obtienen fácilmente a partir de la distribución binomial.

Solución (Ej. 53) — 1. 0.39499

2. 0.305586

3. 0.0894079

Solución (Ej. 55) — $9xe^{-3x}$ para $x > 0$ y 0 en otro caso.

Solución (Ej. 56) — 0 para $x < 0$, $\frac{1}{26}((x+1)^3 - 1)$ para $0 \leq x \leq 2$ y 1 si $x > 2$.

Solución (Ej. 57) — (a)0.1353

(b)0.9950

(c)0.9502

(d)0.3561

Solución (Ej. 58) — (a)$C = 1$.

(b)Sabemos que $\log \frac{1}{1-x} = \sum_{k\geq 1} \frac{x^k}{k}$ y de aquí $C = (\log 2)^{-1}$.

(c)La serie $\{k^{-2}, k \geq 1\}$ es una de las series de Euler y suma $\pi^2/6$ y por tanto $C = 6/\pi^2$.

(d)$C = (e^2 - 1)^{-1}$.

Solución (Ej. 59) — (a)$\frac{1}{2}$; $1-(2\log 2)^{-1}$; $(1-6\pi^{-2}$; $(e^2-3)/(e^2-1)$.

(b)1; 1; 1; 1 y 2.

(c)$P(X\ par) = \sum_{k\geq 1} P(X = 2k)$ y vale $1/3$; $1-(\log 3)/(\log 4)$; $1/4$.

(d) Para la ultima de las funciones de probabilidad nos podemos ayudar de la igualdad

$$\frac{1}{2}\sum_{n\geq 1}\frac{2^n+(-2)^n}{n!}=\sum_{n\geq 1}\frac{2^{2n}}{(2n)!}=\frac{1}{2}(e^2+e^{-2}-2),$$

con lo que

$$P(X\ par)=\frac{e^2+e^{-2}-2}{2(e^2-1)}=0,4323.$$

Solución (Ej. 60) — El resultado es $\binom{n}{k}p^{2k}(1-p^2)^{n-k}$. ¿Por qué?

Solución (Ej. 62) — (a) $C=1/\pi$;

(b) $C=1$;

(c) Al sustituir $v=(1+x^2)^{-1}$ se obtiene la densidad de una $Be(1/2,m-1/2)$ y por tanto $C=\Gamma(m)/[\Gamma(1/2)\Gamma(m-1/2)]$.

Solución (Ej. 63) — $f_Y(y)=f_x(y/a)/a$.

Solución (Ej. 64) — Si $0\leq\alpha\leq 1$, $f\geq 0$ y $g\geq 0$, entonces $\alpha f+(1-\alpha)g\geq 0$. Se comprueba fácilmente que $\int \alpha f+(1-\alpha)g=1$.

Solución (Ej. 65) — $f_Y(k)=P(Y=k)=e^{-\lambda k}(1-e^{-\lambda})$ para $k\in\{0,1,2,\ldots\}$.

Solución (Ej. 67) —

$$F_Y(y)=\begin{cases}0, & y<0;\\ e^{-\lambda(\frac{1-y}{y})}, & 0\leq y\leq 1;\\ 1, & y>1.\end{cases}$$

Solución (Ej. 70) — $Y\sim Exp(\lambda)$.

Solución (Ej. 76) — (a) $f_Y(y)=\frac{1}{\pi(1+y^2)}$, $y\in R$. Conocida como la distribución de cauchy.

(b)

$$F_Z(z)=\frac{e^z}{1+e^z},\quad z\in\mathbf{R}.$$

Solución (Ej. 81) — Y es uniforme en $[0,3]$.

Solución (Ej. 86) — $c = \frac{6}{7}$. La densidad es

$$f_Y(y) = \frac{6}{7}\left(\frac{1}{y^3} + \frac{1}{y^2\sqrt{y}}\right), y > 1,$$

y vale cero en el resto.

Solución (Ej. 89) — 1.$c = \frac{1}{5^4}$.

2.$5 - (32.25)^{\frac{1}{5}}$. La capacidad del depósito de ser de 2996.9 litros.

Solución (Ej. 90) — $\frac{3}{5}$.

Capítulo 3

Vector aleatorio

3.1 Un ejemplo (discreto)

Empecemos con un buen ejemplo que motive lo que sigue.

Ejemplo 3.1 *Supongamos un experimento consistente en lanzar 4 veces una moneda correcta. Sea X el numero de caras en los 3 primeros lanzamientos y sea Y el número de caras en los 3 últimos lanzamientos. Son dos variables que nos describen parcialmente el resultado del experimento. Son dos variables que* observamos simultáneamente. *Además son variables que por su definición sabemos que tienen relación entre ellas. De algún modo conocida una sabemos algo sobre la otra. Por ello, tiene sentido estudiar su comportamiento probabilístico conjunto, lo que llamaremos su distribución conjunta. Si las consideramos simultáneamente tendremos un vector (X,Y) que será* aleatorio *porque sus componentes son variables aleatorias. Se trata de un vector aleatorio que toma valores en un conjunto discreto. En concreto el vector toma valores en $D = \{0,1,2,3\} \times \{0,1,2,3\}$, es decir, el conjunto de valores que puede tomar es un conjunto finito. De un modo análogo al caso de distribuciones discretas podemos plantearnos la probabilidad de que cada uno de los resultados elementales $(x,y) \in D = \{0,1,2,3\} \times \{0,1,2,3\}$. Es claro que $P(X \in D) = 1$ y por ello se habla de vector aleatorio discreto o de distribución discreta. En un experimento de este tipo la herramienta fundamental para trabajar con el vector aleatorio es conocer las probabilidades $P(X = x, Y = y)$, o* función de probabilidad conjunta.[1] *La función de probabilidad conjunta de este vector con distribución discreta aparece en la tabla 3.1.*

Dentro de la tabla tenemos la función de probabilidad conjunta $P(X = x, Y = y)$ pero en la última columna a la derecha tenemos la función de probabilidad marginal. Se obtiene sumando las probabilidades de la fila correspondiente. ¿Por qué? Aplicando la fórmula de la probabilidad total tenemos que

$$P(Y = y) = \sum_{x=0}^{3} P(X = x, Y = y).$$

De un modo similar tenemos en la última fila de la tabla la función de la probabilidad que obtenemos sumando los elementos de la columna.

[1] Menos utilizada pero también correcta es la expresión de *función de cuantía conjunta*.

Tabla 3.1: Lanzamiento de cuatro monedas. X nos indica el número de caras en los tres primeros lanzamientos e Y el número de caras en los tres últimos lanzamientos. Dentro de la tabla tenemos la función de probabilidad conjunta $P(X = x, Y = y)$. En la última columna tenemos la función de probabilidad marginal de Y, $P(Y = y)$. La última fila muestra la función de probabilidad marginal de X, $P(X = x)$.

	X				
Y	**3**	**2**	**1**	**0**	$\boldsymbol{P(Y=y)}$
0	0	0	1/16	1/16	2/16
1	0	2/16	3/16	1/16	6/16
2	1/16	3/16	2/16	0	6/16
3	1/16	1/16	0	0	2/16
$\boldsymbol{P(X=x)}$	2/16	6/16	6/16	2/16	1

Otra vez esto es consecuencia de la fórmula de probabilidad total de la que deducimos

$$P(X = x) = \sum_{y=0}^{3} P(X = x, Y = y).$$

Lo importante es darse cuenta de que si conocemos la función de probabilidad conjunta podemos obtener las funciones de probabilidad marginales. Sin embargo, el recíproco no es cierto. Del conocimiento de las funciones de probabilidad marginales no podremos obtener la función de probabilidad conjunta.

En el ejemplo anterior hemos descrito el experimento no con una sola variable sino con dos variables simultáneamente observadas. En definitiva, dado el resultado del experimento, dado ω, hemos considerado dos variables aleatorias $X_1(\omega)$ y $X_2(\omega)$.

Esta es la opción habitual. Raramente cuando se realiza un experimento aleatorio nos conformamos con observar una sola variable. Pretendemos una descripción más completa de lo que ha pasado. En definitiva, más cantidades asociadas al resultado obtenido.

3.2 Algunas definiciones necesarias

En los ejemplos que acabamos de ver esencialmente *asociamos varias variables aleatorias* simultáneamente a un mismo resultado aleatorio ω. Con una cierta formalidad lo que hacemos es pasar de Ω al espacio $\mathbb{R}^k$. En $\mathbb{R}^k$ hemos de considerar qué sucesos son los relevantes. Hemos de considerar una *buena* familia de sucesos. Sabemos que ha de ser una σ-álgebra. Cuando trabajamos con varias variables aleatorias lo que nos importa fundamentalmente es conocer las siguientes probabilidades

$$P(a_1 < X_1 \leq b_1, \ldots, a_k < X_k \leq b_k)$$

donde los a_i, b_i son números reales con $a_i < b_i$. Por ello, la σ-álgebra que consideremos en $\mathbb{R}^k$ ha de contener a los rectángulos

$$(a, b] = \prod_{i=1}^{k} (a_i, b_i],$$

donde $a = (a_1, \ldots, a_k]$ y $b = (b_1, \ldots, b_k]$. La σ-álgebra más pequeña en $\mathbb{R}^k$ que contiene a los rectángulos es la σ-álgebra de Borel.

Definición 3.1 (σ-álgebra de Borel) *La σ-álgebra de Borel en el espacio $\mathbb{R}^k$ es la mínima σ-álgebra que contiene a los* rectángulos $(a, b]$ *con $a, b \in \mathbb{R}^k$.*

¿Qué entendemos por vector aleatorio?

Definición 3.2 (Vector aleatorio) *Un* vector aleatorio*, $X = (X_1, X_2, \ldots, X_k)$, es una aplicación de un espacio de probabilidad $(\Omega, \mathbb{A}, P)$ en el espacio medible $(\mathbb{R}^k, \beta^k)$, que verifica*

$$X^{-1}(B) \in \mathbb{A}, \quad \forall B \in \beta^k.$$

Tenemos una probabilidad en el espacio medible $(\Omega, \mathbb{A})$. A partir de esta medida de probabilidad podemos definir (inducir) una probabilidad sobre $(\mathbb{R}^k, \beta^k)$.

Definición 3.3 (Probabilidad inducida o distribución) *Dado un vector aleatorio definido de un espacio de probabilidad $(\Omega, \mathbb{A}, P)$ en $(\mathbb{R}^k, \beta^k)$, se define la probabilidad inducida a partir de P mediante X como la medida de probabilidad definida como*

$$P_X(B) = P(X^{-1}(B)), \ \forall B \in \beta^k. \tag{3.1}$$

Es sencillo comprobar que verifica los tres axiomas que definen una probabilidad, por lo que la terna $(\mathbb{R}^k, \beta^k, P_X)$ constituye un espacio de probabilidad con las características de $(\Omega, \mathbb{A}, P)$ heredadas a través de X. Llamaremos a la medida de probabilidad P_X la *distribución* del vector X.

** **Ej. 1** — Demostrar que la función de conjunto definida en 3.1 es una medida de probabilidad.

3.3 Funciones de distribución conjunta y marginales

Definición 3.4 (Función de distribución conjunta) *Dado un vector aleatorio $X = (X_1, X_2, \ldots, X_k)$ definimos su función de distribución asociada (o también la función de distribución asociada a la distribución P_X) en el punto $x = (x_1, \ldots, x_k)$ de $\mathbb{R}^k$ mediante*

$$F_X(x) = F_X(x_1, \ldots, x_k) = P_X(S_x) = P(X \in S_x) = P\left(\bigcap_{i=1}^{k} \{X_i \leq x_i\}\right) \tag{3.2}$$

siendo $S_x = \prod_{i=1}^{k} (-\infty, x_i]$, a los que denominaremos región suroeste de x.[2]

[2] Obviamente contiene a los puntos que están situados al suroeste de x cuando estamos en $\mathbb{R}^2$.

De la definición se derivan las siguientes propiedades:

Proposición 3.1 *La función de distribución definida en 3.4 verifica las siguientes propiedades.*

1. *Es no negativa.*
2. *Es monótona creciente en cada componente.*
3. *Es continua desde arriba: si $x_n \downarrow x$ entonces $F_X(x_n) \downarrow F_X(x)$.*
4. $\lim_{\forall x_i \uparrow +\infty} F_X(x_1, \ldots, x_k) = 1$ *y* $\lim_{\exists x_i \downarrow -\infty} F(x_1, \ldots, x_k) = 0$.

Demostración 3.1 No negatividad.- *Consecuencia inmediata de ser una probabilidad.*

Monotonía en cada componente.- *Si $x \leq y$, es decir, $x_i \leq y_i$, $i = 1, \ldots, k$, $S_x \subset S_y$ y $F_X(x) \leq F_X(y)$.*

Continuidad conjunta por la derecha.- *Si $x^{(n)} \downarrow x$, entonces $F_X(x^{(n)}) \downarrow F_X(x)$,*

Valores límites.- *Al tender a $\pm\infty$ las componentes del punto, se tiene*

$$\lim_{\forall x_i \uparrow +\infty} F_X(x_1, \ldots, x_k) = 1,$$

o bien,

$$\lim_{\exists x_i \downarrow -\infty} F(x_1, \ldots, x_k) = 0.$$

Las propiedades que acabamos de ver en la proposición 3.1 no son más que la versión multidimensional de las propiedades ya conocidas para el caso unidimensional. Existe ahora una quinta propiedad sin la cual no sería posible recorrer el camino inverso, obtener la distribución P_X a partir de F_X y establecer la deseada y conveniente equivalencia entre ambos conceptos.

Supongamos el caso bidimensional, $k = 2$, y consideremos el rectángulo $(a, b] = (a_1, b_1] \times (a_2, b_2]$ tal como lo muestra la figura 3.1.

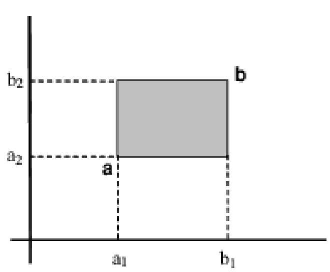


Figura 3.1: Rectángulo $(a, b]$.

Indudablemente $P_X((a, b]) \geq 0$, pero teniendo en cuenta (3.2) podemos escribir,

$$P_X((a, b]) = F_X(b_1, b_2) - F_X(b_1, a_2) - F_X(a_1, b_2) + F_X(a_1, a_2) \geq 0.$$

Este resultado es cierto, obviamente, para cualquier k y necesitamos introducir el operador diferencia a fin de obtener una representación sencilla del resultado en el caso general. Para $a_i \leq b_i$, Δ_{a_i, b_i} representa el operador diferencia que actua de la forma

$$\begin{aligned} &\Delta_{a_i, b_i} F_X(x_1, \ldots, x_k) = \\ &\quad = F_X(x_1, \ldots, x_{i-1}, b_i, \ldots, x_k) - F_X(x_1, \ldots, x_{i-1}, a_i, \ldots, x_k). \end{aligned}$$

Sucesivas aplicaciones del mismo conducen a

$$\Delta_{a,b} F_X(x) = \Delta_{a_1, b_1} \ldots \Delta_{a_k, b_k} F_X(x_1, \ldots, x_k) = P_X((a, b]), \qquad (3.3)$$

lo que permite generalizar el anterior resultado mediante la expresión,

Proposición 3.2 *Se verifica que*

1. $\Delta_{a,b} F_X(x) \geq 0$, $\forall a, b \in \mathbb{R}^k$, *con* $a_i \leq b_i$.

2. $P_X((a,b]) = \Delta_{a,b}F_X(x)$.

De hecho también se puede probar el siguiente resultado.

Proposición 3.3

$$P_X((a,b]) = \sum_{x\in V}(-1)^{sign(x)}F(x), \tag{3.4}$$

siendo $V = \{x = (x_1,\ldots,x_n) : x_i \in \{a_i,b_i\}\} = \prod_{i=1}^n\{a_i,b_i\}$, el conjunto de los vértices del rectángulo $(a,b]$ y $sign(x)$ el número de a_i's (o extremos inferiores de los lados del rectángulo)

Damos los dos resultados previos sin prueba.

Funciones de distribución marginales

Si el vector es aleatorio cabe pensar que sus componentes también lo serán. La siguiente proposición establece una primera relación entre el vector y sus componentes.

Proposición 3.4 *$X = (X_1,\ldots,X_k)$ es un vector aleatorio* sí y solo sí *cada una de sus componentes es una variable aleatoria.*

Demostración 3.2 *Recordemos que la condición de variable aleatoria le viene a una aplicación por el hecho de que las anti-imágenes de conjuntos de Borel son sucesos, están en la σ-álgebra de nuestro espacio de probabilidad original,*

$$X^{-1}(B) \in \mathbb{A},\ B \in \beta. \tag{3.5}$$

Existe un resultado que permite caracterizar esta propiedad utilizando una familia menor de conjuntos, evitándonos así la prueba para cualquier elemento de β. Basta, según dicha caracterización, comprobar la propiedad en una familia que engendre a la σ-álgebra. Pues bien, los intervalos de la forma $(a,b]$, en $\mathbb{R}$, y los conjuntos suroeste, S_x, en $\mathbb{R}^k$, engendran β y β^k, respectivamente. Ahora ya podemos abordar la demostración.

Supongamos que las componentes de $X = (X_1,X_2,\ldots,X_k)$ son variables aleatorias. Para $S_x = \prod_{i=1}^k(-\infty,x_i]$ podemos escribir

$$\{X \in S_x\} = \bigcap_{i=1}^k\{X_i \le x_i\}. \tag{3.6}$$

Si cada componente es aleatoria, $\{X_i \le x_i\} \in \mathbb{A}$, $\forall i$, y X será un vector aleatorio.

Para demostrar el inverso observemos que

$$\{X_j \le x_j\} = \lim_{x_i\uparrow+\infty,\quad i\neq j}\{X \le S_x\},$$

donde para conseguir la numerabilidad los x_i han de tender a $+\infty$ a través de una sucesión, lo que puede conseguirse haciendo $x_i = n$, $i \neq j$. En consecuencia, X_j es medible, pues al tratarse de una sucesión monótona creciente de conjuntos, su límite es la unión de todos ellos que es una operación estable en $\mathbb{A}$.

Si las componentes del vector son variables aleatorias tendrán asociadas sus correspondientes probabilidades inducidas y funciones de distribución. Hay que modificar la nomenclatura que hemos utilizado hasta el momento. Hablaremos de *conjunta* y *marginal*. Puesto que P_X y F_X describen el comportamiento conjunto de las componentes de X, nos referiremos a ellas como *distribución conjunta* y *función de distribución conjunta* del vector X, respectivamente. Cuando, en el mismo contexto, necesitemos referirnos a la distribución de alguna componente lo haremos aludiendo a la *distribución marginal* o a la *función de distribución marginal* de X_i.

La pregunta que surge de inmediato es, *¿qué relación existe entre la distribución conjunta y las marginales?* Estamos en condiciones de dar respuesta en una dirección: cómo obtener la distribución marginal de cada componente a partir de la conjunta. Para ello, basta tener en cuenta que

$$\lim_{x_j \overset{j\neq i}{\to} \infty} \bigcap_{j=1}^{k} \{X_j \leq x_j\} = \{X_i \leq x_i\},$$

y al tomar probabilidades obtendremos

$$F_{X_i}(x_i) = \lim_{x_j \overset{j\neq i}{\to} \infty} F_X(x_1, \ldots, x_k). \tag{3.7}$$

El concepto de *marginalidad* podemos aplicarlo a cualquier subvector del vector original. Así, para $l \leq k$, si $X^l = (X_{i_1}, \ldots, X_{i_l})$ es un subvector de X, podemos hablar de la *distribución conjunta marginal de* X^l, para cuya obtención a partir de la conjunta procederemos de forma análoga a como acabamos de hacer para una componente. Si en la relación (3.6) fijamos $x_{i_1}, \ldots, x_{i_l}$ y hacemos tender a $+\infty$ el resto de componentes, tendremos

$$\{X^l \in S_{x^l}\} = \lim_{x_i \overset{i\neq i_1,\ldots,i_l}{\longrightarrow} \infty} \{X \in S_x\}.$$

Relación que nos permite obtener la función de distribución marginal conjunta de $X^l = (X_{i_1}, \ldots, X_{i_l})$ sin más que tomar probabilidades,

$$F_{X^l}(x_{i_1}, \ldots, x_{i_l}) = \lim_{x_i \overset{i\neq i_1,\ldots,i_l}{\longrightarrow} \infty} F_X(x_1, \ldots, x_k).$$

3.4 Distribución conjunta discreta

Consideremos un vector aleatorio que toma un número finito o infinito numerable de posibles valores. Siendo más formales, existe un conjunto D ($\subset \mathbb{R}$) numerable tal que

$$P(X \in D) = 1.$$

Este conjunto recibe el nombre de soporte de la distribución del vector P_X ya que

$$P_X(D) = 1.$$

Un vector aleatorio cuya distribución verifica lo anterior se dice *vector aleatorio discreto*. Notemos que si X es discreto también lo es cada variable que lo compone o cualquier subvector de X que consideremos. En este tipo de vectores aleatorios tenemos obviamente que $P(X \in D^c) = 0$ y si conocemos las probabilidades $P(X = x)$ para $x \in D$

entonces podemos conocer la probabilidad de cualquier otro suceso ya que

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

Por ello es fundamental la siguiente función.

Definición 3.5 (Función de probabilidad) *Se define la* función de cuantía o probabilidad conjunta *en* $x = (x_1, \ldots, x_k) \in \mathbb{R}^k$ *como*

$$f_X(x_1, \ldots, x_k) = \begin{cases} P(X_i = x_i,\ i = 1, \ldots, k), & si\ x = (x_1, \ldots, x_k) \in D \\ 0, & en\ el\ resto. \end{cases}$$

Proposición 3.5 *Una función de probabilidad verifica que es no negativa y* $\sum_{x \in D} f_X(x) = 1$.

Demostración 3.3 1. *Al tratarse de una probabilidad,* $f_X(x) \geq 0,\ \forall x \in \mathbb{R}^k$,

2. *Como* $P(X \in D) = 1$,

$$\sum_{x \in D} f_X(x) = P(X \in D) = 1.$$

Entre F_X y f_X se establecen relaciones similares a las del caso unidimensional:

$$F_X(x) = \sum_{y \leq x,\ y \in D} f_X(y_1, y_2, \ldots, y_k),$$

y

$$f_X(x_1, x_2, \ldots, x_k) = F_X(x_1, x_2, \ldots, x_k) - F_X(x_1-, x_2-, \ldots, x_k-).$$

Función de probabilidad marginal

Si el vector aleatorio X es discreto también lo serán cada una de sus componentes. Si por D_i designamos el soporte de $X_i,\ i = 1, \ldots, k$, se verifica,

$$\{X_i = x_i\} = \bigcup_{x_j \in D_j,\ j \neq i} \left(\bigcap_{j=1}^{k} \{X_j = x_j\} \right),$$

siendo disjuntos los elementos que intervienen en la unión. Al tomar probabilidades tendremos

$$f_{X_i}(x_i) = \sum_{x_j \in D_j,\ j \neq i} f_X(x_1, \ldots, x_k),$$

que permite obtener la función de cuantía marginal de la X_i a partir de la conjunta. La marginal conjunta de cualquier subvector aleatorio se obtendría de manera análoga, extendiendo la suma sobre todas las componentes del subvector complementario,

$$f_{X^l}(x_{i_1}, \ldots, x_{i_l}) = \sum_{x_j \in D_j,\ j \neq i_1, \ldots, i_l} f_X(x_1, \ldots, x_k).$$

Ejemplo 3.2 (Una prueba con k posibles resultados) *Supongamos que tenemos un experimento en que cuando lo realizamos nos pasa una y solo una cosa de k posibles. De un modo más preciso, tenemos k sucesos mutuamente excluyentes y cuya unión cubre todo el espacio muestral* Ω*. Si estos sucesos los denotamos* $\{A_1, \ldots, A_k\}$ *constituyen una partición de* Ω*. Decimos que ocurre el resultado i si al realizar el experimento se tiene* $\omega \in A_i$*.*[33] *Si denotamos* $p_i = P(A_i)$ *entonces podemos considerar el vector aleatorio* $(X_1, \ldots, X_k)$ *donde* $X_i(\omega) = 1_{A_i}(\omega)$*. Es un vector donde la función de probabilidad conjunta sería*

$$P(X_1 = x_1, \ldots, X_k = x_k) = p_1^{x_1} \ldots p_k^{x_k}, \tag{3.8}$$

con $x_i = 0$ *ó 1 y cero en cualquier otro caso. El soporte de la distribución sería* $D = \{0, 1\}^k$*. De hecho como* $\sum_{i=1}^k X_k = 1$ *tenemos que todos los* x_i *valen cero salvo uno que toma el valor 1.*

[33] Ejemplos clásicos es preguntar a un votante a qué partido político va a votar incluyendo como opciones votar en blanco, no votas y el no sabe no contesta. Otro ejemplo, puede ser responder una pregunta en un cuestionario con respuesta múltiple.

Ejemplo 3.3 (Distribución multinomial) [34] *La versión k-dimensional de la distribución binomial es la distribución* multinomial*. Aparece esta distribución en experimentos aleatorio en el que nos interesamos en la ocurrencia de alguno de k sucesos,* $A_1, A_2, \ldots, A_k$*, que constituyen una partición de* Ω *(ocurre uno y solamente uno de los sucesos* A_i*). Si* $P(A_i) = p_i$ *y repetimos n veces el experimento de manera que las repeticiones son independientes, el vector* $X = (X_1, \ldots, X_k)$*, con* $X_i =$número de ocurrencias de A_i*, decimos que tiene una* distribución multinomial*,* $X \sim M(n; p_1, p_2, \ldots, p_k)$*.* [35] *La función de probabilidad conjunta viene dada por*

[34] Este ejemplo continúa el ejemplo 3.2. Repetimos el experimento allí considerado n veces en las mismas condiciones e independientemente.

[35] Los mismos ejemplos considerados en el ejemplo 3.2 son válidos. Simplemente allí considerábamos una réplica y aquí estamos considerando n réplicas realizadas independientemente unas de otras.

$$P(X_1 = n_1, \ldots, X_k = n_k) = \frac{n!}{n_1! n_2! \ldots n_k!} \prod_{i=1}^k p_i^{n_i}, \tag{3.9}$$

si $0 \leq n_i \leq n$*, con* $i = 1, \ldots, k$ *y* $\sum n_i = n$ *y vale cero en cualquier otro caso.*

Observemos que es una función de probabilidad pues es no negativa y al sumarla para todos los posibles $n_1, n_2, \ldots, n_k$ *obtenemos el desarrollo del polinomio* $(p_1 + p_2 + \ldots + p_k)^n$*, de suma 1 porque los* A_i *constituían una partición de* Ω*. La marginal de* X_i *la observemos con*

$$\frac{n!}{n_1! n_2! \ldots n_k!} \prod_{i=1}^k p_i^{n_i} = \binom{n}{n_i} p_i^{n_i} \frac{(n - n_i)!}{n_1! \ldots n_{i-1}! n_{i+1}! \ldots n_k!} \prod_{j \neq i} p_j^{n_j},$$

y al sumar para el resto de componentes,

$$\begin{aligned} P(X_i = n_i) &= \binom{n}{n_i} p_i^{n_i} \sum \frac{(n - n_i)!}{n_1! \ldots n_{i-1}! n_{i+1}! \ldots n_k!} \prod_{j \neq i} p_j^{n_j}, \\ &= \binom{n}{n_i} p_i^{n_i} (p_1 + \ldots + p_{i-1} + p_{i+1} + \ldots + p_k)^{n - n_i} \\ &= \binom{n}{n_i} p_i^{n_i} (1 - p_i)^{n - n_i}, \end{aligned}$$

llegamos a la conclusión que

$$X_i \sim Bi(n, p_i),$$

como era de esperar, pues al fijar X_i *sólo nos interesamos por la ocurrencia de* A_i *y el experimento que llevamos a cabo puede ser descrito mediante un modelo binomial.*

Ejemplo 3.4 (Distribución binomial negativa bivariante) *Supongamos que A_1, A_2 y A_3 constituyen una partición del espacio muestral Ω, de manera que $P(A_i) = p_i$, $i = 1, 2, 3$ con $p_1 + p_2 + p_3 = 1$. Realizamos un experimento cuyo resultado es, necesariamente, A_1, A_2 o A_3. Repetimos el experimento de manera independiente en cada ocasión hasta que el suceso A_3 haya ocurrido r veces. Ello significa que A_1 habrá ocurrido x veces y A_2 lo habrá hecho en y ocasiones, todas ellas previas a la r-ésima ocurrencia de A_3. La probabilidad de un suceso de estas características vendrá dada por*

$$\binom{x+y+r-1}{x}\binom{y+r-1}{y}p_1^x p_2^y (1-p_1-p_2)^r = \frac{(x+y+r-1)!}{x!y!(r-1)!}p_1^x p_2^y (1-p_1-p_2)^r. \quad (3.10)$$

Podemos ahora definir un vector aleatorio binomial negativo bivariante $(X,Y) \sim BN(r; p_1, p_2)$ *como aquél que tiene por función de probabilidad,*

$$f_{XY}(x,y) = \frac{(x+y+r-1)!}{x!y!(r-1)!}p_1^x p_2^y (1-p_1-p_2)^r, \text{ si } (x,y) \in D_{xy} = [0,1,2,\ldots]^2,$$

y 0 en cualquier otro caso.

Se trata de una función de probabilidad por cuanto $f_{XY}(x,y) \geq 0$, $\forall (x,y) \in \mathbb{R}^2$ y, teniendo en cuenta que

$$(1-p_1+p_2)^{-r} = \sum_{D_{xy}} \binom{x+y+r-1}{x}\binom{y+r-1}{y}p_1^x p_2^y,$$

también verifica

$$\sum_{D_{xy}} f_{XY}(x,y) = (1-p_1+p_2)^r (1-p_1+p_2)^{-r} = 1.$$

La marginal de cualquiera de las componentes, por ejemplo Y, la obtendremos a partir de $f_Y(y) = \sum_{D_x} f_{XY}(x,y)$, con $D_x = \{0,1,2,\ldots\}$,

$$\begin{aligned} f_Y(y) &= \sum_{D_x} \binom{x+y+r-1}{x}\binom{y+r-1}{y}p_1^x p_2^y (1-p_1-p_2)^r \\ &= \binom{y+r-1}{y}p_2^y (1-p_1-p_2)^r \sum_{D_x} \binom{x+y+r-1}{x}p_1^x, \end{aligned}$$

pero recordemos que $(1-x)^{-n} = \sum_{j\geq 0}\binom{n+j-1}{j}x^j$, por tanto,

$$\begin{aligned} f_Y(y) &= \binom{y+r-1}{y}p_2^y (1-p_1-p_2)^r (1-p_1)^{-(y+r)} \\ &= \binom{y+r-1}{y}\left(\frac{p_2}{1-p_1}\right)^y \left(\frac{1-p_1-p_2}{1-p_1}\right)^r \\ &= \binom{y+r-1}{y}p^y (1-p)^r. \end{aligned}$$

Luego $Y \sim BN(r,p)$ con $p = p_2/(1-p_1)$.

3.5 Distribución conjunta continua

Un vector aleatorio se dice continuo o que su distribución de probabilidad es continua cuando las probabilidades vienen dadas por integrales. Supongamos una función f sobre $\mathbb{R}^k$ no negativa e integrable (cuya integral vale uno) y un vector X que verifica

$$P_X((a,b]) = P(X \in (a,b]) = P(a_i < X_i \leq b_i,\ i = 1,\ldots,k) = \int_{a_k}^{b_k} \cdots \int_{a_1}^{b_1} f_X(x_1,\ldots,x_k)\,dx_1 \ldots dx_k, \quad (3.11)$$

para cualesquiera $a, b \in \mathbb{R}^k$ con $a_i \leq b_i$. Entonces decimos que X es un *vector aleatorio continuo* o que la distribución de probabilidad P_X es una *distribución continua*.

Para poder utilizar una función f para calcular probabilidades tiene que verificar las dos propiedades siguientes.

Definición 3.6 (Función de densidad conjunta) *1. $f_X(x)$ es no negativa.*

2. Como $P(X \in \mathbb{R}^k) = 1$ entonces

$$\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f_X(x_1,\ldots,x_k) dx_1 \ldots dx_k = 1.$$

Como consecuencia de esta definición, entre la función de distribución conjunta y la de densidad de probabilidad conjunta se establecen las siguientes relaciones:

$$F_X(x_1,\ldots,x_k) = P_X(S_x) = \int_{-\infty}^{x_k} \cdots \int_{-\infty}^{x_1} f(t_1,\ldots,t_k)\,dt_1 \ldots dt_k, \quad (3.12)$$

y si $x \in \mathbb{R}^k$ es un punto de continuidad de f_X,

$$f_X(x_1,\ldots,x_k) = \frac{\partial^k F_X(x_1,\ldots,x_k)}{\partial x_1 \ldots \partial x_k}. \quad (3.13)$$

Funciones de densidad marginales

Para obtener la densidad marginal de X_i a partir de la función de la densidad conjunta tengamos en cuenta (3.7) y (3.12) y que integración y paso al límite pueden permutarse por ser la densidad conjunta integrable,

$$F_{X_i}(x_i) = \lim_{x_j \overset{j \neq i}{\to} \infty} F_X(x_1,\ldots,x_k) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{x_i} \cdots \int_{-\infty}^{+\infty} f(t_1,\ldots,t_i,\ldots,t_k)\,dt_1 \ldots dt_i \ldots dt_k. \quad (3.14)$$

Pero la derivada de F_{X_i} es la densidad de X_i y como las propiedades de la densidad conjunta permiten también intercambiar derivación e integración se sigue que

$$f_{X_i}(x_i) = \int_{\mathbb{R}^{k-1}} f(t_1,\ldots,x_i,\ldots,t_k)\,dt_1 \ldots dt_{i-1} dt_{i+1} \ldots dt_k. \quad (3.15)$$

Para el caso de un subvector, X^l, la densidad conjunta se obtiene de forma análoga,

$$f_{X^l}(x_{i_1}, \ldots, x_{i_l}) = \\ = \int_{\mathbb{R}^{k-l}} f_X(t_1, \ldots, t_k)\, dt_{i_1} \ldots dt_{i_l} \prod_{j \neq i_1, \ldots, i_l} dt_j.$$

Ejemplo 3.5 (Uniforme en el cuadrado unitario) *Supongamos que tomamos un punto al azar en el cuadrado unitario* $[0, 1] \times [0, 1]$. *Su densidad conjunta sería constante sobre este cuadrado y nula fuera de él. Como, además, ha de integrar uno en el cuadrado necesariamente su densidad conjunta será*

$$f_{XY}(x, y) = 1, \text{ si } (x, y) \in [0, 1] \times [0, 1],$$

y cero en el resto. Obtengamos la marginal de la primera componente, X.

$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x, y) dy = \int_0^1 dy = 1,$$

si $0 \leq x \leq 1$ *y cero en otro caso. Es decir, que la marginal de la variable* X *es una distribución uniforme en el intervalo* $[0, 1]$. *El mismo razonamiento nos lleva a que la variable* Y *también es uniforme en el intervalo* $[0, 1]$.

Ejemplo 3.6 (Uniforme en el disco unitario) *Consideremos el disco centrado en el origen y de radio unidad,* $B(0, 1)$. *Al elegir un punto al azar en este disco* $B(0, 1)$ *podemos definir sobre el correspondiente espacio de probabilidad un vector aleatorio de las coordenadas del punto,* (X, Y). *La elección al azar implica una probabilidad uniforme sobre* $B(0, 1)$, *lo que se traduce en un densidad conjunta constante sobre todo el círculo, pero como por otra parte* $\int_{B(0,1)} f(x, y)\, dxdy = 1$, *la densidad conjunta vendrá dada por*

$$f_{XY}(x, y) = \begin{cases} \dfrac{1}{\pi}, & \text{si } (x, y) \in B(0, 1) \\ \\ 0, & \text{en el resto.} \end{cases}$$

Para obtener la densidad marginal de X,

$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x, y)\, dy = \frac{1}{\pi} \int_{-\sqrt{1-x^2}}^{+\sqrt{1-x^2}} dy = \frac{2}{\pi} \sqrt{1 - x^2},$$

por lo que

$$f_X(x) = \begin{cases} \dfrac{2}{\pi} \sqrt{1 - x^2}, & \text{si } |x| \leq 1 \\ \\ 0, & \text{en el resto,} \end{cases}$$

La marginal de Y, *por simetría, tiene la misma expresión.*

Ejemplo 3.7 *La función de densidad conjunta del vector aleatorio bidimensional* (X, Y) *viene dada por*

$$f_{XY}(x, y) = \begin{cases} 8xy, & \text{si } 0 \leq y \leq x \leq 1 \\ 0, & \text{en el resto.} \end{cases}$$

Si queremos obtener las marginales de cada componente, tendremos para X

$$f_X(x) = \int_0^x f_{XY}(x,v)dv = \int_0^x 8xvdv = 4x^3, \quad 0 \leq x \leq 1,$$

y cero en el resto. Para Y,

$$f_Y(y) = \int_y^1 f_{XY}(u,y)du = \int_y^1 8uydu = 4y(1-y^2), \quad 0 \leq y \leq 1,$$

y cero en el resto.

Obtengamos ahora la función de distribución conjunta, $F_{XY}(x,y)$*. Observemos para ello el gráfico en figura 3.2. La función de densidad es distinta de 0 en la región A por lo que* $F_{XY}(x,y) = 0$ *si* $x \leq 0$ *o* $y \leq 0$.

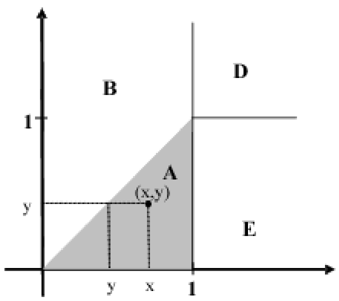


Figura 3.2: Ejemplo 3.7.

Si $(x,y) \in A$,

$$F_{XY}(x,y) = \int\int_{S_{xy}\cap A} f_{XY}(u,v)dudv,$$

pero $S_{xy} \cap A = \{(u,v); 0 \leq v \leq u \leq y\} \cup \{(u,v); y \leq u \leq x, \ 0 \leq v \leq y\}$*, por tanto*

$$F_{XY}(x,y) = \int_0^y \left[\int_0^u 8uvdv\right] du + \int_y^x \left[\int_0^y 8uvdv\right] du = y^2(2x^2 - y^2). \tag{3.16}$$

Si $(x,y) \in B$*, como* $f_{XY}(x,y) = 0$ *en* B*, el rectángulo superior de la figura 3.3 (en negro) no acumula probabilidad y, por tanto,*

$$F_{XY}(x,y) = P(S_{xy} \cap A) = P(S_{xx} \cap A).$$

Así pues,

$$F_{XY}(x,y) = \int_0^x \left[\int_0^u 8uvdv\right] du = x^4. \tag{3.17}$$

Observemos que (3.17) puede obtenerse a partir de (3.16) haciendo $y = x$*. En efecto, de acuerdo con (3.7), (3.17) no es más que* $F_X(x)$*, la función de distribución marginal de* X*. Si* $(x,y) \in E$*, un razonamiento similar conduce a*

$$F_{XY}(x,y) = y^2(2-y^2),$$

que no es más que la función de distribución marginal de Y*, que podríamos haber obtenido haciendo* $x = 1$ *en (3.16). Por último, si* $(x,y) \in D$, $F_{XY}(x,y) = 1$*. Para su obtención basta hacer* $x = 1$ *e* $y = 1$ *en (3.16). Resumiendo*

$$F_{XY}(x,y) = \begin{cases} 0, & si\ x \leq 0\ o\ y \leq 0; \\ y^2(2x^2 - y^2), & si\ (x,y) \in A; \\ x^4, & si\ (x,y) \in B; \\ y^2(2-y^2), & si\ (x,y) \in E; \\ 1, & si\ (x,y) \in D. \end{cases}$$

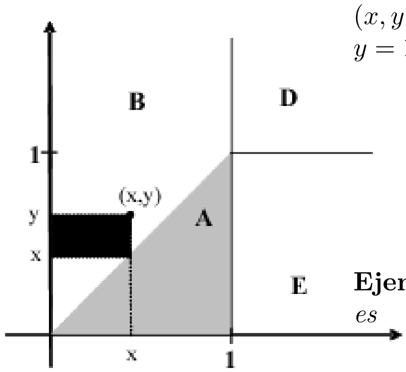


Figura 3.3: Ejemplo 3.7.

Ejemplo 3.8 *La función de densidad conjunta del vector* (X_1, X_2, X_3) *es*

$$f(x_1,x_2,x_3) = \begin{cases} \dfrac{48x_1x_2x_3}{(1+x_1^2+x_2^2+x_3^2)^4}, & si\ x_1,x_2,x_3 \geq 0 \\ 0, & en\ el\ resto. \end{cases}$$

Obtengamos en primer lugar las densidades marginales de cada componente. Dada la simetría de la densidad conjunta bastará con obtener una cualquiera de ellas.

$$\begin{aligned} f_1(x_1) &= \int_0^\infty \int_0^\infty \frac{48x_1x_2x_3}{(1+x_1^2+x_2^2+x_3^2)^4} dx_2 dx_3 \\ &= 48x_1 \int_0^\infty x_2 \left[\int_0^\infty \frac{x_3}{(1+x_1^2+x_2^2+x_3^2)^4} dx_3 \right] dx_2 \\ &= \int_0^\infty \int_0^\infty \frac{8x_1x_2}{(1+x_1^2+x_2^2)^3} dx_2 \qquad (3.18) \\ &= \frac{2x_1}{(1+x_1^2)^2}. \end{aligned}$$

Luego

$$f_i(x_i) = \begin{cases} \dfrac{2x_i}{(1+x_i^2)^2}, & si\ x_i \geq 0 \\ 0, & en\ el\ resto. \end{cases} \qquad (3.19)$$

Por otra parte, en el transcurso de la obtención de $f_i(x_i)$ *el integrando de (3.18) es la densidad marginal conjunta de* (X_1, X_2)*, que por la simetría antes mencionada es la misma para cualquier pareja de componentes. Es decir, para* $i, j = 1, 2, 3$

$$f_{ij}(x_i, x_j) = \begin{cases} \dfrac{8x_ix_j}{(1+x_i^2+x_j^2)^3}, & si\ x_i, x_j \geq 0 \\ 0, & en\ el\ resto. \end{cases}$$

Para obtener la función de distribución conjunta recordemos que

$$F(x_1,x_2,x_3) = P(X_i \leq x_i, i=1,2,3) = \\ P\left(\bigcap_{i=1}^{3} \{X_i \leq x_i\}\right) = \int_0^{x_1} \int_0^{x_2} \int_0^{x_3} f(u,v,z) du dv dz, \qquad (3.20)$$

pero en este caso será más sencillo recurrir a esta otra expresión,

$$F(x_1,x_2,x_3) = 1 - P\left(\left[\bigcap_{i=1}^{3} \{X_i \leq x_i\}\right]^c\right) = 1 - P\left(\bigcup_{i=1}^{3} A_i\right),$$

con $A_i = \{X_i > x_i\}$*. Si aplicamos la fórmula de inclusión-exclusión (1.1),*

$$F(x_1,x_2,x_3) = \\ 1 - \left[\sum_{i=1}^{3} P(A_i) - \sum_{1 \leq i < j \leq 3} P(A_i \cap A_j) + P(A_1 \cap A_2 \cap A_3)\right]. \qquad (3.21)$$

La obtención de las probabilidades que aparecen en (3.21) involucran a las densidades antes calculadas. Así, para $P(A_i)$

$$P(A_i) = P(X_i > x_i) = \int_{x_i}^\infty \frac{2u}{(1+u^2)^2} du = \frac{1}{1+x_i^2}.$$

Para $P(A_i \cap A_j)$,

$$P(A_i \cap A_j) = P(X_i > x_i, X_j > x_j) =$$
$$\int_{x_i}^{\infty}\int_{x_j}^{\infty} \frac{8uv}{(1+u^2+v^2)^3}dudv =$$
$$\frac{1}{1+x_i^2+x_j^2}. \qquad (3.22)$$

Finalmente

$$P(A_1 \cap A_2 \cap A_3) =$$
$$\int_{x_1}^{\infty}\int_{x_2}^{\infty}\int_{x_3}^{\infty} ds\frac{48uvz}{(1+u^2+v^2+z^2)^4}dudvdz = \frac{1}{1+x_1^2+x_2^2+x_3^2}. \qquad (3.23)$$

Sustituyendo en (3.21) tendremos,

$$F(x_1,x_2,x_3) = 1-\sum_{i=1}^{3}\frac{1}{1+x_i^2}+\sum_{1\le i<j\le 3}\frac{1}{1+x_i^2+x_j^2}-\frac{1}{1+x_1^2+x_2^2+x_3^2}.$$

Ejemplo 3.9 *Elegimos un punto al azar sobre el triángulo* T *de vértices* (0,0), (1,0), (0,2) *(figura 3.4).*

Para encontrar la función de distribución conjunta del vector de sus componentes, (X,Y), *observemos la figura y las distintas posiciones del punto. Como la masa de probabilidad está uniformemente repartida sobre el triángulo puesto que la elección del punto es al azar, tendremos que*

$$P((X,Y)\in A) = \frac{|A\cap T|}{|T|},$$

donde $|B|$ *es el área de* B. *Aplicado a la función de distribución dará lugar a*

$$F_{XY}(x,y) = P((X,Y)\in S_{xy}) = |S_{xy}\cap T|, \qquad (3.24)$$

puesto que el área del triángulo vale 1.

Aplicando (3.24) obtenemos

$$F_{XY}(x,y) = \begin{cases} 0, & \text{si } x\le 0 \text{ o } y \le 0; \\ xy, & \text{si } (x,y) \text{ es del tipo } \mathbf{1}\ ; \\ xy-(x+y/2-1)^2, & \text{si } (x,y) \text{ es del tipo } \mathbf{2}\ ; \\ 2x-x^2, & \text{si } (x,y) \text{ es del tipo } \mathbf{3}\ ; \\ y-y^2/4, & \text{si } (x,y) \text{ es del tipo } \mathbf{4}\ ; \\ 1, & \text{si } x\ge 1 \text{ e } y\ge 2; \end{cases}$$

Observemos que las expresiones de $F_{XY}(x,y)$ *correspondientes a puntos del tipo* ***3*** *y* ***4*** *dependen solamente de* x *e* y, *respectivamente. Si recordamos la obtención de la función de distribución marginal veremos que se corresponden con* $F_X(x)$ *y* $F_Y(y)$, *respectivamente.*

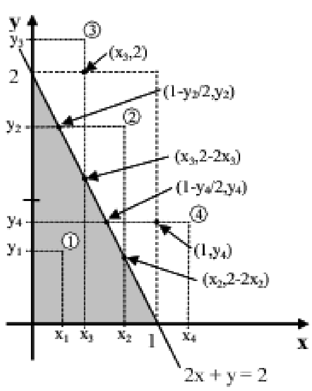


Figura 3.4: Ver ejemplo 3.9.

Distribución normal bivariante

El vector aleatorio bidimensional (X,Y) tiene una distribución normal bivariante de parámetros $\mu_X \in \mathbb{R}$, $\mu_y \in \mathbb{R}$, $\sigma_x > 0$, $\sigma_y > 0$ y ρ, $|\rho|<1$, si su función de densidad conjunta es de la forma,

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}e^{-\frac{q(x,y)}{2}}, \quad (x,y)\in\mathbb{R}^2, \qquad (3.25)$$

donde

$$q(x,y) = \frac{1}{1-\rho^2}\left\{\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right\}. \tag{3.26}$$

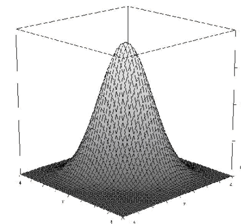


Figura 3.5: Densidad de la normal bivariante con parámetros $\mu_x = \mu_y = 0$, $\sigma_x = \sigma_y = 1$ y $\rho = 0,5$. La gráfica está centrada en (μ_x, μ_y) (parámetros de posición) y su forma depende de σ_x, σ_y y ρ (parámetros de forma). Para ver el efecto de este último la gráfica de la derecha ha sido rotada $-90°$.

En la figura 3.5 tenemos un par de ejemplos de esta función de densidad.

Proposición 3.6 *La función f_{XY} definida en 3.25 es una función de densidad de probabilidad. Además las densidades marginales f_X y f_Y corresponden a las densidades de normales univariantes con parámetros (μ_X, σ_X^2) y (μ_Y, σ_Y^2) respectivamente.*

Demostración 3.4 *Para ver que $f_{XY}(x,y)$ es una densidad es inmediato comprobar que verifica la primera condición, en cuanto a la segunda, $\int_{\mathbb{R}^2} f_{XY}(x,y)dxdy = 1$, observemos que*

$$\begin{aligned}(1-\rho^2)q(x,y) &= \\ &\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 = \\ &\left[\left(\frac{x-\mu_x}{\sigma_x}\right) - \rho\left(\frac{y-\mu_y}{\sigma_y}\right)\right]^2 + (1-\rho^2)\left(\frac{y-\mu_y}{\sigma_y}\right)^2.\end{aligned} \tag{3.27}$$

El primer sumando de (3.27) puede escribirse

$$\begin{aligned}\left(\frac{x-\mu_x}{\sigma_x}\right) - \rho\left(\frac{y-\mu_y}{\sigma_y}\right) &= \\ \left(\frac{x-\mu_x}{\sigma_x}\right) - \frac{\rho\sigma_x}{\sigma_x}\left(\frac{y-\mu_y}{\sigma_y}\right) &= \\ \frac{1}{\sigma_x}\left[x - \left(\mu_x + \rho\sigma_x\frac{y-\mu_y}{\sigma_y}\right)\right] &= \frac{1}{\sigma_x}(x-b),\end{aligned} \tag{3.28}$$

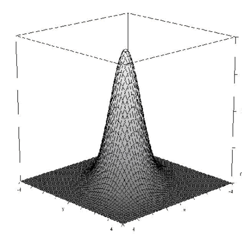


Figura 3.6: Rotación de $-90°$ de la gráfica en figura 3.5.

con $b = \mu_x + \rho\sigma_x\frac{y-\mu_y}{\sigma_y}$. Sustituyendo en (3.27)

$$(1-\rho^2)q(x,y) = \left(\frac{x-b}{\sigma_x}\right)^2 + (1-\rho^2)\left(\frac{y-\mu_y}{\sigma_y}\right)^2$$

y de aquí

$$\begin{aligned}&\int_{\mathbb{R}^2} f_{XY}(x,y)dxdy = \\ &\int_{-\infty}^{+\infty}\frac{1}{\sigma_y\sqrt{2\pi}}e^{\left(-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2\right)}\left[\int_{-\infty}^{+\infty}\frac{1}{\sigma_x\sqrt{2\pi(1-\rho^2)}}e^{\left(-\frac{1}{2(1-\rho^2)}\left(\frac{x-b}{\sigma_x}\right)^2\right)}dx\right]dy = 1,\end{aligned} \tag{3.29}$$

porque el integrando interior es la función de densidad de una $N(b, \sigma_x^2(1-\rho^2))$ e integra la unidad. La integral resultante vale también la unidad por tratarse de la densidad de una $N(\mu_y, \sigma_y^2)$, que es precisamente la densidad marginal de Y (basta recordar la expresión (3.15) que permite obtener la densidad marginal a partir de la conjunta). Por simetría $X \sim N(\mu_x, \sigma_x^2)$.

Esta distribución puede extenderse a n dimensiones, hablaremos entonces de *normal multivariante*. La expresión de su densidad la daremos en el próximo capítulo y utilizaremos una notación matricial que la haga más sencilla y compacta.

3.6 Independencia de variables aleatorias

Hemos definido previamente qué entendemos por sucesos independientes. Dos sucesos A y B son independientes cuando el conocimiento de que uno de ellos se ha producido no modifica la probabilidad del otro. El suceso A tiene probabilidad $P(A)$. Supongamos que *sabemos* que se ha producido B (asumiendo que $P(B) > 0$) entonces la *nueva* probabilidad de A es $P(A|B)$. Si la ocurrencia de B no nos da información sobre la ocurrencia de A entonces A y B son *independientes*. Vimos que esto se traducía en que ha de verificarse la igualdad siguiente

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

También vimos que, si los sucesos A y B son independientes, entonces también lo son A y B^c, A^c y B, A^c y B^c. Dicho de otra forma, son independientes las dos clases $\{A, A^c, \Omega, \emptyset\}$ y $\{B, B^c, \Omega, \emptyset\}$.

Vamos a repetir lo mismo que hemos dicho en el párrafo anterior pero utilizando variables aleatorias. Conocer si un suceso A se ha producido es equivalente a conocer el valor de la variable aleatoria $X(\omega) = 1_A(\omega)$.[3] Si conocemos X conocemos si el resultado ω está en el suceso A esto es sabemos si A se ha producido o no. Tenemos que

$$P(A) = P(X = 1).$$

Si consideramos $Y(\omega) = 1_B(\omega)$ entonces

$$P(B) = P(Y = 1).$$

Además

$$P(A \cap B) = P(X = 1, Y = 1).$$

Y finalmente que los sucesos A y B sean independientes es equivalente a afirmar que

$$P(X = 1, Y = 1) = P(A \cap B) = P(A)P(B) = P(X = 1)P(Y = 1).$$

Pero también tenemos

$$P(X = 1, Y = 0) = P(A \cap B^c) = P(A)P(B^c) = P(X = 1)P(Y = 0),$$

$$P(X = 0, Y = 1) = P(A^c \cap B) = P(A^c)P(B) = P(X = 0)P(Y = 1),$$

y (para ir terminando)

$$P(X = 0, Y = 0) = P(A^c \cap B^c) = P(A^c)P(B^c) = P(X = 0)P(Y = 0),$$

De un modo resumido de la independencia de los sucesos A y B se sigue que

$$P(X = x, Y = y) = P(X = x)P(Y = y)$$

con $x, y = 0, 1$ (notemos que las variables aleatorias X e Y solamente pueden tomar los valores 0 y 1).

[3] Recordemos que $1_A(\omega) = 1$ si $\omega \in A$ y cero en otro caso.

Consideremos ahora dos variables aleatorias discretas X e Y arbitrarias. ¿Qué puede significar que las dos variables son independientes? Siguiendo el ejemplo de los sucesos, el concepto que queremos definir es que conocer el valor de una de ellas no modifica nuestro conocimiento de la otra. Supongamos que *sabemos* que $X = x$. Para cada posible valor de la variable Y la probabilidad de que tome un valor arbitrario y *sin saber nada más* es $P(Y = y)$. Si *conocemos* el valor de X entonces la misma probabilidad vendría dada por $P(Y = y|X = x)$. Si X no contiene información sobre Y lo que debe darse es que

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

que es equivalente a

$$P(X = x, Y = y) = P(X = x)P(Y = y).$$

Llegamos a la misma condición que veíamos para variables dicotómicas o indicatrices. Parece razonable la siguiente definición.

Definición 3.7 (Independencia de dos variables discretas) *Dos variables aleatorias discretas X e Y definidas sobre un mismo espacio de probabilidad $(\Omega, \mathbb{A}, P)$ se dicen independientes cuando*

$$P(X = x, Y = y) = P(X = x)P(Y = y), \forall x, y \in \mathbb{R}. \tag{3.30}$$

Si consideramos la definición 3.7 es clara la siguiente afirmación: dos sucesos aleatorios A y B son independientes si y solo si las variables aleatorias $X(\omega) = 1_A(\omega)$ e $Y(\omega) = 1_B(\omega)$ son independientes. De hecho, estamos enriqueciendo el concepto de independencia.

Ejemplo 3.10 (Lanzamos una moneda cuatro veces) *Lanzamos cuatro veces una moneda. Consideremos las variables X_1, número de caras en los dos primeros lanzamientos; X_2, número de caras en los dos últimos lanzamientos. ¿Son X_1 y X_2 independientes entre sí? Fácilmente vemos que sí porque se verifica la definición 3.7. Modificamos la definición de las variables de modo que X_1, número de caras en los tres primeros lanzamientos; X_2, número de caras en los tres últimos lanzamientos. ¿Son independientes estas variables? No. Por ejemplo:*

$$P(X_1 = 3, X_2 = 3) = \frac{1}{2^4},$$

sin embargo,

$$P(X_1 = 3) = P(X_2 = 3) = \frac{1}{2^3},$$

En general, podemos definir independencia de una colección finita de variables aleatorias discretas.

Definición 3.8 (Independencia de variables aleatorias discretas) *Una colección variables aleatorias discretas $X_1, \ldots, X_n$ definidas sobre un mismo espacio de probabilidad $(\Omega, \mathbb{A}, P)$ se dicen independientes cuando*

$$P(X_1 = x_1, \ldots, X_n = x_n) = P(X_1 = x_1) \ldots P(X_n = x_n), \tag{3.31}$$

para cualesquiera $x_1, \ldots, x_n \in \mathbb{R}$.

¿Y cuando tenemos una variables aleatorias continuas? Damos la definición directamente en el caso general de un vector de dimensión k. Si el vector $(X_1, \ldots, X_k)$ es un vector aleatorio continuo con densidad conjunta $f(x_1, \ldots, x_k)$ de un modo análogo se tiene la siguiente definición.

Definición 3.9 *Supongamos un vector aleatorio continuo $\boldsymbol{X} = (X_1, \ldots, X_k)$ con densidad conjunta f y marginales f_i con $i = 1, \ldots, k$. Las variables $X_1, \ldots, X_k$ se dicen independientes cuando*

$$f(x_1, \ldots, x_k) = f_1(x_1) \ldots f_k(x_k), \tag{3.32}$$

para cualesquiera $x_1, \ldots, x_k \in \mathbb{R}$.

Ejemplo 3.11 (Uniforme en el cuadrado unitario) *Seguimos con el ejemplo 3.5 de un punto al azar en el cuadrado unitario $[0,1] \times [0,1]$. Su densidad conjunta es*

$$f_{XY}(x,y) = 1, \ \textit{si } (x,y) \in [0,1] \times [0,1],$$

y cero en el resto y sus marginales

$$f_X(x) = f_Y(x) = 1,$$

si $0 \leq x \leq 1$ y cero en otro caso. Tanto X como Y tienen densidades uniformes en $[0,1]$. Es obvio que $f(x,y) = f_X(x) f_Y(y)$ para cualquier x e y. Son por tanto variables independientes. Es un resultado esperable. Si vemos la forma del conjunto en donde la densidad conjunta no se anula el conocimiento del valor de x no nos dice nada sobre el valor de y. Siempre y toma valores entre 0 y 1 con densidad uniforme.

Sin embargo, si consideramos un círculo parece que la cosa no debiera de ser así.

Ejemplo 3.12 *Sea (X,Y) con densidad uniforme en círculo unidad, $B(0,1)$. La densidad conjunta es*

$$f_{XY}(x,y) = \begin{cases} \dfrac{1}{\pi}, & \textit{si } (x,y) \in B(0,1) \\ \\ 0, & \textit{en el resto.} \end{cases}$$

Por simetría, las marginales de X e Y son idénticas y tienen la forma,

$$f_X(x) = \begin{cases} \dfrac{2}{\pi}\sqrt{1-x^2}, & \textit{si } |x| \leq 1 \\ \\ 0, & \textit{en el resto.} \end{cases}$$

De inmediato se comprueba que $f_{XY}(x,y) \neq f_X(x) f_Y(y)$ y ambas variables no son independientes.

Las definiciones 3.8 y 3.9 son operativas. Con ellas podemos comprobar si un vector discreto o continuo está compuesto por variables independientes. Sin embargo, son definiciones limitadas. La definición general es la siguiente.

Definición 3.10 (Variables aleatorias independientes) *Decimos que las variables $X_i, \ i = 1, \ldots, k$ son* independientes *si*

$$P(X_1 \in B_1, \ldots, X_n \in B_n) = P(X_1 \in B_1) \ldots P(X_n \in B_n), \tag{3.33}$$

para cualesquiera $B_1, \ldots, B_n \in \beta$.

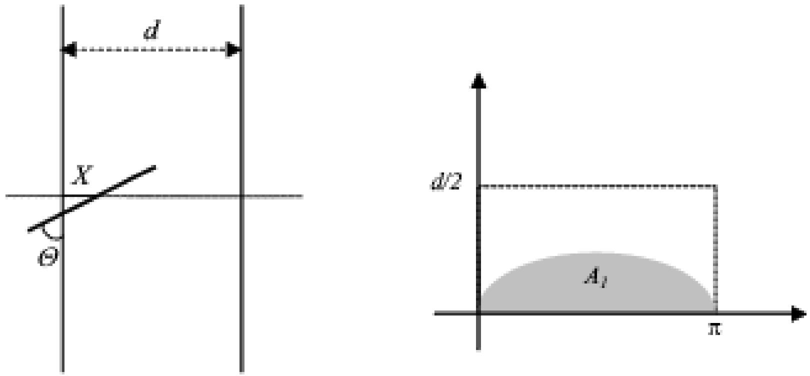


Figura 3.7: Problema de la aguja de Buffon.

Comprobar la independencia de variables aleatorias utilizando la definición 3.10 no es posible. Sin embargo, se puede demostrar que, para variables discretas las definiciones 3.8 y 3.10 equivalen. Análogamente, para variables continuas, las definiciones 3.9 y 3.10 son equivalentes.

Ejemplo 3.13 (La aguja de Buffon) [36]

Sobre una trama de líneas paralelas equidistantes entre sí d unidades, lanzamos al azar una aguja de longitud l unidades, con $l<d$. ¿Cuál es la probabilidad de que la aguja corte alguna de las paralelas?

Si denotamos por X, la distancia del centro de la aguja a la paralela más próxima, y por Θ, el ángulo que la aguja forma con dicha paralela, la posición de la aguja sobre el entramado de paralelas queda unívocamente determinada con ambas variables. **El lanzamiento al azar cabe interpretarlo en el sentido que ambas variables tienen distribuciones uniformes y son independientes**. *En concreto, $X \sim U(0,d)$ y $\Theta \sim U(0,\pi)$ (por razones de simetría, la elección de un ángulo en $]\pi, 2\pi]$ duplicaría las posiciones) y su distribución conjunta tendrá por densidad.*

$$f_{X\Theta}(x,\theta) = \begin{cases} \dfrac{1}{\pi d}, & \textit{si } (x,\theta) \in [0,d]\times[0,\pi], \\ \\ 0 & \textit{en el resto.} \end{cases}$$

Como puede apreciarse en la figura 3.7, para cada θ fijo la aguja cortará a la paralela de su izquierda o de su derecha si,

$$0 \leq x \leq \frac{l}{2}\sin\theta \quad \longrightarrow \quad \textit{corta a la izquierda}$$

$$0 \leq d - x \leq \frac{l}{2}\sin\theta \quad \longrightarrow \quad \textit{corta a la derecha}$$

Si definimos los sucesos $A_1 = \{(x,\theta);\ 0 \leq \theta \leq \pi,\ 0 \leq x \leq \frac{l}{2}\sin\theta\}$ y $A_2 = \{(x,\theta);\ 0 \leq \theta \leq \pi,\ d - \frac{l}{2}\sin\theta \leq x \leq d\}$ la probabilidad de

[36] Este problema ya fue resuelto en §1. Lo revisitamos desde otro punto de vista.

corte vendrá dada por

$$P(Corte) = P((X,\Theta) \in A_1 \cup A_2) = \int_0^{\pi}\int_0^{\frac{l}{2}\sin\theta} \frac{1}{\pi d}\,dx d\theta + \int_0^{\pi}\int_{d-\frac{l}{2}\sin\theta}^{d} \frac{1}{\pi d}\,dx d\theta = \frac{1}{\pi d}\int_0^{\pi}\left[\frac{l}{2}\sin\theta + \frac{l}{2}\sin\theta\right] d\theta = \frac{l}{\pi d}\int_0^{\pi}\sin\theta\,d\theta = \frac{2l}{\pi d}. \quad (3.34)$$

3.7 Distribución condicionada

Tenemos un vector bidimensional (X, Y). Este vector tiene un comportamiento probabilístico que nos describe su distribución conjunta. Cuando nos interesa algo sobre su comportamiento conjunto podemos recurrir a su función de probabilidad o de densidad conjuntas. Si queremos saber la probabilidad de algún suceso en donde solamente interviene uno de ellos podemos utilizar su distribución marginal. Pero qué ocurre si conocemos una de ellas y queremos saber algo sobre la otra. ¿Es rara esta situación? En absoluto. De una persona podemos saber si es diabético o no, ser diabético es la variable X que *conocemos*, y querríamos saber si un cierto tratamiento de control de la diabetes va a ser efectivo en esta persona: esta sería la variable Y. Si $Y = 1$ indice un tratamiento exitoso nuestro interés estaría en las probabilidades de éxito y fracaso en enfermos diabéticos, es decir, en lo que vamos a llamar la distribución condicionada de la variable Y a la variable X. De esto va esta sección. En Estadística constantemente se está trabajando con distribuciones condicionadas. Vamos a considerar, como siempre, los casos discreto (§ 3.7) y continuo (§ 3.7) separadamente.

Distribución condicionada discreta

Ejemplo 3.14 (Ejemplo 3.1 continuado) *Retomamos el ejemplo visto en 3.1. Ahora vamos a suponer que fijamos el valor de la variable Y, esto es, suponemos que conocemos el número de caras en los tres últimos lanzamientos. Por ejemplo, supongamos que sabemos que $Y = 0$. Si esto es así la variable X solamente nos queda por determinar el primer lanzamiento que puede ser cara o cruz con probabilidad $1/2$, es decir, estamos diciendo que $P(X = 0|Y = 0) = P(X = 1|Y = 0) = 1/2$. Esto lo podemos deducir directamente si partimos del conocimiento previo de $P(X = x, Y = 0)$ y de $P(Y = 0)$. Si aplicamos la definición de probabilidad condicionada para sucesos se tiene que*

$$P(X = x|Y = 0) = \frac{P(X = x, Y = 0)}{P(Y = 0)},$$

y podemos comprobar que obtenemos las probabilidades anteriores. Y esto que acabamos de hacer lo podemos hacer para cualquier valor que previamente sepamos que toma Y. Dado *que $Y = y$ entonces las probabilidades condicionadas de que $X = x$ las podemos obtener a partir de la función de probabilidad conjunta y la marginal de Y aplicando la definición de probabilidad condicionada con*

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

Tabla 3.2: Lanzamiento de cuatro monedas. X nos indica el número de caras en los tres primeros lanzamientos e Y el número de caras en los tres últimos lanzamientos. Dentro de la tabla tenemos la función de probabilidad $P(X = x|Y = y)$ en la fila y.

	X				
Y	**3**	**2**	**1**	**0**	
0	0	0	1/2	1/2	1
1	0	2/6	3/6	1/6	1
2	1/6	3/6	2/6	0	1
3	1/2	1/2	0	0	1
$\boldsymbol{P(X = x)}$	2/16	6/16	6/16	2/16	1

En la tabla 3.2 tenemos las cuatro posibles distribuciones condicionadas (una en cada fila de la tabla). Claramente son cuatro funciones de probabilidad distintas. Conociendo el valor de Y no conocemos el valor exacto de X pero la probabilidad de que X tome los distintos valores es distinta. Esta función de probabilidad condicionada representa la incorporación de la información que la variable Y tiene sobre la variable X a la hora de conocer el valor aleatorio de X. Podemos comprobar que las probabilidades condicionadas $P(X = x|Y = y)$ (cada fila de la tabla) no coinciden con la probabilidad marginal $P(X = x)$ (última fila de la tabla).

Consideremos un vector aleatorio bidimensional (X, Y), con soportes para cada una de sus componentes D_X y D_Y, respectivamente, $P(X \in D_X) = 1$ y $P(Y \in D_Y) = 1$.

Definición 3.11 *La* función de probabilidad condicionada *de Y dado $\{X = x\}$, $x \in D_X$, se define mediante,*

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

Proposición 3.7 *La función definida en 3.11 es una función de probabilidad.*

Demostración 3.5 *En primer lugar es no negativa por tratarse de una probabilidad condicionada. Además suma la unidad sobre D_y ya que*

$$\sum_{y \in D_y} f_{Y|X}(y|x) = \frac{\sum_{y \in D_Y} P(X = x, Y = y)}{P(X = x)} = \frac{P(X = x)}{P(X = x)} = 1.$$

Vamos a ver distintos ejemplos donde veamos cómo obtener la distribución condicionada.

Ejemplo 3.15 (Una de cartas) *Cuando se inicia una partida de cartas* la mano *se suele decidir a la carta más alta. Supongamos una partida entre dos jugadores que, antes de iniciarla, extraen al azar sendas cartas para decidir cuál de los dos repartirá inicialmente las cartas. Si por X designamos la mayor de las dos cartas y por Y la menor, vamos encontrar la distribución conjunta del vector (X, Y), las marginales y las condicionadas.*

La baraja española consta de 4 palos con 12 cartas cada uno de ellos, numeradas del 1 (As) al 12 (Rey). Como el As se considera siempre la carta más alta, podemos asignarle el número 13 y suponer las cartas numeradas del 2 al 13. No pueden haber empates, con lo que el problema es equivalente al de extraer al azar, sin reemplazamiento, dos bolas de una urna que contiene 12 bolas numeradas del 2 al 13. Observemos que el orden no cuenta en la extracción, sólo cuál de ellas es la mayor y cuál la menor, por lo que las posibles extracciones son $\binom{12}{2} = 66$. *En definitiva,*

$$f_{XY}(x,y) = P(X=x, Y=y) = \begin{cases} \dfrac{1}{66}, & 2 \leq y < x \leq 13; \\ 0, & \text{en el resto.} \end{cases}$$

La función de probabilidad marginal de X *vale,*

$$f_X(x) = P(X=x) = \sum_{2 \leq y < x} \frac{1}{66} = \frac{x-2}{66}, \quad x \in D_X = \{3, \ldots, 13\},$$

y $f_X(x) = 0$ *si* $x \in D_X^c$. *Para* Y,

$$f_Y(y) = \sum_{y < x \leq 13} \frac{1}{66} = \frac{13-y}{66}, \quad y \in D_Y = \{2, \ldots, 12\},$$

y $f_Y(y) = 0$ *si* $y \in D_Y^c$. *En cuanto a las condicionadas,*

$$f_{X|Y}(x|y) = P(X=x|Y=y) = \frac{1/66}{(13-y)/66} = \frac{1}{13-y}, \quad y < x \leq 13,$$

que es la distribución uniforme sobre el conjunto $D_{X|Y} = \{y+1, y+2, \ldots, 13\}$.

$$f_{Y|X}(y|x) = \frac{1/66}{(x-2)/66} = \frac{1}{x-2}, \quad 2 \leq y < x,$$

que es la distribución uniforme sobre el conjunto $D_{Y|X} = \{2, 3, \ldots, x-1\}$.

Ejemplo 3.16 (Tablas de contingencia y distribuciones condicionadas) *Tomamos una gran muestra de tamaño* n *en una población. Observamos dos variables discretas en cada uno de ellos.*[37] *Tendremos los valores observados* (x_i, y_i) *con* $i = 1, \ldots, n$ *son los valores observados. Consideramos el vector aleatorio* (X, Y) *tal que su distributición asigna una probabilidad de* $1/n$ *a cada uno de los datos bivariantes observados. El vector que tiene por distribución lo que se conoce como la distribución empírica de la muestra. Vamos a ver cuál es la distribución conjunta de este vector, las marginales y las condicionadas. Denotamos por* $\{u_1, \ldots, u_I\}$ *(respectivamente por* $\{v_1, \ldots, v_J\}$ *los valores* **distintos** *entre los* x_i *(respectivamente entre los* y_j. *Por ejemplo, si nuestra muestra es* $\{(1,2), (1,3), (2,3), (2,3)\}$ *entonces los valores distintos de* X *e* Y *son* $\{1,2\}$ *y* $\{2,3\}$ *y distribución conjunta de este vector aleatorio será*

$$P(X = u_i, Y = v_j) = \frac{n_{ij}}{n},$$

donde n_{ij} *es el número de* (x_k, y_k) *tales que* $x_k = u_i$ *e* $y_k = v_j$.[38] *Si*

[37] Por ejemplo, tomamos 1000 individuos con carnet de conducir al menos cinco años en la población de la Comunidad Valenciana. Observamos el número de años que tiene el carnet de conducir y el número de accidentes en los últimos cinco años.

[38] La tabla de doble entrada que en la fila i y columna j tiene el valor n_{ij} se le llama tabla de contingencia.

denotamos $n_{i\cdot} = \sum_{j=1}^{J} n_{ij}$ *y* $n_{\cdot j} = \sum_{i=1}^{I} n_{ij}$ *entonces las marginales de* X *e* Y *son*

$$P(X = u_i) = \frac{n_{i\cdot}}{n},$$

y

$$P(Y = v_j) = \frac{n_{\cdot j}}{n}.$$

Las distribuciones condicionadas serían

$$P(X = u_i|Y = v_j) = \frac{n_{ij}}{n_{\cdot j}},$$

y

$$P(Y = v_j|X = u_i) = \frac{n_{ij}}{n_{i\cdot}}.$$

El siguiente ejemplo muestra una interesante relación entre la distribución de Poisson y la distribución multinomial. Esta relación tiene un gran interés cuando se trabaja con variables discretas en análisis de tablas de contingencia.

Ejemplo 3.17 *Consideremos dos variables aleatorias independientes* X *e* Y*, con distribución de Poisson de parámetros* μ *y* λ*, respectivamente. Queremos encontrar al distribución de la variable condicionada* $X|X + Y = r$*. Recordemos que*

$$f_{X|X+Y}(k|r) = P(X = k|X + Y = r) = \frac{P(X = k, Y = r - k)}{P(X + Y = r)} = \frac{f_{XY}(k, r - k)}{f_{X+Y}(r)}. \quad (3.35)$$

La distribución conjunta del vector (X, Y) *es conocida por tratarse de variables independientes,*

$$f_{XY}(k, r - k) = \frac{\mu^k}{k!} e^{-\mu} \frac{\lambda^{r-k}}{r - k!} e^{-\lambda}. \quad (3.36)$$

La distribución de la variable $X + Y$ *se obtiene de la forma*

$$P(X + Y = r) = P\left(\bigcup_{k=0}^{r} \{X = k, Y = r - k\}\right) = \sum_{k=0}^{r} P(X = k, Y = r - k) = \sum_{k=0}^{r} \frac{\mu^k \lambda^{r-k}}{k! r - k!} e^{-(\mu+\lambda)} = \frac{e^{-(\mu+\lambda)}}{r!} \sum_{k=0}^{r} \frac{r!}{k! r - k!} \mu^k \lambda^{r-k} = \frac{(\mu + \lambda)^r}{r!} e^{-(\mu+\lambda)}. \quad (3.37)$$

Lo que nos dice que $X + Y \sim Po(\mu + \lambda)$*. Sustituyendo (3.36) y (3.37) en (3.35),*

$$f_{X|X+Y}(k|r) = \frac{\frac{\mu^k}{k!} e^{-\mu} \frac{\lambda^{r-k}}{r-k!} e^{-\lambda}}{\frac{(\mu+\lambda)^r}{r!} e^{-(\mu+\lambda)}} = \frac{r!}{k!(r - k)!} \frac{\mu^k \lambda^{r-k}}{(\mu + \lambda)^r} = \binom{r}{k} \left(\frac{\mu}{\mu + \lambda}\right)^k \left(1 - \frac{\mu}{\mu + \lambda}\right)^{r-k}, \quad (3.38)$$

concluimos que la distribución condicionada $X|X + Y = r$ *es una distribución binomial, en concreto,* $X|X + Y = r \sim B(r, \mu/(\mu + \lambda))$*.*

El concepto de distribución condicional se extiende con facilidad al caso k-dimensional. Si $X = (X_1, \ldots, X_k)$ es un vector aleatorio k-dimensional y $X^l = (X_{i_1}, \ldots, X_{i_l})$, $l \leq k$ y $X^{k-l} = (X_{j_1}, \ldots, X_{j_{k-l}})$ son subvectores de dimensiones complementarias, con soportes D_{X^l} y $D_{X^{k-l}}$, respectivamente, la *función de probabilidad condicionada* de X^l dado $X^{k-l} = (x_{j_1}, \ldots, x_{j_{k-l}})$ con $(x_{j_1}, \ldots, x_{j_{k-l}}) \in D_{x^{k-l}}$, se define como

$$f_{X^l|X^{k-l}}(x_{i_1}, \ldots, x_{i_l}|x_{j_1}, \ldots, x_{j_{k-l}}) = \frac{f_X(x_1, \ldots, x_k)}{f_{X^{k-l}}(x_{j_1}, \ldots, x_{j_{k-l}})},$$

donde el argumento del numerador, $(x_1, \ldots, x_k)$, está formado por las componentes $(x_{i_1}, \ldots, x_{i_l})$ y $(x_{j_1}, \ldots, x_{j_{k-l}})$ adecuadamente ordenadas.

Y podemos aplicar esto al caso de la distribución multinomial.

Ejemplo 3.18 (Distribución multinomial) *Si el vector $X = (X_1, \ldots, X_k) \sim M(n; p_1, \ldots, p_k)$ sabemos que la marginal del subvector $X^l = (X_1, \ldots, X_l)$ es una $M(n; p_1, \ldots, p_l, (1-p^*))$, $p^* = p_1 + \cdots + p_l$ (en definitiva la partición de sucesos que genera la multinomial queda reducida a $A_1, \ldots, A_l, A^*$, con $A^* = \left(\cup_{i=1}^l A_i\right)^c$). La distribución condicionada de $X^{k-l} = (X_{l+1}, \ldots, X_k)$ dado $X^l = (n_1, \ldots, n_l)$, viene dada por*

$$f_{X^{k-l}|X^l}(n_{l+1}, \ldots, n_k|n_1, \ldots, n_l) =$$
$$\frac{\dfrac{n!}{n_1!n_2!\ldots n_k!}\displaystyle\prod_{i=1}^k p_i^{n_i}}{\dfrac{n!}{n_1!\ldots n_l!(n-n^*)!}(1-p^*)^{(n-n^*)}\displaystyle\prod_{i=1}^l p_i^{n_i}} =$$
$$\frac{(n-n^*)!}{n_{l+1}!\ldots n_k!}\prod_{i=l+1}^k \left(\frac{p_i}{1-p^*}\right)^{n_i} \tag{3.39}$$

con $n^* = n_1 + \cdots + n_l$ y $\sum_{i=l+1}^k n_i = n - n^*$. Se trata, en definitiva, de una $M(n-n^*; \frac{p_{l+1}}{1-p^*}, \ldots, \frac{p_k}{1-p^*})$.

Ejemplo 3.19 (Binomial negativa bivariante) *La función de probabilidad condicionada de Y dado $X = x$, cuando $(X, Y) \sim BN(r; p_1, p_2)$, vale*

$$f_{Y|X}(y|x) =$$
$$\frac{\binom{x+y+r-1}{y}\binom{x+r-1}{x}p_1^x p_2^y (1-p_1-p_2)^r}{\binom{x+r-1}{x}\left(\dfrac{p_1}{1-p_2}\right)^x\left(\dfrac{1-p_1-p_2}{1-p_2}\right)^r} =$$
$$\binom{x+y+r-1}{y}p_2^y(1-p_2)^{x+r}, \tag{3.40}$$

para $x = 0, 1, \ldots$ $\quad y = 0, 1, \ldots$. Luego $Y|X = x \sim BN(x+r, p_2)$.

Distribución condicionada con vectores continuos

Suponemos ahora que el vector (X, Y) es continuo con densidad conjunta $f_{(X,Y)}$ y marginales f_X y f_Y respectivamente. Conocemos el valor de X, sabemos que $X = x$ pero tenemos un pequeño problema: esto tiene probabilidad cero. La definición que hemos visto en el caso discreto en la sección anterior no nos vale. Una opción simple y válida es definir directamente la función de densidad de la distribución condicionada de la variable Y a $X = x$ que podemos denotar $Y|X = x$. Y lo vamos a hacer.

Definición 3.12 (Función de densidad condicionada) *Definimos la función de densidad condicionada de la variable Y a la variable X como*

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

si $f_X(x) > 0$.

Con el siguiente ejemplo justificamos la sensatez de la definición que acabamos de ver.

Ejemplo 3.20 (Continúa ejemplo 3.6) *Pensemos en la elección de un punto al azar en $B(0, 1)$, círculo unidad. Fijemos la abscisa del punto en $X = x$, $|x| < 1$, y consideremos cómo se distribuirá la ordenada Y sobre la correspondiente cuerda. Estamos hablando de la distribución condicionada de $Y|X = x$, que no sólo tiene sentido, si no que intuimos que será uniforme sobre la cuerda. ¿Cómo comprobar nuestra intuición? Recordando las expresiones de las densidades conjuntas y las marginales obtenidas en 3.6 y el ejemplo 3.12, tendremos*

$$f_{Y|X}(y|x) = \begin{cases} \dfrac{1/\pi}{2\sqrt{1-x^2}/\pi}, & \text{si } |y| \leq \sqrt{1-x^2} \\ \\ 0, & \text{en el resto,} \end{cases}$$

que confirma nuestra intuición, $Y|X = x \sim U(-\sqrt{1-x^2}, +\sqrt{1-x^2})$. Parece lógico pensar que la densidad condicionada sea, efectivamente, la que hemos supuesto.

En el siguiente ejemplo pretendemos una situación donde el vector (X, Y) no tiene una distribución uniforme sobre un dominio como un cuadrado o una bola.

Ejemplo 3.21 *Elegimos al azar X en [0,1] y a continuación Y, también al azar, en $[0, X^2]$. Ver figura 3.8. Es decir*

$$f_X(x) = \begin{cases} 1, & x \in [0, 1]; \\ 0, & \text{en el resto.} \end{cases} \quad \text{y} \quad f_{Y|X}(y|x) = \begin{cases} 1/x^2, & y \in [0, x^2]; \\ 0, & \text{en el resto.} \end{cases}$$

La densidad conjunta de (X, Y) vale

$$f_{(X,Y)}(x, y) = f_X(x) f_{Y|X}(y|x) = \begin{cases} \dfrac{1}{x^2}, & x \in [0, 1],\ y \in [0, x^2]; \\ \\ 0, & \text{en el resto.} \end{cases}$$

La densidad marginal de Y es

$$f_Y(y) = \int_{\sqrt{y}}^{1} \frac{1}{x^2} dx = \frac{1}{\sqrt{y}} - 1, \quad y \in [0, 1],$$

y vale 0 fuera del intervalo.

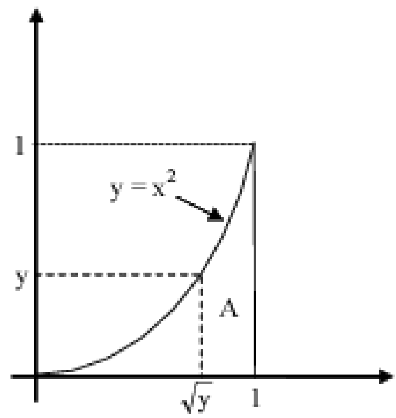


Figura 3.8: Corresponde al ejemplo 3.21.

Cabe preguntarse si la elección de X e Y que hemos hecho se corresponde con la elección al azar de un punto en el recinto A de la figura, determinado por la parábola $y = x^2$ entre $x = 0$ y $x = 1$. La respuesta es negativa, puesto que la densidad conjunta vendría dada en este caso por

$$f^*_{(X,Y)}(x,y) = \begin{cases} \dfrac{1}{\text{área de } A} = 3, & (x,y) \in A \\ \\ 0, & \text{en el resto.} \end{cases}$$

y evidentemente, $f_{(X,Y)}(x,y) \neq f^*_{(X,Y)}(x,y)$. Puede comprobarse que, en este caso,

$$f^*_X(x) = \begin{cases} 3x^2, & x \in [0,1]; \\ 0, & \text{en el resto.} \end{cases} \quad \text{y} \quad f^*_{Y|X}(y|x) = \begin{cases} 1/x^2, & y \in [0,x^2]; \\ 0, & \text{en el resto.} \end{cases}$$

Es decir, elegida la componente X la elección de Y continua siendo al azar en el intervalo $[0, X^2]$, pero a diferencia de cómo elegíamos X inicialmente, ahora ha de elegirse con la densidad $f^*_X(x)$.

Hemos definido la distribución condicionada de una variable a otra. Sin embargo, en las aplicaciones no se suele condicionar a una sola variable sino a varias. Y también no se trabaja con la distribución de una variable a condicionar sino también a varias. En resumen: ¿cómo definimos la distribución condicionada de un subvector al resto de las variables? Tenemos el vector aleatorio $\boldsymbol{X} = (X_1, \ldots, X_k)$. Lo dividimos en dos subvectores que podemos denotar $\boldsymbol{X}_{(i_1,\ldots,i_l)} = (X_{i_1}, \ldots, X_{i_l})$. Si suponemos que $\{j_1, \ldots, j_{k-l}\} = \{1, \ldots, k\} \setminus \{i_1, \ldots, i_l\}$ entonces podemos definir la función de densidad condicionada del subvector $\boldsymbol{X}_{(i_1,\ldots,i_l)}$ al subvector $\boldsymbol{X}_{(j_1,\ldots,j_{k-l})}$ (donde $\{i_1, \ldots, i_l\}$ y $\{j_1, \ldots, j_{k-l}\}$ son una partición de $\{1, \ldots, k\}$) como

$$f_{\boldsymbol{X}_{(i_1,\ldots,i_l)}|\boldsymbol{X}_{(j_1,\ldots,j_{k-l})}}(x_{i_1}, \ldots, x_{i_l}|x_{j_1}, \ldots, x_{j_{k-l}}) = \frac{f_{\boldsymbol{X}}(x_1, \ldots, x_k)}{f_{\boldsymbol{X}_{(j_1,\ldots,j_{k-l})}}(x_{j_1}, \ldots, x_{j_{k-l}})}, \tag{3.41}$$

con $f_V(x_{j_1}, \ldots, x_{j_{k-l}}) > 0$ y donde el argumento de ambos numeradores, $(x_1, \ldots, x_k)$, está formado por las componentes $(x_{i_1}, \ldots, x_{i_l})$ y $(x_{j_1}, \ldots, x_{j_{k-l}})$ adecuadamente ordenadas.

Ejemplo 3.22 (Normal bivariante) *Si* (X,Y) *es un vector normal bivariante,*

$$f_{Y|X}(y|x) = \frac{\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}e^{-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x-\mu_x}{\sigma_x}\right)^2-2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right)+\left(\frac{y-\mu_y}{\sigma_y}\right)^2\right\}}}{\frac{1}{\sigma_x\sqrt{2\pi}}e^{\left(-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2\right)}} =$$

$$\frac{1}{\sigma_y\sqrt{2\pi(1-\rho^2)}}e^{-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{y-\mu_y}{\sigma_y}\right)-\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\right\}^2}$$

$$= \frac{1}{\sigma_y\sqrt{2\pi(1-\rho^2)}}e^{-\frac{1}{2\sigma_y^2(1-\rho^2)}\left\{y-\left(\mu_y+\rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)\right)\right\}^2}. \quad (3.42)$$

Es decir, $Y|X=x \sim N\left(\mu_y+\rho\frac{\sigma_y}{\sigma_x}(x-\mu_x), \sigma_y^2(1-\rho^2)\right)$.

3.8 Transformación de un vector aleatorio

¿Solamente transformamos variables? Habitualmente partiendo de un vector aleatorio $\boldsymbol{X} = (X_1,\ldots,X_k)$ nos interesarán funciones de estas variables aleatorias de modo que tendremos

$$\boldsymbol{Y} : \Omega \xrightarrow{\boldsymbol{X}} \mathbb{R}^k \xrightarrow{\boldsymbol{g}} \mathbb{R}^l,$$

siendo $\boldsymbol{g} = (g_1,\ldots,g_l)$. Y pediremos que $\boldsymbol{g}$ sea medible entre $(\mathbb{R}^k,\beta^k)$ y $(\mathbb{R}^l,\beta^l)$.[4]

Un punto importante a considerar, si las variables que teníamos originalmente eran independientes: ¿siguen siendo independientes variables que obtenemos transformando (marginalmente) cada una de ellas?

Como corolario de la definición (3.10) se comprueba fácilmente que las transformaciones de variables independientes también lo son.

Corolario 3.1 *Si* g_i, $i=1,\ldots,k$, *son funciones de* $\mathbb{R}$ *en* $\mathbb{R}$, *las variables aleatorias* $Y_i = g_i(X_i)$, $i=1,\ldots,k$ *son independientes si las* X_i *lo son.*

Demostración 3.6 *Tenemos*

$$P(Y_1\in B_1,\ldots,Y_k\in B_k) = P(X_1\in g_1^{-1}(B_1),\ldots,X_k\in g_k^{-1}(B_k)) = P(X_1\in g_1^{-1}(B_1))\ldots P(X_k\in g_k^{-1}(B_k)). \quad (3.43)$$

Para $X = (X_1,\ldots,X_k)$, vector aleatorio k-dimensional, abordaremos el problema solamente para el caso continuo. La obtención de la densidad de la nueva variable o vector resultante en función de f_X plantea dificultades en el caso más general, pero bajo ciertas condiciones, equivalentes a las impuestas para el caso univariante, es posible disponer de una expresión relativamente sencilla.

[4]Significando esto que $g^{-1}(B)\in\beta^k$ siendo $B\in\beta^l$. Se puede ver que g es medible si y solo si cada una de las g_i lo es.

Teorema 3.1 *Sea $\boldsymbol{X} = (X_1, \ldots, X_k)$ es un vector aleatorio continuo con soporte D y sea $\boldsymbol{g} = (g_1, \ldots, g_k) : \mathbb{R}^k \to \mathbb{R}^k$ una función vectorial que verifica:*

1. *$\boldsymbol{g}$ es uno a uno sobre D,*
2. *el Jacobiano de $\boldsymbol{g}$, $J = \frac{\partial(g_1,\ldots,g_k)}{\partial(x_1,\ldots,x_k)}$, es distinto de cero $\forall x \in D$, y*
3. *existe $\boldsymbol{h} = (h_1, \ldots, h_k)$ inversa de g.*

Entonces, $\boldsymbol{Y} = g(\boldsymbol{X})$ es un vector aleatorio continuo cuya densidad conjunta, para $\boldsymbol{y} = (y_1, \ldots, y_k) \in \boldsymbol{g}(D)$, viene dada por

$$f_{\boldsymbol{Y}}(y_1, \ldots, y_k) = f_{\boldsymbol{X}}\left(h_1(y_1, \ldots, y_k), \ldots, h_k(y_1, \ldots, y_k)\right) |J^{-1}|, \tag{3.44}$$

donde $J^{-1} = \frac{\partial(h_1,\ldots,h_k)}{\partial(y_1,\ldots,y_k)}$ es el Jacobiano de h.

Este teorema no es más que el teorema del cambio de variable en una integral múltiple y su demostración rigurosa, de gran dificultad técnica, puede encontrarse en cualquier libro de Análisis Matemático.

Veamos el interés del resultado a través de los siguientes ejemplos.

Ejemplo 3.23 (Continuación del ejemplo 3.12) *En 3.6 estudiábamos el vector aleatorio determinado por las coordenadas de un punto elegido al azar en el círculo unidad. La densidad conjunta venía dada por*

$$f_{(X,Y)}(x,y) = \begin{cases} \dfrac{1}{\pi}, & si\ (x,y) \in B(0,1) \\ \\ 0, & en\ el\ resto. \end{cases}$$

Consideremos ahora las coordenadas polares del punto, $R = \sqrt{X^2 + Y^2}$ y $\Theta = \arctan \frac{Y}{X}$. Para obtener su densidad conjunta, necesitamos las transformaciones inversas, $X = R\cos\Theta$ e $Y = R\sin\Theta$. El correspondiente jacobiano vale R y la densidad conjunta es

$$f_{(R,\Theta)}(r,\theta) = \begin{cases} \dfrac{r}{\pi}, & si\ (r,\theta) \in [0,1] \times [0,2\pi] \\ \\ 0, & en\ el\ resto. \end{cases}$$

Con facilidad se obtienen las marginales correspondientes, que resultan ser

$$f_R(r) = \begin{cases} 2r, & si\ r \in [0,1] \\ \\ 0, & en\ el\ resto, \end{cases}$$

y para Θ,

$$f_\Theta(\theta) = \begin{cases} \dfrac{1}{2\pi}, & si\ \theta \in [0,2\pi] \\ \\ 0, & en\ el\ resto. \end{cases}$$

Como $f_{(R,\Theta)}(r,\theta) = f_R(r) f_\Theta(\theta),\ \forall (r,\theta)$, R y Θ son independientes.

Ejemplo 3.24 (Suma de dos variables aleatorias) *Sea $\boldsymbol{X} = (X_1, X_2)$ un vector aleatorio bidimensional con densidad conjunta $f_{\boldsymbol{X}}(x_1, x_2)$ o simplemente $f(x_1, x_2)$. Las marginales de X_1 y X_2 las denotamos f_1 y f_2 respectivamente. Definimos $U = X_1 + X_2$ y queremos obtener su densidad. Para poder utilizar el resultado anterior la transformación* *debe de ser también bidimensional, cosa que conseguimos si consideramos una nueva variable* $V = X_1$. *El nuevo vector aleatorio es* $\boldsymbol{Y} = (U, V)$ *y podemos aplicar el teorema. La transformación inversa viene dada por* $X_1 = V$ *y* $X_2 = U - V$, *cuyo jacobiano es* -1. *Tendremos pues,*

$$f_{\boldsymbol{Y}}(u, v) = f_{\boldsymbol{X}}(v, u - v),$$

y para obtener la densidad marginal de la suma, U,

$$f_U(u) = \int_{-\infty}^{+\infty} f_{\boldsymbol{X}}(v, u - v)dv. \tag{3.45}$$

Ejemplo 3.25 (Producto de dos variables) *Tenemos el vector* $\boldsymbol{X} = (X_1, X_2)$ *con densidad conjunta* $f_{\boldsymbol{X}}(x_1, x_2)$. *Para obtener la densidad de* $W = X_1X_2$, *definimos* $T = X_1$ *y actuamos como en el ejemplo 3.24. Con* $\boldsymbol{Y} = (T, W)$ *y transformaciones inversas* $X_1 = T$ *y* $X_2 = W/T$, *el Jacobiano es* $J^{-1} = 1/T$ *y la densidad conjunta de* $\boldsymbol{Y}$,

$$f_{\boldsymbol{Y}}(t, w) = \frac{1}{|t|} f_X\left(t, \frac{w}{t}\right).$$

La marginal del producto se obtiene integrando respecto de la otra componente,

$$f_W(w) = \int_{-\infty}^{+\infty} \frac{1}{|t|} f_X\left(t, \frac{w}{t}\right) dt. \tag{3.46}$$

En los ejemplos 3.24 y 3.25 hubieramos podido también proceder utilizando la transformación bidimensional $Y = (Y_1, Y_2)$, con $Y_1 = X_1 + X_2$ e $Y_2 = X_1X_2$, lo que en teoría nos hubiera hecho ganar tiempo; pero sólo en teoría, porque en la práctica las inversas hubieran sido más complicadas de manejar que las anteriores.

Ejemplo 3.26 (Suma de variables aleatorias independientes) *En el ejemplo 3.24 hemos considerado variables cualesquiera no necesariamente independientes. ¿Y si asumimos la independencia de las variables? ¿En qué se modifica la densidad que acabamos de obtener? En la ecuación 3.45 podemos factorizar la densidad conjunta* $f_{\boldsymbol{X}}(v, u-v) = f_{X_1}(v)f_{X_2}(u-v)$ *de modo que tendríamos que la densidad de* $X_1 + X_2$ *cuando ambas variables son independientes vendría dada por*

$$f_{X_1+X_2}(u) = \int_{-\infty}^{+\infty} f_{X_1}(v)f_{X_2}(u - v)dv. \tag{3.47}$$

A la expresión de la densidad dada en 3.47 se le suele llamar la convolución de las densidades f_{X_1} *y* f_{X_2}.

Ejemplo 3.27 (Distribución del máximo y el mínimo) *Si* $X_1, X_2, \ldots, X_n$ *son* n *variables aleatorias independientes, vamos a obtener la distribución de probabilidad del máximo y el mínimo de todas ellas. Definimos* $V = \max\{X_1, X_2, \ldots, X_n\}$ *y* $U = \min\{X_1, X_2, \ldots, X_n\}$. *Empezamos con la distribución marginal del máximo. Lo más sencillo es obtener la función de distribución de* V, *puesto que*

$$\{V \leq x\} \iff \bigcap_{i=1}^{n} \{X_i \leq x\},$$

y de aquí

$$F_V(x) = P\left(\cap_{i=1}^n \{X_i \leq x\}\right) = \prod_{i=1}^n F_i(x).$$

La función de densidad puede obtenerse derivando la expresión anterior,

$$f_V(x) = \sum_{i=1}^n \left[f_i(x) \prod_{j \neq i} F_j(x)\right].$$

Un caso especial es aquel en el que las variables tiene todas la misma distribución de probabilidad. Si F y f son, respectivamente, las funciones de distribución y densidad comunes a todas ellas,

$$F_V(x) = [F(x)]^n$$

y

$$f_V(x) = n\,[F(x)]^{n-1}\,f(x).$$

Veamos ahora la distribución del mínimo. Observemos que

$$\{U > x\} \iff \bigcap_{i=1}^n \{X_i > x\},$$

y de aquí

$$F_V(x) = 1 - P(U > x) = 1 - P\left(\cap_{i=1}^n \{X_i > x\}\right) = 1 - \prod_{i=1}^n \left[1 - F_i(x)\right],$$

y

$$f_V(x) = \sum_{i=1}^n \left[f_i(x) \prod_{j \neq i} (1 - F_j(x))\right].$$

Si todas las variables comparten una distribución común con F y f como funciones de distribución y densidad, respectivamente,

$$F_V(x) = 1 - [1 - F(x)]^n$$

y

$$f_V(x) = n\,[1 - F(x)]^{n-1}\,f(x).$$

Finalmente nos planteamos la distribución conjunta del máximo y del mínimo. Para obtener al distribución conjunta de V y U observemos que,

$$\{V \leq x\} = (\{V \leq x\} \cap \{U \leq y\}) \cup (\{V \leq x\} \cap \{U > y\}).$$

Al tomar probabilidades hemos de tener en cuenta que

$$P(\{V \leq x\} \cap \{U > y\}) = \begin{cases} 0, & \text{si } x \leq y; \\ P\left(\cap_{i=1}^n \{y < X_i \leq x\}\right), & \text{si } x > y. \end{cases}$$

En definitiva

$$F_{VU}(x, y) = \begin{cases} \prod_{i=1}^n F_i(x) - \prod_{i=1}^n [F_i(x) - F_i(y)], & \text{si } x > y; \\ \prod_{i=1}^n F_i(x), & \text{si } x \leq y, \end{cases}$$

resultado que nos indica que V y U no son independientes. La función de densidad conjunta la obtendremos mediante (3.13),

$$f_{VU}(x,y) = \begin{cases} \sum_{i \neq j} \left[f_i(x) f_j(y) \prod_{k \neq i,j} [F_k(x) - F_k(y)] \right], & \text{si } x > y; \\ \\ 0, & \text{si } x \leq y. \end{cases}$$

Cuando las variables tiene la misma distribución con F y f como funciones de distribución y densidad comunes, respectivamente,

$$F_{VU}(x,y) = \begin{cases} [F(x)]^n - [F(x) - F(y)]^n, & \text{si } x > y; \\ \\ [F(x)]^n, & \text{si } x \leq y, \end{cases}$$

y

$$f_{VU}(x,y) = \begin{cases} n(n-1) f(x) f(y) [F(x) - F(y)]^{n-2}, & \text{si } x > y; \\ \\ 0, & \text{si } x \leq y. \end{cases}$$

3.9 Problemas

Distribución conjunta, marginal y condicionada

* **Ej. 2** — (*[10, pág. 233]*)Elegimos al azar (sin reemplazamiento) tres bolas de una urna que contiene 3 bolas rojas, 4 bolas blancas y 5 bolas azules. Si denotamos por X e Y el número de bolas rojas y blancas elegidas, determinar la función de probabilidad conjunta de X e Y así como las funciones de probabilidad marginales.

* **Ej. 3** — Sean X e Y dos variables aleatorias independientes ambas con soporte en el conjunto $\{0,1,2\}$, tales que $P(X=0) = 0.3$,$P(X=1) = 0.5$, $P(Y=0) = 0.6$ y $P(Y=1) = 0.1$. Se pide:

1. Construye una tabla en la que se muestre $P(X=x, Y=y)$ para todo par $(x,y) \in \{0,1,2\} \times \{0,1,2\}$.
2. Halla $P(X=Y)$
3. Halla $P(X+Y=2)$
4. Halla la función de probabilidad de la variable aleatoria $Z = X+Y$, y represéntala gráficamente.
5. Halla la función de probabilidad de la variable aleatoria $T = X-Y$, y represéntala gráficamente.

* **Ej. 4** — Sean X e Y dos variables aleatorias independientes, ambas con la misma distribución: geométrica de parámetro p, $0 < p < 1$. Sea Z la variable aleatoria definida como el mínimo de X e Y . Queremos determinar la función de probabilidad de Z, para ello vamos a completar los siguientes pasos:

1. Halla $P(X \geq n)$ para $n \in \{0,1,\ldots\}$
2. Halla $P(Z \geq n)$ para $n \in \{0,1,\ldots\}$ (Pista: $\{Z \geq n\} \Leftrightarrow \{X \geq n \bigcap Y \geq n\}$
3. Halla $f_Z(k) = P(Z=k)$, para $k \in \{0,1,\ldots,\}$

* **Ej. 5** — Una planta produce M semillas, donde M se supone que sigue una distribución binomial con parámetros n y p. Cada semilla germina con probabilidad γ independientemente de las demás. Denotamos por R el número de plantas que realmente germinan. Se pide calcular (y dibujar para el caso particular de $n = 5$, $p = 0.5$ y $\gamma = 0.8$) las siguientes funciones:

1. La función de probabilidad condicional de R dado que $M = m$.
2. La función de probabilidad conjunta de R y M.
3. La función de probabilidad de R.
4. La función de probabilidad de M dado R = r.

* **Ej. 6** — Lanzamos 2 veces un dado especial aunque correctamentamente equilibrado. Por especial entenderemos que, a diferencia de los dados habituales (cuyas caras están numeradas de 1 a 6 mediante puntos) tiene 1 cara marcada con un punto, 2 caras numeradas con 2 puntos y las tres restantes numeradas con 3 puntos cada una de ellas. Denotamos por X_i la puntuación obtenida en el lanzamiento i, $i = 1, 2$, y por Y el producto de las dos puntuaciones, es decir $Y = X_1 * X_2$.

1. Halla la función de distribución de X_1
2. Obtén la función de probabilidad de Y .
3. Halla la función de probabilidad de $Y|X_1$

* **Ej. 7** — La función de densidad conjunta de las variables aleatorias X e Y viene dada por $f(x, y) = x + y$ para $0 < x, y < 1$ y $f(x, y) = 0$ en el resto. Consideremos el círculo centrado en el origen que pasa por el punto *(X, Y)*. ¿Cuál es la probabilidad de que la circunferencia de dicho círculo no sea mayor que 2π?

[39] Ver ejemplo 3.17.

** **Ej. 8** — [39] Supongamos que tenemos n variables aleatorias independientes, $X_1, \ldots, X_n$. La variable X_i sigue una distribución de Poisson con parámetro λ_i (con $\lambda_i > 0$). Se pide demostrar que la distribución condicionada del vector $(X_1, \ldots, X_n)$ a la variable $\sum_{i=1}^n X_i$ es una multinomial. Determinar los parámetros de la distribución condicionada.

* **Ej. 9** — La densidad conjunta de X e Y es $f(x, y) = \lambda^3 x e^{-\lambda(x+y)}$ para $x > 0$ e $y > 0$ (0 en otro caso). Se pide

1. Halla las densidades marginales y demuestra que X e Y son independientes.
2. Halla $P(X \leq a, Y \leq b)$ para cualesquiera a y b números positivos.
3. Halla $P(X \leq a)$ para $a > 0$.
4. Calcula la probabilidad de que $X + Y \leq a$ para $a > 0$
5. Encuentra una densidad para $Z = X + Y$. Represéntala.

* **Ej. 10** — Sea (X, Y) un vector aleatorio cuya función de densidad conjunta es

$$f(x, y) = \begin{cases} 30x^4(1-x)^2 y e^{-xy}, & 0 < x < 1, 0 < y < \infty; \\ 0, & \text{en el resto.} \end{cases}$$

Halla la función de densidad marginal $f_X(x)$ y la función de densidad condicionada $f(y|x)$.

* **Ej. 11** — Sea (X, Y) un vector aleatorio cuya función de densidad conjunta es

$$f(x,y) = \frac{1}{2\sqrt{2\pi}} exp\{-\frac{1}{2}(y-x)^2\}$$

si $-1 < x < 1$, $-\infty < y < \infty$ y 0 en otro caso. Halla $f_X(x)$ y $f(y|x)$.

** **Ej. 12** — Inyectamos a unos ratones microorganismos del tipo A y B en igual proporción. Se supone que los microorganismos efectivos de cada tipo se distribuyen independientemente con arreglo a una misma distribución de Poisson de parámetro λ. Un ratón sobrevive si y sólo si no hay microorganismos efectivos en la dosis inyectada. A los ratones muertos se les examina para ver si contenían microorganismos de uno o de los dos tipos. Encontrar la probabilidad de que un ratón muerto contenga microorganismos de un solo tipo.

* **Ej. 13** — Consideremos la siguiente función

$$f_{X|Y}(x|y) = \begin{cases} \frac{y^x e^{-y}}{x!}, & \text{si } y \geq 0,\ x = 0, 1, \ldots \\ 0, & \text{en el resto.} \end{cases}$$

1. Demostrar que para cada y fijo, $f_{X|Y}(x|y)$ es una función de probabilidad (la de X condicionada a $Y = y$).
2. Si $Y \sim Exp(\lambda)$, con $\lambda = 1$, encontrar la densidad conjunta de X e Y.
3. Comprobar que la marginal de X viene dada por

$$f_X(x) = \begin{cases} \frac{1}{2^{x+1}}, & x = 0, 1, \ldots \\ \\ 0, & \text{en el resto.} \end{cases}$$

** **Ej. 14** — Sean $X_1, X_2, \ldots, X_n$ variables aleatorias con función de distribución conjunta F y con funciones de distribución marginales $F_1, F_2, \ldots, F_n$. Demostrar que

$$1 - \sum_{i=1}^{n}[1 - F_i(x_i)] \leq F(x_1, x_2, \ldots, x_n) \leq \min_{1\leq i\leq n} F_i(x_i).$$

** **Ej. 15** — Las variables aleatorias X e Y tienen una distribución conjunta uniforme en el interior del triángulo T de vértices (0,0), (2,0) y (1,2). Calcular $P(X < 1, Y < 1)$ y $P(Y < 1|X < 1.5)$.

Ej. 16 — La densidad conjunta de las variables A y B viene dada por

$$f(a,b) = \begin{cases} a + b, & (a,b) \in [0,1]^2; \\ 0, & \text{en el resto.} \end{cases}$$

¿Cuál es la probabilidad de que la ecuación $Ax^2 + Bx + 1 = 0$ tenga ambas raíces reales?

* **Ej. 17** — Las puntuaciones obtenidas por los estudiantes del centro A en las pruebas de selectividad se distribuyen $N(6.25, 1)$, mientras que las de los estudiantes del centro B se distribuyen $N(6, 1.5)$. Elegimos al azar 2 estudiantes del centro A y 3 del centro B. Se pide:

1. la probabilidad de que la puntuación media de los 2 estudiantes de A sea superior a la puntuación media de los 3 de B, y
2. la probabilidad de que al escoger al azar uno de los 5 estudiantes, su nota sea superior a 6.5

** **Ej. 18** — Calcular la probabilidad de poder formar un triángulo con dos puntos elegidos en el intervalo $[0,1]$ según los distintos métodos que se enumeran a continuación.

1. Los dos puntos se eligen al azar en el intervalo.
2. Elegimos primero un punto al azar y a continuación elegimos el segundo punto, también al azar, pero sobre el trozo mayor.
3. Elegimos un punto al azar y a continuación uno de los trozos, elegido al azar, lo dividimos en dos partes iguales. Calcular, para este método, la probabilidad de que el triángulo sea obtuso.

*** **Ej. 19** — Elegimos al azar tres puntos A, B, C sobre la circunferencia de un círculo. Calcular la probabilidad de que al menos uno de los ángulos del triángulo que determinan sea mayor que $x\pi$, $0 < x < 1$.

*** **Ej. 20** — Sea $X \sim U(0,1)$. Calcular $P(X \leq x, X^2 \leq y)$ y $P(X \leq x | X^2 = y)$.

** **Ej. 21** — En una urna hay una bola roja. Extraemos tres cartas de una baraja francesa(52 cartas repartidas en 4 palos) y añadimos a la urna tantas bolas verdes como ases hayamos extraído. A continuación lanzamos 2 veces una moneda cuya probabilidad de cara es $p = 1/5$ y añadimos tantas bolas rojas como cruces hayamos obtenido. Finalmente llevamos a cabo 2 extracciones con reemplazamiento de la urna. Si X es el número de bolas verdes añadidas a la urna e Y el número de bolas rojas añadidas a la urna,

1. Obtener la función de probabilidad de X.
2. Obtener la función de probabilidad de Y.
3. Si las dos bolas extraídas con reemplazamiento son rojas, ¿cuál es la probabilidad de no haber obtenido ningún as al extraer las 3 cartas de la baraja francesa?

Ej. 22 — Elegimos al azar tres números X_1, X_2 y X_3 en [0,1], independientemente unos de otros.

1. ¿Cuál es la probabilidad de que $\{X_1 > X_2 + X_3\}$?
2. ¿Cuál es la probabilidad de que el mayor de los tres supere la suma de los otros dos?

1. $\frac{1}{6}$.
2. $\frac{1}{2}$.

Independencia de variables aleatorias

* **Ej. 23** — La distribución conjunta de las variables aleatorias X e Y es uniforme en el rombo con vértices en $(1,0), (-1,0), (0,1)$ y $(0,-1)$.

1. Sin hacer ningún cálculo, decide si X e Y son independientes.

2.Escribe la función de densidad conjunta de X e Y .

3.Calcula la función de densidad marginal de X y represéntala.

4.Utilizando las funciones de densidad conjunta y marginal obtenidas en los apartados anteriores decidir si son independientes las variables.

* **Ej. 24** — Consideremos el espacio de probabilidad $(\Omega, \mathbb{A}, P)$ y sean A_1 y A_2 dos sucesos. Definimos $X_1 = \mathbf{1}_{A_1}$ y $X_2 = \mathbf{1}_{A_2}$, las funciones indicatrices asociadas. Demostrar que X_1 y X_2 son independientes si y sólo si A_1 y A_2 lo son.

** **Ej. 25** — Sean X_1 y X_2 variables aleatorias independientes tales que $X_i \sim Po(\lambda_i)$. Se pide:

1.Demostrar que $X_1 + X_2 \sim Po(\lambda_1 + \lambda_2)$.

2.$X_1|X_1 + X_2 = k \sim Bi(k, \frac{\lambda 1}{\lambda_1+\lambda_2})$

Ej. 26 — Lanzamos una moneda, cuya probabilidad de cara es p, N veces, con $N \sim Po(\lambda)$. Sean X e Y el número de caras y el número de cruces, respectivamente. Se pide encontrar la distributición de X y de Y. ¿Son independientes las variables?

Función de un vector aleatorio

* **Ej. 27** — Sean X e Y variables aleatorias independientes, ambas geométricas con igual parámetro p.

(a)Obtener la distribución de $X + Y$.

(b)Sin efectuar ningún cálculo, ¿qué piensas que valdrá $P(X = i|X + Y = n)$? Puedes pensar en lanzamientos sucesivos de una moneda con probabilidad de cara p. Si la segunda cara ha ocurrido en el lanzamiento $n+2$, ¿cuál es la función de probabilidad del lanzamiento en el que ocurre la primera cara?

(c)Obtén la probabilidad anterior y comprueba la conjetura que hayas hecho.

* **Ej. 28** — Si X e Y son variables aleatorias con función de densidad conjunta $f(x, y) = 2$ si $0 < x < y < 1$ (0 en otro caso), determina la función de densidad de $Z = X+Y$. Resuelve este problema mediante dos procedimientos distintos:

1.Obteniendo primero la función de distribución de Z.

2.Mediante el teorema de transformaciones de variables.

En ambos casos haz todas las representaciones gráficas oportunas.

* **Ej. 29** — Sean X e Y dos v.a. independientes que siguen una distribución exponencial de parámetro λ. Halla la función de densidad de las dos variables aleatorias siguientes (halla primero la función de distribución y después deriva):

1.$U = max(X, Y)$ [40]

2.$V = min(X, Y)$ [41]

[40] Sugerencia: Si u es un número positivo, $U \leq u$ si y solo si tanto $X \leq u$ como $Y \leq u$.

[41] Sugerencia: Si v es un número positivo, halla $P(V > v)$ y a partir de ella determina $P(V \leq v)$).

* **Ej. 30** — Cuando una corriente de I amperios pasa a través de una resistencia de R ohmios, la potencia generada viene dada por $W = I^2R$ vatios. Supongamos que I y R son variables aleatorias independientes con densidades

$$f_I(x) = \begin{cases} 6x(1-x), & \text{si } 0 \leq x \leq 1; \\ \\ 0, & \text{fuera.} \end{cases}$$

y

$$f_R(y) = \begin{cases} 2y, & \text{si } 0 \leq y \leq 1; \\ \\ 0, & \text{fuera.} \end{cases}$$

Hallar la densidad de W.

** **Ej. 31** — Las variables X e Y son independientes y continuas. Sea U una variable dicotómica, independiente de ambas con $P(U = 1) = P(U = -1) = 1/2$. Definimos $S = UX$ y $T = UY$. Comprobar que S y T son dependientes, pero S^2 y T^2 son independientes.

** **Ej. 32** — Sean $\{X_k, k = 1, \ldots, n\}$ variables aleatorias i.i.d. con distribución $U(0, 1)$. Tomemos $0 < x < 1$ y definamos

$$N = \min\{n \geq 1; X_1 + X_2 + \cdots + X_n > x\}.$$

Encontrar $P(N > n)$.

** **Ej. 33** — Sean X, Y, Z variables aleatorias i.i.d. con distribución $U(0, 1)$. Encontrar la densidad conjunta de XY y Z^2 y calcular $P(XY \leq Z^2)$.

* **Ej. 34** — Elegimos al azar dos puntos, X e Y, en el intervalo $[0, a]$. Calcular la función de distribución de la distancia entre ellos.

* **Ej. 35** — Si X_1 y X_2 son variables aleatorias independientes con distribución uniforme en $(0, 1)$, obtén y representa gráficamente la función de densidad de $Y = X_1 + X_2$. Resuelve este problema por dos caminos distintos: mediante la función de distribución de Y , y mediante el teorema de transformaciones de variables. En ambos casos haz todas las representaciones gráficas oportunas.

** **Ej. 36** — Sean X_1 y X_2 variables aleatorias independientes con distribución exponencial siendo sus parámetros respectivos λ_1 y λ_2. Determinar la distribución de la variable aleatoria X_1/X_2. Calcular $P(X_1 < X_2)$.

* **Ej. 37** — Si X_1 y X_2 son variables aleatorias con función de densidad conjunta $f(x_1, x_2) = 2$ en la región $0 < x_1 < x_2 < 1$, determina la función de densidad de $Y = X_1 + X_2$. Resuelve este problema por dos caminos distintos: mediante la función de distribución de Y , y mediante el teorema de transformaciones de variables. En ambos casos haz todas las representaciones gráficas oportunas.

* **Ej. 38** — Si X_1 y X_2 son variables aleatorias independientes con distribución gamma $Ga(\frac{1}{2}, \frac{1}{2})$, determina la distribución conjunta de $Y_1 = \frac{X_1}{X_1+X_2}$ e $Y_2 = X_1 + X_2$ así como sus marginales. Representa gráficamente el soporte de las variables originales y el soporte de las variables transformadas.

Ej. 39 — Las variables aleatorias X e Y son independientes ambas $U(0,1)$. Definimos $V = X + Y$ y $W = X/Y$. Obtener la densidad conjunta de V y W. ¿Son independientes?

* **Ej. 40** — Determina la función de densidad de la suma de dos variables aleatorias independientes, ambas con distribución gamma de parámetros $\alpha = \beta = 2$. Utiliza para ello el teorema de cambio de variables, representando gráficamente el soporte de las variables originales y el soporte de las variables transformadas que hayas utilizado.

* **Ej. 41** — ([13])Sean X e Y variables aleatorias con densidad conjunta dada por $f(x,y) = e^{-y}$ si $0 < x < y < +\infty$ y cero en otro caso. Determinar la densidad de $Z = \max\{X,Y\}$. [42]

[42] Está resuelto en la referencia indicada.

* **Ej. 42** — Sean $X_1, X_2, \ldots X_n$ n variables aleatorias independientes e idénticamente distribuidas $U(0,1)$. Halla la función de densidad de

1.$U = max\{X_1, X_2, \ldots X_n\}$.

2.$V = min\{X_1, X_2, \ldots X_n\}$.

** **Ej. 43** — Sean X e Y variables aleatorias con densidad conjunta

$$f(x,y) = \frac{1}{x^2y^2}$$

si $x, y \geq 1$ y cero en otro caso. Se pide:

1.La densidad conjunta de $U = XY$ y de $V = X/Y$.

2.Las densidades marginales de U y V.

** **Ej. 44** — Sean X e Y variables aleatorias independientes y uniformemente distribuidas en el intervalo unitario $[0,1]$. Se pide determinar la densidad conjunta de $U = X + Y$ y $V = X/Y$ y las correspondientes marginales.

De todo un poco

[5]

* **Ej. 45** — Dos personas lanzan, cada una de ellas, n veces una moneda. Obtener la probabilidad de que ambas obtengan el mismo número de caras.

** **Ej. 46** — Realizamos n pruebas independientes. En cada una de las pruebas podemos tener uno de tres posibles resultados: 0, 1 y 2, con probabilidades respectivas p_0, p_1, p_2. Determinar la probabilidad de que los resultados 1 y 2 ocurren al menos una vez.

* **Ej. 47** — Tenemos dos líneas de comunicación telefónica paralelas de longitud l, separadas por una distancia $d < l$. Sabemos que, al azar y de manera independiente, se producen sendas roturas a lo largo del recorrido de cada una de ellas. Encontrar la probabilidad de que la distancia R entre ambas roturas no sea superior a c.

[5] Lo tocan todo un poco y no hay manera de meterlos en las secciones anteriores.

** **Ej. 48** — **Los sorteos de La Primitiva**Un asiduo de La Primitiva anda un tanto preocupado al comprobar que en los 18 últimos sorteos hay algunos números que no han sido extraídos. Piensa que las 108 extracciones (18×6) suponen un poco más del doble de 49 y que por tanto cada número debería haber aparecido, aproximadamente, unas dos veces. ¿Y si el sorteo no fuera correcto? ¿Habrá números más probables que otros?

* **Ej. 49** — En un almacén tenemos dos tipos de baterías. La duración de las baterías de tipo I (con $i = 1, 2$) sigue una distribución exponencial con parámetro λ_i. Elegimos al azar una batería del almacén. La proporción de baterías del tipo i en el almacén es p_i de modo que $p_1 + p_2 = 1$. Supongamos que la batería elegida está en funcionamiento después de t horas de uso. ¿Cuál es la probabilidad de que siga en funcionamiento s horas adicionales?

** **Ej. 50** — Sea (X, Y) un punto aleatorio del plano cuyas coordenadas son independientes con distribución N(0,1). Obtener la distribución conjunta y las distribuciones marginales de las coordenadas polares (R, Θ) del punto. ¿Son independientes R y Θ?

** **Ej. 51** — Las variables (X, Y) tienen como densidad conjunta

$$f_{XY}(x, y) = \begin{cases} \frac{3}{2}(x + y), & (x, y) \in D = \{(-1, 1) \times (0, 2), -2x < y < 1 - x\}; \\ 0, & \text{en el resto.} \end{cases}$$

Obtener la densidad de $U = X + Y$. ¿Son U e Y independientes?

3.10 Soluciones a los ejercicios

Solución (Ej. 4) — 1.$P(X \geq n) = (1-p)^n$.

2.$P(z \geq n) = (1-p)^{2n}$.

3.$P(Z = k) = (1-p)^{2k}(1-(1-p)^2)$.

Solución (Ej. 5) — 1.$P(R = r|M = m) = \binom{m}{r}\gamma^r(1-\gamma)^{m-r}$.

2.$P(R = r, M = m) = \binom{m}{r}\gamma^r(1-\gamma)^{m-r}\binom{n}{m}p^m(1-p)^{n-m}$.

3.$P(R = r) == \binom{n}{r}(\gamma p)^r((1-\gamma)p + (1-p))^{n-r}$.

4.$P(M = m|R = r) = \frac{\binom{m}{r}\binom{n}{m}}{\binom{n}{r}}\frac{(p(1-\gamma))^{m-r}(1-p)^{n-m}}{(1-\gamma p)^{n-r}}$ cuando $m \geq r$, en otro caso la función vale 0.

Solución (Ej. 7) — $\frac{2}{3}$.

Solución (Ej. 16) — $\frac{11}{160}$.

Solución (Ej. 17) — 1.0.5987.

2.0.3655.

Solución (Ej. 18) — 1.$\frac{1}{4}$.

2.$2\ln(2) - 1$.

3.$3 - 2\sqrt{2}$.

Solución (Ej. 26) — 1.$X \sim Po(\lambda p)$, $Y \sim Po(\lambda(1-p))$

2.Demostrar que la función de probabilidad conjunta se puede factorizar como producto de las marginales que acabamos de determinar.

Solución (Ej. 27) — (a)Si $Z = X + Y$, $f_z(n) = (n+1)p^2(1-p)^n$, $n = \{0, 1, 2, \ldots\}$.

(b)Sin respuesta para que conjeturéis.

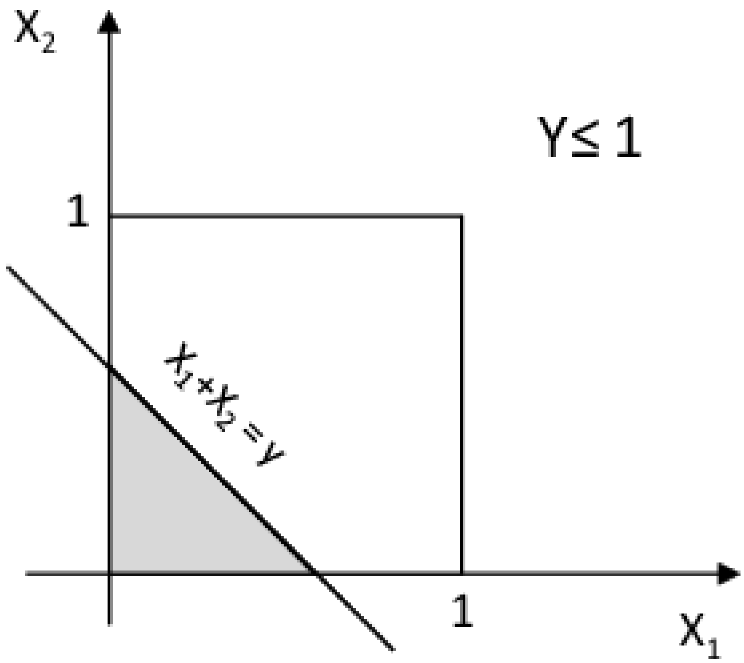
(c)Lo mismo.

Solución (Ej. 29) — (a)$f_U(u) = 2\lambda e^{-\lambda u}(1-e^{-\lambda u})$, $u > 0$ y 0 en el resto.

(b)$f_V(v) = 2\lambda e^{-2\lambda v}$, $v > 0$ y 0 en el resto.

Solución (Ej. 35) — Resolverlo mediante la función de distribución supone hacerlo por medio de probabilidades geométricas. Nos valdremos de la figura y obtendremos

$$f_Y(y) = \begin{cases} y, & 0 \leq y \leq 1; \\ 2-y, & 1 \leq y \leq 2; \\ 0, & \text{en el resto.} \end{cases}$$



Solución (Ej. 36) — La función de densidad del cociente viene dada por

$$f(y) = \frac{\lambda_1 \lambda_2}{(\lambda_1 y_1 + \lambda_2)^2}, \quad y > 0.$$

Solución (Ej. 38) — La función de densidad conjunta viene dada por

$$f_{Y_1 Y_2}(y_1, y_2) = [2e^{-2y_2}][\pi^{-1} y_1^{-1/2} (1-y_1)^{-1/2}], \quad y_2 \geq 0, \quad 0 < y_1 < 1$$

de donde $Y_2 \sim Exp(2)$ e $Y_1 \sim Beta(1/2, 1/2)$ y son además independientes.
El resultado puede generalizarse. Si $X_1 \sim Gamma(\alpha, \lambda)$ y $X_2 \sim Gamma(\beta, \lambda)$ e independientes, entonces

$$Y_1 \sim Beta(\alpha, \beta) \qquad Y_2 \sim Gamma(\alpha + \beta, \lambda).$$

Recordemos que una $Exp(\lambda)$ es una $Gamma(1, 1/\lambda)$.

Solución (Ej. 39) — La densidad conjunta de (X, Y) viene dada por

$$f(x, y) = \begin{cases} 1, & (x, y) \in [0, 1]^2; \\ 0, & \text{en el resto.} \end{cases}$$

por ser independientes.
Para obtener la densidad conjunta de (V, W) utilizaremos el teorema del cambio de variable. La transformación inversa es

$$X = \frac{VW}{1+W}, \qquad Y = \frac{V}{1+W},$$

cuyo jacobiano vale

$$|J| = \frac{V}{(1+W)^2}.$$

La distribución conjunta de (V, W) vale

$$g(v,w) = \begin{cases} \frac{v}{(1+w)^2}, & 0 < v < 2, 0 < w; \\ 0, & \text{en el resto.} \end{cases}$$

La respectivas marginales,

$$g_V(v) = \int_0^{+\infty} \frac{v}{(1+w)^2} dw = \frac{1}{2}v, \quad 0 < v < 2,$$

y vale 0 fuera.

$$g_W(w) = \int_0^2 \frac{v}{(1+w)^2} dv = \frac{2}{(1+w)^2}, \quad 0 < w,$$

y vale 0 fuera.
Son independientes porque $g(v,w) = g_V(v)g_W(w), \forall (v,w)$.

Solución (Ej. 40) — Conviene hacer la misma transformación del problema anterior y aplicar el resultado.

Solución (Ej. 42) — Resuelto en el tema.

Solución (Ej. 43) — La densidad conjunta de (U, V) es

$$f_{UV}(u,v) = \frac{1}{2u^2v}, \quad (u,v) \in D_{UV}$$

con

$$D_{UV} = \{u \geq 1, 1/u < v < u\}.$$

La marginal de U es

$$f_U(u) = \frac{\log u}{u^2}, \quad u \geq 1,$$

y la de V es

$$f_V(v) = \begin{cases} \int_{1/v}^{+\infty} \frac{1}{2u^2v} du = \frac{1}{2}, & v \in [0,1]; \\ \\ \int_v^{+\infty} \frac{1}{2u^2v} du = \frac{1}{2v^2}, & u \in]1, +\infty]. \end{cases}$$

Solución (Ej. 44) — La densidad conjunta de (U, V) viene dada por

$$f_{UV}(u, v) = \frac{u}{(1+v)^2}, \quad (u, v) \in D_{UV}.$$

con

$$D_{UV} = \{[0,1] \times [0, +\infty[\} \cup \{[1,2] \times [U-1, 1/(U-1)]\},$$

Solución (Ej. 46) — $1 - [(p_2 + p_3)^n + (p_1 + p_3)^n - p_0^n] = 1 + p_0^n - (p_2 + p_3)^n - (p_1 + p_3)^n$.

Solución (Ej. 50) — La distribución conjunta tiene por densidad,

$$f_{R\Theta}(r, \theta) = \frac{1}{2\pi} r e^{-\frac{1}{2}r^2}, \quad 0 < \theta < 2\pi, \quad r \geq 0.$$

Las marginales,

$$f_R(r) = r e^{-\frac{1}{2}r^2}, \quad r \geq 0,$$

y

$$f_\Theta(\theta) = \frac{1}{2\pi}, \quad 0 < \theta < 2\pi.$$

De donde se deduce que son independientes.

Solución (Ej. 51) — La densidad conjunta de (U, Y) es

$$f_{UY}(u, y) = \frac{3}{2}u, \quad (u, y) \in D_1 = \{(0,1) \times (0,2), \frac{y}{2} < u < 1\}.$$

Las densidades marginales,

$$f_U(u) = 3u^2, \quad 0 \leq u \leq 1$$

y

$$f_Y(y) = \frac{3}{16}(4 - y^2), \quad 0 \leq y \leq 2.$$

No son independientes.

3.11 Material complementario

En esta sección se obtienen las distribuciones de ciertos estadísticos de un gran interés en Estadística. Se incluye más por su interés en posteriores cursos de Estadística.

Ejemplo 3.28 (Suma de gammas independientes) *Supongamos que X_1 y X_2 son variables aleatorias independientes gamma con parámetros (α_i, β), $i = 1, 2$. La densidad de $U = X_1 + X_2$ la podemos obtener aplicando (3.45). Como las variables son independientes podemos factorizar la densidad conjunta y*

$$\begin{aligned} f_{X_1X_2}(v, u-v) &= f_{X_1}(v)f_{X_2}(u-v) = \\ &\left[\frac{1}{\Gamma(\alpha_1)\beta^{\alpha_1}}v^{\alpha_1-1}e^{-v/\beta}\right]\left[\frac{1}{\Gamma(\alpha_2)\beta^{\alpha_2}}(u-v)^{\alpha_2-1}e^{-(u-v)/\beta}\right] = \\ &\frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)\beta^{\alpha_1+\alpha_2}}e^{-u/\beta}v^{\alpha_1-1}(u-v)^{\alpha_2-1}, \end{aligned} \tag{3.48}$$

y de aquí, teniendo en cuenta que $0 \leq v \leq u$,

$$\begin{aligned} f_U(u) &= \int_0^u f_X(v, u-v)dv = \\ &\frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)\beta^{\alpha_1+\alpha_2}}e^{-u/\beta}\int_0^u v^{\alpha_1-1}(u-v)^{\alpha_2-1}dv. \end{aligned} \tag{3.49}$$

Haciendo el cambio $z = v/u$, $0 \leq z \leq 1$, la integral del último miembro quedará de la forma,

$$\begin{aligned} \int_0^u v^{\alpha_1-1}(u-v)^{\alpha_2-1}dv &= u^{\alpha_1+\alpha_2-1}\int_0^1 z^{\alpha_1-1}(1-z)^{\alpha_2-1}dz = \\ &u^{\alpha_1+\alpha_2-1}\frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1+\alpha_2)}. \end{aligned} \tag{3.50}$$

Sustituyendo en (3.49),

$$f_U(u) = \frac{1}{\Gamma(\alpha_1+\alpha_2)\beta^{\alpha_1+\alpha_2}}u^{\alpha_1+\alpha_2-1}e^{-u/\beta}.$$

Es decir, U sigue una distribución gamma, $U \sim Gamma(\alpha_1+\alpha_2, \beta)$.

Reiterando el proceso para un vector con k componentes independientes, $X_i \sim Gamma(\alpha_i, \beta)$, encontraríamos que la suma de sus componentes, $U = X_1 + X_2 + \cdots + X_k$, es una distribución $Gamma(\sum_{i=1}^k \alpha_i, \beta)$.

Dos consecuencias inmediatas de este resultado:

1. *La suma de k exponenciales independientes con parámetro común λ es una $Gamma(k, 1/\lambda)$.*

2. *Para $X_i \sim N(0,1)$, $i = 1, \ldots, k$, independientes, sabemos por el ejemplo 2.13 que cada $X_i^2 \sim \chi_1^2$, por tanto $Y = \sum_{i=1}^k X_i^2 \sim \chi_k^2$.*

Ejemplo 3.29 (Combinación lineal de normales independientes) *Sean $X_1 \sim N(\mu_1, \sigma_1^2)$ y $X_2 \sim N(\mu_1 2, \sigma_1 2^2)$ variables independientes. Por el Lema 2.3 sabemos que las variables $Y_i = a_iX_i \sim N(a_i\mu_i, a_i^2\sigma_i^2)$,*

$i = 1, 2$. *Por otra parte, el anterior corolario nos asegura la independencia de Y_1 e Y_2 y la distribución conjunta de ambas variables tendrá por densidad el producto de sus respectivas densidades,*

$$f_{Y_1Y_2}(y_1, y_2) = \frac{1}{2\pi a_1\sigma_1 a_2\sigma_2} \exp\left\{-\frac{1}{2}\left[\left(\frac{y_1 - a_1\mu_1}{a_1\sigma_1}\right)^2 + \left(\frac{y_2 - a_2\mu_2}{a_2\sigma_2}\right)^2\right]\right\}. \quad (3.51)$$

Si hacemos $U = Y_1 + Y_2$ y $V = Y_1$ y aplicamos (3.45),

$$f_U(u) = \int_{-\infty}^{+\infty} f_{Y_1Y_2}(v, u - v)\, dv, \quad (3.52)$$

con

$$f_{Y_1Y_2}(v, u - v) = \frac{1}{2\pi a_1\sigma_1 a_2\sigma_2} \exp\left\{-\frac{1}{2}\left[\left(\frac{v - a_1\mu_1}{a_1\sigma_1}\right)^2 + \left(\frac{(u - v) - a_2\mu_2}{a_2\sigma_2}\right)^2\right]\right\}. \quad (3.53)$$

Simplifiquemos la forma cuadrática del exponente,

$$q(u, v) = -\frac{1}{2}\left[\left(\frac{v - a_1\mu_1}{a_1\sigma_1}\right)^2 + \left(\frac{(u - v) - a_2\mu_2}{a_2\sigma_2}\right)^2\right] = -\frac{1}{2}\left[\left(\frac{v - a_1\mu_1}{a_1\sigma_1}\right)^2 + \left(\frac{u - (a_1\mu_1 + a_2\mu_2) - (v - a_1\mu_1)}{a_2\sigma_2}\right)^2\right]. \quad (3.54)$$

Haciendo $t = v - a_1\mu_1$ y $z = u - (a_1\mu_1 + a_2\mu_2)$ y operando, la última expresión puede escribirse

$$q(u, v) = -\frac{1}{2}\cdot\frac{(a_1\sigma_1)^2 + (a_2\sigma_2)^2}{(a_1\sigma_1)^2(a_2\sigma_2)^2}\left[\left(t - \frac{(a_1\sigma_1)^2(a_2\sigma_2)^2}{a_1\sigma_1)^2 + (a_2\sigma_2)^2}\right)^2 + \frac{z^2(a_1\sigma_1)^2(a_2\sigma_2)^2}{(a_1\sigma_1)^2 + (a_2\sigma_2)^2}\right], \quad (3.55)$$

con lo que la exponencial quedará de la forma

$$\exp\{q(z, t)\} = \exp\left\{-\frac{1}{2}\cdot\frac{(a_1\sigma_1)^2 + (a_2\sigma_2)^2}{(a_1\sigma_1)^2(a_2\sigma_2)^2}\left(t - \frac{(a_1\sigma_1)^2(a_2\sigma_2)^2}{(a_1\sigma_1)^2 + (a_2\sigma_2)^2}z\right)^2\right\}\times \exp\left\{-\frac{1}{2}\left(\frac{z}{[(a_1\sigma_1)^2 + (a_2\sigma_2)^2]^{1/2}}\right)^2\right\}. \quad (3.56)$$

Sustituyendo en 3.52,

$$f_U(u) =$$
$$\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{u-(a_1\mu_1+a_2\mu_2)}{[(a_1\sigma_1)^2+(a_2\sigma_2)^2]^{1/2}}\right)^2\right\}\times\int_{-\infty}^{+\infty}\frac{1}{a_1\sigma_1a_2\sigma_2\sqrt{2\pi}}$$
$$\exp\left\{-\frac{1}{2}\cdot\frac{(a_1\sigma_1)^2+(a_2\sigma_2)^2}{(a_1\sigma_1)^2(a_2\sigma_2)^2}\left(t-\frac{(a_1\sigma_1)^2(a_2\sigma_2)^2}{(a_1\sigma_1)^2+(a_2\sigma_2)^2}z\right)^2\right\}dt.$$
$$=\frac{1}{\sqrt{2\pi[(a_1\sigma_1)^2+(a_2\sigma_2)^2]}}\exp\left\{-\frac{1}{2}\left(\frac{u-(a_1\mu_1+a_2\mu_2)}{[(a_1\sigma_1)^2+(a_2\sigma_2)^2]^{1/2}}\right)^2\right\}\times$$
$$\int_{-\infty}^{+\infty}\frac{[(a_1\sigma_1)^2+(a_2\sigma_2)^2]^{1/2}}{a_1\sigma_1a_2\sigma_2\sqrt{2\pi}}$$
$$\exp\left\{-\frac{1}{2}\cdot\frac{(a_1\sigma_1)^2+(a_2\sigma_2)^2}{(a_1\sigma_1)^2(a_2\sigma_2)^2}\left(t-\frac{(a_1\sigma_1)^2(a_2\sigma_2)^2}{(a_1\sigma_1)^2+(a_2\sigma_2)^2}z\right)^2\right\}dt. \tag{3.57}$$

Observemos que el integrando es la función de densidad de una variable aleatoria normal con parámetros $\mu = \frac{z[(a_1\sigma_1)^2(a_2\sigma_2)^2]}{(a_1\sigma_1)^2+(a_2\sigma_2)^2}$ *y* $\sigma^2 = \frac{(a_1\sigma_1)^2+(a_2\sigma_2)^2}{(a_1\sigma_1)^2(a_2\sigma_2)^2}$*, por tanto la integral valdrá 1. En definitiva,*

$$f_U(u) =$$
$$\frac{1}{\sqrt{2\pi[(a_1\sigma_1)^2+(a_2\sigma_2)^2]}}\exp\left\{-\frac{1}{2}\left(\frac{u-(a_1\mu_1+a_2\mu_2)}{[(a_1\sigma_1)^2+(a_2\sigma_2)^2]^{1/2}}\right)^2\right\}, \tag{3.58}$$

lo que nos dice que $U = a_1X_1 + a_2X_2 \sim N(a_1\mu_1 + a_2\mu_2, (a_1\sigma_1)^2 + (a_2\sigma_2)^2)$.

El resultado puede extenderse a una combinación lineal finita de variables aleatoria normales independientes. Así, si $X_i \sim N(\mu_i, \sigma_i^2)$,

$$X = \sum_{1=1}^{n} a_iX_i \sim N\left(\sum_{i=1}^{n} a_i\mu_i, \sum_{i=1}^{n}(a_i\sigma_i)^2\right).$$

Ejemplo 3.30 (La distribución t de Student) *Consideremos las* $n+1$ *variables aleatorias* $X, X_1, X_2, \ldots, X_n$*, independientes y todas ellas* $N(0,1)$*. Definimos las variable*

$$t = \frac{X}{Y}. \tag{3.59}$$

con $Y = \sqrt{\frac{1}{n}\sum_{i=1}^{n} X_i^2}$.

Para obtener la distribución de t *observemos que de acuerdo con el ejemplo anterior,* $U = \sum_{i=1}^{n} X_i^2 \sim \chi_n^2$*. Su densidad es (ver página 72)*

$$f_U(u) = \begin{cases} 0, & si\ u \leq 0 \\ \\ \dfrac{1}{\Gamma(n/2)2^{n/2}}u^{u/2-1}e^{-u/2}, & si\ u > 0. \end{cases}$$

La variable $Y = \sqrt{U/n}$ *tendrá por densidad*

$$f_Y(y) = \begin{cases} 0, & si\ y \leq 0 \\ \\ \dfrac{n^{n/2}}{\Gamma(n/2)2^{n/2-1}} y^{n-1} e^{-y^2/2}, & si\ y > 0. \end{cases}$$

Por otra parte, por el corolario 3.1 X *e* Y *son independientes y su densidad conjunta vendrá dada por*

$$f_{(X,Y)}(x,y) = \left(\frac{2}{\pi}\right)^{1/2} \frac{\left(\frac{n}{2}\right)^{n/2}}{\Gamma(n/2)} e^{-\frac{1}{2}(x^2+ny^2)} y^{n-1},$$

para $x \in \mathbb{R}$ *e* $y > 0$. *Si hacemos el doble cambio*

$$t = \frac{X}{Y}, \qquad U = Y,$$

el Jacobiano de la transformación inversa vale $J = u$ *y la densidad conjunta de* t, U *es*

$$f_{(X,Y)}(x,y) = Ce^{-\frac{1}{2}(t^2u^2+nu^2)} u^n, \qquad t \in \mathbb{R}, \quad u > 0,$$

con $C = \left(\frac{2}{\pi}\right)^{1/2} \frac{\left(\frac{n}{2}\right)^{n/2}}{\Gamma(n/2)}$. *La densidad marginal de* t *se obtiene de la integral*

$$f_t(t) = \int_0^\infty Ce^{-\frac{1}{2}(t^2u^2+nu^2)} u^n du.$$

Haciendo el cambio $u^2 = v$

$$\begin{aligned} f_t(t) &= \int_0^\infty \frac{C}{2} e^{-v\left(\frac{1}{2}(t^2+n)\right)} v^{\frac{n-1}{2}} dv \\ &= \frac{C}{2} \left(\frac{t^2+n}{2}\right)^{-\frac{n-1}{2}} \int_0^\infty e^{-v\left(\frac{1}{2}(t^2+n)\right)} \left[v\left(\frac{1}{2}(t^2+n)\right)\right]^{\frac{n+1}{2}-1} dv \\ &= \frac{C}{2} \left(\frac{t^2+n}{2}\right)^{-\frac{n+1}{2}} \int_0^\infty e^{-z} z^{\frac{n+1}{2}-1} dz \\ &= \frac{C}{2} \left(\frac{t^2+n}{2}\right)^{-\frac{n+1}{2}} \Gamma\left(\frac{n+1}{2}\right). \end{aligned}$$

La distribución recibe el nombre de t de Student *con* n *grados de libertad. Student era el seudónimo de W. L. Gosset, estadístico inglés de siglo XIX, quién descubrió por primera vez esta distribución de manera empírica.*

El interés de la t *de Student reside en que surge de forma natural al estudiar la distribución de algunas características ligadas a una muestra de tamaño* n *de una variable* $N(\mu, \sigma^2)$. *Si representamos gráficamente la densidad para distintos valores de* n *comprobaremos que su forma es muy parecida a la de una* $N(0,1)$.

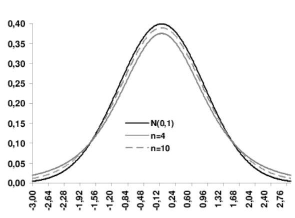


Figura 3.9: Densidades de la t de Student.

Se observa en la figure 3.9 que a medida que n *aumenta la curva de la* t *de Student se asemeja más a la de la normal. No es casual este comportamiento, en efecto, la densidad de* t *puede escribirse de* *la forma,*

$$f_t(t) = \frac{\left(\frac{n}{2}\right)^{n/2}}{\sqrt{2\pi}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(\frac{t^2+n}{2}\right)^{-\frac{n+1}{2}} =$$

$$\frac{1}{\sqrt{2\pi}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(1+\frac{t^2}{n}\right)^{-\frac{n}{2}} \left(\frac{t^2+n}{2}\right)^{-\frac{1}{2}} \quad (3.60)$$

Al pasar al límite cuando $n \to \infty$,

$$f_t(t) = \frac{1}{\sqrt{2\pi}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(1+\frac{t^2}{n}\right)^{-\frac{n}{2}} \left(\frac{t^2+n}{2}\right)^{-\frac{1}{2}} \longrightarrow \frac{1}{\sqrt{2\pi}} e^{-t^2/2},$$

que es la densidad de la $N(0,1)$.

Ejemplo 3.31 (La distribución F de Snedecor) *Sean $(X_1, X_2, \ldots, X_m)$ e $(Y_1, Y_2, \ldots, Y_n)$ sendos vectores aleatorios cuyas componentes son todas ellas variables aleatorias independientes $N(0,1)$. Queremos obtener la distribución de una nueva variable F definida mediante la relación*

$$F = \frac{\frac{1}{m}\sum_1^m X_i^2}{\frac{1}{n}\sum_1^n Y_j^2}.$$

Razonando de manera análoga a como lo hemos hecho en la obtención de la distribución de la t de Student, la densidad de F tiene la expresión,

$$f_F(x) \begin{cases} \dfrac{\Gamma\left(\frac{m+n}{2}\right)\left(\frac{m}{n}\right)^{m/2}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} x^{m/2-1}\left(1+\dfrac{mx}{n}\right)^{-(m+n)/2}, & x \geq 0; \\ \\ 0, & x < 0. \end{cases}$$

Esta distribución se conoce con el nombre de F de Snedecor con m y n grados de libertad *y surge al estudiar la distribución del cociente de las varianzas muestrales asociadas a sendas muestras de tamaño m y n de variables aleatorias $N(\mu_1, \sigma_1^2)$ y $N(\mu_2, \sigma_2^2)$, respectivamente.*

De la F de Snedecor se deriva una nueva variable, la z de Fisher, *mediante la transformación*

$$z = \frac{1}{2}\ln F.$$

Capítulo 4

Esperanza

4.1 Introducción

En el capítulo precedente hemos visto que la descripción completa de una variable o de un vector aleatorio nos la proporciona cualquiera de las funciones allí estudiadas. Es cierto que unas son de manejo más sencillo que otras, pero todas son equivalentes para el cometido citado.

En ocasiones no necesitamos un conocimiento tan exhaustivo y nos basta con una idea general. Ciertas características numéricas ligadas a las variables o los vectores aleatorios pueden satisfacernos. Estas cantidades son muy importantes en Teoría de la Probabilidad y sus aplicaciones, y su obtención se lleva a cabo a partir de las correspondientes distribuciones de probabilidad.

Entre estas constantes, sin duda las que denominaremos *esperanza matemática* y *varianza* son las de uso más difundido. La primera juega el papel de centro de gravedad de la distribución y nos indica alrededor de qué valor se situa nuestra variable o vector. La segunda completa la información indicándonos cuan dispersos o agrupados se presentan los valores alrededor de aquella. Existen también otras constantes que proporcionan información acerca de la distribución de probabilidad, son los llamados *momentos*, de los cuales esperanza y varianza son casos particulares. Los momentos pueden llegar a aportarnos un conocimiento exhaustivo de la variable aleatoria.

La herramienta que nos permite acceder a todas estas cantidades es el concepto de *esperanza*, también conocido como *media* o *valor medio*, del que nos ocupamos a continuación.

Dada una variable aleatoria X: ¿Qué es su esperanza o media? ¿Qué formaliza este concepto? ¿Cómo se llegó a él?

4.2 Esperanza o media

Empezamos definiendo la **esperanza** de una variable aleatoria discreta y luego veremos el caso de variable continua.

Definición 4.1 *Supongamos X una variable discreta no negativa. Definimos la esperanza de esta variable como*

$$EX = \sum_{i \geq 1} x_i P(X = x_i). \tag{4.1}$$

Diremos que existe la esperanza cuando el sumatorio anterior es finito. Si el soporte de la distribución de la variable aleatoria X es finito no tenemos ningún problema y tenemos definida su esperanza. Para un soporte numerable entonces lo que estamos pidiendo en que la correspondiente serie numérica sea convergente. ¿Y el caso general? ¿Cómo lo definimos para una variable aleatoria no necesariamente no negativa?

Sea X una variable discreta cualquiera (no necesariamente no negativa). Definimos las variables no negativas $X^+ = \max\{X, 0\}$ y $X^- = \max\{-X, 0\}$ de modo que $X = X^+ - X^-$. Supongamos que existen las esperanzas de X^+ y X^-. Si esto es así existe la esperanza de X y la definimos como

$$EX = EX^+ - EX^- = \sum_{i \geq 1} x_i P(X = x_i).$$

De un modo inmediato se tiene el siguiente resultado.

Proposición 4.1 *Existe la media de la variable aleatoria X si y solo si existe la media del valor absoluto de X, $|X|$.*

Demostración 4.1 *La prueba es inmediata si tenemos en cuenta que*

$$|X| = X^+ + X^-.$$

La esperanza recibe también el nombre de media o valor medio.[1]

Veamos ejemplos de medias en los casos más simples de distribuciones discretas.

Ejemplo 4.1 (Media de la distribución Bernoulli) *Si X es una Bernoulli entonces toma los valores 1 (éxito) y 0 (fracaso) con probabilidades p y $1-p$ respectivamente. ¿Cuál es su media?*

$$EX = 1 \times p + 0 \times (1-p) = p.$$

Ejemplo 4.2 (Uniforme discreta) *Supongamos que la variable X toma los valores $\{x_1, \ldots, x_n\}$ con la misma probabilidad, es decir, que X es uniforme $X \sim U(\{x_1, \ldots, x_n\})$ entonces su media será*

$$EX = \sum_{i=1}^{n} x_i P(X = x_i) = \sum_{i=1}^{n} x_i \frac{1}{n} = \bar{x}.$$

Vemos que si tenemos la uniforme en $\{x_1, \ldots, x_n\}$ (a esta distribución se le llama también la distribución empírica) entonces la media de esta variable no es más que la media muestral.

Ejemplo 4.3 (Media de la binomial) *El ejemplo de distribución binomial más conocido es, sin duda, cuando contamos el número de caras en n lanzamientos de una moneda correcta, X. Tenemos $X \sim Bi(n, 1/2)$. Cuando lanzamos muchas veces la moneda, por ejemplo 1000 veces, decir que vamos a obtener la mitad de veces una cara y*

[1] De hecho es la nomenclatura habitual en textos de Estadística.

la otra mitad cruz es falso.[2] *El número de caras es aleatorio y no variará cada vez que lancemos 1000 veces la moneda. Sin embargo, sí que es cierta la afirmación de que* esperamos observar 500 caras. *Veamos porqué.*

$$E(X) = \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x} =$$

$$\sum_{x=0}^{n} x \frac{n(n-1)\dots(n-x+1)}{x!} p^x (1-p)^{n-x}$$

$$= np \sum_{x=1}^{n} \frac{(n-1)\dots(n-x+1)}{(x-1)!} p^{x-1} (1-p)^{n-x} =$$

$$np \sum_{y=0}^{n-1} \binom{n-1}{y} p^y (1-p)^{n-y-1} = np. \quad (4.2)$$

Es decir que la media de la binomial es np*, número de pruebas de Bernoulli por la probabilidad de éxito en una prueba.*

Si lanzamos 1000 veces la moneda esperamos observar $1000 \times \frac{1}{2} = 500$ *caras.*[3]

Ejemplo 4.4 (Media de la distribución de Poisson) *Si* $X \sim Po(\lambda)$,

$$E(X) = \sum_{x \geq 0} x e^{-\lambda} \frac{\lambda^x}{x!} = \lambda e^{-\lambda} \sum_{x-1 \geq 0} \frac{\lambda^{x-1}}{(x-1)!} = \lambda.$$

[2]Como *La falsa monea.*
[3]¿Alguien lo dudaba?

Media de una variable continua

Hemos definido la media de una variable discreta. Su definición como hemos visto tiene que ver con sumatorios finitos y sumas de series infinitas. Cuando la variable es continua tenemos una función de densidad de probabilidad. Básicamente lo que vamos a hacer es *sustituir los sumatorios por integración*. Por lo demás la definición es la misma.

Supongamos primero que tenemos una variable continua no negativa X con densidad f. Definimos su media como

$$E[X] = \int_0^{+\infty} xf(x)dx. \tag{4.3}$$

Decimos que existe la media cuando la integral anterior es finita.

Si es una variable aleatoria continua no necesariamente no negativa entonces definimos (del mismo modo que en el caso discreto) las variables no negativas siguientes: $X^+ = \max\{X, 0\}$ y $X^- = \max\{-X, 0\}$. Son variables no negativas para las que hemos definido su media. En el caso de que exista la media tanto de X^+ como de X^- decimos que existe la media de X y la definimos como

$$EX = EX^+ - EX^-.$$

Se tiene que en el caso de que exista la media está vendrá dada por

$$E[X] = \int_{-\infty}^{+\infty} xf(x)dx. \tag{4.4}$$

Notemos que la existencia de la media supone que también existe la media de $E|X|$. De hecho la proposición 4.1 es válida para variables continuas con la misma prueba.

Ejemplo 4.5 (Sobre la media de una distribución exponencial) *Si X tiene una distribución exponencial de parámetro λ[43] entonces*

[43] Con densidad $f(x) = \lambda e^{-\lambda x}$ para $x \geq 0$ y cero en otro caso.

$$EX = \int_0^{+\infty} x\lambda e^{-\lambda x}dx,$$

pero fácilmente comprobamos (integrando por partes) que $\int_0^{+\infty} xe^{-\lambda x}dx = \frac{1}{\lambda^2}$ de donde $EX = \frac{1}{\lambda}$.

De hecho, se tiene un curioso resultado y es que la probabilidad $P(\frac{X}{EX} > x)$ no depende del parámetro λ. Es la misma para cualquier distribución exponencial. Se tiene

$$P\left(\frac{X}{EX} > x\right) = P\left(X > \frac{x}{\lambda}\right) = \int_{\frac{x}{\lambda}}^{+\infty} \lambda e^{-\lambda x}dx = e^{-\lambda\frac{x}{\lambda}} = e^{-x}. \tag{4.5}$$

Media de una función de una variable

Ya hemos definido la media de una variable aleatoria X. Supongamos que transformamos la variable y consideramos la nueva variable $g(X)$. ¿Quién será la media de la nueva variable? Si X es discreta entonces la nueva variable también es discreta y tendremos que

$$Eg(X) = \sum_i g(x_i)P(X = x_i),$$

ya que la variable $g(X)$ toma el valor $g(x_i)$ cuando X toma el valor x_i.

Si X es continua y su transformada también lo es entonces también se verifica que

$$Eg(X) = \int_{-\infty}^{+\infty} g(x)f(x)dx.$$

En el siguiente resultado vemos algunas propiedades básicas de la esperanza.

Proposición 4.2 1. *Tenemos la variable aleatoria X y dos transformaciones de la misma, g y h así como $a, b \in \mathbb{R}$. Se verifica que*

$$E[ag(X) + bh(X)] = aE[g(X)] + bE[h(X)].$$

Y, como caso particular, se sigue que

$$E(aX + b) = aE(X) + b.$$

2. $P(a \leq X \leq b) = 1 \Longrightarrow a \leq E(X) \leq b.$

3. $P(g(X) \leq h(X)) = 1 \Longrightarrow E[g(X)] \leq E[h(X)].$

4. $|Eg(X)| \leq E|g(X)|.$

Ejercicios

* **Ej. 1** — Supongamos X una variable aleatoria discreta y g una función real de variable real. Consideramos la variable aleatoria $Y = g(X)$. Se pide demostrar la siguiente igualdad

$$EY = \sum_j y_j P(Y = y_j) = \sum_i g(x_i)P(X = x_i).$$

** **Ej. 2** — Demostrar la proposición 4.2 tanto para variables discretas como para variables continuas.

Momentos de una variable aleatoria

La media es una medida de localización de la variable aleatoria. *Localizamos* con un número algo que es aleatorio. Obviamente, es una muy simple descripción del valor aleatorio. Dos variables pueden tener la misma media y ser, desde el punto de vista estocástico, muy diferentes. Tener distribuciones de probabilidad muy distintas. Sin embargo, el concepto de valor medio es muy intuitivo para nosotros. Siguiendo con estas dos ideas uno puede pretender describir de un modo más detallado a la variable aleatoria utilizando valores medios pero no solamente de la variable original sino de transformaciones de dicha variable. Además estas variables bien elegidas nos pueden describir distintos aspectos de la variable original.

Definición 4.2 (Momento de orden k) *Dada la variable aleatoria X definimos el momento de orden k o momento no central de orden k como*

$$\mu_k = EX^k.$$

Definición 4.3 (Momento central de orden k) *Dada la variable aleatoria X definimos el momento central de orden k como*

$$E(X - EX)^k.$$

En particular, la **varianza** de la variable X es el momento central de orden 2. Se denota como $var(X)$ o $V(X)$ o $\sigma^2(X)$.

Definición 4.4 (Varianza) *Si la variable aleatoria X es discreta definimos su varianza como*

$$V(X) = var(X) = \sigma^2(X) = \sum_i (x_i - EX)^2 P(X = x_i).$$

Si la variable X es continua entonces su varianza vendrá dada por

$$V(X) = var(X) = \sigma^2(X) = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx.$$

Definición 4.5 (Desviación estándar) *Dada una variable aleatoria X definimos su desviación estándar o desviación típica como*

$$\sigma_X = \sqrt{var(X)},$$

es decir, la definimos como la raiz cuadrada de la varianza.

Y finalmente introducimos los momentos factoriales. Quizás de menos uso en Estadística que los anteriores pero que tienen una gran utilidad en Probabilidad.

Definición 4.6 (Momentos factoriales) *Dada la variable aleatoria X definimos el momento factorial de orden k como $E[X(X - 1)\ldots(X - k + 1)]$.*

En la siguiente proposición mostramos algunas propiedades básicas de la varianza.

Proposición 4.3 (Propiedades de la varianza) *Sea X una variable aleatoria con varianza finita $\sigma_X^2 < \infty$. Se verifica que*

1. $var(X) \geq 0$.
2. $var(aX + b) = a^2 var(X)$.
3. $var(X) = E(X^2) - [E(X)]^2$.
4. $var(X)$ *hace mínima* $E\left[(X - a)^2\right]$.

Demostración 4.2 *Las dos primeras son inmediatas. Para la propiedad 3 tenemos*

$$var(X) = E(X - \mu)^2 = \\ E(X^2 + \mu^2 - 2\mu X) = E(X^2) + \mu^2 - 2\mu\mu = EX^2 - \mu^2. \quad (4.6)$$

donde la tercera igualdad es una consecuencia de la proposición 4.2. Finalmente para la propiedad 4 tenemos

$$E\left[(X - a)^2\right] = E\left[(X - E(X) + E(X) - a)^2\right] = \\ E\left[(X - E(X))^2\right] + E\left[(E(X) - a)^2\right] + 2E\left[(X - E(X)(E(X) - a)\right] = \\ var(X) + (E(X) - a)^2. \quad (4.7)$$

Estandarización o tipificación Una transformación de variable muy utilizada en Estadística es lo que se conoce como *estandarización* o *tipificación* de una variable. Si suponemos una variable aleatoria X con media μ y con varianza σ^2 entonces tipificar o estandarizar la variable consiste en considerar la nueva variable aleatoria

$$Z = \frac{X-\mu}{\sigma}. \tag{4.8}$$

¿Por qué esta transformación? ¿Qué se pretende al transformar X de este modo? Lo que se pretende es conseguir una variable que esté centrada en cero (que tenga media nula) y desviación estándar y varianza unitarias. Es fácil comprobar que se verifican estas propiedades.

$$EZ = E\left(\frac{X-\mu}{\sigma}\right) = E\left(\frac{1}{\sigma}X - \frac{\mu}{\sigma}\right) = \frac{1}{\sigma}\mu - \frac{\mu}{\sigma} = 0. \tag{4.9}$$

En cuanto a la varianza,

$$var(Z) = var\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma^2}\sigma^2 = 1. \tag{4.10}$$

Obviamente la desviación estándar también vale uno.

Funciones generatrices

Las funciones generatrices de probabilidad y de momentos son también valores medios de ciertas transformaciones de la variable aleatoria que tienen una gran utilidad.

Función generatriz de momentos [44] Vamos a tomar el valor medio de e^{tX} considerada como una función de t. Esta función de t tiene una gran utilidad práctica a la hora de obtener momentos como veremos inmediatamente. Además cuando existe[45] tiene una gran utilidad porque nos caracteriza la medida de probabilidad.

[44] https://en.wikipedia.org/wiki/Moment-generating_function

[45] Para ser precisos cuando existe en un intervalo abierto que contiene al origen.

Definición 4.7 (Función generatriz de momentos) *Sea X una variable aleatoria se define su función generatriz de momentos como como*

$$M_X(t) = E\big[e^{tX}\big].$$

Función generatriz de probabilidad [46]

Cuando trabajamos con variables discretas la media de la transformación t^X considerada como una función de t tiene mucho interés. En particular nos permite recuperar la función de probabilidad y, por tanto, caracteriza la distribución de la variable.

[46] https://en.wikipedia.org/wiki/Probability-generating_function

Definición 4.8 (Función generatriz de probabilidad) *Sea X una variable aleatoria discreta se define su función generatriz de probabilidad como*

$$G_X(t) = E\big[t^X\big].$$

Momentos de distribuciones importantes

Ejemplo 4.6 (Momento factorial de la distribución binomial) *Supongamos X con distribución binomial con parámetros n y p, $X \sim Bi(n,p)$. Vamos a calcular el momento factorial de primer orden. Calcularlo es simple porque*

$$E(X(X-1)) =$$
$$\sum_{x=0}^{n} x(x-1)\binom{n}{x}p^x(1-p)^{n-x} =$$
$$\sum_{x=2}^{n} x(x-1)\frac{n(n-1)\dots(n-x+1)}{x!}p^x(1-p)^{n-x} =$$
$$n(n-1)p^2\sum_{x=2}^{n}\frac{(n-2)!}{(x-2)!(n-x)!}p^{x-2}(1-p)^{n-x} =$$
$$n(n-1)p^2\sum_{x=0}^{n-2}\binom{n-2}{x}p^x(1-p)^{n-2-x} = n(n-1)p^2. \quad (4.11)$$

Ej. 3 — Si tenemos $X \sim Bi(n,p)$ determinar todos los momentos factoriales con $k = 1,\dots,n$.

Ejemplo 4.7 (Varianza de la distribución binomial) *Para obtener $var(X)$ es conveniente plantearse primero el momento factorial de primer orden dado por $E[(X(X-1)]$. Lo tenemos calculado en el ejemplo 4.6. Tenemos que*

$$E[(X(X-1)] = E(X^2) - E(X).$$

y de aquí se sigue

$$E[(X(X-1)]+E(X)-[E(X)]^2 = EX^2-EX+E(X)-[E(X)]^2 = var(X).$$

Por lo tanto

$$var(X) = n(n-1)p^2 + np - n^2p^2 = np(1-p).$$

Ejemplo 4.8 (Varianza de la distribución Poisson) *Para calcular la varianza vamos a aplicar un procedimiento similar al caso de la binomial que hemos visto en los ejemplos 4.6 y 4.7. Primero determinamos el momento factorial de primer order de la distribución de Poisson.*

$$E[X(X-1)] = \sum_{x\geq 0} x(x-1)e^{-\lambda}\frac{\lambda^x}{x!} = \lambda^2 e^{-\lambda}\sum_{x\geq 2}\frac{\lambda^{x-2}}{(x-2)!} = \lambda^2.$$

De aquí,

$$var(X) = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Veamos ejemplos de medias y varianzas de variables continuas.

Ejemplo 4.9 (Uniforme en intervalo unitario) *Si $X \sim U(0,1)$,*

$$E(X) = \int_{-\infty}^{+\infty} x f_X\, dx = \int_0^1 x\, dx = \frac{1}{2}$$

Para obtener var(X) utilizaremos la expresión alternativa, var$(X) = E(X^2) - [E(X)]^2$,

$$E(X^2) = \int_0^1 x^2\,dx = \frac{1}{3},$$

y de aquí,

$$var(X) = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}.$$

Ejemplo 4.10 (Normal estándar) *Si $X \sim N(0,1)$, como su función de densidad es simétrica respecto del origen,*

$$E(X^k) = \int_{-\infty}^{+\infty} x^k \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\,dx = \begin{cases} 0, & si\ k = 2n+1 \\ m_{2n}, & si\ k = 2n. \end{cases}$$

Ello supone que $E(X) = 0$ y var$(X) = E(X^2)$. Para obtener los momentos de orden par,

$$m_{2n} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^{2n} e^{-\frac{x^2}{2}}\,dx = \frac{2}{\sqrt{2\pi}} \int_0^{+\infty} x^{2n} e^{-\frac{x^2}{2}}\,dx.$$

Integrando por partes,

$$\begin{aligned} &\int_0^{+\infty} x^{2n} e^{-\frac{x^2}{2}}\,dx = \\ &\qquad - x^{2n-1} e^{-\frac{x^2}{2}}\Big|_0^{+\infty} + (2n-1)\int_0^{+\infty} x^{2n-2} e^{-\frac{x^2}{2}}\,dx = \\ &\qquad\qquad (2n-1)\int_0^{+\infty} x^{2n-2} e^{-\frac{x^2}{2}}\,dx, \end{aligned} \tag{4.12}$$

lo que conduce a la fórmula de recurrencia $m_{2n} = (2n-1)m_{2n-2}$ y recurriendo sobre n,

$$\begin{aligned} m_{2n} &= (2n-1)(2n-3)\cdots 1 = \\ &= \frac{2n(2n-1)(2n-2)\cdots 2\cdot 1}{2n(2n-2)\cdots 2} = \frac{(2n)!}{2^n n!}. \end{aligned}$$

La varianza valdrá por tanto,

$$var(X) = E(X^2) = \frac{2!}{2\cdot 1!} = 1.$$

Ejemplo 4.11 (Normal con parámetros μ y σ^2) *Si $Z \sim N(0,1)$ es fácil comprobar que la variable definida mediante la expresión $X = \sigma Z + \mu$ es $N(\mu, \sigma^2)$. Teniendo en cuenta las propiedades de la esperanza y de la varianza,*

$$E(X) = \sigma E(X) + \mu = \mu, \qquad var(X) = \sigma^2 var(Z) = \sigma^2,$$

que son precisamente los parámetros de la distribución.

No siempre existen los momentos. En todos los ejemplos anteriores de distribuciones importantes esto es así. Pero no siempre lo es. Veamos un ejemplo.

Ejemplo 4.12 (Distribución de Cauchy) [47] *Hemos insistido a lo largo de la exposición que precede en que la $E(X)$ existe siempre que la variable X sea absolutamente integrable. Puede ocurrir que una variable aleatoria carezca, por este motivo, de momentos de cualquier orden. Es el caso de la distribución de Cauchy.*

Decimos que X tiene una distribución de Cauchy si su función de de densidad viene dada por,

$$f_X(x) = \frac{1}{\pi}\frac{1}{1+x^2}, \ -\infty < x < +\infty.$$

Observemos que

$$E(|X|) = \frac{1}{\pi}\int_{-\infty}^{+\infty} \frac{|x|}{1+x^2}dx = \frac{2}{\pi}\int_0^{+\infty} \frac{x}{1+x^2}dx = \frac{1}{\pi}\log(1+x^2)\Big|_0^{+\infty} = +\infty. \quad (4.13)$$

Este resultado supone que no existe $E(X)$ ni ningún otro momento. [48]

[47] No siempre existen los momentos.

[48] El que no existen momentos de orden superior es inmediato a partir de la proposición 4.4.

Respecto de la existencia de los momentos se verifica el siguiente resultado.

Proposición 4.4 *Si $E(X^k)$ existe, existen todos los momentos de orden inferior.*

Demostración 4.3 *Tomamos $j < k$. Recordemos que existe EX^j si y lo solo si existe $E|X|^j$. Pero $|X|^j \leq 1+|X|^k$ y $E|X|^j \leq 1+E X|^k < +\infty$.*

4.3 Media y función de distribución

Ya hemos dicho en la introducción que el interés de los momentos de una variable aleatoria estriba en que son características numéricas que resumen su comportamiento probabilístico. Bajo ciertas condiciones el conocimiento de todos los momentos permite conocer completamente la distribución de probabilidad de la variable.

El siguiente resultado nos ofrece una forma alternativa de obtener la $E(X)$ cuando X es una variable no negativa.

Teorema 4.1 *Si X es una variable aleatoria no negativa ($X \geq 0$ a.s.)* [49] *con $E(X)$ finita. Se verifica la siguiente igualdad.*

$$E(X) = \int_0^{+\infty} P(X > x)dx = \int_0^{+\infty} P(X \geq x)dx \quad (4.14)$$

[49] Esta notación significa que $P(X \geq 0) = 1$ y viene de la expresión *a*lmost surely.

Demostración 4.4 *Veamos una prueba simple para el caso discreto. Si D_X denota el soporte de X (de hecho, $D_X \subset \mathbb{R}_+$) por ser no negativa. Supongamos que $D_X = \{x_1, x_2, \ldots\}$ y que además tenemos ordenados los valores: $x_1 < x_2 < \ldots$. Tenemos que*

$$\begin{aligned}\int_0^{+\infty} P(X > x)dx = \\ x_1 + (x_2 - x_1)P(X > x_1) + (x_3 - x_2)P(X > x_2) + \ldots = \\ x_1(1 - P(X > x_1)) + x_2(P(X > x_1) - P(X > x_2)) + \ldots = \\ x_1 P(X = x_1) + x_2 P(X = x_2) + \ldots \quad (4.15)\end{aligned}$$

[4]

Veamos ahora la prueba para una variable continua. Puesto que $E(X)$ existe, tendremos

$$E(X) = \int_0^{+\infty} x f_X(x)\,dx = \lim_{n\to+\infty} \int_0^n x f_X(x)\,dx.$$

Integrando por partes,

$$\int_0^n x f_X(x)\,dx = x F_X(x)|_0^n - \int_0^n F_X(x)\,dx = nF_X(n) - \int_0^n F_X(x)\,dx,$$

y sumando y restando n, tendremos

$$\int_0^n x f_X(x)dx = -n + nF_X(n) + n - \int_0^n F_X(x)dx = \\ -n(1-F_X(n)) + \int_0^n (1-F_X(x))dx. \quad (4.18)$$

Pero,

$$n(1-F_X(n)) = n\int_n^{+\infty} f_X(x))\,dx < \int_n^{+\infty} x f_X(x))\,dx \overset{n\to+\infty}{\longrightarrow} 0$$

al ser $E(|X|) < +\infty$. En definitiva,

$$E(X) = \int_0^{+\infty} x f_X(x)\,dx = \lim_{n\to+\infty} \int_0^n (1-F_X(x))\,dx = \int_0^{+\infty} (1-F_X(x))\,dx.$$

Finalmente hemos de probar la siguiente igualdad.

$$\int_0^{+\infty} P(X > x)dx = \int_0^{+\infty} P(X \geq x)dx.$$

[4] Una segunda prueba del caso discreto: X es una variable discreta. Si D_X es el soporte de X, $E(X)$ viene dada por: $E(X) = \sum_{x_j\in D_X} x_j f_X(x_j)$. Si $I = \int_0^{+\infty}(1-F_X(x))\,dx$; entonces $I = \sum_{k\geq 1}\int_{(k-1)/n}^{k/n} P(X > x)\,dx$. Pero $P(X > x)$ es decreciente en x y para $(k-1)/n \leq x \leq k/n$ y $\forall n$ se tiene $P(X > k/n) \leq P(X > x) \leq P(X > (k-1)/n)$. Integrando y sumando sobre k, $\frac{1}{n}\sum_{k\geq 1} P(X > k/n) \leq I$ y $\frac{1}{n}\sum_{k\geq 1} P(X > (k-1)/n)$ para todo n. Si llamamos L_n al primer miembro de la desigualdad,

$$L_n = \frac{1}{n}\sum_{k\geq 1}\sum_{j\geq k} P\left(\frac{j}{n} < X \leq \frac{j+1}{n}\right) = \frac{1}{n}\sum_{k\geq 1}(k-1)P\left(\frac{k-1}{n} < X \leq \frac{k}{n}\right). \quad (4.16)$$

Así,

$$L_n = \sum_{k\geq 1}\frac{k}{n}P\left(\frac{k-1}{n} < X \leq \frac{k}{n}\right) - \sum_{k\geq 1}\frac{1}{n}P\left(\frac{k-1}{n} < X \leq \frac{k}{n}\right) = \\ \sum_{k\geq 1}\frac{k}{n}\left[\sum_{\frac{k-1}{n}<x_j\leq\frac{k}{n}} f_X(x_j)\right] - \frac{1}{n}P\left(X > \frac{1}{n}\right) \geq \sum_{k\geq 1}\left[\sum_{\frac{k-1}{n}<x_j\leq\frac{k}{n}} x_j f_X(x_j)\right] - \frac{1}{n} = E(X) - \frac{1}{n}, \quad (4.17)$$

para todo n. Análogamente se comprueba que el tercer miembro de (4.19), U_n, está acotado por $U_n \leq E(X) + \frac{1}{n}$. Reuniendo todas las desigualdades se llega al resultado enunciado,

$$E(X) - \frac{1}{n} \leq \int_0^{+\infty}(1-F_X(x))\,dx \leq E(X) + \frac{1}{n},\ \forall n.$$

Notemos que

$$P(X \geq x) - P(X > x) = P(X = x).$$

Pero si X es continua entonces $P(X = x) = 0$ para cualquier x. Si X es discreta los puntos donde la probabilidad es positiva es un conjunto numerable y dos funciones que difieren en una cantidad numerable de puntos tienen la misma integral.

Teorema 4.2 *Si X es una variable aleatoria tal que existe su media entonces*

$$EX = \int_0^{+\infty} P(X > x)dx - \int_{-\infty}^0 P(X < x)dx.$$

Demostración 4.5 *Recordemos que $EX = EX^+ - EX^-$. Si aplicamos el teorema anterior tenemos que*

$$EX^+ = \int_0^{+\infty} P(X^+ > x)dx = \int_0^{+\infty} P(X > x)dx.$$

$$EX^- = \int_0^{+\infty} P(X^- > x)dx = \int_0^{+\infty} P(-X > x)dx = \int_{-\infty}^0 P(X < x)dx.$$

De hecho, si tenemos en cuenta el comentario final de la prueba del teorema 4.1 podemos afirmar que

$$EX = \int_0^{+\infty} P(X \geq x)dx - \int_{-\infty}^0 P(X \leq x)dx. \tag{4.19}$$

4.4 Desigualdades

Si para $X \geq 0$ a.s. existe su esperanza, sea $\varepsilon > 0$. Tenemos por el teorema 4.1.

$$E(X) = \int_0^{+\infty} P(X \geq x)\,dx = \int_0^{\varepsilon} P(X \geq x)\,dx + \int_{\varepsilon}^{+\infty} P(X \geq x)\,dx.$$

La segunda integral en la igualdad anterior es no negativa y la función $P(X \geq x)$ es decreciente por lo que tenemos

$$E(X) \geq \int_0^{\varepsilon} P(X \geq x)\,dx \geq \int_0^{\varepsilon} P(X \geq \varepsilon)\,dx = \varepsilon P(X \geq \varepsilon),$$

y de aquí,

$$P(X \geq \varepsilon) \leq \frac{E(X)}{\varepsilon}. \tag{4.20}$$

Utilizando la desigualdad 4.20 se siguen varias desigualdades de interés.

Proposición 4.5 (Desigualdad de Markov)

$$P\left(|X| \geq \varepsilon\right) \leq \frac{1}{\varepsilon^k} E\left(|X|^k\right), \tag{4.21}$$

Demostración 4.6 *Se obtiene al sustituir en (4.20) X por $|X|^k$ y ε por ε^k. Haciendo esto tenemos que*

$$P\left(|X| \geq \varepsilon\right) = P\left(|X|^k \geq \varepsilon^k\right) \leq \frac{1}{\varepsilon^k} E\left(|X|^k\right).$$

Un caso especial de (4.21) se conoce como la *desigualdad de Chebyshev.*

Proposición 4.6 (Desigualdad de Chebyshev)

$$P(|X - E(X)| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} Var(X). \tag{4.22}$$

Demostración 4.7 *En 4.21 tomamos $k = 2$ y en lugar de X consideramos $X - E(X)$.*

Un interesante resultado se deriva de esta última desigualdad.

Proposición 4.7 *Si $var(X) = 0$, entonces $X = \mu$ a.s..*

Demostración 4.8 *Supongamos $E(X) = \mu$ y consideremos los conjuntos $A_n = \{|X - \mu| \geq 1/n\}$, aplicando (4.22)*

$$P(A_n) = P\left(|X - \mu| \geq \frac{1}{n}\right) = 0,$$

para cualquier n y de aquí $P(\cup_n A_n) = 0$ y $P(\cap_n A_n^c) = 1$. Pero

$$\bigcap_{n\geq 1} A_n^c = \bigcap_{n\geq 1}\left\{|X - \mu| < \frac{1}{n}\right\} = \{X = \mu\},$$

luego $P(X = \mu) = 1$ o, de otro modo, $X = \mu$ a.s..

La siguiente desigualdad tiene que ver con funciones convexas.[5]

Proposición 4.8 (Desigualdad de Jensen) *Si g es una función convexa entonces*

$$E(g(X)) \geq g(E(X)).$$

Demostración 4.9 *Si g es convexa sabemos que para cualquier a encontramos λ_a tal que $g(x) \geq g(a) + \lambda_a(x - a)$ para cualquier x.*[50] *En particular podemos considerar $a = E(X)$. Se tendrá que*

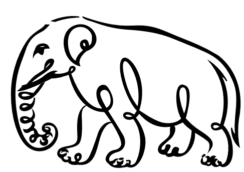
[50] Recordemos que $\{(x, y) : g(x) \leq y\}$ es un conjunto convexo.

$$g(X) \geq g(E(X)) + \lambda_{EX}(X - E(X)),$$

y tomando esperanzas obtenemos la desigualdad de Jensen

$$E(g(X)) \geq g(E(X)).$$

Ejemplo 4.13 *Si tomamos $g(x) = x^2$ que es convexa tenemos por la desigualdad de Jensen que $E(X^2) \geq (EX)^2$.*



[5] Una función g se dice convexa si para $x_1 \leq x_2$ y $0 \leq \alpha \leq 1$ entonces $g(\alpha x_1 + (1-\alpha)x_2) \leq \alpha g(x_1) + (1-\alpha)g(x_2)$. De otro modo, $\{(x, y) : g(x) \leq y\}$ es un conjunto convexo.

4.5 Esperanza de una función de un vector aleatorio

Definición 4.9 *Sea* $X = (X_1, \ldots, X_k)$ *un vector aletorio y sea* g *una función medible de* $\mathbb{R}^k$ *en* $\mathbb{R}$*. Vamos a definir la esperanza del vector en los casos en que el vector sea discreto y en el caso de que sea continuo.*

Vector aleatorio discreto.- *Si* D_X *es el soporte del vector y* f_X *su función de probabilidad conjunta, la esperanza se obtiene a partir de*

$$E(g(X)) = \sum_{(x_1,\ldots,x_k)\in D_X} g(x_1, \ldots, x_k) f_X(x_1, \ldots, x_k).$$

Vector aleatorio continuo.- *Si* f_X *es la función de densidad conjunta,*

$$E(g(X)) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} g(x_1, \ldots, x_k) f_X(x_1, \ldots, x_k)\, dx_1 \ldots dx_k.$$

Momentos de un vector aleatorio

Como ya hemos visto en el caso de una variable aleatoria, determinadas formas de la función g dan lugar a los llamados *momentos* que se definen de forma análoga a como lo hicimos entonces.

El *momento conjunto de orden* $(n_1, \ldots, n_k)$ se obtiene, siempre que la esperanza exista, para

$$g(X_1, \ldots, X_k) = X_1^{n_1} \ldots X_k^{n_k},\ n_i \geq 0. \tag{4.23}$$

Su definición formal sería la siguiente.

Definición 4.10 (Momento conjunto) *Se define el momento conjunto de orden* $(n_1, \ldots, n_k)$ *como* $E(X_1^{n_1} \ldots X_k^{n_k})$*.*

Obsérvese que los momentos de orden k respecto del origen para cada componente pueden obtenerse como casos particulares de (4.23) haciendo $n_i = k$ y $n_j = 0,\ j \neq i$, pues entonces $E(X_1^{n_1} \ldots X_k^{n_k}) = E(X_i^k)$. El *momento conjunto central de orden* $(n_1, \ldots, n_k)$ se obtienen, siempre que la esperanza exista, para $g(X_1, \ldots, X_k) = (X_1 - E(X_1))^{n_1} \ldots (X_k - E(X_k))^{n_k},\ n_i \geq 0$.

Definición 4.11 (Momento conjunto central) *Se define como* $E(X_1 - E(X_1))^{n_1} \ldots (X_k - E(X_k))^{n_k}$ *para* $n_i \geq 0$*.*

En las aplicaciones hay un momento conjunto que tiene una especial importancia.

Definición 4.12 (Covarianza) *Supongamos un vector bidimensional* (X_1, X_2)*. Se define la covarianza entre* X_1 *e* X_2 *como*

$$cov(X_1, X_2) = E(X_1 - EX_1)(X_2 - EX_2).$$

En lo que sigue veremos la razón de su importancia como medida de dependencia entre variables aleatorias. Si, en lugar de un vector bidimensional, tenemos un vector de dimensión k, $\boldsymbol{X} = (X_1, \ldots, X_k)$ se define una matriz tal que en la fila i y columna j tiene la covarianza entre las variables X_i y X_j. Si denotamos $\sigma_{ij} = cov(X_i, X_j)$ entonces podemos definir la matriz de covarianzas.

Definición 4.13 (Matriz de covarianzas) *Dado el vector aleatorio* $\boldsymbol{X} = (X_1, \ldots, X_k)$*, definimos la matriz de covarianzas como*

$$\Sigma = [\sigma_{ij}]_{i,j=1,\ldots,k}$$

Esta matriz es fundamentalmente en el estudio de las dependencias en un vector aleatorio y fundamental en Estadística multivariante.

Momentos y sumas de variables

De las funciones de un vector la que sin duda más interés tendrá es la suma de las variables. En particular tenemos el siguiente resultado básico en lo que sigue.

Teorema 4.3 (Esperanza de la suma de variables aleatorias) *Si tenemos variables* $X_1, \ldots, X_k$ *y* $a_1, \ldots, a_k$ *números reales entonces*

$$E(a_1 X_1 + \ldots + a_k X_k) = a_1 E(X_1) + \ldots + a_k E(X_k). \tag{4.24}$$

Demostración 4.10 *Hacemos la prueba en el caso continuo y para dos variables.*

$$E(a_1X_1 + a_2X_2) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(a_1x_1 + a_2x_2)f(x_1,x_2)dx_1dx_2 =$$
$$\int_{-\infty}^{+\infty} a_1x_1\left(\int_{-\infty}^{+\infty} f(x_1,x_2)dx_2\right)dx_1 + \int_{-\infty}^{+\infty} a_2x_2\left(\int_{-\infty}^{+\infty} f(x_1,x_2)dx_1\right)dx_2 =$$
$$a_1\int_{-\infty}^{+\infty} x_1f_1(x_1)dx_1 + a_2\int_{-\infty}^{+\infty} x_2f_2(x_2)dx_2. \tag{4.25}$$

Y por inducción tenemos la igualdad 4.24.

La covarianza nos permite también obtener la varianza asociada a la suma de variables aleatorias, como vemos en el siguiente resultado.

Teorema 4.4 (Varianza de la suma de variables aleatorias) *Si tenemos variables* $X_1, \ldots, X_k$ *y números reales* $a_1, \ldots, a_k$ *entonces*

$$var(a_1X_1 + \ldots + a_kX_k) = \sum_{i=1}^{k} a_i^2 var(X_i) + 2\sum_{i<j} a_ia_j cov(X_i, X_j). \tag{4.26}$$

Demostración 4.11 *Si denotamos* $S = a_1X_1 + \ldots + a_kX_k$ *entonces*

$$var(S) = E[(S - E(S))^2] = E\left(\sum_{i=1}^{k} a_i(X_i - E(X_i))\right)^2$$
$$= \sum_{i=1}^{k} a_i^2 var(X_i) + 2\sum_{i<j} a_ia_j cov(X_i, X_j). \tag{4.27}$$

Momentos y variables independientes

Es de interés especial el estudio de sumas de distintas variables aleatorias. En particular en esta sección nos planteamos determinar la media y varianza de una suma de variables aleatorias.

Proposición 4.9 1. *Si X_1 y X_2 son variables independientes entonces podemos factorizar el momento conjunto,*

$$E(X_1X_2) = EX_1EX_2.$$

y si g_1 y g_2 son funciones (medibles) entonces tendremos

$$E(g_1(X_1)g_2(X_2)) = Eg_1(X_1)Eg_2(X_2).$$

2. *De hecho, si tenemos $X_1, \ldots, X_k$ independientes y $g_1, \ldots, g_k$ funciones medibles se verifica que*

$$E(g_1(X_1)\ldots g_k(X_k)) = Eg_1(X_1)\ldots Eg_k(X_k).$$

Demostración 4.12 *Empezamos con la afirmación dada en apartado 1. Lo haremos en el caso continuo. Como siempre el caso discreto tiene una prueba análoga.*

$$E(XY) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x_1x_2f(x_1,x_2)dx_1dx_2 = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x_1x_2f_1(x_1)f_2(x_2)dx_1dx_2 = \int_{-\infty}^{+\infty} x_1f_1(x_1)dx_1 \int_{-\infty}^{+\infty} x_2f_2(x_2)dx_2. \quad (4.28)$$

Cuando tomamos las funciones g_1 y g_2 tenemos que

$$E(g_1(X_1)g_2(X_2)) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g_1(x_1)g_2(x_2)f(x_1,x_2)dx_1dx_2 = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g_1(x_1)g_2(x_2)f_1(x_1)f_2(x_2)dx_1dx_2 = \int_{-\infty}^{+\infty} g_1(x_1)f_1(x_1)dx_1 \int_{-\infty}^{+\infty} g_2(x_2)f_2(x_2)dx_2. \quad (4.29)$$

Por inducción finita se sigue el resultado del apartado 2.

Es inmediato a partir de la proposición 4.9 el siguiente resultado.

Corolario 4.1 (Covarianza y variables independientes) *Si las variables X_1 y X_2 son independientes, entonces*

$$cov(X_1, X_2) = 0.$$

Nota 4.1 (Matriz de covarianzas para variables independientes) *Si consideramos el vector aleatorio $\boldsymbol{X} = (X_1, \ldots, X_k)$ tal que las variables son independientes dos a dos se tiene que $cov(X_i, X_j) = 0$ para $i \neq j$ y que $cov(X_i, X_i) = var(X_i) = \sigma_i^2$. En resumen tenemos que la matriz de covarianzas es la matriz diagonal que en la i-ésima posición de la diagonal es igual a σ_i^2.*

Corolario 4.2 (Varianza de la suma de variables independientes) *Si las variables* $X_1, \ldots, X_k$ *son independientes y sus varianzas existen, la varianza de* $\sum_{i=1}^k X_i$ *existe y viene dada por*

$$var\left(\sum_{i=1}^k X_i\right) = \sum_{i=1}^k var(X_i).$$

Tenemos el siguiente corolario.

Corolario 4.3 (Media y varianza de la media muestral) *Sean* $X_1, \ldots, X_k$ *son independientes y con la misma media y varianza. Denotamos* $EX_i = \mu$ *y* $var(X_i) = \sigma^2$*. Consideramos la variable aleatoria definida como la media muestral* $\bar{X} = \sum_{i=1}^k \frac{X_i}{k}$*. Se tiene que la media de* $\bar{X}$ *es* μ *y que la varianza de* $\bar{X}$ *es* $\frac{\sigma^2}{k}$*.*

Demostración 4.13 *Tenemos*

$$E\bar{X} = \sum_{i=1}^k \frac{EX_i}{k} = \sum_{i=1}^k \frac{\mu}{k} = \mu$$

Tenemos por el corolario 4.2 que

$$var\left(\sum_{i=1}^k \frac{X_i}{k}\right) = \sum_{i=1}^k \frac{1}{k^2}\sigma^2 = \frac{k\sigma^2}{k^2} = \frac{\sigma^2}{k}.$$

Covarianza y correlación

Vamos a estudiar con detalle estas dos cantidades y justificar la razón de su importancia en Estadística como medida de dependencia entre pares de variables.

De hecho, teniendo en cuenta que la esperanza de la suma de variables aleatorias es la suma de las esperanzas tendremos que

$$\begin{aligned} cov(X_1, X_2) &= E\big((X_1 - E(X_1)(X_2 - E(X_2)\big) = \\ &E\big(X_1X_2 - X_1E(X_2) - X_2E(X_1) + E(X_1)E(X_2)\big) = \\ &E(X_1X_2) - E(X_1)E(X_2). \end{aligned} \tag{4.30}$$

Por tanto tenemos que

$$cov(X_1, X_2) = E(X_1X_2) - E(X_1)E(X_2).$$

La covarianza nos informa acerca del grado y tipo de dependencia existente entre ambas variables mediante su magnitud y signo, porque a diferencia de lo que ocurría con la varianza, la covarianza puede tener signo negativo.

Ejemplo 4.14 (¿Qué nos indica el signo de la covarianza?) *Para comprender el significado del signo de la covarianza, consideremos dos sucesos* A *y* B *con probabilidades* $P(A)$ *y* $P(B)$*, respectivamente. Las correspondientes funciones características,* $X = \mathbf{1}_A$ *e* $Y = \mathbf{1}_B$*, son sendas variables aleatorias cuya covarianza vale*

$$cov(X, Y) = E(XY) - E(X)E(Y) = P(A \cap B) - P(A)P(B),$$

puesto que $XY = \mathbf{1}_{A\cap B}$ *y* $E(X) = P(A)$ *y* $E(Y) = P(B)$. *Observemos que*

$$\begin{array}{llll} cov(X,Y) > 0 & si & P(A\cap B) > P(A)P(B) & \Longrightarrow & P(A|B) > P(A) \text{ y } P(B|A) > P(B) \\ cov(X,Y) = 0 & si & P(A\cap B) = P(A)P(B) & \Longrightarrow & P(A|B) = P(A) \text{ y } P(B|A) = P(B) \\ cov(X,Y) < 0 & si & P(A\cap B) < P(A)P(B) & \Longrightarrow & P(A|B) < P(A) \text{ y } P(B|A) < P(B) \end{array}$$

Así pues, una covarianza positiva supone que el conocimiento previo de la ocurrencia de B aumenta la probabilidad de A, $P(A|B) > P(A)$, *y análogamente para A. De aquí que hablemos de* dependencia positiva *entre los sucesos. En el caso contrario hablaremos de* dependencia negativa. *La covarianza vale cero cuando ambos sucesos son independientes.*

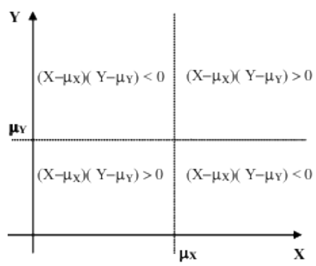


Figura 4.1: Signo de la covarianza

Cuando las variables X e Y son variables aleatorias cualesquiera, ¿qué significado hemos de darle a la expresión dependencia positiva *o* negativa*? En la figura 4.1 hemos representado los cuatro cuadrantes en los que* $\mu_X = E(X)$ *y* $\mu_Y = E(Y)$ *dividen al plano, indicando el signo que el producto* $(X - \mu_X)(Y - \mu_Y)$ *tiene en cada uno de ellos.*

Si los valores de X superiores (inferiores) a μ_X *tienden a asociarse con los valores de Y superiores (inferiores)* μ_Y, *la covarianza será positiva y las variables tendrán una relación creciente. En caso contrario, superior con inferior o viceversa, la covarianza será negativa. En este contexto hemos de entender* tienden a asociarse *como* son más probables.

En definitiva, una covarianza positiva (negativa) hemos de interpretarla como la existencia de una relación creciente (decreciente) entre las variables.

Ejemplo 4.15 (¿Qué nos indica la magnitud de la covarianza?) *La magnitud de la covarianza mide el grado de dependencia entre las variables. A mayor covarianza, medida en valor absoluto, mayor relación de dependencia entre las variables. Veamos un ejemplo.*

Lanzamos tres monedas correctas y definimos el vector aleatorio (X,Y) *con,* $X=\{$número de caras$\}$ *e* $Y=\{$número de cruces$\}$. *La tabla nos muestra la función de cuantía conjunta.*

	$Y=0$	$Y=1$	$Y=2$	$Y=3$
$X=0$	0	0	0	1/8
$X=1$	0	0	3/8	0
$X=2$	0	3/8	0	0
$X=3$	1/8	0	0	0

A partir de ella calculamos la covarianza mediante

$$cov(X,Y) = E(XY) - E(X)E(Y) = \frac{12}{8} - \frac{3}{2}\times\frac{3}{2} = -\frac{3}{4}.$$

Si definimos ahora otro vector aleatorio (X,Z) *con,* $X=\{$número de caras$\}$ *y* $Z=\{$número de cruces en los 2 primeros lanzamientos$\}$, *la función de cuantía conjunta es*

	$Z=0$	$Z=1$	$Z=2$
$X=0$	0	0	1/8
$X=1$	0	2/8	1/8
$X=2$	1/8	2/8	0
$X=3$	1/8	0	0

La covarianza vale ahora

$$cov(X,Z) = E(XZ) - E(X)E(Z) = \frac{8}{8} - \frac{3}{2} \times 1 = -\frac{1}{2}.$$

Ambas covarianzas son negativas, indicando una relación decreciente entre ambos pares de variables. Así es, puesto que se trata del número de caras y cruces en un mismo conjunto de lanzamientos. A más caras, menos cruces. La primera covarianza es mayor en valor absoluto que la segunda, lo que supone un grado de dependencia mayor entre las componentes del primer vector. En efecto, existe una relación lineal entre el primer par de variables, $X + Y = 3$, que no existe para el segundo par, que sin duda están también relacionadas pero no linealmente.

Sin embargo, el hecho de que la covarianza, como ya ocurría con la varianza, sea sensible a los cambios de escala, hace que debamos ser precavidos a la hora de valorar el grado de dependencia entre dos variables aleatorias mediante la magnitud de la covarianza. Si las variables X e Y las sustituimos por $X' = aX$ e $Y' = bY$, fácilmente comprobaremos que

$$cov(X', Y') = a \times b \times cov(X, Y).$$

¿Significa ello que por el mero hecho de cambiar de escala la calidad de la relación entre X e Y cambia? Como la respuesta es obviamente no, es necesario introducir una nueva característica numérica ligada a las dos variables que mida su relación y sea invariante frente a los cambios de escala. Se trata del *coeficiente de correlación*.

El coeficiente de correlación

Empezamos dando un resultado que necesitamos.

Teorema 4.5 (Desigualdad de Cauchy-Schwartz) *Sean X e Y variables aleatorias con varianzas finitas. Entonces $cov(X,Y)$ existe y se verifica*

$$[E(XY)]^2 \leq E(X^2)E(Y^2),$$

verificándose la igualdad si y solo si *existe un real α tal que $P(\alpha X + Y = 0) = 1$.*

Demostración 4.14 *Para cualesquiera números reales a y b se verifica*

$$|ab| \leq \frac{a^2 + b^2}{2},$$

lo que significa que $E(XY) < \infty$ si $E(X^2) < \infty$ y $E(Y^2) < \infty$. Por otra parte, para cualquier real α, se tiene

$$E[(\alpha X + Y)^2] = \alpha^2 E(X^2) + 2\alpha E(XY) + E(Y^2) \geq 0,$$

lo que supone que la ecuación de segundo grado tiene a lo sumo una raíz y su discriminante será no positivo. Es decir,

$$[E(XY)]^2 \leq E(X^2)E(Y^2).$$

Si se diera la igualdad, la ecuación tendría una raíz doble, α_0, y $E[(\alpha_0 X + Y)^2] = 0$. Tratándose de una función no negativa, esto implica que $P(\alpha_0 X + Y = 0) = 1$.

El coeficiente de correlación entre dos componentes cualesquiera de un vector aleatorio, X y Y, se define como la covarianza de dichas variables tipificadas[6].

Definición 4.14 (Coeficiente de correlación de Pearson)

$$\rho_{XY} = E\left[\left(\frac{X - E(X)}{\sigma_X}\right)\left(\frac{Y - E(Y)}{\sigma_Y}\right)\right] = \frac{cov(X,Y)}{\sigma_X \sigma_Y}.$$

Dadas las variables aleatorias X e Y, si aplicamos el teorema 4.5 a las variables $X - EX$ e $Y - EY$ tendremos que

$$[E(X - EX)(Y - EY)]^2 \leq E(X - EX)^2 E(Y - EY)^2,$$

o, equivalentemente, que

$$\rho_{XY}^2 \leq 1,$$

es decir,

$$-1 \leq \rho_{XY} \leq 1.$$

Además hemos visto que cuando $\Big(E(X - EX)(Y - EY)\Big) = E(X - EX)^2 E(Y - EY)^2$ y por lo tanto $\rho^2 = 1$ existe una constante α_0 tal que

$$\alpha_0(X - EX) + (Y - EY) = 0 \text{ a.s.}$$

De otro modo tenemos que

$$Y = \alpha_0 X - \alpha_0 EX - EY \quad \text{a.s.}$$

¿Qué ocurre cuando $\rho_{XY} = 0$? Lo primero es que $\rho_{XY} = 0$ es equivalente a que $cov(X,Y) = 0$. Si $\rho_{XY} = 0$ decimos que las variables están *incorreladas*. Hay entonces una ausencia total de relación lineal entre ambas variables. Podemos decir que el valor absoluto del coeficiente de correlación es una medida del grado de linealidad entre las variables medido, de menor a mayor, en una escala de 0 a 1.

Acabamos de ver que si $\rho_{XY} = 0$ entonces $cov(X,Y) = 0$. La implicación contraria también es cierta, lógicamente. Como la independencia entre dos variables implicaba que su covarianza era nula, deducimos que *independencia* $\rightarrow$ *incorrelación*. La implicación contraria no es cierta como puede comprobarse en el siguiente ejemplo.

Ejemplo 4.16 *Tenemos la variable aletoria X con distribución uniforme en $\{-1, 0, 1\}$. Consideramos la variable $Y = X^2$. Obviamente ambas variables son dependientes. Sin embargo,*

$$cov(X,Y) = E(XY) - EXEY = EX^3 = 0.$$

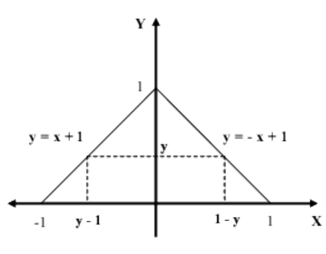


Figura 4.2: Triángulo en donde tenemos distribución uniforme.

Ejemplo 4.17 *Consideremos el vector (X,Y) uniformemente distribuido en el recinto triangular con vértices en (-1,0), (1,0) y (0,1) que podemos ver en figura 4.2.*

Se verifica

$$E(XY) = \int_0^1 y \left[\int_{y-1}^{1-y} x dx\right] dy = 0$$

y

$$E(X) = \int_0^1 \left[\int_{y-1}^{1-y} x dx\right] dy = 0$$

y por tanto $cov(X,Y) = 0$ pero X e Y no son independientes.

[6] Una variable tipificada es la que resulta de transformar la original restándole la media y dividiendo por la desviación típica, $X_t = (X - \mu_X)/\sigma_X$. Como consecuencia de esta transformación $E(X_t) = 0$ y $var(X_t) = 1$.

Momentos de algunas distribuciones

Una aplicación de los anteriores resultados permite la obtención de la esperanza y la varianza de algunas variables de manera mucho más sencilla a como lo hicimos anteriormente. Veamos algunos ejemplos.

Ejemplo 4.18 (Variable binomial como suma de variables Bernoulli) *Si recordamos las características de una variable binomial fácilmente se comprueba que si* $X \sim Bi(n,p)$ *entonces* $X = \sum_{i=1}^{n} X_i$ *con* $X_i \sim Bi(1,p)$ *e independientes. Como* $\forall i$, $E(X_i) = p$ *y* $var(X_i) = p(1-p)$, *tendremos que* $E(X) = \sum_{i=1}^{n} E(X_i) = nE(X_i) = np$, *y*

$$var(X) = \sum_{i=1}^{n} var(X_i) = np(1-p). \tag{4.31}$$

.

Ejemplo 4.19 (Variable hipergeométrica como suma de variables Bernoulli) *Si* $X \sim H(n,N,r)$ *también podemos escribir*

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

con $X_i \sim Bi(1, r/N)$ *pero a diferencia de la variable binomial las variables* X_i *son ahora dependientes. Tendremos pues*

$$E(X) = \sum_{i=1}^{n} E(X_i) = nE(X_i) = n\frac{r}{N},$$

y aplicando (4.27)

$$var(X) = \sum_{i=1}^{n} var(X_i) + 2\sum_{i=1}^{k}\sum_{j>i}^{k} cov(X_i, X_j) =$$
$$n\ \frac{r}{N}\ \frac{N-r}{N} + n(n-1)cov(X_1, X_2). \tag{4.32}$$

Pero todas las covarianzas son iguales. Para obtener $cov(X_1, X_2)$,

$$cov(X_1, X_2) =$$
$$E(X_1X_2) - E(X_1)E(X_2) = P(X_1 = 1, X_2 = 1) - \left(\frac{r}{N}\right)^2 =$$
$$P(X_2 = 1|X_1 = 1)P(X_1 = 1) - \left(\frac{r}{N}\right)^2 = \frac{r-1}{N-1}\ \frac{r}{N} - \left(\frac{r}{N}\right)^2 =$$
$$-\frac{r(N-r)}{N^2(N-1)}. \tag{4.33}$$

Sustituyendo en (4.32)

$$var(X) = n\ \frac{r}{N}\ \frac{N-r}{N}\left(1 - \frac{n-1}{N-1}\right). \tag{4.34}$$

Es interesante comparar (4.34) con (4.31), para $p = r/N$. *Vemos que difieren en el último factor, factor que será muy próximo a la unidad si* $n \ll N$. *Es decir, si la* fracción de muestreo *(así se la denomina en Teoría de Muestras) es muy pequeña. Conviene recordar aquí lo que dijimos en la página 60 al deducir la relación entre ambas distribuciones.*

Ejemplo 4.20 (Variable binomial negativa como suma de variables geométricas) *Si $X \sim BN(r,p)$ y recordamos su definición, podemos expresarla como $X = \sum_{i=1}^{r} X_i$, donde cada $X_i \sim BN(1,p)$ e independiente de las demás y representa las pruebas Bernoulli necesarias después del $(i-1)$-ésimo éxito para alcanzar el i-ésimo.*

Obtendremos primero la esperanza y la varianza de una variable geométrica de parámetro p.

$$E(X_i) = \sum_{n\geq 0} np(1-p)^n = p\sum_{n\geq 1} n(1-p)^n = \frac{1-p}{p}, \ \forall i.$$

Para calcular la varianza necesitamos conocer $E(X_i^2)$,

$$E(X_i^2) = \sum_{n\geq 0} n^2p(1-p)^n = p\sum_{n\geq 1} n^2(1-p)^n = \frac{1-p}{p^2}(2-p), \ \forall i.$$

y de aquí,

$$var(X) = \frac{1-p}{p^2}(2-p) - \left(\frac{1-p}{p}\right)^2 = \frac{1-p}{p^2}.$$

La esperanza y la varianza de la variable binomial negativa de parámetros r y p, valdrán

$$E(X) = \sum_{i=1}^{r} E(X_i) = \frac{r(1-p)}{p}.$$

y

$$var(X) = \sum_{i=1}^{r} var(X_i) = \frac{r(1-p)}{p^2}.$$

Ejemplo 4.21 (Distribución multinomial) *Si $X \sim M(n; p_1, p_2, \ldots, p_k)$, sabemos que cada componente $X_i \sim Bi(n, p_i)$, $i = 1, \ldots, k$, y puede por tanto expresarse como suma de n variables Bernoulli de parámetro p_i. La covarianza entre dos cualesquiera puede expresarse como,*

$$cov(X_i, X_j) = cov\left(\sum_{k=1}^{n} X_{ik}, \sum_{l=1}^{n} X_{jl}\right).$$

Se demuestra fácilmente que

$$cov\left(\sum_{k=1}^{n} X_{ik}, \sum_{l=1}^{n} X_{jl}\right) = \sum_{k=1}^{n}\sum_{l=1}^{n} cov(X_{ik}, X_{jl}).$$

Para calcular $cov(X_{ik}, X_{jl})$ recordemos que

$$X_{ik} = \begin{cases} 1, & \text{si en la prueba } k \text{ ocurre } A_i; \\ 0, & \text{en cualquier otro caso,} \end{cases}$$

y

$$X_{jl} = \begin{cases} 1, & \text{si en la prueba } l \text{ ocurre } A_j; \\ 0, & \text{en cualquier otro caso.} \end{cases}$$

En consecuencia, $cov(X_{ik}, X_{jl}) = 0$ si $k \neq l$ porque las pruebas de Bernoulli son independientes y

$$cov(X_{ik}, X_{jk}) = E(X_{ik}X_{jk}) - E(X_{ik})E(X_{jk}) = 0 - p_ip_j,$$

donde $E(X_{ik}X_{jk}) = 0$ porque en una misma prueba, la k-ésima, no puede ocurrir simultáneamente A_i y A_j. En definitiva,

$$cov(X_i, X_j) = \sum_{k=1}^{n} cov(X_{ik}, X_{jk}) = -np_ip_j.$$

El coeficiente de correlación (ver 4.14) entre ambas variables vale,

$$\rho_{ij} = -\frac{np_ip_j}{\sqrt{np_i(1-p_i)}\sqrt{np_j(1-p_j)}} = -\sqrt{\frac{p_ip_j}{(1-p_i)(1-p_j)}}.$$

El valor negativo de la covarianza y del coeficiente de correlación se explica por el hecho de que siendo el número total de pruebas fijo, n, a mayor número de ocurrencias de A_i, menor número de ocurrencias de A_j.

Ejemplo 4.22 (Normal bivariante) *Si (X, Y) tiene una distribución normal bivariante de parámetros μ_X, μ_y, σ_x, σ_y y ρ, ya vimos en la página 104 que $X \sim N(\mu_x, \sigma_x^2)$ e $Y \sim N(\mu_y, \sigma_y^2)$. Para simplificar la obtención de $cov(X, Y)$ podemos hacer uso de la invarianza por traslación de la covarianza (se trata de una propiedad de fácil demostración que ya comprobamos para la varianza). De acuerdo con ella, $cov(X, Y) = cov(X - \mu_x, Y - \mu_y)$ y podemos suponer que X e Y tienen media 0, lo que permite expresar la covarianza como $cov(X, Y) = E(XY)$. Procedamos a su cálculo.*

$$\begin{aligned}
E(XY) = &\\
&\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyf_{XY}(x,y)dxdy = \\
&\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xye^{-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x}{\sigma_x}\right)\left(\frac{y}{\sigma_y}\right) + \left(\frac{y}{\sigma_y}\right)^2\right\}}dxdy = \\
&\int_{-\infty}^{+\infty}\frac{y}{\sigma_y\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{y}{\sigma_y}\right)^2}\left[\int_{-\infty}^{+\infty}\frac{x}{\sigma_x\sqrt{2\pi(1-\rho^2)}}e^{-\frac{1}{2(1-\rho^2)}\left(\frac{x}{\sigma_x} - \rho\frac{y}{\sigma_y}\right)^2}dx\right]dy.
\end{aligned} \tag{4.35}$$

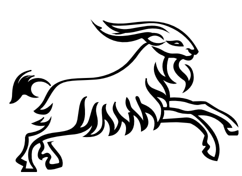
La integral interior en la ecuación 4.35 es la esperanza de una $N(\rho\sigma_x y/\sigma_y, \sigma_x^2(1-\rho^2))$ y su valor será por tanto $\rho\sigma_x y/\sigma_y$. Sustituyendo en la ecuación 4.35 tenemos

$$E(XY) = \frac{\rho\sigma_x}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\left(\frac{y}{\sigma_y}\right)^2 e^{-\frac{1}{2}\left(\frac{y}{\sigma_y}\right)^2}dy = \frac{\rho\sigma_x\sigma_y}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} z^2e^{-\frac{1}{2}z^2}dy = \rho\sigma_x\sigma_y.$$

El coeficiente de correlación (4.14) entre ambas variables es

$$\rho_{XY} = \frac{cov(X,Y)}{\sigma_x\sigma_y} = \rho.$$

Todos los parámetros de la normal bivariante adquieren ahora significado.



4.6 Esperanza condicionada

Definición 4.15 *Sea (X, Y) un vector aleatorio definido sobre el espacio de probabilidad $(\Omega, \mathbb{A}, P)$ y denotemos por $P_{X|Y=y}$ la distribución de probabilidad de X condicionada a $Y = y$.*

(X, Y) **es discreto.-** *Ello supone que el soporte de la distribución, D, es numerable y*

$$E[g(X)|y] = \sum_{x \in D_y} g(x) P(X = x|Y = y) = \sum_{x \in D_y} g(x) f_{X|Y}(x|y),$$

donde $D_y = \{x; (x, y) \in D\}$ es la sección de D mediante y.

(X, Y) **es continuo.-** *Entonces*

$$E[g(X)|y] = \int_{\mathbb{R}} g(x) f_{X|Y}(x|y) dx.$$

Ejemplo 4.23 *Sean X e Y variables aleatorias independientes ambas con distribución $Bi(n, p)$. La distribución de $X + Y$ se obtiene fácilmente a partir de*

$$f_{X+Y}(m) = P\left(\bigcup_{k=0}^{m} \{X = k, Y = m - k\}\right) = \sum_{k=0}^{m} P(X = k, Y = m-k) =$$

$$\sum_{k=0}^{m} P(X = k)P(Y = m-k) = \sum_{k=0}^{m} \binom{n}{k} p^k (1-p)^{n-k} \binom{n}{m-k} p^{m-k} (1-p)^{n-(m-k)}$$

$$= p^m (1-p)^{2n-m} \sum_{k=0}^{m} \binom{n}{k}\binom{n}{m-k} = \binom{2n}{m} p^m (1-p)^{2n-m}, \tag{4.36}$$

de donde $X + Y \sim Bi(2n, p)$. Y ahora podemos plantearnos la distribución condicionada de $Y|X + Y = m$ que vendría dada por

$$P(Y = k|X+Y = m) = \frac{P(Y = k, X + Y = m)}{P(X + Y = m)} = \frac{P(Y = k, X = m - k)}{P(X + Y = m)}$$

$$= \frac{\binom{n}{k} p^k (1-p)^{n-k} \binom{n}{m-k} p^{m-k} (1-p)^{n-(m-k)}}{\binom{2n}{m} p^m (1-p)^{2n-m}} = \frac{\binom{n}{k}\binom{n}{m-k}}{\binom{2n}{m}}, \tag{4.37}$$

es decir, $Y|X + Y = m \sim H(m, 2n, n)$. La $E(Y|X + Y = m)$ valdrá

$$E(Y|X + Y = m) = \frac{nm}{2n} = \frac{m}{2}.$$

La esperanza condicionada posee todas las propiedades inherentes al concepto de esperanza anteriormente estudiadas.

Proposición 4.10 (Propiedades de la esperanza condicionada)

1. *La esperanza condicionada es un operador lineal,*

   $E[(ag_1(X)+bg_2(X))|Y] = aE[g_1(X)|Y]+bE[g_2(X)|Y]$ (P_Y *a.s.*)

   En particular,

   $$E[(aX + b)|Y] = aE(X|Y) + b \ (P_Y \ a.s.)$$

2. *$P(a \leq X \leq b) = 1 \Longrightarrow a \leq E(X|Y) \leq b$ (P_Y a.s.)*

3. *$P(g_1(X) \leq g_2(X)) = 1 \Longrightarrow E[g_1(X)|Y] \leq E[g_2(X)|Y]$ (P_Y a.s.)*

4. *$E(c|Y) = c$ (P_Y a.s.), para c constante.*

En todas las igualdades anteriores se supone son ciertas con probabilidad uno respecto de la distribución marginal de la variable Y, P_Y.

Momentos de todo tipo de la distribución condicionada se definen de forma análoga a como hicimos en el caso de la distribución absoluta y gozan de las mismas propiedades. Existen, no obstante, nuevas propiedades derivadas de las peculiares características de este tipo de distribuciones y de que $E[g(X)|Y)]$, en tanto que función de Y, es una variable aleatoria. Veámoslas.

Proposición 4.11 *Si $E(X)$ existe, entonces*

$$E\left(E[X|Y)]\right) = E(X). \tag{4.38}$$

Demostración 4.15 *Supongamos el caso continuo.*

$$E\left(E[X|Y)]\right) = \int_{-\infty}^{+\infty} E[X|Y = y)]f_Y(y)dy =$$

$$\int_{-\infty}^{+\infty}\left[\int_{-\infty}^{+\infty} x f_{X|Y}(x|y)dx\right] f_Y(y)dy = \int_{-\infty}^{+\infty}\left[\int x\frac{f_{XY}(x,y)}{f_Y(y)}dx\right] f_Y(y)dy =$$

$$\int_{-\infty}^{+\infty} x\left[\int f_{XY}(x,y)dy\right]dx = \int_{-\infty}^{+\infty} x f_X(x)dx = E(X). \tag{4.39}$$

La demostración se haría de forma análoga para el caso discreto.

Es claro que del mismo modo tenemos que

$$E\left(E[g(X)|Y)]\right) = E(g(X)).$$

Ejemplo 4.24 *Consideremos el vector (X, Y) con densidad conjunta*

$$f_{XY}(x,y) = \begin{cases} \dfrac{1}{x}, & 0 < y \leq x \leq 1 \\ \\ 0, & \textit{en el resto.} \end{cases}$$

Fácilmente obtendremos que $X \sim U(0,1)$ y que $Y|X = x \sim U(0,x)$. Aplicando el resultado anterior podemos calcular $E(Y)$,

$$E(Y) = \int_0^1 E(Y|x)f_X(x)dx = \int_0^1 \frac{1}{2}xdx = \frac{1}{4}.$$

También es posible relacionar la varianza absoluta con la varianza condicionada, aunque la expresión no es tan directa como la obtenida para la esperanza.

Proposición 4.12 *Si $E(X^2)$ existe, entonces*

$$var(X) = E(var[X|Y]) + var(E[X|Y]). \tag{4.40}$$

Demostración 4.16 *Haciendo del anterior resultado,*

$$
\begin{aligned}
& var(X) = \\
& \quad E\left[(X - E(X))^2\right] = E\left\{E\left[(X - E(X))^2|Y\right]\right\} = \\
& \quad E\left\{E\left[\left(X^2 + (E(X))^2 - 2XE(X)\right)|Y\right]\right\} = \\
& \quad E\left\{E[X^2|Y] + E(X)^2 - 2E(X)E[X|Y]\right\} = \\
& E\left\{E[X^2|Y] - (E[X|Y])^2 + (E[X|Y])^2 + E(X)^2 - 2E(X)E[X|Y]\right\} = \\
& E\left\{var[X|Y] + (E[X|Y] - E(X))^2\right\} = E\left\{var[X|Y]\right\} + E\left\{(E[X|Y] - E(X))^2\right\} = \\
& \quad E(var[X|Y]) + var(E[X|Y]).
\end{aligned}
\tag{4.41}
$$

De la proposición 4.40 se deduce de un modo inmediato el siguiente corolario.

Corolario 4.4 *Si $E(X^2)$ existe, por la propiedad anterior*

$$var(X) \geq var(E[X|Y]).$$

Si se verifica la igualdad, de (4.40) deducimos que $E(var[X|Y]) = 0$, y como $var[X|Y] \geq 0$, tendremos que $P(var[X|Y = 0) = 1$, es decir

$$P\left(E\left\{(X - E[X|Y])^2|Y\right\} = 0\right) = 1,$$

y como $(X - E[X|Y])^2 \geq 0$, aplicando de nuevo el anterior razonamiento concluiremos que

$$P\left(X = E[X|Y]\right) = 1,$$

lo que supone que, con probabilidad 1, X es una función de Y puesto que $E[X|Y]$ lo es.

4.7 Ejercicios

Propiedades básicas de media y varianza

* **Ej. 4** — Sea X una variable aleatoria para la que existe $E(X)$ y $E(X^2)$.

1. Comprueba que $E((X - E(X))^2) = E(X^2) - (E(X))^2$.
2. Demuestra que para cualquier número real a, se cumple la relación: $E((X - a)^2) = E((X - E(X))^2) + (a - E(X))^2$ por lo que $E((X - a)^2)$ se minimiza cuando $a = E(X)$.

* **Ej. 5** — Calcula la esperanza y varianza de una variable aleatoria X con cada una de las siguientes distribuciones:

1. Distribución discreta cuyos valores con probabilidad no nula son $\{1, 2, \ldots, n\}$, todos ellos equiprobables.
2. Distribución binomial de parámetros $n = 4$ y $p = \frac{1}{2}$.

* **Ej. 6** — Si representamos por μ y σ^2 , respectivamente, la esperanza y varianza de una variable aleatoria X, expresa en términos de estos dos momentos:

1. $E(2X - 3)$ y $V(2X - 3)$.
2. El segundo momento de X y el segundo momento de X respecto de a $= 1$.
3. $V(5 - X)$.
4. $E((X - 2)(X + 1))$

* **Ej. 7** — Sea X una variable aleatoria con distribución Poisson de parámetro λ.

1. Comprueba que $E(X) = \lambda$ y $E(X(X - 1)) = \lambda^2$
2. Utilizando los resultados anteriores, calcula $E(X^2)$ y $E((X - E(X))^2)$.

* **Ej. 8** — Calcula la esperanza y varianza de una variable aleatoria X con densidad:

1. $f(x) = 6x(1 - x)$ si $0 < x < 1$ (y 0 en otro caso).
2. $f(c) = \frac{3}{x^4}$ si $x > 1$ (y 0 en otro caso)
3. $f(x)$ es proporcional a x^2 si $0 < x < 1$ (y 0 en otro caso)

* **Ej. 9** — Sea X una variable aleatoria no negativa. Demostrar que $[E(X)]^{1/2} \geq E(X^{1/2})$.

* **Ej. 10** — La esperanza de una variable aleatoria X no siempre representa un valor típico ni tampoco un valor central de su distribución.

(a) Calcula la esperanza de una variable aleatoria X cuyos únicos valores posibles (con probabilidad no nula) son $1, 2$ y 1000 con probabilidades $\frac{1}{4}, \frac{1}{2}$ y $\frac{1}{4}$.

(b) Si X tiene una distribución exponencial con parámetro λ, calcula la probabilidad de que X sea mayor que su esperanza.

(c)Si X tiene una distribución exponencial con parámetro λ, calcula el valor m para el que $P(X > m) = \frac{1}{2}$.

* **Ej. 11** — Calcula la esperanza y varianza de una variable aleatoria X con distribución

1. Uniforme en el intervalo $[\alpha, \beta]$.
2. Dada por la función de densidad $f(x) = \lambda^2 x e^{-\lambda x}$ si $x > 0$ (y 0 en otro caso)

* **Ej. 12** — Sea X una variable aleatoria con distribución uniforme en el intervalo $(0, 1)$. Calcula $E(e^{5X})$ y demuestra que $E(1/X)$ no existe.

* **Ej. 13** — Una variable aleatoria toma valores enteros positivos con probabilidades decreciendo en progresión geométrica. Elegir el primer término y la razón de la progresión para que $E(X) = 10$, y calcular, bajo dichas condiciones, $P(X \leq 10)$.

* **Ej. 14** — En una fila de 15 butacas de un cine se sientan aleatoriamente 7 mujeres y 8 hombres. Por término medio, ¿cuantas parejas de asientos adyacentes están ocupadas por personas de distinto sexo?

* **Ej. 15** — Un rebaño de ovejas es sometido a examen para detectar aquellas que padecen determinada enfermedad. Sabemos que cada una de ellas la padece con probabilidad p e independientemente de las otras. La detección de la enfermedad requiere una análisis de sangre y si se mezcla la sangre de n ovejas, basta que una de ellas padezca la enfermedad para que el análisis dé positivo. Como el rebaño es grande se plantean dos posibles estrategias:

1. Examinar los animales uno a uno y llevar a cabo tantos análisis como animales tiene el rebaño.
2. Examinar a los animales por grupos de n elementos, si el análisis es negativo todos los del grupo están sanos, pero si es positivo habrá que analizar uno a uno todos los animales del grupo.

Determinar cuál de las dos estrategias es mejor porque conduce a un número menor de análisis.

* **Ej. 16** — Dos jugadores A y B juegan al siguiente juego: lanzan una moneda hasta que aparece un cara, si la cara aparece en el k-ésimo lanzamiento el jugador B paga al jugador A k monedas. ¿Cuantas monedas deberá pagar el jugador A antes de iniciarse el juego para que éste sea equilibrado?

** **Ej. 17** — Sea X el número de pruebas de Bernoulli necesarias para obtener un éxito y un fracaso. Determinar la distribución de probabilidad de X y calcular $E(X)$ y $var(X)$.

** **Ej. 18** — Sea X una variable aleatoria discreta con soporte $D_X = \{x_1, x_2, \ldots, x_m\}$. Calcular

$$\lim_{n\to\infty} \frac{E(X^{n+1})}{E(X^n)}.$$

*** **Ej. 19** — Elegimos un número aleatorio X de la siguiente forma: lanzamos repetidamente una moneda hasta que nos muestre la primera cara, si el número lanzamientos que hemos necesitado es N, elegimos aleatoriamente un entero k en $1, 2, \ldots, 10^N$, valor que asignamos a X. Encontrar la distribución de probabilidad de X y su valor esperado.

** **Ej. 20** — **Si hay cumpleaños no se trabaja** En algunos países socialistas y en algún momento de su historia se estableció una ley laboral por la que en las factorías e instituciones estatales se trabajaba todos los días de la semana. La excepción a esta regla era que algunos de los trabajadores cumpliera años, ese día nadie trabajaba. Suponiendo un año no bisiesto y que los cumpleaños se distribuyen equiprobablemente a lo largo del año, encontrar el número de trabajadores para que el número esperado de hombres-día trabajados sea máximo.

** **Ej. 21** — Elegimos al azar un número del 1 al 10. Hemos de adivinarlo mediante preguntas cuya respuesta sea *si* o *no*. Calcular el número esperado de preguntas que debemos hacer si,

1. la pregunta i-ésima que hacemos es *>se trata del número i?*, o
2. con cada pregunta intentamos eliminar la mitad de los números restantes.

** **Ej. 22** — Un vendedor de periódicos compra cada ejemplar a 10 céntimos y los vende a 15 céntimos. Como no puede devolver los ejemplares no vendidos, trata de saber cuantos le conviene comprar para maximizar lo que espera ganar cada día. Si las demandas diarias siguen una distribución $Bi(10, 1/3)$, ¿cuántos ejemplares debe comprar?

** **Ej. 23** — Un vendedor de periódicos compra cada ejemplar a C céntimos de euro y lo vende a V céntimos de euro ($C < V$). Puede devolver los ejemplares no vendidos pero sólo obtiene R céntimos de euro por ejemplar ($R < C$). Se trata de saber cuantos le conviene comprar para maximizar lo que espera ganar cada día. Si la demanda diaria X sigue una distribución discreta con $P(X = k) = p_k$, $k = 0, 1, 2, \ldots, n$, ¿cuántos ejemplares debe comprar? Aplicar el resultado para el caso en que $V = 100$, $C = 50$, $R = 20$ y la función de probabilidad viene dada por

k	P(X=k)	k	P(X=k)
0	0.05	7	0.08
1	0.08	8	0.05
2	0.10	9	0.04
3	0.15	10	0.03
4	0.15	11	0.03
5	0.10	12	0.03
6	0.08	13	0.03

* **Ej. 24** — **Un camino aleatorio en la recta** Un punto se desplaza sobre el eje de abscisas a derecha o izquierda en saltos unitarios. La probabilidad de que se desplace a la derecha es p y $1 - p$ de que lo haga a la izquierda. Si la variable X representa la posición del punto después de n desplazamientos, partiendo de 0, obtener su distribución de probabilidad y calcular $E(X)$ y $var(X)$.

* **Ej. 25** — La función de densidad de la variable aleatoria X viene dada por

$$f(x) = \begin{cases} \dfrac{1}{2} + cx^3, & x \in]-1,1[; \\ \\ 0, & \text{en el resto.} \end{cases}$$

1. Determinar los valores de c para que $f(x)$ sea una densidad.
2. Calcular, si existen, los momentos de orden 1 y 2 de la variable $Y = |X|^{-\frac{1}{2}}$.

* **Ej. 26** — Definamos la función

$$f(x) = \begin{cases} ax^{-(s+1)}, & \text{si } x > r; \\ 0, & \text{si } x \leq r, \end{cases}$$

con $r, s > 0$.

1. Determinar a para qué $f(x)$ sea una función de densidad de probabilidad.
2. Si la variable aleatoria X tiene por densidad $f(x)$, ¿para qué valores de s existirá su esperanza?

* **Ej. 27** — Elegimos un punto al azar en la circunferencia de centro cero y de radio uno. Supongamos que denotamos por Θ la variable aleatoria que nos da el ángulo aleatorio asociado a este punto y por X la variable aleatoria que nos da la abscisa del punto. Se pide:

1. ¿Cuál es la distribución de Θ? Hay que obtener tanto la función de densidad como la función de distribución.
2. Determinar $\mathcal{P}(\Theta \in [a,b])$ para cualesquiera a y b con $0 \leq a \leq b \leq 2\pi$.
3. Determinar la función de distribución de la variable X.
4. Determinar la función de densidad de la variable X.
5. Calcular $E(X)$.
6. Calcular $E(|X|)$.

** **Ej. 28** — El juego de los *dos dedos* se juega entre dos jugadores A y B. Los jugadores muestran a la vez uno o dos dedos y simultáneamente dicen el número de dedos que su oponente va a mostrar. El que acierta recibe del que pierde tantos euros como dedos hayan mostrado entre ambos. Si ninguno acierta, o lo hacen ambos, nadie gana ni paga. Designemos por (i,j) la jugada de cualquiera de los dos jugadores, consistente en mostrar i dedos y apostar a que el contrario va a mostrar j dedos, $i,j = 1,2$. El jugador A decide jugar sólo las jugadas $(1,2)$ con probabilidad $\frac{4}{7}$ y $(2,1)$ con probabilidad $\frac{3}{7}$. Si el jugador B, independientemente de lo que haga A, juega a cualquiera de las cuatro posibles jugadas con igual probabilidad, $\frac{1}{4}$, calcular la ganancia esperada de A. ¿Podría A mejorar sus ganancias manteniendo sus jugadas?

** **Ej. 29** — Sea X una variable aleatoria tal qu $X \sim Po(\lambda)$, donde $0 < \lambda < 1$. Encuentra $E(X!)$.

* **Ej. 30** — Calcula $E(X)$ si X tiene una función de densidad dada por:

1.
$$f_X(x) = \begin{cases} \frac{1}{4}xe^{-\frac{x}{2}}, & \text{si } x > 0; \\ 0, & \text{en otro caso.} \end{cases}$$

2.
$$f_X(x) = \begin{cases} c(1-x^2), & \text{si } -1 < x < 1; \\ 0, & \text{en otro caso.} \end{cases}$$

3.
$$f_X(x) = \begin{cases} \frac{5}{x^2}, & \text{si } x > 5; \\ 0, & \text{si } x \leq 5. \end{cases}$$

* **Ej. 31** — La función de densidad de X es dada por:

$$f_X(x) = \begin{cases} a + bx^2, & \text{si } 0 \leq x \leq 1; \\ 0, & \text{en otro caso.} \end{cases}$$

Si $E(X) = \frac{3}{5}$, encuentra a y b.

Desigualdades

* **Ej. 32** — Sea X la variable aleatoria que describe el número de caras obtenidas en 100 lanzamientos de una moneda equilibrada,

1. ¿Qué dice la desigualdad de Chebyshev sobre la probabilidad de que X no esté a menos de un 10% de su valor esperado?
2. Repite el apartado anterior suponiendo que se realizan 1000 lanzamientos.
3. Repite el apartado anterior suponiendo que se realizan n lanzamientos.

* **Ej. 33** — Aunque con frecuencia la desigualdad de Chebyshev no proporciona una buena cota, en ocasiones no puede mejorarse. Sea X una variable aleatoria que puede tomar los valores $-a$, 0 y a con probabilidades $p, 1-2p$ y p respectivamente, donde $0 < p < \frac{1}{2}$ y siendo $a > 0$.

1. Calcula la esperanza y la varianza de X.
2. Utilizando la desigualdad de Chebyshev, halla una cota para $P(|X - E(X)| \geq a)$.
3. Calcula el valor exacto de $P(|X - E(X)| \geq a)$ y compáralo con la cota obtenida en el apartado anterior.

** **Ej. 34** — Si X es una variable aleatoria continua con media 0 y varianza σ^2 y función de distribución F, comprobar que se verifica

$$F(x) \geq \frac{x^2}{x^2+\sigma^2}, \quad \text{si } x > 0 \quad y \quad F(x) \leq \frac{\sigma^2}{x^2+\sigma^2}, \quad \text{si } x < 0.$$

** **Ej. 35** — Sea X una variable aleatoria no negativa con función de densidad f. Demostrar que

$$E(X^r) = \int_0^\infty r x^{r-1} P(X > x) dx.$$

** **Ej. 36** — Sea X una variable aleatoria continua con media μ y función de distribución F. Demostrar que

$$\int_{-\infty}^{\mu} F(x)dx = \int_{\mu}^{\infty} (1 - F(x))dx.$$

*** **Ej. 37** — Una caja contiene a bolas *blancas* y b bolas *negras*, con $a > 0$ y $b > 0$. Se extrae una bola al azar y si es negra el proceso termina. Si la bola es blanca se reintegra a la caja junto con otra bola blanca y el proceso continua. Calcular la probabilidad de que el proceso termine en la k-ésima extracción. Calcular la probabilidad de que el número de extracciones sea finito. Obtener el número esperado de extracciones.

* **Ej. 38** — Si $g(x)$ es una función creciente no negativa de x, probar que

$$P(X \geq a) \leq \frac{E(g(X))}{g(a)}.$$

** **Ej. 39** — Sea $X_1, X_2, \ldots$ una sucesión de variables aleatorias continuas independientes e idénticamente distribuidas. Sea $N \geq 2$ el punto en que la sucesión deja de ser decreciente, es decir, $X_1 \geq X_2 \geq \ldots \geq X_{N-1} < X_N$. Demostrar que $E(N) = e$. [51]

[51] Sugerencia: Obtener primero $P(N \geq n)$.

* **Ej. 40** — Sea X una variable aleatoria con soporte $D_X = \{0, 1, \ldots, n\}$ tal que $E(X) = var(X) = 1$. Demostrar que para cualquier natural k,

$$P(X \geq k + 1) \leq \frac{1}{k^2}.$$

* **Ej. 41** — Sea X una variable aleatoria con función de densidad

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

donde n es un natural. Demostrar que se verifica la siguiente desigualdad

$$P(0 < X < 2(n + 1)) > \frac{n}{n + 1}.$$

Momentos de vectores

* **Ej. 42** — Sean X e Y dos variables aleatorias independientes tales que: $E(X) = 2, E(X^2) = 6, E(Y) = 3$ y $E(Y^2) = 13$. Calcula:

1. $E(X^2 - 3X + 2), E((X + 1)^2), E((X - E(X))^2)$
2. $E(X + Y), E(2XY), E((3X - Y)^2)$ y $(E(3X - Y))^2$

* **Ej. 43** — Si X, Y y Z son variables aleatorias y a, b, c y d son números reales, comprueba que:

1. $Cov(aX + b, cY + d) = acCov(X, Y)$.
2. $Cov(X + Y, Z) = Cov(X, Z) + Cov(Y, Z)$
3. $Cov(X, Y) = Cov(Y, X)$

* **Ej. 44** — Si (X,Y) es un vector aleatorio con función de densidad conjunta $f(x, y) = \frac{6}{5}(x^2 + y)$ si $0 < x < 1$ y $0 < y < 1$ (y 0 en otro caso), calcula la covarianza y el coeficiente de correlación entre ambas variables.

* **Ej. 45** — (*[10, Problema 6.19]*)Demostrar que $f(x, y) = \frac{1}{x}$ si $0 < y < x < 1$ y cero en otro caso es una función de densidad conjunta. Asumiendo que efectivamente es una densidad conjunta calcular:

1. Determinar la densidad marginal de X.
2. Determinar la densidad marginal de X.
3. Calcular $E(X)$.
4. Calcular $E(Y)$.
5. Calcular $cov(X, Y)$.

* **Ej. 46** — El vector aleatorio (X, Y) está definido en el triángulo OAB, $O(0, 0)$, $A(0, 1)$, $B(1, 0)$. Su función de densidad es $f(x, y) = ky^2$, y vale 0 fuera del triángulo. Calcular el coeficiente de correlación entre ambas variables.

* **Ej. 47** — La distribución conjunta de las variables aleatorias X e Y es uniforme en el rombo con vértices en $(1, 0), (-1, 0), (0, 1)$ y $(0, -1)$.

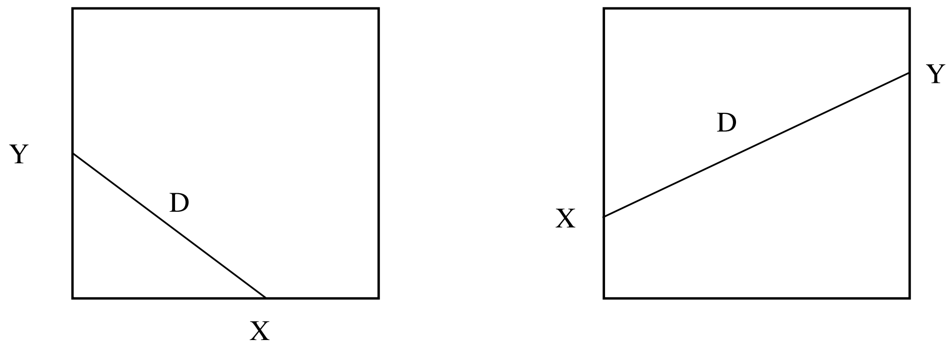
1. Calcula E(X) y V(X).
2. Calcula E(Y) y V(Y).
3. Calcula $E(XY), Cov(X, Y)$ y $\rho(X, Y)$.

* **Ej. 48** — El vector (X, Y) se distribuye uniformemente sobre el triángulo de vértices $(0, 0)$, $(a, 0)$ y $(a, 2a)$. Obtener la densidad condicionada de $X|Y$, las esperanzas y varianzas de X e Y y el coeficiente de correlación entre las dos variables.

* **Ej. 49** — Un suceso, A, asociado con cierto experimento es tal que $P(A) = 0.35$. El experimento se repite 45 veces, y asociada a la repetición i-esima, $i \in \{1, \ldots, 45\}$ se define la variable $X_i = 1$ si ocurre A y $X_i = 0$ en otro caso.

1. Obtén la distribución de $S_{45} = \sum_{i=1}^{45} X_i$.
2. Calcula $E(\overline{X}_{45})$ y $V(\overline{X}_{45})$ siendo $\overline{X}_{45} = \frac{S_{45}}{45}$ la media muestral de las 45 variables anteriores.
3. Utilizando la desigualdad de Chebyshev, calcula una cota para $P(0.3 < \overline{X}_{45} < 0.4)$
4. Calcula el valor exacto de $P(0.3 < \overline{X}_{45} < 0.4)$.

* **Ej. 50** — En dos lados distintos de un cuadrado unidad elegimos al azar e independientemente los puntos X e Y. Si por D designamos la distancia entre ellos, calcular $E(D^2)$ en las dos situaciones posibles descritas en la figura.



* **Ej. 51** — Elegimos al azar dos puntos en el intervalo $[0, a]$. Queremos conocer,

a) Area media del rectángulo que tiene por lados las correspondientes longitudes.

b) Area media de dicho rectángulo si ambos puntos son mayores que $a/2$.

** **Ej. 52** — Sean X e Y variables aleatorias independientes. Demostrar que si X y $X - Y$ son independientes, entonces X es una variable aleatoria degenerada.

* **Ej. 53** — Dos puntos se eligen al azar sobre un segmento de longitud a. Encontrar la esperanza y la varianza de la distancia entre ellos.

* **Ej. 54** — Calcular el coeficiente de correlación entre el mayor y el menor valor obtenidos al lanzar dos dados.

* **Ej. 55** — Obtener el coeficiente de correlación entre las coordenadas de un punto elegido al azar en el círculo unidad.

* **Ej. 56** — Las variables aleatorias X e Y tienen media 0, varianza 1 y su coeficiente de correlación es ρ, Encontrar la media y la varianza de $Z = X - \rho Y$ y sus coeficientes de correlación con X e Y.

* **Ej. 57** — Lanzamos tres veces consecutivas una moneda y definimos las variables aleatorias $X =${*número de caras en los dos primeros lanzamientos*} e $Y =${*número de cruces en los dos últimos lanzamientos*}. Obtener la distribución de probabilidad conjunta de X e Y, sus marginales y el coeficiente de correlación entre ambas.

* **Ej. 58** — Las variables aleatorias $X_1, X_2, \ldots, X_{m+n}$, $n > m$, son independientes e idénticamente distribuidas con varianza común, σ^2, finita. Calcular el coeficiente de correlación de $U = \sum_{i=1}^{n} X_i$ y $V = \sum_{j=m+1}^{m+n} X_j$.

*** **Ej. 59** — Una urna contiene 2^n bolas numeradas de 0 a n, de forma que hay $\binom{n}{i}$ bolas con el número i. Si extraemos m sin reemplazamiento y por S denotamos la suma de sus números, calcular $E(S)$ y $var(S)$.

** **Ej. 60** — **Un camino aleatorio en el plano** El punto se desplaza ahora sobre el plano con movimientos independientes de manera que en cada movimiento la distancia que recorre es siempre la misma, por ejemplo una unidad, pero la dirección es aleatoria determinada por un ángulo cuya orientación respecto la posición del punto se elige al azar en el intervalo $[0, 2\pi]$. Si D es la distancia del punto a su posición original después de n movimientos, calcular $E(D^2)$.

** **Ej. 61** — **Número esperado de coincidencias** Recordemos que el llamado *problema de las coincidencias* consiste, en una de sus múltiples versiones, en n individuos que al abandonar una fiesta recogen sus sombreros al azar. Si por X denotamos el número de individuos que han cogido su propio sombrero *(coincidencias)*, calcular $E(X)$ y $var(X)$.

** **Ej. 62** — **Número esperado de parejas restantes** Se trata de un problema propuesto y resuelto por Daniel Bernoulli en el siglo XVIII. Se proponía averiguar el número esperado de parejas que permanecen completas, de un total de N parejas iniciales, después de la muerte de m sus miembros.

*** **Ej. 63** — Se realizan n pruebas independientes de un experimento que tiene k resultados posibles $A_1, A_2, \ldots, A_k$ con probabilidades $p_1, p_2, \ldots, p_k$, $\sum p_i = 1$. Sea Y la variable *número de resultados que no han aparecido en ninguna de las n pruebas*.

1. ¿Qué distribución tiene cada una de las variables X_j que indica el número de veces que ha sucedido A_j? (Puedes usar el resultado visto en el tema anterior sin demostrarlo)
2. ¿Cuál es el valor esperado de Y?
3. Comprobar que los valores de $p_1, \ldots, p_k$ que hacen mínima $E(Y)$ son $p_1 = p_2 = \cdots = p_k = 1/k$.

* **Ej. 64** — Sean X e Y dos variables aleatorias idénticamente distribuidas, obtener $cov(X + Y, X - Y)$.

* **Ej. 65** — Sean $X_1, X_2, \ldots$ una familia de variables aleatorias independientes con la misma media μ y la misma varianza σ_2, y sea $Y_n = X_n + X_{n+1} + X_{n+2}$. Hallar, para $j \geq 0$, $cov(Y_n, Y_{n+j})$

* **Ej. 66** — Sean X e Y el número de éxitos y fracasos, respectivamente, cuando llevamos a cabo n pruebas Bernoulli con probabilidad de éxito p. Calcular el coeficiente de correlación entre ambas, ρ_{XY}.

** **Ej. 67** — Supongamos que n niños de alturas diferentes se colocan en fila. Elegimos el primer niño y nos desplazamos con él a lo largo de la fila hasta que encontramos un niño más alto o hasta el final de la misma. Si encontramos un niño más alto lo elegimos y continuamos con el desplazamiento y así sucesivamente. ¿Cuál es el número esperado de niños que elegiremos en la fila?

* **Ej. 68** — ¿Cuál es el número esperado de cumpleaños distintos en un grupo de 100 personas elegidas al azar?

Esperanza condicionada

** **Ej. 69** — Un trabajador está encargado del correcto funcionamiento de n máquinas situadas en linea recta y distantes una de otra l metros. El trabajador debe repararlas cuando se averían, cosa que sucede con igual probabilidad para todas ellas e independientemente de una a otra. El operario puede seguir dos estrategias:

1. acudir a reparar la máquina estropeada y permanecer en ella hasta que otra máquina se avería, desplazándose entonces hacia ella, o
2. situarse en el punto medio de la linea de máquinas y desde allí acudir a la averiada, regresando nuevamente a dicho punto cuando la avería está resuelta.

Si X es la distancia que recorre el trabajador entre dos averías consecutivas, ¿cúal de ambas estrategias le conviene más para andar menos?

* **Ej. 70** — Lanzamos una moneda, si sale cruz lanzamos un dado y recibimos tantos puntos como muestra la cara del dado. Si sale cara, lanzamos otras cinco monedas y el número de puntos es ahora el total de caras obtenidas, incluida la inicial. Calcula el número esperado de puntos obtenidos.

** **Ej. 71** — Sean X e Y dos variables aleatorias. Si $E(X|Y) = 10 - Y$ y $E(Y|X) = 7 - X/4$, obtener su coeficiente de correlación.

** **Ej. 72** — Llevamos a cabo N lanzamientos de un dado, donde N sigue una distribución de Poisson de parámetro $\lambda = 10$. Obtener la media y la varianza de la suma de las puntuaciones de las caras obtenidas en los N lanzamientos.

** **Ej. 73** — Una máquina fabrica fibras de longitud aleatoria X. Para medir la desigualdad entre las longitudes de las fibras fabricadas se utiliza el coeficiente
$$\lambda = \frac{a'' - a'}{a},$$
donde $a = E(X)$, $a'' = E(X \mid X > a)$ y $a' = E(X \mid X < a)$. Si $X \sim N(a, \sigma^2)$, encontrar la relación entre λ, a y σ.

* **Ej. 74** — El número de clientes en la cola de la caja de un supermercado sigue una distribución de Poisson con parámetro λ. El tiempo que tarda cada cliente en ser atendido sigue una Exponencial de parámetro λ. Calcular el tiempo medio que debemos esperar en la cola.

* **Ej. 75** — El número de pasajeros que espera el tren en cierta estación en un instante t es una variable aleatoria Poisson de parámetro λt. Si el tiempo de llegada del tren se distribuye uniformemente en el intervalo $[0, T]$, ¿cuál es el número medio de pasajeros que subirá al tren? Obtener también su varianza.

* **Ej. 76** — Un minero atrapado en una mina tiene tres posibles caminos para tratar de escapar de la mina. El primero le conduce al exterior después de 3 horas. El segundo le conduce de nuevo al interior de la mina después de un recorrido de 5 horas. El tercero le conduce también al interior de la mina, pero después de 7 horas de recorrido. Si el minero echa a suertes (igual probabilidad) el camino

por el que tratar de escapar, ¿cuál será el tiempo de medio que tardará en conseguirlo?

*** **Ej. 77** — Sea $\{X_j\}_{j\geq 1}$ una sucesión de variables aleatorias uniformes en el intervalo $[0,1]$. Calcular $E(N)$ siendo

$$N = \min\left\{n; \sum_{j=1}^{n} X_j > 1\right\}.$$

* **Ej. 78** — El vector aleatorio (X,Y) tiene por densidad conjunta

$$f_{XY}(x,y) = \begin{cases} \dfrac{1}{2}ye^{-yx}, & 1 < y < 3,\ x > 0; \\ 0, & \text{en el resto.} \end{cases}$$

Hallar $E(X)$ y $var(X)$ haciendo uso de $E(X|Y)$ y de $var(X|Y)$.

* **Ej. 79** — El vector aleatorio (X,Y) tiene por densidad conjunta

$$f_{XY}(x,y) = \begin{cases} \dfrac{2e^{-2x}}{x}, & 0 \leq x < +\infty,\ 0 \leq y \leq x; \\ 0, & \text{en el resto.} \end{cases}$$

Calcular $cov(X,Y)$.
Sugerencia.- La obtención de $E(Y)$ puede resultar más sencilla a través de $E(Y) = E[E(Y|X)]$.

* **Ej. 80** — Tenemos un lote de N pastillas y antes de comercializarlo lo sometemos al siguiente control de calidad. Fijamos un umbral c ($c < n$) y a continuación tomamos una muestra de n pastillas del total de N que componen el lote. Sea X la variable aleatoria que nos da el número de pastillas en mal estado en la muestra, si $X \leq c$ entonces sustituimos las X pastillas defectuosas por pastillas en buen estado y comercializamos, sin más inspección, el lote. En el caso en que $X > c$ entonces se muestrean todas y cada una de las pastillas que componen el lote y se sustituyen por pastillas en buen estado, comercializándolo a continuación. En este caso el lote no tiene ninguna pastilla en mal estado. Si la probabilidad de que el proceso de fabricación produzca una pastilla en mal estado es p, ¿qué número medio de pastillas en mal estado estamos lanzando al mercado?

* **Ej. 81** — Sobre un círculo cuyo radio R es aleatorio con función de densidad

$$f_R(r) = \begin{cases} \dfrac{r^2}{9}, & r \in [0,3]; \\ 0 & \text{en el resto,} \end{cases}$$

elegimos un punto al azar. Si X designa la distancia del punto al origen, obtener

1. La función de distribución y la función de densidad de $X|R = r$.
2. La media y la varianza de X.

* **Ej. 82** — Lanzamos un dado repetidamente y designamos por X e Y, respectivamente, el número de lanzamientos necesarios para obtener un 6 y un 5. Calcular $E(X)$, $E(X|Y=1)$ y $E(X|Y=5)$.

* **Ej. 83** — La distribución conjunta de (X,Y) tiene por densidad,

$$\begin{cases} \frac{e^{-x/y}e^{-y}}{y}, & 0 < x < \infty,\ 0 < y < \infty; \\ 0, & \text{en el resto.} \end{cases}$$

Calcular $E(X^2|Y)$.

Y los demás[7]

** **Ej. 84** — Consideremos el vector aleatorio (X,Y). La variable Y es positiva con $E(Y)=1$ y $\sigma_Y^2=2$. Por otra parte $X|Y \sim U(1-Y, 1+Y)$. Obtener $E(X)$ y σ_X^2.

** **Ej. 85** — Supongamos $X_1, X_2, \ldots, X_n$ variables aleatorias independientes con distribución común $U(0,1)$. Determinar

1. $E[\max\{X_1,\ldots,X_n\}]$.
2. $E[\min\{X_1,\ldots,X_n\}]$.

*** **Ej. 86** — **Un problema de pastillas**Una caja contiene pastillas de dos tipos: grandes y pequeñas. Cada pastilla grande equivale a dos pequeñas. Cada día el paciente elige debe tomar una de las pastillas, que elige al azar. Si es de las grandes la parte en dos, toma una mitad y devuelve la otra a la caja. Si la pastilla es pequeña la toma sin más. A todos los efectos las mitades devueltas a la caja son consideradas como pastillas pequeñas.
¿Cuál es el número esperado de pastillas pequeñas que quedarán en la caja cuando las grandes se hayan terminado?

** **Ej. 87** — Un juego consiste en lanzar un dado repetidas veces. El juego se detiene bien porque aparece un 1, bien porque decidimos detenerlo en cualquier lanzamiento antes de que haya aparecido el 1. El premio que recibimos es el número de puntos que muestra la cara del dado en el ultimo lanzamiento. ¿Cuál es la estrategia más conveniente para obtener el mayor premio?

** **Ej. 88** — Sea X una variable aleatoria discreta con soporte $D_X = \{x_1, x_2, \ldots, x_m\}$. Calcular

$$\lim_{n\to\infty} \sqrt[n]{E(X^n)}.$$

*** **Ej. 89** — Se repite indefinida e independientemente un experimento cuyo resultado es E (éxito), con probabilidad p, o F (fracaso), con probabilidad $q = 1-p$. Denotamos por ω_n el resultado de la n-ésima repetición. Sea T la variable que designa el número mínimo de repeticiones necesarias para alcanzar r éxitos consecutivos. Demostrar que $P(T=\infty)=0$ y calcular $E(T)$.

[7] Que uno no sabe donde clasificarlos.

** **Ej. 90** — Tenemos 10 pares de zapatos y elegimos al azar 4 zapatos. ¿Cuál es la probabilidad de que no hayamos elegido ningún par? Si X es una variable aleatoria que representa el número de pares elegidos, obtener el número medio de pares entre los 4 zapatos elegidos.

** **Ej. 91** — Un autobús tiene en su recorrido 15 paradas. Supongamos que en la primera parada suben 20 personas. Cada una de ellas elige al azar e independientemente de las otras en cuál de las 14 paradas restantes quiere bajar.

1. Si X_i es una variable aleatoria que vale 1 si alguna de las personas baja en la parada i y 0 en caso contrario, calcular su distribución de probabilidad.
2. Calcular el número medio de paradas que debe realizar el autobús para que bajen todos los pasajeros.

** **Ej. 92** — (*[10, Problema 7.11]*)Realizamos n lanzamientos independientes de una moneda que tiene probabilidad de cara p. Diremos que se produce un cambio en la secuencia de resultados cuando el resultado de un lanzamiento difiere del anterior. Por ejemplo, denotando H cara y T cruz y suponiendo $n = 5$ entonces la secuencia $HHTHT$ tendría tres cambios. Se pide determinar el número esperado de cambios. Sugerencia: Considerar variables Bernoulli.

* **Ej. 93** — Sean U_1 y U_2 dos variables aletorias independientes ambas uniformes en [0,1]. Definimos $X = \sqrt{U_1}$ e $Y = 2XU_2$.

1. Obtener la densidad conjunta de (X, Y)
2. Obtener las densidades marginales de X e Y y sus esperanzas.
3. Si W es una variable Bernoulli tal que $P(W = 1) = P(0 \leq Y \leq \sqrt{X})$, calcular $E(W)$.

*** **Ej. 94** — Un juego consiste en partir al azar un palo en dos trozos y apostar acerca de la razón entre las longitudes del trozo mayor y el menor. Si acertamos ganamos k euros si dicha razón está entre k y *k+1*, con $1 \leq k \leq m - 1$, o m euros si la razón supera m, siendo m un entero positivo dado. ¿Cuanto deberíamos apostar para que el juego sea justo? ¿Cómo se comporta la apuesta a medida que m aumenta?

*** **Ej. 95** — **Spaghetti**Un plato tiene n spaghetti. Si n es suficientemente grande y los spaghetti suficientemente largos no es posible reconocer de una mirada los dos extremos de un mismo spaghetti. Bajo este supuesto, por otra parte razonable, elegimos al azar dos extremos de spaghetti y los unimos, repitiendo el proceso con los restantes extremos libres del plato. Al terminar habremos formado un determinado número de bucles de distinta longitud. Obtener el número esperado de dichos bucles de cualquier longitud y, en particular, el de los bucles de longitud 1, resultado de unir los dos extremos de un mismo spaghetti.

** **Ej. 96** — **Número esperado de extracciones**Extraemos con reemplazamiento bolas numeradas de 1 a n ($n \geq 2$) hasta que su suma exceda por primera vez k, $0 \leq k \leq n$. Encontrar el número esperado de extracciones, E_k.

** **Ej. 97** — Se eligen al azar n numeros en el intervalo *(a,b)*. Sean $X_{(1)}$ y $X_{(n)}$, respectivamente, el menor y el mayor de ellos. Obtener las líneas de regresión de $X_{(1)}$ sobre $X_{(n)}$ y de $X_{(n)}$ sobre $X_{(1)}$.

** **Ej. 98** — Sean X e Y dos variables aleatorias independientes uniformes en $(0,1)$, demuestra que:

$$E[|X-Y|^{\alpha}] = \frac{2}{(\alpha+1)(\alpha+2)}, \text{ para } \alpha > 0.$$

** **Ej. 99** — Sean X e Y variables aleatorias independientes y uniformemente distribuidas en el intervalo unitario $[0,1]$.

1. Obtener la densidad conjunta de $U = X+Y$ y $V = X/Y$, y las correspondientes marginales.
2. Calcular la $E(U)$ y la $E(V)$.

** **Ej. 100** — Un total de n bolas, numeradas de 1 a n son introducidas en n urnas numeradas a su vez de 1 a n. Cada bola es introducida de un modo equiprobable en las n posibles urnas. Determinar:

1. El número esperado de urnas vacías.
2. La probabilidad de que ninguna de las urnas esté vacía.

4.8 Soluciones a los ejercicios

Solución (Ej. 5) — 1.$EX = \bar{X} = \sum_{i=1}^{n} \frac{X_i}{n}$ y $var(X) = \sum_{i=1}^{n} \frac{(X_i-\bar{X})^2}{n}$.
2.$EX = 2$ y $var(X) = 1$.

Solución (Ej. 8) — 1.$E(X) = \frac{1}{2}$, $var(X) = \frac{1}{20}$.
2.$E(X) = 3$, $var(X) = \frac{3}{4}$.
3.$E(X) = \frac{a}{5}$, $var(X) = \frac{16a-5a^2}{80}$.

Solución (Ej. 9) — Consideremos la varianza de la variable $X^{1/2}$.

Solución (Ej. 10) — 1.$E(X) = 251.25$.
2.$P(X > E(X)) = e^{-1} = 0.3678$.
3.$m = m = \frac{Ln(2)}{\lambda}$.

Solución (Ej. 11) — 1.$E(X) = \frac{\alpha+\beta}{2}$, $var(X) = \frac{(\beta-\alpha)^2}{2}$.
2.$E(X) = \frac{\alpha}{\beta}$ y $var(X) = \frac{\alpha}{\beta^2}$.

Solución (Ej. 12) — 1.$E(e^{5X}) = \frac{1}{5}(e^5 - 1)$.
2.$E(\frac{1}{X}) = Ln(1) - Ln(0) = -\infty$.

Solución (Ej. 13) — El primer término es $a_1 = 0.1$ y la razón $r = 0.9$. Además, $P(X \leq 10) = 0.651$.

Solución (Ej. 14) — $E(X) = 14 \times \frac{8}{15}$.

Solución (Ej. 15) — Si $q = 1 - p$ entonces

- si $nq^n < 1$, la *primera* estrategia es la *mejor*,
- si $nq^n > 1$, la *segunda* estrategia es la *mejor*.

Solución (Ej. 16) — Dos monedas.

Solución (Ej. 18) — Si designamos mediante $x_{\max}$ el máximo de los posibles valores que toma X entonces

$$\lim_{n\to\infty} \frac{E(X^{n+1})}{E(X^n)} = x_{\max}.$$

Solución (Ej. 20) — 364.

Solución (Ej. 21) — En el primer método el número esperado es 5.5 mientras que en el segundo método el número esperado es 3.4.

Solución (Ej. 22) — Debe comprar 3 periódicos.

Solución (Ej. 23) — Debe comprar 5 ejemplares.

Solución (Ej. 24) — $E(X) = n(2p-1)$, $var(X) = 4np(1-p)$.

Solución (Ej. 25) —

1. Cualquier c es válido.
2. $E(Y) = 2$ mientras que $E(Y^2)$ no existe.

Solución (Ej. 26) —

1. $a = s/r^{-s}$.
2. $s > 1$.

Solución (Ej. 27) —

1. $\Theta \sim U(0, 2\pi)$.
2. $P(\Theta \in [a,b]) = \frac{b-a}{2\pi}$.
3. $F_X(x) = 2\arccos(x)$, si $-1 \leq x \leq 1$, cero por debajo de -1 y uno por encima.
4. $$f_X(x) = \frac{1}{\pi\sqrt{1-x^2}}, -1 \leq x \leq 1.$$
5. $EX = 0$.
6. $E|X| = \frac{2}{\pi}$.

Solución (Ej. 29) —

$$E(X!) = (1-\lambda)^{-1}e^{-\lambda}.$$

Solución (Ej. 30) —

1. $E(X) = 4$.
2. $E(X) = 0$.
3. $E(X) = \infty$.

Solución (Ej. 31) — $a = \frac{3}{5}$ y $b = \frac{6}{5}$.

Solución (Ej. 38) — Relacionar $P(X \geq a)$ con $P(g(X) \geq g(a))$ y aplicar la desigualdad de Markov.

Solución (Ej. 41) — Es aplicación directa de la desigualdad de Tchebychev.

Solución (Ej. 46) — $\rho_{XY} = -\frac{\sqrt{3}}{2\sqrt{2}}$.

Solución (Ej. 48) — 1.La densidad condicionada vale

$$f_{X|Y}(x|y) = \frac{1}{a - \frac{y}{2}}, \quad y/2 < x < a,$$

con lo que $X|Y \sim U(y/2, a)$.

2.$E(X) = \frac{2a}{3}$, $EY = \frac{2a}{3}$, $\sigma_X^2 = \frac{a^2}{18}$, $\sigma_Y^2 = \frac{2a^2}{9}$.

3.$\rho_{XY} = \frac{3}{\sqrt{18}}$.

Solución (Ej. 61) —

$$E(X) = var(X) = 1.$$

Solución (Ej. 62) —

$$E(X) = \frac{(2N-m)(2N-m-1)}{2(2N-1)}.$$

Solución (Ej. 64) — $cov(X+Y, X-Y) = 0$.

Solución (Ej. 66) —

$$\rho_{XY} = -1.$$

Solución (Ej. 68) —

$$E(X) = 87,6.$$

Solución (Ej. 70) — 3.5

Solución (Ej. 72) — Si S es la variable que nos da la suma de las puntuaciones entonces $ES = 35$ y $var(S) = \frac{455}{3}$.

Solución (Ej. 84) — $E(X) = var(X) = 1$.

Solución (Ej. 85) — $E[\max(X_1, \ldots, X_n)] = \frac{n}{n+1}$. $E[\min(X_1, \ldots, X_n)] = \frac{1}{n+1}$.

Solución (Ej. 92) — $2p(1-p)$

Solución (Ej. 99) — 1.

$$f_{UV}(u,v) = f_{XY}(x(u,v), y(u,v))|J_{g^{-1}}(u,v)| = \frac{u}{(1+v)^2}, \quad (u,v) \in D_{UV}.$$

con

$$D_{UV} = \{[0,1] \times [0,+\infty[\} \cup \{[1,2] \times [U-1, 1/(U-1)]\}.$$

2.$EU = 1$ y no existe la esperanza de V.

Solución (Ej. 100) — 1. $\frac{(n-1)^n}{n^{n-1}}$.

2. $\frac{n!}{n^n}$.

4.9 Material complementario

En esta sección se incluye material de interés pero que no necesario conocer para el desarrollo de la asignatura.

Muestra aleatoria: media y varianza muestrales

En esta sección introduciremos el concepto de muestra aleatoria y algunas variables aleatorias con ella asociadas, particularmente la media y la varianza muestrales. Cuando la muestra proceda de una distribución normal, las distribuciones de probabilidad relacionadas con dichas variables constituyen el eje sobre el que se basa la Inferencia Estadística clásica.

Definición 4.16 *Una muestra aleatoria de tamaño* n *es un vector aleatorio* $X_1, X_2, \ldots, X_n$, *cuyas componentes son* n *variables aleatorias independientes e idénticamente distribuidas* (i.i.d.).

A partir de la muestra aleatoria podemos definir la media y la varianza muestrales, $\overline{X}_n$ y S_n^2, mediante las expresiones

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \text{y} \qquad S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2,$$

respectivamente.

Si las componentes de la muestra aleatoria poseen momentos de segundo orden, y su media y varianza comunes son μ y σ, respectivamente, se verifica que

$$E\left(\overline{X}_n\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \mu \tag{4.42}$$

y

$$var\left(\overline{X}_n\right) = \frac{1}{n^2}\sum_{i=1}^{n} var(X_i) = \frac{\sigma^2}{n}.$$

Por lo que respecta a la varianza muestral, observemos en primer lugar que

$$\begin{aligned}
S_n^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \mu + \mu - \overline{X}_n)^2 \\
&= \frac{1}{n-1}\left\{\sum_{i=1}^{n}(X_i-\mu)^2 + \sum_{i=1}^{n}(\overline{X}_n-\mu)^2 - 2(\overline{X}_n-\mu)\sum_{i=1}^{n}(X_i-\mu)\right\} \\
&= \frac{1}{n-1}\left\{\sum_{i=1}^{n}(X_i-\mu)^2 + n(\overline{X}_n-\mu)^2 - 2(\overline{X}_n-\mu)n(\overline{X}_n-\mu)\right\} \\
&= \frac{1}{n-1}\left\{\sum_{i=1}^{n}(X_i-\mu)^2 - n(\overline{X}_n-\mu)^2\right\}.
\end{aligned}$$

Tomando ahora esperanzas,

$$\begin{aligned}
E(S_n^2) &= \frac{1}{n-1}\left\{\sum_{i=1}^{n} E\left[(X_i-\mu)^2\right] - nE\left[(\overline{X}_n-\mu)^2\right]\right\} \\
&= \frac{1}{n-1}\left(n\sigma^2 - n\times var(\overline{X}_n)\right) \\
&= \sigma^2. \qquad (4.43)
\end{aligned}$$

La $var(S_n^2)$ existe si las componentes de la muestra aleatoria poseen momentos de cuarto orden, si es así, la expresión de $var(S_n^2)$ viene dada por

$$var(S_n^2) = \frac{1}{n}\left(\mu_4 - \frac{n-3}{n-1}\sigma^4\right),$$

con

$$\mu_4 = E\left[(X_i-\mu)^4\right] \qquad \text{y} \qquad \sigma^4 = [\sigma^2]^2,$$

cuya deducción dejamos al cuidado del lector.

Observación 4.1 *La Inferencia Estadística proporciona una serie de técnicas que permiten conocer el comportamiento probabilístico de un fenómeno aleatorio a partir de la observación parcial del mismo. Por ejemplo, la estatura de una población de individuos tiene un comportamiento aleatorio que podemos deducir con un cierto margen de error (probabilístico) de la observación de una muestra de alturas de* n *individuos de esa población. Si postulamos la hipótesis de que dicha altura, X, se distribuye $N(\mu,\sigma^2)$, podemos estimar los valores de μ y σ^2 a partir de sus homólogos muestrales, $\overline{X}_n$ y S_n^2. Esta elección se basa, entre otras, en una propiedad conocida como* insesgadez, *que es la que representan las igualdades (4.42) y (4.43). A saber, que las esperanzas de los estimadores coinciden con el parámetro estimado.*

Algunas aplicaciones de la esperanza condicionada

Ejemplo 4.25 *Un trabajador está encargado del correcto funcionamiento de n máquinas situadas en linea recta y distantes una de otra l metros. El trabajador debe repararlas cuando se averían, cosa que sucede con igual probabilidad para todas ellas e independientemente de una a otra. El operario puede seguir dos estrategias:*

1. *acudir a reparar la máquina estropeada y permanecer en ella hasta que otra máquina se avería, desplazándose entonces hacia ella, o*

2. *situarse en el punto medio de la linea de máquinas y desde allí acudir a la averiada, regresando nuevamente a dicho punto cuando la avería está resuelta.*

Si X es la distancia que recorre el trabajador entre dos averías consecutivas, ¿cuál de ambas estrategias le conviene más para andar menos?

Se trata, en cualquiera de las dos estrategias, de obtener la $E(X)$ y elegir aquella que la proporciona menor. Designaremos por $E_i(X)$ la esperanza obtenida bajo la estrategia $i=1,2$.

> ***Estrategia 1.-*** *Sea A_k el suceso,* el operario se encuentra en la máquina *k. Para obtener $E_1(X)$ recurriremos a la propiedad anterior, pero utilizando como distribución condicionada la que se deriva de condicionar respecto del suceso A_k. Tendremos $E_1(X) = E(E(X|A_k))$.*
>
> *Para obtener $E(X|A_k)$ tengamos en cuenta que si* i *es la próxima máquina averiada, $P(A_i) = 1/n$, $\forall i$ y el camino a recorrer será*
>
> $$X|A_k = \begin{cases} (k-i)l, & si\ i \le k, \\ \\ (i-k)l, & si\ i > k. \end{cases}$$

Así pues,

$$E(X|A_k) = \frac{1}{n}\left(\sum_{i=1}^{k}(k-i)l + \sum_{i=k+1}^{n}(i-k)l\right) = \frac{l}{2n}[2k^2-2(n+1)k+n(n+1)].$$

Utilizando

$$\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6},$$

obtenemos para $E_1(X)$

$$E_1(X) = E(E(X|A_k)) = \frac{1}{n}\sum_k E(X|A_k) = \frac{l(n^2-1)}{3n}$$

Estrategia 2.- *Para facilitar los cálculos supongamos que* n *es impar, lo que supone que hay una máquina situada en el punto medio de la linea, la* $\frac{n+1}{2}$*-ésima. Si la próxima máquina averiada es la* i *la distancia a recorrer será,*

$$X = \begin{cases} 2(\frac{n+1}{2}-i)l, & si\ i \le \frac{n+1}{2}, \\ \\ 2(i-\frac{n+1}{2})l, & si\ i > \frac{n+1}{2}, \end{cases}$$

donde el 2 *se justifica porque el operario en esta segunda estrategia regresa siempre al punto medio de la linea de máquinas. La esperanza viene dada por,*

$$E_2(X) = \frac{2}{n}\sum_{i=1}^{n}\left|\frac{n+1}{2}-i\right| l = \frac{l(n-1)}{2}.$$

Como $E_1(X)/E_2(X) = 2(n+1)/3n \le 1$ *si* $n \ge 2$, *podemos deducir que la primera estrategia es mejor, salvo que hubiera una sola máquina.*

El ejemplo anterior ilustra la utilidad de (4.38) en la obtención de la esperanza absoluta. El ejemplo que sigue nos muestra como, en ocasiones, el rodeo que (4.38) pueda suponer es sólo aparente porque, en ocasiones, la obtención directa de $E(X)$ es mucho más laboriosa.

Ejemplo 4.26 *De una urna con* n *bolas blancas y* m *bolas negras llevamos a cabo extracciones sucesivas sin reemplazamiento. Si queremos conocer el número esperado de bolas negras extraídas antes de la primera blanca deberemos conocer la distribución de la correspondiente variable,* X.

La función de probabilidad de X *vale*

$$f_X(k) = P(X=k) = \frac{\binom{m}{k}}{\binom{m+n}{k}} \times \frac{n}{m+n-k}, \quad k = 0,1,\ldots,m, \tag{4.44}$$

donde el primer factor establece la probabilidad de que las k *primeras bolas sean negras, y el segundo la de que la* $k+1$*-ésima bola sea blanca. Un sencillo cálculo con los números combinatorios permite escribir (4.44) de esta otra forma*

$$f_X(k) = \frac{1}{\binom{m+n}{n}} \times \binom{m+n-1-k}{n-1},$$

más útil a la hora de comprobar que (4.44) es una función de probabilidad. En efecto,

$$\begin{aligned}\sum_{k=0}^{m} f_X(k) &= \sum_{k=0}^{m} \frac{1}{\binom{m+n}{n}} \times \binom{m+n-1-k}{n-1} \\ &= \frac{1}{\binom{m+n}{n}} \sum_{k=0}^{m} \binom{m+n-1-k}{n-1} \quad \textit{cambio } [m-k=j] \\ &= \frac{1}{\binom{m+n}{n}} \sum_{j=0}^{m} \binom{n-1+j}{n-1} = \frac{\binom{m+n}{n}}{\binom{m+n}{n}} = 1. \qquad (4.45)\end{aligned}$$

El valor esperado valdrá,

$$\begin{aligned}E(X) = \sum_{k=0}^{m} k f_X(k) &= \sum_{k=1}^{m} k \times \frac{\binom{m}{k}}{\binom{m+n}{k}} \times \frac{n}{m+n-k} \\ &= \frac{m}{m+n} \sum_{k=1}^{m} k \times \frac{\binom{m-1}{k-1}}{\binom{m+n-1}{k-1}} \times \frac{n}{m-1+n-(k-1)} = \\ &\frac{m}{m+n} \sum_{j=0}^{m-1} (j+1) f_{X_{m-1}}(j), \qquad (4.46)\end{aligned}$$

donde X_{m-1} es la variable asociada al mismo experimento cuando la urna contiene $m-1$ bolas negras y n bolas blancas. Obtenemos en definitiva la fórmula de recurrencia

$$E_m(X) = \frac{m}{m+n}(1 + E_{m-1}(X)), \qquad (4.47)$$

en la que el significado de los subíndices es obvio. Observemos que si $m=0$, $E_0(X)=0$ y a partir de aquí

$$\begin{aligned}E_1(X) &= \frac{1}{n+1}(1+0) = \frac{1}{n+1} \\ E_2(X) &= \frac{2}{n+2}\left(1 + \frac{1}{n+1}\right) = \frac{2}{n+1} \\ E_3(X) &= \frac{3}{n+3}\left(1 + \frac{2}{n+1}\right) = \frac{3}{n+1} \\ \cdots & \quad \cdots \\ E_m(X) &= \frac{m}{n+m}\left(1 + \frac{m-1}{n+1}\right) = \frac{m}{n+1}.\end{aligned}$$

La expresión (4.47) puede obtenerse más fácilmente si utilizamos la variable auxiliar Y definida mediante

$$Y = \begin{cases} 1, & \textit{si la primera bola es blanca;} \\ 0, & \textit{si la primera bola es negra.} \end{cases}$$

Podremos escribir

$$E_m(X) = E[E_m(X|Y)] = E_m(X|Y=1)P(Y=1) + E_m(X|Y=0)P(Y=0),$$

pero $E_m(X|Y=1) = 0$ *y* $E_m(X|Y=0) = 1 + E_{m-1}(X)$. *Sustituyendo,*

$$E_m(X) = \frac{m}{m+n}(1 + E_{m-1}(X)),$$

que coincide con (4.47).

Ejemplo 4.27 *El tiempo de descarga de un barco que transporta cereal se distribuye uniformemente entre* a *y* b *horas,* $T \sim U(a,b)$, *siempre que dicha descarga se lleve a cabo sin interrupciones. La grúa que efectúa la descarga es poco fiable y propensa a las averías. Éstas ocurren según una distribución de Poisson con media* λ *averías por hora. Cuando la grúa está estropeada se requieren* d *horas para arreglarla. Con todas estas circunstancias, queremos conocer el tiempo real medio de descarga del barco.*

Si designamos por T_R *el tiempo real de descarga del barco, lo que se nos pide es* $E(T_R)$. *Observemos que sobre dicho tiempo influyen, tanto la variable* T, *que nos proporciona el tiempo de descarga sin tener en cuenta las posibles interrupciones, como las horas que haya que añadir debido a la averías de la grúa durante el período* T, *averías cuyo número en dicho lapso de tiempo será* $N_T \sim Po(\lambda T)$.

Sabemos que $E(T_R) = E[E(T_R|T)]$. *Para calcular la esperanza condicionada hemos de tener en cuenta las posibles averías, lo que equivale a decir que hemos de condicionar a los posibles valores de* N_T. *Así,*

$$\begin{aligned} E(T_R|T) &= E\left\{E\left[E(T_R|T)|N_T\right]\right\} \\ &= \sum_{n\geq 0} E(T_R|T, N_T = n)P(N_T = n) \\ &= \sum_{n\geq 0}(T + nd)P(N_T = n) \\ &= T\sum_{n\geq 0} P(N_T = n) + d\sum_{n\geq 0} nP(N_T = n) \\ &= T(1 + d\lambda). \end{aligned}$$

Por último

$$E(T_R) = E[E(T_R|T)] = (1 + d\lambda)E(T) = \frac{(a+b)(1+d\lambda)}{2}.$$

El principio de los mínimos cuadrados

Supongamos que entre las variables aleatorias X e Y existe una relación funcional que deseamos conocer. Ante la dificultad de poder hacerlo vamos a tratar de encontrar la función $h(X)$ que mejor aproxime dicha relación. El interés en semejante función es poder predecir el valor que tomará Y a partir del valor que ha tomado X.

Si $E(Y^2)$ y $E(h(X)^2)$ existen, el *principio de los mínimos cuadrados* es uno de los posibles criterios para encontrar $h(X)$. Consiste en elegir aquella $h(X)$ que minimice la cantidad $E\left\{(Y - h(X))^2\right\}$, que no es más que la esperanza del error que cometeremos al sustituir el verdadero valor de Y por su estimación, $h(X)$. Si (X,Y) es un vector aleatorio continuo, tenemos

$$\begin{aligned}
E\left\{(Y-h(X))^2\right\} &= \int_{\mathbb{R}^2} (y-h(x))^2 f_{XY}(x,y)dxdy \\
&= \int_{\mathbb{R}^2} (y-h(x))^2 f_{Y|X}(x,y)f_X(x)dydx \\
&= \int_{\mathbb{R}} \left[\int_{\mathbb{R}} (y-h(x))^2 f_{Y|X}(x,y)dy\right] f_X(x)dx.
\end{aligned}$$

Que sea mínima esta expresión supone que lo es la integral interior, pero de acuerdo con una propiedad de la varianza (PV4 de la página 144) el valor que minimiza dicha integral es precisamente $h(X) = E(X|Y)$. Para el caso discreto obtendríamos un resultado análogo.

Definición 4.17 (Regresión de Y sobre X) *La relación $y = E[Y|x]$ se conoce como la regresión de Y sobre X. Análogamente $x = E[X|y]$ se conoce como la regresión de X sobre Y.*

Recta de regresión

En ocasiones, fundamentalmente por motivos de sencillez, se está interesado en aproximar la relación entre X e Y mediante una línea recta. En esta situación, y haciendo uso del principio de los mínimos cuadrados, elegiremos los parámetros de la recta de forma que

$$L(a,b) = E\left\{(Y-aX-b)^2\right\}$$

sea mínimo.

La obtención de a y b se reduce a un problema de máximos y mínimos y basta igualar a 0 las derivadas parciales $\partial L/\partial a$ y $\partial L/\partial b$. Si lo hacemos obtendremos,

$$a = \frac{cov(X,Y)}{var(X)}, \qquad b = E(Y) - aE(X).$$

La ecuación de la que se conoce como *recta de regresión* de Y sobre X tendrá por expresión,

$$Y - E(Y) = \frac{cov(X,Y)}{var(X)}(X - E(X)).$$

Por simetría, la *recta de regresión* de X sobre Y tendrá por expresión,

$$X - E(X) = \frac{cov(X,Y)}{var(Y)}(Y - E(Y)).$$

En las anteriores expresiones, las cantidades

$$\gamma_{X|Y} = \frac{cov(X,Y)}{var(Y)} \quad y \quad \gamma_{Y|X} = \frac{cov(X,Y)}{var(X)},$$

reciben el nombre de *coeficientes de regresión* de X sobre Y e Y sobre X, respectivamente. Tiene una interesante propiedad,

$$\gamma_{X|Y}\cdot\gamma_{Y|X} = \frac{(cov(X,Y))^2}{var(X)var(Y)} = \rho_{xy}^2. \tag{4.48}$$

Ejemplo 4.28 *Consideremos el vector aleatorio (X, Y) con densidad conjunta,*

$$f_{XY}(x,y) = \begin{cases} e^{-y}, & 0 < x < y < +\infty \\ 0, & \text{en el resto.} \end{cases}$$

Las correspondientes marginales vienen dadas por

$$f_X(x) = \int_x^{+\infty} e^{-y} dy = \begin{cases} e^{-x}, & 0 < x < +\infty \\ 0, & \text{en el resto,} \end{cases}$$

y

$$f_Y(y) = \int_0^{y} e^{-y} dx = \begin{cases} ye^{-y}, & 0 < y < +\infty \\ 0, & \text{en el resto.} \end{cases}$$

Las distribuciones condicionadas se obtienen fácilmente a partir de las anteriores,

$$f_{X|Y}(x|y) = \begin{cases} \frac{1}{y}, & 0 < x < y \\ 0, & \text{en el resto,} \end{cases} \qquad f_{Y|X}(y|x) = \begin{cases} e^{x-y}, & x < y < +\infty \\ 0, & \text{en el resto.} \end{cases}$$

Las esperanzas condicionadas valen

$$E(X|Y) = \int_0^{y} x\frac{1}{y} dx = \frac{y}{2}, \qquad 0 < y < +\infty,$$

$$E(Y|X) = \int_x^{+\infty} ye^{x-y} dy = x + 1, \qquad 0 < x < +\infty.$$

Se trata en ambos casos de rectas, por lo que coincidirán con las rectas de regresión correspondientes. El valor absoluto del coeficiente de correlación podrá obtenerse fácilmente a partir de (4.48),

$$\rho^2_{XY} = \gamma_{X|Y} \cdot \gamma_{Y|X} = \frac{1}{2} \cdot 1 = \frac{1}{2},$$

el signo será el mismo que el de la covarianza, que se comprueba fácilmente que vale 1. Así pues, $\rho_{XY} = \sqrt{1/2}$.

Ejemplo 4.29 (El fenómeno de la regresión a la media) *La recta de regresión nos ayuda a comprender porqué los hijos de padres muy altos tienden a ser más bajos que sus padres y los de padres muy bajos suelen superarlos en estatura. Pensemos en un padre que tiene una altura X cuando nace su hijo y preguntémonos la altura Y que alcanzará su hijo cuando tenga la edad que ahora tiene su padre.*

Si por μ_X, μ_Y, σ^2_X y σ^2_Y designamos, respectivamente, las medias y las varianzas de las alturas de padre e hijo, una hipótesis razonable y lógica es suponer que $\mu_X = \mu_Y = \mu$ y $\sigma^2_X = \sigma^2_Y = \sigma^2$. Por las propiedades de la recta de regresión, sabemos que la mejor estimación lineal de Y es $\hat{Y}$, que verifica

$$\hat{Y} - \mu = \frac{\sigma_{XY}}{\sigma^2}(X - \mu). \tag{4.49}$$

Como

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{\sigma_{XY}}{\sigma^2},$$

de aquí $\sigma_{XY} = \rho_{XY}\sigma^2$ y sustituyendo en (4.49)

$$\hat{Y} - \mu = \rho_{XY}(X - \mu).$$

Puesto que $|\rho_{XY}| \leq 1$ y es de suponer que será positivo, concluiremos que la estatura del hijo estará más próxima a la media que la del padre.

Capítulo 5

Sucesiones de variables aleatorias

5.1 Introducción

Los capítulos anteriores nos han permitido familiarizarnos con el concepto de variable y vector aleatorio, dotándonos de las herramientas que nos permiten conocer su comportamiento probabilístico. En el caso de un vector aleatorio somos capaces de estudiar el comportamiento conjunto de un número finito de variables aleatorias. Pero imaginemos por un momento los modelos probabilísticos asociados a los siguientes fenómenos:

1. lanzamientos sucesivos de una moneda,
2. tiempos transcurridos entre llamadas consecutivas a una misma centralita,
3. sucesión de estimadores de un parámetro cuando se incrementa el tamaño de la muestra...

En todos los casos las variables aleatorias involucradas lo son en cantidad numerable y habremos de ser capaces de estudiar su comportamiento conjunto y, tal y como siempre sucede en ocasiones similares, de conocer cuanto haga referencia al límite de la sucesión. Del comportamiento conjunto se ocupa una parte de la Teoría de la Probabilidad que dado su interés ha tomado entidad propia: la Teoría de los Procesos Estocásticos. En este capítulo nos ocuparemos de estudiar cuanto está relacionado con el límite de las sucesiones de variables aleatorias. Este estudio requiere en *primer lugar*, introducir los tipos de convergencia apropiados a la naturaleza de las sucesiones que nos ocupan, para en *segundo lugar* obtener las condiciones bajo las que tienen lugar las dos convergencias que nos interesan: la convergencia de la sucesión de variables a una constante (Leyes de los Grandes Números) y la convergencia a otra variable (Teorema Central del Límite). El estudio de esta segunda situación se ve facilitado con el uso de una herramienta conocida como *función característica* de la cual nos habremos ocupado previamente.

Dos sencillos ejemplos relacionados con la distribución binomial no servirán de introducción y nos ayudarán a situarnos en el problema.

Ejemplo 5.1 (Un resultado de J. Bernoulli) *Si repetimos n veces un experimento cuyo resultado es la ocurrencia o no del suceso A, tal que $P(A) = p$, y si la repeticiones son independientes unas de otras, la variable X_n =*número de ocurrencias de A, *tiene una distribución $B(n,p)$. La variable X_n/n representa la frecuencia relativa de A y sabemos que*

$$E\left(\frac{X_n}{n}\right) = \frac{1}{n}E(X_n) = \frac{np}{n} = p,$$

y

$$var\left(\frac{X_n}{n}\right) = \frac{1}{n^2}var(X_n) = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}.$$

Si aplicamos la desigualdad de Chebyshev,

$$P\left(\left|\frac{X_n}{n} - p\right| \geq \varepsilon\right) \leq \frac{var(X_n/n)}{\varepsilon^2} = \frac{p(1-p)}{n\varepsilon^2} \xrightarrow{n} 0.$$

Deducimos que la frecuencia relativa de ocurrencias de A converge*, en algún sentido, a $P(A)$.*

Ejemplo 5.2 (Binomial vs Poisson) *El segundo ejemplo ya fue expuesto en la página 58 y no lo repetiremos aquí. Hacía referencia a la aproximación de la distribución binomial mediante la distribución de Poisson. Vimos que cuando tenemos un gran número de pruebas Bernoulli con una probabilidad de éxito muy pequeña de manera que* $\lim_n np_n = \lambda$, $0 < \lambda < +\infty$, *la sucesión de funciones de cuantía de las variables aleatorias $X_n \sim B(n, p_n)$ converge a la función de cuantía de $X \sim Po(\lambda)$.*

Dos ejemplos con sucesiones de variables binomiales que han conducido a *límites* muy distintos. En el primero, el valor límite es la probabilidad de un suceso, y por tanto una constante; en el segundo la función de cuantía tiende a otra función de cuantía.

5.2 Tipos de convergencia

Comenzaremos por formalizar el tipo de convergencia que aparece en el primer ejemplo. Para ello, y también para el resto de definiciones, sobre un espacio de probabilidad $(\Omega, \mathbb{A}, P)$ consideremos la sucesión de variables aleatorias $\{X_n\}$ y la variable aleatoria X.

Definición 5.1 (Convergencia en probabilidad) *Decimos que $\{X_n\}$ converge a X en probabilidad, $X_n \xrightarrow{P} X$, si para cada $\delta > 0$,*

$$\lim_n P\{\omega : |X_n(\omega) - X(\omega)| > \delta\} = 0.$$

No es esta una convergencia puntual como las que estamos acostumbrados a utilizar en Análisis Matemático. La siguiente sí es de este tipo.

Definición 5.2 (Convergencia casi segura o con probabilidad 1) *Decimos que $\{X_n\}$ converge casi seguramente*[1] *a X (o con probabilidad 1), $X_n \xrightarrow{a.s.} X$, si*

$$P(\{\omega : \lim_n X_n(\omega) = X(\omega)\}) = 1.$$

[1] Utilizaremos la abreviatura *a. s.*, que corresponde a las iniciales de **a**lmost **s**urely, por ser la notación más extendida

El último tipo de convergencia involucra a las funciones de distribución asociadas a cada variable y requiere previamente una definición para la convergencia de aquellas.

Definición 5.3 (Convergencia débil) *Sean F_n, $n \geq 1$, y F funciones de distribución de probabilidad. Decimos que la sucesión F_n converge débilmente*[2] *a F, $F_n \xrightarrow{\omega} F$, si $\lim_n F_n(x) = F(x)$, $\forall x$ que sea punto de continuidad de F.*

Definición 5.4 (Convergencia en ley) *Decimos que $\{X_n\}$ converge en ley a X, $X_n \xrightarrow{L} X$, si $F_{X_n} \xrightarrow{\omega} F_X$. Teniendo en cuenta la definición de F_n y F, la convergencia en ley puede expresarse también*

$$X_n \xrightarrow{L} X \Longleftrightarrow \lim_n P(X_n \leq x) = P(X \leq x),$$

$\forall x$ que sea punto de continuidad de F.

Las relaciones entre los tres tipos de convergencia se establecen en el siguiente teorema.

Teorema 5.1 (Relaciones entre convergencias) *Sean X_n y X variables aleatorias definidas sobre un mismo espacio de probabilidad entonces:*

$$X_n \overset{a.s.}{\to} X \Rightarrow X_n \overset{P}{\to} X \Rightarrow X_n \overset{L}{\to} X$$

Demostración 5.1

1) $X_n \overset{a.s.}{\to} X \Rightarrow X_n \overset{P}{\to} X$

Para $\delta > 0$ sean

$$A = \{\omega : \lim_n X_n(\omega) \neq X(\omega)\}$$

y

$$A_n = \{\omega : |X_n(\omega) - X(\omega)| > \delta\},$$

entonces[52] $\limsup A_n \subset A$ *y* $P(\limsup A_n) = 0$, *y de aquí* $\lim P(A_n) = 0$.

[52] Recordemos las definiciones $\liminf_n A_n = \bigcup_{n\geq 1} \bigcap_{k\geq n} A_k$, y $\limsup_n A_n = \bigcap_{n\geq 1} \bigcup_{k\geq n} A_k$. Además: $P(\liminf_n A_n) \leq \liminf_n P(A_n) \leq \limsup_n P(A_n) \leq P(\limsup_n A_n)$.

2) $X_n \overset{P}{\to} X \Rightarrow X_n \overset{L}{\to} X$

Si x es tal que $P(X = x) = 0$, se verifica que

$$\begin{aligned} P(X \leq x - \varepsilon) &= \\ &P(X \leq x - \varepsilon, X_n \leq x) + P(X \leq x - \varepsilon, X_n > x) \leq \\ &P(X_n \leq x) + P(|X_n - X| > \varepsilon). \end{aligned} \tag{5.1}$$

y

$$\begin{aligned} P(X_n \leq x) &= \\ &P(X_n \leq x, X \leq x + \varepsilon) + P(X_n \leq x, X > x + \varepsilon) \leq \\ &P(X \leq x + \varepsilon) + P(|X_n - X| > \varepsilon). \end{aligned} \tag{5.2}$$

Expresando conjuntamente las dos desigualdades tenemos:

$$\begin{aligned} &P(X \leq x - \varepsilon) - P(|X_n - X| > \varepsilon) \leq \\ &\leq P(X_n \leq x) \leq P(X \leq x + \varepsilon) + P(|X_n - X| > \varepsilon), \end{aligned}$$

[2] Utilizaremos la abreviatura ω, que corresponde a la inicial de **w**eakly, por ser la notación más extendida

pero $\lim_n P(|X_n - X| > \varepsilon) = 0$ *para cualquier* ε *positivo por lo que:*

$$P(X \leq x-\varepsilon) \leq \liminf_n P(X_n \leq x) \leq \limsup_n P(X_n \leq x) \leq P(X \leq x+\varepsilon)$$

y, puesto que $P(X = x) = 0$ *es equivalente a que* $F(x) = P(X \leq x)$ *sea continua en* x*, se sigue el resultado.*

Las convergencias *casi segura* y *en probabilidad* tienen distinta naturaleza, mientras aquella es de tipo puntual, esta última es de tipo conjuntista. El ejemplo que sigue ilustra bien esta diferencia y pone de manifiesto que la contraria de la primera implicación no es cierta.

Ejemplo 5.3 (La convergencia en probabilidad $\not\Rightarrow$ la convergencia casi segura) *Como espacio de probabilidad consideraremos el intervalo unidad dotado de la* σ*-álgebra de Borel y de la medida de Lebesgue, es decir, un espacio de probabilidad uniforme en [0,1]. Definimos la sucesión* $X_n = \mathbf{1}_{I_n}$*,* $\forall n$*, con* $I_n = [\frac{p}{2^q}, \frac{p+1}{2^q}]$*, siendo* p *y* q *los únicos enteros positivos que verifican,* $p + 2^q = n$ *y* $0 \leq p < 2^q$*. Obviamente* $q = q(n)$ *y* $\lim_n q(n) = +\infty$*. Los primeros términos de la sucesión son,*

$$\begin{array}{lll}
n=1 & q=0, p=0 & X_1 = \mathbf{1}_{[0,1[} \\
n=2 & q=1, p=0 & X_2 = \mathbf{1}_{[0,\frac{1}{2}[} \\
n=3 & q=1, p=1 & X_3 = \mathbf{1}_{[\frac{1}{2},1[} \\
n=4 & q=2, p=0 & X_4 = \mathbf{1}_{[0,\frac{1}{4}[} \\
n=5 & q=2, p=1 & X_5 = \mathbf{1}_{[\frac{1}{4},\frac{1}{2}[} \\
n=6 & q=2, p=2 & X_6 = \mathbf{1}_{[\frac{1}{2},\frac{3}{4}[} \\
n=7 & q=2, p=3 & X_7 = \mathbf{1}_{[\frac{3}{4},1[} \\
\ldots\ldots & &
\end{array}$$

Observemos que si $X = 0$*,* $\forall \delta > 0$ *se tiene* $\lambda\{\omega : |X_n(\omega)| > \delta\} = \lambda\{\omega : |X_n(\omega)| = 1\} = \lambda(I_n) = 2^{-q}$*,* $2^q \leq n < 2^{q+1}$ *y* $X_n \xrightarrow{\lambda} 0$*; pero dada la construcción de las* X_n *en ningún* $\omega \in [0,1]$ *se verifica* $\lim X_n(\omega) = 0$*.*

La convergencia en ley (débil) no implica la convergencia en probabilidad, como pone de manifiesto el siguiente ejemplo, lo que justifica el nombre de *convergencia débil* puesto que es la última en la cadena de implicaciones.

Ejemplo 5.4 *Consideremos una variable Bernoulli con* $p = 1/2$*,* $X \sim B(1,1/2)$ *y definamos una sucesión de variables aleatorias,* $X_n = X$*,* $\forall n$*. La variable* $Y = 1 - X$ *tiene la misma distribución que* X*, es decir,* $Y \sim B(1,1/2)$*. Obviamente* $X_n \xrightarrow{L} Y$*, pero como* $|X_n - Y| = |2X - 1| = 1$*, no puede haber convergencia en probabilidad.*

Hay, no obstante, una situación en las que la convergencia débil y la convergencia en probabilidad son equivalentes, cuando el límite es a una constante. Veámoslo.

Teorema 5.2 *Si* $X_n \xrightarrow{L} c$ *entonces* $X_n \xrightarrow{P} c$*.*

Demostración 5.2 *Si* F_n *es la función de distribución de* X_n*, la convergencia en Ley implica que* $F_n(x) \to F_c(x)$ *tal que*

$$F_c(x) = \begin{cases} 0, & \text{si } x < c; \\ 1, & \text{si } x \geq c. \end{cases}$$

Sea ahora $\delta > 0$,

$$P(|X_n - c| > \delta) =$$
$$P(X_n - c < -\delta) + P(X_n - c > \delta) = P(X_n < c - \delta) + P(X_n > c + \delta)$$
$$\leq P(X_n \leq c - \delta) + 1 - P(X_n \leq c + \delta) = F_n(c - \delta) + 1 - F_n(c + \delta), \tag{5.3}$$

por ser $c - \delta$ *y* $c + \delta$ *puntos de continuidad de* F_c. *La convergencia débil de* F_n *a* F_c *implica que* $P(|X_n - c| > \delta) \to 0$ *y en consecuencia* $X_n \xrightarrow{P} c$.

5.3 Leyes de los Grandes Números

El nombre de *leyes de los grandes números* hace referencia al estudio de un tipo especial de límites derivados de la sucesión de variables aleatorias $\{X_n\}$. Concretamente los de la forma $\lim_n \frac{S_n - a_n}{b_n}$, con $S_n = \sum_{i=1}^n X_i$ y siendo $\{a_n\}$ y $\{b_n\}$ sucesiones de constantes tales que $\lim b_n = +\infty$. En esta sección fijaremos las condiciones para saber cuando existe convergencia a.s, y como nos ocuparemos también de la convergencia en probabilidad, las leyes se denominarán *fuerte* y *débil*, respectivamente.

Teorema 5.3 (Ley débil) *Sea* $\{X_k\}$ *una sucesión de variables aleatorias independientes tales que* $E(X_k^2) < +\infty$, $\forall k$, *y* $\lim_n \frac{1}{n^2} \sum_{k=1}^n var(X_k) = 0$, *entonces*

$$\frac{1}{n} \sum_{k=1}^n (X_k - E(X_k)) \xrightarrow{P} 0.$$

Demostración 5.3 *Para* $S_n = \frac{1}{n} \sum_{k=1}^n (X_k - E(X_k))$, $E(S_n) = 0$ *y* $var(S_n) = \frac{1}{n^2} \sum_{k=1}^n var(X_k)$. *Por la desigualdad de Chebyshev,* $\forall \varepsilon > 0$

$$P(|S_n| \geq \varepsilon) \leq \frac{var(S_n)}{\varepsilon^2} = \frac{1}{n^2 \varepsilon^2} \sum_{k=1}^n var(X_k),$$

que al pasar al límite nos asegura la convergencia en probabilidad de S_n *a* 0.

Corolario 5.1 *Si las* X_n *son i.i.d. con varianza finita y esperanza común* $E(X_1)$, *entonces* $\frac{1}{n} \sum_{k=1}^n X_k \xrightarrow{P} E(X_1)$.

Demostración 5.4 *Si* $var(X_k) = \sigma^2$, $\forall k$, *tendremos* $\frac{1}{n^2} \sum_{k=1}^n var(X_k) = \frac{\sigma^2}{n}$ *que tiende a cero con* n. *Es por tanto de aplicación la ley débil que conduce al resultado enunciado.*

Este resultado fué demostrado por primera vez por J. Bernoulli para variables con distribución binomial (véase el ejemplo 5.1), versión que se conoce como la *ley de los grandes números de Bernoulli*

El siguiente paso será fijar las condiciones para que el resultado sea válido bajo convergencia a.s.

Teorema 5.4 (Ley fuerte) *Si* $\{X_k\}$ *es una sucesión de variables aleatorias i.i.d. con media finita, entonces*

$$\sum_{k=1}^n \frac{X_k}{n} \xrightarrow{a.s.} E(X_1).$$

Corolario 5.2 *Si $\{X_k\}$ es una sucesión de variables aleatorias i.i.d. con $E(X_1^-) < +\infty$ y $E(X_1^+) = +\infty$, entonces $\frac{S_n}{n} \xrightarrow{a.s.} \infty$.*

La demostración de la *ley fuerte* es de una complejidad, aun en su versión más sencilla debida a Etemadi, fuera del alcance y pretensiones de este texto. La primera demostración, más compleja que la de Etemadi, se debe a Kolmogorov y es el resultado final de una cadena de propiedades previas de gran interés y utilidad en sí mismas. Aconsejamos vivamente al estudiante que encuentre ocasión de hojear ambos desarrollos en cualquiera de los textos habituales (Burrill, Billingsley,...), pero que en ningún caso olvide el desarrollo de Kolmogorov.

5.3.1 El teorema de Glivenko-Cantelli

Para las variables aleatorias $X_1, X_2, \ldots, X_n$ se define la función de distribución empírica mediante

$$F_n(x,\omega) = \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{]-\infty,x]}(X_k(\omega)).$$

Cuando todas las variables tienen la misma distribución, $F_n(x,\omega)$ es el estimador natural de la función de distribución común, $F(x)$. El acierto en la elección de este estimador se pone de manifiesto en el siguiente resultado.

Teorema 5.5 *Sea $\{X_k\}$ una sucesión de variables aleatorias i.i.d. con función de distribución común $F(x)$, entonces $F_n(x,\omega) \xrightarrow{a.s.} F(x)$.*

Demostración 5.5 *Para cada x, $F_n(x,\omega)$ es una variable aleatoria resultante de sumar las n variables aleatorias independientes, $\mathbf{1}_{]-\infty,x]}(X_k(\omega))$, $k = 1,\ldots,n$, cada una de ellas con la misma esperanza, $E(\mathbf{1}_{]-\infty,x]}(X_k(\omega))) = \mathbb{P}(X_k \leq x) = F(x)$. Aplicando la ley fuerte de los grandes números,*

$$F_n(x,\omega) \xrightarrow{a.s.} F(x),$$

que es el resultado buscado.

Este resultado es previo al teorema que da nombre al apartado y que nos permite contrastar la hipótesis de suponer que F es la distribución común a toda la sucesión.

Teorema 5.6 (Glivenko-Cantelli) *Sea $\{X_k\}$ una sucesión de variables aleatorias i.i.d. con función de distribución común $F(x)$. Hagamos $D_n(\omega) = \sup_x |F_n(x,\omega) - F(x)|$, entonces $D_n \xrightarrow{a.s.} 0$.*

La demostración, muy técnica, la omitimos y dejamos al interés del lector consultarla en el texto de Billingsley.

5.3.2 Cálculo aproximado de integrales por el método de Monte-Carlo

Sea $f(x) \in \mathbb{C}([0,1])$ con valores en $[0,1]$. Una aproximación al valor de $\int_0^1 f(x)dx$ puede obtenerse a partir de una sucesión de pares de variables aleatorias distribuidas uniformemente en $[0,1]$, $(X_1,Y_1), (X_2,Y_2), \ldots$.

Para ello hagamos,

$$Z_i = \begin{cases} 1, & si\ f(X_i) \geq Y_i \\ 0, & si\ f(X_i) < Y_i. \end{cases}$$

Así definidas las Z_i son variables Bernoulli con parámetro $p = E(Z_i) = P(f(X_i) \geq Y_i) = \int_0^1 f(x)dx$, y aplicándoles la ley fuerte de los grandes números tendremos que

$$\frac{1}{n}\sum_{i=1}^{n} Z_i \xrightarrow{a.s.} \int_0^1 f(x)dx,$$

lo que en términos prácticos supone simular los pares $(X_i, Y_i),\ i = 1, \ldots, n$, con X_i e $Y_i \sim U(0,1)$, y calcular la proporción de ellos que caen por debajo de la gráfica $y = f(x)$.

5.3.3 Aplicación de la LGN a la aproximación de funciones

Sea g una función acotada definida sobre $[0,1]$, la función B_n definida sobre $[0,1]$ mediante

$$B_n(x) = \sum_{k=0}^{n} g\left(\frac{k}{n}\right)\binom{n}{k}x^k(1-x)^{n-k},$$

es conocida como polinomio de Bernstein de grado n.

El teorema de aproximación de Weierstrass asegura que toda función continua sobre un intervalo cerrado puede ser aproximada uniformemente mediante polinomios. Probemos dicha afirmación para los polinomios de Bernstein.

Si la función g a aproximar es continua en $[0,1]$, será uniformemente continua, entonces

$$\forall \epsilon > 0, \exists \delta > 0 \text{ tal que } |g(x) - g(y)| < \epsilon, \text{ si } |x - y| < \delta.$$

Además g estará también acotada y por tanto $|g(x)| < M, \forall x \in [0,1]$.

Sea ahora un x cualquiera en $[0,1]$,

$$\begin{aligned} |g(x) - B_n(x)| &= \left| g(x)\sum_{k=0}^{n}\binom{n}{k}x^k(1-x)^{n-k} - \sum_{k=0}^{n} g\left(\frac{k}{n}\right)\binom{n}{k}x^k(1-x)^{n-k} \right| \\ &\leq \sum_{k=0}^{n}\left| g(x) - g\left(\frac{k}{n}\right)\right|\binom{n}{k}x^k(1-x)^{n-k} \\ &= \sum_{|k/n-x|<\delta}\left| g(x) - g\left(\frac{k}{n}\right)\right|\binom{n}{k}x^k(1-x)^{n-k} + \\ &\quad + \sum_{|k/n-x|\geq\delta}\left| g(x) - g\left(\frac{k}{n}\right)\right|\binom{n}{k}x^k(1-x)^{n-k} \\ &\leq \epsilon + 2M\sum_{|k/n-x|\geq\delta}\binom{n}{k}x^k(1-x)^{n-k}. \end{aligned}$$

Si $Z_n \sim B(n,x)$, el último sumatorio no es más que

$$P\left(\left|\frac{Z_n}{n} - x\right| > \delta\right) = \sum_{|k/n-x|\geq\delta}\binom{n}{k}x^k(1-x)^{n-k},$$

y tendremos

$$|g(x) - B_n(x)| \leq \epsilon + 2MP\left(\left|\frac{Z_n}{n} - x\right| > \delta\right),$$

pero por la ley de los grandes números

$$\frac{Z_n}{n} \xrightarrow{P} x \quad \text{y por tanto} \quad P\left(\left|\frac{Z_n}{n} - x\right| > \delta\right) \longrightarrow 0,$$

lo que demuestra la convergencia uniforme de B_n a g en $[0,1]$.

5.4 Función característica

La *función característica* es una herramienta de gran utilidad en Teoría de la Probabilidad, una de sus mayores virtudes reside en facilitar la obtención de la distribución de probabilidad de la suma de variables aleatorias y la del límite de sucesiones de variables aleatorias, situaciones ambas que aparecen con frecuencia en Inferencia Estadística.

El concepto de *función característica*, introducido por Lyapunov en una de la primeras versiones del Teorema Central del Límite, procede del Análisis Matemático donde se le conoce con el nombre de *transformada de Fourier*.

Definición 5.5 *Sea X una variable aleatoria y sea $t \in R$. La función característica de X, $\phi_X(t)$, se define como $E(e^{itX}) = E(\cos tX) + iE(\sin tX)$.*

Como $|e^{itX}| \leq 1,\ \forall t$, $\phi_X(t)$ existe siempre y está definida $\forall t \in R$. Para su obtención recordemos que,

Caso discreto.- Si X es una v. a. discreta con soporte D_X y función de cuantía $f_X(x)$,

$$\phi_X(t) = \sum_{x \in D_X} e^{itx} f_X(x). \tag{5.4}$$

Caso continuo.- Si X es una v. a. continua con función de densidad de probabilidad $f_X(x)$,

$$\phi_X(t) = \int_{-\infty}^{+\infty} e^{itx} f_X(x)dx. \tag{5.5}$$

De la definición se derivan, entre otras, las siguientes propiedades:

Pϕ1) $\phi_X(0) = 1$

Pϕ2) $|\phi_X(t)| \leq 1$

Pϕ3) $\phi_X(t)$ es uniformemente continua

En efecto,

$$\phi_X(t+h) - \phi_X(t) = \int_\Omega e^{itX}(e^{ihX} - 1)\,dP.$$

Al tomar módulos,

$$|\phi_X(t+h) - \phi_X(t)| \leq \int_\Omega |e^{ihX} - 1| \, dP, \tag{5.6}$$

pero $|e^{ihX} - 1| \leq 2$ y (5.6) será finito, lo que permite intercambiar integración y paso al límite, obteniendo

$$\lim_{h\to 0} |\phi_X(t+h) - \phi_X(t)| \leq \int_\Omega \lim_{h\to 0} |e^{ihX} - 1| \, dP = 0.$$

Pϕ4) Si definimos $Y = aX + b$,

$$\phi_Y(t) = E(e^{itY}) = E\left(e^{it(aX+b)}\right) = e^{itb}\phi_X(at)$$

Pϕ5) Si $E(X^n)$ existe, la función característica es n veces diferenciable y $\forall k \leq n$ se verifica $\phi_X^{(k)}(0) = i^k E(X^k)$

La propiedad 5 establece un interesante relación entre las derivadas de $\phi_X(t)$ y los momentos de X cuando estos existen, relación que permite desarrollar $\phi_X(t)$ en serie de potencias. En efecto, si $E(X^n)$ existe $\forall n$, entonces,

$$\phi_X(t) = \sum_{k\geq 0} \frac{i^k E(X^k)}{k!} t^k. \tag{5.7}$$

5.4.1 Función característica e independencia

Si $X_1, X_2, \ldots, X_n$ son variables aleatorias independientes y definimos $Y = X_1 + X_2 + \cdots + X_n$. Tendremos

$$\phi_Y(t) = E\left(e^{it(X_1+X_2+\cdots+X_n)}\right) = E\left(\prod_{k=1}^n e^{itX_k}\right) = \prod_{k=1}^n E\left(e^{itX_k}\right) = \prod_{k=1}^n \phi_{X_k}(t), \tag{5.8}$$

expresión que permite obtener con facilidad la función característica de la suma de variables independientes y cuya utilidad pondremos de manifiesto de inmediato.

5.4.2 Funciones características de algunas distribuciones conocidas

Bernoulli.- Si $X \sim B(1,p)$

$$\phi_X(t) = e^0 q + e^{it} p = q + pe^{it}.$$

Binomial.- Si $X \sim B(n,p)$

$$\phi_X(t) = \prod_{k=1}^n (q + pe^{it}) = (q + pe^{it})^n.$$

Poisson.- Si $X \sim P(\lambda)$

$$\phi_X(t) = \sum_{x\geq 0} e^{itx} \frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda} \sum_{x\geq 0} \frac{(\lambda e^{it})^x}{x!} = e^{\lambda(e^{it}-1)}.$$

Normal tipificada.- Si $Z \sim N(0,1)$, sabemos que existen los momentos de cualquier orden y en particular, $E(Z^{2n+1}) = 0$, $\forall n$ y $E(Z^{2n}) = \frac{(2n)!}{2^n n!}$, $\forall n$. Aplicando (5.7),

$$\phi_Z(t) = \sum_{n\geq 0} \frac{i^{2n}(2n)!}{2^n(2n)!n!}t^{2n} = \sum_{n\geq 0} \frac{\left(\frac{(it)^2}{2}\right)^n}{n!} = \sum_{n\geq 0} \frac{\left(-\frac{t^2}{2}\right)^n}{n!} = e^{-\frac{t^2}{2}}.$$

Para obtener $\phi_X(t)$ si $X \sim N(\mu, \sigma^2)$, podemos utilizar el resultado anterior y P4. En efecto, recordemos que X puede expresarse en función de Z mediante $X = \mu + \sigma Z$ y aplicando P4,

$$\phi_X(t) = e^{i\mu t}\phi_Z(\sigma t) = e^{i\mu t - \frac{\sigma^2 t^2}{2}}.$$

Observación 5.1 *Obsérvese que $Im(\phi_Z(t)) = 0$. El lector puede comprobar que no se trata de un resultado exclusivo de la Normal tipificada, si no de una propiedad que poseen todas las v. a. con distribución de probabilidad simétrica, es decir, aquellas que verifican ($P(X \geq x)$=$P(X \leq -x)$).*

Gamma.- Si $X \sim G(\alpha, \lambda)$, su función de densidad de probabilidad viene dada por

$$f_X(x) = \begin{cases} \dfrac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x} & \text{si } x > 0, \\ \\ 0 & \text{si } x \leq 0, \end{cases}$$

por lo que aplicando (5.5),

$$\phi_X(t) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty e^{itx}x^{\alpha-1}e^{-\lambda x}dx,$$

que con el cambio $y = x(1 - it/\lambda)$ conduce a

$$\phi_X(t) = \left(1 - \frac{it}{\lambda}\right)^{-\alpha}.$$

Hay dos casos particulares que merecen ser mencionados:

Exponencial.- La distribución exponencial puede ser considerada un caso particular de $G(\alpha, \lambda)$ cuando $\alpha = 1$. A partir de aquí,

$$\phi_X(t) = \frac{\lambda}{\lambda - it}.$$

ji cuadrado.- Cuando $\alpha = n/2$ y $\lambda = 1/2$, decimos que X tiene una distribución χ^2 con n grados de libertad, $X \sim \chi^2_n$. Su función característica será

$$\phi_X(t) = (1 - 2it)^{-\frac{n}{2}}.$$

5.4.3 Teorema de inversión. Unicidad

Hemos obtenido la función característica de una v. a. X a partir de su distribución de probabilidad, pero es posible proceder de manera inversa por cuanto el conocimiento de $\phi_X(t)$ permite obtener $F_X(x)$.

Teorema 5.7 (Fórmula de inversión de Lévy) *Sean $\phi(t)$ y $F(x)$ las funciones característica y de distribución de la v. a. X y sean $a \leq b$ sendos puntos de continuidad de F, entonces*

$$F(b) - F(a) = \lim_{T\to\infty} \frac{1}{2\pi}\int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it}\phi(t)dt.$$

Que puede también expresarse

$$F(x_0 + h) - F(x_0 - h) = \lim_{T\to\infty} \frac{1}{\pi}\int_{-T}^{T} \frac{\sin ht}{t} e^{-itx_0}\phi(t)dt,$$

donde $h > 0$ y $x_0 + h$ y $x_0 - h$ son puntos de continuidad de F.

Demostración 5.6 *Sea*

$$\begin{aligned} J &= \frac{1}{\pi}\int_{-T}^{T} \frac{\sin ht}{t} e^{-itx_0}\phi(t)dt \\ &= \frac{1}{\pi}\int_{-T}^{T} \frac{\sin ht}{t} e^{-itx_0}\left[\int_R e^{itx}dF(x)\right]dt \\ &= \frac{1}{\pi}\int_{-T}^{T}\left[\int_R \frac{\sin ht}{t} e^{it(x-x_0)}dF(x)\right]dt, \end{aligned}$$

donde $\phi(t) = E(e^{itX}) = \int_R e^{itx}dF(x)$. Pero

$$\left|\frac{\sin ht}{t} e^{it(x-x_0)}\right| \leq \left|\frac{\sin ht}{t}\right| \leq h$$

y como T es finito J será también finito y podemos aplicar el teorema de Fubini que nos permite el cambio en el orden de integración. Es decir

$$\begin{aligned} J &= \frac{1}{\pi}\int_R \left[\int_{-T}^{T} \frac{\sin ht}{t} e^{it(x-x_0)}dt\right] dF(x) \\ &= \frac{1}{\pi}\int_R \left[\int_{-T}^{T} \frac{\sin ht}{t}(\cos t(x-x_0) + i\sin t(x-x_0))dt\right] dF(x) \\ &= \frac{2}{\pi}\int_R \left[\int_{0}^{T} \frac{\sin ht}{t}\cos t(x-x_0)dt\right] dF(x). \end{aligned}$$

Aplicando la fórmula $\sin\alpha\cos\beta = \frac{1}{2}(\sin(\alpha+\beta) + \sin(\alpha-\beta))$, con $\alpha = ht$ y $\beta = t(x - x_0)$,

$$\begin{aligned} J &= \frac{2}{\pi}\int_R \left[\int_{0}^{T} \frac{1}{2t}(\sin t(x-x_0+h) - \sin t(x-x_0-h))dt\right] dF(x) \\ &= \frac{1}{\pi}\int_R \left[\int_{0}^{T} \frac{\sin t(x-x_0+h)}{t}dt - \int_0^T \frac{\sin t(x-x_0-+h)}{t}dt\right] dF(x) \\ &= \frac{1}{\pi}\int_R g(x,T)dF(x). \end{aligned}$$

Por lo que respecta a la función $g(x,T)$ observemos que

$$\lim_{T\to\infty}\int_0^T \frac{\sin x}{x}dx = \frac{\pi}{2}$$

y está acotada para $T>0$. Por otra parte

$$\int_0^T \frac{\sin \alpha x}{x}dx = \int_0^{\alpha T}\frac{\sin u}{u}du$$

y en consecuencia

$$\lim_{T\to\infty}\frac{2}{\pi}\int_0^T \frac{\sin \alpha x}{x}dx = \begin{cases} 1, & \alpha>0; \\ 0, & \alpha=0; \\ -1, & \alpha<0. \end{cases}$$

Aplicándolo a $g(x,T)$,

$$\lim_{T\to\infty}\frac{1}{\pi}g(x,T) = \begin{cases} 0, & x<x_0-h; \\ \frac{1}{2}, & x=x_0-h; \\ 1, & x_0-h<x<x_0+h; \\ \frac{1}{2}, & x=x_0+h; \\ 0, & x>x_0+h. \end{cases}$$

Como $|g(x,T)|$ está acotada podemos hacer uso de los teoremas de convergencia para permutar integración y paso al límite, con lo que tendremos

$$\begin{aligned} \lim_{T\to\infty} J &= \lim_{T\to\infty}\frac{1}{\pi}\int_R g(x,T)dF(x) \\ &= \frac{1}{\pi}\int_R \lim_{T\to\infty} g(x,T)dF(x) \\ &= \int_{x_0-h}^{x_0+h} dF(x) = F(x_0+h)-F(x_0-h). \end{aligned}$$

Este resultado permite obtener $F(x)$ en cualquier x que sea punto de continuidad de F. Basta para ello que en $F(x)-F(y)$ hagamos que $y\to-\infty$ a través de puntos de continuidad. Como $F_X(x)$ es continua por la derecha, la tendremos también determinada en los puntos de discontinuidad sin más que descender hacia ellos a través de puntos de continuidad.

Si la variable es continua, un corolario del anterior teorema permite obtener la función de densidad directamente a partir de la función característica.

Corolario 5.3 *Si $\phi(t)$ es absolutamente integrable en R, entonces la función de distribución es absolutamente continua, su derivada es uniformemente continua y*

$$f(x) = \frac{dF(x)}{dx} = \frac{1}{2\pi}\int_R e^{-itx}\phi(t)dt.$$

Demostración 5.7 *Del desarrollo del teorema anterior tenemos que*

$$\begin{aligned} \left|\frac{F(x_0+h)-F(x_0-h)}{2h}\right| &= \left|\frac{1}{2\pi}\int_R \frac{\sin ht}{ht}e^{-itx_0}\phi(t)dt\right| \\ &\leq \frac{1}{2\pi}\int_R \left|\frac{\sin ht}{ht}e^{-itx_0}\phi(t)\right|dt \\ &\leq \frac{1}{2\pi}\int_R |\phi(t)|dt < +\infty, \end{aligned}$$

y por tanto

$$\lim_{h\to 0}\frac{F(x_0+h)-F(x_0-h)}{2h}=f(x_0)=\frac{1}{2\pi}\int_R e^{-itx_0}\phi(t)dt.$$

Existe por tanto la derivada en todos los puntos de continuidad de F. La continuidad absoluta de F deriva de la desigualdad

$$F(x_0+h)-F(x_0-h)\leq\frac{2h}{2\pi}\int_R|\phi(t)|dt,$$

cuyo segundo miembro podemos hacer tan pequeño como queramos eligiendo h adecuadamente.

Por último, la continuidad uniforme de $f(x)$ se deriva de

$$|f(x+h)-f(x)|=\frac{1}{2\pi}\left|\int_R e^{-ix}(e^{-ith}-1)\phi(t)dt\right|\leq\frac{1}{2\pi}\int_R|e^{-ith}-1||\phi(t)|dt, \tag{5.9}$$

pero

$$\begin{aligned}|e^{-ith}-1| &= |(\cos th-1)-i\sin th|\\ &= \sqrt{(\cos th-1)^2+\sin^2 th}\\ &= \sqrt{2(1-\cos th)}\\ &= 2\left|\sin\frac{th}{2}\right|,\end{aligned}$$

y sustituyendo en (5.9)

$$\begin{aligned}|f(x+h)-f(x)| &\leq \frac{1}{2\pi}\int_R 2\left|\sin\frac{th}{2}\right||\phi(t)|dt\\ &\leq \frac{1}{2\pi}\int_{|t|\leq a}2\left|\sin\frac{th}{2}\right||\phi(t)|dt+\frac{1}{2\pi}\int_{|t|>a}2|\phi(t)|dt,\end{aligned}$$

donde a se elige de forma que la segunda integral sea menor que $\epsilon/2$, lo que es posible por la integrabilidad absoluta de ϕ. La primera integral también puede acotarse por $\epsilon/2$ eligiendo h adecuadamente.

Pero este teorema tiene una trascendencia mayor por cuanto implica la *unicidad* de la función característica, que no por casualidad recibe este nombre, porque *caracteriza*, al igual que lo hacen otras funciones asociadas a X (la de distribución, la de probabilidad o densidad de probabilidad, ...), su distribución de probabilidad. Podemos afirmar que si dos variables X e Y comparten la misma función característica tienen idéntica distribución de probabilidad. La combinación de este resultado con las propiedades antes enunciadas da lugar a una potente herramienta que facilita el estudio y obtención de las distribuciones de probabilidad asociadas a la suma de variables independientes. Veámoslo con algunos ejemplos.

1) Suma de binomiales independientes.- Si las variables $X_k\sim B(n_k,p)$, $k=1,2,\ldots,m$ son independientes, al definir $X=\sum_{k=1}^m X_k$, sabemos por (5.8) que

$$\phi_X(t)=\prod_{k=1}^m\phi_{X_k}(t)=\prod_{k=1}^m(q+pe^{it})^{n_k}=(q+pe^{it})^n,$$

con $n=n_1+n_2+\cdots+n_m$. Pero siendo esta la función característica de una variable $B(n,p)$, podemos afirmar que $X\sim B(n,p)$.

2) Suma de Poisson independientes.- Si nuestra suma es ahora de variables Poisson independientes, $X_k \sim P(\lambda_k)$, entonces

$$\phi_X(t) = \prod_{k=1}^{m} \phi_{X_k}(t) = \prod_{k=1}^{m} e^{\lambda_k(e^{it}-1)} = e^{\lambda(e^{it}-1)},$$

con $\lambda = \lambda_1 + \lambda_2 + \cdots + \lambda_m$. Así pues, $X \sim P(\lambda)$.

3) Combinación lineal de de Normales independientes.- Si $X = \sum_{k=1}^{n} c_k X_k$ con $X_k \sim N(\mu_k, \sigma_k^2)$ e independientes,

$$\begin{aligned} \phi_X(t) &= \phi_{\sum c_k X_k}(t) = \prod_{k=1}^{n} \phi_{X_k}(c_k t) \\ &= \prod_{k=1}^{n} e^{ic_k t\mu_k - \frac{\sigma_k^2 c_k^2 t^2}{2}} \\ &= e^{i\mu t - \frac{\sigma^2 t^2}{2}}, \end{aligned} \tag{5.10}$$

Se deduce de (5.10) que $X \sim N(\mu, \sigma^2)$ con

$$\mu = \sum_{k=1}^{n} c_k \mu_k \qquad \text{y} \qquad \sigma^2 = \sum_{k=1}^{n} c_k^2 \sigma_k^2.$$

4) Suma de Exponenciales independientes.- En el caso de que la suma esté formada por n variables independientes todas ellas $Exp(\lambda)$,

$$\phi_X(t) = \left(\frac{\lambda}{\lambda - it}\right)^n = \left(1 - \frac{it}{\lambda}\right)^{-n},$$

y su distribución será la de una $G(n, \lambda)$.

5) Cuadrado de una $N(0,1)$.- Sea ahora $Y = X^2$ con $X \sim N(0,1)$, su función característica viene dada por

$$\phi_Y(t) = \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{\frac{1}{2}}} e^{-\frac{x^2}{2}(1-2it)} dx = \frac{1}{(1-2it)^{\frac{1}{2}}},$$

lo que nos asegura que $Y \sim \chi_1^2$.

Algunos de estos resultados fueron obtenidos anteriormente, pero su obtención fué entonces mucho más laboriosa de lo que ahora ha resultado.

5.4.4 Distribuciones en el muestreo en una población normal

Si la distribución común a las componentes de la muestra aleatoria de tamaño n es una $N(\mu, \sigma^2)$, se verifica el siguiente teorema:

Teorema 5.8 *Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de tamaño* n *de una población $N(\mu, \sigma^2)$. Entonces, $\overline{X}_n \sim N(\mu, \sigma^2/n)$ y $(n-1)S_n^2/\sigma^2 \sim \chi_{n-1}^2$, y además son independientes.*

Demostración 5.8 *La distribución de* $\overline{X}_n$ *se deduce de (5.10) haciendo* $c_k = 1/n$, $\forall k$. *Tendremos que*

$$\mu = \sum_{k=1}^{n} c_k\mu_k = \mu \qquad y \qquad \sigma^2 = \sum_{k=1}^{n} c_k^2\sigma_k^2 = \frac{\sigma^2}{n}.$$

La obtención de la distribución de S_n^2 *exige primero demostrar su independencia de* $\overline{X}_n$. *Para ello recordemos la expresión de la densidad conjunta de las* n *componentes de la muestra,*

$$f(x_1, x_2, \ldots, x_n) = \frac{1}{\sigma^n(2\pi)^{n/2}} e^{\left\{-\frac{1}{2\sigma^2}\sum_{i=i}^{n}(x_i-\mu)^2\right\}},$$

pero fácilmente se comprueba que

$$\sum_{i=i}^{n}(x_i - \mu)^2 = \sum_{i=i}^{n}(x_i - \overline{x}_n)^2 + n(\overline{x}_n - \mu)^2. \tag{5.11}$$

Sustituyendo en la anterior expresión

$$\begin{aligned} f(x_1, x_2, \ldots, x_n) &= \frac{1}{\sigma^n(2\pi)^{n/2}} e^{\left\{-\frac{1}{2\sigma^2}\sum_{i=i}^{n}(x_i-\overline{x}_n)^2 - \frac{n}{2\sigma^2}(\overline{x}_n-\mu)^2\right\}} \\ &= \frac{1}{\sigma^n(2\pi)^{n/2}} e^{-\left(u/2\sigma^2\right)} e^{-\left(n/2\sigma^2\right)(\overline{x}_n-\mu)^2}, \end{aligned}$$

con $u = \sum(x_i - \overline{x}_n)^2$.

Hagamos ahora el cambio

$$x_i = \overline{x}_n + v_i\sqrt{u}, \quad i = 1, 2, \ldots, n.$$

Como

$$\sum_{i=1}^{n} v_i = 0 \qquad y \qquad \sum_{i=1}^{n} v_i^2 = 1, \tag{5.12}$$

dos de las nuevas variables serán función de las restantes. Resolviendo las ecuaciones de (5.12) para v_{n-1} *y* v_n, *una de las dos soluciones es*

$$v_{n-1} = \frac{A-B}{2} \qquad y \qquad v_n = \frac{A+B}{2},$$

con

$$A = -\sum_{i=1}^{n-2} v_i \qquad y \qquad B = \left(2 - 3\sum_{i=1}^{n-2} v_i^2 - \sum_{i\neq j} v_i v_j\right)^2.$$

Así pues, podemos expresar cada x_i *en función de* $(v_1, \ldots, v_{n-2}, \overline{x}_n, u)$ *de la siguiente forma,*

$$x_i = \overline{x} + v_i\sqrt{u}, \quad i = 1, \ldots, n-2,$$

$$x_{n-1} = \overline{x}_n + \frac{A-B}{2}\sqrt{u}, \quad x_n = \overline{x}_n + \frac{A+B}{2}\sqrt{u}. \tag{5.13}$$

El Jacobiano de (5.13), laborioso pero sencillo de obtener, tiene la forma

$$|J| = u^{(n-2)/2} h(v_1, \ldots, v_{n-2}),$$

donde la expresión exacta de h no es necesaria a los efectos de obtener la densidad conjunta de $(v_1, \ldots, v_{n-2}, \overline{x}_n, u)$. Utilizando del Teorema 3.1 (página 118), podemos escribir

$$g(v_1, \ldots, v_{n-2}, \overline{x}_n, u) = \frac{1}{\sigma^n (2\pi)^{n/2}} u^{n-2} e^{-\left(u/2\sigma^2\right)} e^{-\left(n/2\sigma^2\right)(\overline{x}_n-\mu)^2} h(v_1, \ldots, v_{n-2}),$$

que no es más que el producto de las tres densidades siguientes

$$\begin{aligned} g_1(u) &= c_1 u^{n-2} e^{-\left(u/2\sigma^2\right)} \\ g_2(\overline{x}_n) &= c_2 e^{-\left(n/2\sigma^2\right)(\overline{x}_n-\mu)^2} \\ g_3(v_1, \ldots, v_{n-2}) &= c_3 h(v_1, \ldots, v_{n-2}). \end{aligned}$$

Esta factorización nos permite afirmar que las variables $U = (n-1)S_n^2$, $\overline{X}_n$ y $(V_1, V_2, \ldots, V_{n-2})$ son independientes.

Volvamos ahora a la distribución de S_n^2. De (5.11) podemos escribir,

$$\sum_{i=i}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2 = \sum_{i=i}^{n} \left(\frac{X_i - \overline{X}_n}{\sigma}\right)^2 + \left(\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}\right)^2 = (n-1)\frac{S_n^2}{\sigma^2} + \left(\frac{\sqrt{n}\left(\overline{X}_n - \mu\right)}{\sigma}\right)^2. \tag{5.14}$$

Como $Y_i = (X_i - \mu)/\sigma$, $i = 1, \ldots, n$ y $Z = \sqrt{n}\left(\overline{X}_n - \mu\right)/\sigma$ son todas ellas variables $N(0,1)$, tendremos que

$$\sum_{i=i}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi_n^2 \qquad y \qquad \sqrt{n}\left(\frac{\overline{X}_n - \mu}{\sigma}\right)^2 \sim \chi_1^2,$$

pero como $(n-1)S_n^2$ y $\overline{X}_n$ son independientes, también lo serán $(n-1)S_n^2/\sigma^2$ y $\overline{X}_n$, y al calcular funciones características en ambos miembros de (5.14),

$$(1-2it)^{n/2} = (1-2it)^{1/2} \phi_{(n-1)S_n^2/\sigma^2}.$$

De donde,

$$\phi_{(n-1)S_n^2/\sigma^2} = (1-2it)^{(n-1)/2},$$

y

$$(n-1)\frac{S_n^2}{\sigma^2} \sim \chi_{n-1}^2.$$

5.4.5 Teorema de continuidad de Lévy

Se trata del último de los resultados que presentaremos y permite conocer la convergencia de una sucesión de v. a. a través de la convergencia puntual de la sucesión de sus funciones características.

Teorema 5.9 (Directo) *Sea $\{X_n\}_{n\geq 1}$ una sucesión de v. a. y $\{F_n(x)\}_{n\geq 1}$ y $\{\phi_n(t)\}_{n\geq 1}$ las respectivas sucesiones de sus funciones de distribución y características. Sea X una v. a. y $F_X(x)$ y $\phi(t)$ sus funciones de distribución y característica, respectivamente. Si $F_n \xrightarrow{w} F$ (es decir, $X_n \xrightarrow{L} X$), entonces*

$$\phi_n(t) \longrightarrow \phi(t), \quad \forall t \in R.$$

Resultado que se completa con el teorema inverso.

Teorema 5.10 (Inverso) *Sea* $\{\phi_n(t)\}_{n\geq 1}$ *una sucesión de funciones características y* $\{F_n(x)\}_{n\geq 1}$ *la sucesión de funciones de distribución asociadas. Sea* $\phi(t)$ *una función continua en 0, si* $\forall t \in R$, $\phi_n(t) \to \phi(t)$, *entonces*

$$F_n \xrightarrow{w} F,$$

donde $F(x)$ *en una función de distribución cuya función característica es* $\phi(t)$.

Este resultado permite, como decíamos antes, estudiar el comportamiento límite de sucesiones de v. a. a través del de sus funciones características, generalmente de mayor sencillez. Sin duda una de sus aplicaciones más relevantes ha sido el conjunto de resultados que se conocen como *Teorema Central del Límite* (TCL), bautizados con este nombre por Lyapunov que pretendió así destacar el papel *central* de estos resultados en la Teoría de la Probabilidad.

5.5 Teorema Central de Límite

Una aplicación inmediata es el Teorema de De Moivre-Laplace, una versión temprana del TCL, que estudia el comportamiento asintótico de una $B(n,p)$.

Teorema 5.11 (De Moivre-Laplace) *Sea* $X_n \sim B(n,p)$ *y definamos* $Z_n = \frac{X_n - np}{\sqrt{np(1-p)}}$. *Entonces*

$$Z_n \xrightarrow{L} N(0,1).$$

Demostración 5.9 *Aplicando los resultados anteriores, se obtiene*

$$\phi_{Z_n}(t) = \left((1-p)e^{-it\sqrt{\frac{p}{n(1-p)}}} + pe^{it\sqrt{\frac{(1-p)}{np}}}\right)^n,$$

que admite un desarrollo en serie de potencias de la forma

$$\phi_{Z_n}(t) = \left[1 - \frac{t^2}{2n}(1+R_n)\right]^n,$$

con $R_n \to 0$, *si* $n \to \infty$. *En consecuencia,*

$$\lim_{n\to\infty} \phi_{Z_n}(t) = e^{-\frac{t^2}{2}}.$$

La unicidad y el teorema de continuidad hacen el resto.

Observación 5.2 *Lo que el teorema afirma es que si* $X \sim B(n,p)$, *para n suficientemente grande, tenemos*

$$P\left(\frac{X - np}{\sqrt{np(1-p)}} \leq x\right) \simeq \Phi(x),$$

donde $\Phi(x)$ *es la función de distribución de la* $N(0,1)$.

¿De qué forma puede generalizarse este resultado? Como ya sabemos $X_n \sim B(n,p)$ es la suma de n v. a. i.i.d., todas ellas Bernoulli ($Y_k \sim B(1,p)$), cuya varianza común, $var(Y_1) = p(1-p)$, es finita. En esta dirección tiene lugar la generalización: variables independientes, con igual distribución y con varianza finita.

Teorema 5.12 (Lindeberg) *Sean $X_1, X_2, \ldots, X_n$, v.a. i.i.d. con media y varianza finitas, μ y σ^2, respectivamente. Sea $\overline{X}_n = \frac{1}{n}\sum_{k=1}^n X_k$ su media muestral, entonces*

$$Y_n = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{\overline{X}_n - E(\overline{X}_n)}{\sqrt{var(\overline{X}_n)}} \xrightarrow{L} N(0,1).$$

Demostración 5.10 *Teniendo en cuenta la definición de $\overline{X}_n$ podemos escribir*

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{1}{\sqrt{n}}\sum_{k=1}^n Z_k,$$

con $Z_k = (X_k - \mu)/\sigma$, variables aleatorias i.i.d. con $E(Z_1) = 0$ y $var(Z_1) = 1$. Aplicando Pϕ4 y (5.8) tendremos

$$\phi_{Y_n}(t) = \left[\phi_{Z_1}\left(\frac{t}{\sqrt{n}}\right)\right]^n$$

Pero existiendo los dos primeros momentos de Z_1 y teniendo en cuenta (5.7), $\phi_{Z_1}(t)$ puede también expresarse de la forma

$$\phi_{Z_1}(t) = 1 - \frac{t^2}{2n}(1 + R_n),$$

con $R_n \to 0$, si $n \to \infty$. En consecuencia,

$$\phi_{Y_n}(t) = \left[1 - \frac{t^2}{2n}(1 + R_n)\right]^n.$$

Así pues,

$$\lim_{n\to\infty} \phi_{Y_n}(t) = e^{-\frac{t^2}{2}},$$

que es la función característica de una $N(0,1)$.

Observemos que el Teorema de De Moivre-Laplace es un caso particular del Teorema de Lindeberg, acerca de cuya importancia se invita al lector a reflexionar porque lo que en él se afirma es, ni más ni menos, que sea cual sea la distribución común a las X_i, su media muestral $\overline{X}_n$, adecuadamente tipificada, converge a una $N(0,1)$ cuando $n \to \infty$.

El teorema de Lindeberg, que puede considerarse el teorema central del límite básico, admite una generalización en la dirección de relajar la condición de equidistribución exigida a las variables. Las llamadas condiciones de Lindeberg y Lyapunov muestran sendos resultados que permiten eliminar aquella condición.

Ejemplo 5.5 (La fórmula de Stirling para aproximar n!) *Consideremos una sucesión de variables aleatorias $X_1, X_2, \ldots,$, independientes e idénticamente distribuidas, Poisson de parámetro $\lambda = 1$. La variable $S_n = \sum_{i=1}^n X_i$ es también Poisson con parámetro $\lambda_n = n$. Si $Z \sim N(0,1)$, para n suficientemente grande el TCL nos permite es-*

cribir,

$$\begin{aligned} P(S_n = n) &= P(n-1 < S_n \leq n) \\ &= P\left(-\frac{1}{\sqrt{n}} < \frac{S_n - n}{\sqrt{n}} \leq 0\right) \\ &\approx P\left(-\frac{1}{\sqrt{n}} < Z \leq 0\right) \\ &= \frac{1}{\sqrt{2\pi}} \int_{-1/\sqrt{n}}^{0} e^{-x^2/2} dx \\ &\approx \frac{1}{\sqrt{2\pi n}}, \end{aligned}$$

en donde la última expresión surge de aproximar la integral entre $[-1/\sqrt{n}, 0]$ *de* $f(x) = e^{-x^2/2}$ *mediante el área del rectángulo que tiene por base el intervalo de integración y por altura el* $f(0) = 1$.

Por otra parte,

$$P(S_n = n) = e^{-n} \frac{n^n}{n!}.$$

Igualando ambos resultado y despejando $n!$ *se obtiene la llamada fórmula de Stirling,*

$$n! \approx n^{n+1/2} e^{-n} \sqrt{2\pi}.$$

5.5.1 Una curiosa aplicación del TCL

De Moivre y Laplace dieron en primer lugar una *versión local* del TCL al demostrar que si $X \sim B(n,p)$,

$$P(X = m)\sqrt{np(1-p)} \approx \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \tag{5.15}$$

para n suficientemente grande y $x = \frac{m-np}{\sqrt{np(1-p)}}$. Esta aproximación nos va a servir para estudiar la credibilidad de algunas aproximaciones al número π obtenidas a partir del problema de la *aguja de Buffon*.

Recordemos que en el problema planteado por Buffon se pretende calcular la probabilidad de que una aguja de longitud l, lanzada al azar sobre una trama de paralelas separadas entre si una distancia a, con $a > l$, corte a alguna de las paralelas. Puestos de acuerdo sobre el significado de *lanzada al azar*, la respuesta es

$$P(corte) = \frac{2l}{a\pi},$$

resultado que permite obtener una aproximación de π si, conocidos a y l, sustituimos en $\pi = \frac{2l}{aP(corte)}$ la probabilidad de corte por su estimador natural *la frecuencia relativa de corte*, p, a lo largo de n lanzamientos. Podremos escribir, si en lugar de trabajar con π lo hacemos con su inverso,

$$\frac{1}{\pi} = \frac{am}{2ln},$$

donde m es el número de cortes en los n lanzamientos.

El año 1901 Lazzarini realizó 3408 lanzamientos obteniendo para π el valor 3.1415929 con ¡¡6 cifras decimales exactas!!. La aproximación es tan buena que merece como mínimo alguna pequeña reflexión. Para empezar supongamos que el número de cortes aumenta en una unidad, las aproximaciones de los inversos de π correspondientes a los m y $m+1$ cortes diferirían en

$$\frac{a(m+1)}{2ln} - \frac{am}{2ln} = \frac{a}{2ln} \geq \frac{1}{2n},$$

que si $n \approx 5000$, da lugar a $\frac{1}{2n} \approx 10^{-4}$. Es decir, un corte más produce una diferencia mayor que la precisión de 10^{-6} alcanzada. *No queda más alternativa que reconocer que Lazzarini tuvo la suerte de obtener* exactamente *el número de cortes, m, que conducía a tan excelente aproximación*. La pregunta inmediata es, *>cuál es la probabilidad de que ello ocurriera?*, y para responderla podemos recurrir a (5.15) de la siguiente forma,

$$P(X=m) \approx \frac{1}{\sqrt{2\pi np(1-p)}} e^{-\frac{(m-np)^2}{2np(1-p)}} \leq \frac{1}{\sqrt{2\pi np(1-p)}}.$$

Por ejemplo, si $a = 2l$ entonces $p = 1/\pi$ y para $P(X = m)$ obtendríamos la siguiente cota

$$P(X=m) \leq \sqrt{\frac{\pi}{2n(\pi-1)}}.$$

Para el caso de Lazzarini n=3408 y $P(X = m) \leq 0.0146$, $\forall m$. *Parece ser que Lazzarini era un hombre de suerte, quizás demasiada.*

5.6 Ejercicios

Algunos básicos

* **Ej. 1** — La sucesión de variables aleatorias $\{X_n\}_{n=1}^{\infty}$ converge en probabilidad a una constante b si $\forall \epsilon > 0, \lim_{n\to\infty} P(\mid X_n - b \mid < \epsilon) = 1$

1. Sea $\{Z_n\}_{n=1}^{\infty}$ una sucesión de variables aleatorias, tales que Z_n solo puede tomar los valores 0 y n^2 con probabilidades $1 - \frac{1}{n}$ y $\frac{1}{n}$, respectivamente. Comprueba que la sucesión de variables aleatorias $\{Z_n\}_{n=1}^{\infty}$ converge en probabilidad a 0, pero la sucesión de sus valores esperados diverge.
2. Construye una sucesión de variables aleatorias cuyas medias converjan a 0, pero que no converja en probabilidad.

* **Ej. 2** — La sucesión de variables aleatorias $\{X_n\}_{n=1}^{\infty}$ converge en media cuadrática a una constante b si $\lim_{n\to\infty} E((X_n - b)^2) = 0$

1. Comprueba que la sucesión de variables aleatorias $\{X_n\}_{n=1}^{\infty}$ converge en media cuadrática a b si y solo si sus medias convergen a b y sus varianzas convergen a 0.
2. Comprueba que si $\{X_n\}_{n=1}^{\infty}$ converge en media cuadrática a b, entonces también converge en probabilidad a b.

* **Ej. 3** — Sea $\{X_n\}_{n=1}^{\infty}$ una sucesión de variables aleatorias independientes, todas ellas con la misma distribución cuya media μ y varianza σ^2 existen. Comprueba que la sucesión $\{\overline{X}_n\}_{n=1}^{\infty}$ donde $\overline{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k$ converge en media cuadrática a μ.

* **Ej. 4** — Sea $\{Z_n\}_{n=1}^{\infty}$ una sucesión de variables aleatorias, tales que Z_n solo puede tomar los valores $\frac{1}{n}$ y n con probabilidades $1-\frac{1}{n^2}$ y $\frac{1}{n^2}$, respectivamente.

1. ¿Existe alguna constante b a la que esa sucesión converja en probabilidad?
2. ¿Existe alguna constante b a la que esa sucesión converja en media cuadrática?

* **Ej. 5** — Los clientes de cierto banco efectúan depósitos con media 157.92 euros y desviación típica 30.20 euros. Aparte de esto no se sabe nada más acerca de la distribución de estos depósitos. Como parte de un estudio, se eligieron al azar e independientemente 75 depósitos. ¿Cuál es la probabilidad de que la media de estos 75 depósitos sea 170.00 euros o mayor?

* **Ej. 6** — Los vehículos que cruzan un puente tienen pesos cuya media es de 4675 quilos y cuya desviación estándar es de 345 quilos. Si hay 40 vehículos sobre el puente en un instante dado, hallar el número K tal que la probabilidad (aproximada) de que su peso total no supere a K sea de 0.99.

* **Ej. 7** — La gente que frecuenta cierto pub tiene una probabilidad de 0.001 de salir y cantar con el grupo que está actuando. En una noche determinada hay 150 personas en el pub. ¿Cuál es la probabilidad de que al menos una persona salga y cante con el grupo ? Suponer que cada persona en el pub toma la decisión independientemente del resto. Hallar el verdadero valor y el aproximado por la distribución apropiada con una precisión de 5 dígitos.

* **Ej. 8** — El suceso A tiene una probabilidad de 0.4. Esto significa que esperamos que la frecuencia relativa de A esté cercana a 0.4 en una larga serie de repeticiones del experimento que se está modelizando. ¿Cuál es la probabilidad de que en 1000 experimentos, la frecuencia relativa esté entre 0.38 y 0.42 (inclusive)? Usar una aproximación adecuada.

* **Ej. 9** — Un borracho camina de forma aleatoria de la siguiente forma: Cada minuto da un paso hacia adelante o hacia atrás con igual probabilidad y con independencia de los pasos anteriores. Cada paso es de 50 cm. Utiliza el teorema central del límite para aproximar la probabilidad de que en una hora avance más de 5 metros.

* **Ej. 10** — Una compañía de logística envía paquetes de distintos pesos, con una media de 15 kg y una desviación típica de 10. Teniendo en cuenta que los paquetes provienen de una gran cantidad de clientes diferentes, es razonable modelizar sus pesos como variables aleatorias independientes. Calcular la probabilidad de que el peso total de 100 paquetes exceda de 1700 kg.

* **Ej. 11** — Un jugador de baloncesto encesta un lanzamiento de 3 puntos con probabilidad 0.3. Empleando el teorema central del límite, aproxima la distribución del número de canastas conseguidas en 25 lanzamientos. Calcula la probabilidad de encestar más de 10 canastas.

Y los demás

* **Ej. 12** — Ante la caja de un banco hay 60 personas que esperan recibir su salario, del que sabemos que es una cantidad aleatoria de media $\mu = 100$ *ecus* y desviación típica $\sigma = 30$ euros. Si los salarios son independientes de un asalariado a otro, queremos saber:

1. ¿cuanto dinero debe tener la caja para, con probabilidad 0.95, poder pagar todos los salarios?
2. si la caja cuenta incialmente con 7.000 euros: ¿cúal es la probabilidad de que al final de los pagos queden en caja al menos 500 euros?

* **Ej. 13** — Una empresa produce 10000 bolas de acero para rodamientos, siendo $p = 0.05$ la probabilidad de que una bola sea defectuosa. El proceso de producción garantiza que las bolas son fabricadas independientemente unas de otras. Las bolas defectuosas son arrojadas a un recipiente cuya capacidad queremos determinar para que, con probabilidad 0.99, quepan en él todas las bolas defectuosas del proceso.

* **Ej. 14** — Las bombillas utilizadas por cierto aparato pueden ser de dos clases, A y B. Las de la clase A tienen una duración media $\mu_A = 2000$ horas con devsiación típica $\sigma_A = 400$ horas, mientras que las de la clase B tienen una duración media $\mu_B = 1800$ horas con desviación típica $\sigma_B = 500$ horas. Se compran 200 bombillas de la clase A y 150 de la clase B. Calcular la probabilidad de que la duración media de la muestra de la clase A no supere en más de 100 horas la duración media de la muestra de la clase B.

* **Ej. 15** — Un colegio está preparando la fiesta de graduación de sus 500 estudiantes. Se sabe, por lo ocurrido en otras ocasiones, que el 50% de los estudiantes vienen acompañados de sus padres, el 30% sólo por uno de ellos y el 20% restante vienen solos. >Cuantos asientos hay que disponer para los padres si queremos que, con una probabilidad superior a 0.95, todos ellos puedan sentarse?

* **Ej. 16** — Disponemos de un dado cargado en el que la probabilidad de obtener cualquiera de las caras es proporcional a su número de puntos. Jugamos con él pagando 4 euros por jugada y recibiendo como premio tantos euros como puntos tiene la cara obtenida al lanzarlo. Obtener la probabilidad aproximada de ir ganando al cabo de 100 jugadas.

* **Ej. 17** — Consideremos la variable aleatoria $X \sim U(0,1)$ y definamos la sucesión $X_n = X/n$, $n \geq 1$. Comprobar que $X_n \xrightarrow{L} Y$, siendo Y una variable aleatoria degenerada, $P(Y = 0) = 1$.

* **Ej. 18** — La sucesión de variables aleatorias $\{X_n\}_{n\geq 1}$ tales que $P(X_n = 1 - 1/n) = P(X_n = 1 + 1/n) = 1/2$, convergen en ley a la variable aleatoria X tal que $P(X = 1) = 1$. >Tienden sus funciones de probabilidad a una función de probabilidad?

* **Ej. 19** — Sean X_j, $j = 1, \ldots, n$, variables aleatorias independientes, todas ellas $U(0,1)$. Definimos

$$Y_n = \min(X_1, \ldots, X_n), \quad Z_n = \max(X_1, \ldots, X_n), \quad U_n = nY_n, \quad V_n = n(1-Z_n).$$

Demostrar que cuando $n \to +\infty$,

$$Y_n \xrightarrow{P} 0; \quad Z_n \xrightarrow{P} 1; \quad U_n \xrightarrow{L} U; \quad V_n \xrightarrow{P} V,$$

siendo U y V variables aleatorias exponenciales con parámetro $\lambda = 1$.

* **Ej. 20** — Demostrar que la independencia de las variables en la ley débil de los grandes números puede relajarse exigiendo solamente independencia dos a dos y acotación de las varianzas. Es decir, si X_j, $j = 1, \ldots, n$, verifican que $E(X_j) = \mu_j$ y $var(X_j) = \sigma_j^2$ son finitas, entonces

$$\overline{X}_n - \overline{\mu}_n \xrightarrow{P} 0,$$

donde

$$\overline{\mu}_n = \frac{1}{n}\sum_{j=1}^{n} \mu_j, \quad \overline{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j$$

las X_j son independientes dos a dos y $\sigma_j^2 \leq M$, $\forall j$.

* **Ej. 21** — Aplicar el anterior resultado a las variables aleatorias X_j, $j = 1, \ldots, n$, independientes dos a dos y tales que

$$P(X_j = -a^j) = P(X_j = a^j) = \frac{1}{2}.$$

>Para qué valores de a es aplicable el resultado?

* **Ej. 22** — Sea S_n el número de éxitos en n pruebas Bernoulli con probabilidad de éxito p en cada prueba. Encontrar una cota para $P(|S_n/n - p| \geq \epsilon)$ que no dependa de p.

* **Ej. 23** — Tenemos dos monedas, una correcta y otra con probabilidad de cara $p = 3/4$. Elegimos al azar una de las dos y realizamos una serie de lanzamientos. Después de observar el resultado de un gran número de ellos, >podemos saber la moneda elegida? >Cuál es el mínimo número de lanzamientos para poder saberlo con una probabilidad de al menos 95%?

** **Ej. 24** — **Aplicación de la LGN a la aproximación de funciones** Sea g una función acotada definida sobre $[0, 1]$, la función B_n definida sobre $[0, 1]$ mediante

$$B_n(x) = \sum_{k=0}^{n} g\left(\frac{k}{n}\right)\binom{n}{k} x^k (1-x)^{n-k},$$

es conocida como polinomio de Bernstein de grado n.
El teorema de aproximación de Weierstrass asegura que que toda función continua sobre un intervalo cerrado puede ser aproximada uniformemente mediante polinomios. Probar dicha afirmación para los polinomios de Bernstein.

* **Ej. 25** — Probar que una variable aleatoria X es simétrica sí y sólo sí su función característica es real.

* **Ej. 26** — Encontrar la distribución de las variables aleatorias que tiene por función característica

$$1) \quad \phi_1(t) = \sum_{k\geq 0} a_k \cos kt, \qquad 2) \quad \phi_2(t) = \sum_{k\geq 0} a_k e^{i\lambda_k t}.$$

** **Ej. 27** — Probar con argumentos probabilísticos que

$$e^{-n}\sum_{k=0}^{n}\frac{n^k}{k!} \xrightarrow{n} \frac{1}{2}.$$

* **Ej. 28** — Consideremos $x \in [0,1]$ y sea $\xi_n(x)$={*el n-ésimo dígito de su expresión decimal*}. Definamos $S_n(x) = \sum_{k=1}^{n}\xi_k(x)$ y

$$A_n(y) = \left\{x;\ \frac{2S_n(x) - 9n}{\sqrt{33n}} < y\right\},$$

y sea $\lambda(A_n(y))$ la medida de Lebesgue de $A_n(y)$. Probar que

$$\lim_{n\to\infty}\lambda(A_n(y)) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y} e^{-\frac{u^2}{2}}du.$$

* **Ej. 29** — La fabricación de zumo de naranja se lleva a cabo por lotes de n envases. La caducidad en días de cada envase es una variable aleatoria de media $\mu = 20$ y varianza $\sigma^2 = 9$. Obtener la media y la varianza de la caducidad media del lote y calcular el tamaño mínimo del lote para que con probabilidad 0,99 dicha caducidad media supere los 19 días.

* **Ej. 30** — La variable aleatoria Y se distribuye $Exp(1)$. Definimos

$$X_n = \begin{cases} 1, & \text{si } Y < \ln n; \\ 0, & \text{en caso contario.} \end{cases}$$

Obtener la función de distribución de X_n y estudiar la convergencia en ley de las X_n.

* **Ej. 31** — Sea $T \sim U(-1/c, 1/c)$, $c > 0$, y definamos $Y = cT$. Para $k \in N$, las variables aleatorias X_k se definen mediante,

$$X_k = \begin{cases} -1, & \text{si } -1 < Y < -1/k; \\ 0, & \text{si } -1/k \le Y < 1/k; \\ 1, & \text{si } 1/k \le Y < 1. \end{cases}$$

Demostrar que X_k converge en ley a la variable aleatoria X con función de probabilidad $f_X(-1) = f_X(1) = 1/2$ y $f_X(x) = 0$, $x \notin \{-1, 1\}$.

Capítulo 6

Simulación de variables aleatorias

6.1 Introducción

En el juego del dominó en sus distints versiones, cada jugador elige al azar 7 fichas de entre las 28. Calcular la probabilidad de ciertas composiciones de dichas 7 fichas puede ser tarea sencilla o, en ocasiones, prácticamente imposible por su complejidad. Así, si nos piden la probabilidad de que el jugador no tenga entre sus fichas ningún 6, la fórmula de Laplace nos da como resultado

$$p = \frac{\binom{21}{7}}{\binom{28}{7}} = 0,0982.$$

Si el jugador está jugando a *"la correlativa"*, le interesa saber con qué probabilidad, p_c, elegirá 7 fichas que puedan colocarse correlativamente una tras otra. Dicha probabilidad es matemáticamente intratable.

Un problema de imposibilidad práctica semejante nos ocurriría si quisiéramos calcular la probabilidad de tener un mazo de cartas ordenado de tal forma que permitiera hacer un solitario directamente. >Hemos de conformarnos con no poder dar respuesta a situaciones como las planteadas u otras similares que puedan surgir?

La ley fuerte de los grandes números, LFGN, (ver Teorema 5.4 en la página 195) y la potencia de cálculo de los ordenadores de sobremesa actuales no permiten abordar el problema empíricamente. En efecto, si podemos simular la elección al azar de 7 fichas de entre las 28 y comprobar después si es posible colocarlas correlativamente una tras otra (éxito), repitiendo el proceso de simulación definiremos una variable X_i ligada a la simulación *i-ésima* que tomará valores

$$X_i = \begin{cases} 1, & \text{si la colocación es posible;} \\ 0, & \text{en caso contrario.} \end{cases}$$

Así definida, $X_i \sim B(1, p_c)$, y es independiente de cualquier otra X_j. Si llevamos a cabo n simulaciones, por la LFGN se verificará

$$\sum_{k=1}^{n} \frac{X_i}{n} \xrightarrow{a.s.} p_c,$$

y podremos aproximar el valor de p_c mediante la frecuencia relativa de éxitos que hayamos obtenido.

Un programa adecuado en cualquiera de los lenguajes habituales de programación permitirá averiguar si las 7 fichas son correlativas, pero >cómo simular su elección aleatoria entre las 28 fichas del dominó? Podríamos proceder como sigue:

1. ordenamos las fichas con algún criterio, por ejemplo, comenzar por las blancas ordenadas según el otro valor que las acompaña, continuar por los 1's ordenados según el valor que les acompaña y así sucesivamente hasta finalizar con el 6 doble,

2. las numeramos a continuación del 1 al 28, y

   (a) extraemos al azar, sin repetición, 7 números del 1 al 28, las correspondientes fichas constituirán nuestra elección; o bien,

   (b) permutamos aleatoriamente los 28 números y las fichas correspondientes a los 7 primeros constituirán nuestra elección.

Queda, no obstante, por resolver la forma de simular la extracción de los 7 números o de permutar aleatoriamente los 28. En cualquier caso necesitamos poder generar con el ordenador números aleatorios.

6.2 Generación de números aleatorios

La generación de números aleatorios con el ordenador consiste en una subrutina capaz de proporcionar valores de una variable aleatoria $X \sim U(0,1)$. Cabe preguntarse si una sucesión de valores obtenida de semejante forma es aleatoria. La respuesta es inmediata y decepcionante, no. En realidad, los números generados mediante cualquiera de las subrutinas disponibles a tal efecto en los sistemas operativos de los ordenadores podemos calificarlos como *pseudoaleatorios* que, eso sí, satisfacen los contrastes de uniformidad e independencia.

La mayoría de los generadores de números aleatorios son del tipo *congruencial*. A partir de un valor inicial, X_0, denominado semilla, y enteros fijos no negativos, a, c y m, calculan de forma recursiva

$$X_{i+1} = aX_i + c \quad (\text{módulo } m), \qquad i = 1, 2, \ldots, n. \tag{6.1}$$

Como valores de la $U(0,1)$ se toman los

$$U_i = \frac{X_i}{m}.$$

La expresión (6.1) justifica la denominación de pseudoaleatorios que les hemos otorgado. En primer lugar, porque si iniciamos la generación siempre con el mismo valor de X_0 obtendremos la misma sucesión de valores y, claro está, nada más alejado de la aleatoriedad que conocer de antemano el resultado. En segundo lugar, porque el mayor número de valores distintos que podemos obtener es precisamente m. Se deduce de aquí la conveniencia de que m sea grande.

Puede demostrarse que una elección adecuada de a, c y m proporciona valores que, aun no siendo aleatorios, se comportan como si lo fueran a efectos prácticos porque, como ya hemos dicho, satisfacen las condiciones de uniformidad e independencia.

Si, como es nuestro caso en el ejemplo del dominó, deseamos elegir al azar números entre 1 y 28, recordemos que si $U_i \sim U(0,1)$, $kU_i \sim U(0,k)$ y por tanto,

$$P(j-1 < kU_i \leq j) = \frac{1}{k}, \quad j = 1, 2, \ldots, k.$$

Si definimos $N_k = [kU_i] + 1$, donde $[\cdot]$ representa la parte entera del argumento, N_k es una variable discreta que tiene la distribución deseada, uniforme en $\{1, 2, \ldots, k\}$.

6.3 Técnicas generales de simulación de variables aleatorias

La generación de una variable $U(0,1)$ es un primer paso necesario pero no suficiente, como el ejemplo del dominó ha puesto de manifiesto. Nuestro objetivo era generar valores de una uniforme discreta sobre el soporte $\{1, 2, \ldots, k\}$ y lo hemos conseguido previa simulación de una $U(0,1)$. En muchos procesos industriales o en el estudio de fenómenos experimentales, un estudio previo de su comportamiento requiere simular cantidades aleatorias que siguen, por lo general, una distribución diferente de la uniforme. >Cómo hacerlo? A continuación presentamos tres métodos para simular variables aleatorias.

Antes de responder digamos que el origen de la simulación de variables aleatorias, conocida como *simulación de Montecarlo*, se debe a von Neumann y Ulam que lo utilizaron por primera vez durante la segunda guerra mundial para simular el comportamiento de la difusión aleatoria de neutrones en la materia fisionable. El trabajo, relacionado con la fabricación de la primera bomba atómica, se desarrollaba en el laboratorio de Los Álamos y Montecarlo fue el código secreto que ambos físicos le dieron.

6.3.1 Método de la transformación inversa

Este método se basa en el resultado de la Proposición 2.6 de la página 78. Si $U \sim U(0,1)$, F es una función de distribución de probabilidad y definimos

$$F^{-1}(x) = \inf\{t: \; F(t) \geq x\},$$

la variable $X = F^{-1}(U)$ verifica que $F_X = F$.

Este resultado permite generar valores de una variable aleatoria con función de distribución dada, sin más que obtener la antiimagen mediante F de valores de una $U(0,1)$. Veamos como generar una exponencial de parámetro λ.

Ejemplo 6.1 (Generación de $X \sim Exp(\lambda)$) *La densidad de una Exponencial de parámetro λ es*

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0; \\ 0, & \text{en el resto,} \end{cases}$$

y su función de distribución,

$$F(x) = \begin{cases} 0, & x < 0; \\ 1 - e^{-\lambda x}, & x \geq 0;. \end{cases}$$

De aquí,

$$1 - e^{-\lambda X} = U \quad \Longrightarrow \quad X = -\frac{1}{\lambda}\ln(1-U),$$

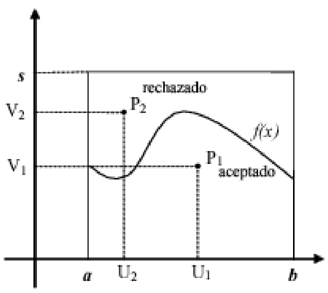
nos permite generar valores de una Exp(λ). Observemos que $1-U$ es también $U(0,1)$ con lo que

$$X = -\frac{1}{\lambda}\ln U,$$

genera también valores de una Exp(λ).

Recordemos que si $Y \sim Ga(n, 1/\lambda)$ entonces Y es la suma de n *exponenciales independientes de parámetro λ (página 133). Aplicando el resultado anterior podremos generar valores de una $Ga(n,1/\lambda)$ a partir de* n *$U(0,1)$, $U_1, U_2, \ldots, U_n$ mediante la expresión*

$$Y = -\sum_{i=1}^{n} \frac{1}{\lambda}\ln U_i = -\frac{1}{\lambda}\ln\left(\prod_{i=1}^{n} U_i\right).$$



6.3.2 Método de aceptación-rechazo

Si queremos generar una variable aleatoria X con densidad $f(x)$, cuyo soporte es el intervalo $[a,b]$, podemos proceder tal como ilustra la figura. Si s es una cota superior de $f(x)$, $f(x) \leq s$, $\forall x \in [a,b]$, generamos aleatoriamente un punto (U,V) en el rectángulo $[a,b] \times [0,s]$. Si $V \leq f(U)$ hacemos $X = U$, en caso contrario rechazamos el punto y generamos uno nuevo.

La primera pregunta que surge es si el número de simulaciones necesarias para obtener un valor de X será finito. Veamos que lo es con probabilidad 1. Para ello, designemos por A la sombra de $f(x)$ y por $\{P_n, n \geq 1\}$ una sucesión de puntos elegidos al azar en $[a,b]\times[0,s]$. La probabilidad de que uno cualquiera de ellos, P_i, no cumpla con la condición $V \leq f(U)$, equivale a $\{P_i \notin A\}$ y vale

$$P(P_i \notin A) = \frac{s(b-a) - |A|}{s(b-a)} = 1 - \frac{1}{s(b-a)},$$

pues $|A| = \int_a^b f(x)dx = 1$. Como la elección es al azar, la probabilidad de que ninguno de ellos esté en A,

$$P\left(\cap_{n\geq 1}\{P_n \notin A\}\right) = \lim_{n\to\infty}\left(1 - \frac{1}{s(b-a)}\right)^n = 0,$$

y por tanto con probabilidad 1 el número necesario de simulaciones para generar un valor de X será finito.

Queda ahora por comprobar que X se distribuye con densidad $f(x)$. En efecto,

$$P(X \leq x) = \sum_{n\geq 1} P(X \leq x, X = U_n),$$

donde

$$\{X \leq x, X = U_n\} = \{P_1 \notin A, \ldots, P_{n-1} \notin A, P_n \in [a,x]\times[0,s]\}.$$

Al tomar probabilidades,

$$\begin{aligned}
P(X \leq x) &= \sum_{n \geq 1} P(X \leq x, X = U_n) \\
&= \sum_{n \geq 1} \left(1 - \frac{1}{s(b-a)}\right)^{n-1} \frac{\int_a^x f(x)dx}{s(b-a)} \\
&= \int_a^x f(x)dx.
\end{aligned}$$

Ejemplo 6.2 *Una variable $X \sim Be(4,3)$ tiene por densidad,*

$$f(x) = \begin{cases} 60x^3(1-x)^2, & 0 \leq x \leq 1; \\ 0, & \textit{en el resto.} \end{cases}$$

Su función de distribución será un polinomio de grado 6, lo que dificulta la generación de una variable aleatoria con esta distribución mediante el método de la transformación inversa. Podemos, en su lugar, recurrir al método de aceptación-rechazo. Tomemos para ello $s = \max_{0 \leq x \leq 1} f(x) = f(0.6) = 2.0736$. El procedimiento consistirá en generar

$$(U, V) \sim U([0,1] \times [0; 2.0736]),$$

pero como $kV \sim U(0,k)$ si $V \sim U(0,1)$, bastará con generar sendas $U(0,1)$ y hacer $X = U$ si

$$2.0736V \leq 60U^3(1-U)^2 \quad \Longrightarrow \quad V \leq \frac{60U^3(1-U)^2}{2.0736}.$$

Generalización del método de aceptación-rechazo

El método anterior puede generalizarse utilizando una función $t(x)$ que mayore a $f(x)$, $\forall x$, en lugar de una cota superior. Si $t(x)$ es tal que

$$1 \leq \int_{-\infty}^{\infty} t(x)dx = c < +\infty,$$

la función $g(x) = t(x)/c$ es una densidad. El método exige que podamos generar con facilidad una variable aleatoria Y con densidad $g(x)$ y, en ese caso, los pasos a seguir son,

1. Generamos Y cuya densidad es $g(y) = t(y)/c$.
2. Generamos $U \sim U(0,1)$.
3. Si $U \leq f(Y)/t(Y)$, hacemos $X = Y$, en caso contrario reiniciamos el proceso desde 1.

La X así generada tiene por densidad $f(x)$ porque

$$\{X \leq x\} = \{Y_N \leq x\} = \{Y \leq x | U \leq f(Y)/t(Y)\},$$

donde N designa el número de la iteración en la que se ha cumplido la condición. Al tomar probabilidades

$$P(X \leq x) = P(Y \leq x | U \leq f(Y)/t(Y)) = \frac{P(Y \leq x, U \leq f(Y)/t(Y))}{P(U \leq f(Y)/t(Y))}.$$

Como Y y U son independientes, su densidad conjunta será

$$h(y,u) = g(y), \quad 0 \leq u \leq 1, \quad -\infty < y < \infty.$$

De aquí,

$$\begin{aligned} P(X \leq x) &= \frac{1}{P(U \leq f(Y)/t(Y))} \int_{-\infty}^{x} g(y) \left[\int_{0}^{f(y)/cg(y)} du \right] dy \\ &= \frac{1}{cP(U \leq f(Y)/t(Y))} \int_{-\infty}^{x} f(y) dy. \end{aligned} \tag{6.2}$$

Pero

$$\lim_{x \to \infty} P(X \leq x) = 1 = \frac{1}{cP(U \leq f(Y)/t(Y))} \int_{-\infty}^{\infty} f(y) dy = \frac{1}{cP(U \leq f(Y)/t(Y))}. \tag{6.3}$$

Sustituyendo en (6.2),

$$P(X \leq x) = \int_{-\infty}^{x} f(y) dy.$$

Respecto al número N de iteraciones necesarias para obtener un valor de X, se trata de una variable geométrica con $p = P(U \leq f(Y)/t(Y))$. De (6.3) deducimos que $P(U \leq f(Y)/t(Y)) = 1/c$ y $E(N) = c$, además

$$P(N = \infty) = \lim_{n \to \infty} P(N > n) = \lim_{n \to \infty} \sum_{j \geq n+1} \frac{1}{c} \left(1 - \frac{1}{c}\right)^{j-1} = \lim_{n \to \infty} \left(1 - \frac{1}{c}\right)^{n} = 0,$$

con lo que N será finito con probabilidad 1.

Ejemplo 6.3 (Simulación de una N(0,1). Método general de aceptación-rechazo) *Recordemos en primer lugar la densidad de $Z \sim N(0,1)$,*

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad -\infty < z < \infty.$$

Si definimos $X = |Z|$, su densidad será

$$f_X(x) = \frac{2}{\sqrt{2\pi}} e^{-x^2/2}, \quad 0 < x < \infty. \tag{6.4}$$

Vamos a utilizar el método general de aceptación-rechazo para simular valores de la N(0,1) utilizando como paso previo la generación de valores de X con densidad (6.4). Para ello tomaremos como función auxiliar $t(x) = \sqrt{2e/\pi} e^{-x}$, $x \geq 0$, porque mayora a $f_X(x)$ en todo su dominio,

$$\frac{f_X(x)}{t(x)} = \frac{\sqrt{2/\pi} e^{-x^2/2}}{\sqrt{2e/\pi} e^{-x}} = \exp\left\{ -\frac{(x-1)^2}{2} \right\} \leq 1, \quad x \geq 0.$$

Por otra parte,

$$\int_{0}^{\infty} t(x) dx = \sqrt{\frac{2e}{\pi}},$$

y por tanto

$$g(x) = \frac{t(x)}{\sqrt{2e/\pi}} = e^{-x},$$

que es la densidad de una Exponencial con parámetro $\lambda = 1$. De acuerdo con el procedimiento antes descrito,

1. *Generaremos sendas variables aleatorias, Y y U, la primera $Exp(1)$ y la segunda $U(0,1)$.*

2. *Si*
$$U \leq \frac{f_X(Y)}{t(Y)} = \exp\left\{-\frac{(Y-1)^2}{2}\right\},$$
haremos $X = Y$, en caso contrario comenzaremos de nuevo.

Una vez obtenido un valor de X el valor de Z se obtiene haciendo $Z = X$ o $Z = -X$ con probabilidad 1/2.

El procedimiento anterior puede modificarse si observamos que aceptamos Y si

$$U \leq \exp\{-(Y-1)^2/2\} \iff -\ln U \geq (Y-1)^2/2.$$

Pero de acuerdo con el ejemplo 6.1 $-\ln U \sim Exp(1)$ con lo que la generación de valores de una $N(0,1)$ se pude llevar a cabo mediante,

1. *Generaremos sendas variables aleatoria Y_1, Y_2, ambas $Exp(1)$.*

2. *Si $Y_2 \geq (Y_1-1)^2/2$, hacemos $X = Y_1$, en caso contrario comenzaremos de nuevo.*

3. *Generamos $U \sim U(0,1)$ y hacemos*
$$Z = \begin{cases} X, & \text{si } U \leq 1/2; \\ -X, & \text{si } U > 1/2. \end{cases}$$

Ejemplo 6.4 (Simulación de una N(0,1). El método polar) *Si $X \sim N(0,1)$ e $Y \sim N(0,1)$ y son independientes, su densidad conjunta viene dada por*

$$f_{XY}(x,y) = \frac{1}{2\pi}\exp\left\{-\frac{x^2+y^2}{2}\right\}.$$

Consideremos ahora las coordenadas polares del punto (X,Y), $R = \sqrt{X^2+Y^2}$ y $\Theta = \arctan Y/X$. Para obtener su densidad conjunta, necesitamos las transformaciones inversas, $X = R\cos\Theta$ e $Y = R\sin\Theta$. El correspondiente jacobiano vale $J_1 = R$ y su densidad conjunta,

$$f_{R\Theta}(r,\theta) = \begin{cases} \dfrac{1}{2\pi} r \exp\left\{-\dfrac{r^2}{2}\right\}, & r \geq 0,\ 0 \leq \theta \leq 2\pi, \\ \\ 0, & \text{en el resto.} \end{cases}$$

De aquí se obtienen fácilmente las densidades marginales de R^2 y Θ que resultan ser $R^2 \sim Exp(1/2)$ y $\Theta \sim U(0,2\pi)$. Haciendo uso del resultado del ejemplo 6.1, si generamos dos variables $U80,1)$, U_1 y U_2,

$$\begin{aligned} R^2 &= -2\ln U_1 \sim Exp(1/2) \\ \Theta &= 2\pi U_2 \sim U(0,2\pi), \end{aligned}$$

y de aquí

$$\begin{aligned} X &= (-2\ln U_1)^{1/2}\cos 2\pi U_2 \\ Y &= (-2\ln U_1)^{1/2}\sin 2\pi U_2, \end{aligned}$$

son $N(0,1)$ e independientes.

6.3.3 Simulación de variables aleatorias discretas

El método más extendido para simular variables discretas es el de la transformación inversa. Si la variable X cuyos valores queremos simular tiene por función de distribución $F(x)$, entonces

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} p_i,$$

con $x_i \in D_X$, el soporte de la variable, y $p_i = P(X = x_i)$.

Si suponemos que los valores del soporte están ordenados, $x_1 < x_2 < \ldots$, el algoritmo consiste en los pasos siguientes:

1. Generamos $U \sim U(0,1)$.
2. Hacemos $X = k$, con $k = \min\{i; U \leq F(x_i)\}$.

La variable así generada se distribuye como deseamos. En efecto,

$$P(X = x_i) = P\left(F(x_{i-1} < U \leq F(x_i)\right) = F(x_i) - F(x_{i-1}) = p_i.$$

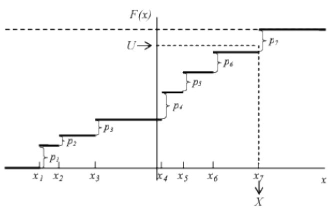


Figura 6.1: Función de distribución de una discreta.

La figura 6.1 ilustra gráficamente el algoritmo para una variable discreta con soporte $D_X = \{x_1, x_2, \ldots, x_7\}$.

A continuación presentamos la adaptación de este algoritmo para la simulación de las variables discretas más conocidas. En algunos casos, el algoritmo no resulta reconocible porque la forma de la función de cuantía permite formas de búsqueda más ventajosas que lo enmascaran.

Ejemplo 6.5 (Simulación de una variable Bernoulli) *Si $X \sim B(1,p)$, el algoritmo que sigue es muy intuitivo y equivale al algoritmo de la transformación inversa si intercambiamos U por $1-U$.*

1. *Generamos $U \sim U(0,1)$.*
2. *Hacemos*

$$X = \begin{cases} 1, & \textit{si } U \leq p; \\ 0, & \textit{si } U > p. \end{cases}$$

Ejemplo 6.6 (Simulación de una variable uniforme) *Para generar X con $D_X = \{x_1, x_2, \ldots, x_n\}$ con $P(X = x_i) = 1/n$, $\forall i$,*

1. *Generamos $U \sim U(0,1)$.*
2. *Hacemos $X = x_k$, con $k = 1 + \lfloor nU \rfloor$ donde $\lfloor s \rfloor$, es la parte entre por defecto de* s.

Ejemplo 6.7 (Simulación de una variable Binomial) *Si $X \sim B(n,p)$, entonces X es la suma de* n *Bernoullis independientes con igual parámetro, bastará por tanto repetir* n *veces el algoritmo del ejemplo 6.5 y tomar $X = \sum_{i=1}^{n} X_i$, la suma de los* n *valores generados*

Ejemplo 6.8 (Simulación de una variable Geométrica) *Recordemos que $X \sim Ge(p)$, X es el número de pruebas Bernoulli necesarias para alcanzar el primer éxito, de aquí, $P(X = i) = p(1-p)^{i-1}$, $i \geq 1$, y $P(X > i) = \sum_{j>i} P(X = j) = (1-p)^i$. Tendremos que*

$$F(k-1) = P(X \leq k-1) = 1 - P(X > k-1) = 1 - (1-p)^{k-1}.$$

Aplicando el algoritmo de la transformación inversa

1. *Generamos* $U \sim U(0,1)$.

2. *Hacemos* $X = k$, *tal que* $1-(1-p)^{k-1} < U \leq 1-(1-p)^k$, *que podemos también escribir* $(1-p)^k < 1-U \leq (1-p)^{k-1}$. *Como* $1-U \sim U(0,1)$, *podemos definir* X *mediante*

$$\begin{aligned} X &= \min\{k; (1-p)^k < U\} \\ &= \min\{k; k\ln(1-p) < \ln U\} \\ &= \min\left\{k; k > \frac{\ln U}{\ln(1-p)}\right\}. \end{aligned}$$

El valor que tomaremos para X *será,*

$$X = 1 + \left\lfloor \frac{\ln U}{\ln(1-p)} \right\rfloor. \tag{6.5}$$

Ejemplo 6.9 (Simulación de una variable Binomial Negativa) *Para generar valores de* $X \sim BN(r,p)$ *recordemos que* X *puede expresarse como suma de* r *variables independientes todas ellas Geométricas de parámetro* p *(véase el ejemplo 4.20 de la página 160). Bastará por tanto utilizar el siguiente algoritmo:*

1. *Simulamos* $X_1, X_2, \ldots, X_r$ *todas ellas* $Ge(p)$, *utilizando para ello la expresión (6.5).*

2. *hacemos* $X = X_1 + X_2 + \ldots + X_r$.

Ejemplo 6.10 (Simulación de una variable Poisson) *La simulación de una variable* $X \sim Po(\lambda)$ *se basa en la propiedad que liga la distribución de Poisson con la Exponencial de igual parámetro. Como señalábamos en la página 71, al estudiar la ocurrencia de determinado suceso a largo del tiempo, por ejemplo las partículas emitidas por un mineral radiactivo durante su desintegración, el tiempo que transcurre entre dos ocurrencias consecutivas es una variable aleatoria* $Y \sim Exp(\lambda)$ *y el número de ocurrencias que se producen en el intervalo [0,t] es* $X_t \sim Po(\lambda t)$, *donde* λ *es el número medio de ocurrencias por unidad de tiempo. De acuerdo con esto, el algoritmo de simulación es el siguiente:*

1. *Simulamos* $U_1, U_2, U_3, \ldots$, *todas ellas* $U(0,1)$.

2. *Hacemos* $X = N - 1$, *donde*

$$N = \min\left\{n; \prod_{i=1}^{n} U_i < e^{-\lambda}\right\}.$$

La variable así obtenida se distribuye como una $Po(\lambda)$. *En efecto,*

$$\begin{aligned} X+1 &= \min\left\{n; \prod_{i=1}^{n} U_i < e^{-\lambda}\right\}. \\ X &= \max\left\{n; \prod_{i=1}^{n} U_i \geq e^{-\lambda}\right\} \\ &= \max\left\{n; \sum_{i=1}^{n} \ln U_i \geq -\lambda\right\} \\ &= \max\left\{n; \sum_{i=1}^{n} -\ln U_i \leq \lambda\right\}, \end{aligned}$$

pero como vimos en el Ejemplo 6.1, $-\ln U_i \sim Exp(1)$*, con lo que* X *puede interpretarse com el mayor número de* $Exp(1)$ *cuya suma es menor que* λ*. Ahora bien, exponenciales independientes de parámetro 1 equivalen a tiempos de espera entre ocurrencias de un suceso con una ratio media de ocurrencias de 1 suceso por unidad de tiempo, lo que permite interpretar* X *como el número y. Así,* $X \sim Po(\lambda)$*.*

Apéndice A

Combinatoria

A.1 Principio básico del conteo

Supongamos que se realizan dos experimentos. Si el primero puede tener m resultados diferentes y por cada resultado del primero hay n resultados del segundo, entonces hay $n \times m$ resultados posibles para los dos experimentos conjuntamente.

Este principio se puede generalizar para r experimentos, cada uno con n_i posibles resultados para cada resultado de los $i-1$ anteriores $(i = 1, \dots r)$, en este caso, el número total de resultados es $n_1 \times n_2 \times \dots \times n_r$

Ejemplo A.1 *¿Cuántas matrículas de automóvil distintas se tendrían si los 3 primeros lugares los ocuparan letras y los 4 últimos números? ¿Y si se prohíbe que se repitan letras y números?*

Respuesta.- $26 \times 26 \times 26 \times 10 \times 10 \times 10 \times 10 = 175.760.000$ y
$26 \times 25 \times 24 \times 10 \times 9 \times 8 \times 7 = 78.624.000$

A.2 Permutaciones

¿De cuántas maneras distintas se pueden ordenar las letras a,b,c? La respuesta es 6: abc, acb, bac, bca, cab y cba. Cada una de éstas ordenaciones es una permutación de las 3 letras.
Si se tienen n objetos distintos, existen $n \times (n-1) \times (n-2) \dots 3 \times 2 \times 1 = n!$ permutaciones de los n objetos.

Ejemplo A.2 *Luis quiere poner 10 libros en una estantería. De los 10 libros, 4 son de matemáticas, 3 de química, 2 de historia y 1 de lengua. El chico quiere ordenarlos de forma que queden juntos los de la misma materia. ¿Cuántas ordenaciones diferentes son posibles?*

Respuesta.- $4! \times 4! \times 3! \times 2!$

A.3 Permutaciones con repetición

¿Cuántas palabras distintas pueden formarse con las letras PEPPER? Si pudiéramos distinguir todas las letras, entonces $P_1E_1P_2P_3E_2R$ no sería lo mismo que $P_2E_1P_1P_3E_2R$ y la respuesta sería que pueden formarse 6! palabras, sin embargo es evidente que todas estas dan lugar a la misma palabra:

$P_1E_1P_2P_3E_2R$	$P_1E_1P_3P_2E_2R$	$P_2E_1P_1P_3E_2R$
$P_2E_1P_3P_1E_2R$	$P_3E_1P_1P_2E_2R$	$P_3E_1P_2P_1E_2R$
$P_1E_2P_2P_3E_1R$	$P_1E_2P_3P_2E_1R$	$P_2E_2P_1P_3E_1R$
$P_2E_2P_3P_1E_1R$	$P_3E_2P_1P_2E_1R$	$P_3E_2P_2P_1E_1R$

Así pues hay $\frac{6!}{3!2!}$= 60 palabras distintas.

En general, si tenemos n objetos de los cuales n_1 son de un tipo, n_2 son de otro, ...y n_r son de otro, hay $\frac{n!}{n_1!n_2!\ldots n_r!}$ permutaciones distintas.

A.4 Combinaciones

En muchas ocasiones interesa saber el número de grupos diferentes de r objetos que pueden formarse si se dispone de n diferentes. Cuando el orden de selección es importante, ya sabemos que éste número es $n \times (n-1) \times \ldots \times (n-r+1)$. Como de esta forma cada grupo de r items aparecería contado $r!$ veces, el número de grupos diferentes de r items elegidos entre n diferentes es

$$\frac{n \times (n-1) \times \ldots \times (n-r+1)}{r!} = \frac{n!}{(n-r)!r!}.$$

Definición A.1 *Definimos*

$$\binom{n}{r} = \frac{n!}{(n-r)!r!}, \ \ con\ r \leq n$$

y decimos que representa el número de combinaciones *posibles de* n *objetos tomados de* r *en* r*, cuando el orden de selección no se considera importante.*

Ejemplo A.3 *A partir de un grupo de 5 hombres y 7 mujeres ¿Cuántos comités que consten de 2 hombres y 3 mujeres se pueden formar? ¿Y si dos de las mujeres no quieren estar juntas en el mismo comité?*

Respuesta.- 350 y 300

Ejemplo A.4 *Supongamos que tenemos* n *componentes electrónicas de las cuales* m *son defectuosas (D) y* $n-m$ *funcionan (F) correctamente. Las componentes son indistinguibles a simple vista, ¿de cuántas formas pueden ordenarse de forma que no haya dos defectuosas consecutivas?*

Si $m > n-m+1$ no hay ninguna porque en cualquier ordenación habría 2 defectuosas consecutivas necesariamente.

Supongamos que $m \leq n-m+1$. La siguiente notación nos ayudará a resolver el problema. Dada una ordenación, por ejemplo $DFDFDFFFF$ $(n=9,\ m=3,\ n-m=6)$, le asociamos biunívocamente una serie en la que indicamos cuantas componentes defectuosas hay antes de cada una que funciona (si la última funciona escribimos un 0 al final y si las últimas son defectuosas indicamos cuántas hay) en este caso $1F1F1F0F0F0F0$. Otros 2 ejemplos, $FDDFDFFFF$, su representante es $0F2F1F0F0F0F0$, y $FDFFDFFFD$, en este caso lo denotaríamos $0F1F0F1F0F0F1$.

Observa que, en general, serán series de $n-m+1$ números y $n-m$ $F's$. Lo que sucede es que, ahora, las F's estan fijas y realmente las podríamos eliminar y asociar a cada ordenación únicamente la serie de números que la caracteriza, en los casos anteriores 1110000 , 0210000 y 0101001 .

Así que nos interesará contar las series de $n-m+1$ números en las que aparezcan m $1's$ (siendo 0's el resto) . ¿Cuántas hay?
Respuesta.- $\binom{n-m+1}{m}$

A.5 Argumentos combinatorios para justificar propiedades

Demostrar la siguiente igualdad

$$\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r} \text{ para } 1 \leq r \leq n.$$

Este resultado puede demostrarse analíticamente o mediante argumentos combinatorios.

Supongamos que tenemos n objetos y formamos grupos de r elementos, esto puede hacerse de $\binom{n}{r}$ maneras diferentes. Si nos fijamos en un objeto en particular, vemos que $\binom{n-1}{r-1}$ de los grupos lo contienen y $\binom{n-1}{r}$ no.

A.6 Coeficientes multinomiales

Consideremos un conjunto de n items distintos que tiene que dividirse en r grupos distintos de tamaños $n_1, n_2, \ldots, n_r$, de manera que $\sum_{i=1}^{r} n_i = n$. ¿Cuántas divisiones diferentes son posibles?

La respuesta es

$$\binom{n}{n_1}\binom{n-n_1}{n_2}\cdots\binom{n-n_1-\ldots-n_{r-1}}{n_r} = \frac{n!}{n_1!n_2!\ldots n_r!}$$

Definición A.2 *Si $\sum_{i=1}^{r} n_i = n$, llamamos coeficientes multinomiales a*

$$\binom{n}{n_1\ n_2 \ldots n_r} = \frac{n!}{n_1!n_2!\ldots n_r!},$$

que representan el número de posibles divisiones diferentes de n *objetos distintos en* r *grupos de tamaños $n_1, n_2 \ldots n_r$.*

Ej. 1 — Una comisaría de policía de un pueblo consta de 10 oficiales. Si 5 deben patrullar las calles, 2 se quedan trabajando en la comisaría y 3 están de reserva, ¿cuántas divisiones diferentes en 3 grupos pueden hacerse con los 10 policías?

A.7 Distribución de bolas en urnas

Es fácil ver que hay r^n posibles resultados cuando n bolas distinguibles se distribuyen en r urnas distinguibles, ya que, cada bola puede caer en cualquiera de las r urnas diferentes.

Vamos a suponer que las n bolas son indistinguibles, ¿cuántos resultados posibles hay en éste caso? Observa que cada resultado puede describirse mediante un vector de r componentes $(x_1, x_2, \ldots x_r)$ donde

x_i denota el número de bolas en la urna i-ésima. Así que el problema se reduce a encontrar el número de vectores diferentes que hay con valores enteros no negativos y de forma que la suma de sus r componentes sea n.

Vamos a resolver primero un caso particular y después generalizaremos. Supongamos que $n = 8$ y $r = 3$, un vector de los que estamos buscando sería, por ejemplo, (2,4,2). Podríamos identificar este vector con la representación 00 | 0000 | 00. Otro vector podría ser (6,0,2) que se corresponde con 000000 | | 00. Observa que lo que diferencia una configuración de otra es *dónde se encuentran las barras* que indican la separación entre cajas. Las barras, en el primer caso, están en las posiciones {3, 8}, en el segundo, en las posiciones {7, 8}. Por tanto el problema se reduce a encontrar cuantas posiciones distintas pueden ocupar las 2 barras separadoras. La respuesta es que tantas como grupos de tamaño 2 podamos construir con los números {1, 2, 3, 4, 5, 6, 7, 8, 9,10}, es decir, $\binom{10}{2}$.

En general si tenemos n bolas y r urnas, habrá $r - 1$ barras separadoras y $n + r - 1$ lugares donde podría aparecer una barra. Entonces la solución en el caso general es $\binom{n+r-1}{r-1}$

Ejemplo A.5 *Un inversor tiene 20 mil euros para invertir en 4 activos. Él desea que cada inversión se haga en unidades de mil euros. Si quiere invertir los 20 mil, ¿de cuántas maneras diferentes puede hacerlo? ¿Y si no tiene que invertirlo todo?*

Respuesta.- $\binom{23}{3}$ y $\binom{24}{4}$

Apéndice B

Algunos resultados necesarios

B.1 Series aritmético-geométricas

Se trata de una serie aritmético-geométrica porque su término general es el producto del término general de una serie aritmética por el término general de una serie geométrica. Es probabilidad nos encontramos con frecuencia son este tipo de serie. Veamos cómo sumarlas. Designemos al término general de la serie como $b_n = a_n p^n$, con $a_{n+1} = a_n + r$ y $p < 1$. Tendremos

$$S = \sum_{n\geq 1} b_n = \sum_{n\geq 1} a_n p^n,$$

que al multiplicarla por p da lugar a

$$\begin{aligned} p\ S = \sum_{n\geq 1} a_n p^{n+1} = \\ \sum_{n\geq 1}(a_{n+1} - r)p^{n+1} = \sum_{n\geq 1} a_{n+1}p^{n+1} - \sum_{n\geq 1} rp^{n+1} = \\ \sum_{m\geq 2} a_m p^m - r\frac{p^2}{1-p} = S - a_1 p - r\frac{p^2}{1-p}. \end{aligned} \tag{B.1}$$

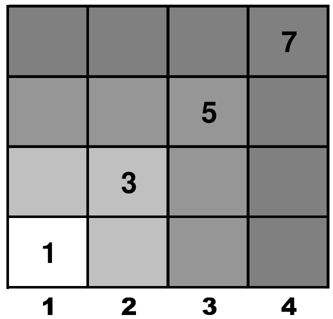
Restando pS de S y operando se obtiene

$$S = \frac{p}{1-p}\left(a_1 + r\frac{p}{1-p}\right). \tag{B.2}$$

B.2 Suma de cuadrados de los n primeros enteros

Una forma sencilla de obtener el valor de $\sum_{k=1}^{n} k^2$, es sumando las áreas correspondientes a los cuadrados de lados $1, 2, 3, \ldots, n$. Si observamos el dibujo comprobaremos que

1. para pasar del cuadrado de lado k al de lado $k+1$ basta con orlar el primero con $O_k = 2k+1$ cuadrados unitarios,
2. al sumar las áreas de todos los cuadrados, la orla $O_k,\ k = 0, 1, \ldots, n-1$, es sumada $n-k$ veces.



Teniendo en cuenta lo anterior,

$$\begin{aligned}
S_n^2 &= 1^2 + 2^2 + \cdots + n^2 \\
&= 1 \times n + 3 \times (n-1) + 5 \times (n-2) + \cdots + (2n-1) \times 1 \\
&= \sum_{i=1}^{n} a_i, \qquad \text{(B.3)}
\end{aligned}$$

donde,

$$\begin{aligned}
a_i &= (n-(i-1))(2i-1) \\
&= 2ni - n - 2i(i-1) + (i-1) \\
&= (2n+3)i - 2i^2 - (n+1).
\end{aligned}$$

Sustituyendo en (B.3)

$$\begin{aligned}
S_n^2 &= \sum_{i=1}^{n} \left\{(2n+3)i - 2i^2 - (n+1)\right\} \\
&= (2n+3)\frac{n(n+1)}{2} - 2S_n^2 - n(n+1) \\
&= n(n+1)\frac{2n+1}{2} - 2S_n^2,
\end{aligned}$$

de donde

$$S_n^2 = \frac{n(n+1)(2n+1)}{6}.$$

Bibliografía

[1] Robert B. Ash. *Basic Probability Theory*. DOVER PUBN INC, 11 de jun. de 2008. 337 págs. ISBN: 0486466280. URL: https://www.ebook.de/de/product/7101948/robert_b_ash_basic_probability_theory.html.

[2] William Feller. *introduction to probability theory and its applications*. eng. Bibliographical footnotes. New York, 1950.

[3] Carlos Sánchez Fernández y Concepción Valdés Castro. *Kolmógorov, el zar de azar*. Nivola Libros y Ediciones, S.L., 11 de mayo de 2003. 192 págs. ISBN: 8495599600. URL: https://www.ebook.de/de/product/8466396/carlos_sanchez_fernandez_concepcion_valdes_castro_kolmogorov_el_zar_de_azar.html.

[4] B.V. Gnedenko. *The Theory of Probability*. Mir Publishers, 1978.

[5] G.R. Grimmett y D.R. Stirzaker. *One Thousand Exercises in Probability*. Oxford University Press, 2001.

[6] C.M. Grinstead y J.L. Snell. *Introduction to Probability*. American Mathematical Society, 2010.

[7] Douglas G. Kelly. *Introduction to Probability*. Pearson, 1993. ISBN: 978-0023631450. URL: https://www.amazon.com/Introduction-Probability-Douglas-G-Kelly/dp/0023631457?SubscriptionId=AKIAIOBINVZYXZQZ2U3A&tag=chimbori05-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=0023631457.

[8] Michael D. Perlman y Michael J. Wichura. «Sharpening Buffon's Needle». En: *The American Statistician* 29.4 (nov. de 1975), pág. 157. DOI: 10.2307/2683484.

[9] J. Pitman. *Probability*. Springer-Verlag, 1993.

[10] Sheldon Ross. *A First Course in Probability (8th Edition)*. Pearson Prentice Hall, 2009. ISBN: 978-0-13-603313-4. URL: https://www.amazon.com/First-Course-Probability-8th/dp/013603313X?SubscriptionId=AKIAIOBINVZYXZQZ2U3A&tag=chimbori05-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=013603313X.

[11] Sheldon M. Ross. *Introduction to Probability Models*. Ninth Edition. Academic Press, 2007.

[12] Deborah Rumsey. *Probability for Dummies*. Wiley, 2006.

[13] David Stirzaker. *Elementary Probability*. CAMBRIDGE UNIV PR, 11 de ago. de 2003. 538 págs. ISBN: 0521833442. URL: https://www.ebook.de/de/product/2988234/david_stirzaker_elementary_probability.html.

[14] G. R. Wood y J. M. Robertson. «Buffon got it straight». En: *Statistics & Probability Letters* 37.4 (mar. de 1998), págs. 415-421. DOI: 10.1016/s0167-7152(97)00145-4.