Comentarios de

Teoría de la Probabilidad

Prólogo

En este documento se incluyen algunos comentarios relacionados con la Teoría de la Probabilidad. El origen de estos comentarios está en mi docencia, como doctorando, en la Universidad Carlos III de Madrid. La mayor parte de estas notas ha sido escrita para los alumnos de la asignaturas de Estadística de la Licenciatura en Documentación y del Grado en Información y Documentación, ambos planes en modalidad semipresencial. La docencia a distancia me ha obligado a escribir gran parte de los comentarios que en su día recogí para los alumnos en varios archivos (diciembre del 2007 y enero del 2008), y que ahora incluyo aquí, revisados y mejorados, por si pudieran ser de utilidad para alguien.

Por su génesis, estos comentarios **no constituyen un material completo** de Teoría de la Probabilidad. Pueden ser, si acaso, **un complemento** a material que sí abarque completamente los puntos que se quieran trabajar. Aquí se recogen sólo algunas –de las muchas posibles– dudas teóricas y prácticas que con más frecuencia tienen los alumnos a la hora de hacer este tipo de estadística. En general, el carácter de las dudas que los alumnos puedan tener depende tanto de las características de los materiales de la asignatura como de los alumnos (carácter, intereses, bagaje en ciencias). Las hay triviales y no tan triviales, pero ninguna «tonta».

La Teoría de la Probabilidad es una rama de las Matemáticas que debe su formalismo (de axiomas y conjuntos) principalmente al matemático ruso Kolmogórov. En este documento se intenta explicar de forma intuitiva parte de ese formalismo, que ayuda a aclarar la diferencia entre conceptos como *experimento* y *variable aleatoria*, *datos* y *modelos* o *variable aleatoria* y *distribución de probabilidad*. Este material es prácticamente independiente de la Estadística, que en general hace más uso de la Probabilidad que viceversa; sin embargo, para no repetir algunas explicaciones, se hará referencia a otros archivos.

Índice

Ejercicios

Más ejercicios

«Variaciones sobre un ejercicio de Teoría de la Probabilidad»

http://www.Casado-D.org/edu/VariacionesEjercicioProbabilidad.pdf

Variable aleatoria: primera aproximación

En esta sección se introducen, sin rigurosidad, los conceptos de la Teoría de la Probabilidad básica. Se ha preferido dejar para la siguiente sección una visión más completa y rigurosa, de lectura prescindible para resolver algunos ejercicios sencillos. El último apartado de esta sección, Experimentos y variables aleatorias, pretende preparar al lector para la lectura de la segunda aproximación a las variables aleatorias.

Variables aleatorias y variables estadísticas

Una variable estadística toma ciertos valores con ciertas frecuencias. Más adelante se da una definición más rigurosa de *variable aleatoria*, pero desde el punto de vista del observador una variable aleatoria es una variable, a veces determinada a partir de un *experimento*, que puede tomar ciertos valores con ciertas probabilidades. La regla por la que estas probabilidades se asignan a los valores se denomina *distribución de probabilidad*.

La interpretación de las frecuencias como probabilidades, aunque sea asintóticamente, permite suponer que las variables aleatorias son los modelos teóricos que generan las variables estadísticas. Léase lo escrito sobre variables estadísticas y variables aleatorias en

«Comentarios de Estadística descriptiva con una variable y dos variables»

http://www.Casado-D.org/edu/ComentariosEstadisticaDescriptiva.pdf

Aunque ahora sólo se puedan comprender en parte, terminamos este pequeño apartado con unas definiciones del Vocabulario científico y técnico, de la Real Academia de Ciencias Exactas, Física y Naturales,

> **variable estadística**. Función real definida sobre una población finita o una muestra, que toma los valores de cada una de las modalidades de un atributo, y a la que hay asociada una distribución de frecuencias. Generalmente se toman como espacio muestral de la variable estadística los posibles valores del atributo. Si en vez de considerar solamente un atributo, se consideran varios, se tiene una variable estadística multidimensional [sería mejor utilizar «multivariante»].
>
> **variable aleatoria**. Cualquier función definida sobre el espacio muestral con valores reales y que verifique la condición siguiente: la contraimagen de cualquier conjunto boreliano es un suceso de la σ--álgebra del espacio muestral. Así, una variable aleatoria establece una correspondencia entre los sucesos elementales y los números reales, midiendo numéricamente alguna característica del resultado del fenómeno aleatorio.

Variables aleatorias discretas y continuas

Respecto a los posibles valores que pueden tomar las variables aleatorias, que determinan si son *discretas* o *continuas*, léase también lo escrito para las variables estadísticas en

«Comentarios de Estadística descriptiva con una variable y dos variables»

http://www.Casado-D.org/edu/ComentariosEstadisticaDescriptiva.pdf

Tiempo: Variables simultáneas y consecutivas

Es conveniente prestar una atención especial al parámetro *tiempo*, dada su importancia en las situaciones en que aparecen varias variables aleatorias. Un conjunto de variables aleatorias, como sucedía con las variables estadísticas, puede estar formado por variables que se consideran *simultánea* o *transversalmente*, en un único instante de tiempo, o, a diferencia, por variables ordenadas *en el tiempo* (si las variables «se parecen mucho», se puede pensar equivalentemente en «una variable aleatoria que cambia con el tiempo»). En este sentido, siguen siendo aplicables conceptualmente las ideas que se mencionan para las variables estadísticas en el documento

«Comentarios de Estadística descriptiva con una variable y dos variables»

En la anterior estadística descriptiva sólo se consideraba el caso de variables estadísticas simultáneas. Ahora, para las variables aleatorias, sí se considerarán con frecuencia tanto situaciones de variables simultáneas como de unas pocas variables que suceden consecutiva o encadenadamente (cuando las variables son muchas —teóricamente infinitas— y están definidas sobre el mismo experimento se consideran los *procesos estocásticos*, que son a las *series temporales* —variables estadísticas a lo largo del tiempo— lo que las variables aleatorias son a las variables estadísticas: sus modelos teóricos).

En el caso de variables simultáneas el subíndice sirve para diferenciar a unas variables de otras, mientras que en el caso de variables en el tiempo denota también una posición en el tiempo (los subíndices en general podrían indicar una posición en el espacio, pero eso no se verá en este documento). Habrá que tener en cuenta la diferencia entre estas dos situaciones al hacer los cálculos.

Misma teoría

Las herramientas teóricas para las dos situaciones antes mencionadas son las mismas, porque la propia Teoría de la Probabilidad así lo hace al considerar estructuras generales: un espacio de posibles resultados o sucesos, una asignación de probabilidad a cada resultado, unos axiomas, unas fórmulas, etcétera. Y aquí, dentro de «espacio de posibles sucesos» caben tanto «sucesos de variables simultáneas» como «sucesos de variables consecutivas». Todo esto se comprende mejor con algunos ejemplos que se dan en los siguientes apartados, pero antes se quiere insistir en que toda la teoría que se verá más adelante (definiciones, axiomas, fórmulas, teoremas, gráficos, etcétera) es válida para ambas situaciones, la diferencia la tenemos que tener en cuenta nosotros al definir bien y con cuidado los conceptos de cada ejercicio concreto.

Dependencia entre las variables

Más adelante se hablará de la definición de *independencia* de sucesos, ahora sólo se van a mencionar de forma intuitiva estos conceptos para aclarar su relación con la simultaneidad de las variables. Dos variables aleatorias son *independientes entre sí* si la forma en que cada una toma valores no depende de la forma en que lo haga la otra, es decir, si sus distribuciones de probabilidad son independientes. Según el factor en que se produce, algunas tipos de dependencia son:

a) Dependencia en el tiempo: Una variable puede depender de otra anterior en el tiempo. Éste es el tipo de dependencia que va a aparecer con más frecuencia.
b) Dependencia en el espacio: Una variable puede depender de otra mediante alguna relación espacial, por ejemplo, que esté situada a la derecha.

Tiempo y dependencia

En apartados anteriores se ha hablado de dos condiciones diferentes: el tiempo en que se consideran las variables y la dependencia entre ellas. El tiempo es una característica «exterior» a la teoría en el sentido de que está relacionado con la forma en que se organizan u ordenan las variables aleatorias, no con la propia teoría de definiciones, axiomas, etcétera. La dependencia, sin embargo, es una definición –condición– entre los sucesos y las variables aleatorias de toda la estructura probabilística que se mencionará más adelante. Se quiere resaltar con lo anterior que no hay ninguna relación general de implicación entre los conceptos de tiempo y dependencia, por lo que pueden darse todas las posibles situaciones que surgen de su combinación. Lo vemos con los siguientes ejemplos.

A) Simultaneidad e independencia: Se lanzan dos dados a la vez y se suman sus resultados.

B) Simultaneidad y dependencia: Se lanzan dos dados a la vez, uno azul y otro rojo. Se hace la siguiente suma: el dado azul aporta a la suma con su propio resultado, mientras que el dado rojo aporta con su resultado multiplicado por el resultado del dado azul.

C) Encadenamiento e independencia: Se lanzan dos dados, uno tras otro, y se suman los resultados.

D) Encadenamiento y dependencia: Se lanza un dado azul y después otro rojo. Se hace la siguiente suma: el dado azul aporta a la suma con su propio resultado, mientras que el dado rojo aporta con su resultado multiplicado por el resultado del dado azul.

En los casos A y C los dados son independientes, mientras que en los casos B y C el dado azul es independiente del rojo pero el rojo depende del azul (en realidad lo que depende en este caso no es el propio resultado del dado sino las consecuencias que este resultado tiene en el experimento general).

Experimentos y variables aleatorias

A veces una variable aleatoria viene determinada o definida a partir de un *experimento*. El conjunto de todos los posibles resultados del experimento se denomina *espacio muestral*. La asignación de una probabilidad a cada uno de los elementos del espacio muestral se conoce como *distribución de probabilidad*.

Los sucesos del espacio muestral pueden describirse equivalentemente con notación numérica en vez de literal. Cuando se hace esto, en el fondo se obtienen un nuevo espacio muestral y una nueva distribución de probabilidad, que no dependen del experimento concreto: pueden asignarse a varios experimentos distintos. Es interesante mencionar que los resultados obtenidos a partir de este nuevo «entorno numérico» son entonces aplicables a todos esos experimentos. Las *variables aleatorias* se encargan de hacer esta traducción del espacio muestral (del experimento) a números. La nueva distribución de probabilidad se expresa numéricamente mediante la *función de probabilidad*, que se llama *función de masa* para las variables discretas y *función de densidad* para las variables continuas.

Se puede pensar que sólo sabemos hacer cálculos con números u objetos definidos sobre ellos (matrices, funciones, funcionales, etcétera). Sucede lo mismo con los ordenadores, que aunque nos muestran letras, en realidad funcionan con números, concretamente sólo con ceros y unos, por lo que es necesaria una interfaz que haga de traductora entre letras y números. Se ha hablado de experimentos aleatorios, cuyos resultados pueden ser «cara», «un número par» o «una terna de números». Entonces, matemáticamente hablando, las variables aleatorias son funciones que nos «codifican» los resultados del espacio muestral para convertirlos en números, que es lo que manejamos cómodamente:

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

Es decir, la variable aleatoria X asigna un suceso en los números reales a cada suceso del espacio muestral. De esta manera, después es fácil trabajar con los números y decir cuál es la media o la varianza, por ejemplo.

La variable aleatoria no trivial más sencilla posible, la variable *de Bernoulli*, asigna un 0 a «cara» y un 1 a «cruz». Por tanto, si antes se podía decir

$$\text{«cara»} \quad \text{y} \quad \text{«cruz»} \qquad \text{(Viejo espacio muestral)}$$

$$P(\text{«cara»}) = p \quad \text{y} \quad P(\text{«cruz»}) = 1 - p \qquad \text{(Vieja distribución de probabilidad)}$$

ahora

$$X(\text{«cara»}) = 0 \quad \text{y} \quad X(\text{«cruz»}) = 1 \qquad \text{(Nuevo espacio muestral)}$$

$$P_X(X = 0) = p \quad \text{y} \quad P_X(X = 1) = 1 - p \qquad \text{(Nueva distribución de probabilidad)}$$

Y esto último nos gusta más. Nótese que esta última línea no depende del experimento; esto implica que todo lo que deduzcamos a partir de ella será válido y aplicable a todos los posibles experimentos que puedan dar lugar a una variable aleatoria de Bernoulli. En otras palabras, convertir a números nos permite obtener resultados generales, independientes de los experimentos concretos, para luego, en un último paso, «descodificar» e interpretar qué significan esos resultados generales en el caso de experimentos concretos.

$$\Omega \qquad \rightarrow X \rightarrow \qquad \mathbb{R}$$

$$\{\text{cara}\} \qquad \rightarrow \qquad \{X = 0\}$$

$$\{\text{cruz}\} \qquad \rightarrow \qquad \{X = 1\}$$

Por tanto, la diferencia entre *experimento* y *variable aleatoria* es que el primero es el mecanismo por el que se obtiene un elemento del espacio muestral, mientras que la segunda representa numéricamente esa información. Como vemos, son conceptos cercanos, de manera que hay ocasiones en que, abusando del lenguaje y cuando no hay lugar a confusión, se habla indistintamente de uno u otra como si fuesen equivalentes.

Por otro lado, la diferencia entre *variable aleatoria* y *distribución de probabilidad* es que la primera es una función mientras que la segunda es una regla que atribuye probabilidades a los sucesos. Hay una distribución de probabilidad en el espacio de probabilidad del experimento, pero la variable aleatoria crea una copia de esta distribución de probabilidad en el espacio de probabilidad de llegada, que es la que más propiamente se llama *distribución de probabilidad*.

Tipos de ejercicios y problemas

Todo lo mencionado hasta ahora sobre la estructura de experimento, espacio muestral, distribución de probabilidad, variable aleatoria, etcétera, hace que haya ocasiones (ejercicios y problemas) en las que podemos oír hablar de «sucesos del espacio muestral del experimento» o, directamente, de «distribuciones de probabilidad» o «variables aleatorias que siguen cierta distribución de probabilidad». En estos dos últimos casos, no hay interés en ningún experimento –real o teórico– concreto, sino sólo en las variables aleatorias o en las distribuciones de probabilidad. Esto da lugar a los distintos tipos de ejercicios, cuya solución se encuentra más fácilmente con todas estas ideas claras. Por ejemplo, no es lo mismo tener que deducir la distribución de probabilidad a partir de un experimento, que, ya dada la función de probabilidad, utilizarla para calcular la probabilidad de un determinado suceso.

Tiempo

Para los experimentos, como parte asociada y subyacente a las variables aleatorias, también puede hablarse de simultaneidad u ordenamiento en el tiempo, como se hizo para las variables aleatorias.

Restitución

Aparte del tiempo, otro factor muy presente en los experimentos, cuando hay varios o uno que se compone de varias etapas (ambas situaciones son equivalentes desde el punto de vista teórico), es la presencia o no de *restitución*. Dentro del experimento, este factor va a restituir o no después de cada *acción* la situación a las condiciones iniciales. Cuando la acción es una *extracción*, la restitución es un *reemplazo* de lo que se había extraído. En resumen, en ejercicios «con paso del tiempo» suele ser necesario prestarle atención a la restitución. (Normalmente se habla casi siempre de «reemplazo», se extraiga o no algo durante el experimento; pero he preferido utilizar «restitución», que es más general.)

Variable aleatoria: segunda aproximación

En esta sección se profundiza en la estructura teórica de la Teoría de la Probabilidad, por lo que para leer esta segunda aproximación es conveniente leer antes la primera.

Experimento y espacio de probabilidad

Según el Vocabulario científico y técnico, de la Real Academia de Ciencias Exactas, Física y Naturales, un *experimento aleatorio* se define

> **experimento aleatorio**. Experimento tal que: a) se puede repetir indefinidamente en análogas condiciones; b) en cada prueba se obtiene un resultado que pertenece al conjunto de resultados posibles del experimento; c) antes de realizar una nueva prueba del experimento, no se puede predecir el resultado que se obtendrá. || Cualquier situación en la cual se puede enumerar de antemano los resultados posibles sin que se conozca cuál de ellos se va a presentar.

Por abstractas que a veces se hagan, el origen de las Matemáticas se sitúa en la realidad. En el caso de la Teoría de la Probabilidad, su origen está en los juegos de azar. Como instrumento general, esta teoría intenta dar solución a los problemas planteados a partir de posibles experimentos aleatorios. Para ello define conceptos cada vez más abstractos, como las variables aleatorias o las distribuciones de probabilidad, que llegan a tener «vida propia», es decir, se puede hablar de estos conceptos sin mencionar la palabra «experimento»; o, para quien lo prefiera, se puede decir que hay detrás un experimento teórico y virtual que es ya igual al experimento de los posibles valores numéricos que la variable puede tomar con ciertas probabilidades (esto equivale a que la variable aleatoria, como función del espacio muestral del experimento en el conjunto de los números reales, es la función identidad).

Los resultados del experimento son los *sucesos*, que juntos forman el *espacio muestral*:

suceso. Cualquier subconjunto del espacio muestral asociado a un experimento aleatorio. Cuando el espacio muestral está dotado de una σ-álgebra (o de un álgebra), se suele llamar sucesos únicamente a los subconjuntos del espacio muestral que son elementos de la σ-álgebra (o del álgebra).

espacio muestral. Conjunto de posibles resultados de un experimento aleatorio; p. ej., asociado al experimento aleatorio de lanzar un dado, el espacio muestral es el constituido por las seis distintas caras del dado.

Pero para que estén incluidas todas las uniones e intersecciones de los posibles sucesos, es necesario ampliar la estructura del espacio muestral a un *álgebra* o una σ-*álgebra* (esta segunda estructura es más general que la primera, porque permite que las uniones de sucesos sean infinitas numerables, en vez de sólo finitas). Esta ampliación se hace considerando los elementos del espacio muestral y todas las posibles uniones e intersecciones de los elementos del espacio muestral, aparte de los complementarios de todos los sucesos. Se da la definición de la primera:

álgebra de sucesos. Familia de subconjuntos del espacio muestral que verifique las siguientes condiciones: a) el conjunto vacío pertenece a la familia; b) la unión de dos elementos cualesquiera de la familia pertenece a ella; c) el complementario de cualquier elemento de la familia pertenece también a la familia. Se deduce de las propiedades anteriores que la intersección de dos elementos cualesquiera de un álgebra es también un elemento del álgebra.

La función que asigna probabilidades a los sucesos de esta estructura ampliada es la *probabilidad*:

probabilidad. Cualquier función P con valores reales, definida para los elementos de una σ-álgebra A sobre un espacio muestral $\Omega,$ que cumple:

a) $P(\Omega)=1$

b) $P(a)\geq 0 \quad \forall\ a\in A$

c) $P(U_{n=1}^{\infty} a_n)=\sum_{n=1}^{\infty} P(a_n)$

para cualquier sucesión $\{a_n\}$ de sucesos disjuntos dos a dos de la σ-álgebra. $P(a)$ se interpreta como la frecuencia relativa con la que se presenta el suceso a en una sucesión de repeticiones del experimento aleatorio al que Ω está asociada. Existen diversas teorías que tratan de dar una normativa para construir la probabilidad asociada a un fenómeno aleatorio; principalmente, se distingue entre teorías objetivistas, que tratan de asignar la probabilidad por consideraciones físicas o por experimentación, y teorías subjetivistas, que utilizan los grados de creencia subjetiva de los individuos.

Ya estamos preparados para definir la estructura completa de *espacio de probabilidad*:

espacio probabilístico. Abstracción matemática de un fenómeno aleatorio, a saber, una terna (Ω, A, P) constituida por un espacio muestral $\Omega,$ una σ-álgebra A y una probabilidad P definida sobre los sucesos de la σ-álgebra. El espacio muestral contiene los posibles resultados, la σ-álgebra contiene los sucesos de interés para los cuales está definida la probabilidad.

En resumen, cuando tenemos un experimento con una probabilidad asignada a cada uno de los posibles resultados, aparece de forma natural la estructura de espacio de probabilidad.

Variable aleatoria y duplicación del espacio de probabilidad

Una *variable aleatoria* es, matemáticamente, una función que asigna sucesos numéricos a los sucesos del espacio muestral del experimento. Como la σ-álgebra se genera a partir de los *sucesos elementales* del espacio muestral, queda así definida la variable aleatoria para todos los sucesos de la σ-álgebra, puesto que los *sucesos compuestos* se pueden descomponer en elementales.

Ahora,

$$(\Omega, A, P) \quad \rightarrow \; X \rightarrow \quad (\mathbb{R}, B, P_X)$$

donde (Ω, A, P), es un *espacio de probabilidad* (*espacio de medida* donde la medida es una probabilidad) con:

- Ω ≡ *Espacio muestral del experimento*
- A ≡ *σ-álgebra de sucesos asociados al experimento*
- P ≡ *Probabilidad*

y, dado que la variable aleatoria *X* es por definición una *función medible*, se crea –*induce*– en $\mathbb{R}$ una copia «exacta» de toda esta estructura.

$(\mathbb{R}, B, P_X)$, otro *espacio de probabilidad*, donde:

- $\mathbb{R}$ ≡ *Conjunto de los números reales*
- B ≡ *σ-álgebra de Borel en* $\mathbb{R}$ $(\{X(a) / a \in A\})$
- P_X ≡ *Probabilidad inducida por X* $\left(P_X(b) := P(X^{-1}(b))\right)$

Es decir, las variables aleatorias duplican toda la estructura. Desde el punto de vista teórico es importante tener en mente este hecho, puesto que dado que las definiciones, propiedades y resultados de la Probabilidad son válidos para toda estructura de espacio probabilístico, lo serán tanto en Ω como en $\mathbb{R}$. Es decir, la definición de *independencia de sucesos* o el *Teorema de la probabilidad total*, por ejemplo, tiene una «copia» en cada uno de los dos espacios de medida. De las definiciones, propiedades, resultados, etcétera, de los espacios de probabilidad se habla en la siguiente sección de este documento. Ahora se insiste una vez más en que lo que allí se diga será válido en los dos espacios de probabilidad que relaciona la variable *X*.

Por otro lado, dado que una variable aleatoria es una transformación de unos sucesos en otros, no es de extrañar que al hablar de algunos conceptos, como independencia o simultaneidad, por ejemplo, se tengan definiciones análogas para sucesos y variables aleatorias, o que a veces se pueda pensar equivalentemente en unos o en otras.

Conviene también hacer hincapié en la diferencia entre la variable aleatoria *X* y su *distribución de probabilidad* P_X. En general, en vez de utilizar la definición conjuntista dada antes para P_X, las distribuciones de probabilidad usuales ya disponen de una expresión que proporciona los valores de las probabilidades. Esta función que expresa la distribución de probabilidad se llama en general *función de probabilidad*, pero en el caso de variables aleatorias discretas se llama *función de masa* y en el caso de las continuas se llama *función de densidad*. Veamos algunos ejemplos de funciones de masa (en ellos a la x sólo se le pueden dar valores de un conjunto discreto):

- Distribución de Bernoulli de parámetro p: $f(x)=p^{1-x}\cdot(1-p)^x \quad \rightarrow \quad f(0)=p$ y $f(1)=1-p$
- Distribución binomial de parámetros n y p: $f(x)=\binom{n}{x}p^{n-x}\cdot(1-p)^x$

y un ejemplo de función de densidad (ahora la x puede tomar valores en un conjunto continuo)

- Distribución normal de parámetros μ y σ: $f(x)=\frac{1}{\sigma\cdot\sqrt{(2\pi)}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

Conocer esta estructura de espacio de probabilidad, fundamental en Teoría de la Probabilidad, es conveniente no sólo para comprender mejor los conceptos sino para resolver ejercicios y problemas, como veremos más adelante.

No unicidad entre experimentos y distribuciones

A experimentos distintos les puede corresponder una misma distribución de probabilidad; todos somos capaces de imaginarnos distintos experimentos a los que les correspondería de forma natural una variable de Bernoulli: cara y cruz, rojo y negro (en la ruleta), etcétera.

Por otra parte, un mismo experimento se puede estudiar considerando variables y distribuciones distintas (aunque frecuentemente haya una candidata natural clara). Por ejemplo, el experimento de lanzar una moneda cuatro veces se puede estudiar utilizando un proceso de cuatro variables aleatorias de Bernoulli o, a diferencia, utilizando una sola variable aleatoria binomial.

Lo que se quiere representar en ambos casos es el hecho de que, en el camino hacia la solución de un ejercicio, hay veces en que se pueden utilizar distribuciones distintas; pero en ambos casos la asignación final de probabilidades a los sucesos es única y debe coincidir. Se ve mejor con la notación de los espacios de probabilidad:

- Dado un experimento, cada suceso de su espacio muestral y de la σ-álgebra tiene asociada una probabilidad, y esta asociación **es única**; es decir, para (Ω, A, P) la probabilidad P sí es única dado Ω, porque cada suceso tiene asignado un único número que indica su probabilidad.
- Dada una variable aleatoria, la distribución de probabilidad que sigue también es única; es decir, para $(\mathbb{R}, B, P_X)$ la distribución de probabilidad P_X sí **es única** dada X.
- **Una situación interesante**: Lo que se quiere decir con lo anterior es que hay ejercicios en los que, a la hora de calcular no toda P sino <u>la probabilidad concreta de un suceso</u>, existe la posibilidad de utilizar experimentos distintos y variables distintas (por tanto, puede que distribuciones distintas), para llegar a la misma solución. Véase el comentario de la solución del *tema original* en «Variaciones sobre un ejercicio de Teoría de la Probabilidad» (el enlace está en el índice).

Variables consecutivas

Como se ha mencionado en Tiempo: Variables simultáneas y consecutivas, toda la teoría relacionada con los espacios de probabilidad es suficientemente general como para incluir el caso de varias variables consecutivas en el tiempo: debemos ser nosotros quienes, con nuestras definiciones, tengamos en cuenta este factor temporal.

Una primera observación es que hablar de «variables aleatorias consecutivas» es equivalente a hablar de «experimentos aleatorios consecutivos», como se representa en la siguiente tabla.

Experimentos	E_1	E_2	$\cdots$	E_n
Sucesos	Ω_1	Ω_2	$\cdots$	Ω_n
Variables aleatorias	X_1	X_2	$\cdots$	X_n

Equivalencia con un proceso estocástico

Por otro lado, nosotros también podemos considerar equivalente hablar de «variables aleatorias consecutivas» y de «proceso estocástico en el tiempo». En rigor para hablar de proceso las variables deben estar definidas sobre el mismo espacio muestral, es decir, sus experimentos deben ser el mismo repetido consecutivamente: esto es lo que permite hacer mucha de la teoría de los procesos estocásticos, basada en la *distribución finitodimensional,* y no es posible hacer teoría general sobre sucesiones de variables aleatorias distintas. Además, en el caso de los procesos el interés suele estar en que se componga de infinitas variables, con el interés en lo que suceda en este límite, mientras que a nosotros nos interesan ahora sólo las primeras variables. Lanzar una moneda y después un dado, incluso con alguna dependencia de ella, puede dar lugar a una sucesión de dos variables aleatorias, pero no suele considerarse como un proceso estocástico (aparte de la poca longitud). Volviendo a la equivalencia, los experimentos pueden verse como etapas consecutivas o encadenadas que forman parte de un experimento compuesto mayor:

Experimentos	E_1	E_2	$\cdots$	E_n
Suceso		$\boldsymbol{\Omega}$		
Proceso estocástico		$\boldsymbol{X}$		

donde $\boldsymbol{\Omega} \in \Omega_1 x \Omega_2 x \cdots \Omega_n = \{(S_1, S_2, \cdots, S_n) / S_i \in \Omega_i\}$ y $\boldsymbol{X} = (X_{1,} X_{2,} \cdots, X_n)$. Aparecen en la teoría conceptos, de los que no se va a hablar ahora, como *conjunto cilíndrico* y *trayectoria*.

Por ejemplo, en el caso de tres variables, se pueden considerar tres variables aleatorias distintas, lo que hace que haya tres espacios muestrales distintos, o considerar un proceso, lo que daría lugar a un solo espacio muestral compuesto por todas las posibles ternas (compatibles) que se pueden formar tomando, en orden, un suceso de cada uno de los tres espacios muestrales distintos antes mencionados. Se puede imaginar un proceso estocástico, además de la interpretación principal como sucesión infinita de variables aleatorias, como «variable compleja cuyos posibles resultados se componen de varias (infinitas) etapas».

Equivalencia con el caso simultáneo

Desde el punto de vista teórico, para el caso de variables consecutivas suele ser posible encontrar un caso equivalente de variables simultáneas (aunque a veces pueda ser muy complicado, porque cada variable debería reflejar toda la estructura de posibles resultados de las anteriores en el tiempo); es decir, suele ser posible imaginarse un conjunto de variables simultáneas que definen una situación equivalente. Lo vemos con un ejemplo sencillo; para las variables

X_1 == Resultado de lanzar un dado

X_2 == Resultado de lanzar otra vez el dado y multiplicar su resultado por el resultado anterior

se pueden definir dos variables simultáneas

Y_1 == Resultado del dado 1

Y_2 == Resultado del dado 2 multiplicado por el resultado del dado 1

donde se lanzan a la vez dos dados que se pueden distinguir por el color o alguna otra marca.

En este ejemplo anterior se ha transformado la «dependencia en el tiempo» en «dependencia en la marca», pero la manera de resolver el ejercicio sería la misma. Y aquí no se obtiene mucha ventaja al aplicar la equivalencia, pero en otros ejemplos la ventaja puede ser que se puedan contar los casos favorables y los posibles con facilidad para después aplicar la regla de Laplace: por ejemplo, si se están extrayendo sin reemplazo bolas de una bolsa, ¿qué más da extraerlas de una en una que todas a la vez?

Un poco más...

En esta sección se incluyen algunos comentarios de un nivel más alto, que pretenden por un lado facilitar la comprensión de lo visto y de los temas siguientes, y por otro satisfacer la curiosidad de quien quiera ir un poco más allá de lo estrictamente necesario según el temario.

Procesos aleatorios o estocásticos

Hemos hablado hasta ahora de experimentos simultáneos o «sin etapas» y experimentos «por etapas». Para cada etapa se puede definir una variable aleatoria, que nos dirá en ese paso lo que puede suceder y con qué probabilidades. A todo este conjunto de variables, cuando están definidas sobre el mismo espacio de probabilidad, se le llama *proceso aleatorio* o *estocástico* (en Variables consecutivas se ha hablado de la diferencia conceptual entre una sucesión de variables aleatorias y un proceso estocástico). Sus variables pueden ser discretas o continuas. Es decir, una cosa es el índice, que puede ser un número de un conjunto discreto o continuo de valores, y otra es cada variable aleatoria, que puede ser discreta o continua.

Clasificación según el tiempo

En el caso concreto en que el índice es discreto, es decir, 1, 2, 3,... el proceso se dice que está *en tiempo discreto*, y se compone de una sucesión de variables aleatorias. Cuando las variables aleatorias son también discretas, se le suele hablar de *cadena*. Cuando el índice es continuo, se tiene una familia de variables aleatorias, no una sucesión; se le llama *proceso en tiempo continuo*.

Clasificación según la dependencia

Los procesos se pueden clasificar según el grado de dependencia entre sus variables:

- Proceso de variables aleatorias independientes, si para cada índice la variable es independiente de las pasadas (no puede depender de las futuras, que todavía no han sucedido).
- Proceso de Markov, cuando una variable depende sólo de la situación anterior y no de ninguna otra pasada, es decir, es independiente del camino que se ha seguido para llegar a esa situación anterior.
- Proceso de memoria corta, si cada variable depende de un número relativamente pequeño de variables anteriores.
- Proceso de memoria larga, si el número de las que depende es grande.

Series temporales

Una *serie temporal* es una realización de un proceso estocástico. Con esto también se está diciendo implícitamente que al igual que las variables estadísticas se modelizan con variables aleatorias, las series temporales se modelizan con procesos estocásticos. Es decir, un proceso se compone de variables aleatorias, mientras que la serie temporal se compone de los valores núméricos que han tomado esas variables aleatorias después de realizado el experimento. Un proceso está en la «parte de la teoría», mientras que una serie está en la «parte de los números o datos».

Vectores aleatorios

Si las variables aleatorias consecutivas pueden formar procesos estocásticos, varias variables simultáneas forman un *vector aleatorio*. Los vectores aleatorios son variables aleatorias multivariantes (este documento trata de las univariantes, tanto simultáneas como de unas pocas consecutivas), que podrían también formar *procesos estocásticos multivariantes*, si se considerasen varios vectores aleatorios consecutivos en el tiempo.

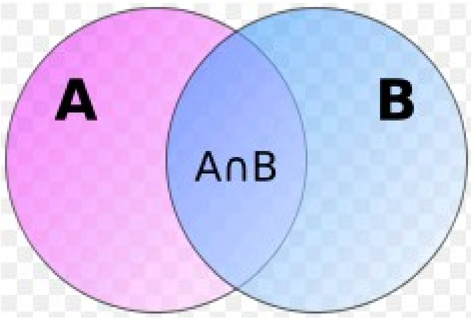
Por otro lado, de la misma manera que a las variables estadísticas univariantes se les podían asignar, como modelo, las variables aleatorias univariantes, a las variables estadísticas multivariantes se les puede asignar una *vector aleatorio multivariante*.

Espacios de probabilidad

Diagrama de Venn

En Matemáticas (más concretamente, en Teoría de la Medida) hay un tipo de funciones que se llaman *medidas*, que serían el equivalente a lo que entendemos en la vida diaria como medida: longitud, área, volumen. La *probabilidad* es una de estas medidas matemáticas. Por tanto, pensar en las medidas usuales de longitud, área y volumen puede ayudar a comprender o recordar las propiedades de la probabilidad; siempre teniendo en cuenta que la causa de que la probabilidad comparta características con esas medidas usuales se debe a que el concepto de medida matemática intenta «imitar» las medidas usuales y no se pueden trasladar propiedades entre ellas y la probabilidad. Esto significa que pensar en las medidas usuales es una ayuda o un truco mnemotécnico, pero nunca una demostración.

Para hacer dibujos en el plano (hoja de papel), el área es la que resulta más útil. Los diagramas de la Teoría de Conjuntos (aplicables a la Teoría de la Probabilidad) más usuales son los *diagramas de Venn*:



http://es.wikipedia.org/wiki/Diagrama_de_Venn

Estos diagramas son especialmente útiles para representar sucesos simultáneos, aunque si uno tiene claro lo que está representando también podrían aplicarse para representar sucesos consecutivos o encadenados: basta imaginar que los círculos van apareciendo consecutivamente.

Diagrama de árbol

Es importante recalcar que, aunque este tipo de gráfico es aplicable al caso de sucesos simultáneos (para representar todos los sucesos del espacio muestral sin dejarse ninguno), es especialmente útil para hacer un esquema de los sucesos en experimentos que tienen varias etapas (se lanza un dado después de otro, se sacan tres bolas de una en una, etcétera). En las flechas del árbol solemos ir indicando la probabilidad de cada paso.

- Debe valer 1 la suma de las probabilidades de todas las flechas que salen de cada nodo concreto.
- Por otro lado, esa probabilidad que expresamos de pasar de un nodo a otro es una probabilidad condicionada a que hayamos llegado hasta el nodo de salida. Esto, que puede parecer una perogrullada, explica por qué para obtener la probabilidad de que vayan pasando los sucesos de un «ruta», «camino» o «rama» concreto vamos multiplicando todas esas probabilidades.

Regla de Laplace

Esta regla es correcta cuando todos los sucesos elementales del espacio muestral tienen la misma probabilidad de ocurrir. Cuando el suceso A que nos interesa está compuesto de varios sucesos elementales, es decir, se cumple A cuando se cumpla alguno de esos sucesos elementales (serán los *casos favorables*),

$$P(A)=\frac{CF}{CP}$$

donde CF denota el número de sucesos elementales favorables y CP el de todos los posibles sucesos del espacio muestral (*casos posibles*). Tenemos entonces que esta regla será aplicable sólo cuando podamos contar fácilmente el número de casos favorables y posibles.

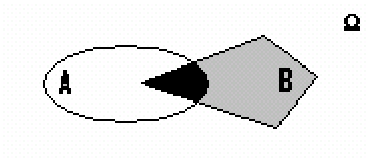
Como aplicación, véanse los ejercicios Ejercicio 3, Ejercicio 7, Ejercicio 8 y los ejercicios finales del documento «Variaciones sobre un ejercicio de Teoría de la Probabilidad» (el enlace está en el índice). En el Ejercicio 3 se puede encontrar una demostración de la regla de Laplace.

Probabilidad condicionada

La fórmula de la probabilidad condicionada de un suceso A a otro B es:

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

La interpretación de esta fórmula es que mide, cuando se cumple B, la probabilidad de que además se cumpla A, es decir, la probabilidad de la intersección; pero normalizando por la probabilidad de B, porque se quiere ver la proporción que esa intersección supone dentro del suceso B. Con el siguiente dibujo se entiende bien la fórmula y la explicación anteriores.



Una cosa que hay que tener en cuenta de esta fórmula es que se puede escribir en la forma:

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

y ésta será la forma en que se utilice en muchos ejercicios. Por ejemplo, en los ejercicios en que hay dos etapas: en la primera se puede cumplir o no B, y en la segunda se puede cumplir o no A. La fórmula en esta forma nos dice entonces la probabilidad de que primero se cumpla B y después se cumpla A.

La fórmula de la probabilidad condicionada relaciona dos sucesos, y es aplicable tanto cuando son simultáneos (el experimento tiene una sola etapa) como cuando no (el experimento consta de dos etapas).

Para su aplicación, véanse los ejercicios Ejercicio 4 y Ejercicio 7.

Independencia de sucesos

Dos sucesos

Hay dos definiciones equivalentes de *independencia entre dos sucesos A* y *B*:

- $P(A\cap B)=P(A)\cdot P(B)$
- $P(A|B)=P(A)$ y $P(B|A)=P(B)$

Esto se interpreta, respectivamente, como que a *A* «le da igual» lo que valga *B* y a *B* «le da igual» lo que valga *A*. Es fácil ver la equivalencia entre las dos definiciones utilizando la fórmula de la probabilidad condicionada y que $P(A\cap B)=P(B\cap A)$. También conviene mencionar que si *A* y *B* son independientes, también lo son los pares de sucesos *A* y B^c, A^c y *B*, y A^c y B^c; además, a partir de las condiciones de arriba se puede demostrar fácilmente que cambiando alguno de los conjuntos por su complementario, también siguen siendo ciertas.

Más de dos sucesos

En una colección de más de dos sucesos, puede suceder que sean independientes dos a dos pero no entre todos. Por tanto, es necesaria la siguiente definición de *independencia de* n *sucesos*:

- $P(A_1\cap A_2\cap\cdots\cap A_n)=P(A_1)\cdot P(A_2)\cdots P(A_n)$

Como aplicación, véase el ejercicio Ejercicio 5.

Otras definiciones

Hay definiciones de *independencia* para otros muchos conceptos. Se puede hablar de la *independencia de (más de dos) sucesos dos a dos*, que no exige tanto como la expresión dada antes. Es necesario hablar de la *independencia de dos variables aleatorias* que, razonablemente, se define como la condición de que los valores que tome una no estén influidos por los que toma la otra (se expresa esta condición a través de la función de probabilidad). También existen la *independencia de* n *variables aleatorias* y la *independencia de* n *variables aleatorias dos a dos*. Por último, también existen definiciones de independencia para vectores aleatorios y procesos estocásticos.

Vuelta al diagrama de árbol

Conviene volver a hablar de este tipo de gráfico, que ya se presentó en un apartado anterior, ahora con los conceptos de la probabilidad condicionada. En Probabilidad condicionada se presentó la fórmula

$$P(A\cap B)=P(B)\cdot P(A|B) \qquad \text{(a)}$$

y se habló de su aplicación a experimentos que se puedan dividir en dos etapas (aunque es aplicable a dos sucesos simultáneos). Resulta que también es cierto que

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_2 \cap A_1) \qquad \text{(b)}$$

cuando hay tres etapas, y que

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1) \cdot P(A_2|A_1) \cdots P(A_n|A_{n-1} \cap A_{n-2} \cap \cdots \cap A_1) \qquad \text{(c)}$$

cuando hay ene etapas.

En el fondo éstas son las fórmulas que se aplican —implícitamente— cuando se calcula la probabilidad de una rama del diagrama de árbol. Todo es consecuencia de la fórmula (a). La demostración de estas fórmulas viene bien para explicar un método de demostración muy utilizado en Matemáticas: la demostración por inducción. Este método de demostración se explica en los apartados avanzados (no entran en la asignatura).

Es importante recalcar que, como se dijo para la fórmula (a), y por deducirse de ella, las fórmulas (b) y (c) son aplicables tanto a sucesos simultáneos como a sucesos consecutivos.

Independencia

Es interesante recalcar que en el caso en que los sucesos son independientes, el condicionamiento no varía la probabilidad de los sucesos, por lo que las fórmulas anteriores se convierten en:

$$P(A_1) \cdot P(A_2|A_1) \cdots P(A_n|A_{n-1} \cap A_{n-2} \cap \cdots \cap A_1) = P(A_1) \cdot P(A_2) \cdots P(A_n) \qquad \text{(d)}$$

Como aplicación, véanse los ejercicios Ejercicio 5 y Ejercicio 8.

Una fórmula útil

Dado que $A \cap B = B \cap A$, se cumple que $P(A \cap B) = P(B \cap A)$ y, por tanto:

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

fórmula que sirve para relacionar las probabilidades de los sucesos $A|B$ y $B|A$. Esto es útil para calcular la probabilidad de uno conocida la del otro.

Teorema de la probabilidad total

La idea de este teorema y su interpretación se introducen mejor con el dibujo:

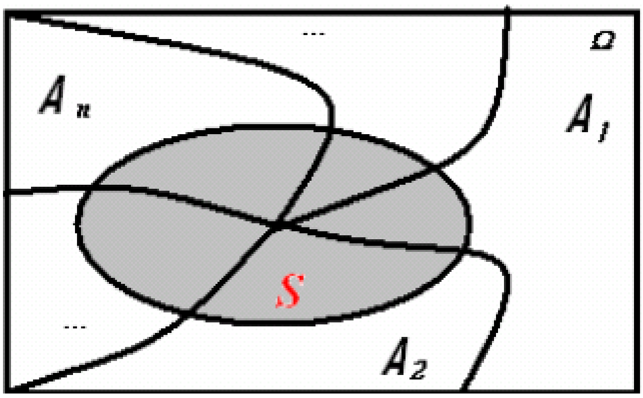


Figura: Descomposición disjunta del suceso S según su intersección con los Ai

La situación es la siguiente: el espacio muestral total Ω está dividido por una partición disjunta y completa de sucesos, es decir, en la que esos sucesos no se intersecan (son incompatibles, no pueden suceder a la vez) y cubre todo el espacio muestral; hay además un nuevo suceso ***S***. Si pensamos en áreas es fácil ver del dibujo que:

$$S=(S\cap A_1)\cup(S\cap A_2)\cup\cdots\cup(S\cap A_n).$$

Por tanto, como la partición es completa y disjunta,

$$P(S)=P(S\cap A_1)+P(S\cap A_2)+\cdots+P(S\cap A_n).$$

Aplicando la fórmula de la probabilidad condicionada a cada suceso $S\cap A_i$,

$$P(S)=P(A_1)\cdot P(S|A_1)+P(A_2)\cdot P(S|A_2)+\cdots+P(A_n)\cdot P(S|A_n)$$

En los ejercicios, de esta fórmula pueden preguntar por cualquiera de sus factores, siempre que haya información suficiente para conocer todos los demás. Suele aplicarse en ejercicios donde hay un suceso y se quiere estudiar su relación con una colección de sucesos que llenan todo el universo de sucesos.

En el caso particular en que la sucesión A_i tiene sólo a A y a su complementario A^c, repetimos los razonamientos informalmente con la palabra «área» en vez de «probabilidad»:

El suceso: B

La partición completa de sucesos: A y A^c

Por otro lado, los diagramas de Venn y pensar en áreas ayuda mucho a entender y recordar las fórmulas. Como A y A^c llenan todo el universo:

área(A) + área(A^c) = 1

Entonces B se puede descomponer en tantas partes como sucesos tenga la colección:

área(B) = área($B \cap A$) + área($B \cap A^c$)

Y para calcular estas dos áreas, se utiliza la «fórmula del área condicionada»

área($B \cap A$) = área(A)área($B|A$)

área($B \cap A^c$) = área(A^c)área($B|A^c$)

Sustituyendo estos resultado en el anteriores área(B) = área(A)área($B|A$) + área(A^c)área($B|A^c$)

Y ésta es ya la fórmula del teorema (para más de dos sucesos se razona igual).

Teorema de Bayes

El teorema de Bayes es una fórmula que nos permite calcular la probabilidad de que las cosas vayan a suceder o hayan sucedido por un «camino» concreto de entre los posibles (por «camino» se entiende una sucesión de etapas, experimentos, sucesos o variables). La demostración es fácil. Supongamos que hay una partición de sucesos disjunta y completa de Ω, y otro suceso S. Queremos calcular la probabilidad $P(A_i|S)$ de que, sabiendo que se cumpla (o haya cumplido) S, se cumpla (o haya cumplido) también A_i. Tenemos que:

$$\begin{cases} P(A_i|S)=\dfrac{P(A_i\cap S)}{P(S)} \\ P(S|A_i)\cdot P(A_i)=P(S\cap A_i) \\ P(S)=P(A_1)\cdot P(S|A_1)+P(A_2)\cdot P(S|A_2)+\cdots+P(A_n)\cdot P(S|A_n) \end{cases}$$

de donde

$$P(A_i|S)=\frac{P(A_i)\cdot P(S|A_i)}{P(A_1)\cdot P(S|A_1)+P(A_2)\cdot P(S|A_2)+\cdots+P(A_n)\cdot P(S|A_n)}$$

Si observamos la fórmula vemos que en el denominador aparece S condicionado a todos los A_j y en el numerador sólo a A_j, mientras que lo que queremos calcular es la probabilidad de A_i una vez que sabemos que ha sucedido S. En los ejercicios será entonces muy importante definir bien quién es la partición de sucesos y quién es el suceso S.

Un poco más...

En esta sección se incluyen algunos comentarios de un nivel más alto, que pretenden por un lado facilitar la comprensión de lo visto y de los temas siguientes, y por otro satisfacer la curiosidad de quien quiera ir un poco más allá de lo estrictamente necesario según el temario.

Definiciones frecuentista y bayesiana de la probabilidad

Según http://es.wikipedia.org/wiki/Probabilidad:

La palabra *probabilidad* no tiene una definición consistente. De hecho, hay dos amplias categorías de **interpretaciones de la probabilidad**: los frecuentistas hablan de probabilidades sólo cuando se trata de *experimentos aleatorios* bien definidos. La frecuencia relativa de ocurrencia del resultado de un experimento, cuando se repite el experimento, es una medida de la probabilidad de ese suceso aleatorio.

Los bayesianos, no obstante, asignan las probabilidades a *cualquier declaración*, incluso cuando no implica un proceso aleatorio, como una manera de representar su verosimilitud subjetiva.

La *Ley de los grandes números* es el apoyo teórico principal de la visión frecuentista de la probabilidad:

$$P(A)=lím\ fr(A)$$

Es decir, la probabilidad de un suceso es el número al que se aproxima la frecuencia a medida que repetimos el experimento indefinidamente.

La fórmula de Bayes es la principal herramienta teórica que utilizan los bayesianos (de ahí su nombre). Introduce la información a priori (a veces subjetiva) a través de una de las probabilidades condicionadas de la definición de probabilidad condicionada.

Condicionamiento de *n* sucesos: demostración por inducción

El *método de inducción* es un método de demostración especialmente indicado para situaciones en que queremos demostrar que algo se cumple para cualquier valor natural de un parámetro (en nuestro caso el número de conjuntos en la intersección). La idea es muy sencilla, consiste en demostrar:

1. Que se cumple para el primer valor de la cadena.
2. Que, si es cierto para *n-1*, es cierto para *n*.

Se va utilizar esta método para demostrar la fórmula

$$P(A_1\cap A_2\cap\cdots\cap A_n)=P(A_1)\cdot P(A_2|A_1)\cdots P(A_n|A_{n-1}\cap A_{n-2}\cap\cdots\cap A_1)$$

[Demostración 1] Una primera demostración, incompleta, es la siguiente. Podría basarse en dar por cierto (se intuye con un dibujo, pero habría que demostrarlo utilizando las propiedades de la teoría de conjuntos y el método de inducción) que:

$$A_1\cap A_2\cap\cdots\cap A_n=A_1\cap\{A_2|A_1\}\cap\{A_3|A_2\cap A_1\}\cap\cdots\cap\{A_n|A_{n-1}\cap A_{n-2}\cap\cdots\cap A_1\}$$

Si nos creemos lo anterior y que esos sucesos son independientes, entonces tiene que cumplirse que

$$P(A_1\cap A_2\cap\cdots\cap A_n)=P(A_1)\cdot P(A_2|A_1)\cdots P(A_n|A_{n-1}\cap A_{n-2}\cap\cdots\cap A_1)$$

[Demostración 2] Otra forma de demostración es aplicando directamente el método de inducción sobre las probabilidades que queremos hallar. En nuestro caso, puesto que el caso de dos conjuntos es una definición, debemos demostrar que, a partir de ella, es cierta para $n=3$. Lo hacemos utilizando el truco de agrupar lo que nos interesa en dos sucesos compuestos, ya que para este caso tenemos que es cierto (por definición):

$$P(A_1\cap A_2\cap A_3)=P(\{A_1\cap A_2\}\cap A_3)=P(A_3)\cdot P(\{A_1\cap A_2\}|A_3)=P(A_3)\cdot\frac{P(A_3|\{A_1\cap A_2\})\cdot P(\{A_1\cap A_2\})}{P(A_3)}\quad=$$

donde hemos aplicado en el último paso la fórmula $P(B)\cdot P(A|B)=P(A)\cdot P(B|A)$. Ahora queda

$$= \cancel{P(A_3)}\cdot\frac{P(A_3|\{A_1\cap A_2\})\cdot P(\{A_1\cap A_2\})}{\cancel{P(A_3)}}=P(A_3|\{A_1\cap A_2\})\cdot P(\{A_1\cap A_2\})=P(A_3|\{A_1\cap A_2\})\cdot P(A_2|A_1)\cdot P(A_1)$$

En el penúltimo paso ya hemos conseguido por un lado que aparezca una probabilidad condicionada que nos interesa y por otro que en la intersección haya un término menos. En el último paso se ha aplicado la definición de probabilidad condicionada. Ya entrevemos que para obtener la fórmula buscada sólo hay que ir aplicando una vez tras otra el procedimiento anterior; en cada paso iremos obteniendo un factor de probabilidad condicionada que multiplica y otro factor que tiene una intersección con un conjunto menos. Vamos a hacer el paso con detalle. Supongamos que es cierto que

$$P(A_1\cap A_2\cap\cdots\cap A_{n-1})=P(A_1)\cdot P(A_2|A_1)\cdots P(A_{n-1}|A_{n-2}\cap A_{n-3}\cap\cdots\cap A_1).$$

Tenemos ahora que

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

y ahora, aplicando que es cierto para *n-1*, se obtiene el resultado deseado:

$$P(A_1\cap A_2\cap\cdots\cap A_n)=P(A_1)\cdot P(A_2|A_1)\cdots P(A_n|A_{n-1}\cap A_{n-2}\cap\cdots\cap A_1)$$

Como se ve, el procedimiento es engorroso pero no es difícil.

Combinatoria

La Combinatoria se encarga de dar fórmulas y métodos que nos permiten calcular el número de formas de organizar los elementos de un conjunto, de tomar subconjuntos de un tamaño concreto, etcétera. Esto lo hace –en nuestro caso– a través de conceptos como las *variaciones*, las *permutaciones* y las *combinaciones*, todas ellas en sus versiones sin y con repetición. Una de las expresiones más utilizada es la de «combinaciones de *m* elementos tomados de *n* en *n*», es decir, la que nos dice cuántos posibles subconjuntos de *n* elementos distintos puedo hacer sin que el orden en que ordeno los elementos en el subconjunto importe y sin que estos elementos se repitan. La fórmula es:

$$C_m^n=\binom{m}{n}=\frac{m!}{n!\cdot(m-n)!}$$

donde $m!=m\cdot(m-1)\cdot(m-2)\cdots 3\cdot 2\cdot 1$.

Hay una información muy breve de combinatoria en

«Combinatoria elemental», http://www.Casado-D.org/edu/CombinatoriaElemental.pdf

y en

http://es.wikipedia.org/wiki/Combinatoria

Distribuciones de probabilidad

Se ven en este documento las distribuciones discretas más importante, la *de Bernoulli* y la *binomial*, y la distribución continua también más importante, la *normal*.

Un distribución de probabilidad es una regla que asigna números a los sucesos para atribuirles una probabilidad. Estas distribuciones no dependen de ningún experimento concreto, sino que pueden aplicarse a experimentos distintos.

Distribución de Bernoulli

Merece la pena dedicar un poco de atención a la *distribución de Bernoulli*, no sólo por el propio interés de esta distribución sino también porque es fácil que aparezca con frecuencia en cualquier texto sobre probabilidad.

Dado un experimento dicotómico, es decir, un experimento en el que sólo puede suceder una cosa o la contraria, podemos asociar a esos sucesos los valores 1 y 0, con lo que tenemos construida una función que es una variable aleatoria con distribución de Bernoulli. Tendremos así que si X sigue una *ley* o *distribución de Bernoulli*:

$$P(X=1)=p \qquad \text{y} \qquad P(X=0)=1-p$$

Como para toda distribución, la suma de las probabilidades de todos los valores tiene que ser 1, como sucede aquí: $p + (1 - p) = 1$. Esta distribución puede tomar valores aislados (dos), por lo que es una distribución discreta.

Ejemplos de experimentos que llevan asociadas variables aleatorias que siguen esta distribución serían: lanzar una moneda, lanzar un dado y mirar si sale *4* o no, etcétera.

Distribución binomial

Se puede pensar que la distribución de Bernoulli es un «contador», ya que toma el valor 1 cuando ocurre el suceso de interés y 0 cuando no. Así, si repetimos el experimento n veces y sumamos los valores de las variables de Bernoulli $X_1, ..., X_n$, estaríamos contando el número de veces que ha ocurrido nuestro suceso de entre n veces. Esto mismo es la definición de la *distribución binomial*, que equivalentemente se puede pensar como construida sumando n variables de Bernoulli. Por tanto, esta variable es también un «contador». Es decir, si Y es una binomial, se puede pensar que:

$$Y=\sum_{i=1}^{n} X_i$$

Por su significado, una variable binomial puede tomar valores en el conjunto $\{0,1,2, ..., n\}$, por lo que es una distribución discreta. Si Y sigue una ley binomial, $p^k \cdot (1-p)^{n-k}$ es la probabilidad de que suceda, en un orden concreto, k veces el suceso de interés y n-k veces su complementario; pero como sólo nos interesa contar el número de veces que ocurre el suceso y no importan las posiciones de las tiradas en que han sucedido esos k sucesos, se utiliza la combinatoria para calcular el número de formas en que podemos distribuir esos k sucesos entre los n. Se obtiene:

$$P(Y=k)=\binom{n}{k}\cdot p^k\cdot(1-p)^{n-k} \qquad (*)$$

Vemos que la distribución binomial tiene en cuenta todas las posibles formas de que ocurra k veces el suceso de interés y n-k veces el complementario; es decir, en el fondo esto es lo mismo que se puede hacer dibujando un diagrama de árbol (para no dejarse ningún caso) y contando el número de «ramas» o «caminos» que incluyen que haya sucedido k veces el suceso de intereś y n-k veces el complementario. De esta manera tenemos que la distribución binomial permite en muchos casos sustituir el diagrama de árbol y hacer los cálculos más fácilmente.

Cada variable aleatoria discreta tiene asociada una probabilidad (*masa*) a cada posible valor que puede tomar. La aplicación que hace esta asociación se llama *función de masa*; en el caso de la binomial viene dada por (*). Se podría hacer entonces un gráfico como el siguiente:

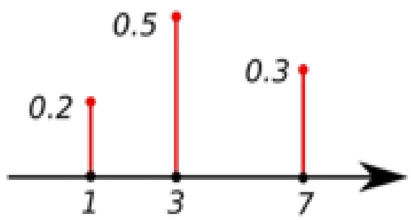


Figura: Ejemplo de función de masa

Para resolver ejercicios

En los ejercicios que tengamos que resolver, nos preguntarán por la probabilidad de sucesos que sean simples (para el experimento en cuestión) o compuestos, pero que se puedan formar fácilmente a partir de sucesos simples, para los que sí sabemos calcular la probabilidad con la fórmula (*). Ejemplos de sucesos por los que nos pueden preguntar son:

- $\{Y=k\}$ $\rightarrow$ Su probabilidad se calcula directamente con la fórmula (*).

- $\{Y\leq k\}$ $\rightarrow$ Descompondremos este suceso en la partición disjunta y completa

  $$\{Y\leq k\}=\{Y=0\}\cup\{Y=1\}\cup\cdots\cup\{Y=k\}$$

  de donde se tiene que

  $$P\{Y\leq k\}=P\{Y=0\}+P\{Y=1\}+\cdots+P\{Y=k\}$$

  y ahora podemos aplicar (*) a cada uno de los sumandos. En algunos ejercicios interesa tener en cuenta que a veces el suceso complementario se descompone de forma más sencilla.

- $\{Y<k\}$ $\rightarrow$ Tendremos en cuenta que

  $$\{Y<k\}=\{Y\leq k-1\}$$

  y aplicaremos la descomposición del apartado anterior

  $$P\{Y<k\}=P\{Y\leq k-1\}$$

- $\{Y>k\}$ → Tendremos en cuenta que ya sabemos calcular la probabilidad de su complementario

  $$\{Y \leq k\}$$

  y haremos

  $$P\{Y>k\}=1-P\{Y \leq k\}$$

- Por último nos pueden preguntar por algún suceso más complicado, por ejemplo $\{Y \leq k_1\} \cup \{Y>k_2\}$. En este caso tendremos que ir de nuevo descomponiendo el suceso complicado en otros más sencillos para los que sí sepamos calcular la probabilidad.

No hace falta decir que no se deben aprender de memoria las descomposiciones anteriores, sino comprenderlas para poder generalizar las ideas a otros ejercicios y distribuciones discretas (y continuas).

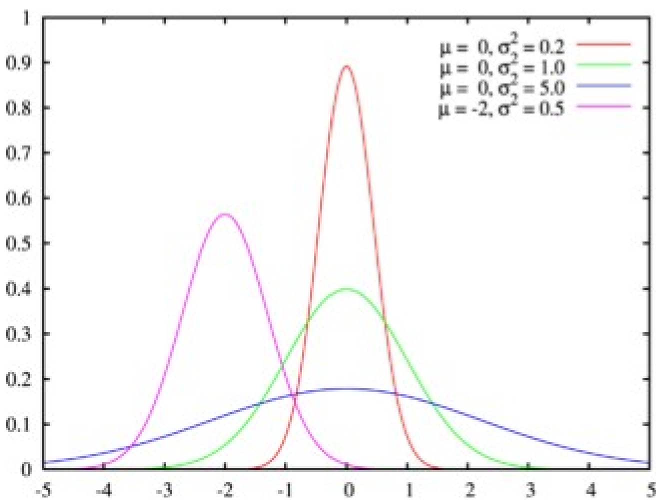
Como aplicación, véanse los ejercicios Ejercicio 10, Ejercicio 14 y Ejercicio 15.

Distribución normal o gaussiana

La *distribución normal* o *gaussiana* es una distribución continua, puesto que puede tomar valores en un conjunto continuo de puntos (infinito no numerable). Como en toda variable continua, puesto que son infinitos, es imposible asignar una probabilidad mayor que cero a cada uno de ellos y esperar que la suma sea 1 (máximo de la escala en que por convenio se miden las probabilidades). Se utiliza entonces un concepto análogo a la función de masa de las variables discretas: para las variables aleatorias continuas se utiliza la *función de densidad*, que da con su altura una idea de lo probable que es cada valor (¡pero esa altura no es su probabilidad!). En esta distribución el nombre de *normal* se debe a la frecuencia con que aparece y al papel central que desempeña en la Teoría de Probabilidades, y el nombre de *gaussiana* al matemático alemán Gauss (hay una pequeña biografía en http://www.Casado-D.org/edu/Gauss.pdf), quien la empezó a utilizar para modelizar los errores en los procesos de medición. Para esta ley, la función de densidad es:

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

cuya gráfica, para distintos valores de los parámetros μ y σ, es:



http://es.wikipedia.org/wiki/Distribuci%C3%B3n_normal

Se ve en el dibujo (mirando sólo a la curva de un color concreto) cuál es el valor más probable (moda), que coincide con la media y la mediana: la distribución es simétrica, es decir, la probabilidad al alejarse de la media una cantidad por defecto o por exceso es la misma. Por último, se ve (mirando a todas las curvas) que cuanto menor es la varianza, mayores probabilidades tienen los valores cercanos al valor central; es decir, hay más «densidad» de probabilidad cerca de ese centro.

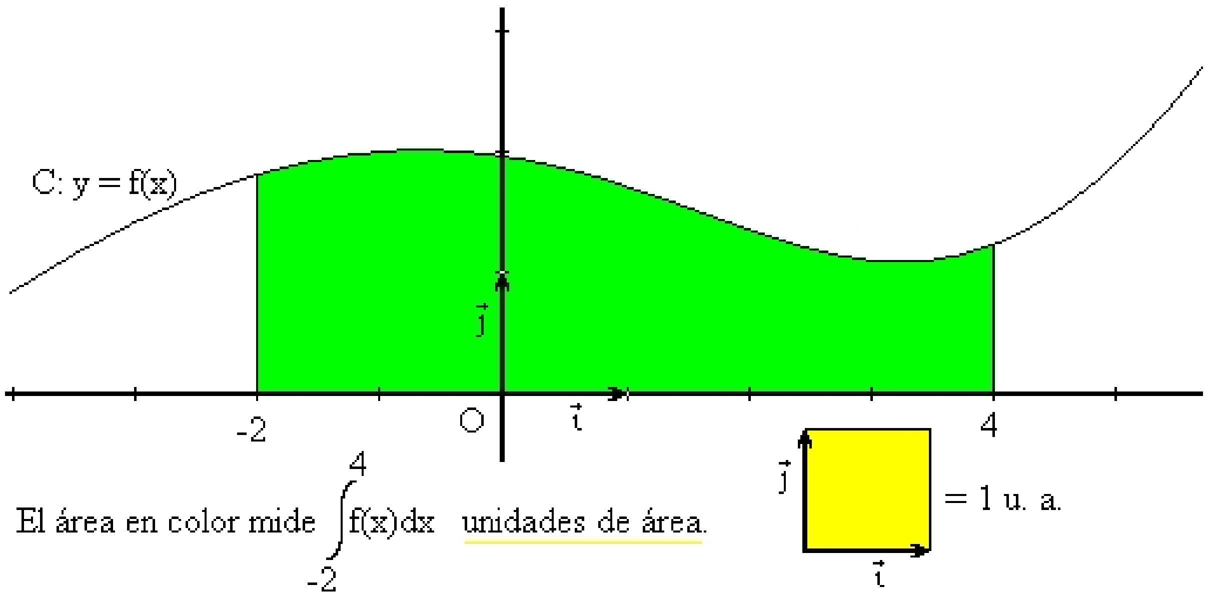
Para las distribuciones continuas no tiene sentido calcular la probabilidad de un valor concreto (porque es 0), sino sólo calcular la de un intervalo de valores. Esta probabilidad se puede calcular utilizando la *función de distribución*, que se define como $F(t)=P(X\leq t)$, y que en este caso es:

$$F(t)=P(X\leq t)=\int_{-\infty}^{t}\frac{1}{\sigma\cdot\sqrt{(2\pi)}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx$$

En Matemáticas la integral definida de una función

$$\int_a^b f(x)dx$$

calcula el área que la función *f(x)* encierra entre ella y el eje horizontal, entre los valores *a* y *b*. Si el área está por debajo del eje le asigna un signo negativo, pero en nuestro caso no tenemos que preocuparnos de esto ahora, porque *f(x)* es siempre positiva o cero.



http://es.wikipedia.org/wiki/Integral_definida

Esto nos facilita las cosas, porque podemos pensar en las probabilidades mirando las áreas.

Dado que siempre se cumple que

$$\{X\leq t\}=\{X<t\}\cup\{X=t\},$$

se tiene que

$$P\{X\leq t\}=P\{X<t\}+P\{X=t\}.$$

Como la probabilidad de un punto aislado es cero para las distribuciones continuas, $P\{X=t\}=0, \forall\, t\in\mathbb{R},$ se cumple que $P(X\leq t)=P(X<t);$ es decir, da igual incluir o no el punto *t*

en la desigualdad (en el suceso). Cuidado: esto no es así con las distribuciones discretas si la probabilidad del punto *t* es distinta de cero.

Ahora, con la función de densidad o la función de distribución, se pueden calcular las probabilidades de todos los sucesos que queramos:

<u>Si conocemos la función de densidad</u>

Si nos interesa el suceso $\{k_1 \leq X \leq k_2\}$, se hace

$$P\{k_1 \leq X \leq k_2\} = \int_{k_1}^{k_2} \frac{1}{\sigma \cdot \sqrt{(2\pi)}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx .$$

<u>Si conocemos la función de distribución</u>

En este caso, como nos dan (sea a través de la expresión de *F*(*x*), de un programa informático o mediante tablas) los cálculos hechos, para el mismo suceso $\{k_1 \leq X \leq k_2\}$, se hace

$$P\{k_1 \leq X \leq k_2\} = F(k_2) - F(K_1).$$

Para resolver ejercicios

En los ejercicios sobre la distribución normal, es frecuente que haya que seguir estos pasos:

1) Lo primero es identificar el valor de los parámetros.
2) Es necesario hacer casi siempre unas operaciones sencillas con las desigualdades. Hay un teorema que dice:

   Si X es $N(\mu, \sigma)$ entonces $(X - \mu)/\sigma$ es $N(0, 1)$

   Este teorema se utiliza para remitir los cálculos siempre a la normal estándar y que baste sólo con la tabla de estos valores de los parámetros. Entonces, en general:

   $$a \leq X \leq b$$

   $$\frac{a-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}$$

   y los dos cocientes de los lados son números, mientras que a lo del centro, la variable tipificada, lo llamamos *Z*. Entonces:

   $$P(a \leq X \leq b) = P\left(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right)$$

3) Ahora, para calcular la probabilidad de la derecha se utiliza:

   $$\left\{\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right\} = \left\{Z \leq \frac{b-\mu}{\sigma}\right\} \setminus \left\{Z < \frac{a-\mu}{\sigma}\right\} = \left\{Z \leq \frac{b-\mu}{\sigma}\right\} \cap \left\{Z < \frac{a-\mu}{\sigma}\right\}^C$$

   $$P\left(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right) = P\left(Z \leq \frac{b-\mu}{\sigma}\right) - P\left(Z < \frac{a-\mu}{\sigma}\right) = F_Z\left(\frac{b-\mu}{\sigma}\right) - F_Z\left(\frac{a-\mu}{\sigma}\right)$$

4) Antes de ir a buscar el valor de estas probabilidades a la tabla de la normal estándar, hay que tener en cuenta algunos detalles. La tabla está dada para algunos valores (no se puede dar para infinitos) y como la normal es simétrica, sólo para los positivos. Así que hay que hacerse un dibujo con la campana de Gauss centrada en su media y ver cómo hay que consultar la tabla. Para números negativos, como la distribución es simétrica, hay que hacer:

$$P(Z \leq -t) = P(Z \geq t)$$

y, si la tabla sólo da áreas inferiores, es decir, para sucesos de la forma $\{Z \leq n\}$, se hace

$$P(Z \leq -t) = P(Z \geq t) = 1 - P(Z < t)$$

Como aplicación, véanse los ejercicios Ejercicio 11, Ejercicio 12, Ejercicio 13, Ejercicio 14 y Ejercicio 15.

Tipificación de una variable

Al igual que las variables estadísticas, también se pueden *tipificar* las variables aleatorias. Tipificar una variable es aplicarle la transformación

$$Z = \frac{X - \mu}{\sqrt{\sigma^2}} = \frac{X - \mu}{\sigma}$$

que da lugar a una nueva variable aleatoria *Z*. En la fórmula anterior μ y σ son, respectivamente, la media y la desviación típica de la distribución que sigue la variable *X*.

En el caso particular en el que la variable aleatoria que tipificamos sigue una ley normal, se cumple que:

Teorema

Cuando X~N(μ, σ), si se define Z := (X - μ)/σ, se cumple que Z~N(0, 1).

Este teorema se utiliza como herramienta para no tener que dar una tabla para cada par de valores (habría infinitos) de los parámetros μ y σ y así poder remitir siempre a la tabla de la *N*(0,1).

Tablas de la distribución normal

Para no tener que calcular cada vez las probabilidades de los sucesos a partir de las funciones de densidad o distribución, se han creado para las distribuciones más utilizadas tablas en las que se recogen esos valores ya calculados, para un número suficiente de valores de la variable (algunas variables pueden tomar infinitos).

Otra cosa que suele simplificar las tablas, por ejemplo, es que la distribución sea simétrica respecto al origen, puesto que es suficiente dar valores positivos (esto sucede en la tabla de la normal estándar). Como tampoco se puede dar una tabla para cada posible combinación de valores de los parámetros de la distribución (son infinitas), lo que se hace es bien darlas para las combinaciones más frecuentes, caso de la binomial, o utilizar algún resultado teórico que remita a estas tablas, caso de la normal.

Con cada tabla se suele incluir un dibujo que aclara si el valor del área/probabilidad que se está dando en la tabla es $F(t) = P(X \leq t)$, que es lo más frecuente, o $P(X > t)$. Es importante mirar esto cuando estamos trabajando con una tabla. Con una tabla podemos:

a) Calcular directamente la probabilidad de que la variable sea menor o igual que un valor *t*, con

$$p=P(X\leq t)$$

b) Dada una probabilidad, calcular el valor de la variable; es decir, buscar *t* dada *p* en la expresión

$$p=P(X\leq t)$$

c) Calcular la probabilidad de un intervalo de valores, con

$$P(t_1\leq X\leq t_2)=P(X\leq t_2)-P(X\leq t_1)$$

puesto que $\{t_1\leq X\leq t_2\}=\{X<t_2\}\setminus\{X\leq t_1\}=\{X\leq t_2\}\cap\{X<t_1\}^C$.

d) En situaciones más complicadas no podemos dar un valor exacto para la probabilidad si el valor de *t* no está en la tabla; pero sí podemos acotarlo, es decir, coger los valores de la tabla que están justo por debajo y por encima de *t*, y decir que la probabilidad buscada está entre estas dos probabilidades. Para algunos ejercicios esto es suficiente.

En nuestro caso es interesante observar el paso siguiente, que tendremos que aplicar en algunos ejercicios:

$$P(Z\leq -t)=P(Z\geq t)=1-P(Z<t)$$

1) El primer cambio lo hemos tenido que hacer porque la tabla de la normal estándar sólo está dada para valores positivos de la variable.
2) El segundo cambio lo hemos tenido que hacer porque la tabla nos informa de las áreas de las colas inferiores, es decir, de $P(Z\leq z)$, no directamente de $P(Z>z)$.

Un poco más...

En esta sección se incluyen algunos comentarios de un nivel más alto, que pretenden por un lado facilitar la comprensión de lo visto y de los temas siguientes, y por otro satisfacer la curiosidad de quien quiera ir un poco más allá de lo estrictamente necesario según el temario.

Teorema del límite central

Se da aquí el *Teorema del límite central* en una versión un poco más general que la que se utilizará, para así sacar más conclusiones y comprender mejor la distribución normal.

Teorema

Si tenemos un conjunto de variables aleatorias independientes, (X_1, ..., X_n) tales que cada X_i tiene media μ_i y desviación típica σ_i, entonces se cumple que

$$X_1+X_2+\cdots+X_n \quad \rightarrow \quad N\left(\mu_1+\mu_2+\cdots+\mu_n, \sqrt{\sigma_1^2+\sigma_2^2+\cdots+\sigma_n^2}\right)$$

ó, equivalentemente,

$$\frac{(X_1+X_2+\cdots+X_n)-(\mu_1+\mu_2+\cdots+\mu_n)}{\sqrt{(\sigma_1^2+\sigma_2^2+\cdots+\sigma_n^2)}} \quad \rightarrow \quad N(0,1)$$

Una función que involucra variables aleatorias se denomina *estadístico* y también es una cantidad aleatoria (podemos imaginar que es «una variable aleatoria más complicada»). Entonces, si se construye el *estadístico* de la izquierda de la expresión superior anterior, en el límite, cuando añadimos más y más variables, tenderá a distribuirse como una variable aleatoria normal cuya media es la suma de las medias y cuya desviación típica es la raíz cuadrada de la suma de las varianzas (¡no la suma de las desviaciones típicas!). Puesto que sólo se ha hablado de la media y de la desviación típica de las variables que sumamos, lo que se quiere decir es que este teorema es válido para cualquier conjunto de variables, sigan la distribución que sigan (incluso pueden no seguir todas la misma). Esto significa que la distribución normal es una distribución «especial», límite de ciertas sumas de las demás (por eso lo de «límite central»). El teorema también está afirmando que siempre que el resultado de un proceso sea la suma de muchos factores pequeños (cada uno es una parte pequeña de la suma total), seguirá aproximadamente una ley normal**.** De aquí que esta distribución aparezca tan frecuentemente en la Naturaleza para explicar la altura o el peso de los seres vivos, los errores de un proceso de medición, etcétera.

Aproximación de la binomial a la normal

Una de las utilidades del teorema del límite central es que se pueden aproximar probabilidades de la suma de variables aleatorias a partir de las probabilidades de la variable aleatoria de la derecha.

En el caso concreto en que cada X_i sigue una ley de Bernoulli, como su suma es una binomial, el teorema se convierte en el teorema

Teorema

Si Y sigue una distribución binomial, como para ella $\mu=np$ *y* $\sigma=\sqrt{npq}$, *en el límite se tiene que*

$$\frac{Y-np}{\sqrt{npq}} \quad \rightarrow \quad N(0,1).$$

Esta versión del teorema es la que se utiliza para aproximar la binomial por una normal (véase el Ejercicio 14). Esta aproximación es buena cuando el valor de *n* es grande y el valor de *p* es tal que *np* es también grande. Para el caso en que *p* es tan pequeño que *np* toma valores pequeños incluso para valores grandes de *n*, la aproximación que se utiliza es la que se describe en Aproximación de la binomial a la Poisson.

Corrección por continuidad

En los teoremas de los apartados Teorema del límite central y Aproximación de la binomial a la normal, el cociente es un estadístico discreto (como una «variable aleatoria complicada» discreta), mientras que su límite es una variable aleatoria continua. Para disminuir el posible error de aproximar un tipo de distribución con el otro, se aplica lo que se conoce como *corrección por continuidad.*

La idea que hay detrás de la corrección por continuidad en este caso es imaginar la función de masa de la variable discreta no como formada por puntos –o líneas– sobre los valores discretos de la variable, como corresponde a una función de masa, sino como barras con la misma altura pero con anchura 1 centradas en los valores (es lo que separa valores enteros contiguos), de aquí que estas barras sobresalgan 0,5 por cada lado. De esta manera, la variable discreta cubre toda la recta real de valores, como hace la continua (normal).

Si nos imaginamos las barras de anchura uno centradas en los valores que puede tomar la variable discreta, podemos deducir las equivalencias que hay que considerar cuando queramos aplicar la corrección, sin necesidad de aprenderlas de memoria:

a) $P(X=a)=P(a-0{,}5\leq X\leq a+0{,}5)$ → Consideramos la barra que corresponde al valor *a*.

b) $P(X\leq a)=P(X\leq a+0{,}5)$ → Porque sí se incluye toda la barra del valor *a*, por estar incluido este valor.

c) $P(X<a)=P(X\leq a-0{,}5)$ → Porque no se incluye la barra del valor *a*, por no estar incluido este valor.

d) $P(a\leq X\leq b)=P(a-0{,}5\leq X\leq b+0{,}5)$ → Porque, dado que la igualdad primera incluye los valores *a* y *b*, es necesario incluir enteras las dos barras correspondientes a estos valores. Como cada barra está centrada y sobresale *0,5* a cada lado del valor, tenemos que en *a* hay que moverse *0,5* puntos hacia la izquierda y en *b* hay que incluir *0,5* puntos a la derecha.

e) $P(a<X<b)=P(a+0{,}5\leq X\leq b-0{,}5)$ → Porque, dado que la desigualdad primera no incluye los valores *a* y *b*, es necesario excluir enteras las dos barras correspondientes a estos valores. Como cada barra está centrada y sobresale *0,5* a cada lado del valor, tenemos que en *a* hay que moverse *0,5* puntos hacia la derecha y en *b* hay que incluir *0,5* puntos a la izquierda.

f) Y así con otros sucesos más complicados, que pueden ser combinaciones de éstos...

Concluimos que el criterio es: incluir toda la barra si su valor discreto está incluido; es decir, si *a* está incluido en el suceso de interés, debemos modificar la expresión del suceso para incluir todo el rango (*a-05, a+0,5*).

Como aplicación, véase el ejercicio Ejercicio 14.

Aproximación de la binomial a la Poisson

En Aproximación de la binomial a la normal se ha mencionado cómo, basándose en un resultado teórico, era posible hacer los cálculos de las probabilidades de la binomial, para valores *n* y *np* grandes, aproximándolos por la distribución normal. Aquí se muestra el resultado teórico que se utiliza para aproximar los cálculos de las probabilidades de la binomial para valores de *p* que hacen que *np* no sea grande, mediante la *distribución de Poisson*.

Teorema

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

$$\lim \binom{n}{k} p^{n-k}(1-p)^k \quad \rightarrow \quad e^{-\lambda}\frac{\lambda^{n-k}}{(n-k)!}$$

para $k \leq n$.

Inecuaciones o desigualdades

Quien necesite repasar las operaciones elementales necesarias a este nivel, puede consultar el archivo

«Repaso no del todo elemental de Matemáticas elementales»

http://www.Casado-D.org/edu/RepasoMatematicas.pdf

Ejercicios

Tipos de ejercicios y problemas

A partir del esquema de teoría presentado en los primeros apartados, se me ocurre la siguiente clasificación general de los posibles tipos de ejercicios y problemas por los que se pregunta a los alumnos más frecuentemente. Para resolverlos es imprescindible identificar qué nos están preguntando.

A) En un espacio de probabilidad cualquiera: verificar alguna propiedad sobre sucesos genéricos, calcular la probabilidad de algunos sucesos genéricos con una probabilidad, etcétera. Son de este tipo los ejercicios 1 y 2.

B) En el espacio de probabilidad del experimento: describir el espacio muestral, descomponer sucesos compuestos hasta expresarlos en función de sucesos elementales, calcular al probabilidad de algunos sucesos o determinar toda la distribución de probabilidad (si el espacio muestral tiene pocos elementos). Son de este tipo los ejercicios 3, 4, 5, 6, 7 y 8.

C) Entre el espacio de probabilidad del experimento y el que induce la variable aleatoria: ejercicios que se pueden resolver en el espacio de probabilidad del experimento, pero que se resuelven más cómodamente definiendo alguna variable aleatoria. Pueden reducirse, por tanto, a ejercicios de los tipos B y D. Son de este tipo el ejercicio 9 y los que se resuelven en el documento «Variaciones sobre un ejercicio de Teoría de la Probabilidad» (el enlace está en el índice).

D) En el espacio de probabilidad inducido por la variable aleatoria (sin hablar ya de experimentos): descomponer sucesos para expresarlos de forma que se puedan aplicar las expresiones de las funciones de probabilidad, calcular la probabilidades de sucesos aplicando las funciones de probabilidad, encontrar las probabilidades de algunos sucesos en las tablas, aproximar la distribución binomial por la normal o la Poisson, aplicar la corrección por continuidad al calcular probabilidades con la distribución normal. Son de este tipo los ejercicios 10, 11, 12, 13, 14 y 15.

De esta posible clasificación general se pueden deducir varias cosas.

- Una es que los ejercicios del tipo A serán poco frecuentes, porque en general están contenidos más o menos implícitamente en los otros tipos. Se limitarán en muchos casos a preguntas sobre propiedades elementales de conjuntos o de los axiomas de la probabilidad.
- Los ejercicios de tipo C son también poco frecuentes, porque si son sencillos se resuelven en el espacio de probabilidad del experimento, tipo B, y si son complicados se utilizan variables aleatorias, tipo D.
- Respecto a dónde hay que ir a buscar la teoría para cada tipo de ejercicio, está claro que en todos se utiliza lo dicho en la sección «Espacio de probabilidad» más otras cosas no incluidas en este documento (los axiomas, por ejemplo); mientras, el material de la sección «Distribuciones de probabilidad» sólo se utiliza en los ejercicios del tipo D.

Ejercicio 1

Hallar P(A) si $A^c \subset B$, $P(B)=0{,}8$ *y* $P(B\cap A)=0{,}2$.

Tomado del libro:

Cálculo de probabilidades I
Ricardo Vélez Ibarrola
Víctor Hernández Morales
UNED, 1ªed (1995), 1ªreimp (1999)

Después de escribir Tipos de ejercicios y problemas he tomado estos dos primeros ejercicios para ponerlos como ejemplos del tipo A, pero no incluyo las soluciones por sencillas (están en el libro, de todas formas). Aunque sólo haya cogido de este libro estos dos primeros ejercicios tan sencillos, es un buen libro con un buen nivel.

Ejercicio 2

Expresar en términos de P(A), P(B) y P(A∩B) las probabilidades de los sucesos:

1. $A^c \cup B^c$
2. $A^c \cap B^c$
3. $A^c \cap B$
4. $A^c \cup B$

Tomado del libro:

Cálculo de probabilidades I
Ricardo Vélez Ibarrola
Víctor Hernández Morales
UNED, 1ªed (1995), 1ªreimp (1999)

Después de escribir Tipos de ejercicios y problemas he tomado estos dos primeros ejercicios para ponerlos como ejemplos del tipo A, pero no incluyo las soluciones por sencillas (están en el libro, de todas formas). Aunque sólo haya cogido de este libro estos dos primeros ejercicios tan sencillos, es un buen libro con un buen nivel.

Ejercicio 3

Al lanzar dos dados, hallar la probabilidad de que su suma sea 4 u 11.

Se trata claramente de un ejercicio sencillo del tipo B.

Experimento: El experimento consiste en lanzar dos dados (vamos a suponer que a la vez).

Espacio muestral: Los posibles resultados son: (1,1), (1,2), ..., (6,6); es decir:

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Se puede poner, más matemáticamente: Los pares (a,b) tales que $a\in\{1,2,\ldots,6\}$ y $b\in\{1,2,\ldots,6\}$. El espacio muestral consta de $6\cdot 6=36$ elementos.

Suceso de interés: $S=\{(a,b)\ /\ a+b\in\{4,11\}\}$.

Características: El dado que proporciona *a* es independiente del que proporciona *b*. Cada elemento del espacio muestral es equiprobable, por lo que su probabilidad es 1/36. Los sucesos {a+b=4} y {a+b=11} son incompatibles, es decir, o la suma vale 4 u 11, pero no las dos cosas a la vez.

Cálculos: Hagamos la descomposición completa en sucesos elementales

$$S=\{a+b\in\{4,11\}\}=\{a+b=4\}\cup\{a+b=11\}=\{(1,3),(2,2),(3,1)\}\cup\{(5,6),(6,5)\}$$

y finalmente, como los pares son incompatibles entre sí,

$$S=\{(1,3)\}\cup\{(2,2)\}\cup\{(3,1)\}\cup\{(5,6)\}\cup\{(6,5)\}.$$

Como tenemos una cadena de igualdades (de conjuntos de sucesos) y *S* se expresa como unión de sucesos incompatibles:

$$P(S)=P(\{(1,3)\})+P(\{(2,2)\})+P(\{(3,1)\})+P(\{(5,6)\})+P(\{(6,5)\})=5\cdot\frac{1}{36}=\frac{5}{36}=\frac{CF}{CP}$$

Los pasos anteriores incluyen implícitamente una demostración de la regla de Laplace (aunque sea en este caso concreto). ¿Dónde se ha utilizado que los sucesos son equiprobables? Sólo al final, al hacer $5\cdot\frac{1}{36}$. Es decir, la descomposición hecha hasta ahí es válida siempre, pero se llega a la regla de Laplace sólo si los sucesos elementales son equiprobables; en otro caso habría que calcular cada probabilidad de los sucesos elementales y sumarlas.

Ejercicio 4

Una urna contiene quince bolas blancas y doce negras, ¿cuál es la probabilidad de sacar dos bolas negras?

a) Sin reintegrar la primera bola que se saca

b) Reintegrando la primera bola

<u>Experimento</u>: Sacar una bola detrás de otra (da igual que las miremos al final las dos o cada una al sacarla).

<u>Espacio muestral</u>: Es {(*B*,*B*), (*B*,*N*), (*N*,*B*), (*N*,*N*)} (el orden queda indicado por la posición en el vector).

<u>Suceso de interés</u>: (*N*,*N*), es decir, $N_1\cap N_2$ (que la primera sea negra y a la vez [intersección] la segunda sea negra).

<u>Características del problema</u>: 1) El suceso de interés es elemental, porque está tal cual en el espacio muestral. 2) Sin embargo, dado que el experimento se lleva a cabo en dos etapas, es conveniente descomponer el suceso en esas etapas. 3) Estas etapas son compatibles e independientes.

Apartado a)

Cálculos: Aplicando la fórmula de la probabilidad condicionada:

$$P(N\,,N)=P(N_1\cap N_2)=P(N_1)P(N_2/N_1)=\frac{11}{27}\cdot\frac{11}{26}$$

porque, por la regla de Laplace

$P(N_1) = 12/27$

$P(N_2 \mid N_1) = 11/26$

Apartado b)

Cálculos: Aplicando la fórmula de la probabilidad condicionada:

$$P(N\,,N)=P(N_1\cap N_2)=P(N_1)P(N_2/N_1)=\frac{12}{27}\cdot\frac{12}{27}$$

porque, por la regla de Laplace

$P(N_1) = 12/27$

$P(N_2 \mid N_1) = P(N_2) = 12/27$

Comentario

Lo único que cambia de un apartado a otro es que, al reinsertar la bola, la situación al inicio de la segunda etapa vuelve a ser idéntica a la que había inicialmente y, por tanto, la segunda etapa del experimento es igual e independiente de la primera. Es decir, el reemplazo hace que las etapas sean independientes.

Ejercicio 5

¿Cuál es la probabilidad de que salga algún uno al lanzar un dado seis veces seguidas?

En este ejercicio conviene pensar un poco antes de hacer cálculos (bueno, en realidad esto conviene en todos, pero en éste especialmente...). Ponerse calcular en seguida puede llevar a coger un camino más largo de lo necesario, con el riesgo de equivocarse y la mayor inversión de tiempo.

Experimento: El experimento consiste en lanzar un mismo dado seis veces seguidas. Pero como cada lanzamiento es independiente, puede pensarse en lanzar seis dados distintos a la vez.

Espacio muestral: Los posibles resultados, si los ordenados en vectores, serían todos los vectores de longitud seis donde en cada posición puede ir cualquier dígito del 1 al 6. Hay en total $6 \cdot 6 \cdots 6 = 6^6 = 46656$ elementos en el espacio muestral. No vamos ahora a enumerarlos todos...

Suceso de interés: El conjunto de los vectores que tengan algún 1.

Características: El suceso de interés es compuesto y podría descomponerse en los vectores que tengan algún 1. Estos vectores son incompatibles entre sí (no hay dos iguales) y cada una de sus componentes es independiente de las demás. Pensando un poco nos damos cuenta de que al decir «algún uno», el suceso complementario «ningún uno» es más cómodo para trabajar con él. Para cada lanzamiento, pensando en la situación dicotómica «sale 1 con probabilidad 1/6 y no sale 1 con probabilidad 5/6», podemos denotar el suceso de interés como

$$S = \{(N, N, N, N, N, N)\}^c = \{N_1 \cap N_2 \cap N_3 \cap N_4 \cap N_5 \cap N_6\}^c$$

Cálculos:

$$P(S) = 1 - P((N_1 \cap N_2 \cap N_3 \cap N_4 \cap N_5 \cap N_6)) = 1 - P(N)^6 = 1 - \left(\frac{5}{6}\right)^6 = 1 - \frac{15625}{46656} = 1 - 0{,}33 = 0{,}67$$

Comentario

En este ejercicio ha sido clave utilizar:

- $P(A^c) = 1 - P(A)$
- Las seis tiradas son independientes

Ejercicio 6

Un fabricante de caramelos produce paquetes de diez caramelos cada uno. Se utilizan dos máquinas para ello. Después de haberse completado la elaboración de un buen número de lotes, se descubre que una de las máquinas, que produce el 40% de la producción total, tiene un defecto que ha conducido a la introducción de impurezas en el 10% de las unidades de caramelos que elabora. De un paquete de caramelos se extrae una unidad al azar y se analiza. Si dicha unidad no contiene ninguna impureza, ¿cuál es la probabilidad de que proceda de un paquete producido por la máquina defectuosa?

Experimento: Extraer un caramelo de uno de los paquetes producidos por las máquinas y mirar si tiene impurezas o no.

Espacio muestral: {«sí», «no»}.

Suceso de interés: Nos están preguntando no por un suceso que esté en el espacio muestral sino por un proceso condicionado a «no». No tendría sentido que nos preguntasen de qué máquina proviene un caramelo con impurezas porque sólo podría provenir de una. El suceso de interés es: «proviene de la máquina 1 sabiendo que no tiene impurezas».

Características: Las máquinas son independientes entre sí.

Cálculos: Definamos formalmente los sucesos

M_1 == Producido por la máquina 1
M_2 == Producido por la máquina 2
S == Tiene impurezas
N == No tiene impurezas

Con estas definiciones, la información del enunciado se traduce en:

$P(M_1) = 0{,}4$ y $P(M_2) = 0{,}6$

$P(S|M_1) = 0{,}1$ y $P(N|M_1) = 0{,}9$

$P(S|M_2) = 0$ y $P(N|M_2) = 1$

Lo que nos preguntan en el enunciado es $P(M_1|N)$. Parece que hay dos caminos de fabricación para el caramelo extraído: la máquina 1 y la 2. Por tanto, el ejercicio parece que se tiene que resolver por el teorema de Bayes.

$$P(M_1|N)=\frac{P(M_1)\cdot P(N|M_1)}{P(M_1)\cdot P(N|M_1)+P(M_2)\cdot P(N|M_2)}=\frac{0{,}4\cdot 0{,}9}{0{,}4\cdot 0{,}9+0{,}6\cdot 1}=0{,}375\,.$$

Comentario

En este ejercicio los paquetes no desempeñan ningún papel. Una máquina produce el 40% de los caramelos (o de los paquetes, da igual), la otra el 60%, y al final extraemos un caramelo. La probabilidad de que provenga el caramelo (o el paquete) de una u otra máquina es la mismas.

Ejercicio 7

El temario de unas oposiciones tiene veinticinco temas, de los cuales se eligen dos *por sorteo. Si un opositor se ha preparado quince de los veinticinco temas, ¿cuál es la probabilidad de que los dos temas que sean elegidos estén entre sus quince?*

Primera forma de resolverlo

Experimento: Se puede pensar en una bolsa con 15 bolas de un color y 10 de otro. Como los temas son distintos, se puede pensar que se extraen las dos a la vez o una tras otra sin reemplazo.

Espacio muestral: {(Sí, Sí), (Sí, No), (No, Sí), (No, No)}

Suceso de interés: (Sí, Sí), que escribimos como (T, T) ó $T_1 \cap T_2$.

Características: Los dos sucesos T_1 y T_2 son dependientes porque no hay reemplazo. Hay que utilizar la fórmula de la probabilidad condicionada.

Cálculos:

$$P(T_1 \cap T_2) = P(T_1)\,P(T_2 / T_1) = \frac{15}{25} \cdot \frac{14}{24} = 0{,}35$$

porque

P(T_1) = 15/25
P(T_2 | T_1) = 14/24

Segunda forma de resolverlo

Esta forma utiliza los números combinatorios y es un poco más difícil. Brevemente,

Casos favorables = Formas de tomar 2 elementos de entre 15 = $\binom{15}{2}$

Casos totales = Formas de tomar 2 elementos de entre 25 = $\binom{25}{2}$

Entonces, aplicando la regla de Laplace

$$P = \frac{CF}{CP} = \binom{15}{2} : \binom{25}{2} = \frac{15 \cdot 14}{25 \cdot 24} = 0{,}35.$$

Comentario

Vemos que a veces convertir un problema real en uno más abstracto facilita su resolución.

Ejercicio 8

En una urna hay 8 bolas blancas, 5 negras y 2 rojas. Se extraen tres bolas al azar sin reemplazo: a) Probabilidad de que las tres sean blancas; b) Probabilidad de que dos sean blancas y una negra.

Primera forma de resolverlo

Experimento: El experimento consiste en extraer tres bolas de la urna, una tras otra, sin reemplazo.

Espacio muestral: Los posibles resultados son todas las ternas: (B,B,B), (B,B,N),..., (R,R,R). Las bolas sólo se distinguen entre sí por el color.

Sucesos de interés: $A=\{(B,B,B)\}$ y $B=\{(B,B,N),(B,N,B),(N,B,B)\}$.

Características: A es un suceso elemental. B se compone de tres sucesos elementales incompatibles entre sí. Los sucesos del espacio muestral no son equiprobables, porque las proporciones iniciales de colores que hay no son iguales.

Cálculos: En cada terna la posición nos dice en qué extracción se obtuvo ese color, pero para escribir las posiciones también podemos utilizar subíndices, de manera que definimos:

B_i == La bola i-ésima es blanca
N_i == La bola i-ésima es negra
R_i == La bola i-ésima es roja

donde i = 1, 2 y 3. Con esta otra notación, se pueden reescribir los sucesos de interés como:

$$A=B_1\cap B_2\cap B_3 \quad \text{y} \quad B=\{B_1\cap B_2\cap N_3\}\cup\{B_1\cap N_2\cap B_3\}\cup\{N_1\cap B_2\cap B_3\}.$$

Tenemos ahora que:

$$P(A)=P(B_1\cap B_2\cap B_3)=P(B_1)\cdot P(B_2|B_1)\cdot P(B_3|B_2\cap B_1)=\frac{8}{15}\cdot\frac{7}{14}\cdot\frac{6}{13}=0{,}12.$$

$$\begin{aligned}
P(B)&=P(B_1\cap B_2\cap N_3)+P(B_1\cap N_2\cap B_3)+P(N_1\cap B_2\cap B_3)\\
&= P(B_1)\cdot P(B_2|B_1)\cdot P(N_3|B_2\cap B_1)\\
&\quad + P(B_1)\cdot P(N_2|B_1)\cdot P(B_3|N_2\cap B_1)\\
&\quad + P(N_1)\cdot P(B_2|N_1)\cdot P(B_3|B_2\cap N_1)\\
&= \frac{8}{15}\cdot\frac{7}{14}\cdot\frac{5}{13}+\frac{8}{15}\cdot\frac{5}{14}\cdot\frac{7}{13}+\frac{5}{15}\cdot\frac{8}{14}\cdot\frac{7}{13}=3\cdot\frac{5\cdot 7\cdot 8}{13\cdot 14\cdot 15}=\frac{7\cdot 8}{13\cdot 14}=\frac{4}{13}=0{,}31.
\end{aligned}$$

Comentario

Al no haber reemplazo, da igual en este caso extraer las bolas de una en una o las tres a la vez; cuando hay reemplazo la situación de la urna es siempre la misma al ir a extraer la siguiente bola. Cuando vamos a extraer una bola y calculamos la probabilidad de un color, lo que hacemos es aplicar la regla de Laplace: número de casos favorables entre número de posibles.

Segunda forma de resolverlo

Experimento: El experimento consiste en extraer tres bolas de la urna, una tras otra sin reemplazo, o a la vez, da igual.

Truco: Supongamos que ponemos un pequeño número a las bolas, del 1 al 15, de manera que tienen ahora un color y un número.

Espacio muestral: El espacio muestral, si atendemos al número, es ahora el conjunto de ternas (*a,b,c*) que se pueden formar con números distintos del 1 al 15. Cada terna de números es equiprobable, mientras que no lo era cada terna de letras.

Sucesos de interés: *A* estará compuesto por todas las ternas posibles formadas con números de bolas blancas. *B* se compondrá de todas las ternas posibles formadas con dos números de los de las bolas blancas y uno de los de las negras.

Características: *A* es un suceso elemental, mientras que *B* se compone de tres sucesos elementales (incompatibles entre sí).

Cálculos: Ahora vamos a aplicar la sencilla regla de Laplace, puesto que cada terna de bolas sí es ahora, con los números, equiprobable.

$CP ==$ Número de subconjuntos de 3 elementos que se pueden formar con 15 bolas

$$= \quad C_{15}^3 = \binom{15}{3} = \frac{15!}{3!\cdot(15-3)!} = \frac{15\cdot 14\cdots 1}{(3\cdot 2\cdot 1)\cdot(12\cdot 11\cdots 1)} = \frac{15\cdot 14\cdot 13}{(3\cdot 2\cdot 1)} = 5\cdot 7\cdot 13 .$$

Empezamos con A:

$CF ==$ Número de ternas que podemos crear con los números de las bolas blancas

$$= \quad C_8^3 = \binom{8}{3} = \frac{8!}{3!\cdot(8-3)!} = \frac{8\cdot 7\cdots 1}{(3\cdot 2\cdot 1)\cdot(5\cdot 4\cdots 1)} = \frac{8\cdot 7\cdot 6}{(3\cdot 2\cdot 1)} = 8\cdot 7 .$$

Por tanto

$$P(A) = \frac{CF}{CP} = \frac{8\cdot 7}{5\cdot 7\cdot 13} = \frac{8}{5\cdot 13} = 0{,}12 .$$

Para *B* la cosa se complica más:

- Hay 5 bolas negras
- Hay $C_8^2 = \binom{8}{2} = \frac{8!}{2!\cdot(8-2)!} = \frac{8\cdot 7\cdots 1}{(2\cdot 1)\cdot(6\cdot 5\cdots 1)} = \frac{8\cdot 7}{2} = 4\cdot 7$ formas de elegir 2 bolas blancas de entre sus 8.

Por tanto, $CF = 5\cdot 4\cdot 7$ y, de nuevo por la regla de Laplace,

$$P(B)=\frac{CF}{CP}=\frac{5\cdot 4\cdot 7}{5\cdot 7\cdot 13}=\frac{4}{13}=0{,}31\,.$$

Ejercicio 9

Una determinada raza de perros tiene cuatro cachorros en cada camada. Si la probabilidad de que un cachorro sea macho es de 0,55, se pide:

(a) La probabilidad de que en una camada dos exactamente sean hembras

(b) La probabilidad de que en una camada al menos dos sean hembras

Solución sin utilizar variables aleatorias

Experimento: Observar el sexo de los cuatro cachorros de una camada y anotar si es macho o hembra.

Espacio muestral: Los posibles resultados son todas las posibles formas de colocar, en un vector de tamaño cuatro una de las dos letras M (macho) o H (hembra); es decir:

Si no hay hembras $\rightarrow$ (M, M, M, M)

Si hay una hembra $\rightarrow$ (H, M, M, M), (M, H, M, M), (M, M, H, M) y (M, M, M, H)

Si hay dos hembras $\rightarrow$ (H, H, M, M), (H, M, H, M) $\cdots$ (M, M, H, H)

Si hay tres hembras $\rightarrow$ (M, H, H, H), (H, M, H, H), (H, H, M, H) y (H, H, H, M)

Si hay cuatro hembras $\rightarrow$ (H, H, H, H)

Yendo con el orden necesario se puede escribir el espacio muestral entero. En total, como en cada posición puede haber dos letras, habría $2\cdot 2\cdot 2\cdot 2 = 2^4 = 16$ casos posibles.

Suceso: Estamos interesados en los sucesos

(a) {Exactamente dos hembras} = {(H, H, M, M), (H, M, H, M) $\cdots$ (M, M, H, H)}

(b) {Al menos dos hembras} = {Exactamente dos hembras} U
U {Exactamente tres hembras} U
U {Exactamente cuatro hembras}

Cálculos: No se puede aplicar la regla de Laplace, puesto que todos los cuartetos de letras no tienen la misma probabilidad. Hay que ir caso por caso. Para el segundo apartado se considera el suceso complementario, lo que simplifica los cálculos:

(a) P{Exactamente dos hembras} = P{(H, H, M, M), (H, M, H, M) $\cdots$ (M, M, H, H)}
= P{(H, H, M, M)} + P{(H, M, H, M)} + $\cdots$ + P{(M, M, H, H)}

= 0,45·0,45·0,55·0,55 + 0,45·0,55·0,45·0,55 + ⋯ + 0,55·0,55·0,45·0,45

= 6·$0,45^2$·$0,55^2$ = 0,37

(b) P{Al menos dos hembras} = 1 - P{Ninguna hembra o una hembra}

= 1 - P{(M, M, M, M), (H, M, M, M), (M, H, M, M), (M, M, H, M), (M, M, M, H)}

= 1 - ($0,55^4$ + 4·0,45·$0,55^3$) = 0,61

Solución utilizando variables aleatorias

Experimento: Observar el sexo de los cuatro cachorros de una camada y contar el número de hembras. Por tanto, se define la variable

X ≡ Número de hembras en la camada

Espacio muestral: Los posibles valores que puede tomar la variable aleatoria; es decir:

Si no hay hembras → $X = 0$

Si hay una hembra → $X = 1$

Si hay dos hembras → $X = 2$

Si hay tres hembras → $X = 3$

Si hay cuatro hembras → $X = 4$

Suceso: Estamos interesados en los sucesos

(a) {Exactamente dos hembras} = {$X = 2$}

(b) {Al menos dos hembras} = $\{X \geq 2\}$ = {$X = 2$} U {$X = 3$} U {$X = 4$}

Cálculos: La variable que hemos definido tiene distribución $X \sim B(n=4, p=0{,}45)$, por lo que:

(a) $P(X=2)=\binom{4}{2}\cdot 0{,}45^2\cdot 0{,}55^2=\frac{4!}{2!\cdot(4-2)!}\cdot 0{,}2025\cdot 0{,}3025=0{,}3675$

(b) $P(X\geq 2)=P(X=2)+P(X=3)+P(X=4)$

= $\binom{4}{2}0{,}45^2 0{,}55^2+\binom{4}{3}0{,}45^3 0{,}55^1+\binom{4}{4}0{,}45^4 0{,}55^0$

= $0{,}3675+4\cdot 0{,}55^1\cdot 0{,}45^3+1\cdot 0{,}45^4\cdot 1$

= $0{,}3675+0{,}2005+0{,}04100=0{,}61$

Comentarios

- La misma solución se obtiene utilizando la variable

  Y ≡ Número de machos en la camada

  y reescribiendo la solución convenientemente para ella.

- Las variables aleatorias con distribución binomial son «contadores» que ya tienen en cuenta todas las posibles posiciones en que pueden darse los casos. Esto hace que no sea necesario que lo hagamos nosotros «a mano», como hemos tenido que hacer en la primera solución.
- Utilizar una variable aleatoria para estudiar el experimento ha facilitado mucho la solución.

Ejercicio 10

Sabiendo que la variable aleatoria X sigue una distribución binomial con n = 3, y que P(X = 0) = 0,3486:

a) Determinar el valor de p

b) Hallar P(X > 1)

Basado en un ejercicio del libro:

Problemas de cálculo de probabilidades y estadística
Vicente Novo Sanjurjo
UNED, 1993

a) $P(X=0)=\binom{3}{0}\cdot p^0\cdot q^3=q^3=0{,}3486 \quad \rightarrow \quad q=\sqrt[3]{0{,}3486}=0{,}7038$

Si $q = 0{,}7038$, entonces $p = 0{,}2962$.

b) $P(X>1)=P(X=2)+P(X=3)$

$$=\binom{3}{2}\cdot 0{,}2962^2\cdot 0{,}7038^1+\binom{3}{3}\cdot 0{,}2962^3\cdot 0{,}7038^0=0{,}1852+0{,}0260=0{,}2112 \quad .$$

Ejercicio 11

Una empresa lleva a cabo una prueba para seleccionar nuevos empleados. Por la experiencia de pruebas anteriores, se sabe que las puntuaciones siguen una distribución normal de media 80 y desviación típica 25. ¿Qué porcentaje de candidatos obtendrá entre 75 y 100 puntos?

$X \equiv$ Puntuación

$X \sim N(\mu=80, \sigma=25)$

Queremos calcular $P(75 \leq X \leq 100)$. Para ello primero hacemos la siguiente manipulación de desigualdades (que más adelante haremos dentro de la cadena de expresiones *P( ) = P( ) = ...*):

$$75 \leq X \leq 100$$

$$75-80 \leq X-80 \leq 100-80$$

$$(75-80)/25 \leq (X-80)/25 \leq (100-80)/25$$

$$(75-80)/25 \leq Z \leq (100-80)/25$$

Es decir, tenemos que

$$P(75 \leq X \leq 100) = P((75-80)/25 \leq Z \leq (100-80)/25)$$

Como (por un teorema ya mencionado en la teoría) Z~N(0,1), podemos ir a la tabla y hacer

$$P((75-80)/25 \leq Z \leq (100-80)/25) = P(Z \leq (100-80)/25) - P(Z \leq (75-80)/25)$$

$$= 0{,}788 - 0{,}421 = 0{,}367$$

Si convertimos esta cantidad a tanto por ciento obtenemos el 36,7.

Ejercicio 12

Tras realizar una encuesta se supo que, de las personas que se muestran de acuerdo con la posibilidad de suprimir el servicio militar, la edad media es de treinta y dos años, y la desviación típica es de doce. Suponiendo que se trata de una distribución normal, se pide calcular:

1. *La probabilidad de encontrar una persona de treinta años o menos que esté de acuerdo.*
2. *Probabilidad de encontrar una persona entre veinte y treinta y cinco años acorde a la desaparición del servicio militar.*
3. *Probabilidad de localizar a uno, con la misma disposición, de cuarenta años o menos.*
4. *Probabilidad de localizar a uno de cuarenta años o más.*

Ejercicio del libro

Estadística aplicada a las Ciencias Sociales: Ejercicios resueltos

Marisa García de Cortázar
José-María Arribas
Consuelo del Val
Luis Camarero
Antonio Vallejos
UNED, 1ªed (1996) 8ªreimp (2004)
(Problema 3.21.)

Apartado 1

$$P(X \leq 30) = P\left(\frac{X-32}{12} \leq \frac{30-32}{12}\right) = P(Z \leq -1/6) = P(Z \geq 1/6)$$
$$= \ 1 - P(Z < 1/6) = 1 - P(Z < 0{,}16) = 1 - 0{,}564 = 0{,}436$$

Apartado 2

$$P(20 \leq X \leq 30) = P\left(\frac{20-32}{12} \leq \frac{X-32}{12} \leq \frac{35-32}{12}\right) = P(-1 \leq Z \leq 1/4) = P(Z \leq 1/4) - P(Z \leq -1)$$
$$= \ P(Z \leq 0{,}25) - P(Z \geq 1) = P(Z \leq 0{,}25) - [1 - P(Z < 1)] = 0{,}599 - 0{,}159 = 0{,}44$$

Apartado 3

$$P(X \leq 40) = P\left(\frac{X-32}{12} \leq \frac{40-32}{12}\right) = P(Z \leq 2/3) = P(Z \leq 0{,}66) = 0{,}745$$

Apartado 4

$$P(X \geq 40) = 1 - P(X \leq 40) = 1 - 0{,}745 = 0{,}255$$

Ejercicio 13

La variable aleatoria X sigue una distribución $N(16, \sqrt{8})$, *entonces:*

a) Obténganse $P(17 \leq X \leq 18)$ *y* $P(14 \leq X \leq 15)$

b) Calcular los valores de n y p tales que la distribución B(n,p) podría aproximarse utilizando la distribución $N(16, \sqrt{8})$

Basado en un ejercicio del libro:

Problemas de cálculo de probabilidades y estadística
Vicente Novo Sanjurjo
UNED, 1993

a) $P(17 \leq X \leq 18) = P\left(\frac{17-16}{\sqrt{8}} \leq Z \leq \frac{18-16}{\sqrt{8}}\right) = P(0{,}35 \leq Z \leq 0{,}70)$

$$= \quad P(Z \leq 0{,}70) - P(Z < 0{,}35) = 0{,}758 - 0{,}637 = 0{,}121$$

$$P(14 \leq X \leq 15) = P\left(\frac{14-16}{\sqrt{8}} \leq Z \leq \frac{15-16}{\sqrt{8}}\right) = P(-0{,}70 \leq Z \leq -0{,}35)$$

$$= \quad P(Z \leq -0{,}35) - P(Z < -0{,}70) = P(Z \geq 0{,}35) - P(Z > 0{,}70)$$

$$= \quad [1 - P(Z < 0{,}35)] - [1 - P(Z \leq 0{,}70)] = 0{,}758 - 0{,}637 = 0{,}121$$

b) Por el teorema que se menciona en Aproximación de la binomial a la normal, para *n* grande, y cuando tanto *p* como *q* no estén próximos a cero, la distribución *B(n, p)* se puede aproximar mediante una distribución normal $N(np, \sqrt{npq})$. Así, en el problema podemos plantear un sistema de ecuaciones:

np = 16

npq = 8

Tenemos un sistema de dos ecuaciones y tres variables, por lo que necesitamos una tercera ecuación, que será:

p = 1 - q

Y despejando, obtenemos que *n = 32* y que *p = 0,5 =1/2 = q, p*or lo que podemos concluir que la $B(32, 0{,}5)$ se podrá aproximar mediante la $N(16, \sqrt{8})$.

Ejercicio 14

Después de realizar varios sondeos sobre una población con escasa cultura, se ha conseguido averiguar que únicamente el 15 % de la misma es favorable a los tratamientos de psicoterapia. Elegida al azar una muestra de 50 personas de dicha población, se desea saber:

A) La probabilidad de que haya más de 5 personas favorables a dichos tratamientos.

B) La probabilidad de que a lo sumo haya 6 personas favorables.

Forma exacta de resolver el ejercicio (con la binomial)

X ≡ Número de personas favorables de entre las 50 entrevistadas

X~ B(n=50, p=0,15)

Las probabilidades en este caso se calculan con:

A) $P(X>5)=1-P(X\leq 5)$

$$= \quad 1-[P(X=0)+P(X=1)+P(X=2)+P(X=3)+P(X=4)+P(X=5)]$$

$$= \quad 1-\left[\binom{50}{0}\cdot 0{,}15^{0}\cdot 0{,}85^{50}+\binom{50}{1}\cdot 0{,}15^{1}\cdot 0{,}85^{49}+\cdots+\binom{50}{5}\cdot 0{,}15^{5}\cdot 0{,}85^{45}\right]$$

$$= \quad 1-[0{,}0002958+0{,}002610+0{,}01128+0{,}03186+0{,}06606+0{,}1072]$$

$$= \quad 1-0{,}2194=0{,}7806$$

B) $P(X\leq 6)=P(X\leq 5)+P(X=6)=P(X\leq 5)+\binom{50}{6}\cdot 0{,}15^{6}\cdot 0{,}85^{44}$

$$= \quad 0{,}2194+0{,}1419=0{,}3613$$

Comentarios

En el primer apartado se calcula la probabilidad del suceso como uno menos la de su complementario, que es más cómodo. En el segundo apartado se aprovechan los cálculos del primero.

Forma aproximada de resolver el ejercicio (con la normal)

$X\equiv$ Número de personas favorables de entre las 50 entrevistadas

X~ B(n=50, p=0,15)

Como $n\cdot p>5,$ se puede utilizar una distribución normal para aproximar a la binomial. Es decir, tendríamos que hacer

$$Y\sim N(\mu=n\cdot p\,,\sigma=\sqrt{n\cdot p\cdot q})$$

donde $\mu=n\cdot p=50\cdot 0{,}15=7{,}5$ y $\sigma=\sqrt{50\cdot 0{,}15\cdot 0{,}85}=2{,}52.$

También está la variable

$$Z=(Y-\mu)/\sigma\sim N(0{,}1)$$

Es decir, podemos aproximar *X* por *Y* que, tipificada para poder consultar las tablas, se convierte en *Z*. Se dejan indicados los cálculos hasta el punto en el que habría que ir a las tablas de la *N(0,1)* para buscar las probabilidades. Nótese que en los pasos hay tanto igualdades estrictas como aproximaciones (que se denotan por $\simeq$).

<u>Sin corrección por continuidad</u> (primero, para comprender mejor la aproximación):

A) $P(X>5)\simeq P(Y>5)=P\left(\frac{Y-\mu}{\sigma}>\frac{5-\mu}{\sigma}\right)=P\left(Z>\frac{5-\mu}{\sigma}\right)=P\left(Z>\frac{5-7,5}{2,52}\right)$

$$= \quad P\left(Z>\frac{-2,5}{2,52}\right)=P(Z>-0,9921)=P(Z<0,9921)=0,839$$

B) $P(X\leq 6)\simeq P(Y\leq 6)=P\left(\frac{Y-\mu}{\sigma}\leq\frac{6-\mu}{\sigma}\right)=P\left(Z\leq\frac{6-\mu}{\sigma}\right)=P\left(Z\leq\frac{6-7,5}{2,52}\right)=P\left(Z\leq\frac{-1,5}{2,52}\right)$

$$= \quad P(Z\leq -0,5952)=P(Z\geq 0,5952)=1-P(Z<0,5952)=1-0,722=0,278$$

Dado que se está aproximando una distribución discreta (la binomial) por una continua (la normal), el resultado mejora si se aplica la llamada corrección por continuidad. Entonces *X* es discreta e *Y* y *Z* son continuas; la corrección se le ha de aplicar a *X* justo antes de pasar a *Y*.

Con corrección por continuidad los cálculos serían:

A) $P(X>5)=P(X>5+0,5)\simeq P(Y>5,5)=P\left(\frac{Y-\mu}{\sigma}>\frac{5,5-\mu}{\sigma}\right)=P\left(Z>\frac{5,5-\mu}{\sigma}\right)$

$$= \quad P\left(Z>\frac{-2}{2,52}\right)=P(Z>-0,7937)=P(Z<0,7937)=0,785$$

B) $P(X\leq 6)=P(X\leq 6+0,5)\simeq P(Y\leq 6,5)=P\left(\frac{Y-\mu}{\sigma}\leq\frac{6,5-\mu}{\sigma}\right)=P\left(Z\leq -\frac{1}{2,52}\right)$

$$= \quad P(Z\leq -0,3968)=P(Z\geq 0,3968)=1-P(Z<0,3968)=1-0,652=0,348$$

Comentarios

- Vemos el efecto de la corrección: estas últimas cantidades son mucho más cercanas a las de la binomial que a las de la aproximación sin corrección. Cuando no hay corrección, en el primer caso la distribución normal «empieza a sumar» área/probabilidad en el valor 5, no en el 5,5, lo que hace que añada una cantidad que le corresponde al 5, no al 6; por eso se obtiene un resultado mayor que el correcto. Para el otro suceso, si no hay corrección la distribución normal «deja de contar» el área entre el 6 y el 6,5, que le corresponde al valor 6, lo que hace que se obtenga un valor menor que el correcto. Por orden de mayor a menor exactitud, las soluciones se ordenan como:

  binomial $>$ normal con corrección $>$ normal sin corrección.

- Es interesante observar el paso siguiente, que hemos tenido que aplicar en dos ocasiones:

$$P(Z\leq -t)=P(Z\geq t)=1-P(Z<t).$$

Ejercicio 15

La nota media de las pruebas de acceso correspondientes a los estudiantes que querían ingresar en una facultad era 5,8 y la desviación típica 1,75. Fueron admitidos los de nota superior a 6.

a)¿Cuál fue el porcentaje de admitidos si la distribución es normal?

b)¿Con qué probabilidad exactamente cuatro de diez estudiantes son admitidos?

Hay situaciones en las que pueden aparecer involucradas distribuciones distintas. El siguiente ejercicio se puede resolver utilizando primero la normal y luego la binomial:

a) Del enunciado sacamos la información siguiente:

$X \equiv$ Nota media y $X \sim N(5{,}8; 1{,}75)$

Suceso de interés: {Admitido} = {X > 6}

$$P(X>6)=P\left(\frac{X-5{,}8}{1{,}75}>\frac{6-5{,}8}{1{,}75}\right)=P\left(Z>\frac{6-5{,}8}{1{,}75}\right)=P(Z>0{,}11)=1-P(Z\leq 0{,}11)$$

$$= \quad 1-0{,}05438=0{,}4562$$

que da 45,62% como solución.

b) Para el segundo apartado podemos imaginar ahora que cada persona puede ser admitida o no con esa probabilidad; es decir, cada persona lleva asociada una variable de Bernoulli:

$Y_i \equiv$ ¿Admitido?

$Y_i \sim Bern(0{,}4562)$

de manera que

$K \equiv$ Número de admitidos entre diez casos analizados

$K \sim B(10; 0{,}4562)$

Lo que nos preguntan es

$$P(K=4)=\binom{10}{4}\cdot 0{,}4562^4\cdot(1-0{,}4562)^6=0{,}235$$

Universidad Complutense de Madrid
∟ Facultad de Ciencias Económicas y Empresariales
∟ Departamento de Estadística e Investigación Operativa II
∟ David Casado de Lucas

15 de febrero del 2012