VARIABLES ALEATORIAS

Julián de la Horra
Departamento de Matemáticas U.A.M.

1 Introducción

Desde un punto de vista formal, las variables aleatorias son la herramienta matemática adecuada para transportar un modelo de probabilidad de un espacio muestral cualquiera a la recta real. En este sentido, las variables aleatorias constituyen un caso particular de las funciones medibles que se estudian en Teoría de la Medida.

Por otro lado, desde un punto de vista más aplicado, las variables aleatorias proporcionan la modelización necesaria para poder hacer Inferencia Estadística, en el siguiente sentido:

Cuando disponemos de una muestra de datos sobre alguna característica numérica (pesos de personas, envergaduras de aves, índices de contaminación, tiempos de vida útil de una pieza, ...) que nos interese estudiar en una población, podemos hacer representaciones gráficas y resúmenes numéricos con estos datos, que pueden resultar muy interesantes para poder manejar y visualizar, de manera sencilla, ese conjunto de datos. Ahora bien, si queremos ir más allá, e intentar obtener (inferir) conclusiones sobre la población en general, a partir de esa muestra particular de datos, necesitamos, en primer lugar, modelizar la característica numérica que estamos estudiando. Las variables aleatorias van a proporcionar el punto de partida para esa modelización.

2 Variables aleatorias

Comenzamos con la definición formal de variable aleatoria:

Definición.- Consideramos un espacio de probabilidad (Ω, S, P). Una *variable aleatoria*, X, es una función $X : \Omega \to \Re$ que verifica:

$X^{-1}(A) \in S$, para todo suceso A de la σ-álgebra de Borel en $\Re$. •

A continuación, veremos que con esta definición se cumple el objetivo formal de trasladar perfectamente un modelo de probabilidad de un espacio muestral cualquiera, Ω, a la recta real:

Teorema.- Consideramos un espacio de probabilidad (Ω, S, P) y una variable aleatoria, X. Definimos:

$P_X(A) = P(X^{-1}(A))$, para todo suceso A de la σ-álgebra de Borel en $\Re$.

Entonces, la función P_X es un modelo de probabilidad sobre la recta real, $\Re$, dotada de la σ-álgebra de Borel.

Demostración.- Evidentemente, $P_X(A) \in [0,1]$, para todo suceso A de la σ-álgebra de Borel en $\Re$. Veamos que se cumplen las dos propiedades que tiene que verificar cualquier modelo de probabilidad:
(a) $P_X(\Re) = P(X^{-1}(\Re)) = P(\Omega) = 1$.
(b) Si $A_1, \ldots, A_n, \ldots, ...$ son sucesos disjuntos de Borel, tenemos:
$P_X(\bigcup_n A_n) = P(X^{-1}(\bigcup_n A_n)) = P(\bigcup_n X^{-1}(A_n)) = \sum_n P(X^{-1}(A_n)) = \sum_n P_X(A_n)$ •

Con la construcción anterior, hemos definido un modelo de probabilidad sobre la recta real, a partir de una variable aleatoria. Además, recordemos que siempre hay una función de distribución asociada a un modelo de probabilidad sobre la recta real. Todo esto, motiva las siguientes definiciones.

Definiciones.- Consideramos un espacio de probabilidad (Ω, S, P) y una variable aleatoria, X. El modelo de probabilidad P_X, definido sobre la recta real por

$$P_X(A) = P(X^{-1}(A)),$$

recibe el nombre de *modelo de probabilidad generado o inducido por la variable aleatoria X.*

La función de distribución F_X definida como

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

recibe el nombre de *función de distribución generada o inducida por la variable aleatoria X.* •

Lo único que realmente nos va a interesar de una variable aleatoria es el modelo de probabilidad que genera sobre la recta real. Es decir, el modelo de probabilidad del que partimos, P, no nos va a interesar en absoluto, una vez que dispongamos del modelo de probabilidad, P_X, que ha generado la variable aleatoria sobre la recta real. Ilustramos esto con un ejemplo sencillo:

Ejemplo.- Consideramos el experimento aleatorio consistente en 3 lanzamientos sucesivos de una moneda equilibrada. El espacio muestral de este sencillo experimento es $\Omega = \{(C,C,C),(C,C,Cr),(C,Cr,C),(Cr,C,C),$ $(C,Cr,Cr),(Cr,C,Cr),(Cr,Cr,C),(Cr,Cr,Cr)\}$. Consideramos partes de Ω como σ-álgebra y, en este caso, todos los sucesos elementales son equiprobables, es decir, la probabilidad de cualquier terna de resultados es $1/8$.

Definimos ahora la variable aleatoria:

$X=$ "Número total de caras obtenidas en los 3 lanzamientos"

Los valores posibles de esta variable aleatoria son 0, 1, 2 y 3, y la probabilidad inducida sobre cada uno de ellos es:

$P_X(0) = P\{(Cr,Cr,Cr)\} = 1/8$

$P_X(1) = P\{(C,Cr,Cr),(Cr,C,Cr),(Cr,Cr,C)\} = 3/8$

$P_X(2) = P\{(C,C,Cr),(C,Cr,C),(Cr,C,C)\} = 3/8$

$P_X(3) = P\{(C,C,C)\} = 1/8$

Una vez que tenemos claramente construida la probabilidad inducida sobre $\Re$, nos podemos olvidar del espacio muestral de partida y de su modelo de probabilidad. •

Finalmente, recordando que la σ-álgebra de Borel en $\Re$ está generada por la clase $C = \{(-\infty, x]\}$, podemos ver el siguiente resultado teórico, que será de utilidad más adelante:

Teorema.- Consideramos un espacio de probabilidad (Ω, S, P) y una función $X : \Omega \to \Re$. Entonces:

X es variable aleatoria si y sólo si $X^{-1}(-\infty, x] \in S$, para todos los intervalos $(-\infty, x]$.

Demostración.- La implicación hacia la derecha es inmediata. Para probar la implicación hacia la izquierda, definimos las clases siguientes:

$$C = \{(-\infty, x]\} \qquad H = \{A : X^{-1}(A) \in S\}$$

Obviamente, $C \subset H$, y comprobar que H verifica las dos propiedades de toda σ-álgebra también es inmediato. Por lo tanto, H es σ-álgebra, $C \subset H$, de modo que la σ-álgebra de Borel está contenida en H. •

A continuación, nos vamos a centrar en el estudio de los dos tipos más importantes de variables aleatorias: discretas y continuas.

3 Variables aleatorias discretas

Definición.- Una variable aleatoria, X, es de *tipo discreto*, cuando el modelo de probabilidad que genera sobre la recta real reparte la probabilidad sobre un conjunto finito o numerable de elementos, $\{x_1, \ldots, x_n, \ldots\}$.

Una variable aleatoria discreta queda caracterizada por su función de masa:

$P_X(x_1) = P(X = x_1), \ldots, P_X(x_n) = P(X = x_n), \ldots$

la cual, obviamente, cumple las dos características que tiene que verificar una función de masa:

(a) $P(X = x_n) \in [0, 1]$, para todo x_n.

(b) $\sum_n P(X = x_n) = 1$. $\bullet$

Las variables aleatorias discretas proporcionan una modelización adecuada para todos aquellos experimentos aleatorios que tienen un número finito o numerable de posibles resultados: número de caras en varios lanzamientos de una moneda, número de crias por camada en una especie de mamíferos, número de siniestros anuales de cierto tipo en una compañía de seguros, ...

La función de distribución de una variable aleatoria discreta es constante a trozos:

$$F_X(x) = P_X(-\infty, x] = P(X \leq x) = \begin{cases} 0 & \text{para } x < x_1 \\ P_X(x_1) & \text{para } x \in [x_1, x_2) \\ P_X(x_1) + P_X(x_2) & \text{para } x \in [x_2, x_3) \\ \ldots & \ldots \end{cases}$$

donde estamos suponiendo que los valores $x_1, \ldots, x_n, \ldots$ están ordenados de menor a mayor. En el caso finito, es decir, si x_n es el último valor posible, entonces $F_X(x) = 1$, para $x \geq x_n$.

Las siguientes definiciones nos proporcionan resúmenes numéricos de los valores centrales de la característica que hemos modelizado mediante la variable aleatoria X. Se suelen llamar medidas de centralización.

Definición.- La *media o esperanza* de una variable aleatoria discreta X se define como:

$$\mu = E[X] = \sum_n x_n P(X = x_n) \qquad \bullet$$

Definición.- La idea intuitiva para definir la mediana de una variable aleatoria es que buscamos el valor que deja la misma cantidad de probabilidad a un lado y a otro. Esta idea se formaliza de la siguiente manera:

La *mediana* de una variable aleatoria X es cualquier valor M que verifica:

$$\begin{aligned} F_X(M^-) = \lim_{h \to 0} F_X(M-h) &\leq 1/2 \\ F_X(M) &\geq 1/2 \end{aligned}$$

En el caso de una variable aleatoria de tipo discreto, la función de distribución es constante a trozos (como vimos antes), y tenemos dos posibilidades para la mediana:

(a) Si ninguno de los peldaños de la función de distribución vale 1/2, la mediana M es única, y será el valor x_i en el que la función de distribución pasa de valer menos de 1/2 a valer más de 1/2.

(b) Si uno de los peldaños de la función de distribución vale 1/2 en el intervalo $[x_i, x_{i+1})$, la mediana M no es única, y será cualquier valor del intervalo $[x_i, x_{i+1}]$. •

Las siguientes definiciones nos proporcionan resúmenes numéricos de la dispersión de los valores de la característica que hemos modelizado mediante la variable aleatoria X. Se suelen llamar medidas de dispersión.

Definiciones.- La *varianza* de una variable aleatoria discreta X se define como:

$$\begin{aligned} \sigma^2 = V(X) &= \sum_n (x_n - E[X])^2 P(X = x_n) = \dots \\ &= \sum_n x_n^2 P(X = x_n) - (E[X])^2 = E[X^2] - (E[X])^2 \end{aligned}$$

La *desviación típica* de una variable aleatoria discreta X se define como la raíz cuadrada (positiva) de la varianza, y sirve para medir la dispersión en las mismas unidades que la característica que se está estudiando. •

La esperanza y la varianza de una variable aleatoria son casos particulares (muy importantes) de unos conceptos más generales:

Definiciones.- El *momento no centrado de orden k* de una variable aleatoria discreta X se define como:

$$E[X^k] = \sum_n x_n^k P(X = x_n)$$

El *momento centrado de orden k* de una variable aleatoria discreta X se define como:

$$E[(X - E[X])^k] = \sum_n (x_n - E[X])^k P(X = x_n) \qquad \bullet$$

4 Variables aleatorias continuas

Definición.- Una variable aleatoria, X, es de *tipo continuo*, cuando el modelo de probabilidad que genera sobre la recta real reparte la probabilidad mediante una función de densidad, que es una función $f_X : \Re \to \Re$ que verifica las dos siguientes propiedades:

(a) $f_X(x) \geq 0$, para todo $x \in \Re$.

(b) $\int_{\Re} f_X(x)dx = 1$.

La probabilidad de cualquier suceso A, se obtiene de la siguiente forma:

$$P_X(A) = P(X \in A) = \int_A f_X(x)dx \qquad \bullet$$

Las variables aleatorias continuas proporcionan una modelización adecuada para todos aquellos experimentos aleatorios cuyos posibles resultados son cualquier valor dentro de un intervalo finito o infinito: peso de una persona, alzada de una especie de mamíferos, envergadura de una especie de aves, tiempo de vida útil de un tipo de piezas, ...

La función de distribución de una variable aleatoria continua es continua (como su nombre hacía suponer). Para comprobar la continuidad de la función de distribución, lo único que tenemos que probar es que se trata de una función continua por la izquierda, ya que la continuidad por la derecha es una propiedad común a todas las funciones de distribución. Para cualquier $x \in \Re$ y $h > 0$, tenemos:

$$\begin{aligned} F_X(x) &= P(X \leq x) = P(X \leq x-h) + P(x-h < X \leq x) \\ &= F_X(x-h) + \int_{x-h}^{x} f_X(x)dx \end{aligned}$$

Tomamos límites y tenemos:

$$F_X(x) = \lim_{h\to 0} F_X(x-h) + \lim_{h\to 0} \int_{x-h}^{x} f_X(x)dx = \lim_{h\to 0} F_X(x-h) + 0 = F_X(x^-) \ ,$$

quedando así probada la continuidad por la izquierda.

Las siguientes definiciones nos proporcionan (igual que en el caso discreto) resúmenes numéricos de los valores centrales de la característica que hemos modelizado mediante la variable aleatoria X, y también se suelen llamar medidas de centralización.

Definición.- La *media o esperanza* de una variable aleatoria continua X se define como:

$$\mu = E[X] = \int_{\Re} x f_X(x)dx \qquad \bullet$$

Definición.- La idea intuitiva para definir la mediana de una variable aleatoria es que buscamos el valor que deja la misma cantidad de probabilidad a un lado y a otro. Esta idea se formaliza de la siguiente manera::

La *mediana* de una variable aleatoria X es cualquier valor M que verifica:

$$\begin{aligned} F_X(M^-) = \lim_{h \to 0} F_X(M-h) &\leq 1/2 \\ F_X(M) &\geq 1/2 \end{aligned}$$

En el caso de una variable aleatoria de tipo continuo, la función de distribución es continua (como vimos antes), y la definición de la mediana se simplifica bastante:

La *mediana* de una variable aleatoria X de tipo continuo es cualquier valor M que verifica:

$$F_X(M) = \int_{-\infty}^{M} f_X(x)dx = 1/2$$

En el caso continuo, la mediana suele ser única. •

Las siguientes definiciones nos proporcionan (igual que en el caso discreto) resúmenes numéricos de la dispersión de los valores de la característica que hemos modelizado mediante la variable aleatoria X, y también se suelen llamar medidas de dispersión.

Definiciones.- La *varianza* de una variable aleatoria continua X se define como:

$$\begin{aligned} \sigma^2 = V(X) &= \int_{\Re} (x - E[X])^2 f_X(x)dx = \ldots \\ &= \int_{\Re} x^2 f_X(x)dx - (E[X])^2 = E[X^2] - (E[X])^2 \end{aligned}$$

La *desviación típica* de una variable aleatoria continua X se define como la raíz cuadrada (positiva) de la varianza, y sirve para medir la dispersión en las mismas unidades que la característica que se está estudiando. •

La esperanza y la varianza de una variable aleatoria son casos particulares (muy importantes) de unos conceptos más generales:

Definiciones.- El *momento no centrado de orden k* de una variable aleatoria continua X se define como:

$$E[X^k] = \int_{\Re} x^k f_X(x)dx$$

El *momento centrado de orden k* de una variable aleatoria discreta X se define como:

$$E[(X - E[X])^k] = \int_{\Re} (x - E[X])^k f_X(x) dx \qquad \bullet$$

5 Desigualdades y funciones de interés

En primer lugar vamos a ver la siguiente desigualdad probabilística:

Desigualdad de Chebychev.- Sea X una variable aleatoria y sea $g : \Re \to \Re$ una función no negativa. Entonces, para todo $\varepsilon > 0$, tenemos:

$$P\{x : g(x) \geq \varepsilon\} \leq \frac{E[g(X)]}{\varepsilon}$$

Demostración.- Utilizaremos la notación de una variable aleatoria continua. Si es discreta, la demostración es análoga.

$$\begin{aligned} E[g(X)] &= \int_{\Re} g(x) f_X(x) dx \\ &= \int_{\{x:g(x)\geq\varepsilon\}} g(x) f_X(x) dx + \int_{\{x:g(x)<\varepsilon\}} g(x) f_X(x) dx \\ &\geq \int_{\{x:g(x)\geq\varepsilon\}} g(x) f_X(x) dx \\ &\geq \varepsilon \int_{\{x:g(x)\geq\varepsilon\}} f_X(x) dx \\ &= \varepsilon\, P\{x : g(x) \geq \varepsilon\} \qquad \bullet \end{aligned}$$

A continuación, veremos una de las aplicaciones más utilizadas de la desigualdad de Chebychev, que nos va a permitir expresar la relación práctica entre la esperanza y la varianza de una variable aleatoria:

Corolario.- Sea X una variable aleatoria con $E[X] = \mu$ y $V(X) = \sigma^2$. Entonces, para cualquier $k > 0$, tenemos:

$$P\{x : |x - \mu| \geq k\sigma\} \leq \frac{1}{k^2}$$

Demostración.- Consideramos $g(x) = (x - \mu)^2$, y tenemos:

$$\begin{aligned} P\{x : |x - \mu| \geq k\sigma\} &= P\{x : (x - \mu)^2 \geq k^2\sigma^2\} \\ &\leq \frac{E[(X - \mu)^2]}{k^2\sigma^2} = \frac{\sigma^2}{k^2\sigma^2} = \frac{1}{k^2} \qquad \bullet \end{aligned}$$

Por ejemplo, tomando $k = 2$, tenemos:

$$P\{x : |x-\mu| \geq 2\sigma\} \leq \frac{1}{4}$$

Es decir, la probabilidad de obtener valores que se alejen más de dos desviaciones típicas de la media es, en cualquier caso, menor o igual que 0,25.

La principal ventaja de esta desigualdad es que vale para cualquier variable aleatoria que tenga esperanza y varianza. El principal inconveniente es que, al servir para todas las variables aleatorias, es una desigualdad poco afinada, en general.

A continuación, definimos y estudiamos dos funciones auxiliares que son muy útiles, tanto para algunos cálculos que surgen con las funciones de densidad, como para definir más adelante algunos modelos de probabilidad:

Definiciones.- La *función gamma* se define, para $\alpha > 0$, como

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$$

La *función beta* se define, para $\alpha > 0$ y $\beta > 0$, como

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx \qquad \bullet$$

Algunas propiedades de las funciones gamma y beta:

(a) Para $\alpha > 1$, $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$.

En efecto, aplicando la integración por partes tenemos:

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx = \int_0^\infty (x^{\alpha-1})(e^{-x} dx) = \int_0^\infty u dv,$$

siendo $u = x^{\alpha-1}$ y $dv = e^{-x} dx$.

Tenemos:

$u = x^{\alpha-1} \Rightarrow \quad u' = (\alpha-1)x^{\alpha-2} = \frac{du}{dx} \Rightarrow \quad du = (\alpha-1)x^{\alpha-2} dx$

$dv = e^{-x} dx \quad \Rightarrow \quad v = \int dv = \int e^{-x} dx = -e^{-x}$

Por tanto:

$$\begin{aligned}
\Gamma(\alpha) &= \int_0^\infty x^{\alpha-1}e^{-x}dx = \int_0^\infty udv = uv - \int_0^\infty vdu \\
&= \left[-e^{-x}x^{\alpha-1}\right]_0^\infty - \int_0^\infty (-e^{-x})(\alpha-1)x^{\alpha-2}dx \\
&= 0 + \int_0^\infty e^{-x}(\alpha-1)x^{\alpha-2}dx \\
&= (\alpha-1)\int_0^\infty e^{-x}x^{\alpha-2}dx = (\alpha-1)\Gamma(\alpha-1)\ .
\end{aligned}$$

(b) $\Gamma(1) = 1$.

En efecto: $\Gamma(1) = \int_0^\infty e^{-x}dx = [-e^{-x}]_0^\infty = 1$.

(c) Para $\alpha = 2, 3, ...$, tenemos: $\Gamma(\alpha) = (\alpha-1)!$

En efecto: $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1) = (\alpha-1)(\alpha-2)\Gamma(\alpha-2) = ... = (\alpha-1)!$

(d) Las funciones *gamma* y *beta* están relacionadas de la siguiente forma:

$$B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}\ .$$

En efecto, tenemos:

$$\begin{aligned}
\Gamma(\alpha)\Gamma(\beta) &= \left[\int_0^\infty t^{\alpha-1}e^{-t}dt\right]\left[\int_0^\infty y^{\beta-1}e^{-y}dy\right] = \int_0^\infty t^{\alpha-1}e^{-t}\left[\int_0^\infty y^{\beta-1}e^{-y}dy\right]dt \\
&= \int_0^\infty t^{\alpha-1}e^{-t}\left[\int_0^\infty (tx)^{\beta-1}e^{-xt}tdx\right]dt \quad \text{(definiendo } x = \tfrac{y}{t}\text{, para } t \text{ fijo)} \\
&= \int_0^\infty x^{\beta-1}\left[\int_0^\infty e^{-t(1+x)}t^{\alpha+\beta-1}dt\right]dx \quad \text{(permutando orden de integración)} \\
&= \int_0^\infty \frac{x^{\beta-1}}{(1+x)^{\alpha+\beta}}\left[\int_0^\infty e^{-z}z^{\alpha+\beta-1}dz\right]dx \quad \text{(def. } z = t(1+x)\text{, para } x \text{ fijo)} \\
&= \int_0^\infty \frac{x^{\beta-1}}{(1+x)^{\alpha+\beta}}\Gamma(\alpha+\beta)dx = \Gamma(\alpha+\beta)\int_0^\infty \left(\frac{x}{1+x}\right)^\beta \frac{1}{x(1+x)^\alpha}dx \\
&= \Gamma(\alpha+\beta)\int_0^1 u^{\beta-1}(1-u)^{\alpha-1}du \quad \text{(definiendo } u = \tfrac{x}{1+x}\text{)} \\
&= \Gamma(\alpha+\beta)\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx \quad \text{(definiendo } x = 1-u\text{)} \\
&= \Gamma(\alpha+\beta)B(\alpha,\beta)
\end{aligned}$$

(e) $B(1/2, 1/2) = \pi$.

En efecto:

$$\begin{aligned} B(1/2,1/2) &= \int_0^1 x^{-1/2}(1-x)^{-1/2}dx = \int_0^1 \frac{1}{\sqrt{x}\sqrt{1-x}}dx \\ &= \int_0^1 \frac{2}{\sqrt{1-y^2}}dy \qquad \text{(definiendo } y = \sqrt{x}\text{)} \\ &= [2\arcsin y]_0^1 = \pi \end{aligned}$$

(f) $\Gamma(1/2) = \sqrt{\pi}$.

En efecto, sabemos que:

$$B(1/2,1/2) = \frac{\Gamma(1/2)\Gamma(1/2)}{\Gamma(1)} = [\Gamma(1/2)]^2 = \pi \ .$$

Por tanto: $\Gamma(1/2) = \sqrt{\pi}$.