14. VARIABLES ALEATORIAS

14.1 Definición de variable aleatoria

Dado un experimento aleatorio, con un espacio muestral asociado E, una variable aleatoria, es cualquier función, X, que asocia a cada suceso elemental un número real

$$\begin{aligned} X : E &\to \mathbb{R} \\ w &\to X(w) \end{aligned}$$

verificando que, para cualquier número real r, el conjunto dado por los sucesos elementales cuyas imágenes son menores que r sea un suceso.

Propiedades:

1. Si X es una variable aleatoria definida sobre un espacio muestral E y c es una constante cualquiera, cX es una variable aleatoria sobre el mismo espacio.

2. Si X e Y son dos variables aleatorias su suma $X+Y$ y su producto $X \cdot Y$ son también variables aleatorias.

Ejemplo 1: Dado el experimento aleatorio consistente en el lanzamiento de dos dados, de seis caras equiprobables, y sea la variable, X, que asocia cada elemento muestral, par (i,j), con su suma $i+j$.

$$E = \{(1,1),(1,2),...,(6,6)\}, \quad X(i,j) = i+j$$

Los posibles valores de la variable aleatoria X son los de la siguiente tabla:

i/j	**1**	**2**	**3**	**4**	**5**	**6**
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

14.1.1 Funcion de distribución de una variable aleatoria

Función de distribución de una variable aleatoria, es una función real, que a cualquier número real x, le asocia la probabilidad de que la variable tome valores menores o iguales a dicho número, esto es, si X es la variable aleatoria:

$$F(x) = P(X \leq x) = P(w \in E \text{ tal que } X(w)) \leq x)$$

Propiedades:
1. $0 \leq F(x) \leq 1$
2. F es no decreciente.
3. $F(+\infty) = 1$
4. $F(-\infty) = 0$
5. F es continua por la derecha.

14.2 Variables aleatorias discretas

Variable aleatoria discreta, es aquella que sólo puede tomar valores dentro de un conjunto finito o infinito numerable. Sea X una variable discreta que tome los valores $x_1, x_2, \ldots, x_n, \ldots$ Al conjunto de probabilidades $\{p_1, p_2, \ldots, p_n, \ldots\}$, de forma que $p_i = P(X = x_i)$, con $\sum p_i = 1$, se le denomina *función de masa de probabilidad o función de probabilidad* de la variable aleatoria discreta X.

La función de distribución para una variable discreta es una función real de variable real, que asocia a cada número x la probabilidad acumulada hasta ése valor. Está dada como suma de la función de masa de probabilidad, del siguiente modo:

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} p_i$$

La función F es escalonada, no decreciente, con saltos de discontinuidad en los puntos x_i iguales a la probabilidad en dichos puntos p_i.

Ejemplo 2: En el experimento aleatorio consistente en lanzar dos dados, se definió la variable $X(i, j) = i + j$, cuya función de masa de probabilidad será:

$$\begin{aligned}
p_1 &= P(X = 2) = P(\{(1,1)\}) = \frac{1}{36} \\
p_2 &= P(X = 3) = P(\{(1,2),(2,1)\}) = \frac{2}{36} \\
p_3 &= P(X = 4) = P(\{(2,2),(1,3),(3,1)\}) = \frac{3}{36} \\
&\vdots \qquad\qquad \vdots \\
p_{11} &= P(X = 12) = P(\{(6,6)\}) = \frac{1}{36}.
\end{aligned}$$

Obsérvese que:

$$\sum_i p_i = \frac{1}{36} + \frac{2}{36} + \frac{3}{36} + \ldots + \frac{1}{36} = 1.$$

$$F(x) = P(X \leq x) = \begin{cases} 0 & \text{si} & x < 2 \\ 1/36 & \text{si} & 2 \leq x < 3 \\ 3/36 & \text{si} & 3 \leq x < 4 \\ 6/36 & \text{si} & 4 \leq x < 5 \\ 10/36 & \text{si} & 5 \leq x < 6 \\ 15/36 & \text{si} & 6 \leq x < 7 \\ 21/36 & \text{si} & 7 \leq x < 8 \\ 26/36 & \text{si} & 8 \leq x < 9 \\ 30/36 & \text{si} & 9 \leq x < 10 \\ 33/36 & \text{si} & 10 \leq x < 11 \\ 35/36 & \text{si} & 11 \leq x < 12 \\ 1 & \text{si} & x \geq 12 \end{cases}$$

14.3 Variables aleatorias continuas

Una *variable aleatoria es continua*, si toma valores en uno o varios intervalos de la recta real.

Son ejemplos de variables aleatorias continuas: “hora de llegada de un profesor a su despacho”,“tiempo de duración de una llamada telefónica”, “cantidad de lluvia caída por metro cuadrado en una determinada zona”, “superficie de planchas metálicas producidas en una factoría”, etc.

Dada una variable aleatoria continua X, *la función de densidad* es la función real de variable real que resulta de derivar la función de distribución, esto es, $f(x) = F'(x)$.

La función de densidad f describe el comportamiento idealizado de la variable aleatoria continua asociada, reflejando para cada intervalo real, sobre el que tome valores la variable, su densidad de probabilidad.

Geométricamente $F(x_0)$ mide el área de la región limitada por la función de densidad, el eje X, y la recta de ecuación $x = x_0$.

Propiedades de la función de densidad

1. $f(x) \geq 0, \quad -\infty < x < \infty$.

2. El área bajo la curva será 1, pues representa la suma de todas las probabilidades asociadas al experimento aleatorio, y puesto que $P(E) = 1$, se tendrá esta propiedad para $f : \int_{-\infty}^{\infty} f(x)dx = 1$

3. La probabilidad del intervalo $[a, b]$ será el área bajo la curva $f(x)$ entre las rectas $x = a$ y $x = b$.

$$P(a \leq X \leq b) = \int_a^b f(t)dt = F(b) - F(a)$$

4. La probabilidad de que una variable continua tome un único valor es nula.

$$P(X = x) = \int_x^x f(t)dt = 0$$

5. En general cualquier función real que verifica las propiedades 1 y 2, será una función de densidad.

Ejemplo 3: Dada una variable aleatoria continua cuya función de densidad viene dada por

$$f(x) = \begin{cases} \frac{3}{8}x^2 & \text{si} \quad 0 \leq x \leq 2 \\ 0 & \text{si} \quad x \notin (0,2) \end{cases}$$

su función de distribución es

$$F(x) = \begin{cases} 0 & \text{si} \quad x < 0 \\ \frac{x^3}{8} & \text{si} \quad 0 \leq x \leq 2 \\ 1 & \text{si} \quad x > 2 \end{cases}$$

La probabilidad de que la variable tome valores entre $1'5$ y 2, está dada por

$$P(1'5 < X < 2) = \int_{1'5}^2 \frac{3}{8}t^2 dt = 0'578$$

14.4 Características de una variable aleatoria

14.4.1 Esperanza matemática de una variable aleatoria

Sea X una variable aleatoria discreta con valores $x_1, x_2, ..., x_n, ...$ y función de masa de probabilidad $p_1, p_2, ..., p_n, ...$ La *media o esperanza matemática* de la vaiable aleatoria X es el número real

$$E(X) = \sum_i x_i p_i.$$

Para el caso de una variable aleatoria continua X con función de densidad $f(x)$ su media o esperanza es el número real

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx.$$

Propiedades de la media:

1. $E(aX + b) = aE(X) + b$

2. $E(X + Y) = E(X) + E(Y)$

Ejemplo 4: Valor esperado de la variable X, suma de las caras al lanzar dos dados:

$$E(X) = \sum_{x=2}^{12} x \cdot P(X = x) = 2\frac{1}{36} + 3\frac{2}{36} + \cdots + 12\frac{1}{36} = 7.$$

Media para la variable aleatoria continua cuya función de densidad está dada por:

$$f(x) = \begin{cases} 1 & \text{si} \quad 8 \leq x \leq 9 \\ 0 & \text{si} \quad x \notin [8, 9] \end{cases}$$

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_{8}^{9} xdx = 8'5$$

14.4.2 Varianza de una variable aleatoria

Sea X una variable aleatoria con media $E(X)$, la *varianza* de X es el valor esperado de los cuadrados de las diferencias con la media:

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

Esta medida de dispersión tiene el inconveniente de no dar los resultados en las mismas unidades que los datos, por lo que se define la *desviación típica* de la variable X como la raíz positiva de la varianza:

$$\sigma = +\sqrt{E[(x - E(X))^2]}$$

presentando la ventaja de dar los resultados en las mismas unidades que los valores de la variable.

Propiedades de la varianza:

1. $Var(aX + b) = a^2 Var(X)$

2. $Var(X) = E(X^2) - E(X)^2$ (fórmula reducida para el cálculo de la varianza)

14.4.3 Tipificación de una variable aleatoria

Una variable aleatoria se dirá que está *estandarizada o tipificada*, si su media es 0 y su varianza es 1.

Para transformar una variable X con media μ y desviación típica σ en otra tipificada, basta con aplicar un cambio de variable de tipo lineal, $Y = \dfrac{X-\mu}{\sigma}$. Esta nueva variable, Y, evidentemente, tiene por media 0 y por varianza 1.

El interés de tipificar una variable se verá en el capítulo 6, donde se estudian variables como la Normal, que por su importancia está tabulada, siendo necesario tipificarla para el cálculo de probabilidades usando la tabla.

14.4.4 Otras medidas características de una variable aleatoria

La varianza es una medida de dispersión que no es invariante a cambios de escala, por lo que resulta útil construir otra medida que mida la dispersión y permanezca inalterada si se produce un cambio en la escala de los datos, así definiremos:

Coeficiente de variación para la variable X, con media μ y desviación típica σ :

$$C.V.(X) = \frac{\sigma}{\mu},$$

siempre que $\mu > 0$.

Medidas de simetría y de apuntamiento

Para medir la forma de una distribución se usa el *coeficiente de asimetría de Fisher* que se define como el cociente,

$$\gamma_1 = \frac{E(X-\mu)^3}{n\sigma^3}.$$

En distribuciones simétricas o con una asimetría muy pequeña se utiliza otra medida, que sirve para comparar la distribución con una distribución conocida y muy utilizada: la variable Normal; definiendo el *coeficiente de apuntamiento o curtosis de Fisher* como el número

$$\gamma_2 = \frac{E(X-\mu)^4}{n\sigma^4} - 3$$

Ejemplo 5: Vamos a calcular la media, varianza y coeficiente de variación para la distribución continua que tiene por función de densidad

$$f(x) = \begin{cases} 0'1 & \text{si} \quad 0 \leq x \leq 10 \\ 0 & \text{en otro caso} \end{cases}$$

Media y varianza

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_{0}^{10} \frac{1}{10}xdx = \ 5$$

$$\sigma^2 = E\left((x-5)^2\right) = \int_{0}^{10} \frac{1}{10}(x-5)^2 dx = 8'3333$$

Coeficiente de variación

$$C.V. = \frac{\sigma}{E(x)} = 0'57735$$

14.5 Principales distribuciones unidimensionales discretas

14.6 Distribución Binomial

Supongamos que se realiza un experimento con dos posibles resultados correspondientes a la ocurrencia o no de un suceso. Llamaremos éxito (E) a la ocurrencia y fracaso $(\overline{E})$ a la no ocurrencia, siendo p la probabilidad de éxito y q $(q = 1-p)$ la probabilidad de fracaso.

La variable aleatoria, X, que estudia el número de éxito en n pruebas sigue una distribución Binomial siempre que el experimento tenga las siguientes características:

- El experimento consta de n pruebas idénticas donde la probabilidad de éxito permanece constante en cada prueba.

- Las pruebas son independientes.

La variable aleatoria X que estudia el "número de éxitos en las n realizaciones del experimento" sigue una distribución binomial de parámetros n y p que se denota como $B(n,p)$.

Su función de probabilidad es:

$$P(X = x) = \binom{n}{x} p^x q^{n-x}.$$

Características:

$$E(X) = np$$

$$Var(X) = npq$$

Ejemplo 6: Se seleccionan, de manera aleatoria, 15 unidades de un proceso de manufactura con el propósito de verificar el porcentaje de unidades defectuosas en la producción. Con base en información pasada, la probabilidad de tener una unidad defectuosa es $0'05$. La gerencia ha decidido detener la producción cada vez que una muestra de 15 unidades tenga dos o más defectuosas. ¿Cuál es la probabilidad de que, un día determinado, el proceso se detenga?

Si X es la variable que representa el nba de unidades defectuosas que se encuentran entre las 15 seleccionadas, $X \sim B(15, 0'05)$.
La probabilidad de que la producción se detenga es igual a la probabilidad de que X sea mayor o igual que dos:

$$\begin{aligned} P(X &\geq 2) = 1 - P(X = 0) - P(X = 1) \\ &= 1 - \begin{pmatrix} 15 \\ 0 \end{pmatrix} 0.05^0 0.95^{15} - \begin{pmatrix} 15 \\ 1 \end{pmatrix} 0.05^1 0.95^{14} = 0.1709. \end{aligned}$$

14.7 Distribución normal

La distribución normal o de Gauss es la más importante y de mayor uso de todas las distribuciones continuas de probabilidad. Hay multitud de experimentos que se ajustan a esta distribución, por ejemplo: el peso, la talla, errores de medición, calificaciones de pruebas de aptitud, etc.

Se dice que una variable aleatoria se encuentra distribuida normalmente con media 0 y desviación típica 1, N(0,1), si su distribución viene dada por la función de densidad:

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

La variable aleatoria N(0,1) se denomina normal tipificada o estandarizada, y las representaremos usualmente por la letra Z. Esta distribución se encuentra tabulada.
Una variable aleatoria tiene una distribución normal de media μ y desviación típica σ si su función de densidad:

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

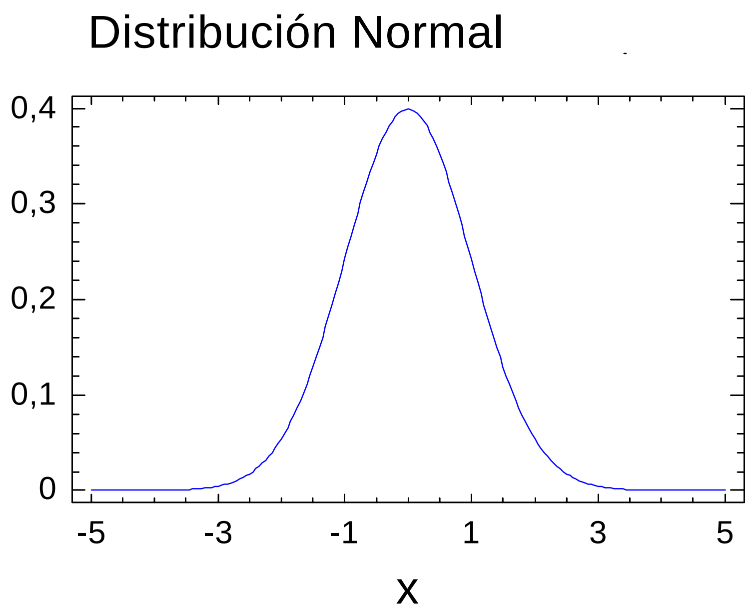

Figura 14-1 Función de densidad de la variable N(0,1)

Puesto que la distribución N(0,1) está tabulada, lo que haremos para calcular las probabilidades de una variable $X \sim N(\mu, \sigma)$ es tipificar la variable para obtener una distribución N(0,1).

Características:

$$E(X) = \mu$$

$$Var(X) = \sigma^2$$

Propiedades:

Si $X \sim N(\mu_1, \sigma_1)$ e $Y \sim N(\mu_2, \sigma_2)$, son variables independientes entonces $X + Y \sim N(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$.

Ejemplo 7: Si el peso de una persona se distribuye según una $N(62, 18)$, ¿cuál es la probabilidad de que 30 personas superen el 2000 kg de peso?

Por la aditividad de la Normal el peso de las 2000 personas seguirá una $N(62 \cdot 2000, \sqrt{2000} \cdot 18)$.

Por tanto $P(X > 2000) = P\left(Z > 1'42\right) = 0'0778$.

14.7.1 Teorema Central del Límite

El Teorema Central del Límite establece que cuando los resultados de un experimento son debidos a un conjunto muy grande de causas independientes que actúan sumando sus efectos, dichos resultados siguen una distribución normal.

14.7.2 Relación entre la binomial y normal

Sea $X \sim B(n,p)$, siempre que $n > 30$ y $npq > 5$, consideraremos la aproximación de la binomial por una $N(np, \sqrt{npq})$.

Para cometer un menor error cuando se trata de aproximar una distribución discreta por una continua, si nos piden la $P(a \leq X \leq b)$, con X distribución discreta, la aproximación se puede hacer calculando $P(a - 0'5 \leq X \leq b + 0'5)$, con X distribución normal.

Ejemplo 8: ¿Qué probabilidad se tiene de obtener más de 60 caras en el lanzamiento de cien tiradas de una moneda no sesgada?

Sea X_i la variable que toma el valor uno si se obtiene cara en el lanzamiento i y 0 si se obtiene una cruz en dicho lanzamiento, $X_i \sim B(1, 1/2) \;\; \forall i = 1, 2, ..., n$.
$E(X_i) = 1/2 \;\; \forall i$
$Var(X_i) = 1/4 \;\; \forall i$

Sea X la variable que mide el número de caras en 100 tiradas, se sigue que X se ajusta a una $B(100, 1/2)$ que podemos aproximar por una:

$$N(50, \sqrt{100 \cdot \frac{1}{2} \cdot \frac{1}{2}}) = N(50, 5)$$

Por tanto:

$$P(X > 60) = P\left(Z > \frac{60 - 50}{5}\right) = P(Z > 2) = 0.0228.$$