

Variables Aleatorias

Nos interesa *asignar probabilidades* a valores numéricos obtenidos a partir de fenómenos aleatorios, es decir a *variables aleatorias*.

Por ejemplo, calcular la probabilidad de que al elegir una muestra de tamaño 20 de un lote que contiene 5% de artículos defectuosos por lo menos 7 artículos de la muestra resulte defectuoso. ¿Cuál es el experimento aleatorio ε , el espacio muestral Ω , la familia de sucesos en Ω , la función de probabilidad P , el suceso de interés y la variable aleatoria en este ejemplo?

Formalmente, dados un experimento aleatorio ε , un espacio muestral Ω , una familia de sucesos en Ω con una probabilidad P , diremos que una función

$X: \Omega \rightarrow \mathbb{R}$, es una variable aleatoria si

$X^{-1}(-\infty, x] = \{ \omega \in \Omega : X(\omega) \leq x \}$ es un suceso

Esta definición asegura que sea posible el cálculo de probabilidades sobre las variables aleatorias. No será necesario utilizar esta definición en lo que sigue!

Veremos dos tipos de variables aleatorias, discretas y continuas, que nos permitirán utilizar la teoría de probabilidades para realizar inferencias estadísticas.

Variables Aleatorias Discretas

Si el conjunto de resultados posibles, R_X , es finito ó infinito numerable, decimos que X es una **v.a. discreta**.

Sea X una v.a. discreta. Indicaremos por $R_X = \{x_i, i \in \text{Naturales}\}$ al conjunto de todos sus valores posibles. Un modelo de probabilidad para X está dado por la asignación de probabilidades, p_i , a estos resultados,

$$P(X=x_i) = p_X(x_i) = p_i .$$

Las probabilidades deben satisfacer

$$1. \ 0 \leq p_i \leq 1$$

$$2. \ \sum p_i = 1$$

La probabilidad $P(X \text{ esté en } A)$ de cualquier suceso se calcula sumando las p_i de los resultados que componen A .

Los valores p_i determinan la *distribución de probabilidades* de la v.a. X mediante la función de *probabilidad puntual* (f.p.p. p_X) que asigna a cada x_i una probabilidad p_i .

Ejemplo: Un profesor califica sus pruebas en una escala de 4 puntos (1, 2, 3, 4). Supongamos que en un curso de 30 alumnos los resultados ordenados fueron:

1 1 1 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4 4 4

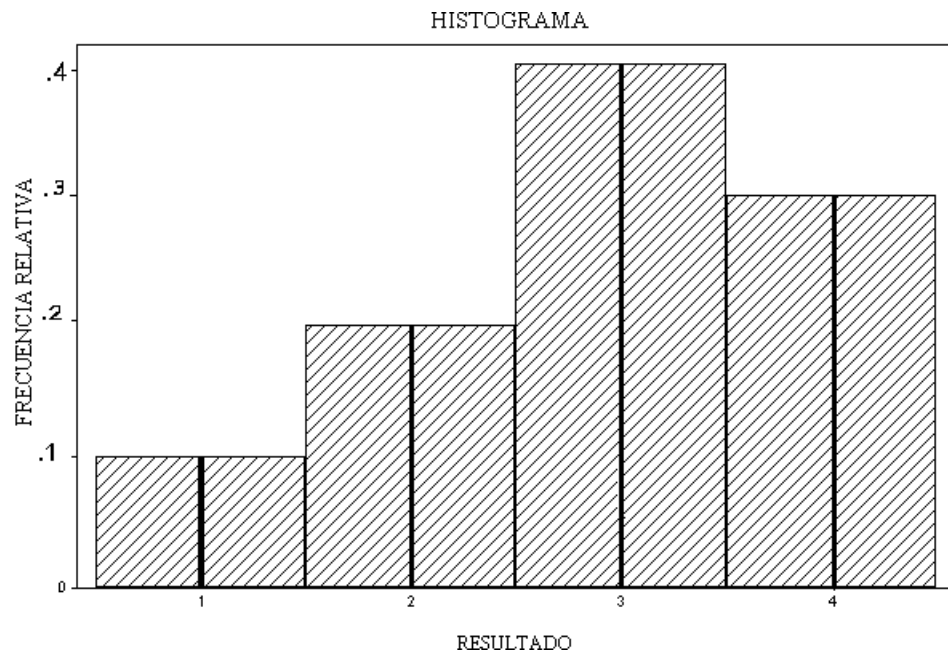
Sea X = resultado de la prueba para un alumno del curso elegido al azar.

Luego $R_X = \{1, 2, 3, 4\}$ X es **una variable aleatoria discreta**

La frecuencia de cada uno de los resultados es, 3, 6, 12 y 9 respectivamente.

resultado (x_i)	1	2	3	4
frecuencia (f_i)	3	6	12	9
frec. relativa (f_i/n)	.1	.2	.4	.3
$p_X(x_i) = p_i$	.1	.2	.4	.3

Diagrama de barras -- Histograma de Probabilidad



Siguiendo con el ejemplo:

¿Cuál es la probabilidad de que un alumno elegido al azar haya tenido a lo sumo un 3?

$$\begin{aligned}
 P(X \leq 3) &= p_X(1) + p_X(2) + p_X(3) \\
 &= p_1 + p_2 + p_3 = .1 + .2 + .4 = .7
 \end{aligned}$$

Si pensamos que los 30 alumnos constituyen la población en estudio, podemos calcular la media, μ , y la varianza, σ^2 , poblacionales de la variable, X (*resultado de la prueba*).

$$\begin{aligned}
 \mu &= \frac{1+1+1+2+2+2+2+2+2+2+2+\dots+4+4+4}{30} \\
 &= 1\frac{3}{30} + 2\frac{6}{30} + 3\frac{12}{30} + 4\frac{9}{30} \\
 &= 1p_X(1) + 2p_X(2) + 3p_X(3) + 4p_X(4)
 \end{aligned}$$

La media poblacional μ_X de una variable X es un promedio pesado de los valores posibles.

En general tenemos:

Si X es una v.a. discreta con valores posibles $R_X = \{x_i, i \text{ en los Naturales}\}$ y f.p.p. $p_X(x_i)$ entonces se define la Esperanza de X (μ_X)

$$E(X) = \sum_{x_i \in R_X} x_i p_X(x_i)$$

Volviendo al ejemplo: La varianza poblacional σ^2 es

$$\begin{aligned}
 \sigma^2 &= \frac{(1-\mu)^2 + (1-\mu)^2 + (1-\mu)^2 + \dots + (4-\mu)^2}{30} \\
 &= (1-\mu)^2 \frac{3}{30} + (2-\mu)^2 \frac{6}{30} + (3-\mu)^2 \frac{12}{30} + (4-\mu)^2 \frac{9}{30} \\
 &= (1-\mu)^2 p_X(1) + (2-\mu)^2 p_X(2) + (3-\mu)^2 p_X(3) + (4-\mu)^2 p_X(4)
 \end{aligned}$$

En general tenemos:

Si X es una v.a. discreta con valores posibles $R_X = \{x_i, i \text{ en los Naturales}\}$ y f.p.p. $p_X(x_i)$ entonces se define la Varianza de X (σ^2)

$$\text{Var}(X) = \sum_{x_i \in R_X} (x_i - \mu)^2 p_X(x_i)$$

Variables aleatorias continuas. Informalmente diremos que:

Una variable aleatoria es continua cuando el conjunto de sus valores posibles son todos los valores de un intervalo o de una unión de intervalos de números reales.

Por ejemplo, la concentración de cromo en el Riachuelo es una variable aleatoria continua.

La distribución de una variable aleatoria continua se describe mediante la *función de densidad de probabilidad*, ó simplemente función de densidad f_X .

La función de densidad, f_X , de una variable aleatoria X satisface:

$$1. f_X(x) \geq 0$$

$$2. \int f_X(x) dx = 1$$

$$3. P(X \in A) = \int_A f_X(x) dx$$

Si X es una v. a. **continua** cualquiera. ¿Cuánto vale la probabilidad de que $X = 160$?

¿Por qué?

Ejemplo: Las alturas de las mujeres jóvenes argentinas están aproximadamente distribuidas normalmente con $\mu = 160$ cm $\sigma = 4$ cm. Sea X = altura de una mujer argentina joven, elegida al azar, X es una v.a. continua y su función de densidad de probabilidad está dada por:

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

En general, si la función de densidad de probabilidad de una variable aleatoria X está dada por la expresión anterior, decimos que X tiene distribución normal de parámetros con μ y σ^2 :

$$X \sim N(\mu, \sigma^2)$$

Propiedades de la distribución Normal

Sea $X \sim N(\mu, \sigma^2)$, entonces

a) $Z = (X - \mu) / \sigma \sim N(0, 1)$ distribución Normal Estándar.

b) $aX + b \sim N(a\mu + b, a^2\sigma^2)$

Siguiendo con el ejemplo: ¿Cuál es la probabilidad de que una mujer joven elegida al azar tenga una altura entre 160 cm y 168 cm?

$$\begin{aligned} P(160 < X < 170) &= P(0 < (X - 160) / 4 < 2) = \Phi(2) - \Phi(0) \\ &= 0.9772 - 0.5 = 0.4772 \end{aligned}$$

Esperanza y varianza de una variable aleatoria continua

Si X es una variable aleatoria continua con función de densidad, f_X entonces la su media o esperanza ($E(X)$) está dada por:

$$\mu_X = \int_{-\infty}^{\infty} t f_X(t) dt$$

y su varianza

$$\sigma_X^2 = \int_{-\infty}^{\infty} (t - \mu_X)^2 f_X(t) dt$$

Compare con el caso discreto.

En base a las propiedades de la esperanza tendremos que $\text{Var}(X) = E[(X - \mu)^2]$

Propiedades de la Esperanza y la Varianza

a) Esperanza de una función de una variable aleatoria: sea $Y = h(X)$ entonces

$$E(Y) = \begin{cases} \sum_{i \geq 1} h(x_i) p_X(x_i) & \text{si } X \text{ es discreta} \\ \int h(x) f_X(x) dx & \text{si } X \text{ es continua} \end{cases}$$

b) Si $P(X = c) = 1$ entonces $E(X) = c$

c) $E(aX + b) = aE(X) + b$

d) $\text{Var}(aX + b) = a^2 \text{Var}(X)$

Función de Distribución acumulada.

La función de distribución acumulada de una variable aleatoria X , cualquiera, es:

$$F_X(x) = P(X \leq x)$$

Si X es una variable aleatoria discreta

$$F_X(x) = \sum_{\substack{x_i \in R_X \\ x_i \leq x}} p_X(x_i)$$

Si X es una variable aleatoria continua

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

Luego

$$F'_X(x) = f_X(x)$$

Propiedades

(i) $0 \leq F_X(x) \leq 1 \quad \forall x \in \mathbf{R}$	porque es una probabilidad
(ii) Es monótona no decreciente	porque es acumulada
(iii) Es continua a derecha	por la forma en que está definida
(iv) $\lim_{x \rightarrow +\infty} F_X(x) = 1$ $\lim_{x \rightarrow -\infty} F_X(x) = 0$	por (i) (ii)
(v) En cada punto x el valor del salto es la <i>probabilidad puntual</i> de ese punto: $\text{salto} = F_X(x) - F_X(x^-)$ $= p_X(x)$	

Ejemplo: Sea X = nota obtenida en una prueba de un alumno elegido al azar

nota (x)	1	2	3	4
$p_X(x)$	0.1	0.2	0.4	0.3

Función de distribución acumulada

$$F_X(x) = \begin{cases} 0 & \text{si } x < 1 \\ 0.1 & \text{si } 1 \leq x < 2 \\ 0.3 & \text{si } 2 \leq x < 3 \\ 0.7 & \text{si } 3 \leq x < 4 \\ 1 & \text{si } 4 \leq x \end{cases}$$

