Introducción a la Probabilidad y Estadística - Probabilidad y Estadística

Grisel Britos

2023

VARIABLE ALEATORIA

Introducción

Una variable aleatoria es una función que permite asociar valores del espacio muestral con números reales.

Ejemplos:

- *Experimento: Se arroja 3 veces una moneda honesta. Una variable aleatoria posible es la variable que cuenta el número de veces que salió cara en las 3 tiradas.*
- *Experimento: Se examinaron los tiempos de duración de una lista de canciones hasta encontrar la primera canción que dura exactamente 4 minutos. Una variable de interés podría ser "el número de canciones que se examinan hasta encontrar la primera canción que dura exactamente 4 minutos".*
- *Experimento: Se tomó un examen que dura a lo sumo 4 horas. Una posible variable a estudiar es el "tiempo que demora un alumno en entregar el examen".*

Definición: Sea $(\Omega, \mathcal{A}, P)$ un modelo probabilístico. Se llama variable aleatoria (v.a.) a cualquier función $X : \Omega \longrightarrow \mathbb{R}$ tal que $\{w \in \Omega : X(w) \leq x\} \in \mathcal{A}$ para todo $x \in \mathbb{R}$.

Notación: Para todo $x \in \mathbb{R}$,

$$[X \leq x] = \{w \in \Omega : X(w) \leq x\}$$
$$[X = x] = \{w \in \Omega : X(w) = x\}$$

Definición: dada X una v.a. entonces se llama **función de distribución acumulada** (f.d.a.) a la función $F : \mathbb{R} \longrightarrow [0, 1]$ definida como

$$F(x) = P([X \leq x]) \quad \forall x \in \mathbb{R}.$$

Clasificación de variables aleatorias:

DISCRETAS: número de caras en tres tiradas de una moneda honesta; número de accidentes en una autopista durante el fin de semana; número de camas ocupadas en la UTI de un hospital diariamente; número de focos de incendios que se iniciaron durante el verano 2020-2021 en las sierras de Córdoba; ...
CONTINUAS: la altura de niños recién nacidos; tiempo que demora un estudiante en resolver un parcial; cantidad de lluvia caída durante el verano; ...

VARIABLE ALEATORIA DISCRETA

Definición: Sea $(\Omega, \mathcal{A}, P)$ un m.p. y $X : \Omega \longrightarrow \mathbb{R}$ una v.a.. Se dice que X es una **variable aleatoria discreta** si toma un número finito o infinito numerable de valores con probabilidad positiva.

Observación: Si X toma sólo dos valores, por ejemplo 0 (fracaso) y 1 (éxito), con probabilidades positivas y $p = P(X = 1)$ entonces se dice que X tiene **distribución Bernoulli de parámetro** p.

Definición: Sea $(\Omega, \mathcal{A}, P)$ un m.p. y $X : \Omega \longrightarrow \mathbb{R}$ una v.a. discreta. Se llama **función de probabilidad de masa** (f.p.m.) a la función $p : \mathbb{R} \longrightarrow [0, 1]$ definida como

$$p(x) = P([X = x]) \quad \forall x \in \mathbb{R}.$$

Observación: Sea $(\Omega, \mathcal{A}, P)$ un m.p., $X : \Omega \longrightarrow \mathbb{R}$ una v.a. discreta y sea $Im(X) = \{a_i\}_{i\in I}$ (el conjunto de todos los valores posibles de la v. a. con probabilidad positiva) con $I \subseteq \mathbb{N}$ entonces

1. $\sum_{i\in I} p(a_i) = 1$
2. $F(a) = P(X \leq a) = \sum_{\substack{x\leq a \\ x\in Im(X)}} P(X = x) = \sum_{\substack{x\leq a \\ x\in Im(X)}} p(x) \quad \forall a \in \mathbb{R}$

Ejemplos: Hallar la f.p.m. y la f.d.a. para las siguientes v.a.

1) *Experimento: arrojar 3 veces una moneda honesta. Sea X la v.a. que cuenta el "número de veces que resultó cara en las 3 tiradas".*

2) *Experimento: se examinan memorias RAM en una línea de producción, independientemente una de otra, hasta obtener la primera que cumple las condiciones de calidad. Sea Y la v.a. que "cuenta el número de memorias RAM examinadas hasta obtener la primera que cumple las condiciones de calidad".*

Propiedades de la f.d.a. de una v.a. discreta

Sea $(\Omega, \mathcal{A}, P)$ un m.p., $X : \Omega \longrightarrow \mathbb{R}$ una v.a. discreta y F su f.d.a.

a) $0 \leq F(x) \leq 1 \quad \forall x \in \mathbb{R}$.

b) F es monótona creciente : $F(x) \leq F(y) \quad \forall x < y$.

c) $\lim_{x\to-\infty} F(x) = 0$ y $\lim_{x\to\infty} F(x) = 1$.

d) F es una función escalonada con discontinuidades en $Im(X)$.

e) F es continua por derecha, o sea

$$\lim_{x\to a^+} F(x) = F(a) \quad \forall a \in \mathbb{R}$$

f) Para todo $a \in \mathbb{R}$,

$$\lim_{x\to a^-} F(x) = P(X < a)$$

g) Si $Im(X) = \{a_i\}_{i\in I}$ tales que $a_i < a_{i+1}$ $\forall i \in I$, entonces

$$p(a_1) = F(a_1) \quad \text{y} \quad p(a_{i+1}) = F(a_{i+1}) - F(a_i)$$

Observación: Si X es una v.a. discreta con f.d.a. F entonces:

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

pero

$$P(a \leq X \leq b) = F(b) - \lim_{x \to a^-} F(x) \quad \forall a < b \quad \text{números reales}$$

Valores asociados a la f.p.m.

ESPERANZA

Definición: Sea X una v.a. discreta con $Im(X) = \{x_i\}_{i \in I}$ y p su f.p.m.. Se define el **valor esperado o valor medio o esperanza** de X como

$$E(X) = \sum_{i \in I} x_i p(x_i)$$

siempre que $\sum_{i \in I} |x_i| p(x_i) < \infty$.
Se denota a la esperanza como $\mu = E(X)$.

Se puede ver en la definición que la esperanza de X es un promedio ponderado de los valores que toma la variable X. La definición del valor esperado se basa en la interpretación frecuentista de las probabilidades. Esta interpretación asume que, si se lleva a cabo un gran número de repeticiones del experimento, la proporción de veces que ocurre el suceso A es igual a $P(A)$.

Proposición: Sea X una v.a. discreta con $Im(X) = \{x_i\}_{i \in I}$, $p : \mathbb{R} \to [0, 1]$ f.p.m. y sea $W = h(X)$ una v.a. discreta. Entonces

$$E(h(X)) = \sum_{i \in I} h(x_i) p(x_i)$$

siempre que $\sum_{i \in I} |h(x_i)| p(x_i) < \infty$.

Proposición: Sea X una v.a. discreta con $E(X) < \infty$. Entonces

$$E(aX + b) = aE(X) + b \quad \forall a, b \in \mathbb{R}$$

VARIANZA Y DESVIACIÓN ESTÁNDAR

Dado que uno espera que la variable aleatoria tome valores alrededor de su media $E(X)$, una forma razonable de medir la variación de X es considerar en qué medida X tiende a separarse de su media. Para esto se podría considerar $E(|X - \mu|)$, donde $\mu = E(X)$ y $|X - \mu|$ es el valor absoluto de la diferencia entre X y μ. Sin embargo, a los fines prácticos del cálculo, resulta más conveniente no considerar el valor absoluto sino el cuadrado de la diferencia.

Definición: Sea X una v.a. discreta con $Im(X) = \{x_i\}_{i \in I}$, f.p.m. p y $E(X) = \mu$. Entonces si $E(X^2) < \infty$, se define la **varianza** de X como:

$$V(X) = E(X - \mu)^2$$

y la **desviación estándar** de X como

$$\sqrt{V(X)}$$

Denotaremos a la varianza con $\sigma^2 = V(X)$ y al desvío estándar con $\sigma = \sqrt{V(X)}$.

Proposición: Sea X una v.a. discreta, p su f.p.m. y $E(X) = \mu$. Entonces si $E(X^2) < \infty$ se tiene

a) $0 \leq V(X) = E(X^2) - \mu^2$

b) $V(aX + b) = a^2 V(X)$ para todo $a, b \in \mathbb{R}$

Observación: Sea X una v. a. Bernoulli de parámetro p, donde $p = P(X = 1)$ y $1 - p = P(X = 0)$. Entonces $E(X) = p$ y $V(X) = p(1 - p)$.

Ejemplos:

1) *Experimento: arrojar 3 veces una moneda honesta. Sea X la variable que cuenta el número de veces que resultó cara en las 3 tiradas. Hallar la $E(X)$ y $V(X)$.*

2) *Experimento: se examinan memorias RAM en una línea de producción, independientemente una de otra, hasta obtener la primera que cumple las condiciones de calidad. Considerando que $p^* = P($una memoria RAM cumpla la condiciones de calidad$)$ y la v.a. Y definida como el número de memorias RAM examinadas hasta obtener la primera que cumple las condiciones de calidad. Hallar la $E(Y)$ y $V(Y)$.*

3) *Experimento: arrojar un dado honesto dos veces. Sea X la suma de los dos valores obtenidos. Hallar la $E(X)$ y $V(X)$.*

Distribución Binomial

Para definir una v.a. con distribución binomial se deben cumplir las siguientes condiciones experimentales:

i) El experimento consta de n pruebas o ensayos idénticos.

ii) Cada ensayo tiene sólo dos resultados posibles, que se llamaran éxito (E) o fracaso (F).

iii) Los ensayos son independientes uno de otro.

iv) La probabilidad de éxito es constante en cada ensayo

Un experimento que cumple con todas estas condiciones se lo llama experimento Binomial.

Definición: Consideremos un experimento Binomial que consta de n ensayos independientes y probabilidad de éxito $p = P(E)$ en cada uno. Si la v.a. X que cuenta el número de éxitos en n ensayos independientes, se dice que X tiene **distribución Binomial de parámetros n y p.**
Se denota $X \sim B(n, p)$.

Proposición: Si $X \sim B(n, p)$ entonces:

a)

$$p(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \forall x \in \{k \in \mathbb{Z} : 0 \leq k \leq n\} \\ 0 & \text{c.c.} \end{cases}$$

b) $E(X) = np$ y $V(X) = np(1 - p)$.

Ejemplos:

a) *Si X tiene distribución Bernoulli de parámetro p entonces $X \sim B(1, p)$.*

b) *Sea W la variable que cuenta el número de veces que resultó cara en las 3 tiradas de una moneda honesta entonces $W \sim B(3, 0{,}5)$.*

c) *Ver el enunciado de Ejercicio 13 de la Guía No2. Sea Y la variable que cuenta el número de remaches defectuosos en la costura de un avión. Justificar que esta v.a. tiene distribución Binomial e identifique sus parámetros. Resolver ahora el problema usando esta información.*

Distribución de Poisson

Definición: Se dice que una variable aleatoria X tiene **distribución de Poisson de parámetro λ** (con $\lambda > 0$) si la función de probabilidad de masa (f.p.m.) está dada por:

$$p(k) = \begin{cases} e^{-\lambda}\frac{\lambda^k}{k!} & \forall k \in \mathbb{N} \cup \{0\} \\ 0 & \text{c.c.} \end{cases}$$

Notación: $X \sim P(\lambda)$.

Proposición: Si $X \sim P(\lambda)$ entonces $E(X) = V(X) = \lambda$.

Aproximación de una Poisson por límite de Binomiales

Sean $X_n \sim B(n, p)$. Si se tiene que $np \longrightarrow \lambda$ cuando $n \longrightarrow \infty$ y $p \longrightarrow 0$ entonces:

$$\lim_{\substack{n\to\infty \\ p\to 0}} P(X_n = k) = e^{-\lambda}\frac{\lambda^k}{k!} \quad \forall k \in \mathbb{N} \cup \{0\}$$

Regla práctica para la aproximación binomial por la Poisson

En cualquier experimento binomial donde n sea grande y p pequeño, se puede aproximar

$$P(X_n = k) \cong e^{-\lambda}\frac{\lambda^k}{k!}$$

donde $\lambda = np$, $n \geq 100$, $p \leq 0{,}01$ y $np \leq 20$.

Ejemplo 4: Un editor de novelas se esfuerza por asegurar que sus libros estén libres de errores tipográficos, teniendo una probabilidad de contener por lo menos un error en una página de 0.005. Suponiendo que los errores son independientes de una página a otra,

i) *¿cuál es la probabilidad de que una de sus novelas de 400 páginas contenga exactamente una página con error? y*

ii) *¿por lo menos cuatro páginas con errores?*

Distribución Hipergeométrica

Para definir una v.a. con distribución hipergeométrica se debe cumplir las siguientes condiciones experimentales:

i) la población o conjunto de donde se obtiene la muestra tiene N objetos o individuos,

ii) cada objeto o individuo puede ser clasificado como Éxito (E) o Fracaso (F), siendo M los éxitos y $(N - M)$ los fracasos,

iii) cada subconjunto de n objetos o individuos tiene igual probabilidad de ser seleccionado.

Definición: Dada una población de tamaño N cuyos elementos se clasifican en M éxitos y $(N - M)$ fracasos. Si X es la variable aleatoria que cuenta el número de éxitos (E) en una muestra de tamaño n $(n \leq N)$ se dice que tiene **distribución Hipergeométrica de parámetros n, M y N**.
Se denota $X \sim H(n, M, N)$

Proposición: Si $X \sim H(n, M, N)$ entonces:

a) La f.p.m. de X está dada por

$$p(k) = \begin{cases} \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} & \{k \in \mathbb{N} : max(0, n+M-N) \leq k \leq min(M, n)\} \\ 0 & \text{c.c.} \end{cases}$$

b) $E(X) = n\left(\frac{M}{N}\right)$ y $V(X) = n\left(\frac{M}{N}\right)\left(1 - \frac{M}{N}\right)\left(\frac{N-n}{N-1}\right)$

Ejemplo 5: En la biblioteca de una facultad hay 20 ejemplares del libro "Probabilidad y estadística para ingeniería y ciencias" de Devore, de los cuales 8 son de la quinta edición y 12 de la sexta. El director del Departamento de estadística ha pedido que cinco ejemplares sean reservados por dos horas. Si las elecciones son al azar y cada subconjunto de cinco ejemplares es igualmente probable de ser elegido, entonces ¿cuál es la probabilidad que

i) *no haya ninguno de la sexta edición?*

ii) *exactamente uno sea de sexta edición?*

iii) *por lo menos dos sean de sexta edición?*

Ejemplo 6: una urna contiene 20 bolillas de las cuales 12 son negras y 8 blancas. Se extraen simultáneamente y sin reposición 5 bolillas de la urna y se cuentasn cuantas bolillas negras fueron extraidas. Calcular:

i) *¿Cuál es la probabilidad de que haya exactamente una bolilla negra*

ii) *¿Cuál es la probabilidad de que exactamente uno sea de sexta edición?*

iii) *¿ Cuál es la probabilidad de que en la muestra haya por lo menos dos bolillas negras?*

Aproximación a la Hipergeométrica por una Binomial

El uso de ésta aproximación es adecuada si el tamaño de la muestra (n) no excede el 5 % de la población (N) y tanto $\frac{M}{N}$ como $1 - \frac{M}{N}$ no sean próximos a 0. Entonces, bajo estas condiciones, si $X \sim H(n, M, N)$ se tiene

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

donde $p = \frac{M}{N}$ y $1 - p = 1 - \frac{M}{N}$ no son próximos a 0.

Ejemplo 7: Se ha estimado que para cierta población de 10000 individuos, 8000 personas presentan problemas de concentración. Se seleccionó al azar una muestra sin reposición de 5 personas de esta población. Se quiere calcular:

a) La probabilidad de que exactamente dos individuos de la muestra presenten problemas de concentración.

b) La probabilidad de que a lo sumo dos personas de la muestra presenten problemas de concentración.

Distribución Binomial Negativa

Para definir una v.a. con distribución binomial negativa se deben cumplir las siguientes condiciones experimentales:

i) consta de una secuencia de ensayos independientes,

ii) cada ensayo puede ser Éxito (E) o Fracaso (F),

iii) la probabilidad de E es constante en cada ensayo ($P(E) = p$),

iv) los ensayos se continúan hasta obtener r éxitos, con $r \in \mathbb{N}$ valor fijo.

Definición: Dado un experimento que cumpla las condiciones antes enumeradas, entonces la v.a. X que cuenta el número de fracasos que preceden al r- ésimo éxito se dice que tiene **distribución Binomial Negativa de parámetros r y p.**
Se denota $X \sim BN(r, p)$ o $X \sim B^{-}(r, p)$.

Proposición: Si $X \sim BN(r, p)$ entonces:

a) la f.p.m. de X está dada por

$$p(k) = \begin{cases} \binom{r+k-1}{r-1} p^r (1-p)^k & k \in \mathbb{N} \\ 0 & \text{c.c.} \end{cases}$$

b) $E(X) = \frac{r(1-p)}{p}$ y $V(X) = \frac{r(1-p)}{p^2}$

Ejemplos:

2) *El experimento consiste en examinar memorias RAM en una línea de producción, independientemente una de otra, hasta obtener la primera que cumple con las condiciones de calidad. Considerando el evento $E = \{$seleccionar una memoria que cumple las condiciones de calidad$\}$, $p^* = P(E)$ y que la v.a. Y definida como el número de memorias RAM examinadas hasta obtener la primera que cumple las condiciones de calidad, sea la v.a. X número de fracasos que preceden al primer éxito. Calcular $E(X)$ y $V(X)$.*

8) *Un estudio geológico indica que en un pozo exploratorio podría hallarse petróleo con una probabilidad de 0.2. Suponiendo que las exploraciones son independientes una de otra. ¿Cuál es la probabilidad que*

   i) *el primer descubrimiento ocurra en la tercera perforación?*

   ii) *el tercer descubrimiento ocurra en la quinta perforación?*