Lección 3: Variables aleatorias

Módulo 2: Probabilidades y decisiones bajo incertidumbre

Magdalena Cornejo

Variable aleatoria

Definición

Una **variable aleatoria** es una variable que toma valores numéricos determinados por el resultado de un experimento que tiene asociado una probabilidad.

Notación: en general describimos la variable aleatoria con mayúscula (X) y el resultado (realización) con minúscula (x).

Ejemplo

Se tira una moneda al aire tres veces (**experimento**), definimos X (**variable aleatoria**) como "el número de caras obtenidas". Entonces X es una variable aleatoria que puede tomar los siguientes valores (x): 0, 1, 2 y 3. Cada uno de esos valores va a tener una probabilidad asociada.

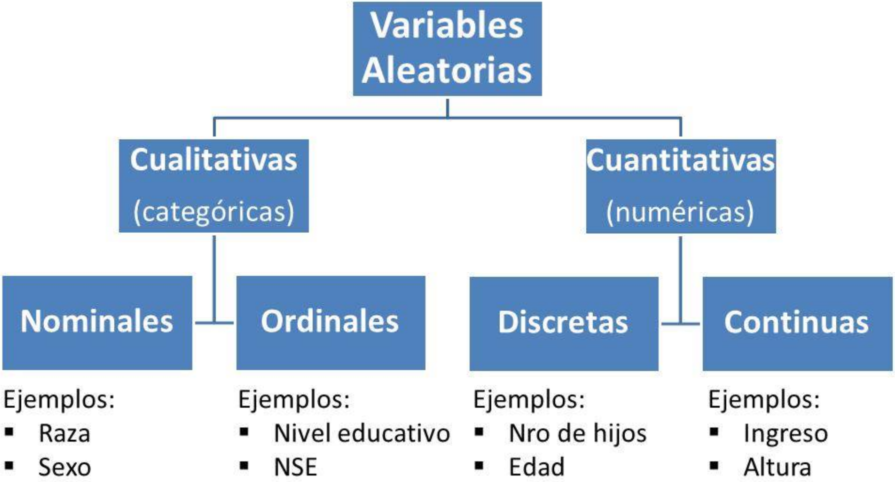
¿Se animan a calcularla?

Ejemplo (Resolución)

- Se lanza 3 veces una moneda
- X = número de caras obtenidas. Entonces, $X = 0, 1, 2, 3$
- Llamemos H a sacar una cara y T a sacar una ceca.
- Los resultados posibles son 8: HHH, HHT, HTH, THH, HTT, THT, TTH, TTT.
- Entonces se puede construir la siguiente tabla:

X	$P(X = x)$
0	1/8=0,125
1	3/8=0,375
2	3/8=0,375
3	1/8=0,125

Tipos de variables aleatorias



Volviendo a las variables aleatorias

Entonces, una **variable aleatoria** es una variable para la cual

- no sabemos (con certeza)
- cuál va a ser (ex-ante) su valor.

Vamos a centrarnos principalmente en el estudio de variables aleatorias **cuantitativas**:

- Discretas
- Continuas

Variables Aleatorias Discretas

Si X es una **variable aleatoria discreta** y x es uno de sus posibles valores, la probabilidad de que X tome un valor específico x se escribe: $p(x)$ o $P(X = x)$. Esto se conoce como la **función de probabilidad**.

Ejemplo

Se tira un dado. X = variable aleatoria que indica el número resultante. Si el dado es equilibrado, $P(X = 1) = ... = P(X = 6) = 1/6$, entonces su función de probabilidad será:

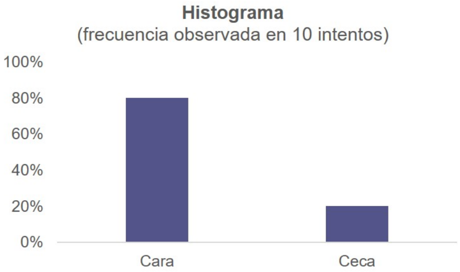
$$P(X = x) = 1/6 \text{ para } x = 1, 2, 3, ..., 6$$

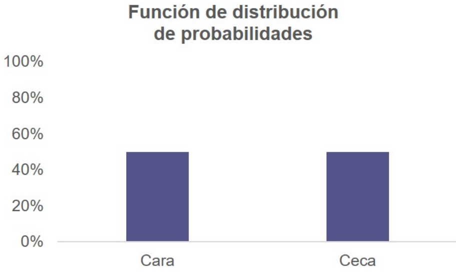
¿Cómo se vería gráficamente su distribución de probabilidades?

Esto nos permite calcular distintas probabilidades:

- $P(2 \leq X \leq 4) = P(X = 2) + P(X = 3) + P(X = 4) = 1/2$
- $P(X \leq 5) = 1 - P(X > 5) = 1 - P(X = 6) = 5/6$

Histograma vs Distribución de Probabilidades





Valor esperado (Esperanza Matemática)

Definición

El **valor esperado** (o esperanza matemática) es una medida de lo que ocurre con más frecuencia o en promedio.

El **valor esperado** de una variable aleatoria discreta X es:

$$\mu = E(X) = p_1 x_1 + p_2 x_2 + \ldots + p_n x_n = \sum_i p_i x_i$$

Es decir, el valor esperado es el **promedio ponderado** de todos los posibles valores que la misma puede adoptar, donde los ponderadores son las probabilidades asociadas a cada x. Notar que el promedio simple es el caso particular en que $p = 1/n$.

Ejemplo

- X=cantidad de autos que compra una familia en el lapso de 5 años.
- Supongamos que conocemos la probabilidad asociada a cada x:

X	P(X=x)
0	0,30
1	0,27
2	0,20
3	0,13
4	0,06
5	0,03
6	0,01

¿Cuál es el valor esperado?

Ejemplo

- X=cantidad de autos que compra una familia en el lapso de 5 años.
- Supongamos que conocemos la probabilidad asociada a cada x:

X	P(x)	x.P(x)
0	0,30	0.0,30=0
1	0,27	1.0,27=0,27
2	0,20	2.0,20=0,40
3	0,13	3.0,13=0,39
4	0,06	4.0,06=0,24
5	0,03	5.0,03=0,15
6	0,01	6.0,01=0,06
Suma		**1,51**

Propiedades de la esperanza

Si a es una constante (un número) y X una variable aleatoria, entonces:

- $E(a) = a$
- $E(aX) = aE(X)$

Si tenemos dos variables aleatorias (X e Y). Entonces,

- $E(X + Y) = E(X) + E(Y)$
- $E(X - Y) = E(X) - E(Y)$
- ¡Multiplicarlas y dividirlas es distinto!

¡Conocer las propiedades de la esperanza nos facilita realizar ciertos cómputos!

Ejemplo

- X= ingreso mensual del jefe de hogar
- Y= ingreso mensual del cónyuge

Se sabe que E(X)=8000 y E(Y)=6000.

Ahora, supongamos que todos los jefes de hogar reciben un incremento del 20% en su salario, mientras que sus cónyuges una suma fija de $2000. ¿Cuál es el valor esperado del nuevo ingreso mensual del hogar (jefe + cónyuge)?

Ejemplo

- $X=$ ingreso mensual del jefe de hogar
- $Y=$ ingreso mensual del cónyuge

Se sabe que E(X)=8000 y E(Y)=6000.

Ahora, supongamos que todos los jefes de hogar reciben un incremento del 20% en su salario, mientras que sus cónyuges una suma fija de $2000. ¿Cuál es el valor esperado del nuevo ingreso mensual del hogar (jefe + cónyuge)?

- $T=$nuevo ingreso total familiar.

Ejemplo

- X= ingreso mensual del jefe de hogar
- Y= ingreso mensual del cónyuge

Se sabe que E(X)=8000 y E(Y)=6000.

Ahora, supongamos que todos los jefes de hogar reciben un incremento del 20% en su salario, mientras que sus cónyuges una suma fija de $2000. ¿Cuál es el valor esperado del nuevo ingreso mensual del hogar (jefe + cónyuge)?

- T=nuevo ingreso total familiar.
- Entonces, $E(T) = E(1,2X + Y + 2000)$.

Ejemplo

- X= ingreso mensual del jefe de hogar
- Y= ingreso mensual del cónyuge

Se sabe que E(X)=8000 y E(Y)=6000.

Ahora, supongamos que todos los jefes de hogar reciben un incremento del 20% en su salario, mientras que sus cónyuges una suma fija de $2000. ¿Cuál es el valor esperado del nuevo ingreso mensual del hogar (jefe + cónyuge)?

- T=nuevo ingreso total familiar.
- Entonces, $E(T) = E(1,2X + Y + 2000)$.
- Aplicando las propiedades:
  $E(T) = 1,2E(X) + E(Y) + 2000 = 1,2.8000 + 6000 + 2000$=$17600

Varianza

- Como vimos en estadística descriptiva, no podemos limitarnos solamente a fijarnos en medidas de tendencia central como la media.
- No tendríamos una idea acabada de cómo se distribuyen los datos.
- Podemos calcular también la **varianza** (o el desvío) de una variable aleatoria como medida de dispersión.

Varianza

- Como vimos en estadística descriptiva, no podemos limitarnos solamente a fijarnos en medidas de tendencia central como la media.
- No tendríamos una idea acabada de cómo se distribuyen los datos.
- Podemos calcular también la **varianza** (o el desvío) de una variable aleatoria como medida de dispersión.

Si X es una variable aleatoria discreta, la esperanza del desvío al cuadrado respecto de la media $(X - E(X))^2$ se denomina **varianza** ($V(X)$ o σ_X^2):

$$\sigma^2 = V(X) = E[(X-\mu)^2] = \sum_i (x-\mu)^2 p(x)$$

Notar que: (Fórmula alternativa)

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

Intuición

$$\sigma^2 = V(X) = E[(X-\mu)^2] = \sum_i (x-\mu)^2 p(x)$$

- La varianza (generalmente denotada como σ^2) es un promedio ponderado (por su probabilidad) de la distancia cuadrática entre la media (μ) y cada x.
- Es mayor cuanto más lejos estén los x de la media, y cuanto mayor sea el peso (probabilidad de ocurrencia) de esos valores.
- Usualmente se lo asocia con variabilidad y riesgo.

Ejemplo

En el ejemplo de la cantidad de autos que compra una familia en un lapso de 5 años, si queremos calcular la varianza:

X	P(x)	x-μ	(x-μ)^2	(x-μ)^{2}P(x)
0	0,30	0-1,51=-1,51	2,28	2,28.0,30=0,68
1	0,27	1-1,51=-0,51	0,26	0,26.0,27=0,07
2	0,20	2-1,51=0,49	0,24	0,24.0,20=0,05
3	0,13	3-1,51=1,49	2,22	2,22.0,13=0,29
4	0,06	4-1,51=2,49	6,20	6,20.0,06=0,37
5	0,03	5-1,51=3,49	12,18	12,18.0,03=0,37
6	0,01	6-1,51=4,49	20,16	20,16.0,01=0,20
			Suma	**2,03**

Entonces, $\sigma^2 = 2,02$ y $\sigma = 1,42$.

Propiedades de la varianza - Creer o reventar

Si a y b son dos números (constantes, no variables) y X una variable aleatoria, entonces:

- $V(a) = 0$
- $V(aX) = a^2 V(X)$
- $V(a + bX) = b^2 V(X)$

Si X e Y son dos variables aleatorias:

- $V(X \pm Y) = V(X) + V(Y) \pm 2COV(X, Y)$
- $V(X \pm Y) = V(X) + V(Y)$ **si** X e Y son variables aleatorias **independientes** (no tienen relación lineal, $COV(X, Y) = 0$)

Ejemplo (propiedades de la varianza)

- $X=$ ingreso mensual del jefe de hogar
- $Y=$ ingreso mensual del cónyuge

Se sabe que $\sigma(X) = 150$ y $\sigma(Y) = 100$.

Ahora, supongamos que todos los jefes de hogar reciben un incremento del 20% en su salario, mientras que sus cónyuges una suma fija de $2000. ¿Cuál es el el desvío estándar del nuevo ingreso mensual del hogar (jefe + cónyuge)? Asuma que el ingreso del jefe de hogar es **independiente** del de su cónyuge

Ejemplo (propiedades de la varianza)

- X= ingreso mensual del jefe de hogar
- Y= ingreso mensual del cónyuge

Se sabe que $\sigma(X) = 150$ y $\sigma(Y) = 100$.

Ahora, supongamos que todos los jefes de hogar reciben un incremento del 20% en su salario, mientras que sus cónyuges una suma fija de \$2000. ¿Cuál es el el desvío estándar del nuevo ingreso mensual del hogar (jefe + cónyuge)? Asuma que el ingreso del jefe de hogar es **independiente** del de su cónyuge

- T=nuevo ingreso total familiar.

Ejemplo (propiedades de la varianza)

- $X=$ ingreso mensual del jefe de hogar
- $Y=$ ingreso mensual del cónyuge

Se sabe que $\sigma(X) = 150$ y $\sigma(Y) = 100$.

Ahora, supongamos que todos los jefes de hogar reciben un incremento del 20% en su salario, mientras que sus cónyuges una suma fija de \$2000. ¿Cuál es el el desvío estándar del nuevo ingreso mensual del hogar (jefe + cónyuge)? Asuma que el ingreso del jefe de hogar es **independiente** del de su cónyuge

- $T=$nuevo ingreso total familiar.
- Entonces, $V(T) = V(1,2X + Y + 2000)$.

Ejemplo (propiedades de la varianza)

- $X=$ ingreso mensual del jefe de hogar
- $Y=$ ingreso mensual del cónyuge

Se sabe que $\sigma(X) = 150$ y $\sigma(Y) = 100$.

Ahora, supongamos que todos los jefes de hogar reciben un incremento del 20% en su salario, mientras que sus cónyuges una suma fija de \$2000. ¿Cuál es el el desvío estándar del nuevo ingreso mensual del hogar (jefe + cónyuge)? Asuma que el ingreso del jefe de hogar es **independiente** del de su cónyuge

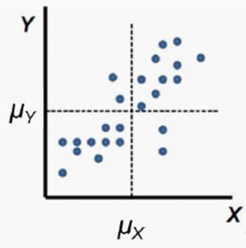
- $T=$nuevo ingreso total familiar.
- Entonces, $V(T) = V(1,2X + Y + 2000)$.
- Aplicando las propiedades: $V(T) = 1,2^2 V(X) + V(Y) = 1,2^2.150^2 + 100^2 = 42400$
- $\sigma(T)=$\$205.9

Covarianza

La **covarianza** entre dos variables aleatorias X e Y es una medida de la asociación que existe entre ambas. Está dada por:

$$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] = \sum_i \sum_j (x-\mu_X)(y-\mu_Y)p(x,y)$$

- Para calcularla hay que conocer la **probabilidad conjunta** ($p(x,y)$).
- Una covarianza positiva (negativa) implica que existe una asociación lineal positiva (negativa) entre el par de variables aleatorias.



Intuición

- $Cov(X, Y) > 0$: si $\uparrow X$, Y tiende a subir (o si $\uparrow Y$, X tiende a subir)
- $Cov(X, Y) < 0$: si $\uparrow X$, Y tiende a bajar (o si $\uparrow Y$, X tiende a bajar)
- $Cov(X, Y) = 0$: **no** hay asociación **lineal** entre X e Y

Intuición

- $Cov(X, Y) > 0$: si $\uparrow X$, Y tiende a subir (o si $\uparrow Y$, X tiende a subir)
- $Cov(X, Y) < 0$: si $\uparrow X$, Y tiende a bajar (o si $\uparrow Y$, X tiende a bajar)
- $Cov(X, Y) = 0$: **no** hay asociación **lineal** entre X e Y

- En la lección anterior vimos que las variables aleatorias X e Y son **independientes** si $p(x, y) = p(x).p(y)$ para todos los valores de X e Y.
- Bajo independencia, también se verifica que $E(XY) = E(X).E(Y)$.
- Entonces, si X e Y son independientes, $COV(X, Y) = 0$.
- Sin embargo, la inversa no es necesariamente verdadera (puede haber otro tipo de relación distinta de la lineal).

Correlación

Como la covarianza depende de las unidades de medición de X e Y, muchas veces utilizamos el **coeficiente de correlación** (ρ_{XY}) que lo escribimos de la siguiente manera:

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

- ρ_{XY} mantiene el signo de la covarianza
- ρ_{XY} es 0 cuando la covarianza es 0.
- $-1 \leq \rho_{X,Y} \leq 1$

El coeficiente de correlación es una medida MUY utilizada en la práctica y que retomaremos cuando veamos regresión.

Correlación