# Unidad 3. Probabilidad

Javier Santibáñez

17 de agosto de 2018

## 1. Introducción

**Definición 1.** La probabilidad es una medida subjetiva del grado de creencia que se tiene acerca de que algo desconocido sea verdadero u ocurra.

La clave es la ignorancia, incertidumbre o desconocimiento.

**Ejemplo 1.1.** Al lanzar un dado, no se sabe con anticipación cuál será el resultado, entonces tiene sentido asignar probabilidades a los posibles resultados.

**Ejemplo 1.2.** Si en una encuesta se pregunta la edad de las personas, la edad de un individuo en específico no es aleatoria, es fija pero desconocida, entonces tiene sentido asignar una probabilidad a cada respuesta posible.

En los ejemplos anteriores, la asignación de probabilidades es subjetiva, pues depende del conocimiento o experiencia previas.
El modelo matemático para estudiar la probabilidad recibe el nombre de espacio de probabilidad.

**Definición 2.** Un espacio de probabilidad se conforma de los siguientes tres elementos.

1. El conjunto de todos los posibles resultados del experimento o proceso de interés, denotado por $\Omega$ y llamado espacio muestral.

2. Una familia de eventos de interés $\mathcal{A}$. Los eventos son combinaciones u agrupaciones (conjuntos) de posibles resultados. *$\mathcal{A}$ debe ser un $\sigma$-álgebra de subconjuntos de $\Omega$.

3. Una medida de probabilidad $P$. Esta función va de $\mathcal{A}$ al intervalo $[0,1]$ y asigna probabilidades a los eventos de interés. $P$ debe cumplir con los siguientes propiedades:

*a*) $P(\Omega) = 1$,

*b*) Si $E_1, E_2, E_3, \ldots \mathcal{A}$ y $E_i \cap E_j = \emptyset$ para $i \neq j$, entonces

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i).$$

**Ejemplo 1.3.** Se lanzan tres volados con una moneda balanceada. Los resultados se pueden escribir de la forma $XYZ$, donde $X, Y, Z$ representan el resultado del primer, segundo y tercer volado, respectivamente, $X, Y, Z \in \{A = \text{águila}, S = \text{sol}\}$. El espacio muestral de este experimento tiene ocho elementos

$$\Omega = \{AAA, AAS, ASA, ASS, SAA, SAS, SSA, SSS\}.$$

Se puede tomar como familia de eventos de interés $\mathcal{A} = 2^{\Omega}$. Finalmente, dada la información sobre la moneda, se puede pensar que es igualmente probable que ocurra $A$ o $S$ en cualquier volado, de manera que los ocho elementos del espacio muestral son igualmente probables. Entonces $P(\omega) = 0.125$, para todo $\omega \in \Omega$.

En la mayoría de los experimentos, no es de interés estudiar las probabilidades de los resultados, sino de una característica numérica de los mismos. En el ejemplo anterior, podría pensarse que el interés es el número de águilas que ocurren en los tres volados.

**Definición 3.** Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad. Se dice que una función $X$ que va de $\Omega$ a $\mathbb{R}$ es una variable aleatoria si cumple que

$$\{\omega \in \Omega \,|\, X(\omega) \leq a\} \in \mathcal{A}, \qquad \text{para todo } a \in \mathbb{R}.$$

**Notación.** Definimos $\{X \leq a\}$ como $\{\omega \in \Omega \,|\, X(\omega) \leq a\}$.

**Ejemplo 1.4.** En el Ejemplo 1.3, se define la función

$$X(\omega) = \text{número de águilas en } \omega.$$

Entonces

$$X(AAA) = 3$$
$$X(AAS) = X(ASA) = X(SAA) = 2$$
$$X(SSA) = X(SAS) = X(ASS) = 1$$
$$X(SSS) = 0.$$

A partir de lo anterior se puede verificar que

$$\{X \leq a\} = \begin{cases} \emptyset, & \text{si } a < 0, \\ \{SSS\}, & \text{si } 0 \leq a < 1, \\ \{SSS, SSA, SAS, ASS\}, & \text{si } 1 \leq a < 2, \\ \{SSS, SSA, SAS, ASS, SAA, ASA, AAS\}, & \text{si } 2 \leq a < 3, \\ \Omega, & \text{si } a \geq 3. \end{cases}$$

En cualquier caso se cumple que $\{X \leq a\} \in \mathcal{A} = 2^{\Omega}$. Por lo tanto, la función $X$ es una variable aleatoria en espacio del Ejemplo 1.3.

¿Cómo se asignan probabilidades a las variables aleatorias? Al igual que en la definición de variable aleatoria, los eventos $\{X \leq a\}$, con $a \in \mathbb{R}$, desempeñan un papel central para asignar probabilidades.

**Definición 4.** Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad y $X$ una variable aleatoria. Se define la función de distribución acumulada (fda) de $X$ como la función $F$ que va de $\mathbb{R}$ al intervalo $[0,1]$ dada por

$$F(a) = P\{X \leq a\}, \qquad \text{para } a \in \mathbb{R}.$$

**Ejemplo 1.5.** La fda de la variable aleatoria $X$ del Ejemplo 1.4 está dada por

$$F(a) = \begin{cases} 0, & \text{si } a < 0, \\ 1/8, & \text{si } 0 \leq a < 1, \\ 1/2, & \text{si } 1 \leq a < 2, \\ 7/8, & \text{si } 2 \leq a < 3, \\ 1, & \text{si } a \geq 3. \end{cases}$$

**Proposición 1.** Si $F$ es una función de distribución acumulada, entonces se cumple lo siguiente.

a) $F$ es no decreciente.

b) $F$ es continua por la derecha.

c) $\lim_{a \to -\infty} F(a) = 0$.

d) $\lim_{a \to \infty} F(a) = 1$.

A partir de la fda podemos asignar probabilidades a los eventos determinados a partir de una variable aleatoria, pero la forma de hacerlo depende de las características de la variable.

## 2. Variables aleatorias discretas

**Definición 5.** Sea $X$ una variable aleatoria definida en un espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Si el rango de $X$, denotado por $X(\Omega)$, es un conjunto numerable, entonces se dice que $X$ es una variable aleatoria discreta.

**Definición 6.** Sea $X$ una variable aleatoria discreta definida en un espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Se define la función masa de probabilidades de $X$ como la función $p$ de $\mathbb{R}$ a $[0,1]$ tal que

$$p(x) = \begin{cases} P\{X = x\}, & \text{si } x \in X(\Omega), \\ 0, & \text{d.o.m.} \end{cases}, \qquad \text{para todo } x \in \mathbb{R}.$$

**Notación.** Para $a, b$ constantes y $B \subset \mathbb{R}$, se definen $\{X = a\}$, $\{a \leq X \leq b\}$ y $\{X \in B\}$ de manera análoga a cómo se hizo con $\{X \leq a\}$.

La fda de una variable aleatoria discreta $X$ está completamente determinada por su fmp. Por ejemplo, si $X(\Omega) = \{x_1, x_2, x_3, \ldots\}$, entonces

$$F(a) = P\{X \leq a\} = \sum_{x_i \leq a} p(x_i), \qquad \text{para todo } a \in \mathbb{R},$$

donde la suma de las probabilidades se realiza sobre todos los $x_i$ que son menores o iguales a $a$. Las fmp cumplen con la siguiente propiedad

$$\sum_{x \in X(\Omega)} p(x) = 1.$$

Cuando $X(\Omega)$ es finito y tiene pocos elementos, lo usual es representar la fmp de $X$ en una tabla como la siguiente

| $x$ | $x_1$ | $x_2$ | $\ldots$ | $x_n$ |
|---|---|---|---|---|
| $p(x)$ | $p(x_1)$ | $p(x_2)$ | $\ldots$ | $p(x_n)$ |

**Ejemplo 2.1.** La fmp de la variable aleatoria del Ejemplo 1.4 está dada por

| $x$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| $p(x)$ | 1/8 | 3/8 | 3/8 | 1/8 |

En la siguiente sección se presentarán algunas de las variables aleatorias discretas más utilizadas.

Existe un resultado teórico muy importante que establece que las variables aleatorias quedan completamente caracterizadas por su función de distribución, de manera que se usarán de manera indistinta estos dos términos para hacer referencia a los modelos de probabilidad.

## 2.1. Distribuciones discretas más comunes

**Definición 7.** Se dice que una variable aleatoria $X$ tiene una distribución de Bernoulli con parámetro $\theta$, $\theta \in (0,1)$, si el rango de $X$ es $X(\Omega) = \{0,1\}$ y su fmp está dada por

$$p(x) = \begin{cases} \theta^x (1-\theta)^{1-x}, & \text{si } x = 0,1, \\ 0, & \text{d.o.m.} \end{cases}, \qquad \text{para todo } x \in \mathbb{R}.$$

Lo anterior se denota como $X \sim B(\theta)$.

Este distribución se utiliza para modelar el resultado de ensayos binarios (dos posibles resultados, genéricamente denominados éxito/fracaso), también llamados ensayos de Bernoulli. $X$ asigna 1 al éxito y 0 al fracaso.

**Definición 8.** Se dice que una variable aleatoria $X$ tiene una distribución binomial con parámetros $n$ y $\theta$, $n \in \{1,2,3,\ldots\}$ y $\theta \in (0,1)$, si el rango de $X$ es $X(\Omega) = \{0,1,\ldots,n\}$ y su fmp está dada por

$$p(x) = \begin{cases} \binom{n}{x} \theta^x (1-\theta)^{n-x}, & \text{si } x = 0,1,\ldots,n, \\ 0, & \text{d.o.m.} \end{cases}, \qquad \text{para todo } x \in \mathbb{R}.$$

donde $\binom{n}{x} = \frac{n!}{(n-x)!x!}$. Lo anterior se denota como $X \sim Bin(n,\theta)$.

La distribución binomial se utiliza para modelar experimentos que consisten de la repetición de ensayos de Bernoulli idénticos e independientes. Si se llevan a cabo $n$ ensayos de Bernoulli independientes y cada uno de ellos tiene probabilidad de éxito $\theta$, entonces el número total de éxitos obtenidos tiene una distribución $Bin(n,\theta)$.

**Definición 9.** Se dice que una variable aleatoria $X$ tiene una distribución geométrica con parámetro $\theta$, $\theta \in (0,1)$, si su rango es $X(\Omega) = \{1,2,3,\ldots\}$ y su fmp es

$$p(x) = \begin{cases} \theta(1-\theta)^{x-1}, & \text{si } x = 1,2,3,\ldots, \\ 0, & \text{d.o.m.} \end{cases}, \qquad \text{para todo } x \in \mathbb{R}.$$

Lo anterior se denota como $X \sim Geo(\theta)$.

El modelo geométrico también se utiliza en experimentos que surgen de la repetición de ensayos de Bernoulli idénticos e independientes. SI se llevan a cabo ensayos de Bernoulli independientes, cada uno con probabilidad de éxito $\theta$, el experimento termina cuando se observa el primer éxito, entonces el número de ensayos necesarios para observar el primer éxito tiene una distribución $Geo(\theta)$.

Existe un modelo relacionado con la distribución geométrica que, en términos de la interpretación anterior, resulta de contar el número de ensayos necesarios para observar $r$ éxitos, $r = 1, 2, 3, \ldots$. En este caso el modelo recibe el nombre de distribución binomial negativa y tiene como parámetros a $r$ y $\theta$.

**Definición 10.** Se dice que una variable aleatoria $X$ tiene una distribución binomial negativa con parámetros $r$ y $\theta$, $r = 1, 2, 3, \ldots$ y $\theta \in (0, 1)$, si su rango es $X(\Omega) = r, r+1, r+2, \ldots$ y su fmp es

$$p(x) = \begin{cases} \binom{x-1}{x-r}\theta^r (1-\theta)^{x-r}, & \text{si } x = r, r+1, r+2, \ldots, \\ 0, & \text{d.o.m.} \end{cases}, \qquad \text{para todo } x \in \mathbb{R}.$$

Lo anterior se denota como $X \sim BinNeg(r, \theta)$.

**Definición 11.** Se dice que una variable aleatoria $X$ tiene una distribución de Poisson con parámetro $\lambda$, $\lambda \in (0, \infty)$, si $X(\Omega) = \{0, 1, 2, \ldots\}$ y su fmp está dada por

$$p(x) = \begin{cases} \dfrac{e^{-\lambda}\lambda^x}{x!}, & \text{si } x = 0, 1, 2, \ldots, \\ 0, & \text{d.o.m.} \end{cases}, \qquad \text{para todo } x \in \mathbb{R}.$$

Lo anterior se denota como $X \sim Poi(\lambda)$.

El modelo de Poisson se utiliza generalmente para describir experimentos en donde el resultado es el conteo de eventos que ocurren en un periodo específico. En este contexto, $\lambda$ se interpreta como la tasa de ocurrencia de los eventos. Por ejemplo, el número de temblores/terremotos que ocurren en una determinada región en un periodo de una semana, se puede modelar con la distribución de Poisson. En este caso $\lambda$ representa la tasa con la que ocurren terremotos en una semana, es decir, el promedio de terremotos que ocurren en una semana.

## 3. Variables aleatorias continuas

**Definición 12.** Sea $X$ una variable aleatoria definida en un espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Se dice que $X$ es una variable aleatoria continua (absolutamente continua*) si su rango $X(\Omega)$ es no numerable y existe una función $f$ no negativa tal que

$$P\{X \in B\} = \int_B f(x)dx, \qquad B \subset \mathbb{R}.$$

La función $f$ se conoce como función de densidad de probabilidad (fdp) de $X$.

La propiedad de las fdp análoga a los fmp es

$$\int_{-\infty}^{\infty} f(x)dx = 1.$$

De la definición de variable aleatoria continua, se sigue que

$$P\{a \leq X \leq b\} = \int_a^b f(x)dx, \qquad \text{para todo } a, b \int \mathbb{R}, a < b;$$
$$P\{X = a\} = \int_a^a f(x)dx = 0, \qquad \text{para todo } a \in \mathbb{R}.$$

En esta segunda igualdad está la principal diferencia entre las variables aleatorias continuas y discretas. Si $X$ es continua, entonces $P\{X = a\} = 0$, para todo $a \in \mathbb{R}$, mientras que si $X$ es discreta, entonces existe al menos un $a \in \mathbb{R}$ tal que $P(X = a) > 1$.

La fda de $X$ se puede expresar como

$$F(a) = \int_{-\infty}^{a} f(x)dx, \qquad \text{para todo } a \in \mathbb{R}.$$

**Proposición 2.** Sea $X$ una variable aleatoriaa continua con fda $F$ y fdp $f$. Para todo $a \in \mathbb{R}$ se cumple que

$$\frac{d}{dx}F(x)\Big|_{x=a} = f(a).$$

El resultado anterior se puede probar fácilmente utilizando el Teorema Fundamental del Cálculo.

Una interpretación importante de la fdp es la siguiente, para $\epsilon > 0$ pequeño, se cumple que

$$P\{a - \frac{\epsilon}{2} \leq X \leq a + \frac{\epsilon}{2}\} = \int_{a-\frac{\epsilon}{2}}^{a+\frac{\epsilon}{2}} f(x)dx \approx \epsilon f(a).$$

Esto nos dice que la probabilidad de que $X$ esté alrededor de $a$ en un intervalo de longitud $\epsilon$, es aproximadamente $\epsilon f(a)$. En otras palabras, $f(a)$ es una medida de qué tan verosímil es que $X$ esté cerca de $a$.

## 3.1. Distribuciones discretas más comunes

**Definición 13.** Se dice que una variable aleatoria continua $X$ tiene una distribución uniforme en el intervalo $[\alpha, \beta]$ si su fdp está dada por

$$f(x) = \frac{1}{\beta - \alpha} \mathbb{1}_{[\alpha,\beta]}(x), \qquad \text{para todo } x \in \mathbb{R}.$$

Lo anterior se denota como $X \sim U(\alpha, \beta)$.

**Notación.** La función $\mathbb{1}_A(x)$ se conoce como función indicadora del conjunto $A$, la cual se define como

$$\mathbb{1}_A(x) = \begin{cases} 1, & \text{si } x \in A, \\ 0, & \text{si } x \notin A. \end{cases} \qquad \text{para todo } x \in \mathbb{R}.$$

Es fácil verificar que si $X \in U(\alpha, \beta)$, entonces su fda es

$$F(x) = \begin{cases} 0, & \text{si } x < 0, \\ \dfrac{x - \alpha}{\beta - \alpha}, & \text{si } \alpha \leq x < \beta, \\ 1, & \text{si } x \leq \beta. \end{cases} \qquad \text{para todo } x \in \mathbb{R}.$$

**Definición 14.** Se dice que una variable aleatoria $X$ tiene una distribución exponencial con parámetro $\lambda$, $\lambda \in (0, \infty)$, si su fdp está dada por

$$f(x) = \lambda e^{-\lambda x} \mathbb{1}_{(0,\infty)}(x), \qquad \text{para todo } x \in \mathbb{R}.$$

Esto se denota como $X \sin Exp(\lambda)$.

Es fácil verificar que la fda de una variable aleatoria $Exp(\lambda)$ está dada por

$$F(x) = \begin{cases} 0, & \text{si } x < 0, \\ 1 - e^{-\lambda x}, & \text{si } x \leq 0, \end{cases} \qquad \text{para todo } x \in \mathbb{R}.$$

Esta distribución generalmente se utiliza para modelar tiempos de vida, aunque hay más distribuciones disponibles.

**Definición 15.** Se define la función gamma, denotada por $\Gamma$, como la función dada por

$$Gamma(x) = \int_0^\infty y^{x-1} e^{-x} dy, \qquad x > 0.$$

La función gamma cumple con las siguientes propiedades:

a) $\Gamma(\alpha + 1) = \alpha \Gamma(\alpha)$,

b) $\Gamma(n+1) = n!$, si $n \in \{1, 2, 3, \ldots\}$,

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

**Definición 16.** Se dice que una variable aleatoria continua $X$ tiene una distribución Gamma con parámetros $\alpha$ y $\lambda$, con $\alpha \in (0, \infty)$ y $\lambda \in (0, \infty)$, si su fdp está dada por

$$f(x) = \frac{\lambda^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \mathbb{1}_{(0,\infty)}(x), \qquad \text{para todo } x \in \mathbb{R}.$$

Lo anterior se denota por $Ga(\alpha, \lambda)$. La fda de está distribución no tiene una expresión cerrada o bien, depende de la función gamma incompleta.

Cuando $\alpha \in \{1, 2, 3\ldots\}$, esta distribución se utiliza para modelar el tiempo de vida de un sistema o dispositivo con $\alpha$ componentes idénticos e independientes con tiempos de fallo modelados como $Exp(\lambda)$.

**Definición 17.** Se define la función beta, denotada por $B$, como la función dada por

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\, dx, \qquad \alpha, \beta > 0.$$

La función beta cumple con las siguientes propiedades

a) $B(\alpha, \beta) = B(\beta, \alpha)$.

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

## 4. Momentos

## 5. Distribuciones derivadas del muestreo de poblaciones normales