**Clase 6: Probabilidad**

**Experimento**: Cualquier proceso que produce una **observación** o **resultado**.

**Experimento Determinístico**: Al realizarse en condiciones similares, da el mismo resultado.

**Experimento Aleatorio**: No se sabe qué resultado dará hasta realizarlo. Se pueden conocer con anticipación sus **posibles resultados**, pero no qué resultado concreto dará. IE: Lanzamiento de dado, lanzamiento de moneda, extracción de una carta.

**Espacio Muestral:** Conjunto de Resultados posibles de un experimento aleatorio: Ω. A un resultado de un experimento puntual se lo denomina con ω. IE: Lanzamiento de un dado Ω = {1, 2, 3, 4, 5, 6}

**Evento**: Se define como evento el conjunto A = {1, 3, 5}; que es cuando sale un resultado impar en el lanzamiento de un dado. Si sale impar, el evento **ocurre**, caso contrario, **no ocurre**.

**Variable Aleatoria:** Transformación X del espacio de resultados Ω al conjunto de números reales. El valor que toma una variable aleatoria está asociado al resultado de un experimento aleatorio. Es un número determinado por la realización del experimento.

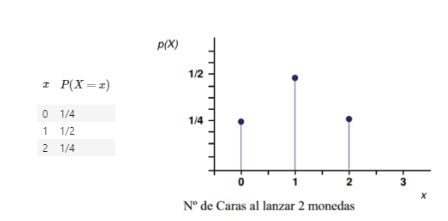
**Función de Distribución de Probabilidad.**

La **Distribución de Probabilidad de una Variable Aleatoria X,** (**función de Distribución de X)**, función que asigna a cada evento definido sobre la variable aleatoria X una probabilidad:

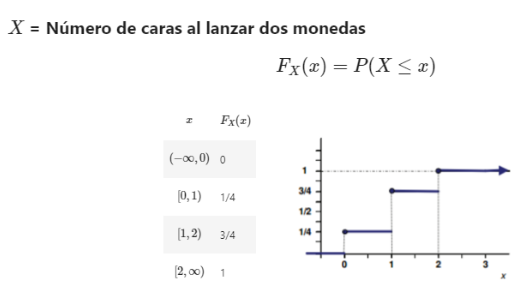
El valor en cada x real es la probabilidad de que la variable aleatoria sea menor o igual que x.

**Variables Aleatorias Discretas:** Los valores que toma una variable son **finitos o infinitos numerables**.IE: resultado del lanzamiento de un dado, cuántas veces hay que lanzar una moneda para obtener cara.

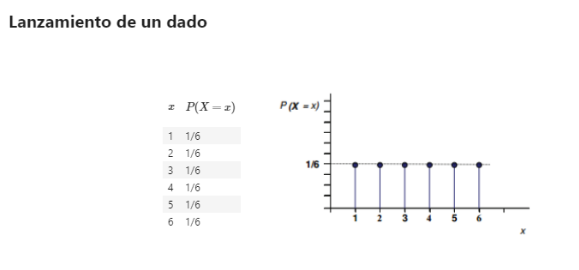
En este ejemplo grafica la probabilidad de ocurrencia de que no salga cara ninguna vez, una vez o dos veces.

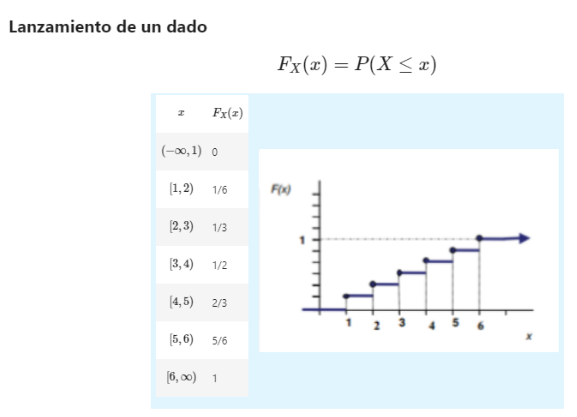


En este ejemplo grafica la probabilidad de ocurrencia acumulada de que no salga cara ninguna vez, o ninguna vez + una vez o ninuna vez, más una vez más dos veces (100% porque son todos los casos posibles).



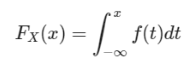
Ejemplo análogo con un dado:





**Variables Aleatorias Continuas:** Cuando los valores que puede tomar una variable aleatoria puede ser cualquier valor en un intervalo continuo dado (números reales), se dice que es **continua**. IE: Altura de un alumno del curso. Una variable aleatoria es continua si su función de distribución es continua.

En forma análoga a las variables aleatorias discretas, las variables aleatorias continuas tienen **funciones de densidad de probabilidad de variables aleatorias continuas.** Describen la probabilidad relativa y nos permiten conocer cómo se distribuyen las probabilidades de un suceso o evento en relación al resultado.

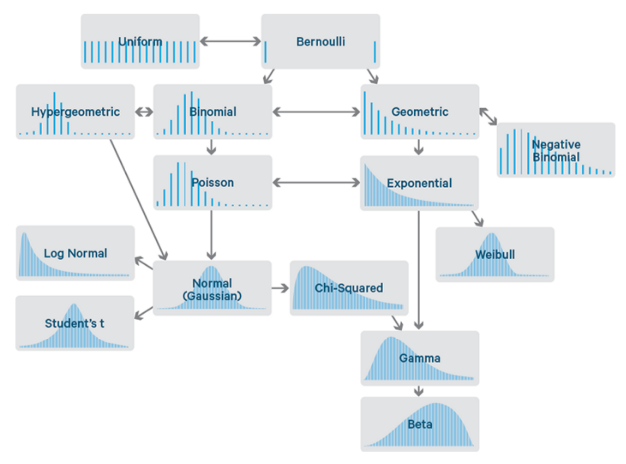


Fx: Función de probabilidad.

f(x): función de densidad.

La integral de la densidad de probabilidad en una región específica del espacio de posibilidades nos termina dando la probabilidad de que la variable caiga en la misma. A diferencia de con las variables discretas, no tiene sentido hablar de probabilidades puntuales en las variables continuas, porque la misma es infinitesimalmente pequeña; tiende a 0.

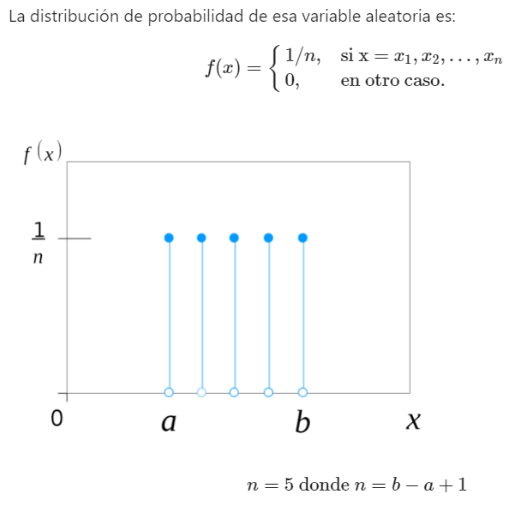
**Distribuciones de Probabilidad**



Existen distribuciones de probabilidad conocidas, tanto de variables discretas como de variables continuas. Las distribuciones pueden depender de uno o más **parámetros**. Para cada uno de estos parámetros hay una distribución de probabilidad. **Familias de distribuciones paramétricas**.

**Distribución Uniforme Discreta**: Aquella en la que la probabilidad de que X tome cualquiera de sus valores posibles es 1/n, siendo n el total de valores posibles. IE: Un dado no cargado. Tiene **espacios de probabilidad equiprobables**.





**En Python**:

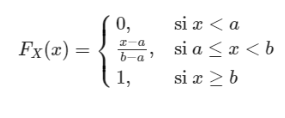
random\_generator = np.random.default\_rng()

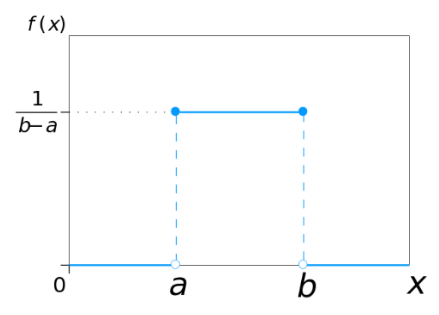
possible\_values = [3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16]

random\_uniform\_disc\_data = random\_generator.choice(possible\_values, size = 100000)

**Distribución Uniforme Continua**:







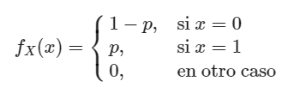
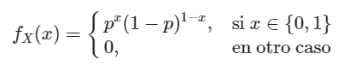
**En Python**:

random\_generator = np.random.default\_rng()

random\_uniform\_cont\_data = random\_generator.uniform(low = 3, high = 17, size = 100000)

**Distribución de Bernoulli**:

Distribución de Probabilidad **discreta**. Es un experimento aleatorio con sólo 2 resultados posibles: **éxito o fracaso**. Probabilidad de éxito: p; Probabilidad de Fracaso: 1-p.

 o bien 

Ejemplos: Prueba de si un tratamiento médico es o no efectivo. Lanzamiento de una moneda que de cara. Lanzar un dado y que salga 3.

**En Python**:

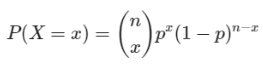
random\_generator = np.random.default\_rng()

sample\_size = 1000

random\_bernoulli\_data = random\_generator.binomial(n=1, p = 0.7, size = sample\_size)

**Distribución Binomial:**

Distribución de Probabilidad **Discreta**. Mide el número de éxitos en una secuencia de ensayos independientes entre sí, con una probabilidad de ocurrencia fija p del éxito entre los ensayos. Si hacemos n intentos y buscamos tener x éxitos, para una probabilidad de éxito p en cada prueba:



**En Python**:

random\_generator = np.random.default\_rng()

sample\_size = 1000

random\_bernoulli\_data = random\_generator.binomial(n=1, p = 0.7, size = sample\_size)

**Notar que la distribución Bernoulli es una distribución Binomial con n = 1**.

IE: Pacientes en un grupo de 10 para los que un tratamiento fue efectivo; caras al lanzar una moneda 5 veces; 3 al lanzar un dado 10 veces.

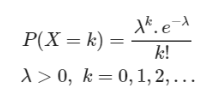
**Distribución de Poisson**:

Cuenta la **cantidad de eventos en un período de tiempo dado**. Probabilidad de que ocurra un determinado número de eventos en cierto período de tiempo. X es el número de veces en que un evento determinado ocurre dentro de un intervalo de tiempo definido. Es una distribución de probabilidad **discreta**.

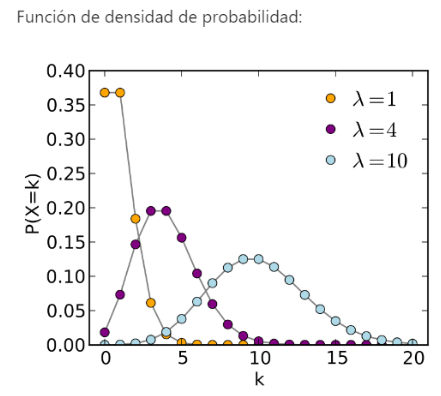
Implica 3 condiciones:

1. El **número de eventos** que ocurren en períodos de tiempo, **sin superposición** entre períodos, es **independiente.**
2. La probabilidad de exactamente un evento en un intervalo de duración corto h = 1/n es aproximadamente h \* λ; siendo n la cantidad de intervalos dentro del período considerado.
3. La probabilidad de exactamente 2 o más eventos en un intervalo de tiempo corto es esencialmente cero.

X es una variable aleatoria que sigue un proceso de Poisson aproximado con parámetro λ > 0.



λ es la media y la varianza de una variable de Poisson.



Ejemplos: Pacientes que llegan a la guardia de un hospital en una hora; autos que pasan por una cabina de peajes en un día; llamados que llegan a un operador en un call center durante la mañana.

**En Python**:

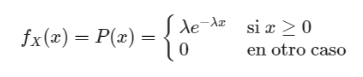
random\_generator = np.random.default\_rng()

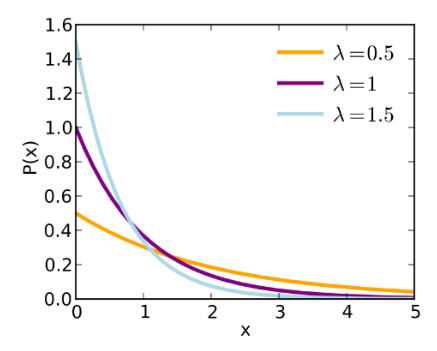
lambda\_value = 10

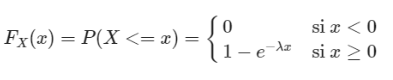
sample\_size = 10000

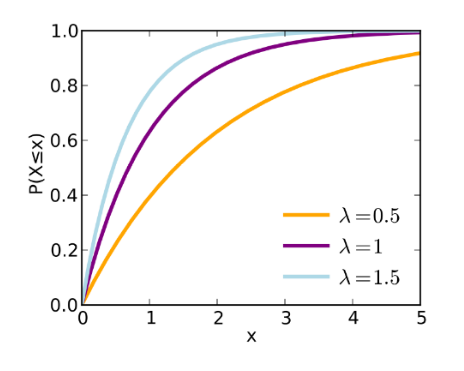
random\_poisson\_data = random\_generator.poisson(lam=lambda\_value, size = sample\_size)

**Distribución Exponencial**: Es una distribución de probabilidad **continua**. Describe el **tiempo** entre eventos de un evento de Poisson.









Ejemplos: Tiempo entre ingreso de pacientes seguidos en una guardia de un hospital en una hora; Tiempo entre llegada de dos autos consecutivos a una estación de peaje; tiempo entre llamados consecutivos que llegan a un operador de call center por la mañana.

**En Python**:

random\_generator = np.random.default\_rng()

lambda\_value = .5

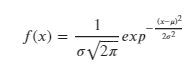
beta\_value = 1.0 / lambda\_value

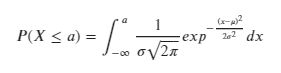
sample\_size = 100000

random\_exponential\_data\_1 = random\_generator.exponential(scale=beta\_value, size = sample\_size)

**Distribución Normal**: Las distribuciones de muchos fenómenos naturales están, al menos aproximadamente, distribuidos normalmente.

**Historia:** Empezó aplicándose al medir errores de medición en observaciones astronómicas. Galileo (S XVII) descubrió que dichos errores eran simétricos, y que los errores pequeños ocurrían con mayor frecuencia que los grandes. Recién en el S XIX se descubrió que estos errores seguían una distribución normal. En 1808 y 1809, Adrian y Gauss desarrollaron por separado la fórmula para la distribución normal y mostraron que los errores se ajustaban bien a la misma.

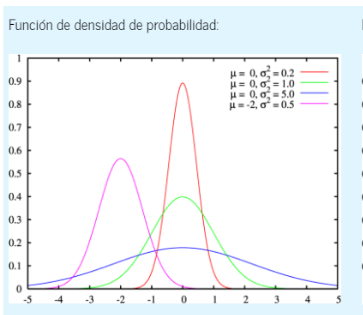


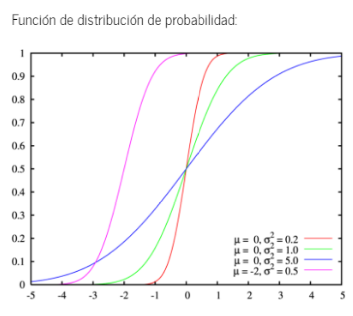


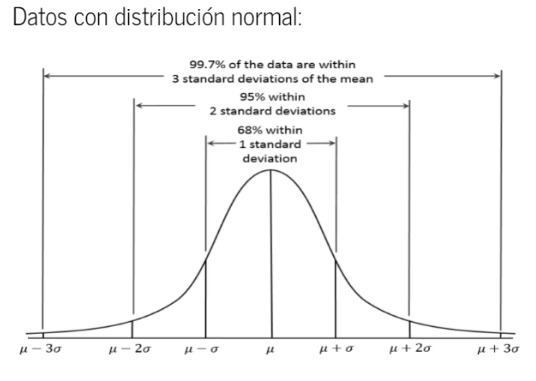
Cuando la media es 0 y el desvío estándar es 1, decimos que se trata de una normal estándar.

**Características:**

* Cualquier función lineal de una variable aleatoria normal también será normal. Dada X como una variable aleatoria con parámetros µ y , podemos definir Y = ax + b; Y tendrá también una distribución normal.
* La variable aleatoria generada por una suma finita de variables aleatorias normales independientes es normal.
* En caso de sumar infinitas variables aleatorias, independientes y tales que tanto la suma de las medias como de las varianzas sean finitas; la variable resultante tendrá una distribución normal.







µσ abarca el 68% de los casos

µσ abarca el 95% de los casos

µσ abarca el 99,7% de los casos

Ejemplos de distribuciones normales: Longitud de clavos producidos en una fábrica; altura de alumnos del curso de DH; Tiempo de vida de un modelo de lámpara; Tiempo que usa un empleado para hacer una tarea determinada en una empresa.

**En Python**:

random\_generator = np.random.default\_rng()

n = 100000

media = 1000

varianza = 200

random\_normal\_data = random\_generator.normal(loc=media, scale=np.sqrt(varianza), size=n)

**Relación entre Distribuciones**

**Bernoulli y Binomial:** La distribución Bernoulli es una distribución Binomial con n = 1.

**Poisson y Binomial:** Poisson es una distribución binomial con **n muy grande**; con probabilidad de éxito de cada repetición **p** siendo la misma e **infinitesimalmente pequeña** y con **np = λ finito**.

**Normal y Binomial:** La distribución normal es el límite de una distribución binomial con **n muy grande**, **p (probabilidad de éxito) y q (probabilidad de fracaso) no infinitamente pequeñas** simultáneamente. Se tiende a una distribución normal de parámetros

µ = n \* p

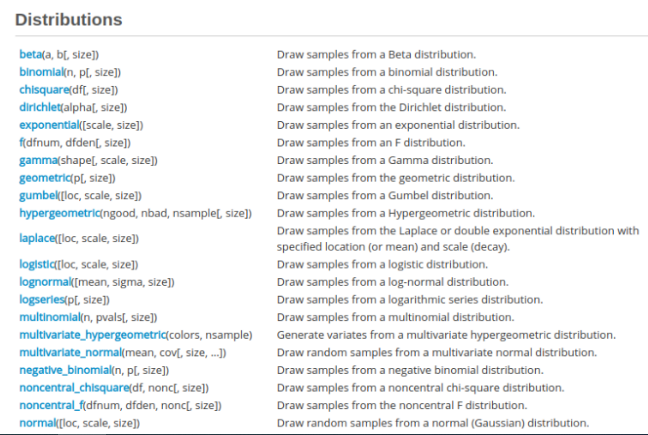
**Normal y Poisson:** La Normal es el límite de una distribución de Poisson con λ tendiendo a infinito.

**Exponencial y Poisson:** Si el tiempo entre eventos sigue una distribución exponencial de parámetro λ, entonces la cantidad de eventos en un período de tiempo t sigue una distribución de Poisson con parámetro λt.

Las distribuciones son modelos estadísticos probabilísticos cuyas propiedades se usan para analizar datos empíricos: Se recopilan los datos y se comparan con cualquier distribución teórica conocida. En caso de que sean similares, se transfieren las propiedades del modelo teórico a los datos empíricos con las conclusiones correspondientes (cálculos de intervalos de confianza, comparación de valores promedio, verificación de la importancia de los parámetros, etc). Si no coinciden, se recomienda **NO** usar estos modelos estadísticos.

Recordemos que con Python, random\_generator = np.random.default\_rng(###) genera números aleatorios. Si definimos un número puntual en ###, dicho número aleatorio será siempre el mismo.

Se pueden usar diferentes métodos sobre el objeto random\_generator para generar distribuciones:



IE: Una distribución de Poisson con λ = 10.

Random\_Poisson\_data\_10 = random\_generator.poisson(lam=10, size = 10000) Al hacer λ cada vez más grande puede verse que se va aproximando cada vez más a una distribución normal.

**Función Distribution Plotter:**

import seaborn as sns

**Definición de Distribution Plotter:**

def distribution\_plotter(data, label, color = 'steelblue',

bins='auto', binrange=None, binwidth=0.5 ):

sns.set(rc={"figure.figsize": (10, 7)})

sns.set\_style("white")

dist = sns.histplot(data, bins= bins, stat = 'density', kde = True, line\_kws={'linewidth':5},

color = color, thresh = None,

binrange = binrange, binwidth=binwidth)

dist.set\_title('Distribucion ' + label + '\n', fontsize=16)

**Ejemplo de uso de Distribution Plotter:**

distribution\_plotter(random\_uniform\_cont\_data, "uniforme continua", binrange=(0,20))