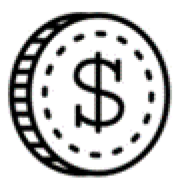
Distribuciones discretas

Distribución de Bernoulli

Variables con ocurrencias binarias.
De aquí podemos regresar a los ejemplos con lanzamiento de monedas. También entonces dependiendo de si la moneda es justa; es decir 50% y 50%o si tiene otra probabilidad.

$$P(X = 1) = p$$

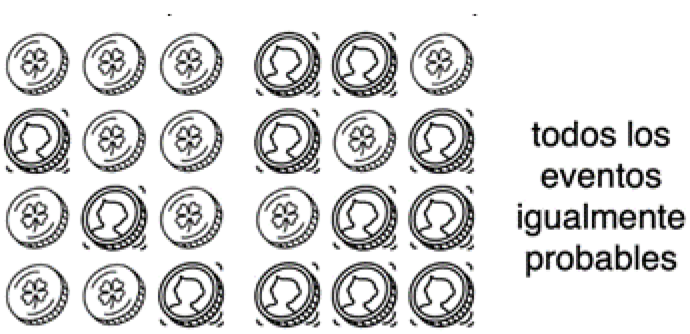


$$P(X = 0) = 1 - p$$

Así que en resumen describimos una función de Bernoulli como una función que a la variable le asigna 2 valores. Si X=1 siendo 1 igual a cara, entonces hay una probabilidad p. Que típicamente seria 0.5 en el caso equilibrado y cuando X=0, que sería el caso opuesto, entonces la probabilidad se define como 1-p; porque la suma de las probabilidades tiene que dar 1.

Luego podemos empezar a considerar situaciones mas complejas con base al ejercicio anterior. ¿cual ejercicio? Lanzar monedas.

Lanzando mas monedas, se vuelve mas compleja la situación. Y entonces cuando tenemos secuencias repetitivas de varios eventos binarios son llamadas eventos tipo **Bernoulli**, entonces ahora es cuando hablamos de la tan famosa **Distribución Binomial**



$$P(2 \text{ caras}|3 \text{ lanzamientos}) = 3/8$$

Aquí explicaremos el contenido esencial de este tipo de distribución binomial.

Tenemos supongamos 3 lanzamientos consecutivos, entonces ¿cuales son los posibles resultados? Este experimento tiene un espacio muestral de 8 posibilidades y son los que se ven en la diapositiva.

Otra situación que nos podríamos preguntar es: ¿cual es la probabilidad de que de esos 3 lanzamientos 2 sean cara? Es decir de cuantas maneras pueden suceder esto, del espacio muestral que tenemos (8) solo tenemos 3 eventos en los cuales tenemos 2 caras, entonces la probabilidad es $\frac{3}{8}$. Pero claro también estoy asumiendo que cada una de las 8 secuencias es igualmente probable, es decir de las que venimos trabajando desde hace unas clases; es decir hay probabilidades axiomáticas y que son igualmente probables. Y en el caso en que en las distribuciones no suceda este hecho de eventos igualmente probables, es cuando definimos un numero p (p minúscula) que lo que hace es decirme la probabilidad de uno o de otro, esto es una aplicación de de darle un ajuste a los datos de manera cuando trabajamos en la realidad.

Si complicamos mas la situación generalizando, es decir no solo preguntándome por 3 lanzamientos, sino por n lanzamientos y de esos lanzamientos tener k cantidad de caras. Tenemos que

$$P(\text{k caras} \mid \text{n lanzamientos})$$

Aquí ya complicamos mas la cosas, porque entre mas lanzamientos, el espacio muestral crece también y crece mucho más rápido de lo exponencial, de aquí nos preguntamos si existe una formula para contar este tipo de estados. Claro que si existe y de eso se trata específicamente la distribución binomial.

de 3 lanzamientos k caras

Recordemos nuevamente que cuando tenemos variables discretas tenemos una representación gráfica de un histograma, donde cada barra representa la frecuencia relativa de cada evento posible.

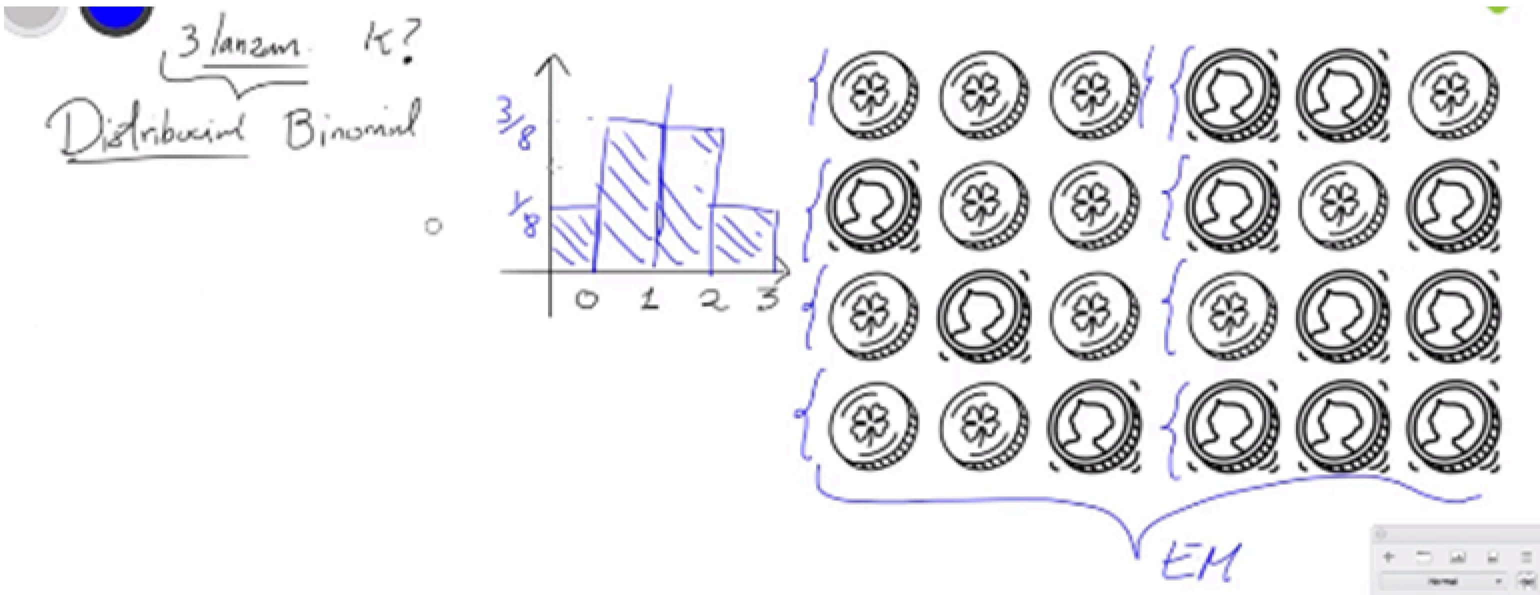
Entonces los eventos posibles son, que obtenga:

- 0 caras
- 1 cara
- 2 caras
- 3 caras

Esas son las 4 posibilidades sobre las cuales yo debo calcular las frecuencias relativas, entonces de todos los eventos (8). Ahora analicemos o desglosemos como se comporta el espacio muestral con respecto de los eventos de caras:

- 0 caras: 1/8
- 1 cara: 3/8

- 2 caras: 3/8
- 3 caras: 1/8



Entonces la fórmula esta dada por:

$$p(k,n) = ?$$

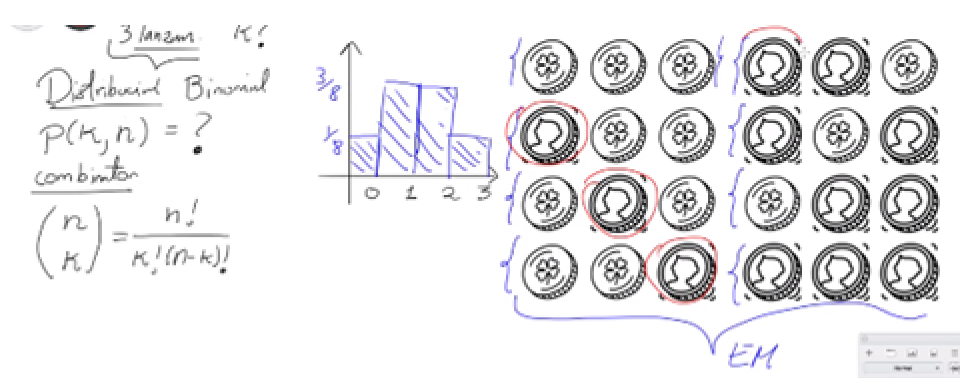
- n = lanzamientos
- k = éxitos

Combinatorio

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Dados "n" lanzamientos que yo obtenga "k" éxitos a partir de las opciones y que esa manera de éxitos puede suceder de todas las maneras posibles, porque no tiene que suceder siempre de la misma manera ordenada.

¿Que quiere decir esto? Lo que sucede es que, el mismo eventos sucede de muchas formas, entonces ¿cual evento? por ejemplo que yo obtenga 1 cara.



Lo mismo sucede con 2 caras, puede suceder que salga consecutivas, en los 2 primeros lugares, en los 2 últimos. Pero sin tener en cuenta el orden, los eventos que se presentan son básicamente el mismo y por eso se cuentan dentro de la misma probabilidad y a su vez se ve reflejado en el histograma. Entonces la formula anterior nos da esa forma de conteo de los sucesos, donde implica el factorial.
Donde $n!$ es:

$$n! = 1 \cdot 2 \cdot 3 \cdot 4 \ldots n$$

Y así podemos contar los estados.
Por ejemplo quiero obtener la probabilidad de que me salga **1** cara dados **3** lanzamientos $p(1,3)$

$$\binom{3}{1} = \frac{3!}{1!(3-1)!} = \frac{1 \cdot 2 \cdot 3}{1 \cdot (1 \cdot 2)} = 3$$

Entonces de 3 formas puedo obtener el mismo resultado y la probabilidad será $p(1,3) = \frac{3}{8}$ Ocho, porque es el espacio muestral o estados posibles.

Formula de distribución binomial

probabilidad del suceso $\rightarrow$ # estados del suceso x probabilidad de cada estado

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

- n = lanzamientos
- k = éxitos

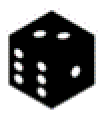
$$P(\text{k caras}|\text{n lanzamientos}) = P(k; n, p)$$

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

Así la distribución binomial es donde puede calcular de una secuencia de eventos tipo Bernoulli, cuantos éxitos yo puedo tener.

Esta no es la única distribución para una variable aleatoria binaria, pueden existir otras como **Distribución multinomial**

$$P(X_1, \ldots, X_n) = \frac{n!}{k_1! \ldots k_2!} p_1^{k_1} \ldots p_n^{k_n}$$



Distribución multinomial: Es aquella que es la generalización natural de la distribución binomial. ¿En donde se puede usar? en el caso del lanzamiento de un dado.

Pero también lo importante es entender que existen otras distribuciones discretas como:

- Poisson.
- Geométrica
- Hipergeométrica
- Binomial negativa
- ...

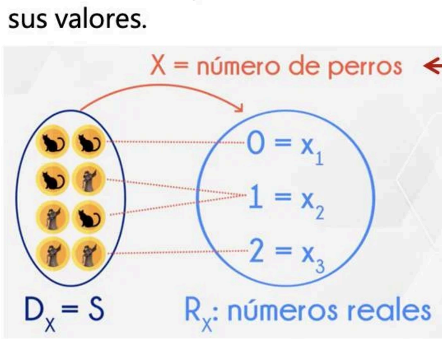
Te preguntaras en que caso utilizo cada distribución. La experiencia nos demuestra que hay ciertos experimentos aleatorios donde ciertas distribuciones aplican y en las próximas clases veras que cuando tienes un conjunto de datos particular, donde tu a priori no sabes cual es la distribución, existen técnicas para ajustar la mejor distribución de probabilidad a un conjunto de datos que conozcamos, porque en la vida real, nosotros no conocemos en principio las distribuciones de probabilidad de un conjunto de datos, tenemos que aprenderlas. Entonces tenemos un algoritmo que aprende esas distribuciones, para nosotros a partir de los datos, eso es Machine Learning solo que con un enfoque probabilístico. Entonces en la siguiente clase entraremos en código, usando Google Colab y profundizando sobre aspectos muy importantes de la distribución binomial, como ejemplo de las distribuciones discretas

EXTRA

VARIABLE ALEATORIA:

-Es una función que asocia un número real (valor numérico) con cada elemento del espacio muestral (resultado de un experimento aleatorio).

-Se utiliza "X" para denotar una variable aleatoria y "x" para cada uno de sus valores.

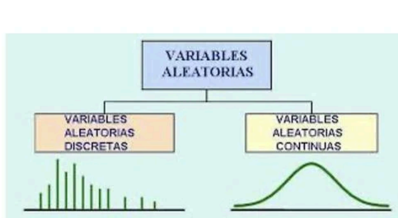


Plan de muestreo

Nota: La variables aleatoria en la que se eligen 0 y 1 para los 2 posibles valores se denomina "Varibale aleatoria de Bernulli"

-Las variables aleatorias discretas: Son aquellas que podemos cuantificar en un intervalo finito de números, las podemos cuantificar con números naturales.

-Las variables aleatorias continuas: Son aquellas que cuantificamos en un intervalo infinito de números, las podemos cuantificar con todo el conjunto de números reales.

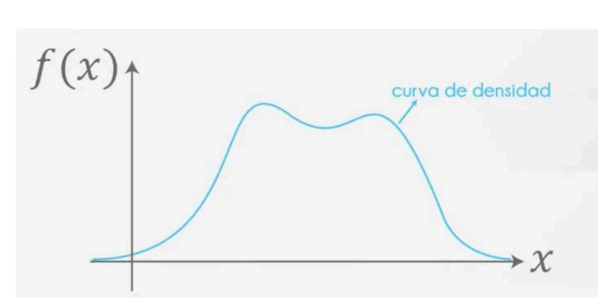


DISTRIBUCIONES DE PROBABILIDAD

Es una función que toma una variable aleatoria y a cada uno de sus posibles estados en un espacio muestral, le asigna una probabilidad.

-f(x) = P(X=x), represnta es probabilidad de ocurrencia de los valores x.

-El conjunto de pares oredenados (x, f(x)) forman la Función de densidad de probabilidad o Distribución de probabiliada



Toda variable aleatoria continua ***X*** tiene una función de densidad de probabilidad ***f***. Esto significa que la probabilidad de que ***X*** se encuentre entre ***a*** y ***b*** se encuentra integrando ***f*** desde ***a*** hasta ***b:***

Condiciones:

1. *f(x) ≥ 0, para toda x ∈ R.*
2. $\int_{-\infty}^{+\infty} f(x)\,dx = 1$

$$P(a \leq X \leq b) = \int_a^b P(X)dx = P(b) - P(a) = C(X)$$

Distribución Binomial

$$P(r) = \frac{n!}{r!(n-r)!}p^r q^{n-r}$$

n=Número de ensayos/experimentos

r=Número de éxitos

p=Probabilidad de éxito

q = Probabilidad de fracaso (1-p)

@DIEGOARAQUE

Propiedades

- En cada ensayo, experimento o prueba solo son posibles dos resultados (éxito o fracaso).
- La probabilidad del éxito ha de ser constante. Esta se representa mediante la letra p.
- La probabilidad de fracaso ha de ser también constate. Esta se representa mediante la letra q = 1-p.
- El resultado obtenido en cada experimento es independiente del anterior. Por lo tanto, lo que ocurra en cada experimento no afecta a los siguientes.
- Los sucesos son mutuamente excluyentes, es decir, no pueden ocurrir los 2 al mismo tiempo.
- Los sucesos son colectivamente exhaustivos, es decir, al menos uno de los 2 ha de ocurrir.
- La variable aleatoria que sigue una distribución binomial se suele representar como X~(n,p), donde n representa el número de ensayos o experimentos y p la probabilidad de éxito.

Ejemplo

Imaginemos que un 80% de personas en el mundo ha visto el partido de la final del último mundial de fútbol. Tras el evento, 4 amigos se reúnen a conversar, ¿Cuál es la probabilidad de que 3 de ellos hayan visto el partido?

n = 4 (es el total de la muestra que tenemos)
x = 3
p = probabilidad de éxito (0,8)
q = probabilidad de fracaso (0,2). Este resultado se obtiene al restar 1-p.

$$P(3)=0.4096 \longleftarrow P_{(3)} = \frac{4!}{3!\,(4-3)!}0{,}8^3 0{,}2^{4-3}$$