Análisis y Visualización de Datos

Clase 1

Conceptos de Probabilidades

Docentes : Soledad Palacios (UNLP)

Milagro Teruel (UNC)

¿De qué hablamos cuando decimos que algo es probable?

La **Teoría de Probabilidades** estudia los llamados experimentos aleatorios.
Un **experimento aleatorio** tiene las siguientes características:

1- Se lo puede repetir bajo las mismas condiciones tantas veces como se desee.

2- No se puede predecir con exactitud el resultado de dicho experimento, pero se puede decir cuáles son los posibles resultados del mismo.

¿De qué hablamos cuando decimos que algo es probable?

3- A medida que el experimento se repite, los resultados individuales parecen ocurrir en forma caprichosa. Pero si el experimento se repite un gran número de veces, y registramos la proporción de veces que ocurre un determinado resultado, veremos que esa proporción tiende a estabilizarse en un valor determinado a medida que aumenta el número de veces que se repite el experimento.

“No pretendamos que las cosas cambien si siempre hacemos lo mismo”

Albert Einstein

¿Ejemplos?

a) tirar un dado y observar el número en la cara de arriba.

b) El pronóstico meteorológico

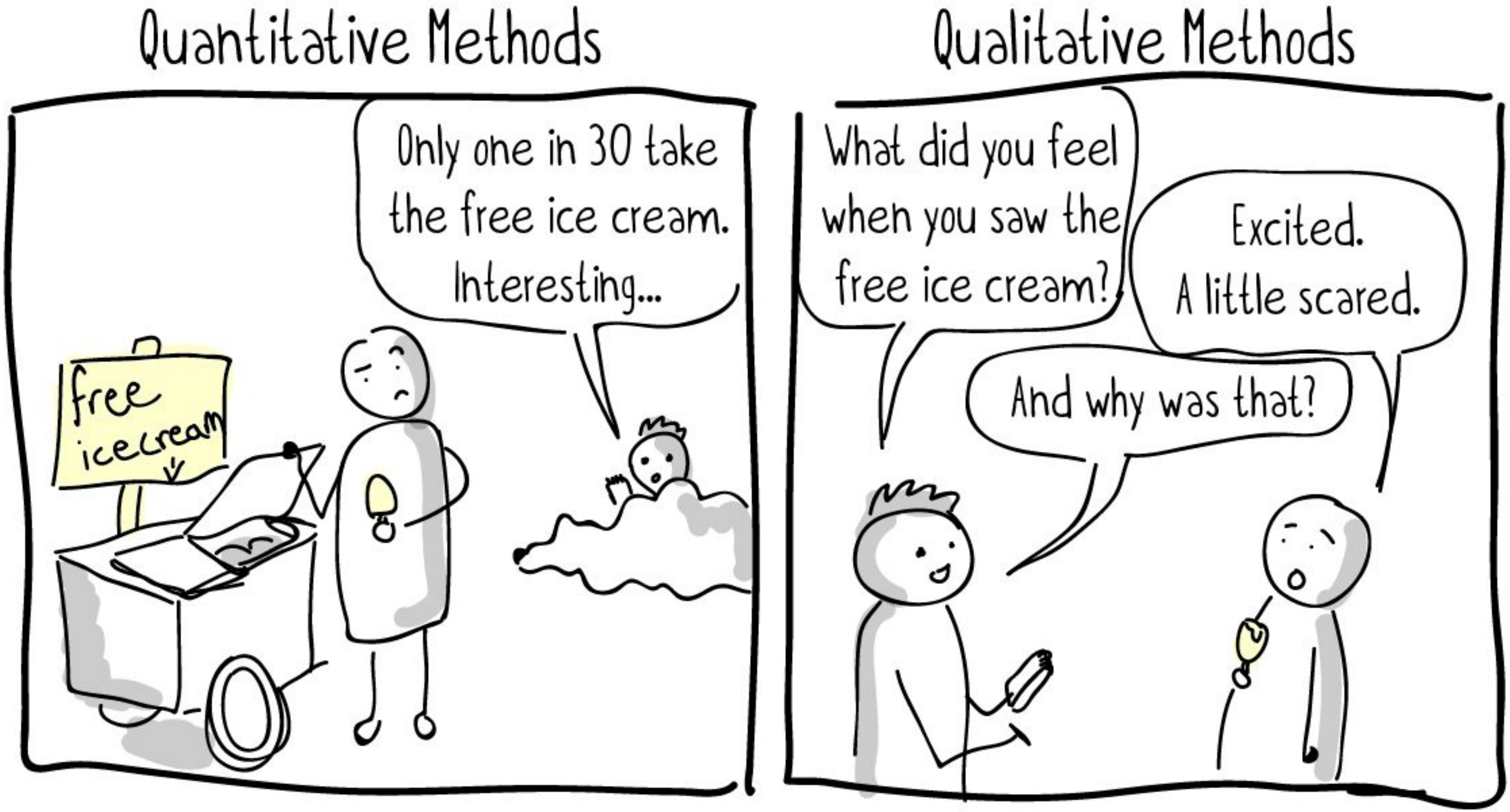
En los experimentos no aleatorios o deterministas se puede predecir con exactitud el resultado del experimento, es decir, las condiciones en las que se verifica un experimento determinan el resultado del mismo

¿Ejemplos?

A veces sucede que un experimento no es aleatorio estrictamente, pero resulta mucho más sencillo estudiarlo como si fuera aleatorio

¿Qué variables deberíamos conocer con anterioridad para predecir de qué lado cae la moneda?

Tipos de datos



Credit: [A crash course in UX design research]

Tipos de datos

- Numéricos (discretos y continuos)
- Categóricos
- Ordinales

Datos Cuantitativos

Los **datos cuantitativos** son datos que miden o calculan un algo para llegar a un punto en su investigación. Estos datos nos dicen a través de números una explicación para alguna tendencia o resultados de algún experimento.

Con los datos cuantitativos, se puede hacer todo tipo de tareas de procesamiento de datos numéricos, tales como sumarlos, calcular promedios, o medir su variabilidad.

Datos Cuantitativos

Los **datos discretos** solo van a poder asumir un valor de una lista de números específicos.

Representan ítems que pueden ser contados; todos sus posibles valores pueden ser listados.

Suele ser relativamente fácil trabajar con este tipo de dato.

Los **datos continuos** representan mediciones; sus posibles valores no pueden ser contados y sólo pueden ser descritos usando intervalos en la recta de los números reales.

Datos Cualitativos o Categóricos

Si los datos nos dicen en cual de determinadas categorías no numéricas nuestros ítems van a caer, entonces estamos hablando de datos cualitativos o categóricos; ya que los mismos van a representar determinada cualidad que los ítems poseen

Datos Ordinales

Una categoría intermedia entre los dos tipos de datos anteriores, son los datos ordinales. En este tipo de datos, va a existir un orden significativo, vamos a poder clasificar un primero, segundo, tercero, etc. es decir, que podemos establecer un ranking para estos datos, el cual posiblemente luego tenga un rol importante en la etapa de análisis. Los datos se dividen en categorías, pero los números colocados en cada categoría tienen un significado..

Ej: Puntuación de estrellas

Nos acercamos a los datos

```
dataset = pandas.read_csv('../violencia-institucional-2018-01.csv', encoding='utf8')
```

```
dataset[:3]
```

	area	organismo_origen	via_acceso	year	provincia	contexto	contexto1	circunstancia	alojamiento	violencia_fisica	violencia_psiquica
0	DNPCVI	SECRETARIA DE DDHH	Telefónica	2017.0	Buenos Aires	Situaciones de Detención	Penal / Complejo Penitenciario PROVINCIAL	NaN	NaN	NaN	NaN
1	DNPCVI	SECRETARIA DE DDHH	Telefónica	2017.0	Buenos Aires	Situaciones de Detención	Penal / Complejo Penitenciario PROVINCIAL	NaN	Malas condiciones de alojamiento (higiene), Hu...	NaN	NaN
2	DNPCVI	SECRETARIA DE DDHH	Telefónica	2017.0	Buenos Aires	Situaciones de Detención	Penal / Complejo Penitenciario PROVINCIAL	NaN	Malas condiciones de alojamiento (higiene)	NaN	NaN

Datos categóricos - Ordinales?

Leemos el dataset

```
dataset = pandas.read_csv('../violencia-institucional-2018-01.csv', encoding='utf8')
```

```
dataset[:3]
```

	area	organismo_origen	via_acceso	year	provincia	contexto	contexto1	circunstancia	alojamiento	violencia_fisica	violencia_psiquica
0	DNPCVI	SECRETARIA DE DDHH	Telefónica	2017.0	Buenos Aires	Situaciones de Detención	Penal / Complejo Penitenciario PROVINCIAL	NaN	NaN	NaN	NaN
1	DNPCVI	SECRETARIA DE DDHH	Telefónica	2017.0	Buenos Aires	Situaciones de Detención	Penal / Complejo Penitenciario PROVINCIAL	NaN	Malas condiciones de alojamiento (higiene), Hu...	NaN	NaN
2	DNPCVI	SECRETARIA DE DDHH	Telefónica	2017.0	Buenos Aires	Situaciones de Detención	Penal / Complejo Penitenciario PROVINCIAL	NaN	Malas condiciones de alojamiento (higiene)	NaN	NaN

Datos numéricos - Continuos o discretos?

Leemos el dataset

```
In [41]: poblacion[:3]
```

Out[41]:

	Provincia	Población 2001	Población 2010	Variación absoluta	Variación relativa (%)
0	Ciudad de Buenos Aires	2.776.138	2.890.151	114.013	4,1
1	Buenos Aires	13.827.203	15.625.084	1.797.881	13,0
2	Catamarca	334.568	367.828	33.260	9,9

Datos categóricos

Datos numéricos

Discretos

Datos numéricos

Continuos

El Espacio Muestral

El conjunto de todos los resultados posibles de un experimento aleatorio es el espacio muestral. Al espacio muestral lo anotamos con la letra S.

1- La elección de S no es única, depende de lo que se quiera observar del experimento aleatorio. Ejemplos? pensemos en un simple dado...

2- El espacio muestral puede ser un conjunto finito, o infinito. A su vez si es infinito puede ser infinito numerable o no numerable.

Que implican los infinitos?

Los conjuntos infinitos tienen propiedades muy especiales, pero, entre otras, la que atenta contra la intuición es que un subconjunto “más pequeño”, “incluido” dentro de un conjunto, puede contener el mismo número de elementos que el todo.

Ejemplos de conjuntos infinitos numerables

Los números naturales: no sabemos cuántos hay (sabemos que son infinitos pero infinito no es un numero), aunque podemos decir cuantos naturales hay entre un numero natural y otro

Hay variables aleatorias cuyo rango son todos los números reales de un intervalo dado, (es decir es un conjunto infinito no numerable). Ejemplos de variables continuas podrían ser

- X: “tiempo que tarda en llegar un colectivo a una parada”
- Y: “tiempo de vida de un fusible”

El espacio Muestral

a) Si ε : tirar un dado y observar el número en la cara de arriba, entonces podemos tomar como espacio muestral a S = { 1 , 2 , 3 , 4 , 5 , 6 }

b) Si ε : tirar una moneda, entonces S = { c , s }

c) Si ε : lanzar una moneda tres veces y contar el número total de caras obtenidas entonces podemos considerar S = { 0 , 1 , 2 , 3 }

El espacio Muestral

d) Si ε : lanzar una moneda tres veces y observar la sucesión de caras y cecas

e) Si ε : tirar un dado las veces necesarias hasta que sale un 6 por primera vez, y contar el número de tiros realizados, entonces $S = \{ 1 , 2 , 3 , 4 ,..... \} = N$, donde N es el conjunto de los números naturales.

f) Si ε : medir el tiempo de vida de una lamparita eléctrica, entonces $S = \{ t \in R , t \geq 0 \}$ donde R es el conjunto de los números reales.

Variables Aleatorias

En muchas situaciones experimentales se quiere asignar un número real a cada uno de los elementos del espacio muestral.

Definición

Sea ε un experimento aleatorio y S un espacio muestral asociado a él. Una variable aleatoria es una función que asigna a cada elemento de S un número real.

Notación se anota a una variable aleatoria con letras mayúsculas X, Y, Z, W,...

Rango o recorrido

Dada una v.a. X a su imagen se la anota R_X y se la denomina rango o recorrido de X

Las variables aleatorias se clasifican según su rango.

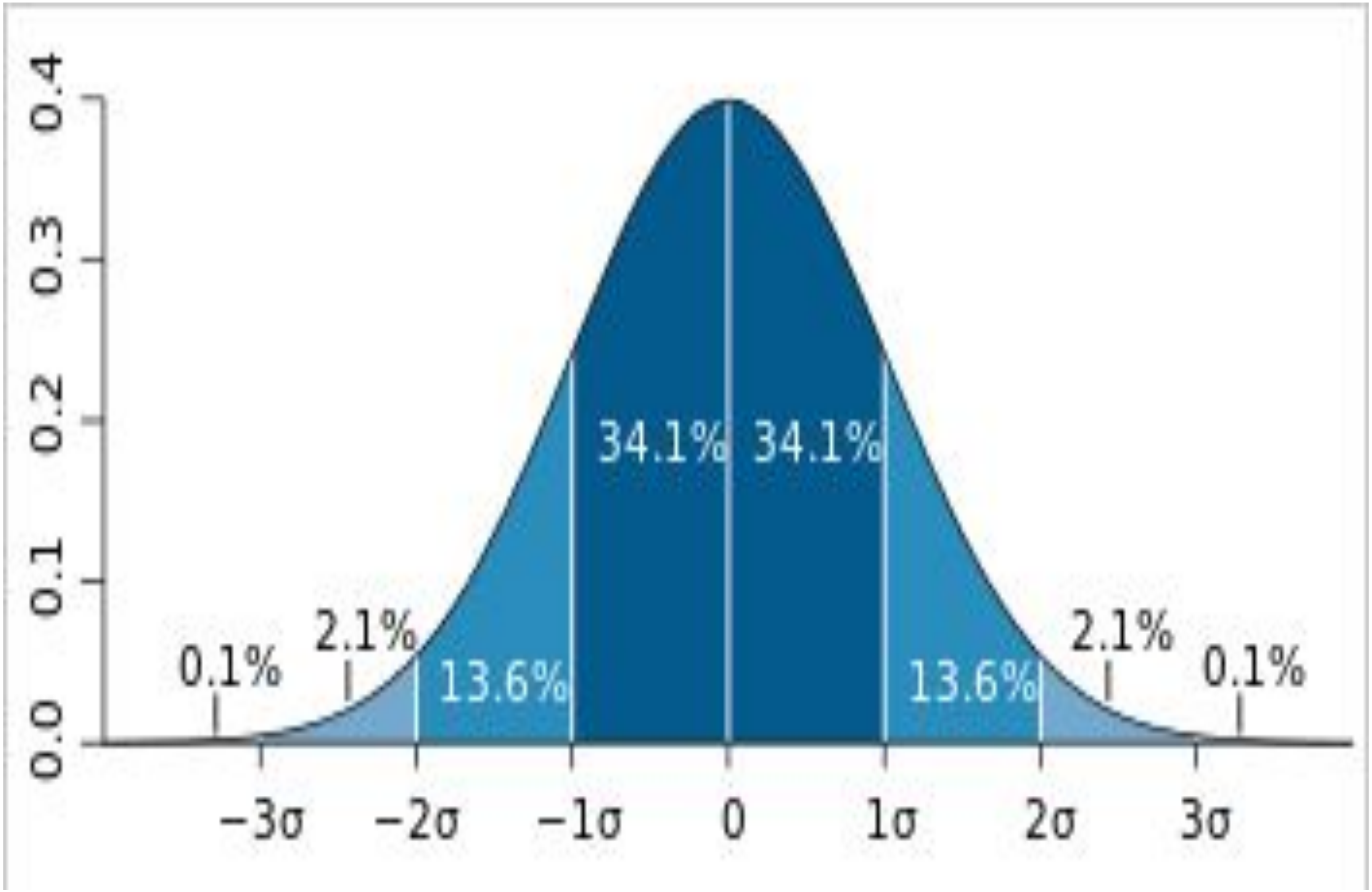
Sea X es una v.a. con rango R_X. Si R_X es un conjunto finito o infinito numerable entonces se dice que X es una v.a. discreta. Si R_X es un conjunto infinito no numerable entonces X es una v.a. continua.

El rango R_X es considerado un nuevo espacio muestral , y sus subconjuntos son eventos .

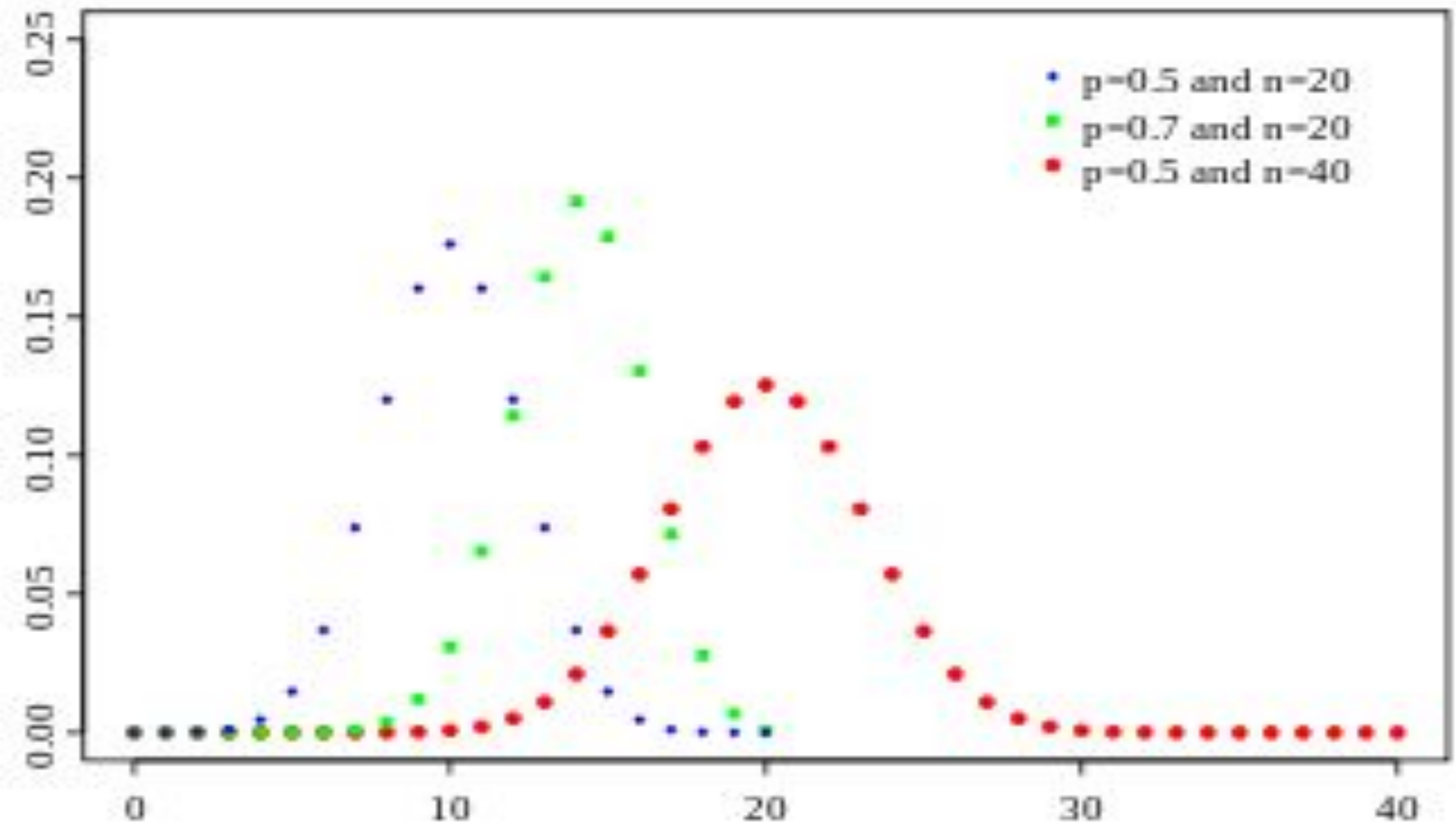
Frecuencia de Distribución de Probabilidad

La **distribución de probabilidad** de una variable aleatoria es una función que asigna a cada suceso definido sobre la variable la probabilidad de que dicho suceso ocurra. La distribución de probabilidad está definida sobre el conjunto de todos los sucesos y cada uno de los sucesos es el rango de valores de la variable aleatoria.

Frecuencia de Distribución de Probabilidad



Frecuencia de Distribución de Probabilidad



¿Cómo computarlo a partir de una muestra?

```
C, S = 'c', 's'
SAMPLE_SPACE = ['-'.join(x) for x in
                itertools.product([C, S], repeat=3)]
SAMPLE_SPACE
```

```
['c-c-c', 'c-c-s', 'c-s-c', 'c-s-s', 's-c-c', 's-c-s', 's-s-c', 's-s-s']
```

```
sampled_values = [
    x.count(C) for x in numpy.random.choice(SAMPLE_SPACE, 1000)]
```

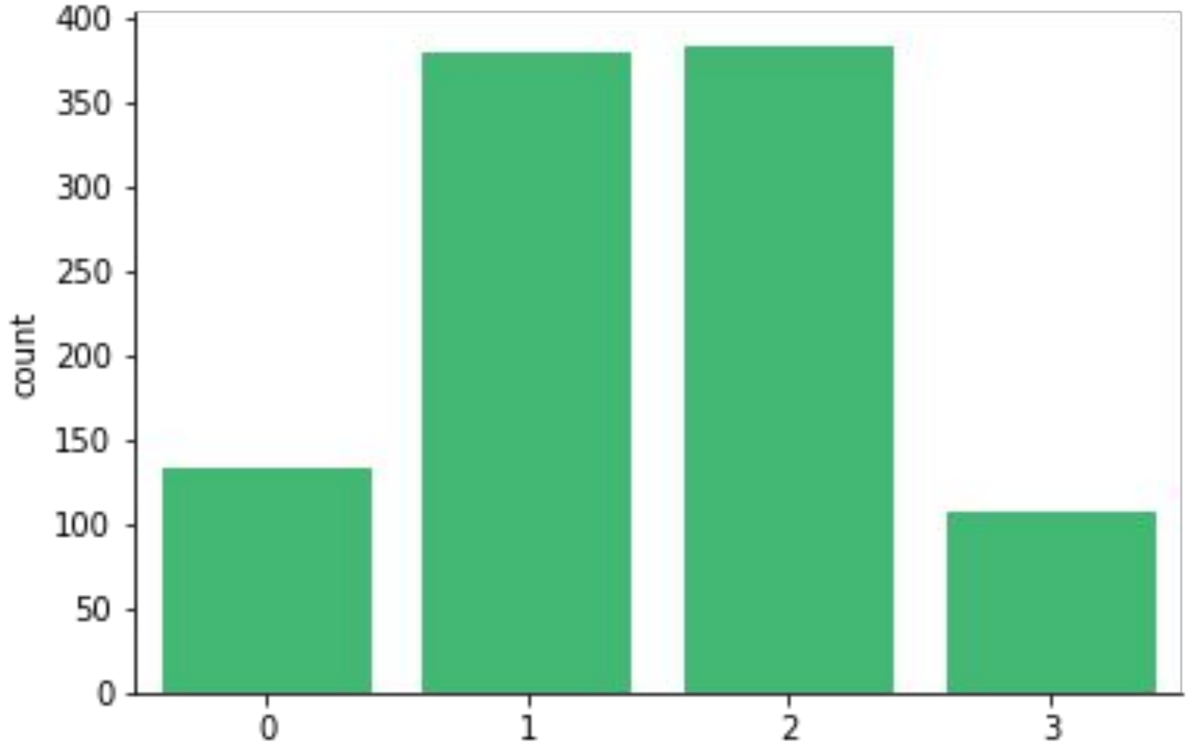
```
result = numpy.unique(sampled_values, return_counts=True)
[(label, count/1000.0) for label, count in zip(*result)]
```

```
[(0, 0.132), (1, 0.379), (2, 0.383), (3, 0.106)]
```

¿Cómo computarlo a partir de una muestra?

```
result = numpy.unique(sampled_values, return_counts=True)
[(label, count/1000.0) for label, count in zip(*result)]
```

```
[(0, 0.132), (1, 0.379), (2, 0.383), (3, 0.106)]
```



¿Cómo computarlo a partir de una muestra?

Utilizando el mismo dataset que vimos antes, queremos calcular la fdp de la v.a. **area**

	area	organismo_origen	via_acceso	year	provincia	contexto	contexto1
0	DNPCVI	SECRETARIA DE DDHH	Telefónica	2017.0	Buenos Aires	Situaciones de Detención	Penal / Complejo Penitenciario PROVINCIAL
1	DNPCVI	SECRETARIA DE DDHH	Telefónica	2017.0	Buenos Aires	Situaciones de Detención	Penal / Complejo Penitenciario PROVINCIAL
2	DNPCVI	SECRETARIA DE DDHH	Telefónica	2017.0	Buenos Aires	Situaciones de Detención	Penal / Complejo Penitenciario PROVINCIAL

¿Cómo computarlo a partir de una muestra?

Utilizando el mismo dataset que vimos antes, queremos calcular la fdp de la v.a. **area**

```
In [51]: areas, counts = numpy.unique(vinstitucional.area.values.astype(str),
                                      return_counts=True)
         total = counts.sum()
         fdps = [x/float(counts.sum()) for x in counts]
         [x for x in zip(areas, fdps)]

Out[51]: [('Centro de Denuncias', 0.3405142460041696),
          ('DNPCVI', 0.09937456567060458),
          ('Juridicos Internacional', 0.018068102849200834),
          ('Juridicos Nacional', 0.02154273801250868 7),
          ('Otros', 0.03613620569840167),
          ('PRONALCI', 0.4558721334259026),
          ('Salud Mental', 0.006254343293954135),
          ('Ulloa', 0.021542738012508687),
          ('nan', 0.0006949270326615705)]
```

¿Cómo computarlo a partir de una muestra?

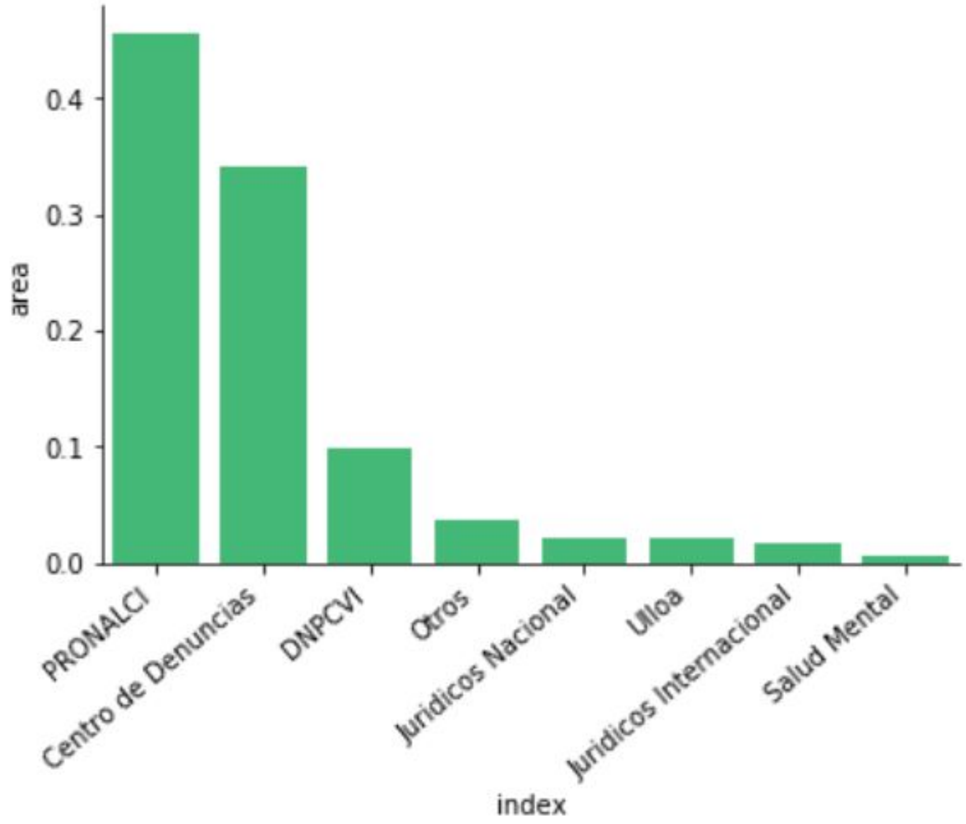
Utilizando el mismo dataset que vimos antes, queremos calcular la fdp de la v.a. **area**

```
In [48]: fdps = vinstitucional.area.value_counts(normalize=True)
         fdps
```

```
Out[48]: PRONALCI                   0.456189
         Centro de Denuncias        0.340751
         DNPCVI                     0.099444
         Otros                      0.036161
         Juridicos Nacional         0.021558
         Ulloa                      0.021558
         Juridicos Internacional    0.018081
         Salud Mental               0.006259
         Name: area, dtype: float64
```

¿Cómo computarlo automáticamente?

```
ax = seaborn.barplot(data=fdps.to_frame().reset_index(),
    x='index', y='area', color='#2ecc71')
ax.set_xticklabels(ax.get_xticklabels(), rotation=40, ha="right")
seaborn.despine()
```



Probabilidad Condicional

Supongamos el experimento aleatorio de extraer al azar sin reemplazo dos bolillas de una urna que contiene 7 bolillas rojas y 3 blancas.
Consideramos los eventos

A: “la primer bolilla extraída es blanca”

B: “la segunda bolilla extraída es blanca”.

Probabilidad Condicional

Calculamos la p(A) = 3 / 10

calculamos p (B)... aunque ahora ya no es tan directo.

¿Que cambió?

Podemos calcular la probabilidad de B sabiendo que A ocurrió :es igual a 2/9, ya que si A ocurrió, entonces en la urna quedaron 9 bolillas de las cuales 2 son blancas. La probabilidad anterior la anotamos P (B / A) y se lee:

“probabilidad condicional de B dado A. Es decir P (B / A) = 2/9”

Teorema de la multiplicación

Si A y B son dos eventos entonces

$P (B / A) = P (A \cap B) * P (A)$

entonces

$P (A \cap B) = P (A / B) * P(B)$

si $P (A) \neq 0$

Análogamente

$P (A \cap B) = P (A / B) * P(B)$

si $P (B) \neq 0$

Si A 1 , A 2 , A 3 son tres eventos entonces

$P (A 1 \cap A 2 \cap A 3) = P (A 1) P (A 2 / A 1) P (A 3 / A 1 \cap A 2)$

Se lee como la probabilidad que pase A1 y A2 y A3 es igual a la probabilidad que ocurra A1 por la probabilidad que ocurra A2 dado que ocurrió A1 por la probabilidad que ocurra A3 dado que ocurrieron A1 y A2

Teorema de la multiplicación

Esta regla es importante porque a menudo se desea obtener $P(A \cap B)$, en tanto que $P(B)$ y $P(A \mid B)$ pueden ser especificadas a partir de la descripción del problema.
La regla de multiplicación es más útil cuando los xperimentos se componen de varias etapas en sucesión. El evento condicionante B describe entonces el resultado de la primera etapa y A el resultado de la segunda, de modo que $P(A \mid B)$, condicionada en lo que ocurra primero, a menudo será conocida. La regla es fácil de ser ampliada a experimentos que implican más de dos etapas.

Ejemplos

Una clase tiene 12 niños y 4 niñas. Si se escogen tres estudiantes de la clase al azar, ¿cuál es la probabilidad de que sean todos niños?

Solución:

Anotamos A_i : "el i-ésimo estudiante elegido es un niño"

$i = 1,2,3$

Entonces la probabilidad pedida es

$P(A_1 \cap A_2 \cap A_3) =$

$P(A_1)\, P(A_2 / A_1)\, P(A_3 / A_1 \cap A_2) =$

$12/16 * 11/15 * 10/14$

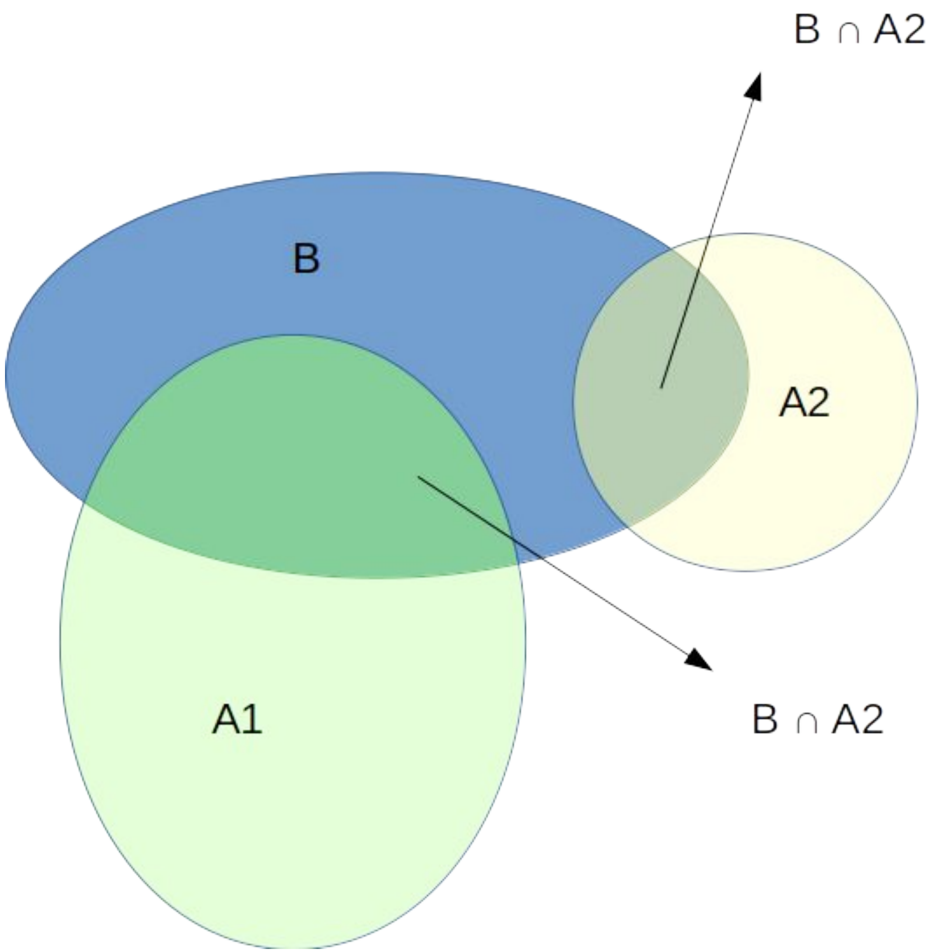
Teorema de la probabilidad total

Sean $A_1, A_2, \ldots, A_n$ eventos de un espacio muestral S que cumplen:

a) $A_1 \cup A_2 \cup \ldots \cup A_n = S$
b) $A_i \cap A_j = \varnothing$ si $i \neq j$
c) $P(A_i) > 0 \ \forall \ i = 1, 2, \ldots, n$ Se dice que $A_1, A_2, \ldots, A_n$ forman una partición de S Entonces para cualquier evento B de S

$$\begin{aligned} \mathbf{P(B)} = \ & \mathbf{P(B/A_1)P(A_1)} + \\ & \mathbf{P(B/A_2)P(A_2)} + \\ & \ldots + \mathbf{P(B/A_n)P(A_n)} \end{aligned}$$

Teorema de la probabilidad total



Ejemplo

El teorema responde a interrogantes como el siguiente

Tres máquinas A, B, y C producen respectivamente 60%, 30% y 10% del número total de artículos de una fábrica. Los porcentajes de desperfectos de producción de estas máquinas son respectivamente 2%, 3% y 4%. Se selecciona un artículo al azar

¿Cuál es la probabilidad de que sea defectuoso?

Solución

Sean los eventos:
A: “el artículo seleccionado fue producido por la máquina A”
B: “el artículo seleccionado fue producido por la máquina B”
C: “el artículo seleccionado fue producido por la máquina C”
D: “el artículo seleccionado es defectuoso”

Los datos que tenemos son los siguientes

$P(A) = 0.6$
$P(B) = 0.3$
$P(C) = 0.1$

$P(D/A) = 0.02$
$P(D/B) = 0.03$
$P(D/C) = 0.04$

Se pide hallar la $P(D)$

Probabilidad Total

Se aplica el teorema de la probabilidad total tomando como partición de S a los eventos A, B y C.
Entonces

$$P(D) =$$
$$P(D/A)P(A) + P(D/B)P(B) + P(D/C)P(C) =$$
$$0.02 \times 0.6 + 0.03 \times 0.3 + 0.04 \times 0.1 = \mathbf{0.025}$$

y finalmente aparece Bayes

De simple, la teoría resulta casi ridícula. Ayuda a la gente a evaluar sus ideas iniciales, actualizarlas y modificarlas con nueva información y a tomar mejores decisiones. En resumen, la regla de Bayes es muy breve y sencilla:

Creencias iniciales + datos objetivos recientes
=
Una nueva creencia mejorada.

Teorema de Bayes

Sean $A_1, A_2, \ldots, A_n$ eventos de un espacio muestral S que cumplen:

a) $A_1 \cup A_2 \cup \ldots \cup A_n = S$

b) $A_i \cap A_j = \varnothing$ si $i \neq j$

c) $P(A_i) > 0 \ \forall \ i = 1, 2, \ldots, n$

Entonces para cualquier evento B de S tal que $P(B) > 0$

Teorema de Bayes

$$P(A_i/B) = \frac{P(A_i) \cdot P(B/A_i)}{P(B)}$$

Donde:

$P(A_i)$ = Probabilidad a priori

$P(B/A_i)$ = Probabilidad condicional

$P(B)$ = Probabilidad Total

$P(A_i/B)$ = Probabilidad a posteriori

Aplicaciones

Aplicaciones

- filtrar correo basura,
- evaluar riesgos médicos o de otro tipo,
- buscar las páginas que nos interesan en Internet y descubrir lo que quizás nos interesaría comprar, basándonos en lo que hemos buscado en el pasado.
- el ejército lo usa para mejorar las imágenes generadas durante el vuelo de los drones,
- los médicos, para mejorar nuestras resonancias magnéticas y estudios PET.
- Se utiliza en Wall Street,
- en traducción automática de lenguas extranjeras,
- genética
- bioinformática.

Ejemplos

2- Se nos dan tres urnas como sigue:
Una urna 1 contiene 3 bolas rojas y 5 blancas.
Una urna 2 contiene 2 bolas rojas y 1 blanca.
Una urna 3 contiene 2 bolas rojas y 3 blancas
Se selecciona una urna al azar y se saca una bola de la urna. Si la bola es roja, ¿cuál es la probabilidad de que proceda de la urna 1?

$$P(A1/B) = \frac{P(A1 \cap B)}{P(B)} = \frac{P(B/A1)P(A1)}{P(B/A1)P(A1)+P(B/A2)P(A2)+P(B/A3)P(A3)}$$

Anotamos quien es cada una y haciendo un poco de álgebra…

$$= \frac{\frac{3}{8}\times\frac{1}{3}}{\frac{3}{8}\times\frac{1}{3}+\frac{2}{3}\times\frac{1}{3}+\frac{2}{5}\times\frac{1}{3}} = \frac{\frac{1}{8}}{\frac{173}{360}} = \frac{45}{173}$$

Teorema de Bayes

Con la Fórmula de Bayes , lo que se hace precisamente es calcular la probabilidad a posteriori dado que se ha presentado una evidencia, P (H/E) , a partir de la probabilidad a priori P(H) y de una probabilidad que normalmente es más fácil conocer,

P(E/H), que es la probabilidad de la evidencia E si la hipótesis H es cierta y se conoce como verosimilitud , ya que representa lo verosímil o creíble que sería la evidencia E que hemos, efectivamente, observado, si la hipótesis H fuese cierta.

$$P(E/H) = \frac{P(H/E).P(E)}{P(H)}$$

Teorema de Bayes

Esta manera de razonar de la inferencia bayesiana, radicalmente diferente a la inferencia clásica o frecuentista (que desdeña en lo formal toda información previa de la realidad que examina), es sin embargo muy cercana al modo de proceder cotidiano, e inductivo. Debe subrayarse que esta metodología, a diferencia del enfoque frecuentista, no tiene como finalidad producir una conclusión dicotómica (significación o no significación, rechazo o aceptación, etc.) sino que cualquier información empírica, combinada con el conocimiento que ya se tenga del problema que se estudia, "actualiza" dicho conocimiento, y la trascendencia de dicha visión actualizada no depende de una regla mecánica.

Tests de embarazo sucesivos

Una mujer sospecha que puede estar embarazada. Para estar segura de su estado compra un test del cual se conoce que tiene una eficacia del 90% en detectar embarazos. La mujer se realiza el test y obtiene un resultado positivo.

Pregunta: Cual es la probabilidad de que dicha mujer esté embarazada?

Adicionalmente supongamos que el test da falsos positivos el 50% de las veces y que, sin ninguna información adicional, la probabilidad de concepci´on luego de mantener una relación sexual es del 15%.

Tests de embarazo sucesivos

Según Bayes

$$P(emb / +) = \frac{P(+ / emb) * P(emb)}{P(+)}$$

__Data__

P(emb)=0.15
P(!emb)=0.85
P(+/emb)=0.9
P (+ /! emb)=0.5

Probabilidad Total = todos los resultados que dieron positivos

P(+)=P (+ / emb)*P(emb) +P (+ /! emb)*P(!emb)

Tests de embarazo sucesivos

Haciendo un poco de álgebra obtenemos el siguiente resultado:

$$P\ (\ emb\ /+)=0.241 = 24.1\%$$

Pero como somos ansiosos queremos repetir la experiencia. Bayes nos permite usar este resultado ya que ahora no tenemos el 15% de posibilidad de embarazo sino que contamos con nueva informacion. Entonces ahora lo podemos usar en una nueva iteración y realizar un nuevo test que también nos da positivo

Tests de embarazo sucesivos

Según Bayes - 2°iteración

$$P(emb / +) = \frac{P(+ / emb) * P(emb)}{P(+)}$$

P(emb / +)=0.364

3°iteración => 0.507
4°iteración => 0.649

Data

P(emb)=0.241
P(!emb)=0.759
P(+/emb)=0.9
P (+ /! emb)=0.5

Independencia

- Dados dos eventos A y B , puede ocurrir que $P(B / A)$ y $P(B)$ sean diferentes, eso significa saber que A ocurrió y modifica la probabilidad de ocurrencia de B

- Entonces, dos eventos A y B son independientes si $P(B / A) = P(B)$, y son dependientes de otro modo

- Notar que por el teorema de la multiplicación $P(A \cap B) = P(B / A)\, P(A)$ si $P(A) > 0$

- Entonces A y B son independientes $\Rightarrow$

$$P(A \cap B) = P(B / A)\, P(A) = P(B)\, P(A)$$

Ejemplo de independencia

1- Las probabilidades de que tres hombres peguen en el blanco son, respectivamente, $\frac{1}{6}$, $\frac{1}{4}$, y $\frac{1}{3}$.

Cada uno dispara una vez al blanco.

a) ¿Cuál es la probabilidad de que exactamente uno de ellos pegue en el blanco?

b) Si solamente uno pega en el blanco, ¿cuál es la probabilidad de que sea el primer hombre?

<u>Solución:</u>

a) consideremos los eventos A_i: "el hombre i-ésimo pega en el blanco" $i = 1, 2, 3$

$$P(A_1) = \frac{1}{6} \qquad P(A_2) = \frac{1}{4} \qquad P(A_3) = \frac{1}{3}$$

Sea el evento B: "exactamente un hombre pega en el blanco"

Entonces $B = \left(A_1^C \cap A_2^C \cap A_3\right) \cup \left(A_1^C \cap A_2 \cap A_3^C\right) \cup \left(A \cap A_2^C \cap A_3^C\right)$

Por lo tanto $P(B) = P\left(A_1^C \cap A_2^C \cap A_3\right) + P\left(A_1^C \cap A_2 \cap A_3^C\right) + P\left(A \cap A_2^C \cap A_3^C\right)$

Y por independencia $P(B) = P\left(A_1^C\right)P(A_2^C)P(A_3) + P\left(A_1^C\right)P(A_2)P)(A_3^C) + P\left(A_1\right)P(A_2^C)P)(A_3^C) =$

Continuación

$$=\left(1-\frac{1}{6}\right)\left(1-\frac{1}{4}\right)\frac{1}{3}+\left(1-\frac{1}{6}\right)\frac{1}{4}\left(1-\frac{1}{3}\right)+\frac{1}{6}\left(1-\frac{1}{4}\right)\left(1-\frac{1}{3}\right)=\frac{1}{12}+\frac{5}{36}+\frac{5}{24}=\frac{31}{72}$$

b) Se pide calcular $P(A_1/B)$

$$P(A_1/B)=\frac{P(A_1\cap B)}{P(B)}=\frac{P(A_1\cap A_2^C\cap A_3^C)}{P(B)}=\frac{\frac{1}{6}\left(1-\frac{1}{4}\right)\left(1-\frac{1}{3}\right)}{\frac{31}{72}}=\frac{6}{31}$$

Conclusiones sobre independencia

Dos (o más) eventos son independientes si la ocurrencia de un evento no cambia la probabilidad de que otro evento ocurra. Existen dos tipos de situaciones cuando esto sucede:

1) Cuando la acción aleatoria no elimina un resultado (procesos estocásticos)

2) cuando la acción aleatoria sí elimina un resultado posible, pero el resultado es reemplazado antes de que la acción vuelva a suceder

Conclusiones sobre independencia

Cuando los eventos son independientes, la probabilidad de que todos ocurran es igual a la multiplicación de las probabilidades de que ocurran los eventos individuales.

Variables Aleatorias Discretas

Una variable aleatoria discreta es una variable aleatoria cuyos valores posibles o constituyen un conjunto finito o bien pueden ser puestos en lista en una secuencia infinita en la cual existe un primer elemento, un segundo elemento, y así sucesivamente (“contablemente” infinita).

Distribuciones de probabilidad para variables aleatorias discretas

Supóngase que p(x) depende de la cantidad que puede ser asignada a cualesquiera de varios valores posibles y cada valor determina una distribución de probabilidad diferente. Tal cantidad se llama parámetro de distribución. El conjunto de todas las distribuciones de probabilidad con diferentes valores del parámetro se llama familia de distribuciones de probabilidad.

Distribución de probabilidad Binomial

Características de un experimento binomial

1. El experimento consta de una secuencia de n experimentos más pequeños llamados ensayos, donde n se fija antes del experimento.
2. Cada ensayo puede dar por resultado uno de los mismos dos resultados posibles (ensayos dicotómicos), los cuales se denotan como éxito (E) y falla (F).
3. Los ensayos son independientes, de modo que el resultado en cualquier ensayo particular no influye en el resultado de cualquier otro ensayo.
4. La probabilidad de éxito es constante de un ensayo a otro; esta probabilidad se denota por p.

Distribución de probabilidad Binomial

En la mayoría de los experimentos binomiales, lo que interesa es el número total de los éxitos (E), en lugar del conocimiento de qué ensayos dieron los éxitos.

La variable aleatoria binomial X asociada con un experimento binomial que consiste en n ensayos se define como

X = el número de los éxitos E entre los n ensayos

Notación : X ~ Bin (n , p)

donde n es el númer de ensayos y p es la probabilidad de éxitos

Distribución de probabilidad Binomial

En la mayoría de los experimentos binomiales, lo que interesa es el número total de los éxitos (E), en lugar del conocimiento de qué ensayos dieron los éxitos.

La variable aleatoria binomial X asociada con un experimento binomial que consiste en n ensayos se define como

X = el número de los éxitos E entre los n ensayos

Notación : $X \sim Bin(n, p)$

donde n es el númer de ensayos y p es la probabilidad de éxitos

Distribución de probabilidad Binomial

Resultados y probabilidades de un experimento binomial con cuatro ensayos

Resultado	x	Probabilidad	Resultado	x	Probabilidad
$EEEE$	4	p^4	$FEEE$	3	$p^3(1-p)$
$EEEF$	3	$p^3(1-p)$	$FEEF$	2	$p^2(1-p)^2$
$EEFE$	3	$p^3(1-p)$	$FEFE$	2	$p^2(1-p)^2$
$EEFF$	2	$p^2(1-p)^2$	$FEFF$	1	$p(1-p)^3$
$EFEE$	3	$p^3(1-p)$	$FFEE$	2	$p^2(1-p)^2$
$EFEF$	2	$p^2(1-p)^2$	$FFEF$	1	$p(1-p)^3$
$EFFE$	2	$p^2(1-p)^2$	$FFFE$	1	$p(1-p)^3$
$EFFF$	1	$p(1-p)^3$	$FFFF$	0	$(1-p)^4$

En este caso especial, se desea b(x; 4, p) con x 0, 1, 2, 3 y 4.

Distribución de probabilidad Binomial

$$b(x; n, p) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0, 1, 2, \ldots n \\ 0 & \text{de lo contrario} \end{cases}$$

[distro binomial jupyter]

Distribución de probabilidad Binomial

La distribución acumulada vendrá dada por la siguiente ecuación:

Para $X \sim \text{Bin}(n, p)$, la función de distribución acumulativa será denotada por

$$P(X \leq x) = B(x; n, p) = \sum_{y=0}^{x} b(y; n, p) \qquad x = 0, 1, \ldots, n$$

Distribución de probabilidad Hipergeométrica y binomiales negativas

Las distribuciones hipergeométricas y binomiales negativas están relacionadas con la distribución binomial.

En tanto que la distribución binomial es el modelo de probabilidad aproximada de muestreo sin reemplazo de una población dicotómica finita (E–F), la distribución hipergeométrica es el modelo de probabilidad exacta del número de éxitos (E) en la muestra.

Distribución hipergeométrica

1. La población o conjunto que se va a muestrear se compone de N individuos, objetos o elementos (una población finita).

2. Cada individuo puede ser caracterizado como éxito (E) o falla (F) y hay M éxitos en la población.

3. Se selecciona una muestra de n individuos sin reemplazo de tal modo que cada subconjunto de tamaño n es igualmente probable de ser seleccionado.

Distribución hipergeométrica

Tipos de problemas al que responde

Se capturaron, etiquetaron y liberaron cinco individuos de una población de animales que se piensa están al borde de la extinción en una región par que se mezclen con la población. Después de haber tenido la oportunidad de mezclarse, se selecciona una muestra aleatoria de 10 de estos animales. Sea X el número de animales etiquetados en la segunda muestra. Si en realidad hay 25 animales de este tipo en la región, ¿cuál es la probabilidad de que

a) X = 2?
b) ¿X <= 2?

Distribución hipergeométrica

Si X es el número de éxitos (E) en una muestra completamente aleatoria de tamaño n extraída de la población compuesta de M éxitos y (N - M) fallas, entonces la distribución de probabilidad de X llamada distribución hipergeométrica, es

$$P(X = x) = h(x; n, M, N) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

con x un entero que satisface
máx(0, n - N + M) <= x <= mín(n, M).

Otras distros discretas

- Geométrica:
 “número de repeticiones de ε hasta que ocurre A por primera vez inclusive”

- Binomial Negativa:
 “número de repeticiones de ε hasta que ocurre A por r-ésima vez, incluyendo la r-ésima
vez que ocurre A” (es una extensión de la geométrica)

Otras distros discretas

- Poisson

La distribución de Poisson se aplica a varios fenómenos discretos de la naturaleza (esto es, aquellos fenómenos que ocurren 0, 1, 2, 3,... veces durante un periodo definido de tiempo o en un área determinada) cuando la probabilidad de ocurrencia del fenómeno es constante en el tiempo o el espacio.

Se puede aplicar en las siguientes situaciones

Otras distros discretas

- El número de autos que pasan a través de un cierto punto en una ruta (suficientemente distantes de los semáforos) durante un periodo definido de tiempo.
- El número de errores de ortografía que uno comete al escribir una única página.
- El número de llamadas telefónicas en una central telefónica por minuto.
- El número de servidores web accedidos por minuto.
- El número de animales muertos encontrados por unidad de longitud de ruta.
- El número de mutaciones de determinada cadena de ADN después de cierta cantidad de radiación.

Variables Aleatorias Continuas

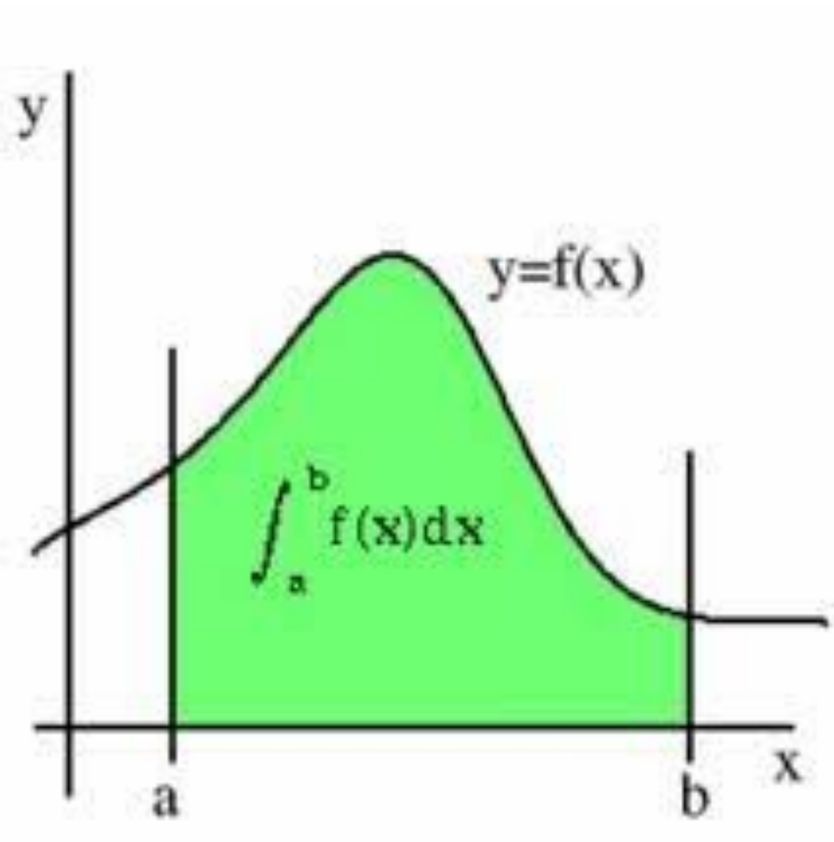
Una variable aleatoria continua es una variable aleatoria con un conjunto de valores posibles (conocido como el rango) que es infinito y no se puede contar.

Sea X una v.a.. Decimos que es continua si existe una función no negativa f , definida sobre todos los reales $x \in (-\infty, \infty)$, tal que para cualquier conjunto B de números reales

$$\mathbf{P(X \in B) = \int f(x)\,dx}$$

O sea que la probabilidad de que X tome valores en B se obtiene al integrar la función f sobre el conjunto B. A la función f la llamamos función densidad de probabilidad (f.d.p.).

Función de Distribución de Probabilidad



¿Por qué pierde significado hablar de p(x=c) si a<c<b?

Observemos que el área bajo la curva corresponde a un punto en particular del eje x

Frecuencia de Distribución Acumulada

Veamos con un ejemplo su uso,

Supongamos que X es una v.a. continua con f.d.p. dada por

$$f(x) = \begin{cases} C\,(4x - 2x^2) & \text{si } 0 \leq x \leq 2 \\ 0 & \text{caso contrario} \end{cases}$$

a) ¿Cuál es el valor de C?
b) Hallar P (X > 1)
c) Hallar la F.d.a. de X

Frecuencia de Distribución Acumulada

se debe cumplir que **$1 = \int_{-\infty}^{\infty} f(x)\,dx$**

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

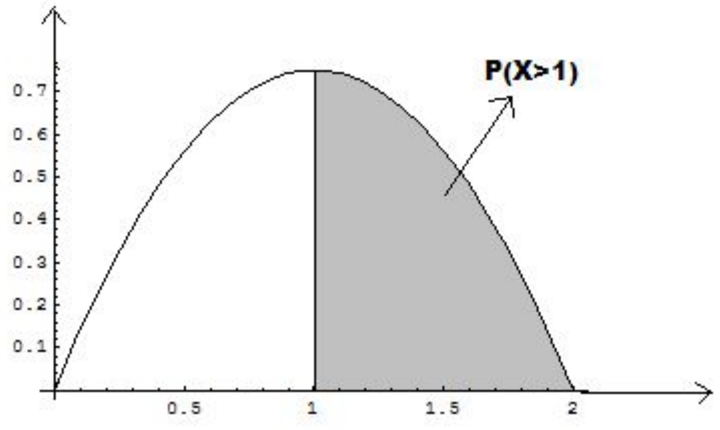
$\int_{-\infty}^{0} 0\,dx + \int_{0}^{2} C\,(4x - 2x^2)\,dx + \int_{2}^{\infty} 0\,dx =$

$C \cdot 8/3 = 1$ $\Rightarrow$ $C = 3/8$

Seguimos con el ejemplo

Para calcular la probabilidad que X sea mayor que 1, planteamos

$$P(X>1) = \int_1^{\infty} 3/8(4\,x - 2\,x^2)\,dx = 1/2$$



Variables aleatorias continuas importantes

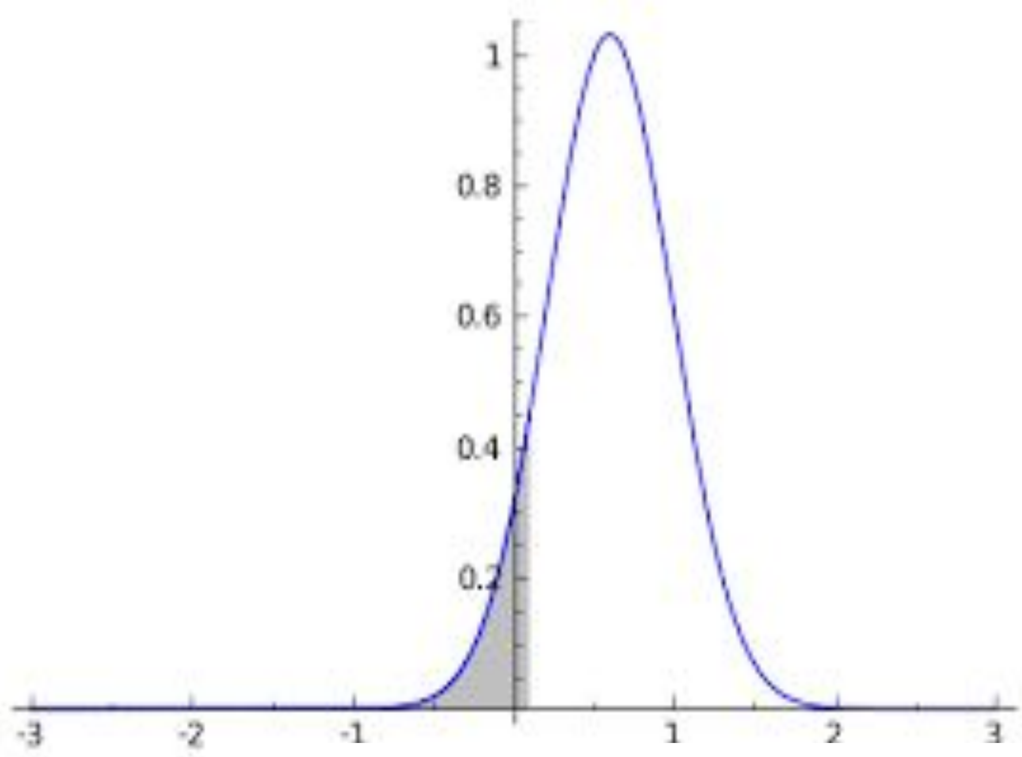
Distribución normal o gaussiana

Se llama distribución normal, distribución de Gauss, distribución gaussiana o distribución de Laplace-Gauss, a una de las distribuciones de probabilidad de variable continua que con más frecuencia estadística aparece.

Fenómenos muy normales

- Caracteres morfológicos de individuos de una especie. Por ejemplo: tallas, pesos, envergaduras, diámetros, perímetros
- Caracteres fisiológicos, por ejemplo: efecto de una misma dosis de un fármaco, o de una misma cantidad de abono.
- Caracteres sociológicos, por ejemplo: cociente intelectual, grado de adaptación a un medio.
- Errores cometidos al medir ciertas magnitudes.
- Valores estadísticos muestrales, por ejemplo: la media.
- Otras distribuciones como la binomial o la Poisson son aproximaciones normales.

Gráfica de una distribución normal

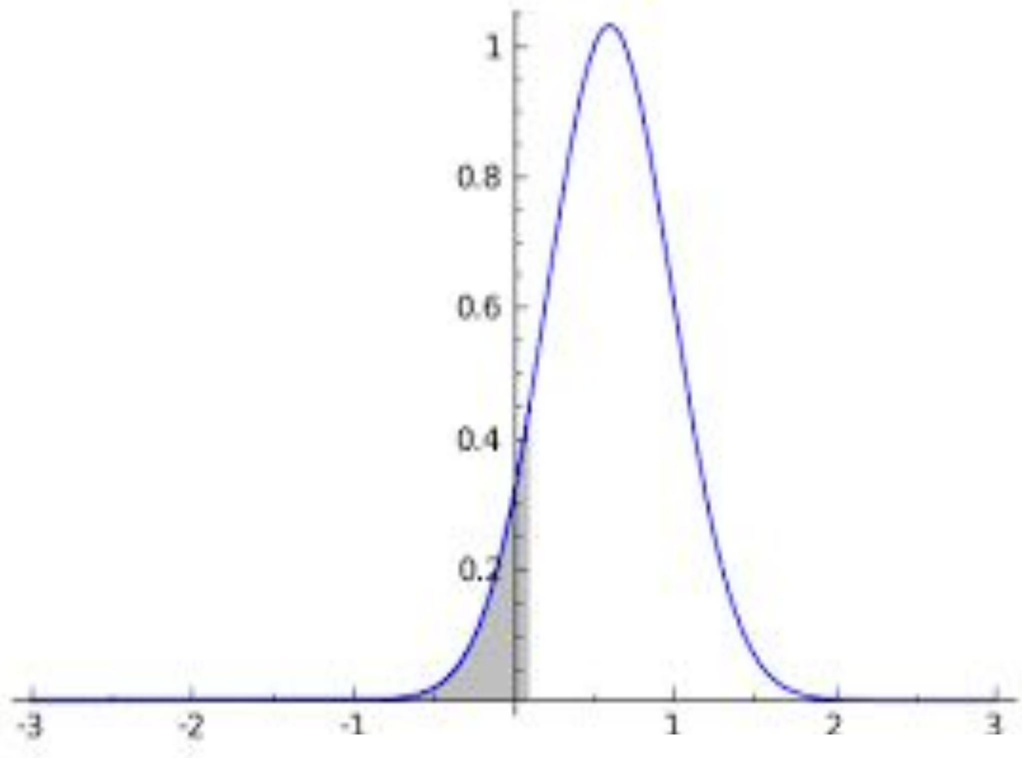


fdp= $\int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt$

Características

- Es una distribución que tiene forma de campana, es simétrica y puede tomar valores entre menos infinito y más infinito
- Media, mediana y moda son iguales
- Es simétrica

Gráfica de una distribución normal



fdp= $\int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt$

- Presenta como asíntota el eje de abscisas al que se va aproximando sin llegar a cortarse
- Presenta puntos de inflexión en los puntos de abscisas μ +σ y μ-σ, donde cambia de concavidad (lo que determina que cuánto mayor sea σ más achatada sea la curva)

Los parámetros de la normal

Cuando μ varía la gráfica de la función se traslada, es un parámetro de posición.
Cuando σ aumenta, la gráfica se "achata", cuando σ disminuye la gráfica se hace más "puntiaguda",
se dice que es un parámetro de escala.

¿Quienes son μ y σ?

- μ es la media
- σ es la desviación típica

Normal estándar

En las familias representadas por las distribuciones normales ocupa un lugar especial la distribución que tiene de media cero ($\mu= 0$) y por desviación típica la unidad ($\sigma= 1$). Esta distribución se llama la distribución normal estándar, o reducida.

Recordemos las frecuencias acumuladas de una variable aleatoria continua, como calculamos la probabilidad

$P(X<a)$
$P(a<X<b)$
$P(X<b)$

la fda de una normal se anota como **$\phi(x) = P(X <= x)$**

Ej $P(X<2)=\phi(2)$

y por simetría
$P(X>2)=1-P(X<2)=1- \phi(2)$

Normal estándar y cálculo fda

¿Cómo calculamos la fda?

Calculamos el área bajo la curva de su fdp. Aquí su fdp empieza a complicarse un poco y la integral de esa ecuacion sólo se resuelve numéricamente!

Esto ya está hecho, existen tablas de la función de distribución acumulada de la normal estándar para valores de x que oscilan en general entre -4 y 4, pues para valores de x menores que -4, $Pb(x)$ 0 , y para valores de x mayores que 4, (x) 1

Tabla de Fda de la normal estandarizada

Normal Deviate z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-4.0	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
-3.9	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
-3.8	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
-3.7	.0001	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
-3.6	.0002	.0002	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
-3.5	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002
-3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
-3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
-3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
-3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
-3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
-2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
-2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
-2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
-2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
-2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
-2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
-2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
-2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
-2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
-2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
-1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
-1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
-1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
-1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
-1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
-1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
-1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
-1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
-1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
-1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
-.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
-.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
-.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
-.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
-.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
-.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
-.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483

Estimar los parámetros de una normal

```
import numpy
from scipy import stats

# Generamos una muestra de 500 elementos de una
# distribucion normal con media 5.0 y ds de 3.5
data = numpy.random.normal(5.0, 3.5, size=500)
# Estimamos los parámetros
loc, scale = stats.norm.fit(data)
print('media', loc, ' - ds', scale)
# Generamos una nueva muestra a partir de nuestra distribución
n = stats.norm(loc=loc, scale=scale)
n.rvs(10)
```

```
media 5.00895613598222  - ds 3.385781993768077

array([ 4.21941632,  4.12794948,  4.89123763,  8.89941538, -0.51757976,
        3.00981439,  7.22050595,  2.62793101,  5.64303333,  4.2471182
6])
```

Distribución exponencial

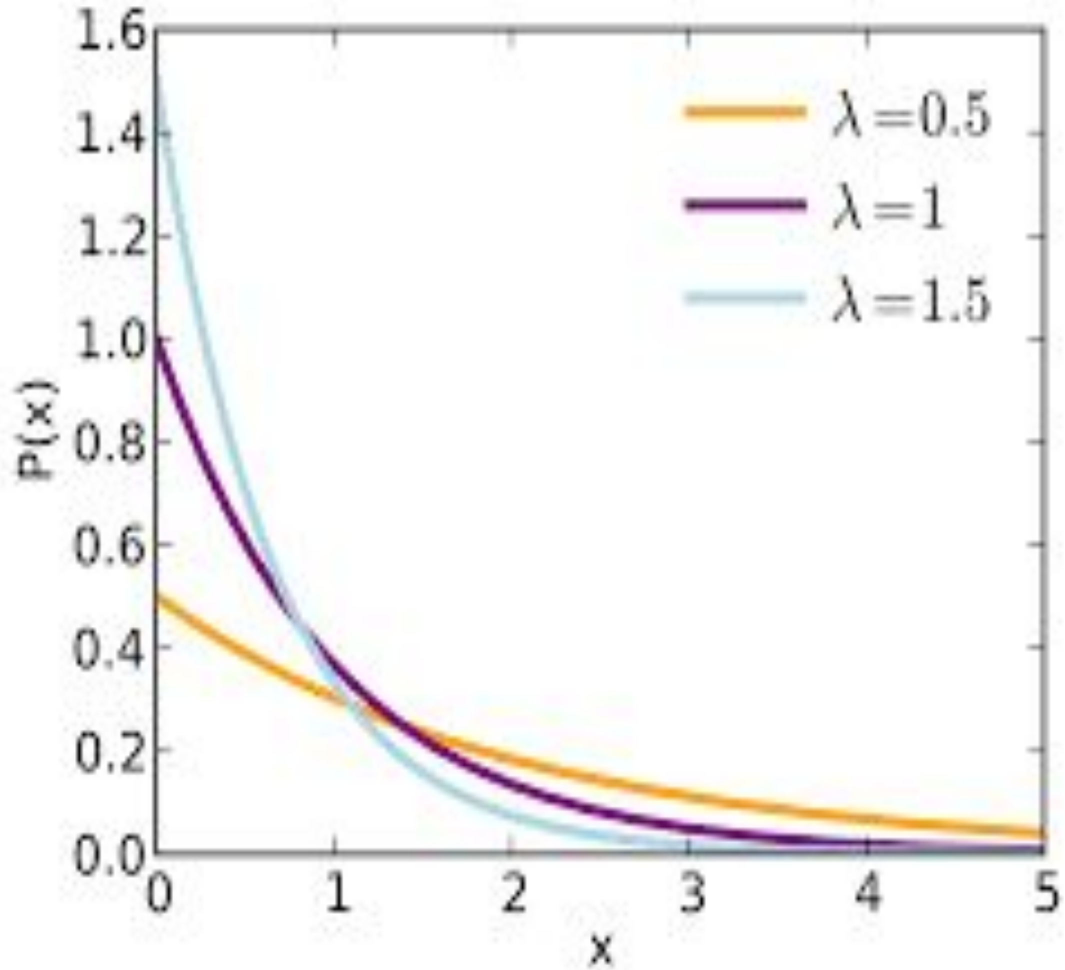
Sea X una v.a. continua. Se dice que tiene distribución exponencial con parámetro λ si su f.d.p. es de la forma

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & si \quad x \geq 0 \\ 0 & si \quad x < 0 \end{cases}$$

Características

La distribución exponencial se utiliza algunas veces para modelar el tiempo que transcurre antes de que ocurra un evento. A menudo se lo llama tiempo de espera.

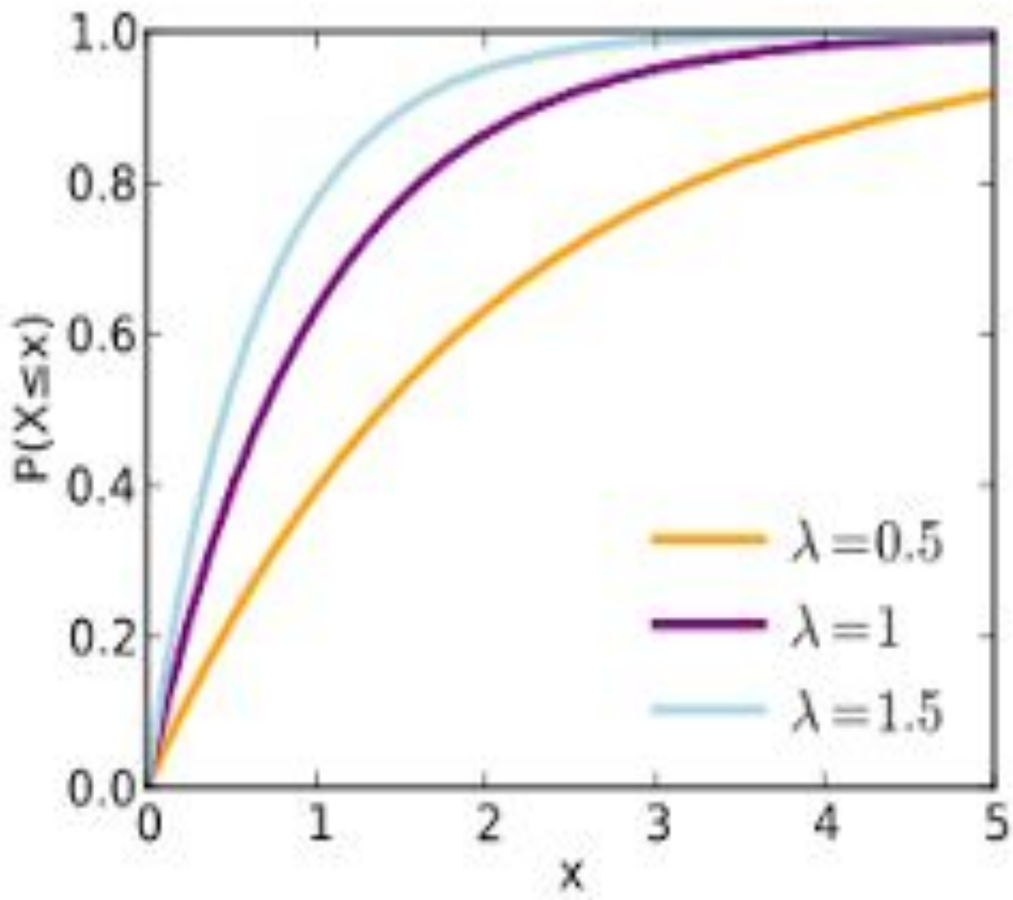
Notación: X ~ Exp (λ)



Fda de la exponencial

Recordemos que la acumulada es el área bajo la curva de la fdp, analíticamente el resultado de la integral.

$$f(x) = \begin{cases} 1 - e^{-\lambda x} & si \quad x \geq 0 \\ 0 & si \quad x < 0 \end{cases}$$



Ejemplo

Supongamos que el tiempo, en segundos, de respuesta en cierta terminal de computadora en línea (es decir el tiempo transcurrido entre el fin de la consulta del usuario y el principio de la respuesta del sistema a esa consulta) se puede considerar como una variable aleatoria con distribución exponencial con parámetro $\lambda = 0,2$. Calcular

a) la probabilidad de que el tiempo de respuesta sea a lo sumo 10 segundos.

b) la probabilidad de que el tiempo de respuesta esté entre 5 y 10 segundos.

Solución

a) Sea X la v.a. entonces X ~ Exp (0.2)

$$P(X \leq 10) =$$

$$F(10) =$$

$$1 \text{ - } e^{-(10*0.2)} =$$

$$1\text{-}0.135 =$$

$$0.865$$

Solución

b)

$$P(5 \leq X \leq 10) =$$

$$F(10) - F(5) =$$

$$1 - e^{-(10*0.2)} - (1 - e^{-(5*0.2)}) =$$

$$0.233$$

La falta de memoria de la exponencial

Propiedad falta de memoria La distribución exponencial tiene una propiedad conocida como falta de memoria, que se muestra en el siguiente ejemplo:

El tiempo de vida, en años, de un circuito integrado particular tiene una distribución exponencial con parámetro 0.5. Encuentre la probabilidad de que el circuito dure más de tres años

Ejemplo de la propiedad falta de memoria

Supongamos ahora que actualmente un circuito tiene cuatro años y aún funciona. Se quiere hallar la probabilidad de que funcione tres años más.
Por lo tanto planteamos una probabilidad condicional

$$P(X \geq 4 \mid X \geq 7) =$$

$$\frac{P(X \geq 4 \cap X \geq 7)}{P(X \geq 4)} =$$

idem para P(X>=4)

$$\text{como } \mathrm{P}(\mathrm{X} \geq 4) \subset P(X \geq 7) => \frac{P(X \geq 7)}{P(X \geq 4)} =$$

$$\text{como } \mathrm{P}(\mathrm{X} \geq 7) = 1 - (X \leq 7) = 1 - (1 - e^{-(7*0.5)})$$

$$\mathrm{e}^{-0.5*(7-4)} = 0.223$$

Ejemplo de la propiedad falta de memoria

$$X \sim Exp(0.5)$$

$$P(X \geq 3) =$$

$$1\text{-}F(3) =$$

$$1 - e^{-(3*0.5)} =$$

$$0.223$$

Observaciones

En general, la probabilidad que se tenga que esperar t unidades adicionales, dado que ya se han esperado s unidades, es la misma que la probabilidad de que se tenga que esperar t unidades desde el inicio.

La distribución exponencial no “recuerda” cuánto tiempo se ha esperado. En particular, si el tiempo de vida de un componente sigue una distribución exponencial, entonces la probabilidad de que un componente que tiene s unidades de tiempo dure t unidades de tiempo adicionales es la misma que la probabilidad de que un componente nuevo dure t unidades de tiempo.

Conclusiones

- Probabilidad Clásica: Si un suceso puede ocurrir de N maneras mutuamente excluyentes e igualmente probables, y m de ellas poseen una característica A "**estimando**" el valor de una probabilidad desconocida por medio de un estudio de la conducta de las frecuencias relativas del hecho o suceso correspondiente.
- Se refiere a la probabilidad de ocurrencia de un suceso basado en la experiencia previa, la opinión personal o la intuición del individuo. En este caso después de estudiar la información disponible, se asigna un valor de probabilidad a los sucesos basado en el grado de creencia de que el suceso pueda ocurrir

Conclusiones

- es muy imporante reconocer cual es el suceso que queremos observar y si tiene dependencia de otros que influyan en su resultado
- las distribuciones de la variables aleatorias nos permiten predecir ciertos comportamientos de la situación que estemos analizando
- los gráficos de estas distribuciones nos ayudan a comunicar de manera más clara el resultado de nuestro análisis

¿Dudas?