Modelos probabilísticos y análisis estadístco

Introducción a probabilidad

Rodrigo Gil Castañeda

rodrigo.gil@utadeo.edu.co

Departamento de Ciencias Básicas
Facultad de Ciencias Naturales e Ingeniería
Universidad de Bogotá Jorge Tadeo Lozano

Tabla de contenidos

Introducción

Modelo

En investigación es común que se requiera describir diferentes tipos de fenómenos en términos matemáticos. En este contexto, un modelo corresponde a una idealización (simplificación) matemática usada para aproximarse a un fenómeno observable. El modelo será satifactorio siempre y cuando logre representar de la manera más fiel posible la realidad. Con el fin de determinar la validez del modelo *se realizan experimentos que permitan el registro de observaciones*.

Modelo determinístico

En un modelo deterministico, el resultado está dado por las condiciones que establecen al momento de realizar el experimento.

* La distancia d (km) recorrida por un automóvil en t (h) a una velocidad constante v $(km \cdot h^{-1})$ está determinada por la relación $d = v \times t \Longrightarrow v$ y t determinan el valor de d.

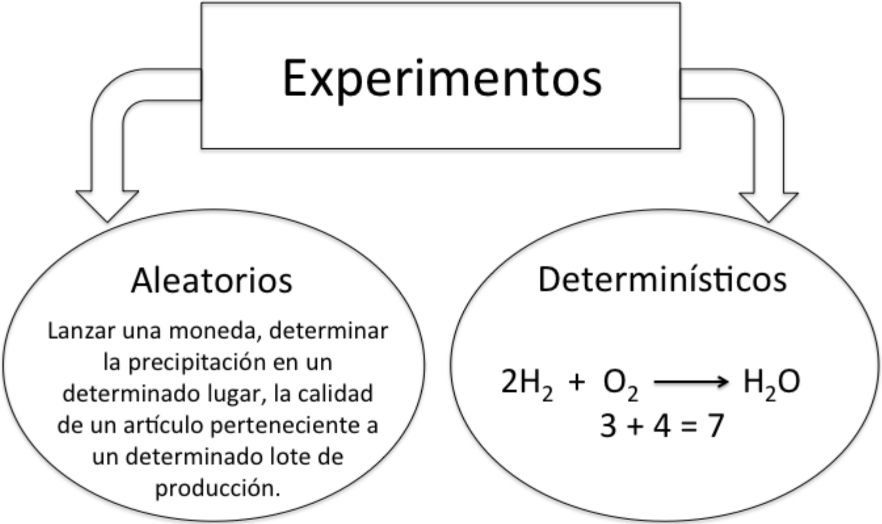
Introducción

Modelo estocástico

En un modelo estocástico, es aquel en el cual la información disponible no permite generar una relación o regla para determinar el resultado de un experimento. Muchos fenómenos (naturales o artificiales) son considerados **aleatorios** debido a que el resultado exacto no puede ser predicho; en este caso los resultados de mediciones repetidas sobre circunstancias similares generan un patrón.

* En un experimento aleatorio E, el conjunto de todos los posibles resultados de E se llama espacio muestral y se denota por la letra griega Ω. Ejemplo: en el experimento de lanzar una moneda al aire, Ω estaría compuesto por los posibles resultados: **cara** o **sello**; $\Omega = \{Ca, Se\}$. En la ejecución del experimento aletorio, aunque el resultado es impredecible, se sabe que debe corresponder a una de las opciones fijadas por el espacio muestral.

Introducción



Introducción: Espacio muestral

En la mayor parte de este módulo se emplearán los paquetes (librerias) `TeachingDemos` y `animation`.

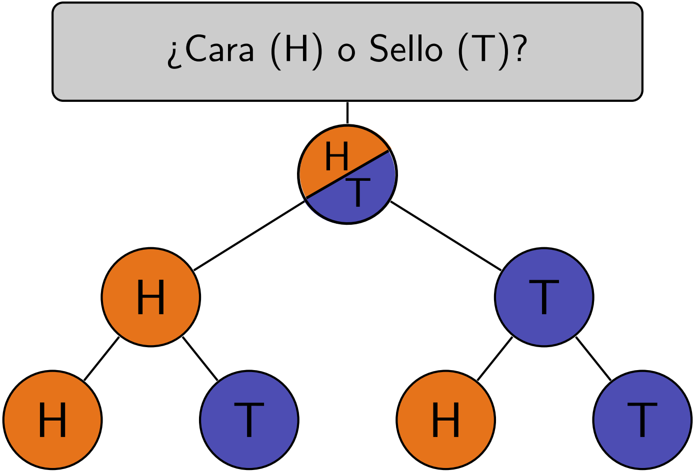
Ejemplo

Considere un experimento aleatorio consistente en lanzar una moneda al aire. Los posibles resultados son H (cara) y T (sello). Usando `R` se puede simular fácilmente este tipo de experimentos con el fin de ver su naturaleza aleatoria.

En R

- `library(animation)`
- `ani.options(nmax = 6, interval = 0.3)`
- `flip.coin(faces = c("Tail", "Head"), type = "n",`
- `prob = c(0.5, 0.5),col = c(2, 4))`

Introducción: Diagrama de árbol



$$\Omega = \{HH, HT, TH, TT\}$$

Introducción: Lanzamiento de dados

Otro ejemplo clásico para introduccir el concepto de experimento aleatorio es lanzar un dado. En `R` se emplea la función `dice`, que por defecto emplea 2 dados de 6 caras, para simular los lanzamientos.

En `R`

- Dado de 6 caras: `dice(2, 1, sides=6)`
- Dado de 6 caras: `dice(10, 2, sides=6, plot.it=T)`
- Dado de 10 caras: `dice(2, 1, sides=10)`
- Dado de 12 caras: `dice(2, 1, sides=12)`

Eventos

Un evento **A** es una colección de resultados, lo que equivale a decir que es un subconjunto del espacio muestral. Se dice que un conjunto de eventos $\{A_1, A_2, A_3, \dots\}$ son mutuamente excluyentes o disyuntos si $A_i \cap A_j = \varnothing$, para cualquier par distinto A_i, A_j.

Por ejemplo, para el experimento de lanzar la moneda al aire los eventos $A = Cara$ y $B = Sello$ serían eventos mutuamente excluyentes.

Axiomas de Probabilidad

Cuadro: Una función P con dominio en A y rango el intervalo $[0,1]$ se llamará medida de probabilidad o simplemente probabilidad si satisface los siguientes axiomas(axiomas de Kolmogorov):

Axioma I	Para cualquier evento $A_i \in A$, una probabilidad no puede ser negativa, $0 \leq P(A) \leq 1$
Axioma II	La probabilidad asociada al espacio muestral es $P(\Omega) = 1$. Ω es un espacio muestral asociado a un experimento estadístico
Axioma III	Si $\{A_i\}$ es una secuencia de eventos mutuamente excluyentes $A_i \cap A_j = \varnothing$. Con $A_i \in A, i = 1, 2, \ldots$. $P(A_1 \cup A_2 \cup \ldots) = \sum P(A_i)$. Según este axioma se puede calcular la probabilidad de conjunto de alternativas mutuamente excluyentes sumando las probabilidades

Intuición detrás de los axiomas:

Primero, la probabilidad de un evento nunca puede ser negativa. Y puesto que el espacio de muestra contiene todos los posibles resultados, su probabilidad debe ser uno (1). El axioma final puede parecer intimidante, pero simplemente significa que para una secuencia de eventos disjuntos, su probabilidad total (medida) debe ser igual a la suma de sus partes. Por ejemplo, la posibilidad de sacar un 1 o un 2 en un dado es la probabilidad de sacar un 1 más la probabilidad de sacar un 2.

Propiedades

Para cualquier eventos A y B,

1. $P(A^c) = 1 - P(A)$

   Prueba. Puesto que $A \cup A^c = S$ y $A \cap A^c = \varnothing$, se tiene

   $1 = P(\Omega) = P(A \cup A^c) = P(A) + P(A^c)$
2. $P(\varnothing) = 0$

   Prueba. Note que $\varnothing = \Omega^c$, ahora se aplica la primera propiedad.

Propiedades

Propiedades

Para cualquier eventos A y B,

3. Si $A \subset B$, entonces $P(A) \leq P(B)$

   Prueba. $B = A \cup (B \cap A^c)$, y se nota que $A \cap (B \cap A^c) = \varnothing$, entonces:

   $P(B) = P(A \cup (B \cap A^c)) = P(A) + P(B \cap A^c) \geq P(A)$

   ya que $P(B \cap A^c) \geq 0$

4. $0 \leq P(A) \leq 1$

   Prueba. La desigualdad de la izquierda se deriva del primer axioma. La desigualdad de la derecha se ajusta a la propiedad $A \subset \Omega$

5. Regla general de la adición.

   $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

. Teorema de la probabilidad total. Sean $B_1, B_2, \ldots, B_n$ eventos exhaustivos y mutuamente excluyentes, entonces.

   $P(A) = P(A \cap B_1) + P(A \cap B_2) + \cdots + P(A \cap B_n)$

Distribuciones de probabilidad

Una de las ventajas de `R` es la capacidad de simular muestras que se ajustan a varias distribuciones de probabilidad .

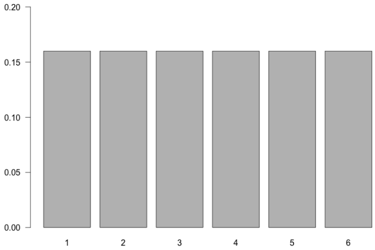
Un estudio de simulación generalmente empieza con un modelo de probabilidad para los datos y la simulación de respuestas a partir de este modelo. `R` cuenta con un conjunto de funciones para varias distribuciones de probabilidad, a las que se les suele llamar: de la familia d-p-q-r. Esto se debe a que se emplean para evaluar la función de densidad de probabilidad (para distribuciones continuas y la función de masa de probabilidad para distribuciones discretas), la función de distribución acumulada o la función cuantil (inversa de la CDF) y para la simulación de una muestra aleatoria.

Distribución uniforme

Todos y cada uno de los resultados del experimento tienen la misma probabilidad de ocurrir .

Ejemplo

En el lanzamiento de un dado de 6 caras, es espacio muestral es $\Omega = \{1, 2, 3, 4, 5, 6\}$ y la ocurrencia de cada uno de estos resultados tiene una probabilidad de 0.1666667.



x	$P(x)$
1	0.16
2	0.16
3	0.16
4	0.16
5	0.16
6	0.16

Distribución binomial

Es una de las más empleadas debido a sus multiples aplicaciones. En el caso de las distribuciones binomiales, los experimentos tienen las siguientes características: Se realizan n intentos (ensayos) independientes y en cada uno tienen dos posibles resultados: éxito (p: probalilidad de éxito) o fracaso (q: probabilidad de fracaso). Éxito y fracaso son mútuamente excluyentes, por lo tanto: $p + q = 1$. x : número de éxitos que se desean tener.
Si una variable aleatoria X sigue una distribución binomial con parámetros n y p, escribimos: $X \sim B(x; n, p)$. La probabilidad de obtener exactamente k éxitos en n pruebas está dada por la función de masa de probabilidad:

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x} \qquad \text{(Función de probabilidad: Binomial)}$$

Con media $= n \times p$ y varianza $= n \times p \times (1 - p)$

Distribución de Probabilidad Binomial

Propiedades:

1. El experimento consiste de una secuencia de n ensayos idénticos.
2. Cada ensayo tiene únicamente dos posibles resultados (éxito o fracaso). $\Omega = \{Exito, Fracaso\}$.
3. Ni la probabilidad (p) de un éxito ni la de un fracaso $(1 - p)$ cambia entre ensayos.
4. Los ensayos del experimento son independientes.

Ejemplo

Considere la decisión de compra de los tres siguientes clientes que entran a una tienda de ropa. A partir de experiencias pasadas, el gerente de la tienda estima la probabilidad de compra de un cliente en 0.30. Cuál es la probabilidad de que dos de los tres siguientes clientes que entren a la tienda hagan una compra?

$$P(X = x) = \binom{n}{x} \times p^x \times (1 - p)^{n-x}$$

$$P(X = x) = \binom{3}{2} \times 0.3^2 \times (1 - 0.3)^{3-2}$$

$$P(X = x) = 3 \times 0.09 \times 0.70 = 0.189$$

La probabilidad de que dos de los siguientes tres clientes que entren a la tienda hagan una compra es de 0.189

En R sería así $\Longrightarrow$ `dbinom(x=2,size=3, prob=0.3)`

Distribución binomial en R

Se creará un vector que contenga 100 números aleatorios con una districución binomial con parámetros $n = 7$ y $p = 0.3$ en un objeto vectorial llamado X. Donde n representa el número de ensayos, llamado size en la función de R.

```
X <- rbinom(n=100, size=7, prob=0.3)
barplot(table(X))
```

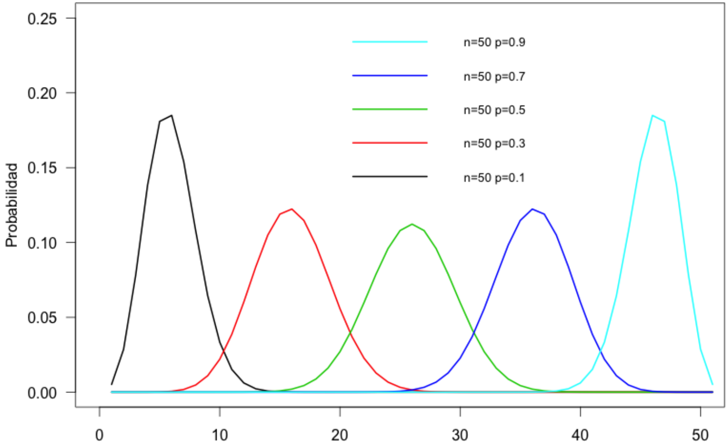
El diagrama de barras representa una estimación de la distribución de probabilidad $P(X = x)$. El diagrama de barras es más apropiado que un histograma, ya que los datos son discretos, no continuos.

La media, la desviación estándar y la varianza se pueden calcular directamente a partir de X.

```
mean(X)
var(X)
sd(X)
```

Distribución binomial

Figura: En esta figura se muestran seis variables aleatorias Binomiales con $n = 50$ y $p = 0.1$, $p = 0.3$, $p = 0.5$, $p = 0.7$, y $p = 0.9$.



Distribución de Probabilidad Poisson

Distribución de variables aleatorias discretas útil para estimar la ocurrencia de eventos dentro de un intervalo específico de tiempo o espacio.
$X \sim pois(x; \lambda)$

$$f(x) = \frac{\lambda^x \times e^{-\lambda}}{x!}$$

- $f(x)$: Probabilidad de x ocurrencia en un intervalo
- λ: Valor esperado o número promedio de ocurrencias dentro de un intervalo
- $e = 2.71828$

Distribución de probabilidad Poisson

Propiedades de la distribución de probabilidad Poisson

1. La probabilidad de una ocurrencia es la misma para cualquier par de intervalos de igual longitud
2. La ocurrencia o no ocurrencia en cualquier intervalo es independiente de la ocurrencia o no ocurrencia en cualquier otro intervalo.

Número de clientes atendidos en una caja de un banco durante un periodo de 15 minutos en las mañanas de días entre semana

Asumiendo que:

1 La probabilidad de que una persona llegue en cualquier intervalo de tiempo es la misma para dos intervalos de tiempo iguales y que

2 la llegada o no llegada del cliente en cualquier periodo de tiempo es independiente de la llegada o no en cualquier otro periodo de tiempo.

Función de probabilidad Poisson

Ejemplo

Número de clientes atendidos en la ventanilla de un banco durante un periodo de 15 minutos, en las mañanas de días entre semana.

Suponiendo que los datos históricos indican que el promedio de clientes que llegan a la vetanilla del banco en un periodo de 15 minutos es 10; y que el gerente quiere conocer la probabilidad de que exactamente lleguen 5 clientes en un periodo de 15 minutos, entonces:

$$f(x=5) = \frac{10^5 \times e^{-10}}{5!} = 0.0378$$

En R sería así $\Longrightarrow$ `dpois(5, lambda=10)`

Valor esperado y Varianza

En la distribución de probabilidad Poisson la media (y la varianza son IGUALES

$$Media(X) = \lambda = Var(X)$$

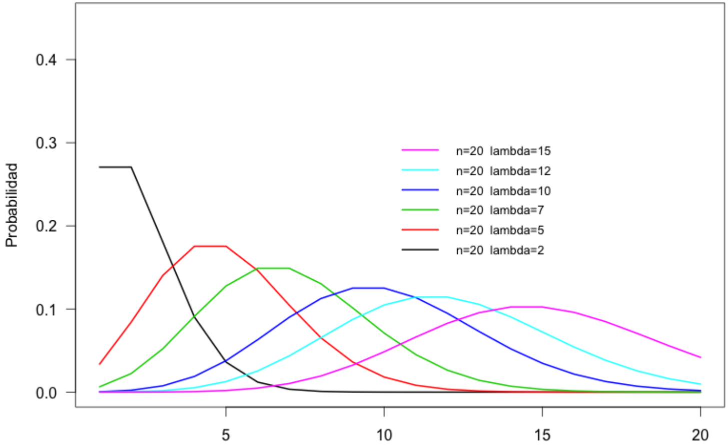
Siguiendo con el ejemplo anterior:

$$Media(X) = \lambda = 10$$

$$\sigma = 3.16$$

Distribución Poison

Figura: En la figurase muestran varias variables aleatorias Poisson con diferentes valores de λ.



Distribución de Probabilidad Hipergeométrica

Distribución utilizada para calcular la probabilidad de que en una selección aleatoria de n elementos, seleccionados sin reemplazamiento, se obtengan x elementos calificados como éxitos y $n - x$ elementos calificados como fracasos

$$f(x) = \frac{\begin{pmatrix} r \\ x \end{pmatrix} \begin{pmatrix} N - r \\ n - x \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}}$$

$f(x)$: Probabilidad de x éxitos en n ensayos

x: Número de éxitos

n: Número de ensayos

N: Número de elementos en la población

r: Número de elementos en la población calificados como éxitos

Distribución de Probabilidad Hipergeométrica

$\left(\begin{array}{c} r \\ x \end{array} \right)$: Número de maneras en que x éxitos pueden ser seleccionados de un total de r éxitos en la población

$\left(\begin{array}{c} N-r \\ n-x \end{array} \right)$: Número de maneras en que $n-x$ fracasos pueden ser seleccionados de un total de $N-r$ fracasos en la población

$\left(\begin{array}{c} N \\ n \end{array} \right)$: Número de maneras en que n elementos pueden ser seleccionados de una población N

Características

La distribución de probabilidad hipergeométrica es cercana a la distribución de probabilidad binomial

- Binomial: Ensayos independientes
- Hipergeométrica: Ensayos NO son independientes

- Binomial: La probabilidad de éxito NO cambia entre ensayos
- Hipergeométrica: La probabilidad de éxito SI cambia entre ensayos

Condiciones

La distribución es válida solo cuando:

1. El número de éxitos observados es menor o igual que el número de éxitos presentes en la población

$$x \leq r$$

2. El número de fracasos es menor o igual que el número de fracasos presentes en la población

$$n - x \leq N - r$$

Ejemplo

Una compañía fabrica bombillos ahorradores de energía y los empaca en cajas de 12 unidades cada una. Un inspector elige al azar una caja para probarlos . Si la caja contiene exactamente 5 bombillos defectuosos, ¿cuál es la probabilidad de que el inspector encuentre exactamente un bombillo defectuoso al inspeccionar los tres primeros bombillos?

$$N = 12 \qquad n = 3 \qquad r = 5 \qquad x = 1$$

$$f(x=5) = \frac{\binom{5}{1}\binom{12-5}{3-1}}{\binom{12}{3}} = \frac{5 \times 21}{220} = 0.4773$$

La probabilidad de encontrar exactamente un bombillo defectuoso al inspeccionar los primeros tres bombillos será de 0.4773.

En R sería así $\Longrightarrow$ `dhyper(x=1, m=5, n=7, k=3)`

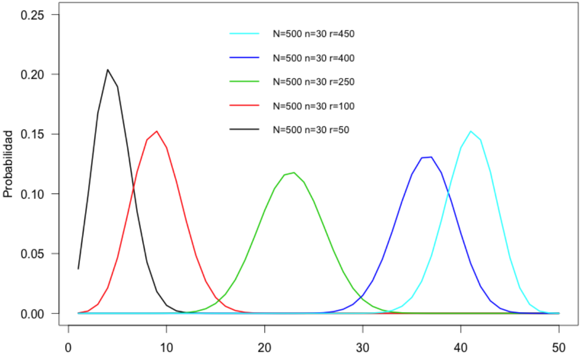
Valor Esperado y Varianza

$$Media = \mu = n\left(\frac{r}{N}\right)$$

$$Var(x) = \sigma^2 = n\left(\frac{r}{N}\right)\left(1 - \frac{r}{N}\right)\left(\frac{N-n}{N-1}\right)$$

Distribución Hipergeométrica

Figura: La distribución hipergeométrica describe el número de eventos k de una muestra n extraída de una población total de N sin reemplazo. En una población de N objetos de los cuales r son **defectuosos** y Nr son **no-defectuosos** (**éxito** o **fracaso**). Si tomamos muestras n elementos sin reemplazo entonces ¿cuál es la probabilidad de que exactamente k artículos en la muestra (n) sean defectuosos?.



Es la distribución de probabilidad más importante para describir una variable aleatoria continua

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

μ: Promedio - media
σ: Desviación estándar

Características

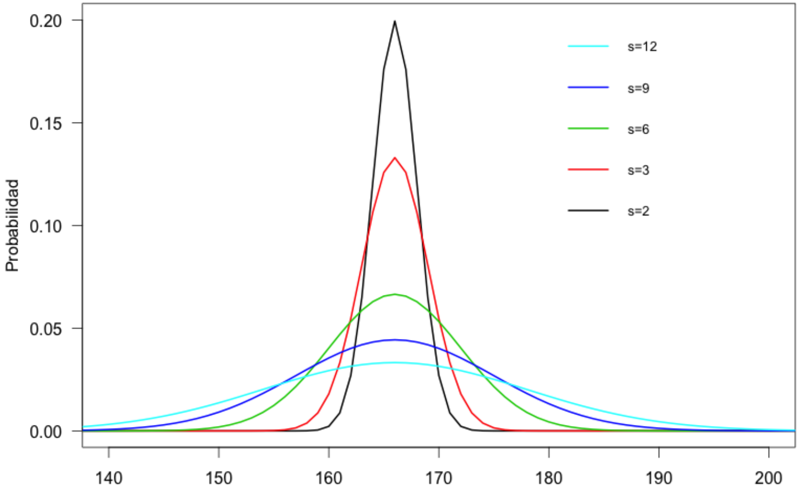
- Está caracterizada por dos parámetros: la media y la desviación estándar
- El punto más alto de la curva normal representa la media, la cual también puede indicar la mediana y la moda de la distribución
- La media de una distribución normal puede ser cualquier valor: negativo, cero o positivo

Características

- La distribución normal es simétrica. Las colas de la curva normal se extienden hasta el infinito en ambas direcciones y teóricamente nunca tocan el eje horizontal
- La desviación estándar determina el ancho y la altura de la curva normal. Valores altos de desviación estándar indican curvas anchas y achatadas indicando una mayor variabilidad de los datos
- Las probabilidad de una variable aleatoria normal está dada por el área bajo la curva. El área total bajo la curva para la distribución normal es 1. Al ser simétrica, el área tanto a la derecha como a la izquierda de la curva es 0.5

Distribución Normal

Figura: La figura muestra la función de densidad de probabilidad para una normal de media constante y diferentes desviaciones estándar



¿Son mis datos normales?

Primero es bueno echar un vistazo a los datos mediante una gráfica como el histograma, el boxplot......

Prueba formal de normalidad

- La hipótesis **nula** de esta prueba es que los datos están normalmente distribuidos.
- La hipótesis alternativa es que los datos **no** están normalmente distribuidos.
- La prueba rechaza la hipótesis de normalidad cuando los valores de p (p-value) son menores que 0.05.
- Si el valor de p es mayor que 0.05, se puede asumir que los datos están normalmente distribuidos.

Distribución normal

Prueba de Shapiro-Wilk

Se encuentra implementado en R usando la función `shapiro.test()`.

Prueba de Anderson-Darling

Esta implementado en R usando la función `ad.test()`, la cual se encuentra en el paquete `nortest`.

Funciones en R asociadas a distribuciones de probabilidad

Cuadro: En la siguiente tabla se muestran los nombres de las funciones de R asociadas a distribuciones de probabilidad

`d`	Función de densidad de probabilidad (o función de masa de probabilidad).
`p`	Función de probabilidad acumulada (los valores están siempre en el intervalos de 0 a 1).
`q`	Función cuantil - el inverso (más o menos) de la función p.
`r`	Simulación de una muestra aleatoria de la distribución.

Distribuciones continuas más comunes

Cuadro: Las distribuciones continuas más comunes son:

Exponencial	`exp`	El parámetro es la tasa (por defecto es 1). La media de la distribución es $1/Tasa$
Normal (Gaussiana)	`norm`	Es la distribución más famosa en estadística, y se le conoce como la curva de campana. Los parámetros por defecto son media $= 0$ y la desviación estándar $= 1$.
Uniforme	`unif`	Los parámetros son `min` (por defecto es 0) y `max` (por defecto es 1).

Distribuciones discretas más comunes

Cuadro: Las distribuciones discretas más comunes son:

Binomial	`binom`	Los parámetros son `size`: el número de ensayos y `prob`: es la probabilidad de éxito en un único ensayo (No hay valores por defecto).
Geométrica	`geom`	El parámetro es `prob`, la probabilidad de éxito es cada intento independiente. Tenga en cuenta que la distribución se define en términos del número de fallos antes del primer éxito.
Poisson	`pois`	El parámetro es lambda, la media.

Distribuciones menos comunes y derivadas

Cuadro: Algunas distribuciones continuas menos comunes son:

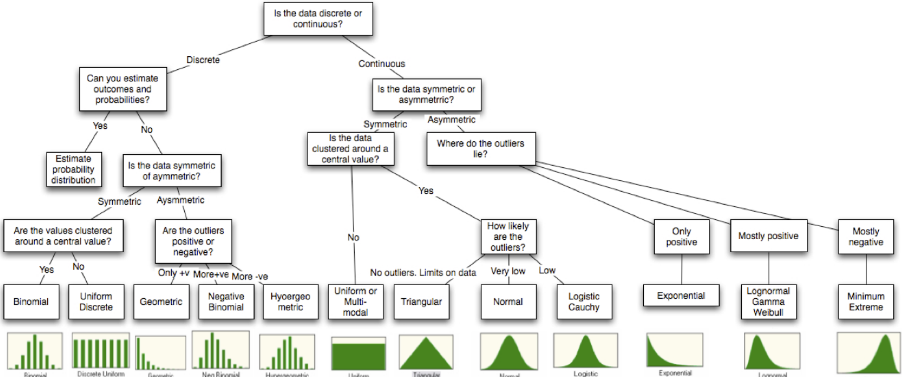
Beta	`Beta`	Los parámetros son `shape1` y `shape2` (sin valores por defecto) y `ncp` (parámetro de no-centralidad) cuyo valor predeterminado es 0, que corresponde a la beta =0.
Cauchy	`Cauchy`	Los parámetros son: `location` (0 por defecto) y `scale` (0 por defecto).
Gamma	`Gamma`	Los parámetros son `shape` y uno entre `rate` o `scale`. Los dos últimos parámetros son inversos entre si y ambos por defecto tienen un valor de `1` .
Logistica	`Logistic`	Los parámetros son `location` y `scale`.
log-normal	`Lognormal`	Los parámetros son `meanlog` y `sdlog`.
Weibull	`Weibull`	Los parámetros son `shape` y `scale` .

Distribuciones menos comunes y derivadas

Cuadro: Algunas distribuciones continuas menos comunes son:

Chi-cuadrado	`Chisquare`	Los parámetros son `df` y `ncp` parámetro de no-centralidad.
F	`FDist`	Los parámetros son `df1` y `df2`, el numerador y el denominador de los grados de libertad. Un parámetro opcional es `ncp`, el parámetro de no-centralidad. Se se omite este parámetro se asume una función `F` central.
t	`TDist`	Los parámetros son `df` y `pnc`; si se omite el parámetro de no-centralidad se asume una distribución `textbft` central.

Opciones de distribuciones, adaptado de: Probabilistic approaches to risk de Aswath Damodaran



Tomado de . . .