

### 3.- Algunas distribuciones de probabilidad importantes

#### 3.1- Familias de distribuciones

Para poder llevar a cabo la simulación de un sistema cuyas entradas son variables aleatorias, debemos conocer las distribuciones de probabilidad de estas variables aleatorias. Sabido que las variables aleatorias de entrada siguen una determinada distribución, la simulación genera los valores de estas variables a partir de las distribuciones. El objetivo de este apartado es analizar cómo, a partir de los valores experimentales aleatorios, puede establecerse cual es su distribución de probabilidad.

Por ejemplo, en el caso de la cola atendida por un único operario, deberían recogerse datos experimentales de las dos variables aleatorias de entrada: el intervalo entre llegadas sucesivas de clientes y el tiempo que tarda el empleado en atender a cada cliente. A partir de estos datos, deberemos ser capaces de estimar a que distribución de probabilidad obedece cada una de ellas.

Suponiendo que se dispone de datos acerca de las variables aleatorias, puede procederse de dos maneras:

- Ajustar los datos experimentales a diferentes tipos de distribuciones teóricas "estándar" (exponencial, normal, de Poisson, etc...) obteniendo para cada una de ellas los parámetros que proporcionan el ajuste óptimo de acuerdo a un determinado criterio. A continuación, decidir cual de los ajustes obtenido, uno para cada tipo de distribución, es el mejor. La distribución obtenida será la que se use para generar los valores de las variables aleatorias durante la simulación.
- Definir una distribución empírica, a partir de los valores experimentales, específica para nuestro problema, diferente de las distribuciones de probabilidad teóricas "estándar" comúnmente usadas.

Siempre que sea posible conviene usar el primer método, ya que:

- Los datos experimentales recogidos son, en sí, aleatorios, con lo cual, una distribución empírica, obtenida a partir de un conjunto de observaciones, puede diferir mucho de la obtenida a partir de otro conjunto de observaciones del mismo proceso. Al realizar el ajuste de una distribución teórica "estándar" a los datos experimentales estamos tratando de extraer información acerca de la naturaleza de la distribución a que obedecen; las distribuciones obtenidas de este modo suelen ser menos sensibles a las variaciones entre diferentes conjuntos de observaciones de un mismo proceso.
- Si las distribuciones empíricas se definen en el modo en que habitualmente se hace (y que veremos más adelante), no pueden generarse a partir de ellas variables aleatorias que caigan fuera del rango de los datos observados, cosa que no ocurre cuando se ajusta a una distribución teórica.

En algunas ocasiones, la información de que disponemos acerca del sistema, nos hace suponer que una variable aleatoria de entrada obedece a una determinada distribución teórica. En estos casos, es una medida aconsejable contrastar la hipótesis con los datos experimentales.

En lo sucesivo supondremos que nuestros datos son observaciones independientes y que bajo todos ellos subyace una misma distribución. Cuando esto no sucede, por ejemplo, cuando los datos no son independientes, la mayoría de las técnicas que discutiremos no son aplicables directamente.

Para una misma familia de distribuciones (por ejemplo, la familia de las distribuciones normales) existen diferentes maneras de definir, o parametrizar, la densidad de probabilidad o la distribución de probabilidad acumulada. La mayoría de los parámetros empleados para definir una distribución pueden clasificarse, de acuerdo con su interpretación física o geométrica, en uno de estos tres tipos: parámetros de posición, de escala o de forma.

Un *parámetro de posición*,  $\gamma$  especifica la posición en el eje de abscisas (eje x) del rango de valores de la distribución; normalmente,  $\gamma$  es el punto medio del rango o su extremo inferior. Una variación en  $\gamma$  supone un desplazamiento de la distribución, a la derecha o a la izquierda, sin variar ninguna otra de sus características.

Un *parámetro de escala*,  $\beta$ , determina la escala (o unidades) de medida de los valores en el rango de la distribución. Cuando  $\gamma$  está fijo a 0, un cambio en  $\beta$  comprime o expande la distribución sin alterar su forma básica.

Un *parámetro de forma*,  $\alpha$ , determina, con independencia de la posición y la escala, la forma básica de la distribución dentro de la familia general de distribuciones de interés. Un cambio en  $\alpha$  generalmente altera las propiedades de la distribución más profundamente que un cambio en su posición o escala. Algunas distribuciones no tienen ningún parámetro de forma (por ejemplo, la exponencial y la normal), mientras que otras poseen varios (por ejemplo, la distribución beta tiene dos).

Dos variables aleatorias  $X$  e  $Y$ , de la misma familia de distribuciones, se dice que difieren sólo en la posición cuando existe un número real,  $\gamma$ , tal que  $\gamma + X$  e  $Y$  tienen la misma distribución.

Similarmente, se dice que  $X$  e  $Y$  difieren sólo en la escala cuando existe algún real positivo,  $\beta$ , tal que  $\beta X$  tiene la misma distribución que  $Y$ .

Finalmente,  $X$  e  $Y$  difieren en la posición y en la escala cuando existen dos reales  $\gamma$  y  $\beta$  tales que  $\gamma + \beta X$  tiene la misma distribución que  $Y$ .

Sin embargo, si las distribuciones de  $X$  e  $Y$  poseen diferente valor de un parámetro de forma, no podrá hacerse que posean la misma distribución mediante un cambio de escala y de posición.

A continuación incluimos una relación de algunas de las distribuciones teóricas, continuas y discretas, que más se emplean en simulación. Seguidamente veremos qué formas hay de especificar una distribución empírica.

### 3.2- Distribuciones continuas

#### 3.2.1- Distribución uniforme, U(a,b)

Suele emplearse como modelo en "primera aproximación" cuando se sabe que la variable aleatoria puede tomar valores entre a y b pero no se tiene más información acerca de ella. Como veremos más adelante, la distribución U(0,1) es esencial en la generación de números aleatorios a partir de cualquier distribución.

La función densidad de probabilidad es:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{si } a \leq x \leq b \\ 0 & \text{en cualquier otro caso} \end{cases}$$

La distribución de probabilidad acumulada es:

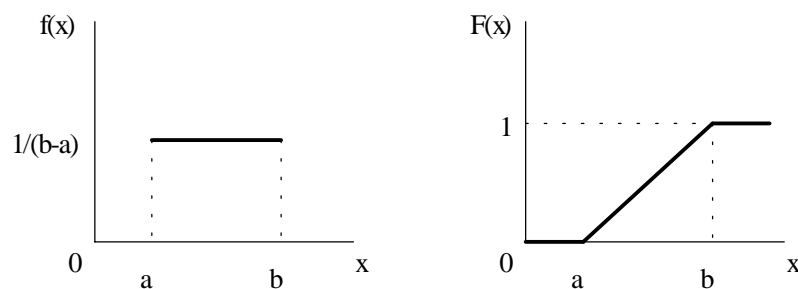
$$F(x) = \begin{cases} 0 & \text{si } x < a \\ \frac{x-a}{b-a} & \text{si } a \leq x \leq b \\ 1 & \text{si } x > b \end{cases}$$

La familia de distribuciones posee dos parámetros a y b reales, con  $a < b$ . a es el parámetro de posición y (b-a) es el parámetro de escala. El rango de la distribución es el intervalo [a,b]. La media y la varianza valen:

$$\text{Media: } \frac{a+b}{2} \qquad \text{Varianza: } \frac{(b-a)^2}{12}$$

Si X esta distribuida U(0,1), la probabilidad de que X adquiera un valor en el intervalo  $[x, x + \Delta x] \subset [0,1]$  es igual a  $\Delta x$ :

$$P\{X \in [x, x + \Delta x]\} = \int_x^{x+\Delta x} 1 dy = (x + \Delta x) - x = \Delta x$$



### 3.2.2- Distribución exponencial, $\text{expo}(\beta)$

Suele emplearse para modelar intervalos de tiempo entre eventos independientes que ocurren a frecuencia constante.

La función densidad de probabilidad es:

$$f(x) = \begin{cases} \left(\frac{1}{\beta}\right)e^{-\frac{x}{\beta}} & \text{si } x \geq 0 \\ 0 & \text{en cualquier otro caso} \end{cases}$$

La distribución de probabilidad acumulada es:

$$F(x) = \begin{cases} 1 - e^{-\frac{x}{\beta}} & \text{si } x \geq 0 \\ 0 & \text{en cualquier otro caso} \end{cases}$$

$\beta$  es un parámetro de escala,  $\beta > 0$ . El rango de la distribución es  $[0, \infty)$ . La media y la varianza valen:

$$\text{Media: } \beta \qquad \text{Varianza: } \beta^2$$

Se dice que una variable aleatoria  $X$  no tiene memoria cuando:

$$P\{X > t + s | X > t\} = P\{X > s\} \quad \text{para todo } t, s > 0$$

donde la probabilidad condicionada  $P\{X > t + s | X > t\}$  es la probabilidad de que ocurra el evento  $\{X > t + s\}$  supuesto que previamente se ha satisfecho  $\{X > t\}$ .

La distribución exponencial es la única distribución continua que posee la propiedad de no tener memoria:

$$P\{X > t + s | X > t\} = \frac{P\{X > t + s\}}{P\{X > t\}} = \frac{e^{-\frac{t+s}{\beta}}}{e^{-\frac{t}{\beta}}} = e^{-\frac{s}{\beta}} = P\{X > s\} \quad \text{para todo } t, s > 0$$

### 3.2.3- Distribución Gamma, $\text{gamma}(\alpha, \beta)$

Suele emplearse para modelar el tiempo necesario para completar una tarea, por ejemplo el tiempo de servicio de un cliente, el tiempo necesario para reparar una máquina, la duración de la vida del equipo industrial, etc.

La función densidad de probabilidad es:

$$f(x) = \begin{cases} \frac{x^{\alpha-1} e^{-\frac{x}{\beta}}}{\beta^\alpha \Gamma(\alpha)} & \text{si } x > 0 \\ 0 & \text{en cualquier otro caso} \end{cases}$$

donde  $\Gamma(\alpha)$  es la función gamma, definida:

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt \quad \text{para cualquier real } z > 0 \text{ (ver Apéndice).}$$

Se cumple:

$$\lim_{x \rightarrow 0} f(x) = \begin{cases} \infty & \text{si } \alpha < 1 \\ \frac{1}{\beta} & \text{si } \alpha = 1 \\ 0 & \text{si } \alpha > 1 \end{cases}$$

La distribución de probabilidad acumulada es, en el caso en que  $\alpha$  sea un entero positivo:

$$F(x) = \begin{cases} 1 - e^{-\frac{x}{\beta}} \sum_{j=0}^{\alpha-1} \frac{\left(\frac{x}{\beta}\right)^j}{j!} & \text{si } x > 0 \\ 0 & \text{en cualquier otro caso} \end{cases}$$

en el caso en que  $\alpha$  no sea un entero positivo no existe una expresión analítica.

Posee un parámetro de forma  $\alpha > 0$  y uno de escala  $\beta > 0$ . El rango de la distribución es  $[0, \infty)$ . La media y la varianza valen:

Media:  $\alpha\beta$

Varianza:  $\alpha\beta^2$

Si  $X_1, X_2, \dots, X_m$  son independientes distribuidas  $\text{expo}(\beta)$ , entonces  $X_1 + X_2 + \dots + X_m$  esta distribuida  $\text{gamma}(m, \beta)$ .  $\text{gamma}(m, \beta)$ , con  $m$  entero positivo, se llama distribución m-Erlang( $\beta$ ).

$\text{expo}(\beta)$  y  $\text{gamma}(1, \beta)$  son la misma distribución.

La distribución *chi-cuadrado* con  $k$  grados de libertad es la  $\text{gamma}(k/2, 2)$ . Véase Tabla II.

### 3.2.4- Distribución normal o de Gauss, $N(\mu, \sigma^2)$

Representa cantidades que son suma de un gran número de otras cantidades (en virtud de los teoremas del límite central).

La función densidad de probabilidad es:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$

Posee un parámetro de posición,  $\mu \in (-\infty, \infty)$ , y un parámetro de escala,  $\sigma > 0$ . El rango en el cual la variable aleatoria puede tomar valores es  $(-\infty, \infty)$ . La media y la varianza valen:

Media:  $\mu$

Varianza:  $\sigma^2$

La función densidad de probabilidad es simétrica respecto al valor de la media,  $f(\mu - x) = f(\mu + x)$ , y tiene forma de campana, tanto más estrecha y puntiaguda cuanto menor sea la varianza. El valor máximo de  $f(x)$  ocurre para  $x = \mu$ .

La función probabilidad acumulada:

$$F(x_1) = P\{X \leq x_1\} = \int_{-\infty}^{x_1} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx, \quad -\infty < x_1 < \infty$$

no tiene expresión analítica, debiéndose usar métodos numéricos para evaluarla. Parecería que es preciso calcular la integral para cada par de valores  $(\mu, \sigma^2)$ , sin embargo, el cambio de variables  $z = \frac{x-\mu}{\sigma}$  permite hacer la evaluación independiente de  $\mu$  y  $\sigma$ .

Si  $X$  esta distribuida  $N(\mu, \sigma^2)$ , entonces  $Z = \frac{X-\mu}{\sigma}$  tiene distribución normal estándar,  $N(0,1)$ :

$$\begin{aligned} F(x) &= P\{X \leq x\} = P\left\{\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right\} = P\left\{Z \leq \frac{x-\mu}{\sigma}\right\} = \\ &= \int_{-\infty}^{\frac{x-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = \int_{-\infty}^{\frac{x-\mu}{\sigma}} \phi(z) dz = \Phi\left(\frac{x-\mu}{\sigma}\right) \end{aligned}$$

donde:

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

es la función densidad de probabilidad  $N(0,1)$ . En la Tabla III están tabulados los valores de la integral de  $\phi(z)$  entre 0 y  $z$ , para  $z > 0$ .

**Ejemplo.** X esta distribuida  $N(50,9)$ . Calcular  $F(56) = P\{X \leq 56\}$ .

De la ecuación:  $F(x) = P\{X \leq x\} = \Phi\left(\frac{x-\mu}{\sigma}\right)$

obtenemos:  $F(56) = P\{X \leq 56\} = \Phi\left(\frac{56-50}{3}\right) = \Phi(2)$

de la Tabla III vemos que el valor de la integral de la función densidad de probabilidad  $N(0,1)$  entre 0 y 2 vale 0.4772. Por consiguiente, el valor entre  $-\infty$  y 2 es  $0.5+0.4772=0.9772$ .

En general, un *punto crítico*  $z_\gamma$  de la variable aleatoria Z se define como aquel valor de la variable que cumple  $P\{Z \leq z_\gamma\} = \gamma$ . Existen tablas de puntos críticos para diversos tipos de distribuciones de Z. En el caso en que Z sea una variable aleatoria normal estándar,  $N(0,1)$ , los puntos  $z_\gamma$  se llaman puntos críticos normales y  $\gamma = P\{Z \leq z_\gamma\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_\gamma} e^{-\frac{y^2}{2}} dy$ . Las tablas, para un valor  $\gamma$  de la probabilidad acumulada, dan el valor del punto crítico  $z_\gamma$  y viceversa.

$\gamma$	0.600	0.700	0.800	0.900	0.933	0.950	0.960	0.967	0.975	0.980	0.983	0.987	0.990	0.992	0.994	0.995
$z_\gamma$	0.253	0.524	0.842	1.282	1.501	1.645	1.751	1.834	1.960	2.054	2.127	2.241	2.326	2.395	2.501	2.576

Tabla: Puntos críticos normales  $z_\gamma$ , donde Z es la variable aleatoria normal estándar. Corresponde con la última fila de la Tabla I.

Si dos distribuciones normales tienen la misma media, pero la desviación estándar de una de ellas es la mitad que la de la otra, habrá tanta probabilidad acumulada dentro de un intervalo central de amplitud  $L/2$  para la primera distribución como dentro de un intervalo central de amplitud  $L$  para la segunda distribución.

En este sentido, la desviación estándar de una variable normal mide el grado de concentración de la variable aleatoria con respecto a su media. Puesto que la desviación estándar de una media muestral  $\bar{X}$  normal mide la concentración de las  $\bar{X}$  muestrales con respecto a  $\mu$  y puede, por lo tanto, considerarse como una medida de la precisión en la estimación de  $\mu$ , es necesario tomar una muestra cuatro veces más grande para duplicar la precisión de la estimación.

Si dos variables aleatorias normales no están correlacionadas, entonces son independientes. Esta implicación no es válida, en general, para el resto de distribuciones.

El *teorema del límite central* establece que si las variables aleatorias  $X_1, X_2, \dots, X_n$  son muestras aleatorias de una misma distribución de cualquier tipo, de media  $\mu$  y varianza  $\sigma^2$ , entonces, si  $n$  es un número suficientemente grande, la variable aleatoria  $\bar{X}$  esta distribuida aproximadamente por una distribución normal de media  $\mu$  y varianza  $\frac{\sigma^2}{n}$ ,  $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ .

El valor que debe tomar  $n$  para poder ser considerado "suficientemente grande" depende de la precisión que se requiera en la aproximación y de lo parecida que sea la distribución de las muestras aleatorias  $X_1, X_2, \dots, X_n$  a una distribución normal. Si la distribución es aproximadamente simétrica y con un pico central, como la normal, quizá con valores de  $n$  del orden de 5 sean adecuados. En cambio, si la forma de la distribución difiere considerablemente de la normal, por ejemplo, si es exponencial, puede ser preciso un valor de  $n$  mayor que 30 para que la aproximación sea útil.

### 3.2.5- Distribución beta, $\text{beta}(\alpha_1, \alpha_2)$

Suele usarse como modelo en "primera aproximación" en ausencia de datos experimentales.

La función densidad de probabilidad es:

$$f(x) = \begin{cases} \frac{x^{\alpha_1-1}(1-x)^{\alpha_2-1}}{B(\alpha_1, \alpha_2)} & \text{si } 0 < x < 1 \\ 0 & \text{en cualquier otro caso} \end{cases}$$

donde  $B(\alpha_1, \alpha_2)$  es la función beta, definida:

$$B(z_1, z_2) = \int_0^1 t^{z_1-1} (1-t)^{z_2-1} dt \quad \text{para todo } z_1 > 0, z_2 > 0 \text{ reales}$$

y cuyo cálculo se basa en la propiedad:

$$B(z_1, z_2) = \frac{\Gamma(z_1)\Gamma(z_2)}{\Gamma(z_1 + z_2)}$$

Se cumple:

$$\lim_{x \rightarrow 0} f(x) = \begin{cases} \infty & \text{si } \alpha_1 < 1 \\ \alpha_2 & \text{si } \alpha_1 = 1 \\ 0 & \text{si } \alpha_1 > 1 \end{cases} \quad \lim_{x \rightarrow 1} f(x) = \begin{cases} \infty & \text{si } \alpha_2 < 1 \\ \alpha_1 & \text{si } \alpha_2 = 1 \\ 0 & \text{si } \alpha_2 > 1 \end{cases}$$

No existe una expresión analítica general para la distribución de probabilidad acumulada.

$\alpha_1 > 0$  y  $\alpha_2 > 0$ , son parámetros de forma. El rango en el cual la variable aleatoria puede tomar valores es  $[0,1]$ . La media y la varianza valen:

$$\text{Media: } \frac{\alpha_1}{\alpha_1 + \alpha_2} \quad \text{Varianza: } \frac{\alpha_1 \alpha_2}{(\alpha_1 + \alpha_2)^2 (\alpha_1 + \alpha_2 + 1)}$$

Obsérvese que  $U(0,1)$  y  $\text{beta}(1,1)$  son la misma distribución.

Una variable aleatoria  $X$ , distribuida  $\text{beta}(\alpha_1, \alpha_2)$  en  $[0,1]$ , puede reescalar y desplazarse para obtener una variable aleatoria  $\text{beta}(\alpha_1, \alpha_2)$  en  $[a,b]$ , con los mismos parámetros de forma, mediante la transformación  $a+(b-a)X$ .

Si  $X$  esta distribuida  $\text{beta}(\alpha_1, \alpha_2)$ , entonces  $1-X$  esta distribuida  $\text{beta}(\alpha_2, \alpha_1)$ .

La densidad de probabilidad es simétrica respecto a  $x=0.5$  sólo en el caso en que  $\alpha_2 = \alpha_1$ .



### 3.2.6- Distribución triangular, $\text{triang}(a,b,c)$

Suele usarse como modelo "en primera aproximación" en ausencia de datos experimentales.

La función densidad de probabilidad vale:

$$f(x) = \begin{cases} 0 & \text{si } x < a \\ \frac{2(x-a)}{(b-a)(c-a)} & \text{si } a \leq x \leq c \\ \frac{2(b-x)}{(b-a)(b-c)} & \text{si } c < x \leq b \\ 0 & \text{si } x > b \end{cases}$$

La distribución de probabilidad acumulada vale:

$$F(x) = \begin{cases} 0 & \text{si } x < a \\ \frac{(x-a)^2}{(b-a)(c-a)} & \text{si } a \leq x \leq c \\ 1 - \frac{(b-x)^2}{(b-a)(b-c)} & \text{si } c < x \leq b \\ 1 & \text{si } b < x \end{cases}$$

Los parámetros  $a$ ,  $b$  y  $c$  deben ser números reales, satisfaciéndose  $a < c < b$ .  $a$  es el parámetro de posición,  $(b-a)$  es el parámetro de escala y  $c$  es el parámetro de forma. El rango en el cual la variable aleatoria puede tomar valores es  $[a,b]$ . La media y la varianza valen:

$$\text{Media: } \frac{a+b+c}{3} \qquad \text{Varianza: } \frac{a^2 + b^2 + c^2 - ab - ac - bc}{18}$$

### 3.2.7- Ejercicio

Representar gráficamente las siguientes densidades de probabilidad:

$\text{expo}(1)$ ,  
 $\text{gamma}(0.5,1)$ ,  $\text{gamma}(1,1)$ ,  $\text{gamma}(2,1)$ ,  $\text{gamma}(3,1)$ ,  
 $N(0,1)$ ,  
 $\text{beta}(1,1)$ ,  $\text{beta}(1,2)$ ,  $\text{beta}(2,1)$ ,  $\text{beta}(2,2)$ ,  $\text{beta}(0.8,0.2)$ ,  $\text{beta}(0.2,0.8)$ ,  $\text{beta}(0.5,0.5)$ ,  $\text{beta}(1.5,5)$ ,  
 $\text{beta}(5,1.5)$ ,  
 $\text{triang}(0,1,0)$ ,  $\text{triang}(0,1,0.5)$

### 3.3- Distribuciones discretas

#### 3.3.1- Distribución de Bernoulli, Bernoulli(b)

La variable aleatoria es el resultado de un experimento con dos únicos posibles resultados. Por ejemplo, una variable aleatoria de Bernoulli(b) puede representar el hecho de que un evento suceda o no, si la probabilidad de que suceda es b. En este caso, el espacio muestra sería {no sucede , sucede}. La variable aleatoria asocia un número a cada uno de los elementos del espacio muestra: si no sucede,  $X=0$ , y si sucede,  $X=1$ . La función de probabilidad asocia un número (probabilidad) a cada uno de los posibles valores de la variable aleatoria: asocia  $(1-b)$  a  $X=0$  y  $b$  a  $X=1$ .

Suele emplearse para generar otras variables aleatorias discretas, como la binomial, la geométrica y la binomial negativa.

Supongamos que una variable aleatoria de Bernoulli(b) representa el hecho de que en el experimento suceda un evento ( $X=1$ ) o no ( $X=0$ ), por ejemplo, el experimento consiste en lanzar una moneda y el evento en que salga cara. La probabilidad de que el evento suceda es b (en el ejemplo de la moneda,  $b=0.5$ ) . Supongamos que repetimos el experimento el número de veces necesarias hasta que sucede el evento. Entonces, el número de experimentos "fallidos" antes del primero en que sucede el evento es una variable aleatoria discreta con distribución geométrica de parámetro b.

Si repetimos el experimento hasta que suceda el evento un determinado número entero s de veces, entonces el número de intentos "fallidos" (en que no sucede el evento) antes de que produzca el evento s-ésimo es una variable aleatoria de distribución binomial negativa con parámetros s y b.

El número de eventos que suceden en s experimentos, donde la probabilidad de que se produzca el evento en cada experimento es b, es una variable aleatoria de distribución binomial. De esto se deduce que la distribución de Bernoulli(p) es un caso especial de la binomial, ya que es idéntica a la binomial con  $s=1$  y con el mismo valor de p.

La función de probabilidad es:

$$p(x) = \begin{cases} 1-b & \text{si } x = 0 \\ b & \text{si } x = 1 \\ 0 & \text{en cualquier otro caso} \end{cases}$$

La función de distribución acumulada es:

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1-b & \text{si } 0 \leq x < 1 \\ 1 & \text{si } x \geq 1 \end{cases}$$

El parámetro b debe estar comprendido en el intervalo (0,1). El rango de valores que puede tomar la variable aleatoria es  $\{0,1\}$ , es decir, puede valer 0 ó 1. La media y la varianza valen:

Media: b

Varianza:  $b(1-b)$

### 3.3.2- Distribución uniforme discreta, DU(i,j)

La variable aleatoria es el resultado de un experimento con varios posibles resultados, todos ellos igualmente probables. Suele emplearse como modelo en "primera aproximación" para una cantidad entera que varía entre los enteros  $i$  y  $j$  ( $i < j$ ) y sobre la cual no se tiene más información.

La función de probabilidad es:

$$p(x) = \begin{cases} \frac{1}{j-i+1} & \text{si } x \in \{i, i+1, \dots, j\} \\ 0 & \text{en cualquier otro caso} \end{cases}$$

La distribución de probabilidad acumulada es:

$$F(x) = \begin{cases} 0 & \text{si } x < i \\ \frac{\lfloor x \rfloor - i + 1}{j - i + 1} & \text{si } i \leq x \leq j \\ 1 & \text{si } x > j \end{cases}$$

donde  $\lfloor x \rfloor$  nota el mayor entero menor o igual que  $x$ .

Los parámetros  $i$  y  $j$  son enteros.  $i$  es el parámetro de posición. El rango de la distribución son los valores  $\{i, i+1, \dots, j\}$ . La media y la varianza son:

$$\text{Media: } \frac{i+j}{2} \qquad \text{Varianza: } \frac{(j-i+1)^2 - 1}{12}$$

### 3.3.3- Distribución binomial, bin(t,b)

Número de eventos que se producen en  $t$  experimentos de Bernoulli (experimentos con 2 posibles resultados: sucede el evento o no) independientes, donde la probabilidad de que suceda el evento en cada experimento es  $b$ .

Típicamente surge una distribución binomial cuando estamos interesados en el número de miembros de un grupo de individuos, elegido aleatoriamente, que poseen cierta característica. Por ejemplo, el número de personas, de un grupo de 1000, que simpatizan con cierta idea política, el número de piezas defectuosas que se encuentran en una caja de 200 piezas, el número de veces que se obtiene cara si se lanza 10 veces una moneda, etc.

En todos los casos se satisfacen 3 condiciones:

- cada individuo tiene la misma probabilidad de poseer la característica.
- el hecho de que un individuo posea la característica es independiente del hecho de que otro la posea. El hecho de saber que ciertos individuos posean la característica no nos ayuda a decidir si otro individuo en concreto la tiene.
- el número de individuos es fijo y conocido a priori.

En general, dada una muestra aleatoria de  $t$  individuos, cada uno de los cuales tiene una probabilidad  $b$  de poseer la característica, si la variable aleatoria de interés  $X$  es el número de individuos de la muestra que poseen la característica, entonces se dice que  $X$  tiene una distribución binomial con parámetros  $t$  y  $b$ ,  $\text{bin}(t,b)$ .

La función de probabilidad es:

$$p(x) = \begin{cases} \binom{t}{x} b^x (1-b)^{t-x} & \text{si } x \in \{0,1,\dots,t\} \\ 0 & \text{en cualquier otro caso} \end{cases}$$

donde el coeficiente binomial  $\binom{t}{x}$  se define:  $\binom{t}{x} = \frac{t!}{x!(t-x)!}$

La distribución de probabilidad acumulada es:

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ \sum_{i=0}^{\lfloor x \rfloor} \binom{t}{i} b^i (1-b)^{t-i} & \text{si } 0 \leq x \leq t \\ 1 & \text{si } x > t \end{cases}$$

donde  $\lfloor x \rfloor$  nota el mayor entero menor o igual que  $x$ .

$t$  es un entero positivo y  $b \in (0,1)$ . El rango de valores que puede tomar la variable aleatoria es  $\{0,1,\dots,t\}$ . La media y la varianza valen:

Media:  $tb$

Varianza:  $tb(1-b)$

Veamos unos ejemplos:

Si un dado perfecto se lanza 5 veces, ¿cual es la probabilidad de que en 2 tiradas salga un 1?. En este caso el evento consiste en obtener un 1,  $b=1/6$  el número de experimentos es  $t=5$  y el número de eventos debe ser  $x=2$ . La probabilidad es:

$$p(2) = \binom{5}{2} \left(\frac{1}{6}\right)^2 \left(1 - \frac{1}{6}\right)^{5-2} = \frac{5!}{2!3!} \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^3 = 0.16$$

Calcular la probabilidad de conseguir por lo menos 2 dianas en 20 disparos hechos contra un blanco, si la probabilidad de hacer diana en un solo disparo es 0.1. En este caso  $b=0.1$ ,  $t=20$ , con lo cual la probabilidad es:

$$P\{X \geq 2\} = 1 - p(0) - p(1) = 1 - F(1) = 0.608$$

Si el número de pruebas  $t$  es grande, los cálculos que conlleva el aplicar la fórmula de la probabilidad se hacen muy largos, con lo cual, en estos casos, sería útil disponer a una aproximación conveniente de la distribución binomial. Para  $t$  grande hay dos funciones de densidad que se aproximan mucho a la distribución binomial: una cuando  $b$  (probabilidad de que suceda el evento en 1 experimento) es muy pequeña (Poisson) y otra cuando no lo es (normal).

Si la probabilidad  $b$  de que suceda el evento en un experimento tiende a cero mientras que el número de experimentos,  $t$ , tiende a infinito de una forma tal que la media  $\lambda = bt$  permanezca constante, la distribución binomial se aproxima a la distribución de Poisson con media  $\lambda$ . Aunque la distribución de Poisson se ha introducido por medio de su propiedad de aproximarse a la binomial, es un modelo muy útil para tratar ciertos tipos de problemas no relacionados con la binomial.

Si  $X$  representa el número de eventos acaecidos en  $t$  experimentos independientes, donde  $b$  es la probabilidad de que suceda un evento en un experimento, entonces la proporción de eventos,  $X/t$ , tendrá una distribución aproximadamente normal, con media  $\mu = b$  y desviación estándar  $\sigma = \sqrt{\frac{b(1-b)}{t}}$ , si  $t$  es suficientemente grande.

Esto es equivalente a considerar que la binomial  $\text{bin}(t, b)$  puede aproximarse por la normal  $N(tb, \sqrt{tb(1-b)})$ . La aproximación suele considerarse aceptable si se cumple  $tb > 5$ .

### 3.3.4- Distribución geométrica, $\text{geom}(b)$

Si repetimos un experimento de Bernoulli hasta que sucede el evento, el número de experimentos "fallidos" antes del primero en que sucede el evento es una variable aleatoria discreta con distribución geométrica de parámetro  $b$ . Por ejemplo, el número de personas preguntadas de un grupo de 1000 antes de encontrar una que simpatice con determinada idea política, el número de piezas revisadas de una caja de 200 antes de encontrar una defectuosa, el número de lanzamientos de la moneda antes de que salga cara, etc.

La función probabilidad es:

$$p(x) = \begin{cases} b(1-b)^x & \text{si } x \in \{0, 1, \dots\} \\ 0 & \text{en cualquier otro caso} \end{cases}$$

La distribución de probabilidad acumulada es:

$$F(x) = \begin{cases} 1 - (1-b)^{\lfloor x \rfloor + 1} & \text{si } x \geq 0 \\ 0 & \text{en cualquier otro caso} \end{cases}$$

El parámetro  $b$  debe tomar un valor en el intervalo  $(0, 1)$ . La media y la varianza valen:

$$\text{Media: } \frac{1-b}{b} \qquad \text{Varianza: } \frac{1-b}{b^2}$$

### 3.3.5- Distribución negativa binomial, $\text{negbin}(s,b)$

Si se repite un experimento de Bernoulli hasta que suceda el evento un determinado número entero  $s$  de veces, el número de intentos "fallidos" (en que no sucede el evento) antes de que produzca el evento  $s$ -ésimo es una variable aleatoria de distribución binomial negativa con parámetros  $s$  y  $b$ . Por ejemplo, el número de personas preguntadas de un grupo de 1000 que no simpatizan con una determinada idea política antes de encontrar a  $s$  que si simpaticen, el número de piezas buenas revisadas de una caja de 200 antes de encontrar  $s$  piezas defectuosas, el número de lanzamientos de la moneda en que sale cruz antes de que salga cara  $s$  veces, etc.

La función probabilidad es:

$$p(x) = \begin{cases} \binom{s+x-1}{x} b^s (1-b)^x & \text{si } x \in \{0,1,\dots\} \\ 0 & \text{en cualquier otro caso} \end{cases}$$

La distribución de probabilidad acumulada es:

$$F(x) = \begin{cases} \sum_{i=0}^{\lfloor x \rfloor} \binom{s+i-1}{i} b^s (1-b)^i & \text{si } x \geq 0 \\ 0 & \text{en cualquier otro caso} \end{cases}$$

$s$  es un entero positivo y  $p \in (0,1)$ . La media y la varianza valen:

$$\text{Media: } \frac{s(1-b)}{b} \qquad \text{Varianza: } \frac{s(1-b)}{b^2}$$

La distribución geométrica es un caso especial de la binomial negativa, ya que  $\text{geom}(b)$  es la misma distribución que  $\text{negbin}(1,b)$ .

### 3.3.6- Distribución de Poisson, $\text{Poisson}(\lambda)$

La distribución de Poisson surge normalmente cuando estamos interesados en el número de eventos aleatorios que suceden en un intervalo fijo. Son problemas en que una variable aleatoria se distribuye a lo largo del tiempo o del espacio. Por ejemplo, el número de llamadas recibidas en una centralita telefónica durante un periodo de 30 segundos, el número de pequeños defectos de fabricación encontrados en una fibra de vidrio de 1 km. de longitud, el número de meteoritos encontrados en una hectárea de tierra desierta.

En todos los casos, las condiciones necesarias para que se trate de una distribución de Poisson son:

- los eventos de interés deben ocurrir independientemente unos de otros.
- la probabilidad de que suceda un evento en un intervalo depende de la longitud del intervalo y no de su posición.

La función probabilidad es:

$$p(x) = \begin{cases} \frac{e^{-\lambda} \lambda^x}{x!} & \text{si } x \in \{0, 1, \dots\} \\ 0 & \text{en cualquier otro caso} \end{cases}$$

La distribución de probabilidad acumulada es:

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ e^{-\lambda} \sum_{i=0}^{\lfloor x \rfloor} \frac{\lambda^i}{i!} & \text{si } x \geq 0 \end{cases}$$

El parámetro  $\lambda$ , mayor que cero, es la media y la varianza de la distribución.

Como ejemplo supongamos que se sabe que la cifra media de llamadas que recibe una central telefónica es 5 por minuto. Si la central puede manejar un máximo de 8 llamadas por minuto. ¿Cual es la probabilidad de que sea incapaz de manejar todas las llamadas que le lleguen en un minuto?. La probabilidad deseada puede obtenerse calculando la probabilidad de recibir 8 o menos llamadas y después restando esta probabilidad de 1. Tomando  $\lambda = 5$  en la distribución de probabilidad acumulada:

$$P\{X \leq 8\} = F(8) = e^{-5} \sum_{i=0}^8 \frac{5^i}{i!} = 0.932$$

y la probabilidad de que la central telefónica sea desbordada es  $P\{X > 8\} = 1 - 0.932 = 0.068$

### 3.3.7- Ejercicio

Representar gráficamente las siguientes funciones de probabilidad: Bernoulli(0.6), DU(0,5), bin(5,0.1), bin(10,0.1), bin(5,0.5), bin(10,0.5), geom(0.25), geom(0.5), negbin(2,0.5), negbin(5,0.5), Poisson(0.5), Poisson(1), Poisson(2), Poisson(6).

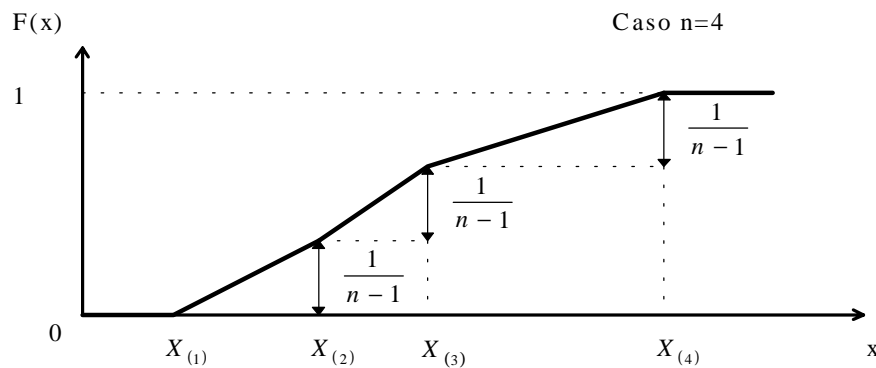
### 3.4-Distribuciones empíricas

En algunas ocasiones, puede ser necesario usar los propios datos aleatorios experimentales en sí para definir directamente una distribución, llamada distribución empírica, en vez de ajustar una distribución teórica a los datos experimentales.

Los tipos de distribuciones empíricas que pueden definirse para variables aleatorias continuas dependen de si disponemos de los datos originales de las observaciones,  $X_1, X_2, \dots, X_n$ , o bien sólo disponemos de un histograma de los datos, es decir, del número de datos que caen dentro de cada uno de una serie de intervalos especificados.

Si se dispone de los datos originales, puede definirse una función de distribución de probabilidad acumulada  $F$  continua y lineal a tramos. Para ello deben ordenarse los datos  $X_i$  en orden creciente. Notaremos  $X_{(i)}$  la  $i$ -ésima menor muestra, es decir,  $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ . Entonces,  $F$  viene dada por:

$$F(x) = \begin{cases} 0 & \text{si } x < X_{(1)} \\ \frac{i-1}{n-1} + \frac{x - X_{(i)}}{(n-1)(X_{(i+1)} - X_{(i)})} & \text{si } X_{(i)} \leq x < X_{(i+1)} \text{ para } i = 1, 2, \dots, n \\ 1 & \text{si } x \geq X_{(n)} \end{cases}$$



Una desventaja de este tipo de distribución empírica es que las variables aleatorias generadas durante la simulación no podrán tener valores menores que  $X_{(1)}$  ni mayores que  $X_{(n)}$ .

Si se dispone de los datos agrupados en forma de histograma debe seguirse otro método, ya que no se dispone de los valores individuales de  $X_1, X_2, \dots, X_n$ . Supongamos que los datos experimentales,  $X_1, X_2, \dots, X_n$ , están agrupados en  $k$  intervalos adyacentes  $[a_0, a_1), [a_1, a_2), \dots, [a_{k-1}, a_k)$ , de modo que el intervalo  $j$ -ésimo contiene  $n_j$  observaciones, con lo cual  $n_1 + n_2 + \dots + n_k = n$ . Normalmente las  $a_j$  están equiespaciadas, si bien, no imponemos esa condición.



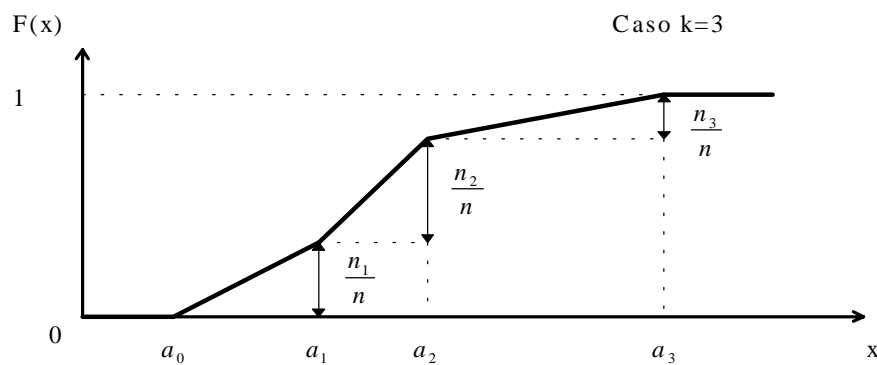
Puede definirse una posible función de distribución de probabilidad acumulada,  $G$ , lineal a tramos, asignando:

$$G(a_0) = 0 \quad G(a_j) = \frac{n_1 + n_2 + \dots + n_j}{n} \quad \text{para } j: 1, 2, \dots, k$$

Interpolando linealmente entre las  $a_j$  definimos:

$$G(x) = \begin{cases} 0 & \text{si } x < a_0 \\ G(a_{j-1}) + \frac{x - a_{j-1}}{a_j - a_{j-1}} [G(a_j) - G(a_{j-1})] & \text{si } a_{j-1} \leq x \leq a_j \text{ para } j: 1, 2, \dots, k \\ 1 & \text{si } a_k \leq x \end{cases}$$

Igual que en el caso anterior, las variables aleatorias generadas mediante estas distribuciones no pueden tomar valores menores que  $a_0$  ni mayores que  $a_k$ .



Estos dos métodos son sólo dos de los muchos posibles para definir distribuciones empíricas de variables aleatorias continuas. Existen gran variedad de métodos, entre ellos los que introducen en uno o a ambos lados una "cola exponencial" para eliminar el problema de la acotación en los valores de la variable aleatoria.

Cuando los datos son discretos, y se dispone de ellos, es muy sencillo definir una distribución empírica: para cada posible valor  $x$  puede definirse la probabilidad  $p(x)$  como la proporción de los datos que son iguales a  $x$ .

Cuando los datos son discretos, pero se dispone de ellos agrupados en forma de histograma, puede definirse una función distribución tal que la suma de las  $p(x)$  sobre todos los posibles valores de  $x$  en el intervalo sea igual a la proporción de las medidas que se encuentra en el intervalo. Es arbitrario como se asignan valores a los  $p(x)$  de los posibles  $x$  en el intervalo.

### 3.5- Desplazamiento y truncado de las distribuciones

Algunas de las distribuciones continuas (exponencial, gamma, etc.) tienen rango  $[0, \infty)$ . Con lo cual, si  $X$  tiene cualquiera de estas distribuciones, puede tomar valores positivos arbitrariamente pequeños. Sin embargo, es frecuente que  $X$  represente el tiempo necesario para realizar una tarea (tal como el tiempo de servicio a un cliente), y es imposible que ésta pueda realizarse en un tiempo inferior a uno dado. Por ejemplo, en un banco, posiblemente sea imposible atender a un cliente en menos de 30 segundos. Esto debe reflejarse en las observaciones experimentales recogidas del tiempo de servicio. Así pues, en realidad  $P\{X \leq 30 \text{ segundos}\} = 0$ , sin embargo, la distribución ajustada estándar puede que tenga una probabilidad distinta de cero de que el tiempo de servicio sea inferior a 30 segundos. Como vemos, es conveniente realizar modificaciones en estas distribuciones para que se ajusten a la realidad más fielmente. Hay, al menos, dos posibles modos de realizar esta modificación.

El primer método consiste en desplazar la distribución cierta distancia hacia la derecha. Por ejemplo, la función densidad de probabilidad de la distribución  $gamma(\alpha, \beta)$ , desplazada hacia la derecha una cantidad  $\gamma > 0$ , obtenida sustituyendo  $x$  por  $(x - \gamma)$ , es:

$$f(x) = \begin{cases} \frac{(x - \gamma)^{\alpha-1} e^{-\frac{x-\gamma}{\beta}}}{\beta^{\alpha} \Gamma(\alpha)} & \text{si } x > \gamma \\ 0 & \text{en cualquier otro caso} \end{cases}$$

Si se emplea una distribución desplazada, el parámetro de posición  $\gamma$  deberá estimarse de la misma manera que los demás. Para la exponencial desplazada, es fácil estimar  $\hat{\gamma}, \hat{\beta}$  empleando el método de máxima verosimilitud, sin embargo, para la distribución gamma es preciso emplear métodos numéricos para estimar los 3 parámetros.

El segundo método consiste en truncar la distribución de alguna manera. Si  $f$  es la densidad  $gamma(\alpha, \beta)$  sin desplazar, sea:

$$a(\gamma) = \int_{\gamma}^{\infty} f(x) dx$$

que es menor que 1 si  $\gamma > 0$ . Entonces, se define:

$$f^*(x) = \begin{cases} \frac{f(x)}{a(\gamma)} & \text{si } x > \gamma \\ 0 & \text{en cualquier otro caso} \end{cases}$$

de modo que la densidad  $f^*(x)$  es discontinua en  $x = \gamma$ , saltando de 0 a  $\frac{f(\gamma)}{a(\gamma)}$ . Dependiendo de

los datos, una distribución desplazada proporcionará mejor ajuste que una truncada o viceversa. En este caso, igual que en el anterior, la estima de  $\gamma$  puede ser problemática.

Un método práctico, aunque no riguroso, de especificar  $\gamma$  es mediante la inspección visual a partir del histograma de las observaciones experimentales.