

## Capítulo 4

# Modelos de probabilidad univariantes

- 
1. Introducción
  2. El proceso de Bernoulli
  3. Variables aleatorias asociadas al proceso de Bernoulli
  4. El proceso de Poisson
  5. Variables aleatorias asociadas al proceso de Poisson
  6. Fiabilidad
  7. La variable aleatoria normal
- 

---

<sup>0</sup>Apuntes realizados por Ismael Sánchez. Universidad Carlos III de Madrid.

## 4.1. Introducción

En este tema introduciremos algunos modelos de probabilidad de variables aleatorias que serán de gran utilidad en la práctica. El objetivo es suministrar una especie de catálogo al que acudir para asignar un modelo de probabilidad para describir el comportamiento de una variable real (función de probabilidad para variables aleatorias discretas o función de densidad para las continuas). En la práctica, será muy difícil construir un modelo de probabilidad a la medida de un problema concreto, con lo que la opción más realista será la selección de uno de los modelos ya existentes en la literatura. Un modelo será de utilidad si supone una buena aproximación del comportamiento de una variable aleatoria real. En temas posteriores aprenderemos a valorar si el modelo seleccionado es adecuado para explicar nuestros datos.

En el tema anterior ya se mencionaron dos variables aleatorias con las que abrir este catálogo. La variable aleatoria **uniforme discreta** y la variable aleatoria **uniforme continua**. Usualmente se emplean indistintamente el término 'variable aleatoria' y el término 'distribución'. Por ejemplo se puede decir que el resultado de cierto experimento es una variable aleatoria uniforme o que es una variable de distribución uniforme.

## 4.2. El proceso de Bernoulli

Supongamos un experimento aleatorio cuyo resultado puede resumirse en comprobar la presencia o ausencia de cierto atributo. Por ejemplo, analizar si una pieza es defectuosa o aceptable, averiguar si una persona simpatiza con el candidato A o no, descubrir si una persona ha desarrollado cierta enfermedad o no, conseguir terminar cierto procedimiento con éxito o no conseguirlo, comprobar si una conexión telefónica se ha realizado con éxito o se ha bloqueado, etc. El experimento tendrá, por tanto, dos únicos resultados: se observa/no se observa dicho atributo. Ambos sucesos son complementarios. Muchos experimentos de interés en ingeniería pueden ser descritos en términos de observar cierto atributo o no observarlo. Este tipo de experimento aleatorio recibe el nombre de **experimento de Bernoulli**.

Llamemos  $X$  al resultado de un experimento de Bernoulli. Entonces  $X$  tendrá dos valores posibles, que denotaremos por  $x = 1$  si se observa el atributo de interés, y  $x = 0$  si no observamos el atributo de interés. Es frecuente que a estos dos estados se les denomine éxito/fracaso. Antes de realizar el experimento aleatorio no sabremos qué valor tomará  $X$ , por lo que  $X$  es una variable aleatoria. Esta variable aleatoria recibe el nombre de **variable aleatoria de Bernoulli**. Vamos a describir esta variable aleatoria de Bernoulli. Si llamamos  $p$  a la probabilidad de que al realizar un experimento de Bernoulli observemos el valor  $x = 1$ , es decir  $P(X = 1) = p$ , y llamamos  $q = 1 - p$  a la probabilidad del suceso contrario, es decir  $P(X = 0) = q$ , tendremos que la **función de probabilidad** de esta variable de Bernoulli es

$$p(x) = \begin{cases} p & \text{si } x = 1; \\ q = 1 - p & \text{si } x = 0. \end{cases}$$

Es fácil deducir que si  $X$  es una **variable de Bernoulli**

$$\begin{aligned} E(X) &= 1 \times p + 0 \times q = p \\ \text{Var}(X) &= (1 - p)^2 p + (0 - p)^2 q = p(1 - p) = pq. \end{aligned}$$

Es importante ver que una variable de Bernoulli queda totalmente definida en cuanto se conoce  $p$ . Si repitiésemos un experimento de Bernoulli tendríamos una sucesión de variables aleatorias

de Bernoulli  $X_1, X_2, \dots$  todas ellas con la misma función de probabilidad (mismo valor de  $p$ ). En estadística, a una sucesión de variables aleatorias se le denomina **proceso estocástico**, o simplemente **proceso**. A una sucesión de variables aleatorias de Bernoulli procedentes de repetir un mismo experimento de Bernoulli (es decir, con la misma  $p$ ), se le denomina **proceso de Bernoulli**. Un **proceso de Bernoulli**, debe tener las siguientes características:

1. El experimento sólo tiene **dos resultados posibles**, que pueden interpretarse como ausencia o presencia de cierto atributo de interés.
2. El resultado de cada experimento es **independiente** del resultado obtenido en repeticiones anteriores. Esta hipótesis de independencia también puede interpretarse como **ausencia de memoria**: el proceso no recuerda los resultados de los anteriores experimentos ni condicionará a los siguientes.
3. El proceso es **estable**, en el sentido de que en cada repetición  $P(X = 1) = p$ . Por tanto,  $p$  no cambia con el tiempo. Esta propiedad puede también expresarse como que las variables de Bernoulli son todas idénticas.

Cuando se posee una sucesión de variables aleatorias independientes e idénticas se les suele denotar por **iid** (*independientes e idénticamente distribuidas*).

### 4.3. Variables aleatorias asociadas al proceso de Bernoulli

La sucesión de resultados que se obtienen al repetir un experimento de Bernoulli puede analizarse desde varios puntos de vista, dando lugar a diferentes variables aleatorias. Si analizamos una sola de las variables del proceso de Bernoulli, tenemos la ya conocida variable de Bernoulli. Por el contrario, si sumamos los valores de  $n$  variables de Bernoulli tendremos una variable aleatoria llamada **binomial**. Si contamos el número de repeticiones de un experimento de Bernoulli que realizamos hasta que obtenemos una repetición en la que se observa el atributo por primera vez, tendremos la llamada **variable geométrica**. A continuación veremos estas dos distribuciones con más detalle.

#### 4.3.1. Variable aleatoria binomial

Supongamos, por ejemplo, un proceso productivo que en condiciones normales de funcionamiento produce artículos con un porcentaje de defectuosos  $p$ . Supongamos que el proceso es estable (el porcentaje de artículos defectuosos no aumenta ni disminuye a lo largo del tiempo; lo que ocurriría, por ejemplo, si la maquinaria se fuese deteriorando o el proceso productivo fuese mejorando por cambios tecnológicos). Supongamos asimismo que la aparición de un artículo defectuoso es impredecible, y por tanto independiente de la producción anterior. Si asignamos a cada artículo con una variable de Bernoulli  $X_i$  que toma el valor 1 si el artículo  $i$ -ésimo es defectuoso y 0 en caso contrario tendremos que la sucesión  $X_1, X_2, \dots, X_i, \dots$ , será, como ya se ha mencionado anteriormente, un proceso de Bernoulli.

Supongamos ahora que extraemos  $n$  artículos de la producción. Queremos saber cuántos artículos defectuosos tendremos en esos  $n$  artículos seleccionados. A priori, no sabremos cuántos artículos serán defectuosos hasta que no les hagamos las pruebas pertinentes. Aunque el proceso produzca una proporción de defectuosos  $p$ , en una muestra concreta puede haber más o menos artículos defectuosos de dicha proporción, dependiendo del azar en la selección de los  $n$  artículos. Sea  $X =$

**número de artículos defectuosos en una muestra de tamaño  $n$ .** La variable  $X$  será una variable aleatoria que puede tomar valores entre 0 y  $n$ . Esta variable  $X$  recibe el nombre de **variable aleatoria binomial** (o que sigue una distribución binomial) y se simboliza como

$$X \sim B(n, p).$$

En un contexto más general, una variable aleatoria binomial  $X \sim B(n, p)$  se define como **el número de sucesos, o éxitos, observados en  $n$  realizaciones de un experimento de Bernoulli** (estable e independiente). Se puede demostrar que la función de probabilidad de una variable binomial es:

$$p(r) = P(X = r) = \binom{n}{r} p^r (1-p)^{n-r}; r = 0, 1, \dots, n; \quad (4.1)$$

donde

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

El cálculo de la media y la varianza de una variable binomial se puede obtener aplicando la definición de estas medidas características en (4.1). Otra forma más sencilla de calcular estas medidas características es escribiendo la variable binomial como la suma de las  $n$  variables de Bernoulli  $X_1, X_2, \dots, X_n$ :

$$X = X_1 + X_2 + \dots + X_n.$$

Al valer  $X_i = 1$  ó  $X_i = 0$  en función de que en esa repetición del experimento se observe o no el atributo, la suma de los  $X_i$  nos proporcionará el número de sucesos observados en el total de  $n$  repeticiones. Como  $E(X_i) = p$ ,  $\text{Var}(X_i) = pq$ ,  $i = 1, \dots, n$  y por la independencia de estas variables de Bernoulli  $x_i$  es fácil deducir que

$$E(X) = E(X_1 + X_2 + \dots + X_n) = \sum_{i=1}^n E(X_i) = np,$$

$$\text{Var}(X) = \text{Var}(X_1 + X_2 + \dots + X_n) = \sum_{i=1}^n \text{Var}(X_i) = npq.$$

**Ejemplo 1** Se sabe por información histórica que cuando la máquina  $M$  opera en condiciones normales produce un 0.5 % de artículos defectuosos. Se toman 200 artículos producidos por dicha máquina y 5 de ellos resultan ser defectuosos. ¿Puede concluirse que la máquina funciona según lo esperado?

Si suponemos que en condiciones normales, los artículos defectuosos aparecen de forma fortuita (es decir, independientemente de los artículos anteriores) y con probabilidad estable, tendremos que la secuencia de artículos aceptables/defectuosos puede modelizarse como un proceso de Bernoulli. La población serían las infinitas piezas que, idealmente, podría fabricar la máquina. A cada artículo se le puede entonces asociar una variable de Bernoulli  $X_i$ ,  $i = 1, \dots, 200$  que tomaría los valores  $X_i = 1$  si el artículo es defectuoso y  $X_i = 0$  si es aceptable. Cuando la máquina opera correctamente, se tiene que  $P(X_i = 1) = p = 0,005$ . Por tanto, el número  $X$  de artículos defectuosos en una muestra de tamaño  $n = 200$  será

$$X \sim B(200, 0,005).$$

El número medio de artículos defectuosos según este modelo de probabilidad es  $E(X) = np = 200 \times 0,005 = 1$  artículo. Sin embargo, en la muestra analizada se han obtenido 5. ¿Es preocupante

obtener tantos artículos defectuosos cuando  $X \sim B(200, 0,005)$ ? Aplicando la fórmula (4.1) para  $r = 0, 1, \dots, 5$  se tiene que

$$p(0) = P(X = 0) = \binom{200}{0} 0,005^0 (1 - 0,005)^{200} = 0,367$$

$$p(1) = P(X = 1) = \binom{200}{1} 0,005^1 (1 - 0,005)^{199} = 0,369$$

$$p(2) = P(X = 2) = \binom{200}{2} 0,005^2 (1 - 0,005)^{198} = 0,184$$

$$p(3) = P(X = 3) = \binom{200}{3} 0,005^3 (1 - 0,005)^{197} = 0,061$$

$$p(4) = P(X = 4) = \binom{200}{4} 0,005^4 (1 - 0,005)^{196} = 0,015$$

y por tanto

$$\begin{aligned} P(X \geq 5) &= 1 - P(X \leq 4) \\ &= 1 - (0,367 + 0,369 + 0,184 + 0,061 + 0,015) \\ &= 0,004 \end{aligned}$$

De esta forma se concluye que encontrar 5 ó más artículos defectuosos en una muestra de tamaño 200 es un suceso muy improbable si de verdad tuviésemos que  $p = 0,005$ . Es un resultado preocupante. No es un suceso imposible, pues la operación anterior dice que, si  $p = 0,005$  en 4 de cada 1000 muestras habrá 5 ó más artículos defectuosos. Pero la prudencia nos debe llevar a reaccionar al observar un suceso tan poco probable. Por tanto, lo más conveniente sería dar la alarma en la línea de producción, pues es muy posible que se haya producido un desajuste en la máquina que haya provocado que  $p > 0,005$ .

A partir de un modelo de probabilidad podemos entonces atribuir probabilidades a los distintos sucesos. Esa asignación de probabilidades sigue un razonamiento puramente deductivo (¿por qué?). Si los datos finalmente observados resultan ser un suceso que, de acuerdo a dicho modelo, tiene muy poca probabilidad de haberse observado, lo más sensato será pensar que el modelo ha dejado de ser válido (razonamiento inductivo). En este ejercicio de comparar lo observado con la probabilidad de haberlo observado estamos haciendo el **ciclo deducción-inducción**, que se mencionaba en el Tema 1, y que es la base del **aprendizaje a través de la observación**. A partir del modelo Binomial  $B(200, 0,005)$  **deducimos** la probabilidad de encontrar muestras con un determinado número de artículos defectuosos. Posteriormente, observamos la realidad y la comparamos con lo que nos decía el modelo. Si discrepa mucho, **inducimos** o **inferimos** que el modelo que explica la realidad ha debido cambiar. En nuestro ejemplo, que la máquina tiene una mayor propensión a fabricar artículos defectuosos.

Claro está, que al trabajar de esta forma corremos el riesgo de equivocarnos. Obsérvese que al comparar lo observado con su probabilidad no podemos concluir que el modelo sea incorrecto, sino que lo observado es muy poco probable de acuerdo al modelo. Pudiera ser que el modelo fuese correcto y que, por casualidad (azar), hayamos observado una muestra con muchos artículos defectuosos. Según los valores del Ejemplo 1, si la máquina  $M$  funciona correctamente, en 4 de cada 1000 muestras de tamaño  $n = 200$  habrá más de 4 artículos defectuosos. Por tanto, si decidimos que el modelo ya no es válido y damos la alarma en el proceso productivo, en 4 de cada 1000

muestras estaríamos dando falsas alarmas. El riesgo de dar una falsa alarma podría, sin embargo, compensar al riesgo de no darla cuando realmente la máquina se ha desajustado. Estos riesgos serán los que en la práctica lleven a decidir cuándo dar la alarma.

### 4.3.2. Variable aleatoria geométrica

Supongamos un proceso de Bernoulli, y sean  $X_1, X_2, \dots$  la sucesión de variables de Bernoulli que forman dicho proceso. Supongamos que repetimos el experimento de Bernoulli hasta que observamos un éxito por primera vez, es decir, hasta que  $X_i = 1$  por primera vez. Sea  $Y =$  **número de repeticiones de un experimento de Bernoulli para observar el primer éxito**. Un ejemplo muy sencillo de experimento de Bernoulli es el lanzamiento de una moneda, en el que podemos asignar el resultado de cara al éxito y cruz a fracaso. Entonces  $P(X = 1) = 0,5$ . Supongamos una sucesión de lanzamientos de dicha moneda hasta que obtuviésemos cara por primera vez. Entonces tendríamos que  $Y =$  **número de lanzamientos para observar cara por primera vez** será una variable aleatoria, pues dependerá de la suerte que se tenga al repetir el experimento. Esta variable  $Y$  se denomina **variable aleatoria geométrica** (o distribución geométrica) y puede tomar los valores  $y = 1, 2, \dots$  hasta, idealmente, infinito ( $y = 1$  representa que en el primer experimento se observa el éxito, o en el primer lanzamiento sale cara). Más genéricamente,  $Y$  es el **número de repeticiones de un experimento de Bernoulli que hay que ejecutar para observar el atributo de interés por primera vez**. Se puede demostrar que la función de probabilidad de una variable geométrica es

$$P(Y = y) = q^{y-1}p; y = 1, 2, \dots$$

La variable aleatoria geométrica tiene las siguientes propiedades

$$\begin{aligned} E(Y) &= \frac{1}{p}, \\ \text{Var}(Y) &= \frac{q}{p^2}. \end{aligned}$$

**Ejemplo 2** Supongamos que lanzamos una moneda de curso legal hasta que obtenemos cara por primera vez. ¿Cuántas veces hay que lanzarla por término medio?. Como  $p = 0,5$  tenemos que

$$E(Y) = \frac{1}{0,5} = 2 \text{ lanzamientos.}$$

Si lanzamos una moneda 20 veces y aún no ha salido cara ninguna vez ¿deberíamos sospechar de su legalidad? Para responder a esta pregunta podemos calcular  $P(Y > 20 | p = 0,5)$  y resulta

$$P(Y > 20) = 1 - P(Y \leq 20) = 1 - \sum_{y=1}^{20} 0,5^{y-1}0,5 = 1 - \sum_{y=1}^{20} 0,5^y \approx 0$$

luego lo más plausible es que la moneda esté desequilibrada.

**Ejemplo 3** Una máquina produce por término medio un 1 % de artículos defectuosos. ¿Cuántos artículos aceptables se producirán por término medio hasta que se produzca un artículo defectuoso? El número de artículos producidos para obtener un artículo defectuosos por primera vez es una variable geométrica con  $p = 0,01$ . Por tanto, por término medio se producirán  $1/p = 100$  artículos (99 aceptables y el último defectuoso). El número medio de repeticiones del experimento ANTES de encontrar el primer atributo será

$$E(Y) - 1 = 1/p - 1 = q/p.$$

## 4.4. El proceso de Poisson

Un proceso de Poisson consiste en la aparición de sucesos a lo largo de sucesivas unidades de medida. Por ejemplo, llegada de clientes a la hora en cierto puesto de servicio, la aparición de defectos por área de superficie de material, la aparición de defectos por metro de cable, averías diárias de un sistema al día, llamadas de clientes a una centralita por minuto, etc. Una diferencia importante entre un proceso de Poisson y un proceso de Bernoulli es que en el caso de un proceso de Poisson, los sucesos se observan sobre un **soporte continuo** (tiempo, area, longitud,...), mientras que en un proceso de Bernoulli el soporte era discreto: cada elemento podía **tener o no** el atributo. En un proceso de Poisson se tiene, por tanto, un soporte continuo dividido en unidades de medida y en cada una de ellas pueden observarse sucesos (desde cero sucesos a, idealmente, infinitos). Es necesario, al igual que se definió para el proceso de Bernoulli, que exista estabilidad, produciéndose el mismo número medio de sucesos por unidad de medida a lo largo del tiempo. Por ejemplo, si los sucesos observados son averías por unidad de tiempo, éstas deberán producirse por fallos fortuitos y no por envejecimiento, lo que llevaría a un aumento progresivo del número medio de averías por unidad de tiempo. Si hay envejecimiento, la aparición de averías ya no sería un proceso de Poisson, pues la probabilidad de observar una avería aumentaría con el tiempo. Tendríamos que utilizar un modelo de probabilidad diferente que tuviese en cuenta ese envejecimiento.

Para que podamos hablar de un proceso de Poisson, es necesaria también la hipótesis de independencia en la aparición de los sucesos. Por ejemplo, que los clientes que acuden a un puesto de servicio lo hagan sin información sobre si el puesto de servicio está libre o colapsado, y que les anime o desanime a acudir.

Se puede decir entonces que en un proceso de Poisson hay **ausencia de memoria**: puesto que los sucesos aparecen de forma fortuita, el proceso no recuerda cuántos sucesos se han observado en la unidad de medida anterior. Vamos a estudiar dos variables aleatorias que pueden obtenerse de este esquema.

## 4.5. Variables asociadas al proceso de Poisson

### 4.5.1. La variable aleatoria de Poisson

Una **variable aleatoria de Poisson** (o distribución de Poisson) se define como  $X = \text{número de sucesos en un intervalo de longitud fija}$ . Además se debe cumplir que los sucesos ocurran de forma independiente y con media estable. Ejemplos interesantes de variables de tipo Poisson son:

- Número de defectos por unidad de superficie en un proceso donde los defectos se deben al azar.
- Número de personas que acude a un mostrador por unidad de tiempo.

Sea  $X$  una variable Poisson en la que por término medio se observan  $\lambda$  sucesos por unidad de medida. La probabilidad de que  $X = r$  (que se encuentren  $r$  defectos en un metro cuadrado de superficie, o que lleguen  $r$  personas a cierto mostrador en una unidad de tiempo) es:

$$P(X = r) = \frac{\lambda^r}{r!} e^{-\lambda}; r = 0, 1, \dots$$

La media y varianza de esta distribución son:

$$\begin{aligned} E(X) &= \lambda, \\ \text{Var}(X) &= \lambda. \end{aligned}$$

Por tanto  $\lambda$  es el número medio de sucesos que sucederán en dicho intervalo. En el ejemplo del número de defectos en una superficie,  $\lambda = 5$  significaría que las superficies tendrán, por término medio 5 defectos por unidad de superficie. En el ejemplo de las personas que acuden al mostrador, significaría que, por término medio, acuden cinco personas en una unidad de tiempo.

Una característica que tienen las variables de Poisson, es que si sumamos varias variables de Poisson independientes, el resultado es también una variable de Poisson. Es decir, sea  $X_i \sim \mathcal{P}(\lambda_i)$ ,  $i = 1, \dots, K$ , un conjunto de variables de Poisson independientes, entonces  $Y = \sum_{i=1}^K X_i \sim \mathcal{P}(\lambda^*)$ , con  $\lambda^* = \sum_{i=1}^K \lambda_i$ . El resultado es intuitivo, pues si cada variable aleatoria  $X_i$  está generando datos de forma independiente y con media constante, su suma será también una sucesión de datos independiente y de media constante. Este resultado es útil para entender que cuando hay independencia y estabilidad, las propiedades estadísticas del número de sucesos observado por unidad de medida no depende de las unidades que utilicemos. Por ejemplo, si las llamadas telefónicas que recibe una centralita son sucesos independientes (el que una persona llame es independiente de las demás personas) y aparecen con una media de 5 llamadas al minuto, la variable  $X$  = número de llamadas en un minuto será  $X \sim \mathcal{P}(\lambda = 5)$ . Y si sumamos las llamadas en 60 minutos, tendremos que la variable  $Y$  = número de llamadas en una hora, seguirá siendo una Poisson, pues las llamadas seguirán siendo sucesos independientes y con media constante. Ahora tendremos que  $Y \sim \mathcal{P}(\lambda^* = 5 \times 60)$ .

**Ejemplo 4** *Un servidor de una pequeña red recibe una media de 7 accesos al minuto. Suponiendo que los accesos a dicho servidor suceden de forma independiente y con un ritmo medio constante, se quiere calcular la probabilidad de que reciba más de 10 accesos en un minuto, que es el número de accesos a partir del cual el servidor tendría un rendimiento deficiente. Con las hipótesis de estabilidad e independencia se tiene que la variable  $X$  = **número de accesos en un minuto** será una Poisson de media 20:  $X \sim \mathcal{P}(7)$ . Entonces*

$$\begin{aligned} P(X > 10) &= 1 - P(X \leq 10) = 1 - \sum_{r=0}^{10} P(X = r) \\ &= 1 - \sum_{r=0}^{10} \frac{7^r}{r!} e^{-7} = 0,099, \end{aligned}$$

*luego más del 90 % de los minutos recibe 10 accesos o menos. Por tanto, el 10 % del tiempo, el servidor está trabajando por encima de su capacidad.*

#### 4.5.2. La variable aleatoria exponencial

En la variable de Poisson se cuenta el número de sucesos que se observa en una unidad de medida. En la variable aleatoria exponencial, por el contrario, lo que se cuenta es la distancia (en unidades de medida) entre dos sucesos consecutivos. La **variable aleatoria exponencial** (o distribución exponencial) es, por tanto,  $T$  = *intervalo entre dos sucesos consecutivos en un proceso de Poisson*. Por ejemplo, si tenemos llegada de clientes a cierto puesto de servicio, el número de clientes por hora será una variable de Poisson, pero el tiempo entre dos clientes consecutivos será



una variable exponencial. La variable exponencial es una variable aleatoria continua. Su función de densidad es

$$f(t) = \lambda e^{-\lambda t}; \lambda, t > 0;$$

donde  $\lambda$  es el parámetro del proceso de Poisson (número medio de sucesos por unidad de medida). Las medidas características son

$$\begin{aligned} E(T) &= \frac{1}{\lambda}, \\ \text{Var}(T) &= \frac{1}{\lambda^2}. \end{aligned}$$

Es fácil comprobar que la función de distribución es

$$F(t) = P(T \leq t) = 1 - e^{-\lambda t},$$

así como

$$P(T > t) = e^{-\lambda t}. \quad (4.2)$$

Estos resultados sirven para ilustrar la propiedad de independencia, o de ausencia de memoria. Por ejemplo, si en cierta maquinaria la aparición de averías fortuitas por unidad de tiempo forman un proceso estable e independiente tendremos que el número de averías por unidad de tiempo será una variable de Poisson, y el intervalo de tiempo entre dos averías consecutivas será una exponencial. Supongamos que la máquina ha estado sin averiarse  $t_1$  unidades de tiempo ¿cuál es la probabilidad de que esté  $t$  unidades de tiempo adicionales sin averiarse (es decir  $P(T > t_1 + t | T > t_1)$ )? Por la hipótesis de independencia, que implica ausencia de memoria, el proceso de Poisson que forma la secuencia de averías no recuerda si en el instante anterior ha habido avería o no, por tanto, la probabilidad de durar  $t$  unidades de tiempo adicionales dado que se ha durado  $t_1$ , es la misma que durar  $t$  unidades de tiempo empezando a contar desde el instante actual, es decir,

$$P(T > t_1 + t | T > t_1) = P(T > t).$$

Por tanto, el que haya llevado ya  $t_1$  unidades de tiempo en funcionamiento no aporta ninguna información sobre la probabilidad de durar  $t$  unidades adicionales. Esta propiedad puede probarse analíticamente:

$$\begin{aligned} P(T > t_1 + t | T > t_1) &= \frac{P[(T > t_1 + t) \cap (T > t_1)]}{P(T > t_1)} \\ &= \frac{P(T > t_1 + t)}{P(T > t_1)} = \frac{e^{-\lambda(t_1+t)}}{e^{-\lambda t_1}} = e^{-\lambda t} \equiv P(T > t). \end{aligned}$$

Esta propiedad de ausencia de memoria permite enunciar la variable exponencial de otra forma que es útil para modelizar duraciones. La **variable aleatoria exponencial** es también  $T = \text{intervalo desde la posición actual hasta el próximo suceso en un proceso de Poisson}$ .

**Ejemplo 5** Siguiendo con los datos del Ejemplo 2 utilizado anteriormente, se quiere calcular la probabilidad de que el servicio de mantenimiento no tenga que hacer ninguna reparación de ningún ascensor durante más de dos jornadas (16 horas). El suceso del que se quiere calcular la probabilidad tiene entonces que ver con la variable aleatoria  $T = \text{tiempo (horas) entre dos averías consecutivas en el conjunto de los 12 ascensores}$ , y se quiere calcular  $P(T > 16)$ . Si el número de averías en una unidad de tiempo es una variable de Poisson, entonces el tiempo entre dos

**averías consecutivas será una variable exponencial.** Sea  $X$  = número de averías por hora. Si cada ascensor tiene por término medio  $\lambda = 1/50 = 0,02$  averías por hora, el conjunto de los 12 ascensores formaran un sistema de media  $\lambda_{(12)} = 0,02 \times 12 = 0,24$  averías por hora. Entonces  $T \sim \exp(\lambda_{(12)})$ . La probabilidad de que estén más de 16 horas sin averías es

$$P(T > 16) = 1 - P(T \leq 16) = 1 - \int_0^{16} \lambda_{(12)} e^{-\lambda_{(12)} t} dt = e^{-\lambda_{(12)} \times 16} = e^{-0,24 \times 16} = 0,021$$

**Ejemplo 6** El proceso de llegadas de clientes a un puesto de servicio (por ejemplo, accesos de ordenadores a una red) se produce de manera estable e independiente, siendo el intervalo transcurrido entre la llegada de dos clientes consecutivos una variable aleatoria exponencial. Por término medio llega un cliente cada minuto.

¿Cuál es la probabilidad de que no lleguen clientes en 3 minutos? ¿Cuál es la probabilidad de que el intervalo de tiempo entre dos clientes consecutivos sea inferior a un minuto?

#### SOLUCIÓN:

Sea  $X$  = número de clientes que llegan cada tres minutos. Si  $T$  = tiempo transcurrido entre dos clientes, es una exponencial, el número de llegadas al puesto de servicio por unidad de tiempo será una Poisson de media  $\lambda = 1$  cliente/minuto. Si llegan por término medio 1 cliente por minuto, llegarán por término medio 3 clientes cada 3 minutos. Por tanto

$$X \sim \mathcal{P}(\lambda^* = 3)$$

La probabilidad que se pide es, por tanto,

$$P(X = 0) = \frac{\lambda^0}{0!} e^{-\lambda^*} = \frac{3^0}{0!} e^{-3} = 0,05$$

Otra posibilidad para calcular esta probabilidad es calcular  $P(T > 3)$ , que será, por (4.2)

$$P(T > 3) = e^{-1 \times 3} = 0,05,$$

resultando el mismo valor. Para responder a la segunda cuestión tenemos que si  $T$  = tiempo entre clientes. y según el enunciado  $T \sim \exp(\lambda)$ , con  $\lambda = 1$  clientes/minuto.

$$P(T \leq 1) = \int_0^1 \lambda e^{-\lambda t} dt = 1 - e^{-\lambda} = 1 - e^{-1} = 0,63$$

## 4.6. Fiabilidad (avanzado)

### 4.6.1. Tasa de fallos

La fiabilidad es la rama de la estadística que modeliza variables relacionadas con la duración. Por ejemplo, el tiempo que tarda un sistema en averiarse, el tiempo que vive una persona, o la duración de una espera hasta que se observa el acontecimiento de interés. La fiabilidad es una disciplina muy extensa, y aquí sólo veremos algunos de sus conceptos más fundamentales.

La distribución exponencial estudiada en la sección anterior, relacionada con el proceso de Poisson, sería un ejemplo de modelo de probabilidad para modelizar una duración. Dadas las características del proceso de Poisson, el sistema deberá ser estable y sin memoria. Por tanto la

variable exponencial es útil para modelizar la duración de un sistema hasta que falla de forma fortuita. Este tipo de modelos no es útil cuando hay envejecimiento, pues éste implicaría que a medida que pasa el tiempo la probabilidad de fallo aumenta. Tampoco es útil para sistemas que mejoran con el tiempo. Es necesario entonces otros modelos que puedan abarcar un mayor abanico de posibilidades.

Supondremos que la variable de interés es el tiempo de vida de un elemento, aunque esta definición pueda también extenderse a tiempos de espera. Definiremos **tasa de fallo** a la proporción de elementos que fallan por unidad de tiempo, o en el caso de tiempo de espera, a la proporción de individuos que dejan de esperar. Supongamos un elemento que ha sobrevivido después de  $t$  unidades de tiempo. Entonces, la probabilidad de que falle antes de  $t + \Delta t$  es

$$\begin{aligned} P(\text{Falle en el próximo } \Delta t | \text{Sobrevive a } t) &= P(T < t + \Delta t | T > t) \\ &= \frac{P[(T < t + \Delta t) \cap (T > t)]}{P(T > t)} \\ &= \frac{P(t < T < t + \Delta t)}{P(T > t)} \\ &= \frac{F(t + \Delta t) - F(t)}{1 - F(t)}. \end{aligned}$$

Por otra parte,  $P(T < t)$  es la función de distribución de  $T$ . Por tanto, podemos escribir

$$P(T < t + \Delta t | T > t) = \frac{F(t + \Delta t) - F(t)}{1 - F(t)}. \quad (4.3)$$

A  $P(T > t) = 1 - F(t)$  se le suele denominar con frecuencia la **función de supervivencia**, pues proporciona la probabilidad de sobrevivir después de  $t$  unidades de tiempo. La probabilidad (4.3) nos proporciona la probabilidad de fallar en el momento  $\Delta t$  siguiente. Si la dividimos por  $\Delta t$  nos dará la probabilidad de fallo por unidad de tiempo, es decir

$$\frac{P(T < t + \Delta t | T > t)}{\Delta t} = \frac{F(t + \Delta t) - F(t)}{(1 - F(t)) \Delta t},$$

y en el límite cuando  $\Delta t \rightarrow 0$  tendremos la tasa de fallo, también denominada tasa instantánea de fallo o **hazard rate**. A esta tasa de fallo la denotaremos como  $\lambda(t)$ , y será, aplicando que la función de densidad es la derivada de la función de distribución,

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{F(t + \Delta t) - F(t)}{(1 - F(t)) \Delta t} = \frac{f(t)}{1 - F(t)}. \quad (4.4)$$

**Ejemplo 7** Sea  $T$  la duración de un componente hasta que falla por causas fortuitas, tal que  $T \sim \exp(\lambda)$ . ¿Cuál es su tasa de fallos? En la distribución exponencial se tiene que  $f(t) = \lambda e^{-\lambda t}$ , y  $(1 - F(t)) = e^{-\lambda t}$ . Por tanto

$$\lambda(t) = \frac{f(t)}{1 - F(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda, \quad (4.5)$$

es decir, como esperábamos, su tasa de fallos es constante a lo largo de  $t$ .

La tasa de fallos tiene las unidades de número de fallos por unidad de tiempo, y se interpreta como la tasa a la que fallarán los elementos que han sobrevivido hasta  $t$ , en el instante siguiente.

Vemos en (4.4) que hay una relación entre la tasa de fallos y la función de densidad, de manera que **sabiendo la función de densidad se puede obtener la tasa de fallos**. La relación inversa puede también obtenerse, de manera que **a partir de una determinada tasa de fallos podemos derivar la función de densidad que le corresponda**. Para ello denominaremos **Tasa acumulada de fallos** a

$$H(t) = \int_0^t \lambda(x) dx. \quad (4.6)$$

Integrando en (4.4) se obtiene que

$$H(t) = \int_0^t \frac{f(x)}{1 - F(x)} dx = -\ln [1 - F(x)]_0^t = -\ln [1 - F(t)],$$

y por tanto

$$F(t) = 1 - e^{-H(t)},$$

y derivando tenemos que para una tasa de fallos  $\lambda(t)$  dada, la función de densidad será

$$f(t) = \lambda(t)e^{-H(t)} \quad (4.7)$$

**Ejemplo 8** (continuación) *Sabiendo que la distribución exponencial tiene una tasa de fallos constante, podemos deducir su función de densidad. Si queremos deducir la función de densidad de una variable aleatoria que describa la duración de un sistema con tasa de fallos constante tendremos que*

$$\begin{aligned} \lambda(t) &= \lambda, \\ H(t) &= \int_0^t \lambda dx = \lambda x \Big|_0^t = \lambda t, \end{aligned}$$

y por tanto

$$f(t) = \lambda(t)e^{-H(t)} = \lambda e^{-\lambda t}.$$

#### 4.6.2. La distribución Weibull

Como ya hemos dicho anteriormente, la distribución exponencial no es un modelo útil para modelizar la duración, o el tiempo de espera, de sistemas que envejecen o rejuvenecen. Sería más realista plantear sistemas en los que la tasa de fallos sea una función del tipo

$$\lambda(t) = kt^{c-1}, \quad (4.8)$$

donde  $k$  es una constante de proporcionalidad y  $c$  es una constante que serviría para modelizar la evolución de la tasa de fallo con el tiempo. Si  $c = 1$  tendríamos el caso de tasa de fallo constante, que es el caso de la exponencial. Si  $c > 1$  tendríamos el caso de tasa de fallos que aumenta con el tiempo, lo que serviría para modelizar sistemas que envejecen. Si  $c < 1$  tendríamos sistemas que mejoran con el tiempo. Utilizando (4.6) y (4.7) tenemos que una variable aleatoria con tasa de fallo (4.8) tendrá como función de densidad

$$f(t) = kt^{c-1} \exp\left(-\frac{k}{c}t^c\right) \quad (4.9)$$

A la distribución con función de densidad (4.9) se la conoce como **distribución de Weibull**. En la mayoría de los textos y aplicaciones informáticas, la distribución de Weibull se escribe de

forma ligeramente diferente a (4.9). La forma más frecuente para definir la distribución Weibull (equivalente a la anterior) es con la siguiente notación

$$f(t) = \frac{\beta}{\alpha^\beta} t^{\beta-1} \exp \left\{ - \left( \frac{t}{\alpha} \right)^\beta \right\}, \quad (4.10)$$

$$F(t) = 1 - \exp \left\{ - \left( \frac{t}{\alpha} \right)^\beta \right\} \quad (4.11)$$

donde  $\beta$  recibe el nombre de parámetro de forma (shape), y  $\alpha$  recibe el nombre de parámetro de escala (scale). La tasa de fallo con la notación (4.10) es

$$\lambda(t) = \frac{\beta}{\alpha^\beta} t^{\beta-1} = \frac{\beta}{\alpha} \left( \frac{t}{\alpha} \right)^{\beta-1}.$$

Cuando  $\beta = 1$ , la distribución Weibull se convierte en la distribución exponencial con  $\lambda = 1/\alpha$ . Con  $\beta > 1$ , el sistema tiene una tasa de fallos creciente con  $t$ . La media de una distribución de Weibull es

$$\mu = \alpha \Gamma \left( 1 + \frac{1}{\beta} \right), \quad (4.12)$$

donde  $\Gamma(\cdot)$  es la función gamma, que tiene las siguientes propiedades

$$\begin{aligned} \Gamma(r) &= \int_0^\infty x^{r-1} e^{-x} dx \\ \Gamma(r) &= (r-1)\Gamma(r-1) \\ \Gamma(1/2) &= \pi^{1/2}. \end{aligned} \quad (4.13)$$

Para el caso de  $r$  un número entero la función gamma es muy sencilla de calcular y es igual a

$$\Gamma(r) = (r-1)!$$

**Ejemplo 9** Se tiene un cluster formado por un conjunto de ordenadores, y sea  $T$ =tiempo hasta que se detecta un fallo en el cluster (un ordenador necesita por alguna razón ser reiniciado). El tiempo entre reinicios (Time To Reboot, o TTB) es una medida de la fiabilidad de un sistema de ordenadores. Se considera que la variable aleatoria  $T$  se distribuye como una Weibull de parámetros:  $\alpha=99$  horas,  $\beta=1/3$ . ¿El sistema aumenta o disminuye su fiabilidad con  $T$ ?

Como  $\beta < 1$  se tiene que el sistema aumenta su fiabilidad a medida que pasa más tiempo entre reinicios. La tasa de fallos es

$$\lambda(t) = \frac{\beta}{\alpha} \left( \frac{t}{\alpha} \right)^{\beta-1} = 0,072 t^{-0,67},$$

que disminuye al aumentar  $t$ . La razón podría ser el que al reiniciar, el sistema no sea estable y tarde tiempo en alcanzar estabilidad. Durante ese periodo de inestabilidad podría bloquearse y necesitar reiniciarse de nuevo.

¿Cual es la media del tiempo entre fallos?

La media es, de acuerdo a (4.12)

$$\mu \approx 99\Gamma(1 + 3) = 99 \times 3! = 594 \text{ horas}$$

¿Cuál es la probabilidad de que esté más de 100 horas sin fallar?

Usando (4.11), tenemos que

$$\begin{aligned} P(T > 100) &= 1 - F(100) \\ &= 1 - \left( 1 - \exp \left\{ - \left( \frac{100}{\alpha} \right)^\beta \right\} \right) \\ &= \exp \left\{ - \left( \frac{100}{\alpha} \right)^\beta \right\} = \exp \left\{ - \left( \frac{100}{99} \right)^{0,33} \right\} = 0,37 \end{aligned}$$

¿Cuál es la probabilidad de, habiendo estado 100 horas sin fallar, esté al menos otras 100?

Queremos calcular  $P(T > 200 | T > 100)$ . Por la regla de la probabilidad condicionada tenemos que

$$\begin{aligned} P(T > 200 | T > 100) &= \frac{P[(T > 200) \cap (T > 100)]}{P(T > 100)} = \frac{P(T > 200)}{P(T > 100)} \\ &= \frac{\exp \left\{ - \left( \frac{200}{\alpha} \right)^\beta \right\}}{\exp \left\{ - \left( \frac{100}{\alpha} \right)^\beta \right\}} = \exp \left\{ \left( \frac{100}{\alpha} \right)^\beta - \left( \frac{200}{\alpha} \right)^\beta \right\} = 0,77 \end{aligned}$$

Por tanto, le cuesta menos sobrevivir a estas segundas 100 horas que a las primeras. ¿Cuánto sería esta segunda probabilidad en el caso en que  $T$  siguiese una distribución exponencial?

### 4.6.3. La distribución Gamma

Una variable aleatoria  $X$  sigue una **distribución Gamma**, si su función de densidad es

$$f(x) = \frac{\lambda^\alpha x^{\alpha-1}}{\Gamma(\alpha)} e^{-\lambda x}, \quad x > 0, \quad (4.14)$$

con  $\lambda > 0$  y  $\alpha > 0$  parámetros del modelo. A esta distribución la denotaremos por  $X \sim \text{Gamma}(\alpha, \lambda)$ . Las medidas características de la distribución Gamma son

$$\begin{aligned} \mu &= \frac{\alpha}{\lambda}, \\ \sigma^2 &= \frac{\alpha}{\lambda^2}. \end{aligned}$$

Lo más habitual en la práctica es utilizar la distribución Gamma en (4.14) para valores  $\alpha$  enteros. En ese caso, la distribución recibe también el nombre de **distribución Erlang**. La distribución Erlang está relacionada con el proceso de Poisson y la distribución exponencial. Veíamos anteriormente que la distancia entre dos sucesos consecutivos de un proceso de Poisson sigue una distribución exponencial. Puede demostrarse entonces, que la distancia entre  $\alpha$  sucesos consecutivos de un proceso de Poisson sigue una distribución Gamma (Erlang) de parámetro  $\alpha$ . Es fácil ver en (4.14) que para  $\alpha = 1$  se obtiene la distribución exponencial. La distribución Erlang puede entonces interpretarse como la suma de  $\alpha$  distribuciones exponenciales de parámetro  $\lambda$ . la integración de la distribución Gamma es en general complicada y requiere de integración por partes. Es frecuente llamar a  $\alpha$  parámetro de escala (scale) y  $\lambda$  el parámetro de forma (shape).

**Ejemplo 10** Un sistema para regular la cola de impresión de una impresora colectiva sólo admite un máximo de 4 ficheros en la cola de impresión. Por tanto, si llegan 5 o más ficheros, la cola o bien se bloquea o se pierden los ficheros que pasen del cuarto. Si la llegadas de ficheros a esa impresora sigue un proceso de Poisson de media 1 fichero por minuto, ¿Cuál es la probabilidad de que en un minuto lleguen más de 4 ficheros?

Si llamamos  $X$  = número de ficheros que llegan a la cola de la impresora en un minuto, tenemos que  $X \sim \mathcal{P}(1)$ , y entonces

$$\begin{aligned} P(X > 4) &= 1 - P(X \leq 4) = 1 - [P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)] \\ &= 1 - \sum_{r=0}^4 \frac{1^r}{r!} e^{-1} = 0,0037. \end{aligned}$$

Este problema también lo podemos resolver llamando a  $Y$  = intervalo entre 5 llegadas consecutivas, por lo que  $Y \sim \text{Gamma}(5, 1)$ . Entonces la probabilidad de que lleguen más de cuatro ficheros en una hora será la misma que la probabilidad de que el intervalo entre el primer fichero y el quinto sea inferior a una hora, es decir

$$P(Y < 1) = \int_0^1 \frac{1^5 x^{5-1}}{\Gamma(5)} e^{-x} dx = \frac{1}{24} \int_0^1 x^4 e^{-x} dx = 0,0037$$

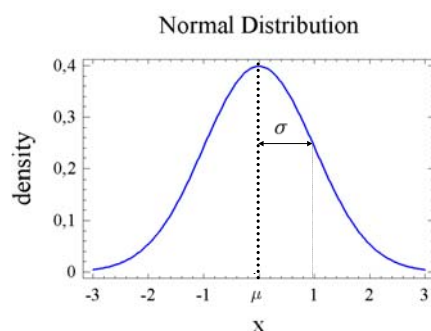
## 4.7. La variable aleatoria normal

### 4.7.1. La distribución normal

Esta distribución es uno de los modelos más frecuentes para describir variables reales continuas. La distribución normal es una distribución continua, simétrica y su rango incluye a todos los números reales. Su función de densidad es

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), -\infty < x < \infty. \quad (4.15)$$

donde los parámetros  $\mu$  y  $\sigma$  son la media y desviación típica respectivamente. La variable aleatoria normal se representa como  $X \sim N(\mu, \sigma^2)$ , aunque algunos textos también usan la notación  $X \sim N(\mu, \sigma)$ . La figura siguiente muestra la función de densidad de una distribución normal de media  $\mu$  y desviación típica  $\sigma$ .



Como puede verse en esta figura, la función de densidad de la normal es simétrica alrededor de la media y tiene forma de campana. **El punto de inflexión de su curvatura está a distancia  $\sigma$  del eje de simetría.** Entre la media y  $\pm 3\sigma$  está la mayoría de la probabilidad (el área bajo la curva entre  $\mu \pm 3\sigma$  es 0.997). El coeficiente de asimetría de la normal es 0, y el coeficiente de curtosis es 3. Puede demostrarse que **cualquier transformación lineal de una variable normal da como resultado otra variable normal.** Este resultado no se cumple, en general, con otras distribuciones. De esta forma, cualquier variable aleatoria normal de media  $\mu$  y varianza  $\sigma^2$  puede transformarse en una  $N(0,1)$  restándole su media y dividiendo por su desviación típica. A esta transformación se le llama **tipificación** o **estandarización**. A la normal tipificada se le denomina **normal estándar** y se denota con la letra  $Z$ . Por tanto, si  $X \sim N(\mu, \sigma^2)$

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1).$$

Los valores de la función de probabilidad de la normal estándar están tabulados. De esta forma, los problemas de probabilidad de una normal cualquiera  $x \sim N(\mu, \sigma^2)$  suelen resolverse sobre la normal estandarizada. A la función de distribución de la normal estándar se le denota con la letra  $\Phi$ ; es decir  $P(x \leq x_0) = F(x_0) \equiv \Phi(x_0)$ . Por ejemplo

$$\begin{aligned} P(a < X < b) &= P\left(\frac{a - \mu}{\sigma} < \frac{X - \mu}{\sigma} < \frac{b - \mu}{\sigma}\right) = P\left(\frac{a - \mu}{\sigma} < Z < \frac{b - \mu}{\sigma}\right) \\ &= P\left(Z < \frac{b - \mu}{\sigma}\right) - P\left(Z < \frac{a - \mu}{\sigma}\right) \\ &= \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right), \end{aligned}$$

y estos valores de  $\Phi$  pueden encontrarse en las tablas de la normal  $N(0, 1)$ . Otra propiedad interesante que cumple la distribución normal es que **la combinación lineal de variables normales es también normal.**

**Ejemplo 11** La longitud  $L$  en milímetros, de las piezas fabricadas en un proceso es una variable aleatoria que se distribuye según una  $N(32, 0,3^2)$ , considerándose aceptables aquellas cuya medida se encuentra dentro del intervalo  $(31,1, 32,6)$ . Calcular la probabilidad de que una pieza elegida al azar sea aceptable.

La probabilidad de que una pieza elegida al azar sea aceptable es

$$P(31,1 < L < 32,6).$$

Llamando  $Z$  a la variable aleatoria  $(L - 32)/0,3$ , se cumple que  $Z \sim N(0, 1)$ , de manera que

$$P(31,1 < L < 32,6) = P\left(\frac{31,1 - 32}{0,3} < Z < \frac{32,6 - 32}{0,3}\right) = 0,976.$$

**Ejemplo 12** Sean  $X_1 \sim N(\mu_1, \sigma_1^2)$  y  $X_2 \sim N(\mu_2, \sigma_2^2)$ , dos variables aleatorias normales independientes la una de la otra. Entonces  $Y = aX_1 + bX_2$  es también una distribución normal. Además se tendrá que

$$\mu_y \equiv E(Y) = E(aX_1 + bX_2) = a\mu_1 + b\mu_2.$$



Por ser ambas variables aleatorias independientes se tiene también que

$$\sigma_y^2 \equiv \text{Var}(Y) = \text{Var}(aX_1 + bX_2) = a^2\sigma_1^2 + b^2\sigma_2^2,$$

y entonces

$$Y \sim N(\mu_y, \sigma_y^2).$$

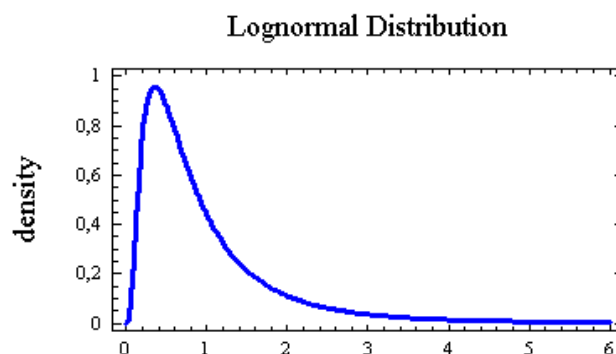
Una variable relacionada con la distribución normal es la llamada distribución **lognormal**. Se denomina lognormal a la variable aleatoria cuyo logaritmo neperiano es normal. Es decir, la variable aleatoria  $Y$  es lognormal si  $X = \ln Y$  es normal. la función de densidad de  $Y$  será, por tanto

$$f(y) = \frac{1}{y\sqrt{2\pi\sigma^2}} \exp \left[ -\frac{1}{2} \left( \frac{\ln y - \mu}{\sigma} \right)^2 \right], y > 0$$

Se verifica que

$$\begin{aligned} E(Y) &= e^{\mu + \frac{1}{2}\sigma^2} \\ \text{Var}(Y) &= e^{2\mu + \sigma^2} (e^{\sigma^2} - 1) \end{aligned}$$

La figura siguiente muestra la función de densidad de una distribución lognormal de parámetros  $\mu = \sigma^2 = 1$ .



En esta figura puede apreciarse que la distribución lognormal es asimétrica positiva.

#### 4.7.2. El Teorema Central del Límite

El teorema central del límite es una ley estadística básica para entender por qué la distribución normal es tan frecuente para explicar datos reales. Este teorema dice que, en condiciones bastantes generales, *la suma de un conjunto de variables aleatorias se aproxima, a medida que el número de variables aumenta, a una variable aleatoria normal, independientemente de cuál sea la distribución de las variables que se promedian, siempre y cuando ninguna de ellas predomine sobre las demás.* Es decir, sean  $X_1, X_2, \dots, X_n$  un conjunto de variables aleatorias cualesquiera tal que

$$\begin{aligned} E(X_i) &= \mu_i < \infty \\ \text{Var}(X_i) &= \sigma_i^2 < \infty \end{aligned}$$

Entonces la variable  $Y = \sum_{i=1}^n X_i$  tendrá una distribución en general desconocida. Sin embargo, si  $n \rightarrow \infty$ , verificará que

$$\begin{aligned} Y &\sim N[E(Y), \text{Var}(Y)], \\ Z &= \frac{Y - E(Y)}{\sqrt{\text{Var}(Y)}} \sim N(0, 1), \end{aligned}$$

donde

$$E(Y) = \sum_{i=1}^n \mu_i.$$

Si las variables  $X_i$  son independientes se tiene además que

$$\text{Var}(Y) = \sum_{i=1}^n \sigma_i^2,$$

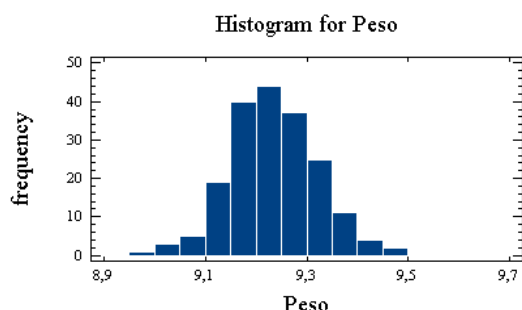
y por tanto, en el caso de suma de variables independientes, si  $n \rightarrow \infty$

$$Y \sim N \left[ \left( \sum_{i=1}^n \mu_i \right), \left( \sum_{i=1}^n \sigma_i^2 \right) \right].$$

Nótese que si las  $X_i$  son normales entonces  $Y$  será normal para cualquier valor de  $n$ , pues la combinación lineal de normales es siempre una distribución normal. El teorema central del límite sirve para dar una explicación al hecho constatado de que muchas variables observadas en la naturaleza o en experimentos físicos sean aproximadamente normales. Por ejemplo, las medidas del cuerpo humano tales como la estatura, peso, longitud de los huesos, siguen aproximadamente una distribución normal dentro de una misma raza y sexo. Por ejemplo, en la estatura de un individuo afectan muchos factores, cada uno ejerciendo un efecto pequeño: estatura de los padres, de los abuelos, etc, así como calidad de la alimentación, tipo de vida, etc, de manera que el efecto total será la suma de esos efectos individuales. Por esa razón, un histograma de estaturas de individuos de un mismo sexo es aproximadamente normal.

Este mismo planteamiento puede aplicarse a la variabilidad de un proceso industrial. Si sobre un proceso ejercen muchos factores y cada uno de ellos tiene un peso pequeño, el resultado final seguirá una distribución parecida a la normal. Por esa razón, la distribución de las magnitudes de un proceso industrial: longitudes, pesos, diámetros, etc, son con mucha frecuencia normales. La siguiente figura ilustra este resultado. Esta figura representa el peso de 200 monedas de 100 pesetas (las últimas monedas circulantes, conseguidas a través de un profesional de la banca poco después de la retirada de la peseta). El peso de una moneda será una variable aleatoria que dependerá de muchos factores, tanto en el momento de su elaboración (homogeneidad del material, precisión del molde, temperatura de fundición, etc) como del desgaste a que fue sometida durante su circulación (unas se habrán desgastado más que otras). Por esa razón, no resulta extraño que su histograma

se asemeje mucho al de una distribución normal.



Un resultado similar puede enunciarse con la distribución lognormal, pero en lugar de sumas de variables aleatorias utilizamos productos. *El producto de un conjunto de variables aleatorias se aproxima, a medida que el número de variables aumenta, a una variable aleatoria lognormal, independientemente de cuál sea la distribución de las variables que se promedian siempre y cuando ninguna de ellas predomine sobre las demás.* Es decir, sean  $X_1, X_2, \dots, X_n$  un conjunto de variables aleatorias cualesquiera tal que

$$\begin{aligned} E(X_i) &= \mu_i < \infty \\ \text{Var}(X_i) &= \sigma_i^2 < \infty \end{aligned}$$

Entonces, la variable  $Y = \prod_{i=1}^n X_i$ , tendrá una distribución en general desconocida. Sin embargo, si  $n \rightarrow \infty$ , verificará que

$$\begin{aligned} Y &\sim \text{Lognormal} \\ Z &= \frac{\ln Y - E(\ln Y)}{\sqrt{\text{Var}(\ln Y)}} \sim N(0, 1), \end{aligned}$$

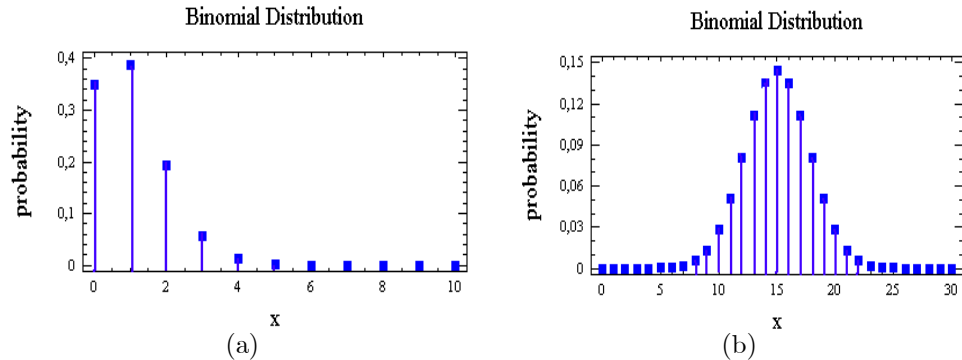
### 4.7.3. Relación entre la normal y otras variables aleatorias

#### Relación entre la binomial y la normal

Supongamos un proceso que produce una proporción de artículos defectuosos  $p$ . Supongamos que tenemos una muestra de tamaño uno (un solo artículo). Llamemos  $X$  al número de artículos defectuosos en dicha muestra. Esa variable así definida será una variable aleatoria binomial  $B(1, p)$  (también llamada de Bernoulli) de media  $E(X) = p$  y varianza  $\text{Var}(X) = p(1 - p)$ . Claramente la variable  $X$  sólo tomará dos valores posibles  $X = 0$  (con probabilidad  $(1 - p)$ ) y  $X = 1$  (con probabilidad  $p$ ). Supongamos ahora que en lugar de un artículo tenemos  $n$  artículos. Como se vio antes, el número de artículos defectuosos seguirá una distribución binomial  $B(n, p)$ . Llamemos  $X_1$  a la variable Bernoulli correspondiente al primer artículo,  $X_2$  a la del segundo, y así hasta  $X_n$  a la correspondiente al último artículo. Tendremos entonces  $n$  variables aleatorias  $B(1, p)$ . Esas variables tomarán los valores uno o cero dependiendo de si es defectuoso (uno: un defecto) o no lo es (cero: cero defectos). Por tanto el número de artículos defectuosos  $Y$  en esa muestra de tamaño  $n$  puede verse como la suma de las  $n$  variables  $B(1, p)$ :  $Y = X_1 + X_2 + \dots + X_n$ . De esta forma, si  $n$  es elevado, el teorema central del límite nos permite utilizar la aproximación

$$Y = \text{número de defectuosos} \sim B(n, p) \approx N(np, np(1 - p)).$$

En la práctica, esta aproximación es buena si  $np(1-p) > 5$ . En la figura siguiente se muestra la función de probabilidad de dos binomiales. La Figura (a) muestra la función de probabilidad de la binomial  $B(10, 0,1)$ , donde puede verse que, al ser  $np(1-p) = 0,9 < 5$  dicha función no tiene mucho parecido con la normal. La Figura (b) representa a la binomial  $B(30, 0,5)$  donde  $np(1-p) = 7,5 > 5$  y, como puede apreciarse, guarda mucho parecido con la normal. Este resultado es útil para calcular probabilidades, pues nos dice que obtendremos resultados muy parecidos si usamos la aproximación a la normal.



**Ejemplo 13** Se tiene un lote de 5000 artículos fabricados por la misma máquina  $M$ . Se sabe que en condiciones habituales de funcionamiento, la máquina produce un 1 % de artículos defectuosos. ¿Cuál es la probabilidad de que en ese lote haya más de 50 artículos defectuosos? El número de artículos defectuosos en ese lote de tamaño 5000 será  $X \sim B(5000, 0.01)$ . Como  $np(1-p) = 5000 \times 0.01 \times 0.99 = 49.5$  tendremos que  $X \sim N(50, 49.5)$ , y por tanto, por la simetría de la normal es fácil deducir que,  $P(X > 50) = 0.5$

#### Relación entre la Poisson y la normal

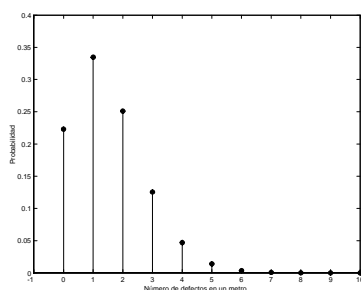
Supongamos ahora que tenemos cable que procede de un proceso productivo que produce, por término medio,  $\lambda$  defectos por metro de longitud del cable. Ya vimos anteriormente que el número de sucesos en un intervalo fijo, si los sucesos aparecen con media estable y de forma independiente, sigue una distribución de Poisson. Por tanto la variable: *número de defectos en un metro de cable* seguirá una Poisson de parámetro  $\lambda$ . Vamos a ver que esta distribución puede aproximarse, bajo ciertas condiciones, a la distribución normal. Para ello aplicaremos el teorema central del límite a una suma de procesos de Poisson.

Supongamos que estamos interesados en medir el número de defectos por metro de longitud, pero la inspección se realiza con un aparato que detecta defectos por centímetro. Sea  $X_1$  el número de defectos en el primer centímetro,  $X_2$  en el segundo, así hasta  $X_{100}$  en el centímetro 100. Cada una de estas variables  $X_i$  (*número de fallos en un centímetro*) seguirá también una variable aleatoria de tipo Poisson. El número de fallos en un metro puede verse entonces como la suma de los fallos en cada centímetro:  $Y = X_1 + X_2 + \dots + X_{100}$ . Por tanto una variable Poisson que representa el número de sucesos en un intervalo puede verse como la suma de variables Poisson que miden el número de sucesos en intervalos más pequeños. El teorema central del límite aplicado a variables tipo Poisson dice que cuanto mayor sea el número de variables Poisson que sumemos, la variable  $y$  resultante (que será también Poisson) se parecerá cada vez más a una normal. En la práctica, para que podamos considerar que una Poisson es la suma de variables Poisson evaluadas en intervalos más pequeños, el número medio de defectos  $\lambda$  ha de ser suficientemente grande, para

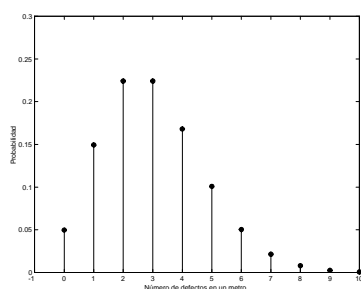
que al reducir el intervalo de observación sigamos detectando defectos. Un valor de  $\lambda$  elevado se puede conseguir eligiendo una unidad de medida suficientemente grande. Por ejemplo, puede que el número de defectos por metro sea pequeño, pero tal vez el número de defectos por cada diez metros sea una cifra suficientemente grande. Se considera que si  $\lambda > 5$  la aproximación a la normal es buena. En ese caso, se puede considerar que, aproximadamente

$$y \sim N(\lambda, \lambda).$$

Por ejemplo, supongamos que el proceso anterior produce cable con un número de defectos medio de 1.5 por metro lineal. Con este dato podemos construirnos variables Poisson utilizando distinta longitud de intervalo. Si llamamos  $Y_1$  al número de defectos por metro tenemos que  $\lambda_1 = 1,5$  defectos/metro. La probabilidad de obtener 0,1,2,... defectos en un metro puede verse en el siguiente gráfico



Si llamamos  $Y_2$  al número de defectos por cada dos metros tendremos que  $Y_2 = Y_1 + Y_1$  y será una variable Poisson de parámetro  $\lambda_2 = 3$  def/dos metros. La probabilidad de obtener 0,1,... defectos en dos metros de cable puede verse en el siguiente gráfico, que se parece más a la normal que el anterior.



Si llamamos  $Y_{10}$  al número de defectos por cada diez metros tendremos que  $Y_2 = 10Y_1$  y será una variable Poisson de parámetro  $\lambda_2 = 15$  def/diez metros. La probabilidad de obtener 0,1,2, ... defectos en diez metros de cable puede verse en el siguiente gráfico. Como puede verse, eligiendo la unidad de medida adecuadamente podemos obtener una variable que tenga una distribución aproximadamente normal.

