Estadística climatológica con R

4. Distribuciones: aplicación en climatología.
4.1 Distribución normal. Ejemplos climatológicos
4.2 Ajustes de datos a las principales funciones de distribución climatológicas.

4. Distribuciones de probabilidad

Normalmente, un investigador trata de analizar unas características en los individuos de una población. Este análisis puede ser *unidimensional* si solo se observa una sola variable o característica o *multidimensional*, si se observan varias variables unidimensionales a la vez y se trata de ver su relación, como por ejemplo estudiar la temperatura en dos estaciones próximas.

Más concretamente, lo que hará el investigador es estudiar algún parámetro relacionado con esa variable, como su temperatura media mensual o temperatura máxima anual en el ejemplo, tratando, bien de inferir un valor para dicho *parámetro poblacional* o bien de construir *un intervalo de confianza* para él, o bien tratando de decidir, mediante un *contraste de hipótesis*, entre dos conjuntos de valores del parámetro.

Todas estas técnicas dependen de ***la distribución de probabilidad*** o ***modelo probabilístico*** supuesto como ley que rige el *fenómeno aleatorio* en estudio. Vamos a ver una serie de modelos probabilísticos, viendo sus propiedades y cuáles son los fenómenos aleatorios típicos a los que asociar estos modelos.

Definamos un poco teóricamente el concepto de ***distribución de probabilidad***.

Una ***variable aleatoria X*** es una función $\boldsymbol{X}: \boldsymbol{\Omega} \rightarrow \boldsymbol{R}$ que a cada elemento del espacio muestral (correspondiente al experimento aleatorio que estemos estudiando) la hace corresponder un número real.

La idea es que *X* representa *la característica* que nos interesa estudiar. Lo que más nos interesa es conocer la probabilidad de los diferentes sucesos correspondientes a una variable aleatoria, es decir su *modelo* o *función de probabilidad*.

Definimos ***la distribución de probabilidad P*** como

$$P(A) = P(X \in A) = P\{\omega \in \Omega : X(\omega) \in A\}$$

ω es un individuo de la población, aunque realmente estamos interesados en una característica $X(\omega)$ como por ejemplo el peso de una población de individuos. Ω es el *espacio muestral* y *A* es el *conjunto de los sucesos*.

Si por ejemplo en una población total de 50 millones de individuos, hay 20 millones con un peso entre 60 y 75 kg, la transformación *X* debe de ser tal que

$$P_X\{[60\,,\,75]\} = P\{\omega \in \Omega : 60 \leq X(\omega) \leq 75\} = \frac{2}{5}.$$

o lo que es lo mismo la probabilidad de que *X* este entre 60 y 75 es de 2/5 (20/50).

La función *X* es la variable aleatoria y P_X su distribución de probabilidad.

Si consideramos el experimento aleatorio del lanzamiento de un dado. El espacio muestral Ω es el compuesto por las 6 posibilidades que se pueden obtener: sacar un 1, sacar un 2, etc. y la probabilidad es igual a 1/6 para cada uno de ellos.

La variable aleatoria a considerar es X = numero de puntos de la cara superior del dado. La distribución de probabilidad de X es, para x = 1, 2,…, 6

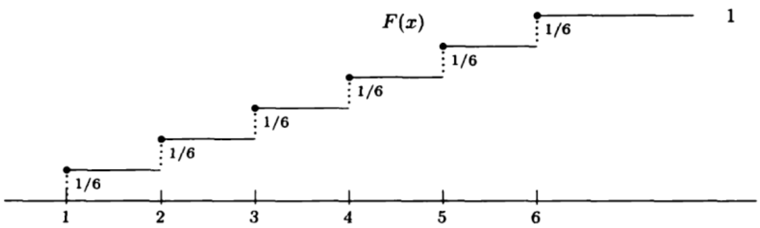
$$P_X(\{x\}) = P\{\omega : X(\omega) = x\} = \frac{1}{6}$$

Asociada a toda variable aleatoria existe una función *F(x)*, denominada ***función de distribución*** de X , la cual va midiendo la probabilidad acumulada por X hasta el punto x. Es decir

$$F(x) = P\{\omega \in \Omega : X(\omega) \leq x\}.$$

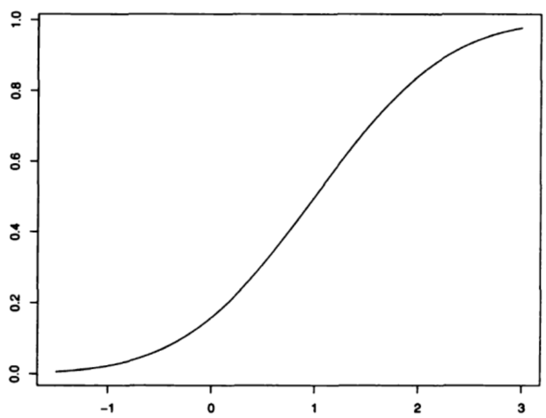
Esta función tiene la propiedad de caracterizar la distribución de probabilidad de X. Es decir a partir de una de ellas se obtiene la otra, siendo más cómodo de trabajar con la función de distribución.

En el caso del ejemplo del dado la función de distribución seria:



Como vimos, si la variable aleatoria toma valores discretos como en el ejemplo del dado se denomina *discreta*. Si por el contrario puede tomar cualquier valor de un intervalo como el peso o la talla recibe el nombre de *continua*. Estos calificativos se aplican también a su distribución.

La función de distribución de una variable aleatoria continua es una función continua como por ejemplo:



Las funciones de distribución son:

- No decrecientes
- Los limites por la izquierda y derecha son 0 y 1.

Las variables aleatorias discretas tienen asociadas una función, denominada ***función de masa***, $p_X(x)$, la cual da la probabilidad de los valores de dicha variable aleatoria:

$$p_X(x) = P_X(\{x\}) = P\{\omega : X(\omega) = x\}.$$

La relación entre la función de distribución y la función de masa es

$$F(x) = \sum_{y \leq x} p_X(y)$$

De manera análoga, las variables aleatorias continuas X tienen asociadas una función, denominada ***función de densidad***, $f_X(x)$, la cual indica la velocidad a la que crece la función de distribución, siendo

$$f_X(x) = \frac{d}{dx} F(x)$$

o inversamente

$$F(x) = \int_{-\infty}^{x} f_X(y)\, dy.$$

Así pues la distribución de una variable aleatoria se puede caracterizar por su distribución de probabilidad, por su función de distribución o por su función de masa o densidad (según sea discreta o continua).

Asociamos a una variable en estudio una determinada distribución de probabilidad, o mejor, un determinado ***modelo probabilístico***. Esto habrá que suponerlo con objeto de hacer inferencias sobre X, y si nuestros datos presentan, por ejemplo en el caso continuo, un histograma tal que cuando las bases de los rectángulos que la forman tienden a cero a medida que la frecuencia total aumenta, la curva resultante se ajusta bien al modelo propuesto (tiene la misma forma), las inferencias que hagamos serán aceptables. En caso contrario debemos cambiar el modelo.

Como los modelos probabilísticos representan básicamente un ideal de las distribuciones de frecuencias estudiadas en capítulo 2, tendrán también medidas de tendencia central, de dispersión, etc., definidas de un modo similar.

Veremos las principales distribuciones en relación a la climatología.

4.1 Distribución normal.

Cuando la curva de distribución de frecuencias o probabilidades, o bien, cuando la función de densidad de probabilidad, es continua y simétrica respecto al valor o probabilidad máxima recibe el nombre de ***Curva de Gauss*** o ***distribución normal***. La importancia que tuvo al principio esta distribución como modelo probabilístico se debía, en gran medida, a que los matemáticos pensaban que este modelo era el habitual o normal para explicar los fenómenos naturales. Muchos experimentos se distribuyen aproximadamente con esta forma, como por ejemplo algunas series de temperatura.

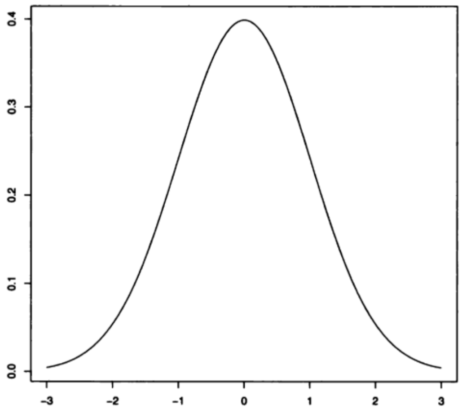
El modelo de distribución normal se consideraba en el siglo XIX el modelo habitual de la mayoría de los fenómenos de la naturaleza. Esta idea tiene un fundamento teórico en el *teorema central del límite*, que nos dice que si tenemos una sucesión de N variables aleatorias independientes $X_1, X_2, \ldots$, idénticamente distribuidas y con varianza común σ^2 finita, se puede demostrar que con un numero N suficientemente grande, la suma de todas estas variables aleatorias se comporta aproximadamente con la *distribución normal*.

Es decir que si tenemos una gran cantidad de variables aleatorias que se refieren al mismo aspecto (temperatura, precipitación, humedad relativa, presión, etc.), con la misma media y varianza, su suma es una variable aproximadamente normal.

Hoy en día se considera una modelo probabilístico mas aunque gran parte de la estadística está desarrollada suponiendo esta distribución como modelo.

Es una función continua que teóricamente toma valores desde menos infinito hasta más infinito.

La distribución normal para el caso de media $\mu=0$ y desviación típica $\sigma = 1$, se la denomina ***normal standard***, ***N(0,1)*** y tiene la forma



La distribución Normal se define como aquella distribución cuya función de densidad es

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

Donde $\boldsymbol{\mu}$ es la media y $\boldsymbol{\sigma}$ la desviación típica. Como la distribución Normal depende de estos dos parámetros se suele expresar de la forma ***N(μ,σ)***. El caso de la *normal estándar* ***N(0,1)*** es un caso particular con $\mu=0$ y desviación típica $\sigma=1$.

El área bajo la curva es la unidad, el coeficiente de asimetría es 0 y el de curtosis es 3. Esta es una rápida indicación para ver si los datos se ajustan a una distribución normal.

A esta curva se aproximan más o menos fielmente las frecuencias o probabilidades de ocurrencia de los datos de las series de ***temperatura***, ***precipitación anual, humedad***, ***presión, las componentes del viento por separado***, etc., cuando el número de observaciones crece indefinidamente, en especial cuando se refieren a observaciones o datos obtenidos a una misma hora del día, un mismo día del año en un mismo lugar.

Los datos de ***precipitación*** se distribuyen presentando una cola o rama de curva más larga hacia la derecha, al igual que ocurre con las ***temperatura máximas***, es decir tienen asimetría positiva. En el caso de las ***temperaturas mínimas*** lo es hacia la izquierda o asimetría negativa. De hecho las ***temperaturas máxima absoluta*** y ***mínima absoluta***, así como la mínima de las máximas o máxima de las mínimas se ajustan más a una ley general de valores extremos que veremos más adelante.

La precipitación máxima diaria así como ***la intensidad máxima de precipitación y la racha máxima de viento*** se aproximan a la *distribución de Gumbel*.

Funciones básicas de R

Existen cuatro funciones básicas en R que permiten calcular, respectivamente, los valores de la función de distribución, la función de masa o densidad (según caso discreto o continuo), los cuantiles y la obtención de números aleatorios de determinados modelos probabilísticos. Supongamos un modelo de distribución ***distri*** (normal, binomial, etc.)

- **p*distri* (x,par)** : calcula el valor de *la función de distribución* del modelo ***distri*** en el punto x. Es decir, *F(x)*, siendo *F* la función de distribución de *distri*. (Se sustituye *distri* por la función concreta que sea, por ejemplo ***norm*** para la distribución normal)
- **d*distri* (x,par)** : calcula el valor de la *función de masa o densidad* de la distribución *distri* en el punto x.
- **q*distri* (p,par)** : calcula el *p-cuantil* de la distribución *distri*. Es decir, $F^{-1}(p)$, siendo *F* la función de distribución de *distri*.
- **r*distri* (n,par)** : consigue *n valores al azar* según el modelo *distri*.

El segundo argumento ***par*** indica el parámetro o parámetros de la distribución considerada.

Ejemplo 4.1

Así por ejemplo si queremos obtener el valor de *la función de distribución* de una distribución normal de parámetros 1 y 2 (media $\mu=1$ y desviación típica $\sigma = 2$), ***N(1,2)***, en el punto x=1,5, ejecutaríamos

> pnorm(1.5,1,2)

obteniendo *F(x)=0,5987*. O si, como es habitual en estadística, queremos determinar el valor de la abscisa de una normal (0,1), que deje a la derecha un área de probabilidad 0,025 (es decir, ***el cuantil 0,975*** que normalmente se representa por $z_{0,025}$) ejecutaríamos

> qnorm(0.975, 0, 1)

obteniendo el resultado 1,96.

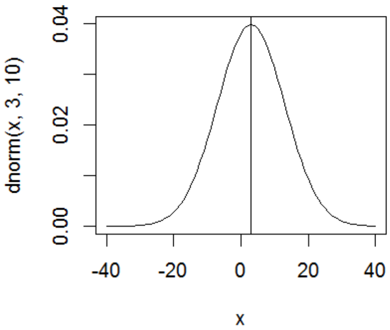
Si no indicamos los parámetros de la distribución por defecto cogerá la distribución *N(0,1)*.

Una forma gráfica de ver si una variable sigue una distribución normal es dibujar sobre el histograma de la variable, la distribución normal propuesta y ver si coinciden razonablemente.

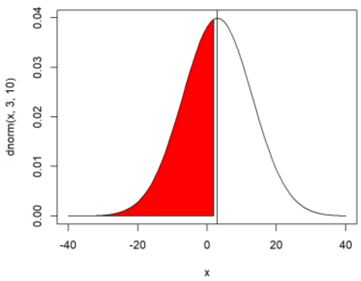
Para dibujar la gráfica de una distribución normal de media μ=3 y σ=10, *N(3,10):*
> x <- seq(-40,40,1)   # creamos primero un vector con los puntos del eje de las x
> plot(x, dnorm(x,3,10),type='l')   # ploteamos la densidad de distribución

También se puede dibujar más fácilmente mediante la función ***curve (función, limite_inferior_x, limite_superior_x)***, sin necesidad de escribir el vector *x*:
> curve (dnorm(x,3,10), -40, 40)

Para ver que el eje vertical de la distribución es 3 (media), podemos dibujar una línea vertical en x=3
> abline(v=3)



> y<-seq(-40,2.03,1)
> polygon(c(y,2), c(dnorm(y,3,10),0), col='red')



Para calcular la probabilidad de que una variable que ajustamos a una distribución normal estándar *N(0,1)* (o de cualquier otros parámetros) sea menor o igual que un valor determinado, por ejemplo 2,03
> pnorm (2.03, 0 ,1)

La probabilidad de que sea mayor que un determinado valor, por ejemplo 2.03
> 1- pnorm(2.03,0,1)

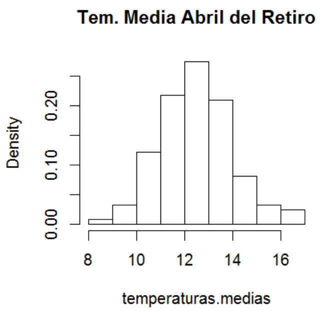
La probabilidad de que este comprendida entre 1.1 y 4.5 se calcularía restando la probabilidad de que sea menor que estos valores
> pnorm (4.5, 0 ,1) - pnorm (1.1, 0 ,1)

Ejemplo 4.2

Tenemos los datos de la temperatura media mensual del mes de Abril del Retiro de su serie histórica en un fichero ***Tem.media.Abril.Retiro.txt*** (faltan algunos años que no están completos). Primero vamos a leer los datos de la serie y hacer un histograma por defecto, dejando a R que coja los intervalos, luego a partir de la media $\overline{x}$ y desviación típica ***s*** de la serie, vamos a compararla con una distribución normal **N($\overline{x}$, *s*)** representando encima del histograma su función de densidad. Calcular también la asimetría y curtosis de la serie para ver si se ajusta a una distribución normal.

> datos <- read.table('Tem.media.Abril.Retiro.txt',header=TRUE) # lee el fichero de texto
> temperaturas.medias <- datos$abril # cogemos solo la columna de las temperaturas

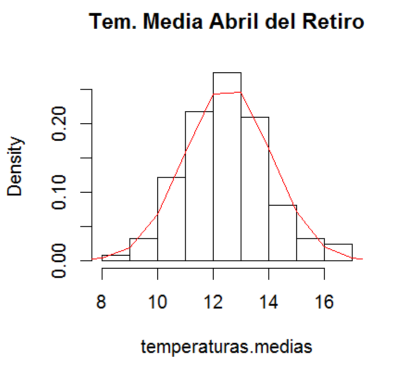
> num.datos <- length(temperaturas.medias) # número de datos o años completos (129)
> media <- mean(temperaturas.medias) # calculamos la media de todos los años (12.58)
> desv.tipica <- sd(temperaturas.medias) #calculamos la cuasi desviación típica de la serie (1.58)



Dibujamos el histograma de la serie
> hist (temperaturas.medias,prob=T, main='Tem. Media Abril del Retiro')

Ahora a partir de la media y la desviación típica superponemos la distribución ***N(media, desv.tipica)***

> x <- seq(4,20,1) # creamos primero un vector con los puntos del eje de las x
> lines (x, dnorm(x, media, desv.tipica), type='l',col='red') # ploteamos la función de la densidad de la distribución encima



Vemos que gráficamente se parece mucho la distribución normal calculada con la de los datos.

Calculamos el coeficiente de asimetría y de curtosis por medio del paquete ***moments*** como vimos en el capitulo 2.

> library('moments')   # carga el paquete moments
> coeficiente.asimetria <- skewness(temperaturas.medias) # coef. asimetria
> coeficiente.curtosis <- kurtosis(temperaturas.medias) # coef. curtosis

El Coeficiente de asimetría de los datos es casi *0 (0.4411)* (distribución simétrica) aunque ligeramente positiva, que nos indica que tiene un poco de *sesgo o asimetría positiva* (cola hacia la derecha) y ***el coeficiente de kurtosis*** nos da casi *3 (3.351461)*, que nos indica que es casi normal aunque ligeramente afilada o apuntada (*leptocúrtica*). Como resultado vemos que *la distribución normal es un buen ajuste para la temperatura media*.

Conociendo la media y desviación típica de una serie o de una muestra muy grande del colectivo, se puede calcular la probabilidad o frecuencia con que se presenta un cierto valor x de la variables, comprendido en un intervalo x –*a*/2 y x + *a*/2, con a un valor cualquiera.

Por ejemplo en el ejemplo anterior podemos pedir la probabilidad de que la temperatura media del mes de Abril en el retiro sea de 13º $\pm$ 1º C

> pnorm ((13+1), media, desv.tipica) - pnorm ((13-1), media, desv.tipica)

Obtenemos 0.46, es decir en un 46% de las veces se encontrara en ese intervalo.

Ejercicio 4.1

Dejamos como ejercicio repetir el proceso pero para las precipitaciones medias anuales de la misma estación, tratando de ajustarlas a una distribución normal, fichero: '***Precip.Retiro.txt'.*** La columna con los datos tiene ahora de cabecera: ***Media***

4.2 Ajustes de datos a las principales funciones de distribución climatológicas.

Distribución binomial o de Bernoulli

Este modelo probabilístico discreto modeliza el número de éxitos en experimentos denominados *pruebas de Bernoulli*. Estas pruebas consisten en la realización de ensayos repetidos e independientes, existiendo en cada ensayo dos resultados posibles, éxito o fracaso, y manteniendo constante la probabilidad de éxito a lo largo de los ensayos.

Si consideramos dos sucesos aleatorios mutuamente excluyentes, como, por ejemplo, "cielo con nubes" y "cielo sin nubes" y llamamos p a la probabilidad del primer suceso, la probabilidad del segundo suceso será $q=1-p$. Si ahora consideramos n días consecutivos, la probabilidad de que habiéndose dado el primer día con nubes, el segundo también lo sea, es una probabilidad compuesta, $p \times p$ y la probabilidad de que el *n-esimo* también lo sea será

$$p^n = p^n q^0$$

La probabilidad de que haya *x* días seguidos "con nubes" y el resto *n-x* "sin nubes" es

$$p^x q^{n-x}$$

Si no se toma en cuenta el orden en que producen los n sucesos, la probabilidad de tener *x* días con nubes y *n-x* sin nubes de los *n* días, en cualquier orden es

$$C_x^n p^x q^{n-x} = \binom{n}{x} p^x q^{n-x} = \binom{n}{x} p^x (1-p)^{n-x} \qquad \text{con } x=0,1,\ldots,n$$

Donde C_x^n son las combinaciones de n elementos tomados de *x* en *x*.

$$C_x^n = \binom{n}{x} = \frac{n!}{x!\,(n-x)!}$$

La función de masa es entonces

$$p_X(x) = P\{X = x\} = \binom{n}{x} p^x (1-p)^{n-x} \quad , \quad x = 0, 1, \ldots, n$$

En este caso diremos que *X* sigue una distribución (o tiene un modelo) ***binomial*** de parámetros ***n*** y ***p*** y lo representaremos por $X \to B(n,p)$

Su media y su varianza son

$$E[X] = np \qquad\qquad V(X) = np(1-p)$$

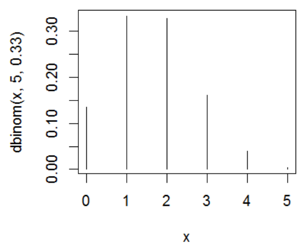
Para calcular factoriales en *R* se usa: ***factorial(x)*** y para combinaciones de N elementos tomados de r en r se usa: ***choose(N,r)***

Para calcular probabilidades binomiales con *R*, usamos el nombre ***binom*** en las funciones

```
> pbinom(x,n,p)     # Valor de la función de distribución en x de la binomial (n,p)
> dbinom(x,n,p)     # Valor de la función de masa en x de la binomial (n,p)
> qbinom(q,n,p)     # Cuantil de orden q de la binomial (n,p)
> rbinom(m,n,p)     # muestra aleatoria de tamaño m de la binomial (n,p)
```

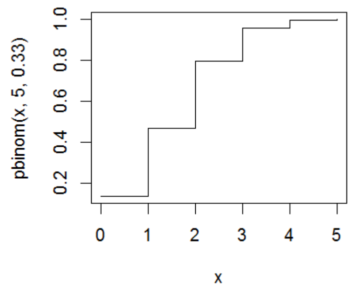
Si queremos por ejemplo representar la función de masa de una distribución binomial ***B(5,0.33)*** podemos pintarla con un ***plot*** y con la opción ***type='h'*** para pintar un histograma

```
>  x <- seq(0,5,len=6)     # Vector del eje x
>  plot (x, dbinom(x, 5, 0.33), type= 'h')
```



Para pintar la función de distribución:

```
>  plot (x, pbinom(x,5,0.33), type= 's')
```



Ejemplo 4.3

De los temporales del NW que llegan a la península Ibérica, supongamos que tienen una probabilidad de que afecten a Madrid de p=0.8.

a) ¿Cuál será la probabilidad de que los ocho próximos que se presenten, a lo sumo tres afecten a Madrid?
b) La probabilidad de que afecten al menos dos (dos o más).

a) Tendremos que buscar la probabilidad de que afecten 3 o 2 o 1 o 0 temporales.

Probabilidad de 3 o menos: proba. de 0 + probab. de 1 + probab. de 2 + probab. de 3:
(3 exitos)

P(3 o menos) = $\binom{8}{0}0.8^0 \times 0.2^8 + \binom{8}{1}0.8^1 \times 0.2^7 + \binom{8}{2}0.8^2 \times 0.2^6 + \binom{8}{3}0.8^3 \times 0.2^5 = 0.0104$

Con R se haría simplemente:

```
>  pbinom(3,8,0.8)
```

b) La probabilidad de que afecten al menos dos (dos o más) se puede plantear como suceso complementario de tener éxito 1 vez

p(éxito $\geq$ 2) = 1 – p(éxito cero veces) – p(éxito una vez) =

$$= 1 - \binom{8}{0} 0.8^0 \text{ x } 0.2^8 + \binom{8}{1} 0.8^1 \text{ x } 0.2^7 = 0.9999$$

En R simplemente

> 1 - pbinom(1,8,0.8)

Ejemplo 4.4

Si en un régimen anticiclónico tan posible es un día de niebla como un día despejado. En un anticiclón que dure 8 días cual es la probabilidad de que haya 5 nieblas.

> dbinom(5,8,0.5) # calculamos solo la probabilidad de que exactamente haya 5 nieblas. Hay que usar la función de densidad (masa).

Si se pidiese la probabilidad de que haya cero, uno, dos, tres o cuatro días de niebla, o sea a lo máximo cuatro, habría que sumar las probabilidades de todos los casos. En R

> dbinom(0,8,0.5) + dbinom(1,8,0.5) + dbinom(2,8,0.5) + dbinom(3,8,0.5) + dbinom(4,8,0.5)

O más fácilmente como antes usando la función de distribución:

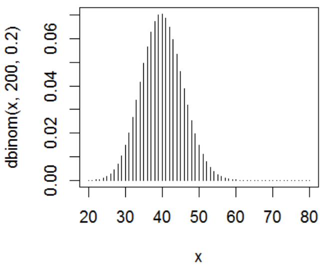
> pbinom(4,8,0.5)

Cuando el número de pruebas, crece bastante la distribución de probabilidad binomial se aproxima a la distribución normal.

Si pintamos una distribución para n elevado:

> x <- seq(20,80,1) # Vector del eje x

> plot (x, dbinom(x,200,0.2), type= 'h')



Ejercicio 4.2

Supongamos que en un aeropuerto el 30% de los días de invierno hay niebla al amanecer; cual es la probabilidad de que de 4 amaneceres de invierno, elegidos al azar, a lo más haya tres de ellos en los que se presente niebla.

Nota: la probabilidad de que haya niebla es $p=0.3$ y de no niebla $q=0.7$

Calcúlese también la media y la varianza de la distribución binomial.

(*Solución*: *99.2%, 1.2* y *0.84*)

Ejercicio 4.3

En una estación la frecuencia de temperaturas máximas anuales iguales o superiores a 40º C es de 1%. Cuál será la probabilidad de que en 1995 se den: ninguno, uno, dos o tres días de estas temperaturas suponiendo aplicable la distribución binomial. (En este caso el número de pruebas es 365 y p=0.01)

(*Solución*: 50.4%)

Distribución de Poisson

Esta distribución discreta introducida en 1837 por Simeon Denis Poisson se usa en general para modelizar el número de veces (0, 1, 2, …) que ocurren sucesos raros, como por ejemplo, accidentes de tráfico, crecidas extraordinarias, muy altas precipitaciones, etc. En general los días de fenómenos poco frecuentes.

La función de masa o de densidad de la variable número de éxitos de esta distribución es

$$p_X(x) = \frac{e^{-\lambda}\,\lambda^x}{x!} \qquad\qquad x = 0,1,2,\ldots$$

Siendo $\lambda > 0$ el único parámetro del que depende esta distribución. Este modelo se llama ***Distribución de Poisson*** de parámetro $\boldsymbol{\lambda}$. Este parámetro es a la vez su media y su varianza, E[X] = λ = V(X).

La distribución de Poisson es más aceptable que la de Bernoulli cuando la probabilidad de ocurrencia de un suceso, *p*, es muy pequeña y puede deducirse de la de Bernoulli cuando n crece indefinidamente, viniendo dado λ por el producto *np*.

Esta aproximación es buena cuando p es muy pequeño con respecto a *np* y a su vez *np* es muy pequeño con respecto a *n*. Como indicación se toma por buena la aproximación cuando al menos $np < 5$ y $p < 0.1$.

Para calcular probabilidades de Poisson de parámetro $\lambda=a$ con *R*, usamos el nombre ***pois*** en las funciones

```
> ppois(x,a)    # Valor de la función de distribución en x de la Poisson(a)
> dpois(x,a)    # Valor de la función de masa en x de la Poisson (a)
> qpois(p,a)    # Cuantil de orden p de la Poisson (a)
> rpois(m,a)    # muestra aleatoria de tamaño m de la Poisson (a)
```

Ejemplo 4.5

De un estudio estadístico de las nieblas ocurridas en el mes de marzo en un aeropuerto durante 50 años se obtiene: 25 años sin ninguna niebla ese mes, 15 años con una niebla; 6 años con dos nieblas; tres años con tres niebla; y un año con 4 nieblas. Determínese la distribución de Poisson que represente el fenómeno; y las probabilidades teóricas respectivas para cada suceso.

Se sabe que la media o esperanza de los datos, es el parámetro λ de esta distribución, es decir:

$$\lambda = \frac{n_1 s_1 + n_2 s_2 + n_3 s_3 + \cdots}{n_1 + n_2 + n_3 + \cdots} = \frac{25\cdot 0 + 15\cdot 1 + 6\cdot 2 + 3\cdot 3 + 1\cdot 4}{50} = 0.8 = n\cdot p$$

Así pues tenemos la función de masa es

$$p(x) = \frac{0.8^x e^{-0.8}}{x!}$$

Que aplicada a cada uno de los casos

$p(0) = e^{-0.8}$, $p(1) = 0.8\,e^{-0.8}$, $p(2) = \frac{0.8^2}{2}\,e^{-0.8}$, $p(3) = \frac{0.8^3}{6}\,e^{-0.8}$, $p(4) = \frac{0.8^4}{24}\,e^{-0.8}$

> dpois(x,0.8) ***# para x= 0,1,2,3 y 4***

Ejemplo 4.6

Los días de niebla que pueden ocurrir en un lugar de 100 días de invierno tienen una probabilidad de ocurrencia de p=0.01. ¿Podrían representarse mediante una distribución de Poisson? ¿Cuál será la probabilidad de que ocurra un día de niebla?

Según lo visto anteriormente siempre que *np*, media de la distribución binomial, sea inferior a 5 es aplicable. Conocida la media λ=1 (*np*), la distribución de Poisson es

$$p(x) = \frac{1^x e^{-1}}{x!}$$

La probabilidad de un día de niebla será:

$$p(1) = \frac{1}{e} = 0.368$$

> dpois(1,1)

Ejemplo 4.7

Un 5% de las borrascas que llegan del SW dan una precipitación media superior a 120 mm. ¿Cuál es la probabilidad de que de una muestra de 6 de ellas, dos den precipitación superior a 120 mm.?

La probabilidad p es 0.05. Se puede resolver por la distribución de Bernoulli o por la de Poisson.

Según Bernoulli:

$$p(2) = \binom{6}{2} 0.05^2 \cdot 0.95^4 = 0.0305$$

> dbinom(2,6,0.05)

Según Poisson:

$$\lambda = Np = 6\,x\,0.05 = 0.3 \qquad p(2) = \frac{0.3^2\,e^{-0.3}}{2!} = 0.0333$$

> dpois(2,0.3)

Ejercicio 4.4

En una cuenca la probabilidad de que precipite una media de 200 mm en dos horas es 1/1000. ¿Cuál es la probabilidad de que esta precipitación ocurra cinco veces en 1000 años? Comprobar primero que es aplicable la distribución de Poisson.
(*Solución*: 0.003)