Durante la segunda guerra mundial, 535 bombas fueron lanzadas sobre el sur de Londres. Esta área ha sido dividida en una cuadrícula de 576 pequeños cuadrados de 0,25 metros cuadrados cada uno. Asumiendo que el objetivo es aleatorio: cada bomba impacta en un solo lugar a la vez, cada lugar tiene la misma probabilidad de ser impactada y, los impactos son sucesos independientes.
Si queremos analizar el número de bombas que impactan en un cuadrado en particular. ¿Qué modelo de distribución sigue dicha variable?

4.3. Distribuciones de probabilidad discretas más comunes

Distribución binomial

Muchos experimentos consisten en la repetición del ensayo, con dos posibles resultados, que pueden marcarse como exitoso o fallido (ensayo dicotómico). Si la probabilidad de éxito (p) es la misma en cada ensayo y estos son independientes, se denominan *ensayos de Bernoulli*. Si una variable discreta aleatoria (VAD) indica el número de éxitos en n ensayos de Bernoulli, con una probabilidad de éxito p, decimos que esta variable sigue una distribución binomial con los parámetros n y p. Su notación es $X \hookrightarrow B(n,p)$. Su función de densidad, su media y su varianza vienen dadas por:

$$f(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad \text{para} \qquad x = 0, 1, 2, ..., n.$$

$$E(X) = np, \qquad V(X) = np(1-p).$$

Distribución geométrica

Si una VAD indica el número de ensayos de Bernoulli necesarios hasta el primer suceso, con una probabilidad de éxito p, decimos que esta variable sigue una distribución geométrica con un parámetro p y su notación es $X \hookrightarrow G(p)$. Su función de densidad, su media y su varianza vienen dadas por:

$$f(x) = P(X = x) = p(1-p)^{x-1} \qquad \text{para} \qquad x = 1, 2, ...$$

$$E(X) = \frac{1}{p}, \qquad V(X) = \frac{(1-p)}{p^2}.$$

Durante la segunda guerra mundial, 535 bombas fueron lanzadas sobre el sur de Londres. Esta área ha sido dividida en una cuadrícula de 576 pequeños cuadrados de 0,25 metros cuadrados cada uno. Asumiendo que el objetivo es aleatorio: cada bomba impacta en un solo lugar a la vez, cada lugar tiene la misma probabilidad de ser impactada y, los impactos son sucesos independientes.
Si queremos analizar el número de bombas que impactan en un cuadrado en particular. ¿Qué modelo de distribución sigue dicha variable?

Distribución binomial negativa

Si una VAD indica el número de ensayos de Bernoulli necesarios hasta obtener r éxitos, con una probabilidad de éxito p, podemos decir que esta variable sigue la distribución binomial negativa con los parámetros r y p. Su notación es $X \hookrightarrow NB(r,p)$. Su función de densidad, su media y su varianza vienen dadas por:

$$f(x) = P(X = x) = \binom{x-1}{r-1} p^r (1-p)^{x-r} \qquad \text{para} \qquad x = r, r+1, r+2, \ldots$$

$$E(X) = \frac{r}{p}, V(X) = r\frac{(1-p)}{p^2}.$$

Distribución hipergeométrica

Si una VAD indica el número de éxitos en n ensayos dependientes dicotómicos, con una probabilidad de éxito que cambia en cada ensayo (población que consiste en N éxitos y $N-k$ fracasos), podemos decir que esta variable sigue la distribución hipergeométrica con parámetros n, N y k. Su notación es $X \hookrightarrow H(n,N,k)$. Su función de densidad, su media y su varianza vienen dadas por:

$$f(x) = P(X = x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} \quad \text{para} \quad x = max\{0, n+k-N\}, \ldots, min\{k,n\}$$

$$p = \frac{k}{N}, E(X) = np, V(X) = np(1-p)\frac{N-n}{N-1}.$$

Distribución de Poisson

Los experimentos que tienen como resultado el número de eventos que ocurren durante un intervalo de tiempo dado o en una región específica se denominan *ensayos de Poisson*. Si una VAD indica el número de *ensayos de Poisson*, con una frecuencia de ocurrencia media λ, podemos decir que esta variable sigue la distribución de Poisson con parámetro λ. Su notación es $X \hookrightarrow P(\lambda)$ y su función de densidad, su media y su varianza vienen dadas por:

$$f(x) = P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} \qquad \text{para} \qquad x = 0, 1, 2, \ldots$$

$$E(X) = \lambda, \qquad V(X) = \lambda.$$

Distribución de Poisson con $\lambda = \dfrac{535}{576} = 0.929$, $Pois(0.929)$

```
x2 <- 0:6
f2 <- dpois(x2,lambda=0.929)
plot(x2, f2, type="h", col="red", lwd=3, main="Función de probabilidad",
xlab="X", ylab="f(x)", xlim=c(-0.5,6.5), ylim=c(0,0.5))
points(x2, f2, col="red", lwd=8);
```

```
x2 <- 0:6
F2 <- ppois(x2,lambda=0.929)
plot(c(-1,x2,7), c(0,F2,1), type="s", col="red", lwd=3, main="Función de distribución", xlab="X", ylab="F(x)")
points(x2, F2, col="red", lwd=8)
```

¿Cuál es la probabilidad de que exactamente 2 bombas impacten en una zona en particular?

```
dpois(2,lambda=0.929)
```

¿Cuál es la probabilidad de que una zona concreta sea bombardeada (al menos caiga una bomba)?

```
ppois(0,lambda=0.929,lower.tail = F)
```

```
1-dpois(0,lambda=0.929)
```

Lowe.tail Si TRUE, las probabilidades son P(X <= x), FALSE P(X > x) (defecto TRUE)

Si queremos analizar el número de zonas que reciben exactamente 2 impactos. ¿Qué modelo de distribución sigue dicha variable?

Binomial con $n = 576$ y $p = 0.1704284$, $B(576, 0.1704284)$

```
x3 <- 0:576
f3 <- dbinom(x3,size=576,prob=0.1704284)
plot(x3, f3, type="h", col="red", lwd=1, main="Función de probabilidad", xlab="X", ylab="f(x)", xlim=c(50,150), ylim=c(0,0.05))
points(x3, f3, col="red", lwd=3);
```

```
F3 <- pbinom(x3,size=576,prob=0.1704284)
plot(c(-1,x3,578), c(0,F3,1), type="s", col="red", lwd=1, main="Función de distribución", xlab="X", ylab="F(x)", xlim=c(50,150))
points(x3, F3, col="red", lwd=1)
```

¿Cuál es la probabilidad de que menos de 100 zonas reciban exactamente 2 impactos?

```
pbinom(99,size=576,prob=0.1704284)
```

¿Cuántas zonas se espera que sufran exactamente dos impactos?

```
576*0.1704284
```

Si queremos analizar el número de zonas que deben ser inspeccionadas para encontrar 10 que hayan sido bombardeadas. ¿Qué modelo de distribución sigue dicha variable?

Distribución Binomial Negativa con $r = 10$ y $p = 0.6050515$, $BN(10, 0.6050515)$

Realiza la gráfica de su función de densidad (Manipula los límites de los ejes hasta encontrar una gráfica apreciable)

```
x4 <- 10:40
f4 <- dnbinom(x4-10,size=10,prob=0.6050515) # Se ha de poner el número de fallos que ocurren antes de cumplir el objetivo de éxitos
plot(x4, f4, type="h", col="red", lwd=1, main="Función de probabilidad", xlab="X", ylab="f(x)", xlim=c(0,40.5), ylim=c(0,0.15))
points(x4, f4, col="red", lwd=3);
```

Realiza la gráfica de su función de distribución

```
F4 <- pnbinom(x4-10,size=10,prob=0.6050515)
plot(c(0,x4,40), c(0,F4,1), type="s", col="red", lwd=1, main="Función de distribución", xlab="X", ylab="F(x)")
points(x4, F4, col="red", lwd=1)
```

¿Cuál es la probabilidad de que se deban inspeccionar al menos 20 zonas para encontrar 10 que hayan sido bombardeadas?

```
pnbinom(19-10,size=10,prob=0.6050515, lower.tail = F)
```

¿Cuántas zonas han de ser inspeccionadas para encontrar 10 que hayan sido bombardeadas?

Si una zona es inspeccionada ......0.6050 bombardejada, entonces n inspeccionades ....10 bombardeadas....... 10/0.6050515

5.3. Distribuciones de probabilidad continuas más comunes

Distribución uniforme

Si una VAC toma cualquier valor en un intervalo $[a, b]$ con la misma probabilidad, decimos que esta variable sigue una distribución uniforme continua. Su función de densidad, su media y su varianza vienen dadas por:

$$f(x) = \frac{1}{b-a} \qquad \forall x \in [a, b]$$

$$E(X) = \frac{a+b}{2}$$

$$V(X) = \sigma^2 = \frac{(b-a)^2}{12}$$

Distribución exponencial

La familia de distribuciones exponenciales proporciona modelos que se utilizan mucho en la ciencia y en la ingeniería. La VAC, que es igual a la distancia en los sucesivos eventos de un proceso de Poisson con una media de eventos $\lambda > 0$ por intervalo la unidad, sigue una distribución exponencial con parámetro λ. Su función de densidad, su media y su varianza vienen dadas por:

$$f(x) = \lambda e^{-\lambda x} \qquad \forall x \in \mathbb{R}^+$$

$$E(X) = \frac{1}{\lambda}$$

$$V(X) = \sigma^2 = \frac{1}{\lambda^2}$$

Distribución normal

La distribución de probabilidad continua más importante en todo el campo de la estadística es la distribución normal. Su representación gráfica, denominada **curva normal**, es la curva con forma de campana, la cual describe aproximadamente el fenómeno que se presenta en la naturaleza, en la industria y en la investigación. Las medidas físicas en áreas tales como los experimentos meteorológicos o los estudios pluviales, y las medidas de partes manufacturadas suelen quedar mejor explicadas utilizando una distribución normal.

Si una VAC X tiene una distribución normal con parámetros μ y σ $(X \hookrightarrow N(\mu, \sigma))$ o $(N(\mu, \sigma^2))$, donde $-\infty < \mu < \infty$ y $0 < \sigma$, su función de densidad, su media y su varianza vienen dadas por:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad \forall x \in \mathbb{R}$$

$$E(X) = \mu, \qquad V(X) = \sigma^2$$

Las ventas diarias de un supermercado siguen una distribución normal, con media 9000 euros y desviación estándar 2000 euros.
Realiza la gráfica de la función de densidad de las ventas diarias.

```
mu = 9000
sigma = 2000
curve(dnorm(x, mean=mu, sd=sigma), xlim = c(mu-4*sigma, mu+4*sigma), col="red", lwd=3, main="N(9000,2000²)", xlab="X", ylab="f(x)")
```

Realiza la gráfica de la función de distribución de las ventas diarias.

```
curve(pnorm(x, mean=mu, sd=sigma), xlim = c(mu-4*sigma, mu+4*sigma), col="red", lwd=3, main="N(9000,2000²)", xlab="X", ylab="F(x)")
```

Calcula la probabilidad de que un día las ventas sean mas pequeñas que 10000 euros. Representa dicha probabilidad en la gráfica de la función de distribución.

```
F10000 <- pnorm(10000, mean=mu, sd=sigma); F10000
```

```
curve(pnorm(x, mean=mu, sd=sigma), xlim = c(mu-4*sigma, mu+4*sigma), col="red", lwd=3, main="N(9000,2000²)", xlab="X", ylab="F(x)")
abline(v=10000, col="blue");
abline(h=F10000, col="blue");
text(10000, F10000,expression(P(X<=10000)), pos=2, col="blue")
```

Calcula la probabilidad de que un día las ventas sean mayores de 12000 euros. Representa dicha probabilidad en la gráfica de la función de densidad.

```
FC12000 <- pnorm(12000, mean=mu, sd=sigma, lower.tail = FALSE); FC12000
```

```
curve(dnorm(x, mean=mu, sd=sigma), xlim = c(mu-4*sigma, mu+4*sigma), col="red", lwd=3, main="N(9000,2000²)", xlab="X", ylab="f(x)")
cord.x=c(12000,seq(12000,15000,length=100),15000) # Vector de vértices en x para elpolígono
cord.y=c(0,dnorm(seq(12000,15000,length=100),mean=mu,sd=sigma),0) # Vector de vértices en y
polygon(cord.x, cord.y, col='skyblue')
```

Calcula la probabilidad de que un día las ventas sean exactamente de 7000 euros. 0

El 90% de los días, las ventas superan un valor, ¿Cuál es ese valor?

```
qnorm(0.9, mean=mu, sd=sigma, lower.tail = FALSE)
```

El 30% de los días, las ventas no superan un valor, ¿Cuál es ese valor?

```
qnorm(0.3, mean=mu, sd=sigma)
```

Realiza 10000 simulaciones de las ventas diarias. (Utiliza como semilla el número 123)

```
set.seed(123) sim.X <- rnorm(10000, mean=mu, sd=sigma)
```

Representa gráficamente la frecuencia de los resultados obtenidos y compáralo con el diagrama de la función de densidad.

```
hist(sim.X, freq=FALSE);
curve(dnorm(x, mean=mu, sd=sigma), xlim = c(mu-4*sigma, mu+4*sigma), col="red", lwd=3, main="N(9000,2000²)", xlab="X", ylab="f(x)",add=TRUE)
```

¿Cuál es el valor de la media de los resultados de las simulaciones?

```
mean.sim <- mean(sim.X); mean.sim
```

¿Cuál es la mediana de los resultados de las simulaciones?

```
med.sim <- median(sim.X); med.sim
```

¿Cuál es la desviación típica de los resultados de las simulaciones?

```
sd.sim <- sd(sim.X); sd.sim
```

Sea X la distancia en metros que recorre un animal desde el lugar donde nace hasta el primer territorio desocupado que encuentra. Según el artículo "Competition and dispersal from multiple nests" publicado en la revista científica Ecology, para las ratas canguro, la distancia X se puede modelar mediante una variable aleatoria con distribución exponencial con parámetro 0.01005 $m^1 - 1$.

Realiza la gráfica de la función de densidad de X.

```
lambda = 0.01005
curve(dexp(x, rate=lambda), xlim = c(0, 4/lambda), col="red", lwd=3, main="Exp(0.01005)", xlab="X", ylab="f(x)")
```

Realiza la gráfica de la función de distribución de X.

```
curve(pexp(x, rate=lambda), xlim = c(0, 4/lambda), col="red", lwd=3, main="Exp(0.01005)", xlab="X", ylab="F(x)")
```

Calcula la probabilidad de que un animal recorra menos de 100 metros hasta encontrar el primer territorio desocupado. Representa dicha probabilidad en la gráfica de la función de distribución.

```
F100 <- pexp(100, rate=lambda); F100
```

```
curve(pexp(x, rate=lambda), xlim = c(0, 4/lambda), col="red", lwd=3, main="Exp(0.01005)", xlab="X", ylab="F(x)")
abline(v=100, col="blue");
abline(h=F100, col="blue");
text(100, F10000,expression(P(X<=100)), pos=2, col="blue")
```

Calcula la probabilidad de que un animal recorra más de 80 pero menos de 110 metros hasta encontrar el primer territorio desocupado. Representa dicha probabilidad en la gráfica de la función de densidad.

```
pexp(110, rate=lambda) - pexp(80, rate=lambda)
```

```r
curve(dexp(x, rate=lambda), xlim = c(0, 4/lambda), col="red", lwd=3, main="Exp(0.01005)", xlab="X", ylab="f(x)")
cord.x=c(80,seq(80,110,length=100),110) # Vector de vértices en x para elpolígono
cord.y=c(0,dexp(seq(80,110,length=100),rate=lambda),0) # Vector de vértices en y
polygon(cord.x, cord.y, col='skyblue')
```

¿Cuál es el valor de la distancia mediana que recorre un animal desde el lugar donde nace hasta el primer territorio desocupado que encuentra?

```r
qexp(0.5, rate=lambda)
```

Realiza 10000 simulaciones de la distancia que recorre un animal. (Utiliza como semilla el número 123)

```r
set.seed(123)
sim.X <- rexp(10000, rate=lambda)
```

```r
hist(sim.X, freq=FALSE);
curve(dexp(x, rate=lambda), xlim = c(0, 4/lambda), col="red", lwd=3, main="Exp(0.01005)", xlab="X", ylab="f(x)", add=TRUE)
```

¿Cuál es el valor de la media de los resultados de las simulaciones?

```r
mean.sim <- mean(sim.X); mean.sim
```

¿Cuál es la mediana de los resultados de las simulaciones?

```r
med.sim <- median(sim.X); med.sim
```

¿Cuál es la desviación típica de los resultados de las simulaciones?

```r
sd.sim <- sd(sim.X); sd.sim
```