Control Estadístico de Procesos

Gustavo Ahumada

Distribuciones de probabilidad univariada

En esta sección examinamos las distribuciones de probabilidad univariadas que son empleadas frucuentemente para modelar fénomenos aleatorios. Primero, introducimos las distribuciones de probabilidad discreta, seguido por las distribuciones de probabilidad continuas. Las distribiciones discretas pueden ser empleadas para modelar número de fallos hasta encontrar un seis en el lanzamiento de un dado, el número de estudiantes con una nota de siete en una clase, o el número de taxis que pasan por una calle. Las distribuciones continuas son usadas para modelar variables tales como peso, estatura, y tiempo.

Distribuciones de probabilidad univaridas discretas

Cuando existen un número contable de elementos en un espacio muestral para un solo experimento, una **distribución discreta de probabilidad** es el resultado. La probabilidad para la ocurrencia de un valor dado de una variable aleatoria puede ser igual para cada valor o puede ser más o menos probable para ciertos valores, lo cual depende de la estructura del experimento.

Distribución uniforme discreta

La más simple de todas las distribuciones de probabilidad discreta es una en la cual la variable aleatoria toma cada uno de sus valores con una misma probabilidad. Esta distribución de probabilidad es conocida como **distribución uniforme discreta**.

Definición: Si la variable aletarga X toma valores $x_1, x_2, ..., x_n$, con idénticas probabilidades, entonces la distribución uniforme discreta está dada por

$$f(x; n) = \frac{1}{n}, \quad x = x_1, x_2, ..., x_n,$$

Teorema La media y la varianza de la distribución uniforme discreta $f(x; n)$ son

$$E(X) = \mu = \frac{\sum_{i=1}^{n} x_i}{n}, \quad y \quad \sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n}$$

Ejemplo. Un foco se selecciona al azar de una caja que contiene un foco de 40 watts, uno de 60 watts, uno de 75 watts, uno de 100 watts y uno de 120 watts. Escribir la función de probabilidad para la variable aleatoria que representa la selección aleatoria de un foco y determinar la media y la varianza de esa variable aleatoria.

Solución. La variable aleatoria X puede asumir el siguiente conjunto de valore $S = \{40, 60, 75, 100, 120\}$. La función de densidad de probabilidad para la variable aleatoria X es

$$P(X = x/n = 5) = 1/5 \quad para \quad x = 40, 60, 75, 100, 120.$$

Los calculos aritmeticos son realizados con el siguiente código

```r
Focos <- c(40, 60, 75, 100, 120)
n <- length(Focos)
media_focos <- (1/n) * sum(Focos)
var_focos <- (1/n) * sum((Focos - media_focos)^{2})
resultado <- c(media_focos, var_focos)
resultado
```

```
## [1]  79 804
```

Distribuciones Bernoulli y Binomial

Cuando la misma moneda es lanzada n veces por la misma persona bajo las mismas condiciones, por tal razón que cada lanzamiento de la monera resultará en uno dos posibles resultados (cara o sello), y que la probabilidad de obtener un cara en cualquier intento es una constante $\frac{1}{2}$. Lanzar una moneda una sola vez es un ejemplo de un **proceso Bernoulli**. Cada ensayo se llama **experimento Bernoulli** con solamente dos posible resultados. Los resultados son mutuamente excluyentes y exhaustivos; por ejemplo, exito o fracaso, cierto o falso, vivo o muerto, hombre o mejur, etc. Una variable aleatoria Bernulli, X, puede tomar tomar dos valores, donde $X(exito) = 1$ y $X(fracaso) = 0$. La probabilidad que X sea un exito es p, y la probabildad de que X sea un fracaso es de $1 - p$. La **pdf**, la media y la varianza de la distribución de una variable aleatoria Bernoulli son:

$$
\begin{aligned}
f(X = x/p) =& p^x(1-p)^{1-x}, x = 0, 1 \\
E(X) = p& \\
Var(X) = p&(1-p)
\end{aligned}
$$

Proceso Bernoulli Si se habla con exactitud, el proceso de Bernoulli debe tener las siguientes propiedad: 1. El experimento consiste en n pruebas que se repiten. 2. Cada prueba produce un resultado que se puede clasificar con éxito o fracaso. 3. La probabilidad de un éxito, que denota por p, permanece constante en cada prueba. 4. Las pruebas que se repiten son independientes.

El número de X éxitos en n experimentos de Bernoulli se denomina **variable aleatoria binomial**. La distribución de probabilidad de esta variable aleatoria discreta se llama **distribución binomial**, y sus valores se denotarán como $b(x; n, p)$, pues depended de número de pruebas y de la probabilidad de éxito en una prueba dada

Definición. Un experimento de <bernoulli puede terner como resultado un éxito con probabilidad de p y fracaso con probabilidad $q = 1 - p$. Entonces la distribución de probabilidad de la variable aleatoria binomial X, el núermo de éxtios en n pruebas independientes, es

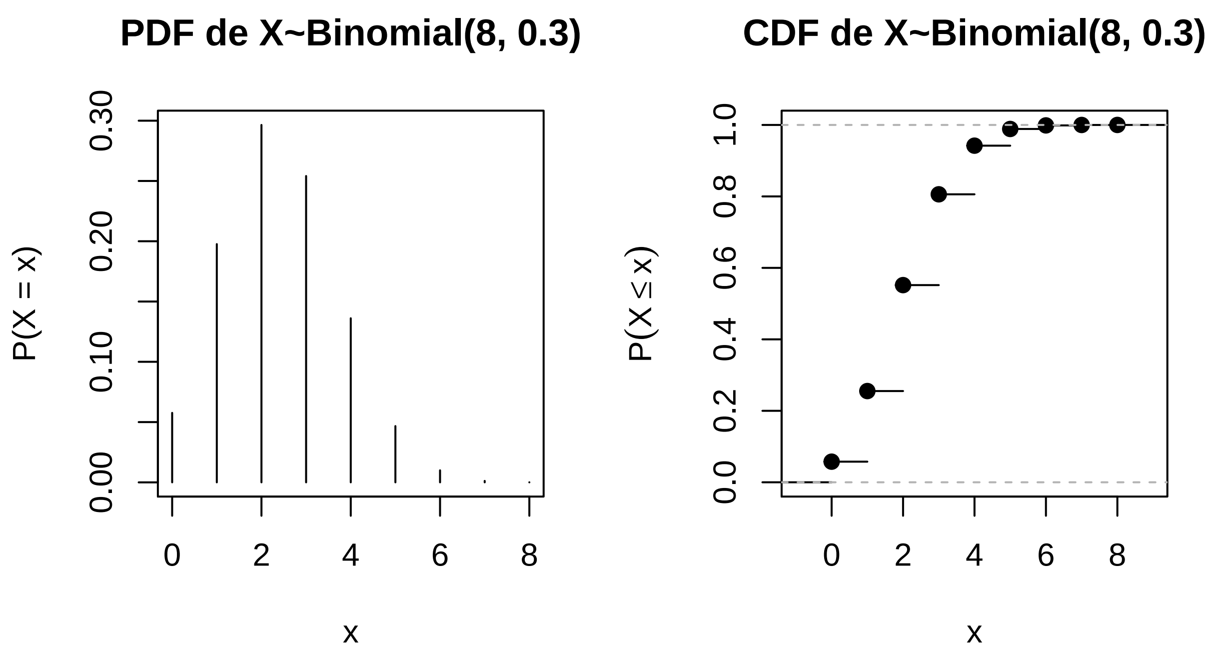
$$b(x; n, p) = \binom{n}{x} p^x q^{n-x}, \quad x = 0, 1, 2, ..., n. \tag{1}$$

Teorema. La media y la varianza de la distribución binomial $b(x; n, p)$ son

$$E(X) = \mu = np \quad y \quad \sigma^2 = npq. \tag{2}$$

El código para crear las gráficas que representen la **pdf** y la **cdf** para una variable aleatoria binomial $b(x; 8.0.3)$ se presenta a continuación

```r
opar <- par(no.readonly = TRUE)
par(mfrow = c(1, 2), pty = "s")
x <- 0:8
px <- dbinom(x, 8, 0.3)
plot(x, px, type = "h", xlab = "x", ylab = "P(X = x)",
    main = "PDF de X~Binomial(8, 0.3)")
xs <- rep(0:8, round(dbinom(0:8, 8, 0.3)*100000, 0))
plot(ecdf(xs), main = "CDF de X~Binomial(8, 0.3)",
    ylab = expression(P(X<=x)), xlab = "x" )
```




```r
par(opar)
```

Ejemplo. **Simulación proceso Bernoulli** Considerar el problema de simular m muestras repetidas de n procesos de Bernoulli. a. Escribir la función que generará m repetidas muestras de n procesos de Bernoulli cada una con un probabilidad de ocurrencia de p. La función debería crear un histograma con los valores teoricos superpuestos sobre los valores simulados.

b. Emplear la función para generar un muestra de tamaño $n = 5$ con $p = 0.5$ para simular la distribución binomial.

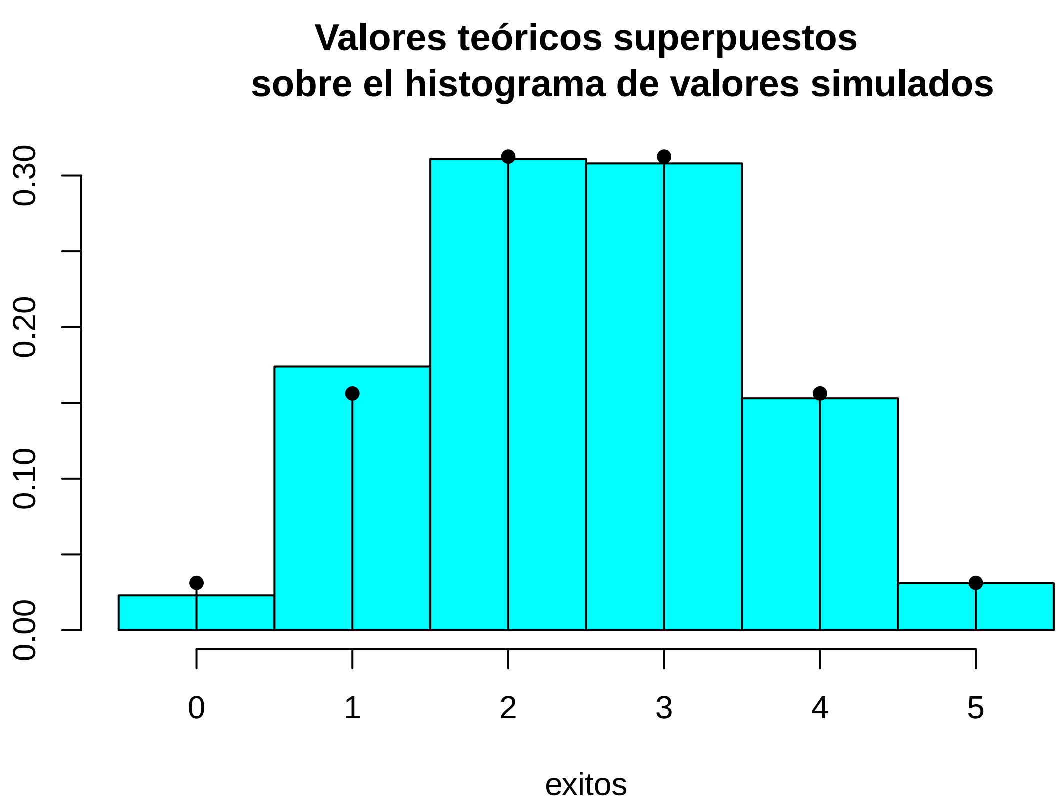
c. **Solución**: Creamos la función *simular.bino*()

```r
simular.bino <- function(samples = 10000, n =20, pi = 0.5)
{
```

```
  valores <- sample(c(0,1), samples * n, replace = TRUE,
                      prob = c(1 - pi, pi))
  matr.valores <- matrix(valores, ncol = n)
  exitos <- apply(matr.valores, 1, sum) # Apply aplica una función a una matriz
  a1 <- round((table(exitos)/samples), 3)
  b1 <- round(dbinom(0:n, n, pi), 3)
  names(b1) <- 0:n # Obtener el nombre de los objetos de la tabla
  hist(exitos, breaks = c((-0.5+0):(n+0.5)), freq = FALSE,
       ylab = "", col = 13, ylim = c(0, max(a1, b1)),
       main = "Valores teóricos superpuestos
       sobre el histograma de valores simulados")
  x <- 0:n
  fx <- dbinom(x, n, pi)
  lines(x, fx, type = "h")
  lines(x, fx, type = "p", pch = 16)
  list(distribucion.simulada = a1, distribucion.teorica = b1)
}
```

b. **Solución**

```
set.seed(31)
simular.bino(samples = 1000, n = 5, pi = 0.5)
```




```
## $distribucion.simulada
```

```
## exitos
##     0     1     2     3     4     5
## 0.023 0.174 0.311 0.308 0.153 0.031
##
## $distribucion.teorica
##     0     1     2     3     4     5
## 0.031 0.156 0.312 0.312 0.156 0.031
```

En la figura anterior tenemos un histograma de 1000 muestras simuladas donde \$n $= 5$ \$ y $p = 0.5$ superpuesta sobre la distribución teorica para una variable aletatoria que sigue una distribución $b(n = 5, p = 0.5)$.

Distribución de Poisson

Loa experimentos que dan valores números de una variable aleatoria X, el número de resultados que ocurren durante un intervalo dado o en un región específica, se laman **experimentos de Poisson**. El intervalo dado puede tener cualquier longitud, como un minuto, un día, una semana, un mes, o incluso un año. Por lo tanto, podemos representar mediante un proceso de Poisson un variable aleatoria X que represente el número de llamadas telefonicas reportando casos de Covid-19 por durante una hora en un hospital, el número de días que una escuela está cerrada por la temporada de invierno o el número de juegos de la futbol que están suspendidos debido a la lluvia.

Un experimento de Poisson se deriva del **proceso de Poisson** y posee las siguientes propiedades: 1. El número de resultados que ocurre en un intervalo o región específica es independiente del numero que ocurre en cualquier otro intervalo o región del espacio disyunto. 2. La probabilidad de que ocurra un solo resultado en un intervalo muy corto es proporcional a la longitud del intervalo o al tamaña de la región y ni depende del número de resultados que ocurren fuera de este intervalo. 3. La probabilidad de que ocurra más de un resultado en un intervalo corto o que caiga en tal región pequeña es insignificante.

Definición: La distribución de probabilidad de la variable aleatoria de Poisson X, que representa el número de resultados que ocurren en un intervalo dado o región específica que se denota con t, es

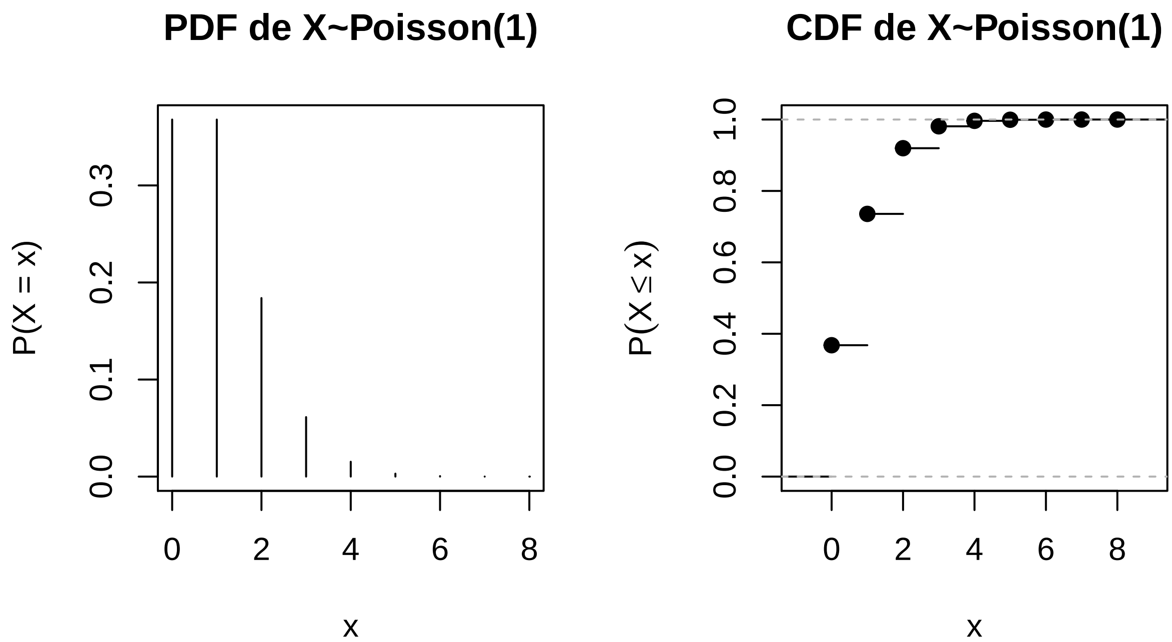
$$p(x; \lambda t) = \frac{e^{-\lambda t}(\lambda t)^x}{x!}, \quad x = 0, 1, 2, ...,$$

donde λ es el número promedio de resultados por unidad de tiempo o región y $e = 2.71828...$

Teorema. La media y la varianza de la distribución de Poisson $p(x; \lambda t)$ tiene el valor λt.

El siguiente código representa una función de densidad y una función de distribución acumulada para una distribución $p(\lambda = 1)$

```r
opar <- par(no.readonly = TRUE)
par(mfrow = c(1, 2), pty = "s")
x <- 0:8
px <- dpois(x, 1)
plot(x, px, type = "h", xlab = "x", ylab = "P(X = x)",
    main = "PDF de X~Poisson(1)")
xs <- rep(0:8, round(dpois(0:8, 1)*100000, 0))
plot(ecdf(xs), main = "CDF de X~Poisson(1)",
    ylab = expression(P(X<=x)), xlab = "x" )
```




```r
par(opar)
```

Ejemplo. En un proceso de fabricación donde se manufacturan productos de vidrio ocurren defectos o burbujas, lo que deja ocasionalmente a la pieza indeseada para su venta. Se sabe que, en promedio, un de cada 1000 casos de estos artículos que se producen tiene una o más burbujas. ¿Cuál es la probabilidad de que una muestra aleatoria de 800 tenga menos de siete articulos con burbujas?

Solución. Éste es en esencia un experimento binomial con $n = 800$ y $p = 0.001$. Como p es muy cercano a cero y n es bastante grande, haremos la aproximación con la distribución de Poisson utilizado

$$\mu = (800)(0.001) = 8.$$

De aquí, si X representa el número de burbujas, tenemos

$$P(X < 7) = \sum_{x=0}^{6} b(x; 8000, 0.001) = \sum_{x=0}^{6} p(x; 8) = 0.3134.$$

Ejemplo. Las llamadas telefonicas a el número local 911 son conocidas por seguir una distribución de Poisson con una media de dos llamadas por minuto. Calcular las probabilidad que:

a. Habrán cero llamadas durante un periodo de un minuto.
b. Habrán menos de cinco llamadas durante un periodo de un minuto.
c. Habrán menos de seis llamadas en una hora.

Solución a. $P(X = 0/\lambda = 2) = \frac{e^{-2}(2)^0}{0!} = 0.1353$

```r
dpois(x = 0, lambda = 2)
```

```
## [1] 0.1353353
```

b. Notar que $P(X < 5) = P(X \leq 4)$. $P(X \leq 4/\lambda = 2) = \sum_{x=0}^{4} \frac{e^{-2}(2)^x}{x!} = e^{-2}1 + 2 + \frac{2^2}{2!} + \frac{2^3}{3!} + \frac{2^4}{4!} = 0.9473$

```r
ppois(q = 4, lambda = 2)
```

```
## [1] 0.947347
```

c. NOtar que el periodo de tiempo cambia de un minuto a una hora (60 minutos). Consecuentemente, el número promedio de llamadas en una hora es $2 * (60) = 120$ $(X \leq 5/\lambda = 120$. $P(X \leq 4/\lambda = 2) = \sum_{x=0}^{5} \frac{e^{-120}(120)^x}{x!} = 0$

```r
ppois(q = 5, lambda = 120)
```

```
## [1] 1.658476e-44
```

Distribución Geométrica

La Distribución geométrica, como la distribución binomial, está basada en el proceso de Bernoulli; sin embargo, ésta no fija el número de intentos a priori al experimento. La distribución geométrica calcula la probabilidad que el primer éxito ocurra después de n fallos en vez de calcular la probabilidad de obsevar x exitos en n intentos.

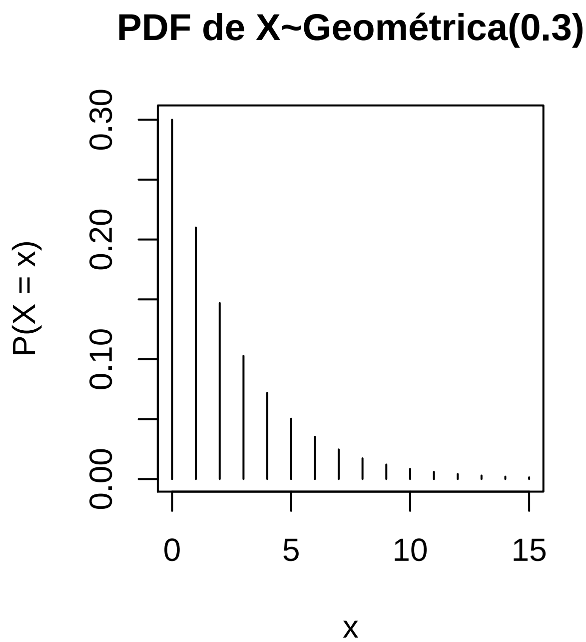
Definición. Si pruebas independientes repetidas pueden tener como resultado un éxito con probabilidad p y un fracaso con probabilidad $q = 1 - p$, entonces la distribución de la variable aleatoria X, el número de las prueba en el que ocurre el primer éxito, es

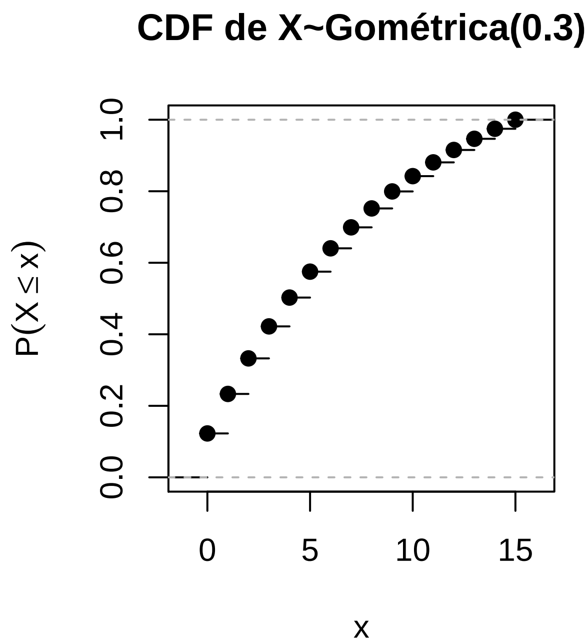
$$g(x; p) = pq^{x-1}, \quad x = 1, 2, 3...,$$

Teorema. La media y la varianza de una variable aleatoria que sigue una distribución geométrica son

$$\mu = \frac{1}{p}, \quad \sigma^2 = \frac{1-p}{p^2}.$$

```r
opar <- par(no.readonly = TRUE)
par(mfrow = c(1, 2), pty = "s")
x <- 0:15
px <- dgeom(x, 0.3)
plot(x, px, type = "h", xlab = "x", ylab = "P(X = x)",
    main = "PDF de X~Geométrica(0.3)")
xs <- rep(0:15, round(dgeom(0:15, 0.1)*100000, 0))
plot(ecdf(xs), main = "CDF de X~Gométrica(0.3)",
    ylab = expression(P(X<=x)), xlab = "x" )
```






```r
par(opar)
```

Distribución Binomial negativa

Definición. Si pruebas independientes repetidas pueden tener como resultado un éxtio con probabilidad p y un fracaso con probabilidad $1-p$, entonces la disgribcuión de la vartiable aleatoria X, el número de la prueba en la que ocurre el *k-ésimo* éxito, es

$$b*(x;k,p) = \binom{x-1}{k-1} p^k q^{x-k}, \quad x = k, k+1, k+2, ...,$$

Ejemplo. Encuentre la probabilidad de que una persona que lanza tres monedas obtenga sólo caras o sólo cruces por seguna vex en el quinto lanzamiento.

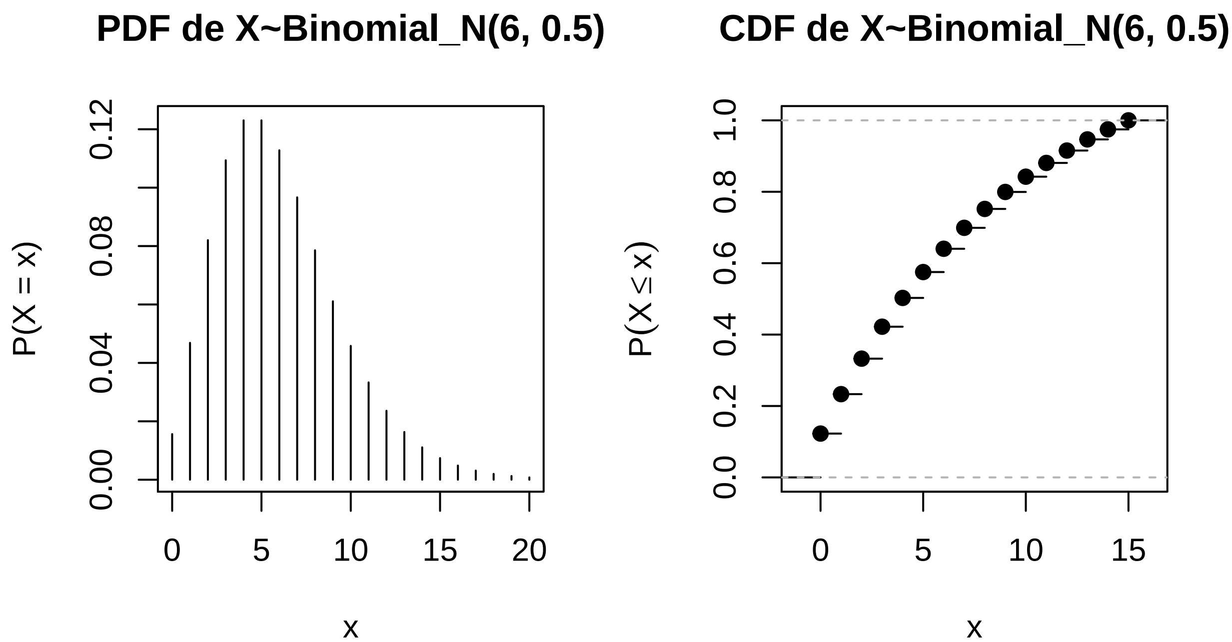
Solución. AL utilizar la distribución binomial negativa con $x=5$, $k=2$ y $p=1/4$, tenemos

$$b^*\left(5;2,\frac{1}{4}\right) = \binom{4}{1}\binom{1}{4}^2\binom{3}{4}^3 = \binom{3}{4}^3 = \frac{4!}{1!3!} * \frac{3^3}{4^5} = \frac{27}{256}.$$

Media y varianza de las distribución binomial negativa

$$\mu = k\frac{1}{p}, \quad \sigma^2 = k\frac{1}{p^2}.$$

```r
opar <- par(no.readonly = TRUE)
par(mfrow = c(1, 2), pty = "s")
x <- 0:20
px <- dnbinom(x, size = 6, 0.5)
plot(x, px, type = "h", xlab = "x", ylab = "P(X = x)",
    main = "PDF de X~Binomial_N(6, 0.5)")
#xs <- rep(0:15, round(dnbinom(0:20, size = 6, 0.5)*100000, 0))
plot(ecdf(xs), main = "CDF de X~Binomial_N(6, 0.5)",
    ylab = expression(P(X<=x)), xlab = "x" )
```




```r
par(opar)
```