

# Control Estadístico de Procesos

Gustavo Ahumada

## Variables aleatorias discretas y continuas

### Variable aleatoria discreta

Una variable aleatoria discreta asume cada uno de sus valores cierta probabilidad. Cuando dos dados son lanzados, la probabilidad de que su suma sea 7, escrita  $f(x) = P(X = x)$ , es igual  $1/6$ . La función que asigna la probabilidad a los valores de las variables aleatorias es llamada **función de densidad de probabilidad (pdf)** o **distribución de probabilidad**. Toda pdf debe satisfacer las siguientes condiciones:

1.  $f(x) \geq 0$ .
2.  $\sum_x f(x) = 1$ .
3.  $P(X = x) = f(x)$ .

Hay muchos problemas en los cuales deseamos calcular la probabilidad de que el valor observado de una variable aleatoria  $X$  sea mayor o igual que algún número real  $x$ . Al escribir  $F(X) = P(X \leq x)$  para cualquier número real  $x$ , definimos la **distribución acumulada (cdf)**  $F(X)$  de una variable aleatoria discreta  $X$  con distribución  $f(x)$ . Se define como:

$$F(X) = P(X \leq x) = \sum_{t \leq x} f(t) \text{ para } -\infty < x < \infty.$$

Tiene las siguientes propiedades:

1.  $0 \leq F(X) \leq 1$ .
2. Si  $a < b$ , entonces  $F(a) < F(b)$  para cualquier números reales  $a$  y  $b$ . Por lo tanto,  $F(X)$  es una función no decreciente de  $x$ .
3.  $\lim_{x \rightarrow \infty} F(X) = 1$ .
4.  $\lim_{x \rightarrow -\infty} F(X) = 0$ .

**Ejemplo 1** Lanzar una moneda equilibrada tres veces y sea la variable aleatoria  $X$  el número de caras observadas en los tres intentos.

**Solución** El espacio muestral para el experimento es:

$$S = \{TTT, HTT, THT, HHT, TTH, HTH, THH, HHH\}$$

La variable aleatoria  $X$  puede tomar los valores de 0, 1, 2, y 3 con probabilidades  $\frac{1}{8}$ ,  $\frac{3}{8}$ ,  $\frac{3}{8}$ , y  $\frac{1}{8}$ , respectivamente. Definir la cdf para  $X$ ,  $F(x) = P(X \leq x)$  como sigue:

$$f(n) = \begin{cases} 0 & \text{si } x < 0, \\ 1/8 & \text{si } 0 \leq x < 1, \\ 4/8 & \text{si } 1 \leq x < 2, \\ 7/8 & \text{si } 2 \leq x < 3, \\ 1 & \text{si } x \geq 3 \end{cases}$$

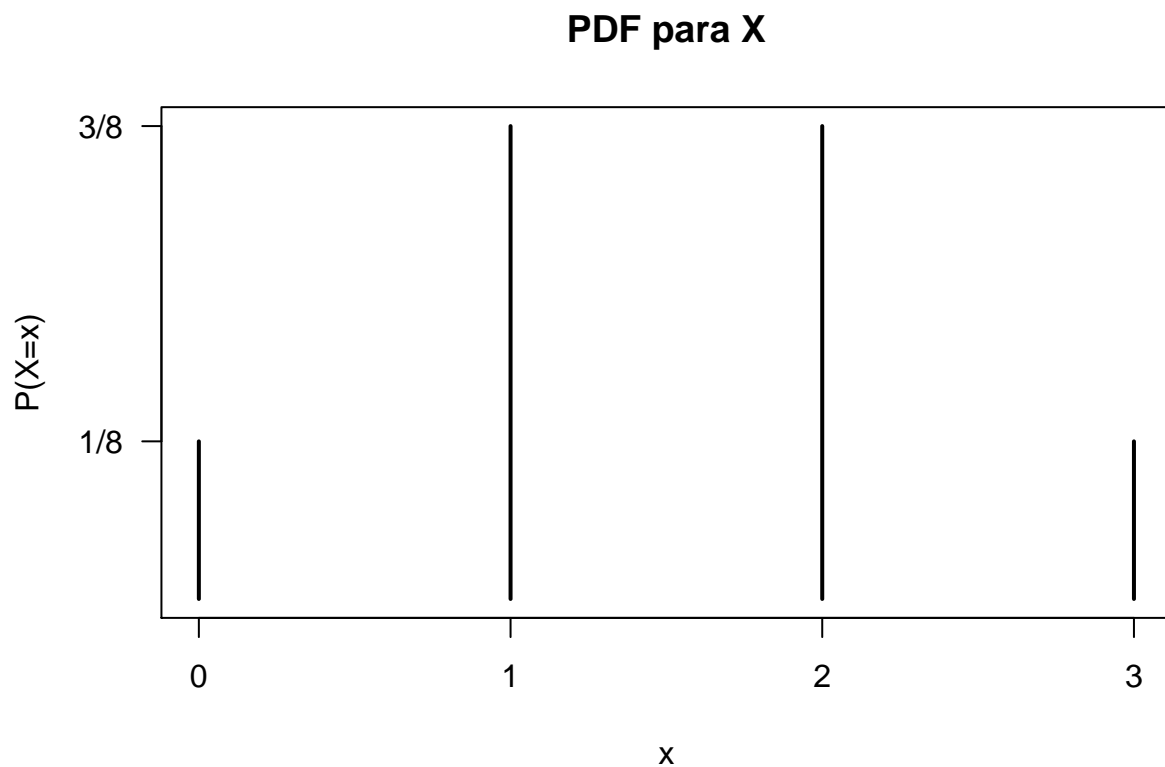
```
opar <- par(no.readonly = TRUE)
library(MASS)
par(mfrow=c(1,2), pty = "s")
S <- expand.grid(moneda1 = 0:1, moneda2 = 0:1, moneda3 = 0:1)
n.heads <- apply(S, 1, sum)
cbind(S, n.heads)
```

```
##   moneda1 moneda2 moneda3 n.heads
## 1      0      0      0      0
## 2      1      0      0      1
## 3      0      1      0      1
## 4      1      1      0      2
## 5      0      0      1      1
## 6      1      0      1      2
## 7      0      1      1      2
## 8      1      1      1      3
```

```
T1 <- table(n.heads)/length(n.heads)
fractions(T1)
```

```
## n.heads
##  0  1  2  3
## 1/8 3/8 3/8 1/8
```

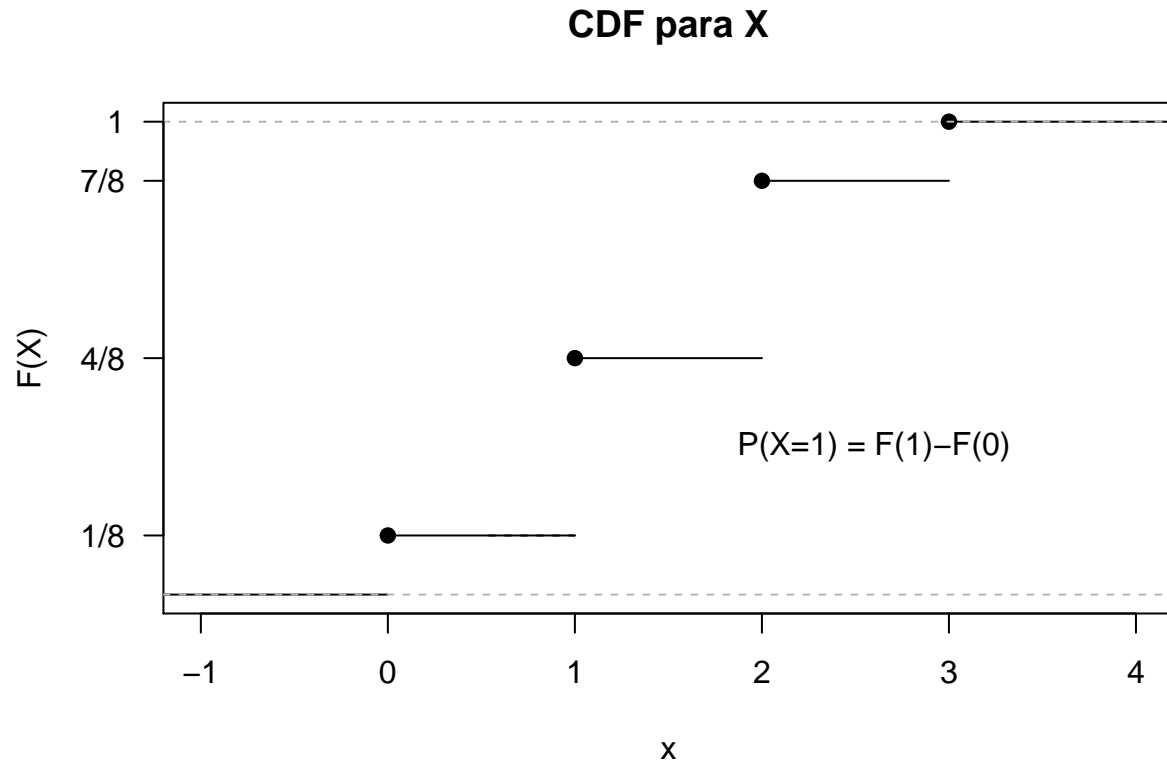
```
plot(T1, xlab = "x", ylab = "P(X=x)", yaxt = "n", main = "PDF para X") +
axis(2, at =c(1/8, 3/8), labels = c("1/8", "3/8"), las = 1)
```



```
## [1] 0.125 1.375 2.125 3.375
```

```
plot(ecdf(n.heads), main = "CDF para X", ylab = "F(X)", xlab = "x",
     yaxt = "n") +
```

```
axis(2, at = c(1/8, 4/8, 7/8, 1), labels = c("1/8", "4/8", "7/8", "1"),
     las = 1) +
segments(1, 1/8, 4/8, lty = 2) +
text(2.6, 2.5/8, "P(X=1) = F(1)-F(0)")
```



```
## numeric(0)
```

### Valor esperado

Una de las más importantes ideas acerca de resumir la información provista en una **pdf** es el valor esperado. Dada una variable aleatoria  $X$  con una **pdf**  $f(x)$ , el **valor esperado** de la variable aleatoria  $X$ , escrito  $E(X)$ , es

$$\mu = E(X) = \sum_x x f(x)$$

$E(X)$  puede ser denotada como  $\mu$ , porque  $E(X)$  es la media de una variable aleatoria  $X$ .

**Ejemplo.** Un inspector de calidad muestreo un lote que contiene siete componentes; el lote contiene cuatro componentes buenos y tres defectuosos. El inspector toma una muestra de tres componentes. Encuentre el valor esperado del número de componentes buenos en esta muestra.

**Solución.** Sea  $X$  el número de componentes buenos en la muestra. La distribución de probabilidad de  $X$  es

$$f(x) = \frac{\binom{4}{x} \binom{3}{3-x}}{\binom{7}{3}}, \quad x = 0, 1, 2, 3.$$

Unos simples cálculos dan  $f(0) = 1/35, f(1) = 12/35, f(2) = 18/35, f(3) = 4/35$ . Por lo tanto,

$$E(X) = (0)\binom{1}{35} + (1)\binom{12}{35} + (2)\binom{18}{35} + (3)\binom{4}{35} = \frac{12}{7} = 1.7.$$

```
x <- c(0,1,2,3)
px <- c(0.029,0.343,0.514,0.114)
EX <- sum(x*px)
MP <- weighted.mean(x, px)
c(EX, MP)
```

```
## [1] 1.713 1.713
```

## Varianza

El segundo momento de una alrededor de la media es llamado la **varianza** de la distribución de  $X$ , o simplemente la varianza de  $X$ :

$$\sigma_x^2 = E[(X - \mu)]^2 = \sum_x (x - \mu)^2 f(x)$$

La raíz cuadrada de la varianza,  $\sigma$ , se llama **desviación estándar** de  $X$ .

## Variable aleatoria continua

Una variable aleatoria se llama **variable aleatoria continua** si puede tomar valores en una escala continua. A menudo los posibles valores de una variable continua son precisamente los mismos valores el espacio muestral continuo. Por ejemplo, la medición de la distancia que cierta marca de automóvil recorre en una pista de prueba con 5 litros de gasolina.

Una variable aleatoria continua tiene una probabilidad cero de tomar *exactamente* cualquiera de sus valores. Por lo tanto, su distribución de probabilidad no puede dar una forma tabular. Este tipo de variables aleatorias se caracterizan por:

- Toman valores en un intervalo (abierto o cerrado), es decir en un continuo.
- No tiene sentido calcular la probabilidad de que la v.a. adopte un valor particular, ya que es cero.
- Pero calculamos probabilidades en intervalos.
- Estas probabilidades son siempre áreas bajo curvas, estas curvas son las funciones de densidad.

La función de densidad de una variable aleatoria es una función que determinará una curva bajo la cual se calcularán las áreas = probabilidades en un intervalo.

**Definición:** Sea  $X$  una variable aleatoria continua  $x$  un número perteneciente al rango posible de valores de  $X$ , la función  $f(x)$  es una **función de densidad de probabilidad** definida en el conjunto de los números reales  $\mathbb{R}$ , si

1.  $f(x) \geq 0$ , para todo  $x \in \mathbb{R}$ .
2.  $\int_{-\infty}^{\infty} f(x)dx = 1$ .
3.  $P(a < X < b) = \int_a^b f(x)dx$ .

**Definición:** La **distribución acumulada (cdf)**  $F(X)$  de una variable aleatoria  $X$  con función de densidad  $f(x)$  es

$$F(x) = P(X \leq x) = \int_{-\infty}^{\infty} f(t)dt \quad \text{para} \quad -\infty < x < \infty$$

Dada la anterior definición, Siempre se pueden calcular probabilidades en un intervalo, y hay, por lo tanto dos maneras de hacerlo que son equivalentes: Sean dos valores,  $a < b$

$$P(a < X < b) = \int_a^b f(x)dx \quad \text{para} \quad -\infty < x < \infty$$

$$P(a < X < b) = F_x(b) - F_x(a) \quad \text{para} \quad -\infty < x < \infty$$

Es decir, directamente integrando la función de densidad o bien utilizando la cdf.

**Ejemplo.** Suponer que  $X$  es una variable aleatoria continua con **pdf**  $f(x)$ , donde

$$f(x) = \begin{cases} c(1-x^2) & \text{si } -1 < x \leq 1, \\ 0 & \text{en otros casos.} \end{cases}$$

- Encontrar la constante  $c$  que  $f(x)$  es una **pdf** de la variable aleatoria  $X$ .
- Encontrar la **cdf** para  $X$ .
- Calcular  $P(-0.5 \leq X \leq 1)$ .
- Gráficar la **pdf** y **cdf**.

**Solución a.** Empleando la propiedad 2 de la definición de **pdf**

$$1 = \int_{-\infty}^{\infty} f(x)dx$$

$$\begin{aligned} 1 &= c \left( x - \frac{x^3}{3} \right) \Big|_0^1 \\ &= c \left( 1 - \frac{1}{3} - 0 - 0 \right) \\ &= c \left( \frac{2}{3} \right) \\ c &= \frac{3}{2} \end{aligned}$$

b. Empleando la definición de **cdf** se verifica que

$$F(x) = \begin{cases} 0 & \text{si } x \leq -1, \\ \int_{-1}^x \frac{3}{2}(1-t^2)dt = \frac{-x^3}{4} + \frac{3x}{4} + \frac{1}{2} & \text{si } -1 < x \leq 1, \\ 1 & \text{si } x > 1. \end{cases}$$

c. Empleando la definición de distribución acumulada para la **pdf** de una variable continua, tenemos

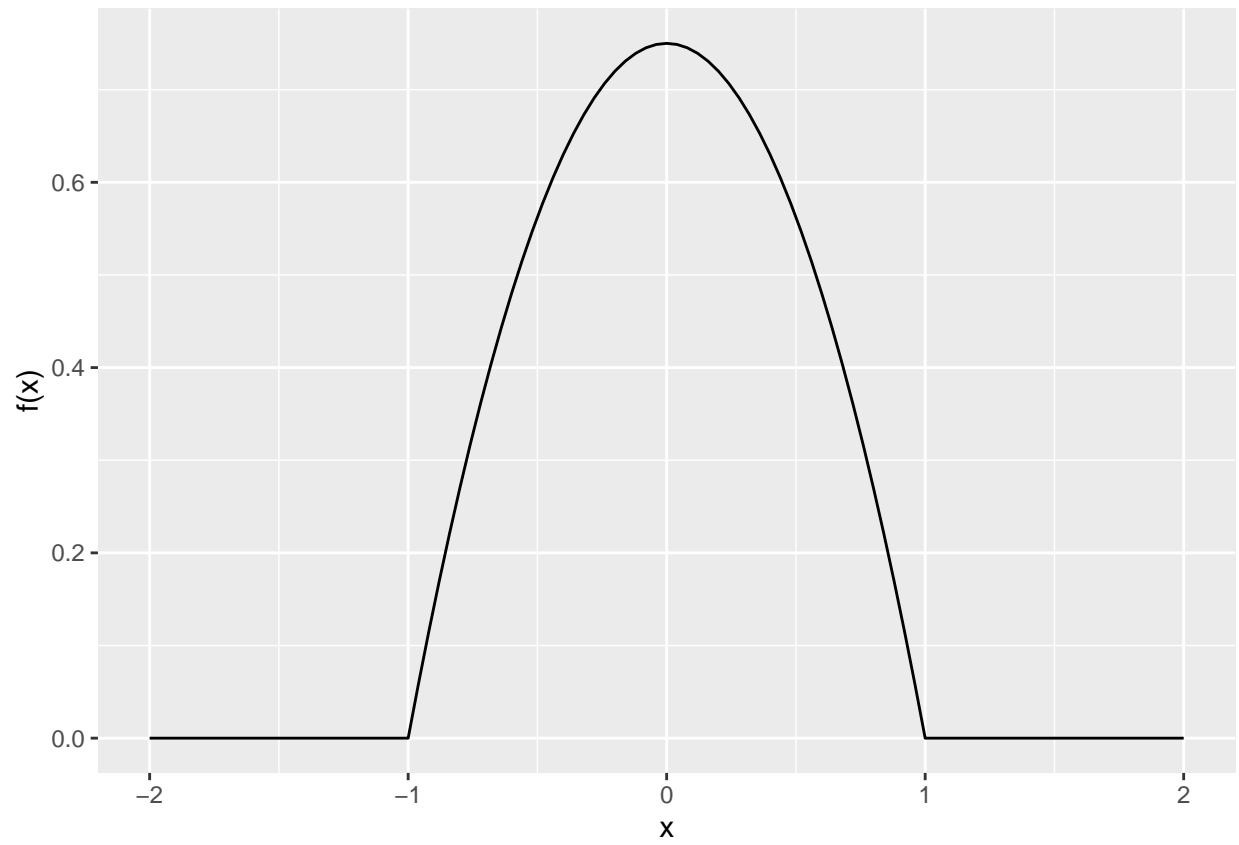
$$\begin{aligned}
P(-0.5 \leq X \leq 1) &= F(1) - F(-0.5) \\
&= \left( \frac{-1^3}{4} + \frac{3 * 1}{4} + \frac{1}{2} \right) - \left( -\frac{-0.5^3}{4} + \frac{3 * -0.5}{4} + \frac{1}{2} \right) \\
&= c \left( \frac{2}{3} - \frac{-2}{3} \right) \\
c &= \left( \frac{-1}{4} + \frac{3}{4} + \frac{1}{2} \right) - \left( \frac{1}{32} + \frac{-3}{8} + \frac{1}{2} \right) \\
&= c \left( \frac{2}{3} - \frac{-2}{3} \right) \\
&= 1 - \frac{5}{32} = \frac{27}{32} = 0.84375
\end{aligned}$$

d. El siguiente código genera la **pdf** y la **cdf** de  $X$ .

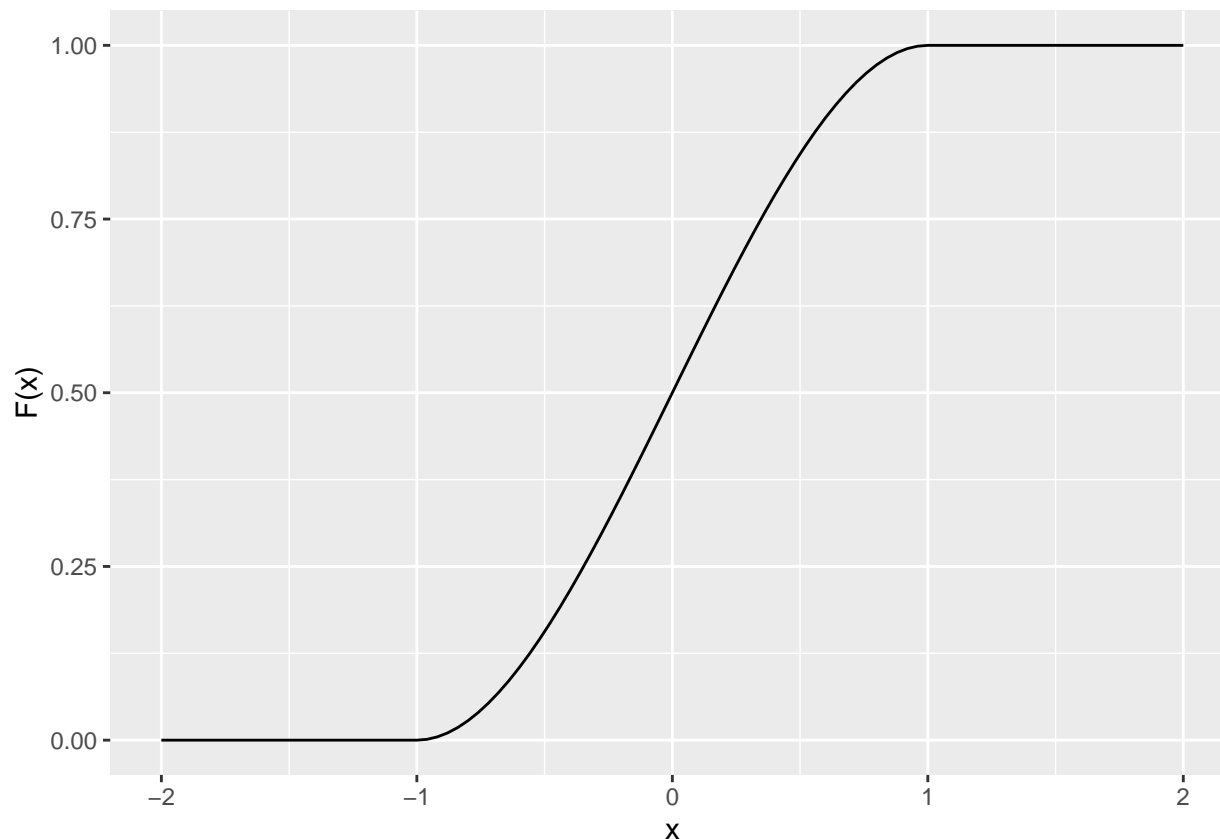
```
f <- function(x) {
  y <- 3/4 * (1 - x^2)
  y[x < -1 | x > 1] <- 0
  return(y)
}
```

```
F <- function(x) {
  y <- -x^3/4 + 3 * x/4 + 1/2
  y[x <= -1] <- 0
  y[x > 1] <- 1
  return(y)
}
```

```
library(ggplot2)
graf <- ggplot(data.frame(x = c(-2, 2)), aes(x = x))
graf + stat_function(fun = f) + labs(x = "x", y = "f(x)", title = "PDF para X")
```



```
graf + stat_function(fun = F) + labs(x = "x", y = "F(x)", tittle = "CDF para X")
```



### Valor esperado

Para variables aleatorias continuas, la definición asociada con la esperanza de una variable aleatoria  $X$  o una función, digase  $g(x)$ , de  $X$  son iguales para las variables aleatorias discretas, excepto que las sumatorias son reemplazadas con integrales. El **valor esperado** de una variable aleatoria continua  $X$  es

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx.$$

### Varianza

$$Var(X) = \sigma_x^2 = E[(X - \mu)]^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx.$$

**Ejemplo.** Dada la función

$$f(x) = k, \quad -1 < x < 1$$

de la variable aleatoria  $X$ , a. Encontrar el valor de  $k$  para hacer  $f(x)$  una **pdf**. Emplear este  $k$  para la parte (c) y (d).

- b. Gráficar la **pdf** con ggplot2.
- c. Encontrar la media de la distribución
- d. Encontrar la varianza de la distribución.



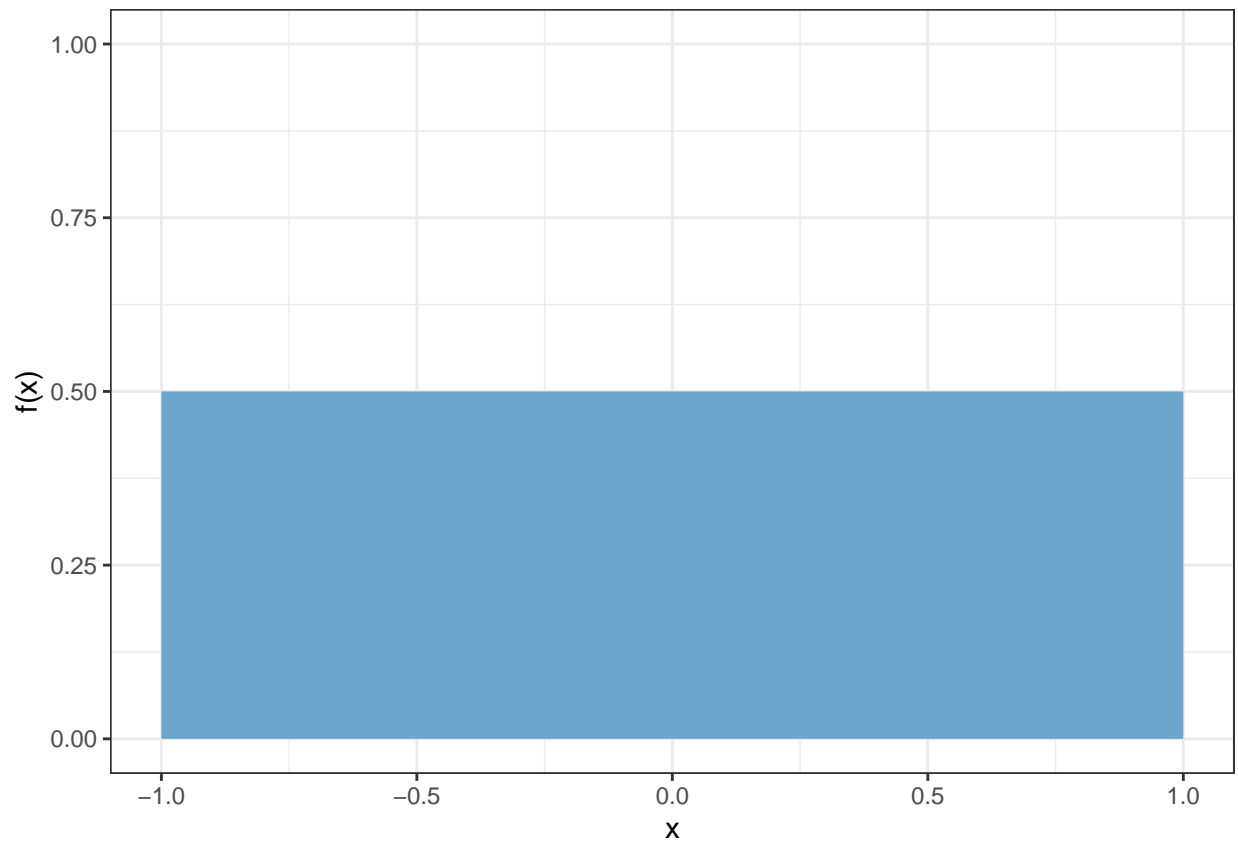
## Solución

a. Desde  $\int_{-\infty}^{\infty} x f(x) dx$  debería ser 1 para que  $f(x)$  sea una **pdf**, establecemos  $\int_{-1}^1 k dx = 1$

$$\int_{-1}^1 k dx = 1$$
$$2k = 1; \quad k = \frac{1}{2}$$

b. Código para el gráfico

```
x <- seq(-1, 1, length = 500)
y <- dunif(x, -1, 1)
DF <- data.frame(fx = y)
previous_theme <- theme_set(theme_bw())
ggplot(data = DF, aes(x = x, y = fx)) +
  geom_area(fill = "skyblue3") +
  labs(x = "x", y = "f(x)") +
  ylim(c(0, 1)) +
  theme_set(previous_theme)
```



## Distribuciones de probabilidad univariada

En esta sección examinamos las distribuciones de probabilidad univariadas que son empleadas frecuentemente para modelar fenómenos aleatorios. Primero, introducimos las distribuciones de probabilidad discreta, seguido por las distribuciones de probabilidad continuas. Las distribuciones 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 univariadas 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 aleatoria  $X$  toma valores  $x_1, x_2, \dots, x_n$ , con idénticas probabilidades, entonces la distribución uniforme discreta está dada por

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

**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 valores  $S = \{40, 60, 75, 100, 120\}$ . La función de densidad de probabilidad para la variable aleatoria  $X$  es

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

Los cálculos aritméticos son realizados con el siguiente código

```
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 de 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 posibles resultados. Los resultados son mutuamente excluyentes y exhaustivos; por ejemplo, éxito o fracaso, cierto o falso, vivo o muerto, hombre o mujer, etc. Una variable aleatoria Bernoulli,  $X$ , puede tomar dos valores, donde  $X(\text{éxito}) = 1$  y  $X(\text{fracaso}) = 0$ . La probabilidad que  $X$  sea un éxito es  $p$ , y la probabilidad 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 propiedades: 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 depende del número de pruebas y de la probabilidad de éxito en una prueba dada.

**Definición.** Un experimento de Bernoulli puede tener 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 número de éxitos en  $n$  pruebas independientes, es

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

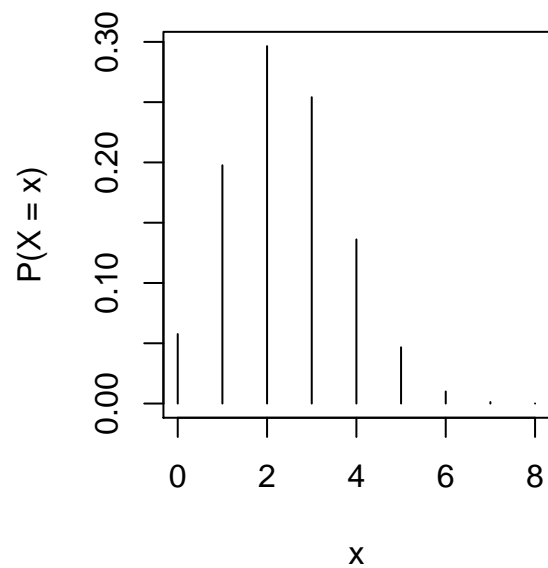
**Teorema.** La media y la varianza de la distribución binomial  $b(x; n, p)$  son

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

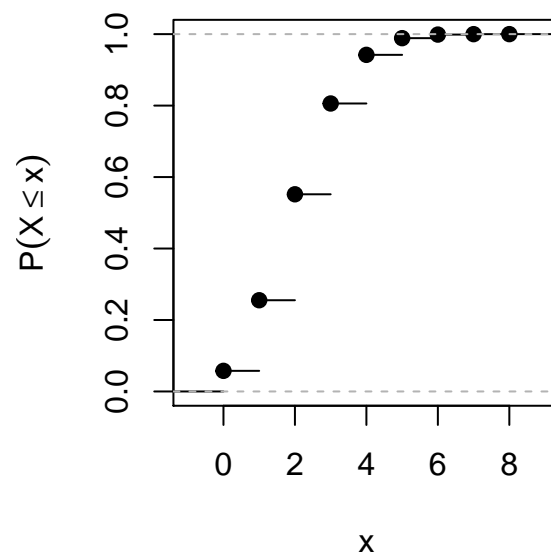
El código para crear las gráficas que representen la **pdf** y la **cdf** para una variable aleatoria binomial  $b(x; 8, 0.3)$  se presenta a continuación

```
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" )
```

**PDF de  $X \sim \text{Binomial}(8, 0.3)$**



**CDF de  $X \sim \text{Binomial}(8, 0.3)$**



`par(opar)`