Control Estadístico de Procesos

Gustavo Ahumada

Variables aleatorias discretas y continuas

Variable aleatoria discreta

Una variable aletatoria discreta asume cada uno de sus valorescon 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:

$F(X) = P(X \leq x) = \sum_{t \leq x} f(x)$ 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. $lim_{x \to \infty} F(X) = 1$.
4. $lim_{x \to -\infty} F(X) = 0$.

Ejemplo 1 Lanzar una moneda equilibrada tres veces y sea las 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:

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

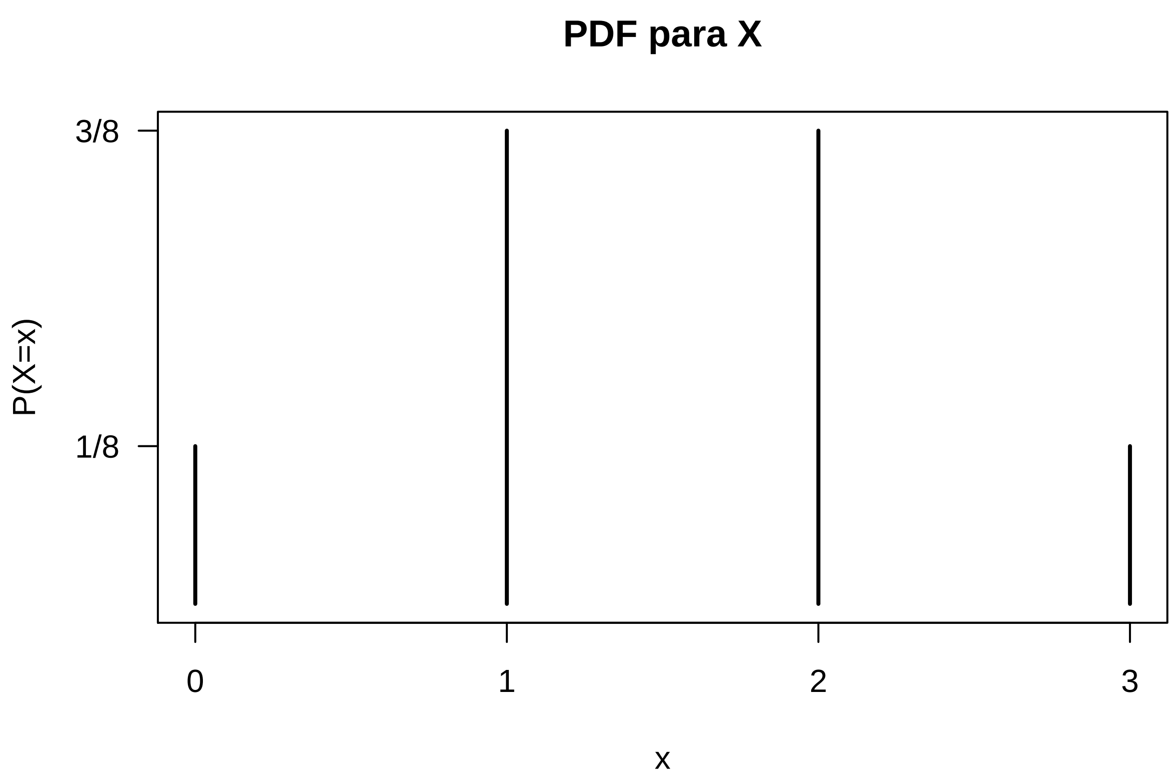
```
opar <- par(no.readonly = TRUE)
library(MASS)
par(mfrow=c(1,2), pty = "s")
S <- expand.grid(moneda1 = 0:1, moneda2 = 0:1, moneda3 = 0:1)
n.heads <- apply(S, 1, sum)
cbind(S, n.heads)
```

```
##   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
```

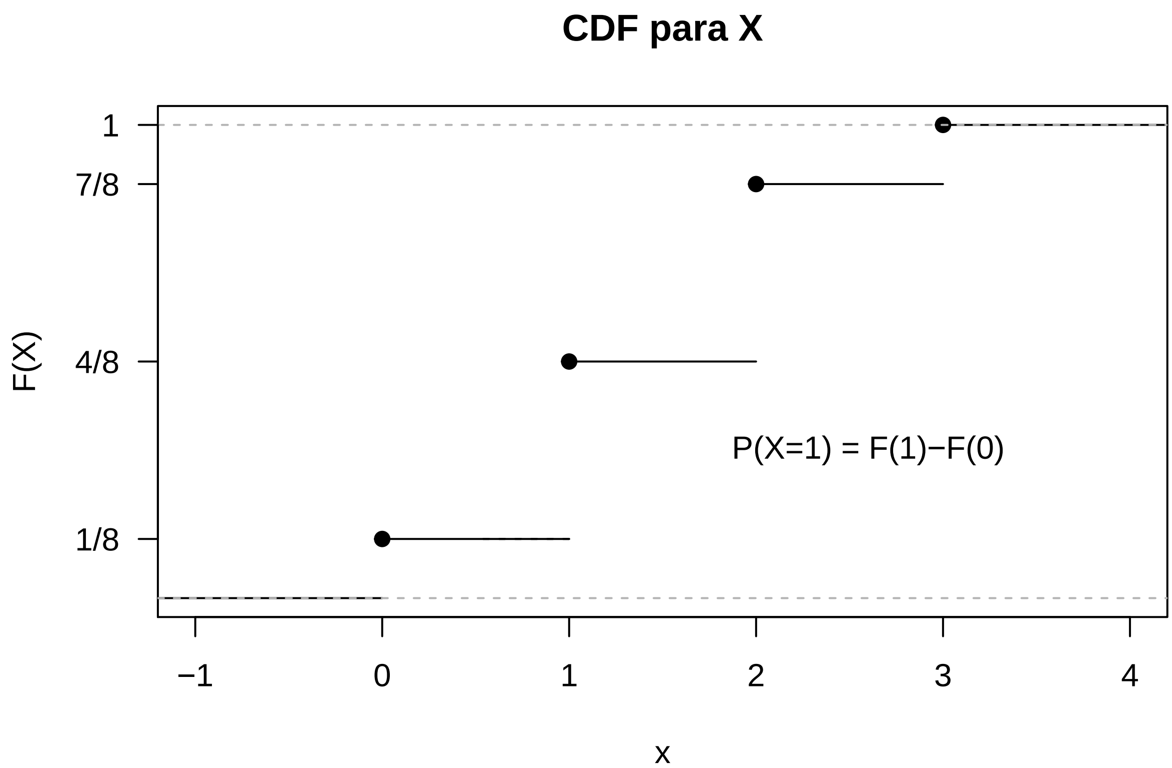
```
plot(T1, xlab = "x", ylab = "P(X=x)", yaxt = "n", main = "PDF para X") +
axis(2, at =c(1/8, 3/8), labels = c("1/8", "3/8"), las = 1)
```




```
## [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)
```

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 · reas 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 dustribución acumulada (cdf)** $F(X)$ de una variable aleatoria X con función de densidad $f(x)$ es

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

.

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, $a < b$

$$P(a < X > b) = \int_a^b f(x)dx \quad para \quad -\infty < x < \infty$$
$$P(a < X > b) = F_x(b) - F_x(a) \quad para \quad -\infty < x < \infty$$

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

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

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

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

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

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

$$\begin{aligned} 1 &= c\left(x - \frac{x^3}{3})\right|_0^1 \\ &= c\left(1 - \frac{1}{3} - -1 - \frac{-1}{3}\right) \\ &= c\left(\frac{2}{3} - \frac{-2}{3}\right) \\ c &= \frac{3}{4} \end{aligned}$$

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

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

.

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

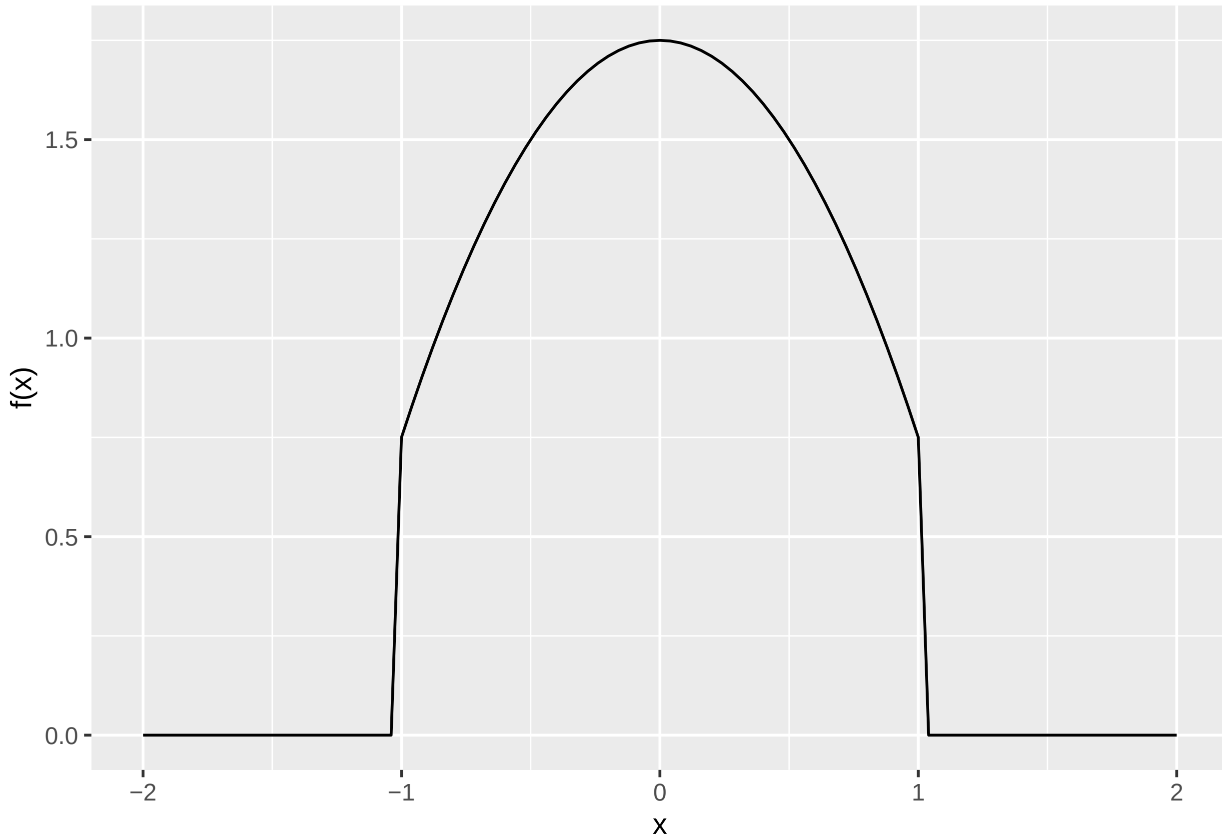
$$\begin{aligned} P(-0.5 \leq X \leq 1) &= F(1) - F(-0.5) \\ &= \left(\frac{-1^3}{4} + \frac{3*1}{4} + \frac{1}{2}\right) - \left(-\frac{-0.5^3}{4} + \frac{3*-0.5}{4} + \frac{1}{2}\right) \\ &= c\left(\frac{2}{3} - \frac{-2}{3}\right) \\ c &= \left(\frac{-1}{4} + \frac{3}{4} + \frac{1}{2}\right) - \left(\frac{1}{32}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

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

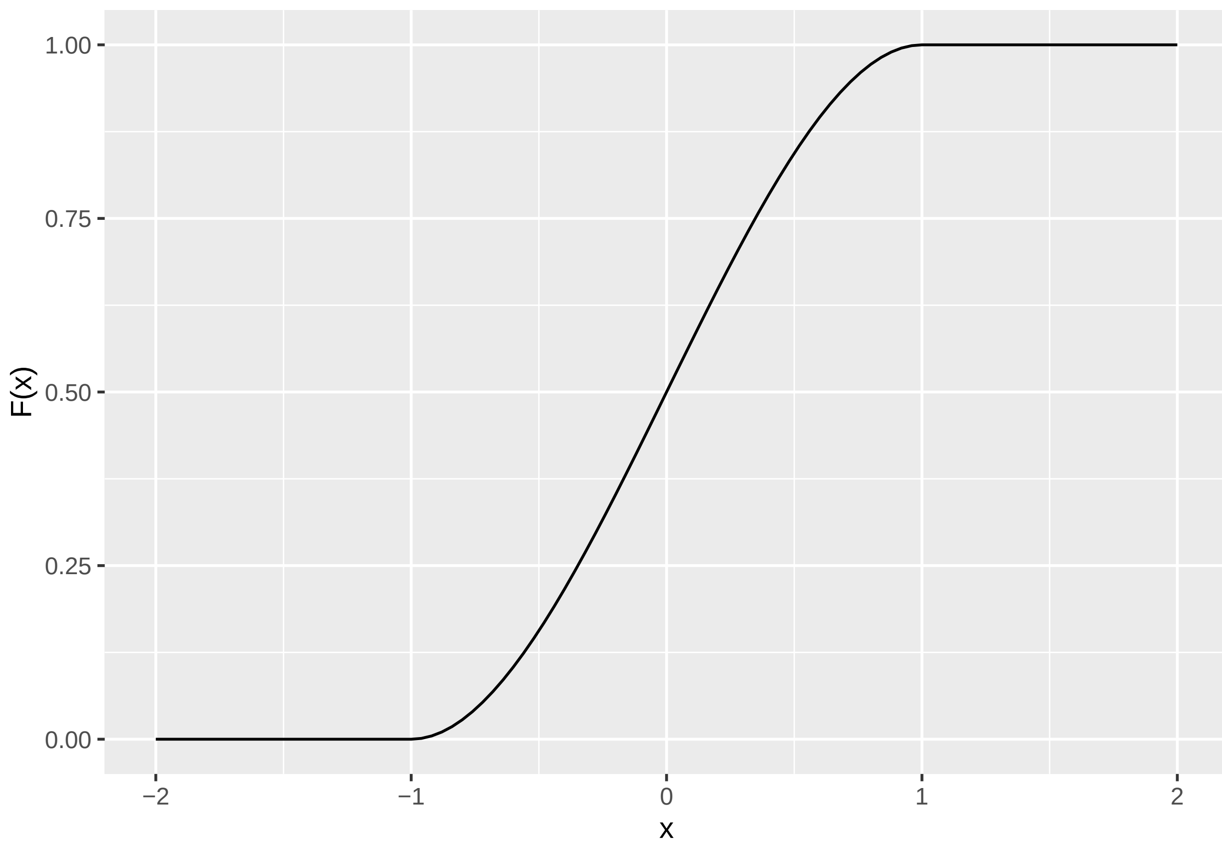
```
f <- function(x) {
  y <- 3/4 + (1 - x^2)
  y[x < -1 | x > 1] <- 0
  return(y)
}
```

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

```
library(ggplot2)
p1 <- ggplot(data.frame(x = c(-2, 2)), aes(x = x))
p1 +  stat_function(fun = f) + labs(x = "x", y = "f(x)", tittle = "PDF para X")
```




```
p2 <- ggplot(data.frame(x = c(-2, 2)), aes(x = x))
p2 + stat_function(fun = F) + labs(x = "x", y = "F(x)", tittle = "CDF para X")
```


1.00
0.75
0.50
0.25
0.00
F(x)
−2
−1
0
1
2
x