Capítulo 6

Convergencia (débil) de variables aleatorias

En esta última práctica vamos a vemos cómo podemos estudiar de forma empírica la convergencia débil de una sucesión de variables aleatorias. Empezaremos la práctica realizando un breve repaso sobre la convergencia débil y a continuación veremos cómo utilizar la simulación para tratar de intuir la posible convergencia débil de una sucesión de variables aleatorias.

6.1. Convergencia débil de variables aleatorias

Partimos de una sucesión de variables aleatorias que denotaremos por $\{X_n\}_n$, donde todas las X_n están definidas en el mismo espacio probabilístico y son independientes.
Definición: Una sucesión de variableas aleatorias $\{X_n\}$ independientes *convergence en ley* (o en distribución o en convergencia débil) a X si para todo punto x de continuidad de F_X se cumple que:

$$F_{X_n}(x) \longrightarrow F_X(x),$$

donde F_{X_n} y F_X denotan las funciones de distribución de X_n y X, respectivamente. Se denotará por $X_n \xrightarrow{\mathcal{D}} X$.

6.2. Uso de la simulación para el estudio de la convergencia débil

En ocasiones, el estudio analítico de la convergencia débil puede resultar muy tedioso. En esos casos, el uso de la simulación nos puede ayudar a hacernos una idea de la posible convergencia débil. Evidentemente, hay que recalcar que la simulación nunca da lugar a una demostración matemática, pero sí que nos puede ayudar a intuir el resultado.

A continuación, vamos a analizar dos ejemplos donde podemos analizar la convergencia débil utilizando la simulación.

Ejemplo 6.1. Consideremos la sucesión de variables aleatorias independientes $\{X_m\}_m$ donde el soporte de la variable X_m está formado por los puntos $\frac{k}{m}$, donde $k = 1, 2, \ldots, m$, y su probabilidad viene dada por $P(X_m = \frac{k}{m}) = \frac{1}{m}$. Es decir, X_m distribuye la probabilidad de manera uniforme sobre los puntos $\frac{k}{m}$. Por ejemplo, para los primeros valores de m, las variables X_m vienen dadas por:

$$\begin{array}{c|c} X_1 & 1 \\ \hline p_{X_1} & 1 \end{array} \qquad \begin{array}{c|cc} X_2 & \frac{1}{2} & 1 \\ \hline p_{X_2} & \frac{1}{2} & \frac{1}{2} \end{array} \qquad \begin{array}{c|ccc} X_3 & \frac{1}{3} & \frac{2}{3} & 1 \\ \hline p_{X_3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{array} \qquad \begin{array}{c|cccc} X_4 & \frac{1}{4} & \frac{1}{2} & \frac{3}{4} & 1 \\ \hline p_{X_4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{array}$$

De cara a estudiar la posible convergencia débil de $\{X_m\}$, vamos a simular valores de X_m para, por ejemplo, $m = 5$, $m = 10$, $m = 50$ y $m = 1000$, y dibujaremos en cada caso la función de distribución empírica. De cara a simular los valores de las variables, vamos a utilizar la instrucción `sample`, que permite simular valores aleatorios de una distribución discreta. En esta instrucción, tendremos que indicar, por este orden:

- Un vector con los valores que toma la variable. Para $m = 5$, sería `c(1/5,2/5,3/5,4/5,1)` o, más sencillo, `(1:5)/5`.
- El número de valores aleatorios deseados.
- Si queremos reemplazamiento (en tal caso, `replace=TRUE`) o no (en tal caso, `replace=FALSE`).
- Un vector con las probabilidades de cada valor que toma la variable. Para $m = 5$, sería `c(1/5,1/5,1/5,1/5,1/5)`. Sin embargo, por defecto supone que los puntos son equiprobables, como ocurre en este caso, así que podemos omitir este vector.

Comenzamos realizando la representación para $m = 5$ simulando 1000 valores:

```
n<-1000
x5<-sample((1:5)/5,n,replace=TRUE)
library(ggplot2)
g1<-ggplot(data=NULL, aes(x5))+
  stat_ecdf(col="red",size=1.2)+
  ylab("Función de distribución empírica")+
  xlab(TeX("Variable $x_5$"))
```

Con la segunda instrucción generamos un vector con 1000 valores que provienen de una distribución con valores en $\frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, 1$ y cada uno de esos valores es equiprobable. A continuación, representamos gráficamente la función de distribución de empírica como hemos hecho en prácticas previas.

Repetimos los pasos anteriores con $m = 10$, $m = 50$ y $m = 1000$:

```
x10<-sample((1:10)/10,n,replace=TRUE)
g2<-g1+
  stat_ecdf(aes(x10),col="blue",size=1.2)+
  ylab("Función de distribución empírica")+
  xlab(TeX("Variable $x_{10}$"))
x50<-sample((1:50)/50,n,replace=TRUE)
g3<-g2+
  stat_ecdf(aes(x50),col="green",size=1.2)+
  ylab("Función de distribución empírica")+
  xlab(TeX("Variable $x_{50}$"))
x1000<-sample((1:1000)/1000,n,replace=TRUE)
```

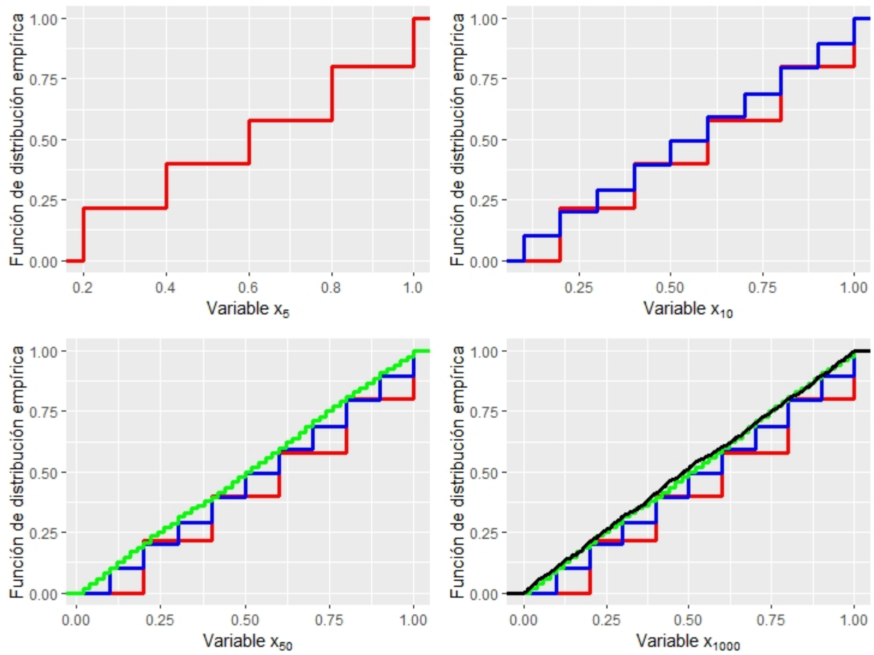

Figura 6.1: Representación gráfica de las funciones de distribución empírica en el Ejemplo 6.1.

```
g4<-g3+
  stat_ecdf(aes(x1000),col="black",size=1.2)+
  ylab("Función de distribución empírica")+
  xlab(TeX("Variable $x_{1000}$"))
```

De esta forma, obtenemos los gráficas mostradas en la Figura 6.1.

Como podemos ver, esas funciones de distribución cada vez se parecen más a la función de distribución $F(x) = x$ para $x \in [0, 1]$, que se corresponde con la función de distribución de una distribución $\mathcal{U}(0, 1)$. De hecho, se puede demostrar que, efectivamente, $\{X_m\} \xrightarrow{\mathcal{D}} \mathcal{U}(0, 1)$ (ver ejercicio 4 del tema de convergencias).

Ejemplo 6.2. Considera una sucesión de variables aleatorias independientes $\{X_m\}$ de manera que todas ellas siguen una distribución $\mathcal{U}(0, 10)$. A continuación, definimos la sucesión de variables $\{Y_m\}_m$ dada por $Y_m = \text{máx}\{X_1, \ldots, X_m\}$. ¿Podemos intuir la convergencia en ley de la sucesión $\{Y_m\}_m$?

Para resolver este ejercicio, vamos a dibujar la función de distribución empírica de la variable Y_m para $m = 10$, $m = 100$ y $m = 1000$, al igual que hicimos en el ejercicio anterior. Comencemos trabajando sobre la variable Y_{10}. Para ello, lo primero que haremos será simular los valores de las distribuciones uniformes.

```
n<-1000
x10<-matrix(runif(10*n,0,10),ncol=10)
y10<-apply(x10,1,max)
g1<-ggplot(data=NULL, aes(y10))+
  stat_ecdf(col="red",size=1.2)+
  ylab("Función de distribución")+
  xlab(TeX("Variable $y_{10}$"))
```

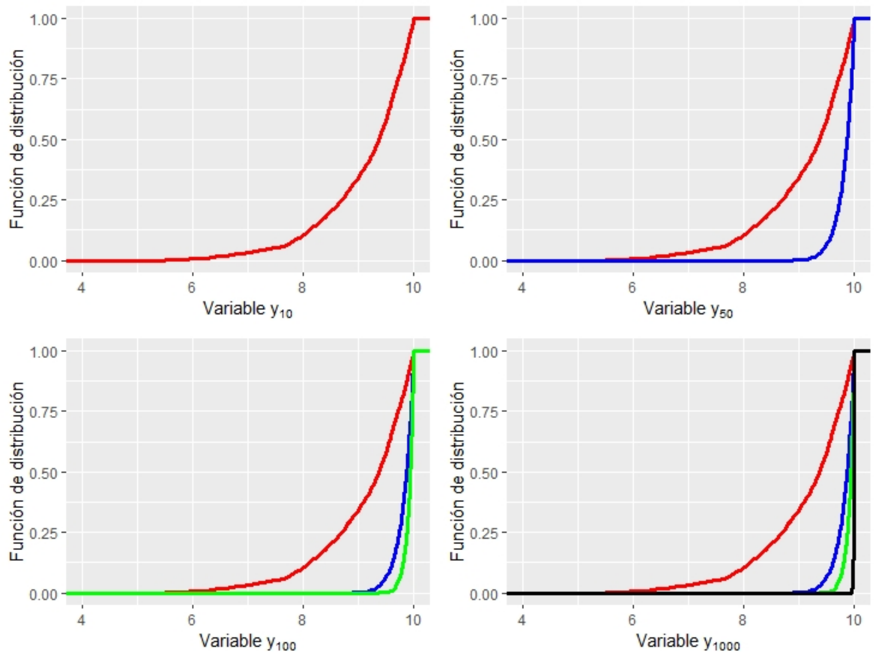

Figura 6.2: Representación gráfica de las funciones de distribución empíricas del Ejemplo 6.2.

En este caso, hemos simulado una matriz 1000×10, para posteriormente calcular los valores de Y_{10} haciendo la suma por filas de esa matriz. Finalmente, dibujamos su función de distribución empírica.

Repetimos el mismo procedimiento para Y_{50}, Y_{100} e Y_{1000}:

```
x50<-matrix(runif(50*n,0,10),ncol=50)
y50<-apply(x50,1,max)
g2<-g1+
  stat_ecdf(aes(y50),col="blue",size=1.2)+
  ylab("Función de distribución")+
  xlab(TeX("Variable $y_{50}$"))
x100<-matrix(runif(100*n,0,10),ncol=100)
y100<-apply(x100,1,max)
g3<-g2+
  stat_ecdf(aes(y100),col="green",size=1.2)+
  ylab("Función de distribución")+
  xlab(TeX("Variable $y_{100}$"))
x1000<-matrix(runif(1000*n,0,10),ncol=1000)
y1000<-apply(x1000,1,max)
g4<-g3+
  stat_ecdf(aes(y1000),col="black",size=1.2)+
  ylab("Función de distribución")+
  xlab(TeX("Variable $y_{1000}$"))
```

La representación gráfica obtenida paso a paso se puede ver en la Figura 6.2.

En estas gráficas se puede ver cómo las funciones de distribución se van aproximando a la función de distribución de una variable degenerada en el punto 10. De hecho, se puede demostrar que efectivmente $Y_n \xrightarrow{\mathcal{D}} Y$, donde Y es una variable degenerada en el 10.

6.3. Ejercicios propuestos

Ejercicio 6.1. Repite los pasos dados en el ejercicio 1. Dibuja en una misma gráfica la función de distribución empírica en el caso de $m = 1000$ y la función de distribución de una uniforme en el (0,1) para comprobar su parecido.

Ejercicio 6.2. Sea $\{X_m\}_m$ una sucesión de variables aleatorias independientes con distribución $\mathcal{U}(0, 10)$. Definimos $Y_m = \text{mín}\{X_1, \ldots, X_m\}$ y $W_m = m \cdot Y_m$. Se sabe que la sucesión $\{W_m\}_m$ converge débilmente a una distribución exponencial de parámetro $\frac{1}{10}$.

1. Realiza simulaciones para ver cómo va evolucionando la función de distribución empírica de W_m para $m = 10$, $m = 50$, $m = 100$ y $m = 1000$.

2. Dibuja la función de distribución de una exponencial de parámatro $\frac{1}{10}$ y la función de distribución empírica de W_{1000}.

Ejercicio 6.3. Sea $\{X_m\}_m$ una sucesión de variables aleatorias independientes donde $X_m \equiv \mathcal{B}(m, p_m)$, donde $p_m = \frac{1}{m}$. ¿Podríamos suponer que $X_m \xrightarrow{\mathcal{D}} \mathcal{P}(1)$?