

# CONVOLUCIÓN, PRUEBA $\chi^2$ , VARIANZA Y COVARIANZA

FABIOLA VÁZQUEZ

17 de noviembre de 2020

---

## 1. Introducción

El presente trabajo busca presentar resultados elementales de la varianza y covarianza de variables aleatorias. Se demuestran un par de resultados teóricos que son justificados de manera computacional. Además, se muestran aplicaciones de la técnica de convolución de funciones, y una prueba  $\chi^2$  de bondad de ajuste.

## 2. Convolución

La convolución se aplica en el procesamiento de imágenes, específicamente cuando se desea realizar alguna transformación como podría ser la aplicación de algún filtro de difuminado o de acentuación. Para este proceso es necesario definir la convolución discreta para dos matrices [3], la matriz asociada a la imagen de entrada y la matriz de kernel (filtro).

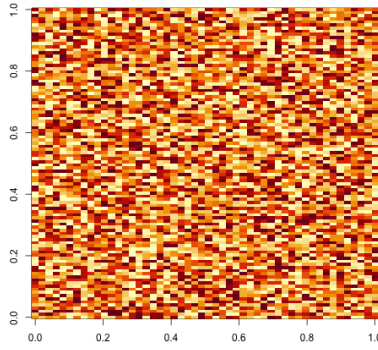
**Definición 1** (Convolución discreta). *La fórmula matemática para la convolución en 2D, está dada por*

$$k[m, n] = x[m, n] * h[m, n] = \sum_{j=-\infty}^{\infty} \sum_{i=-\infty}^{\infty} x[i, j] \cdot h[m - i, n - j], \quad (1)$$

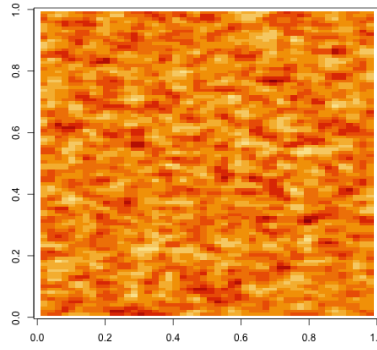
*donde  $x$  representa la matriz de la imagen de entrada a la que se le aplica la convolución con la matriz de kernel  $h$ ,  $k$  representa la matriz de la imagen de salida.*

La librería `image` [6] de R [4] tiene una implementación para realizar los cálculos de la convolución de matrices con la función `convolution2D`. Para ello es necesario una matriz  $x$  y una matriz kernel. En el código 1 se considera una matriz  $x$  de entrada de tamaño  $50 \times 100$ , donde los valores de las entradas de la matriz son generados con la función `runif`, la figura 1a muestra la imagen asociada a dicha matriz. Como kernel se utiliza la matriz de  $3 \times 3$  asociada al filtro de Gauss [5], es decir cuando este filtro se aplica a una imagen lo que hace es suavizarla, obteniendo una nueva imagen con difuminado.

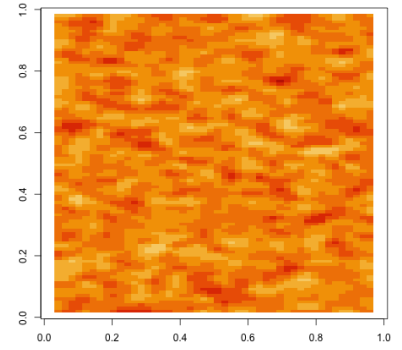
Al realizar la convolución de la matriz `Entrada` y la matriz `Kernel`, se obtiene la matriz `Salida1`. La figura 1b muestra la imagen asociada a esta matriz, como se puede ver esta es la imagen original difuminada. Si se vuelve a aplicar el filtro, pero ahora a la imagen de salida de la primera convolución se obtiene la matriz `Salida2` cuya imagen se aprecia en la figura 1c la cual tiene un mayor difuminado que la figura 1b.



(a) Imagen original.



(b) Aplicación del filtro de Gauss a la imagen original.



(c) Aplicación del filtro de Gauss a la imagen 1b.

Figura 1: Convolución en el procesamiento de imágenes.

```
library("image")
filas <- 50
columnas <- 100

Entrada <- matrix(runif(filas*columnas, 0, 100), nrow = filas, ncol = columnas)
Kernel <- matrix(c(1/16, 2/16, 1/16, 2/16, 4/16, 2/16, 1/16, 2/16, 1/16),
                 nrow=3, ncol=3)

Salida1 <- convolution2D(Entrada, kernel)
Salida2 <- convolution2D(Salida1, kernel)
```

Código 1: Implementación de la convolución en R.

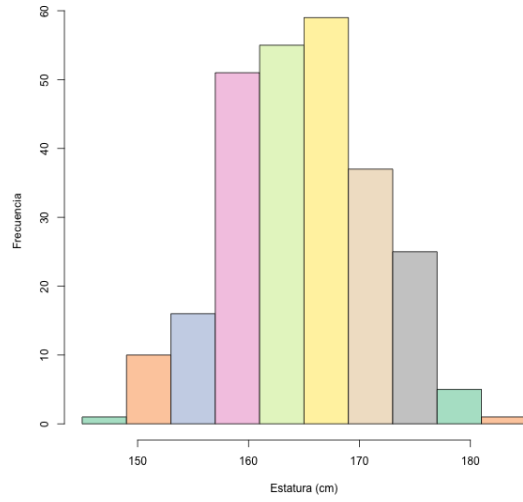
### 3. Prueba $\chi^2$

Una de las aplicaciones de la prueba  $\chi^2$  es realizar pruebas de bondad de ajuste para verificar si un conjunto de datos proviene de una población que se ajusta a un tipo de distribución de probabilidad. Como ejemplo, se trabajan con los [datos](#) recopilados por Heinz [1], considerando solamente la altura de las 260 mujeres.

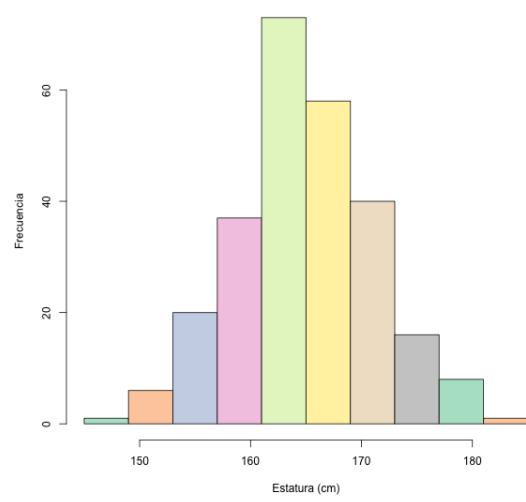
La figura [2a](#) muestra el histograma de los datos y por la forma de este se piensa que los datos siguen una distribución normal. Se generan 260 números con `rnorm` con media 165 y varianza 6.5, (que son la media y varianza de los datos observados). En el cuadro [1](#) se tienen los datos observados y los esperados. Se calcula

$$\sum_1^{10} \frac{(\text{esperado} - \text{observado})^2}{\text{esperado}}, \quad (2)$$

y se obtiene un valor aproximado de 20.83, con lo que se tiene un valor  $p \approx 0.013$ , el cual es menor que 0.05, por lo que se concluye que los datos no provienen de una distribución normal.



(a) Histograma de las alturas de las 260 mujeres.



(b) Histogramas de datos generados con `rnorm(260, mean = 165, sd = 6.5)`.

Figura 2: Comparación visual de los datos observados y los datos esperados.

Cuadro 1: Datos observados contra los datos generados.

Estatura (cm)	Observados	Esperados
145 - 148	1	1
149 - 152	10	6
153 - 156	16	20
157 - 160	51	37
161 - 164	55	73
165 - 168	59	58
169 - 172	37	40
173 - 176	25	16
177 - 180	5	8
181 - 185	1	1

Cuadro 2: Resultados de las pruebas computacionales.

(a) Teorema 2			(b) Teorema		
$X$	$Y$	Resultados	$X$	$Y$	Resultados
$N(0, 1)$	$U(0, 1)$	100	$N(0, 1)$	$U(0, 1)$	100
$N(0, 1)$	$\text{Exp}(1)$	100	$N(0, 1)$	$\text{Exp}(1)$	100
$\text{Exp}(1)$	$\text{Pois}(1)$	100	$\text{Exp}(1)$	$\text{Pois}(2)$	100

## 4. Varianza y covarianza

Se presentan dos resultados teóricos referentes a la varianza y covarianza de dos variables aleatorias. Primeramente, se analiza dichas propiedades computacionalmente generando variables con distintas distribuciones de probabilidad y comprobando la igualdad buscada. Para cada variable aleatoria se generan 5000 números y se realizaron 100 réplicas. El cuadro 2 muestra las distribuciones generadas para las dos variables aleatorias y la cantidad de resultados en los que la igualdad es cierta, como se aprecia en dicho cuadro, de las 100 réplicas todas cumplieron la igualdad, por lo que se procede a comprobar analíticamente los teoremas. Para las demostraciones, se ocupa recordar una definición y un teorema [2] vistos anteriormente.

**Definición 2.** Sean  $X$  y  $Y$  variables aleatorias, se define la covarianza de  $X$  y  $Y$  como,

$$\text{Cov}[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X] \mathbb{E}[Y]. \quad (3)$$

**Teorema 1.** Sea  $X$  una variable aleatoria, se define la varianza de  $X$  como,

$$\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2. \quad (4)$$

**Teorema 2.**  $\text{Cov}[aX + b, cY + d] = ac \text{Cov}[X, Y]$ .

*Demostración.*

$$\begin{aligned}
\text{Cov}[aX + b, cY + d] &= \mathbb{E}[(aX + b)(cY + d)] - \mathbb{E}[aX + b] \mathbb{E}[cY + d] \\
&= \mathbb{E}[acXY + adX + bcY + bd] - (a \mathbb{E}[X] + b)(c \mathbb{E}[Y] + d) \\
&= ac \mathbb{E}[XY] + ad \mathbb{E}[X] + bc \mathbb{E}[Y] + bd - (ac \mathbb{E}[X] \mathbb{E}[Y] + ad \mathbb{E}[X] + bc \mathbb{E}[Y] + bd) \\
&= ac \mathbb{E}[XY] + ad \mathbb{E}[X] + bc \mathbb{E}[Y] + bd - ac \mathbb{E}[X] \mathbb{E}[Y] - ad \mathbb{E}[X] - bc \mathbb{E}[Y] - bd \\
&= ac \mathbb{E}[XY] - ac \mathbb{E}[X] \mathbb{E}[Y] \\
&= ac(\mathbb{E}[XY] - \mathbb{E}[X] \mathbb{E}[Y]) \\
&= ac \text{Cov}[X, Y].
\end{aligned}$$

□

**Teorema 3.**  $\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2 \text{Cov}[X, Y]$ .

*Demostración.*

$$\begin{aligned}
\text{Var}[X + Y] &= \mathbb{E}[(X + Y)^2 - (\mathbb{E}[X + Y])^2] \\
&= \mathbb{E}[X^2 + 2XY + Y^2] - (\mathbb{E}[X] + \mathbb{E}[Y])^2 \\
&= \mathbb{E}[X^2] + 2 \mathbb{E}[XY] + \mathbb{E}[Y^2] - \mathbb{E}[X]^2 - 2 \mathbb{E}[X] \mathbb{E}[Y] - \mathbb{E}[Y]^2 \\
&= (\mathbb{E}[X^2] - \mathbb{E}[X]^2) + (\mathbb{E}[Y^2] - \mathbb{E}[Y]^2) + 2(\mathbb{E}[XY] - \mathbb{E}[X] \mathbb{E}[Y]) \\
&= \text{Var}[X] + \text{Var}[Y] + 2 \text{Cov}[X, Y].
\end{aligned}$$

□

# Referencias

- [1] Roger W. Johnson Grete Heinz, Louis J. Peterson. Exploring relationships in body dimensions. *Journal of Statistics Education*, 2003. <http://jse.amstat.org/v11n2/datasets.heinz.html>.
- [2] Charles M. Grinstead and J. Laurie Snell. *Introduction to Probability*. 2006.
- [3] Steve Marschner Peter Shirley. *Fundamentals of Computer Graphics*. A K Peters, Natick, Massachusetts, 2000.
- [4] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020.
- [5] Richard A. Haddad and Ali N. Akansu. A class of fast Gaussian binomial filters for speech and image processing. <https://web.njit.edu/~akansu/PAPERS/Haddad-AkansuFastGaussianBinomialFiltersIEEE-TSP-March1991.pdf>.
- [6] Wencheng Lau Medrano. IMAGIng engINEs, Tools for Application of Image Filters to Data Matrices. <https://cran.r-project.org/web/packages/imagine/imagine.pdf>.