Tarea 11 de Modelos Probabilistas Aplicados

Convolución, χ^2, covarianza

5271

17 de noviembre de 2020

1. Introducción

En este documento se presentan los resultados del análisis de varias propiedades y conceptos de la convolución, χ^2 y covarianza de forma analítica, así como los resultados alcanzados numéricamente mediante simulación. Para la simulación se utiliza lenguaje R [1].

2. Convolución

Una convolución es un operador matemático que transforma dos funciones f_1 y f_2 en una tercera función f_c que representa la magnitud en la que se superponen f_1 y una versión trasladada e invertida de f_2. Para el caso discreto se tiene la expresión:

$$\begin{array}{rcl} f_c & = & f_1 * f_2 \\ f_c(i) & = & \sum_j f_1(j) \times f_2(i-j) \end{array}.$$

Y para el caso continuo se cuenta con la siguiente ecuación:

$$(f*g)(z) = \int_{-\infty}^{\infty} f(z-y) \times g(y)\mathrm{d}y = \int_{-\infty}^{\infty} f(z-x) \times g(x)\mathrm{d}x.$$

2.1. Aplicaciones

La convolución tiene muchas aplicaciones prácticas como son:

- Procesamiento de imágenes.
- Filtrado de señales.
- Multiplicación de polinomios.
- Procesamiento de audio.

- Inteligencia artificial.
- Óptica.
- Teoría de la probabilidad.
- Mercado financiero.

Teoría de la probabilidad: Si se considera una situación en la que se tienen dos variables aleatorias independientes, X e Y, con funciones de densidad de probabilidad f y g respectivamente. Y se desea calcular la función de densidad $(X+Y)$, podemos usar la convolución de f y g. Por lo que se puede calcular la suma de cualquier número de variables independientes. Esto es importante porque muchas de las distribuciones estándar se caracterizan por patrones de convolución simples, lo que significa que podemos encontrar sus funciones de densidad de probabilidad mediante convolución.

Para realizar una prueba numérica de lo antes mencionado se utilizan los valores de rendimiento del oro y del platino obtenidos del sitio *Yahoo!finanzas* en la liga: https://es.finance.yahoo.com/materias-primas. Los valores de rendimientos fueron calculados a partir de los datos obtenidos como el valor de cierre menos el valor de apertura, el mismo es una variable aleatoria. Se pretende calcular la probabilidad conjunta de las variables X (oro) e Y (platino), para lo cual se utiliza la función *convolve* de la librería de R que utiliza la transformada rápida de Fourier para calcular la convolución de dos secuencias. En la figura 1 de la página 3 se puede observar los resultados obtenidos.

En la figura 2.1 y 2.1 de la página 3, muestran un comportamiento similar a la distribución normal, aunque no es tan evidente como en la 2.1 que muestra que la distribución de la convolución de las variables se distribuye normalmente. Lo anterior se corrobora aplicando la prueba de Shapiro-Wilk, que calcula un W estadístico que prueba si una muestra aleatoria $x1, x_2, ..., x_n$ proviene de una distribución normal. Al aplicar la prueba con un valor del estadístico $W = 0{,}99112$ y un valor $p = 0{,}0643 > 0{,}05$ no se tiene suficiente evidencia para rechazar la hipótesis nula, que plantea que la variable sigue una distribución normal, esto se puede afirmar con un intervalo de confianza del 95 %.

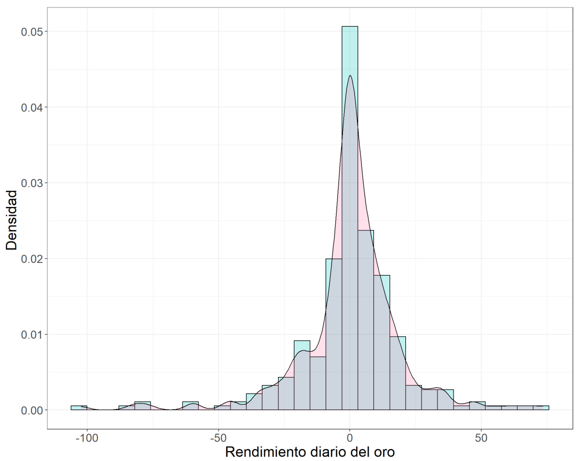
prueba.txt

```

	Shapiro-Wilk normality test

data:  re$resul
W = 0.9912, p-value = 0.06439

```

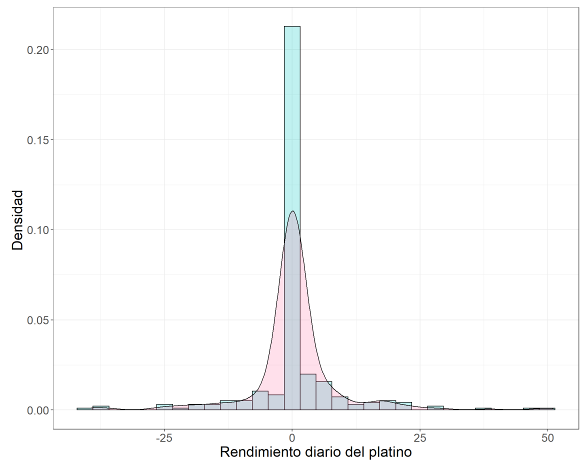
3. Chi cuadrada (χ^2)

La distribución χ^2 se utiliza para examinar tiene si un conjunto de datos presenta una diferencia estadísticamente significativa de lo que se espera. Para la misma necesitamos conocer la distribución que se espera ver y las frecuencias observadas de cada valor posible. Para el caso de estudio en este trabajo, se tiene el valor de densidad de empaquetamiento de seis tipos de figuras (triángulos,rectángulos, cuadrados,pentágonos, cuadriláteros mixtos y hexágonos) en un contenedor circular. Los valores de densidad se clasifican en tres categorías:

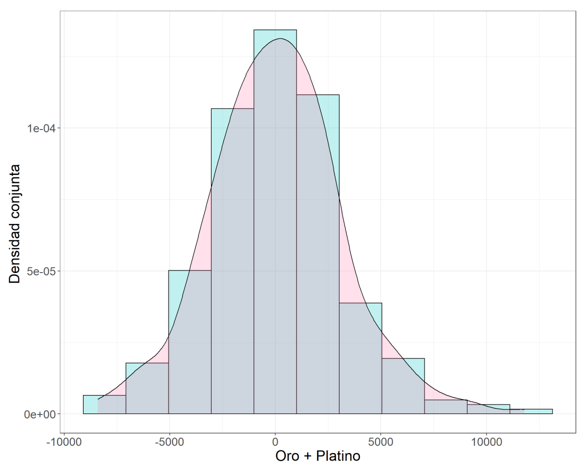
- Malos (densidad $\leq 0{,}40$).



(a) Histograma de densidad de la variable aleatoria X



(b) Histograma de densidad de la variable aleatoria Y



(c) Histograma de densidad conjunta de las variable aleatoria X e Y

Figura 1: Histogramas de la variables aleatorias X e Y y su convolución

- Buenos (0,40 $>$ densidad $\leq$ 0,70).
- Muy buenos (densidad $>$ 0,70).

Esta clasificación se puede ver en el cuadro 1 de la página 4, teniendo esto se plantea la hipótesis nula (H_0) que el tipo de figura no afecta en que la densidad de empaquetamiento sea mala, buena o muy buena. Para probar esta H_0 se realiza la prueba de *chisq.test* de la librería de R como se muestra a continuación.

prueba.txt

```
        Chi-squared test for given probabilities

data:  tabla
X-squared = 16.171, df = 2, p-value = 0.0003079
```

Con el estadístico $\chi^2 = 16{,}171$ y un valor $p = 0{,}0003 < 0{,}05$, se tiene suficiente evidencia para rechazar la H_0, por lo que se puede decir que el tipo de figura si influye en la calidad de la densidad de empaque, con un intervalo de confianza del 95 %.

Cuadro 1: Clasificación de los resultados de densidad para cada figura

Figuras	Malos	Buenos	Muy buenos	**Total**
Triángulo	1	20	14	**35**
Rectángulo	2	22	11	**35**
Cuadrados	5	20	10	**35**
Pentágono	7	16	7	**30**
Cuadrilátero mix	7	23	5	**35**
Hexágono	5	24	6	**35**
Total	**27**	**125**	**53**	**205**

4. Covarianza

En probabilidad y la estadística, la covarianza es una medida del grado en que dos variables aleatorias (X, Y) varían conjuntamente entorno a los valores de sus medias. Si las variables tienden a mostrar un comportamiento similar, la covarianza es positiva. En el caso contrario, cuando son inversamente proporcionales, la covarianza es negativa. Por lo que se puede carcular como se muestra en la ecuación 1 de la página 4.

$$\mathrm{Cov}[X,Y] = E[(X-\mu_X)(Y-\mu_Y)]. \tag{1}$$

4.1. Propiedades de la covarianza

En esta sección se comprobará numéricamente y analíticamente dos de las propiedades de la covarianza.

Primera: Se tiene:

$$\operatorname{Cov}[aX + b, cY + d] = ac \cdot \operatorname{Cov}[X, Y] \tag{2}$$

Para demostrar numéricamente la igualdad planteada en la ecuación 2 de la página 5, se crean diez pares de variables aleatorias X con distribución normal e $Y = \left(\frac{X}{2}\right) * \frac{d}{a}$ como se puede ver en el cuadro 2 de la página 5. Posteriormente se les asignan valores a las constantes a, b, c, d. Lo anterior se realiza con el código 4.1 que se muestra a continuación.

```
mi=numeric()
md=numeric()
normal= seq(100,190,10)
for (i in c(1:10)) {

  X = rnorm(sample(normal,1))
  Y = (X/2)+d/a
  a = sample(2:6,1)
  b = sample(1:8,1)
  c = sample(2:10,1)
  d = sample(3:6,1)

  mi =c(mi,a*c*cov(X,Y))
  md = c(md,cov(((a*X)+b), ((c*Y)+d)))
}
```

R/Tarea11.R

Cuadro 2: Fragmento de uno de los pares de variables aleatorias X e Y creadas

	X	Y
1	1.41293	1.20647
2	-0.44490	0.27755
3	1.02916	1.01458
4	0.54957	0.77478
5	-1.35893	-0.17946
54	-0.13546	0.43227
55	1.06661	1.03331
56	0.65222	0.82611
57	1.04822	1.02411
58	-1.37364	-0.18682

Cuadro 3: Resultados de ambos miembros de la ecuación 2

	M. izquierdo(M.i)	M. derecho(M.d)	M.i = M.d
1	10.28066	10.28066	Si
2	6.25041	6.25041	Si
3	14.60058	14.60058	Si
4	15.01817	15.01817	Si
5	6.16484	6.16484	Si
6	16.68462	16.68462	Si
7	14.00431	14.00431	Si
8	16.71537	16.71537	Si
9	7.37975	7.37975	Si
10	3.11481	3.11481	Si

Como se observa en el cuadro 3 de la página 5, para todas las variables aleatorias X e Y creadas los valores del miembro izquierdo de la ecuación 2 son iguales al miembro derecho de dicha ecuación. De esta manera queda demostrado numéricamente la igualdad planteada. Para esta comprobar la igualdad de manera analítica es conveniente recordar las dos propiedades siguientes.

Sea a y c dos constantes,

$$\mathrm{Cov}(aX, Y) = a \cdot \mathrm{Cov}(X, Y) \tag{3}$$

$$\mathrm{Cov}(X + c, Y) = \mathrm{Cov}(X, Y). \tag{4}$$

Desarrollando el lado izquierdo de la ecuación 2 se tiene:

$$\begin{aligned} \mathrm{Cov}[aX + b, cY + d] &= ac \cdot \mathrm{Cov}[X + b, Y + d], && \text{por la propiedad 3,} \\ &= ac \cdot \mathrm{Cov}[X, Y], && \text{por propiedad 4.} \end{aligned} \tag{5}$$

Segunda: Se tiene:

$$\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2 \cdot \mathrm{Cov}[X, Y] \tag{6}$$

Para la comprobación numérica de la igualdad 6 de la página 6, se utiliza el código 4.1, los valores de las variables aleatorias y las constantes tienen las mismas características que en el caso anterior.

```
mi=numeric()
md=numeric()
normal= seq(100,190,10)
for (i in c(1:10)) {

  X = rnorm(sample(normal,1))
  Y = (X/2)+d/a
  a = sample(2:6,1)
  b = sample(1:8,1)
  c = sample(2:10,1)
  d = sample(3:6,1)

  mi =c(mi,var(X+Y))
  md = c(md,var(X)+ var(Y)+2*cov(X,Y))
}
```

R/Tarea11.R

Cuadro 4: Resultados de ambos miembros de la ecuación 6

	M. izquierdo(M.i)	M. derecho(M.d)	M.i = M.d
1	2.49469	2.49469	Si
2	2.23400	2.23400	Si
3	2.44015	2.44015	Si
4	2.20168	2.20168	Si
5	2.37420	2.37420	Si
6	2.40887	2.40887	Si
7	2.17660	2.17660	Si
8	1.95016	1.95016	Si
9	2.60644	2.60644	Si
10	1.98039	1.98039	Si

En los resultados que se muestran en el cuadro 4 de la página 6, queda comprobado de forma numérica la igualad planteada en la ecuación 6. Para la corporación analítica de ecuación en cuestión, no apoyamos de las expresiones siguientes:

$$\mathrm{Var}[X] = E[X^2] - E[X]^2. \tag{7}$$

$$\mathrm{Cov}(X,Y) = E[XY] - E[X]E[Y]. \tag{8}$$

Desarrollando el lado izquierdo de la ecuación 6 se tiene:

$$\begin{aligned}
\mathrm{Var}[X+Y] &= E[(X+Y)^2] - E[X+Y]^2 \\
&= E[(X+Y)^2] - E[X+Y]^2 \\
&= E[X^2] - E[X]^2 + E[Y^2] - E[Y]^2 + 2\cdot E[XY] - 2\cdot E[X]E[Y] \\
&= \mathrm{Var}[X] + \mathrm{Var}[Y] + 2(E[XY] - E[X]E[Y]) \\
&= \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\mathrm{Cov}(X,Y).
\end{aligned} \tag{9}$$

El código general se encuentra disponible en el repositorio. https://github.com/Albertomnoa/Tareas

Referencias

[1] R Core Team. R: Un lenguaje y un entorno para la informática estadística, 2020.