Tarea 11

Convolución, varianza y covarianza

Erick Cervantes Mendieta
Matrícula: 2032430

Modelos Probabilistas Aplicados

17/11/2020

1. Convolución

La *convolución* es una operación matemática que combina dos señales para producir una tercera señal, algunas de sus aplicaciones se encuentran en la ingeniería, en esta sección se describe una aplicación en el pronóstico de demanda, ya que al tener una serie de tiempo, se necesita conocer valores futuros, por lo que se describe el **promedio móvil ponderado**.

Esta técnica es una variante del método de promedios móviles, solo que en este caso se tiene la posibilidad de modificar las ponderaciones que tiene cada uno de los datos en el cálculo del promedio, por lo general, el dato con mayor ponderación es el más reciente, sin embargo, la asignación de las ponderaciones puede ser arbitrario o con base a la experiencia.

La ecuación a utilizar es la siguiente:

$$\hat{X}_t = \sum_{i=1}^{n} C_i x_{t-i},$$

con $\hat{X}_t$ = promedio de la demanda en el periodo t, C_i es el factor de ponderación, x_{t-i} es la demanda real de los periodos anteriores a t y n es el número de datos.

En [1] se define a la convolución de $m_1(x)$ y $m_2(x)$ como la función de distribución m_3 dada por:

$$m_3(j) = \sum_{k} m_1(k) \cdot m_2(j-k),$$

de esta forma se podría decir que $\hat{X}_t$ se considera como $m_3(j)$, C_i como $m_1(k)$ y a x_{t-i} como $m_2(j-k)$. Por ejemplo, suponga que desea que desea prónosticar el valor de compra del dólar para la próxima semana, para esto en el cuadro 1 se presenta el valor de compra del dólar del 9 al 16 de noviembre del 2020, por lo que si se considera un periodo igual a 4, con

Cuadro 1: Valor de compra del dólar en la semana del 9 al 15 de noviembre del 2020.

Día	Compra ($)
09/11/2020	20.0833
10/11/2020	20.0604
11/11/2020	20.1647
12/11/2020	20.2442
13/11/2020	20.1172
14/11/2020	20.1151
15/11/2020	20.111
16/11/2020	¿?

Cuadro 2: Tabla de contigencia para la cantidad de alumnos matriculados en los diferentes niveles de escolaridad de algunos estados (INEGI).

	Primaria	**Secundaria**	**Media Superior**	**Superior**
CDMX	805,705	436,827	466,232	560,680
Nuevo León	580,968	271,699	191,339	229,086
Tlaxcala	152,880	74,396	56,597	37,521
Veracruz	834,274	388,304	316,662	197,118

$C_1 = 0.15$, $C_2 = 0.20$, $C_3 = 0.30$ y $C_4 = 0.35$, se tiene que, para prónosticar el valor del día viernes 13 de noviembre es necesario realizar las siguientes operaciones:

$$\begin{aligned}
\text{viernes}_{13} &= 0.15(\text{lunes}_{11}) + 0.20(\text{martes}_{10}) + 0.30(\text{miércoles}_{11}) + 0.35(\text{jueves}_{12}) \\
&= 0.15(20.08833) + 0.20(20.0604) + 0.30(20.1647) + 0.35(20.2442) \\
&= 20.1602,
\end{aligned}$$

este procedimiento se vuelve a repetir, considerando siempre los valores reales de compra del dólar, y en la cuarta iteración se obtendría que el valor de compra del dólar el día lunes 16 de noviembre del 2020 va a ser de $20.1335.

2. Prueba χ^2

Se está interesado saber si la cantidad de alumnos matriculados de la República Mexicana en los diferentes niveles de escolaridad (primaria, secundaria, media superior y superior) tiene que ver con los diferentes estados del país, ya que algunos dirigentes aseguran que se está apoyando a los niños y jóvenes con becas para que se sigan preparando y no tengan que emigrar a otro lugar. En este caso se plantea la hipótesis nula: No existe relación alguna entre el número de alumnos matriculados con algún estado del país. La tabla de contigencia se muestra en el cuadro 2.

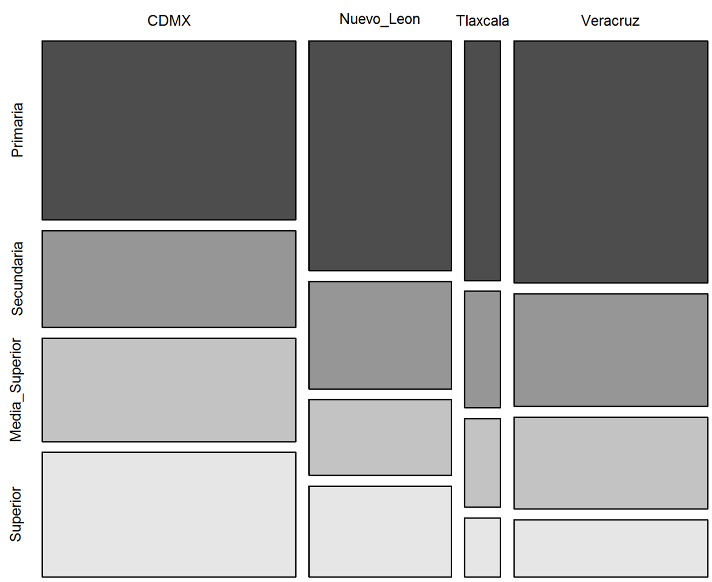

Figura 1: Gráfica de mosaico para la cantidad de alumnos matriculados en los diferentes niveles de escolaridad de algunos estados.

Al implementar la prueba χ^2 en el entorno R (Versión 4.0.2) [2], se obtuvo un valor $p < 2.2 \times 10^{-16}$, por lo que hay suficiente evidencia para decir que si *hay relación* entre el número de matricula en los diferentes niveles de edución con algunos estados de la República Mexicana, es decir, parece observarse un mismo comportamiento en la matricula, sin importar si el estado tiene más/menos población (ver figura 1).

3. Propiedades covarianza

i) Demuestre que $\mathrm{Cov}[aX+b, cY+d] = ac\mathrm{Cov}[X,Y]$. Para demostrar lo anterior, recurrimos a la definición de covarianza: $\mathrm{Cov}[X,Y] = \mathbb{E}[XY] - \mathbb{E}(X)\mathbb{E}[Y]$ (Véase [3]), y considerando a a, b, c y d como constantes, tenemos entonces:

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

En [3] se enuncia el siguiente ejemplo: Se seleccionan al azar dos repuestos para bolígrafo de una caja que contiene tres repuestos azules, dos rojos y tres verdes. Si X es el número de repuestos azules y Y es el número de repuestos rojos seleccionados, entonces la función de probabilidad conjunta esta dada por:

$$
f(x,y)=\frac{\begin{pmatrix}3\\x\end{pmatrix}\begin{pmatrix}2\\y\end{pmatrix}\begin{pmatrix}3\\2-x-y\end{pmatrix}}{\begin{pmatrix}8\\2\end{pmatrix}},
$$

para $x=0,1,2$; $y=0,1,2$ y $0\leq x+y\leq 2$, luego entonces, la distribución de probabilidad conjunta se muestra en el cuadro 3. Los valores esperados de X, Y y XY se muestran en seguida:

$$
\mathbb{E}(X)=0\left(\frac{5}{28}\right)+1\left(\frac{15}{28}\right)+2\left(\frac{3}{28}\right)=\frac{3}{4},
$$

$$
\mathbb{E}(Y)=0\left(\frac{15}{28}\right)+1\left(\frac{3}{7}\right)+2\left(\frac{1}{28}\right)=\frac{1}{2},
$$

$$
\begin{aligned}
\mathbb{E}(XY) &= \sum_{x=0}^{2}\sum_{x=0}^{2}xyf(x,y) \\
&= (0)(0)f(0,0)+(0)(1)f(0,1)+(0)(2)f(0,2)+(1)(0)f(1,0) \\
&\quad +(1)(1)f(1,1)+(2)(0)f(2,0)=\frac{3}{14}.
\end{aligned}
$$

Luego, definimos a $f(X)=3X-1$ y a $g(Y)=2Y+5$, por lo que se tiene lo siguiente:

$$
\mu_{3X-1}=3\mathbb{E}(X)-1=3\left(\frac{3}{4}\right)-1=\frac{5}{4},
$$

$$
\mu_{2Y+5}=2\mathbb{E}(Y)+5=2\left(\frac{1}{2}\right)+5=6,
$$

Cuadro 3: Distribución de probabilidad conjunta para el ejemplo del bolígrafo.

$f(x,y)$		x			Totales por renglón
		0	1	2	
	0	3/28	9/28	3/28	15/28
y	1	3/14	3/14	0	3/7
	2	1/28	0	0	1/28
Totales por columna		5/28	15/28	3/28	1

luego, para calcular $\mathrm{Cov}[3X-1, 2Y+5]$ tenemos que:

$$\begin{aligned}
\mathrm{Cov}[3X-1,2Y+5] &= \mathbb{E}[((3X-1)-\mu_{3X-1})((2Y+5)-\mu_{2Y+5})] \\
&= \mathbb{E}\left[\left((3X-1)-\frac{5}{4}\right)((2Y+5)-6)\right] \\
&= \mathbb{E}\left[\left(3X-\frac{9}{4}\right)(2Y-1)\right] \\
&= \mathbb{E}\left[6XY-3X-\frac{9}{2}Y+\frac{9}{4}\right] \\
&= 6\mathbb{E}[XY]-3\mathbb{E}[X]-\frac{9}{2}\mathbb{E}[Y]+\frac{9}{4} \\
&= 6\left(\frac{3}{14}\right)-3\left(\frac{3}{4}\right)-\frac{9}{2}\left(\frac{1}{2}\right)+\frac{9}{4} \\
&= -\frac{27}{28}.
\end{aligned}$$

Se tiene por definición que $ac\mathrm{Cov}[X,Y] = ac(\mathbb{E}[XY]-\mathbb{E}[X]\mathbb{E}[Y])$, entonces:

$$\begin{aligned}
3(2)\mathrm{Cov}[X,Y] &= 3(2)\left(\frac{3}{14}-\left(\frac{3}{4}\right)\left(\frac{1}{2}\right)\right) \\
&= 6\left(\frac{3}{14}-\frac{3}{8}\right) = 6\left(-\frac{9}{56}\right) = -\frac{27}{28},
\end{aligned}$$

luego entonces: $\mathrm{Cov}[3X-1,2Y+5] = 3(2)\mathrm{Cov}[X,Y]$.

ii) Demuestre que $\mathrm{Var}[X+Y] = \mathrm{Var}[X]+\mathrm{Var}[Y]+2\mathrm{Cov}[X,Y]$. Para demostrar lo anterior, se define a $\mathrm{Cov}[X,Y] = \mathbb{E}[(X-\mu_X)(Y-\mu_Y)]$ y $\mathrm{Var}[X] = \mathbb{E}[(X-\mu_X)^2]$ (Véase [3]). Luego:

$$\begin{aligned}
\mathrm{Var}[X+Y] &= \mathbb{E}\left[\{(X+Y)-(\mu_X+\mu_Y)\}^2\right] \\
&= \mathbb{E}\left[\{(X-\mu_X)+(Y-\mu_Y)\}^2\right] \\
&= \mathbb{E}\left[(X-\mu_X)^2+2(X-\mu_X)(Y-\mu_Y)+(Y-\mu_Y)^2\right] \\
&= \mathbb{E}\left[(X-\mu_X)^2\right]+2\mathbb{E}\left[(X-\mu_X)(Y-\mu_Y)\right]+\mathbb{E}\left[(Y-\mu_Y)^2\right] \\
&= \mathrm{Var}[X]+\mathrm{Var}[Y]+2\mathrm{Cov}[X,Y].
\end{aligned}$$

Del ejercicio de los *bolígrafos* mencionado con anterioridad, en el cual sse definió a X como el número de repuestos azules y Y como el número de repuestos rojos, se tiene que el valor esperado de X^2, Y^2 y $X+Y$ es:

$$\mathbb{E}(X^2) = 0^2\left(\frac{5}{28}\right) + 1^2\left(\frac{15}{28}\right) + 2^2\left(\frac{3}{28}\right) = \frac{27}{28},$$

$$\mathbb{E}(Y^2) = 0^2\left(\frac{15}{28}\right) + 1^2\left(\frac{3}{7}\right) + 2^2\left(\frac{1}{28}\right) = \frac{4}{7},$$

$$\begin{aligned}
\mathbb{E}(X+Y) &= \sum_{x=0}^{2}\sum_{x=0}^{2}(x+y)f(x,y) \\
&= (0+0)f(0,0) + (0+1)f(0,1) + (0+2)f(0,2) + (1+0)f(1,0) \\
&\quad +(1+1)f(1,1) + (2+0)f(2,0) \\
&= f(0,1) + 2f(0,2) + f(1,0) + 2f(1,1) + 2f(2,0) \\
&= \frac{9}{28} + 2\left(\frac{3}{28}\right) + \frac{3}{14} + 2\left(\frac{3}{14}\right) + 2\left(\frac{1}{28}\right) \\
&= \frac{9}{28} + \frac{6}{28} + \frac{3}{14} + \frac{6}{14} + \frac{2}{28} \\
&= \frac{5}{4},
\end{aligned}$$

luego se tiene que:

$$\begin{aligned}
\mathrm{Var}[X+Y] &= \mathbb{E}\left[(X+Y)^2\right] - \mu_{x+y}^2 \\
&= \mathbb{E}[X^2 + 2XY + Y^2] - \left(\frac{5}{4}\right)^2 \\
&= \mathbb{E}[X^2] + 2\mathbb{E}[XY] + \mathbb{E}[Y^2] - \left(\frac{25}{16}\right) \\
&= \frac{27}{28} + 2\left(\frac{3}{14}\right) + \frac{4}{7} - \left(\frac{25}{16}\right) \\
&= \frac{45}{112}.
\end{aligned}$$

Por otra parte, la varianza de X, la varianza de Y y la covarianza de X y Y se determinan como sigue:

$$\mathrm{Var}[X] = \mathbb{E}[X^2] - \left(\frac{3}{4}\right)^2 = \frac{45}{112},$$

$$\mathrm{Var}[Y] = \mathbb{E}[Y^2] - \left(\frac{1}{2}\right)^2 = \frac{9}{28},$$

$$\mathrm{Cov}[X,Y] = \mathbb{E}[XY] - \mu_X\mu_Y = -\frac{9}{56},$$

por lo tanto $\mathrm{Var}[X] + \mathrm{Var}[Y] + 2\mathrm{Cov}[X,Y] = 45/112 = \mathrm{Var}[X+Y]$.

Referencias

[1] . S. J. L. Grinstead, C. M. *Introduction to probability.* American Mathematical Soc., 2012.

[2] R Core Team. R: A Language and Environment for Statistical Computing. https://www.R-project.org/, 2020.

[3] R. E. Walpole and R. H. Myers. *Probabilidad y estadística para ingeniería y ciencias.* Pearson Educación de México, 2012.