CALCULO DE LAS PROBABILIDADES DE QUE EN UN INTERVALO DE *n* DIAS, HAYA AL MENOS UN DIA DE PRECIPITACION O UN DIA SECO

Autores: Carlos Pérez Manrique; M.ª Isabel Garmendia Rodríguez; Concepción Rodríguez Puebla y José Garmendia Iraundegui

Departamento de Física del Aire. Facultad de Ciencias. Universidad de Salamanca

RESUMEN

Utilizando Cadenas de Markov de 1.er orden, se ha hecho el cálculo de la probabilidad, R_n, de que en un período de *n* días consecutivos haya al menos un día de precipitación, y de la probabilidad, G_n, de que en un intervalo de *n* días consecutivos haya al menos un día seco.

Los cálculos se han hecho para las estaciones meteorológicas de Gijón y San Sebastián y para cada uno de los meses del año.

Mediante el test de la χ^2 de Pearson se han comparado las frecuencias teóricas con las observadas durante los 30 años que transcurren desde 1946 a 1975, obteniéndose en ambas poblaciones y en cada una de las estaciones meteorológicas, unos resultados óptimos, excepto en el cálculo de R_n en el mes de marzo en Gijón. Para este mes se ha repetido el cálculo de R_n, utilizando Cadenas de Markov de 2.º orden.

1. Probabilidad de que en un intervalo de *n* días, haya al menos un día de precipitación

Se define R_n, como la probabilidad de que en un período de *n* días consecutivos, al menos haya un día con precipitación apreciable.

Se define R'_n como la probabilidad de que en un período de *n* días, no haya ningún día con precipitación apreciable, es decir, todos los días sean secos.

Se considera día lluvioso, cuando la precipitación registrada entre las 0 y 24 horas TMG, de dicho día, sea igual o superior a 0,1 mm. En caso contrario, se dice que el día ha sido seco.

Estas dos probabilidades son complementarias y por tanto $R_n = 1 - R'_n$ y para su cálculo seguimos la idea de B. Erickson (1965):

$X_i = 1$ significa que el día i-esimo fue lluvioso.

$X_i = 0$ significa que el día i-esimo fue seco.

Ahora bien, $R'_n = P(X_1 = X_2 = = \ldots X_n = 0) = P(X_1 = 0) \times P(X_2 = 0/X_1 = X_2 = 0) \times \times P(X_4 = = 0/X_1 = X_2 = X_3 = 0) \ldots P(X_n = 0/X_1 = = X_2 = X_3 = \ldots X_{n-1} = 0)$.

Si el cálculo de R'_n se hace aplicando el modelo de Cadena de Markov de 1.er orden, como $P(X_n = 0/X_1 = X_2 = \ldots X_{n-1} = 0) = P(X_n = 0/X_{n-1} = 0)$ para todo *n*, se tiene que:

$R'_n = P(X_1 = 0) \times P(X_2 = 0/X_1 = 0)\ P(X_3 = = 0/X_2 = 0) \ldots P(X_n = 0/X_{n-1} = 0)$, y como se hace la suposición de que todas las Cademas de Markov con que trabajamos en nuestro estudio, son homogéneas dentro de cada mes, es decir que a lo largo de todos los días del mes no varían los valores de $P(0)$, probabilidad de día seco y $P_0(0)$, probabilidad de día seco condicionada a que el día precedente fue seco, se puede concluir que:

$$R'_n = P(0) \times \underbrace{P_0(0) \ldots P_0(0)}_{n - 1 \text{ veces}} = P(0)\, P_0^{n-1}(0) \text{ y por tanto } R_n = 1 - P(0) \times P_0^{n-1}(0)$$

En el caso en que para el cálculo de R'_n se emplee una Cadena de Markov de 2.º orden, puesto que

$P(X_n = 0/X_1 = X_2 = ... X_{n-1} = 0) =$
$= P(X_n = 0/X_{n-2} = X_{n-1} = 0)$ para todo *n* se tiene que $R'_n = P(X_1 = 0) \times P(X_2 = 0/X_1 =$
$= 0) \times P(X_3 = 0/X_1 = X_2 = 0) \times P(X_4 = 0/X_2 =$
$= X_3 = 0) ... P(X_n = 0/X_{n-2} = X_{n-1} = 0)$ y por suponer que dentro de cada mes la Cadena de Markov es homogénea, se puede concluir $R'_n =$
$P_{00}(0) ... P_{00}(0) =$ n − 2 veces
$= P(0)\ P_0(0)\ P_0\ P_{00}^{n-2}(0)$ y por tanto $R_n = P(0)\ P_0 \times$
y por tanto $R_n = 1 - P(0) \times P_0(0) \times P_o \times P_{00}^{n-2}(0)$
siendo válida esta fórmula para todo *n* mayor o igual que 2.

$P_{00}(0)$ significa la probabilidad de día seco condicionada a que los dos días precedentes fueron secos.

Si n = 1, $R'_1 = P(X_1 = 0) = P(0)$ y por tanto $R_1 =$
$= 1 - P(0)$.

Por todo lo deducido disponemos para el cálculo de R_n de estas dos fórmulas:

(1) $R_n = 1 - P(0)\ P_0^{n-1}(0)$, cuando el modelo utilizado es una Cadena de Markov de primer orden.

(2) $R_n\ 1 - P(0)\ P_0(0)\ P_{00}^{n-2}(0)$ $\quad n \geqslant 2$ cuando el modelo utilizado es Cadena de Markov de 2.º orden.

$R_1 = 1 - P(0)$.

En la tabla 1 se describen los valores de R_n calculados por meses en Gijón y San Sebastián utilizando la fórmula (1) y los resultados se ha expresado en tanto por ciento de probabilidad. Es decir, que a las probabilidades R_n las multiplicamos por 100.

El período mayor de días que se ha considerado es de 21, puesto que para ese día en todos los meses R_n, alcanza al menos un valor del 99 %.

Observando la tabla 1 en la que / significa más del 99 % se sacan las siguientes conclusiones:

TABLA 1.— Probabilidad Rn de que en un intervalo de *n* días haya al menos un día de precipitación en Gijón

M	Enero	Febr.	Mar.	Abr.	Mayo	Jun.	Jul.	Agost.	Sept.	Oct.	Nov.	Dic.
1	51	51	48	49	47	36	26	36	42	45	53	53
2	66	65	63	66	63	51	41	52	60	59	67	67
3	78	74	73	78	74	63	52	65	73	69	78	77
4	84	81	81	85	82	72	61	75	81	77	84	84
5	89	87	86	90	87	78	69	82	87	83	89	89
6	92	90	90	93	91	84	75	87	91	87	93	93
7	95	93	93	96	94	87	80	91	94	91	95	95
8	96	95	95	97	95	90	84	93	96,7	93	96	96
9	97	96	96	98	97	93	87	95	97	95	97,6	97,5
10	98	97	97	98,8	98	94	89	96	98	96	98,3	98,3
11	98,3	98	98	99,2	98,4	96	91	97,4	98,6	97	98,8	98,8
12	98,9	99	98,7	/	98,9	97	93	98,1	99	97,9	99,2	99,2
13	99	99	99	/	99,2	97,5	94	98,6	/	98,4	/	/
14	/	/	/	/	/	98,1	96	99	/	98,9	/	/
15	/	/	/	/	/	98,5	96,4	/	/	99,1	/	/
16	/	/	/	/	/	98,9	97,1	/	/	/	/	/
17	/	/	/	/	/	99,2	97,2	/	/	/	/	/
18	/	/	/	/	/	/	98,2	/	/	/	/	/
19	/	/	/	/	/	/	98,5	/	/	/	/	/
20	/	/	/	/	/	/	98,8	/	/	/	/	/
21	/	/	/	/	/	/	99	/	/	/	/	/

La probabilidad R_n de que al menos llueva un día en un intervalo de *n* días consecutivos, obtenida utilizando Cadenas de Markov de 1.er orden, crece rápidamente al aumentar *n*, en las dos estaciones meteorológicas, y se alcanza el 99 %, en todos los meses del año, antes en San Sebastián que en Gijón, y donde más se acentúa esa diferencia es en el mes de julio, puesto que en Gijón se necesitan 21 días para que R_n alcance el 99 %, mientras que en San Sebastián ya para el día 11, R_n supera el 99 % de probabilidad.

En Gijón el mes en que antes alcanza R_n el 99 % de probabilidad es abril que lo hace el día undécimo y el último julio que lo hace el día 21, mientras que en San Sebastián el primero es mayo que lo consigue el día noveno y el último octubre que lo alcanza el día 14. Se puede observar que en ambas estaciones meteorológicas apenas se producen rachas secas de larga duración en cualquier mes del año.

A partir de los datos de la precipitación regitrada en Gijón y en San Sebastián (Monte Igueldo) durante los 30 años que van desde 1946 a 1975 se han estimado las probabilidades P(0), $P_0(0)$, etc., necesarias para nuestro estudio. Asimismo con estos mismos datos se calculan las frecuencias observadas, las cuales se comparan con las frecuencias teóricas obtenidas a partir de los valores de la tabla 1. Esta comparación se hace mediante el test de la χ^2 de Pearson.

El cálculo de R_n, hecho en base al modelo teórico de Cadenas de Markov de 1.er orden, obtiene unos ajustes excelentes entre las frecuencias teóricas y las observadas, en las dos estaciones meteorológicas, para todos los meses del año, excepto marzo en Gijón, siendo el nivel de significación de los test superior al 90 % en todos ellos. En marzo el nivel de significación sólo es del 0,5 %.

El desajuste que existe en el mes de marzo, es debido a que en los años 1948 y 1953 se producen, respectivamente, una racha seca de 28 y 30 días de duración. Si se prescinde de los datos de estos dos años, el modelo de Cadenas de Markov de 1.er orden se ajusta perfectamente a lo observado, consiguiéndose en el test de la χ^2 un nivel de significación de 99,5 %.

En el mes de marzo, se ha repetido el cálculo de R_n, utilizando el modelo de Cadenas de Markov de 2.º orden, comprobándose que este modelo mejora el de 1.er orden en este mes, puesto que el nivel de significación de marzo pasa de 0,5 % a un valor entre el 2,5 % y el 5 %, pero de todas maneras no llega al 5 % del nivel de significacion, debido a las 2 rachas secas de larga duración anteriomente mencionadas.

TABLA 1.—Probabilidad Rn de que en un intervalo de *n* días haya al menos un día de precipitación en San Sebastián

M	Enero	Febr.	Mar.	Abr.	Mayo	Jun.	Jul.	Agos.	Sept.	Oct.	Nov.	Dic.
1	52	52	49	56	57	51	47	51	51	44	54	55
2	67	67	65	72	74	70	65	69	68	59	68	69
3	77	78	75	82	84	81	77	81	79	70	78	78
4	84	85	83	89	91	88	85	88	87	78	85	85
5	89	90	88	93	94	93	90	93	91	84	90	90
6	92	93	92	95	97	95	93	96	94	88	93	93
7	94	95	94	97	98	97	96	97	96	91	95	95
8	96	97	96	98	98,7	98,3	97	98	98	94	97	97
9	97	98	97	98,8	99,2	98,9	98,1	99	98,4	95	98	97,6
10	98	98,6	98	99,2	/	99,3	98,7	/	99	97	98,5	98,4
11	99,7	99	98,6	/	/	/	99,2	/	/	97,4	99	99
12	/	/	99	/	/	/	/	/	/	98	/	/
13	/	/	/	/	/	/	/	/	/	98,6	/	/
14	/	/	/	/	/	/	/	/	/	99	/	/

Este último hecho de la idea de que las Cadenas de Markov de 1.er y 2.o orden tienen una mayor validez, para el cálculo de R_n, en las estaciones meteorológicas donde escasean las rachas secas de larga duración.

2. Probabilidad de que haya al menos un día seco en un período de *n* días consecutivos

La resolución de este problema se obtiene de forma totalmente similar a como se hizo el cálculo de la probabilidad R_n de que al menos haya un día de lluvía en un período de *n* días consecutivos.

Se define G_n como la probabilidad de que haya al menos un día seco en un período de *n* días consecutivos.

$G_n = 1 - G'_n$, siendo G'_n la probabilidad de *n* días seguidos de lluvia, por tanto $G'_n = P(1) \times P_1^{n-1}(1)$ cuando se utiliza el modelo de Cadena de Markov de 2.o orden, de donde se deducen las fórmulas:

(3) $G_n = 1 - P(1) \times P_1^{n-1}(1)$ cuando el modelo utilizado es una Cadena de Markov de 1.er orden.

(4) $G_n = 1 - P(1)\, P_1(1)\, P_{11}^{n-2}(1) \quad n \geq 2$ Cuando el modelo utilizado es una Cadena de Markov de 2.o orden.

$G_1 = 1 - P(1)$

$P(1)$ = probabilidad de que un día sea lluvioso.

$P_1(1)$ = probabilidad de qe un día que lluvioso condicionado a que el día anterior fue seco.

$P_{11}(1)$ = probabilidad de que un día sea lluvioso, condicionado a que los dos días inmediatamente anteriores, fuesen lluviosos.

Los valores de G_n, expresados en tanto por ciento, se escriben en la tabla 2, significando más del 99 %.

Al observar la tabla 2, se sacan las siguientes conclusiones:

La probabilidad G_n de que al menos sea seco un día en un período de *n* días consecutivos, obtenida utilizando Cadenas de Markov de 1.er orden crece rápidamente al crecer *n*, en las dos estaciones meteorológicas, de manera que para cada mes, en ambas, se supera el 99 % de probabilidad antes de que transcurran 15 días.

La variación de G_n es muy similar en Gijón y en San Sebastián en todas las estaciones del año, excepto en el verano, que en Gijón se sobrepasa antes el 99 % de probabilidad que en San Sebastián, y concretamente en el mes de julio en Gijón se necesitan 5 días para $G_n \geq 99$ %, mientras que en San Sebastián se necesitan 9. Es un resultado dual de la consecuencia anteriormente expuesta en el estudio de R_n.

TABLA 2.—Probabilidad G_n de que haya al menos un día seco, en un período de *n* días consecutivos en Gijón

n	Enero	Febr.	Mar.	Abr.	Mayo	Jun.	Jul.	Agost.	Sept.	Oct.	Nov.	Dic.
1	49	49	52	51	53	64	74	64	58	55	47	47
2	64	62	67	68	68	79	88	82	76	69	62	61
3	75	72	77	79	79	88	95	91	86	79	72	72
4	82	80	84	87	86	93	97,6	95,5	92	86	80	79
5	87	85	89	91	91	96	99	97,7	95	90	86	85
6	91	89	92	94	94	97,7	/	98,9	97,4	93	90	89
7	94	92	95	96	96	98,7	/	99,4	98,5	95	93	92
8	96	94	96	97,6	97,2	99,2	/	/	99,2	97	95	94
9	97	96	97,4	98,5	98,1	/	/	/	/	97,9	96	96
10	97,8	97	98,2	99	98,7	/	/	/	/	98,5	97,2	97
11	98,4	97,6	98,7	99,3	99,1	/	/	/	/	99	98	97,7
12	98,9	98,3	99,1	/	/	/	/	/	/	/	98,5	98,3
13	99,2	98,7	/	/	/	/	/	/	/	/	98,9	98,7
14	/	99,1	/	/	/	/	/	/	/	/	99,2	99,1

TABLA 2.—Probabilidad G_n de que en un intervalo de *n* días haya al menos un día seco en San Sebastián.

n	Enero	Febr.	Mar.	Abr.	Mayo	Jun.	Jul.	Agos.	Sept.	Oct.	Nov.	Dic.
1	48	48	51	44	43	49	53	49	49	56	46	45
2	62	64	66	60	60	68	72	68	66	71	61	59
3	73	75	76	72	72	79	83	80	77	80	72	69
4	80	83	83	80	81	87	90	88	85	87	79	77
5	86	88	88	86	87	92	94	92	90	91	85	82
6	90	92	92	90	90	91	95	96	95	93	94	89
7	93	94	94	93	93	97	97,6	97	95	96	92	90
8	95	96	96	95	95	97,8	98,6	98	97	97	94	93
9	96	97	97	96	97	98,6	99,1	98,8	98	98,3	96	94
10	97	98	98	97	97,8	99,1	/	99,2	98,6	98,9	97	96
11	98	98,6	98,7	98	98,4	/	/	/	99,1	99,3	97,8	96,8
12	98,5	99	99,1	98,7	98,9	/	/	/	/	/	98,4	97,6
13	99	/	/	99,1	99,2	/	/	/	/	/	98,8	98,2
14	/	/	/	/	/	/	/	/	/	/	99,1	98,6
15	/	/	/	/	/	/	/	/	/	/	/	99

Esto es consecuencia de que la probabilidad de lluvia en Gijón y San Sebastián es bastante similar en todas las estaciones del año excepto en verano. En San Sebastián llueve en verano casi con la misma probabilidad que en cualquier otra estación del año, mientras que en Gijón la probabilidad de lluvia en verano es sensiblemente inferior a la media de las otras épocas del año. En otro trabajo nos extenderemos más sobre las causas.

De los resultados de G_n se puede deducir que en ambas estaciones meteorológicas la probabilidad de racha lluviosa de larga duración es insignificante.

Los test de la χ^2 que comparan las frecuencias teóricas con las observadas, de que haya al menos un día seco entre *n* días consecutivos, son óptimos en las dos estaciones meteorológicas para todos los meses del año, superándose en todos los meses el 98 % del nivel de significación.

Puesto que con el modelo de Cadenas de Markov de 1.er orden se obtiene un ajuste excelente entre frecuencias teóricas y observadas no consideramos necesario utilizar el modelo de Cadenas de Markov de 2.º orden.

Como un ejemplo gráfico del buen ajuste obtenido entre frecuencias teóricas y observadas en los dos problemas que se han estudiado, hemos elegido la representación gráfica del mes de julio en Gijón para el primer problema la del mes de octubre en San Sebastián para el segundo.

BIBLIOGRAFIA

COX, D.R. y MILLER, H.D. (1976): *The Theory of Stochastic processes.* Edit. Chapman and Hall, London, pp. 398.

ERICKSON, B. (1965): *Climatological study of persistency and probability of precipitation in Sweden.* Tellus, XVII, pp. 484-497.

GABRIEL, K.R. and NEUMANN, J. (1961): *A Markov chain model for daily rainfall occurrencd at Tel Aviv.* Quart. J. Royal Meteorol. Soc., 88, pp. 90-95.

KEMENY, J.G y SNOLL, J.L. (1961): *Finite Markov Chains.* Edit. D. Van Nostrand Company, Inc., London, pp. 210.

LAMBAS, M. (1977): *Aplicación de la cadena de Markov al estudio de la precipitación en Salamanca.* Universidad de Salamanca, pp. 58.

MATEO, P. (1965): *Persistencia de los días con precipitación y sin precipitación en Gijón (Costa Cantábrica de España).* Pub. Serie A, n.º 40, Serv. Meteor. Nac., Madrid, pp. 88.

VIEDMA, J.A. (1980): *Bioestadística* (Métodos estadísticos en Medicina y Biología). Gráfica Internacional. San Dalmacio, 3. Madrid, pp. 350.

AGRADECIMIENTO

Agradecemos a los jefes del Centro Meteorológico Zonal de San Sebastián y del Observatorio Meteorológico de Oviedo, D. José Ignacio Alvarez Usabiaga y D. Pedro Mateo González, respectivamente, por la amabilidad de proporcionarnos los datos meteorológicos necesarios para este estudio.